

The Oceanic Optics Book

Curtis D. Mobley

Primary Author and Editor

Vice President and Senior Scientist, Emeritus

Sequoia Scientific, Inc.

Bellevue, Washington, USA

Copyright © 2022 by Curtis D. Mobley

Permission is granted to copy and distribute this book under a

[Creative Commons BY-NC-ND 4.0 License](#)

DOI: [10.25607/OBP-1710](https://doi.org/10.25607/OBP-1710)

Version 1.0

January 24, 2022

This book was written and edited by Curtis D. Mobley with contributions from several contributing authors (see next page). It is published by the International Ocean Colour Coordinating Group (IOCCG) to ensure that foundational knowledge is available to the ocean colour community free of charge. The IOCCG is an international group of experts in the field of satellite ocean colour, acting as a liaison and communication channel between users, managers and agencies in the ocean colour arena.



International Ocean Colour Coordinating Group
P.O. Box 1006
Dartmouth, Nova Scotia, B2Y 4A2
Canada
<https://ioccg.org/>

Suggested citation formats:

Mobley, C. D. (Editor), 2022. *The Oceanic Optics Book*, International Ocean Colour Coordinating Group (IOCCG), Dartmouth, NS, Canada, 924pp. DOI: [10.25607/OBP-1710](https://doi.org/10.25607/OBP-1710)

or

Author Name, 2022. Chapter or Section Name, In *The Oceanic Optics Book*, C. D. Mobley (editor), International Ocean Colour Coordinating Group (IOCCG), Dartmouth, NS, Canada, 924pp. DOI: [10.25607/OBP-1710](https://doi.org/10.25607/OBP-1710)

Contributing Authors

The *Ocean Optics Web Book* is a continually growing community resource that invites anyone to write a page on relevant topics in their specialty. Many scientists have contributed to one or more pages of the Web Book, and consequently to the corresponding sections of the *The Oceanic Optics Book*. Each Web Book page and section of the PDF book shows the contributing authors. (Any web page or book section not explicitly showing authorship was written by Curtis Mobley.) The authors contributing to the Web Book and consequently to the PDF book are as follows, in alphabetical order by last name:

Ziauddin Ahmad, NASA Goddard Space Flight Center
David Antoine, Curtin University, Australia
Sean Bailey, NASA Goddard Space Flight Center
Paul Bissett, Florida Environmental Research Institute and WeoGeo
Emmanuel Boss, University of Maine
David Bowers, Bangor University, United Kingdom
Ivona Cetinić, NASA Goddard Space Flight Center
Bryan Franz, NASA Goddard Space Flight Center
Howard Gordon, University of Miami (Professor Emeritus)
Nils Haëntjens, University of Maine
Victoria Hill, Old Dominion University
Curtis Mobley, Sequoia Scientific, Inc. (Vice President Emeritus)
Bryan Monosmith, NASA Goddard Space Flight Center
Collin Roesler, Bowdoin College
Kenneth Voss, University of Miami
Jeremy Werdell, NASA Goddard Space Flight Center
Xiaodong Zhang, University of Southern Mississippi
Richard Zimmerman, Old Dominion University

Preface

Since the initial content went online in 2007, the *Ocean Optics Web Book* beginning at <https://www.oceanopticsbook.info/> has become a widely used resource for information on optical oceanography and ocean color remote sensing. A number of Web Book users have asked for a PDF version of the entire web site. Ask and ye shall receive.

History

The idea for a web book grew out of the 2004 Summer Course in Optical Oceanography and Ocean Color Remote Sensing, which was taught at the University of Maine Darling Marine Center. Funding to develop the website software and initial content was generously provided by the Ocean Biology and Biogeochemistry Program of the National Aeronautics and Space Administration. The first year of that work involved developing sophisticated content-management software (called SiteTurbine, developed by [Rainstorm Consulting](#)) that could convert LaTeX files into web pages. LaTeX was (and remains) the standard software for creating printed documents with complicated mathematics. However, at the start of the work, there was no software to conveniently convert a LaTeX file into web pages that would display the mathematics on any computer system or browser. SiteTurbine converted any symbol (such as a Greek letter) or equation into a PNG image, so that the symbol or equation displayed properly on any web browser. The initial content (version 1.0) went online in time for use during the 2007 Summer Course at the University of Maine. Many more pages were added over the next decade.

After a decade of use, SiteTurbine was showing its age. In particular, it did not satisfy modern standards for web security and was no longer supported, and RainStorm Consulting had sold its web development work of this type to [Sozo Technologies](#). The content-management software underlying the web site was therefore completely redone in 2019 and 2020 by Sozo Technologies. The present web book, which can be thought of as the Web Book version 2.0, uses the [Concrete5](#) content-management software and [MathJax](#) for conversion of LaTeX files to web pages. Both of these are open-source software and bring the web book up to modern standards.

Porting the Web Book content from its original host and software to the new platform gave me the opportunity to review all of the content. I did that during 2020 and early 2021, cleaning up archaic LaTeX code, bringing uniformity of notation to web pages written at

different times (and sometimes by different authors), resizing figures as needed for viewing on most monitors, and adding many new pages. Once that task was complete, the obvious next step was to combine the new Web Book 2.0 content into the single PDF file that you have before you.

The *Ocean Optics Web Book* is organized into topical chapters, with Level 1, 2, and 3 material within each chapter. The Level 1 web pages cover introductory and basic material for the chapter topic. The intended audience is people new to the field and those who want just a summary presentation of the subject matter. Level 2 provides an in-depth examination of the Level 1 material at the level of texts such as *Light and Water* (Mobley, 1994) and *Light and Photosynthesis in Aquatic Ecosystems* (Kirk, 1994). The Level 3 pages contain ancillary material such as links to the websites of commercial instrument makers, tabulated numerical data, and software to be used for data processing.

This Level 1, 2, 3 format works well for a web site, but not for a printed text. I have therefore in a few cases rearranged the order of the web pages when creating a PDF book chapter, and I have omitted a few of the Level 3 Web Book pages. However, for the most part, the chapters of the PDF book correspond to the chapters of the Web Book, and the web pages within a Web Book chapter appear as sections within the PDF chapters.

Although this PDF book contains essentially the same material as the Web Book, some features cannot be reproduced. In particular, some Web Book pages have animations or video clips, or links to download data files or computer programs. In the PDF book, a frame of an animation or video is used along with a note to see the online Web Book for the animation or video, or a note to go online to download files.

Overview

Hydrologic optics is the quantitative study of the interactions of light with the Earth's oceans, lakes, rivers and other water bodies. Most of the scientific interest (and funding!) in hydrologic optics is directed towards understanding light in the ocean, but the general theory to be developed in this book is applicable to any water body, and in many cases is applicable to any medium—a water body, the atmosphere, a sea of liquid methane on Saturn's moon Titan, or human tissue.

The development of the general theory requires many different concepts: radiometric variables, inherent optical properties (IOPs), apparent optical properties (AOPs), and the mathematical relations connecting them. The radiometric variables are various measures of the light itself (how much light energy is present, what direction it is traveling, what wavelengths are present). The IOPs describe the optical properties of the medium through which the light propagates. In particular, IOPs describe how light is absorbed (light energy is converted to other forms, such as heat or the energy in a chemical bond) and scattered (how it changes direction and, perhaps, wavelength) when it interacts with the medium. The equations of radiative transfer theory connect the various pieces and enable the prediction of light propagation through a medium given the properties of the medium and the light incident onto the medium.

Figure 1 shows an “organization chart” with the most commonly used IOPs, AOPs, radiometric variables, and their relationships. This chart serves as a road map for the initial chapters of this book, which work through the various boxes in the figure until all of the terms have been defined and illustrated by examples. The later chapters of the book discuss applications of the basic concepts to problems such as remote sensing of the oceans from satellites or the prediction of underwater visibility.

The 16 chapters develop the standard material of optical oceanography and remote sensing of the oceans at visible wavelengths: radiometry, inherent and apparent optical properties, absorption, elastic and inelastic scattering, the optical properties of sea water constituents and of surfaces, radiative transfer and electromagnetic theory, remote sensing and the associated atmospheric correction, and visibility. This is the material presented in the famous summer classes in optical oceanography and ocean color remote sensing, which I have helped teach since 1995. The 7 appendices contain more advanced and more mathematical topics: Fourier transforms and their use in modeling sea surfaces given a wave energy variance spectrum or autocovariance function, Monte Carlo simulation as a way to compute radiances and irradiances, and the prediction of the appearance of an object as seen through an absorbing and scattering medium like water. It can be argued that some of these appendices cover material that is outside the domain of optics and would be better placed in a text on physical oceanography. That may be true, but I had to learn this material in order to carry out optical calculations such as the prediction of reflectance and transmission of light by wind-blown sea surfaces, and the prediction of underwater image degradation due to absorption and scattering. The existing literature never—absolutely never—gave me enough detail to allow the writing of computer programs to implement the general theory and, consequently, in some cases it took me of months of frustration to get everything working exactly correctly in my computer programs. My appendices present the relevant theory and in addition give the many details needed for correct numerical calculations. I am unaware of any other publication that covers some of these details. Believe me, if you ever have to write computer programs to model sea surfaces or predict the appearance of an underwater object, you will be thanking me from the bottom of your heart for including this material here.

The contents of the chapters and appendices are briefly summarized as follows:

Chapter 1, Light and Radiometry, defines the geometric and radiometric concepts needed to describe light itself.

Chapter 2, Light Fields in Nature, illustrates radiometric variables and describes light from blackbodies, the Sun, and bioluminescence.

Chapter 3, Inherent Optical Properties, defines and illustrates the inherent optical properties, which describe the optical properties of the medium through which the light propagates.

Chapter 4, Apparent Optical Properties, defines and illustrates apparent optical properties, which are hybrids of the inherent optical properties and the radiometric variables, and which also can be used to describe the medium.

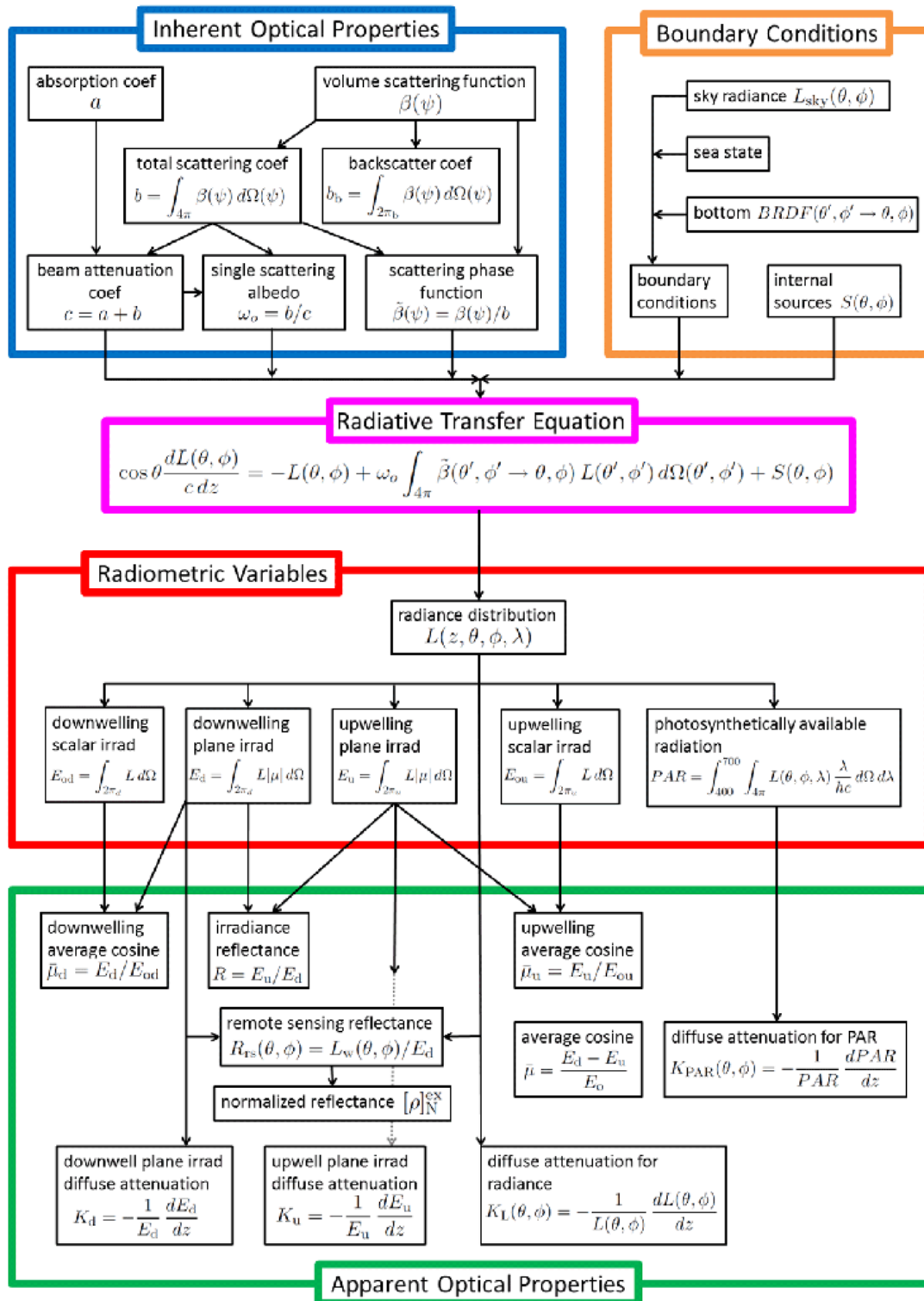


Figure 1: An ocean optics organization chart.

- Chapter 5, Absorption**, discusses the physics of absorption and how it is measured.
- Chapter 6, Elastic Scattering**, discusses the physics of scattering of light, and surveys empirical data and models for scattering by particles, surfaces, and turbulence.
- Chapter 7, Inelastic Scattering**, discusses inelastic scattering processes in the ocean, in particular Raman scattering by water and fluorescence by chlorophyll and colored dissolved organic matter.
- Chapter 8, Optical Constituents of the Ocean**, surveys the optical properties of the most important constituents of oceanic waters: the water itself, phytoplankton, dissolved organic matter, mineral particles, and air bubbles. Develops bio-optical models for absorption and scattering, and discusses particle size distributions.
- Chapter 9, Radiative Transfer Theory**, derives the equations that govern light propagation in the ocean.
- Chapter 10, Radiative Transfer Solutions**, derives several approximate analytical solutions of the scalar radiative transfer equation.
- Chapter 11, Maxwell's Equations**, discusses light as an electromagnetic phenomenon governed by Maxwell's equations. Wave propagation and dispersion are discussed in detail.
- Chapter 12, Mie Theory**, describes Mie's solution of Maxwell's equations for light incident onto a homogeneous sphere. Example output and approximations to the exact solution are presented.
- Chapter 13, Surfaces**, describes reflection and refraction by level and wind-blown sea surfaces. Defines the bidirectional reflectance distribution function (BRDF).
- Chapter 14, Ocean Color Remote Sensing**, defines the basic concepts used in remote sensing of the oceans at visible wavelengths.
- Chapter 15, Atmospheric Correction**, describes in detail the algorithms currently used by NASA for atmospheric correction of radiances measured by satellite-borne sensors.
- Chapter 16, Photometry and Visibility**, describes the basic concepts of photometry, human vision, and how visibility is predicted.
- Appendix A, Fourier Transforms**, collects the results from Fourier transform theory needed for the appendices on sea surface modeling and image prediction.
- Appendix B, Wave Variance Spectra**, collects the results from the theory of wave variance spectra for wind-blown sea surfaces as needed for the appendices on sea surface modeling.
- Appendix C, Modeling Sea Surfaces: Wave Variance Spectra Techniques**, shows how sea surfaces are modeled using wave variance spectra and Fourier transforms.
- Appendix D, Modeling Sea Surfaces: Autocovariance Techniques**, shows how sea surfaces are modeled using autocovariance functions and Fourier transforms.

Appendix E, Monte Carlo Simulation, reviews basic probability theory and its use in Monte Carlo ray tracing to solve the radiative transfer equation.

Appendix F, Monte Carlo Ray Tracing, develops the mathematical techniques needed for Monte Carlo ray tracing, including backward Monte Carlo techniques.

Appendix G, Image Prediction, develops the Fourier transform techniques needed for the prediction of images seen through absorbing and scattering media.

Usage

It is intended that this book be viewed using a PDF viewer such as Adobe Acrobat. However, the pages are formatted so that they can be printed double sided on 8.5 x 11 inch paper with extra space at the left margin for binding and with a layout typical of printed texts. When viewed as a PDF, the equations, tables, and figures are numbered and reference to them is displayed in blue. Bibliographic references are displayed in maroon. If you place your cursor over a colored equation, table, or bibliographic reference, or on an item in the table of contents or in the index, and click, the viewer will take you to the page where the selected equation, table, figure, bibliographic reference, or index entry appears. To return to the original position in the text, use “Alt” + the “left arrow” if you are using Adobe Acrobat on a Windows system. (There is some variation here so your computer may be different. Some computers require only the left arrow to return. On the MacOS operating system use “command” + “left arrow” in Acrobat and “command” + “[” in the Mac application Preview). Links to external websites are in green; clicking on a green link opens the website.

Since this PDF book is created from almost the same LaTeX code as the Web Book pages, it is my intention to update the PDF book occasionally as new web pages are added. However, this PDF book is a retirement project for an old man who would rather spend his last few good years sea kayaking or trekking in some distant land, so I make no guarantees about future updates. However, please do send comments, corrections, and suggestions for additional sections to curtis.mobley@gmail.com.

Curtis Mobley
The Comptorium
Sammamish, Washington

Felix, qui potuit rerem cognoscere causas
(Happy is the person who understands the causes of things)
—Publius Vergilius Maro (Virgil) in *Georgics*, book 2, verse 490

Notation and Conventions

Vectors in 3-D space are denoted by bold face, e. g. \mathbf{x} or \mathbf{E} . Unit vectors in 3-D space are given a “hat”, e.g., $\hat{\mathbf{z}}$ or $\hat{\boldsymbol{\xi}}$. Initial or unscattered directions are primed and final or scattered directions are unprimed, e.g., initial direction $\hat{\boldsymbol{\xi}}'$ is scattered into final direction $\hat{\boldsymbol{\xi}}$. The LaTeX “blackboard” font is used for 4×1 and 4×4 matrices, e.g., the 4×1 Stokes vector \mathbb{S} and the associated 4×4 rotation (\mathbb{R}) and Mueller (\mathbb{M}) matrices. Latin letters that are mathematical variables are italicized; unitalicized letters are abbreviations. Thus $\hat{\boldsymbol{\xi}}_i$ is one of possibly many vectors, $\hat{\boldsymbol{\xi}}_i, i = 1, \dots, N$, whereas $\hat{\boldsymbol{\xi}}_i$ might stand for an incident direction onto a surface. Exceptions to this formatting follow standard mathematical typesetting conventions, e.g. dy/dx for a derivative, rather than dy/dx , because dx and dy can be viewed as differentials that represent single mathematical quantities. Likewise, the x component of $\hat{\boldsymbol{\xi}}$ is written as ξ_x and not ξ_x because ξ_x represents a single variable. Physical constants like the speed of light c and Planck’s constant h are by convention italicized. I use the symbol \triangleq to denote “is defined as.”

Tables 1 to 3 list the symbols used throughout the book. The Reference column links to the location of the definition, first usage, or where the term is discussed in detail. For multiple uses of the same symbol, context should make the intended usage clear, e.g., c for the speed of light versus c for the beam attenuation coefficient. Symbols used only within one section are not shown. For additional symbols and historical usage, see Tables 1.4, 3.1, 4.1, 11.1, 15.1, A.1, and B.1. Table 4 lists the abbreviations and acronyms used throughout the book.

Symbol	Definition	Reference
a	absorption coefficient	§3.0.1
b	subscript for backward (e.g., b_b)	§3.1.2
b	scattering coefficient	§3.0.1
b_b	backscattering coefficient	§3.1.2
\tilde{b}_b	backscattering fraction (same as B)	§3.1.2
c	– beam attenuation coefficient	§3.0.1
	– speed of light <i>in vacuo</i>	Tab. 1.1
d	subscript for downward (e.g., E_d , K_d)	§1.4.1.1
f	temporal frequency	§1.5.9
g	asymmetry parameter of a phase function	§3.1.3
h	Planck constant	Tab. 1.1
k	angular wavenumber or spatial frequency	§1.5.9
$\hat{\mathbf{n}}$	unit normal to a surface	§13.1.1
n	– real index of refraction	§8.2.1.1
	– number size distribution	§8.12
p	subscript for particles (e.g., a_p , b_p)	§8.1
r	geometric path length in direction of light propagation	§E.2
τ	the value of a normally distributed random variable	§F.3
u	subscript for upward (e.g., E_u , L_u)	§1.4.1.1
u	the value of a uniformly distributed random variable	§E.1
v	– subscript for viewing direction (e.g., θ_v , ϕ_v)	§1.4.1.1
	– subscript for visual (e.g., L_v , E_{dv})	§16.2
w	subscript for water (e.g., a_w , L_w)	§3.4
w	weight of a ray in Monte Carlo simulations	§E.4
z	– geometric depth	§1.4.1.1
	– sea surface elevation	§C
\hat{z}	Fourier transform of z	§C
$\hat{\mathbf{z}}$	unit vector in the $+z$ direction	§1.4.1.1

Table 1: Lower case Latin letters used throughout the book. The Reference column gives the location of first usage or of detailed discussion.

Symbol	Definition	Reference
A	absorptance	§3.0.1
B	backscattering fraction (same as \tilde{b}_b)	§3.1.2
Chl	chlorophyll concentration	§8.3
\mathcal{D}	discrete Fourier transform	§A.3
E	– irradiance	§1.5.2
	– electric field magnitude	§1.6.2
\mathcal{E}	irradiance (in sections using E for the electric field)	§11.3
\mathbf{E}	electric field vector	§11.1, 1.6
\mathcal{F}	continuous Fourier transform	§A.2
I	– intensity	§1.5.6
	– first element of Stokes vector (same as L)	§1.6.2
K	diffuse attenuation function (e.g., K_d , K_{Lu})	§4.1
L	radiance	§1.5.1
\mathbb{M}	Mueller matrix	§1.7.2
\mathcal{N}	a normal probability distribution function	§E
Q	– efficiency factor (e.g., Q_a , Q_b)	§12.2
	– energy	§1.5.8
	– second element of Stokes vector	§1.6.2
R	subscript for Rayleigh (e.g., L_R)	§12.5.1
R	irradiance reflectance	§4.2.1
\mathbb{R}	(4×4) rotation (transformation) matrix	§1.7.2
R_{rs}	remote-sensing reflectance	§4.2.2
\mathbb{S}	Stokes vector	§1.6.2
T	– temporal wave period	§1.5.9
	– temperature	§2.3
U	– wind speed	§13.4
	– third element of Stokes vector	§1.6.2
\mathcal{U}	a uniform probability distribution function	§E.1
V	fourth element of Stokes vector	§1.6.2

Table 2: Upper case Latin letters used throughout the book.

Symbol	Definition	Reference
α	azimuthal scattering angle in a local coordinate system	§1.4.1.2, E.3, Fig. 1.21
β	volume scattering function	§3.1
$\tilde{\beta}$	scattering phase function	§3.1.3
γ	– instrument field-of-view half angle – emission polar angle in Monte Carlo simulations	§1.5.1 §F.6
ζ	optical depth in water	§1.4.1.1
θ	polar angle in a global (ocean) coordinate system	§1.4.1.1
$\tilde{\kappa}$	complex wavenumber: $\tilde{\kappa} = \kappa' + i\kappa''$	§11.3
λ	wavelength of light <i>in vacuo</i>	§1.3
μ	– $\mu = \cos \theta$ – slope parameter in particle size distributions	§1.4.1.1 §6.7
$\bar{\mu}$	average (mean) cosine of a radiance distribution (e.g., $\bar{\mu}_d$)	§4.3
ν	spatial frequency, wavenumber	§1.5.9
ϕ	azimuthal angle in a global (ocean) coordinate system	§1.4.1.1
ψ	polar scattering angle in a local coordinate system	§1.4.1.2, Eq. (9.3)
$[\rho_w(\lambda)]_N^{\text{ex}}$	exact normalized water-leaving reflectance	§15.3
τ	– optical path length in direction of light travel – optical depth in the atmosphere	§E.2 §15.4
$\hat{\xi}$	unit direction vector	§1.4.1.1
ω	angular temporal frequency	1.5.9
ω_o	albedo of single scattering	§3.0.1
Ω	solid angle	§1.4.2

Table 3: Greek letters used throughout the book.

Abbreviation	Meaning	Reference
AOP	apparent optical property	§4, Table 4.1
BMC	backward Monte Carlo	§E.7
BRDF	bidirectional reflectance distribution function	§13.6
BSF	beam spread function	§9.5.1
CDF	cumulative distribution function	§E.1
CDOM	colored dissolved organic matter	8.4
CIE	Commission Internationale de l’Eclairage	16.3
CZCS	Coastal Zone Color Scanner	§14
det	subscript for detritus (e.g., a_{det})	§8.1, 8.5.1
DFT	discrete Fourier transform	§A.3
FOV	field of view	§1.5.1
FFT	fast Fourier transform	§A.3
FWHM	full width at half maximum	§1.5.1
IOCCG	International Ocean Color Coordinating Group	§14
IOP	inherent optical property	§3
IR	infrared	§2.4.2
Lamb	subscript for Lambertian (e.g., $BRDF_{\text{Lamb}}$)	§13.7
lidar	light detection and ranging	§9.6
min	subscript for mineral (e.g., a_{min})	§8.5
MODIS	MODERate-resolution Imaging Spectroradiometer	§14
MODTRAN	MODERate resolution atmospheric TRANsmission	§15.1
NAP	non-algal particles	§8.1, 8.5
NASA	National Aeronautics and Space Administration	§15
NIR	near-infrared	§15.5
NIST	National Institute of Standards and Technology	§1.3
NOMAD	NASA bio-Optical Marine Algorithm Dataset	§4.1
OPBG	Ocean Biology Processing Group	§15
PAR	photosynthetically available radiation	§1.5.7
PDF	probability distribution function	§E.1
phy	subscript for phytoplankton (e.g., a_{phy})	§8.3
PSD	particle size distribution	§8.12
PSF	point spread function	§9.5.2
QSSA	quasi-single scattering approximation	§10.3
RGB	red-green-blue	§16.1
RSR	remote-sensing ratio	§13.8
RTE	radiative transfer equation	§9
SeaWiFS	Sea-viewing Wide Field-of-view Sensor	§14
SD	subscript for Secchi depth (z_{SD})	§16.7
SOS	successive order of scattering	§10.2.1
SRTE	scalar radiative transfer equation	§9.4
SSA	single-scattering approximation	§10.2
TOA	top of the atmosphere	§14.3.2, 15
UV	ultraviolet	§2.4.2
VIIRS	Visible Infrared Imaging Radiometer Suite	§15.3.2
VRTE	vector radiative transfer equation	§9.2.3
VSF	volume scattering function	§3.1

Table 4: Abbreviations and acronyms used throughout the book.

Contents

Authorship	i
Preface	viii
Notation	xiv
1 Light and Radiometry	1
1.1 A Brief History of Light	1
1.2 Measuring Radiant Energy	12
1.3 Units	15
1.3.1 The Fundamental Physical Constants	15
1.3.2 The SI Base Units	15
1.3.3 The Fundamental Photon Properties	18
1.4 Geometry	19
1.4.1 Coordinate Systems and Directions	19
1.4.1.1 Global coordinate systems	19
1.4.1.2 Local coordinate systems	22
1.4.2 Solid Angle	24
1.4.3 Dirac Delta Functions	25
1.5 Geometrical Radiometry	27
1.5.1 Radiance	27
1.5.2 Plane Irradiance	30
1.5.3 Scalar Irradiance	32
1.5.4 Vector Irradiance	34
1.5.5 Example: Irradiances of an Isotropic Radiance Distribution	34
1.5.6 Intensity	34
1.5.7 Photosynthetically Available Radiation	35
1.5.8 Terminology and Notation	36
1.5.9 Spectral Quantities Expressed per Unit Frequency Interval	37
1.6 Polarization: Stokes Vectors	39
1.6.1 What is Polarization?	40
1.6.2 Stokes Vectors	42
1.7 Polarization: Scattering Geometry	46
1.7.1 Coordinate Systems	47
1.7.2 Scattering Geometry	49

1.8	The Limitations of Radiance	54
2	Light Fields in Nature	57
2.1	Irradiances	57
2.2	Radiances	60
2.2.1	Visualizing Radiances	64
2.3	Blackbody Radiation	71
2.4	Light from the Sun	76
2.4.1	The Proton-Proton Cycle	76
2.4.2	The Solar Spectrum	77
2.4.3	The Lunar Spectrum	80
2.5	A Common Misconception	82
2.5.1	Point Functions and Density Functions	82
2.6	Bioluminescence	85
2.6.1	Modeling Bioluminescence in Radiative Transfer Calculations	87
2.7	Cherenkov Radiation	89
3	Inherent Optical Properties	97
3.0.1	Conceptual Definitions of IOPs	98
3.0.2	Historical IOP Notation	100
3.1	The Volume Scattering Function (VSF)	100
3.1.1	Conceptual Definition of the Volume Scattering Function	101
3.1.2	Scattering Coefficients	102
3.1.3	The Scattering Phase Function	102
3.2	Visualizing VSFs	103
3.3	Measuring IOPs	106
3.3.1	Measuring The Beam Attenuation Coefficient	106
3.3.2	Measuring The Absorption Coefficient	109
3.3.3	Measuring The Scattering Coefficient	111
3.3.4	Measuring The Volume Scattering Function	111
3.4	Additivity of IOPs	113
3.5	Absorption and Scattering Are Equally Important	114
4	Apparent Optical Properties	117
4.1	Diffuse Attenuation Coefficients (K Functions)	120
4.1.1	Dependence of K Functions on IOPs and Environmental Conditions	120
4.1.2	Beam Attenuation versus Diffuse Attenuation	125
4.1.3	Gordon's Normalization of K_d	127
4.1.4	Models for K_d	128
4.2	Reflectances	132
4.2.1	The Irradiance Reflectance R	132
4.2.2	The Remote-Sensing Reflectance R_{rs}	134
4.2.3	Dependence of R and R_{rs} on IOPs and Environmental Conditions	135
4.3	Average Cosines of the Radiance Distribution	138
4.4	Optical Classification of Water Types	140
4.4.1	The Forel-Ule Color Classification	141

4.4.2	The Jerlov Classification	142
4.4.3	Case 1 versus Case 2 Water Types	143
4.4.4	Ternary Diagrams	145
4.4.5	Other Classification Schemes	147
5	Absorption	149
5.1	Why is Absorption Important?	149
5.2	Absorption by Oceanic Components	150
5.3	The Physics of Absorption	153
5.3.1	Quantum Mechanics Terminology	153
5.3.2	Quantum Mechanics of Absorption	157
5.4	Measurement of Absorption: From Theory to Reality	161
5.5	Benchtop Spectrophotometry	163
5.5.1	Solutions in Transmission Mode (T-mode)	163
5.5.2	Particulate Matter in Cuvettes in an Integrating Sphere (IS-mode)	165
5.5.3	Particulate Matter on Glass Fiber Filters in an Integrating Sphere (QFT in IS-mode)	165
6	Elastic Scattering	169
6.1	Why is Scattering Important?	169
6.1.1	Review of Definitions	171
6.2	The Physics of Scattering	172
6.2.1	The Index of Refraction	172
6.3	Models for Scattering	173
6.3.1	Reflection and Refraction: Snell's Law	173
6.3.2	Scattering by Homogeneous Spheres: Mie Theory	175
6.3.3	Scattering by Irregular Particles	177
6.3.4	Scattering by Pure Water: Einstein-Smoluchowski Theory	178
6.3.5	Scattering by Turbulence	180
6.4	Backscattering	182
6.4.1	Measurement of the Backscattering Coefficient	183
6.4.2	Models for the Particle Backscattering Coefficient	185
6.4.3	Models for the Particle Backscattering Fraction	189
6.4.4	Backscattering at 180 degrees	190
6.5	Petzold's VSF Measurements	192
6.6	The Henyey-Greenstein Phase Function	196
6.6.1	The Reynolds-McCormick Phase Function	198
6.7	The Fournier-Forand Phase Function	199
6.8	Effect of Particle Shape on Optical Properties	205
7	Inelastic Scattering	211
7.1	General Formalism for Inelastic Scattering	211
7.2	Raman Scattering	213
7.2.1	The Raman Scattering Coefficient	213
7.2.2	The Raman Wavelength Redistribution Function	214
7.2.3	The Raman Phase Function	216

7.2.4	Incorporation of Raman Scattering into Radiative Transfer Calculations	218
7.2.5	Interpretation of Raman emission profiles	218
7.2.6	Temperature and Salinity Dependence of Raman Scattering	220
7.3	Theory of Fluorescence and Phosphorescence	221
7.3.1	The Physics of Fluorescence and Phosphorescence	222
7.3.2	Incorporation of Fluorescence into Time-independent Radiative Transfer Theory	225
7.3.3	Comments on Terminology	226
7.4	Chlorophyll Fluorescence	227
7.4.1	The Chlorophyll Fluorescence Scattering Coefficient	227
7.4.2	The Chlorophyll Fluorescence Wavelength Redistribution Function	227
7.4.3	The Quantum Efficiency of Chlorophyll Fluorescence	228
7.4.4	The Chlorophyll Excitation Function	231
7.4.5	The Chlorophyll Emission Function	231
7.4.6	The Chlorophyll Fluorescence Phase Function	232
7.5	CDOM Fluorescence	232
7.5.1	The CDOM Fluorescence Scattering Coefficient	232
7.5.2	The CDOM Fluorescence Wavelength Redistribution Function	233
7.5.3	The CDOM Fluorescence Phase Function	236
7.6	Examples of Inelastic Scattering Effects	236
7.6.1	Effect of Raman Scattering on R_{rs}	237
7.6.2	Effect of Raman Scattering on Upwelling Plane Irradiance	237
7.6.3	Effect of Chlorophyll Fluorescence on Downwelling Plane Irradiance	238
7.6.4	Effect of Chlorophyll Fluorescence Quantum Efficiency on R_{rs}	241
7.6.5	Effect of CDOM Fluorescence on R_{rs}	241
7.6.6	Effect of Inelastic Scattering on K Functions	242
8	Optical Constituents of the Ocean	247
8.1	Classifying the Optically Important Marine Particles	247
8.2	Water	252
8.2.1	The Index of Refraction of Water	252
8.2.1.1	The real part of the index of refraction	253
8.2.1.2	The imaginary part of the index of refraction	255
8.2.2	Absorption by Water	256
8.2.3	Elastic Scattering by Water	257
8.3	Phytoplankton	259
8.3.1	Pigment Packaging	260
8.3.2	Absorption by Phytoplankton	262
8.3.3	Scattering by Phytoplankton	263
8.3.4	Fluorescence by Phytoplankton	265
8.4	Colored Dissolved Organic Matter (CDOM)	266
8.4.1	Absorption by CDOM	266
8.4.2	Elastic Scattering by CDOM	268
8.4.3	Inelastic Scattering by CDOM	269
8.5	Non-algal Particles	269

8.5.1	Absorption by Non-algal Particles	269
8.5.2	Scattering by Non-algal Particles	269
8.5.3	Absorption by Minerals	271
8.5.4	Scattering by Minerals	272
8.6	Air Bubbles	274
8.7	Aggregates	280
8.8	A Classic IOP Model for Case 1 Waters	281
8.8.1	Absorption	282
8.8.1.1	Absorption by particles	282
8.8.1.2	Absorption by colored dissolved organic matter	283
8.8.2	Scattering	283
8.9	A New IOP Model for Case 1 waters	284
8.9.1	Absorption by Particles	284
8.9.2	Absorption by CDOM	288
8.9.3	Scattering	289
8.9.4	Scattering Phase Function	290
8.10	Commonly Used Models for IOPs and Biogeochemistry	292
8.10.1	Models Linking CDOM Absorption to Biogeochemical Parameters	293
8.10.2	Non-algal Particles	293
8.10.3	Particulate Organic Carbon (POC)	294
8.10.4	Particulate Matter or Total Suspended Matter	294
8.10.5	Global Particulate Scattering	294
8.10.6	Global Particulate Absorption	295
8.11	Optical Properties of Shelf Seas and Estuaries	295
8.11.1	CDOM in Shelf Seas	296
8.11.2	Effect of Variations in Source Concentration	296
8.11.3	Using CDOM to Trace Water Masses in Shelf Waters	297
8.11.4	Remote Sensing of CDOM and Salinity	298
8.11.5	Suspended Mineral Particles in Shelf Seas	299
8.11.6	Measuring the Mass Concentration of Mineral Particles	300
8.11.7	Light Absorption by Mineral Particles	300
8.11.8	Light Scattering by Mineral Particles	301
8.11.9	Theoretical Ideas About the Optical Properties of Mineral Particles	302
8.11.10	Relationship Between Scattering and Mean Particle Size and Density	302
8.11.11	Relationship Between Absorption and Mean Particle Size and Density	303
8.11.12	Remote Sensing of Mineral Particles in Shelf Seas	304
8.12	Particle Size Distributions	305
8.12.1	Non-uniqueness of Particle Size	305
8.12.2	Cumulative and Particle Number Size Distributions	307
8.12.3	Area and Volume Size Distributions	308
8.12.4	Models for Particle Size Distributions	309
8.12.5	Single-parameter Measures of Particle Size	314
8.12.6	Comments on Terminology	315
8.12.7	Comments on Units	316
8.13	Creating Particle Size Distributions from Data	317

8.13.1	Empirical PSDs	317
8.13.2	Parametric Description of a PSD	318
8.13.3	Ecological Size Spectra	320
8.13.4	Correctly Fitting a Power Law to a PSD	320
8.13.5	Building PSDs with a Different Rule for the Bin Size	321
9	Radiative Transfer Theory	323
9.1	The Scalar Radiative Transfer Equation; Heuristic Derivation	324
9.1.1	Radiative Processes	324
9.1.2	Standard Forms of the RTE	328
9.2	The General Vector Radiative Transfer Equation	330
9.2.1	Quantum Electrodynamics	331
9.2.2	Maxwell's Equations	332
9.2.3	The General Vector Radiative Transfer Equation	333
9.3	The VRTE for Mirror-symmetric Media	335
9.4	The Scalar Radiative Transfer Equation; Rigorous Derivation	340
9.4.1	From the VRTE for Total Radiance to the SRTE	340
9.4.2	Summary	343
9.5	Beam and Point Spread Functions	344
9.5.1	The Beam Spread Function (BSF)	344
9.5.2	The Point Spread Function (PSF)	344
9.5.3	Equivalence of the BSF and PSF	347
9.6	A Lidar Equation for Oceanography	348
9.6.1	Preliminaries	349
9.6.2	Derivation of the Lidar Equation	351
9.6.3	Example Lidar Calculation	353
10	Radiative Transfer Solutions	355
10.1	Solving the Radiative Transfer Equation	355
10.1.1	Exact Analytical Solutions	356
10.1.2	The Bouguer-Lambert-Beer Law	356
10.1.3	Approximate Analytical Solutions	357
10.1.4	Numerical Solutions	357
10.2	The Single-Scattering Approximation (SSA)	359
10.2.1	The Successive-Order-of-Scattering (SOS) Solution Technique	359
10.2.1.1	Solution for the unscattered radiance	361
10.2.1.2	Solution for the singly-scattered radiance	362
10.2.2	Assembling the SSA solution	364
10.3	The Quasi-Single-Scattering Approximation (QSSA)	366
10.4	The Asymptotic Radiance Distribution	370
10.4.1	An Integral Equation for the Asymptotic Radiance Distribution	371
10.4.2	Dependence of Asymptotic Values on Inherent Optical Properties	373
10.4.3	Rate of Approach to Asymptotic Values	375
10.5	An Analytical Asymptotic Solution for Internal Sources	377
10.6	HydroLight	379
10.6.1	The HydroLight Physical Model	381
10.6.2	The HydroLight Computational Model	382

10.6.3	Ways in Which HydroLight Can Be Used	383
10.6.4	Inputs to HydroLight	384
10.6.5	Outputs from HydroLight	386
10.6.6	Documentation	386
10.7	Gershun's Law	387
10.7.1	Heating Rates in the Upper Ocean	389
10.8	Energy Conservation	389
10.8.1	Energy Conservation Within the Water	390
10.8.2	Energy Conservation Across the Air-Water Surface	390
10.8.2.1	Energy conservation in terms of plane irradiances	391
10.8.2.2	Energy conservation in terms of scalar irradiances	392
10.8.2.3	Energy conservation in terms of radiance	394
10.9	Closure	395
11	Maxwell's Equations	401
11.1	Maxwell's Equations in Vacuo	402
11.1.1	Physical Preliminaries: Electric and Magnetic Fields	402
11.1.2	Mathematical Preliminaries: Divergence and Curl	403
11.1.3	Maxwell's Equations in a Vacuum	405
11.1.4	Light as an Electromagnetic Phenomenon	407
11.2	Maxwell's Equations in Matter	408
11.2.1	Dielectrics	408
11.2.2	Maxwell's Equations Inside Matter	410
11.2.3	Comments on Terminology	413
11.3	Plane Wave Solutions of Maxwell's Equations	414
11.3.1	Plane Waves in a Dielectric	414
11.3.2	The Absorption Coefficient	419
11.3.3	Generalizations	419
11.4	Dispersion: Phase and Group Speeds	420
11.4.1	Phase and Group Speeds for a Two-component System	421
11.4.2	Dispersion Relations	422
11.4.3	The General Development	426
11.5	Anomalous Dispersion: The Lorentz Model	428
11.6	Kramers-Kronig Relations	432
12	Mie Theory	437
12.1	Statement of the Problem	438
12.1.1	Geometry	439
12.2	The Solution	441
12.2.1	Computational Issues and Derived Quantities	445
12.3	Mie Theory with Particle Size Distributions	447
12.4	Mie Theory Examples	448
12.4.1	Example 1: Bohren and Huffman Fig. 4.9(b)	450
12.4.2	Example 2: Oceanic Particles	451
12.4.3	The Extinction Paradox	454
12.5	Mie Theory Approximations	458
12.5.1	Small Particles, $x \ll 1$: Rayleigh's Approximation	459

12.5.2	Rayleigh's Approximation Obtained from Mie Theory	460
12.5.3	Applicability of Rayleigh's Approximation	461
12.5.4	Weakly Scattering or "Soft" Particles: The Rayleigh-Gans Approximation	463
12.5.5	Applicability of the Rayleigh-Gans Approximation	464
12.5.6	Large Particles, $x \gg 1$: Geometric Optics	466
12.5.7	A Geometric Optics Ray Tracing Example	468
12.5.8	Closing Thoughts	469
13	Surfaces	471
13.1	The Level Sea Surface	471
13.1.1	Geometric Relations	471
13.1.2	The n^2 Law for Radiance	474
13.2	Fresnel's Equations for Unpolarized Light	476
13.3	Fresnel's Equations for Polarized Light	477
13.4	The Cox-Munk Sea Surface Slope Statistics	486
13.5	Sea Surface Simulations	488
13.5.1	Optical Differences in Sea Surfaces	491
13.6	The Bidirectional Reflectance Distribution Function (BRDF)	494
13.7	Lambertian Surfaces	498
13.8	Other Measures of Reflectance	500
13.8.1	Albedos	501
13.8.2	The Irradiance Reflectance versus The Bi-Hemispherical Reflectance	501
13.8.3	The Remote-Sensing Reflectance	502
13.8.4	The Reflectance Factor and the Radiance Factor	502
13.8.5	The Fresnel Reflectance	502
14	Ocean Color Remote Sensing	505
14.1	Remote-sensing Terminology	507
14.1.1	Active versus Passive Remote Sensing	507
14.1.2	Data Resolution	508
14.1.3	Data Processing Levels	509
14.1.4	Validation	510
14.2	Inverse Problems	512
14.2.1	Classification of Inverse Problems	514
14.3	Counting Photons for Remote Sensing	515
14.3.1	Radiance from the Sea Surface	515
14.3.2	Photons Detected at the Top of the Atmosphere	516
14.4	Thematic Mapping	520
14.4.1	Supervised Classification	520
14.4.2	Covariance and Correlation Matrices	521
14.4.3	Spectrum Matching versus Statistical Classification	526
15	Atmospheric Correction	529
15.1	The Atmospheric Correction Problem	530
15.2	Problem Formulation	538
15.3	Normalized Reflectances	541

15.3.1	Normalized Radiances and Reflectances	541
15.3.2	The BRDF Effect	545
15.3.3	Summary	550
15.4	Atmospheric Transmittances	552
15.4.1	Direct Transmittance	552
15.4.2	Diffuse Transmittance	552
15.5	Vicarious Calibration	555
15.6	Algorithm Introduction	560
15.7	Non-absorbing Gases	562
15.7.1	Wind Speed and Surface Reflectance Effects	563
15.7.2	Pressure Effects	563
15.8	Absorbing Gases	565
15.8.1	Absorption by Ozone	567
15.8.2	Absorption by Nitrogen Dioxide	567
15.9	Sun Glint	569
15.10	Whitecaps	571
15.11	Aerosols	572
15.11.1	Aerosol Properties	572
15.11.2	Black-pixel Calculations	575
15.11.3	Non-black-pixel Calculations	578
15.11.4	Strongly Absorbing Aerosols	580
15.12	Out-of-band Response	581
15.13	Polarization	587
15.14	Empirical Line Fits	590
15.15	Radiative Transfer Techniques	594
16	Photometry and Visibility	597
16.1	Human Color Vision	598
16.2	Luminosity Functions	602
16.3	Chromaticity	606
16.3.1	Color Matching	606
16.3.2	CIE Chromaticity Coordinates	607
16.3.3	Metameric Spectra	610
16.4	From XYZ to RGB	612
16.5	The Luminance Transfer Equation	616
16.5.1	Dependence of c_v on the Ambient Radiance	617
16.6	Contrast	620
16.6.1	The Luminance Difference Law	620
16.6.2	The Law of Contrast Reduction	622
16.7	The Secchi Disk	623
16.7.1	The Classical Secchi Depth Model of Preisendorfer	624
16.7.2	The Secchi Depth Model of Lee et al.	626
16.7.3	Black-and-White Secchi Disks	627

A	Fourier Transforms	629
A.1	Notation for Fourier Transforms	629
A.2	Continuous Fourier Transforms	631
A.3	Discrete Fourier Transforms	632
A.3.1	Frequency Ordering	635
A.3.2	Interpretation of Negative Frequencies	636
A.4	Parseval's Relations	637
A.5	Discrete Convolution Theorems	637
A.5.1	The 1-D Convolution Theorem	637
A.5.2	The 2-D Convolution Theorem	639
A.6	Orthogonality of Sines and Cosines	640
B	Wave Variance Spectra	643
B.1	Sinusoidal Wave Representations	643
B.1.1	Sampling	646
B.2	Theory of Wave Variance Spectra	648
B.2.1	Wave Energy	648
B.2.2	Wave Elevation Variance Spectra	648
B.2.3	Wave Slope Variance Spectra	653
B.3	Examples of Wave Variance Spectra	655
B.3.1	The Pierson-Moskowitz Omnidirectional Gravity Wave Spectrum	656
B.3.2	The ECKV Directional Gravity-Capillary Wave Spectrum	659
B.3.3	Spreading Functions	662
C	Modeling Sea Surfaces: Wave Variance Spectra Techniques	665
C.1	Surfaces to Spectra: 1-D	665
C.2	Spectra to Surfaces: 1-D	670
C.2.1	Theory for 1-D Surfaces	671
C.2.2	Example: A Roundtrip Calculation	674
C.3	Surfaces to Spectra: 2-D	677
C.3.1	Example: A Random Sea Surface	678
C.3.2	Example: A Sea Surface of Crossing Sinusoids	680
C.4	Spectra to Surfaces: 2-D	682
C.4.1	Theory	682
C.4.2	Example: A Two-Dimensional Sea Surface	687
C.5	Spreading Function Effects	687
C.6	Numerical Resolution	691
C.7	Time-Dependent Surfaces	698
D	Modeling Sea Surfaces: Autocovariance Techniques	703
D.1	Autocovariance	704
D.2	The Wiener-Khinchin Theorem	705
D.2.1	Example	706
D.3	Numerical Example of the Wiener-Khinchin Theorem	709
D.4	Sampling Strategy and Computational Details	715
D.4.1	Idle Speculations	717
D.4.2	Lessons Learned	718

D.5	Turbulence-Generated Water Surfaces	721
D.5.1	Two-dimensional Water Surfaces	723
E	Monte Carlo Simulation	727
E.1	Probability Functions	728
E.2	Determining Ray Path Lengths	730
E.3	Determining Scattering Angles	731
E.4	Ray Tracing Options	735
E.4.1	Numerical Comparison of Tracking Types	737
E.5	Error Estimation	745
E.5.1	Probability Theory	746
E.5.2	Numerical Examples	748
E.5.3	Error Estimation	751
E.6	Importance Sampling	753
E.6.1	Theory of Importance Sampling	753
E.6.2	Example for an Embedded Point Source	754
E.6.3	Example for Backscattering	760
E.7	Backward Monte Carlo Ray Tracing	764
E.7.1	The Essence of Backward Monte Carlo Simulation	764
E.7.2	Example BMC Simulations	765
F	Monte Carlo Ray Tracing	769
F.1	General Ray-Plane Intersections	770
F.1.1	Ray-Bottom Intersections	771
F.1.2	Ray-Sea Surface Intersections	772
F.2	Reflection by Lambertian Surfaces	773
F.3	Reflection and Transmission by the Sea Surface	774
F.4	In-water Ray Scattering	776
F.5	Use of the Preceding Equations	780
F.6	Ray Initialization in Backward Monte Carlo Simulations	780
F.7	Processing Ray Weights in Backward Monte Carlo Calculations	782
F.8	The BRDF as a PDF	788
F.8.1	Computing the Reflected Ray Weight and Direction from a BRDF	788
F.8.2	A Simple BRDF Example	789
G	Image Prediction	791
G.1	The Importance of the PSF in Image Prediction	792
G.2	Image Boundary Conditions	794
G.3	Convolution	795
G.3.1	Serial Convolution	795
G.3.2	Cyclic Convolution	796
G.4	Image Prediction Using the Convolution Theorem and FFTs	799
G.5	Computational Issues	801
G.5.1	Converting $PSF(R, \psi)$ to $PSF(r, s, R)$	801
G.5.2	Image Shifting	803
G.5.3	Summary of the FFT Algorithm	806
G.6	Understanding Image Predictions	807

G.6.1	Example Point Spread Functions	807
G.6.2	Boundary Effects	808
G.6.3	Point Spread Function Effects	811
G.6.4	Timing	814
G.6.5	Image Formats	814
G.6.5.1	JPEG files	814
G.6.5.2	PNG files	815
G.7	Example Image Simulations	817
G.7.1	Secchi Disk Gray-Scale Image Simulations	817
G.7.2	Color Image Simulation	822
G.8	The Modulation Transfer Function (MTF)	824
G.8.1	Measuring the MFT	824
G.8.2	Relation Between the MTF and the PSF	827
G.9	Numerical Examples of MTFs	828
Bibliography		833
Index		874

List of Figures

1	An oceanic optics organization chart	vi
1.1	Pinhole camera	2
1.2	Young's double-slit experiment	3
1.3	Fresnel's spot	4
1.4	Compton scattering	6
1.5	Interference by single photons in a double slit experiment	9
1.6	A global, ocean coordinate system	21
1.7	A local coordinate system for scattering of in-water rays	23
1.8	Geometry for plane and solid angles	24
1.9	Geometry for an element of solid angle in spherical coordinates	25
1.10	Gershun tube radiometer	28
1.11	Modern design of a well collimated radiometer	29
1.12	Instrument design for measuring plane irradiance	31
1.13	Design of a plane irradiance sensor head	31
1.14	Schematic instrument design for measuring hemispherical scalar irradiance	33
1.15	The hierarchy of radiometric concepts	38
1.16	Electric field for linear polarization	40
1.17	Right-hand-twist drill bit analogy to right-circular polarization	41
1.18	Right-circular polarization	42
1.19	Coordinate systems for specification of the state of linear polarization	44
1.20	Coordinate system for Stokes vectors	48
1.21	The scattering plane and the incident and final meridian planes	50
1.22	Computation of rotation angles	51
2.1	Photographic film measurement of light	58
2.2	Spectral irradiances measured in Crater Lake, Oregon	59
2.3	Spectral irradiances measured in San Vicente Reservoir, California	60
2.4	Spectral radiances measured by Johnson and Liljequist (1938)	61
2.5	The Scripps radiometer used by J. E. Tyler	62
2.6	Radiance distributions for a clear day	63
2.7	Radiance distributions for clear versus overcast days	64
2.8	Radiance in air as a function of polar viewing direction and wavelength	66
2.9	Radiance at depth 0 as a function of polar viewing direction and wavelength	66
2.10	Radiance at 10 m as a function of polar viewing direction and wavelength	67
2.11	Radiance at 50 m as a function of polar viewing direction and wavelength	68

2.12	Radiance in air as a function of polar and azimuthal viewing directions	68
2.13	Radiances in selected directions as a function of depth	69
2.14	Downwelling radiance as a function of wavelength at selected depths	70
2.15	Blackbody radiation spectra	75
2.16	The extraterrestrial solar spectrum, 0 to 3000 nm	78
2.17	The extraterrestrial solar spectrum, 300 to 800 nm	79
2.18	Total sea-level plane irradiances	80
2.19	Sea-level irradiances for the Sun, Moon, and nighttime sky	82
2.20	Energy and photon irradiances as functions of wavelength, temporal frequency, and wavenumber.	84
2.21	Shapes of bioluminescence spectra	86
2.22	Bioluminescence energy spectrum for a bacterium	88
2.23	Cherenkov radiation in a nuclear reactor core	89
2.24	Water index of refraction, 0 to 700 nm	92
2.25	Cherenkov radiation spectra	92
2.26	Measured bioluminescence and Cherenkov radiation photon counts	95
3.1	Geometry used to define inherent optical properties	98
3.2	Geometry used to define the volume scattering function	101
3.3	Example volume scattering functions	104
3.4	Example scattering phase functions	104
3.5	A typical oceanic phase function viewed in various ways	105
3.6	The phase function for pure water	107
3.7	Geometry for measurement of beam attenuation	108
3.8	Schematic instrument design for measurement of the beam attenuation coefficient	108
3.9	Schematic instrument design for measurement of the absorption coefficient .	109
3.10	Geometry for measurement of the volume scattering function	112
3.11	Illustration of absorption and scattering effects on visibility	115
4.1	Irradiance profiles for selected solar zenith angles and sky conditions	118
4.2	K functions for Crater Lake and San Vicente Reservoir	121
4.3	Computed K functions for highly scattering water; SZA = 40, wind = 0 . .	122
4.4	Computed K functions for highly scattering water; SZA = 40, wind = 15 .	124
4.5	Computed K functions for highly scattering water; SZA = 0, wind = 0 . . .	124
4.6	Computed K functions for highly absorbing water; SZA = 40, wind = 0 . .	125
4.7	Illustration of the difference in beam and diffuse attenuation	126
4.8	Spectral response of the Wratten #75 filter	129
4.9	CZCS model for $K_d(490)$	130
4.10	Comparison of the Morel K_d model with measurements	131
4.11	Comparison of the Lee K_d model with measurements	132
4.12	Illustration of light rays contributing to the irradiance reflectance	132
4.13	Irradiance reflectances for Crater Lake and San Vicente Reservoir	133
4.14	Illustration of light rays contributing to the remote-sensing reflectance . . .	134
4.15	Illustration of light rays contributing to above-surface L_u	135
4.16	Illustration of the depth dependence of R on IOPs and sky conditions . . .	135
4.17	Dependence of R on chlorophyll concentration, sky condition, and wind speed	136

4.18	Dependence of R_{rs} on chlorophyll concentration, sky condition, and wind speed	137
4.19	Dependence of R_{rs} on chlorophyll concentration and viewing direction . . .	137
4.20	Illustration of average cosines of the radiance distribution	139
4.21	Illustration of the depth and IOP dependence of the average cosines of the radiance distribution	140
4.22	Using a Forel-Ule color scale to determine water color	141
4.23	The 21 Forel-Ule colors on a CIE chromaticity diagram	142
4.24	Jerlov classification based on transmittance and K_d	143
4.25	Layout of a ternary diagram	146
4.26	Example of a ternary diagram	146
5.1	Absorption spectrum for pure water, 400-700 nm	151
5.2	Typical absorption spectrum for phytoplankton, 400-700 nm	151
5.3	Typical absorption spectrum for NAP, 400-700 nm	152
5.4	Typical absorption spectrum for CDOM, 400-700 nm	152
5.5	Typical component and total absorption spectra for clear open-ocean waters and for eutrophic coastal waters	153
5.6	Electron orbital shapes	155
5.7	Conceptual layout of a Jablonski energy level diagram	156
5.8	Electron energy levels in a molecule	157
5.9	Illustration of electron energy levels and absorption of blue and red light. .	158
5.10	Illustration of molecular vibrational modes	159
5.11	Illustration of electronic and vibrational energy levels in a molecule	160
5.12	Illustration of electronic, vibrational, and rotational energy levels in a molecule	161
5.13	Illustration of attenuation by a thin layers of non-attenuating, absorbing, and absorbing and scattering material	162
5.14	Schematic diagram of single- and dual-beam spectrophotometers	163
5.15	Schematic diagram of a center-mounted integrating sphere	166
5.16	Diagrammatic representation of the geometric pathlength of a filtered sample	167
5.17	Relationship between samples measured in suspension versus on a filter pad	167
6.1	The world with and without scattering	170
6.2	The real index of refraction of pure water at room temperature	173
6.3	Illustration of Snell's law and the law of reflection	174
6.4	Example of turbulence effects on underwater imagery	182
6.5	Backscattering sensor design	184
6.6	Conversion factors for backscattering instruments	185
6.7	Particle backscattering coefficient as a function of chlorophyll	187
6.8	Particle backscattering coefficients as a function of wavelength and chlorophyll	188
6.9	Particle backscattering as a function of particulate organic carbon	189
6.10	Particle backscattering fraction as a function of the chlorophyll concentration	190
6.11	Particle backscattering fraction as a function of wavelength	191
6.12	Enhanced backscattering near 180 deg	192
6.13	Log-log plots of Petzold's measured volume scattering functions	193
6.14	Log-linear plots of Petzold's measured volume scattering functions	194
6.15	Log-log plots of the phase functions for the Petzold VSFs	194

6.16	Log-linear plots of the phase functions for the Petzold VSFs	195
6.17	62 phase functions measured in coastal waters	195
6.18	Log-log comparison of the Petzold average-particle and Henyey-Greenstein phase functions	197
6.19	Log-linear comparison of the Petzold average-particle and Henyey-Greenstein phase functions	198
6.20	Log-log plot of Fournier-Forand phase functions	200
6.21	Log-linear plot of Fournier-Forand phase functions	200
6.22	Log-log plot of 4 measured phase functions and the corresponding Fournier-Forand phase functions	201
6.23	Log-linear plot of 4 measured phase functions and the corresponding Fournier-Forand phase functions	202
6.24	The Fournier-Forand backscattering fraction as a function of index of refraction and Junge slope	203
6.25	Log-log plots of Fournier-Forand phase functions having the same backscattering fraction but different size distributions	204
6.26	Log-linear plots of Fournier-Forand phase functions having the same backscattering fraction but different size distributions	204
6.27	Measured and Mie-computed phase functions and degree of linear polarization for mineral particles in air	206
6.28	Phase functions for prolate and oblate spheroids versus area-equivalent spheres	206
6.29	Extinction cross sections for prolate and oblate spheroids compared to volume-equivalent spheres	207
6.30	Photograph of a dinoflagellate with multiple size scales	208
7.1	The Raman wavenumber redistribution function	215
7.2	The Raman emission function for excitation at 488 nm, and the excitation function for emission at 583 nm	216
7.3	The Raman emission functions for four excitation wavelengths	217
7.4	Center of the Raman emission band for a given excitation wavelength	217
7.5	Simulations of downwelling Raman-scattered radiance as a function of depth and wavelength for 488 nm excitation	219
7.6	Simulations of downwelling Raman-scattered radiance for wavelength-independent IOPs	220
7.7	Dependence of the Raman wavelength redistribution function on temperature and salinity	221
7.8	Illustration of singlet and triplet states	222
7.9	Jablonski diagram illustrating fluorescence and phosphorescence	223
7.10	Measured chlorophyll fluorescence emission spectra	228
7.11	Measured depth profiles of chlorophyll fluorescence quantum efficiency	229
7.12	Chlorophyll fluorescence quantum efficiency as a function of PAR	230
7.13	A measured CDOM spectral fluorescence quantum efficiency function	233
7.14	Best-fit parameter values for a CDOM model and the quantum efficiency	235
7.15	The modeled spectral fluorescence quantum efficiency function	235
7.16	Two fluorescence excitation-emission functions from different source materials	236
7.17	Effect of Raman scattering on remote-sensing reflectance R_{rs}	237
7.18	Simulations of upwelling plane irradiance with and without Raman scattering	238

7.19	Simulated fluorescence effects for Case 1 water with a high chlorophyll concentration and a high quantum efficiency	240
7.20	Effect of chlorophyll quantum efficiency on the chlorophyll emission band	241
7.21	Effects of CDOM fluorescence on the remote-sensing reflectance R_{rs}	243
7.22	K at 10 m depth for Case 1 water with inelastic scattering	244
7.23	K at 10 m depth for Case 1 water without inelastic scattering	244
7.24	Absorption measured and determined by Gershun's law	245
8.1	Size ranges for different components of natural waters	250
8.2	Approximate size distribution of biological particles in the ocean	251
8.3	Measured particle size distributions in coastal waters	251
8.4	The real index of refraction of pure water	253
8.5	The imaginary index of refraction of pure water	256
8.6	Dependence of the water absorption coefficient on temperature and salinity	257
8.7	Dependence of the water scattering coefficient on temperature and salinity	258
8.8	Comparison of models for scattering by pure and saline water	258
8.9	Qualitative idea of pigment packaging	260
8.10	Absorbance spectra for in vivo and sonicated cells	261
8.11	Particulate absorption spectra for Case 1 water	262
8.12	Mass-normalized absorption of phytoplankton pigments	263
8.13	Chlorophyll-normalized phytoplankton absorption	263
8.14	Normalized VSFs for different phytoplankton species	264
8.15	CDOM absorption spectra	267
8.16	Measured distribution of S_{CDOM}	268
8.17	Measurements of detritus and NAP absorption spectra	270
8.18	Measured distribution of S_{NAP}	270
8.19	Scattering coefficient spectra $b_{NAP}(\lambda)$ from Great Lake (Taihu), China	271
8.20	Measured mass-specific absorption coefficients for four types of minerals	272
8.21	Measured mass-specific absorption coefficients for mineral dust in sea water	273
8.22	Measured mass-specific scattering coefficients for four types of minerals	273
8.23	Measured mass-specific scattering coefficients for mineral dust suspended in sea water	274
8.24	Schematic drawing of light incident on and scattered by a coated bubble	275
8.25	Mie scattering calculations for clean bubbles	276
8.26	Mie scattering calculations for clean and coated bubbles	277
8.27	Size distributions of bubbles measured under breaking waves and in quiescent seas	278
8.28	Phase functions measured for bubble populations	279
8.29	Non-dimensional chlorophyll-specific absorption coefficient	282
8.30	Particulate absorption at 440 nm as a function of chlorophyll for Case 1 water	285
8.31	Comparison of particulate absorption spectra	286
8.32	Comparison of deduced absorption spectral functions	286
8.33	Comparison of averaged absorption spectral functions	287
8.34	Particle absorption coefficients computed by the "New Case 1" model	288
8.35	Modeled absorption, scattering, and beam attenuation for low to high chlorophyll	290
8.36	Phase functions for small and large particles	291

8.37	Comparison of R_{TS} as computed by the Classic and New Case 1 IOP models	292
8.38	Color of shelf sea waters	295
8.39	Typical variation of CDOM concentration with salinity in an estuary	297
8.40	Flasks of clear water and water with CDOM	298
8.41	Particles in a drop of shelf sea water photographed through a microscope	299
8.42	GF/F filters after filtration of water samples	300
8.43	SeaWiFS false color image of the shelf seas of northwest Europe	304
8.44	168 particle number size distributions measured in Arctic waters	310
8.45	Distribution of the best-fit exponents of a power-law PSD model	310
8.46	Two sets of PSDs from the deep chlorophyll maximum at two locations	311
8.47	Measured number size distributions for 18 classes or species of microbes	312
8.48	Typical open-ocean aerosol distributions	314
8.49	Differences in the mode, median, mean, and Sauter mean diameter for a log-normal distribution	315
9.1	Illustration of a single beam of radiance and the processes that affect it	326
9.2	Illustration of a beam of radiance in direction (θ', ϕ') generating radiance in the direction of interest (θ, ϕ) by elastic scattering.	327
9.3	Illustration of mirror-symmetric particles	336
9.4	Diatom shapes	337
9.5	Photograph of <i>Chaetoceros debilis</i>	338
9.6	Measurements of an oceanic scattering matrix	339
9.7	Geometry for definition of the Beam Spread Function	344
9.8	Geometry for definition of the Point Spread Function	345
9.9	The point spread function as the glow of light around a distance street light seen through fog	346
9.10	Example point spread functions	347
9.11	The BSF and PSF measured in very clear Bahamas water	348
9.12	Geometry of a lidar system for detection of a scattering layer	350
10.1	Collimated incident radiance onto the sea surface and the refracted radiance just below the surface	361
10.2	Single-scattering contributions	363
10.3	Comparison of the SSA and HydroLight solutions	366
10.4	The Petzold average particle phase function plotted on log-linear and linear-linear axes	367
10.5	Predictions of R_{TS} for the SSA, QSSA, and HydroLight as a function of ω_o	369
10.6	Predictions of R_{TS} for the SSA, QSSA, and HydroLight as a function of wavelength for Case 1 water	369
10.7	Dependence of κ_∞ on ω_o	373
10.8	Shape of the asymptotic radiance distribution $L_\infty(\theta)$ as a function of ω_o	374
10.9	Asymptotic values of the mean cosines and of the irradiance reflectance, as a function of ω_o .	374
10.10	Approach of K_d to K_∞ for various conditions	376
10.11	Approach of R to R_∞ for various conditions	376
10.12	Approach of the mean cosines to their asymptotic values for various conditions	377
10.13	Approach of measured and modeled radiances to L_∞	378

10.14	Radiances in selected directions for an internal source	379
10.15	Radiances in selected directions for an isotropic scattering phase function	380
10.16	The HydroLight standard quad layout	383
10.17	Geometry needed for expressing conservation of energy across the sea surface	391
10.18	Illustration of the HydroLight radiance quads as needed to express energy conservation	394
10.19	Example of closure for absorption measurements	397
10.20	Example of instrument closure for backscattering measurements	397
10.21	Example of model closure	398
10.22	Example of model-data closure	399
11.1	Photograph of James Clerk Maxwell	401
11.2	Concepts of electric polarization in dielectrics	409
11.3	Sinusoidal electric and magnetic plane waves in a vacuum	415
11.4	Wave propagation with phase speed greater than group speed	423
11.5	Wave propagation with phase speed less than group speed	424
11.6	Wave propagation with phase speed and group speed in opposite directions	425
11.7	Phase speed and group speeds for water	429
11.8	Lorentz line shapes for index of refraction and absorption	433
12.1	Photography of Gustav Mie	437
12.2	Suspensions of gold nanoparticles of various sizes showing a range of colors	438
12.3	Geometry for Mie theory	440
12.4	Two of the pages of Mie's 1908 paper	442
12.5	Geometry for interpreting azimuthal dependence of Mie scattering	445
12.6	Figure 4.9(b) from Bohren and Huffman (1983)	450
12.7	Mie calculation of a phase function for unpolarized light for a small phytoplankton	451
12.8	The first 8 of the Mie angular-dependent functions	452
12.9	Graphical representation of the first few spherical harmonics	453
12.10	Normalized phase functions for unpolarized light for simulated phytoplankton	454
12.11	Efficiencies as a function of the size parameter for highly absorbing soot particles	455
12.12	Measurements and Mie predictions of the extinction efficiency of soot particles	456
12.13	Efficiencies as a function of the size parameter for non-absorbing particles	457
12.14	Efficiencies as a function of the size parameter for phytoplankton	458
12.15	Comparison of exact Mie and approximate Rayleigh scattering efficiencies	462
12.16	Comparison of exact Mie and approximate Rayleigh phase functions	462
12.17	Comparison of exact Mie, Rayleigh-Gans, and Rayleigh efficiencies for phytoplankton	465
12.18	Comparison of exact Mie, Rayleigh-Gans, and Rayleigh efficiencies for soot	466
12.19	Comparison of exact Mie and Rayleigh-Gans phase functions	467
12.20	Illustration of rays passing through the sides and ends of a solid hexagonal ice crystal	468
12.21	Scattering phase function for hexagonal column cirrus cloud ice crystals	469

13.1	Illustration of incident, reflected, and transmitted rays for air- and water-incident light	472
13.2	Graphical representations of relations among incident, reflected, and transmitted directions	473
13.3	Illustration of Snell's Cone	473
13.4	Geometry for deriving the n^2 law for radiance	475
13.5	Fresnel reflectance for unpolarized light and selected water indices of refraction	477
13.6	Reflectance and transmittance matrices as functions of the incident angle for air-incident radiance	481
13.7	Reduced reflectance and transmittance matrices for air-incident radiance	483
13.8	Reflectance and transmittance matrices as functions of the incident angle for water-incident radiance	484
13.9	Reduced reflectance and transmittance matrices for air-incident radiance	485
13.10A	glitter pattern photograph looking toward the Sun	486
13.11A	glitter pattern photograph with an illustration of a tilted wave facet	487
13.12	Example of a Cox-Munk sea surface for a wind speed of $U = 10 \text{ m s}^{-1}$	489
13.13	Example realization of a 10 m s^{-1} sea surface including wave elevations	490
13.14	Monte Carlo simulation of a glitter pattern	491
13.15	HydroLight-computed surface-reflected radiances	492
13.16	Surface-reflected radiances in the plane of the Sun	493
13.17	Water-leaving radiances in the plane of the Sun	493
13.18	Geometry for discussion of surface reflectance	495
13.19	Quantities used in the definition of the BRDF	496
13.20	Mental images corresponding to the two descriptions of Lambertian surfaces	499
13.21	Resolution of the paradoxical statements about how Lambertian surfaces reflect light	500
14.1	Wavelength bands used by various ocean color remote sensors	506
14.2	Illustration of the difference in sampling interval and band width	509
14.3	Illustration of terms used in validating measurements	511
14.4	The conceptual process involved in solving a forward radiative transfer problem	512
14.5	The conceptual process involved in solving a remote-sensing inverse radiative transfer problem	514
14.6	Normalized upwelling radiances at 90 deg to the Sun	516
14.7	Quantum efficiencies for various light detectors	518
14.8	Example R_{rs} spectra defining four thematic classes	522
14.9	Class spectra as re-sampled at 50 nm intervals	522
15.1	Contributions to the total upwelling radiance above the sea surface	531
15.2	Example at-sensor radiances for different sensor altitudes	532
15.3	Upwelling radiance at 3000 m partitioned into the contributions by water, sea surface, and atmosphere	533
15.4	Sun and viewing geometry for a mid-day HypIRI pass over Station ALOHA on 21 June	534
15.5	Example radiances contributing to the TOA radiance	535
15.6	Processes contributing to the TOA radiance	536
15.7	The fraction of L_t due to various processes	537

15.8 TOA radiances for various environmental and viewing conditions	538
15.9 Comparisons of L_u and L_w for a zenith Sun, with and without an atmosphere	543
15.10 Comparison of exact normalized water-leaving reflectances with unnormalized reflectances	550
15.11 Example reflectances contributing to the TOA reflectance	551
15.12 Sun glint and water-leaving radiance as seen from the TOA	553
15.13 Flowchart of the atmospheric correction process	561
15.14 Rayleigh optical thickness and depolarization ratio	564
15.15 Transmittance by O_2 and H_2O for a tropical atmosphere	565
15.16 Transmittance by O_3 as a function of concentration	566
15.17 Transmittance by NO_2 as a function of concentration	566
15.18 Band-averaged Rayleigh optical depth and cross sections	568
15.19 Whitecap reflectance	572
15.20 Illustration of aerosol volume, number, and particle size distributions.	574
15.21 Relative humidity effects on an aerosol volume and particle size distributions.	574
15.22 Dependence of $\epsilon(\lambda, 865)$ on aerosol model	577
15.23 Remote-sensing reflectances in the NIR	579
15.24 Ocean regions for non-black-pixel corrections	581
15.25 Relative sensor response functions for the MODIS-Aqua bands, linear ordinate	582
15.26 Relative sensor response functions for the MODIS-Aqua bands, logarithmic ordinate	583
15.27 MODIS 412 nm sensor response versus an idealized 10 nm FWHM response.	584
15.28 R_{rs} spectra as functions of the chlorophyll concentration	584
15.29 $R_{rs}(555 \text{ nm})$ as a function of chlorophyll	585
15.30 Out-of-band correction factor for the SeaWiFS 555 nm band	586
15.31 Geometry for polarization	588
15.32 RGB image created from a WorldView-2 image	592
15.33 Normalized WV2 spectral response functions	592
15.34 A measured hyperspectral $R_{rs}(\lambda)$ and the 8 values corresponding the the WV2 multispectral bands	593
15.35 R_{rs} values for 10 ground stations, and the corresponding TOA radiances in digital counts	593
15.36 Example of conversion of a TOA spectrum in digital counts (DC) to a sea-level R_{rs} in sr^{-1}	594
16.1 Alexander the Great in his diving bell	598
16.2 Color patches for illustration of afterimage colors	599
16.3 Sensitivities of the three different types of cone cells	600
16.4 Visual appearances of spectra for a normal person and a deuteranop	601
16.5 Change in the appearance of “red” as the intensity of the signal decreases .	602
16.6 Luminosity functions for bright and dim light	604
16.7 The experimental layout of a color-matching experiment	607
16.8 The CIE 1931 two-degree tristimulus functions	608
16.9 The CIE 1931 two-degree chromaticity diagram	609
16.10 Color names for different regions of the CIE chromaticity diagram	610
16.11 Metameric spectra	611
16.12 The CIE chromaticity diagram with the sRGB gamut	614

16.13	Two color patches with the same chromaticities but with different luminances	614
16.14	Environmental effects on photopic beam attenuation for chlorophyll = 0.5	618
16.15	Environmental effects on photopic beam attenuation for chlorophyll = 10	619
16.16	The chlorophyll profile used in the simulations of the previous figures	619
16.17	Environmental effects on photopic beam attenuation for depth-dependent chlorophyll	620
16.18	Geometry for developing the law of contrast reduction	621
16.19	Photograph of Angelo Secchi	624
16.20	A Secchi disk being lowered into greenish water	624
16.21	A black-and-white Secchi disk as used in limnology	627
B.1	Illustration of sampling the harmonics of a given wave and of aliasing	647
B.2	The Pierson-Moskowitz elevation variance spectrum as a function of wind speed, k and ω	658
B.3	The Pierson-Moskowitz slope variance spectrum as a function of wind speed and k	658
B.4	The omnidirectional part of the ECKV elevation and slope spectra for fully developed seas	661
B.5	The omnidirectional part of the ECKV elevation and slope spectra as a function of wave age for a wind speed of $U_{10} = 10 \text{ m s}^{-1}$	661
B.6	Spreading functions for directional wave spectra	663
C.1	A 1-D random sea surface and its Fourier transform and variance spectrum	667
C.2	A 1-D surface composed of cosine waves	669
C.3	A 1-D surface composed of sine waves	670
C.4	Example of a 1-D random sea surface generated from the Pierson-Moskowitz spectrum for $U_{10} = 5 \text{ m s}^{-1}$	675
C.5	A two-dimensional random sea surface and its Fourier transform and two-sided variance spectrum	679
C.6	A two-dimensional sea surface composed of two crossing sinusoids, and the resulting Fourier amplitudes and variance	681
C.7	Pseudocode for looping over non-zero frequencies	686
C.8	Pseudocode for looping over all k_y values for the special case of $k_x = 0$ at frequency array index $u = 0$	686
C.9	Pseudocode for looping over all k_x values for the special cases of $k_y = 0$ at frequency array index $v = 0$	686
C.10	Example two-dimensional sea surface generated from a 2-D, one-sided variance spectrum	688
C.11	The cosine-2S spreading functions for $S = 2$ and 20	689
C.12	A sea surface generated with the ECKV omnidirectional spectrum and a cosine-2S spreading function with $S = 2$	690
C.13	A sea surface generated with the ECKV omnidirectional spectrum and a cosine-2S spreading function with $S = 20$	691
C.14	Example sampling of elevation and slope spectra for $N = 1024$	694
C.15	Example sampling of elevation and slope spectra for $N = 65536$	695
C.16	Example 1-D surfaces and slopes created using true and adjusted variance spectra	697

C.17	A time-dependent sequence of 2-D sea surfaces for a 10 m s^{-1} wind speed	701
D.1	The Horoshenkov autocovariance and elevation variance spectral density	708
D.2	Illustration of the Wiener-Khinchin theorem for a single realization of a random sea surface	710
D.3	Fig. 7 from de Boer (1969)	711
D.4	Illustration of the Wiener-Khinchin theorem for 100 realizations of a random sea surface	713
D.5	Illustration of the Wiener-Khinchin theorem for 1000 realizations of a random sea surface	714
D.6	Illustration of sampling strategy	716
D.7	Illustration of inadequate sampling	720
D.8	Example of a 1-D turbulence-generated water surface	722
D.9	Example of a 2-D turbulence-generated water surface	724
D.10	A slice through the surface of Fig. D.9	724
D.11	A wind-generated surface for a 5 m s^{-1} wind speed	725
E.1	Scattering directions for a collimated beam	733
E.2	A Fournier-Forand cumulative distribution function	734
E.3	Illustration of Type 1 ray tracing	736
E.4	Illustration of Types 2 and 3 ray tracing	736
E.5	Example distributions of the optical distances between interactions	738
E.6	Distribution of rays reaching the target plane as a function of the number of scatterings	739
E.7	Distribution of rays reaching the target plane at $z_T = 5$ for Type 1 ray tracing	740
E.8	Distribution of rays reaching the target plane at $z_T = 5$ for Type 2 ray tracing	740
E.9	Distribution of rays reaching the target plane at $z_T = 5$ for Type 3 ray tracing	741
E.10	Distribution of rays reaching the target plane at $z_T = 15$ for Type 1 ray tracing	742
E.11	Distribution of rays reaching the target plane at $z_T = 15$ for Type 2 ray tracing	742
E.12	Distribution of rays reaching the target plane at $z_T = 15$ for Type 3 ray tracing	743
E.13	Distributions of number of rays and power at the target plane at $z_T = 5$ and $N_{\text{emit}} = 10^4$ rays, for the three types of tracing	743
E.14	Distributions of number of rays and power at the target plane at $z_T = 15$ and $N_{\text{emit}} = 10^4$ rays, for the three types of tracing	744
E.15	Distributions of number of rays and power at the target plane at $z_T = 5$ and $N_{\text{emit}} = 10^6$ rays, for the three types of tracing	745
E.16	Source and detector geometry used for numerical simulations	746
E.17	Estimates of the fraction of detected power for four sets of runs with $N = 10^4$ emitted rays in each run	749
E.18	Estimates of the fraction of detected power for four sets of runs with $N = 10^5$ emitted rays in each run	750
E.19	The Gaussian or normal distribution	752

E.20	Geometry of irradiance at the sea surface due to an imbedded isotropic point light source in the water	754
E.21	Gordon's probability distribution functions for polar angle of ray emission .	756
E.22	Comparison of ray counts, fraction of emitted power detected, and normalized irradiance for a point source at 5 optical depths	757
E.23	Variance in the estimated fraction of emitted power detected by the innermost ring	758
E.24	Numbers of rays detected in various weight bins for three runs	759
E.25	Histograms of the 100 estimates of the fraction of power detected by the innermost ring	759
E.26	Illustration of a source emitting a collimated beam at the center of a circular detector array	761
E.27	Illustration of an emitted ray being backscattered by a biased phase function	761
E.28	Example phase functions used for normal scattering and biased backscattering	762
E.29	Comparison of the number of detected rays and the percent of detected power for biased and unbiased first scatterings	763
E.30	Essence of backward Monte Carlo Simulation	764
E.31	Coral wall reef 3-D geometry	767
E.32	Effect of a 3-D coral structure on various measures of PAR	768
F.1	Coordinate systems for scattering of in-water rays	776
F.2	Forward and adjoint problems for backward Monte Carlo simulation	784
G.1	Use of the PSF in image formation	793
G.2	Illustration of a shark in the ocean	795
G.3	The object plane as modeled by a truncated discrete serial convolution . . .	797
G.4	The object plane as modeled by cyclic convolution and Fourier transforms .	797
G.5	Examples of discrete serial and cyclic convolutions	798
G.6	Coordinate systems for labeling pixels and defining point spread functions .	802
G.7	Effect of convolution type on image shifting	804
G.8	Illustration of image shifting caused by 1-D circular convolution	805
G.9	Example point spread functions	808
G.10	A uniformly gray image at 0 and 5 optical distances	809
G.11	Cross sections through the degraded gray image	810
G.12	Oceanic vs Gaussian PSFs	811
G.13	USAF resolution chart degraded by different PSFs	812
G.14	Cross sections through the resolution chart	813
G.15	The same image processed two different ways	817
G.16	Geometry for simulation of a Secchi disk when viewed horizontally	818
G.17	Secchi disk viewed horizontally with Sun in front	820
G.18	Secchi disk viewed horizontally with Sun behind	821
G.19	Brightness cross sections through the Secchi disk image	821
G.20	PSFs used for the RGB bands of the sponge color image	822
G.21	Sponge viewed at 0 and 5 m	823
G.22	Cross sections of RGB values through the sponge image	824
G.23	Resolution chart line pattern at range 0	825
G.24	Resolution chart line pattern at range R	826

G.25 Example MFT showing FFT frequency order	829
G.26 PSF and MTF for three optical distances	829
G.27 Image degradation for different optical distances	830

List of Tables

1	Lower case Latin letters used throughout the book	x
2	Upper case Latin letters used throughout the book	xi
3	Greek letters used throughout the book	xii
4	Abbreviations and acronyms used in the book	xiii
1.1	The seven fundamental physical constants	16
1.2	SI base units	17
1.3	Derived units useful in radiative transfer theory	17
1.4	Terms, units, and symbols for radiometric quantities	37
1.5	Stokes vector patterns for various polarization states	46
2.1	Blackbody radiation formulas	74
2.2	Distribution of the total solar irradiance in various wavelength bands	79
2.3	Typical irradiances at sea level in the visible wavelength band	81
2.4	Quantities needed for Cherenkov radiation calculations of ^{40}K decay	90
3.1	Terms, units, and symbols for inherent optical properties	100
4.1	Commonly used apparent optical properties	119
4.2	Correspondence between the chlorophyll concentration and the Jerlov open-ocean water type	143
6.1	The size distribution parameters for Fournier-Forand phase functions	201
7.1	Parameter values for the Raman wavenumber redistribution function	215
7.2	Best-fit parameter values for the model of CDOM fluorescence	234
8.1	Best-fit coefficients for index of refraction calculations	254
8.2	Water index of refraction for the extreme ranges of ocean conditions	255
8.3	Absorption, scattering, and backscattering cross sections for detritus and minerals	274
9.1	Variables occurring in the lidar equation.	349
10.1	Irradiances computed within the water body	390
10.2	Plane and scalar irradiances for a collimated incident beam and no scattering within the water	393

10.3	Plane and scalar irradiances for a diffuse incident sky radiance and no scattering within the water.	393
10.4	Plane and scalar irradiances for a diffuse incident radiance, with scattering in the water	393
11.1	Quantities involved in Maxwell's equations	404
13.1	Comparison of reflectance notation	498
14.1	MODIS sensor characteristics	517
15.1	Radiance notation	539
15.2	Transmittance notation	541
15.3	Ancillary data needed for atmospheric correction	562
15.4	NIR bands used for aerosol correction	576
16.1	Typical luminances	605
16.2	Typical illuminances	606
A.1	Notation for Fourier transforms.	630
B.1	Summary of wave variance spectral quantities.	656
C.1	Comparison of Cox-Munk mean square slopes and values for a DFT-generated 2-D surface	690
C.2	Spatial frequencies and wavelengths corresponding to peak variance for the Pierson-Moskowitz spectrum $\mathcal{S}_{PM}(k)$	692
E.1	Formulas for randomly choosing scattering angles for commonly used phase functions	734
E.2	Comparison of detected power for biased and unbiased first scatterings.	763
F.1	Formulas for randomly choosing sensor emission angles in BMC simulations	783
F.2	Emission patterns and scale factors for various radiometric quantities	787
G.1	Run times for serial convolution versus FFT	814
G.2	Secchi disk radiometric quantities	819
G.3	IOPs used for generation of wavelength-dependent PSFs	822

CHAPTER 1

Light and Radiometry

This chapter begins the study of the radiometric variables section of the organization chart of Fig. 1. The chapter opens with a brief history of how the nature of light has been understood (or, more often, misunderstood) from ancient times until the present. The following sections then develop the concepts of radiometry, the science of the quantitative measurement and description of light. This requires discussion of how the direction of light propagation is specified. Then the concepts of radiance and irradiance are developed, first for unpolarized light, and then for polarized light. The next chapter will show examples of the radiances and irradiances defined in this chapter.

1.1 A Brief History of Light

Writing this section is a no-win effort. No matter what I say about the “nature of light” or the question “What is a photon?”, there will be those who tell me (quite correctly!) that I am wrong. However, the question “What is light?” is perfectly legitimate and deserves discussion, even if, as will be seen, there is no answer in everyday, human, classical physics terms.

This section is structured as a chronological “history of light” organized around the debate about whether light is a particle or a wave. Some of the milestones in our understanding of light warrant just a sentence or two; others will be discussed in some detail.

Ancient Egypt: Light is “ocular fire” from the eye of the Sun God Ra.

Ancient Greece: Democritus (c. 500 BCE) speculated that everything, light and the soul included, is made of particles, which he called atoms. Other Greeks thought that light was rays that emanate from the eyes and return with information. I haven’t seen any explanation of how they explained why everyone’s eyes quit emanating rays when the Sun went down, or when a person entered a dark cave. Somewhat later, Aristotle thought that “light is the activity of what is transparent.” In his view, light is a “form,” not a “substance.” I don’t know if he thought that things became nontransparent when it got dark at night.

Ancient China: Mo Zi (c. 468 - c 391 BCE) discovers basic rules of optics. At about the same time as Democritus, the Chinese scholar Mo Zi worked out many of the basic concepts of optics including the rectilinear propagation of light; how plane, concave, and convex mirrors reflect light; and how a pinhole camera creates an image, as seen in Fig. 1.1. Unfortunately, much of his writing in the Book of Mo Zi (compiled by his disciples) is fragmentary and he is almost unknown in the West. [Wu et al. \(2015\)](#) give a brief history of optics in ancient China.

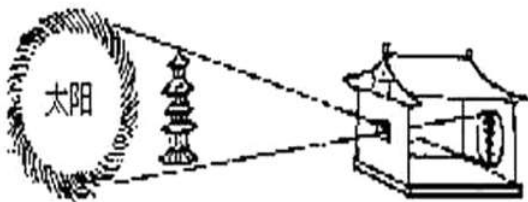


Figure 1.1: Pinhole camera as understood by Mo Zi. The characters mean Sun. Figure reproduced from [Wu et al. \(2015\)](#).

c. 1000 CE: Ibn al-Haytham: Light is rays that travel in straight lines. Little known today in the West, Ibn al-Haytham (965-1039; Latinized as Alhazen) was something of an Arab Isaac Newton. He wrote a seven-volume *Book of Optics* (as well as many other works on astronomy, mathematics, medicine, philosophy, and theology). He clearly understood the “scientific method” and he based his conclusions on observation and clever experiments, rather than on abstract reasoning. He disproved the Greek idea that light emanates from eyes, and he showed that light travels in straight lines.

Late 1600s: Newton: Light is particles. Isaac Newton conducted a series of experiments in the late 1600’s which, among other things, showed that white light is a mixture of all colors. This directly contradicted Aristotle, who claimed that “pure” light (like the light from the Sun) is fundamentally colorless. Because he was able to separate white light into colors with a prism, and because light did not seem to travel around corners (as do sound waves), Newton concluded that light must be made of particles, which he called “corpuscles.” He published his results in his treatise *Opticks* in 1704. Newton’s particle explanation of refraction required light to travel faster in water than in air, and his explanation of “Newton’s rings” (easily explained by wave interference) was rather incoherent (pun intended). In spite of a few errors like these, Newton was a pretty good scientist and *Opticks* is one of the seminal books of science.

1676: Ole Rømer measured speed of light by timing eclipses of Jupiter’s moon Io.

1678: Christiaan Huygens published the first credible wave theory of light. Part of his theory says that at each moment each point of an advancing wave front serves as a point source of secondary spherical waves emanating from that point. The position of the wave front at a later time is then the tangent surface of the secondary waves from each of the point sources. This known as Huygen’s principle.

1803: Young: Light is waves. In 1803 Thomas Young conducted a classic experiment (published in 1807) in which coherent light was incident onto two narrow parallel

slits in an opaque screen as illustrated in Fig. 1.2. According to Huygen’s principle, each slit is the source of secondary waves, which then interfere with each other as they propagate further. The light that passed through the slits formed an interference pattern on a viewing screen, just as do water waves passing through holes in a board. This is easily explained by assuming that light is a wave phenomenon. Young’s double-slit experiment was taken as conclusive proof that light is a wave and that Newton was wrong. As will be seen below, this conceptually simple double-slit experiment will reveal one of the most profound mysteries of light.

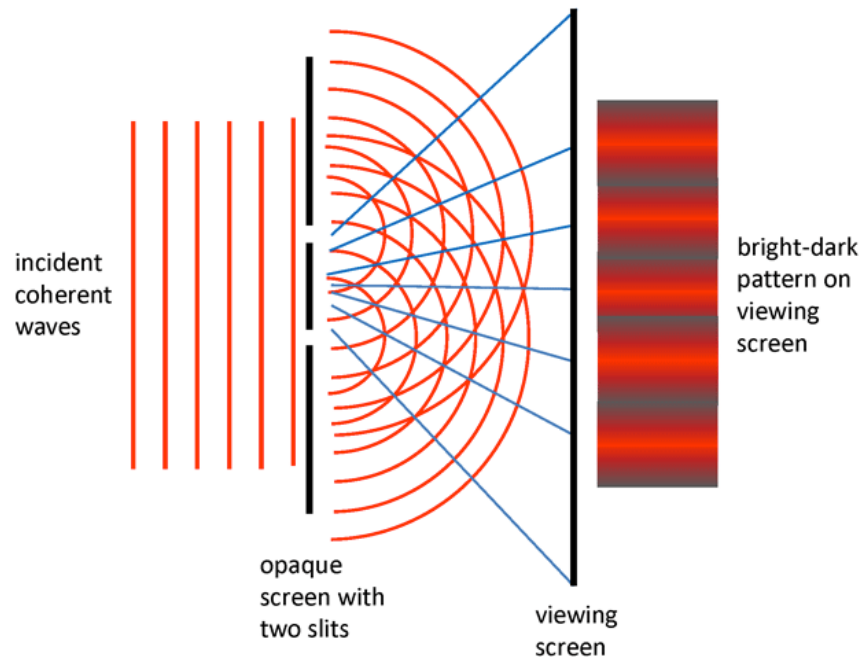


Figure 1.2: The essence of Young’s double-slit experiment. Coherent means that the monochromatic incident light has the wave crests “in step.” The blue lines show the points along which the crests of the red light waves add together to create bright bands on the viewing screen.

1819: Poisson, Fresnel, and Arago: Light is waves. In 1818, Augustin-Jean Fresnel presented a new theory of diffraction. Siméon Poisson, who favored Newton’s corpuscular theory of light, analyzed Fresnel’s equations and concluded that if they were correct, then the shadow of a sphere illuminated by a point light source would show a spot of light at the center of the shadow. Poisson considered this an absurd result, thereby disproving Fresnel’s assumption of the wave nature of light. However, when Dominique Arago did the experiment, the spot was there just as predicted, as seen in Fig. 1.3. Poisson conceded. This point of light is now known as Fresnel’s spot, Arago’s spot, or—no doubt to his chagrin—Poisson’s spot. Again, as in Young’s experiment, light must be understood as a wave.

1865: Maxwell: Light is electric and magnetic fields propagating as a wave. In 1865 James Clerk Maxwell published *A Dynamical Theory of the Electromagnetic*

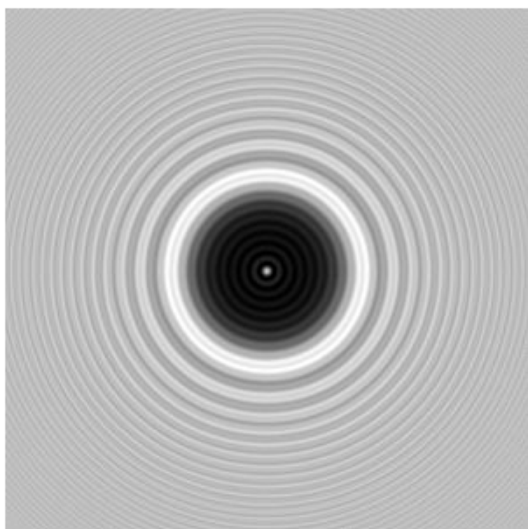


Figure 1.3: Fresnel's Spot at the center of the shadow of a 2 mm diameter sphere at a distance of 1 m from the sphere. The sphere was illuminated by a 633 nm laser. Image from [Wikipedia](#)

Field in which he tied together electric and magnetic fields via his famous equations. He then showed that each component of the electric and magnetic fields obeys a wave equation with a speed of propagation numerically equal to that of light. He concluded “This velocity is so nearly that of light that it seems we have strong reason to conclude that light itself (including radiant heat and other radiations) is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws.” (This has to be one of the greatest sentences ever written.) Maxwell’s equations are discussed in Chapter 11.

late 1880s: Hertz: Discovery of radio waves. Between 1886 and 1889 Heinrich Hertz conducted a series of experiments designed to test Maxwell’s predictions of propagating electromagnetic waves. In these experiments Hertz discovered what are now called radio waves, and he also discovered the photoelectric effect.

Young’s double-slit experiment, Arago’s confirmation of the Fresnel diffraction predictions, and Hertz’s confirmation of Maxwell’s predictions of propagating electromagnetic waves were sufficient to convince everyone that light is a wave. Newton was clearly wrong, and the matter seemed settled once and for all.

late 1800s: Blackbody radiation. One of the final problems of late nineteenth century physics was to explain the spectral distribution of energy emitted by a “blackbody.” Attempts to do this using Maxwell’s concept of electromagnetic radiation led to the prediction that a blackbody would emit an infinite amount of energy as the frequency increased. This unphysical result was called “the ultraviolet catastrophe.”

1901: Planck: The idea that light is quantized. Max Planck was able to derive the formula for the *blackbody radiation spectrum* 2.3, but only if he assumed that light comes in discrete packages, or “quanta.” He had to assume that the energy E of each light quantum is proportional to its frequency f or wavelength λ according to

$$E = hf = \frac{hc}{\lambda}.$$

The proportionality constant h , which occurs both in this equation and in Planck's formula for the energy distribution of blackbody radiation, was a free parameter that was adjusted so that Planck's blackbody spectrum would fit the measurements. The parameter h is now called Planck's constant and is one of the fundamental physical constants. Planck himself could not say why the radiation in the blackbody cavity should come in discrete pieces and, at the time, he thought that this assumption was perhaps just "a mathematical artifice." Planck received the 1918 Nobel Prize in Physics for this work; the Nobel citation credits him with the "discovery of energy quanta." His discovery was the beginning of modern physics.

1905: Einstein: showed that light really is quantized and is absorbed in discrete amounts. The photoelectric effect discovered by Hertz can not be explained if light comes in continuous waves as proposed by Maxwell. Albert Einstein was able to explain the photoelectric effect by assuming that light does indeed come in discrete packages as hypothesized by Planck and that these quanta are absorbed (or emitted) "all at once," rather than being "soaked up" bit by bit as a continuous light wave arrives at the surface of the photoelectric material. In other words, Einstein claimed that energy quanta were real physical quantities, and not just a mathematical trick. As he worded it in his 1905 paper, "Energy, during the propagation of a ray of light, is not continuously distributed over steadily increasing spaces, but it consists of a finite number of energy quanta localized at points in space, moving without dividing and capable of being absorbed or generated only as entities."

Einstein's claim that light comes in discrete packets was not well received at the time because the wave theory of light was so well established and had been so successful in most applications—blackbody radiation and the photoelectric effect being the exceptions. In the same year, Einstein also published his famous paper presenting the special theory of relativity; an explanation of Brownian motion, which helped established the reality of atoms at a time when many scientists still did not accept their existence; and a paper presenting the equivalence of mass and energy via his most famous equation, $E = mc^2$. His claims of light as particles, the mixing of time and space, matter as particles, and the equivalence of matter and energy would have relegated Einstein to the realm of crackpots had his radical ideas not been so successful in explaining so many physical phenomena. When he received the Nobel Prize in 1921 special relativity was still so controversial that the award was given to him for "his discovery of the law of the photoelectric effect."

1923: Compton scattering: light is a particle. In 1923 Arthur Compton did an experiment in which he bombarded electrons with x-rays. Figure 1.4 shows the basic idea. Compton found that the incident x-ray wavelength λ' and the wavelength λ of the scattered x-ray were related to the mass of the electron m and the angle θ of the scattered x-ray from the initial direction by the formula seen in the figure. This formula is derived by assuming that the x-ray is a particle of zero rest mass and energy given by Planck's formula $E = hf$, and then applying the equations for relativistic conservation of energy and momentum.

Compton scattering cannot be explained by a wave theory of light. This result therefore was taken to be direct experimental evidence that light is a particle. Compton received the 1929 Nobel Prize in Physics for this work.

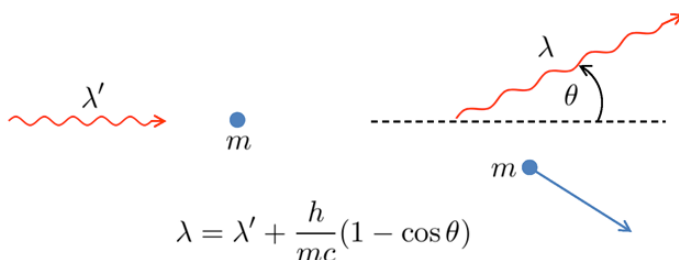


Figure 1.4: Compton scattering. The left panel shows an x-ray quantum approaching a stationary electron (the blue dot). The right panel shows the scattered x-ray and electron.

1926: The name “photon” was coined by the chemist G. N. Lewis to describe a hypothetical particle that transmitted energy from one atom to another. The word caught on as a name for Einstein’s quantum of energy, although that is not what Lewis intended.

The 1910s to 1930s: The development of quantum mechanics. These decades were a time of great excitement (and confusion!) in physics. On the one hand, there were convincing experiments showing that light is a wave: Young’s double-slit experiment, Fresnel’s spot, and Hertz’ discovery of electromagnetic waves just as predicted by Maxwell’s wave theory of light. On the other hand, there were equally convincing experiments showing that light must be a particle of some kind: the photoelectric effect, Compton scattering, and the need to assume light quantization in order to explain blackbody radiation.

The physics vocabulary now began to include phrases such as “the wave-particle duality of light” (and, indeed, of all matter). The idea is that light has both wave and particle properties and that you detect one or the other depending on the type of measurement being made. That is, if you set up an experiment that is designed to detect wave properties (e.g., a double-slit apparatus), then you will detect light as a wave. But if you set up an experiment that absorbs or emits light (e.g., the photoelectric effect), then you will detect it as discrete quanta or particles. You will also see statements such as

- Light behaves as a wave at macroscopic scales (e.g., in a laboratory interference experiment), but it behaves as a particle at atomic scales (e.g., in Compton scattering).
- Light behaves as a wave at low energies (no radio engineer ever talks about radio photons, just radio waves), but it behaves as a particle at high energies (those working with gamma rays always talk about gamma-ray photons, never gamma-ray waves).
- Light propagates as a wave (according to Maxwell’s equations), but it interacts with matter as a particle (e.g., in the photoelectric effect or in Compton scattering).

There is an element of truth to all of these statements, but they also all oversimplify the true nature of light by forcing it to fit into classical categories of wave or particle.

During these decades the great physicists Bohr, Schrödinger, Heisenberg, Pauli, Dirac and many others developed an entirely new kind of physics—quantum mechanics—to describe the internal workings of atoms. This quantum mechanics is a theory of how *matter* behaves at the atomic scale. Energy levels in atoms and molecules are quantized, and atoms and molecules therefore absorb and emit energy only at specific frequencies determined by differences in the quantized energy levels of each kind of atom or molecule. Electromagnetic radiation was thus absorbed or emitted only at the discrete frequencies determined by the quantized energy levels of matter, but the radiation itself did not need to be treated as inherently quantized. A good layman’s history of this era is *Thirty Years that Shook Physics* (Gamow, 1985).

1946: An unexpected result in the Hydrogen spectrum. Willis Lamb and Robert Retherford measured an extremely small difference in the energies of the $2S_{1/2}$ and $2P_{1/2}$ states of the Hydrogen atom (see the the Physics of Absorption, Section 5.3 for a discussion of energy levels and this notation). This energy difference corresponds to a wavelength of about 30 cm, which is in the microwave region of the electromagnetic spectrum. Now known as “the Lamb shift,” this difference in energy levels could not be explained either by the non-relativistic quantum mechanics of Schrödinger and Heisenberg, or by the relativistic quantum mechanics developed by Dirac. This experiment was one of the driving forces behind the development of quantum electrodynamics. Lamb received the Nobel Prize in 1955 “for his discoveries concerning the fine structure of the hydrogen spectrum.”

late 1940s: The development of Quantum Electrodynamics (QED). In QED, light is particles, but very strange particles they are. In part to explain the Lamb shift, Richard Feynman, Julian Schwinger, Shinichiro Tomonaga, and several others developed what is now known as Quantum Electrodynamics or QED. Feynman, Schwinger, and Tomonaga shared the 1965 Nobel Prize in Physics for their development of QED. In this theory, the electromagnetic field itself is quantized. Moreover, in QED the electric field of an electron, for example, is caused by the electron spontaneously emitting and reabsorbing enormous numbers of energy quanta, which are called *virtual photons*. These photons are called virtual photons because they are associated with undetectable energy states of the electron-photon system.

The Heisenberg uncertainty principle can be written as $\Delta t \Delta E \geq h/(4\pi)$. To detect an energy change of size ΔE , you must observe the system for a time Δt that is greater than $h/(4\pi \Delta E)$. The emission of a virtual photon of energy ΔE by an electron violates conservation of energy, but this is allowed by the Heisenberg uncertainty principle so long as the virtual photon is reabsorbed within a time of $\Delta t < h/(4\pi \Delta E)$. That is, you can violate conservation of energy so long as you don’t do it long enough to get caught. Or in another view, you are not really violating conservation of energy if the violation isn’t observable. A photon can travel a distance $c\Delta t$ before it is reabsorbed by the electron. Thus low energy virtual photons with a small ΔE can live longer and “reach out” farther from the electron before reabsorption than can higher energy virtual photons, which exist for a shorter time. This gives rise to the $1/\text{distance}^2$ strength of the classical electrical field as seen in Coulomb’s law. Similarly, a photon can turn into an electron-positron pair so long as the electron and positron reunite before the time limit imposed by Heisenberg’s relation. Note that the

emission of virtual photons is a different process than the emission of photons when an atom changes energy levels and emits a photon with energy equal to the difference of the atomic energy levels. In that case, there is no energy violation, Heisenberg's relation does not come into play, and the emitted photon can live forever (or until it is absorbed by a different atom somewhere else).

In QED, a charged particle is surrounded by a cloud of virtual photons, which are constantly flickering into and out of existence. When two charged particles approach each other, some of these virtual photons are exchanged between the particles, which is what gives the repulsive or attractive force of between the particles. The electric force is said to be “mediated” (i.e., transmitted) by the exchange of virtual photons. (In modern physics, all “action at a distance” forces (except perhaps gravity) are mediated by some type of particle. For electric forces the particles are photons.) Indeed, even “empty space” is seething with virtual particles that come into existence and then promptly disappear.

QED is able to explain the Lamb shift because the $2S_{1/2}$ and $2P_{1/2}$ states interact slightly differently with the cloud of virtual photons surrounding the hydrogen nucleus and electron. This quantitative explanation of the Lamb shift was one of the first tests of QED. There have been many more since, and some of QED's predictions agree with experiment to within about one part in 10^{12} . QED is therefore considered the most successful and well tested theory in physics, and it is the starting point for all of elementary particle physics.

late 20th century: The Fundamental Mystery: single-photon interference in a double-slit experiment. The interference patterns seen in Young's double-slit experiment can be understood in terms of classical wave theory when the incident light is a coherent wave. But what happens if only one photon at a time is incident onto the double-slit screen? Amazingly, you still get the same interference pattern! It takes longer to build up the pattern one photon at a time, but after many photons have been detected (e.g., on a CCD array), the pattern becomes obvious. There is an excellent video of an 1981 verification of this performed at Hamamatsu Photonics K.K., a company that makes optical sensors and instruments. This video is on the [Hamamatsu web site](https://www.hamamatsu.com/) and also on Youtube at <https://www.youtube.com/watch?v=I9Ab8BLW3kA>. This video is well worth ten minutes of your time. It also shows how the experiment was actually conducted. (I don't know who first did an experiment like the one on the Hamamatsu web site. However, in 1909 G. I. Taylor conducted an experiment in which he showed that a very faint light source, equivalent to “a candle burning at a distance slightly exceeding a mile,” gave interference fringes.)

Figure 1.5 shows four frames from the Hamamatsu video. In Fig. 1.5(a), photons have been collected one at a time for 3 minutes. Only about two dozen photons have been detected, and the pattern of dots, showing where each photon was detected on the detector screen (as illustrated in Fig. 1.2) appears to be random. A few minutes later (panel b) more photons have been detected, but no pattern is yet obvious. However, after 25 minutes (panel c) enough photons have been detected that an interference pattern is taking shape. After 6 hours, the pattern of individual photon detection locations clearly shows exactly the same interference pattern as is obtained

for a bright source of coherent monochromatic light.

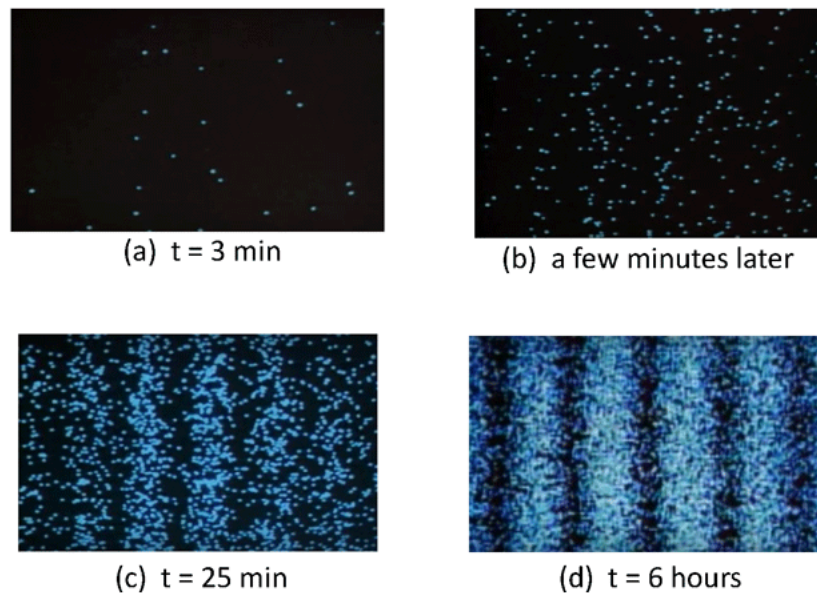


Figure 1.5: Locations of photon detections on the observing screen of a double-slit apparatus showing the build-up of an interference pattern by single photons. Frames captured from the [Hamamatsu video](#).

In classical wave theory (such as in Young’s original experiment), part of the incident wave passes through each slit, and each slit then becomes a point source for waves that propagate further (Huygen’s principle) and interfere with each other as illustrated in Fig. 1.2. ***The fact that single photons also show interference patterns after enough are collected implies that the individual photons must also simultaneously pass through both slits and then interfere with themselves!*** Indeed, if you modify the experiment so that the photon can pass through just one slit, or that you can in some way even *know* which slit it went through, then the interference pattern disappears. The fact that single photons show an interference pattern is so surprising and incomprehensible from the viewpoint of classical physics that the great physicist and teacher Richard Feynman called this “The Fundamental Mystery” (of quantum mechanics). There is no explanation for this other than to say “this is just how photons behave.”

There are two utterly profound consequences of single-photon interference:

- ***We are forced to abandon the idea that photons are localized particles*** in the classical sense of having a well-defined (even small) size. A localized particle could not pass through both slits at the same time and then interfere with itself.
- ***We are forced to abandon the idea that photons take a particular path from one point to another.*** In QED calculations (using so-called Feynman

path integrals), a *photon simultaneously takes all possible paths from one point to another*. Only after all of the calculations are done for all possible paths and the results for the different paths are combined does the final result look like the classical idea of a light ray traveling from one point to the next by a single path.

For a photon, concepts like size, position, and path are undefined and meaningless. All you can say is that a photon was created at point A (e.g., at a spot on the surface of a tungsten filament in a light bulb) and it was detected at point B (e.g., at a particular pixel of a CCD array). You can *nothing* about the path it took from A to B. (In quantum mechanics, there is no position operator for photons, as there is for material particles like electrons. Instead, photons have creation and annihilation operators, which create and destroy them.)

To make matters even more mysterious, *material particles* such as electrons and atoms also display the same interference behavior as light in a double slit apparatus. Single-electron interference was first demonstrated in 1989 (Tonomura et al., 1989). The interference patterns in their experiment look exactly like the ones in Fig. 1.5, except that the points show where the individual electrons were detected rather than where individual photons were detected. This experiment has since been repeated with molecules of more than 800 atoms and molecular weights greater than 10^4 amu (Eibenberger et al., 2013). These experiments are strong verifications of the correctness of quantum mechanics as currently formulated.

Feynman wrote a delightful and highly recommended book, *QED: The Strange Story of Light and Matter* (Feynman, 1985). This book explains, as only Feynman can, the fundamental ideas of QED without the math. He clearly considers light to be particles. For example, (on page 15 of my edition) he states “It is very important to know that light behaves like particles, especially for those of you who have gone to school, where you were probably told something about light behaving like waves. I’m telling you the way it *does* behave—like particles.” However, he also shows that these mysterious particles actually do take all possible paths from one point to another. Thus a photon goes through both slits because each slit is a possible path from the point of the photon’s creation to the point where it is detected. (If you want to see the mathematical horrors of how QED calculations are performed, the best book I’ve found is *Introduction to Elementary Particles* (Griffiths, 2008). However, that book presumes you have spent some serious years in physics and math classes.) In 1979 Feynman also delivered a series of non-technical lectures on QED at the University of Auckland, which were the origin of his QED book. These are well worth viewing and are on-line in various places, e.g., at <http://www.vega.org.uk/video/programme/45>.

Most physicists today seem quite happy talking about photons as one member of the pantheon of “elementary particles.” In their language, photons are zero-rest-mass, stable, spin-one bosons, which always travel at the speed of light and have energy $q = hc/\lambda$, momentum of magnitude $p = h/\lambda$, and angular momentum of magnitude $\ell = h/(2\pi)$.

However, some people view things differently. Willis Lamb, of Lamb shift fame, wrote a paper “Anti-photon” (Lamb, 1995) in which he states “In his [the author Lamb’s] view, there is no such thing as a photon. Only a comedy of errors and historical

accidents led to its popularity among physicists and optical scientists. There are good substitute words for ‘photon’, (e.g., ‘radiation’ or ‘light’)...” In closing he says, “It is high time to give up the use of the word ‘photon’, and of a bad concept which will shortly be a century old. Radiation does not consist of particles, and the classical, i.e., non-quantum, limit of QTR [the quantum theory of radiation, or QED] is described by Maxwell’s equations for the electromagnetic fields, which do not involve particles.”

Lamb’s paper is worth reading, and he makes some valid points. However, I’m afraid his battle to banish the word “photon” is lost. It is just too convenient. Biologists are going to continue to measure the light available for photosynthesis in units of photons per square meter per second. One einstein is going to retain its definition as “one mole of photons.” Optica (previously called The Optical Society of America) is going to continue to publish *Optics & Photonics*. (Indeed, that magazine devoted the entire issue of October 2003 to six articles on the topic of “What is a Photon?”) It certainly would have been fun to get Lamb and Feynman together in a room and watch them argue about the reality or non-reality of photons. Perhaps the greatest danger inherent in the use of the word “photon” is that it makes it easy to think of light as little balls of energy behaving like particles in the every-day sense, which simply is not correct, as we have seen above.

My own concession to Lamb and to the inability to say that a photon takes a particular path from point A to point B is that my Monte Carlo codes no longer “trace photons;” they now “trace rays.” The concept of a light ray is well accepted in the limit of geometrical optics, and all camera lenses are designed with sophisticated ray tracing codes that give perfectly good predictions of what light does, as do my Monte Carlo codes. My Monte Carlo calculations remain unchanged, I’m just more careful in describing what they do.

The present day: Enough has been said. The above discussion has reviewed the long and confused history of ideas about the nature of light. Our understanding of light has gone through “It’s a particle.”; “No, it’s a wave.”; “No, it’s simultaneously both a particle and a wave.”; and finally “It’s neither a particle nor a wave; it’s something much more mysterious.” This whole business reminds me of Nargarjuna’s Tetralemma in Buddhist philosophy. In Western philosophy we think of a statement as being either true or false. But the Buddhist philosopher Nargarjuna (c. 100 CE) posited the tetralemma that a statement can be true, or it can be false, or it can be both true and false at the same time, or it can be neither true nor false.

Planck at the close of his [1918 Nobel Prize lecture](#) raised a fundamental question:

“What becomes of the energy of a photon after complete emission? Does it spread out in all directions with further propagation in the sense of Huygens’ wave theory, so constantly taking up more space, in boundless progressive attenuation? Or does it fly out like a projectile in one direction in the sense of Newton’s emanation theory? In the first case, the quantum would no longer be in the position to concentrate energy upon a single point in space in such a way as to release an electron from its atomic bond, and in the second case, the main triumph of the Maxwell theory

— the continuity between the static and the dynamic fields and, with it, the complete understanding we have enjoyed, until now, of the fully investigated interference phenomena — would have to be sacrificed, both being very unhappy consequences for today’s theoreticians.

Be that as it may, in any case no doubt can arise that science will master the dilemma, serious as it is, and that which appears today so unsatisfactory will in fact eventually, seen from a higher vantage point, be distinguished by its special harmony and simplicity. Until this aim is achieved, the problem of the quantum of action will not cease to inspire research and fructify it, and the greater the difficulties which oppose its solution, the more significant it finally will show itself to be for the broadening and deepening of our whole knowledge in physics.”

As Planck predicted, science has learned much more about the nature of light and the role it plays in the universe, and the mystery of how photons behave continues to deepen (just Google “photon entanglement”). The word “photon” has itself evolved to mean different things to different people, as reviewed by [Kidd et al. \(1989\)](#). In any case, we still cannot say what light or a photon *is* in everyday language. We can only describe what it *does*. This situation is really no different from that of the electron. No one has any idea or model of what an electron “really is,” but that does not prevent electrical engineers from using the known properties of electrons to light our homes and run our computers.

I’ll close this review with two quotes

“All the fifty years of conscious brooding have brought me no closer to the answer to the question: What are light quanta? Of course today every rascal thinks he knows the answer, but he is deluding himself.” —Albert Einstein, quoted in [Zajonc \(2003\)](#)

“No one knows what a photon is, and it’s best not to think about it.”
—Attributed to Richard Feynman

1.2 Measuring Radiant Energy

The normal human eye is a sensitive detector of radiant energy, and it has an extremely wide *dynamic range* (the ratio of maximum to minimum detectable signal). However, the precise work of hydrologic optics requires a more objective means of measuring the flow of radiant energy. Two main classes of light detectors have been developed to detect and measure radiant energy: *thermal* and *quantum* detectors. Thermal detectors measure the *energy* of the detected photons. In thermal detectors, radiant energy is absorbed and converted into heat energy, and the detector responds to the consequent change in temperature of the absorbing medium. Thermal detectors include ordinary thermometers, thermocouples, bolometers, and pyranometers. Quantum detectors respond to the *number* of incident photons, rather than to the cumulative energy carried by the photons, although the photon energy generally affects detector performance. Quantum detectors include photographic film and various photovoltaic, photoemissive, and photoconductive detectors.

A brief description of the latter three detectors, collectively called *photoelectric* devices, is worthwhile.

A. E. Becquerel discovered the *photovoltaic effect* in 1839 in a liquid electrolyte containing two immersed electrodes connected through a galvanometer. When one of the electrodes was irradiated by sunlight, a current flowed through the circuit. This arrangement is now called a *photovoltaic cell*, and the current is measured by a current meter included in the circuit. When no light is incident on the electrode, no current flows. Generally, the greater the number of photons incident on the electrode element of the cell, the greater is the ensuing current in the circuit, which is the foundation for using a photovoltaic cell as a light meter.

The *photoemissive* effect (often called the photoelectric effect) was discovered in crude form in 1887 by Hertz in the very same experiment in which he verified the existence of electromagnetic waves. The basic photoemissive cell consists of an evacuated tube containing a negatively charged electrode (the photocathode, usually made of an alkali metal such as cesium, sodium, or potassium) and a positively charged electrode (the anode). When light is incident on the photocathode, the photons dislodge electrons from the surface of the electrode. These *photoelectrons* are drawn across a gap to the anode, thus generating a current in a series circuit containing the cell, a current meter, and a seat of electromotive force, which replenishes the supply of electrons on the photocathode and maintains the potential difference across the electrodes. In principle, no current would flow if no light were incident on the photocathode, but in practice a small *dark current* flows because of electrons spontaneously emitted by random thermal motions in the cathode.

A *photomultiplier tube* (PMT) is a specialized photoemissive cell. Rather than having only one photocathode and one anode, a PMT has a series of anodes (called dynodes), each of which is held at higher positive voltage than the previous one. The electrons liberated from the photocathode by the incident light are attracted to the first dynode. When these original electrons strike the first dynode, they knock loose additional electrons, which are then attracted to the second dynode. The electrons striking the second dynode liberate still more electrons, which are attracted to the third dynode, and so on. This electron cascade enables a PMT to greatly amplify (typically by a factor of one million) the current which would result from the photoelectrons alone. Commercially available PMT's have up to 15 dynodes and are extremely sensitive light detectors. However, PMT response is very sensitive to temperature, the response is not stable with time (owing to changes in the dynodes caused by electron bombardment), and stable high-voltage power supplies are required for operation. For these and other reasons, PMT's have been supplanted in many oceanographic instruments by solid-state detectors.

It was found experimentally in 1873 that the electrical conductivity of the metal selenium increases when light falls upon it. This effect can be exploited for measuring radiant energy by constructing a series electrical circuit consisting of the selenium (or a similarly behaving substance), a seat of electromotive force (e.g. a battery), and a current meter. The greater the number of photons falling on the *photoconductive cell* containing the selenium, the greater is the cell's conductivity, hence the greater is the current flowing in the circuit. Some dark current flows even if no light is incident on the cell, since the photoconductive substance has a nonzero conductivity even in the absence of light.

Semiconductor diodes can serve as light detectors, in which case they are called *photodiodes*. For example, in a typical pn-junction silicon photodiode, light incident on the

junction frees electrons from the silicon atoms (but does not eject the electrons from the diode). The resulting positively charged silicon ions are held fixed in position by the crystal lattice, whereas the free electrons can move in response to an applied electromotive force. These electrons thus generate a current when the photodiode is included on a series circuit with a seat of electromotive force and a current meter. The diode thus functions as a photoconductive cell. Note that a photodiode does not amplify the photocurrent as does a PMT, and consequently photodiodes are much less sensitive detectors than are PMTs. However, photodiodes have good stability, are easy to calibrate, require little power, and are quite rugged and inexpensive.

When operated as light detectors, diode junctions have the external electromotive force applied so as to separate the photoelectrons and their parent ions, thus generating the measured current. However, if the electrons are allowed to recombine with the ions, then photons are emitted from the junction. These photons have the same energy as the photons required to liberate electrons from the semiconductor atoms. When operated in this fashion, the diode is called a *light-emitting diode* (LED). LED's have the same general characteristics (stability, low cost, etc.) as photodiodes, and are often employed as light sources in oceanographic instruments (such as beam transmissometers) that require an internal light source.

Another type of photoelectric detector is the *charge-coupled device* (CCD), which is the heart of modern electro-optic cameras (digital cameras and camcorders). CCD's consist of linear or area arrays of small (of order $10\ \mu\text{m}$ diameter) spots of silicon. When light is incident on the array, electrons are released from each silicon spot in proportion to the radiant energy falling on the spot. The charge released by each spot is measured. Since the location of the silicon spots is accurately known, the pattern of released charge provides a map of the energy falling on the CCD array. When coupled with a standard camera lens, a CCD array (replacing the normal film) can record an image of the scene seen by the camera.

Theoretical understanding of the photoemissive effect came from Einstein in 1905 in a revolutionary paper (translated in [Arons and Peppard, 1969](#)) in which he introduced the concept of a quantum of light along with its energy equation $q = hf$. This work was a major milestone in the history of physics, and it was for his explanation of the photoemissive effect that Einstein received the Nobel Prize in 1921 “for his services to Theoretical Physics, and especially for his discovery of the law of the photoelectric effect¹”. Full understanding of the photovoltaic and photoconductive effects requires the quantum theory of the structure of matter. The photoconductive effect, for example, occurs when photons transfer electrons into the conduction band of the semiconductor, rather than completely ejecting the electrons from the material.

A thorough discussion of the physics and engineering of all types of radiation detectors can be found in various chapters of [Bass \[Ed.\] \(1995\)](#) and in the texts by [Budde \(1983\)](#) and [Dereniak and Crowe \(1984\)](#).

¹Einstein was passed over for the Nobel for many years because of the prejudices of the times; this sordid tale is told in [Friedman \(2001\)](#). The theory of relativity was still too controversial for the Nobel Committee to mention it by name, so the committee finally awarded him the prize for the photoelectric effect.

1.3 Units

Almost any human endeavor from subsistence farming to modern science requires an agreed-upon set of standards for measuring things as diverse as the amount of grain harvested or the wavelength of a particular color of light. Over the centuries many different measurement standards have been used. These standards were often arbitrary and imprecise. The English inch, for example, was defined in 1324 by King Edward II of England to be “three grains of barley, dry and round, placed end to end, lengthwise.” The metric system was proposed in 1790 with the goal to be “for all times, for all people” because the measurement units would be related to natural physical quantities. Thus the unit of distance, the meter, was defined to be one ten-millionth of the distance from the Earth’s north pole to the equator, measured along a great circle. Such a definition was adequate for a century or so, but is too imprecise for modern needs.

Modern science uses the International System of Units, or SI units, from the French *Système International d’Unités*. These units for the commonly needed measures of distance, time, electrical charge, etc. are now all defined in terms of fundamental physical constants such as the speed of light and the charge of the electron.

1.3.1 The Fundamental Physical Constants

In 2018, after many years of careful measurement and discussion, representatives of 60 nations unanimously agreed on values for seven fundamental physical constants, from which seven SI base units can be defined. Table 1.1 shows these seven fundamental physical constants. Note that the numerical values shown in the table are by definition exact.

Most of the quantities in Table 1.1 should be familiar from introductory physics and chemistry, but a couple warrant comment. The hyperfine transition frequency of the cesium-133 atom refers to the frequency of microwave radiation that corresponds to an electron jump between two closely spaced energy levels of a neutral cesium-133 atom (in “field-free space,” that is, in the absence of gravitational, electrical, or magnetic fields, which can change the atom’s internal energy levels). The value of $\Delta\nu_{\text{Cs}}$ gives a fundamental standard for specifying frequency, measured in Hertz, which is cycles (or periods) per second. The strangest of these constants is the luminous efficacy K_{cd} , which is a measure of how well a light source using a given power (in Watts) produces visible light (as seen by a normal human eye), measured in lumens. K_{cd} is defined to be the luminous efficacy of monochromatic radiation of frequency $540 \cdot 10^{12}$ Hertz, which is green light. Exact definitions of these quantities and further discussion can be found in the National Institute of Standards and Technology (NIST²) [Special Publication 330](#).

1.3.2 The SI Base Units

The seven fundamental physical constants seen in Table 1.1 can be used to define seven SI *base units*, which are more convenient for practical applications. Thus one second is defined via the fundamental $\Delta\nu_{\text{Cs}}$ as the duration of 9 192 631 770 periods of the radiation

²Founded in 1901 as the National Bureau of Standards, the Bureau was renamed the National Institute of Standards and Technology in 1988. Their mission centers on “...advancing measurement science, standards, and technology...”. To date, five NIST employees have received Nobel Prizes, and NIST has been associated with many others, including providing calibrated radium samples to Marie Curie.

Physical Constant	Symbol	Numerical Value	Unit
speed of light in vacuo	c	299 792 458	m s^{-1}
Planck constant	h	$6.626\,070\,15 \cdot 10^{-34}$	J Hz^{-1}
elementary electrical charge	e	$1.602\,176\,634 \cdot 10^{-19}$	C
Boltzman constant	k	$1.380\,649 \cdot 10^{-23}$	J K^{-1}
Avagadro constant	N_A	$6.022\,140\,76 \cdot 10^{23}$	mol^{-1}
hyperfine transition frequency of ^{133}Cs	$\Delta\nu_{\text{Cs}}$	9 192 631 770	Hz
luminous efficacy	K_{cd}	683	lm W^{-1}

Table 1.1: The seven fundamental physical constants used to define the SI base units. The numerical values are by definition exact.

corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom, or

$$1 \text{ second} \triangleq \frac{9\,192\,631\,770}{\Delta\nu_{\text{Cs}}}.$$

Similarly, the meter is defined³ using the speed of light and the fundamental frequency by

$$1 \text{ meter} \triangleq \frac{9\,192\,631\,770}{299\,792\,458} \frac{c}{\Delta\nu_{\text{Cs}}}.$$

The Planck constant h has units of J s or $\text{kg m}^{-2} \text{ s}$. Thus the kilogram can be defined using h and the definitions of the meter and second:

$$1 \text{ kg} \triangleq \frac{(299\,792\,458)^2}{(6.626\,070\,15 \cdot 10^{-34})(9\,192\,631\,770)} \frac{h\Delta\nu_{\text{Cs}}}{c^2}.$$

The definitions of the remaining base units are defined in similar ways as given in NIST Special Publication 330, cited above.

Table 1.2 shows the seven SI base units, plus two supplementary units that are convenient for measurement of plane and solid angle. All other quantities are derivable from these units. With the exception of the candela, which is needed only for the discussion of photometry, the reader should be familiar with these SI units from basic physics and chemistry.

The nomenclature and symbols most widely used today in optical oceanography follow the recommendations of the Committee on Radiant Energy in the Sea of the International Association of Physical Sciences of the Ocean (IAPSO; see [Morel and Smith, 1982](#)). However, neither the SI units nor the recommended IAPSO notation are entirely satisfactory. In particular, they are sometimes inconvenient for measurements and mathematical manipulations; consequently occasional minor deviations from the IAPSO recommendations will be made. Several derived units that we shall need are shown in Table 1.3.

Unless specifically noted otherwise, the wavelength of light, denoted by λ , always refers to the wavelength in vacuo. The wavelength in a medium like water will be less by a factor

³Suppose, for the sake of argument, that the speed of light changes with time as the universe ages. Since the second is fixed by the value of the fundamental constant $\Delta\nu_{\text{Cs}}$, a change in c would result in a change in the length of the meter, so that the speed of light will remain $299\,792\,458 \text{ m s}^{-1}$, now and forever.

Physical quantity	Base Unit	Symbol
length	meter	m
mass	kilogram	kg
time	second	s
electric current	ampere	A
temperature	kelvin	K
amount of substance	mole	mol
luminous intensity	candela	cd
Supplementary units		
plane angle	radian	rad
solid angle	steradian	sr

Table 1.2: SI base units.

Physical quantity	Derived Unit	Symbol	Definition
wavelength of light	nanometer	nm	10^{-9} m
energy	joule	J	$1 \text{ kg m}^2 \text{ s}^{-2}$
power	watt	W	$1 \text{ kg m}^2 \text{ s}^{-3}$
number of photons	einstein	einst	1 mole of photons

Table 1.3: Derived units useful in radiative transfer studies.

of the real index of refraction of the medium, n_{medium} : $\lambda_{\text{medium}} = \lambda_{\text{vacuum}}/n_{\text{medium}}$. Thus light of wavelength 550 nm in vacuo will have a wavelength of about $550/1.34 = 410$ nm in water, but it is still referred to for convenience as 550 nm light.

There are other non-SI units that are commonly used and acceptable. These include minutes, hours, and days, which are multiples of the second; the liter, which is one-thousandth of a cubic meter; degrees, minutes, and seconds of angles, which are fractions of a radian; and so on. Again, it is assumed that the reader is familiar with these units⁴.

⁴The United States is one of only three countries that do not use the SI system (the other two are Myanmar and Liberia). If you live in the USA, you have to learn, for example, that there are 5280 feet in a mile. Where does such a number come from? In England, a (statute) mile was originally defined as 8 furlongs. A furlong (a furrow long) was defined as the distance a team of oxen could plow without resting. A furlong was divided into 40 rods, and a rod was 16.5 feet, where a foot was defined as the average length of the left feet of 16 men chosen at random as they left church on Sunday. Seriously, you can't make this stuff up! So a mile is $8 \times 40 \times 16.5 = 5280$ feet. No wonder the rest of the world makes fun of Americans for not converting to metric units. Fortunately, American scientists have enough sense to use SI units, even if the average American does not.

1.3.3 The Fundamental Photon Properties

As was noted in Section 1.1, most physicists view photons as elementary particles whose energy q , linear momentum p , and angular momentum ℓ are given by

$$q = \frac{hc}{\lambda} \quad \left[\frac{\text{kg m}^2}{\text{s}^2} \text{ or } \frac{\text{J}}{\text{photon}} \right] \quad (1.1)$$

$$p = \frac{h}{\lambda} \quad \left[\frac{\text{kg m}}{\text{s}} \right] \quad (1.2)$$

$$\ell = \frac{h}{2\pi} \quad \left[\frac{\text{kg m}^2}{\text{s}} \right], \quad (1.3)$$

where h is Planck's constant, c is the speed of light, and λ is the wavelength.

It is instructive to consider the number of photons needed to generate typical irradiances and the environmental effects of these irradiances. Equation (1.1) shows that a single photon carries an energy of $q = hc/\lambda$. A typical clear-sky solar irradiance is about 400 W m^{-2} , i.e., $Q = 400 \text{ J}$ on each square meter each second (Table 2.2). If we use $\lambda = 550 \text{ nm}$ for the average wavelength of the solar irradiance, this irradiance corresponds to

$$N = \frac{Q}{q} = \frac{Q\lambda}{hc} = \frac{(400 \text{ J})(550 \cdot 10^{-9} \text{ m})}{(6.63 \cdot 10^{-34} \text{ Js})(3.00 \cdot 10^8 \text{ m/s})} \approx 10^{21}$$

photons per second on every square meter of surface. Even the irradiance from the faintest star corresponds to tens of millions of photons per second per square meter.

The heating rate of the water due to absorbed irradiance is (Eq. (10.35) from the Gershun's law Section 10.7)

$$\frac{\Delta T}{\Delta t} = \frac{1}{c_v \rho} \frac{\Delta E}{\Delta z} \quad \left[\frac{\text{deg C}}{\text{sec}} \right],$$

where

- T is the temperature in deg C
- t is the time in seconds
- $c_v = 3900 \text{ J (kg deg C)}^{-1}$ is the specific heat of sea water at constant volume
- $\rho = 1025 \text{ kg m}^{-3}$ is the density of sea water

If this 400 W m^{-2} irradiance is absorbed by the upper one meter of water (as can be the case in very turbid water), then the heating rate within that 1 m layer of water is

$$\frac{\Delta T}{\Delta t} = \frac{1}{(1025 \text{ kg m}^{-3})(3900 \text{ J (kg deg C)}^{-1})} \frac{400 \text{ W m}^{-2}}{1 \text{ m}} \approx 1 \cdot 10^{-4} \frac{\text{deg C}}{\text{sec}}.$$

A 10 hour day at this heating rate gives an increase of 3.6 deg C, which is very large.

The linear momentum imparted to the 1 m layer of water by absorption of these photons is, by Eq. (1.2),

$$p = N \frac{h}{\lambda} = \frac{(10^{21})(6.63 \cdot 10^{-34} \text{ J s})}{550 \cdot 10^{-9} \text{ m}} \approx 10^{-6} \frac{\text{kg m}}{\text{s}}.$$

This is comparable to the linear momentum of only 1 cubic millimeter of water moving at a speed of 1 m s^{-1} . Thus the linear momentum transported to the ocean by incident sunlight is completely negligible compared to that transported by wind and other process that drive ocean currents.

Equation (1.3) gives the maximum angular momentum that can be carried by these photons:

$$\ell = N \frac{h}{2\pi} = \frac{(10^{21})(6.63 \cdot 10^{-34} \text{ J s})}{2\pi} \approx 10^{-13} \frac{\text{kg m}^2}{\text{s}}.$$

This is comparable to the angular momentum of a small sand grain rotating at one revolution per second. This angular momentum is much less than that of even the smallest turbulent eddies generated by current shears or breaking waves. Moreover, the maximum angular momentum computed here is transferred only if all photons are in the same angular momentum state, i.e., only if the light is 100% circularly polarized. For unpolarized light, the net transfer of angular momentum is zero (Hecht, 1989).

This we see that the physical importance of the sunlight incident on the upper ocean lies in its energy transport and not in its momentum transport. This is consistent with everyday experience: sunlight heats us up, but it does not push us around.

1.4 Geometry

This section develops the mathematical tools needed to specify directions and angles in three-dimensional space. These mathematical concepts are fundamental to the specification of how much light there is and what direction it is traveling.

1.4.1 Coordinate Systems and Directions

Locations and directions are specified with reference to particular coordinate systems chosen to be convenient for the problem at hand. These coordinate systems can be “global” or “ocean” systems that are fixed once and for all in their spatial orientation, or they can be “local” systems determined by the instantaneous direction of light propagation.

1.4.1.1 Global coordinate systems

The most common ocean system is defined as follows. Let let $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ be three mutually perpendicular unit vectors that define a right-handed Cartesian coordinate system. Let the depth z be measured positive downward from 0 at the mean sea surface, as is customary in oceanography. The unit direction vector $\hat{\mathbf{z}}$ thus points downward. For problems involving sea surface waves it is convenient to resolve the waves into “along-wind” and “cross-wind” components (see Section 13.4 and Appendix C). This is most easily done if $\hat{\mathbf{x}}$ is chosen to be in the direction that the wind is blowing over the ocean surface, i.e. $\hat{\mathbf{x}}$ points downwind. The cross-wind direction is then given by the cross (or vector) product $\hat{\mathbf{y}} = \hat{\mathbf{z}} \times \hat{\mathbf{x}}$. This defines the “wind-based” coordinate system used in the HydroLight radiative transfer code (Section 10.6). If the problem at hand required the analysis of Sun glint on the sea surface as seen from an airplane, then it would be convenient to use a “Sun-based” system with $\hat{\mathbf{x}}$ pointing towards or away from the Sun and $\hat{\mathbf{z}}$ pointing upward, away from the mean sea surface (see Fig. 13.11). No matter how chosen, the $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ directions remain fixed in space for the given problem.

Given a global coordinate system, a direction in space is specified as follows. Let $\hat{\boldsymbol{\xi}}$ denote a unit vector pointing in the desired direction. The vector $\hat{\boldsymbol{\xi}}$ has components ξ_x, ξ_y and ξ_z in the $\hat{\mathbf{x}}, \hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$ directions, respectively. We can therefore write

$$\hat{\boldsymbol{\xi}} = \xi_x \hat{\mathbf{x}} + \xi_y \hat{\mathbf{y}} + \xi_z \hat{\mathbf{z}},$$

or just $\hat{\boldsymbol{\xi}} = (\xi_x, \xi_y, \xi_z)$ for notational convenience. Note that because $\hat{\boldsymbol{\xi}}$ is of unit length, its components satisfy $\xi_x^2 + \xi_y^2 + \xi_z^2 = 1$.

An alternative description of $\hat{\boldsymbol{\xi}}$ is given by the angles θ and ϕ , defined as shown in Fig. 1.6. The *polar angle* θ is measured from the direction of $\hat{\mathbf{z}}$, and the *azimuthal angle* ϕ is measured positive counterclockwise from $\hat{\mathbf{x}}$, when looking toward the origin along $\hat{\mathbf{z}}$ (i.e. when looking in the $-\hat{\mathbf{z}}$ direction). For downward directions (the red arrow in the figure), $\theta < \pi/2$; for upward directions (the green arrow in Panel (b)), $\theta > \pi/2$. The connection between $\hat{\boldsymbol{\xi}} = (\xi_x, \xi_y, \xi_z)$ and $\hat{\boldsymbol{\xi}} = (\theta, \phi)$ is obtained by inspection of Panel (a) of Fig. 1.6:

$$\begin{aligned}\xi_x &= \sin \theta \cos \phi \\ \xi_y &= \sin \theta \sin \phi \\ \xi_z &= \cos \theta,\end{aligned}\tag{1.4}$$

where θ and ϕ lie in the ranges $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$. The inverse transformation is

$$\begin{aligned}\theta &= \cos^{-1}(\xi_z) \\ \phi &= \tan^{-1}\left(\frac{\xi_y}{\xi_x}\right).\end{aligned}\tag{1.5}$$

The polar coordinate form of $\hat{\boldsymbol{\xi}}$ could be written as $\hat{\boldsymbol{\xi}} = (r, \theta, \phi)$, but since the length of r is 1, we can drop the radial coordinate for brevity.

In radiative transfer theory, the direction $\hat{\boldsymbol{\xi}} = (\theta, \phi)$ always denotes the direction the light is traveling. However, experimentalists often use the *viewing direction*, which is the direction an instrument points in order to measure the light propagating in the opposite direction. Likewise, it is often convenient to use the viewing direction to plot functions of direction. The viewing direction $\hat{\boldsymbol{\xi}}_v = (\theta_v, \phi_v)$, where the subscript v denotes “viewing”, can be written as

$$\begin{aligned}\hat{\boldsymbol{\xi}}_v &= -\hat{\boldsymbol{\xi}} \\ &= (\theta_v, \phi_v) \\ &= (\pi - \theta, \phi + \pi).\end{aligned}\tag{1.6}$$

Another useful description of $\hat{\boldsymbol{\xi}}$ is obtained using the cosine parameter

$$\mu \triangleq \cos \theta = \xi_z.\tag{1.7}$$

The components of $\hat{\boldsymbol{\xi}} = (\xi_x, \xi_y, \xi_z)$ and $\hat{\boldsymbol{\xi}} = (\mu, \phi)$ are related by

$$\begin{aligned}\xi_x &= \sqrt{1 - \mu^2} \cos \phi \\ \xi_y &= \sqrt{1 - \mu^2} \sin \phi \\ \xi_z &= \mu,\end{aligned}\tag{1.8}$$

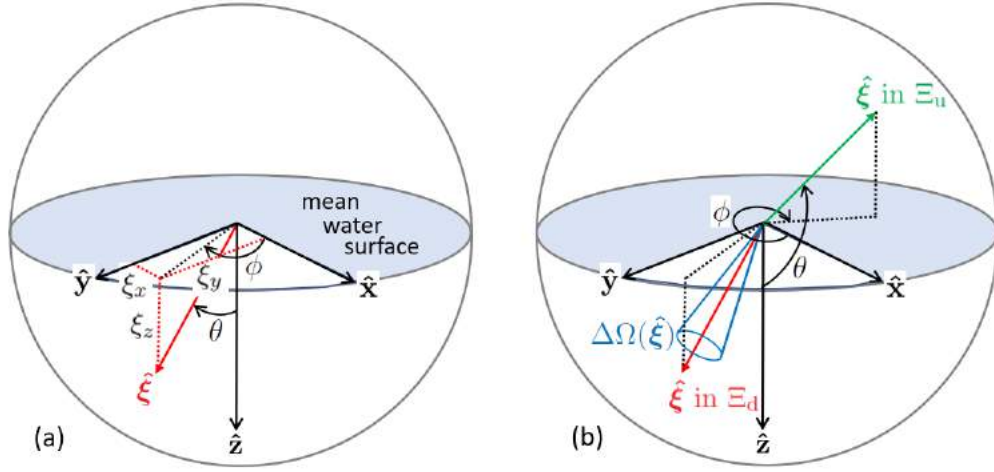


Figure 1.6: A global, ocean coordinate system. Panel (a) shows a unit direction vector $\hat{\xi}$ heading downward into the ocean (red arrow) and the corresponding Cartesian components (ξ_x, ξ_y, ξ_z) and angular coordinates (θ, ϕ) . In panel (b), the blue cone represents an element of solid angle $\Delta\Omega(\hat{\xi})$ centered on direction $\hat{\xi}$ of Panel (a). The green arrow shows an upward directed unit vector and the corresponding (θ, ϕ) .

with μ and ϕ in the ranges $-1 \leq \mu \leq 1$ and $0 \leq \phi < 2\pi$. Hence a direction $\hat{\xi}$ can be represented in three equivalent ways: as (ξ_x, ξ_y, ξ_z) in Cartesian coordinates, and as (θ, ϕ) or (μ, ϕ) in polar coordinates.

The scalar (or dot) product between two direction vectors $\hat{\xi}'$ and $\hat{\xi}$ can be written as

$$\hat{\xi}' \cdot \hat{\xi} = |\hat{\xi}'| |\hat{\xi}| \cos \psi = \cos \psi,$$

where ψ is the angle between directions $\hat{\xi}'$ and $\hat{\xi}$, and $|\hat{\xi}|$ denotes the (unit) length of vector $\hat{\xi}$. The scalar product expressed in Cartesian-component form is

$$\hat{\xi}' \cdot \hat{\xi} = \xi'_x \xi_x + \xi'_y \xi_y + \xi'_z \xi_z.$$

Equating these representations of $\hat{\xi}' \cdot \hat{\xi}$ and recalling Eqs. (1.4) and (1.8) leads to

$$\begin{aligned} \cos \psi &= \xi'_x \xi_x + \xi'_y \xi_y + \xi'_z \xi_z \\ &= \cos \theta' \cos \theta + \sin \theta' \sin \theta \cos(\phi' - \phi) \\ &= \mu' \mu + \sqrt{1 - \mu'^2} \sqrt{1 - \mu^2} \cos(\phi' - \phi). \end{aligned} \quad (1.9)$$

Equation (1.9) gives very useful connections between the various coordinate representations of $\hat{\xi}'$ and $\hat{\xi}$, and the included angle ψ . In particular, this equation allows us to compute the scattering angle ψ when light is scattered from an incident direction $\hat{\xi}'$ to a final direction $\hat{\xi}$.

The set of all directions $\hat{\xi}$ is called the *unit sphere of directions*, which is denoted by Ξ . Referring to polar coordinates, Ξ therefore represents all (θ, ϕ) values such that $0 \leq \theta \leq \pi$ and $0 \leq \phi < 2\pi$. Two subsets of Ξ frequently employed in optical oceanography are the

downward (subscript d) and *upward* (subscript u) *hemispheres* of directions, Ξ_d and Ξ_u , defined by

$$\begin{aligned}\Xi_d &\triangleq \text{all } (\theta, \phi) \text{ such that } 0 \leq \theta \leq \pi/2 \text{ and } 0 \leq \phi < 2\pi, \\ \Xi_u &\triangleq \text{all } (\theta, \phi) \text{ such that } \pi/2 < \theta \leq \pi \text{ and } 0 \leq \phi < 2\pi.\end{aligned}$$

As shown in Fig. 1.6, it is common to measure depth as positive downward from zero at the mean sea surface. For oceanographers, it is convenient to measure depth as *physical depth* z in meters⁵. However, the relevant depth for computations of light attenuation is the *optical depth* ζ , which is computed as

$$\zeta(z, \lambda) = \int_0^z c(z', \lambda) dz', \quad (1.10)$$

where $c(z, \lambda)$ is the beam attenuation coefficient, which is in general a function of depth and wavelength. Beam attenuation has units of m^{-1} , so ζ is nondimensional. The beam attenuation coefficient is defined in Section 3.0.1, and the role of optical depth in radiative transfer calculations is seen in Sections 9.1.2 and E.2. Note that the optical depth corresponding to a given physical depth depends on wavelength. This is another way of saying that light of different wavelengths attenuates with physical depth at different rates. For directions not perpendicular to the sea surface, the nondimensional distance traveled by a collimated beam of light is usually called the optical distance or the optical path length, denoted by τ . (Atmospheric optics uses τ to denote the optical depth or optical thickness of the atmosphere measured normal to the Earth's surface; see Section 15.4.)

1.4.1.2 Local coordinate systems

In Monte Carlo ray tracing calculations, for example, it is necessary to specify the polar and azimuthal scattering angles of a scattered ray relative to the direction of the unscattered ray (see Section F.4). This requires defining a local (at the end point of the unscattered ray) Cartesian system “centered” on the direction of the unscattered ray, rather than on a fixed $\hat{\mathbf{z}}$ direction. A local system that meets the needs of Monte Carlo ray tracing can be defined as follows.

Referring to Fig. 1.7, let $\mathbf{R}' = R' \hat{\boldsymbol{\xi}}'$ denote a light ray with starting point $\mathbf{r}_1 = (x_1, y_1, z_1)$ and ending point $\mathbf{r}_2 = (x_2, y_2, z_2)$ in a global coordinate system. R' is the length of the vector \mathbf{R}' , which is traveling in direction $\hat{\boldsymbol{\xi}}' = (\theta', \phi')$ in the global system. Now suppose that this ray scatters at point \mathbf{r}_2 to create a new ray \mathbf{R} , which ends at point \mathbf{r}_3 . To describe that scattering, it is necessary to define a local (at the point of scattering \mathbf{r}_2) coordinate system for the scattering calculations. The scattering angles ψ and α (computed as described in Section E.3) will then be applied in this system to determine the direction of the scattered ray \mathbf{R} relative to the unscattered ray \mathbf{R}' .

In the *ocean* system, ray \mathbf{R}' has components (R'_x, R'_y, R'_z) :

$$\mathbf{R}' = R'_x \hat{\mathbf{x}} + R'_y \hat{\mathbf{y}} + R'_z \hat{\mathbf{z}}$$

⁵An in-water instrument normally measures pressure, which is then converted to physical depth using a salinity- and temperature-dependent equation of state for sea water; see Saunders (1981). As an approximation accurate to a few percent, one decibar of pressure is equivalent to one meter of depth in water.

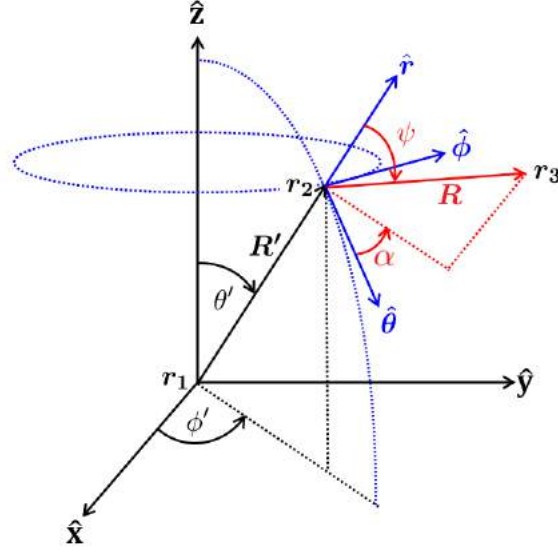


Figure 1.7: A local coordinate system used to describe the scattering of an in-water ray. The initial or unscattered ray is \mathbf{R}' ; the scattered ray \mathbf{R} is in red. The ocean coordinate system and angles measured in this system are in black, the local coordinate system is in blue, and angles measured in the local system are in red. The blue dotted lines are lines of constant θ' and constant ϕ' ; $\hat{\phi}$ and $\hat{\theta}$ are respectively tangent to these lines at point \mathbf{r}_2 .

$$= R' \sin \theta' \cos \phi' \hat{\mathbf{x}} + R' \sin \theta' \sin \phi' \hat{\mathbf{y}} + R' \cos \theta' \hat{\mathbf{z}}, \quad (1.11)$$

where the Cartesian components R'_x etc., in terms of spherical coordinates come from Eq (1.4).

A convenient *local* coordinate system $(\hat{\theta}, \hat{\phi}, \hat{r})$ for scattered rays is constructed as follows. The radial unit vector

$$\hat{\mathbf{r}} = \frac{\mathbf{R}'}{R'} = \frac{\mathbf{r}_2 - \mathbf{r}_1}{|\mathbf{r}_2 - \mathbf{r}_1|}$$

is in the same direction as the initial ray \mathbf{R}' . The azimuthal unit vector $\hat{\phi}$ is defined by the cross product of the ocean coordinate system $\hat{\mathbf{z}}$ and the incident vector's direction:

$$\hat{\phi} = \frac{\hat{\mathbf{z}} \times \hat{\mathbf{r}}}{|\hat{\mathbf{z}} \times \hat{\mathbf{r}}|}.$$

This vector points in the direction of increasing ϕ' values. (If the unscattered vector is in the same direction as $\hat{\mathbf{z}}$, the direction of $\hat{\phi}$ can be chosen at random.) The polar unit vector is then given by

$$\hat{\theta} = \hat{\phi} \times \hat{\mathbf{r}}.$$

This vector points in the direction of increasing θ' values. (If you think of point \mathbf{r}_2 as being on the surface of the Earth, then $\hat{\theta}$ points south, $\hat{\phi}$ points east, and $\hat{\mathbf{r}}$ is straight up.) The $(\hat{\theta}, \hat{\phi}, \hat{\mathbf{r}})$ system is then an orthogonal system of coordinates in which the scattering angles ψ and α can be applied to define the direction of the scattered ray. However, these directions are not fixed in the ocean system; they depend on the direction of the unscattered ray \mathbf{R}' .

After the direction of the scattered ray has been determined in the local $(\hat{\theta}, \hat{\phi}, \hat{r})$ system, the scattered ray direction must be specified in terms of (θ', ϕ') in the global system, so that the scattering process can be repeated. This transformation is not trivial; it is described in Section F.4.

Similarly, in calculations involving polarization, a local Cartesian system is needed to resolve the components of the polarization relative to some reference plane. This reference plane can be either the meridian plane, which is the plane containing \hat{z} and the direction of propagation; or the scattering plane, which is the plane containing the incident and final directions of the scattered light (see Figs. 1.20 and 1.21). Again, the appropriate local coordinate system depends on the direction of propagation of the light.

1.4.2 Solid Angle

Closely related to the specification of directions in three-dimensional space is the concept of *solid angle*, which is an extension of two-dimensional plane angle measurement. As illustrated in the left panel of Fig. 1.8, the plane angle θ between two radii of a circle of radius r is

$$\theta \triangleq \frac{\text{arc length}}{\text{radius}} = \frac{\ell}{r} \quad (\text{rad}).$$

The angular measure of a full circle is therefore 2π rad. In the right panel of Fig. 1.8, a patch of area A is shown on the surface of a sphere of radius r . The boundary of A is traced out by a set of directions $\hat{\xi}$. The solid angle Ω of the set of directions defining the patch A is by definition

$$\Omega \triangleq \frac{\text{area}}{\text{radius squared}} = \frac{A}{r^2} \quad (\text{sr}).$$

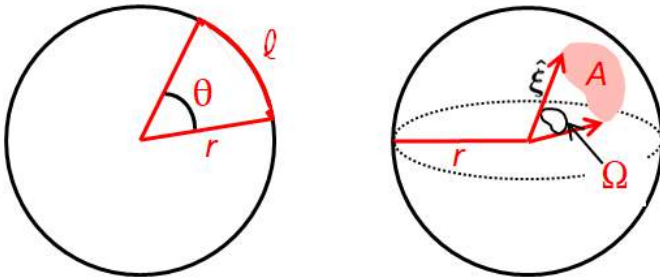


Figure 1.8: Geometry associated with the definition of plane angle (left) and solid angle (right).

Since the area of a sphere is $4\pi r^2$, the solid angle measure of the set of all directions is $\Omega(\Xi) = 4\pi$ sr. Note that both plane angle and solid angle are independent of the radii of the respective circle and sphere. Both plane and solid angle are dimensionless numbers. However, they are given “units” of radians and steradians, respectively, to remind us that they are measures of angle.

Consider a simple application of the definition of solid angle and the observation that a full sphere has 4π sr. The area of Brazil is $8.5 \cdot 10^6$ km² and the area of the earth’s surface is $5.1 \cdot 10^8$ km². The solid angle subtended by Brazil as seen from the center of the earth is then $4\pi 8.5 \cdot 10^6 / 5.1 \cdot 10^8 = 0.21$ sr.

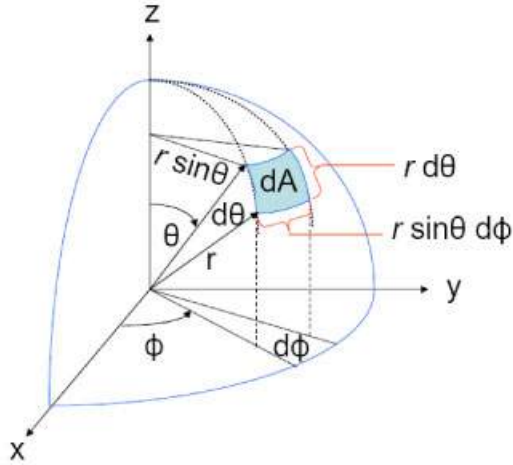


Figure 1.9: Geometry used to obtain an element of solid angle in spherical coordinates.

The definition of solid angle as area on the surface of a sphere divided by radius of the sphere squared gives us a convenient form for a *differential element of solid angle*, as needed for computations. The blue patch shown in Fig. 1.9 represents a differential element of area dA on the surface of a sphere of radius r . Simple trigonometry shows that this area is $dA = (r \sin \theta d\phi)(r d\theta)$. Thus the element of solid angle $d\Omega(\hat{\xi})$ about the direction $\hat{\xi} = (\theta, \phi)$ is given in polar coordinate form by⁶

$$d\Omega(\hat{\xi}) = \frac{dA}{r^2} = \frac{(r \sin \theta d\phi)(r d\theta)}{r^2} = \sin \theta d\theta d\phi = d\mu d\phi \quad (\text{sr}). \quad (1.12)$$

To illustrate the use of Eq. (1.12), let us compute the solid angle of a “polar cap” of half angle θ , i.e. all (θ', ϕ') such that $0 \leq \theta' \leq \theta$ and $0 \leq \phi' < 2\pi$. Integrating the element of solid angle over this range of (θ', ϕ') gives

$$\Omega_{\text{cap}} = \int_{\phi'=0}^{2\pi} \int_{\theta'=0}^{\theta} \sin \theta' d\theta' d\phi' = 2\pi(1 - \cos \theta), \quad (1.13)$$

or

$$\Omega_{\text{cap}} = \int_{\phi'=0}^{2\pi} \int_{\mu'=\mu}^1 d\mu' d\phi' = 2\pi(1 - \mu). \quad (1.14)$$

Note that Ξ_{d} and Ξ_{u} are special cases of a spherical cap (having $\theta = \pi/2$), and that $\Omega(\Xi_{\text{d}}) = \Omega(\Xi_{\text{u}}) = 2\pi$ sr.

1.4.3 Dirac Delta Functions

It is sometimes convenient to specify directions using the *Dirac delta function*, $\delta(\hat{\xi} - \hat{\xi}_0)$. This peculiar mathematical construction is defined (for our purposes) by

$$\delta(\hat{\xi} - \hat{\xi}_0) \triangleq 0 \quad \text{if } \hat{\xi} \neq \hat{\xi}_0, \quad (1.15)$$

⁶This equation is correct even though $d\mu = d\cos \theta = -\sin \theta d\theta$. When the differential element is used in an integral and variables are changed from (θ, ϕ) to (μ, ϕ) , the Jacobian of the transformation involves an absolute value.

and

$$\int_{\Xi} f(\hat{\xi}) \delta(\hat{\xi} - \hat{\xi}_o) d\Omega(\hat{\xi}) \triangleq f(\hat{\xi}_o). \quad (1.16)$$

Here $f(\hat{\xi})$ is any function of direction. Note that $\delta(\hat{\xi} - \hat{\xi}_o)$ simply “picks out” the particular direction $\hat{\xi}_o$ from all directions in Ξ . Note also in Eq. (1.16) that because the element of solid $d\Omega(\hat{\xi})$ has units of steradians, it follows that $\delta(\hat{\xi} - \hat{\xi}_o)$ must have units of inverse steradians.

Equations (1.15) and (1.16) are a symbolic definition of δ . The *mathematical representation* of $\delta(\hat{\xi} - \hat{\xi}_o)$ in spherical coordinates (θ, ϕ) is

$$\delta(\hat{\xi} - \hat{\xi}_o) = \frac{\delta(\theta - \theta_o) \delta(\phi - \phi_o)}{\sin \theta} \quad (\text{sr}^{-1}), \quad (1.17)$$

where $\hat{\xi} = (\theta, \phi)$, $\hat{\xi}_o = (\theta_o, \phi_o)$, and

$$\begin{aligned} \int_0^\pi f(\theta) \delta(\theta - \theta_o) d\theta &\triangleq f(\theta_o) \\ \int_0^{2\pi} f(\phi) \delta(\phi - \phi_o) d\phi &\triangleq f(\phi_o). \end{aligned}$$

Note that the $\sin \theta$ in the denominator of Eq. (1.17) is necessary to cancel the $\sin \theta$ factor in the element of solid angle when integrating in polar coordinates. Thus

$$\begin{aligned} \int_{\Xi} f(\hat{\xi}) \delta(\hat{\xi} - \hat{\xi}_o) d\Omega(\hat{\xi}) &= \\ \int_0^{2\pi} \int_0^\pi f(\theta, \phi) \frac{\delta(\theta - \theta_o) \delta(\phi - \phi_o)}{\sin \theta} \sin \theta d\theta d\phi &= \\ f(\theta_o, \phi_o) &= f(\hat{\xi}_o). \end{aligned}$$

Likewise, we can write

$$\delta(\hat{\xi} - \hat{\xi}_o) = \delta(\mu - \mu_o) \delta(\phi - \phi_o) \quad (\text{sr}^{-1}), \quad (1.18)$$

where

$$\int_{-1}^1 f(\mu) \delta(\mu - \mu_o) d\mu \triangleq f(\mu_o).$$

Although Dirac delta functions entered the present discussion as a way to select directions, they can be formulated for almost any variable. For example, if $f(x)$ is a function of location x , then $\delta(x - x_o)$ picks out the value of $f(x_o)$ via

$$\int_{-\infty}^{\infty} f(x) \delta(x - x_o) dx = f(x_o).$$

If x is measured in meters, then $\delta(x - x_o)$ has units of 1/meters.

1.5 Geometrical Radiometry

[Curtis Mobley and Kenneth Voss contributed to this section.]

By housing one or more radiant energy detectors in watertight assemblies and by appropriately channeling the direction of the radiant energy arriving at the detector, we can measure the flow of radiant energy as a function of direction at any location within a water body. By adding appropriate filters to the instrument, we can also measure the wavelength dependence and state of polarization of the light field. From such measurements we can develop precise descriptions of light propagation in natural waters. Thus we are led to the science of *geometrical radiometry*, the union of Euclidean geometry and radiometry.

We first define *radiance*, the fundamental quantity that describes light in radiometric terms. We then define various *irradiances* and other quantities that are derivable from the radiance, and which are often easier to measure and of more relevance to a particular problem.

1.5.1 Radiance

Figure 1.10 shows the design of a *Gershun tube radiometer*. A circular hole at one end of the housing or collecting tube and a system of internal light baffles allows the detection of light that enters the hole only at angles of γ or less (measured from the axis of the tube). Angle γ is the *field of view (FOV) half angle*. The blue arrow $\hat{\xi}$ in the figure represents light entering the tube within the field of view. Light entering the tube at angles larger than γ , represented by the red arrow $\hat{\xi}'$ in the figure, is blocked by the baffles or absorbed by the inner walls of the tube and is not detected. In most instruments, γ is around 5 deg (at most 10 deg). The solid angle seen by the detector is $\Delta\Omega = 2\pi(1 - \cos\gamma)$.

If the state of polarization is of interest, a polarizing filter is placed in the tube so that the light passes first through the polarizer. This filter will be chosen to measure the desired component of the polarization, e.g, horizontal or vertical plane polarization, or left or right circular polarization. In practice, a sequence of measurements must be made to determine the four components of the Stokes vector; see Sections 1.6 and 1.6 on polarization for further discussion.

A wavelength filter is normally used to select a narrow band of wavelengths. In modern instruments this is usually an interference filter that passes light of some bandwidth $\Delta\lambda$ centered at the nominal wavelength λ . The bandwidth is usually specified by the full-width at half maximum (FWHM) of the filter transmission function. In the early days of optical oceanography, the filter was often a gelatin filter that passed certain wavelengths and absorbed others. In most instruments $\Delta\lambda$ is around 10 nm (at most, 20 nm), but can be only a few nanometers for hyperspectral instruments.

Finally, the light, which is still well collimated after passing through the polarizer and wavelength filter, passes through a translucent diffuser of collection area ΔA . The diffuser makes the light field spatially homogeneous in the region near the detector, so that it is necessary to sample only a part of the internal light field in order to measure the total energy entering the instrument. To the accuracy with which $\cos\gamma = 1$, the solid angle $\Delta\Omega$ of the hole as seen by any point on the diffuser surface is the same.

In order to measure an entire spectrum, an instrument may have a rotating wheel holding many different wavelength filters, or it may use a prism to spread the different wavelengths out along a CCD array. Similarly, an instrument may have a rotating wheel

holding different polarization filters, so as to obtain all components of the Stokes vector by a sequence of measurements as the wheel rotates. These engineering details do not concern us in the present discussion, but they are of great importance in the construction of actual instruments.

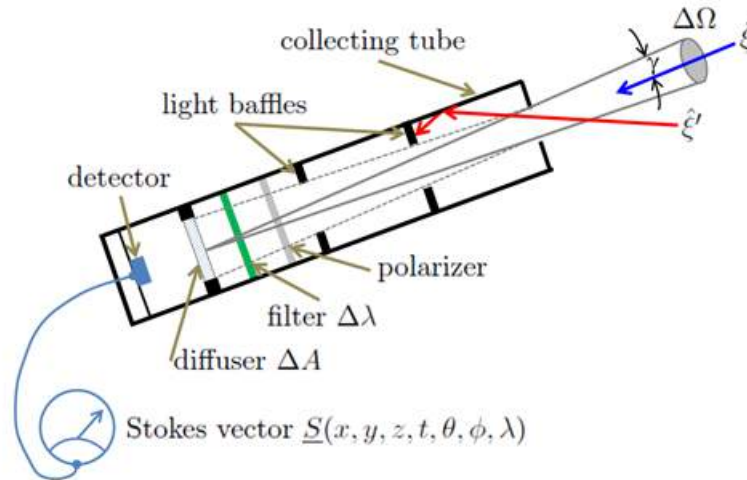


Figure 1.10: Conceptual design of a Gershun tube radiometer.

The instrument design of Fig. 1.10 traces back to the Russian physicist and pioneer of optical oceanography A. A. Gershun in the 1930s. Instruments implementing variants of his design are still in production. Figure 1.11 shows a more modern design that uses lenses and a pinhole in a diaphragm to select the directions of light propagation that can be detected. As shown, an objective lens focuses the incoming light onto an opaque diaphragm, which contains a small hole on the optical axis. Light propagating in a narrow range of directions (the FOV) will pass through the pinhole and be detected. This is illustrated by the blue arrow $\hat{\xi}$ and the bluish shading in the figure. Light propagating in directions outside the FOV will not pass through the pinhole. This light is illustrated by the red arrow $\hat{\xi}'$ and the reddish shading. The light that does pass through the pinhole is then collimated by a relay lens and directed towards a detector with detecting area ΔA . Again, polarizers or wavelength filters can be part of the design. As drawn, the polarizer is placed after the two lenses. This layout is acceptable for glass lenses, which do not significantly affect the state of polarization. If the lenses are made of plastic or another material that alters the state of polarization, then the polarizer needs to be in front of the objective lens. The FOV and corresponding solid angle seen by the detector are determined by the lens and pinhole geometry.

The designs of both Figs. 1.10 and 1.11 are termed “well collimated radiometers” because they detect light propagating in only a narrow set of nearly collimated directions.

If the instrument is pointing in the $-\hat{\xi}$ direction, it collects light traveling in a set of directions of solid angle $\Delta\Omega$ centered on direction $\hat{\xi}$. We assume that the instrument is small compared to the scale of spatial (positional) changes in the light field, so that we can think of the instrument as being located at a point $\mathbf{x} = (x, y, z) = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$ within a water body. Suitable calibration of the current or voltage output of the detector gives the amount of radiant energy ΔQ entering the instrument during a time interval Δt centered

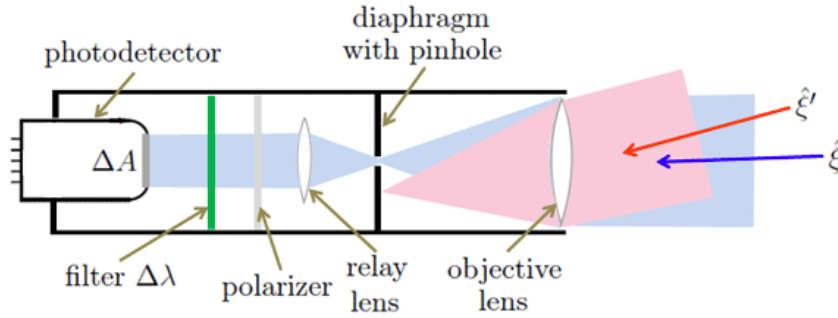


Figure 1.11: Modern design of a well collimated radiometer. Figure based on [Mishchenko \(2014, Fig. 12\(e\)\)](#).

on the time t . For simplicity, consider the total radiance without regard for the state of polarization (that is, remove the polarizer in either of the instrument designs); this is the first component of the Stokes vector and is commonly called the radiance. An *operational definition*⁷ of the *unpolarized spectral radiance* is then

$$L(\mathbf{x}, t, \hat{\boldsymbol{\xi}}, \lambda) \triangleq \frac{\Delta Q}{\Delta t \Delta A \Delta \Omega \Delta \lambda} \quad (\text{J s}^{-1} \text{ m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}). \quad (1.19)$$

In practice the intervals Δt , ΔA , $\Delta \Omega$, and $\Delta \lambda$ are taken small enough to get a useful resolution of L over the various parameter domains, but not so small as to encounter diffraction effects or fluctuations from photon shot noise at low light levels. Typical values are $\Delta t \approx 10^{-2}$ to 10^2 s (depending on whether one wishes an “instantaneous” measurement or wishes to average out sea surface wave effects), $\Delta A \approx 10^{-4}$ m², $\Delta \Omega \approx 0.01$ to 0.1 sr, and $\Delta \lambda \approx 1$ to 10 nm.

Since sampling times Δt are generally long compared to the time ($\approx 10^{-6}$ s) required for the light field in a water body to reach steady state after a change in the environment, time-independent radiative transfer theory is usually sufficient for hydrologic optics studies. In this case, the time is implicitly understood and the argument t can be omitted. A notable exception to the use of time-independent radiative transfer theory occurs with the use of pulsed lasers to determine water depth or to detect underwater objects. In this application, the laser pulses last only for nanoseconds and time-dependent theory must be used. In addition, time-averaged (over seconds to minutes) horizontal variations (on a scale of meters to kilometers) in the environment and in the optical properties of natural water bodies illuminated by the Sun are usually much less than vertical variations. In that case, underwater light fields depend spatially only on the depth $x_3 = z$ to a good approximation. Thus, for example, we often can refer to just “the radiance $L(z, \theta, \phi, \lambda)$ ” without generating confusion. An important exception of this situation occurs with artificial light sources, such as an underwater light or a laser being used for bathymetric mapping. In these cases the horizontal variations in the radiance can be large, and the three-dimensional spatial variations of the light field must be considered.

In the conceptual limit of infinitesimal parameter intervals, the spectral radiance is

⁷An operational definition of something is a definition in terms of quantities or procedures that can be implemented by an instrument.

given by

$$L(\mathbf{x}, t, \hat{\boldsymbol{\xi}}, \lambda) = \frac{\partial^4 Q}{\partial t \partial A \partial \Omega \partial \lambda}. \quad (1.20)$$

This definition of radiance is convenient for the mathematical development of radiative transfer theory. The operational definition is what is used for actual measurements.

Spectral radiance is the fundamental radiometric quantity of interest in hydrologic optics because all other radiometric quantities can be derived from the radiance. It specifies the spatial (\mathbf{x}), temporal (t), directional ($\hat{\boldsymbol{\xi}}$), and wavelength (λ) structure of the light field. In radiative transfer theory, the direction $\hat{\boldsymbol{\xi}} = (\theta, \phi)$ always denotes *the direction of propagation of the radiance*. The *viewing direction* $\hat{\boldsymbol{\xi}}_v$, defined in Eq. , is the direction an instrument is pointed to detect $L(\mathbf{x}, t, \hat{\boldsymbol{\xi}}, \lambda)$.

Although radiance is an extremely useful concept and is adequate for most needs of optical oceanography, it is an approximation to the exact description of light in terms of electric and magnetic fields. We therefore anticipate that there are situations for which radiance fails to give an adequate description of the light field. When that happens, we must resort to Maxwell’s Equations (Chapter 11) and compute the electric and magnetic fields themselves. Some of the limitations of radiance are discussed in Section 1.8. Moreover, both because of instrumental difficulties and because such detailed directional information often is not needed for specific applications, the most commonly measured radiometric quantities are various irradiances.

1.5.2 Plane Irradiance

If the collecting tube is removed from the instrument of Fig. 1.10, then light from an entire hemisphere of directions can reach the detector, as illustrated in Fig. 1.12. Such an instrument, when pointed “straight up” (in the $-\hat{\mathbf{z}}$ direction) so as to detect light headed *downward* (all $\hat{\boldsymbol{\xi}}$ in Ξ_d) measures the *spectral downwelling plane irradiance* E_d :

$$E_d(\mathbf{x}, t, \lambda) \triangleq \frac{\Delta Q}{\Delta t \Delta A \Delta \lambda} \quad (\text{W m}^{-2} \text{ nm}^{-1}). \quad (1.21)$$

Implicit in this definition is the assumption that each *point* of the collector surface is equally sensitive to light incident onto the surface from any angle. If this is the case, however, the collector *as a whole* is *not* equally sensitive to light headed in all downward directions. Imagine a collimated beam of light headed straight downward (e.g. from the Sun straight overhead). This beam, assumed to be larger than the collector surface, sees the full area ΔA of the collector surface. However, the same large beam traveling at an angle θ relative to the instrument axis sees a collector surface of effective area $\Delta A |\cos \theta|$ (the area ΔA as projected onto a plane perpendicular to the beam direction). Otherwise identical collimated light beams therefore generate detector responses that are proportional to the cosines of the incident directions. Such instruments are called *cosine collectors*.

The *cosine law for irradiance* is simply the statement that a collimated beam of light intercepting a plane surface produces an irradiance that is proportional to the cosine of the angle between the incident directions and the normal to the collector surface.

There are subtleties in the construction of instruments that require a cosine response. No real material used for the collector actually obeys the requirement that every point of the material surface be equally efficient at collecting energy from any direction. In practice, surfaces often look a bit “shiny” near grazing angles (θ near 90 deg). Thus more of the

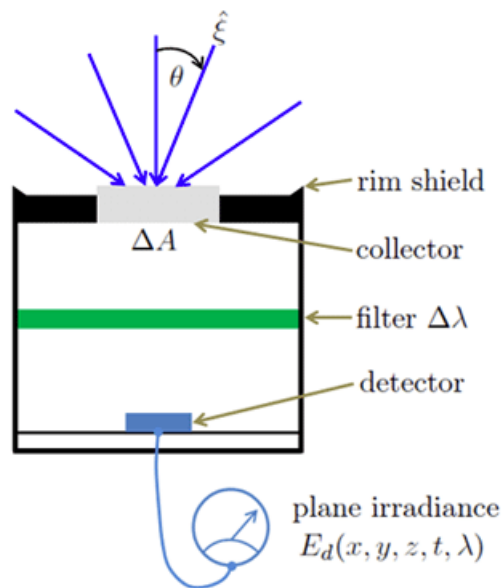


Figure 1.12: Schematic design of an instrument for measuring spectral plane irradiance.

incident energy is reflected by the collector surface at large θ values. To correct for this less efficient energy collection at large θ , the collector itself often extends a small distance above the instrument housing, as seen in Fig. 1.12. The vertical walls of this “button” then receive the light at more nearly normal directions to the vertical edge of the collector. However, that vertical collector side would collect light propagating at $\theta = 90$ deg, which should give no response according to the cosine law. To block light incident when θ is within a few degrees of 90, the rim of the housing has a low “wall” that blocks light at very nearly 90 deg. An actual implementation of this idea is seen in Fig. 1.13. Modern instruments with similar designs achieve an in-water cosine response to within a percent or two.

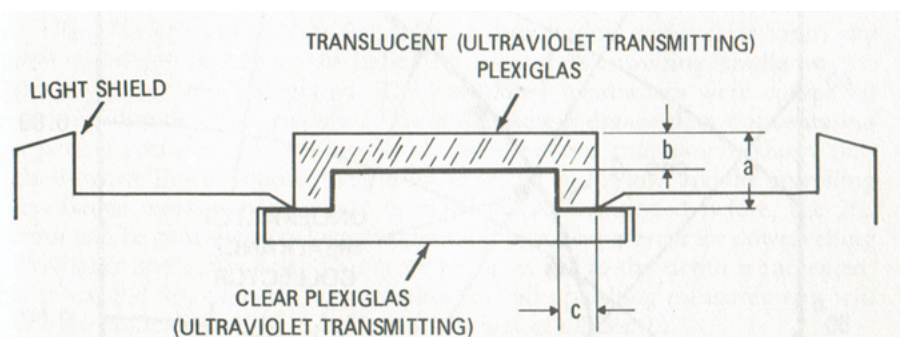


Figure 8 Cross sectional sketch showing construction of an underwater irradiance collector.

Figure 1.13: Design of the plane irradiance sensor head constructed by J. E. Tyler and R. C. Smith at the Scripps Institution of Oceanography in the 1960s. From Tyler and Smith (1970).

Another very important issue with the construction an in-water cosine collector is that

the reflective properties of collector materials such as glass, plastic, or Spectralon are much different in water than in air. This is because of the differences in the collector-air and collector-water indices of refraction. For example, Spectralon, which is often used in optical instruments, has a refractive index of 1.35. This is very close to that of water, which is about 1.33 (with some wavelength variation in each). Air has a refractive index of almost exactly 1. At normal incidence the reflectance of an air-Spectralon surface is

$$r_F = \left(\frac{1.35 - 1.00}{1.35 + 1.00} \right)^2 \approx 0.022$$

For a water-Spectralon surface the corresponding reflectance is only 0.00006. Thus more light gets “into” the collector made of Spectralon when the instrument is in water than when it is in air. Conversely, more of the collected light can escape “back out” from the Spectralon collector in water than in air. Therefore different fractions of the incident energy are collected and reach the detector for the same instrument when used in air and in water. This difference is called the *immersion effect*.

The consequence of these observations is that if a sensor is to be used in water, the design (such as seen in Fig. 1.13) must be such that the sensor has a cosine response in water, and the sensor calibration must account for the immersion effect. These matters are crucial to the design of irradiance sensors, but the details can be left to the optical engineers.

Since the instrument of Fig. 1.12 collects light traveling in all downward directions, but with detector’s surface area weighted by the cosine of the light’s incident angle θ , the instrument is in essence integrating $L(\mathbf{x}, t, \hat{\boldsymbol{\xi}}, \lambda) |\cos \theta|$ over all incident directions in the hemisphere seen by the sensor. The spectral downwelling plane irradiance is therefore related to the spectral radiance by

$$\begin{aligned} E_d(\mathbf{x}, t, \lambda) &= \int_{\hat{\boldsymbol{\xi}} \in \Xi_d} L(\mathbf{x}, t, \hat{\boldsymbol{\xi}}, \lambda) |\cos \theta| d\Omega(\hat{\boldsymbol{\xi}}) \\ &= \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi/2} L(\mathbf{x}, t, \theta, \phi, \lambda) |\cos \theta| \sin \theta d\theta d\phi. \end{aligned} \quad (1.22)$$

If the same instrument is oriented *downward*, so as to detect light heading *upward*, then the quantity being measured is the *spectral upwelling plane irradiance* E_u :

$$E_u(\mathbf{x}, t, \lambda) = \int_{\phi=0}^{2\pi} \int_{\theta=\pi/2}^{\pi} L(\mathbf{x}, t, \theta, \phi, \lambda) |\cos \theta| \sin \theta d\theta d\phi \quad (1.23)$$

Note that it is necessary to take the absolute value of $\cos \theta$ in Eq. (1.23) because, with our choice of coordinates, $\cos \theta < 0$ for $\hat{\boldsymbol{\xi}}$ in Ξ_u . The absolute value is superfluous in Eq. (1.22).

1.5.3 Scalar Irradiance

Now consider an instrument that is designed to be *equally* sensitive to all light headed in the downward direction. Such an instrument is shown in Fig. 1.14. The spherical shape of the collector and diffuser insures that the instrument is equally sensitive to light from any direction. If each *point* on the diffuser surface behaves like a cosine collector, then

the effective area of the collector is $\Delta A = \pi r^2$, where r is the radius of the diffuser. In addition, the radiance distribution in the hollow interior of the collector is isotropic, so that only a small part of the interior light field needs to be measured. The large shield blocks upward-traveling light. In principle, this shield extends to infinity in all directions, but in reality may be only a few centimeters in diameter. The upper surface of the shield is assumed to be completely absorbing, so that it cannot reflect downward traveling light back upward onto the collector. These cases are illustrated by the red arrows in the figure. This instrument, when oriented upward as shown in Fig. 1.14, measures the *spectral downwelling scalar irradiance* E_{od} , which is related to the spectral radiance by

$$\begin{aligned} E_{od}(\mathbf{x}, t, \lambda) &= \int_{\hat{\xi} \in \Xi_d} L(\mathbf{x}, t, \hat{\xi}, \lambda) d\Omega(\hat{\xi}) \\ &= \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi/2} L(\mathbf{x}, t, \theta, \phi, \lambda) \sin \theta d\theta d\phi \end{aligned} \quad (1.24)$$

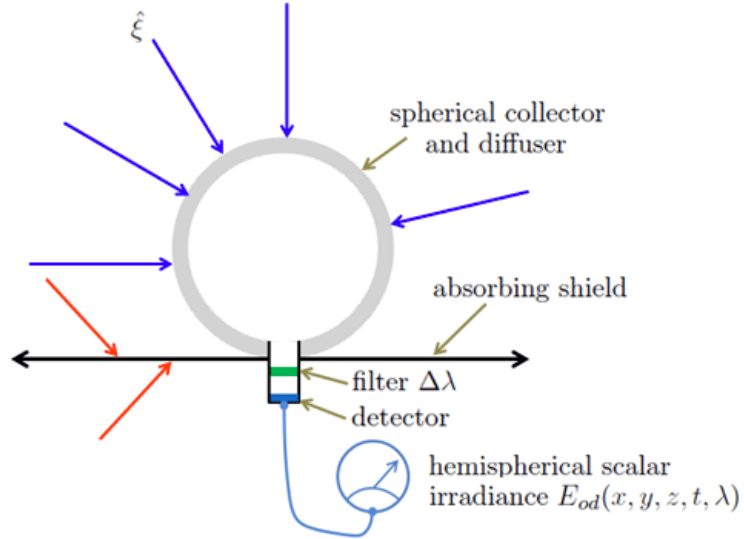


Figure 1.14: Schematic design of an instrument for measuring spectral hemispherical scalar irradiance.

If the instrument of Fig. 1.14 is inverted, so as to collect only upward traveling light, then it measures the *spectral upwelling scalar irradiance* E_{ou} . If this shield is removed, light traveling in all directions is collected. The quantity then measured is the *spectral total scalar irradiance* E_o , which is related to the radiance by

$$\begin{aligned} E_o(\mathbf{x}, t, \lambda) &= \int_{\hat{\xi} \in \Xi} L(\mathbf{x}, t, \hat{\xi}, \lambda) d\Omega(\hat{\xi}) \\ &= E_{od}(\mathbf{x}, t, \lambda) + E_{ou}(\mathbf{x}, t, \lambda). \end{aligned} \quad (1.25)$$

Other possible instrument designs are discussed in Højerslev (1975). However, the three basic types shown above are sufficient for most of the needs of hydrologic optics. Such instruments are commercially available.

1.5.4 Vector Irradiance

The *spectral vector irradiance* \vec{E} is defined as

$$\vec{E}(\mathbf{x}, t, \lambda) = \int_{\hat{\xi} \in \Xi} L(\mathbf{x}, t, \hat{\xi}, \lambda) \hat{\xi} d\Omega(\hat{\xi}). \quad (1.26)$$

Recalling that the vertical component of $\hat{\xi}$ is just $\hat{\mathbf{z}} \cdot \hat{\xi} = \cos \theta$, we can write the vertical component of the vector irradiance as

$$\begin{aligned} (\vec{E})_z &= \hat{\mathbf{z}} \cdot \vec{E} \\ &= \int_{\Xi} L(\mathbf{x}, t, \hat{\xi}, \lambda) \cos \theta d\Omega(\hat{\xi}) \\ &= E_d - E_u \end{aligned} \quad (1.27)$$

In developing Eq. (1.27), we have noted that $\cos \theta > 0$ in Ξ_d and $\cos \theta < 0$ in Ξ_u . The quantity $E_d - E_u$ is called the *net downward irradiance*. This net downward irradiance often is often called the “vector” irradiance, although strictly speaking it is only the vertical component of the vector irradiance. If the radiance distribution is horizontally homogeneous, the horizontal components of the vector irradiance are zero.

1.5.5 Example: Irradiances of an Isotropic Radiance Distribution

Consider an *isotropic*, or directionally uniform, radiance distribution: $L(\mathbf{x}, t, \hat{\xi}, \lambda) = L_o(\mathbf{x}, t, \lambda)$ for all $\hat{\xi}$ in Ξ . Then by Eq. (1.22), the downward plane irradiance is

$$\begin{aligned} E_d(\mathbf{x}, t, \lambda) &= \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi/2} L_o(\mathbf{x}, t, \lambda) |\cos \theta| \sin \theta d\theta d\phi \\ &= \pi L_o(\mathbf{x}, t, \lambda). \end{aligned}$$

In the last equation, π carries units of steradian from the integration over solid angle; thus E_d has units of irradiance when L_o has units of radiance. Likewise, $E_u = \pi L_o$, so that the net downward irradiance is zero. The scalar irradiance E_{od} is given by Eq. 1.24

$$\begin{aligned} E_{od}(\mathbf{x}, t, \lambda) &= \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi/2} L_o(\mathbf{x}, t, \lambda) \sin \theta d\theta d\phi \\ &= 2\pi L_o(\mathbf{x}, t, \lambda) \end{aligned}$$

In the same manner we find $E_{ou} = 2\pi L_o$, so that the total scalar irradiance $E_o = 4\pi L_o$.

1.5.6 Intensity

Another family of radiometric quantities can be defined from the measurements employed in the operational definition of radiance. The *spectral intensity* I is defined as

$$I(\mathbf{x}, t, \hat{\xi}, \lambda) \triangleq \frac{\Delta Q}{\Delta t \Delta \Omega \Delta \lambda} \quad (\text{W sr}^{-1} \text{ nm}^{-1}), \quad (1.28)$$

or

$$I(\mathbf{x}, t, \hat{\xi}, \lambda) = \int_{\Delta A} L(\mathbf{x}, t, \hat{\xi}, \lambda) dA.$$

In Eq. (1.28), ΔA is the surface of the collector that sees the solid angle $\Delta\Omega$, and dA is an element of area. The concept of intensity is useful in the radiometry of point light sources. We will use it in the definition of the volume scattering function 3.1.1.

Note that in radiometry “intensity” has the specific meaning of power per steradian. The word “intensity” should not be used as a synonym for magnitude, brightness, or irradiance. Such misuse has even prompted a publication (Palmer, 1993), the abstract of which begins “The misuse of the term *intensity* in physics and optics is deplored.”

1.5.7 Photosynthetically Available Radiation

Photosynthesis is a quantum process. That is to say, it is the *number* of photons absorbed rather than their total energy that is relevant to the chemical transformations. This is because a photon of, say, wavelength 400 nm, *if absorbed* by a chlorophyll molecule, induces the same chemical change as does a less energetic photon of wavelength 500 nm. (However, photons of different wavelengths are not equally likely to be absorbed.) Only a part of the photon’s energy goes into photosynthesis; the excess appears as heat or is re-radiated. Moreover, chlorophyll is equally able to absorb light regardless of the light’s direction of travel.

Now recall that the spectral total scalar irradiance $E_o(\mathbf{x}, \lambda)$ is the total radiant power per square meter at wavelength λ coursing through point \mathbf{x} owing to light traveling in all directions. The number of photons generating $E_o(\mathbf{x}, \lambda)$ is $E_o(\mathbf{x}, \lambda)\lambda/hc$. Therefore, in studies of phytoplankton biology, a useful measure of the underwater light field is the *photosynthetically available radiation*, PAR or E_{PAR} , which is defined by

$$PAR(\mathbf{x}) \triangleq \int_{400\text{ nm}}^{700\text{ nm}} E_o(\mathbf{x}, \lambda) \frac{\lambda}{hc} d\lambda \quad (\text{quanta s}^{-1} \text{ m}^{-2}). \quad (1.29)$$

Note that PAR is by definition a broadband quantity. Bio-optical literature often states PAR values in units of $\text{mol quanta s}^{-1} \text{ m}^{-2}$ or $\text{einst s}^{-1} \text{ m}^{-2}$. Recall from the table of derived units 1.3 units that one einstein is one mole of photons ($6.023 \cdot 10^{23}$ photons).

Morel and Smith (1974) found that over a wide variety of water types from very clear to turbid, the conversion factor for energy to photons varies by only $\pm 10\%$ about the value

$$2.5 \cdot 10^{18} \text{ quanta s}^{-1} \text{ W}^{-1} = 4.2 \mu\text{einst s}^{-1} \text{ W}^{-1}.$$

PAR is usually estimated using only the visible wavelengths, 400-700 nm, although some investigators include near-ultraviolet wavelengths in the integral of Eq. (1.29). PAR is also sometimes estimated using E_d rather than E_o . This usually underestimates PAR by 30% or more because E_d is always less than E_o .

Instruments for the direct measurement of PAR , often called *quanta meters*, can be constructed along the lines of Fig. 1.14 by the incorporation of broadband wavelength filters. Such instruments are commercially available. Engineering details can be found in Jerlov and Nygard (1969) and in Kirk (1994), which also discusses PAR and photosynthesis in great detail.

Although PAR has a venerable history as a simple parameterization of available light in phytoplankton growth models, it should be noted that PAR is an imperfect measure of how light contributes to photosynthesis. This is because different species of phytoplankton, or even the same species under different environmental conditions, have different suites of

pigments. Phytoplankton with different pigments absorb light differently as a function of wavelength. Thus phytoplankton with different pigments use the same $E_o(\lambda)$ with different efficiencies, thus giving one an advantage over the other. Recent ecosystem models therefore replace PAR by spectral scalar irradiance $E_o(\lambda)$ and also account for different pigments in different functional classes of phytoplankton. Such models can better simulate how light is utilized by different components of the ecosystem (e.g., Bissett et al., 1999; Fujii et al., 2007).

1.5.8 Terminology and Notation

The adjective “spectral,” as in spectral radiance, as commonly used can mean either “as a function of wavelength” or “per unit wavelength interval.” Committees on international standards recommend an argument λ , e.g. $L(\lambda)$, for the first meaning and a subscript λ , e.g. L_λ , for the second meaning (Meyer-Arendt, 1968). Thus $L_\lambda(\lambda)$ would denote the spectral radiance as a function of wavelength. However, this subscript convention is seldom used in optical oceanography, perhaps because the symbols already are cluttered with subscripts. Most authors seem to write $L(\lambda)$ and consider it to mean radiance per unit wavelength interval, as a function of wavelength. The adjective spectral and argument λ are often omitted for brevity, although strictly speaking, a term without the adjective “spectral” refers to a quantity integrated or measured over a finite band of wavelengths, as in the computation of PAR in Eq. (1.29). Most radiative transfer *theory* assumes the energy to be monochromatic, i.e., radiance per unit wavelength interval at a particular wavelength. However, most *measurements* are made over a wavelength band of 1 to 20 nm, which complicates the comparison of theory and observation.

Although this book uses the terminology and notation commonly seen in the current optical oceanography literature, much published work uses a different nomenclature. The recommended (modern) notation came into use in the late 1970s and is employed, for example, in Jerlov (1976) and further refined in Morel and Smith (1982). Prior to then, different symbols were often employed for the radiometric concepts just defined. This historical notation is seen, for example, in the monumental treatise *Hydrologic Optics* (Preisendorfer, 1976). Table 1.4 gives a comparison of historical and modern notation. Additional discussion of units and conversions is given in Palmer (2000).

Much of the radiative transfer work in atmospheric, astrophysical, and biomedical optics, and in neutron transport theory, is relevant to hydrologic optics. However, to a considerable extent, these fields have developed independently and each has its own nomenclature. For example, in atmospheric and astrophysical optics, radiance is often called “intensity” or “specific intensity” and is given the symbol I ; this use of “intensity” is not to be confused with intensity as defined above. The classic text *Light Scattering by Small Particles* by van de Hulst (1957) uses “intensity” for irradiance. The letter “ I ” often is used as the symbol for irradiance. The word “flux” is sometimes used to mean irradiance and is sometimes used to mean power. Those who use “flux” for power generally use “flux density” for irradiance. The ambiguity associated with “flux” can be avoided simply by using “power” or “irradiance,” as is appropriate. Matters are further complicated by the occasional misuse of photometric terms such as “brightness” and “luminance” for radiometric quantities; these matters are discussed in Chapter 16 on Photometry and Visibility.

It is occasionally convenient to distinguish conceptually between light leaving a surface and light arriving at a surface. In radiative transfer theory, *field radiance* L^- refers to the

Quantity	SI Units	Recommended Symbol	Historic Symbol
radiant energy	J	Q	U
radiant power	W	Φ	P
radiant intensity	W sr^{-1}	I	J
radiance	$\text{W m}^{-2} \text{sr}^{-1}$	L	N
plane irradiance	W m^{-2}	E	H
downward plane irradiance	W m^{-2}	E_d	$H(-)$
upward plane irradiance	W m^{-2}	E_u	$H(+)$
scalar irradiance	W m^{-2}	E_o	h
downward scalar irradiance	W m^{-2}	E_{od}	$h(-)$
upward scalar irradiance	W m^{-2}	E_{ou}	$h(+)$
vector irradiance	W m^{-2}	\vec{E}	\vec{H}
vertical net irradiance	W m^{-2}	$E_d - E_u$	—
emittance	W m^{-2}	M	W
photosynthetically available radiation	$\text{photons s}^{-1} \text{m}^{-2}$	PAR or E_{PAR}	—

Table 1.4: Terms, units, and symbols for radiometric quantities commonly used in hydrologic optics. The quantities as shown represent broadband measurements. For narrow band (monochromatic) measurements, add the adjective “spectral” to the term, add nm^{-1} to the units, and add a wavelength index λ to the symbol, e.g., spectral radiance, L_λ or $L(\lambda)$, with units of $\text{W m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$. PAR is inherently broadband.

radiance of light arriving at a surface; this is the quantity measured by a radiance meter. *Surface radiance* L^+ is the radiance attributed to a real or imaginary surface emitting light. *Irradiance* E refers to light incident onto a surface; the corresponding measure of light leaving a surface is denoted by *radiant exitance* or *emittance* M . Likewise, intensity can be subdivided into field intensity I^- and surface intensity I^+ . Figure 1.15 summarizes this hierarchy of radiometric concepts.

1.5.9 Spectral Quantities Expressed per Unit Frequency Interval

The spectral radiance and irradiances were defined above in terms of a unit wavelength interval, i.e., “per nanometer.” This is certainly convenient for instrument design because wavelength filters can select a narrow wavelength range of the incident light. However, it is sometimes convenient to express spectral radiometric variables in terms of a unit frequency interval rather than a unit wavelength interval, which brings us to our first crisis of notation. Depending on the problem—electromagnetic wave propagation, sea surface wave generation, image analysis, etc.—“frequency” may refer to either a temporal or a spatial frequency. We will encounter both meanings in this book, so care will be taken to specify either temporal or spatial frequency as needed. Depending on the scientific field—quantum physics, chemistry, astronomy, atmospheric optics, optical oceanography,

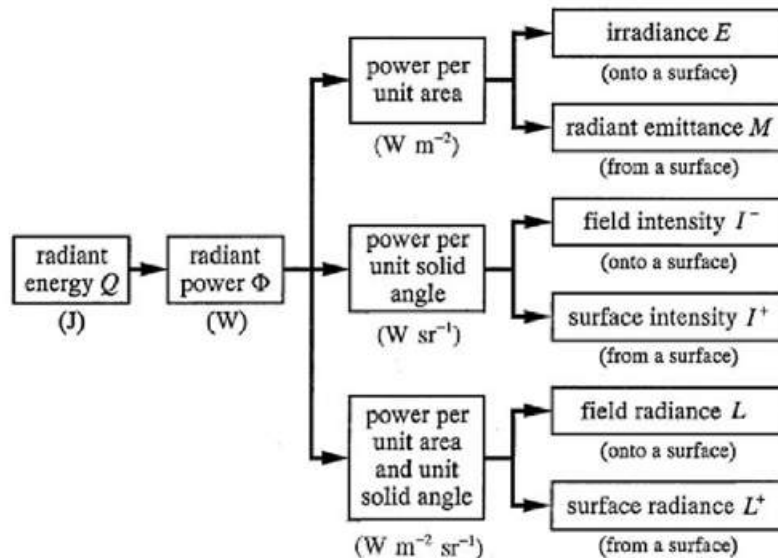


Figure 1.15: The hierarchy of radiometric concepts.

physical oceanography, etc.—one symbol or another will be used. You find the same symbol used for two different concepts, and different symbols used for the same concept. The notation used in this book is as follows.

- Let λ be the spatial wavelength in meters. This could be the wavelength of light, say $500 \cdot 100^{-9}$ m, or the wavelength of a water wave on the sea surface, say 50 m.
- Let T be the temporal period of the wave in seconds. This might be the period of one wavelength of light, say 10^{-14} s, or the period of a water wave on the sea surface, say 7 s.
- $f \triangleq 1/T$ is the *temporal frequency* (units of cycles per second or Hertz), often called just the frequency. Some authors use f and some use ν or σ' for temporal frequency.
- $\omega \triangleq 2\pi/T = 2\pi f$ is the *angular temporal frequency* (units of radians per second), usually called the angular frequency. Some authors use σ rather than ω .
- $\nu \triangleq 1/\lambda$ is the *spatial frequency* or *wavenumber* (cycles per meter). Some authors use κ , κ' , \varkappa , $\tilde{\nu}$, or σ , rather than ν .
- $k \triangleq 2\pi/\lambda = 2\pi\nu$ is the *angular spatial frequency* or *angular wavenumber* (radians per meter). Some authors use κ rather than k , and some call ν the wavenumber and k the spatial frequency.

Temporal frequency f is expressed in units of s^{-1} or Hertz; one Hertz is one cycle per second. Angular frequency ω is expressed in radians per second. One cycle of a wave is 2π radians, so $1 \text{ Hz} = 2\pi \text{ rad s}^{-1}$. Spatial frequency is in cycles per meter, or m^{-1} . However, in image analysis, the spatial frequency is often expressed in lines per millimeter (or something similar) as a measure of the spatial frequency that can be resolved in an image.

There is no uniformity in the literature. The books in my library have examples of each of the above notations. When a paper refers to “frequency,” it usually means temporal

frequency, but it is often left to the reader to figure out from the context if f or ω is meant; and “wavenumber” or “spatial frequency” may mean either ν or k . Older physics books seem to use ν for temporal frequency and write the energy of a photon (Eq. 1.1) as $E = h\nu$; newer texts tend to use f for temporal frequency⁸ and write $E = hf$ or $E = \hbar\omega$, where $\hbar = h/(2\pi)$ is the “reduced” Planck constant. The whole business of terminology and notation for frequencies is a confusing mess, and you just have to figure it out on a case-by-case basis.

Suppose we wish to express the spectral radiance, defined above as power per unit area per unit solid angle per unit wavelength interval (units of $\text{W m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$), as radiance per unit spatial frequency, i.e., power per unit area per unit solid angle per unit spatial frequency interval. To establish the conversion, consider the radiant energy $Q(\lambda)d\lambda$ contained in a wavelength interval $d\lambda$. The same amount of energy is contained in a corresponding spatial frequency interval $Q(\nu)d\nu$. Since an increase in wavelength ($d\lambda > 0$) implies a decrease in spatial frequency ($d\nu < 0$), and vice versa, we can write

$$Q(\lambda)d\lambda = -Q(\nu)d\nu.$$

Using $\lambda = 1/\nu$ we then get

$$Q(\nu) = -Q(\lambda)\frac{d\lambda}{d\nu} = \frac{1}{\nu^2}Q(\lambda = 1/\nu) = \lambda^2Q(\lambda),$$

which is the desired connection between $Q(\lambda)$ and $Q(\nu)$, or between any other radiometric quantities. A wavelength interval $d\lambda$ therefore corresponds to a spatial frequency interval $d\nu$ of size

$$d\nu = \frac{1}{\lambda^2}d\lambda.$$

If $Q(\nu)$ is radiance, the units are now $\text{W m}^{-2} \text{sr}^{-1} (1/\text{nm})^{-1}$ if λ is measured in nanometers.

To convert radiance per unit wavelength to radiance per unit temporal angular frequency, the same process, starting with $\lambda = c/f = 2\pi c/\omega$, where c is the speed of light, gives

$$Q(\omega) = -Q(\lambda)\frac{d\lambda}{d\omega} = \frac{2\pi c}{\omega^2}Q(\lambda = 2\pi c/\omega) = \frac{\lambda^2}{2\pi c}Q(\lambda).$$

If $Q(\omega)$ is radiance, the units are now $\text{W m}^{-2} \text{sr}^{-1} (\text{rad/s})^{-1}$.

1.6 Polarization: Stokes Vectors

The preceding sections have discussed unpolarized or “natural” light. The prime example of interest in oceanography is unscattered light from the Sun. However, unpolarized sunlight scattered by atmospheric molecules is partially polarized, with the degree of polarization depending on the direction of the scattering. Similarly, reflection by and transmission through the air-water surface generates polarized light even if the incident light is unpolarized. Therefore, light in the atmosphere and oceans is usually partially polarized. Although the state of polarization may be irrelevant for processes such as photosynthesis for heating of the water, the state of polarization carries information and is increasingly being used in

⁸The use of ν versus f in older versus newer texts may be an erroneous conclusion due to the high uncertainty in small-sample statistics; I have only a few dozen physics books in my office.

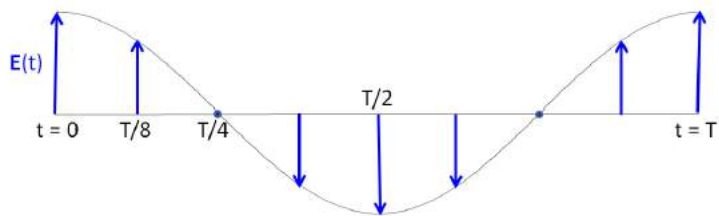


Figure 1.16: Illustration of the behavior of the electric field $\mathbf{E}(t) = \mathbf{E}_0 \cos(2\pi t/T)$ for linearly polarized light. See the Web Book for an animation.

remote sensing and other applications. It is therefore worthwhile to discuss polarization before proceeding with other topics.

Light consists of propagating electric and magnetic fields, which are described by Maxwell's Equations (Chapter 11). If the time- and space-dependent electric field vector $\mathbf{E}(\mathbf{x}, t)$ is known, then the magnetic field vector $\mathbf{B}(\mathbf{x}, t)$ can be computed from Maxwell's equations, and vice versa. It is thus sufficient to discuss just one of these fields, which is customarily chosen to be the electric field vector. Polarization then refers to the plane in which the electric field vector is oscillating.

1.6.1 What is Polarization?

Suppose you are looking toward a light source, or “into the beam.” In the simplest case, called linear (or plane) polarization, the electric field $\mathbf{E}(\mathbf{x}, t)$ lies in, or oscillates in, a fixed plane. Figure 1.16 illustrates how the electric field varies with time as the light wave passes through a fixed reference plane normal to the direction of propagation. See the Web Book for an animation. For visible wavelengths, these oscillations are at a frequency of around 10^{14} times per second and cannot be directly measured because of instrumentation limits.

However, the plane in which the electric field lies may also rotate with time as the beam of light passes. This is called circular polarization if the maximum value of $\mathbf{E}(\mathbf{x}, t)$ is independent of time but the orientation of the plane rotates. There is an intermediate state, elliptical polarization, in which the plane rotates and the amplitude of $\mathbf{E}(\mathbf{x}, t)$ also changes as the plane rotates, so that the maximum value of $\mathbf{E}(\mathbf{x}, t)$ traces out an ellipse. The final possibility is that, as you observe $\mathbf{E}(\mathbf{x}, t)$, the plane of oscillation changes rapidly (on the order of the frequency of the light) and randomly. This is random polarization, which is often called “unpolarized” or “natural” light.

Circular or elliptical polarization is called “right” or “left” depending on the direction the plane of the electric field rotates as the wave passes a fixed reference plane, but understanding the exact meaning these terms as they relate to light can be confusing. Figure 1.17 makes an analogy to a common “right-hand-twist” drill bit used to cut holes in wood. The edges of this bit form right-handed helices as commonly defined. As this bit bores into the wood, the cutting tips trace out right-handed helices. Now think of the outer edge of the bit as being the tip of the electric field vector of circularly polarized light. The light is propagating upward in the figure, in the direction the drill goes into the wood. The reddish plane labeled $z = 0$ is held fixed in space as the light propagates past this reference plane. The red arrow labeled $t = 0$ indicates the electric field vector lying in the reference

plane at time zero. Now think of moving the drill bit upward *without rotating it*. At some time $t = \Delta t > 0$ later the drill bit/light wave will have moved upward, and the red arrow labeled $t = \Delta t$ will lie in the reference plane. The $t = \Delta t$ vector will have rotated from the direction of the $t = 0$ vector. At time $t = 2\Delta t$ the light wave will have moved further upward, and the $t = 2\Delta t$ vector will now be crossing the reference plane at $z = 0$. As time progresses, the *direction of the electric field vectors in the reference plane will appear to rotate in a clockwise direction as viewed looking into the beam* (or counterclockwise when looking along the beam). This is the electric field rotation direction for right-circularly polarized (RCP) light. If the electric field rotates in a counterclockwise direction when looking into the beam (or clockwise looking along the beam), the light is left-circularly polarized (LCP).

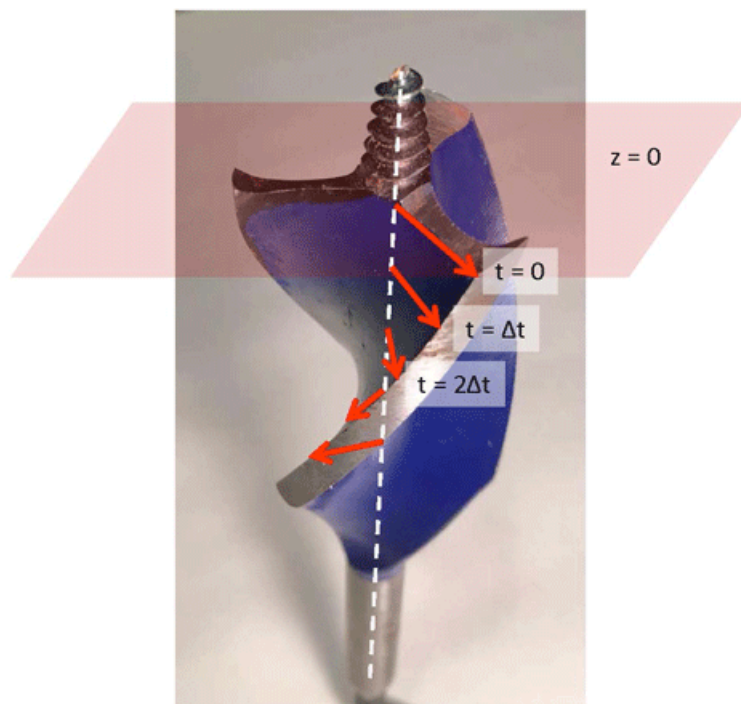


Figure 1.17: A right-hand-twist drill bit as an analogy to right-circular polarization. The edges of the bit are right-handed helices, and as the bit cuts into material, each cutting tip traces out a right-handed helix. If the bit is moved upward, without rotation, through the plane at $z = 0$, the intersection of the cutting edge and the plane at $z = 0$ appears to rotate clockwise when viewed looking into the beam.

Note that the description of the electric field as rotating clockwise or counterclockwise depends on whether you are looking into the beam or along the beam. However, the concept of right-handed versus left-handed helices is independent of the viewing perspective. The definition just described—RCP corresponds to clockwise electric field rotation in a reference plane when looking into the beam as the light propagates through the plane, and to the pattern of electric field vectors lying along a right-hand helix in space—is what is used by

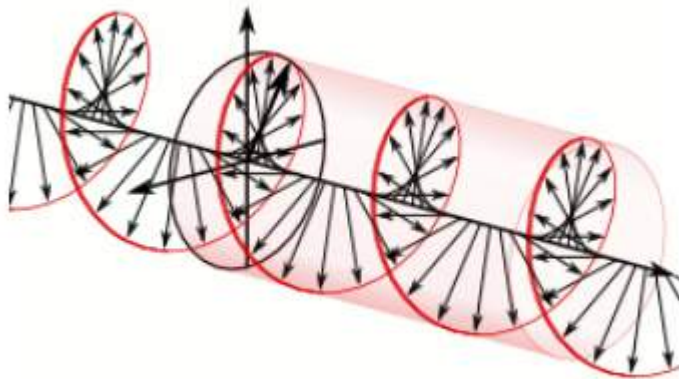


Figure 1.18: A single frame from an animation of a RCP electric field propagating from left to right. See the Web Book for the animation. (From https://en.wikipedia.org/wiki/Circular_polarization)

Bohren and Huffman (1983) and Hecht (1989). I personally like that convention because I can remember the analogy with the moving right-hand-helix drill bit. However, others, e.g., Kattawar (1994) and Jackson (1962), use the opposite convention of RCP meaning that the electric field appears to rotate counterclockwise with looking into the beam (or clockwise looking along the beam); in this choice LCP corresponds to a right-handed helix. The convention for how to define RCP vs LCP often seems to depend on the field of the user—physics versus astronomy versus chemistry, etc. Fortunately it does not matter which one you use, so long as you make a choice and stick with it during the solution of your problem. Problems arise only if you compare your results with someone else’s results, in which case different descriptors like parallel versus perpendicular and right versus left may be referring to the same thing by different names.

Figure 1.18 shows a single frame of an animation of a propagating electric field of RCP light as defined here, at one instant of time, but the animator calls it LCP, which was the choice for the creator of that Wikipedia page.

1.6.2 Stokes Vectors

We now need a quantitative way to specify the state of polarization of light. This is given by the *Stokes vector*, which is an array of four real numbers usually written as

$$\mathbb{S} = \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix}.$$

Note that \mathbb{S} is just an array with four elements; it is not a vector in the geometric sense.

To define the Stokes vector, first pick an $(\hat{x}, \hat{y}, \hat{z})$ coordinate system that is convenient for your problem. In a laboratory setting, this system might have \hat{x} parallel to an optical

bench top, $\hat{\mathbf{y}}$ perpendicular to the bench top, and $\hat{\mathbf{z}} = \hat{\mathbf{x}} \times \hat{\mathbf{y}}$ in the direction of propagation. In this lab setting, $\hat{\mathbf{x}}$ might then be called the “parallel” (to the bench top) direction, and $\hat{\mathbf{y}}$ would then be the “perpendicular” direction. Or $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ might be called “horizontal” and “vertical”, respectively. The next section on scattering of polarized light shows another coordinate system commonly used in oceanography.

The electric field vector in this coordinate system is resolved into x and y components as $\mathbf{E} = E_x \hat{\mathbf{x}} + E_y \hat{\mathbf{y}}$, where the components E_x and E_y depend on position and time. For light propagating in a vacuum, the electric field is transverse to the direction of travel, so the z component of \mathbf{E} is zero. If the light is linearly polarized in the x plane, then $E_x \neq 0$ and $E_y = 0$. For linear polarization in the y plane, $E_x = 0$ and $E_y \neq 0$.

At optical frequencies instruments cannot measure the instantaneous value of the fluctuating electric field $\mathbf{E}(t)$ itself, but they can make time-averaged (over many wave periods) measurements of the corresponding irradiance \mathcal{E} . Here \mathcal{E} is used for irradiance to avoid confusion with the electric field magnitude E . As discussed in detail in Section 11.3, it is convenient for intermediate calculations to write sinusoidal functions as complex exponentials; the real part is then taken to recover the sinusoid. The time-averaged irradiance corresponding to $\mathbf{E}(z, t) = \cos(kz - \omega t) = \Re\{\mathbf{E}_o \exp(ikz - i\omega t)\}$ is (Eq. 11.42)

$$\mathcal{E} = \frac{1}{2} \sqrt{\epsilon_m / \mu_m} |\mathbf{E}_o|^2. \quad (1.30)$$

Here ϵ_m is the electrical permittivity of the medium, which has units of Farad m^{-1} or $\text{C}^2 \text{N}^{-1} \text{m}^{-2}$ or $\text{A}^2 \text{s}^4 \text{kg}^{-1} \text{m}^{-3}$. μ_m is the magnetic permeability of the medium, which has units of Henry m^{-1} or N A^{-2} or $\text{kg m s}^{-2} \text{A}^{-2}$. Electric fields have units of N C^{-1} or V m^{-1} or $\text{kg m s}^{-2} \text{C}^{-1}$. Thus \mathcal{E} has units of kg s^{-3} or Watt m^{-2} , i.e. of irradiance. The factor of $\frac{1}{2}$ comes from the average of the sinusoidal dependence of $|\mathbf{E}(t)|^2$ over a wave period (i.e., $\frac{1}{2\pi} \int_0^{2\pi} \cos^2 x dx = \frac{1}{2}$). We therefore base the definition and actual measurements of Stokes vectors on the measurable time-averaged irradiances if we are working with a collimated monochromatic beam of light.

The Q Stokes parameter is then defined follows. Let \mathcal{E}_x be the time averaged irradiance measured with a linear polarizing filter placed in the beam and oriented in the x (or parallel or horizontal in the lab setting) direction. Let \mathcal{E}_y be the time average measured with the linear polarizer oriented in the y (or perpendicular or vertical) direction. Then Q is defined as

$$Q \triangleq \mathcal{E}_x - \mathcal{E}_y.$$

Thus $Q > 0$ if the polarization lies in the x plane, and $Q < 0$ if it lies in the y plane.

This choice of $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ can distinguish between linear polarization lying in the x or y planes. But suppose that the plane of polarization is intermediate between the x or y planes, as illustrated by the either of the red arrows in the left panel of Fig. 1.19. These are different states of polarization, but both have the same projections onto the x and y planes, hence the same Q value. Thus the Q parameter cannot distinguish between the solid and dashed planes of polarization seen in the left panel of the figure.

The state of linear polarization can be uniquely specified by the choice of a second set of axes, $(\hat{\mathbf{x}}', \hat{\mathbf{y}}')$, chosen at a 45 deg angle to the $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ axes, as shown in the right panel of Fig. 1.19. The solid and dashed red arrows have different projections on the $(\hat{\mathbf{x}}', \hat{\mathbf{y}}')$ axes and are thus distinguished. The Stokes U parameter is non-zero for planes of polarization

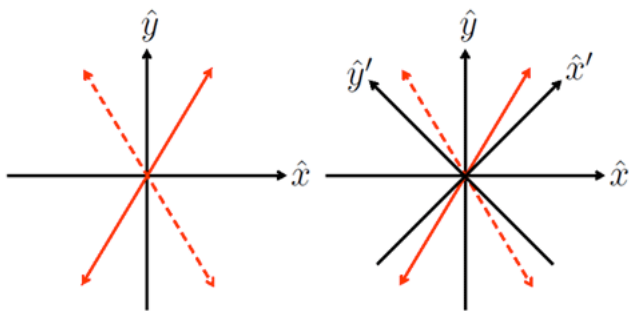


Figure 1.19: Coordinate systems for specification of the state of linear polarization. The red arrows represent two possible planes of linear polarization.

like the red arrows in the figures and is defined by

$$U \triangleq \mathcal{E}_{x'} - \mathcal{E}_{y'}$$

where $\mathcal{E}_{x'}$ and $\mathcal{E}_{y'}$ are the time averages of the irradiances measured with the linear polarizer oriented in the x' and y' planes. Thus if the plane of polarization lies at 45 deg to the x plane, parallel to $\hat{\mathbf{x}}'$, $U > 0$ and $Q = 0$. For polarization in the plane at -45 deg to the x plane, parallel to $\hat{\mathbf{y}}'$, $U < 0$ and $Q = 0$. For planes of linear polarization not lying in either the x, y or x', y' planes (as illustrated by the red arrows in Fig. 1.19), both Q and U will be non-zero and either positive or negative, depending on the inclination of the polarization plane to these two sets of axes.

The Stokes parameters Q and U together specify the state of polarization if the light is linearly polarized. Another parameter, V , is needed to specify the state of circular polarization. The time-averaged amounts of right and left circularly polarized irradiance, call them \mathcal{E}_R and \mathcal{E}_L respectively, can be measured by use of circular polarizers. The V component is then defined as their difference:

$$V \triangleq \mathcal{E}_R - \mathcal{E}_L.$$

Thus $V > 0$ for right circular polarization, and $V < 0$ for left circular polarization.

Finally, consider the case of randomly polarized light. All of the above time averages will be equal because of the rapid fluctuations of the electric fields with all directions and helicities, in which case $Q = U = V = 0$. To account for this case, let I be the total irradiance without regard for the state of polarization. This is measured without the use of any polarization filters in the beam. This is also given in terms of the time averages by

$$I = \mathcal{E}_x + \mathcal{E}_y = \mathcal{E}_{x'} + \mathcal{E}_{y'} = \mathcal{E}_R + \mathcal{E}_L.$$

I is always positive and equal to the total irradiance.

It is important to note that the values of the Q and U parameters depend on the choice of the $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ axes, but the values of I and V are independent of this choice. In the coordinate system described above, the $\hat{\mathbf{x}}$ axis is parallel to the horizontal laboratory bench top, and $\hat{\mathbf{y}}$ is perpendicular to the bench top. As noted, it is common to refer to the corresponding polarizations as being “parallel” or “horizontal” and “perpendicular” or “vertical”, respectively. Terms like parallel and perpendicular or horizontal and vertical always refer to some reference plane—the bench top in this case. However, a different choice of the reference plane changes the meaning of these terms. One person’s parallel polarization can be another person’s perpendicular polarization. You have to figure out

the meanings on a case by case basis for whatever reference coordinate system is being used.

If the beam is perfectly polarized (in whatever state of polarization), then

$$I^2 = Q^2 + U^2 + V^2.$$

If the beam is unpolarized, or is a mixture of polarized and unpolarized light, then this relation becomes an inequality:

$$I^2 > Q^2 + U^2 + V^2.$$

The *degree of polarization*, expressed in percent, is defined by

$$DoP = 100 \frac{\sqrt{Q^2 + U^2 + V^2}}{I}.$$

The *degree of linear polarization* is defined by

$$DoLP = 100 \frac{\sqrt{Q^2 + U^2}}{I},$$

and the *degree of circular polarization* is defined by

$$DoCP = 100 \frac{V}{I}.$$

DoCP is positive for RCP and negative for LCP. These measures of the degree of polarization do not depend on the choice of coordinate system.

The definitions of $Q, U,$ and V above were made in terms of measurable irradiances. There is much more that can be said, in particular about the theoretical formulation of the Stokes vector in terms of the solution of Maxwell's equations for a propagating wave. An excellent and entertaining presentation of those details is given in [Bohren and Clothiaux \(2006, Chapter 7\)](#). Suffice it to say that in discussions of Stokes vectors you may see them defined by equations such as

$$\mathbb{S} = \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = \begin{bmatrix} \mathcal{E}_{\parallel} + \mathcal{E}_{\perp} \\ \mathcal{E}_{\parallel} - \mathcal{E}_{\perp} \\ \mathcal{E}_{+45} - \mathcal{E}_{-45} \\ \mathcal{E}_{\text{R}} - \mathcal{E}_{\text{L}} \end{bmatrix} \quad (1.31)$$

$$= \sqrt{\frac{\epsilon_m}{\mu_m}} \begin{bmatrix} \langle E_{\parallel}(t)E_{\parallel}^*(t) + E_{\perp}(t)E_{\perp}^*(t) \rangle \\ \langle E_{\parallel}(t)E_{\parallel}^*(t) - E_{\perp}(t)E_{\perp}^*(t) \rangle \\ \langle E_{\parallel}(t)E_{\perp}^*(t) + E_{\perp}(t)E_{\parallel}^*(t) \rangle \\ i\langle [E_{\parallel}(t)E_{\perp}^*(t) - E_{\perp}(t)E_{\parallel}^*(t)] \rangle \end{bmatrix} \quad (1.32)$$

$$= \frac{1}{2} \sqrt{\frac{\epsilon_m}{\mu_m}} \begin{bmatrix} E_{\text{o}\parallel}E_{\text{o}\parallel}^* + E_{\text{o}\perp}E_{\text{o}\perp}^* \\ E_{\text{o}\parallel}E_{\text{o}\parallel}^* - E_{\text{o}\perp}E_{\text{o}\perp}^* \\ E_{\text{o}\parallel}E_{\text{o}\perp}^* + E_{\text{o}\perp}E_{\text{o}\parallel}^* \\ i[E_{\text{o}\parallel}E_{\text{o}\perp}^* - E_{\text{o}\perp}E_{\text{o}\parallel}^*] \end{bmatrix} = \frac{1}{2} \sqrt{\frac{\epsilon_m}{\mu_m}} \begin{bmatrix} |E_{\text{o}\parallel}|^2 + |E_{\text{o}\perp}|^2 \\ |E_{\text{o}\parallel}|^2 - |E_{\text{o}\perp}|^2 \\ -2\Re\{E_{\text{o}\parallel}E_{\text{o}\perp}^*\} \\ 2\Im\{E_{\text{o}\perp}E_{\text{o}\parallel}^*\} \end{bmatrix} \quad (1.33)$$

$$= \frac{1}{2} \sqrt{\frac{\epsilon}{\mu}} \begin{bmatrix} E_{o\theta} E_{o\theta}^* + E_{o\phi} E_{o\phi}^* \\ E_{o\theta} E_{o\theta}^* - E_{o\phi} E_{o\phi}^* \\ -E_{o\theta} E_{o\phi}^* - E_{o\phi} E_{o\theta}^* \\ i(E_{o\phi} E_{o\theta}^* - E_{o\theta} E_{o\phi}^*) \end{bmatrix} = \frac{1}{2} \sqrt{\frac{\epsilon}{\mu}} \begin{bmatrix} |E_{o\theta}|^2 + |E_{o\phi}|^2 \\ |E_{o\theta}|^2 - |E_{o\phi}|^2 \\ -2\text{Re}\{E_{o\theta} E_{o\phi}^*\} \\ 2\text{Im}\{E_{o\theta} E_{o\phi}^*\} \end{bmatrix}. \quad (1.34)$$

The first form of definition is in terms of measurable irradiances as discussed above, with an obvious change in notation to show the orientations of the polarizing filters in the chosen coordinate system. The second form is written in terms of the complex, time-dependent, electric field vectors after describing the light beam in terms of a plane-wave solution to Maxwell's equations. Thus $E_{\parallel}(t) = E_{o\parallel} \exp(-i\omega t)$, etc. The $\langle \dots \rangle$ notation indicates the time average of the argument. After the time averages are taken, the magnitudes $E_{o\parallel}$ etc. are left, and there is an additional factor of $\frac{1}{2}$ resulting from the average of products of the sinusoidal electric fields over a wave period. The version in terms of $E_{o\theta}$ and $E_{o\phi}$ refer to the electric field components in the $\hat{\theta}$ and $\hat{\phi}$ directions seen in Fig. 1.20. The forms showing absolute values and real and imaginary parts make clear that the Stokes parameters are real numbers. The discussion of Eq. (1.30) shows that the definitions in terms of electric fields still have units of irradiance. The irradiance form is what you will use in the lab; the electric-field forms are what you will use for theory.

Table 1.5 shows the pattern of Stokes parameters for various states of polarization.

general	unpolarized	parallel	perpendicular
$\begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}$
+45	-45	RCP	LCP
$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$	$\begin{bmatrix} 1 \\ 0 \\ 0 \\ -1 \end{bmatrix}$

Table 1.5: Patterns of Stokes vectors for various polarization states.

1.7 Polarization: Scattering Geometry

This section discusses the geometrical considerations underlying computations of the scattering of polarized light using Stokes vectors.

As is shown in the preceding section, the state of polarization of a light field is specified by the four-component Stokes vector, whose elements are related to the complex amplitudes of the electric field vector E resolved into directions that are parallel (E_{\parallel}) and perpendicular (E_{\perp}) to a conveniently chosen reference plane.

However, there are two versions of the Stokes vector seen in the literature, and these two versions have different units and refer to different physical quantities. The *coherent* Stokes vector describes a quasi-monochromatic plane wave propagating in one exact direction, and the vector components have units of power per unit area (i.e., irradiance) on a surface perpendicular to the direction of propagation. To be specific, the coherent Stokes vector can be defined as shown in Eq. (1.33) (e.g., Mishchenko et al., 2002):

$$\mathbb{S} = \begin{bmatrix} I \\ Q \\ U \\ V \end{bmatrix} = \frac{1}{2} \sqrt{\frac{\epsilon_m}{\mu_m}} \begin{bmatrix} E_{o\parallel} E_{o\parallel}^* + E_{o\perp} E_{o\perp}^* \\ E_{o\parallel} E_{o\parallel}^* - E_{o\perp} E_{o\perp}^* \\ E_{o\parallel} E_{o\perp}^* + E_{o\perp} E_{o\parallel}^* \\ i[E_{o\parallel} E_{o\perp}^* - E_{o\perp} E_{o\parallel}^*] \end{bmatrix}. \quad (1.35)$$

Here E^* denotes complex conjugate, hence the components of the Stokes vector are real numbers.

The *diffuse* Stokes vector is defined as in Eq. (1.35) but describes light propagating in a small set of directions surrounding a particular direction and has units of power per unit area per unit solid angle (i.e., radiance). It is the diffuse Stokes vector that appears in the general Vector Radiative Transfer Equation discussed in Section 9.14. The differences in coherent and diffuse Stokes vectors are rigorously presented in Mishchenko (2008b).

Authors often omit the $\frac{1}{2}\sqrt{\epsilon_m/\mu_m}$ factor seen in Eq. (1.35) because they are interested only in relative values such as the degree of polarization, not in absolute magnitudes, but this omission is both confusing and physically incorrect. *Units and magnitudes matter!* The different units of coherent and diffuse Stokes vectors, and whether or not the $\frac{1}{2}\sqrt{\epsilon/\mu_m}$ factor is included in the definition of the Stokes vector, have subtle but very important consequences in how light propagation across a dielectric interface such as the air-water surface is formulated. The paper by Zhai et al. (2012) gives a definitive discussion of these matters.

1.7.1 Coordinate Systems

To describe the scattering of polarized light, we must first choose coordinate systems and show in detail how to resolve Stokes vectors in these coordinate systems as needed for scattering calculations.

Figure 1.20 shows the coordinate systems commonly used to resolve Stokes vectors as needed for geophysical scattering calculations. In oceanography, depth and direction are defined in a 3-D Cartesian coordinate system with depth measured positive downward from 0 at the mean sea surface. Polar angle θ is defined from 0 in the $+\hat{z}$ (or downwelling direction) to π in the $-\hat{z}$ (upwelling) direction. $+\hat{x}$ is chosen for convenience, e.g., pointing toward the Sun or pointing in the downwind direction (in which case $\pm\hat{y}$ is the cross-wind direction. As shown in Mobley (2014), using a wind-centered coordinate system makes it easier to model a random sea surface with different along-wind and cross-wind slope statistics.) Azimuthal angle ϕ is measured counterclockwise from $+\hat{x}$ when looking in the

$-\hat{\mathbf{z}}$ direction. If the Sun is placed in the $+\hat{\mathbf{x}}$ azimuthal direction at $\phi_{\text{Sun}} = 0$, unscattered rays from the Sun are then traveling in the $-\hat{\mathbf{x}}$ direction at $\phi = 180$ deg.

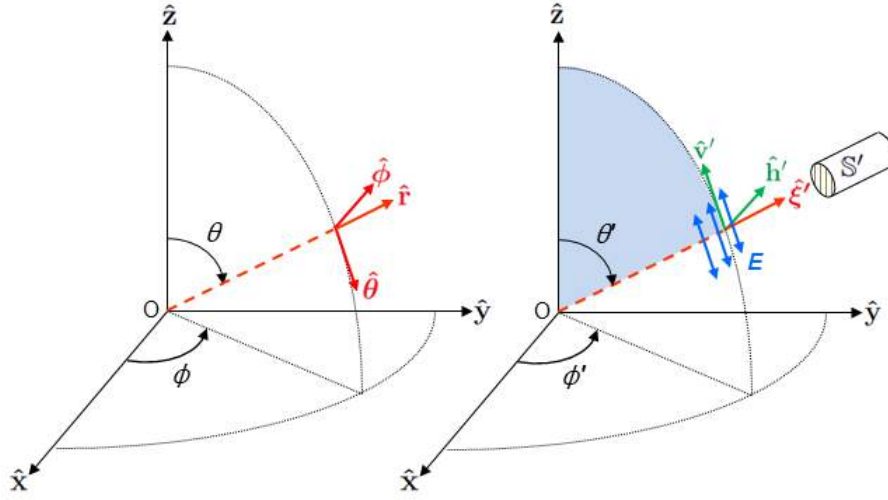


Figure 1.20: Coordinate system for Stokes vectors. The left panel shows an $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ Cartesian system, which is fixed in space and used to define a spherical (r, θ, ϕ) coordinate system. The right panel shows a direction of light propagation $\hat{\boldsymbol{\xi}}'$ whose Stokes vector is S' . The $\hat{\mathbf{h}}'-\hat{\mathbf{v}}'-\hat{\boldsymbol{\xi}}'$ system is a local system (changing with θ', ϕ') for resolving Stokes vectors in directions that are perpendicular and parallel to the incident meridian plane, part of which is shaded in blue. In the oceanographic setting, this figure can be turned “upside down”, with the x - y plane being the mean sea surface, $\hat{\mathbf{x}}$ chosen for convenience (e.g. pointing downwind or east or in the Sun’s direction), and $\hat{\mathbf{z}}$ pointing downward into the water. The blue arrows indicate light that is polarized parallel to the meridian plane.

The unit vectors in the directions of increasing (r, θ, ϕ) are

$$\hat{\mathbf{r}} = \sin \theta \cos \phi \hat{\mathbf{x}} + \sin \theta \sin \phi \hat{\mathbf{y}} + \cos \theta \hat{\mathbf{z}}, \quad (1.36a)$$

$$\hat{\boldsymbol{\theta}} = \cos \theta \cos \phi \hat{\mathbf{x}} + \cos \theta \sin \phi \hat{\mathbf{y}} - \sin \theta \hat{\mathbf{z}}, \quad (1.36b)$$

$$\hat{\boldsymbol{\phi}} = -\sin \phi \hat{\mathbf{x}} + \cos \phi \hat{\mathbf{y}}. \quad (1.36c)$$

Let $\hat{\boldsymbol{\xi}}$ denote a unit vector pointing in the direction of light propagation, as given by angles (θ, ϕ) . (In Fig. 1.20, $\hat{\boldsymbol{\xi}}' = \hat{\mathbf{r}}$.) The components of $\hat{\boldsymbol{\xi}}$ are given by

$$\hat{\boldsymbol{\xi}} = \xi_x \hat{\mathbf{x}} + \xi_y \hat{\mathbf{y}} + \xi_z \hat{\mathbf{z}} = (\xi_x, \xi_y, \xi_z) \quad (1.37)$$

where

$$\xi_x = \sin \theta \cos \phi, \quad (1.38)$$

$$\xi_y = \sin \theta \sin \phi, \quad (1.39)$$

$$\xi_z = \cos \theta. \quad (1.40)$$

There is frequent need to resolve Stokes vectors in directions perpendicular and parallel to a given plane, which can be a meridian⁹ plane, a scattering plane in the water volume,

⁹The name “meridian” comes from the Latin for “mid-day.” The Sun crosses the astronomical meridian

or the plane of incident, reflected, and transmitted light for light incident onto an air-water or bottom surface. The following conventions define these unit vectors. Vector $\hat{\mathbf{p}}$ denotes a unit vector parallel to a plane, and $\hat{\mathbf{s}}$ denotes a unit vector perpendicular to a plane (“s” for senkrecht, German for perpendicular). The perpendicular vector $\hat{\mathbf{s}}$ is chosen to be in the direction given by the vector cross product of the incident direction crossed with the final direction. The parallel vector $\hat{\mathbf{p}}$ is then defined as the direction of propagation cross the perpendicular direction. Thus the perpendicular cross parallel directions give the direction of light propagation: $\hat{\mathbf{s}} \times \hat{\mathbf{p}} = \hat{\boldsymbol{\xi}}$, where \times denotes the vector cross product.

For an incident direction $\hat{\boldsymbol{\xi}}'$ and the associated Stokes vector \mathbf{S}' specified in the incident meridian plane, the first vector is taken to be $\hat{\mathbf{z}}$ and the second is the direction of propagation. Thus $\hat{\mathbf{h}}' \triangleq \hat{\mathbf{z}} \times \hat{\boldsymbol{\xi}}' / |\hat{\mathbf{z}} \times \hat{\boldsymbol{\xi}}'|$ as seen in the right panel of Fig. 1.20. In this case $|\hat{\mathbf{z}} \times \hat{\boldsymbol{\xi}}'| = \sin \theta'$, and $\hat{\mathbf{h}}' = \hat{\boldsymbol{\varphi}}'$. Similarly, $\hat{\mathbf{v}}' \triangleq \hat{\boldsymbol{\xi}}' \times \hat{\mathbf{h}}' = -\hat{\boldsymbol{\theta}}'$, and $\hat{\mathbf{h}}' \times \hat{\mathbf{v}}' = \hat{\boldsymbol{\xi}}'$. Meridian planes are perpendicular to the mean sea surface. The $\hat{\mathbf{h}}'$ vector as just defined is therefore parallel to the mean sea surface and therefore is often referred to as the “horizontal” direction; $\hat{\mathbf{v}}'$ lies in a vertical plane and is correspondingly called the “vertical” direction. For a final direction $\hat{\boldsymbol{\xi}}$ and its Stokes vector \mathbf{S} in the final meridian plane, the first vector is the direction $\hat{\boldsymbol{\xi}}$ and the second is $\hat{\mathbf{z}}$. This vector cross product algorithm for specifying perpendicular and parallel directions will be convenient for sea surface reflectance and transmission calculations in which light can propagate from one tilted wave facet to another without reference to meridian planes, except for the incident and final directions when a photon enters or leaves the region of the sea surface.

The Q and U components of a Stokes vector describe linear polarization with the plane of polarization specified relative to a particular coordinate system. The I component is the total radiance, and V describes circular polarization; these quantities do not depend on the coordinate system and are invariant under a rotation of the coordinate system. The blue arrows in Fig. 1.20 represent the plane of oscillation of the electric field vector E parallel to the meridian plane, i.e. for vertical plane polarization.

1.7.2 Scattering Geometry

The elements of input and output Stokes vectors are defined relative to meridian planes, as described above. However, scattering from an incident direction $\hat{\boldsymbol{\xi}}'$ to a final direction $\hat{\boldsymbol{\xi}}$ is defined in terms of the included scattering angle ψ and the *scattering plane*, as illustrated in Fig. 1.21. Using Eqs. (1.37)-(1.40) to express the incident direction $\hat{\boldsymbol{\xi}}'$ and scattered direction $\hat{\boldsymbol{\xi}}$ in terms of the incident and scattered polar and azimuthal angles gives the scattering angle ψ :

$$\cos \psi = \hat{\boldsymbol{\xi}}' \cdot \hat{\boldsymbol{\xi}} = \cos \theta' \cos \theta + \sin \theta' \sin \theta \cos(\phi - \phi'). \quad (1.41)$$

Consider an incident beam of light propagating in direction $\hat{\boldsymbol{\xi}}' = \hat{\mathbf{r}}$ as in Fig. 1.20. Direction $\hat{\boldsymbol{\xi}}'$ is specified by polar and azimuthal directions (θ', ϕ') . The $\hat{\mathbf{z}}$ axis and the direction of light propagation $\hat{\boldsymbol{\xi}}'$ define the *incident meridian plane*, part of which is shaded in blue in the right panel of Fig. 1.20. The 4×1 (diffuse) Stokes vector $\mathbf{S}' = [I', Q', U', V']^T$ for this beam of light is described with reference to “horizontal” and “vertical” directions,

plane and is highest in the sky half way between sunrise and sunset. This is the origin of A.M. (ante-meridiem, before mid-day) and P.M. (post-meridiem, after mid-day) used in specifying time with a 12-hour clock system.

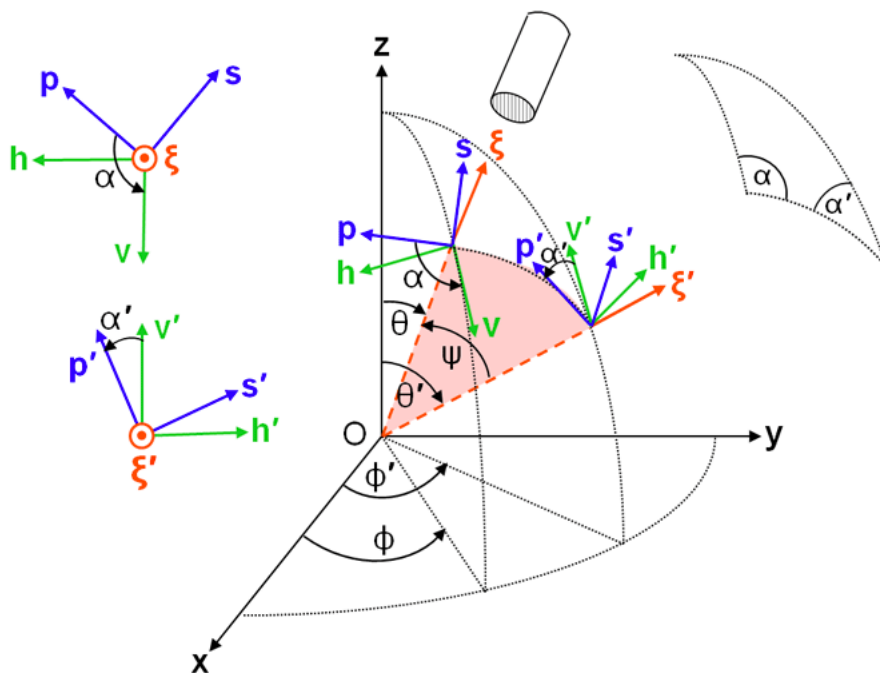


Figure 1.21: The scattering plane and the incident and final meridian planes showing the coordinate systems and rotations used to specify the scattering of polarized light. The scattering plane is partly shaded in red. (Unit vectors in this figure do not have hats.)

$+\hat{\mathbf{h}}'$ and $+\hat{\mathbf{v}}'$ respectively, which were defined above; the superscript T denotes transpose. Note that the horizontal unit vector $+\hat{\mathbf{h}}'$ is perpendicular to the meridian plane, and the vertical vector $+\hat{\mathbf{v}}'$ is parallel to the meridian plane.

To compute how an incident Stokes vector \mathbb{S}' is scattered to a final vector \mathbb{S} , the horizontal and vertical components of \mathbb{S}' in the incident meridian plane must first be transformed (“rotated”) into components parallel and perpendicular to the scattering plane. The coordinate system after rotation of $\hat{\mathbf{v}}'$ and $\hat{\mathbf{h}}'$ about the $\hat{\boldsymbol{\xi}}'$ axis is labeled $\hat{\mathbf{p}}'$ (parallel to the scattering plane) and $\hat{\mathbf{s}}'$ (perpendicular to the scattering plane). Note that $\hat{\mathbf{s}}' \times \hat{\mathbf{p}}'$ still gives the direction of propagation $\hat{\boldsymbol{\xi}}'$. As shown in Fig. 1.21, rotation angle α' takes $\hat{\mathbf{v}}'$ into $\hat{\mathbf{p}}'$ (and $\hat{\mathbf{h}}'$ into $\hat{\mathbf{s}}'$). In the present discussion, *rotation angles are defined as positive for counterclockwise rotations when looking “into the beam,”* e.g. in the $-\hat{\boldsymbol{\xi}}'$ direction. This is similar to rotations about the $\hat{\mathbf{z}}$ axis of Fig. 1.20 having positive angles ϕ for counterclockwise rotations when looking in the $-\hat{\mathbf{z}}$ direction.

When computing single scattering with both $\hat{\boldsymbol{\xi}}'$ and $\hat{\boldsymbol{\xi}}$ being expressed in their respective meridian planes, the rotation angles can be obtained from spherical trigonometry applied to the triangle defined by $\hat{\mathbf{z}}$, $\hat{\boldsymbol{\xi}}'$, and $\hat{\boldsymbol{\xi}}$, which is shown in the inset in Fig. 1.21. Given θ' , ϕ' , θ , ϕ , spherical trigonometry gives the rotation angles α' and α as (e.g. [van de Hulst \(1980, Vol. 2, page 499\)](#) or [Mishchenko et al. \(2002, page 90\)](#))

$$\cos \alpha' = (\cos \theta - \cos \theta' \cos \psi) / (\sin \psi \sin \theta') \quad (1.42)$$

or

$$\sin \alpha' = -\sin \theta \sin(\phi - \phi') / \sin \psi. \quad (1.43)$$

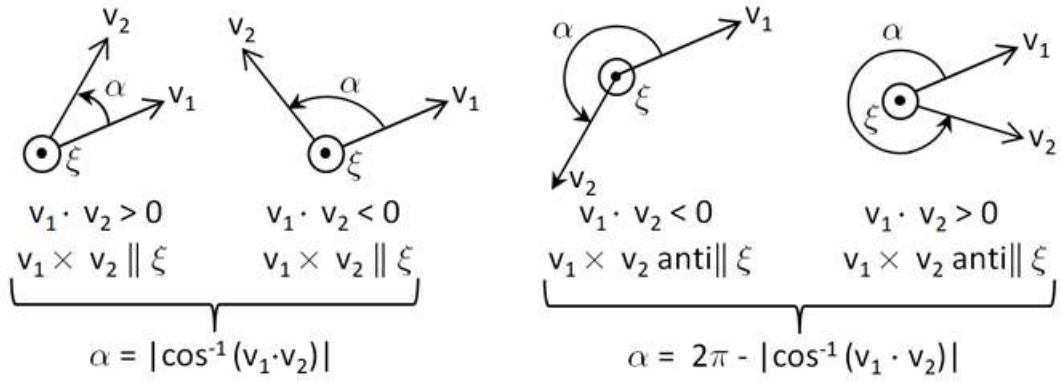


Figure 1.22: Computation of rotation angles for various orientations of initial \mathbf{v}_1 and final \mathbf{v}_2 vectors.

and

$$\cos \alpha = (\cos \theta' - \cos \theta \cos \psi) / (\sin \psi \sin \theta) \quad (1.44)$$

or

$$\sin \alpha = -\sin \theta' \sin(\phi - \phi') / \sin \psi \quad (1.45)$$

for $\psi \neq 0$ or π and for $0 \leq \phi - \phi' \leq \pi$. If $\pi < \phi - \phi' < 2\pi$, then α' and α are given by the negatives of these equations. The scattering angle ψ is given by Eq. (9.3). The rotation and scattering angles depend only on the difference in azimuthal angles via $\phi - \phi'$. Special cases are required when θ' , θ , or ψ are zero. For $\psi = 0$, set $\alpha' = \alpha = 0$ since $\phi' = \phi \implies \sin(\phi - \phi') = 0$. For $\psi = \pi$, set $\alpha' = \alpha = 0$ since $\phi = \phi' + \pi \implies \sin(\phi - \phi') = 0$. If $\sin \theta' = 0$, replace Eqs. (1.42) and (1.44) with (Hu et al., 2001)

$$\cos \alpha' = -\cos \theta' \cos(\phi - \phi') \quad (1.46)$$

$$\cos \alpha = \cos \theta'. \quad (1.47)$$

If $\sin \theta = 0$, replace Eqs. (1.42) and (1.44) with

$$\cos \alpha' = \cos \theta \quad (1.48)$$

$$\cos \alpha = -\cos \theta \cos(\phi - \phi'). \quad (1.49)$$

When doing calculations of multiple scattering between sea surface wave facets, a light ray can reflect from one wave facet to another several times before the incident ray finally leaves the surface region and needs to be rotated into the final meridian plane. In this case, it is more convenient to obtain the rotation angles from the perpendicular (or parallel) axes as determined for the incident ray direction onto a facet and the normal to the tilted wave facet. The details of these calculations are given in Mobley (2015) and will not be repeated here because they are not of general interest. However, it can be seen from Fig. 1.21 that the rotation angles can be obtained from $\cos^{-1}(\hat{\mathbf{v}}' \cdot \hat{\mathbf{p}}')$ and $\cos^{-1}(\hat{\mathbf{p}} \cdot \hat{\mathbf{v}})$, where the dot denotes the vector dot or inner product. Figure 1.22 illustrates this for general initial \mathbf{v}_1 and final \mathbf{v}_2 vectors.

Once the incident Stokes vector is specified in the scattering plane, the scattering matrix is applied to obtain the final Stokes vector, which is then expressed in the $\hat{\mathbf{s}}\text{-}\hat{\mathbf{p}}\text{-}\hat{\boldsymbol{\xi}}$

scattering plane coordinate system defined for the final direction: $\hat{\mathbf{s}} \times \hat{\mathbf{p}} = \hat{\boldsymbol{\xi}}$. Finally, the parallel and perpendicular components of the final Stokes vector must be expressed as horizontal and vertical components in the final meridian plane as specified by the $\hat{\mathbf{h}}\text{-}\hat{\boldsymbol{\xi}}$ system. As illustrated in Fig. 1.21, this requires a counterclockwise rotation through an angle of α , where α is the ‘‘interior’’ angle of the spherical triangle illustrated in the figure. If $\mathbb{R}(\gamma)$ represents a counterclockwise (positive) rotation through angle γ and $\mathbb{M}(\psi)$ represents scattering through scattering angle ψ , then this scattering process is symbolically represented by

$$\mathbb{S} = \mathbb{R}(\alpha)\mathbb{M}(\psi)\mathbb{R}(\alpha')\mathbb{S}' . \quad (1.50)$$

For the choice of a positive rotation being counterclockwise when looking into the beam, the Stokes vector rotation matrix is (e.g., [Mishchenko et al., 2002](#), page 25)

$$\mathbb{R}(\gamma) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos 2\gamma & -\sin 2\gamma & 0 \\ 0 & \sin 2\gamma & \cos 2\gamma & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} . \quad (1.51)$$

These rotation matrices have several obvious but important properties:

$$\mathbb{R}(\pi) = \mathbb{I} , \quad (1.52)$$

$$\mathbb{R}(\pi + \gamma) = \mathbb{R}(\gamma) , \quad (1.53)$$

$$\mathbb{R}(\pi - \gamma) = \mathbb{R}(-\gamma) , \quad (1.54)$$

$$\mathbb{R}(-\gamma) = \mathbb{R}^{-1}(\gamma) = \mathbb{R}^T(\gamma) , \quad (1.55)$$

$$\mathbb{R}(\gamma_1)\mathbb{R}(\gamma_2) = \mathbb{R}(\gamma_1 + \gamma_2) . \quad (1.56)$$

Equation (1.52) shows that rotating a coordinate system through an angle of π , which turns $\hat{\mathbf{s}}$ and $\hat{\mathbf{p}}$ into $-\hat{\mathbf{s}}$ and $-\hat{\mathbf{p}}$ leaves the Stokes vector unchanged, consistent with the previous remark about directions $-\hat{\mathbf{s}}$, $-\hat{\mathbf{p}}$ being equivalent to $\hat{\mathbf{s}}$, $\hat{\mathbf{p}}$ as regards Stokes vectors. That is, Stokes vectors are referred to a plane, not to a particular direction in that plane.

Although not needed here, it is noted for completeness that a rotation in 3-D space by angle γ about an axis specified by unit vector $\hat{\mathbf{u}} = u_x\hat{\mathbf{x}} + u_y\hat{\mathbf{y}} + u_z\hat{\mathbf{z}} = (u_x, u_y, u_z)$ is

$$\mathbf{R}_{3-D}(\hat{\mathbf{u}}, \gamma) = \begin{bmatrix} \cos \gamma + u_x^2(1 - \cos \gamma) & u_x u_y(1 - \cos \gamma) - u_z \sin \gamma & u_x u_z(1 - \cos \gamma) + u_y \sin \gamma \\ u_y u_x(1 - \cos \gamma) + u_z \sin \gamma & \cos \gamma + u_y^2(1 - \cos \gamma) & u_y u_z(1 - \cos \gamma) - u_x \sin \gamma \\ u_z u_x(1 - \cos \gamma) - u_y \sin \gamma & u_z u_y(1 - \cos \gamma) + u_x \sin \gamma & \cos \gamma + u_z^2(1 - \cos \gamma) \end{bmatrix} \quad (1.57)$$

for a counterclockwise rotation when looking in the $-\hat{\mathbf{u}}$ direction. This matrix can be used to check quantities determined by other means. For example, $\hat{\mathbf{s}}' = \mathbf{R}_{3-D}(\hat{\boldsymbol{\xi}}', \alpha')\hat{\mathbf{h}}'$ as seen in Fig. 1.21.

The choice of coordinate systems and rotation angles is not unique. [Kattawar and Adams \(1989\)](#), [Kattawar \(1994\)](#), [Zhai et al. \(2012\)](#), and [Mishchenko et al. \(2002\)](#) all choose the reference plane to the be meridian plane. (These authors use somewhat different

notation; our $\hat{\mathbf{h}}$ and $\hat{\mathbf{v}}$ are Kattawar’s \mathbf{r} and \mathbf{l} , respectively. Mishchenko et al. (2002, page 16) on the other hand use the spherical coordinate system unit vectors $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\varphi}}$ for the vertical and horizontal axes.) Thus in the ocean setting they regard the “horizontal” direction (parallel to the mean sea surface) as being the “perpendicular” direction (relative to the meridian plane), and “vertical” to the mean sea surface as being the “parallel” direction. Note in Eq. (1.35) that if the parallel component of the electric field is zero, then the second element of the Stokes vector is proportional to $-E_{\perp}E_{\perp}^* = -E_{\perp}^2$, which is a negative number. A Stokes vector for light reflected from the sea surface with horizontal linear polarization is therefore proportional to $\mathbb{S} = [1, -1, 0, 0]^T$, which they refer to as perpendicular polarization. Their choice of the reference plane means that the scattering (Mueller) matrix that transmits only horizontally polarized light has the form

$$\tilde{\mathbb{M}}_{\text{h}} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (1.58)$$

However, Bohren and Huffman (1983) and Hecht (1989) choose their “parallel” direction to be parallel to a horizontal direction, such as a laboratory bench top or the mean sea surface, and their vertical direction is perpendicular to the bench top or mean sea surface. Thus their Stokes vector for horizontal polarization (parallel to the mean sea surface) is $\mathbb{S} = [1, 1, 0, 0]^T$ and their Mueller matrix that transmits only horizontally polarized light has the form

$$\tilde{\mathbb{M}}_{\text{h}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

The difference choices arise perhaps from the viewpoints of describing polarization in a convenient way for a laboratory experiment with reference to a table top, versus modeling light incident onto the sea surface with reference to meridian planes.

The forms of scattering matrices as commonly used to model scattering within the ocean and atmosphere are discussed in Section 9.3.

Similar confusion is found in the choice of rotation angles. Kattawar (and his students in their papers) and Bohren and Huffman (1983) define a positive rotation as being *clockwise* when looking into the beam. Since a clockwise rotation through angle γ is the same as a counterclockwise rotation through $-\gamma$, Kattawar’s rotation matrix is the transpose of the one in Eq. (1.51) (see Eq. 1.55). Thus Kattawar (e.g., Kattawar and Adams, 1989, Eq. 10) writes Eq. (1.50) as $\mathbb{S} = \mathbb{R}(-\alpha)\mathbb{M}(\psi)\mathbb{R}(-\alpha')\mathbb{S}'$ (again, with minor differences in notation; α' here is Kattawar’s Φ , etc.). Others often write a rotation as $\mathbb{R}(\pi - \alpha)$ rather than $\mathbb{R}(-\alpha)$; these are equivalent because by Eq. (1.51) shows that $\mathbb{R}(\pi - \alpha) = \mathbb{R}(-\alpha)$. Chandrasekhar (1960) also defines a positive rotation as being clockwise when looking into the beam. However, he uses a different definition for the Stokes vector for which only the fourth component is independent of coordinate system, so his rotation matrix is more complicated. Mishchenko defines a positive rotation as being clockwise when looking in

the direction of propagation. This is equivalent to counterclockwise when looking into the beam as used here; thus his rotation matrix is the same as that in Eq. (1.51). [van de Hulst \(1980\)](#) also uses the same rotation convention as is used here. All of this is considered “well known,” so the details are often omitted in publications, with confusing apparent differences being the price of brevity. Fortunately, the only real requirement for Stokes vectors, coordinate systems, and rotations is consistency in usage once a choice has been made¹⁰.

1.8 The Limitations of Radiance

There are both philosophical and numerical problems with the use of radiance. These problems are often related to the fundamental limitations of the concept of radiance. This section briefly addresses those limitations.

First, recall the operational definition (1.19) of radiance:

$$L(\mathbf{x}, t, \hat{\xi}, \lambda) \triangleq \frac{\Delta Q}{\Delta t \Delta A \Delta \Omega \Delta \lambda}.$$

There is nothing wrong with this definition, although the value of the measured radiance will depend on the sizes of the time interval, area, solid angle, and wavelength band being considered. The problems arise if this definition is squeezed too hard, e.g., by taking the limit as the solid angle goes to zero. The situation is analogous to the operational definition of the mass density of a substance as being $\Delta M/\Delta V$, where ΔM is the mass of substance in a volume ΔV . You cannot let the volume ΔV go to zero because, physically, the volume eventually becomes smaller than even a single molecule of the material and the ratio becomes meaningless. The trick, in practice, is not to let ΔV be so small that the $\Delta M/\Delta V$ ratio begins to fluctuate because the number of molecules in ΔV is noticeably affected by random thermal fluctuations, or even becomes less than one. If you keep in mind that ΔV can be “small” but not go to zero, the concept of density is very useful, as is that of radiance.

A question then arises: “What is the radiance of a collimated beam?” The answer is that the radiance of a perfectly collimated beam is not defined because $\Delta \Omega$ would be zero while the energy in the beam, ΔQ , remains finite. (Or, if you wish, the radiance becomes mathematically a Dirac delta function, but Dirac delta functions are not measurable physical quantities.) You can define (and measure) the radiance of a beam of light only if it has some divergence in direction.

Likewise, you cannot define the radiance emitted by the surface of a point source because ΔA becomes zero even though the point source is emitting a finite amount of energy. (Here ΔA is the area of the surface emitting the energy. See Section 1.5.8 for the distinction between *surface radiance* for radiance emitted by a surface and *field radiance* for radiance incident onto a surface.) That is why point sources are described by their intensity, which is power emitted per unit solid angle.

When doing Monte Carlo calculations you always have to deal with finite solid angles and finite surface areas when collecting the simulated “photons” or rays as they bounce

¹⁰I’ll be the first to admit that this discussion of Stoke’s vector rotations is complicated and confusing. However, if you think it is bad reading about this business, just wait until you have to write a computer program to do these calculations.

around. You cannot tally the number of rays hitting a point or traveling within a solid angle of size zero. (However, the backward Monte Carlo techniques discussed in Section E.7 can allow you to simulate point detectors and collimated detectors under certain circumstances.)

If you are a hard-core physicist, radiance does not exist. Peruse, for example, standard texts like *Introduction to Electrodynamics* by Griffiths (1981), *Absorption and Scattering of Light by Small Particles* by Bohren and Huffman (1983), or *Optics* by Hecht (1989) and you will find no mention of either radiance or solid angle. This is because what exists in nature is not radiance but electric and magnetic fields, which are described by Maxwell's equations (Chapter 11). Quantities of practical interest, such as the electromagnetic energy crossing a surface or scattered by a particle, can be computed using Maxwell's equations and derived quantities such as the Poynting vector. Philosophically speaking, any problem solved by thinking about radiance can be solved with better accuracy and without fundamental limitations by working with electromagnetic fields.

Why, then, do people use radiance? The reason is simple: It is usually exceedingly difficult if not impossible to compute the electromagnetic fields for situations of practical interest. The electric and magnetic fields are vector quantities, and solving Maxwell's equations with appropriate sources and boundary conditions for natural water bodies is almost always beyond the realm of reasonable computation. Moreover, much of the information contained in the electric and magnetic field vectors, such as the phases of the fields, is not needed unless diffraction or coherent scattering are of interest.

You can, however, get a good-enough answer for most (but not all) practical problems by working with the rather contrived but simpler concept of radiance. You just have to remember, for example, not to let solid angles or detector areas go to zero. Likewise, you have to remember that radiance cannot be used for solving problems that depend on the phase of electromagnetic waves, e.g., for problems such as diffraction or coherent backscatter. (However, effects such as diffraction and coherent backscatter can be included within radiance-based calculations to the extent that they can be parameterized by the volume scattering function.) It is also wise to remember that *radiance is defined within the conceptual framework of geometric optics*. You can solve a lot of, but not all, optics problems using geometric optics and ray tracing, but now and then geometric optics is inadequate for the task at hand. Whenever geometric optics or radiative transfer theory fail, you have to get out Maxwell's equations and start calculating electric and magnetic fields, keeping track of both phases and amplitudes, and that is not easy.

Light Fields in Nature

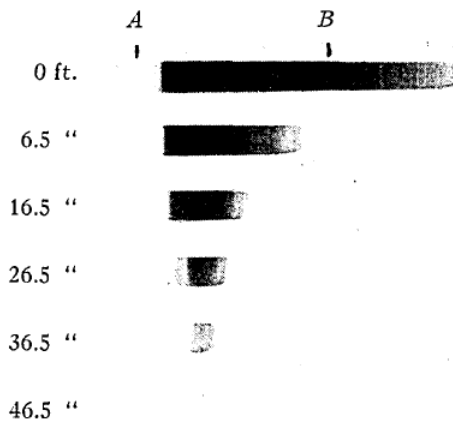
The previous chapter defined several abstract quantities used to describe light—radiance and various irradiances in particular. This chapter now illustrates those concepts with measurements and numerical calculations. The first section shows examples of measured irradiances for the range of very clear to very turbid waters typically encountered in nature. The next section begins with examples of measured radiances. Radiance is always hard to display graphically because it depends, at a minimum, on depth, polar angle, azimuthal angle, and wavelength. Section 2.2.1 therefore also shows different ways radiance is commonly visualized. The following section discusses blackbody radiation, which although not of interest in most optical oceanography is nevertheless important for understanding light from the Sun. A section on the difference in point and density functions follows because these differences are needed to understand spectral functions such as the solar radiance. The chapter finishes with sections on two internal sources of light in the ocean: bioluminescence and Cherenkov radiation.

2.1 Irradiances

Of the radiometric variables, plane irradiance is the most commonly measured and the easiest to visualize, so we will begin with irradiance.

The first person to put a spectrograph in the ocean was likely R. Bertel in 1911. Although his instrument is preserved in the Musée Océanographique in Monaco, his report (Bertel, 1912) cannot be found online. In 1933, H. A. Erikson used a quartz prism to disperse light onto photographic film, which he then analyzed to get semi-quantitative values of the “intensity” of light as a function of depth and wavelength (Erikson, 1933). An example of his photographic data from Gunflint Lake, Minnesota, USA is shown in Fig. 2.1. About this figure, he says “It is there seen how the spectrum shortens at both ends, showing a marked absorption by the water of the longer and shorter wave-lengths. It is also seen that at a depth of 46.5 ft. only the portion in the yellow regions of the visible part of the spectrum is present with sufficient intensity to perceptibly affect the photographic plate in the 30 sec. of exposure.” That is a good description of how the spectrum of light changes with depth in productive waters. He was able to convert the degree of exposure of the film to numerical values and thereby plot curves of intensity versus depth for different

wavelengths, and values of intensity versus wavelength at selected depths, which are correct in all ways, except for the accurate calibration of magnitudes and exact specification of the wavelengths.



AB = Visible Spectrum.
 FIG. 1. Showing the shortening of the light spectrum at different depths, using 30 second exposure.

Figure 2.1: Figure 1 from Erikson (1933) showing strips of film as exposed in his instrument. Red is at the left and blue is at the right of each strip of film.

The first measurements of spectral irradiances underwater that meet modern standards for optical instrument design and radiometric calibration were made in the 1960s by J. E. Tyler and R. C. Smith at the Scripps Institution of Oceanography Visibility Laboratory. Their instrument was optically and mechanically complex, but it was very carefully tested and characterized regarding matters such as stray light sensitivity, electronic stability, and cosine response (the cosine collector of this instrument is shown in Fig. 1.13). It was radiometrically calibrated using lamps traceable to the National Bureau of Standards (now NIST). The radiometer measured from 362 to 736 nm with a bandwidth (full width at half maximum) of 5.3 to 5.9 nm. Measurement of one spectrum (either E_d or E_u at all wavelengths for a given depth) typically required about 25 minutes. Although a modern instrument would be smaller by a factor of 100, would require only a laptop computer, and could measure in minutes what they could measure in hours, it is unlikely than any modern instrument could obtain data significantly better than what was measured by these two founding fathers of optical oceanography. Their measurements were so important that an entire book, Tyler and Smith (1970), was written to describe the instrument and to show data collected from nine different water bodies. Good data have a long half life, and their measurements are still useful half a century later.

Crater Lake, Oregon, USA is a deep (almost 600 m) lake in the caldera of a dormant volcano. Its only water supply comes from rain and snow, so there is almost no input of nutrients. The water is famous for its deep blue color and is considered to be as close to optically pure water as can be found in nature. Secchi depths (Section 16.7) are typically 30 to 40 meters. Figure 2.2 shows downwelling and upwelling spectral irradiances measured in Crater Lake¹ on 5 August 1966. The sky was clear.

¹Ray Smith tells me that he had nightmares about the boat sinking during this field work. The small boat that was available had only 5 cm of freeboard when all of the gear was on board. He figured out a way to keep all of the heavy cable in the water, which increased the freeboard to 10 cm. Even the smallest wave would have swamped the boat and sent everything to the bottom of the deepest lake in the United

Depth $z = 0$ is in the water, just below the surface. Note that as the depth increases, absorption by water itself rapidly removes the downwelling light for wavelengths longer than about 600 nm. At blue wavelengths around 450 nm, E_d decreases by less than one half between the surface and 25 m, but it decreases by over four orders of magnitude at 700 nm. The upwelling irradiance has its maximum value near 410 nm, consistent with the visual appearance of the water as deep blue. (The visual appearance of the lake is best based on upward radiance in air for the given viewing direction, but that could not be measured by this instrument. E_u at 5 m depth is a reasonable proxy for the inherent color of the water.)

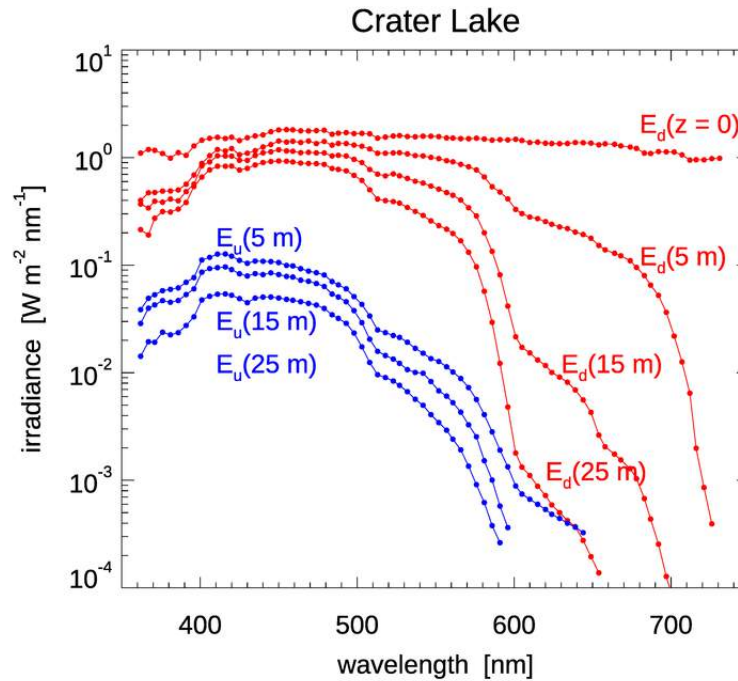


Figure 2.2: Spectral downwelling and upwelling irradiances measured in Crater Lake, Oregon, USA. Plotted from data tabulated in [Tyler and Smith \(1970\)](#).

San Vicente Reservoir, California, USA is a man-made lake whose primary purpose is to supply drinking water to San Diego, California. Its primary water supply comes from water taken from the Colorado River and delivered to the reservoir by an aqueduct. Surface chlorophyll values are typically 5 to 20 mg m^{-3} , but can go over 100 in bloom conditions. Secchi depths are typically 1 to 5 m. This reservoir is an example of turbid, highly productive waters, but it is by no means an extreme limit of what can be found in nature. [Figure 2.3](#) shows irradiance spectra at selected depths measured in San Vicente Reservoir on 20 January 1967. No chlorophyll concentration or other water-quality parameters were measured at the time of the optical measurements. In these waters, at green wavelengths around 550 to 600 nm, the downwelling irradiance decreases by two orders of magnitude between the surface and 10 m depth. The red wavelengths are removed by pure water absorption, just as for Crater Lake. Now, however, blue wavelengths are also removed just

as strongly, almost certainly by high concentrations of colored dissolved organic matter in the water. In Crater Lake, at the wavelengths where the water is clearest, one half of the surface E_d remains at 25 m; in San Vicente Reservoir, one half of the surface irradiance is gone by just 1 m in depth, even at the clearest wavelengths. In San Vicente, the upwelling irradiance at 1 m is greatest in the 530 to 580 nm band, which is yellowish green to greenish yellow. Visually, San Vicente would appear as a turbid greenish yellow color.

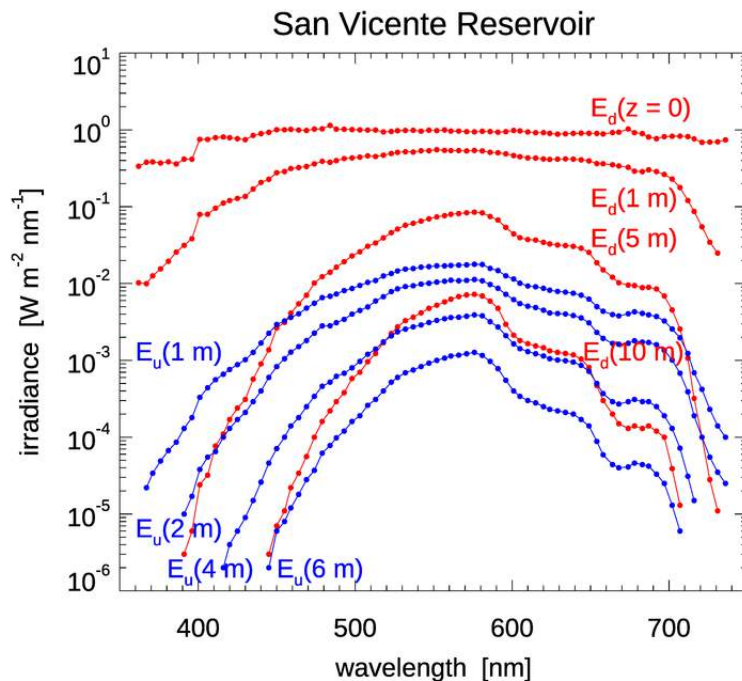


Figure 2.3: Spectral downwelling and upwelling irradiances measured in San Vicente Reservoir, California, USA. Plotted from data tabulated in [Tyler and Smith \(1970\)](#).

These two data sets illustrate the range of irradiance depth and wavelength behaviors commonly encountered in natural waters. There are almost no natural water bodies clearer than Crater Lake, but waters with very heavy sediment loads (sections of the Amazon river, for example) can have Secchi depths of centimeters, not meters, and highly eutrophic lakes can have chlorophyll concentrations in excess of 100 mg m^{-3} and similarly low visibility.

We will use these spectra again in Chapter 4 to illustrate diffuse attenuation functions and reflectances, after those quantities are defined.

2.2 Radiances

Although A. Gershun gets credit for the invention of the Gershun tube radiometer (Fig. 1.10), I can find no papers that show data taken by him with such an instrument. Perhaps the first measurements of radiances underwater were made by [Johnson and Liljequist \(1938\)](#), who² made measurements using red, green, and blue filters for broadband wave-

²N. G. Johnson later became known as N. G. Jerlov, whose name was honored by the Office of Naval Research when they established The Jerlov Award, which is bestowed every two years to recognize scientists

length selection. They measured the radiance at selected polar angles (0, 10, 20, 30, 45, 60, 90, and 180 deg relative to the zenith) and at azimuthal angles of 0, 45, 90, and 180 deg relative to the Sun. They presented their results as polar plots, an example of which is seen in Fig. 2.4. Note that as the depth increases, the direction of maximum radiance shifts toward the zenith.

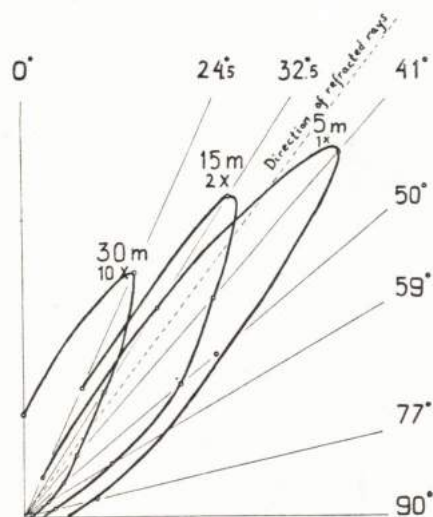


Fig. 2.

Figure 2.4: Spectral radiances measured in the plane of the Sun by Johnson and Liljequist (1938). Note the shift of the direction of the maximum radiance toward the zenith as the depth increases. Also note that the values at 15 and 30 m are rescaled by factors of 2 and 10, respectively, compared to the values at 5 m.

The first underwater measurements of a full radiance distribution were made by J. E. Tyler and colleagues at the Scripps Visibility Laboratory in the late 1950s. Tyler, S. Q. Duntley, R. W. Austin, and several others developed a submersible radiance meter over a six-year timespan with the goal of obtaining radiance distributions that could be used for testing models of underwater light fields, which were being developed by R. W. Preisendorfer and others. The instrument is shown in Fig. 2.5. The two “radiance tubes” seen at the right are Gershun tubes with a field of view of 6.6 deg; the two tubes point in opposite directions. The head on which the tubes are mounted rotates so that the tubes cover the full range of polar angles. The fin and propeller at the other end of the instrument were used to hold the instrument at the desired azimuthal angle while the radiance sensors rotated. The Gershun tubes are each 0.5 m long, so the entire instrument was about 1.5 m long (fin tip to radiance head). The details of the instrument design were presented at a conference (Duntley et al., 1955), but unfortunately I can find no written record of the instrument specifications and operation. However, a photograph of the “control panel and power supply” shows that the control panel was the size of a large kitchen refrigerator (say a meter wide and two meters tall). The control panel shows “space for strip chart recorders,” so data apparently were recorded on paper strip charts. The data report comments that the original data “...are very nearly linearly proportional of the log of the radiance....” Each Gershun tube passed its light to a separate photomultiplier tube. Wratten #45 gelatin filters gave a wavelength response centered at 480 nm with a 64 nm FWHM bandwidth. Neutral density filters could be rotated into the beam to control the dynamic range as needed for different depths.

who make significant contributions to oceanic optics.

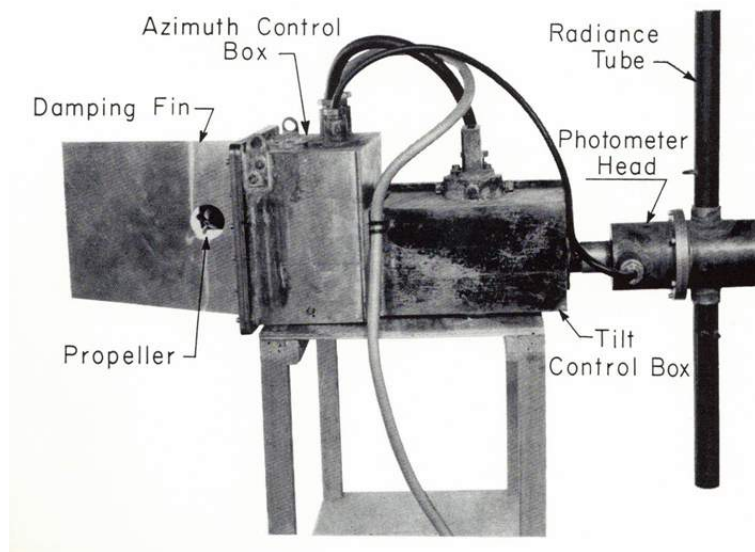


Figure 2.5: The radiometer used to obtain the radiances seen in Figs. 2.6 and 2.7. Photograph from Tyler (1960).

This instrument was deployed in Lake Pend Oreille, Idaho, USA in 1957. Somewhat surprisingly, this lake in the mountains of northern Idaho, 500 km from the nearest salt water, has been the home of a U. S. Navy facility since 1942. The lake is ideal for acoustic experiments because the underwater environment is very quiet (no singing whales!). The lake equally well met Tyler's need for homogeneous, optically deep water, along with the Navy's large barge with a crane for instrument deployment. The data report (Tyler, 1960) tabulates measured radiance distributions for both clear and overcast sky conditions at 7 depths from 4.24 to 66.1 m for the clear day and 5 depths from 6.1 to 55.0 m for the overcast day. The polar angle resolution was 10 deg from $\theta_v = 0$ (nadir viewing) to 180 deg (zenith viewing), and azimuthal angles at 20 deg resolution from $\phi_v = 0$ (looking toward the Sun) to 180 deg (looking opposite to the Sun). Data acquisition for the sunny day began at 08:52 and ended at 14:41 local time. During these five hours, the Sun's zenith angle varied from 49 deg to 34.5 at noon to 45 at the end. The data were all normalized to values from a vertical profile of zenith and nadir radiances taken within a few minutes of noon. (By comparison, the NURADS system of Voss and Chapin (2005) has a fisheye lens that captures the radiance for an entire hemisphere on a CCD with an angular resolution of less than 0.5 deg, and at six wavelengths. One measurement including all wavelengths requires about 2 minutes.) In spite of improved instrumentation, the Tyler data set is still useful and serves the purpose of this section.

Figure 2.6 shows profiles at three depths taken on 28 April 1957, a day that was very clear and with a glassy calm water surface. The radiances in this plot are shown as functions of the viewing directions (θ_v, ϕ_v) defined in Eq. (1.6). Remember that θ is measured from 0 in the $+\hat{z}$ or nadir direction. Thus $\theta_v = 0$ is looking straight down at the nadir radiance traveling toward the zenith; $\theta_v = 180$ is looking straight up at the zenith radiance, which is heading straight down. Azimuthal angle $\phi_v = 0$ is looking toward the Sun, and $\phi_v = 180$ is looking away from the Sun.

The red dots in the figure are in the azimuthal plane of the Sun; the green dots are almost at a right angle to the Sun’s azimuthal direction (the 20 deg ϕ resolution had one Gershun tube pointing at $\phi = 80$ deg when the other was at 100 deg). All values are normalized to 1 at the nadir-viewing direction for a depth of 4.2 m. Tyler estimated that the Sun’s maximum radiance would have been at least twice the largest measured value if the instrument had allowed for better angular resolution. The green dots from the polar angle sweep at (nearly) right angles to the Sun do not “see” the Sun’s direct beam, so the maximum at 4.2 m is only 3% of the maximum when looking at the Sun. Note that as the depth increases, the polar angle of maximum radiance shifts from about 30 deg off-zenith ($\theta_v \approx 150$ as plotted) to the zenith for the measurement in the plane of the Sun. The radiances for both azimuthal orientations are almost identical at 66.1 m. This was hard evidence that the radiance approaches an asymptotic shape at great depths, as predicted by Preisendorfer (1959) and others. The asymptotic radiance distribution depends only in the inherent optical properties of the water body, as will be seen in Section 10.4.

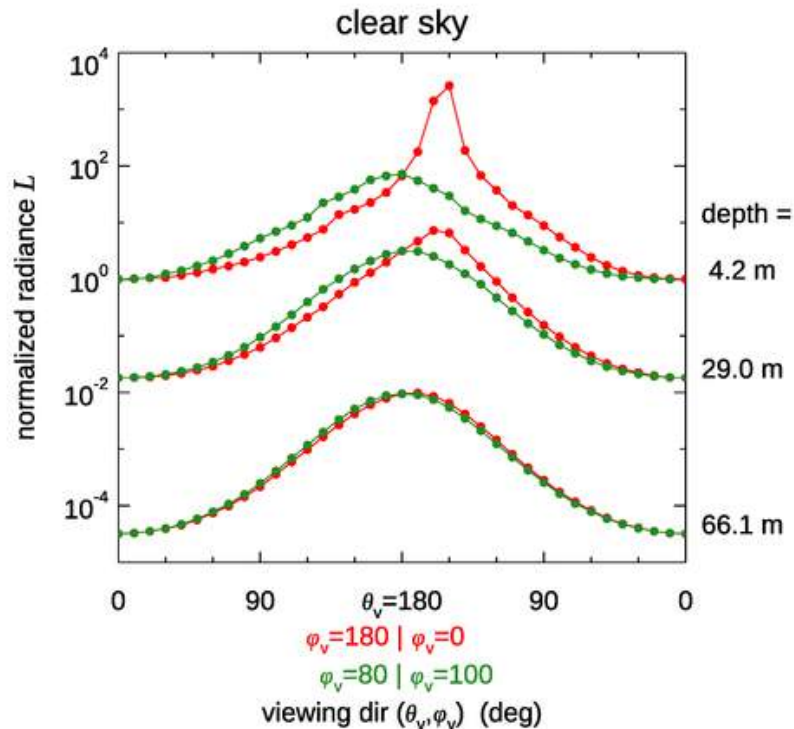


Figure 2.6: Radiance distributions for the clear day in the plane of the Sun (red) and at nearly right angles to the Sun (green). Plotted from data tabulated in Tyler (1960).

Figure 2.7 compares Tyler’s measurements made on the clear day of Fig. 2.6 with data taken on an overcast day (16 March 1957). The clear-day data are the same as in the previous figure at 4.2 m depth; the overcast-day data are at a depth of 6.1 m at a time when the solar zenith angle was 50 deg. All curves in the plot have been normalized to 1 for the nadir-viewing direction so as to compare the shapes of the radiance distributions for clear versus overcast conditions. It is seen that the radiance on the overcast day does not vary much for θ_v within about 40 deg of the zenith-viewing direction. This is the

angular region where Snell’s law maps the entire sky hemisphere into a cone of half-angle 48 deg—the so-called “optical manhole.” This angular region thus sees the fairly uniform sky radiance, so the radiance distribution is fairly uniform in the manhole. Tyler describes the day as “overcast,” but there is still some difference in the curves in the plane of the Sun versus at right angles to the Sun, so the location of the Sun may have been faintly visible.

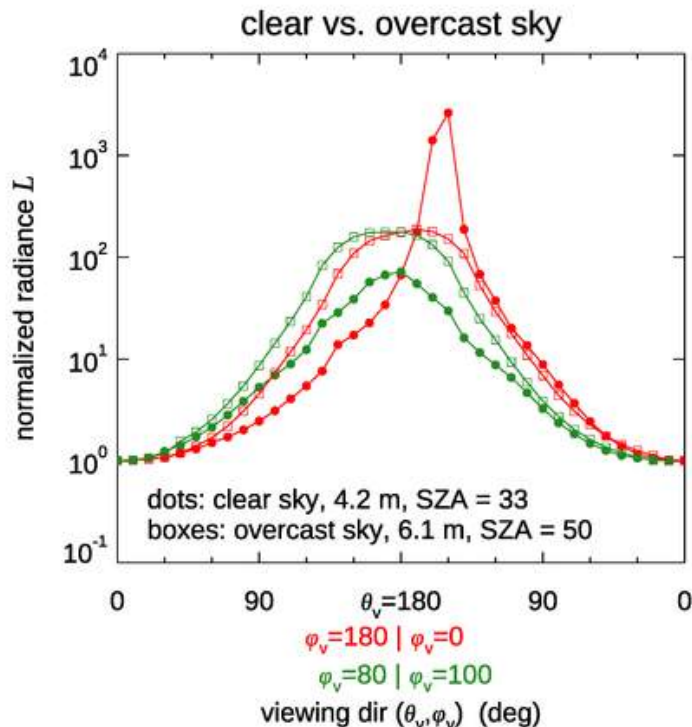


Figure 2.7: Radiance distributions for the clear day (at depth 4.2 m; solid dots) and for the overcast day (at 6.1 m; open boxes) in the plane of the Sun (red) and at nearly right angles to the Sun (green). Plotted from data tabulated in Tyler (1960).

2.2.1 Visualizing Radiances

Figures 2.6 and 2.7 show one way to plot a radiance distribution: as a function of polar angle for a given wavelength, given azimuthal plane, and at selected depths.

Even in the simplest case of horizontally homogeneous water and time independence, the radiance distribution is a function of four variables: depth z , polar angle θ , azimuthal angle ϕ , and wavelength λ . This makes it hard to display radiances graphically and to understand the wealth of information contained in $L(z, \theta, \phi, \lambda)$ for given environmental conditions. The following figures show some of the ways commonly used to present radiance distributions. These figures also give us another chance to study the nature of oceanic radiance distributions.

To obtain a complete radiance distribution for these figures, the HydroLight numerical model (Section 10.6) was run for very simple conditions:

- A bio-optical model for homogeneous Case 1 water was used with a chlorophyll value of 0.5 mg m^{-2} (a typical value for open-ocean water) to generate the absorption and scattering properties of the water.
- The Sun was placed at a solar zenith angle of 40 deg in a clear sky with typical values for marine atmospheric conditions.
- The sea surface was flat (wind speed of 0).
- The water was infinitely deep.
- Fluorescence by chlorophyll and colored dissolved organic matter (CDOM), and Raman scatter by water, were included.
- Radiances were computed on a 10 by 15 deg θ - ϕ angular grid; the computed radiances are thus the average radiances over each 10 by 15 degree angular “window.”
- Radiances were computed at 10 nm wavelength resolution between 350 and 700 nm; computed radiances are thus the averages over 10 nm bands from 350 to 360 nm, etc.

Figure 2.8 shows how this HydroLight-computed radiance distribution in air just above the sea surface depends on polar viewing direction and wavelength, in the azimuthal plane containing the Sun. This figure requires explanation. In HydroLight the polar angle is measured from 0 in the $+z$ or downward direction to 180 deg in the $-z$ or upward direction. Thus light heading straight down has a polar angle of $\theta = 0$. This plot uses the *viewing* direction (θ_v, ϕ_v) , which is the direction an instrument would point when measuring the radiance. Thus the polar viewing angle of $\theta_v = 0$ means looking straight down (nadir-viewing) and seeing the radiance heading straight up (the upwelling radiance L_u ; i.e., light traveling upward in the $\theta = 180$ deg direction). The Sun is located at $\phi_v = 0$ (solar photons travel in the $\phi = 180$ deg direction). This figure shows the full range of polar angles in the $\phi = 0$ –180 deg plane of the Sun. The plotting convention is that positive polar angles ($\theta_v > 0$) are looking toward the Sun (the $\phi_v = 0$ half plane) and negative polar angles ($\theta_v < 0$) are looking away from the Sun (the $\phi_v = 180$ deg half plane). Thus $\theta_v = +90$ deg corresponds to looking horizontally towards the Sun, and $\theta_v = -90$ deg corresponds to looking horizontally away from the Sun. The range of θ_v values between 0 and +90 and between 0 and -90 correspond to looking downward. The range of θ_v values from 90 to 180 to -90 on the θ_v plot axis correspond to looking upward. The thin black lines show the θ - λ computational grid. The colors correspond roughly to the visual colors of the various wavelengths. Note that the radiance values are plotted on a logarithmic scale.

The Sun’s radiance is the large spike at $\theta_v = 140$ deg, which corresponds to looking upward at a zenith angle of 40 deg in the $\phi_v = 0$ azimuthal direction, where the Sun was located in this simulation. The smaller spike at $\theta_v = 40$ deg is the Sun’s specular reflection, seen looking downward and towards the Sun. The ratio of the Sun’s specular reflectance at 355 nm ($0.2586 \text{ W m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$), obtained from the radiance values tabulated in the data file) to the Sun’s direct beam ($9.9546 \text{ W m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$) is 0.026, which is consistent with the Fresnel reflectance for this incidence angle and water index of refraction ($n = 1.34$). The broader radiance peak near $\theta_v = -90$ deg is the relatively bright near-horizon sky radiance and its reflection by the sea surface, seen looking away from the Sun.

Figure 2.9 shows the radiance in the water at depth 0, i.e., just below the level sea surface. The plotting conventions are the same as for Fig. 2.8. Note that the large spike of the Sun’s direct beam has shifted to a smaller zenith angle near 30 deg (i.e., a larger θ_v ,

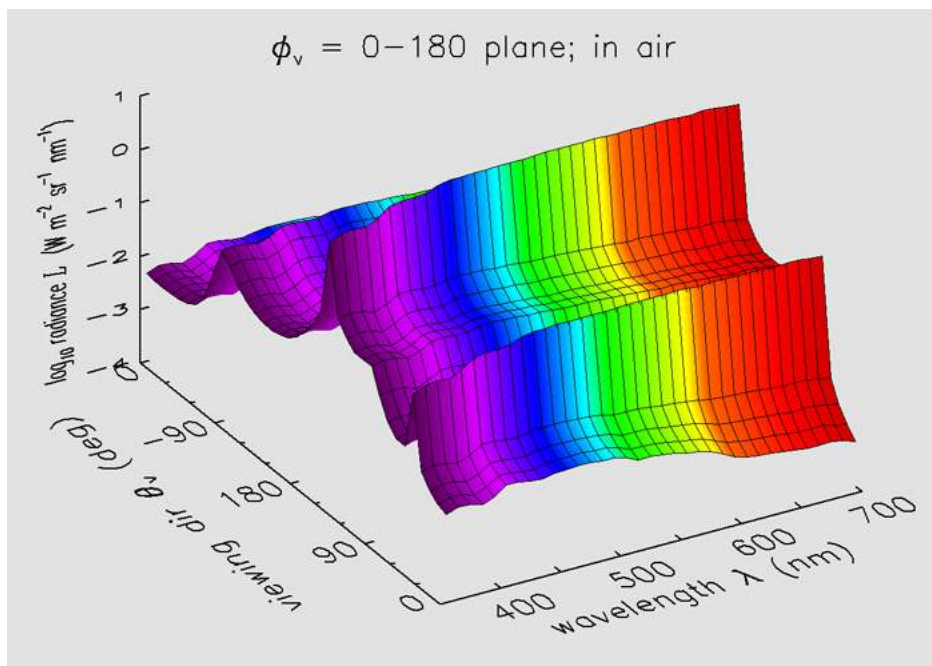


Figure 2.8: Radiance as a function of polar viewing direction and wavelength, in the azimuthal plane of the Sun, in air just above the level sea surface.

now near 150 deg) underwater, in accordance with Snell's law for refraction across a level air-water surface. The specular reflectance spike in the plot has of course disappeared.

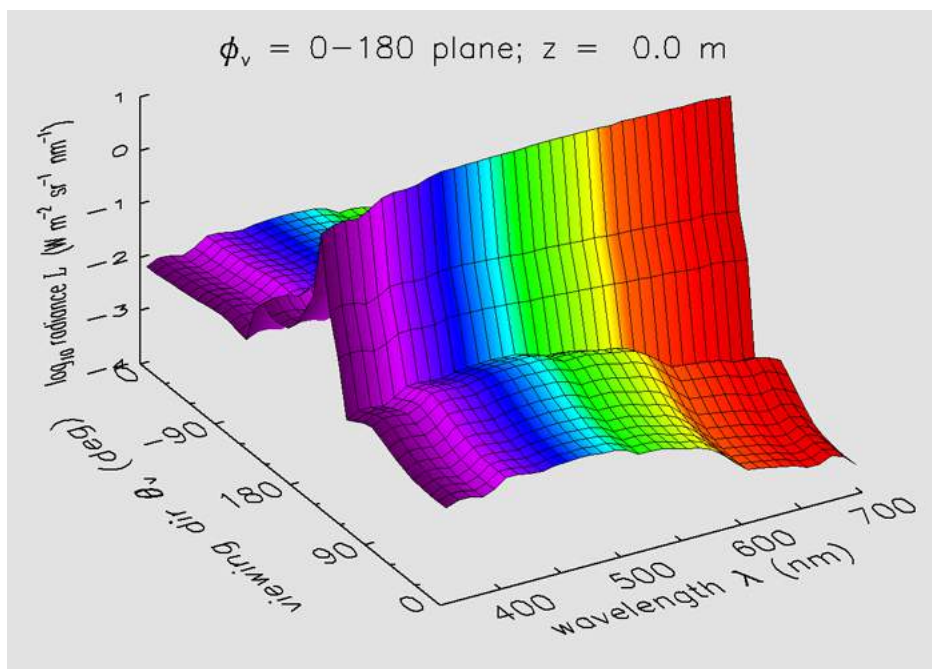


Figure 2.9: Radiance as a function of polar viewing direction and wavelength, in the azimuthal plane of the Sun, at depth 0 just below the level sea surface.

Figure 2.10 shows the radiance at a depth of 10 m. The radiance distribution is now “smoothing out” because of the effects of scattering, which redirects the original directions of the rays. However, the direction of the Sun’s direct beam is still obvious. The effect of chlorophyll fluorescence on the upwelling radiance near 685 nm is quite obvious as the “bump” in the values near $\theta_v = 0$ and $\lambda = 685$ nm. The contribution of fluorescence to the downwelling radiance near 685 nm is not as obvious because much of the downwelling radiance is still from sunlight that has penetrated to this depth.

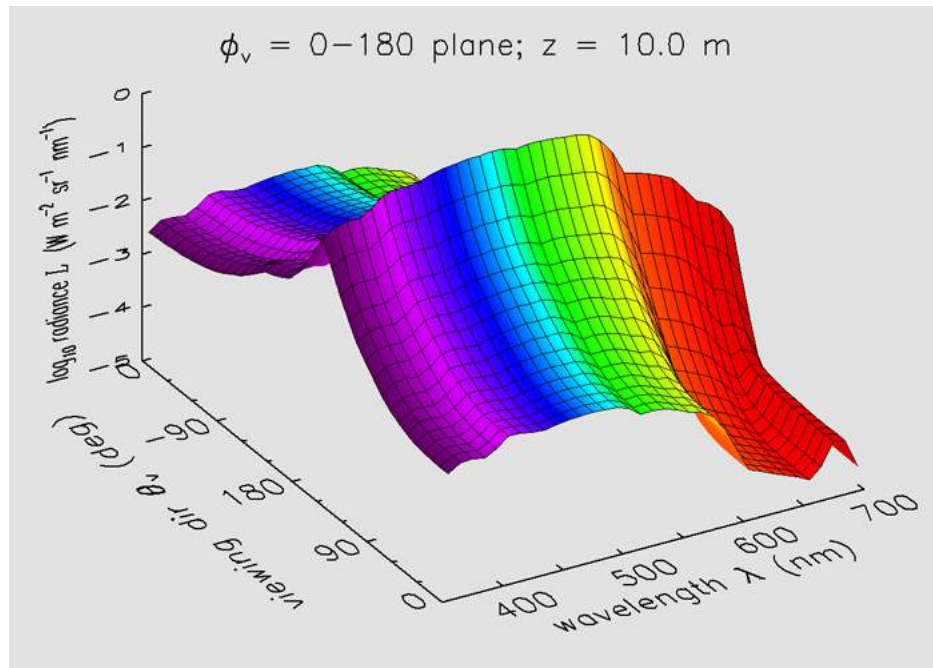


Figure 2.10: Radiance as a function of polar viewing direction and wavelength, in the azimuthal plane of the Sun, at depth 10 m.

Figure 2.11 shows the radiance at a depth of 50 m. Now multiple scattering has removed almost all information about the Sun’s zenith angle. This radiance is close to the asymptotic radiance distribution (Section 10.4), which is determined only by the water’s absorption and scattering properties, and not by the boundary conditions (the solar location in particular) at the sea surface. Chlorophyll fluorescence is now responsible for almost all of the red light near 685 nm in all directions, since the Sun’s beam does not penetrate this deep at red wavelengths because of absorption by water.

The information seen in Figs. 2.8-2.11 does not show the azimuthal dependence of the radiance, except in the plane containing the Sun. Figure 2.12 illustrates both the polar and azimuthal dependence of the radiance at one depth, in air, and one wavelength, 555 nm. To interpret this plot, imagine that you are at the center of the sphere. The solid black line around the “equator” is looking horizontally ($\theta_v = 90$ deg). The solid black line drawn from “pole to pole” is the $\phi_v = 0$ azimuthal direction. The dotted lines show the HydroLight θ - ϕ computational grid. The radiances computed as averages over each dotted-line θ - ϕ window have been spline interpolated to a 5x5 degree grid for generation of the contours in this figure. Values of \log_{10} of the radiance in $\text{W m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$ are contoured and color-coded for display as a function of θ_v and ϕ_v . The colors now represent

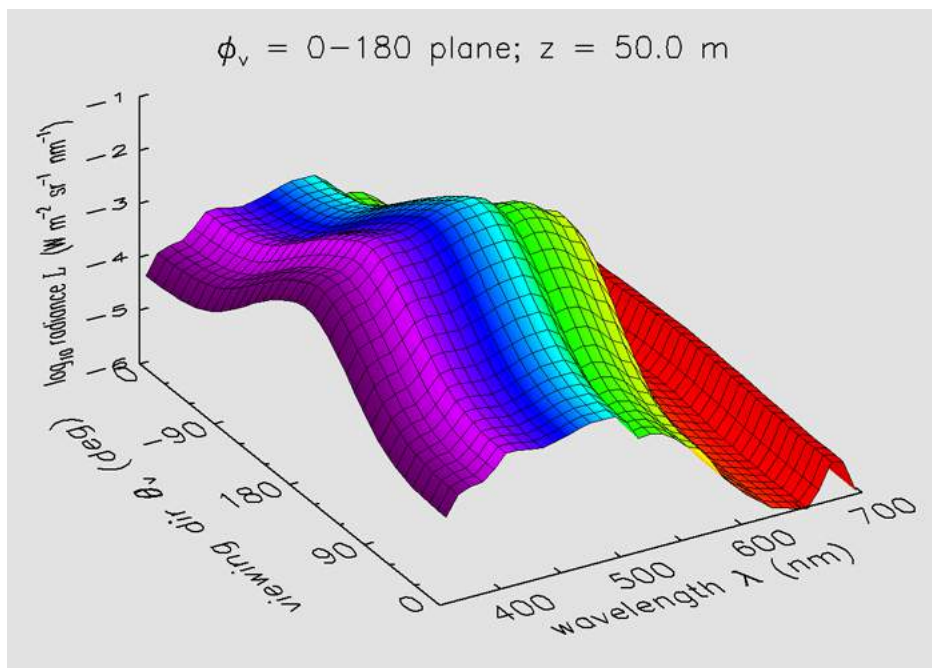


Figure 2.11: Radiance as a function of polar viewing direction and wavelength, in the azimuthal plane of the Sun, at depth 50 m.

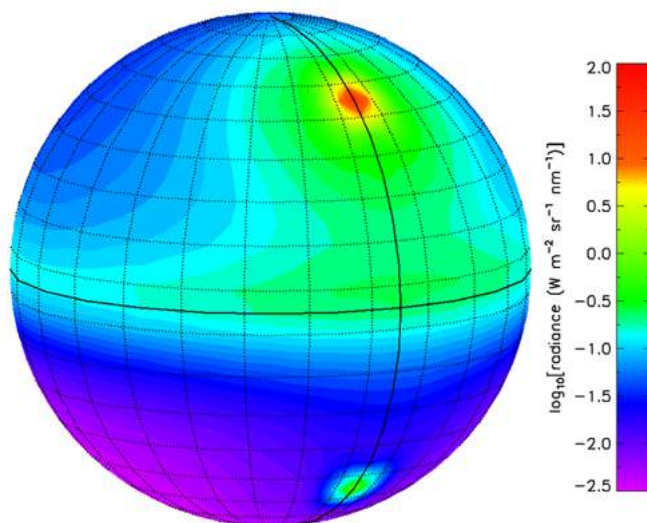


Figure 2.12: Radiance as a function of polar and azimuthal viewing directions, just above the sea surface and at wavelength 555 nm.

the magnitude of the \log_{10} (radiance), with violet being lowest magnitude and red highest.

In this figure the “northern hemisphere” represents looking upward toward the sky, and the “southern hemisphere” represents looking downward toward the sea surface. Thus the Sun is the red spot seen looking upward at a zenith angle of 40 deg in the $\phi_v = 0$ direction. (The Sun is centered in the θ computational window that extends from 35 to 45 deg in

zenith angle, and in a 15 deg wide ϕ window centered on $\phi = 0$.) The corresponding green spot “below the equator” at $\phi_v = 0$ is the Sun’s specular reflectance seen looking toward the Sun but downward at a nadir angle of 40 deg.

The detailed directional information seen in Figs. 2.8-2.12 is seldom necessary in optical oceanography. It is often sufficient to display the radiance in a few selected directions as a function of depth and wavelength, as in the next two figures.

Figure 2.13 shows the radiance in five directions: upwelling or nadir-viewing, L_u ; downwelling or zenith-viewing, L_d ; and horizontal directions L_h looking at azimuthal angles toward, at a right angle to, and away from the Sun’s azimuthal direction. These radiances are plotted as functions of depth at one wavelength, 555 nm. Although radiances and irradiances are generally thought of as decreasing with depth, this is true only away from the surface. Note that several of these radiances actually increase with depth in the upper few meters of the water column, below which they then decrease. This increase is caused by scattering (path radiance) into the direction being plotted, as the Sun’s direct (unscattered) beam is redirected by scattering within the water.

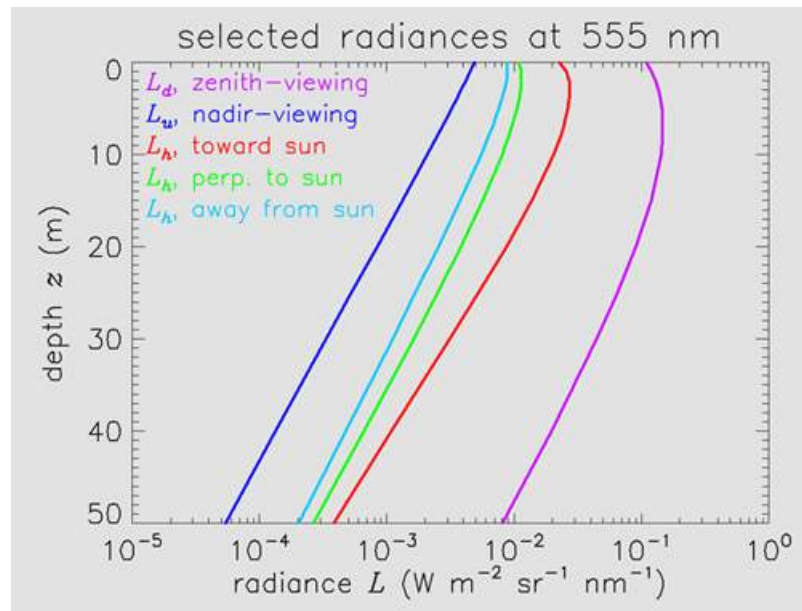


Figure 2.13: Radiances in selected directions as a function of depth, at wavelength 555 nm.

Figure 2.14 shows the downwelling or zenith-viewing radiance L_d plotted as a function of wavelength for selected depths. The downwelling radiance in air, which is the zenith sky radiance, is plotted in black. The in-water curves are plotted in color using the same wavelength coding as in Figs. 2.8-2.10. Note that the downwelling radiance in water at depth $z = 0$ is *greater* than the downwelling radiance in air just above the sea surface. This may seem counterintuitive, since some of the downwelling radiance in air is lost to surface reflection when passing into the water. This increase in L_d when going from air to water is a consequence of the “ n^2 law for radiance” (Section 13.1.2). This law, also called “the fundamental theorem of radiometry,” states that the radiance divided by the square of the index of refraction, L/n^2 , remains constant as light travels through regions of different n , to the extent that absorption and scattering can be neglected. Because of Snell’s law of

refraction, a given solid angle in air, $\Delta\Omega$, becomes $\Delta\Omega/n^2$ in water. Although about 2.6% of the downwelling photons from the sky are Fresnel-reflected back upward for $n = 1.34$, most enter the water. Those transmitted rays then travel in a smaller solid angle, and thus the associated underwater radiance is greater by a factor of roughly n^2 . Some of the downwelling radiance at $z = 0$ is also due to upwelling radiance L_u being reflected back downward by the sea surface, but this contribution to L_d is small since L_u is typically one to two orders of magnitude less than L_d . Note that L_d at 10 m is greater than the value at $z = 0$ for blue to green wavelengths. This is the same behavior as was seen at 555 nm in Fig. 2.13.

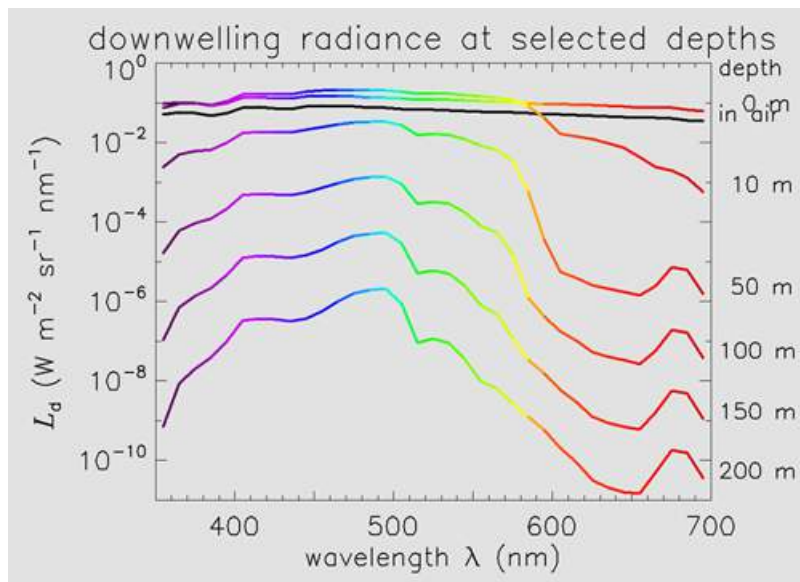


Figure 2.14: Downwelling radiance L_d as a function of wavelength at selected depths

By 200 m depth the blue-green downwelling radiance has decreased by 5 orders of magnitude from the surface value, and other wavelengths have decreased even more. The dominant wavelength is near 500 nm. To get a rough feeling for how this would appear visually, assume that the radiance distribution is isotropic and of magnitude $L = 10^{-6} \text{ W m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$ over the 50 nm band from 475 to 525 nm, with other wavelengths being negligible. The corresponding plane irradiance is then $E \approx 50\pi 10^{-6} \approx 10^{-4} \text{ W m}^{-2}$. As seen in Table 2.3, this is comparable to a moonlit night. Thus the available light at 200 m in this simulated ocean would appear as a very faint bluish green. This is reminiscent of what explorer William Beebe reported during his pioneering bathysphere dives in the 1930's. In one dive near Bermuda (in water that was likely lower in chlorophyll than the 0.5 used here, hence bluer and more transparent) he reported (Beebe, 1934)

“The green faded imperceptibly as we went down, and at 200 feet [61 m] it was impossible to say whether the water was greenish-blue or bluish-green....At 600 feet [183 m] the color appeared to be a dark, luminous blue, and this contradiction shows the difficulty of description. As in former dives, it seemed bright, but was so lacking in actual power that it was useless for reading and writing.”

If you wish to explore this radiance distribution more, you can download the data file `VisualizingRadiances.zip` and the IDL plot programs used to create the figures above from the Visualizing Radiances page of the Web Book.

2.3 Blackbody Radiation

The birth date of modern physics can be regarded as December 14, 1900, when Max Planck presented his derivation of the spectral distribution of radiant energy in thermodynamic equilibrium with matter at a given temperature. The derivation of this distribution is both conceptually and mathematically subtle, and Max well deserved his subsequent Nobel Prize. Planck's function is commonly called the *blackbody radiation spectrum*.

As derived in most physics books, e.g., [Liboff \(1980, Eq. 2.3\)](#) or [Eisberg and Resnick \(1985, Chapter 1\)](#), Planck's function is expressed as a spectral energy density³:

$$U_E(f) = \frac{8\pi h f^3}{c^3} \frac{1}{e^{hf/kT} - 1}, \quad (2.1)$$

where f is temporal frequency in s^{-1} , h is Planck's constant, c the speed of light in vacuo, k is Boltzmann's constant, and T is the temperature in Kelvin. (See [Table 1.1](#) for the values of h , c , and k .) $U_E(f)$ thus has units of $J/(m^3 s^{-1}) = J/(m^3 Hz)$, or energy per unit volume per unit temporal frequency interval (with temporal frequency measured in Hertz = cycles per second).

For ease of comparison with the Sun's irradiance, or with the irradiance measured at the entrance of a blackbody cavity, [Eq. \(2.1\)](#) can be converted to spectral plane irradiance as a function of wavelength. Proceeding as in [Section 1.5.9](#), the energy contained in a unit frequency interval df must equal the energy contained in the corresponding wavelength interval $d\lambda$, i.e.,

$$U_E(f)|df| = U_E(\lambda)|d\lambda|.$$

Recalling that $f = c/\lambda$ gives $df = -(c/\lambda^2)d\lambda$, and [Eq. \(2.1\)](#) becomes

$$U_E(\lambda) = U_E(f) \left| \frac{df}{d\lambda} \right| = \frac{8\pi hc}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1}, \quad (2.2)$$

which has units of $J/(m^3 m)$, or energy per unit volume per unit wavelength interval (with wavelength measured in meters).

The scalar irradiance E_o is related to the energy density by $E_o = Uc$. One way to see this is to think of the many photons making up the energy density. How many photons "hit" a small spherical detector per unit time, there to be recorded as scalar irradiance, equals how many photons there are times how fast they are moving, i.e. $E_o = Uc$. Radiation in thermodynamic equilibrium is isotropic and unpolarized. For isotropic radiance, $E_o = 4E_d$, where E_d is the plane irradiance. Thus [Eq. \(2.2\)](#) can be converted to spectral plane irradiance by a factor of $c/4$:

$$E_d(\lambda) = \frac{c}{4} U_E(\lambda) = \frac{2\pi hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1}. \quad (2.3)$$

This is the form of Planck's law seen, for example, in [Leighton \(1959, page 65\)](#).

³Note that both of these texts use ν for temporal frequency in Hertz.

Two final transformations of Eq. (2.3) are needed for comparison with the Sun's solar irradiance as measured at the top of the earth's atmosphere, as seen Section 2.4. First, in accordance with the r^2 law for irradiance, the irradiance emitted at the Sun's surface (presumed to be a blackbody in the present discussion) is reduced by a factor of $(R_{\text{Sun}}/R_{\text{Earth}})^2$ to obtain the irradiance at the mean distance of the earth's orbit. Here $R_{\text{Earth}} = 1.496 \cdot 10^8$ km is the radius of the Earth's orbit, and $R_{\text{Sun}} = 6.95 \cdot 10^5$ km is the Sun's radius. Finally, a factor of 10^{-9} is applied to Eq. (2.3) to convert the wavelength spectral interval from meters to nanometers. The resulting equation is

$$E_d(\lambda) = \left(\frac{R_{\text{Sun}}}{R_{\text{Earth}}} \right)^2 \frac{2\pi hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1} 10^{-9}, \quad (2.4)$$

where $E_d(\lambda)$ is now in $\text{W m}^{-2} \text{ nm}^{-1}$, although the wavelength is still measured in meters on the right-hand side of the equation for consistency with the SI units for h, c and k .

Integrating Eq. (2.3) over all wavelengths gives the total plane irradiance emitted by a black body:

$$E_d = \sigma T^4, \quad (2.5)$$

where $\sigma = (2\pi^5 k^4)/(15h^3 c^2) = 5.6703 \cdot 10^{-8} \text{ W m}^{-2} \text{ K}^{-4}$ is the Stefan-Boltzmann constant. The Sun's total (over all wavelengths) irradiance as measured at the top of the atmosphere is approximately 1368 W m^{-2} . Carrying this value back to the Sun's surface via a factor of $(R_{\text{Earth}}/R_{\text{Sun}})^2$ and inserting the result into Eq. (2.5) gives a corresponding black body temperature of $T = 5,782 \text{ K}$. That is, a black body at this temperature emits the same total irradiance as does the Sun. This temperature is then used in Eq. (2.4) to generate the blackbody spectra seen in the figures of Section 2.4.

Other forms of the blackbody spectrum are sometimes useful. As already noted, blackbody radiation is isotropic. For isotropic radiance L_o , $E_d = \pi L_o$, where π has units of steradian. Thus formula (2.3) for plane irradiance can be converted to a formula for blackbody radiance L_{BB} by dividing by π :

$$L_{\text{BB}}(\lambda) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1}. \quad (2.6)$$

For some applications it is useful to know the photon density or photon irradiance. The photon density U_Q is obtained from the energy density by dividing the energy density U_E by the energy hf of a single photon. Thus Eq. (2.1) gives

$$U_Q(f) = \frac{8\pi f^2}{c^3} \frac{1}{e^{hf/kT} - 1}, \quad (2.7)$$

where U_Q has units of photons/($\text{m}^3 \text{ Hz}$). Similarly, Eq. (2.3) can be divided by the energy per photon in wavelength units, hc/λ , to obtain the photon plane irradiance

$$Q_d(\lambda) = \frac{2\pi c}{\lambda^4} \frac{1}{e^{hc/\lambda kT} - 1}, \quad (2.8)$$

where Q_d has units of photons/($\text{s m}^2 \text{ m}$). Integrating this equation over all wavelengths gives the total number of photons emitted per second per unit area by a blackbody:

$$Q_d = \sigma_Q T^3, \quad (2.9)$$

where $\sigma_Q = (4.808\pi k^3)/(h^3 c^2) = 1.520 \cdot 10^{15}$ photons $s^{-1} m^{-2} K^{-3}$ is the photon equivalent of the Stefan-Boltzmann constant. Thus the total *energy* emitted by a blackbody is proportional to T^4 , but the total *number of photons* emitted is proportional to T^3 . As the temperature increases, the blackbody spectrum shifts toward the blue, and relatively fewer but more-energetic short-wavelength photons are needed to keep up with the increasing energy output.

It is also common to use the spatial frequency or wavenumber $\nu = 1/\lambda$ as the spectral variable. A change of variables based on $U_E(\nu)|d\nu| = U_E(\lambda)|d\lambda|$ and $d\lambda/d\nu = -\lambda^2$ then gives

$$U_E(\nu) = U_E(\lambda) \left| \frac{d\lambda}{d\nu} \right| = 8\pi h c \nu^3 \frac{1}{e^{hc\nu/kT} - 1},$$

which has units of $J/(m^3 m^{-1})$, or energy per unit volume per unit wavenumber interval (with wavenumber measured in 1/meters). Other formulas in terms of wavenumber are obtained as before.

Table 2.1 summarizes various formulas for blackbody radiation. These cover everything needed for optical oceanography. However, the [Spectral Calculations](#) website has much additional information about blackbody radiation, including such esoterica as how the spectrum shifts if the blackbody source is moving at relativistic speeds.

Figure 2.15 shows the energy and photon densities, and energy and photon irradiances, for a temperature of $T = 5782$ K, corresponding approximately to the Sun's surface temperature. These curves were computed using the first four formulas in Table 2.1. It should be noted that the energy spectra have their maxima at about 500 nm for this temperature, whereas the photon spectra have their maxima at about 635 nm. That is, where the Sun's output is a maximum depends on what measure of the output is used, as well as on which variable is used for the spectral density. This important matter is discussed below Section 2.5.

Quantity	Spectral Variable	Units	Formula
Energy density	wavelength	$\frac{\text{J}}{\text{m}^3 \text{ m}}$	$U_E(\lambda) = \frac{8\pi hc}{\lambda^5} \left(\frac{1}{e^{hc/\lambda kT} - 1} \right)$
Photon density	wavelength	$\frac{\text{photons}}{\text{m}^3 \text{ m}}$	$U_Q(\lambda) = \frac{8\pi}{\lambda^4} \left(\frac{1}{e^{hc/\lambda kT} - 1} \right)$
Energy irradiance	wavelength	$\frac{\text{W}}{\text{m}^2 \text{ m}}$	$E_d(\lambda) = \frac{2\pi hc^2}{\lambda^5} \left(\frac{1}{e^{hc/\lambda kT} - 1} \right)$
Photon irradiance	wavelength	$\frac{\text{photons}}{\text{s m}^2 \text{ m}}$	$Q_d(\lambda) = \frac{2\pi c}{\lambda^4} \left(\frac{1}{e^{hc/\lambda kT} - 1} \right)$
Energy density	temporal frequency	$\frac{\text{J}}{\text{m}^3 \text{ Hz}}$	$U_E(f) = \frac{8\pi hf^3}{c^3} \left(\frac{1}{e^{hf/kT} - 1} \right)$
Photon density	temporal frequency	$\frac{\text{photons}}{\text{m}^3 \text{ Hz}}$	$U_Q(f) = \frac{8\pi f^2}{c^3} \left(\frac{1}{e^{hf/kT} - 1} \right)$
Energy irradiance	temporal frequency	$\frac{\text{W}}{\text{m}^2 \text{ Hz}}$	$E_d(f) = \frac{2\pi hf^3}{c^2} \left(\frac{1}{e^{hf/kT} - 1} \right)$
Photon irradiance	temporal frequency	$\frac{\text{photons}}{\text{s m}^2 \text{ Hz}}$	$Q_d(f) = \frac{2\pi f^2}{c^2} \left(\frac{1}{e^{hf/kT} - 1} \right)$
Energy density	wavenumber	$\frac{\text{J}}{\text{m}^3 \text{ m}^{-1}}$	$U_E(\nu) = 8\pi hc\nu^3 \left(\frac{1}{e^{hc\nu/kT} - 1} \right)$
Photon density	wavenumber	$\frac{\text{photons}}{\text{m}^3 \text{ m}^{-1}}$	$U_Q(\nu) = 8\pi\nu^2 \left(\frac{1}{e^{hc\nu/kT} - 1} \right)$
Energy irradiance	wavenumber	$\frac{\text{W}}{\text{m}^2 \text{ m}^{-1}}$	$E_d(\nu) = 2\pi hc^2\nu^3 \left(\frac{1}{e^{hc\nu/kT} - 1} \right)$
Photon irradiance	wavenumber	$\frac{\text{photons}}{\text{s m}^2 \text{ m}^{-1}}$	$Q_d(\nu) = 2\pi c\nu^2 \left(\frac{1}{e^{hc\nu/kT} - 1} \right)$

Table 2.1: Blackbody radiation formulas for energy and photon density and for energy and photon plane irradiance, in spectral units of wavelength λ , temporal frequency f , and wavenumber ν . Formulas require wavelength in meters and wavenumber in 1/meters. Divide the E_d and Q_d formulas by π to obtain formulas for blackbody radiances.

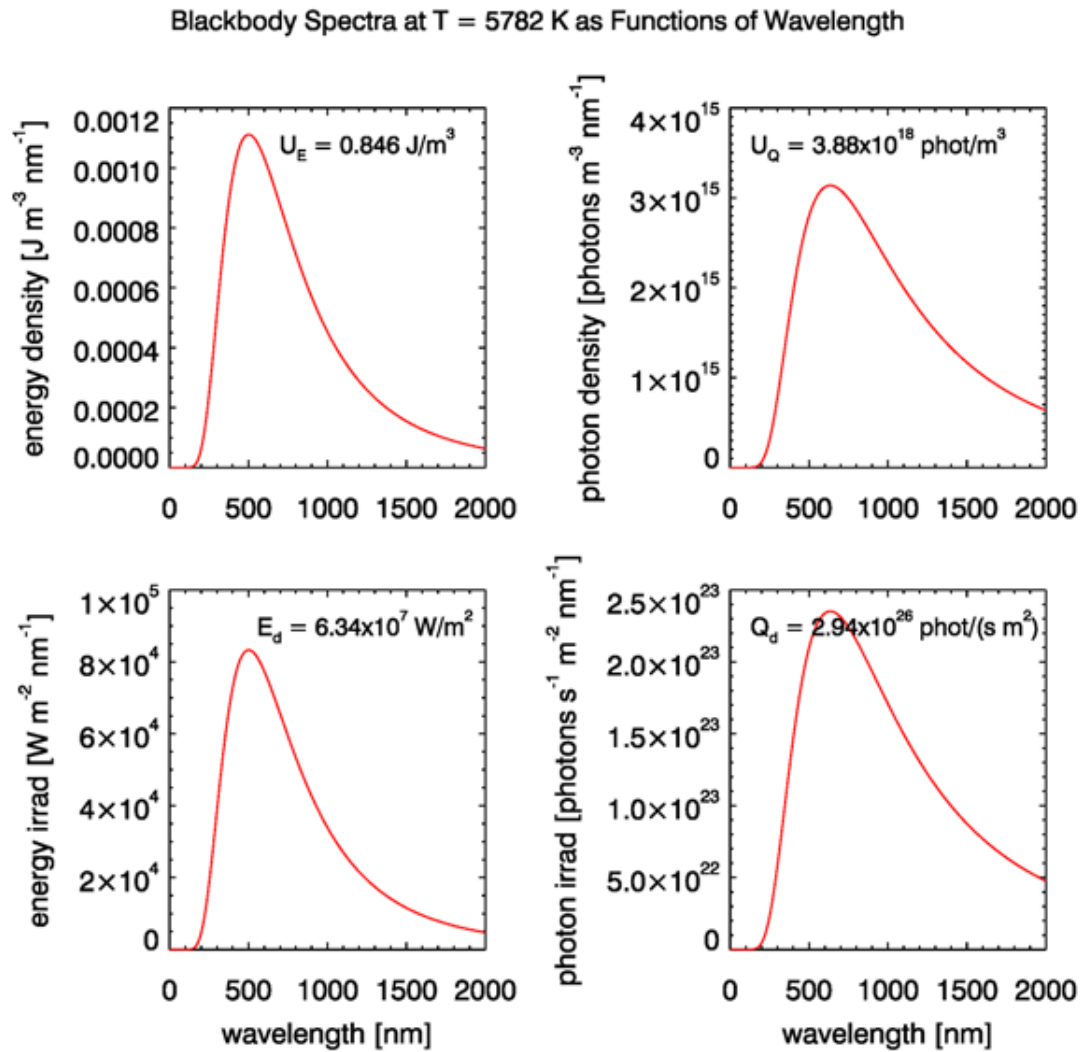


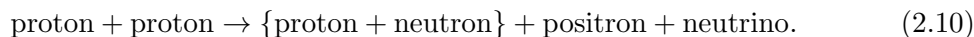
Figure 2.15: Blackbody radiation spectra for energy and photon densities, and for energy and photon irradiances, for a temperature of 5782 K. The inset values give the totals over all wavelengths.

2.4 Light from the Sun

This section briefly surveys how light is produced by the Sun, and characterizes the sunlight reaching the Earth.

2.4.1 The Proton-Proton Cycle

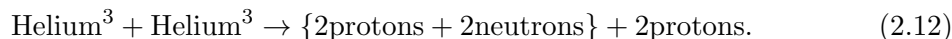
Since most of the light falling upon the earth originates in the Sun, it is worthwhile to start with a brief look at how this light is generated deep within the Sun in a process known as the proton-proton cycle. The core of the Sun is primarily a mixture of completely ionized hydrogen and helium. At the center the temperature is approximately $1.5 \cdot 10^7$ K, and the density is about 150 times that of water. Under these extreme conditions the hydrogen nuclei, or protons, occasionally collide with sufficient energy to overcome their electrostatic repulsion and fuse together according to the reaction



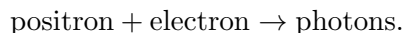
This reaction releases energy, which is carried off by the created particles. The bound state of a proton and a neutron, indicated by the $\{\dots\}$ is called a deuteron. The positron is identical to an electron except for its electric charge, which is positive. The neutrino is an uncharged, nearly massless packet of energy that is in some ways similar to a photon. (There is now experimental evidence that neutrinos, once thought to be massless, do have a very small mass on the order of about 0.1 eV, which is less than one-millionth of the electron mass.) The deuterons are able to undergo further fusion with protons to form an isotope of helium:



This reaction also creates a photon, or particle of light, which carries off most of the energy released in the reaction. The bound state of 2 protons and 1 neutron is a Helium³ nucleus. Helium³ nuclei are in turn able to fuse with each other in the final step of the proton-proton cycle:



The bound state of 2 protons and 2 neutrons is a Helium⁴ nucleus (an alpha particle). In order to create the two Helium³ nuclei needed in step (2.12), there must be two each of reactions (2.10) and (2.11). Thus if we tally the total input, we find that six protons have been converted to a total output of one Helium⁴ nucleus, two protons, two positrons, two neutrinos and two photons. The positrons each soon encounter electrons and undergo mutual annihilation to create two or more photons:



Therefore the net result of the proton-proton cycle is the conversion of four hydrogen atoms into one helium atom plus energy in the form of photons and neutrinos. The mass of one Helium nucleus is about 0.7% less than the mass of the four Hydrogen nuclei (protons); the mass difference is the energy of the photons and neutrines according to the famous relation $E = mc^2$ between energy and mass. The Sun emits about $3.85 \cdot 10^{26}$ W, which

corresponds to a conversion of $4.26 \cdot 10^9 \text{ kg s}^{-1}$ of mass into energy. This corresponds to the mass about 70 large aircraft carriers being converted to energy every second!

Neutrinos interact only very weakly with matter, and they escape from the Sun immediately after their creation in step (2.10). The various photons created in the above reactions are all extremely energetic (in the gamma ray region of the electromagnetic spectrum). They also interact strongly with matter, and therefore the gamma rays undergo repeated scattering, absorption and re-emission by the solar matter as they work their way toward the surface of the Sun. The photons lose energy in these interactions, so that they are predominantly in the visible and infrared parts of the spectrum by the time they arrive at the Sun's surface and escape into space.

There are exothermic stellar nuclear reactions other than the ones just described. However, the proton-proton cycle is responsible for about 98% of the energy generation within our own star.

2.4.2 The Solar Spectrum

At the mean distance of the earth from the Sun, the solar irradiance from photons of all wavelengths, E_s , is

$$E_s = 1368 \pm 5 \text{ W m}^{-2}$$

Although E_s is historically called the *solar constant*, its value varies by a fraction of a percent on time scales of minutes to decades. A better term is therefore the *total solar irradiance*. Moreover, the total solar irradiance received by the earth varies from about 1322 to 1413 W m^{-2} over the course of a year, owing to the ellipticity of the Earth's orbit about the Sun.

As seen in Eq. (1.1), the energy of a photon is inversely proportional to its wavelength. Furthermore, the number of solar photons per wavelength interval is not uniform over the electromagnetic spectrum. Figure 2.16 shows the measured wavelength dependence of the solar spectral irradiance $E_s(\lambda)$. The sharp dips in the $E_s(\lambda)$ curve are *Fraunhofer lines*, which are due to selective absorption of solar radiation by elements in the Sun's outer atmosphere. These lines are typically less than 0.1 nm wide, and they are much "deeper" (the irradiances within the lines are much less) than indicated in Fig. 2.16, which gives $E_s(\lambda)$ values averaged over much wider bands. The resolution of Fig. 2.16 is 1 nm below 630 nm, 2 nm between 630 and 2,500 nm, and 20 nm beyond 2,500 nm. For example, the prominent line centered at $\lambda = 486.13 \text{ nm}$ decreases to 0.2 of the $E_s(\lambda)$ values just outside the line at $\lambda = 486.05$ and 486.25 ; the *line depth* is thus said to be 0.2.

The blue curve in Fig. 2.16 shows the blackbody irradiance for a temperature of 5,782 K. As shown in Section 2.3 on blackbody radiation, this is the temperature for which a perfect absorber and emitter, or blackbody, emits the same total irradiance as the total solar irradiance of 1368 W m^{-2} . The blackbody spectrum is a reasonably good approximation for the Sun's spectral irradiance at infrared (IR) wavelengths, where the solar spectrum is never more than 25% different from the blackbody curve. However, the solar spectrum differs greatly from the black body curve at ultraviolet (UV) and visible wavelengths. Solids and liquids emit radiation that is well approximated by the blackbody curve at the appropriate temperature. Gases, however, show selective absorption and emission over very narrow wavelength ranges, as seen for the Fraunhofer lines. The gases in the solar atmosphere thus do not absorb and emit like a blackbody.

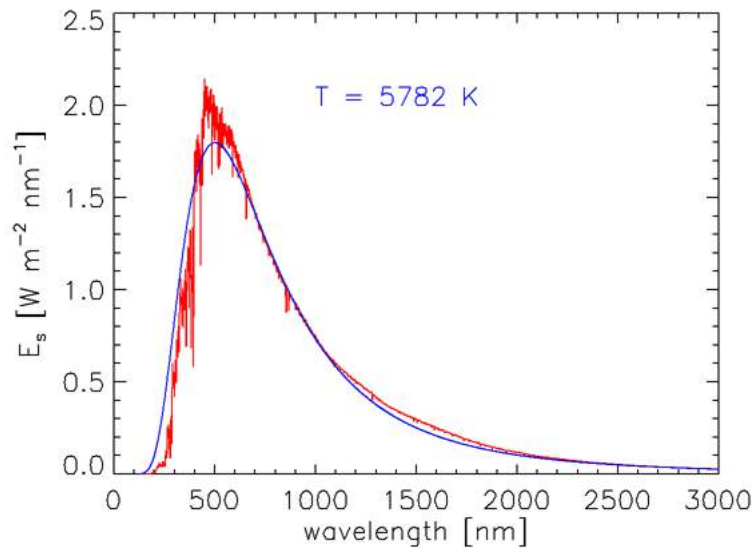


Figure 2.16: The extraterrestrial solar spectrum. Plotted values (red) are the 2000 American Society of Testing and Materials (ASTM) E-490 Air Mass 0 spectrum, which is available from the [ASTM](#). The blue curve is the blackbody spectrum for a temperature of 5,782 K.

Figure 2.17 shows $E_s(\lambda)$ over the wavelength range of most interest in optical oceanography and ocean color remote sensing. The elements in the Sun’s atmosphere causing the most prominent Fraunhofer lines are identified. The Fraunhofer line at 587.5 nm is the line that was originally used in 1868 to deduce the existence of a new element, named Helium (after Helios, the Greek god who personified the Sun), in the Sun’s atmosphere 30 years before it was discovered on earth.

It is seldom necessary for optical oceanographers to concern themselves with the detailed wavelength dependence of $E_s(\lambda)$ seen in Fig. 2.17. It is usually sufficient to deal with $E_s(\lambda)$ values averaged over bandwidths of size $\Delta\lambda \approx 5$ to 20 nm, which correspond to the bandwidths of the optical instruments routinely used in underwater measurements and remote sensing. Table 2.2 gives the distribution of E_s in several broad wavelength bands. About two thirds of the Sun’s energy is in the near-ultraviolet to near-infrared bands relevant to optical oceanography.

It is not the solar irradiance at the top of the atmosphere, but rather the sunlight that actually reaches the sea surface, that is relevant to oceanography. The magnitude and spectral dependence of the solar radiation reaching the earth’s surface are highly variable functions of the solar angle from the zenith (i.e., of the time of day, date, and latitude) and of atmospheric conditions (cloud cover, humidity, aerosols, ozone concentration, etc.). Figure 2.18 illustrates the variability in how much of the total solar irradiance reaches the sea surface. The red curve in that figure is the extraterrestrial solar irradiance E_s averaged over 10 nm bands. The other curves show the total (diffuse plus direct) sea-level downwelling plane irradiance E_d for a range of Sun and sky conditions, also averaged over 10 nm bands. These curves were computed using an the RADTRAN atmospheric radiative transfer model ([Gregg and Carder, 1990](#), as extended to 300 and 1000 nm).

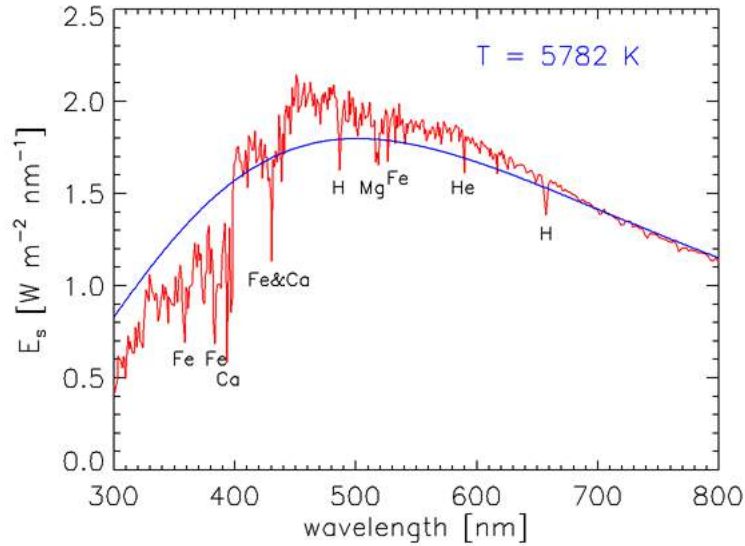


Figure 2.17: Expanded view of Fig. 2.16 for the wavelengths of greatest interest in optical oceanography. The most prominent Fraunhofer lines are labeled by the elements causing them.

Band	Wavelength Interval (nm)	Irradiance (W m^{-2})	Fraction of E_s (percent)
ultraviolet and beyond	< 350	62	4.5
near ultraviolet	350-400	57	4.2
visible	400-700	522	38.2
near infrared	700-1000	309	22.6
infrared and beyond	> 1000	417	30.5
totals		1367	100.0

Table 2.2: Distribution of the total solar irradiance in various wavelength bands.

The $E_d(0, \text{clear})$ curve in Fig. 2.18 is for the Sun at the zenith and a clear marine atmosphere (marine aerosols, sea-level relative humidity 80%, horizontal sea-level visibility 15 km, etc.). Rarely would this much of the solar irradiance actually reach the earth's surface. The $E_d(60, \text{clear})$ curve is for the same atmospheric conditions, but with the Sun at a 60 deg zenith angle. The reduction in the surface irradiance is due primarily to the increased path length through the atmosphere. The $E_d(60, \text{hazy})$ curve shows the effect of going from a clear atmosphere to a cloud-free atmosphere but with increased humidity (95%) and aerosol concentration (mixed marine and continental aerosols), leading to a hazy atmosphere with reduced visibility (5 km). The green curve is for a heavy overcast, for which no Sun is visible.

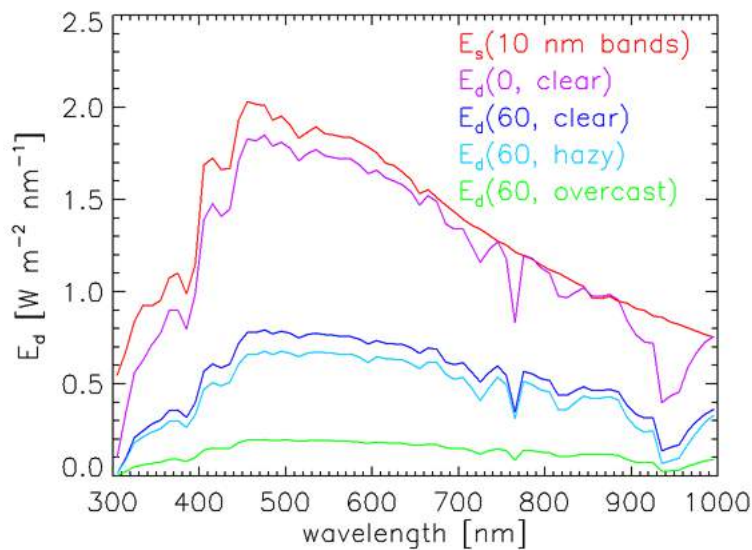


Figure 2.18: Total sea-level plane irradiances E_d for four combinations of solar zenith angle and atmospheric conditions, averaged over 10 nm bands. The red curve is the band-averaged solar irradiance E_s for comparison.

The sea-level irradiances seen in Fig. 2.18 also show the effects of absorption by gases in the earth's atmosphere. Ozone (which was 300 Dobson units in these computations) greatly reduces the irradiance reaching the surface near 300 nm. The prominent dip near 765 nm is due to oxygen, and the broad dip starting near 930 nm is due to water vapor.

Table 2.3 shows typical sea-level irradiances over the visible wavelength band from 400 to 700 nm. These values can show considerable variability depending on cloudiness and atmospheric conditions.

2.4.3 The Lunar Spectrum

It might be supposed that the spectrum of moonlight has the same shape as that of sunlight because moonlight is just sunlight reflected by the lunar surface. That is not the case because the reflectance of the lunar surface depends on wavelength. The left panel of Fig. 2.19 shows that the irradiance of a full Moon is about six orders of magnitude less bright than sunlight, consistent with the typical values seen in Table 2.3. The right panel shows that the moonlight spectrum is shifted towards the red because the lunar surface material absorbs somewhat more blue light than red.

Suppose that you point a telescope at a small, very dark patch of nighttime sky where no nearby stars are in the field of view. As seen in the left panel of Fig. 2.19, there is still a small amount of light present, which comes from starlight scattered by interstellar gas and dust. What is the spectrum of that diffuse background starlight? Since it comes from the light of millions of stars, and the stars you see on a dark night look more or less white, one might expect that the background sky spectrum is also roughly white. That is not the case, as seen in the right panel of Fig. 2.19: the spectrum increases from blue to red. Although it is true that this light is scattered starlight, around 75% of all stars are faint

Environment	Irradiance (W m⁻²)
solar irradiance above the atmosphere (for comparison)	522
very clear atmosphere, Sun at the zenith	500
clear atmosphere, Sun at 60 deg	250
hazy atmosphere, Sun at 60 deg	175
hazy atmosphere, Sun near the horizon	50
heavy overcast, Sun at the zenith	125
heavy overcast, Sun near the horizon	10
clear atmosphere, full moon near the zenith	$1 \cdot 10^{-3}$
clear atmosphere, starlight only	$3 \cdot 10^{-6}$
cloudy night	$3 \cdot 10^{-7}$
clear atmosphere, light from a single, bright star	$3 \cdot 10^{-9}$
clear atmosphere, light from a single, barely visible star	$3 \cdot 10^{-11}$

Table 2.3: Typical total (direct plus diffuse) irradiances at sea level in the visible wavelength band (400-700 nm).

red dwarf stars, which are much smaller and cooler than the Sun. Not a single red dwarf is visible to the naked eye, and their prevalence was not realized until infrared-sensitive telescopes could be placed above the Earth's atmosphere. These stars emit most of their energy in the infrared, and the visible part of their spectra is very red. Thus much of the scattered starlight is at reddish wavelengths. If you compute the visual color of the starlight spectrum of Fig. 2.19 as described in Section 16.3, the color is reddish-orange, somewhat like a pumpkin. The prominent spikes in the stellar spectrum are due to airglow in the Earth's upper atmosphere. Airglow results from recombination of atoms which were ionized by sunlight during the day, or which are excited by cosmic rays at night. In the figure, the airglow emission line at 558 nm is due to excited oxygen atoms about 100 km above the Earth's surface; the emission at 589 nm is due to sodium. When observing deep-space objects, astronomers routinely subtract the background starlight spectrum from measured spectra to obtain the spectrum of the object being observed.

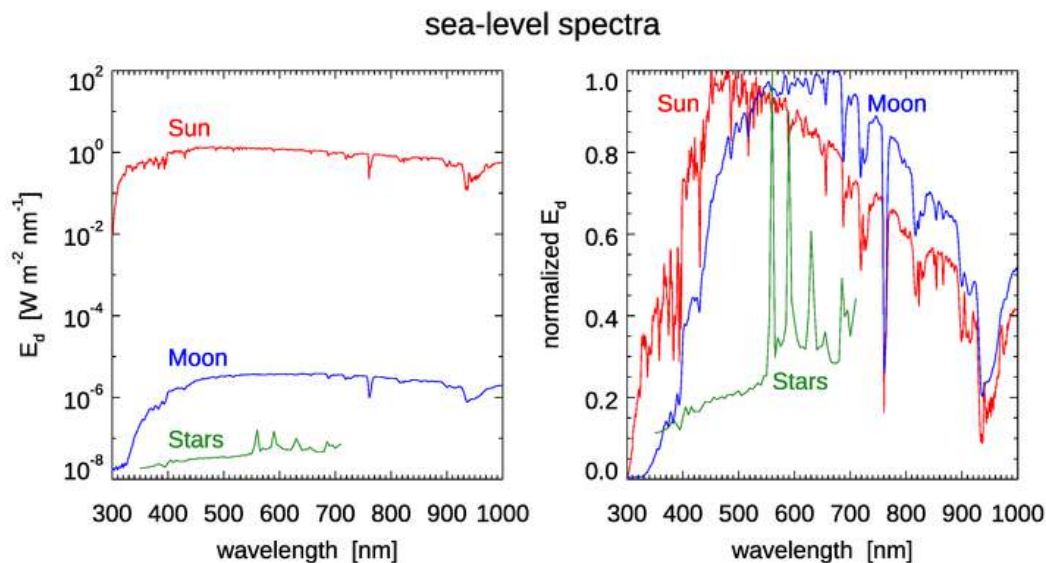


Figure 2.19: Measured sea-level irradiances E_d for the Sun, a full Moon, and the background nighttime sky.

2.5 A Common Misconception

One sometimes sees statements like “The human eye has evolved to take advantage of the Sun’s maximum output at visible wavelengths,” or “Most plants are green because the Sun’s maximum output is at green wavelengths.” At first glance, this idea seems plausible given the plots of solar irradiance in the preceding section, which show that the Sun’s spectral irradiance peaks near 500 nm when the irradiance is plotted as a function of wavelength. However, these speculations about the relation between eye or plant evolution and solar output are simply wrong.

2.5.1 Point Functions and Density Functions

Suppose that we have measured the temperature at one-minute intervals over the course of a day. We then have a *point function*: given a point in time to the nearest minute, the function returns the temperature at that time. If you ask, “When was the hottest part of the day?”, you can answer the question simply by plotting the function and seeing where it is a maximum, say at 3:35 PM. It doesn’t matter if you measured the temperature in Celsius, Fahrenheit, or Kelvin, the time of the temperature maximum will still be 3:35 PM.

Now consider the question, “Where is the maximum of the Sun’s output?” The answer is not so simple, because the Sun’s spectral output, or that of a blackbody, is not a point function. Spectra are *density functions* or *distributions*. That is, they show the distribution of energy or photon numbers *per unit wavelength*, *temporal frequency*, or *wavenumber interval*. If you change the “per unit” spectral variable, you change the overall shape of the function so that it will have its maximum at a different location.

Figure 2.20 shows plots of blackbody energy and quantum irradiances (the red curves) as functions of wavelength λ , temporal frequency f , and wavenumber ν . The spectra are

for a temperature of 5782 K, reduced by the distance of the Earth from the Sun as in Eq. (2.4); the formulas come from Table 2.1. The blue lines are the solar irradiances at the top of the atmosphere. These spectra are the ASTM spectrum of Fig 2.16, converted to photon radiances and to temporal frequency or wavenumber. There is a fairly good overall fit between the Sun’s irradiance and a blackbody spectrum T 5782 K, at least if you ignore the Fraunhofer absorption lines in the solar spectrum and do not look too closely in the ultraviolet.

As seen in the upper left plot, for energy irradiance per unit of wavelength, $E_d(\lambda)$, the peak of the blackbody spectrum is at $\lambda_{\max} = 501$ nm for a temperature of 5782 K. However, the upper right plot for the quantum irradiance $Q_d(\lambda)$ has its maximum at 635 nm. Thus the Sun’s energy “maximum” is at a different wavelength than its peak output of the number of photons. The middle row of plots shows the energy and quantum spectra with temporal frequency f as the variable. The energy spectrum $E_d(f)$ peaks at $f_{\max} = 3.4 \cdot 10^{14}$ Hz, which corresponds to a wavelength of $\lambda_{\max} = c / f_{\max} = 882$ nm. The photon spectrum $Q_d(f)$ in the right plot of the middle row has $f_{\max} = 1.92 \cdot 10^{14}$ Hz, which corresponds to a wavelength of $\lambda_{\max} = 1563$ nm. Thus we get still other wavelengths of “maximum” output if we use temporal frequency as the spectral variable. The bottom row of the figure shows the corresponding plots for wavenumber ν as the spectral variable. The difference in f and ν is just a rescaling by a factor of c , so the corresponding values of λ_{\max} are the same as for temporal frequency f .

In other words, for $T = 5782$ K, *when plotted as a function of wavelength, the energy irradiance is a maximum near 501 nm, in the visible, whereas the maximum is at 882 nm, in the near infrared, when the spectrum is plotted as a function of frequency or wavenumber.* This occurs because the relationship between wavelength and frequency is not linear, so that a unit wavelength interval corresponds to a different size of frequency interval for each wavelength: $|df| = |c/\lambda^2 d\lambda|$. The wavelength maximum of the photon irradiance is at 635 nm for wavelength as the spectral variable, and at 1563 nm when temporal frequency or wavenumber is the spectral variable.

Thus the answer to the question, “Where is the Sun’s output a maximum?” is, without much exaggeration, “Wherever you would like it to be: just pick whether you want to measure output as energy or numbers of photons and pick an appropriate spectral variable.” *Density functions in general do not have unique maxima.*

There are no doubt good evolutionary reasons why eyes see best at visible wavelengths and why plants are green. However, those reasons are most likely related to wavelengths where water is most transparent (a point function, as is the response function of the human eye), and to the thermodynamic instability of molecules that could be used as receptors for infrared radiation in either eyes or plants. The apparent correlation between the wavelengths where we see best and solar maximum output is a false one resulting simply from a choice of spectral variable for how most people plot solar spectral irradiance. Interestingly, no one seems to argue that eyes should have evolved to take advantage of the maximum of solar photon irradiance in the infrared when plotted as a function of frequency, even though eye pigments absorb individual photons based on the frequencies corresponding to transitions between “molecular energy levels.”

The delightful paper by [Soffer and Lynch \(1999\)](#) goes into more detail on these matters and cites various papers that have misunderstood the nature of spectral density functions in the context of vision and evolution.

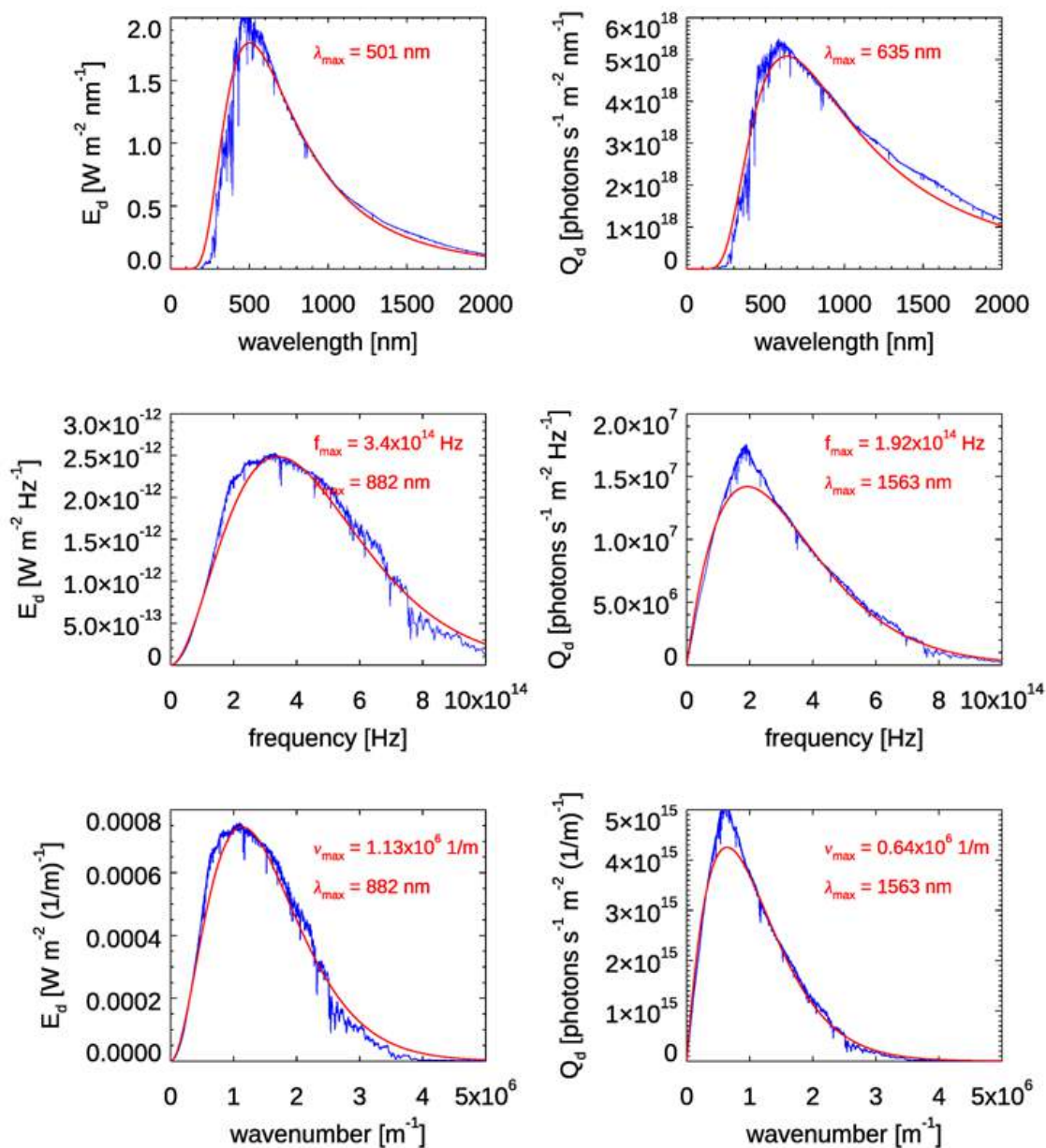
TOA Solar and Blackbody Irradiance Spectra at $T = 5782$ K

Figure 2.20: Solar energy and photon irradiances at the top of the atmosphere as functions of wavelength, temporal frequency, and wavenumber (blue curves). The red curves are the corresponding “best-fit” blackbody spectra for $T = 5782$ K. The insets give the maxima of the blackbody curves, and the corresponding wavelengths for the frequency and wavenumber spectra.

2.6 Bioluminescence

Sunlight is obviously the main source of light in the upper ocean, but there is no sunlight at night or below the 1000 meter limit of the “twilight zone.” However, fish that live in even the deepest parts of the ocean have eyes, and often very sophisticated ones. This suggests that there must be at least some light in the deepest parts of the ocean, even if the light is too faint to be seen by human eyes. And indeed there is. This light comes from two sources: light emitted by living organisms and visible radiation emitted by the decay of naturally occurring potassium.

This section discusses bioluminescence, which refers to the ability of an organism to generate light by chemical reactions. The next section discusses Cherenkov radiation, which is light generated by radioactive decay.

Deep in terrestrial caves, where sunlight never reaches, there are species of animals that have completely lost their sight, even to the point of having no eyes at all. These include fish, salamanders, arthropods, and crustaceans⁴. Eyes are energetically expensive to grow and image processing can require a considerable part of an animal’s brain, so it is not surprising that animals that live in lightless environments have lost their eyes over countless generations and have evolved to use other senses.

On the other hand, light-emitting marine organisms are ubiquitous in the world’s oceans. They range in size from bacteria to fish, and they are found from the sea surface to its bottom and from polar waters to the equator. The many possible ecological roles for self-emitted light include communication for courtship or schooling, attraction of prey, escape from predators, and camouflage by counterillumination. See <https://biolum.eemb.ucsb.edu/functions.html> and Haddock et al. (2010) for a discussion of the possible roles of bioluminescence in the ocean.

Bioluminescence also occurs in terrestrial organisms from fungi to fireflies, but is known to occur in only one freshwater species (*Latia neritoides*, a mollusk found only on the North Island of New Zealand). There does not appear to be any species that spends its entire life in terrestrial caves that bioluminesces. The famous “glowworms” found in some caves are the larval stage of several types of fungus gnats, which leave the caves as adults. It thus seems that the true cave dwellers have gone the route of having no eyes at all, whereas the denizens of the deep ocean have evolved to generate their own light. It is estimated that 75% of all marine organisms can generate light. Several reviews have surveyed the biological, chemical, and ecological aspects of marine bioluminescence (Widder, 2010; Haddock et al., 2010; Widder, 2002). Widder (2006) gives a short history of research on bioluminescence.

Widder et al. (1983) examined the shapes of the emission spectra for 70 marine species ranging from bacteria to fish. Figure 2.21 illustrates the range of shapes encountered in their study. Their instrument counted photons detected by a linear CCD array for 1024 wavelength bins. These spectra thus describe the distribution of numbers of photons emitted. For each photon spectrum there is an associated energy spectrum as discussed in Section 2.6.1. The wavelengths of maximum emission, λ_{\max} , and the full-width at

⁴A few examples from around the world are as follows. Fish: the Mexican Blind Cave Fish, *Astyanax mexicanus*. Salamanders: the Slovenian olm, *Proteus anguinus*; and the Texas Blind Salamander, *Eurycea rathbuni*. Arthropods: the Kauai Cave Wolf Spider, *Adelocosa anopsthe*; the Narrow-necked Blind Cave Beetle, *Leptodirus hochenwartii*, also found in Slovenia; and *Trechinae* beetles of the genus *Dongodytes* in China. Crustaceans: the Blind Albino Cave Crab, *Munidopsis polymorpha*, of the Canary Islands; and the venomous, centipede-like remipedes of the globally distributed Class *Remipedia*.

half-maximum of the emission band, $\Delta\lambda_{\text{FWHM}}$, are indicated in the figure caption. The range of λ_{max} was 439 nm to 574 nm, with an average λ_{max} over all species of 483 nm. The $\Delta\lambda_{\text{FWHM}}$ bandwidths ranged from 26 nm to 100 nm, with an average of about 75 nm. Not surprisingly, these wavelength bands usually coincide with the wavelengths where sea water is most transparent.

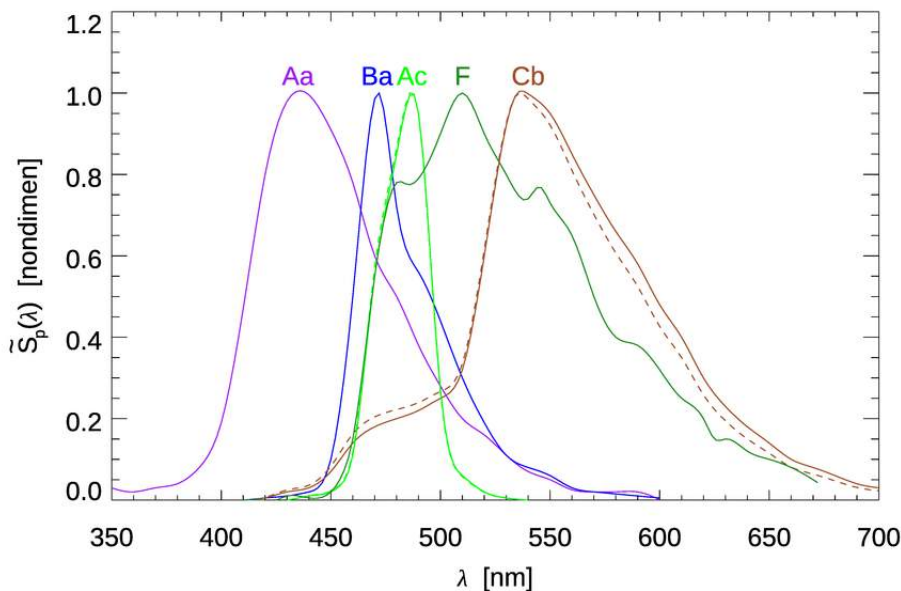


Figure 2.21: Normalized shapes $\tilde{S}_p(\lambda)$ of bioluminescence photon spectra for selected marine organisms: Aa, the arthropod *Scine rattrayi*, $\lambda_{\text{max}} = 439$ nm, $\Delta\lambda_{\text{FWHM}} = 70$ nm; Ba, the dinoflagellate *Pyrocystis noctiluca*, $\lambda_{\text{max}} = 472$ nm, $\Delta\lambda_{\text{FWHM}} = 35$ nm; Ac, the fish *Argyropelecus affinis*, $\lambda_{\text{max}} = 487$ nm, $\Delta\lambda_{\text{FWHM}} = 26$ nm; F, the brittle star *Ophiopholis longispina*, $\lambda_{\text{max}} = 512$ nm, $\Delta\lambda_{\text{FWHM}} = 102$ nm; Cb, the bacterium *Vibrio fischeri* Y-1 strain, $\lambda_{\text{max}} = 540$ nm, $\Delta\lambda_{\text{FWHM}} = 81$ nm. The dashed line for curve Cb shows the shape of the corresponding energy spectrum. [Photon spectra are redrawn from Fig 4 of [Widder et al. \(1983\)](#); the spectrum labels correspond to that figure.]

The magnitude of emission shows more variability than its spectral shape. We first note that most bioluminescent organisms emit light only when they are disturbed. The most common light-inducing disturbance is mechanical stimulation, as when an organism is entrained into a ship's turbulent wake or feels a pressure wave caused, perhaps, by an approaching fish. Flashing lights, electrical fields, earthquakes, and chemical irritants have been known to induce bioluminescence. [Lapota et al. \(1986\)](#) give an interesting description of bioluminescence induced by an ordinary flashlight. When disturbed, organisms emit a flash of light that may last from tens of milliseconds to several seconds. The exception to this statement is certain bacteria, which are able to emit light continuously.

A common measure of the strength of bioluminescence is the number of photons emitted per second by a disturbed organism. Typical values are 10^4 photons s^{-1} cell $^{-1}$ for bacteria, and 10^9 to 10^{11} photons s^{-1} cell $^{-1}$ for dinoflagellates; [Bachelder and Swift \(1989\)](#) give an average value of $5 \cdot 10^{10}$ photons s^{-1} cell $^{-1}$ for the dinoflagellate *Pyrocystis noctiluca*.

See [Lynch III \(1978\)](#) for a tabulation of emission strength for 58 marine species.

A quick calculation gives a feeling for the light levels that are possible. For the dinoflagellate spectrum Ba of [Fig. 2.21](#), the average wavelength is

$$\langle \lambda \rangle = \frac{\int_0^\infty \lambda S_p(\lambda) d\lambda}{\int_0^\infty S_p(\lambda) d\lambda} = 487 \text{ nm} . \quad (2.13)$$

Assuming a typical concentration of 2000 cells m^{-3} for a bioluminescent dinoflagellate, with each cell emitting $5 \cdot 10^{10}$ photons s^{-1} , with an average wavelength of 487 nm, and converting the photon count to energy units via [Eq. \(1.1\)](#), gives

$$\begin{aligned} S_o &= (2 \cdot 10^3 \text{ m}^{-3})(5 \cdot 10^{10} \text{ s}^{-1}) \frac{(6.63 \cdot 10^{-34} \text{ J s})(3 \cdot 10^8 \text{ m s}^{-1})}{487 \cdot 10^{-9} \text{ m}} \\ &= 4.1 \cdot 10^{-5} \text{ W m}^{-3} . \end{aligned}$$

If this power is generated in a sphere of volume 1 m^3 (for example, in a turbulent eddy), then the surface of the sphere receives an irradiance of order 10^{-5} W m^{-2} . [Reference to Table 2.3](#) shows this broad-band irradiance to be greater than that of a clear, starry night, but much less than that of a bright, moonlit night.

2.6.1 Modeling Bioluminescence in Radiative Transfer Calculations

Although disturbed organisms can provide enchanting light shows, some of the most spectacular displays are caused by bacteria near the sea surface. The large numbers of bacteria and their ability to emit light continuously more than make up for their low photon emittance per cell. The Indian Ocean and Arabian Sea are known for their horizon-to-horizon displays of luminous “milky seas,” which can generate enough light for reading on the deck of an otherwise dark ship. Mariners’ logs contain many such reports; see [Kelly and Tett \(1978\)](#) for examples. [Lapota et al. \(1988\)](#) give quantitative observations of a milky sea event. Milky seas can cover as much as $100,000 \text{ km}^2$ and have even been detected from space ([Miller et al., 2005, 2021](#)).

Horizontally uniform milky seas are a one-dimensional geometry, so they can be simulated using HydroLight. A simulation is worthwhile as an example of constructing an internal source term for the radiative transfer equation.

Bioluminescence is included in the scalar radiative transfer equation ([9.21](#)) via a source term $S(z, \theta, \phi, \lambda)$, which has units of $\text{W m}^{-3} \text{ sr}^{-1} \text{ nm}^{-1}$. This function gives the strength of the bioluminescence as a function of depth, direction, and wavelength. Small organisms such as bacteria or dinoflagellates can be assumed to emit light isotropically, in which case S can be written as

$$S(z, \theta, \phi, \lambda) = \frac{S_o(z, \lambda)}{4\pi} , \quad (2.14)$$

where $S_o(z, \lambda)$ gives the emitted spectral power in units of $\text{W m}^{-3} \text{ nm}^{-1}$.

As noted above, the spectra of [Fig. 2.21](#) are for the number of photons emitted, denoted by \tilde{S}_p (subscript “p” for photon). By [Eq. \(1.1\)](#), the energy of a photon of wavelength λ is hc/λ . Thus the corresponding energy spectrum is given by

$$S_o(\lambda) = S_p(\lambda) \frac{hc}{\lambda} . \quad (2.15)$$

We can write S_p as the product of a maximum value and a normalized spectral shape:

$$S_p(z, \lambda) = S_{\max}(z) \tilde{S}_p(\lambda).$$

Lapota et al. (1988) measured values of $1\text{--}5 \cdot 10^{14}$ photons $\text{s}^{-1} \text{m}^{-3}$ during periods of high bioluminescence in the Arabian Sea. Let us take the high value to set the magnitude and use the spectral shape of curve Cb in Fig. 2.21 for $\tilde{S}_p(\lambda)$. The integral of this $\tilde{S}_p(\lambda)$ is

$$\int_0^\infty \tilde{S}_p(\lambda) d\lambda = 98 \text{ nm}.$$

Dividing $5 \cdot 10^{14}$ photons $\text{s}^{-1} \text{m}^{-3}$ by 98 nm sets the magnitude:

$$S_{\max} = \frac{5 \cdot 10^{14}}{98} = 5.1 \cdot 10^{12} \quad \left[\frac{\text{photons}}{\text{s m}^3 \text{ nm}} \right].$$

The resulting $S_p(\lambda) = 5.1 \cdot 10^{12} \tilde{S}_p(\lambda)$ then integrates to give back $5 \cdot 10^{14}$ photons $\text{s}^{-1} \text{m}^{-3}$ as required to describe the number of photons emitted over all wavelengths.

The corresponding energy spectrum as computed by Eq. (2.15) is plotted in Fig. 2.22. The dashed curve in Fig. 2.21 shows the shape of this $S_o(\lambda)$ compared to the shape of the photon spectrum $\tilde{S}_p(\lambda)$. Because photons with shorter wavelengths carry more energy than those with longer wavelengths, the energy spectrum is shifted towards the blue. In the present case, the normalized photon and energy spectral shapes differ by as much as 31% (near 700 nm).

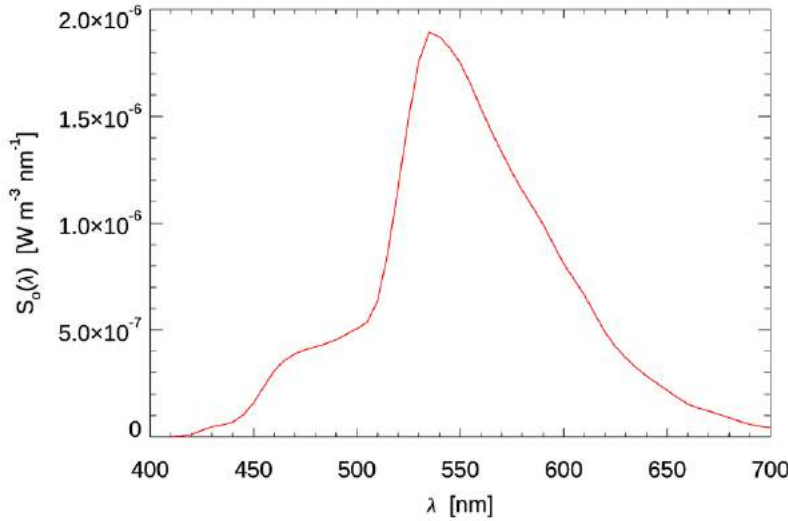


Figure 2.22: The energy spectrum $S_o(\lambda)$ of the bacterium *Vibrio fischeri* Y-1 strain, modeled as described in the text.

This energy spectrum is now used in Eq. (2.14) to define the source function. For the present simulation, it is assumed that the water column is homogeneous⁵. The satellite

⁵There is very little in situ data on milky seas. The value of $5 \cdot 10^{14}$ photons $\text{s}^{-1} \text{m}^{-3}$ measured by Lapota et al. (1988) came from an intake below the ship's hull at 3 m below the sea surface. Miller et al. (2021) point out "...many mariner accounts of the uniform glow persisting under wind-roughened seas (where slicks would break up), light emanating from below the surface even as bucket samples of water were drawn, and the lack of a dark ship wake (which would disrupt a slick)." Thus, for the present simulation, the source term is taken to be constant with depth.

observations in [Miller et al. \(2021\)](#) are in waters with chlorophyll concentrations of 0.5 to 1.5 mg Chl m⁻³, so a value of 1.0 mg Chl m⁻³ was used in HydroLight’s “new Case 1” IOP model. The sky was totally black, so the only light source was the bioluminescence. The resulting HydroLight simulation gives an in-air illuminance of $E_{\text{uv}} = 0.12$ lux. Table 16.2 shows this to be almost as large as the illuminance of a full moon in a clear sky. Thus, just as the mariners claim, you can indeed read by the light of a milky sea.

2.7 Cherenkov Radiation

Cherenkov radiation ([Cherenkov, 1934](#)) is electromagnetic radiation emitted by charged particles traveling faster than the phase speed of light in a dielectric medium such as water⁶. It can be qualitatively thought of as the optical equivalent of the acoustic shock wave (“sonic boom”) generated by an airplane flying faster than the speed of sound in air. Cherenkov radiation is the cause of the beautiful blue glow around the core of a water-cooled nuclear reactor as shown in Fig. 2.23 .



Figure 2.23: Photograph of Cherenkov radiation in a nuclear reactor core. [Public domain photograph from the U. S. Nuclear Regulatory Commission]

[Bradner et al. \(1987\)](#) examined physical sources of light in the ocean. They estimated that near the sea surface, Cherenkov radiation from cosmic rays generates a photon irradiance of order 10^7 photons s⁻¹ m⁻²; this value decreases exponentially with depth with an e-folding distance of about 1 km. However, another source of Cherenkov radiation is due to the decay of radioactive potassium-40, which is distributed uniformly throughout the oceans as part of the dissolved salts that make up salinity.

The Earth’s crust contains potassium at a concentration of about 2.6% by mass. This potassium occurs as three isotopes: stable ³⁹K (93.2581%) and ⁴¹K (6.7302%), and radioactive ⁴⁰K (0.0117%). ⁴⁰K decays 89.1% of the time to ⁴⁰Ca by emission of an electron and an electron anti-neutrino (beta decay), and 10.9% of the time to ⁴⁰Ar by capture of

⁶The emission of light by a charged particle moving faster than the speed of light in a medium was first predicted by Oliver Heavyside in a series of papers starting in 1888. Arnold Sommerfeld independently predicted the effect in 1904, and Marie Curie observed the glow of light in radium solutions. However, these earlier results were unappreciated and forgotten until Cherenkov’s observations in the 1930s.

an inner-shell electron, followed by emission of a gamma ray and a neutrino⁷. The half-life of ⁴⁰K is 1.25 Gy.

When ⁴⁰K decays to ⁴⁰Ca, the emitted electron and anti-neutrino carry a combined kinetic energy of 1.31 MeV or $2.1 \cdot 10^{-13}$ J. There is a continuous distribution of the electron kinetic energy⁸ ranging from 0 to a maximum of 1.31 MeV, with the peak at about 0.55 MeV (Kelly et al., 1959). The associated speed of the electron for a given kinetic energy can be obtained from the formula for relativistic kinetic energy (Halliday and Resnick, 1988, Section 42-14):

$$KE = m_o c^2 \left(\frac{1}{\sqrt{1 - (v/c)^2}} - 1 \right),$$

where m_o is the rest mass of the particle, c is the speed of light, and v is the speed of the particle. In relativity theory, it is customary to let $\beta = v/c$ be the speed of a particle relative to the speed of light in a vacuum. Solving this equation for $\beta^2 = (v/c)^2$ gives

$$\beta^2 = 1 - \left(\frac{KE}{m_o c^2} + 1 \right)^{-2}. \quad (2.16)$$

Using the values in Table 2.4 gives $\beta^2 = 0.918$, or $v = 0.958c$, for a 1.31 MeV electron. The phase speed of light in water is c/n , which is approximately $0.75c$ at visible wavelengths. Thus the emitted electron is traveling faster than the speed of light in water and will therefore emit Cherenkov radiation until the electron slows down to less than c/n through loss of radiated energy and other interactions with the water.

Symbol	Quantity	Value
m_o	rest mass of the electron	$9.109 \cdot 10^{-31}$ kg
e	charge of the electron	$1.602 \cdot 10^{-19}$ C
c	speed of light in vacuo	$2.998 \cdot 10^8$ m s ⁻¹
h	Planck constant	$6.626 \cdot 10^{-34}$ J s
μ_o	magnetic permeability of free space	$1.2566 \cdot 10^{-6}$ N s ² C ⁻²
KE	kinetic energy of emitted electron	≤ 1.31 MeV = $2.099 \cdot 10^{-13}$ J
n	real index of refraction of water	see Fig. 2.24

Table 2.4: Quantities needed for Cherenkov radiation calculations of ⁴⁰K decay.

The energy radiated by a single ⁴⁰K electron per unit of distance traveled (x , in meters) and per unit angular frequency (ω , in radians per second) of the emitted light is given by

⁷Earth's atmosphere is 0.94% argon, of which 99.6% is ⁴⁰Ar. Spectroscopy shows that the argon in stars is 85% ³⁶Ar, which is created by fusion of two alpha particles (Helium nuclei) with one silicon-32 nucleus during supernova explosions, and 15% is ³⁸Ar. It is thus thought that the argon in Earth's atmosphere has accumulated up over billions of years from the decay of ⁴⁰K.

⁸It is a characteristic of beta decay that the emitted electrons have a continuous spectrum of energies from 0 to some maximum. This contrasts with alpha decay, in which the emitted alpha particles have a single energy determined by the quantized energy levels of the decaying nucleus.

the celebrated formula⁹ of Frank and Tamm (1937):

$$\frac{\partial^2 E}{\partial x \partial \omega} = \frac{1}{4\pi} e^2 \mu_w(\omega) \omega \left(1 - \frac{1}{\beta^2 n^2(\omega)} \right) \left[\frac{\text{J}}{\text{m s}^{-1}} \right]. \quad (2.17)$$

Here $\mu_w(\omega)$ is the frequency-dependent magnetic permeability of the medium. Using $\omega = 2\pi c/\lambda$ to convert the Frank and Tamm formula to energy emitted per unit distance per unit wavelength gives

$$\frac{\partial^2 E}{\partial x \partial \lambda} = \pi c^2 e^2 \mu_w(\lambda) \frac{1}{\lambda^3} \left(1 - \frac{1}{\beta^2 n^2(\lambda)} \right) \left[\frac{\text{J}}{\text{m m}} \right]. \quad (2.18)$$

Converting this formula from energy emitted to number N of photons emitted via $E = Nhc/\lambda$ gives

$$\frac{\partial^2 N}{\partial x \partial \lambda} = \frac{\pi c}{h} e^2 \mu_w(\lambda) \frac{1}{\lambda^2} \left(1 - \frac{1}{\beta^2 n^2(\lambda)} \right) \left[\frac{\text{photons}}{\text{m m}} \right]. \quad (2.19)$$

The magnetic permeability μ_w is a function of frequency (or wavelength), but for water its value equals that of a vacuum, μ_o , the permeability of free space, to within a 8 parts per million. Therefore we can replace μ_w with μ_o in these equations with a negligible error. The observant reader will then note that

$$\alpha = \frac{c e^2 \mu_o}{2\pi h} = 0.007297 \approx \frac{1}{137}$$

is the dimensionless fine-structure constant of quantum theory. The last equation therefore can be succinctly written as

$$\frac{\partial^2 N}{\partial x \partial \lambda} = 2\pi\alpha \frac{1}{\lambda^2} \left(1 - \frac{1}{\beta^2 n^2(\lambda)} \right). \quad (2.20)$$

In these formulas, x and λ are in meters.

These formulas show a remarkable feature of Cherenkov radiation, namely that the emission is broad-band and increases rapidly going from visible to ultraviolet (UV) wavelengths. As seen in Fig. 2.24, the real index of refraction of water decreases rapidly between 130 nm, where $n \approx 1.63$, and 71 nm, where n drops to less than 1¹⁰. This gives a sharp radiation cut-off at wavelengths less than about 100 nm because as n approaches 1, the speed of light in water approaches the speed in vacuo, in which case the electron is always traveling slower than light in the water, and there is no emitted radiation. For infrared and longer wavelengths, the emission is small and goes to zero as the wavelength increases. The red curves in Fig. 2.25 show Eqs. (2.18) and (2.20) using the wavelength-dependent $n(\lambda)$ seen in Fig. 2.24 for the initial electron energy of 1.31 MeV. For wavelengths in the visible range, Table 8.2 shows that $n \approx 1.36$ even for the extreme case of cold (0 deg C), saline (35 PSU), deep-ocean (depth of 10,000 m) water, compared to about 1.34 for pure water at atmospheric pressure. This difference would have only a small effect on the spectra plotted in Fig. 2.25, so these curves for pure water are representative of all parts of the ocean.

⁹Although the end result is simple, the derivation of Eq. (2.17) starting from Maxwell's equations is quite difficult (e.g. Jackson, 1962, Sections 13.4 and 14.9). Pavel Cherenkov, Ilya Frank, and Igor Tamm shared the 1958 Nobel Prize in Physics "for the discovery and the interpretation of the Cherenkov effect."

¹⁰The discussion in Section 11.4.3 explains that although $n < 1$ gives the *phase* speed c/n greater than the speed of light *in vacuo*, this is not a violation of special relativity

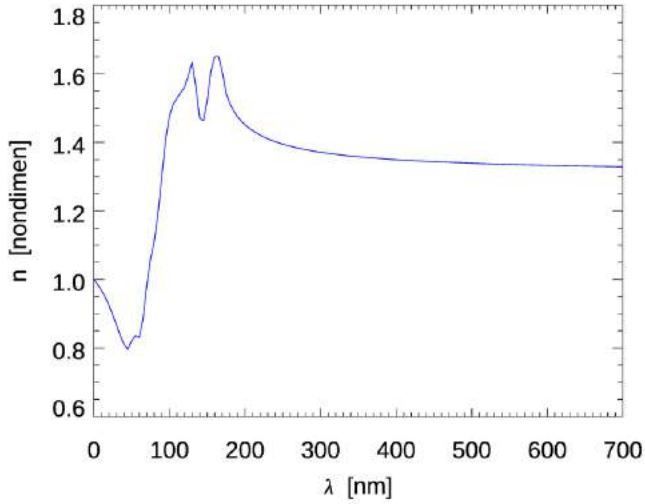


Figure 2.24: The real index of refraction of pure water for wavelengths between 0 and 700 nm (redrawn from Fig. 8.4).

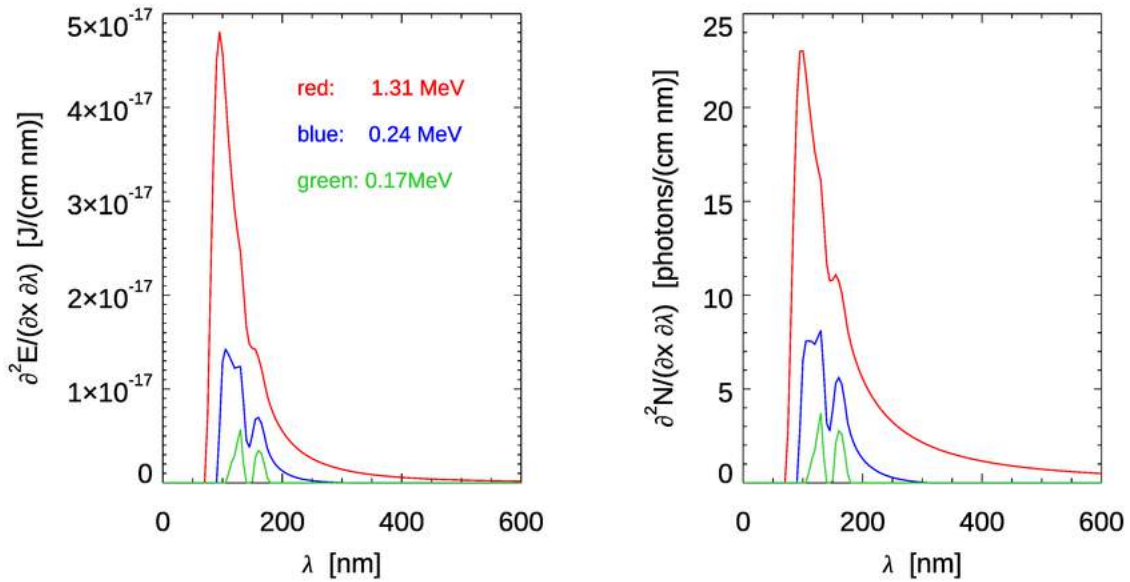


Figure 2.25: Cherenkov radiation spectra for a single ^{40}K electron in optically pure water. The left panel shows the emitted energy per centimeter of travel and per nanometer of wavelength. The right panel shows the corresponding spectra for the number of emitted photons. The red curves are for the initial electron with an energy of 1.31 MeV. The blue curves are for an energy of 0.27 MeV, below which there is no more emission at wavelengths longer than 300 nm. The green curves are for an energy of 0.17 MeV, $v^2/c^2 = 0.434$, for which there is only a small amount of emission at the ultraviolet wavelengths where the index of refraction greater than 1.52.

Numerically integrating Eq. (2.18) over wavelength, for a given value of β , gives the energy emitted over all wavelengths per unit distance traveled. Further integrating over distance gives the total energy emitted. Corresponding integrations of Eq. (2.20) give the numbers of emitted photons. As the electron travels through the water, its kinetic energy decreases. When the kinetic energy decreases to a value such that the electron's speed, given by Eq. (2.16), results in $v = c/n(\lambda)$, photon emission ceases for that wavelength. After the kinetic energy has decreased to 0.24 MeV, $\beta^2 = 0.533$. This gives $v < c/n(\lambda)$ for $n(\lambda) \leq 1.37$. There is thus no more radiation for wavelengths greater than 300 nm, where $n \leq 1.37$. The spectra at this energy are shown in blue in Fig. 2.25. The green curves in the figure are for an energy of 0.17 MeV, $\beta^2 = 0.434$, for which there is only a small amount of emission at the ultraviolet wavelengths where the index of refraction greater than 1.52. To compute the total amount of Cherenkov radiation, the above integrations over wavelength and distance must be repeated for each energy of the distribution of energies of the emitted electrons; thus there is a triple integration over energy, wavelength, and distance.

Equations (2.17)-(2.20) are correct in that they give the energy or number of photons emitted as Cherenkov radiation for a given electron energy. What they do not tell you is that less than one percent of the kinetic energy of of an electron emitted by a ^{40}K nucleus results in Cherenkov radiation. Almost all of the electron's energy goes into ionizing water molecules as the electron travels through the water. This loss of energy to ionization is given by an equation known as the Bethe-Bloch formula¹¹ Discussion of that calculation takes us beyond the needs of optical oceanography and will not be given here because the end result has been calculated by the physicists who use Cherenkov radiation for the detection of neutrinos in the deep ocean.

When a neutrino interacts with matter (which is extremely rare), the result can be a charged particle such as an electron or muon traveling in almost the same direction as the neutrino. Those charged particles also cause Cherenkov radiation, which can be detected as a function of time and direction and used to determine the direction and energy of the initial neutrino. Several neutrino detectors based on this idea have been built at the bottom of the ocean and deep within the ice at the geographic South Pole¹². These detectors are arrays of thousands of photomultiplier tubes (PMTs), occupying as much as a cubic kilometer of space, that track the movement of the Cherenkov "light cone" as the particle travels through the detector. In deep-ocean measurements, Cherenkov radiation from ^{40}K decay is background noise imposed on the signal of interest. The magnitude of this Cherenkov background therefore has been carefully calculated and measured as part of the deep-ocean neutrino detector designs.

Use of the above formulas to compute the number of Cherenkov photons per square meter per second in the ocean must account for number of decaying ^{40}K atoms per cubic meter (about $12000 \text{ m}^{-3} \text{ s}^{-1}$, of which 89.1% result in electron emission) and the clarity of the water (usually close to that of pure water in the deep ocean). Calculations predict

¹¹For a derivation and discussion see Arya (1966, Chapter VII, Section 5). The Bethe-Bloch formula involves two additional parameters: the energy required to ionize a water molecule and the density of electrons in water (the number of water-molecule electrons per cubic meter). My evaluation of the Bethe-Bloch formula shows that for a 1.31 MeV electron, the energy lost to ionization is 153 times that lost to Cherenkov radiation.

¹²To learn more, the keywords to search for are DUMAND (Deep Underwater Muon and Neutrino Detector; 1976-1995) and ANTARES (Astronomy with a Neutrino Telescope and Abyss environmental RESearch project; operational since 2008) in the ocean. AMANDA (Antarctic Muon And Neutrino Detector Array) and IceCube Neutrino Observatory are at the South Pole.

about $1.2 \cdot 10^6$ photons $\text{m}^{-2} \text{s}^{-1}$ at visible wavelengths (400-700 nm) in clear ocean water (Learned, 1981; Bradner et al., 1987). Nighttime measurements to depths of 4300 m in clear waters near Hawaii showed a typical background “glow” of order 10^7 photons $\text{m}^{-2} \text{s}^{-1}$, which includes bioluminescence as well as Cherenkov radiation. Aoki et al. (1986) report $2.18_{-28\%}^{+9\%} \cdot 10^6$ photons $\text{m}^{-2} \text{s}^{-1}$ at the same location. Tamburini et al. (2013) recorded a 2.5 year times series of light at 2,500 m depth in the Mediterranean Sea. Each of their photomultiplier tubes (PMTs) recorded a steady background of $37,000 \pm 3000$ photons s^{-1} attributable to ^{40}K decay, plus another $40,000 \pm 3000$ photons s^{-1} attributable to background bioluminescence. During periods of strong currents, which triggered bioluminescence in the wakes of the PMTs, the bioluminescence signal increased by as much as a factor of 100; Fig. 2.26 shows some of their data. Their PMTs¹³ had a collection area of 0.038 m^2 , so $37,000$ photons s^{-1} received by a PMT corresponds to about 10^6 photons $\text{m}^{-2} \text{s}^{-1}$, consistent with the predictions in the other papers. The measured total of $67,000$ photons s^{-1} for Cherenkov and background bioluminescence corresponds to $1.8 \cdot 10^6$ photons $\text{m}^{-2} \text{s}^{-1}$. Thus it is never completely dark at even the greatest depths, even though no solar photons are present.

Most vertebrates, humans in particular, see only black-and-white in dim light using the rod cells in their retinas. (Cone cells are used for color vision in bright light; see Section 16.1.) The optically sensitive part of these rod cells is a single kind of photopigment known as an opsin (generally rhodopsin-1). A recent genetic study of 101 species of deep-sea fish (Musilova et al., 2019) found that some fish express genes for coding multiple types of opsins, as many as 2 cone and 38 rod opsins in one species. The peaks of the wavelength sensitivities of these multiple opsins cover a range of blue to near-UV wavelengths. Thus it is hypothesized that some fish may be able to “see color” even in the dimmest light of the deep ocean. Giant squid have the largest eyes of any animal, around 30 cm in diameter. It is thought that they may be able to detect bioluminescence generated by their arch-enemies the sperm whales and know it’s time to leave. It may be that squid and other deep-ocean animals just see blobs of light or shadows against the faint background. If they see a small blob of light/shadow, eat it; if it’s a big blob/shadow, flee. In addition, Frank and Widder (1996) found that the eyes of certain deep-sea crustaceans are equally sensitive to near-UV and blue-to-green visible wavelengths. Their paper discussed the penetration of near-UV solar radiation to depth. However, it can be speculated that these animals may have evolved eyes capable of seeing the low level of Cherenkov ultraviolet light. Assuming an average wavelength of 420 nm, the deep-ocean measurements of $1.8 \cdot 10^6$ photons $\text{m}^{-2} \text{s}^{-1}$ (Tamburini et al., 2013) and $2.2 \cdot 10^6$ photons $\text{m}^{-2} \text{s}^{-1}$ (Aoki et al., 1986) correspond to an irradiance of approximately $10^{-12} \text{ W m}^{-2}$, which is the threshold sensitivity of the eyes of some deep-sea fish as estimated from comparative anatomy studies (Denton and Warren, 1957). This is probably not a coincidence.

¹³The PMTs were Hamamatsu model R7081-20. The data sheet gives a “minimum effective area” of 220 mm. Taking this as the diameter of the PMT collector gives a collector area of 0.038 m^2 . The PMT was sensitive to wavelengths from 300 to 650 nm with peak sensitivity at 420 nm.

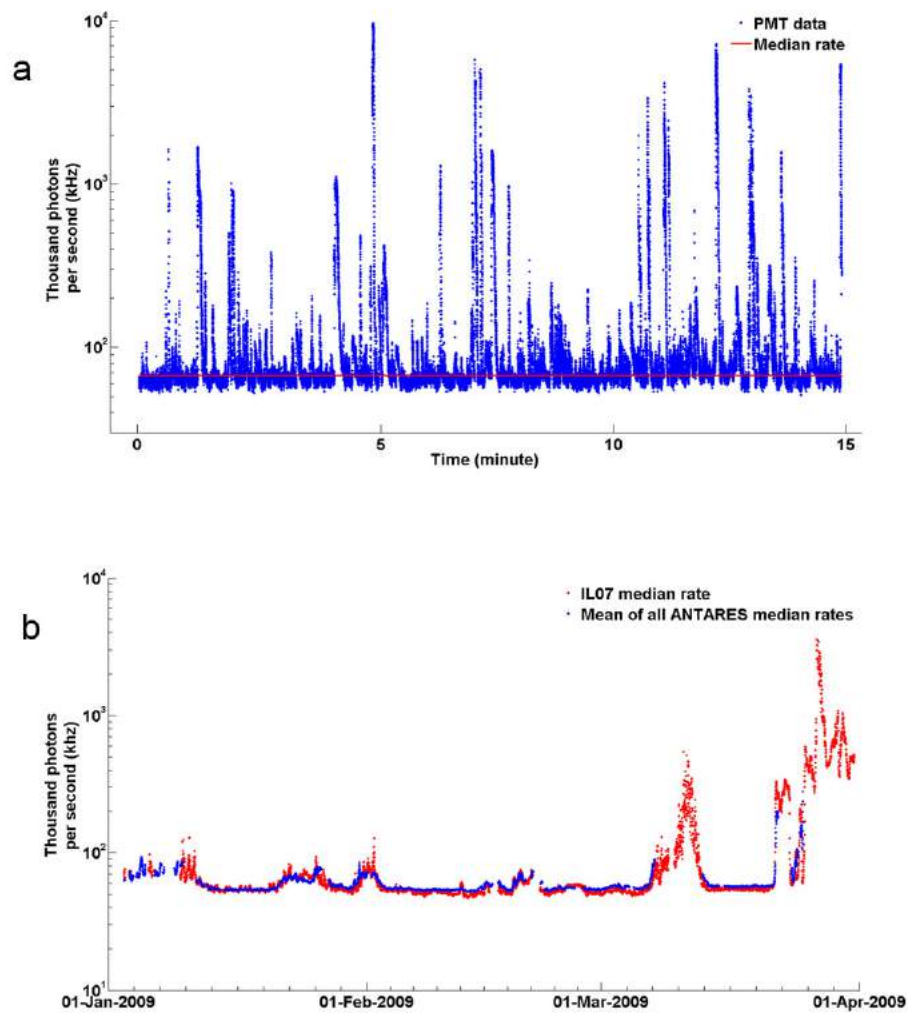


Figure 2.26: Measured bioluminescence and Cherenkov radiation photon counts in the Mediterranean Sea at 2500 m depth. Top panel: the photon counts in thousands per second for one PMT over a time of 15 minutes. Bottom panel: median photon counts from one particular PMT (red) and the average of the median rates for the 885 detectors in the array (blue) over a period of 3 months. Reproduced from Fig. S2 of [Tamburini et al. \(2013\)](#) under a Creative Commons Attribution License.

Inherent Optical Properties

The Light and Radiometry chapter 1 explains how to describe *light* quantitatively. This chapter now explains how to describe and measure the optical properties of the *material* through which the light propagates. This is accomplished by the variables in the inherent optical properties box of Fig. 1.

Natural waters, both fresh and saline, are a witch's brew of dissolved and particulate matter. These solutes and particles are both optically significant and highly variable in kind and concentration. Consequently, the optical properties of natural waters show large temporal and spatial variations and seldom resemble those of pure water.

The great variability in the optical properties of natural waters is the bane of those who desire precise and easily tabulated data or simple models. However, the coupling between constituent physical properties (e.g., particle size, shape, index of refraction, and concentration) and optical properties suggests that optical measurements can be used to deduce information about the constituents of aquatic ecosystems. Indeed, it is the connections between the optical properties and the biological, chemical and geological constituents of natural waters that define the critical role of optics in aquatic research. For just as optical oceanography utilizes results from the biological, chemical, geological, and physical sub-disciplines of limnology and oceanography, so do those subdisciplines incorporate optics. This synergism is seen in such areas as biological oceanography, marine photochemistry, mixed-layer thermodynamics, lidar bathymetry, underwater visibility, and "ocean color" remote sensing of biological productivity, sediment load, pollutants, or bathymetry and bottom type.

The goal of this chapter is to define the bulk inherent optical properties of natural waters. These are the absorption coefficient, the volume scattering function, and various quantities derived from these two. The chapter closes with an overview of various schemes for classifying natural waters. The reasons why the various optical properties have their observed values are discussed in later chapters. However, the unseverable connections between optics and biology, chemistry, and geology will still be obvious.

When light interacts with matter one of three things can happen. The light can disappear, with its energy being converted to another form such as heat or the energy contained in a chemical bond. This process is called *absorption*. The light can also change its direction without a change of wavelength, and the light can change its wavelength, usually also

with a change of direction. Either of these processes is called *scattering*.

The absorption and scattering properties of a medium such as sea water are described by its *inherent optical properties*, or IOPs. ***IOPs are optical properties of the medium and do not depend on the ambient light field.*** That is, a volume of water has well defined absorption and scattering properties whether or not there is any light there to be absorbed or scattered. This means that IOPs can be measured in the laboratory on a water sample, as well as in situ in the ocean.

3.0.1 Conceptual Definitions of IOPs

The *absorption coefficient* is the fundamental IOP that describes how a medium absorbs light. The *volume scattering function* likewise describes how the medium scatters light. If you know these two IOPs, then you know everything there is to know about the medium interacts with unpolarized light. Other IOPs are sometimes convenient, and they can be defined in terms of the absorption coefficient and the volume scattering function. Figure 3.1 is the key to defining the IOPs.

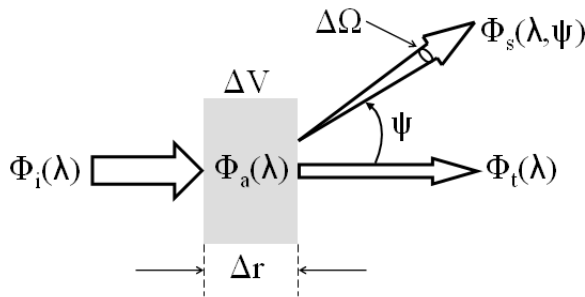


Figure 3.1: Geometry used to define inherent optical properties.

Consider a small volume ΔV of water, of thickness Δr , illuminated by a collimated beam of monochromatic light of spectral radiant power at some wavelength λ , $\Phi_i(\lambda)$ in W nm^{-1} , as illustrated in Fig. 3.1. Some part $\Phi_a(\lambda)$ of the incident power $\Phi_i(\lambda)$ is absorbed within the volume of water. Some part $\Phi_s(\lambda, \psi)$ is scattered out of the beam at an angle ψ , and the remaining power $\Phi_t(\lambda)$ is transmitted through the volume with no change in direction. Let $\Phi_s(\lambda)$ be the total power that is scattered into all directions. Furthermore, assume that the light does not undergo a change in wavelength during the scattering process. Then by conservation of energy,

$$\Phi_i(\lambda) = \Phi_a(\lambda) + \Phi_s(\lambda) + \Phi_t(\lambda).$$

The *absorptance* $A(\lambda)$ is the *fraction* of incident power that is absorbed within the volume:

$$A(\lambda) \triangleq \frac{\Phi_a(\lambda)}{\Phi_i(\lambda)}. \quad (3.1)$$

Likewise, the *scatterance* $B(\lambda)$ is the fractional part of the incident power that is scattered out of the beam into all directions,

$$B(\lambda) \triangleq \frac{\Phi_s(\lambda)}{\Phi_i(\lambda)}, \quad (3.2)$$

and the *transmittance* $T(\lambda)$, the fraction of the incident power that passes through the volume without interacting with the medium, is

$$T(\lambda) \triangleq \frac{\Phi_t(\lambda)}{\Phi_i(\lambda)}. \quad (3.3)$$

Clearly, $A(\lambda) + B(\lambda) + T(\lambda) = 1$.

The inherent optical properties usually employed in optical oceanography are the absorption and scattering *coefficients*, which are respectively the absorptance and scatterance *per unit distance* in the medium. In the geometry of Fig. 3.1, let the thickness Δr approach zero, in which case only a decreasingly small fraction $\Delta A(\lambda)$ of the incident power is absorbed. The *absorption coefficient* $a(\lambda)$ is then defined as

$$a(\lambda) \triangleq \lim_{\Delta r \rightarrow 0} \frac{\Delta A(\lambda)}{\Delta r} = \frac{dA(\lambda)}{dr} \quad (\text{m}^{-1}). \quad (3.4)$$

Likewise the *scattering coefficient* $b(\lambda)$ is defined as

$$b(\lambda) \triangleq \lim_{\Delta r \rightarrow 0} \frac{\Delta B(\lambda)}{\Delta r} = \frac{dB(\lambda)}{dr} \quad (\text{m}^{-1}), \quad (3.5)$$

where $\Delta B(\lambda)$ is the fraction of scattered power.

The *beam attenuation coefficient* $c(\lambda)$ is defined as

$$c(\lambda) \triangleq a(\lambda) + b(\lambda) \quad (\text{m}^{-1}). \quad (3.6)$$

Another commonly used IOP is the *single-scattering albedo* $\omega_o(\lambda)$, defined by

$$\omega_o(\lambda) \triangleq \frac{b(\lambda)}{c(\lambda)}. \quad (3.7)$$

In waters and at wavelengths where the beam attenuation is due primarily to scattering, ω_o is near one. For waters and wavelengths where the beam attenuation is due primarily to absorption, ω_o is near zero. As we shall see elsewhere, the single-scattering albedo is the probability that a light ray will be scattered (rather than absorbed) in any given interaction, hence $\omega_o(\lambda)$ is also known as the *probability of photon survival* in any given interaction between a ray or photon and matter.

Equations (3.4) and (3.5) are conceptual, mathematical definitions in terms of limits. These definitions are not suitable for implementation in measurement instruments, which always have a finite path length r and finite A and B . Section 3.3 on measuring IOPs shows how to recast the conceptual definitions into forms suitable for making measurements.

The IOPs depend on the composition, morphology, and concentration of the particulate and dissolved substances in the ocean. Composition refers to what materials make up the particle or dissolved substance, in particular to the index of refraction of that material relative to that of the surrounding water. Morphology refers to the sizes and shapes of particles. Concentration refers the number of particles in a given volume of water, which is described by the particle size distribution, or to the amount of a dissolved substance in the water. As we will see, different materials absorb much differently as a function of wavelength. Particles with different shapes scatter light differently, even if the particles have the same volume. Particles with different volumes scatter light differently even if they

have the same shape. Consider, for example, atmospheric visibility through rain or fog. Even if the total water content per cubic meter of air is the same, you can “see through” a few large raindrops, whereas many small fog droplets greatly reduce visibility.

Because the physical characteristics of the dissolved and particulate substances in the ocean vary by orders of magnitude, so do the IOPs. For example, both the absorption and scattering coefficients of pure water are less than 0.01 1/m at 440 nm. However, in turbid coastal waters with high concentrations of phytoplankton, mineral particles, and dissolved organic matter, the absorption and scattering coefficients can be four orders of magnitude larger. The volume scattering function (at a given scattering angle and wavelength) can also vary by orders of magnitude between open ocean and coastal waters. Understanding how variability in IOPs is determined by the various constituents of the ocean is a fundamental problem of optical oceanography.

3.0.2 Historical IOP Notation

Just as with the radiometric quantities, there is an “historical” set of symbols in addition to the IAPSO recommended symbols used above. Table 3.1 summarizes the IOP’s as they are commonly employed in optical oceanography.

Quantity	SI Units	Recommended Symbol	Historical Symbol
absorption coefficient	m^{-1}	a	a
scattering coefficient	m^{-1}	b	s
backward scattering coefficient	m^{-1}	b_b	b
forward scattering coefficient	m^{-1}	b_f	f
beam attenuation coefficient	m^{-1}	c	α
volume scattering function	$\text{m}^{-1} \text{sr}^{-1}$	β	p
scattering phase function	sr^{-1}	$\tilde{\beta}$	σ
single-scattering albedo	dimensionless	ω_o	$\rho, \tilde{\omega}$

Table 3.1: Terms, units, and symbols for inherent optical properties.

3.1 The Volume Scattering Function (VSF)

We now take into account the angular distribution of the scattered power, with two assumptions. We first assume that *the medium is isotropic*, i.e., its influence on light is the same in all directions at a given point. This is a reasonable assumption for natural waters in which the particles are randomly oriented by turbulence. We also assume that the incident light is unpolarized. If these two assumptions are true, then the scattering process is *azimuthally symmetric*. This means that the scattering depends only on the *scattering angle* ψ , which is measured from the direction of the unscattered beam as shown in Fig. 3.2. Clearly $0 \leq \psi \leq \pi$.

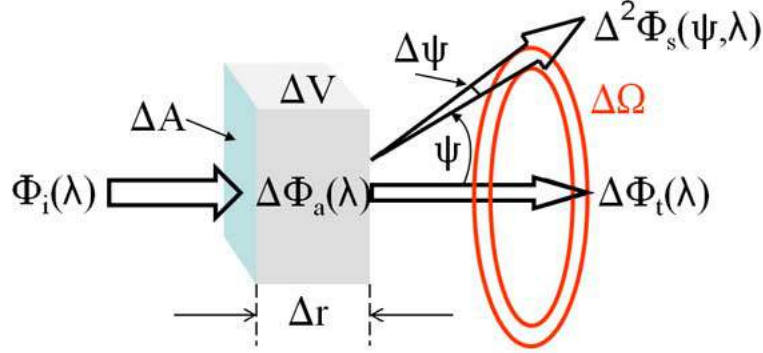


Figure 3.2: Geometry used to define the volume scattering function. $\Phi_i(\lambda)$ is the incident power onto an area ΔA . An amount $\Delta\Phi_a(\lambda)$ of the incident power is absorbed in distance Δr , and $\Delta^2\Phi_s(\psi, \lambda)$ is the power scattered into the annular ring of solid angle $\Delta\Omega$. $\Delta\Phi_t(\lambda)$ is the transmitted power.

3.1.1 Conceptual Definition of the Volume Scattering Function

With these two assumptions, the *fraction* of incident power scattered out of the beam through an angle ψ into a solid angle $\Delta\Omega$ centered on ψ , is $\Delta^2\Phi_s(\psi, \lambda)/\Phi_i(\lambda)$. The notation $\Delta^2\Phi_s(\psi, \lambda)$ is adopted from [Gordon \(1994a\)](#) to indicate that $\Delta^2\Phi_s(\psi, \lambda)$ is a second-order small quantity, that is, it is small because both Δr and $\Delta\Omega$ are small. The solid angle $\Delta\Omega$ now includes all directions within the two red rings shown in the figure, corresponding to all directions between scattering angles ψ and $\psi + \Delta\psi$. The volume scattering function $\beta(\psi, \lambda)$ is defined as the limit of this fraction as $\Delta r \rightarrow 0$ and $\Delta\Omega \rightarrow 0$:

$$\beta(\psi, \lambda) \triangleq \lim_{\Delta r \rightarrow 0} \lim_{\Delta\Omega \rightarrow 0} \frac{\Delta^2\Phi_s(\psi, \lambda)}{\Phi_i(\lambda) \Delta r \Delta\Omega} = \frac{1}{\Phi_i(\lambda)} \frac{\partial^2\Phi_s(\psi, \lambda)}{\partial r \partial\Omega} \quad (\text{m}^{-1} \text{sr}^{-1}). \quad (3.8)$$

Recalling the definition of *spectral radiant intensity* 1.5.6 as scattered power per unit solid angle, the corresponding intensity scattered into the given solid angle $\Delta\Omega$ is $\Delta I_s(\psi, \lambda) = \Delta^2\Phi_s(\psi, \lambda)/\Delta\Omega$. Moreover, if the incident power $\Phi_i(\lambda)$ falls on an area ΔA , then the corresponding incident irradiance is $E_i(\lambda) = \Phi_i(\lambda)/\Delta A$. Noting that $\Delta V = \Delta r \Delta A$ is the volume of water that is illuminated by the incident beam allows Eq. (3.8) to be rewritten as

$$\beta(\psi, \lambda) = \lim_{\Delta V \rightarrow 0} \frac{\Delta I_s(\psi, \lambda)}{E_i(\lambda) \Delta V} = \frac{1}{E_i(\lambda)} \frac{\partial I_s(\psi, \lambda)}{\partial V}. \quad (3.9)$$

This form of $\beta(\psi, \lambda)$ suggests the name *volume scattering function* (commonly abbreviated as VSF) and the physical interpretation of scattered intensity per unit incident irradiance per unit volume of water. In the language of physics, the VSF also can be interpreted as the differential scattering cross section per unit volume.

The forms of Eqs. (3.8) and (3.9) involving partial derivatives are *conceptual* definitions. Instruments that measure the VSF always have to deal with finite path lengths Δr and finite solid angles $\Delta\Omega$. The transition from a definition as a derivative to an instrument design is discussed on Section 3.3 on measuring IOPs.

3.1.2 Scattering Coefficients

Integrating $\beta(\psi, \lambda)$ over all directions (solid angles) gives the total scattered power per unit incident irradiance and unit volume of water, in other words the scattering coefficient:

$$b(\lambda) = \int_{\Xi} \beta(\psi, \lambda) d\Omega = 2\pi \int_0^\pi \beta(\psi, \lambda) \sin \psi d\psi . \quad (3.10)$$

The 2π in the last equation follows from our assumption that the scattering is azimuthally symmetric about the incident direction. This integration is often divided into forward scattering, $0 \leq \psi < \pi/2$, and backward scattering, $\pi/2 \leq \psi \leq \pi$, parts. The corresponding forward and backward¹ scattering coefficients are, respectively,

$$b_f(\lambda) \triangleq 2\pi \int_0^{\pi/2} \beta(\psi, \lambda) \sin \psi d\psi \quad (3.11)$$

$$b_b(\lambda) \triangleq 2\pi \int_{\pi/2}^\pi \beta(\psi, \lambda) \sin \psi d\psi . \quad (3.12)$$

The backscattering fraction, defined by

$$B(\lambda) \triangleq \tilde{b}_b(\lambda) \triangleq \frac{b_b(\lambda)}{b(\lambda)} , \quad (3.13)$$

gives the fraction of scattered light that is deflected through scattering angles greater than 90 deg. This quantity is fundamental to remote sensing because most of the light leaving the ocean in an upward direction comes from sunlight that was originally heading downward, but which was backscattered into upward directions.

3.1.3 The Scattering Phase Function

The volume scattering *phase function*², $\tilde{\beta}(\psi, \lambda)$, is defined by

$$\tilde{\beta}(\psi, \lambda) \triangleq \frac{\beta(\psi, \lambda)}{b(\lambda)} \quad (\text{sr}^{-1}) . \quad (3.14)$$

Writing the VSF $\beta(\psi, \lambda)$ as the product of the scattering coefficient $b(\lambda)$ times the phase function $\tilde{\beta}(\psi, \lambda)$ partitions $\beta(\psi, \lambda)$ into a factor giving the *magnitude* of the total scattering, $b(\lambda)$ with units of m^{-1} , and a factor giving the *angular distribution* of the

¹It would be hard to design worse notation than b_b , in which the same letter “b” stands for both “scattering” and “backward” in the same term. However, we’re stuck with this for historical reasons.

²The term “phase function” has its historical origins in astronomy, where the phase angle is the angle between the directions of the light incident onto and reflected from an object. In the case of sunlight being reflected by the Moon and seen on Earth, this angle depends on the phase of the Moon—hence the name phase function for the function that gives the phase angle as a function of time. In our terminology, the lunar phase angle is π minus the scattering angle. The term phase function is rather misleading because the phase function $\tilde{\beta}(\psi, \lambda)$ defined here has nothing whatsoever to do with the phase of an electromagnetic wave or with its historical origin tied to the phase of the Moon. The Russian literature calls $\tilde{\beta}$ the “scattering diagram,” which is more descriptive of what it really is. You also see the term “indicatrix.” Neither of these terms is commonly used in the English-language literature.

scattered light, $\tilde{\beta}(\psi, \lambda)$, with units of sr^{-1} . Combining Eqs. (3.10) and (3.14) gives the *normalization condition for the phase function*:

$$2\pi \int_0^\pi \tilde{\beta}(\psi, \lambda) \sin \psi \, d\psi = 1. \quad (3.15)$$

This normalization implies that the backscattering fraction can be computed from

$$B(\lambda) = 2\pi \int_{\pi/2}^\pi \tilde{\beta}(\psi, \lambda) \sin \psi \, d\psi. \quad (3.16)$$

The *asymmetry parameter* g of the phase function is the average over all scattering directions of the cosine of the scattering angle ψ , namely

$$g \triangleq \langle \cos \psi \rangle = 2\pi \int_0^\pi \tilde{\beta}(\psi) \cos \psi \sin \psi \, d\psi. \quad (3.17)$$

The asymmetry parameter is a convenient measure of the “shape” of the phase function. For example, if $\tilde{\beta}(\psi)$ is very large for small ψ , then g is near one. If $\tilde{\beta}(\psi)$ is symmetric about $\psi = 90$ degrees, then $g = 0$. Typical ocean waters have g values in the range of 0.8 to 0.95.

We have assumed above that the scattering is azimuthally symmetric, so that the directional pattern or angular shape of the VSF depends only on the scattering angle ψ . This is not the case for polarized incident light, even if the medium is isotropic. Thus a linearly polarized laser beam will scatter differently for different azimuthal directions (measured relative to the plane of polarization), even if the medium is isotropic. Description of polarized light requires the Stokes vector formalism of Section 1.6 and a more complicated description of scattering as described in Section 1.7.

If the medium contains non-spherical particles that are not randomly oriented, even unpolarized light will scatter differently for different azimuthal directions. This is sometimes the case, for example, in cirrus clouds and ice fogs, which can have non-spherical ice crystals that become oriented in a particular way as the crystals fall through a calm atmosphere. The atmosphere is then an optically anisotropic medium, and scattering is not azimuthally symmetric. Scattering of sunlight in such an atmosphere gives rise to phenomena such as “Sun dogs” or [parhelia](#), which are bright spots to either side of the Sun.

3.2 Visualizing VSFs

Oceanic volume scattering functions typically vary by 5 or 6 orders of magnitude over the range of very small to large scattering angles. Figure 3.3 illustrates the range of VSFs that can be found in various oceanic waters. Each of these VSFs varies by over five orders of magnitude between a scattering angle of $\psi = 0.1$ deg and their minima at large angles. The different VSFs vary by over two orders of magnitude at a given scattering angle.

The phase functions computed from these three VSFs are plotted in Fig. 3.4. The left panel shows the backscatter fractions and asymmetry parameters for each phase function. Although these phase functions appear similar in shape, at least when plotted with a logarithmic ordinate, there are significant differences. Note in particular that they vary

by almost an order of magnitude at large scattering angles, and their backscatter fractions vary by over a factor of two. Thus the different shapes of these phase functions would give factor-of-two or greater differences in upwelling radiances as measured by remote-sensing instruments. The actual differences in upwelling radiances would be even larger because of the magnitude differences in the VSFs.

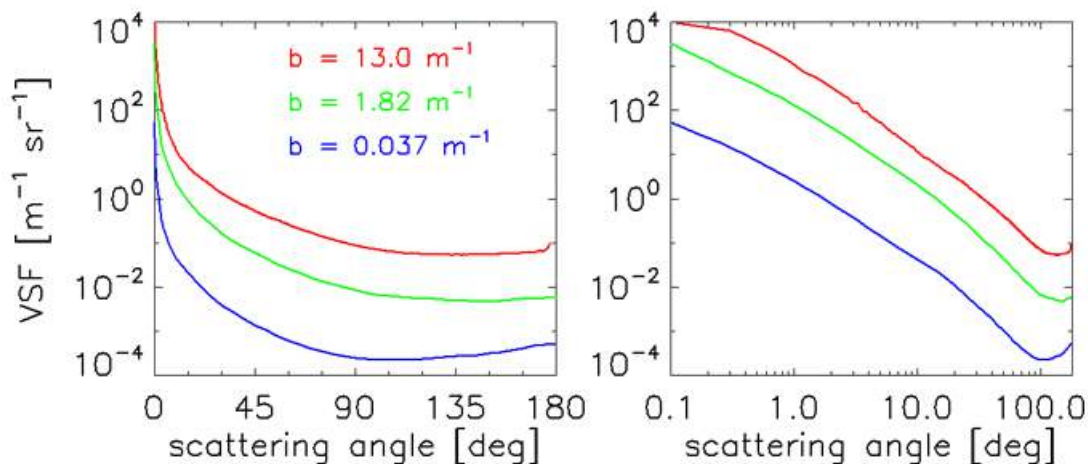


Figure 3.3: Example volume scattering functions. The blue curve was measured in clear, open ocean water; the wavelength was 514 nm. The green curve was measured in a harbor (514 nm), and the red curve in very productive coastal waters (530 nm). The corresponding scattering coefficients b are shown in the left panel.

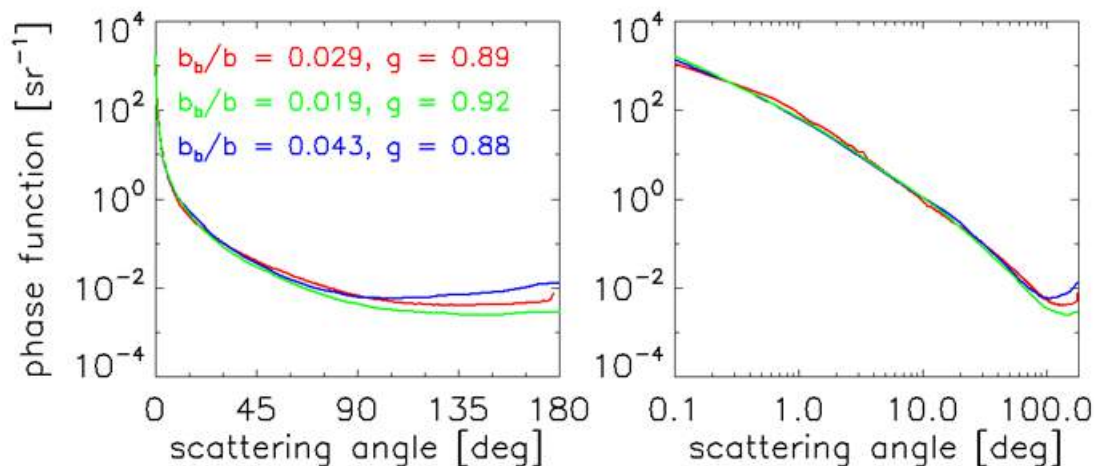


Figure 3.4: Example scattering phase functions. The color coding is the same as for Fig. 3.3

VSFs and phase functions are usually plotted as functions of the scattering angle ψ , as seen in the preceding figures. The ordinate is almost always logarithmic because of the wide ranges of VSF and phase function magnitudes. The abscissa can be linear (as in the left panels) or logarithmic (right panels); a logarithmic abscissa highlights the smallest

scattering angles. However, it should be remembered that scattering is an inherently three-dimensional process involving both polar and azimuthal scattering angles relative to the direction of the unscattered light.

Another typical oceanic phase function is displayed in four different ways in Fig. 3.5, which is worthy of some discussion. This particular phase function has an asymmetry parameter or average cosine of $g = 0.9255$ and a backscatter fraction of $B = 0.0183$. (This phase function was generated using the equations for the Fournier-Forand phase function, which is discussed in Section 6.7. The parameter values were $n = 1.1$ and $\mu = 3.5835$, which give a good fit to the Petzold average-particle phase function of Section 9.5.2.) In the upper two plots, the magnitude of the phase function has been color coded for comparison with the three-dimensional displays seen in the lower panels of the figure.

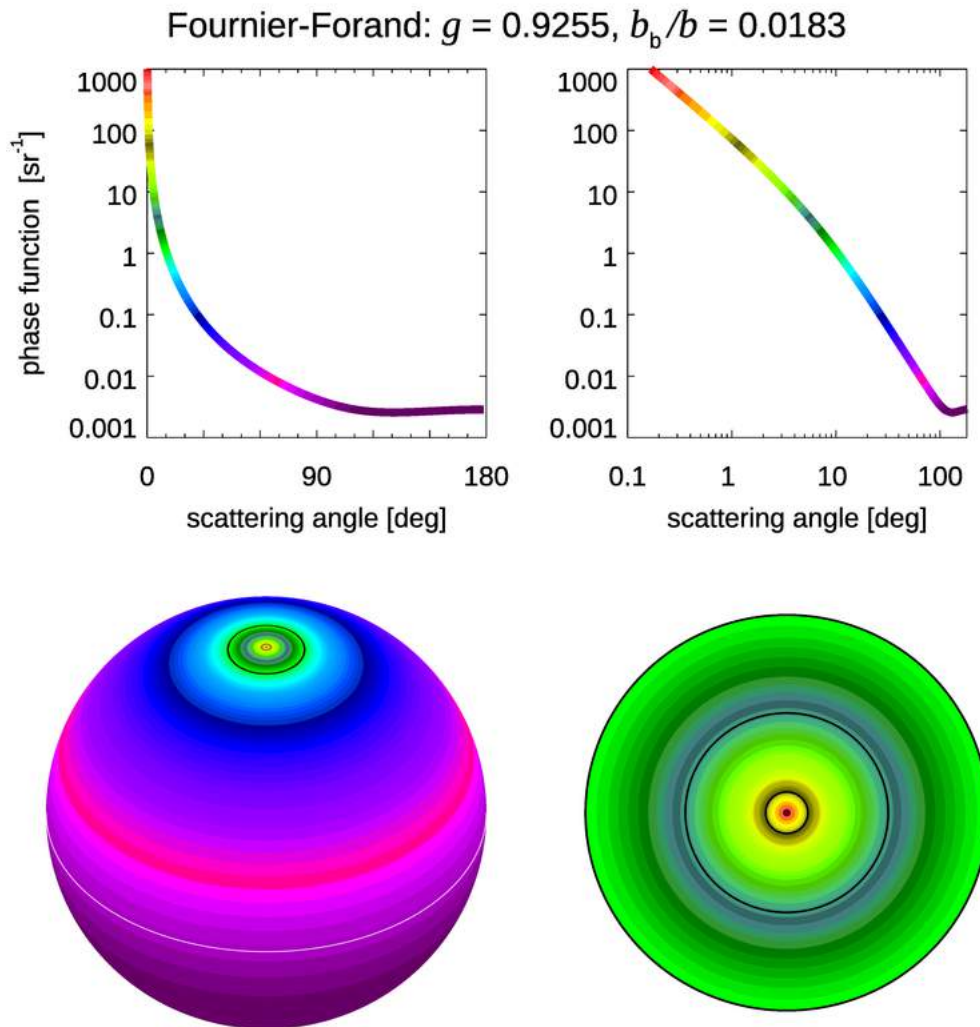


Figure 3.5: A typical oceanic phase function viewed in various ways. The magnitude is color coded the same in all plots. The lower right figure shows the near-forward scattering region: the inner black ring is $\psi = 5$ deg, and the outer ring is $\psi = 10$ deg.

The lower-left panel gives a 3-D perspective of the phase function, with the “north pole” of the sphere being the direction of the unscattered light. This point is colored dark red to indicate the large magnitude of the phase function, corresponding to the color scheme of the line drawings. The polar angle or “latitude” of the sphere is the scattering angle, and the “longitude” is the azimuthal angle of the scattering. Because we have assumed that the scattering is independent of the azimuthal angle, there is no azimuthal dependence in this 3-D plot. The black circle in the green region near the pole is $\psi = 10$ deg scattering angle; the dark red circle at the “equator” is $\psi = 90$ deg. The dark purple color in the backscatter directions at the bottom of the sphere indicates the low magnitudes of the phase function as seen in the line plots.

The lower-right plot is an expanded view of the scattering angles from 0 to 10 deg, as contained within the black circle of the lower-left sphere plot. The view is looking “into the beam,” or toward the light source. The inner black circle is $\psi = 1$ deg., the next black circle is $\psi = 5$ deg, and the outer black circle is the $\psi = 10$ deg value seen in the sphere plot.

These four plots display the same information in different ways. The two plots on the left are linear in the scattering angle, and the two plots on the right emphasize the small-angle scattering. The top two plots are the usual way of showing VSFs or phase functions when the scattering is azimuthally symmetric; the bottom two plots illustrate the 3-D character of the scattering process.

Figure 3.6 shows the phase function for pure water, which is given by the simple formula (see Eq. 6.7)

$$\tilde{\beta}_w(\psi) = 0.0608 (1 + 0.925 \cos^2 \psi) . \quad (3.18)$$

The pure water phase function has only a small range of magnitudes and is symmetric about the $\psi = 90$ deg scattering angle. Therefore the asymmetry parameter or average cosine of this phase function is $g = 0$, and the backscatter fraction is $B = 0.5$.

In the discussion of IOPs in the previous sections, we assumed that the medium, i.e. the water, was isotropic. Note that an isotropic medium does not have isotropic scattering, which would be a VSF that is independent of the scattering angle. Isotropic scattering is an idealization that does not exist in nature. Scattering by pure water is the closest you can come to isotropic scattering in the ocean.

3.3 Measuring IOPs

The definitions of the absorption and scattering coefficients in Section 3.0.1 were in terms of an infinitesimally thin slab of water. We now must reformulate those conceptual definitions so that they can be applied to actual measurements on finite thickness of water in a measuring instrument.

3.3.1 Measuring The Beam Attenuation Coefficient

Just as was done for the absorption and scattering coefficients (recall Eqs. 3.4 and 3.5), we can define the beam attenuation coefficient c as

$$c(\lambda) = \frac{dC(\lambda)}{dr} \quad (\text{m}^{-1}), \quad (3.19)$$

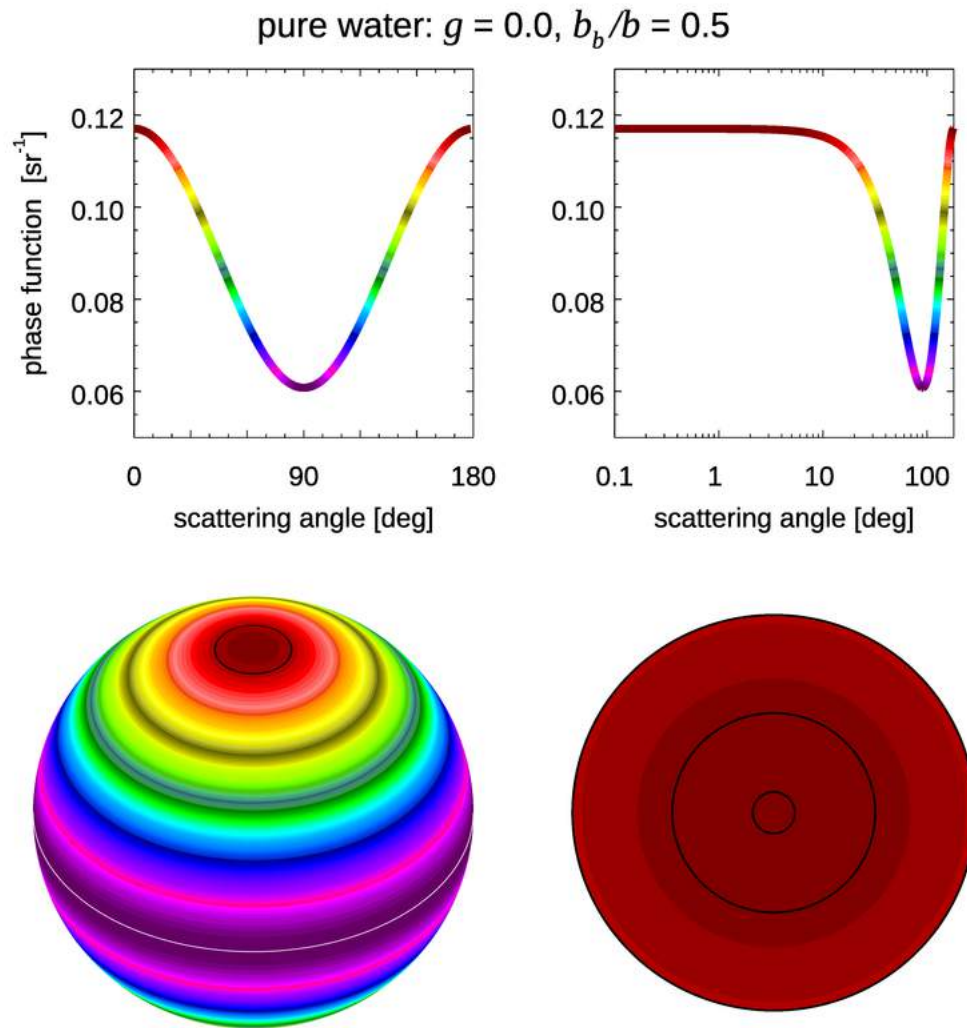


Figure 3.6: The phase function for pure water.

where $dC(\lambda) = dA(\lambda) + dB(\lambda)$ is the attenuation, or fraction of power absorbed or scattered as light passes through the slab of infinitesimal thickness dr . However, any instrument must make its measurements on a slab of some finite thickness R , which is typically between 0.01 and 1 m.

Figure 3.7 shows a finitely thick slab of water with an incident power Φ_i and transmitted power Φ_t , where the wavelength argument has been omitted for brevity. Within the slab the power has magnitude Φ after passing through a thickness r of material. In going from thickness r to $r + dr$, the power changes from Φ to $\Phi + d\Phi$, where $d\Phi < 0$ because the transmitted power decreases as the thickness traversed increases.

Recalling that the attenuation is the fractional change in power, we can rewrite Eq. (3.19) as

$$c = \frac{dC}{dr} = -\frac{d\Phi}{\Phi},$$

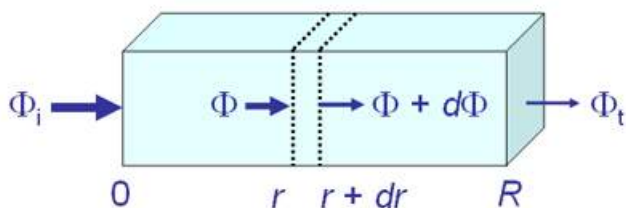


Figure 3.7: Geometry for measurement of beam attenuation c .

or

$$c dr = -\frac{d\Phi}{\Phi}, \quad (3.20)$$

where the minus sign accounts for $d\Phi$ being negative, whereas all other quantities are positive. Assuming that the medium within the slab is uniform, so that c is independent of r , we can integrate Eq. 3.20 from $r = 0$ to $r = R$, which corresponds respectively to powers Φ_i and Φ_t . This gives

$$\int_0^R c dr = cR = -\int_{\Phi_i}^{\Phi_t} \frac{d\Phi}{\Phi} = -\ln\left(\frac{\Phi_t}{\Phi_i}\right),$$

which can be rewritten as

$$c = -\frac{1}{R} \ln\left(\frac{\Phi_t}{\Phi_i}\right). \quad (3.21)$$

This equation gives the beam attenuation coefficient in terms of the measurable incident and transmitted powers and the finite thickness of the slab.

Equation (3.21) is the key to measuring the beam attenuation. However, there are additional subtleties in this measurement related to the instrument design, which is shown in more detail in Fig. 3.8. The development above implicitly assumed that the incident light was perfectly collimated (all light rays travel in exactly the same direction) and that the detector omits all scattered light. Neither requirement can be fully satisfied in a real instrument. Even a laser beam has some divergence, and any detector has a finite field of view (FOV) or acceptance angle. If the detector FOV is 1 deg, for example, then the detector will detect rays that have been scattered by 1 deg or less, as well as unscattered light. This undesired detection of scattered light will make the transmitted power too

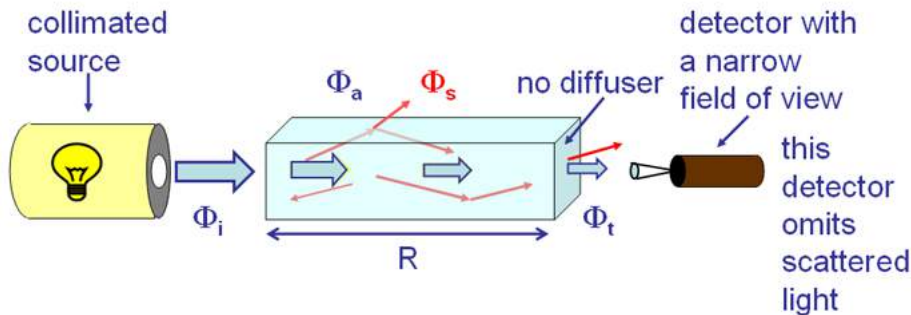


Figure 3.8: Schematic instrument design for measurement of the beam attenuation coefficient c . Φ_s represents the power scattered into all directions, as indicated by the red arrows.

large, and thus c too small. The magnitude of this error depends both on the detector FOV and on the volume scattering function of the water, which determines how much light is scattered through angles smaller than the FOV. Although these errors in beam attenuation measurements have been studied in many papers (see, most recently, Boss et al. (2009b) and references therein), it is difficult to correct for the error in a particular measurement because the VSF is usually unknown, especially at very small scattering angles. Unfortunately, as Boss et al. show, the differences in c values as measured by different instruments can be tens of percent. Indeed, Boss et al. end their paper with the comment, “We conclude that more work needs to be done with the oldest (in terms of availability of commercial in-situ instruments) and simplest (so we thought...) optical property, the beam attenuation.”

3.3.2 Measuring The Absorption Coefficient

If there were no scattering, then the instrument design used to measure beam attenuation would give the absorption coefficient a . However, at least some scattering is always present in sea water, which requires modification of the instrument design shown in Fig. 3.8. When measuring a , any light lost from the beam because of scattering will be attributed to a loss of light due to absorption. Thus it is desirable that the detector collect as much scattered light as possible. Because most scattering is through small angles, a common instrument design is to use as large a detector as possible at the end of the measurement chamber to collect forward scattered light. Such an instrument is shown schematically in Fig. 3.9.

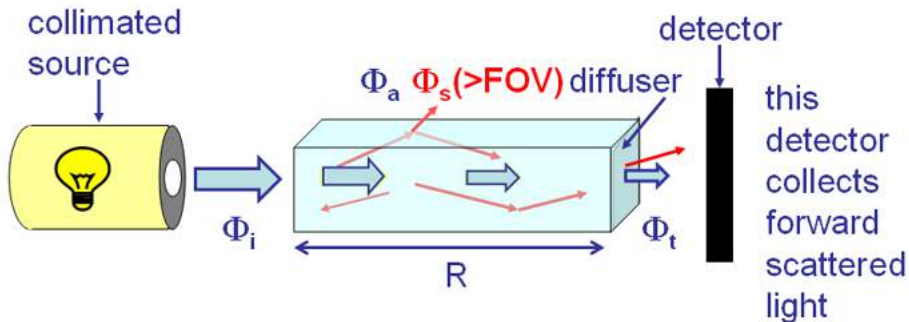


Figure 3.9: Schematic instrument design for measurement of the absorption coefficient a . $\Phi_s(>FOV)$ represents the power lost to the detector by scattering into angles greater than the detector field of view.

An instrument of this design typically collects light forward scattered through angles of a few tens of degrees (depending on the design of the particular instrument), which does account for most of the scattered light. However, some light is still lost “out the side” of the instrument by scattering through larger angles, as represented by $\Phi_s(>FOV)$ in the figure. To get an accurate absorption measurement, it is necessary to perform a scattering correction, i.e., to estimate $\Phi_s(>FOV)$ and to account for this loss when computing a . The equation for computation of a then becomes

$$a = -\frac{1}{R} \ln \left(\frac{\Phi_t + \Phi_s(>FOV)}{\Phi_i} \right). \quad (3.22)$$

In this equation, Φ_t now includes the measured transmitted power due both to unscattered light and to light scattered through angles less than the detector FOV. The $\Phi_s(>FOV)$ term adds back the unmeasured loss of power due to scattering through angles larger than the FOV. Estimating the value of $\Phi_s(>FOV)$ is of course difficult because the amount of scattering at angles greater than the instrument field of view depends on the (usually unknown) volume scattering function. Techniques for doing this scattering correction are given in Zaneveld et al. (1994).

In spite of the fundamental difficulties associated with the instrument designs of Figs. 3.8 and 3.9, the problems are not insurmountable and instruments based on these designs are commercially available. For example, the pioneering and widely used ac-9 instrument (and its successor, the ac-s) uses these designs for measurement of a and c . In the ac-9, the detector for measurement of c has an acceptance angle of about 1 deg, and the detector for measurement of a has an acceptance angle of about 40 deg.

Another way to account for scattering in absorption measurements is to place the entire measurement chamber inside an integrating sphere. An *integrating sphere* is a hollow sphere coated on the inside with a highly reflective white material (usually Spectralon, which has a reflectance greater than 99% at visible wavelengths and is a diffuse reflector). The inner surface of the sphere reflects incident light into all directions, which after many reflections makes the interior light field homogeneous and isotropic. Thus it is possible to measure the power at only one location within the sphere in order to compute the total power within the sphere. Integrating spheres are commonly used in laboratory spectrophotometers to collect the scattered light, and flow-through integrating-cavity absorption meters are now commercially available (e.g., PSICAM).

Equation 3.22 is often rewritten as

$$a = -\frac{1}{R} \ln \left(\frac{\Phi_t + \Phi_s(>FOV)}{\Phi_i} \right) - \frac{1}{R} \ln \left(\frac{\Phi_i - \Phi_a}{\Phi_i} \right) = -\frac{1}{R} \ln(1 - A) ,$$

after recalling the definition of the absorbance A . Using $\log_e x = \log_{10} x / \log_{10} e$ to convert from natural (base e) to common (base 10) logarithms gives

$$a = -\frac{2.303}{R} \log_{10}(1 - A) = \frac{2.303Abs}{R} , \quad (3.23)$$

where $Abs = -\log_{10}(1 - A)$ is called the *absorbance* or *optical density*³ Spectrophotometers usually output their measurements as a value of Abs , from which a can be computed by Eq. (3.23) for the given thickness R of the sample volume.

³Older literature often writes $D = -\log_{10}(1 - A)$ and calls D the “optical density.” The term optical density is now considered obsolete and its use is discouraged by the International Union of Pure and Applied Chemistry; see <https://goldbook.iupac.org/terms/view/004306>. Some authors use the term absorbance if the measurement does not include scattering, and optical density if the measurement includes scattering, but this is nonstandard. In the spectrophotometer literature, you will sometimes see the absorbance defined as $-\log_{10}(T)$, where T is the transmittance. This is equivalent only if there is no scattering in the sample. That is the case for chemists working with pure solutions, but it is seldom so for oceanographers whose samples usually contain scattering particles such as phytoplankton. Many authors used A for absorbance; this book uses A for absorbance and Abs for absorbance.

3.3.3 Measuring The Scattering Coefficient

It is possible to measure the scattering coefficient b , for example by collecting all scattered power in an integrating cavity and then using

$$b = -\frac{1}{R} \ln \left(\frac{\Phi_s}{\Phi_i} \right) .$$

It is also possible to measure the volume scattering function and then obtain b by integration of the VSF over all scattering angles. However, neither of these paths to b is often taken. The scattering coefficient b is usually obtained from measurements of the beam attenuation c and absorption coefficient a , and the definition of c :

$$b = c - a .$$

Note, however, that there is a philosophical problem here, even if the practical difficulties in measuring a and c can be overcome well enough to obtain usefully accurate values. The philosophical problem is that to obtain good values of a and c , you need to know the VSF in order to correct for the finite-FOV errors in c and to perform the scattering correction for the a measurement. Because the VSF is seldom measured, you must make assumptions about the scattering in order to obtain values for a and c , in order to obtain a value for the scattering b . Thus the value obtained for the scattering coefficient depends on the *a priori* assumptions made about scattering in the water being studied. This circular reasoning is tolerated in optical oceanography simply because of the difficulties of designing instruments to make accurate measurements of each IOP independent of assumptions about the other IOPs.

Measurement of the backscattering coefficient b_b is discussed in Section 6.4.

3.3.4 Measuring The Volume Scattering Function

As with the measurement of beam attenuation, measurement of the volume scattering function employs a narrow collimated beam and a narrow-field-of-view detector. Now, however, the detector views the incident beam at a given scattering angle. The basic geometry is shown in Fig. 3.10.

The intersection of the incident beam and the detector FOV define a volume of water ΔV in which the scattering from the incident direction $\hat{\xi}'$ into the scattered direction $\hat{\xi}$ takes place. The scattering angle ψ is defined as shown by the angle between $\hat{\xi}'$ and $\hat{\xi}$ (recall Eq. (1.9)). The incident power Φ_i traveling in direction $\hat{\xi}$ into volume ΔV generates scattered power $\Delta^2 \Phi_s(\psi)$ in the detector FOV of solid angle $\Delta \Omega$ centered on direction $\hat{\xi}'$ over a path length of Δr . The definitions of the VSF in Section 3.1.1 give an operational formula for the VSF in terms of measurable quantities:

$$\beta(\psi) = \frac{\Delta^2 \Phi_s(\psi)}{\Phi_i \Delta r \Delta \Omega} , \quad (3.24)$$

which as previously noted is equivalent to

$$\beta(\psi) = \frac{\Delta I_s(\psi)}{E_i \Delta V} . \quad (3.25)$$

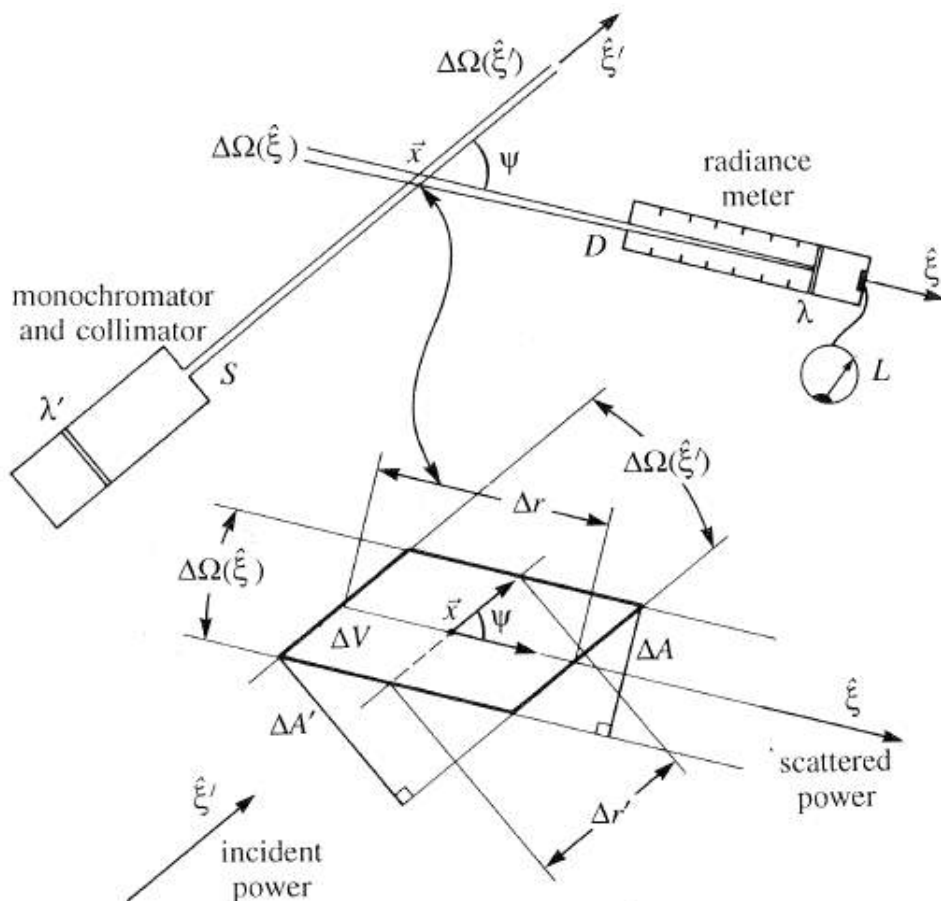


Figure 3.10: Geometry for measurement of the volume scattering function.

As always, there are implicit assumptions and subtleties involved with these equations. It is assumed that the scattering volume ΔV is small enough that only single scattering takes place within the volume, but large enough to contain a representative sample of scattering particles. It is necessary to correct for attenuation of the incident and scattered beams along the entire path from the source to the scattering volume to the sensor, which presumes knowledge of the beam attenuation coefficient c . (Note again our problem: to measure the VSF we need to know c , but we need to know the VSF in order to make the FOV correction for the measurement of c .) In the pioneering General Angle Scattering Meter (GASM) developed by [Petzold \(1972\)](#), it was difficult to accurately determine the scattering volume ΔV , which changes with scattering angle, because of the difficulty of maintaining mechanical alignment of the moving parts as they swept the detector through the full range of scattering angles. The calculations and engineering considerations needed to implement Eqs. [3.24](#) or [3.25](#) in GASM are given in Petzold's classic report.

More recent instruments have used other designs to avoid various problems. For example, the Volume Scattering Meter (VSM) of [Lee and Lewis \(2003\)](#) used a novel design based on a rotating periscope to scan through scattering angles from 0.6 to 177.3 deg. A more recent in situ instrument, the [LISST-VSF](#) can measure the VSF at 515 nm in the range of $\psi = 0.094$ to 150 deg, along with the scattering coefficient and beam attenuation

and selected components of the Mueller matrix for polarized light scattering. See [Slade et al. \(2013\)](#).

In sea water both absorption and scattering are always important. Depending on the water constituents and the wavelength, one process can dominate the other, but neither can be ignored. Thus one IOP cannot be measured independently of the others. Clearly, there is need for continued development of IOP instruments with novel designs that minimize the measurement problems mentioned above.

3.4 Additivity of IOPs

In natural waters many different dissolved and particulate substances contribute to the total IOPs of the water body. As will be seen in Chapter 9 on radiative transfer theory, is it the total absorption and total VSF due to all water components that determine light propagation in a water body. Fortunately, the total IOPs are the sum of the IOPs of the individual components. Thus the total absorption is the sum of the absorption by water (subscript w), by the various species of phytoplankon (phy), by CDOM, by minerals, by pollutants, and so on. If the water body contains N different components, then

$$a_{\text{total}} = \sum_{i=1}^N a_i, \quad (3.26)$$

where i labels the individual components. When modeling IOPs (e.g. Section 8.8) it is common to consider three or four components: the water itself; a generic phytoplankon, often parameterized by the chlorophyll concentration; dissolved organic matter (CDOM); and perhaps mineral particles. In this case (3.26) can be written

$$a_{\text{total}} = a_w + a_{\text{phy}} + a_{\text{CDOM}} + a_{\text{min}}. \quad (3.27)$$

Of course, each component absorption will vary with wavelength, and all but water generally vary with location and time as well.

The same partitioning of totals into sums of component contributions holds for the volume scattering function:

$$\beta_{\text{total}} = \sum_{i=1}^N \beta_i. \quad (3.28)$$

This implies (by integration over scattering angle) that

$$b_{\text{total}} = \sum_{i=1}^N b_i, \quad (3.29)$$

and

$$b_{b,\text{total}} = \sum_{i=1}^N b_{b,i}. \quad (3.30)$$

Note, however, that the scattering phase functions are not directly additive; they must be weighted by the relative amount of scattering for each component. Recall from Eq.

(3.14) that the VSF β can be written as the product of the scattering phase function $\tilde{\beta}$ and the scattering coefficient b . Applying this to each component VSF gives

$$\tilde{\beta}_{\text{total}} b_{\text{total}} = \sum_{i=1}^N \tilde{\beta}_i b_i$$

or

$$\tilde{\beta}_{\text{total}} = \sum_{i=1}^N \left(\frac{b_i}{b_{\text{total}}} \right) \tilde{\beta}_i.$$

Note that if the phase function of each component obeys the normalization condition (3.15), then so does the total.

It is usually assumed when modeling the total IOPs that the component phase functions are determined by the *type* of component, and the spatial variability is contained in the scattering coefficient via the depth-dependent *concentration* of the component. In that case, the last equation can be written as

$$\tilde{\beta}_{\text{total}}(\psi, z, \lambda) = \left(\frac{b_{\text{water}}(\lambda)}{b_{\text{total}}(z, \lambda)} \right) \tilde{\beta}_{\text{water}}(\lambda) + \sum_{i=2}^N \left(\frac{b_i(z, \lambda)}{b_{\text{total}}(z, \lambda)} \right) \tilde{\beta}_i(\lambda)$$

It is also common practice to regard the water contribution as a known quantity and to remove the water contribution from measured absorption and scattering coefficients. However, when doing radiative transfer calculations, the water contribution must always be included.

3.5 Absorption and Scattering Are Equally Important

For a particular problem, knowing just the absorption coefficient or the volume scattering function may be sufficient to solve the problem. For example, phytoplankton photosynthesis depends only on the light (number of photons) absorbed by chlorophyll molecules, which begins the exceedingly complex process of photosynthesis. Light scattered by a cell is unavailable to drive photosynthesis and is irrelevant to the process. For this reason, phytoplankton biologists have made countless measurements of absorption spectra without making accompanying measurements of scattering. On the other hand, only light scattered by the sea surface contributes to the Sun's glitter pattern seen from above the surface, which can be used to deduce the wave state of the surface. Light absorbed by the ocean cannot contribute to the glitter pattern and is irrelevant to remote sensing of surface conditions. Consequently, Fresnel's equations for reflectance (Sections 13.2 and 13.3) are sufficient to understand angular pattern of the surface-reflected radiance, and no knowledge of absorption is needed.

However, as will be seen in Chapter 9, you must know both the absorption coefficient and the volume scattering function in order to solve the radiative transfer equation and predict the in-water radiance distribution. Moreover, knowing just one of a or β gives you no information about the other; both must be measured or modeled. A given parcel of water will have certain absorption and scattering properties. If you add CDOM to the water, you can greatly increase the absorption without any effect on the scattering. If you add non-absorbing quartz or calcite particles to the water, you can greatly increase the

scattering without any increase in absorption. However, the two processes do affect the importance of each other. If there is a lot of scattering, the path traveled by a beam of light between one depth and another in the water increases, so there is more opportunity for the light to be absorbed. If there is more absorption, there is less light available to be scattered.

For many problems, both absorption and scattering simultaneously determine the answer. For example, both processes affect visibility in ways that will be discussed in Chapter 16. Absorption removes light from a beam and thus makes a distant image darker. Scattering through small angles by particles blurs the edges of an image, and scattering by turbulence causes twinkling or mirages. Thus the visual appearance of a distance object is determined by the relative amounts of absorption and scattering, and even by what processes cause the scattering. The effects of absorption and scattering on an image are illustrated in Fig. 3.11. Wavelength-dependent absorption also alters the color of a distant image, as seen in Fig. G.21.

Clearly, *both absorption and scattering are equally important to the prediction and understanding of light propagation.*

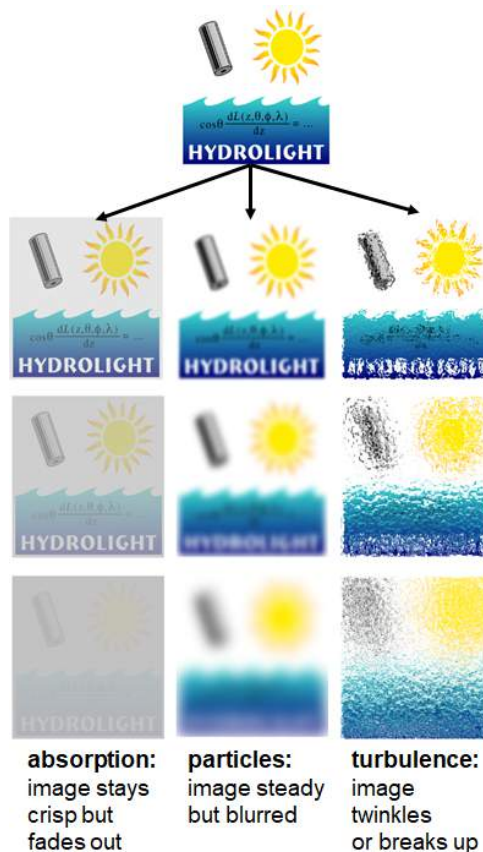


Figure 3.11: Illustration of absorption and scattering by particles and by turbulence on visibility. Absorption and scattering (or distance from the object) increase from top to bottom in each column.

Apparent Optical Properties

One of the primary goals of optical oceanography is to learn something about a water body, e.g., its chlorophyll or sediment concentration, from optical measurements. Ideally one would measure the absorption coefficient and the volume scattering function, which tell us everything there is to know about the bulk optical properties of a water body. Those IOPs do indeed tell us a lot about the types and concentration of the water constituents. However, in the early days of optical oceanography, it was difficult to measure in situ IOPs other than the beam attenuation coefficient. Indeed, there were no commercial instruments for measuring absorption and scattering coefficients in situ until the 1990s.

On the other hand, it was relatively easy to measure radiometric variables such as the upwelling and downwelling plane irradiances. This led to the use of *apparent* optical properties (AOPs) rather than IOPs to describe the bulk optical properties of water bodies. As will be seen, a “good” AOP will give useful information about a water body, e.g., the types and concentrations of the water constituents, from easily made measurements of the light field. This chapter defines and illustrates the AOPs commonly used in optical oceanography. The chapter closes with a section of optical classifications for water bodies.

Apparent optical properties are those properties that

1. depend both on the medium (the IOPs) and on the geometric (directional) structure of the radiance distribution, and that
2. display enough regular features and stability to be useful descriptors of a water body.

Figure 4.1 helps understand the idea of an AOP. This figure shows depth profiles of E_d and E_u for selected solar zenith angles and clear versus overcast sky conditions. These irradiances were computed by HydroLight for Case 1 water with a chlorophyll concentration of 1 mg m^{-3} . Note first that each of these irradiances varies by an order of magnitude for the range of solar zenith angles and sky conditions. Irradiances do not satisfy condition (2) above because of this strong dependence on sky conditions and are therefore not AOPs. An irradiance profile does not give much information about the water IOPs because it is difficult to separate the effects of IOPs from the effects of the external environment on the value of the irradiance. A measurement of spectral E_d might show that the water is blue and thus perhaps low in chlorophyll, versus green and thus perhaps high

in chlorophyll. However, we cannot expect to deduce the chlorophyll concentration from a measurement of E_d because of its sensitivity to external environmental effects. **Radiances and irradiances themselves are not AOPs.**

Note, however, that the *slopes* of the irradiance profiles are very similar and almost independent of the Sun and sky conditions. This suggests that the depth derivative (slope) of $E_d(z)$ and $E_u(z)$ (or of radiances in particular directions) might be almost independent of sky conditions but still be dependent on the IOPs. This is the idea behind K functions. Note also that the *ratio* of $E_u(z)$ to $E_d(z)$ looks to be about the same for each particular solar zenith angle or cloudiness. This suggests that ratios of irradiances (or of radiances) might be almost independent of the external environment but still be dependent on the IOPs. This is the idea behind reflectances.

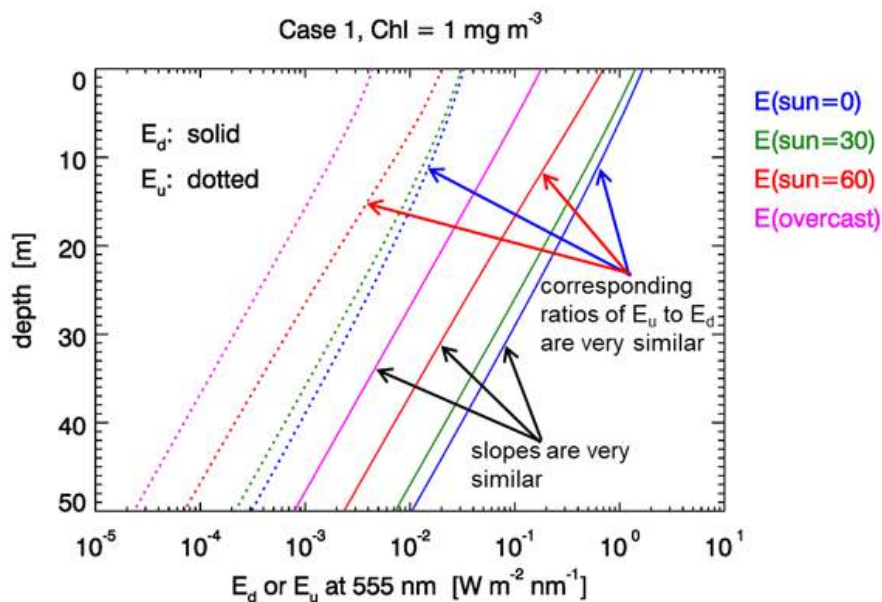


Figure 4.1: HydroLight-computed depth profiles of E_d and E_u for selected solar zenith angles and clear versus overcast sky conditions.

Building on the above ideas, the normalized or logarithmic depth derivative of E_d ,

$$K_d(z, \lambda) \triangleq -\frac{1}{E_d(z, \lambda)} \frac{dE_d(z, \lambda)}{dz} = -\frac{d \ln E_d(z, \lambda)}{dz} \quad (\text{m}^{-1}), \quad (4.1)$$

is a candidate AOP. $K_d(z, \lambda)$ is called the *diffuse attenuation coefficient for downwelling plane irradiance*. If E_d changes suddenly, the first form of the derivative shows that the change in magnitude of E_d will cancel out, leaving the value of K_d unchanged. K_d thus satisfies the stability requirement for an AOP. K_d should also depend on the IOPs because changing them will change now rapidly the irradiance changes with depth. The diffuse attenuation function K_d is thus a candidate worthy of consideration as an AOP.

Now consider the *ratio* of upwelling to downwelling plane irradiances,

$$R(z, \lambda) = \frac{E_u(z, \lambda)}{E_d(z, \lambda)},$$

which is called the *irradiance reflectance*. If E_d changes because of a change in Sun location, cloud cover, or surface waves, E_u will change proportionately because E_u arises mostly from upward scattering of the same downwelling radiance that determines E_d . Thus the E_u/E_d ratio should be much less influenced by changes in the external environment than are E_u and E_d individually. The reflectance R is thus worthy of further investigation as a descriptor of the water body itself. However, we still expect at least a small change in E_u/E_d as Sun zenith angle changes, for example, because different parts of the VSF will contribute to the upward scattering of downwelling radiance.

It is easy to think of other such ratios and depth derivatives that can be formed from radiometric variables: L_u/E_d , L_u/L_d , $-d \ln E_u/dz$, $-d \ln L_u/dz$, and so on. Each of these candidate AOPs must be investigated to see which ones provide the most useful information about water bodies and which ones are least influenced by external environmental conditions such as Sun location or sky conditions. Table 4.1 lists the most commonly used AOP's. These will be examined in the following sections.

AOP name	Symbol	Definition	Units
diffuse attenuation coefficients			
(K functions)			
of radiance in any direction $L(\theta, \phi)$	$K(\theta, \phi)$	$-d \ln L(\theta, \phi)/dz$	m^{-1}
of upwelling radiance L_u	K_{L_u}	$-d \ln L_u/dz$	m^{-1}
of downwelling irradiance E_d	K_d	$-d \ln E_d/dz$	m^{-1}
of upwelling irradiance E_u	K_u	$-d \ln E_u/dz$	m^{-1}
of scalar irradiance E_o	K_o	$-d \ln E_o/dz$	m^{-1}
of PAR	K_{PAR}	$-d \ln PAR/dz$	m^{-1}
reflectances			
irradiance reflectance	R	E_u/E_d	nondim
remote-sensing reflectance	R_{rs}	$L_w(\text{in air})/E_d(\text{in air})$	sr^{-1}
remote-sensing ratio	RSR	L_u/E_d	sr^{-1}
average cosines			
of the radiance distribution	$\bar{\mu}$	$(E_d - E_u)/E_o$	nondim
of the downwelling radiance	$\bar{\mu}_d$	E_d/E_{od}	nondim
of the upwelling radiance	$\bar{\mu}_u$	E_u/E_{ou}	nondim

Table 4.1: Commonly used apparent optical properties. R_{rs} is a function of wavelength only; K_{PAR} is a function of depth only; all other AOPs are functions of both depth and wavelength.

4.1 Diffuse Attenuation Coefficients (K Functions)

This section further examines the diffuse attenuation coefficients, which are commonly called “ K functions.” Recall, for example, the definition (4.1) of the diffuse attenuation coefficient for downwelling plane irradiance:

$$K_d(z, \lambda) \triangleq -\frac{d \ln E_d(z, \lambda)}{dz}.$$

Solving this for the irradiance E_d gives

$$E_d(z, \lambda) = E_d(0, \lambda) \exp \left[-\int_0^z K_d(z', \lambda) dz' \right]. \quad (4.2)$$

This equation is exact. If $K_d(z, \lambda)$ were independent of depth—which it never is, as will be seen below—the last equation would reduce to

$$E_d(z, \lambda) = E_d(0, \lambda) \exp [-K_d(\lambda) z]. \quad (4.3)$$

This equation is always an approximation.

Corresponding equations can be written for all other radiometric variables and their corresponding K functions. For example, we can define a K function for upwelling radiance as seen in Table 4.1,

$$K_{Lu}(z, \lambda) \triangleq -\frac{d \ln L_u(z, \lambda)}{dz}.$$

and then write

$$L_u(z, \lambda) = L_u(0, \lambda) \exp \left[-\int_0^z K_{Lu}(z', \lambda) dz' \right].$$

Under typical oceanic conditions, for which the incident lighting is provided by the Sun and sky, and *in homogeneous water, when far enough below the surface (and far enough above the bottom, in optically shallow water) to be free of boundary effects*, the K functions approach a common value (i.e., $K_d \approx K_u \approx K_{Lu}$, etc.) and do become almost independent of depth, in which case the various radiances and irradiances all decrease *approximately* exponentially with depth. The assumption of an exponential decrease of the light field with depth is often good enough for a back-of-the-envelope estimation, but it must always be remembered that reality is more complicated. The the value of the common K function and the rate of approach of K functions to a common asymptotic value in deep, homogeneous water depends on the IOPs. This is discussed in detail in Section 10.4.

4.1.1 Dependence of K Functions on IOPs and Environmental Conditions

As explained in the previous section, in order to be useful for relating light measurements to water properties, the K functions should depend strongly on the water IOPs but only weakly on external environmental conditions like Sun location, sky condition, or surface waves.

As an initial illustration of the dependence of K functions on water IOPs, recall the irradiances in Crater Lake and San Vicente Reservoir, which were used in Section 2.1 to illustrate downwelling and upwelling radiances in very clear and in very turbid waters. The

irradiance spectra of Fig. 2.2 were used to compute the average K_d and K_u between depths 5 and 25 m in Crater Lake via

$$K_d(5-25) = -\frac{\ln[E_d(25)/E_d(5)]}{25 - 5},$$

with a corresponding equation for $K_u(5-25)$. These K spectra are plotted in the left panel of Fig. 4.2. The green curves are the corresponding K spectra for optically pure water as computed by HydroLight, including the ever-present Raman scattering, which is discussed in Section 7.2. As Tyler and Smith (1970) comment, Crater Lake is as close to pure water as you can find in nature, and the closeness of the K_d spectra support this qualitative statement. The right panel of the figure shows the K spectra for the turbid San Vicente reservoir; these are tabulated in Tyler and Smith (1970, page 96) with the comment that the spectra are an “average for all depths.” The San Vicente K_d functions are over 100 times larger than the Crater Lake values at wavelengths less than 415 nm, and 10 to 100 times larger at wavelengths between 415 and 525 nm. These differences are no doubt due to the exponentially increasing absorption by CDOM going from green to blue to near-UV wavelengths. At red wavelengths, the water itself is the dominant absorber, and the San Vicente values are only twice the Crater Lake values. It is clear that the values of K_d and K_u are strongly determined by the water IOPs, as is desired of an AOP.

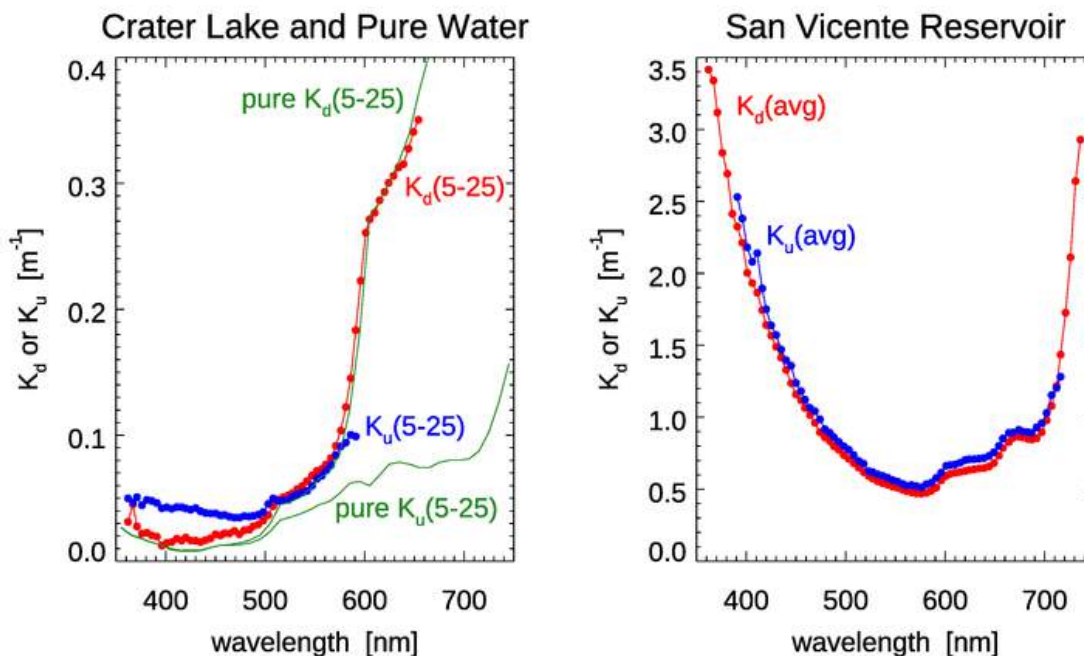


Figure 4.2: The red curve in the left panel is the average K_d between 5 and 25 m in Crater Lake; the blue curve is the average K_u between 5 and 25 m. The green curves are for optically pure water, including Raman scattering effects. The right panel shows K_d and K_u for San Vicente Reservoir.

To further illustrate these dependencies, various K functions were numerically computed using the HydroLight radiative transfer numerical model (Section 10.6). In many

situations it is preferable to work with real data. However, use of this model gives us the ability to simulate different environmental conditions at will and to control things that cannot be controlled in nature, such placing the Sun at various zenith angles while all else is held the same, or turning chlorophyll fluorescence on or off to see its effect. This can be very useful for understanding the interdependence of various quantities.

HydroLight simulations were first performed at one wavelength for homogeneous idealized water bodies dominated by either scattering or absorption. For the “highly scattering” water, the absorption coefficient was set to $a = 0.2 \text{ m}^{-1}$, and the scattering coefficient was $b = 0.8 \text{ m}^{-1}$, so that the albedo of single scattering was $\omega_o = b/(a + b) = 0.8$. These values correspond roughly to what might be found at blue or green wavelengths in Case 1 water with a chlorophyll concentration of 5 mg m^{-3} . An “average-particle” scattering phase function was used, the Sun was placed in a clear sky, and the water was infinitely deep. Note that since $a + b = c = 1 \text{ m}^{-1}$, the nondimensional optical depth $\zeta = cz$ is numerically equal to the geometric depth z in meters.

Figure 4.3 shows various K functions as a function of depth for the highly scattering water when the Sun was placed at a zenith angle of 40 degrees and the surface was level (wind speed of $U = 0$). As is conventional in radiative transfer theory, the $K(\theta, \phi)$ curves are for radiance propagating in the (θ, ϕ) direction. HydroLight measures depth positive downward from the mean sea surface, and polar angle θ is measured from the $+\hat{z}$ or nadir direction. Thus $\theta = 0$ refers to light heading straight down into the water (corresponding to K_{Ld}), $\theta = 180$ refers to light heading straight up (corresponding to K_{Lu}), and $\theta = 90$ refers to light traveling horizontally through the water. Azimuthal angle $\phi = 0$ refers to light heading towards the Sun, which was placed at $\phi = 0$; $\phi = 180$ thus refers to light heading away from the Sun. In this simulation, $K_{\text{Sun}}(30, 180)$ corresponds to looking into the Sun’s refracted beam underwater, which is light heading downward and away from the Sun.

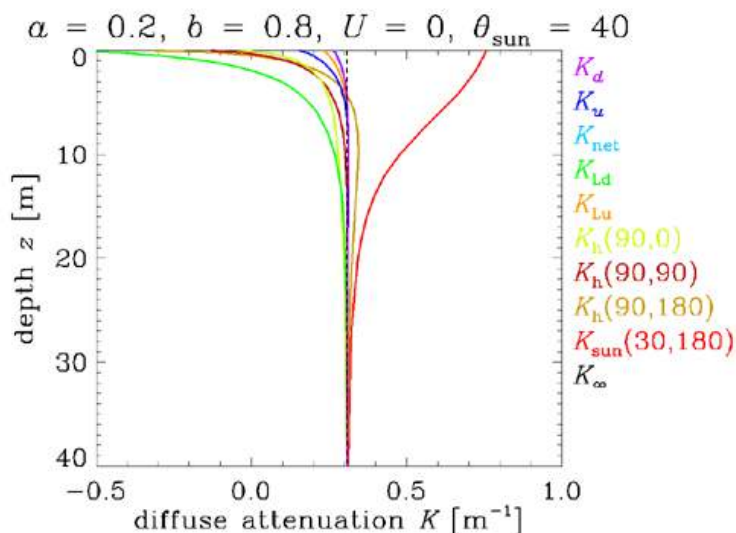


Figure 4.3: Computed K functions for “highly scattering” water. The Sun was at a zenith angle of 40 deg and the sea surface was level. The optical depth is numerically equal to the geometric depth.

There are several important features to note in Fig. 4.3:

- *The various K functions can differ greatly near the sea surface.* This is due to boundary effects on the solution of the radiative transfer equation. The surface boundary affects radiances in different directions in different ways depending on the relative location of the Sun. The large near-surface values of $K_{\text{Sun}}(30, 180)$ indicate that the Sun's direct beam is decreasing rapidly with depth due to absorption and scattering out of the beam. On the other hand, K_{Lu} , which is looking straight downward at radiance propagating upward, is almost constant with depth.
- *K functions can be positive or negative near boundaries.* A negative K means that the radiometric variable is increasing with depth. K_{Ld} , which is looking in the zenith direction at radiance propagating downward, is negative in the first few meters below the surface. At depth 0 just below the sea surface, the downwelling radiance L_{d} comes mostly from the zenith sky radiance transmitted through the level surface. Going deeper into the water, L_{d} increases with depth as scattering from the Sun's strong direct beam contributes more path radiance to L_{d} than is lost by absorption and scattering out of the downward beam. Eventually the Sun's direct beam becomes weak enough that the path radiance contribution to L_{d} is less than the attenuation due to absorption and scattering out of the beam, and K_{Ld} becomes positive. The same effect is seen less dramatically in $K_{\text{h}}(90, 180)$, which corresponds to looking horizontally toward the Sun.
- *K functions are not constant with depth even in homogeneous water.* Again, this is a manifestation of the surface boundary effects. If the water IOPs depend on depth, then the K functions also vary with depth, even far from a boundary. Thus radiometric variables never decrease *exactly* exponentially with depth, although this is often a good approximation for homogeneous water.
- *Far from boundaries (i.e., very deep in the ocean and very far from the bottom), all K functions approach a common value, the "asymptotic K value" K_{∞} , that depends only on the IOPs.* Its value for the IOPs of this simulation was $K_{\infty} = 0.3082 \text{ m}^{-1}$. Thus at depths great enough for boundary effects to be negligible, all K functions are the same and these AOPs become an IOP. In the present simulation, the K functions are all the same to within 3% by 30 m depth; K_{d} is within 0.2% of K_{∞} by 30 m depth. The asymptotic radiance distribution and the associated asymptotic AOPs are discussed in Section 10.4.

Figure 4.4 shows the K functions corresponding to the same conditions as Fig. 4.3, except that the wind speed was $U = 15 \text{ m s}^{-1}$. We see that there is very little difference between Figs. 4.3 and 4.4. Thus, as hoped, the K functions are almost unaffected by the surface waves.

Figure 4.5 shows the K functions for a level surface and the Sun at the zenith, rather than at 40 deg. Again, the irradiance K functions are almost unchanged. However, the radiance K_{Ld} function now corresponds to looking straight upward into the Sun's direct beam. Thus K_{Ld} now looks very similar to $K_{\text{Sun}}(30, 180)$ in the previous figures. Similarly, $K(30, 180)$ now looks much like K_{Ld} in the previous figures. This is because moving the Sun from 40 deg in air (28 deg in water) to the zenith gives $K(30, 180)$ almost the same scattering angle relation to the Sun's direct beam as K_{Ld} had in the previous figures.

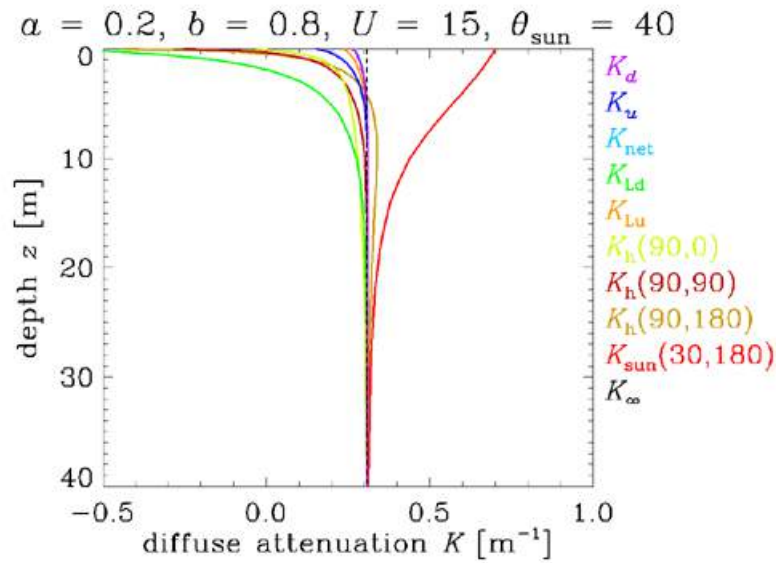


Figure 4.4: Computed K functions for “highly scattering” water. The conditions were the same as for Fig. 4.3, except that the wind speed was 15 m s^{-1} , so that the sea surface was not level.

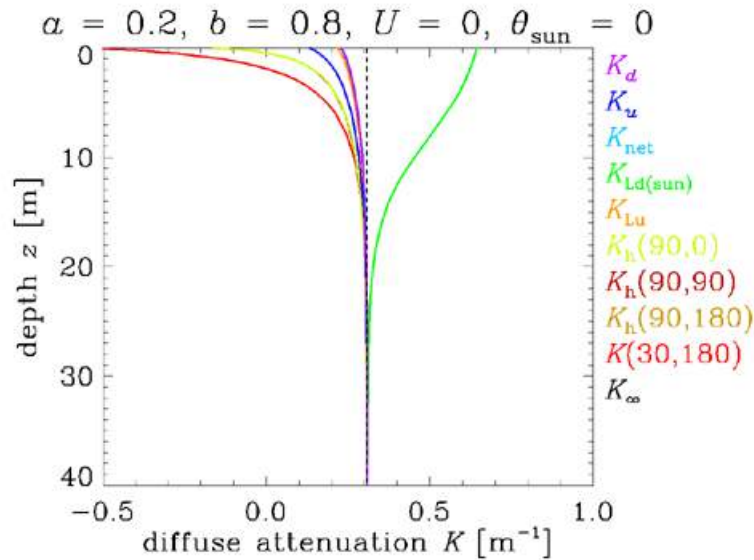


Figure 4.5: Computed K functions for “highly scattering” water. The conditions were the same as for Fig. 4.3, except that the Sun was at the zenith.

Figure 4.6 shows K functions for “highly absorbing” water: the absorption coefficient was $a = 0.8 \text{ m}^{-1}$, and the scattering coefficient was $b = 0.2 \text{ m}^{-1}$, so that the albedo of single scattering was $\omega_o = b/(a+b) = 0.2$. These values correspond roughly to what might be found at red wavelengths, where absorption by the water itself usually dominates the IOPs. Other conditions were the same as for Fig. 4.3.

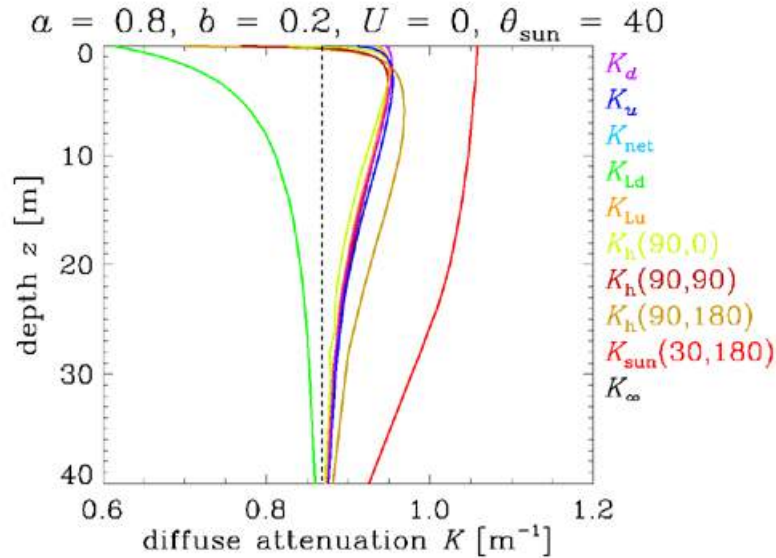


Figure 4.6: Computed K functions for “highly absorbing” water. The Sun was at a zenith angle of 40 deg and the sea surface was level. The optical depth is numerically equal to the geometric depth.

Comparing Figs. 4.3 and 4.6, we note that *The rate of approach to the asymptotic value depends on the IOPs.* In highly scattering water, the approach to K_∞ is much faster than in highly absorbing water. This is because the near-surface radiance distribution must be “redistributed” by multiple scattering into the shape of the *asymptotic radiance distribution* L_∞ in order for the K ’s to approach K_∞ . The more scattering, the faster the initial ray directions are changed by multiple scattering into their asymptotic distribution, which depends only in the IOPs.

Comparing Figs. 4.3 and 4.6 also shows that the K ’s have changed greatly because of the change in IOPs, which is what is desired in any AOP. For the highly absorbing water, $K_\infty = 0.8681 \text{ m}^{-1}$.

4.1.2 Beam Attenuation versus Diffuse Attenuation

The distinction between *beam* and *diffuse* attenuation is important. The beam attenuation coefficient c is defined in terms of the radiant power lost from a collimated beam of light. The downwelling diffuse attenuation function $K_d(z, \lambda)$, for example, is defined in terms of the decrease with depth of the downwelling irradiance $E_d(z, \lambda)$, which comprises light heading in all downward directions (a diffuse, or uncollimated, radiance distribution). In the above simulations, $c = a + b = 1.0 \text{ m}^{-1}$ at all depths, but all K functions except the near-surface $K_{\text{Sun}}(30, 180)$ in the high absorption case are less than c . Radiative transfer

theory shows (e.g., Mobley, 1994, Eq. 5.71) that in general $a \leq K_d \bar{\mu}_d \leq c$, where $\bar{\mu}_d$ is the average cosine of the downwelling radiance (Section 4.3). These inequalities are seen to hold true in the above simulations.

The left panel of Fig. 4.7 illustrates two different radiance distributions. The first, shown in blue, is directed more toward the nadir, and the second, shown in red, has a greater angle from the nadir. Even if the downwelling plane irradiances from both distributions are equal at depth z_1 , $E_{d1}(z_1) = E_{d2}(z_1)$, they will not be equal at depth z_2 . The reason is that the red rays travel a greater distance in going from depth z_1 to z_2 . More energy will be lost to absorption from the red rays than from the blue rays because the path length of the red rays through the water is greater. Thus $E_{d1}(z_2) > E_{d2}(z_2)$, which is equivalent to $K_{d1}(z_1) < K_{d2}(z_1)$. In both cases, the beam attenuation c is the same and does not depend on the depth if the water is homogeneous. Beam c is an IOP and does not depend on the radiance distribution, whereas the AOP K_d depends on the angular distribution of the radiance.

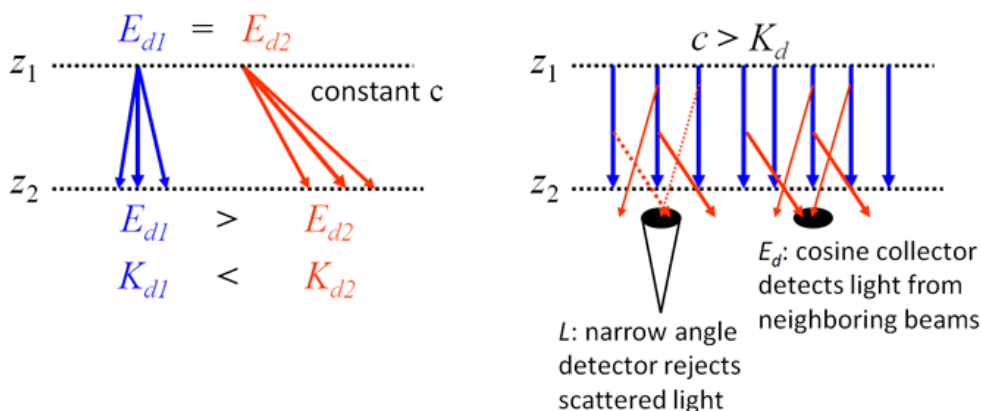


Figure 4.7: Illustration of the difference in beam attenuation c and diffuse attenuation K_d

The right panel of the Figure shows why beam attenuation c is always greater than diffuse attenuation K_d . The blue arrows represent the rays of a collimated radiance in the nadir direction (L_d). As the light propagates downward, some of the radiance will be scattered out of the beam at a particular horizontal location, as illustrated by the red arrows. A narrow-field-of-view radiance meter looking upward (measuring $L_u(z_2)$) will detect the radiance that is transmitted from depth z_1 to z_2 with only small-angle scattering (scattering at angles less than the FOV; there will also be some loss to absorption). This detector will not detect radiance scattered from “neighboring” beams, which is illustrated by the dotted red arrows, because these rays are in directions outside the instrument FOV. The difference in $L_d(z_1)$ and $L_d(z_2)$ determines the beam attenuation averaged over the water column from z_1 to z_2 via

$$c = -\frac{1}{z_2 - z_1} \ln \frac{L_d(z_2)}{L_d(z_1)}.$$

Now consider a plane irradiance detector measuring $E_d(z_2)$. This detector receives light from the beam directly above it, as did the $L_d(z_2)$ detector, but it also receives some of the light scattered out of neighboring beams that would otherwise miss the detector. That

is to say, part of the energy lost from one beam can be replaced by light scattered from nearby beams if the detector has a wide FOV. Thus the decrease in E_d going from depth z_1 to z_2 , which determines K_d , will be relatively less than the change in the radiances L_d over the same distance, which determine c . In other words, $c > K_d$. $c = K_d$ only in the idealized case of a perfectly collimated downward radiance (such as the Sun at the zenith in a black sky) in water with no scattering and no internal sources.

As seen in these figures, radiative transfer theory shows that K functions are very “absorption like,” meaning that *the K functions are strongly correlated with the total absorption coefficient when inelastic scatter effects are negligible*. For K_d , the approximate relation $K_d \approx a/\bar{\mu}_d$ gives close agreement between the exact (computed by HydroLight) K_d and the value estimated from the absorption coefficient and the downwelling average cosine $\bar{\mu}_d$ of the radiance distribution, which was also obtained from the HydroLight simulation.

These few simulations are enough to establish the salient features of diffuse attenuation functions. Their use has a venerable history in optical oceanography. [Smith and Baker \(1978\)](#) listed some of their virtues:

- The K s are defined as normalized depth derivatives and therefore do not require absolute radiometric measurements.
- The K s are strongly correlated with phytoplankton chlorophyll concentration (via the absorption coefficient) in Case 1 waters. Thus they provide a connection between biology and optics.
- About 90% of the diffusely reflected light from a water body comes from a surface layer of water of depth $1/K_d$, which is called the “penetration depth.” Thus K_d has implications for remote sensing ([Gordon and McCluney, 1975](#)). (Note that the penetration depth depends on wavelength.)
- Radiative transfer theory provides several useful relations between the K s and other quantities of interest, such as the absorption and beam attenuation coefficients and other AOP’s.

4.1.3 Gordon’s Normalization of K_d

As seen above, K_d does have some dependence on the Sun location and sky conditions, even though the dependence is weak. [Gordon \(1989\)](#) developed a simple way to normalize measured K_d values. His normalization for all practical purposes *removes the effects of the sea state and incident sky radiance distribution from K_d , so that the normalized K_d can be regarded as an IOP*. The theory behind the normalization is given in his paper; the mechanics of the normalizing process are as follows.

Let $E_d(\text{Sun})$ be the irradiance incident onto the sea surface due to the Sun’s direct beam, and let $E_d(\text{sky})$ be the irradiance due to diffuse background sky radiance. Then the fraction f of the direct sunlight in the incident irradiance that is transmitted through the surface into the water is

$$f = \frac{t(\text{Sun})E_d(\text{Sun})}{t(\text{Sun})E_d(\text{Sun}) + t(\text{sky})E_d(\text{sky})}.$$

Here $t(\text{Sun})$ and $t(\text{sky})$ are respectively the fractions of the direct beam and of the diffuse irradiance transmitted through the surface; these quantities can be computed using methods described in [Mobley \(1994, Chapter 4, where they are denoted by \$t\(a, w\)\$ \)](#). However,

if the solar zenith angle in air, θ_{sa} , is less than 45 degrees, then $t(\text{Sun}) \approx 0.97$. If the sky radiance distribution is roughly uniform (as it is for a clear sky), then $t(\text{sky}) \approx 0.94$. In this case, we can accurately estimate f from measurements made just above the sea surface:

$$f \approx \frac{E_{\text{d}}(\text{Sun})}{E_{\text{d}}(\text{Sun}) + E_{\text{d}}(\text{sky})}.$$

The Sun and sky irradiances are easily obtained from an instrument on the deck of a ship. When both direct and diffuse light fall onto the instrument, it records $E_{\text{d}}(\text{Sun}) + E_{\text{d}}(\text{sky})$. When the direct solar beam is blocked, the instrument records $E_{\text{d}}(\text{sky})$. (Advanced technology is not required here: just hold your hat so that its shadow falls on the instrument.)

Next compute the nadir angle of the transmitted solar beam in water, θ_{sw} , using Snell's law (Eq. 13.2):

$$\theta_{\text{sw}} = \sin^{-1} \left(\frac{\sin \theta_{\text{sa}}}{1.34} \right).$$

Finally, compute the quantity

$$D_{\text{o}} = \frac{f}{\cos \theta_{\text{sw}}} + 1.197(1 - f).$$

This value of D_{o} is valid for flat or rough sea surfaces as long as $\theta_{\text{sa}} \leq 50$ deg. For larger values of θ_{sa} , or for an overcast sky, a correction must be applied to D_{o} to account for surface wave effects on the transmitted light; the correction factors are given in [Gordon \(1989, his Fig. 6\)](#). Gordon's normalization then consists simply of dividing the measured K_{d} by D_{o} :

$$K_{\text{d}}(\text{normalized}) = \frac{K_{\text{d}}(\text{measured})}{D_{\text{o}}}.$$

Physically, D_{o} is a function (essentially $1/\bar{\mu}_{\text{d}}$) that *reduces K_{d} values to the values that would be measured if the Sun were at the zenith, if the sea surface were level, and if the sky were black* (i.e., if there were no atmosphere). The zenith-Sun, level-surface, black-sky case is the only physical situation for which $D_{\text{o}} = 1$. In other words, normalization by D_{o} removes the influence of incident lighting and sea state on K_{d} . The same normalization can be applied to depth-averaged values of K_{d} .

It is recommended that experimentalists routinely make the simple measurements necessary to determine D_{o} , because normalization of K_{d} enhances its value in the recovery of IOP's from irradiance measurements.

4.1.4 Models for K_{d}

Because E_{d} is defined as a logarithmic derivative with depth, it can be computed from a profile of $E_{\text{d}}(z, \lambda)$ measurements, and any overall calibration factor for the E_{d} instrument cancels out (Eq. 4.1). This calibration-independence was an important consideration in the early days of optical oceanography when there were no well-calibrated commercial instruments available. Researchers made their own instruments, and sophisticated radiometric calibration facilities were available at only the largest laboratories. This was the situation until the 1960s. Even in the 1960s, spectral $E_{\text{d}}(z, \lambda)$ measurements were so difficult to make that an entire book ([Tyler and Smith, 1970](#)) was devoted to describing a spectral

irradiance instrument made at Scripps Institution of Oceanography in the 1960s and showing some of the irradiance data collected by this instrument in a variety of waters. Some of these data were shown in Section 2.1.

Historically, the value of K_d lay in Eqs. (4.2) and (4.3). If a measurement or model of K_d were available along with a solar spectrum at the sea surface, then the irradiance within the water could be computed. Thus there was an early interest in developing models of $K_d(\lambda)$ parameterized by either the chlorophyll concentration or a value of $K_d(\lambda_o)$ at some reference wavelength.

After the invention of the Argon-ion laser (emitting at 488 nm) and the Nd:YAG laser (emitting at 532 nm when frequency doubled), both in 1964, it did not take long for their application to oceanographic military problems. These first two of these applications were detection of submarines via airborne lidar bathymetry and communication with submarines using airborne or spaceborne lasers. For clear open-ocean waters, 488 nm was the wavelength of choice; for greener coastal waters, 532 nm gave deeper penetration. The same wavelengths are used today for lidar bottom bathymetry and for numerous other uses, especially in medicine, including surgery.

An accurate prediction of the return power from a laser pulse requires knowledge of several quantities, including the VSF at 180 deg ($\beta(\psi = \pi)$, Section 6.4.4) and the beam spread function (Section 9.5.2). However, before there were instruments for measurement of those quantities, a rough estimate of the depth the laser light would penetrate could be obtained from $1/K_d$. For a 488 nm laser, one thus needs the value of $K_d(488)$ for the water body under consideration. By good fortune, the Wratten #75 gelatin filter has a spectral transmission that peaks at 490 nm, as seen in Fig. 4.8. An instrument with this filter gives an approximation of K_d centered on 490 nm with a FWHM of about 20 nm. This was close enough to 488 for the lidar applications, so “Kd490” became a standard AOP to be measured or modeled.

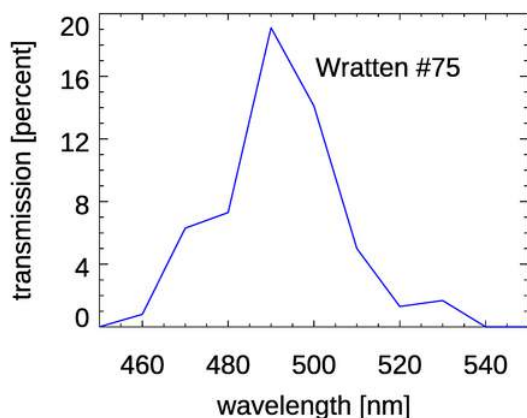


Figure 4.8: Spectral response of the Wratten #75 filter at 10 nm resolution. Data from Eastman Kodak Co. (1920).

As soon as the Coastal Zone Color Scanner was launched in 1978 (Chapter 14), there was interest in developing an algorithm to retrieve Kd490 from space. Austin and Petzold (1981) used the CZCS bands at 443 and 550 nm to develop an empirical band-ratio algorithm,

$$K_d(490) = 0.0883 \left[\frac{L_u(443)}{L_u(550)} \right]^{-1.491} + 0.022 \quad [\text{m}^{-1}]. \quad (4.4)$$

Here the 0.022 is the value of $K_d(490)$ for pure water. Figure 4.9 shows the 88 data points used to develop Eq. (4.4) and the values predicted by the model. They also developed an algorithm of the same form for $K_d(520)$; only the values of the best-fit coefficients are different.

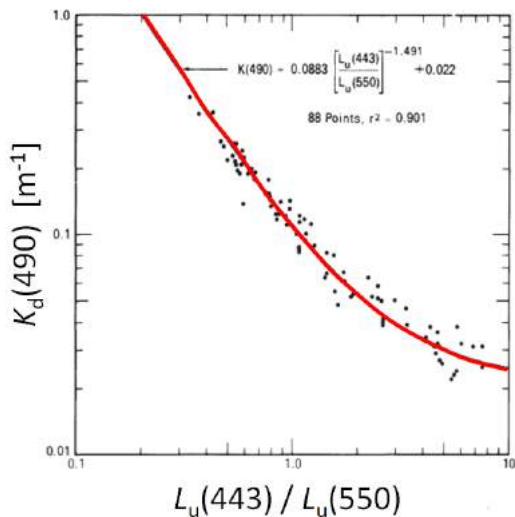


Figure 4.9: Dependence of measured $K_d(490)$ values on the ratio of $L_u(443)$ to $L_u(550)$. The red line is Eq. (4.4). Redrawn from Austin and Petzold (1981, Fig. 1).

Austin and Petzold (1986) presented a model for $K_d(\lambda)$ at any wavelength between 350 and 700 nm, given the value at a reference wavelength λ_o :

$$K_d(\lambda) = \frac{M(\lambda)}{M(\lambda_o)} [K_d(\lambda_o) - K_w(\lambda_o)] + K_w(\lambda). \quad (4.5)$$

The values of $M(\lambda)$ and $K_w(\lambda)$ are tabulated for 350(5)700 nm in their paper. They comment that this model is applicable for “oceanic or moderately clear coastal water, for example Jerlov type I through III oceanic or type 1 coastal, or if $K_d(490) < 0.16 \text{ m}^{-1}$.” Picking $\lambda_o = 490$ nm, a satellite-derived value of $K_d(490)$ (initially obtained from the CZCS algorithm 4.4) enables mapping of $K_d(\lambda)$ for the global ocean.

Morel (1988) developed a model of $K_d(\lambda)$ in Case 1 water as a function of the chlorophyll concentration:

$$K_d(\lambda) = \chi(\lambda) Chl^{e(\lambda)} + K_w(\lambda). \quad (4.6)$$

Here K_d is in m^{-1} when Chl is in mg m^{-3} . The values of the fitting coefficients $\chi(\lambda)$ and $e(\lambda)$ are tabulated in his paper for 400(5)700 nm. Figure 4.10 gives examples of the performance of Eq. (4.6) when compared with measured data.

At 490 nm, the Morel model gives

$$K_d(490) = 0.0690 Chl^{0.702} + 0.0217. \quad (4.7)$$

Morel did not study the depth dependence of K_d as given by Eq. (4.6) when the chlorophyll concentration is depth-dependent. However, he did develop a formula for K_{PAR} averaged over the depth of the euphotic zone, z_{eu} . If taken to be the 1% PAR level, Morel found that the depth of the euphotic zone is

$$z_{eu} = 38.0 Chl^{-0.428}. \quad (4.8)$$

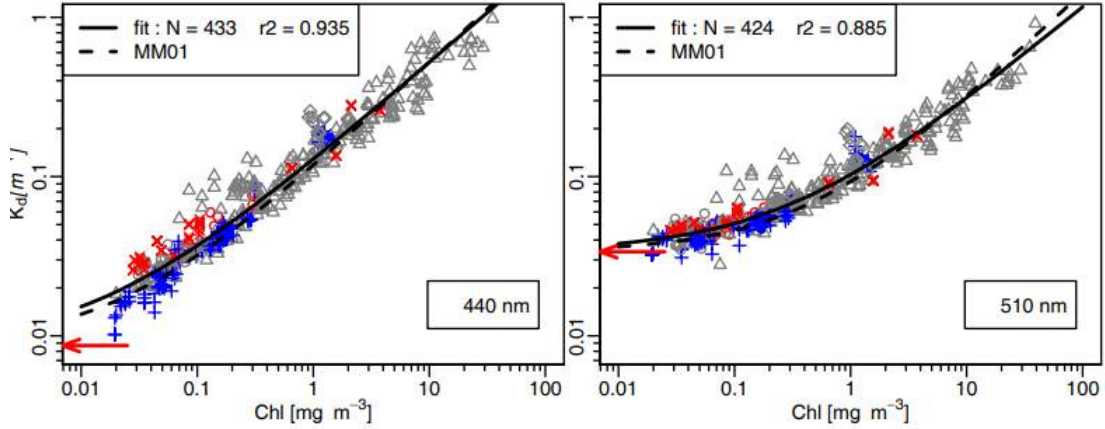


Figure 4.10: Comparison of Eq. (4.6) with measurements. The red arrows are the values of K_w . Figures extracted from Morel et al. (2007, Fig. 1) and reproduced by a Creative Commons license.

This in turn implies that

$$\langle K_{\text{PAR}} \rangle = 0.212 \text{ Chl}^{0.428}, \quad (4.9)$$

where $\langle \dots \rangle$ indicates an average from depth 0 to z_{eu} . It must be remembered that Morel's models are valid only for Case 1 waters (Section 4.4.3).

Lee et al. (2013) developed a model for K_d as a function of the absorption and backscattering coefficients and the solar zenith angle:

$$\begin{aligned} K_d(\lambda) &= (1 + m_0 \theta_{\text{Sun}}) a(\lambda) + m_1 \left[1 - \gamma \frac{b_{\text{bw}}}{b_{\text{b}}} \right] \left[1 - m_2 e^{-m_3 a(\lambda)} \right] b_{\text{b}} \\ &= (1 + 1.005 \theta_{\text{Sun}}) a(\lambda) + 4.259 \left[1 - 0.265 \frac{b_{\text{bw}}}{b_{\text{b}}} \right] \left[1 - 0.52 e^{-10.8 a(\lambda)} \right]. \end{aligned} \quad (4.10)$$

The values of the five model coefficients γ and m_0 to m_3 were determined by finding the best fit of the model to a large database of HydroLight-generated K_d spectra. The model was then tested by application to global data sets of measured IOPs and K_d profiles. Figure 4.11 shows example results.

$K_d(490)$ retains its importance as a key parameter for understanding the state of the ocean ecosystem, and it is a standard output in NASA's suite of ocean color products. Just as for CZCS with Eq. (4.4), the estimation of $K_d(490)$ still uses a ratio of blue to green wavelengths. The current NASA $K_d(490)$ algorithm has the form (NASA $K_d(490)$ algorithm)

$$\begin{aligned} \log_{10}(K_{\text{bio}}(490)) &= a_0 + \sum_{i=1}^4 a_i \left[\log_{10} \left(\frac{R_{\text{rs}}(\lambda_{\text{blue}})}{R_{\text{rs}}(\lambda_{\text{green}})} \right) \right]^i \\ K_d(490) &= K_{\text{bio}}(490) + 0.0166. \end{aligned}$$

The exact wavelengths used and the values of the five fitting coefficients a_i depend on the particular sensor. Their values are given on the NASA website.

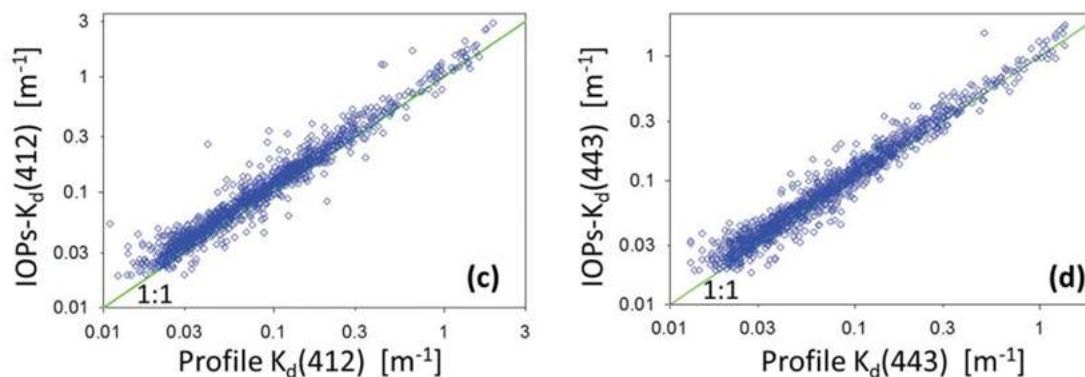


Figure 4.11: Comparison of Eq. (4.10) with measurements from the NOMAD data set (NASA bio-Optical Marine Algorithm Dataset). The abscissa is the value of K_d from the dataset, and the ordinate is the value predicted by Eq. (4.10). Figures extracted from Lee et al. (2013, Fig. 5).

4.2 Reflectances

Various reflectances are probably the most commonly used apparent optical properties because they are fundamental to remote sensing of the oceans. In the early days of ocean color remote sensing, algorithms were developed to relate the irradiance reflectance R to quantities such as absorption and backscatter coefficients and chlorophyll concentrations (e.g., Morel and Prieur, 1977; Gordon and Morel, 1983). More recently, the remote-sensing reflectance R_{rs} has become the AOP of choice for remote sensing of ocean properties (e.g., O'Reilly et al., 1998; Mobley et al., 2005). This section considers each of these reflectances.

4.2.1 The Irradiance Reflectance R

The *spectral irradiance reflectance* (or *irradiance ratio*), $R(z, \lambda)$, is defined as the ratio of spectral upwelling to downwelling plane irradiances:

$$R(z, \lambda) \triangleq \frac{E_u(z, \lambda)}{E_d(z, \lambda)}.$$

$R(z, \lambda)$ is thus a measure of how much of the radiance traveling in all downward directions is reflected upward into any direction, as measured by a cosine collector. This is illustrated in Fig. 4.12. Depth z can be any depth within the water column, or in the air just above the sea surface.

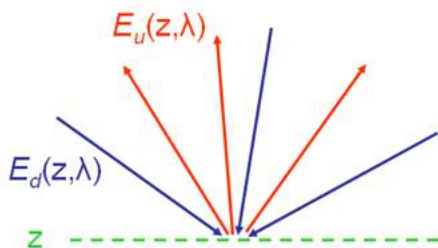


Figure 4.12: Illustration of light rays contributing to the irradiance reflectance R .

Irradiance reflectance has the virtue that it can be measured by a single, uncalibrated, plane irradiance detector. The downwelling irradiance E_d can be measured, and then the detector can be turned “upside down” to measure E_u . The calibration factor needed to convert from detector units (voltage, current, or digital counts) to irradiance units ($\text{W m}^{-2} \text{nm}^{-1}$) cancels out.

The irradiance measurements of Tyler and Smith (1970) give good examples of irradiance reflectances. Figure 4.13 shows R at 5 m depth computed from the Crater Lake irradiances of Fig. 2.2, R at 1 m depth computed from the San Vicente Reservoir data of Fig. 2.3, and the HydroLight-computed R at 5 m for optically pure water. The Crater Lake spectrum is largest near 400 nm and greater than 0.1, which indicates that this water is a very “bright” and deep blue. Again, we see that Crater Lake is similar to optically pure water. The San Vicente reflectance peaks near 590 nm, where the magnitude is only about 0.03. Thus this water appears as a dark yellowish green. The San Vicente reflectance is low at blue wavelengths because of the high absorption by colored dissolved organic matter.

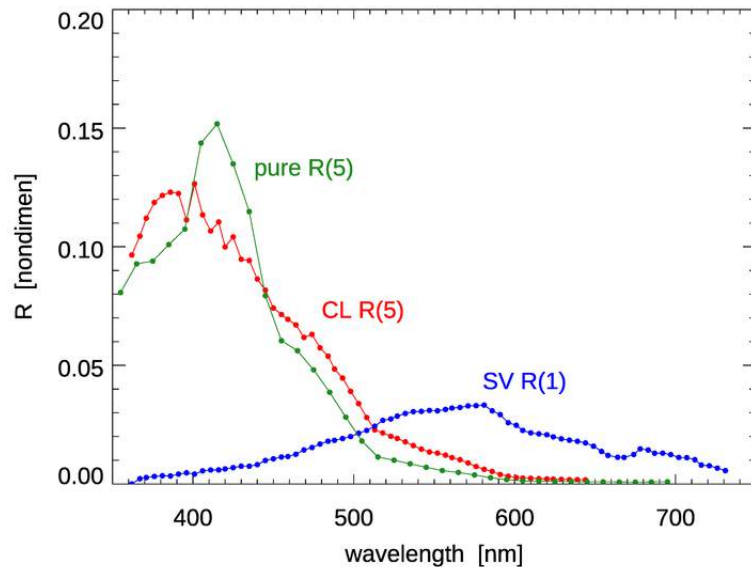


Figure 4.13: Irradiance reflectances for Crater Lake at 5 m depth (red), for San Vicente Reservoir at 1 m (blue), and for optically pure water at 5 m (green) as computed by HydroLight. The Crater Lake and San Vicente spectra are computed from the irradiances in Tyler and Smith (1970).

It is clear from Fig. 4.13 that the irradiance reflectance depends on the water IOPs. One of the pioneering papers on the use of R spectra to obtain IOPs is Roesler and Perry (1995). They first developed a model for $R(\lambda) \propto b_b/a$, where b_b and a are the total backscatter and absorption coefficients. These IOPs were then written as sums of contributions by water, phytoplankton, dissolved substances, and non-living particles. The resulting model was then forced to fit measured R spectra, whereby the best fit determined the concentrations of the various components.

4.2.2 The Remote-Sensing Reflectance R_{rs}

The *spectral remote-sensing reflectance* R_{rs} is defined as

$$R_{rs}(\theta, \phi, \lambda) \triangleq \frac{L_w(\text{in air}, \theta, \phi, \lambda)}{E_d(\text{in air}, \lambda)} \quad (\text{sr}^{-1}).$$

Here the depth argument of “in air” indicates that R_{rs} is evaluated just above the water surface using the water-leaving radiance L_w and E_d in the air. Water-leaving radiance refers to downwelling light that has entered the water body from the air, been scattered into upward directions within the water, and then been transmitted through the water surface back into the air. The remote-sensing reflectance is thus a measure of how much of the downwelling radiance that is incident onto the water surface in any direction (as measured by a plane irradiance sensor) is eventually returned through the surface into a small solid angle $\Delta\Omega$ centered on a particular direction (θ, ϕ) , as illustrated in Fig. 4.14.

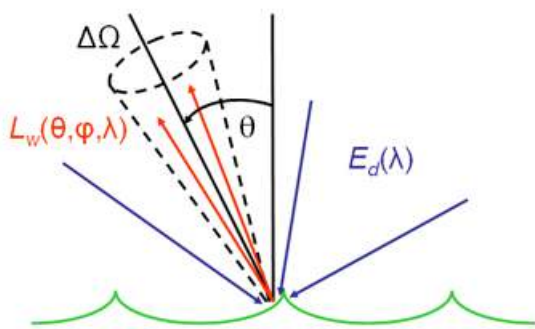


Figure 4.14: Illustration of light rays contributing to the remote-sensing reflectance R_{rs} .

Although R_{rs} is often computed for nadir-viewing directions only, in actual remote sensing it is usually an off-nadir direction that is being observed by an airborne or satellite remote sensor. As shown next, R_{rs} has the virtue that it is less sensitive than R to environmental conditions such as Sun zenith angle or sky conditions. This is the reason that R_{rs} has replaced R for remote sensing. However, determination of R_{rs} is more difficult than R . First, the measurements of L_u and E_d require different sensors, which must be accurately calibrated. Second, the water leaving radiance L_w cannot be measured directly. Only the total upwelling radiance L_u above the surface can be measured. This L_u is the sum of the water-leaving radiance L_w and the downward Sun and sky radiance that is reflected upward by the sea surface, L_r , as illustrated in Fig. 4.15. L_w therefore must be *estimated* either from a measurement of the total upwelling radiance L_u made above the sea surface, or from a measurement of L_u made at some distance below the sea surface and extrapolated upward through the surface. Each of these estimation methodologies has arguments for and against its use (e.g., Mobley, 1999; Toole et al., 2000).

$$L_u(\text{in air}, \theta, \phi, \lambda) = L_u(\theta, \phi, \lambda) + L_r(\theta, \phi, \lambda)$$

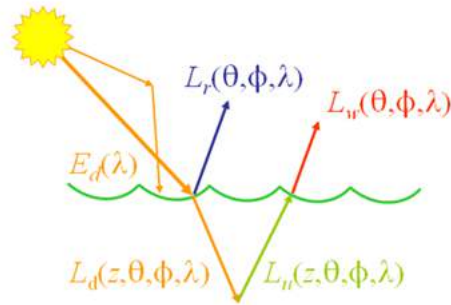


Figure 4.15: Illustration of light rays contributing to L_u as measured above the sea surface.

4.2.3 Dependence of R and R_{rs} on IOPs and Environmental Conditions

As a first illustration of the dependence of R on IOPs and sky conditions, the HydroLight (Section 10.6) radiative transfer numerical model was run using an IOP model for homogeneous Case 1 waters with chlorophyll concentrations of $Chl = 1$, and 5 mg m^{-3} . For each chlorophyll concentration, runs were made for solar zenith angles of 0, 30, and 60 deg in a clear sky, and for a heavily overcast sky. Figure 4.16 shows the resulting depth profiles. The $R(z)$ profiles for the different chlorophyll concentrations are well separated. Near the sea surface, there is a weaker dependence of R on sky conditions. However, those differences disappear as the depth increases and multiple scattering “erases” the effects of different Sun locations or of the background sky. Just as for the K functions, at great depth the R values for the different sky conditions approach an asymptotic value, R_∞ , which is determined only by the IOPs.

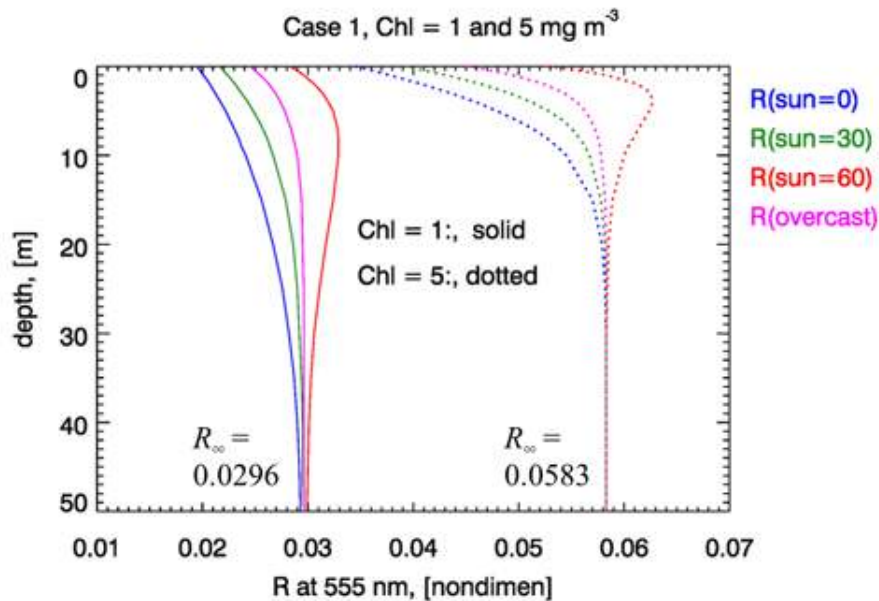


Figure 4.16: Illustration of the depth dependence of R on IOPs and sky conditions.

To illustrate the dependence of near-surface R and R_{rs} on IOPs and external environmental conditions, HydroLight was run using an IOP model for Case 1 waters with chlorophyll concentrations of $Chl = 0.1, 1.0,$ and 10.0 mg m^{-3} . For each chlorophyll concentration, runs were made for three sets of sky conditions: (1) a level sea surface (wind-speed $U = 0$) and the Sun at the zenith ($\text{Sun} = 0$) in a clear sky; (2) a rough sea surface with a wind speed of $U = 10 \text{ m s}^{-1}$ and the Sun at a 50 deg zenith angle ($\text{Sun} = 50$) in a clear sky; (3) a wind speed of 10 m s^{-1} and a heavily overcast sky (overcast) for which the Sun's location cannot be discerned.

Figure 4.17 shows the resulting R spectra at depth $z = 0$, which is in the water just below the mean sea surface. This is the quantity most often used to develop remote-sensing algorithms relating R to IOPs or chlorophyll concentrations. The curves for the different chlorophyll concentrations group together, showing that the shapes of the R spectra are determined primarily by the different IOPs associated with the different chlorophyll concentrations. However, there is a significant effect of the sky conditions on the R spectra within each of the three chlorophyll groups, as was seen in Fig. 4.16 just below the surface.

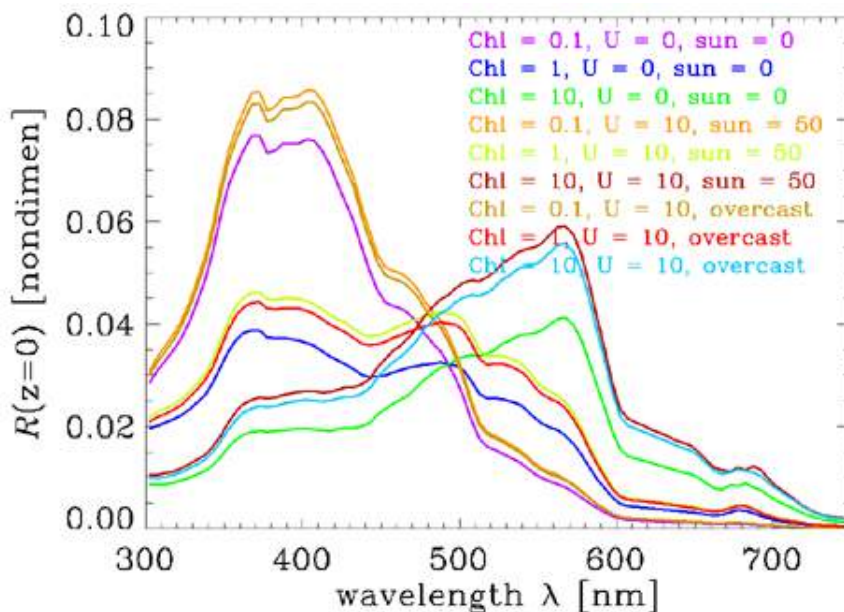


Figure 4.17: Dependence of R on chlorophyll concentration, sky condition, and wind speed for selected conditions in Case 1 water.

Figure 4.18 shows the nadir-viewing R_{rs} spectra for the same set of HydroLight runs. The three chlorophyll groups are similar in shape to the corresponding R spectra, but there is much less variability in the R_{rs} spectra due to the external environmental conditions. R_{rs} is thus a better AOP than is R , because R_{rs} is less sensitive to the sky conditions while remaining very sensitive to the different IOPs corresponding to the different chlorophyll concentrations.

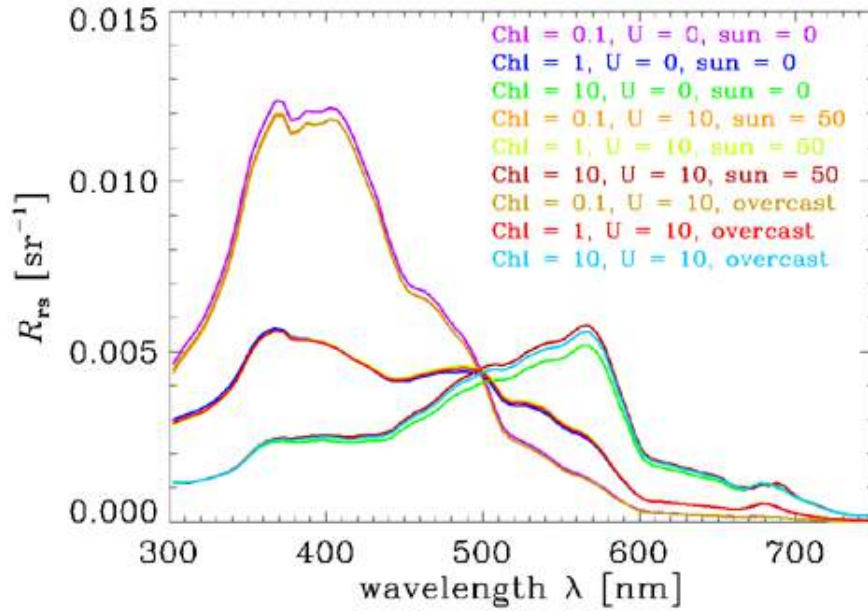


Figure 4.18: Dependence of R_{rs} on chlorophyll concentration, sky condition, and wind speed for selected conditions in Case 1 water.

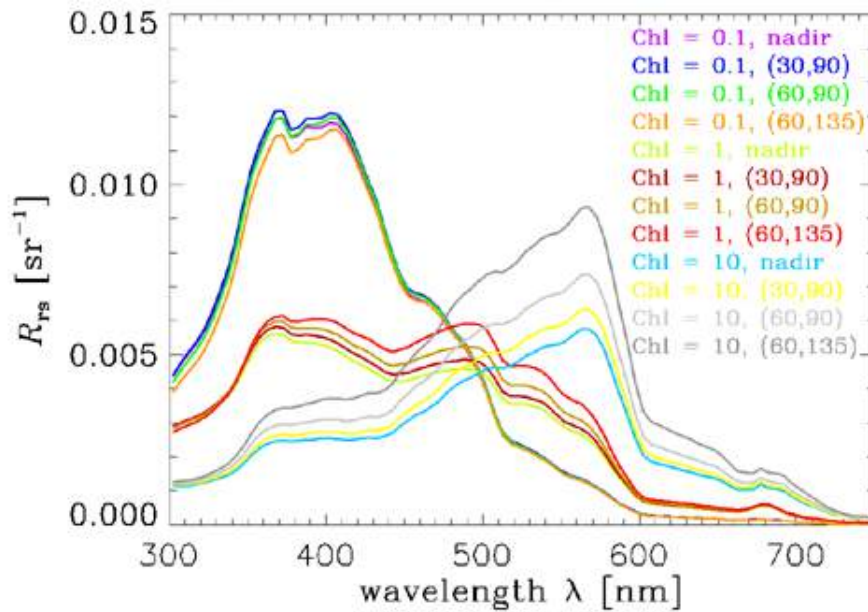


Figure 4.19: Dependence of R_{rs} on chlorophyll concentration in Case 1 water and viewing direction, for a wind speed of 10 m s^{-1} and a solar zenith angle of 50 deg.

Figure 4.19 shows the R_{rs} spectra for nadir ($\theta_v = 0$) and various off-nadir (θ_v, ϕ_v) viewing directions. Azimuthal angle $\phi_v = 90$ deg corresponds to looking at right angles to the Sun's azimuthal direction, and $\phi_v = 135$ deg is looking half-way between normal to the Sun and away from the Sun. This range of ϕ_v values is what is usually observed in remote

sensing to avoid Sun glint from the sea surface. At the lowest chlorophyll concentration, there is not much difference in the spectra for the different viewing directions. However, the differences increase with increasing chlorophyll concentration, and are quite significant for the ($\theta_v = 60$) curves when $Chl = 10 \text{ mg m}^{-3}$. These differences in off-nadir directions for different chlorophyll values are a consequence of the changes in shape and relative importance of the scattering phase functions for the small and large chlorophyll-bearing particles versus that of water as the chlorophyll concentration increases.

As Figs. 4.18 and 4.19 show, R_{rs} is much more sensitive to water IOPs than to external environmental conditions and viewing direction. However, R_{rs} still does depend somewhat on solar zenith angle (Fig. 4.18) and viewing direction (Fig. 4.19). An even better AOP would be obtained if these remaining dependencies can be removed. The resulting AOP is called the *exact normalized water-leaving reflectance*, denoted by $[\rho_w(\lambda)]_N^{ex}$. This reflectance is based on the concept of the normalized water-leaving radiance, which is defined to be “...the radiance that could be measured by a nadir-viewing instrument, if the Sun were at the zenith in the absence of any atmospheric loss, and when the Earth is at its mean distance from the Sun” (Morel and Gentili, 1996, page 4852). (Earlier papers often used phrases like “in the absence of an atmosphere,” implying that the atmosphere is completely removed. This was found to be too extreme, so the current definition and calculations are based on a standard but non-attenuating atmosphere.) The computation and interpretation of $[\rho_w(\lambda)]_N^{ex}$ can be rather subtle. These matters are discussed in detail in Section 15.3 on normalized reflectances.

When processing satellite ocean color imagery, measured top-of-the-atmosphere radiances are converted by the process of *atmospheric correction* to $[\rho_w(\lambda)]_N^{ex}$ spectra, which can then be used in algorithms to retrieve geophysical quantities such as the chlorophyll concentration. However, when running HydroLight, $[\rho_w(\lambda)]_N^{ex}$ can be obtained by putting the Sun at the zenith, in which case $[\rho_w(\lambda)]_N^{ex}$ is π times the nadir-viewing R_{rs} :

$$[\rho_w(\lambda)]_N^{ex} = \pi R_{rs}(\text{HydroLight}; \theta_s = 0, \theta_v = 0). \quad (4.11)$$

The remote-sensing reflectance reported by NASA as a standard output for sensors such as MODIS or VIIRS is sometimes described as $[\rho_w(\lambda)]_N^{ex}/\pi$, which is equivalent to the HydroLight-computed $R_{rs}(\theta_s = 0, \theta_v = 0)$.

$[\rho_w(\lambda)]_N^{ex}$ or its equivalent R_{rs} are now used for most remote sensing. However, there are many other measures of reflectance, which have other applications. These are discussed in Sections 13.6 on the BRDF and 13.8 on Other Measures of Reflectance.

4.3 Average Cosines of the Radiance Distribution

A third family of AOPs (after K functions and reflectances) is the *average cosines of the radiance distribution*, also called the mean cosines. The average cosine of the downwelling radiance distribution is the average of the cosine of the polar angle θ weighted by the radiance in direction θ . This is computed as as

$$\bar{\mu}_d \triangleq \frac{\int_0^{2\pi} \int_0^{\pi/2} L(\theta, \phi) \cos \theta \sin \theta d\theta d\phi}{\int_0^{2\pi} \int_0^{\pi/2} L(\theta, \phi) \sin \theta d\theta d\phi} = \frac{E_d}{E_{od}}. \quad (4.12)$$

The second form follows from the definitions of E_d and E_{od} seen in Eqs. (1.22) and (1.24), respectively. A corresponding equation with the θ integration from $\pi/2$ to π leads to

$\bar{\mu}_u = E_u/E_{ou}$, the mean cosine of the upwelling radiance distribution. The mean cosine of the entire radiance distribution is defined by

$$\bar{\mu} \triangleq \frac{\int_0^{2\pi} \int_0^\pi L(\theta, \phi) \cos \theta \sin \theta d\theta d\phi}{\int_0^{2\pi} \int_0^\pi L(\theta, \phi) \sin \theta d\theta d\phi} = \frac{E_d - E_u}{E_o}. \quad (4.13)$$

The minus sign on E_u arises because $\cos \theta < 0$ for upwelling directions, $\pi/2 < \theta \leq \pi$. Note that although $E_o = E_{od} + E_{ou}$, $\bar{\mu} \neq \bar{\mu}_d + \bar{\mu}_u$.

The average cosines of the radiance distribution give a crude measure of the directional nature of the radiance distribution, as illustrated in Fig. 4.20. It is easy to see that the *distribution function* $D_d \triangleq 1/\bar{\mu}_d$ is the average increase in the distance rays travel in going from one depth to another. A large $\bar{\mu}_d$ indicates that the radiance is heading mostly in the nadir direction and will therefore penetrate more deeply than a radiance with a small $\bar{\mu}_d$, for which more of the radiance is in off-nadir directions. Older models for irradiance penetration into the ocean sometimes used approximations such as (Sathyendranath and Platt, 1988)

$$K_d \approx \frac{a + b_b}{\bar{\mu}_d}$$

to compute in-water irradiances for use in ecosystem models. There was thus interest in modeling $\bar{\mu}_d$ as a function of the IOPs (Sathyendranath and Platt, 1991). However, the average cosines are seldom used today, given the ease of numerically computing in-water light fields, e.g., with the EcoLight-S numerical model, which is designed for use in ecosystem models (Mobley, 2011).

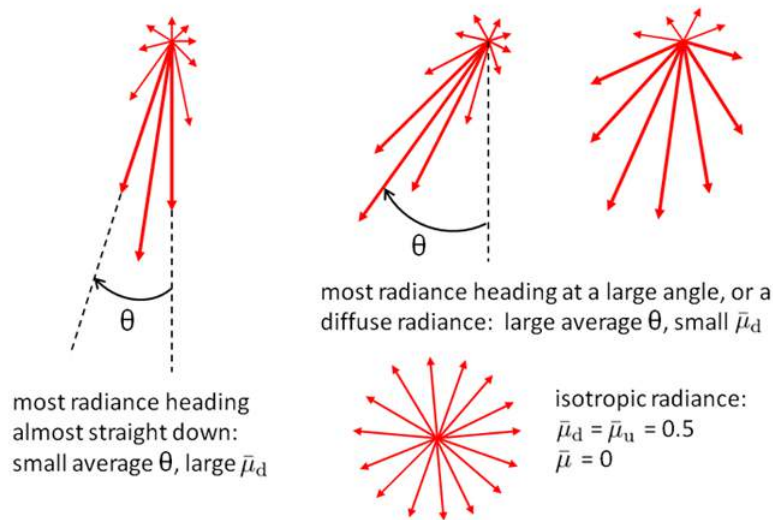


Figure 4.20: Illustration of average cosines as a measure of the directional nature of the radiance distribution.

Figure 4.21 gives an example of the dependence of the average cosines on depth for given chlorophyll concentrations in Case 1 water and on solar zenith angle. These profiles are from the same HydroLight runs that generated Fig. 4.16. Just as was seen for K functions and the irradiance reflectance R , the average cosines approach asymptotic values

at large depths. As a very rough rule, for deep waters and typical sky conditions, $\bar{\mu}_d$ is usually in the range of 0.6 to 0.8, and $\bar{\mu}_u$ is in the range of 0.3 to 0.4. However, values outside these ranges can occur. Consider, for example, shallow water with a Lambertian bottom (Section 13.7), for which the bottom-reflected, upwelling radiance is isotropic. In this case, $\bar{\mu}_u = 0.5$ very near the bottom.

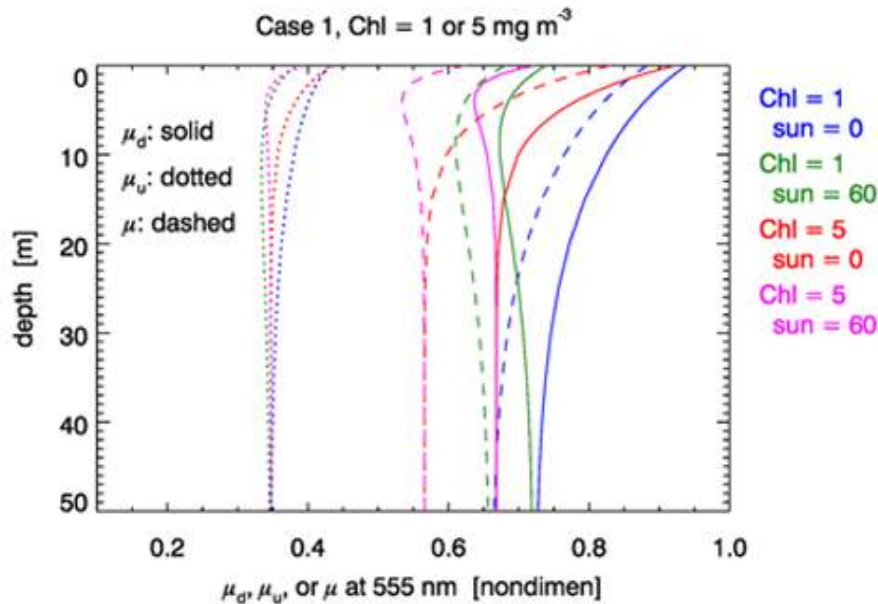


Figure 4.21: Illustration of the depth dependence of the average cosines of the radiance distribution on chlorophyll concentration in Case 1 water and on solar zenith angle.

4.4 Optical Classification of Water Types

[Emmanuel Boss and Curtis Mobley contributed to this section.]

The classification of natural waters into optical water types, as with many other taxonomies, is done in order to generalize and systematize the science of ocean color. [Arnone et al. \(2004\)](#), in a review paper on water mass classification, credits the Secchi disk (Section 16.7), invented around 1865, as being the first quantitative optical instrument used to measure water transparency and hence provide the ability to differentiate water masses based on their optical properties. Classification schemes that are based on water color are primarily based on absorption (scattering contributes to the overall brightness of the water, but less to its color). The most common classification schemes for water types are based on

- The human eye: visually compare the water color to the color of a standard color sample (Forel-Ule color scale)
- Analysis of AOP spectra (Jerlov water types)
- The relative importance of Chlorophyll in determining the optical properties of the water (Case 1 versus Case 2)

- Analysis of inherent optical properties (ternary diagrams)

4.4.1 The Forel-Ule Color Classification

The Forel-Ule Scale classifies water bodies based on the visual color of the water as seen by a normal human eye.

Different inorganic compounds with standard formulas are used to produce a color palette with twenty one fluid vials ranging from blue to green to brown. The analysis consists of finding the vial whose color best matches the color of the water body; the result is an index from 1 to 21. The method is often used in conjunction with the Secchi disk and is included in data bases such as [NOAA's World Ocean Database](#). The method was developed in the 1890s by Francois-Alphonse Forel and William Ule. Although now replaced by instruments for quantitative work, the Forel-Ule color measurements are still of interest because of the ease of making measurements and the large database of historical observations, which has value for studies of long-term changes in ocean ecosystems (e.g., [Wernand, 2011](#)). There is now even a [smart-phone app for making Forel-Ule measurements](#). Figure 4.22 shows the use of a Forel-Ule color scale to determine the color of the ocean by visually determining the closest match of the water color to one of the 21 Forel-Ule colors.

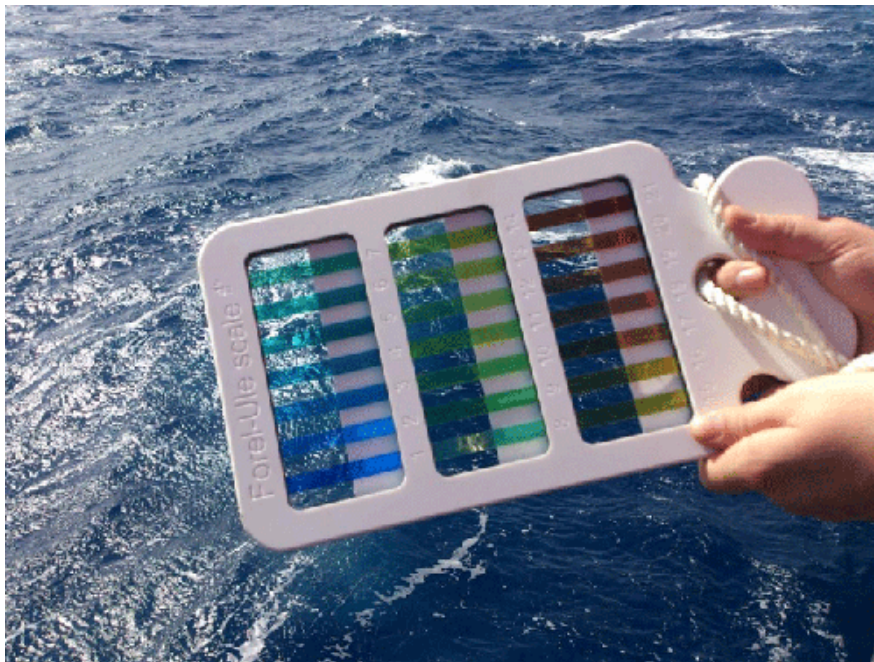


Figure 4.22: Using a Forel-Ule color scale to determine the color of the ocean. Photo from [ICBM Uni. Oldenburg](#).

Figure 4.23 shows the locations of the 21 Forel-Ule colors on a CIE chromaticity diagram (Section 16.3). The open circle is the white point. The open diamond is the CIE color of the remote-sensing reflectance R_{rs} as calculated by HydroLight using the “new Case 1” IOP model (Section 8.9) with a chlorophyll concentration of 5 mg m^{-3} . The closest Forel-Ule value is 6.

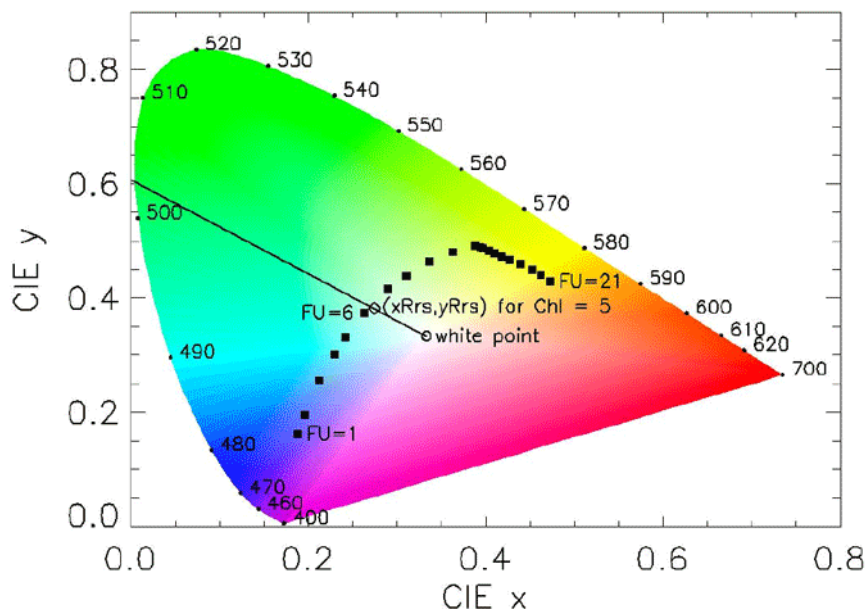


Figure 4.23: The locations of the 21 Forel-Ule colors on a CIE chromaticity diagram (black squares). The open diamond is the color of R_{rs} as calculated by HydroLight for a chlorophyll concentration of 5 mg m^{-3} .

Another comparative scale is the Hazen Platin-Cobalt-Scale, adopted as the American Public Health Association (APHA) color scale (described in [ASTM D1209](#), “Standard Test Method for Color of Clear Liquids (Platinum Cobalt Scale)”). It is based on diluting a standard stock solution (500ppm of PtCo) and finding the dilution whose color best matches the fluid compared.

4.4.2 The Jerlov Classification

The Jerlov scheme classifies water bodies based on the shape of AOP spectra, most commonly the shape of $K_d(\lambda)$.

[Jerlov \(1968\)](#) introduced a classification of water bodies based on the the transmittance per meter of downwelling spectral plane irradiance in surface waters. This is equivalent to using the shape of the spectral diffuse attenuation $K_d(\lambda)$ averaged over some depth interval. Since ocean color is proportional to $1/a \propto 1/K_d$ these spectra are linked to the color observed. Jerlov discretized his observations into a set of five typical open-ocean spectra, labeled I, IA, IB, II, and III, and nine typical coastal spectra, labeled 1 to 9 (e.g., Tables XXVI and XXVII in [Jerlov \(1968\)](#); a spreadsheet of these data can be downloaded from the Web Book.). The left panel of Fig. 4.24 shows the spectra defining the Jerlov types in terms of the irradiance transmission through 1 m of water. The right panel shows the corresponding spectra as the averages of K_d over depths 0 to 10 m.

[Morel \(1988\)](#) gives a very approximate correspondence between the open-ocean Jerlov types and the chlorophyll concentration, which is seen in Table 4.2.

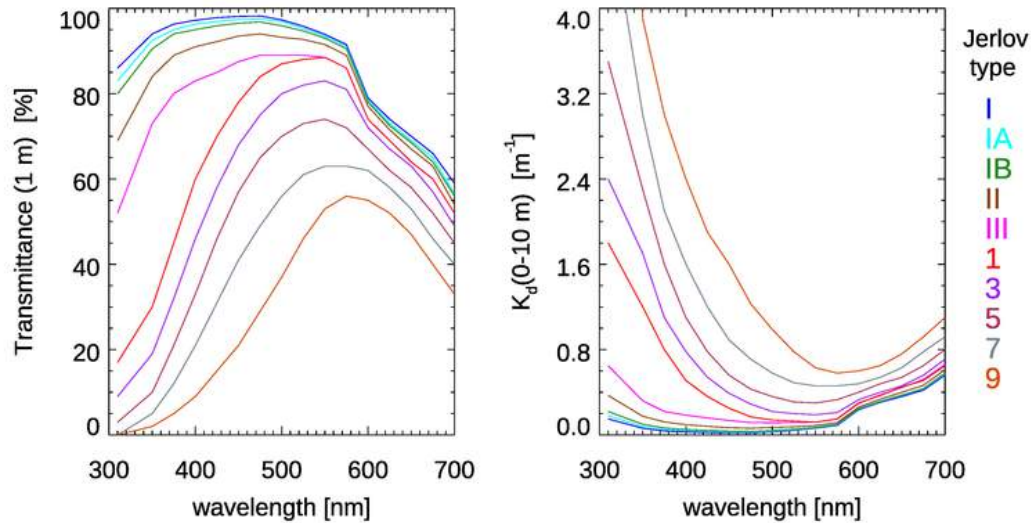


Figure 4.24: Jerlov’s water mass classification based on transmittance through 1 m of sea water (left pane) or based on the average of K_d over depths 0 to 10 m. Data from Jerlov (1976, Tables XXVI and XXVII).

Chlorophyll (mg m^{-3})	0-0.01	0.05	0.1	0.5	1.5-2.0
jerlov water type	I	IA	IB	II	III

Table 4.2: Approximate correspondence between the chlorophyll concentration and the Jerlov open-ocean water type.

Solonenko and Mobley (2015) give tables relating the Jerlov water types to a consistent set of absorption and scattering coefficients (consistent in the sense that the IOPs, when used in HydroLight, generate K_d spectra that correspond to the different Jerlov water types).

4.4.3 Case 1 versus Case 2 Water Types

The Case 1 vs Case 2 classification is based on the importance of chlorophyll versus other components in determining the optical properties of the water.

Morel and Prieur (1977) described the spectral shape of $R = E_u/E_d$ and its changes as separating two types of waters. They found that “...two extreme cases can be identified and separated. Case 1 is that of a concentration of phytoplankton high compared to other particles. In contrast, the inorganic particles are dominant in Case 2. In both cases dissolved yellow substance is present in variable amounts. An ideal Case 1 would be a pure culture of phytoplankton and an ideal Case 2 a suspension of nonliving material with a zero concentration of pigments.”

Morel and Prieur emphasized that these ideal cases are not encountered in nature, and they suggested the use of high or low values of the ratio of pigment concentration to

scattering coefficient as a basis for discriminating between Case 1 and Case 2 waters.

The definitions of Case 1 and Case 2 have evolved over the years into the ones commonly used today (Gordon and Morel, 1983; Morel, 1988):

- Case 1 waters are those waters whose optical properties are determined primarily by phytoplankton and co-varying colored dissolved organic matter (CDOM) and detritus degradation products.
- Case 2 waters are everything else, namely waters whose optical properties are significantly influenced by other constituents such as mineral particles, CDOM, or microbubbles, whose concentrations do not covary with the phytoplankton concentration.

In Case 1 waters, several changes occur as the pigment concentration increases:

- The R values in the blue-violet region decrease progressively and a minimum is formed around 440 nm, which corresponds to the maximum absorption of chlorophyll. The maximum shifts toward 565-570 nm, which is the wavelength where simultaneously the absorption due to pigments is at its minimum and the absorption due to the water itself rapidly increases.
- The second maximum of absorption by chlorophyll *a* in vivo creates an R minimum near 665 nm. This minimum is only slightly marked because the increase in absorption due to the presence of chlorophyll remains weak compared with the absorption due to water itself.
- At 685 nm a second reflectance maximum appears due to fluorescence by chlorophyll superimposed on absorption/scattering interaction near the 676 nm absorption band.
- A hinge point in R values in the 560-640 nm band is observed with relatively little change as chlorophyll varies.

In Case 2 waters inorganic particles are relatively more dominant than phytoplankton. As the turbidity increases the following modifications appear:

- The reflectance values are generally higher than for Case 1 throughout the spectrum and of a different shape. There is no longer, as in Case 1, a minimum at 440 nm. On the contrary, the curves become convex between 400 and 560 nm (inverted “U” shape rather than inverted “V”). The maximum is flatter than in Case 1, but located at the same wavelength, 560 nm.
- R values become higher as turbidity increases. This is opposite of what was observed in Case 1, particularly for wavelengths above 550 nm.
- As a result of the flat shape of the R curves, the dominant wavelength does not shift beyond 510 nm. These waters are blue-green or green with a bright and milky appearance due to the combined effects of high irradiance values and of low purity values.

Although the Case 1 versus Case 2 distinction has proven very useful, it can also be misleading. It must be understood that natural waters are not simply either Case 1 or Case 2; there is a continuous gradation of water optical properties (whether IOPs or AOPs) and

their dependence on chlorophyll and other water constituents. It must also be remembered that “Case 1” and “Case 2” are *not* synonyms for “open ocean” and “coastal” waters.

Mobley *et al.* (2004) have even suggested that it is perhaps time to drop this type of binary classification because it sometimes obfuscates more than it helps. For example, the same water can be Case 1 for absorption but Case 2 for scattering (e.g., if the water contains non-absorbing but highly scattering mineral particles), or it can be Case 2 for absorption and Case 1 for scattering (e.g., if the water contains a high concentration of highly absorbing terrigenous CDOM, which does not covary with the chlorophyll, and which does not affect the scattering). Of course, water in the same area may be Case 1 during some time of the year and Case 2 in other times. Waters with similar chlorophyll concentrations may have large variations in ocean color in open ocean environment (e.g. Eastern Mediterranean versus South Pacific gyre).

4.4.4 Ternary Diagrams

Ternary diagrams are a way to show the relative contributions of 3 variables to a total value using a 2-D plot. In biological oceanography they are often used to show the relative contributions of absorption by phytoplankton, colored dissolved organic matter (CDOM), and non-algal particles (NAP) to the total absorption at a given wavelength (after subtracting out the water contribution). Figure 4.25 shows the layout of a ternary diagram to be used for this purpose. In that figure, the red triangle at the lower left represents all of the absorption being due to CDOM. The blue triangle at the lower right is the point for all absorption being due to NAP, and the green triangle at the top represents all absorption being due to phytoplankton. The red dotted lines are lines of constant CDOM contribution. That is, moving back and forth along one of the red lines leaves the CDOM contribution constant while the relative contributions of phytoplankton and NAP vary. Similarly, the green lines are lines of constant phytoplankton contribution, and the blue lines are “contours” of constant NAP contribution. To determine the three contributions for a plotted point, follow lines parallel to the dotted lines of constant contributions back to the respective axes. The black triangle and the colored arrows show a point with relative concentrations of phytoplankton as 0.65, CDOM as 0.2, and NAP as 0.15 of the total value. (Of course, which side of the triangle corresponds to which component does not matter, nor does the “counterclockwise” versus “clockwise” direction of increasing values on each of the three axes.)

Figure 4.26 shows an example of ternary plots used to show the relative importance of CDOM, NAP, and phytoplankton to absorption at two wavelengths. At 380 nm, most points are in the “CDOM corner” at the lower left because of the high absorption by CDOM at near-UV wavelengths. At 665 nm, phytoplankton dominate the absorption as a consequence of the exponential decrease of absorption by CDOM and NAP as the wavelength increases.

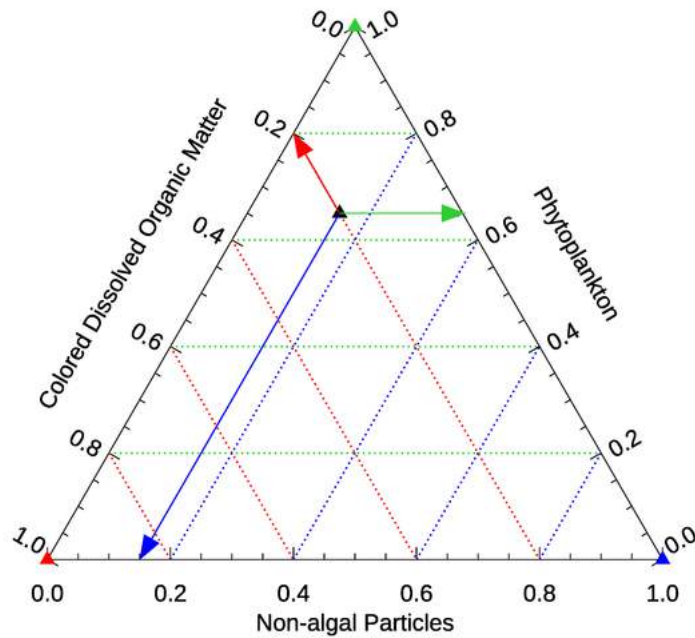


Figure 4.25: Example layout of a ternary diagram for displaying the relative contributions of phytoplankton, CDOM, and NAP to the total absorption.

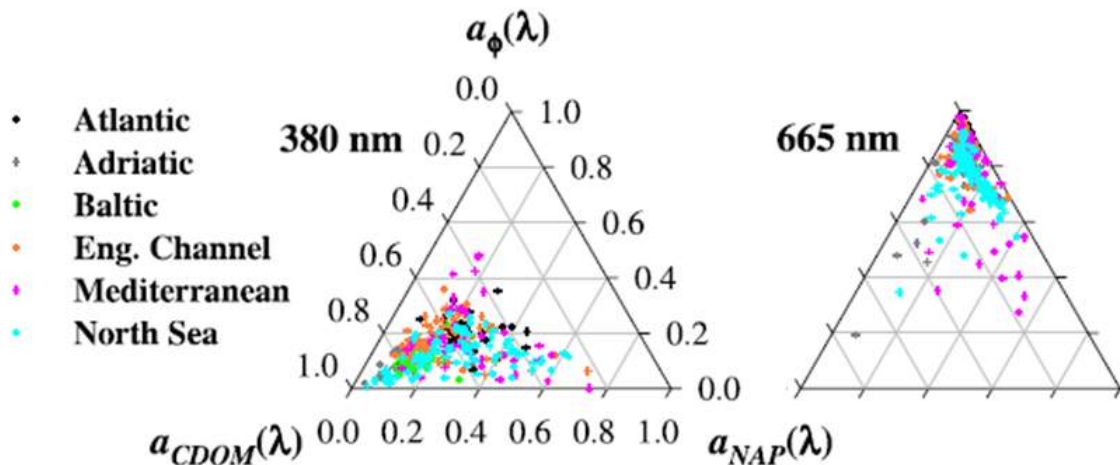


Figure 4.26: Ternary diagrams showing the relative concentrations of phytoplankton, CDOM, and NAP to the total absorption coefficient at 380 and 665 nm. Figures extracted from Fig. 16 of [Babin et al. \(2003b\)](#), which shows plots for 10 wavelengths.

4.4.5 Other Classification Schemes

The preceding discussion has outlined the most commonly used classification schemes for natural waters. However, a number of other classifications have been developed and used for various applications. Representative schemes have been based on

- Cluster Analysis of R_{rs} spectra: [Eleveld et al. \(2017\)](#); [Melin and Vantrepotte \(2015\)](#); [Prasad and Agarwal \(2016\)](#); [Wei et al. \(2016\)](#)
- Fuzzy Logic Classification of R_{rs} spectra: [Moore et al. \(2001, 2009, 2014\)](#)
- Principle Component Analysis/Empirical Orthogonal Functions: [Lubac and Loisel \(2007\)](#); [Avouris and Ortiz \(2019\)](#)
- Maximum wavelength of R_{rs} spectra: [Ye et al. \(2016\)](#)
- Weighted means of R_{rs} spectra: [Vandermeulen et al. \(2020\)](#)
- Trophic classification based on surface chlorophyll values: [Uitz et al. \(2006\)](#)

Absorption

When light interacts with matter one of three things can occur:

Absorption: The light can disappear, with its radiant energy being converted into other forms such as the energy of a chemical bond or heat.

Elastic Scattering: The light can change direction without a change of wavelength.

Inelastic Scattering: The light can undergo a change of wavelength, usually to a longer wavelength, and usually also with a change in direction.

This chapter begins the discussion of the first of these processes. Elastic and inelastic scattering are described in the next two chapters. Absorption by particular substances is discussed in Chapter 8.

5.1 Why is Absorption Important?

If you are interested in phytoplankton physiology, absorbed light is what matters. Light absorbed by phytoplankton provides the energy that drives photosynthesis; light elastically scattered by phytoplankton has no effect on photosynthesis and is of less interest to a plankton biologist. Therefore many more papers have been published on absorption by phytoplankton than on scattering by phytoplankton.

Phytoplankton absorption spectra (Section 8.3.2) are determined by the pigments in phytoplankton, hence measurement of phytoplankton absorption spectra gives information about what pigments are present and in what amounts. Thus absorption is fundamental to understanding phytoplankton physiology. Similarly, measurement of total absorption can be used to extract the amounts of phytoplankton, colored dissolved organic matter (CDOM), non-algal particles, or pollutants (if present) in the water column.

Some optical quantities are determined almost entirely by absorption. For example, the diffuse attenuation coefficient for downwelling plane irradiance, $K_d(\lambda)$, is a very “absorption-like” apparent optical property. The depth dependence of the downwelling plane irradiance, $E_d(z, \lambda)$, can be written as

$$E_d(z, \lambda) = E_d(0, \lambda) \exp[-\langle K_d(\lambda) \rangle z],$$

where $\langle K_d(\lambda) \rangle$ is the average of $K_d(z, \lambda)$ over depths 0 to z . To first order,

$$K_d(z, \lambda) \approx \frac{a(z, \lambda) + b_b(z, \lambda)}{\bar{\mu}_d(z, \lambda)} \approx \frac{a(z, \lambda)}{\cos \theta_{sw}}.$$

Here b_b is the backscatter coefficient and $\bar{\mu}_d$ is the average cosine of the downwelling radiance distribution. In most waters, $b_b \ll a$; and near the sea surface the average cosine can be approximated by the cosine of the Sun's direct beam angle, θ_{sw} , as transmitted through a level sea surface. Thus a measured profile of the absorption coefficient $a(z, \lambda)$ allows for an approximate computation of $E_d(z, \lambda)$, which in turn is directly related to the heating rate of the upper ocean by Gershun's Law (Section 10.7).

Although ocean-color remote sensing measures scattered light leaving the oceans, the goal of that remote sensing is often to deduce how absorption in the water column changed sunlight into the scattered light that was measured. Knowing the effect of absorption gives a wealth of information about the ecological state of the water column.

5.2 Absorption by Oceanic Components

[Collin Roesler wrote this section.]

Absorption by most oceanic constituents depends strongly on wavelength. The overall magnitude and spectral features of oceanic absorption depend upon the concentration and composition of the particulate and dissolved constituents and water itself. The inherent optical properties (IOPs) are conservative properties, and the magnitude of the absorption coefficient varies linearly with the concentration of the absorbing material. Theoretically, the absorption coefficient can be expressed as the sum of the absorption coefficients of each component:

$$a(\lambda) = \sum_{i=1}^N a_i(\lambda), \quad (5.1)$$

where N is the total number of components and a_i indicates the i^{th} component, and all absorption is spectral, indicated by (λ) .

Practically, however, it is not possible to measure the absorption properties of each individual absorbing component and thus the individual components can be grouped into similarly absorbing constituents based upon similarity in their optical properties and/or analytically-based groupings:

$$a(\lambda) = a_w(\lambda) + a_{phy}(\lambda) + a_{NAP}(\lambda) + a_{CDOM}(\lambda) \quad (5.2)$$

where subscripts w, phy, NAP and CDOM indicate water, phytoplankton, non-algal particles, and colored dissolved organic matter, respectively.

This section gives only a brief introduction to the absorption properties of these ocean constituents. The absorption and scattering properties of individual components are discussed in detail in Chapter 8.

Absorption by water (Fig. 5.1) is weak in the blue and strong in the red and varies somewhat with temperature and salinity. See Section 8.2 on water for further discussion.

Absorption by particles is separated into phytoplankton and non-algal particles (NAP) using spectrophotometry and an extractive technique. Phytoplankton absorption (Fig. 5.2)

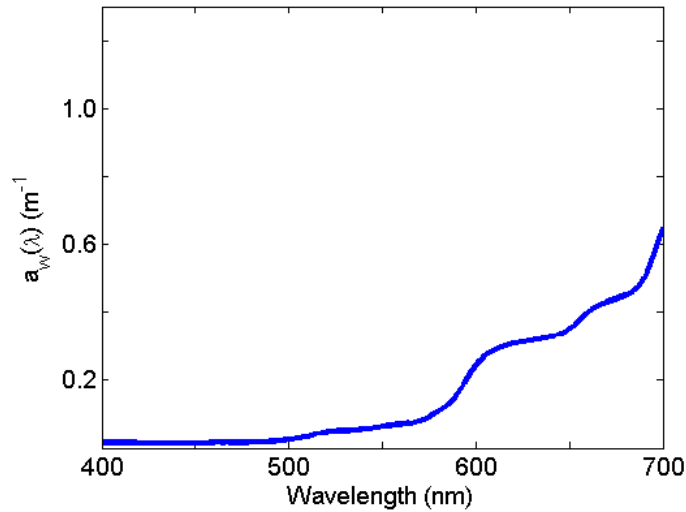


Figure 5.1: Absorption spectrum for pure water (data from Pope and Fry (1997)).

demonstrates the most spectral variations of any of the components due to the individual pigment absorption spectra. In general $a_{\text{phy}}(\lambda)$ exhibits peaks in the blue and red regions of the spectrum due to the ubiquitous presence of chlorophyll *a*. See Section 8.3 for further discussion.

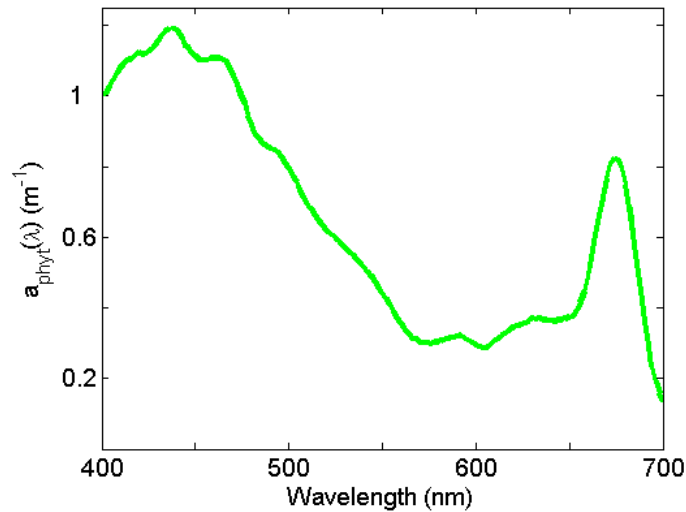


Figure 5.2: Generic phytoplankton absorption spectrum for mixed algal composition (modified from Roesler et al. (1989)).

Non-algal particle absorption (Fig. 5.3) is strongest in the blue, decreasing approximately exponentially to the red. This component, operationally-defined, includes living zooplankton and bacteria, as well as the non-pigmented parts of phytoplankton (cell walls, membranes etc), and detrital material as well as inorganic particles. See Section 8.5 on non-algal particles for further discussion.

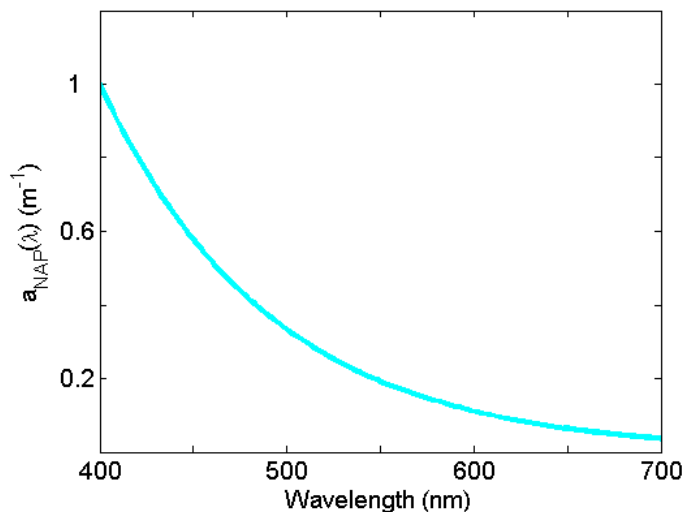


Figure 5.3: Generic non-algal particle absorption spectrum for mixed composition.

Colored dissolved organic matter (CDOM) absorption (Fig. 5.4) is very similar to that of NAP due in part to the similarity in composition (organic material), but generally exhibits a steeper exponential slope. This material is operationally separated from NAP by filtration; CDOM is measured on the filtrate passing through 0.2mm or 0.7 mm nominal pore sized filters. See Section 8.4 on CDOM for further discussion.

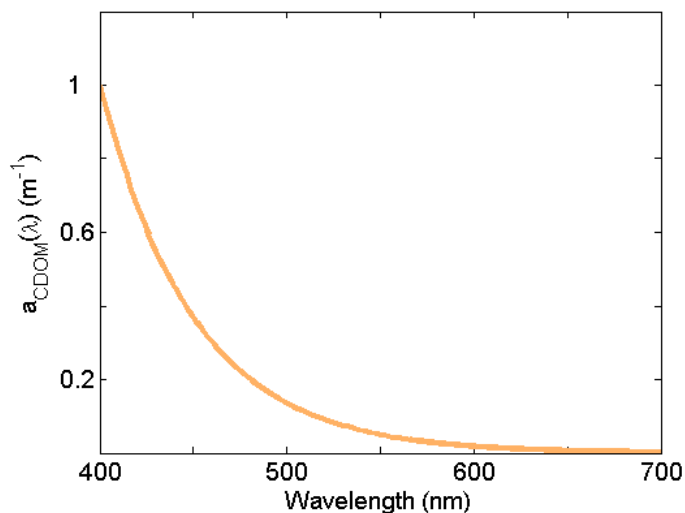


Figure 5.4: Generic colored dissolved organic matter (CDOM) absorption spectrum for mixed composition.

For oligotrophic environments with very low concentrations of suspended and dissolved material, the absorption coefficient is dominated by water (Fig. 5.5A) and the wavelength of minimum absorption is in the blue, hence the blue color of the seawater. For eutrophic and/or coastal environments with high concentrations of suspended and dissolved material

(Fig. 5.5B), the absorption coefficient is dominated by that material and the wavelength of minimum absorption shifts to the green, lending green color to that environment.

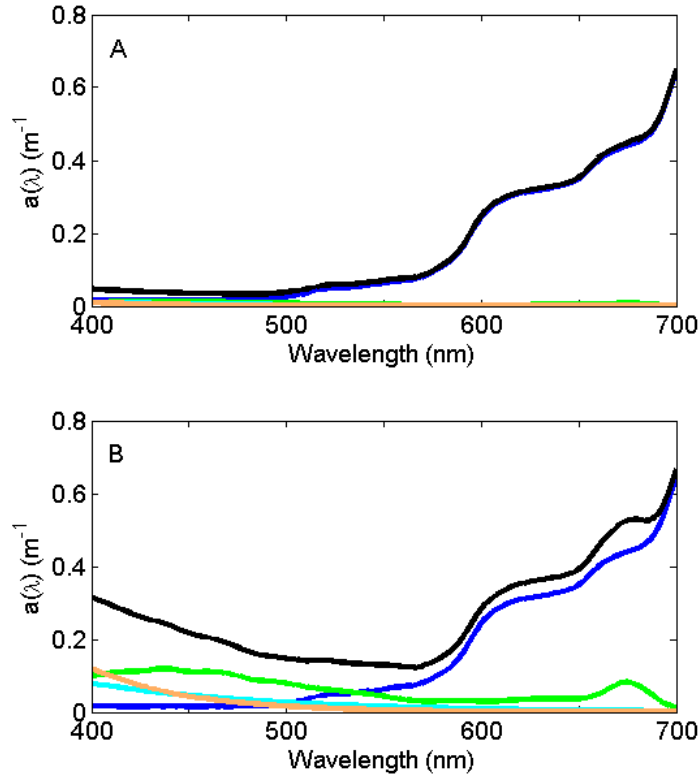


Figure 5.5: Component and total absorption spectra for clear open ocean waters where water dominates the absorption (panel A) and for eutrophic coastal waters where particulate and dissolved organic matter dominate the blue and green portions of the spectrum (panel B).

5.3 The Physics of Absorption

[Curtis Mobley and Collin Roesler contributed to this section]

This section introduces the physical processes that lead to absorption of light by an atom or molecule. Beginning with a few basic concepts from quantum mechanics, we qualitatively show why a molecule like chlorophyll has an absorption spectrum that is unique to that molecule.

5.3.1 Quantum Mechanics Terminology

The terminology of quantum mechanics can be confusing. Physicists tend to speak of quantum numbers (principle, azimuthal, magnetic, and spin); chemists talk about orbitals and shells and subshells. Both groups are describing the same thing in ways that best fit

their respective needs. There are also various graphical ways to represent the arrangement and energies of electrons in an atom or molecule.

An electron *orbital* refers to the size and shape of the three-dimensional spatial region where an electron is likely to be found, say with $\geq 90\%$ probability, at any moment. The term emphatically does *not* imply that electrons buzz around nuclei in well defined orbits like planets going around the Sun. Such an image from classical physics is completely wrong as a way to visualize atoms. If an analogy is needed, it is perhaps permissible to think of “orbital” as loosely corresponding to “energy level.” However, in many cases, different orbitals correspond to the same energy, in which case the energy levels are said to be “degenerate.”

The *principal quantum number* $n = 1, 2, 3, \dots$ describes the overall physical size of the orbital; i.e., the average distance of an electron from the nucleus. The *azimuthal* or *angular quantum number* $l = 0, 1, \dots, n-1$ describes the shape of the orbital. The *magnetic quantum number* $m = -l, -l+1, \dots, 0, 1, \dots, l$ describes orbital’s orientation in space (originally with respect to an externally applied magnetic field, hence the name). The *spin quantum number* s can have values of $+1/2$ or $-1/2$ and defines the direction (usually called “up” or “down”) of the component of an electron’s intrinsic angular momentum (spin) along the direction of an external field. Nature does not allow two electrons in the same atom to have the same set of quantum numbers n, l, m, s , which is known as the Pauli exclusion principle.

Orbitals with the same value for the principal quantum number form a *shell*. Shells are often labeled with letters starting at K (for historical reasons rooted in spectroscopy). Electrons with a principal quantum number of $n = 1$ are in the K shell; electrons with $n = 2$ are in the L shell; those with $n = 3$ are M-shell electrons, and so on. *Subshells* are orbitals in the same shell that have the same magnetic quantum number. Subshells are labeled s for $l = 0$, p for $l = 1$, d for $l = 2$, and so on for f, g, h, ... in that order (again, these letters arose in spectroscopy and stood for “sharp,” “principal,” and “diffuse” spectral lines). Thus an electron with quantum numbers $n = 1, l = 0$ is said to be in a 1s orbital (spoken as “one ess”); an electron in a 2p (two pee) orbital has $n = 2, l = 1$. All s orbitals are spherically symmetrical. The three p orbitals consist of 2 lobes. Four of the five d orbitals are four-lobed somewhat like a clover leaf, and one is lobed with a torus around its “equator.” There is no spatial orientation for spherically symmetric s orbitals, but p orbitals can have their lobes oriented along the x , y , or z axes of some coordinate system, and so on for the other non-spherical orbitals. Figure 5.6 shows a qualitative visualization of orbital shapes for s, p, d, and f orbitals.

Because of the attraction between a negative electron and the positive nucleus, the physically smallest, spherically symmetric orbital, which on average over time has the electron closest to the nucleus, has its electron most tightly bound to the nucleus. This is the 1s orbital. The corresponding *ground state energy* is represented as a negative value (usually in units of electron volts, eV; $1 \text{ eV} \approx 1.60 \times 10^{-19} \text{ J}$). To remove a ground-state electron from the molecule requires this amount of energy to “raise” the electron to 0 binding energy, i.e., no binding to the nucleus. Moving an electron from the ground state to an orbital with $n > 1$ moves the electron further from the nucleus and therefore requires energy to overcome the electrical attraction. Similarly, an electron in a 2p orbital will have more energy than one in a 2s orbital, because a 2p electron spends more time further from the nucleus than does a 2s electron. Finally, a particular orbital can hold at most two electrons, with spin quantum numbers of $s = \pm 1/2$ (the Pauli exclusion principle). Note

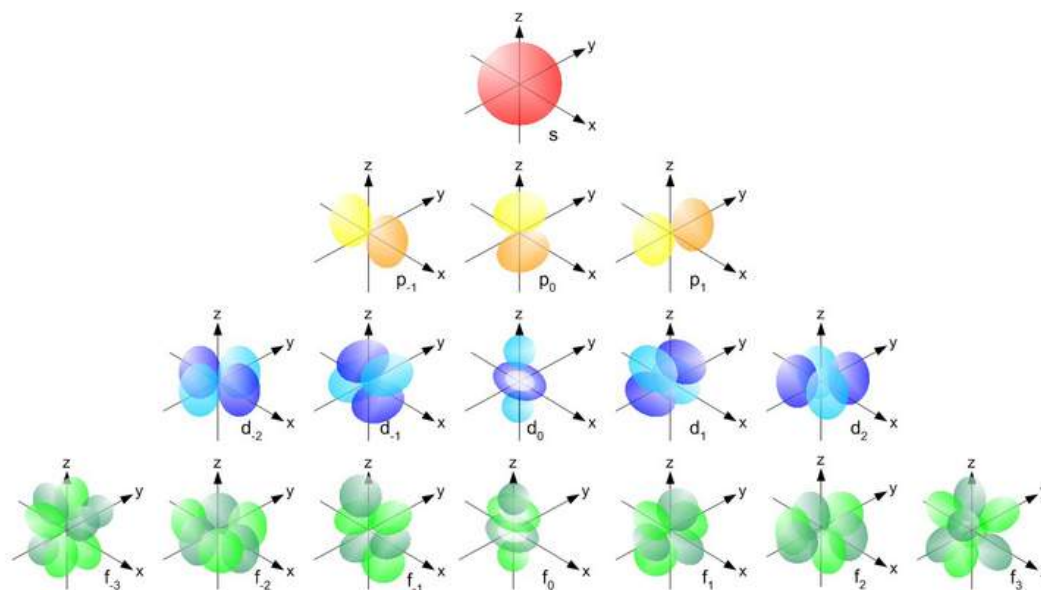


Figure 5.6: Electron orbital shapes. The surfaces drawn can be thought of as surfaces of constant probability of finding an electron at that location. From [WikiMedia](#)

that the s orbital label is not to be confused with the s spin quantum number.

The allowed electron energy levels in an atom or molecule are often displayed on a diagram with binding energy on the y axis and the quantum numbers on the x axis. Such diagrams are sometimes called Jablonski diagrams. Figure 5.7 shows the conceptual layout of an energy level diagram with electron quantum numbers and orbitals labeled. The exact values of the energy levels for a given atom or molecule can be computed from the laws of quantum mechanics. Figure 5.8 shows another common way to display electron energy levels.

The previous discussion refers to the quantum numbers for individual electrons in an atom or molecule. In multi-electron atoms it is also common to write quantum numbers for the total quantum state of all electrons in the atom or molecule. That is, each electron in the molecule has an azimuthal quantum number, which represents its angular momentum with respect to the nucleus. Angular momentum is a vector, so the total angular momentum of all electrons is the vector sum of the angular momentum vectors of the individual electrons, computed using the quantum mechanical rules for adding quantized angular momenta. The resulting total angular momentum quantum number is denoted by a capital letter L with values of $L = 0, 1, 2, \dots$. Likewise, the vector sum of the spin angular momenta of the individual electrons gives the spin quantum number for the whole atom or molecule, which is denoted by S . For atoms with an even number of electrons, the resulting possible values for S are $0, 1, 2, \dots$. For atoms with an odd number of electrons, the resulting S values can be $1/2, 3/2, 5/2, \dots$. In the same way, the total orbital and total spin angular momentum vectors can be combined to obtain a total angular momentum, described by a new quantum number J , whose values are either integral (for an even number of electrons) or half-integral (an odd number of electrons) from $|L - S|$ to $L + S$. The z component of the total angular momentum is specified by quantum number M , which can have either integral (if J is

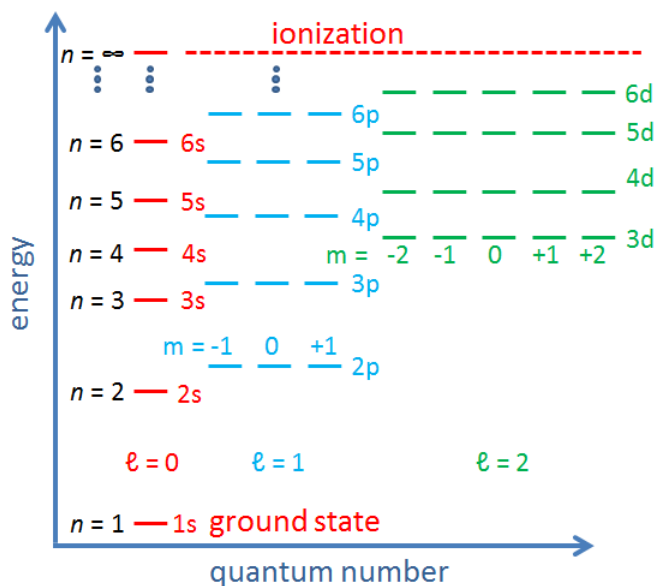


Figure 5.7: Conceptual layout of a Jablonski energy level diagram. The short horizontal lines show the allowed energy levels. The energy levels are not to scale for any actual atom or molecule. The $n = 4$ shell also has seven f orbitals, which are not shown, and so on for the higher shells.

integral) or half-integral (if J is half-integral) values from $-J$ to $+J$. This gives four total or “spectroscopic” quantum numbers L, S, J, M .

When the L quantum number is 0, 1, 2, 3, ..., the atomic or molecular quantum state is labeled with letters S, P, D, F, ..., in analogy with s, p, d, f, ... for electron states with $l = 0, 1, 2, 3, \dots$. The value of $2S + 1$ is written as a preceding superscript, and the J value as a following subscript: $^{2S+1}L_J$, e.g. 1S_0 , $^2P_{1/2}$, 3P_0 , and so on (read as “singlet ess zero”, “doublet pee one half”, “triplet pee zero”, and so on). Again note the difference in the S state labeling the value of the quantum number L versus the total spin quantum number S . The value of quantum numbers L and M are not to be confused with the L and M electron shells corresponding to electron principal quantum numbers $n = 2$ and $n = 3$. Note also that a $3p$ electron state is different from a 3P atomic or molecular state. As was said, the notation can be confusing.

Fortunately this is already more than we need to know for the following qualitative discussion of absorption spectra. However, this material will be needed later, for example in understanding fluorescence (Section 7.3). In any case, the basics of quantum-mechanical terminology are outlined here in hopes of minimizing confusion when the lower-case and upper-case notations are seen elsewhere. The excellent [ChemWiki](#) and the atomic spectroscopy chapter of J. B. Tatum’s online [Stellar Atmospheres](#) give more detailed but still qualitative presentations of these matters, which are treated in their full quantum-mechanical glory in texts on quantum physics and chemistry.

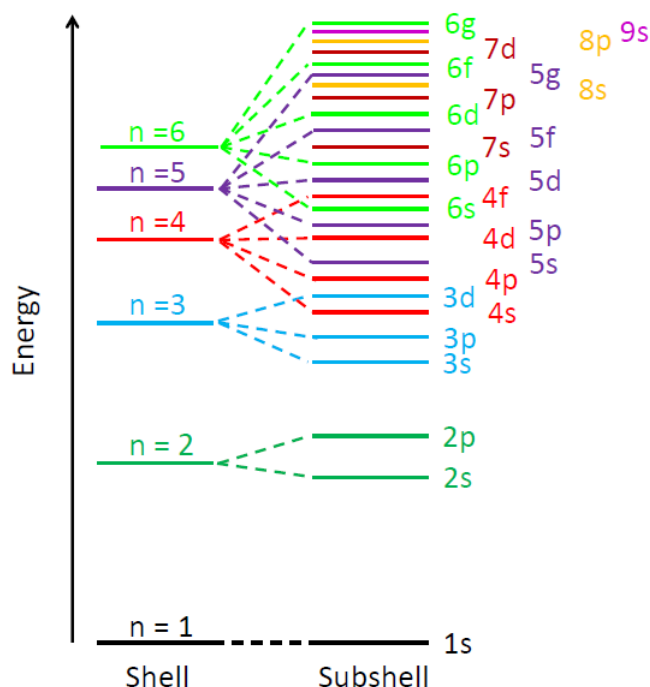


Figure 5.8: Another way to show electron energy levels. This display makes it easy to see the order of the subshell energy levels, which overlap between the higher shells. This figure is complete through all subshells of the $n = 5$ shell, and the order of the subshells is complete through the 6g subshell. The spacing of the energy levels is not to scale.

5.3.2 Quantum Mechanics of Absorption

For the present, it is sufficient to know that the electrons in an atom or molecule have very specific energies—that is, the allowed electron energies are *quantized*. This physical reality is reflected mathematically in the discrete (integer or half-integer) values of the quantum numbers introduced above.

Recall that light is a propagating electromagnetic field. For the moment, we can think (rather naively) of a photon as a region of space with a propagating and rapidly fluctuating electric field. If the electric field is oscillating with a temporal frequency f (units of Hertz, i.e. cycles per second or s^{-1}), then the photon has energy

$$E = hf = \frac{hc}{\lambda},$$

where h is Planck's constant, c is the speed of light, and λ is the wavelength (in a vacuum).

As a photon approaches an atom or molecule, the electrons in the atom or molecule begin to “feel” the photon's electric field. Let E_1 be the energy of an electron in one of the subshells, and let E_2 be the higher energy of one of the subshells that does not contain an electron. If the photon frequency corresponds to the energy difference between these two energy levels, i.e., if

$$f = \frac{E_2 - E_1}{h},$$

then there is a chance that the electron will absorb the photon and use the photon's energy to “jump” from its current subshell with energy E_1 to the vacant higher-energy subshell with energy $E_2 = E_1 + hf$. The atom or molecule is then said to be in an excited energy state. The time scale for this absorption-excitation process is about 10^{-15} s. This is the fundamental way that matter absorbs light. If the photon frequency does not correspond to any of the subshell energy differences, then the light cannot be absorbed and the photon continues on its way.

Figure 5.9 illustrates three electron energy levels of a molecule. Which shells or subshells these levels correspond to in a particular molecule is irrelevant for the present discussion. The usual situation for a molecule is that the lower-energy subshells are all occupied by electrons. It is then likely that visible light will raise one of the outermost (highest energy) electrons to an unoccupied energy level. (Raising an electron from the 1s shell to a higher shell typically requires ultraviolet light.) The red and blue arrows in the left panel of the figure represent electrons absorbing blue- and red-wavelength photons and jumping from a low energy level to higher ones. The right panel of the plot shows the corresponding absorption lines in the absorption spectrum of the molecule.

The probabilities for photon absorption are generally different for different transitions. Thus some absorption lines will be “stronger” than others. This is represented by a higher magnitude of the blue absorption line in the right-hand plot. In emission—the reverse of absorption—a photon is emitted when an electron falls from an excited state to an unoccupied lower energy level. In that case, some emission lines will be “brighter” than others. The visual nature of these emission lines was the origin of the spectroscopic labels of “sharp,” “principle,” “diffuse,” and “fine,” which were later related to the quantum numbers and orbitals as described above.

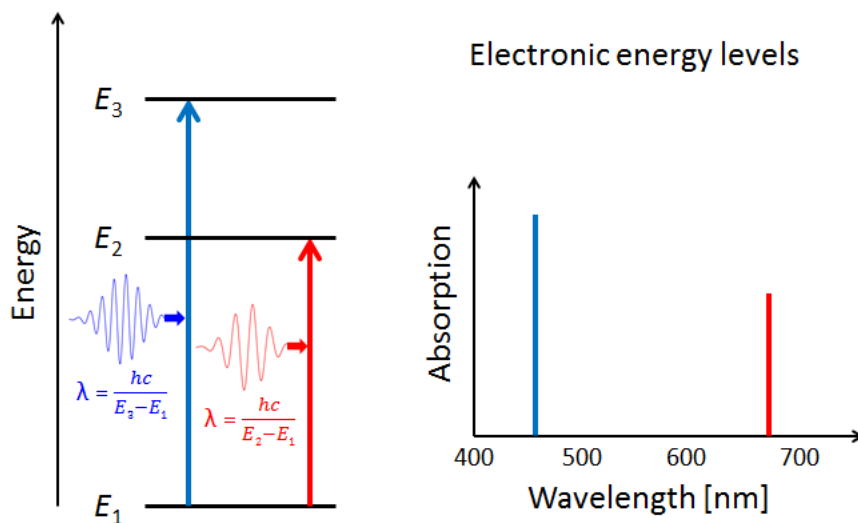


Figure 5.9: Illustration of three electron energy levels in a molecule and absorption of blue and red light.

The situation of Fig. 5.9 is typical of atoms and simple molecules in near isolation (e.g., in gases). However, the situation becomes more complicated for molecules containing many

atoms. In addition to the electron energy levels, molecules also have vibrational modes, as illustrated in Fig. 5.10. This figure is a frame grab from animations, which can be seen on the [Physics of Absorption](#) page of the Web Book. The blue spheres represent atoms attached to the side of a larger molecule. These vibrational modes also have quantized energy levels. That is, the molecules can vibrate only at specific frequencies, which are determined by the molecule's structure and the atoms involved.

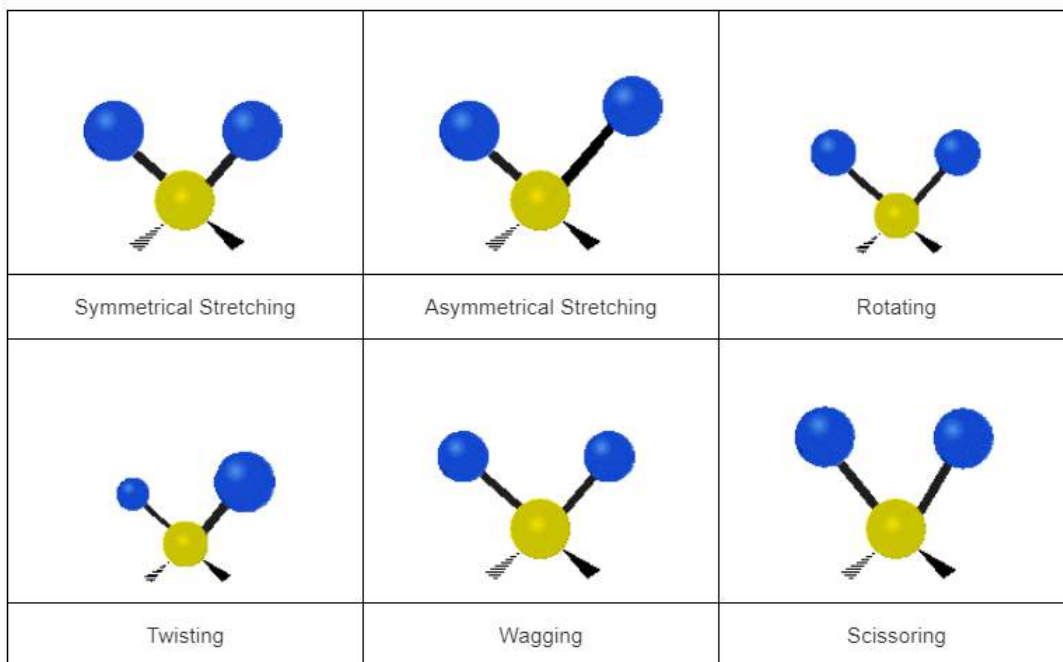


Figure 5.10: Molecular vibration modes. These figures are frame grabs from Wikimedia animations, e.g. [symmetrical stretching](#). See the Web Book page [Physics of Absorption](#) for animations.

It usually requires less energy to excite a vibrational mode than to excite an electron from one subshell to another. The spacing between the quantized vibrational energy levels is thus less than between electron orbitals. A quick calculation is instructive. Molecular vibrational frequencies fall in the 10^{12} to 10^{14} s^{-1} range. For $f = 10^{13} \text{ s}^{-1}$, $\lambda = c/f$ gives a wavelength of $30 \mu\text{m}$, which is in the mid-infrared. Thus infrared light can excite a molecule from one vibrational mode to another (with no change in the electronic energy level). The corresponding vibrational energy difference is $\Delta E_v = hf = hc/\lambda = 0.04 \text{ eV}$. (A useful relation for these sorts of calculations is $hc \approx 1.24 \text{ eV } \mu\text{m}$.)

Now suppose that we are interested in an electronic transition corresponding to a blue wavelength of $440 \text{ nm} = 0.440 \mu\text{m}$. The corresponding electronic energy shift is $\Delta E_e = hf = 2.82 \text{ eV}$. If the vibrational modes now allow energy levels that are equal to the electronic transition \pm a vibrational level, the energy shift can be $\Delta E_e \pm \Delta E_v$. The corresponding wavelengths are then $\lambda = hc/(\Delta E_e \pm \Delta E_v)$, which in the current example gives $\lambda = 0.434$ and $0.446 \mu\text{m}$, or a wavelength shift of 6 nm to either side of the 440 nm electronic transition line.

The thin horizontal lines in the left panel of Fig. 5.11 represent the additional molecular energy levels when both electronic and vibrational levels are included in the energy

diagram. The thin blue and red arrows in the left panel illustrate how the presence of vibrational modes allows for more possible electronic transitions to occur in the wavelength neighborhoods of the electronic transitions between subshells. These transitions add more spectral lines to the absorption spectrum, as shown by the thin lines in the right-hand panel.

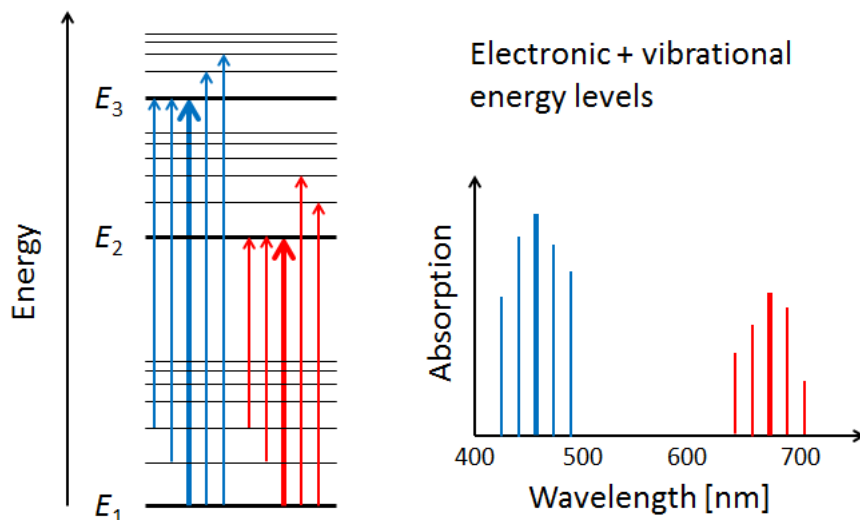


Figure 5.11: Illustration of electronic and vibrational energy levels in a molecule. The thin horizontal lines in the left panel represent the vibrational energy levels within each electronic subshell. The thin arrows represent excitations between energy levels that include both electronic and vibrational excitations.

In addition to the vibrational modes, molecules also have rotational modes. That is, the entire molecule can rotate about an axis with some frequency, which is again quantized. Exciting a molecule from one rotational mode to another typically requires very little energy (in the 10^{-3} to 10^{-6} eV range), so that microwave radiation¹ (λ in the millimeter to meter range) is adequate. When the additional quantized energy states associated with rotational modes are included in the energy diagram, the allowed energy levels become very closely spaced, as illustrated by the dashed lines in the left panel of Fig. 5.12. The corresponding changes in absorption line wavelengths are in the sub-nanometer range when compared to electronic or electronic-vibrational transitions. The net result of having electronic, vibrational, and rotational modes is that the resulting molecular absorption spectrum is such a dense collection of absorption lines that it appears as a continuous function of wavelength when measured by instruments with spectral responses greater than a nanometer.

We have now seen in a qualitative way how absorption spectra like those of chlorophyll arise. The final comment to make is that the absorption spectrum of chlorophyll *a*, for example, will be slightly different when measured *in vivo* (in a living cell) versus *in vitro* (literally “in glass,” i.e. after extraction from a cell). The reason is that the environment of

¹Microwave ovens operate at a frequency of 2.45GHz, which is a wavelength of 12.23 cm. Most foods are 65-90% water, and it is absorption of this wavelength by water molecules that heats the food. See Vollmer (2004) for the details.

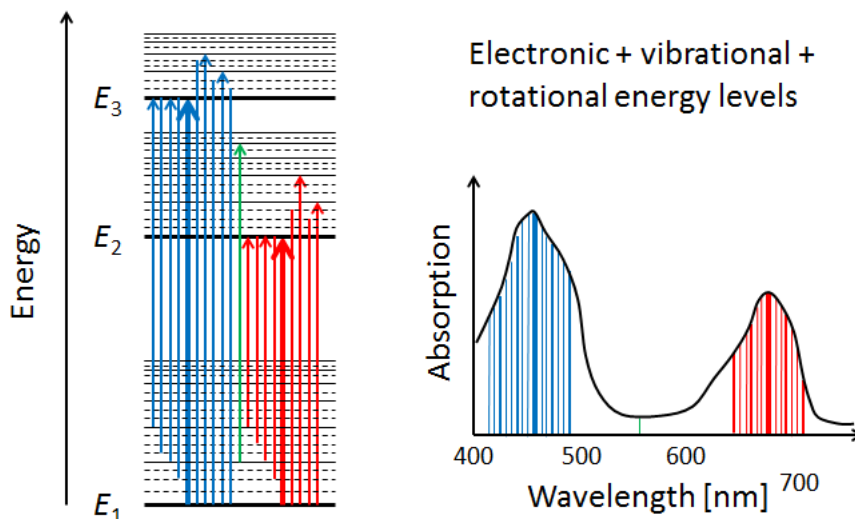


Figure 5.12: Illustration of electronic, vibrational, and rotational energy levels in a molecule. The dashed lines in the left panel represent the closely spaced rotational levels. The thin vertical lines represent excitations between energy levels that include both electronic, vibrational, and rotational excitations. The resulting absorption spectrum appears continuous at the resolution of most instruments.

the chlorophyll molecule effects how it can vibrate and rotate. Thus different environments change the allowed energy bands somewhat, and thus change the shape of the absorption curve.

5.4 Measurement of Absorption: From Theory to Reality

[This section was written by Collin Roesler.]

Consider a scenario where the goal is to measure the absorption spectrum of a thin layer of material (Figure 5.13A). The incident radiant power is given by Φ_o , in the form of a collimated beam. The radiant power transmitted through the layer, Φ_t , is detected. If $\Phi_t = \Phi_o$, there is no loss of radiant power and therefore no attenuation. If however the medium absorbs some quantity of radiant power, Φ_a , then $\Phi_t < \Phi_o$, and $\Phi_o = \Phi_t + \Phi_a$ (Figure 5.13B). In the case of material that both absorbs and scatters (Figure 5.13C), the scattered radiant power is given by Φ_b , and $\Phi_o = \Phi_t + \Phi_a + \Phi_b$.

To quantify the absorbed radiant power only, it is necessary to measure both the transmitted and scattered radiant power. This is a requirement for an absorption meter. Consider first a nonscattering material. Recalling the definitions of Section 3.0.1, the measured dimensionless transmittance, T , is the fraction of incident power transmitted through the layer:

$$T = \frac{\Phi_t}{\Phi_o}.$$

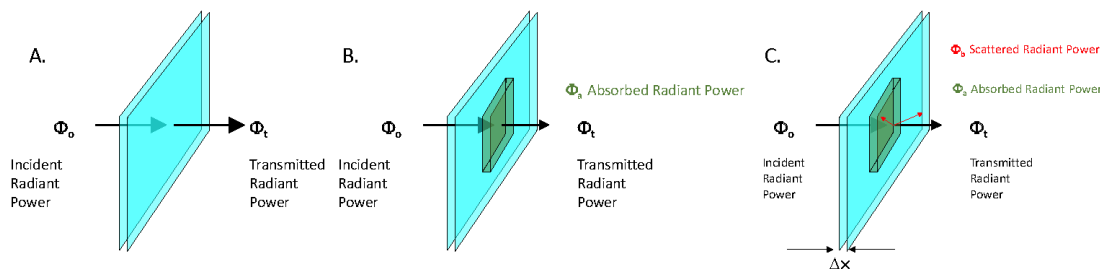


Figure 5.13: Illustration of theoretical attenuation by thin layers of non-attenuating (panel A), absorbing (panel B), and absorbing and scattering (panel C) material. The thickness of the layer is given by Δx .

The absorptance, A , is the fraction of incident radiant power that is absorbed ($1 - T$):

$$A = \frac{\Phi_a}{\Phi_o} = \frac{\Phi_o - \Phi_t}{\Phi_o}.$$

The absorption coefficient a (with units of m^{-1}) is the absorptance per unit distance

$$a = \frac{A}{\Delta x},$$

which for an infinitesimally thin layer can be expressed as

$$a = \frac{\frac{\Delta\Phi}{\Phi}}{\Delta x} = \frac{\Delta\Phi}{\Phi\Delta x}.$$

Rearranging this expression and taking the limit as $\Delta x \rightarrow 0$ yields

$$a\Delta x = \lim_{\Delta x \rightarrow 0} \left(\frac{\Delta\Phi}{\Phi} \right).$$

Assuming that the absorption coefficient is constant over the layer of thickness x and integrating gives

$$\begin{aligned} \int_0^x a \, dx' &= - \int_{\Phi_o}^{\Phi_t} \frac{d\Phi}{\Phi} \\ ax' \Big|_0^x &= - \ln \Phi \Big|_{\Phi_o}^{\Phi_t} \\ ax &= - [\ln(\Phi_t) - \ln(\Phi_o)] = - \ln \left(\frac{\Phi_t}{\Phi_o} \right) \\ a &= - \frac{1}{x} \ln \left(\frac{\Phi_t}{\Phi_o} \right). \end{aligned}$$

This equation provides a guide toward designing instruments to accurately measure absorption. The next sections give the specifics on techniques to measure absorption by dissolved and particulate constituents in seawater.

5.5 Benchtop Spectrophotometry

[This section was written by Collin Roesler.]

The benchtop spectrophotometer was invented by Arnold Beckman and his colleagues at the National Technologies Laboratories in 1940. Beckman, a professor at California Institute of Technology, started NTL in the 1930s, dedicated to designing, producing and selling sophisticated scientific instruments. NTL would later become Beckman Instruments. The development of the earliest model spectrophotometer, the DU, was led by project leader Howard Cary. The invention of the spectrophotometer has been described as one of the most significant in the biosciences. Cary, with two colleagues, later started the Applied Physics Corporation, renamed Cary Instruments (subsequently purchased by Varian Associates). They are also a major supplier of benchtop spectrophotometers. This section shows how these spectrophotometers can be used to measure absorption by dissolved and particulate material.

5.5.1 Benchtop Spectrophotometry of Solutions in Transmission Mode (T-mode)

Designs of spectrophotometers vary between manufacturers and the differences can impact the accuracy of the derived absorption due to variability in the measurements of Φ_o and Φ_t . Because these instruments were developed for non-scattering solutions, users make modifications to optimize scattering collection and to correct for scattering not detected. Spectrophotometers come in single beam- and dual-beam models (Figure 5.14). The incident radiant power, generated by a lamp, passes through a monochromator (not shown) so that light within a narrow spectral band enters the sampling chamber. In the single-beam mode, the incident and transmitted radiant power are determined from two separate scans, the first with a reference material in the sample chamber (e.g., pure water) and the second with the sample in the sample chamber (e.g., filtered seawater). In the dual-beam mode, the incident and transmitted radiant power are measured nearly simultaneously using a beam splitter between the diffraction monochromator and the sample chamber. Similarly in the dual beam model, there is a mirror system to direct the incident and transmitted beams to the detector in an alternating fashion as the instrument scans.

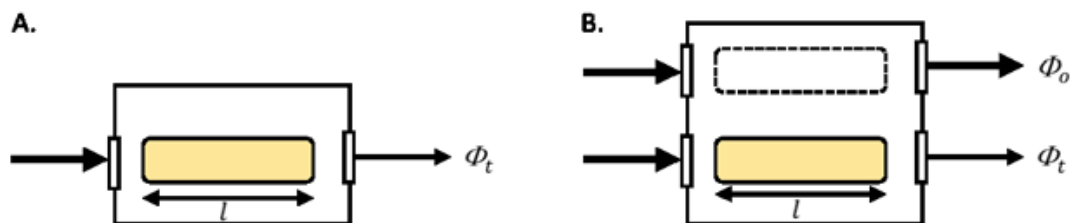


Figure 5.14: Schematic diagram of single- (panel A) and dual-beam (panel B) spectrophotometers showing the configuration of incident and transmitted radiant power. The shaded rectangle indicates a cuvette of some geometric pathlength ℓ filled with sample. The dashed rectangle represents a cuvette with reference material.

Spectrophotometers output their signal as absorbance Abs , from which the absorption coefficient can be determined by Eq. (3.23):

$$a = \frac{2.303Abs}{\ell}, \quad (5.3)$$

where ℓ is the geometric pathlength of the cuvette in units of meters. In order to maintain the adherence to Beer's Law so that the absorbance is linearly related to both the concentration of the absorbing material and to the geometric pathlength (i.e., there is no self shading), it is recommended that the absorbance from the spectrophotometer be maintained within the range 0.1 to 0.4. This is equivalent to transmittances of approximately 80% and 40%, respectively.

The best practice for spectrophotometry of solutions is to employ a dual beam model to minimize uncertainty in the lamp intensity (Φ_o , Fig. 5.14B). When running multiple samples, the material in the reference cuvette can undergo changes such as increases in temperature due to the lamp relative to the sample. Rather than changing the reference material with every sample, which comes with its own set of handling uncertainties, it is recommended to run the dual beam mode without a cuvette in the reference path (i.e., remove the dashed line cuvette in Fig. 5.14B) and follow either:

1. Automatic baseline correction, which consists of the following scan order:
 - 1.1. Baseline scan with nothing in the reference beam and reference material in the sample cuvette (most spectrophotometers will subtract this automatically from the sample scans)
 - 1.2. Without opening the sample compartment, immediately scan again as a sample; this is essentially running the baseline against itself (a zero scan). The expectation is that the result is close to zero (\pm the uncertainty of the instrument, 0.0001) and spectrally flat.
 - 1.3. Make 3-5 blank scans (again with nothing in the reference beam and with new reference material in the sample cuvette).
 - 1.4. Alternate between sets of 5-6 sample scans and a blank scan.
2. Post processing baseline correction, which consists of the following scan order:
 - 2.1. Make 3-5 blank scans, each with new reference material in the sample cuvette.
 - 2.2. Alternate between sets of 5-6 sample scans and a blank scan.

The upside of method (1) is the ability to easily view the variability in the sample scans without having to visually remove the baseline in real time. Additionally, it allows for quantification of the instrument resolution with the zero scan. The downside is that a single scan is used as a baseline, and the average of the 3-5 initial blank scans provides the uncertainty in that baseline. The upside of method (2) is that the mean blank is used as the baseline correction but the downside is that it can be difficult to visually analyze the sample scans as they contain the blank signal in them. Regardless of method, the time series of blank scans is used to quantify the drift of the instrument. If significant, the time course of the blank signal must be removed from the samples as part of the complete baseline correction.

5.5.2 Benchtop Spectrophotometry of Suspended Particulate Matter Measured in Internally-mounted Cuvettes inside an Integrating Sphere (IS-mode)

Measuring absorption of suspended particulate matter in benchtop spectrophotometers is much more challenging because they are not optimized to collect the scattered radiant power. Thus the measured transmitted radiant power is reduced by the amount of scattering, which is incorrectly attributed to absorption. The absorption coefficients therefore will be overestimated. For this approach an integrating sphere is employed. Interestingly, the history of the integrating sphere precedes that of the spectrophotometer by about 50 years. Integrating spheres are hollow spherical cavities coated with highly scattering interior surfaces designed to scatter light in a diffuse pattern (equal in all directions) while minimally absorbing light. Thus the result is to remove the effect of the directionality of the incident light so that transmitted radiant power can be measured from any small portion of the sphere (Fig. 5.15) and extrapolated to all of the light using geometric relationships. Scattered radiant power has the same likelihood of detection as transmitted radiant power; the only loss is due to absorption.

The challenge of the integrating sphere is to measure only the loss of incident radiant power that goes through the sample (thick arrows in Fig. 5.15A, called the sample beam) and not of the radiant power that has reflected from the sphere and re-enters the sample (thin arrows). This is achieved by slightly offsetting the sample cuvette from the reference beam (large arrows in Fig. 5.15B). In this way the reference beam and the sample beam generate identical light fields within the sphere with the exception of the directly transmitted sample beam. Thus by difference, the contribution to the loss of the sample beam due to multiple scattering through the sample is removed (as it shows up in both the transmitted reference and transmitted sample beams).

Similar to the protocol for transmission mode measurements of non-scattering solutions, either a baseline scan is collected with reference material in the cuvette, a zero scan is collected, followed by 3-5 blank scans for automatic baseline correction, or the 3-5 blank scans are first collected and the samples are corrected in post processing. In either case, the alternating sequence of 5-6 sample scans and a blank scan are collected. Absorption coefficients are computed as before.

5.5.3 Benchtop Spectrophotometry of Particulate Matter Collected on Glass Fiber Filters (the Quantitative Filter Technique) Measured Internally-mounted in the Integrating Sphere (the QFT in IS-mode)

The size and shape of commercially available integrating spheres limits the geometric path-length of cuvettes that can be internally mounted. Thus to obtain a strong signal to noise ratio, the concentration of suspended matter in a 1-cm cuvette has to be much higher than is generally observed in aquatic systems. For this reason, discrete water samples are filtered onto glass fiber filters (Whatman GF/F) and the absorbance of the filters is measured. The protocol is identical to that outlined previously with the exception that the baseline/blank scans are performed with a blank filter through which a comparable volume of pure water has been filtered (to flatten the fibers in a manner comparable to that of the sample filter). The filter is placed in the sample holder (same location as the cuvette), perpendicular to

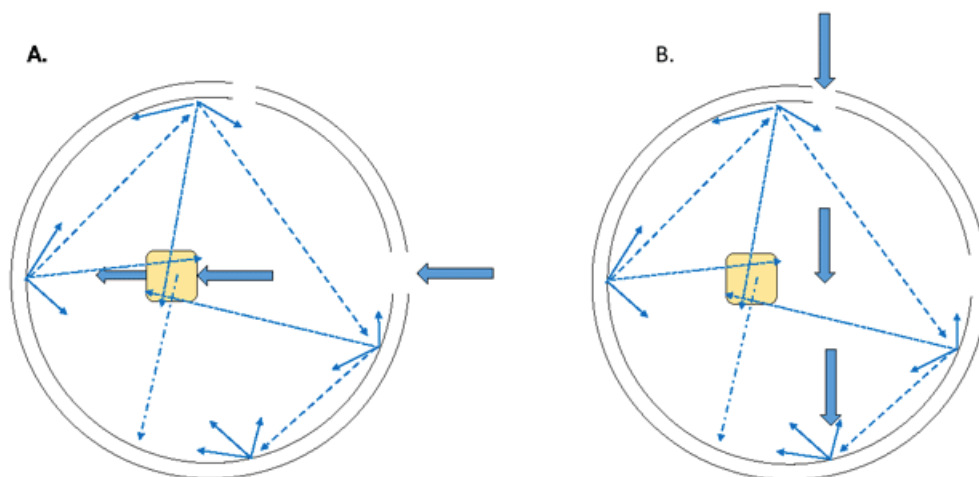


Figure 5.15: Schematic diagram of a center-mounted integrating sphere from the top view (the detector typically lies at the bottom of the sphere and is not shown here) showing the pathways of the A. sample and B. reference beams (large arrows) and the diffuse reflected beams off the integrating sphere (small solid arrows), multiply reflected beams (dashed lines) and multiply reflected beams that pass through the sample (dotted lines) and scattered beam from the sample (dash-dot arrow).

the sample beam with the top of the filter (particle side) facing the incoming sample beam. Thus the sample beam interacts with the particles prior to interacting with the filter. The computation of absorption is

$$a = 2.303 \frac{Abs}{\ell},$$

where the geometric pathlength ℓ is computed from the measured sample volume filtered (ml), and the effective area of the filter (cm^2). The geometric pathlength can be thought of as a cylinder of water of cross section equal to the effective filter area and length ℓ sufficient to account for the volume filtered (Fig. 5.16). The geometric pathlength of the sample on the filter is equivalent to the length of the cylinder of the original suspension.

Comparisons between samples measured in cuvettes with those measured on filter pads indicate that there is an amplification of the mean light ray path through the filter in the original sample beam over that of the geometric path. This is called *pathlength amplification*. It arises as light rays scatter within the filter and particles before exiting into the sphere. It is not accounted for by the reference beam. Careful and extensive paired measurements have demonstrated that the amplification is not negligible (Fig. 5.17).

The pathlength amplification must be corrected before the computation of absorption. The correction, derived empirically from the data in Fig. 5.17A, is

$$Abs_s = 0.323 Abs_f^{1.0867},$$

where Abs_s is the absorbance measured in suspension and Abs_f is the absorbance measured on the filter pad. The absorption coefficient is computed from the absorbance measured

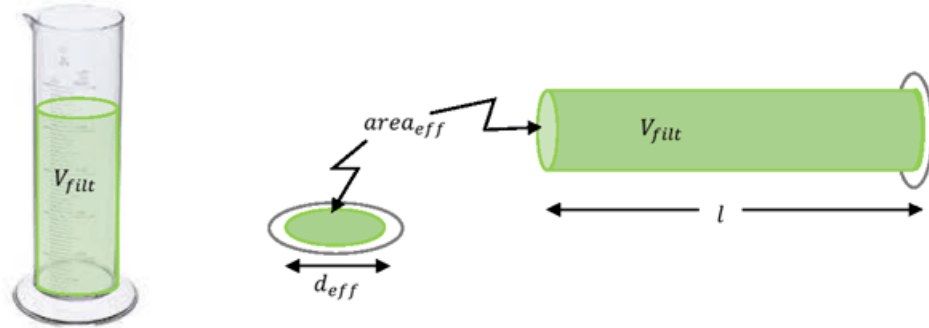


Figure 5.16: Diagrammatic representation of the geometric pathlength of the filtered sample. V_{filt} is the filtered volume, d_{eff} is the diameter of the circular distribution of particles on the filter, and is used to compute the effective area, area_{eff} . The volume filtered can be expressed as a cylinder of area area_{eff} and length, l . The geometric pathlength of the sample is l .

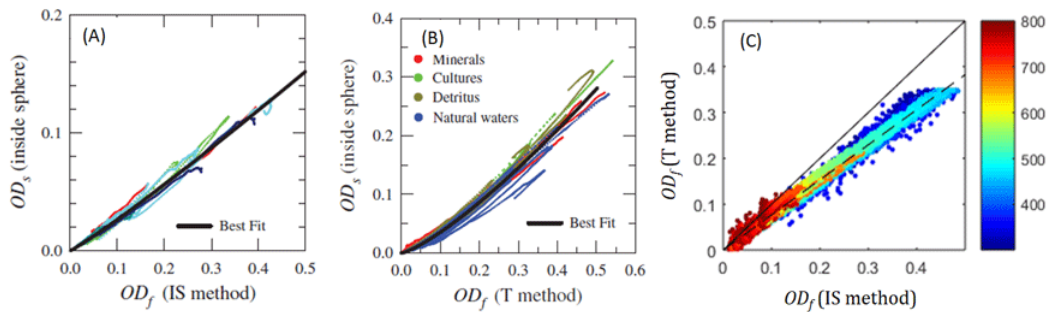


Figure 5.17: Relationship between paired samples measured in suspension in a cuvette mounted inside the integrating sphere (OD_s) versus that measured on the filter pad: A. mounted inside the integrating sphere (OD_f), B. measured in transmission mode. From [Stramski et al. \(2015\)](#). C. Relationship between filters measured in transmission mode compared to filters measured inside the integrating sphere (C. Roesler, unpub. data).

on the filter pad:

$$a = 2.303 \frac{0.323 Abs_f^{1.0867}}{\frac{V_{\text{filt}}}{\pi r_{\text{eff}}^2}} .$$

Historically, the QFT was employed measuring the filters in transmission mode (as in Fig. 5.14, with the cuvette replaced by the filter). In addition to the pathlength amplification within the filter, there was an additional error associated with the loss of nearly half the incident radiant power scattering in the backward direction (away from the detector). The correction for both the scattering loss and pathlength amplification are determined from paired suspension measurements in internally-mounted cuvettes in the integrating sphere and filtered particles on filtered configured in transmission mode,

Abs_{f-T} (Fig. 5.17B):

$$Abs_s = 0.679 Abs_{f-T}^{1.2804} .$$

A simple linear relationship has been found relating the absorption measured on filters in transmission mode, Abs_{f-T} , to that measured in the integrating sphere, Abs_{f-IS} :

$$Abs_{f-IS} = 1.29 Abs_{f-T} + 0.00205 .$$

The linearity is only found for absorbance values less than 0.35 in the transmission mode, and 0.45 in integrating sphere mode (Fig. 5.17C). Beyond that range, the relationship becomes non-linear.

While measuring the absorption by particles on the filter pad comes with measurement challenges and increased uncertainty, it does allow absorption to be determined on optically dilute samples. In addition, the particulate matter can be extracted with methanol (Kishino et al., 1985) to remove the extractable phytoplankton pigments from the particulate matter. The scan of the extracted filter yields the absorbance by the non-algal particle fraction. After computing the absorption spectra for the particulate scan and the non-algal particle scan, the absorption by in vivo phytoplankton pigments (as they were, bound to proteins in the light harvesting complexes in the chloroplasts) can be computed by difference.

Elastic Scattering

As listed in the Absorption Chapter 5, three things can occur when light interacts with matter: absorption, elastic scattering, and inelastic scattering. This chapter discusses the second of these processes, elastic scattering, which refers to any change of direction of the light without a change of wavelength.

Any interaction of light with matter can cause a change in direction. It does not matter whether the change in direction comes from reflection or refraction by a surface (such as sunlight reflected or transmitted by the sea surface), from reflection or refraction by a particle that is much larger than the wavelength of light, or from the light's propagating electric field causing electrons in an atom or molecule to oscillate and radiate light into all directions. Although you will sometimes see terms like “surface scattering” (for reflection by a surface) or “volume scattering” (for interactions within a volume of water), these process are all scattering and are all caused by essentially the same underlying physics, as will be discussed in Section 6.2

This chapter begins with an overview of the role scattering plays in determining the optical environment. The following sections first discuss the physical cause of scattering, namely a spatial change in the real index of refraction. Scattering by air-water surfaces, by particles, and by turbulence are then discussed. Data and models for volume scattering functions and the associated phase functions are then surveyed. Backscattering is given special attention because of its importance to remote sensing.

Mie theory, which describes elastic scattering by homogeneous spheres of any size, is a solution of Maxwell's equations. It is therefore discussed in Chapter 12, after Chapter 11 on Maxwell's Equations.

6.1 Why is Scattering Important?

Ocean-color remote sensing measures light leaving the oceans to determine how absorption in the water column changed sunlight in what was measured. The measured light comes from scattering within the water column—from either backscattering of downwelling sunlight at angles greater than 90 degrees or from multiple forward scattering at angles of a few tens of degrees or greater. Without scattering there would be no remote sensing of the oceans to understand how absorption has modified the in-water light field. To be more

specific, without elastic scattering, downwelling sunlight would continue to head downward into the ocean depths, eventually to be absorbed. The ocean would appear a dim red color due to chlorophyll fluorescence (Section 7.4). Without inelastic scatter, the ocean would be black because there would be no Raman scatter or fluorescence to create even a small amount of upwelling light.

Scattering also enhances the effect of absorption. If light is scattered many times, the distance traveled by a light ray in going from depth z_1 to depth z_2 is greater than just the straight-line geometric distance that would be traveled by the initial ray. This gives the light a greater chance of being absorbed before reaching z_2 . Hence the saying, “A little bit of absorption goes a long ways if you have a lot of scattering.”

The left panel of Fig. 6.1 shows what the world looks like with scattering. The bright light near the Sun (the Sun’s aureole) results from small-angle scattering by atmospheric particles. The halo is caused by light being refracted through hexagonal ice crystals in the cirrus clouds. The blue sky comes from light being scattered through large angles by atmospheric gases (mostly nitrogen and oxygen molecules). The view of the land comes from sunlight being reflected by the surfaces of the rock and ice. The right panel of the figure shows what the same scene would look like without scattering. Without scattering, the only light seen would be the Sun’s unscattered direct beam reduced by a small amount of atmospheric absorption.



Figure 6.1: Left panel: A spectacular scene taken from Annapurna South Base Camp in Nepal showing a 22-deg halo around the Sun. Right panel: what the same scene would look like without scattering. Photo by Anton Yankovyi from [Wikimedia](#), reproduced under the Wikimedia Commons license.

6.1.1 Review of Definitions

For completeness of this chapter, we begin with a quick review of the quantities commonly used to describe scattering. For additional discussion, see the Inherent Optical Properties Chapter 3.

The volume scattering function (VSF) at a particular location \mathbf{x} and wavelength λ , $\beta(\mathbf{x}, \hat{\xi}' \rightarrow \hat{\xi}, \lambda)$, describes the angular distribution of unpolarized light scattered from its initial direction $\hat{\xi}'$ into direction $\hat{\xi}$. In many of the following equations, the location \mathbf{x} and wavelength λ arguments will be omitted for brevity. However, it must be remembered that all IOPs generally depend on location and wavelength. The VSF can be defined as the radiant intensity, $dI(\psi, \alpha)[\text{W sr}^{-1} \text{ nm}^{-1}]$, emanating in direction (ψ, α) centered on $\hat{\xi}'$, from an infinitesimal volume element $dV[\text{m}^3]$, for a given incident irradiance, $E(0)[\text{W m}^2 \text{ nm}^{-1}]$ (recall Eq. 3.9 of the VSF section):

$$\beta(\psi, \alpha) = \frac{1}{E(0)} \frac{\partial I(\psi, \alpha)}{\partial V} \quad [\text{m}^{-1} \text{ sr}^{-1}].$$

The polar (ψ) and azimuthal (α) scattering angles are defined relative to the unscattered direction $\hat{\xi}'$ in a local coordinate system like that of Fig. 1.7.

A measure of the overall magnitude of the scattered light, without regard to its angular distribution, is given by the scattering coefficient, $b[\text{m}^{-1}]$, which is the integral of the VSF over all scattered directions ($4\pi \text{ sr}$) for a given incident direction:

$$b(\hat{\xi}') \triangleq \int_{4\pi} \beta(\hat{\xi}' \rightarrow \hat{\xi}) d\Omega(\hat{\xi}) = \int_0^{2\pi} \int_0^\pi \beta(\psi, \alpha) \sin \psi \, d\psi \, d\alpha,$$

where $d\Omega(\hat{\xi})$ is an element of solid angle centered on direction $\hat{\xi}$. If the medium is isotropic, the scattering does not depend on the incident direction and the scattered light is azimuthally symmetric about the incident direction. Then b is independent of direction and the scattering depends only on the polar angle ψ between the incident ($\hat{\xi}'$) and final ($\hat{\xi}$) directions. The assumption of azimuthal symmetry is valid for spherical particles or randomly oriented non-spherical particles, if the incident irradiance is unpolarized. This assumption is almost always valid for the turbulent aquatic environment with sunlight as the light source. In this case, the previous equation reduces to

$$b = 2\pi \int_0^\pi \beta(\psi) \sin \psi \, d\psi.$$

(Keep in mind that if the light source is a linearly polarized laser, then the scattering depends on the azimuthal angle relative to the plane of polarization.)

Scattering is often described by the phase function, which is the VSF normalized to the total scattering. The phase function provides information on the shape of the VSF regardless of the magnitude of the intensity of the scattered light:

$$\tilde{\beta}(\psi) \triangleq \frac{\beta(\psi)}{b}.$$

Other parameters that define the scattered light include the backscattering coefficient, b_b , which is defined as the total light scattered into the hemisphere from which light has originated (i.e., scattered in the backward hemisphere of directions, denoted by $2\pi_b$):

$$b_b \triangleq \int_{2\pi_b} \beta(\hat{\xi}) d\Omega(\hat{\xi}) = 2\pi \int_{\pi/2}^\pi \beta(\psi) \sin \psi \, d\psi,$$

and the backscattering ratio, which is defined as

$$B = \tilde{b} \triangleq \frac{b_b}{b}.$$

6.2 The Physics of Scattering

This section describes the physics of elastic scattering. Chapter 7 on inelastic scattering discusses the physics of the inelastic processes of Raman scattering and fluorescence.

To say that light undergoes elastic scattering simply means that it changes direction from its initial direction of propagation. This can happen in many seemingly different ways, but fundamentally, *elastic scattering occurs when there is a change in the real part of the index of refraction from one spatial location to another.*

6.2.1 The Index of Refraction

As shown in Section 11.3 on Maxwell's equations in matter, the phase speed of light in a medium is given by $v = 1/\sqrt{\mu\epsilon}$, where μ is the magnetic permeability and ϵ is the electric permittivity of the medium. The real index of refraction n is the ratio of the speed of light in a vacuum to the speed in a medium:

$$n = \frac{c}{v} = \sqrt{\frac{\mu\epsilon}{\mu_0\epsilon_0}},$$

where μ_0 and ϵ_0 are the values in a vacuum. (See Table 11.1 for their values and Section 11.1 for further discussion.) For a dielectric like water, $\mu = \mu_0$. The nondimensional ratio $K = \epsilon/\epsilon_0$ is called the *dielectric constant* even though the value of K depends on frequency and thus is not really a constant. Thus n and K are equivalent via $n = \sqrt{K}$. (A dielectric is a material that conducts electricity poorly; examples are air, water, and glass. Dielectrics are in contrast to conductors like metals, which conduct electricity well. Phase and group speeds are discussed in Section 11.4.)

Pay attention to the frequency dependence when comparing n and K values. If you look up the values of n and K for water in a freshman physics text, it will probably show the values as $n \approx 1.33$ (e.g., Halliday and Resnick (1988, page 867)) and $K \approx 78.5$ (ibid. page 627). Figure 8.4 shows the index of refraction of pure water as a function of wavelength. At the long wavelength/low frequency end of the spectrum, $n \rightarrow 8.85$ corresponding to $K = 78.3$, which is the value for room temperature water. (The value of K is temperature dependent and K at low frequencies decreases from about 88 at 0 C to about 56 at 100 C.) At visible wavelengths, K is around 1.77 and n is around 1.33, consistent with $n = \sqrt{K}$. There are small temperature, salinity, and pressure effects on n at the visible wavelengths of interest here; see Section 8.2 for formulas and tables.

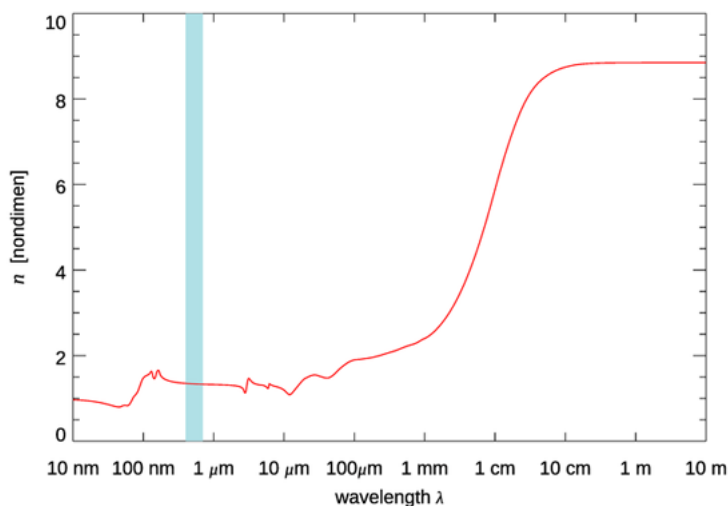


Figure 6.2: The real index of refraction of pure water at room temperature. The blue band shows the visible wavelengths from 400 to 700 nm. Data from [Segelstein \(1981\)](#).

6.3 Models for Scattering

As explained in the previous section, elastic scattering occurs when light travels from a region with one index of refraction into a region with a different index of refraction. In this sense, all scattering is the same. However, there are many ways in which the index of refraction can change, which leads to many ways to *model* the resulting scattering. You will sometimes see terms like “surface scattering” or “volume scattering.” Surface scattering refers to scattering caused by a change in index of refraction at the boundary between two media, such as at the air-water interface. Volume scattering refers to scattering caused by a change in index of refraction within a medium, such as a water body. Volume scattering can be caused by the presence of discrete particles embedded in the medium, by thermal fluctuations in density, or by turbulent mixing of fluids with different physical properties. Those classifications can be useful, but they also can be misleading because they make it seem like there are many unrelated types of scattering, when in reality the terms just describe different ways to change the index of refraction.

6.3.1 Reflection and Refraction: Snell’s Law

The easiest scattering to model is reflection and refraction at a dielectric interface, such as a level air-water surface. At the level of freshman physics, this is usually explained in terms of a plane wave incident onto a boundary between two dielectrics with indices of refraction n_1 and n_2 as illustrated in Fig. 6.3. The red lines in the figure represent lines of constant wave phase, and the red arrow shows the direction of light propagation. The light in medium 1 propagates at a phase speed of $v_1 = c/n_1$, where c is the speed of light in a vacuum; in medium 2 the propagation speed is $v_2 = c/n_2$. If $n_2 > n_1$, the wave slows down as the incident wave front enters medium 2. This causes the wave front to change direction as it enters medium 2, as shown by the green lines and arrow. The *frequency*

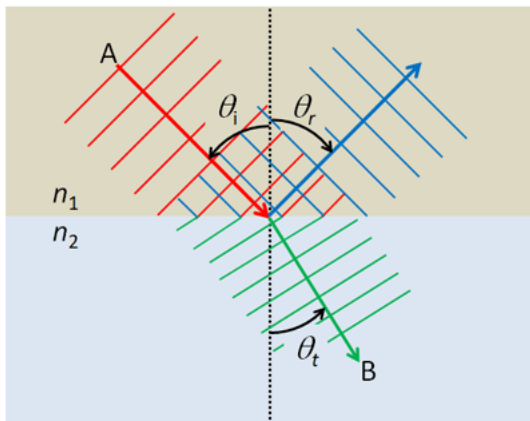


Figure 6.3: Illustration of Snell's law and the law of reflection for an incident plane wave (red) and transmitted (green) and reflected (blue) waves. The lines perpendicular to the arrows represent lines of constant wave phase.

of the wave remains unchanged as the wave passes from medium 1 to 2. Therefore, the wavelength in medium 2 must be less than the wavelength in medium 1: $\lambda_2 = \frac{n_1}{n_2} \lambda_1$.

The relationship between the indices of refraction and the angles of incidence, θ_i , and refraction or transmission, θ_t , is given by Snell's law¹:

$$n_1 \sin \theta_1 = n_2 \sin \theta_2 . \quad (6.1)$$

The surface also reflects some of the incident light back into medium 1, as shown by the blue lines and arrow in Fig. 6.3. For reflection,

$$\theta_r = \theta_i , \quad (6.2)$$

which is known as the law of reflection. A more detailed development of these equations is given in Section 13.1.1.

Snell's law can be derived from fundamental physics at various levels of understanding. One derivation follows from Fermat's Principle of Least Time, which says that light traveling from point A to point B will follow the path that takes the least time. Applied to the geometry of Fig.6.3, this gives the ray path that obey's Snell's law².

Snell's law can be rigorously derived from Maxwell's equations by solving the equations for a plane wave incident onto a dielectric interface. This problem is formulated in terms of the electric fields of the incident, reflected, and transmitted plane waves, and full account is made of the state of polarization of the plane waves. That is a nasty bit of physics and mathematics (e.g., Griffiths (1981, §8.2.5) or Feynman et al. (1964, Chapter 33)).

¹This law was empirically discovered by the Persian Abu Ibn Sahl who reported his discovery in a book *On Burning Mirrors and Lenses* published in Baghdad in 984 CE. Ibn Sahl even used his discovery to design the shapes of lenses that would focus parallel light rays without aberration. The same result was rediscovered in 1621 by Willebrord Snel van Royen (1580-1626; note the spelling of his name). Unfortunately, Ibn Sahl's book was unknown in the west until its rediscovery in the 1930s, so the law became known as Snell's law, although the Ibn Sahl-Snel law would be a more appropriate name. Snel published in Latin under the name Snellius. After Latin was replaced first by German and then English as the common language of western science, Snellius became Snell. There is a move to restore Snel's name to its proper spelling, which I fully support. However, after having been told several times by reviewers that I had misspelled Snell's name, I decided this is a battle not worth fighting at my age.

²To be honest, Fermat cooked up his principle as a way to explain Snell's law, but the principle turns out to have greater applicability. I will refrain from discussing the question of how does the light leaving point A know which path to take to arrive at point B as quickly as possible.

However, the effort required to solve Maxwell's equations pays off well because the solution gives not just Snell's law, which gives only the directions of light propagation, but also Fresnel's equations (Sections 13.2 and 13.3), which give the magnitudes of the reflected and transmitted fields (i.e., of the irradiances) accounting for the state of polarization of the incident light. This solution also shows that the incident, reflected, and transmitted rays must lie in the same plane. The three results of (1) Snell's law, (2) the law of reflection, and (3) the coplanar geometry of the rays are the essence of geometric or ray-tracing optics, which is used in fields such as lens design and Monte Carlo solution of the radiative transfer equation (see Appendix E).

The index of refraction is a bulk material property that parameterizes the accumulated effects of how light interacts with the individual atoms that make up the material. For the most fundamental understanding, Feynman (1985) outlines how Snell's law arises from the properties of individual photons and atoms interacting according to the laws of Quantum Electrodynamics.

Fortunately, for oceanographic purposes, we can just accept Eqs. (6.1) and (6.2) and Fresnel's equations as given and carry on with their applications such as predicting the reflectance and transmittance of wind-blown sea surfaces, as described in Chapter 13 on surfaces.

6.3.2 Scattering by Homogeneous Spheres: Mie Theory

Another way to change the index of refraction is to embed a particle of some index of refraction within a medium with a different index of refraction. If the imbedded particle is a homogeneous sphere, the solution of Maxwell's Equations for a plane wave incident onto the sphere is called Mie theory. The problem is formulated as follows.

- We have given a single, homogeneous sphere of radius ρ , whose material has a complex index of refraction $m_s = n_s + ik_s$. Here n_s is the real index of refraction, and k_s is the complex index of refraction. The complex index is related to the absorption coefficient a_s of the sphere material by $a_s(\lambda) = 4\pi k_s(\lambda)/\lambda$, where λ is the wavelength in vacuo corresponding to the frequency f of an electromagnetic wave (Section 11.3.2).
- The sphere is imbedded in a non-absorbing, homogeneous, infinite medium whose real index of refraction is $m_m = n_m$.
- A plane electromagnetic wave of frequency f is incident onto the sphere. The wavelength on the incident light in the medium is thus $\lambda_m = c/(n_m f) = \lambda/n_m$, which corresponds to a wavelength in vacuo of $\lambda = c/f$.
- We wish to find the electric field within the sphere and throughout the surrounding medium. That is, we wish to determine how the incident light is absorbed and scattered by the sphere, including the angular distribution of the scattered light and its state of polarization.

Mie's solution to this problem is discussed in detail in Chapter 12.

Mie theory is exact and valid for all sizes of spheres, indices of refraction, and wavelengths. Unfortunately, the solution is in the form of infinite series of complicated mathematical functions (Eq. 12.7). The terms in these series depend on a size parameter x ,

$$x \triangleq \frac{2\pi\rho}{\lambda_m} = \frac{2\pi\rho n_m}{\lambda}, \quad (6.3)$$

and on the complex refractive index of the sphere relative to that of the surrounding medium,

$$m = \frac{n_s}{n_m} + \frac{k_s}{n_m}. \quad (6.4)$$

The size parameter x is a measure of the sphere's size relative to the wavelength of the incident light in the surrounding medium. This parameter shows why oceanographers tend to use wavelength rather than frequency as the measure of light's oscillations: it is particle size relative to wavelength that is important for scattering (whether or not the particle is spherical). Note that the real part of the relative refractive index m can be less than 1, for example if the spherical particle is an air bubble ($n_s \approx 1$) in water ($n_m \approx 1.33$).

Because of its computational complexity, Mie's solution was of little use until computers became readily available in the 1960s. However, once it became possible to compute the sums of the infinite series with good accuracy, Mie theory found applications in all fields where scattering by particles is important. One of the first of computer codes, and still perhaps the most widely used, is the BHMIE (Bohren and Huffman Mie) code given in Appendix A of [Bohren and Huffman \(1983\)](#). There are now programs available in all commonly used scientific computer languages (Fortran, C, MATLAB, IDL, and Python). See, for example, the websites at [SCATTERLIB](#) and [Codes for Electromagnetic Scattering by Spheres](#). There are also online Mie calculators that are useful if you only want to obtain a few solutions for small size parameters. See for example [Scott Prahl's Mie Scattering Calculator](#), which is an excellent place to explore Mie Theory. If you need to get serious about doing lots of Mie calculations, download one of the Fortran codes. As Prahl notes on his website, "Let me state that the best Mie codes are in Fortran. Period."

The output of Mie codes is usually given as various absorption and scattering *efficiencies*. The absorption efficiency Q_a , for example, gives the fraction of radiant energy incident on the sphere that is absorbed by the sphere. The term "energy incident on the sphere," means the energy of the incident plane wave passing through an area equal to the cross-sectional (projected, or "shadow") area of the sphere, $A_s = \pi\rho^2$. Likewise, the total scattering efficiency Q_b gives the fraction of incident energy that is scattered into all directions. Other efficiencies can be defined: $Q_c = Q_a + Q_b$ for total attenuation, Q_{bb} for backscattering, and so on.

Mie solutions also can be presented in terms of absorption and scattering *cross sections*. The physical interpretation of these cross sections is simple. The absorption cross section σ_a , for example, is the cross sectional area of the incident plane wave that has energy equal to the energy absorbed by the sphere. The absorption and scattering cross sections are therefore related to the corresponding efficiencies by the geometrical cross section of the sphere. Thus

$$\sigma_a = Q_a A_s = Q_a \pi \rho^2 \quad (\text{m}^2).$$

Likewise, $\sigma_b = Q_b A_s$, and similarly for σ_c and σ_{bb} .

Mie codes also output the scattering phase function $\tilde{\beta}(\psi)$. However, Mie codes usually output *unnormalized* phase functions, so you must always integrate a phase function to verify its normalization. Keep in mind that a phase function must satisfy the normalization $2\pi \int \tilde{\beta}(\psi) \sin \psi \, d\psi = 1$ before it can be used in radiative transfer calculations. This is discussed in more detail in Section [12.4](#).

Understanding the Mie output sometimes involves the nondimensional *phase-shift parameter*

$$\delta = 2x \left(\frac{n_s}{n_m} - 1 \right),$$

and the nondimensional *absorption thickness*

$$\delta_a = 2\rho a_s.$$

It is important to note, when computing the absorption thickness or the imaginary part of the index of refraction, that a_s is the absorption coefficient of the sphere material, not the bulk absorption coefficient of the medium (e.g., of seawater plus phytoplankton). Phytoplankton cells typically have a_s values of 10^4 - 10^5 m^{-1} at visible wavelengths. (Think of the highly absorbing chlorophyll in a phytoplankton cell as the chlorophyll in a can of chopped spinach; you do not see very far through solid spinach.) As an example of the Mie parameter values, consider a phytoplankton cell with $\rho = 4 \text{ }\mu\text{m}$, $n_s = 1.4$, and $a_s = 4 \cdot 10^5 \text{ m}^{-1}$. If the cell is floating in water with $n_m = 1.34$, and if the incident light has $\lambda = 500 \text{ nm}$ in vacuo, then $x = 67.3$, $m = 1.045 + i0.0159$, $\delta = 6.3$, and $\delta_a = 3.2$.

It must be emphasized that Mie's solution assumes that the scattering particle is spherical and homogeneous. Very few particles in nature meet those criteria (an example would be fog droplets in the atmosphere, which are small enough for surface tension to keep the droplet spherical). Most natural particles—phytoplankton, mineral particles, ice crystals, aggregates of spherical particles—are nonspherical and are often also inhomogeneous due to internal structures like chloroplasts or cell walls. The applicability of Mie theory to such particles is always questionable even if numerical results from Mie's equations are within some acceptable error of the correct value, which depends on the problem and the level of accuracy needed. The differences between Mie and reality are often greatest for a computation of interest to oceanographers, namely the computation of backscattered light. [Bohren and Singham \(1991\)](#) have given an excellent review of the inapplicability of Mie theory to backscattering by nonspherical particles.

6.3.3 Scattering by Irregular Particles

Strictly speaking, Mie theory is applicable only to homogeneous spherical particles. That of course does not keep it from being misapplied to non-spherical and/or non-homogeneous particles. Such particles are common in nature: chain-forming or non-spherical diatoms, ice crystals in cirrus clouds, atmospheric dust, and resuspended sediments in water. Fortunately, Mie theory often gives a useful approximate solutions to scattering by irregular particles so long as the particles are close to spherical, such as prolate or oblate spheroids. There are many applications where people use Mie theory with an "equivalent" spherical particle, where the equivalent particle is given the same volume (or cross sectional area) as the irregular particle of interest. For equivalent-volume, roughly spherical particles, Mie may work well. However, for very irregular particles, Mie predictions may disagree greatly with measurements made on such particles. The differences in phase functions between spherical and non-spherical particles are often greatest at backscattering directions, which are of great interest for remote sensing.

In one sense, every spherical particle is geometrically the same; the spheres differ only in size and index of refraction. Thus Mie theory requires only the size parameter x and the relative index of refraction m as inputs. However, irregular particles can have any

imaginable shape and composition. This makes it impossible to have a solution technique for Maxwell's equations that requires only simple inputs.

Many numerical techniques have been developed to compute scattering by irregular particles, all of which are exceedingly mathematical; see [Computational Electromagnetics](#) for an introduction. The two most widely used techniques are probably the *T-matrix Method* (e.g., [Mishchenko et al., 2002](#), Chapter 5) and the *Discrete Dipole Approximation* ([Purcell and Pennypacker, 1973](#)). Although computer codes are available to perform the calculations (e.g., from [SCATTERLIB](#)), they are not trivial to use. In particular, you must first build in the geometry of the particle shape of interest, which in itself can be a major undertaking. For a collection of randomly oriented particles, scattering must first be computed for each possible particle orientation, and then the results must be averaged over all possible particle orientations.

If the irregular particle is much larger than the wavelength of light (e.g., an ice crystal in a cirrus cloud or a large sand grain in the ocean), then geometric optics and ray tracing can be used as the core of Monte Carlo simulations of many rays incident onto randomly oriented particles. As always, you must first program the logic defining the particle shape and how to determine when a ray intersects the surface of the particle. In principle, determining where a ray intersects a particle just amounts to finding the intersection of a line and a surface. However, this is tedious analytical geometry at best and is non-trivial even for simple particle shapes like cubes. But after that logic is programmed, it is just a matter of computer time until enough rays have been traced to give good statistical results.

Enough has now been said about scattering by particles for this overview. The point to be remembered is that a particle is just a discontinuity in a medium's index of refraction. If you insert a relative index of refraction of $m = 1$ into Mie or T-matrix theory, you get no scattering because the particle's properties then match those of the surrounding medium. A non-absorbing particle becomes invisible if its index of refraction matches that of the surrounding medium.

6.3.4 Scattering by Pure Water: Einstein-Smoluchowski Theory

When computing scattering by particles, the surrounding medium is viewed as a continuous, homogeneous medium that itself does not scatter. However, even the purest of water has some scattering. This is because, at the atomic level, pure water is composed of particles, namely water molecules. In water at room temperature, viewed microscopically, the water molecules are moving on average at about 600 m s^{-1} and are continually bumping into each other and flying apart in random directions. These molecular motions are continually creating microscopically small volumes of water that for a very short time have more or fewer molecules in a small volume of space than on average. These fluctuations in the number density of water molecules create small-scale fluctuations in the index of refraction.

The phenomenon of *critical opalescence* refers to a transparent gas or liquid becoming foggy and opaque near a continuous (second-order) phase transition (such when a gas and liquid can coexist at sufficiently high temperature and pressure). An explanation of this phenomenon was lacking until the papers of [v. Smoluchowski \(1908\)](#) and [Einstein \(1910\)](#). They showed that large thermodynamic fluctuations in the material's density, hence in the index of refraction, are responsible for the optical scattering that makes the substance become cloudy. Their explanation required a sophisticated combination of thermodynamics, statistical mechanics, and electromagnetic theory (Maxwell's equations).

See [Morel and Smith \(1974\)](#) for a discussion of the Einstein-Smoluchowski development.

The same process applies to ordinary scattering by water molecules, for which the density fluctuations are much smaller but still non-zero. In this case, the Einstein and Smoluchowski formula for the volume scattering function, with later modifications to account for polarization effects, is

$$\begin{aligned}\beta(\psi) &= \frac{2\pi^2 kT}{\lambda^4 \beta_{\mathbf{T}}} n^2 \left(\frac{\partial n}{\partial P} \right)_{\mathbf{T}}^2 \frac{6 + 6\rho}{6 - 7\rho} \left(1 + \frac{1 - \rho}{1 + \rho} \cos^2 \psi \right) \\ &= \beta(90) \left(1 + \frac{1 - \rho}{1 + \rho} \cos^2 \psi \right).\end{aligned}\quad (6.5)$$

The terms in this equation are as follows, with values shown for a temperature of 20 C, pressure of one atmosphere, and wavelength of 546 nm, as obtained from the formulas of [Buiteveld et al. \(1994\)](#):

k Boltzman's constant; $k = 1.3806 \cdot 10^{-23} \text{ J K}^{-1}$

T Temperature; $T = 293.15 \text{ K}$

n Index of refraction; $n = 1.33477$

$\beta_{\mathbf{T}}$ Isothermal compressibility of water: $4.5893 \cdot 10^{-10} \text{ Pa}^{-1}$

$\left(\frac{\partial n}{\partial P} \right)_{\mathbf{T}}$ Isothermal pressure derivative of the index of refraction; $1.5127 \cdot 10^{-10} \text{ Pa}^{-1}$

ρ depolarization ratio; 0.039 (from [Zhang et al. \(2019\)](#))

λ wavelength; $546 \text{ nm} = 5.46 \cdot 10^{-7} \text{ m}$

P pressure; $1 \text{ atm} = 1.01325 \cdot 10^5 \text{ Pa}$

ψ polar scattering angle

Further discussion of the best values for the various inputs to Eq.(6.5) is given in [Zhang and Hu \(2009\)](#) and [Zhang et al. \(2019\)](#).

The temperature-dependent fluctuations in the index of refraction appear in Eq. (6.5) via the various thermodynamic quantities. The equation shows that scattering will be stronger for a higher temperature; this is because the random molecular motions are then greater, and consequently the density fluctuations are larger. The more compressible a fluid is, the greater the density fluctuations, and the greater the scattering. In particular, if the index of refraction did not depend on the pressure, $\left(\frac{\partial n}{\partial P} \right)_{\mathbf{T}}$ would be zero and there would be no scattering. The scattering is strongly dependent on wavelength, with shorter wavelengths scattering much more strongly than long wavelengths. The $1/\lambda^4$ wavelength dependence seen in (6.5) is the same as for Rayleigh scattering, which describes scattering by spheres that are much smaller than the wavelength of light (and which can be obtained from Mie theory when the particle size is much less than the wavelength of the light). The same wavelength dependence occurs here because in ordinary water the spatial size of the density fluctuations is much less than the wavelength of visible light.

Inserting the values just listed into Eq. (6.5) gives, for a scattering angle of $\psi = 90$ deg,

$$\beta(90, 546 \text{ nm}) = 8.69 \cdot 10^{-5} \text{ m}^{-1} \text{ sr}^{-1}.$$

Writing $f(\rho) = (1 - \rho)/(1 + \rho) = 0.925$, the phase function for pure water is

$$\tilde{\beta}(\psi, 546 \text{ nm}) = \frac{1}{4\pi[1 + f(\rho)/3]} [1 + f(\rho) \cos^2 \psi] \quad (6.6)$$

$$= 0.0608[1 + 0.925 \cos^2 \psi]. \quad (6.7)$$

Finally, it must be pointed out that scattering by pure water (or pure sea water if salinity effects are included in the index of refraction) is one of the few IOPs that can be computed from fundamental physics. (Another such IOP is scattering by a homogeneous sphere, computed using Mie theory.) Other IOPs, absorption coefficients in particular, are obtained from measurements and not from fundamental physics.

6.3.5 Scattering by Turbulence

Yet another way to change the index of refraction is via turbulence-induced fluctuations in temperature and salinity, with the water being viewed as a continuous but slightly inhomogeneous medium. The temperature and salinity fluctuations generate small-scale but spatially continuous changes in the index of refraction, which in turn cause small-angle deviations in the direction of light propagation. Such deviations cause “beam wander” in laser beams and can degrade visibility, which is affected by small-angle scattering. Such effects are often called “optical turbulence.” In the atmosphere, optical turbulence is responsible for the twinkling of stars or the shimmering image of a distant object viewed on a hot day.

We can get an order-of-magnitude estimate of the size of the turbulence-induced fluctuations in the real index of refraction n as follows. Hold the wavelength λ and pressure P constant, so that n is a function only of the temperature T and salinity S . (Water is almost incompressible, which makes the pressure dependence of n much less than that of temperature and salinity.) Then taking differentials of $n(T, S)$ gives

$$\Delta n = \left(\frac{\partial n}{\partial T} \right)_S \Delta T + \left(\frac{\partial n}{\partial S} \right)_T \Delta S,$$

where ΔT and ΔS are small fluctuations in T and S , and Δn is the corresponding change in n . If we square this equation, average it over many turbulent fluctuations, and assume that the fluctuations in temperature and salinity are uncorrelated, we obtain

$$\langle (\Delta n)^2 \rangle = \left(\frac{\partial n}{\partial T} \right)_S^2 \langle (\Delta T)^2 \rangle + \left(\frac{\partial n}{\partial S} \right)_T^2 \langle (\Delta S)^2 \rangle. \quad (6.8)$$

Here $\langle \dots \rangle$ represents the average of the enclosed quantity. The root-mean-square (rms) value of the index-of-refraction fluctuations is then $\sqrt{\langle (\Delta n)^2 \rangle}$. The derivatives $\partial n/\partial T$ and $\partial n/\partial S$ must be evaluated at particular values of T, S, λ and P . We can estimate typical values from the tabulated data in [Austin and Halikas \(1976\)](#). (Some of their data are shown in [Table 8.2](#).) Let us take $T = 20 \text{ C}$, $S = 35 \text{ PSU}$, $\lambda = 546 \text{ nm}$, and $P = 1 \text{ atm}$. Then from the tabulated data of [Austin and Halikas \(1976, page A-6\)](#)

$$\left(\frac{\partial n}{\partial T} \right)_{S=35} = \frac{1.34179 - 1.34442}{30 - 0} = -8.8 \cdot 10^{-5} \text{ C}^{-1}$$

$$\left(\frac{\partial n}{\partial S}\right)_{T=20} = \frac{1.33643 - 1.34295}{0 - 35} = 1.9 \cdot 10^{-4} \text{ PSU}^{-1}$$

Now let us suppose that the temperature and salinity fluctuations are of magnitude $\Delta T \approx 0.005$ C and $\Delta S \approx 0.005$ PSU. Then Eq. (6.8) gives

$$\sqrt{\langle(\Delta n)^2\rangle} = [(-8.8 \cdot 10^{-5})^2 (0.005)^2 + (1.9 \cdot 10^{-4})^2 (0.005)^2]^{1/2} \approx 1 \cdot 10^{-6}.$$

The turbulence-induced fluctuations in the index of refraction are therefore on the order of parts per million.

Such small changes in n are negligible compared to the 3% “fluctuation” in n when light encounters a plankton cell with $n = 1.03$ (relative to the water). However, in the case of particle scattering, we envision the light traveling in straight lines between the occasional encounter with a particle, at which time there may be a large change in direction. In the case of turbulence, we envision many turbulent blobs of water, of many sizes, with the light slightly but continuously changing direction as it passes through water with a continuously varying index of refraction. It is then possible for the cumulative effect of the turbulent fluctuations to change a light’s direction by a fraction of a degree. These turbulence-induced deviations manifest themselves in time-averaged scattering measurements as large values of the VSF $\beta(\psi)$ at very small scattering angles.

Turbulence-induced scatter can significantly degrade the quality of underwater images. The effects of turbulence are especially noticeable in video photography, since the time dependence of the random fluctuations is then apparent, just as it is with the twinkling of stars caused by atmospheric turbulence. Near the boundary between distinctly different water masses, the fluctuations in T and S can be much larger than those assumed above, and even still images can be badly degraded. Fluctuations in T and S can be very large— ΔT of many degrees and ΔS of many PSU—when cold and/or fresh water mixes with warm and/or salty water. This can occur when rain falls on the sea surface, where rivers enter the ocean, and near hydrothermal vents at the sea floor. In these extreme cases, optical turbulence near the boundary of the two mixing water masses can cause striking visual effects such as shimmering of objects seen through the water mass boundary. Figure 6.4 shows an example of this. The fluid being expelled by a hydrothermal vent can be as hot as 350 C, in which case the turbulence-induced fluctuation in n is comparable to the change in n caused by a phytoplankton.

Turbulence effects on optics have been intensively studied by the atmospheric optics community, but less studied by the hydrologic optics community. This is often justified because the turbulence-induced changes in n within water bodies usually are very small and do not significantly effect the distribution of radiant energy within the water. The exception is found in the community of researchers interested in high-resolution underwater visibility and imaging, and in the behavior of coherent light beams. Because the time-averaged effects of turbulence are included in *measured* volume scattering functions, we can assume that those effects are accounted for in the highly peaked VSFs used for oceanic radiative transfer modeling.

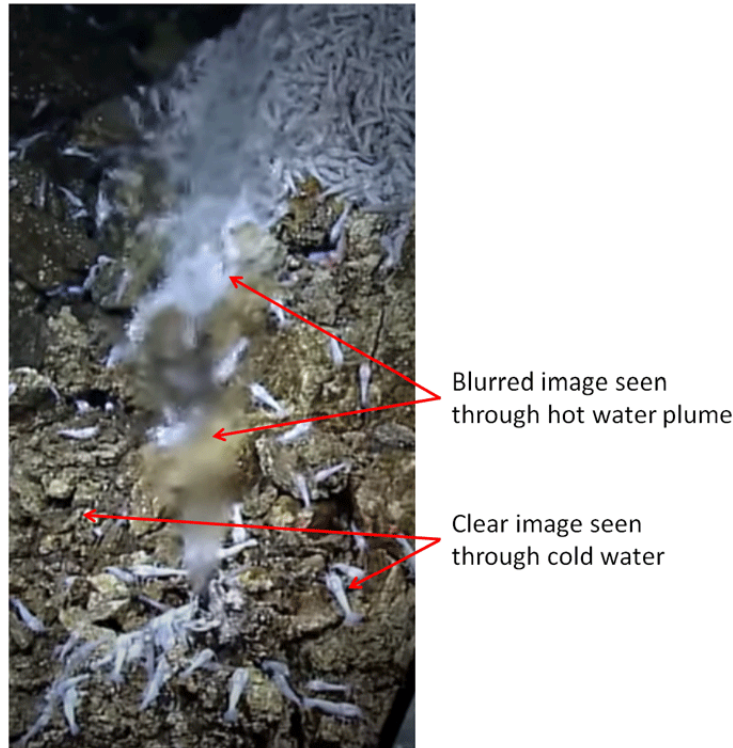


Figure 6.4: Image of rocks and shrimp seen through clear cold water and through a hydrothermal vent of clear very hot water. Frame grab from [Life on a Vent](#); original video from [NOAA National Ocean Service](#).

6.4 Backscattering

As seen in the organization chart of Fig. 1, the absorption coefficient and the volume scattering function are the two fundamental inherent optical properties from which all other IOPs can be obtained. However, for many purposes, the two most *important* IOPs are the absorption (a) and backscattering (b_b) coefficients. This is because the quasi-single-scattering approximation (QSSA; [Gordon \(1994a\)](#) or Section 10.3) shows that to first order the remote-sensing reflectance (R_{rs}) and the near-surface downwelling diffuse attenuation (K_d) depend on a and b_b as

$$R_{rs} \propto \frac{b_b}{a + b_b}$$

$$K_d \propto a + b_b.$$

These two apparent optical properties are the basis for most ocean-color remote sensing and parameterization of in-water light fields for ecosystem modeling.

In addition, as pointed out in [Stramski et al. \(2004a\)](#) and [Boss et al. \(2004b\)](#), the backscattering coefficient gives insight into many ocean ecosystem processes. However, there is both bad and good news. The bad news is that although b_b is to first order determined by the mass of suspended particles in the water, it is also strongly affected by particle size, shape, index of refraction (i.e., composition; organic or inorganic), and

internal structure. This makes it very hard to model b_b with just a single parameter such as the chlorophyll or particulate organic carbon concentration. The good news is that because b_b depends on particle size, shape, index of refraction, and structure, it may be possible to extract information about those parameters from measurements of b_b . The connections between backscattering and other quantities of interest has led to a great deal of research on ways to measure backscattering, either in-situ or remotely by active (lidar) or passive (ocean color) remote sensing; see, for example, [Sauzède et al. \(2016\)](#) and [Bisson et al. \(2021\)](#). [Stramski et al. \(2004a\)](#) give an in-depth review (citing over 180 papers on backscattering) of the contributions to backscattering made by different types of oceanic particles, including colloids, bacteria, phytoplankton, biogenic detritus, minerogenic particles, and bubbles.

6.4.1 Measurement of the Backscattering Coefficient

In principle the backscattering coefficient is obtained by integration of the volume scattering function $\beta(\psi)$ over scattering angles ψ from $\pi/2$ to π (90 to 180 deg):

$$b_b = 2\pi \int_{\pi/2}^{\pi} \beta(\psi) \sin \psi d\psi .$$

In practice this is seldom done because only a few instruments exist for in-situ measurement of the VSF over a sufficient range of scattering angles, and even then usually at just one wavelength.

In the early 1980s a series of backscatter instruments was developed at SRI International ([Moore et al., 1984](#), and references therein) and deployed as part of an underwater lidar experiment known as LIDEX. These instruments and their history are reviewed in [Maffione and Dana \(1997\)](#). The SRI instruments used the basic design seen in the left panel of [Fig. 6.5](#). The first instruments used an LED light source centered at 880 nm because sufficiently bright, visible-wavelength LEDs were not available. When bright, visible-wavelength LEDs became available, the optics and electronics were redesigned and multi-wavelength prototype instruments were made in the early 1990s. At about the same time, [Oishi \(1990\)](#) used Mie theory to investigate the optimum scattering angle for use in estimating b_b from a measurement of the VSF at one scattering angle.

The first commercial instrument based on the SRI design was the HydroScat-6 ([HO-BILabs](#)), followed shortly thereafter by the BB series (initially at 3 wavelengths, now at 9; [Sea-Bird Scientific, Inc.](#)), and most recently by the Hyper-bb (hyperspectral; [Sequoia Scientific, Inc.](#)) and the SC6 (6 wavelengths; [In-Situ Marine Optics](#)).

For each of the instruments just mentioned, a light source sends light into the water in a narrow range of angles about some central direction, which is shown by the red arrow in panel (a) of [Fig. 6.5](#). The detector has a field of view that intersects the cone of emitted light, creating a volume (shaded) where light can be scattered through a nominal scattering angle ψ into the detector, as shown by the blue arrow. The instrument is actually measuring the volume scattering function $\beta(\psi)$ over a narrow range of scattering angles as weighted by the ψ -dependent scattering volume. This measurement is then converted to an estimate of the total (water plus particles) backscattering coefficient via

$$b_b = 2\pi\chi(\psi)\beta_{\text{meas}}(\psi) \tag{6.9}$$

$$= 2\pi\chi_p(\psi)[\beta_{\text{meas}}(\psi) - \beta_{\text{water}}(\psi)] + b_{\text{bw}} . \tag{6.10}$$

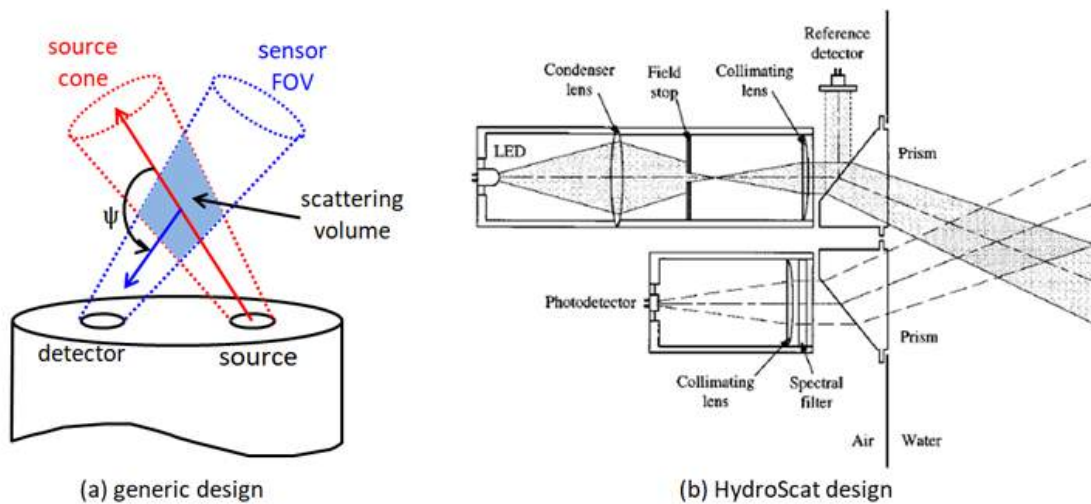


Figure 6.5: Backscattering sensor design. The left panel shows the generic sensor design. The right panel shows the implementation in the HydroScat instruments. (Panel (b) figure reproduced from [Maffione and Dana \(1997\)](#) under Optica’s Fair Use License.)

Here χ is a conversion factor that converts a measurement of the volume scattering function at one wavelength into an estimate of the backscattering coefficient.

The choice of an optimum scattering angle and associated χ value has been the subject of several studies. [Oishi \(1990\)](#) found, based on Mie calculations for assumed particle size distributions, that a scattering angle near 120 deg gave the most data points (94%) within $\pm 10\%$ of the correct value, although 140 deg gave the smallest maximum error (17.7%). [Maffione and Dana \(1997\)](#) argued that measurement at $\psi \approx 140$ deg is optimum because there is less variability in χ for various volume scattering functions. [Boss and Pegau \(2001\)](#) argue that $\psi \approx 120$ deg is optimum because then χ for water is the same as the χ for particle phase functions, in which case the total b_b can be determined by Eq. (6.9) without the need to explicitly remove the water contribution as seen in Eq. (6.10). A recent study by [Zhang et al. \(2021\)](#) also finds $\psi \approx 120$ deg to be optimum, and they find that χ_p is independent of wavelength (as was predicted by Oishi). [Zhang et al. \(2017\)](#) provide a physical understanding of why a scattering angle near 120 deg shows minimum variance in χ_p . In principle, the optimum scattering angle and the corresponding value of χ depend on the shape of the particle phase function. However, measurements for many VSFs in both open-ocean and coastal waters show that, for a given ψ , $\chi(\psi)$ varies by only a few percent, so the choice is not critical; see [Sullivan et al. \(2013\)](#) for further discussion and additional references. The HydroScat instruments have a nominal scattering angle of $\psi = 140$ deg; the bb-9 instrument uses $\psi = 117$ deg; the Hyper-bb uses $\psi = 135$ deg, and the SC6 uses $\psi = 120$ deg. Figure 6.6 shows the values of $\chi(\psi)$ for pure water and for particle phase functions.

Other instrument designs have been proposed. The ECO-VSF 3 instrument uses the same generic design but measures the VSF at 3 angles, $\psi = 104, 130$ and 150 deg, and at 3 wavelengths. The values for the three scattering angles are fit with a function, which is then integrated from 90 to 180 deg to get the backscattering coefficient. An entirely different instrument design is described in [Haubrich et al. \(2011\)](#), but this design has not

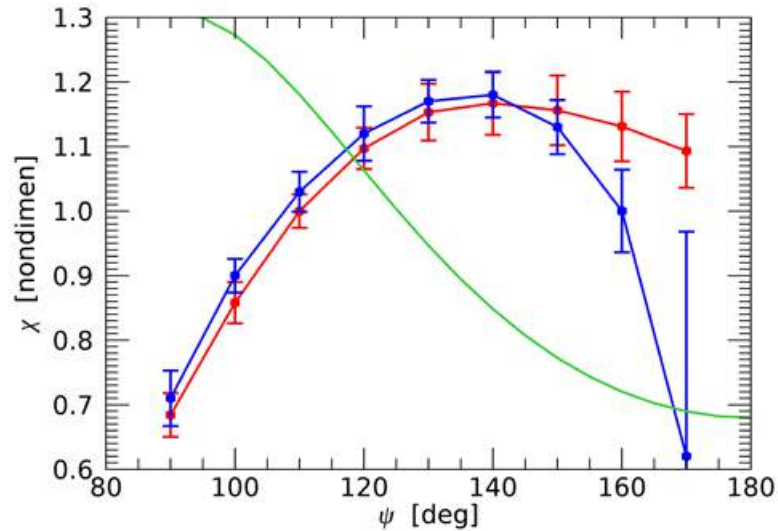


Figure 6.6: Conversion factors $\chi(\psi)$. Green: $\chi_w(\psi)$ for pure water. Red: Data from Sullivan and Twardowski (2009, Table 3). Blue: Data from Boss and Pegau (2001, Table 1). Similar curves are found in the papers by Zhang et al. (2017) and Zhang et al. (2021).

been commercialized.

The path lengths of these instruments are at most a few tens of centimeters, and consequently the backscattering signal is very small: typically about 10^{-8} of the emitted power is scattered into the detector. Sensitivity, stability, and dynamic range in the electronics are therefore critical, as is the instrument calibration. There are three ways to calibrate these instruments. The first is to measure backscattering in a liquid with known backscattering properties; this is sometimes used for laboratory instruments. The second way is to measure the reflectance of a plaque of known reflectance; the details are given in Maffione and Dana (1997). The third way is to measure backscattering in a solution containing spherical beads of known properties; the details are given in Sullivan et al. (2013).

6.4.2 Models for the Particle Backscattering Coefficient

A number of early investigations used Mie theory (Chapter 12) to study the effects of particle size distributions (PSDs; Section 8.12) and indices of refraction on backscattering (e.g., Brown and Gordon, 1974; Morel and Bricaud, 1981b; Bricaud and Morel, 1986; Ahn et al., 1992; Ulloa et al., 1994). These calculations gave great insight into particle optical properties as determined by particle index of refraction and size distribution. However, the Mie-predicted backscattering was generally lower than measurements. For example, Brown and Gordon (1974) used oceanic PSDs measured with a Coulter counter down to a particle diameter of $0.65 \mu\text{m}$ combined with a range of particle complex indices of refraction to compute particle volume scattering functions (VSFs) and absorption and scattering coefficients. The Mie-computed IOPs were compared with absorption and beam attenuation coefficients and VSFs measured simultaneously with the PSDs. As stated in Brown and Gordon (1974, page 2875), “The model fits the data poorly for $\theta \gtrsim 10^\circ$ and in fact is almost 2 orders of magnitude too low for $\theta \gtrsim 90^\circ$. We attempted to increase

the backscattering by adding particles smaller than $0.65 \mu\text{m}$ and found that when the backscattering became the same order as the observations, the forward scattering was about an order of magnitude too large.” [“The model” is Mie theory and θ is the scattering angle.] They then experimented with various combinations of high (mineral) and low (organic) indices of refraction in various size ranges, but they were unable to get agreement between Mie and observations over all scattering angles and wavelengths simultaneously. They concluded that there must be large numbers of very small particles, which were not measured by the Coulter counter but which could increase the amount of predicted backscattering if included in the Mie calculation input. They concluded (page 2881), “The prediction of the existence of vast quantities of small organic particles cannot be verified at this time, since little is known about seawater organics in these small sizes.” This disagreement between Mie and measurement became known as the “missing backscatter problem.”

Brown and Gordon’s prediction of large numbers of particles less than $0.65 \mu\text{m}$ in size turned out to be correct (e.g. the photosynthetic bacteria *Prochlorococcus*, discovered in 1986 by Chisholm et al. (1988), and sub-micrometer particles measured by Koike et al. (1990)). However, such particles were not the cause of the missing backscatter in Mie calculations. It was soon recognized that Mie theory underestimates backscattering for particles that are nonspherical and/or inhomogeneous. Indeed, Bohren and Singham (1991) even wrote a delightful review article on the inapplicability of Mie theory for backscattering calculations.

The next step in modeling realism was the use of a “coated sphere” (describing, for example, a cell wall and interior cytoplasm, or a microbubble with an organic film on its surface) or a “three-layer” particle geometry. Bricaud et al. (1992) used a three-layer model of coccolithophorids as a high-index ($n = 1.22$, non-absorbing) calcite shell, an intermediate absorbing layer representing chloroplasts, and a low-index ($n = 1.015$, non-absorbing) cytoplasm core. They found that “While the internal structures induce insignificant modification in absorption and only weak modifications in total scattering, they appear to be able to increase the backscattering efficiency by a factor as high as 50, depending mainly on the calcite shell thickness. The internal structures also induce spectral changes in backscattering.” Kitchen and Zaneveld (1992) used a three-layer model to study backscattering by generic phytoplankton and likewise found increased backscattering compared to a homogeneous sphere. It is now recognized (e.g., Organelli et al., 2018) that there never was any “missing backscattering.” What was missing was the use of models capable of parameterizing the morphological complexity of oceanic particles.

As seen in Section 8.9 (e.g., Fig. 8.30), the phytoplankton absorption coefficient correlates fairly well with the chlorophyll concentration in Case 1 waters; this is the basis for bio-optical models of absorption as a function of the chlorophyll concentration. However, the backscattering coefficient does not correlate nearly so well with the chlorophyll concentration because of the effects of particle shape and internal structure. Figure 6.7 shows the particle backscattering coefficient b_{bp} at 700 nm as a function of the chlorophyll concentration (Chl) for both surface and euphotic zone waters. For a given chlorophyll value, there is an order-of-magnitude spread of $b_{\text{bp}}(700)$ values. In that figure, b_{bp}^{k} is the background value of NAP backscattering that does not covary with Chl. The blue line in the plots is the best fit model of the form of Eq. (6.11).

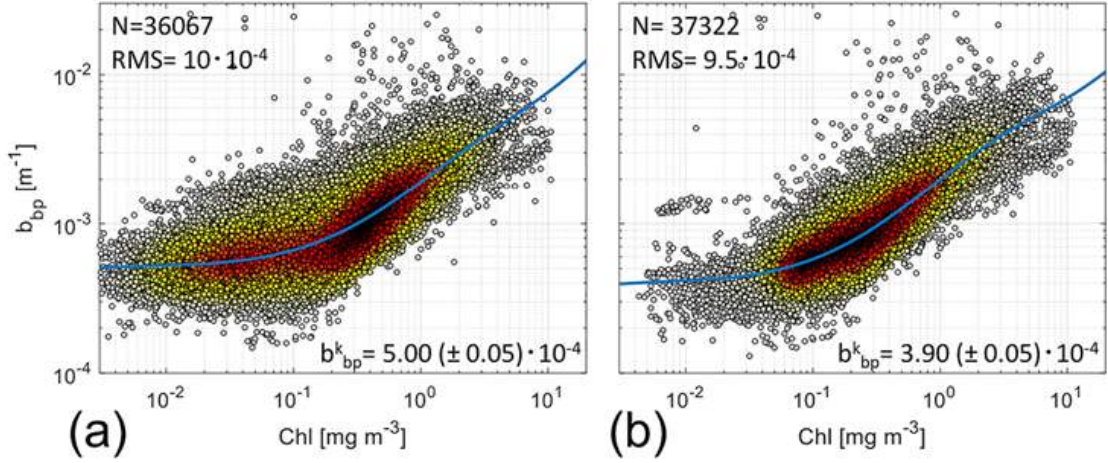


Figure 6.7: Particle backscattering coefficient at 700 nm as a function of the chlorophyll concentration from a global Biogeochemical-Argo dataset. Panel (a), data from only within the surface layer; panel (b), data from the entire euphotic zone. The blue line is the best fit from Eq. (6.11). b_{bp}^k is the background value of NAP backscattering. (Reproduced from Fig. 2 of [Bellacicco et al. \(2019\)](#) under AGU’s reproduction policy.)

Given data like those of Fig. 6.7, one can fit functions of various forms. Power-law functions of the form

$$b_{bp}(\lambda) = A(\lambda)Chl^{B(\lambda)}$$

are often used for such modeling. For example, [Huot et al. \(2008, Eq. 8\)](#) uses this form with $A(\lambda) = A_1 + A_2(\lambda - 550)$ and $B(\lambda) = B_1 + B_2(\lambda - 550)$. This leaves four parameters (A_1, A_2, B_1, B_2) to be determined by an error minimization to get the best fit for a given data set.

[Brewin et al. \(2012\)](#) modeled backscattering as a sum of three terms representing pico-, nano-, and microphytoplankton (subscripts 1, 2, 3, respectively, in the following equations). Analysis of their data justified combing the pico- and nanoplankton terms and adding a constant (independent of chlorophyll) background term. Their model then took the form

$$b_{bp}(\lambda) = C_{1,2}^m [b_{bp,1,2}^*(\lambda) - b_{bp,3}^*(\lambda)] [1 - \exp(-S_{1,2} Chl)] + b_{bp,3}^*(\lambda) Chl + b_{bp}^k(\lambda).$$

In this equation, subscript 1,2 indicates the combined values for pico- and nanoplankton. $C_{1,2}^m$ and C_3^m are the maximum possible values of the chlorophyll concentrations for these components; these were determined from independent data and are held fixed in the model. The slope parameter $S_{1,2}$ is also held fixed. They then modeled the wavelength-dependent parameters as power-law functions. After further analysis (e.g., the wavelength dependence of $b_{bp,3}^*(\lambda)$ is not statistically significant), they eventually obtained ([Brewin et al., 2012, Eq. 22](#))

$$\begin{aligned}
b_{\text{bp}}(\lambda) = & b_{\text{bp},1,2}^*(\lambda_o) \left(\frac{\lambda}{\lambda_o} \right)^{-\gamma_{1,2}} C_{1,2}^m [1 - \exp(-S_{1,2} Chl)] \\
& + b_{\text{bp},3}^*(\lambda_o) [Chl - C_{1,2}^m [1 - \exp(-S_{1,2} Chl)]] \\
& + b_{\text{bp}}^k(\lambda_o) \left(\frac{\lambda}{\lambda_o} \right)^{-\gamma_k}.
\end{aligned} \tag{6.11}$$

This leaves five parameters to be determined for a given data set: $b_{\text{bp},1,2}^*(\lambda_o)$, $b_{\text{bp},3}^*(\lambda_o)$, $b_{\text{bp}}^k(\lambda_o)$, $\gamma_{1,2}$ and γ_k .

Figure 6.8 shows the wavelength dependence of b_{bp} for low to high chlorophyll values and the chlorophyll dependence for blue to red wavelengths, as computed by Eq. (6.11) and typical values of the model parameters ($b_{\text{bp},1,2}^* = 0.003$, $b_{\text{bp},3}^* = 0.0003$, $b_{\text{bp}}^k = 0.0005$, $\gamma_{1,2} = 1.05$, $\gamma_k = 2.7$, $C_{1,2}^m = 0.780 \text{ mg m}^{-3}$, $C_3^m = 0.147 \text{ mg m}^{-3}$, $S_{1,2} = 1.14$, and $\lambda_o = 700$; see the tables in [Brewin et al. \(2012\)](#)).

The model of Eq. (6.11) is based on sound biology and it does a good job *on average* of fitting the swarm of data points seen in Fig. 6.7. However, it must be noted that, for a given chlorophyll value, the actual data can differ from the model value by a factor of 2 or more. Such bio-optical models can give good results on average, and they do give insight into the ecosystem behavior. However, in any particular instance, the model-predicted backscattering coefficient may differ from a measured value by a large amount.

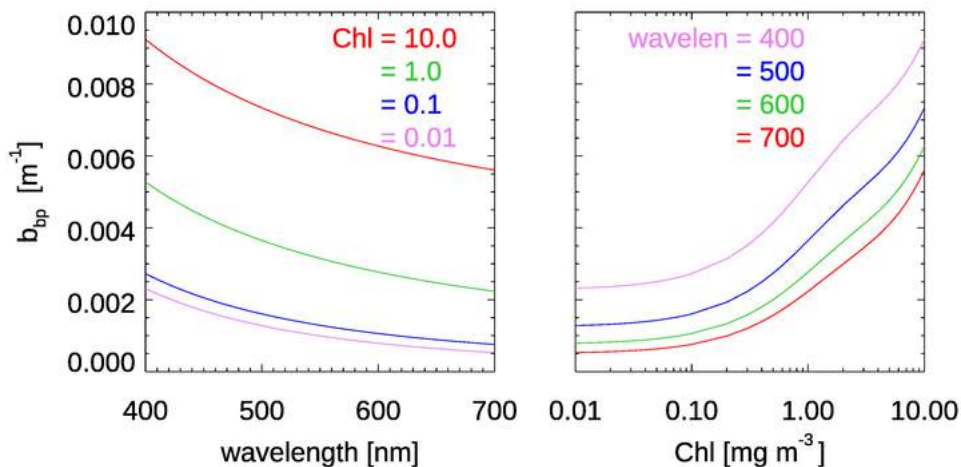


Figure 6.8: Left panel: particle backscattering coefficients as functions of wavelength for selected chlorophyll concentrations from Eq. (6.11) and typical values of the best-fit parameters. Right panel: as functions of chlorophyll for selected wavelengths.

[Zhang et al. \(2020\)](#) studied backscattering in the oligotrophic North Pacific partitioned by size classes for particles less than $0.2 \mu\text{m}$, less than $0.7 \mu\text{m}$ (both as determined by what passes through filters of a given size), and bulk (all sizes of particles). They found that for $Chl < 0.1 \text{ mg m}^{-3}$, sub-micrometer-sized particles contribute as much as one-half of the total backscattering. They found that the smallest size class gave a chlorophyll-

independent background contribution to b_{bp} , similar to the constant background term in the Brewin model.

There is a somewhat better correlation between b_{bp} and particulate organic carbon (POC) than with chlorophyll. An example is seen in Fig. 6.9. Such correlations have led to a number of studies to develop algorithms to retrieve POC from remotely-sensed b_{bp} values (e.g. Stramski et al., 1999). However, as seen by the various best-fit curves in the figure, data-derived algorithms relating POC to b_{bp} are often specific to particular regions of the global ocean. b_{bp} correlates best with total suspended particulate matter (SPM) (e.g. Boss et al., 2009c). Reynolds et al. (2016) found, for data from Arctic waters, that the Chl- $b_{bp}(550)$ relation had $R^2 = 0.37$, POC- $b_{bp}(550)$ had $R^2 = 0.60$, and SPM- $b_{bp}(550)$ had $R^2 = 0.88$.

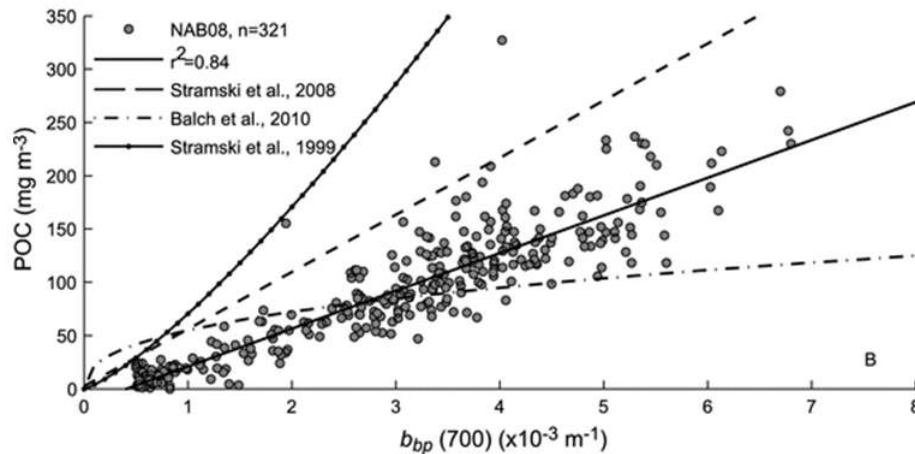


Figure 6.9: Particle backscattering at 700 nm as a function of particulate organic carbon for North Atlantic waters. The labeled curves are the values predicted by various models for $POC(b_{bp})$ in other areas of the global ocean. Reproduced from Fig. 6B of Cetinić et al. (2012) under Wiley’s license for non-commercial uses.

6.4.3 Models for the Particle Backscattering Fraction

The particle backscattering fraction

$$\tilde{b}_{bp} = \frac{b_{bp}}{b_p}$$

is frequently used in bio-optical models of scattering. If \tilde{b}_{bp} is known, then the backscattering coefficient can be determined from easily made measurements of $b_p = c_p - a_p$. Mobley et al. (2002) showed that using a phase function with the correct backscattering fraction is crucial in obtaining agreement between modeled and measured in-water light fields. Twardowski et al. (2001) and Boss et al. (2004a) showed how the backscattering fraction can be used in studies of particle composition.

Unfortunately, \tilde{b}_{bp} is almost uncorrelated with the chlorophyll concentration, even in Case 1 waters. This can be expected because both b_{bp} and b_p are, to first order, determined by the concentration of material in the water. Therefore, their ratio is roughly independent

of concentration, which means that particle size, shape, and composition play a greater role in determining the value of \tilde{b}_{bp} than for either of its components. Figure 6.10 shows shows over 9,000 values of $\tilde{b}_{bp}(555 \text{ nm})$ as determined from HydroScat-6 measurements of the backscattering coefficient and ac-9 measurements of the scattering coefficient (computed from attenuation minus absorption) versus the corresponding chlorophyll values determined from measured chlorophyll fluorescence. Clearly, there is very little correlation between $\tilde{b}_{bp}(555 \text{ nm})$ and Chl , even in Case 1 waters.

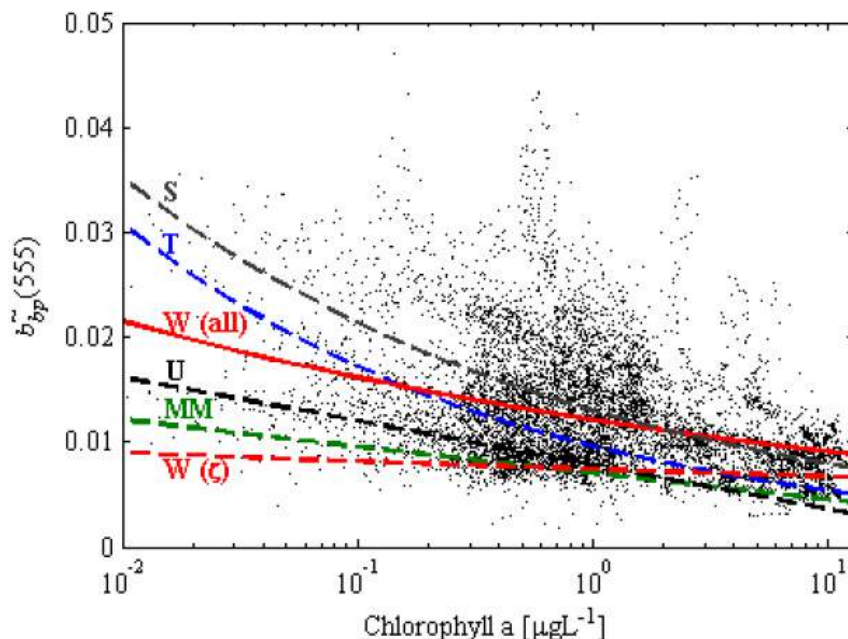


Figure 6.10: Particle backscattering fraction at 555 nm as a function of the chlorophyll concentration. The labeled curves are the values predicted by various models for $\tilde{b}_{bp}(Chl)$. Reproduced from Fig. 5 of [Whitmire et al. \(2007\)](#) under Optica’s Fair Use License.

Mie theory shows that for homogeneous spherical particles with a Junge size distribution and a wavelength-independent index of refraction, the scattering and backscattering coefficients have the same wavelength dependence ([Ulloa et al., 1994](#)). This will not hold true for particles in general, however the wavelength dependence of \tilde{b}_{bp} is not strong. Figure 6.11 shows the spectral dependence of the data shown in Fig. 6.10. Although some situations such as a high mineral particle load may give a significant wavelength dependence of the backscattering fraction, in most cases the error bars are consistent with the assumption of a wavelength-independent backscattering fraction.

6.4.4 Backscattering at 180 degrees

There is yet another measure of backscattering that is fundamental to one particular application: lidar remote sensing. Lidar can be used to remotely sense in-water properties such as the depth of a scattering layer (e.g., [Hoge et al., 1988](#)). *Monostatic* lidar refers to a configuration in which the transmitting and receiving optics (often a telescope) are at the same spatial location. (In a bistatic system, the transmitter and receiver are different opti-

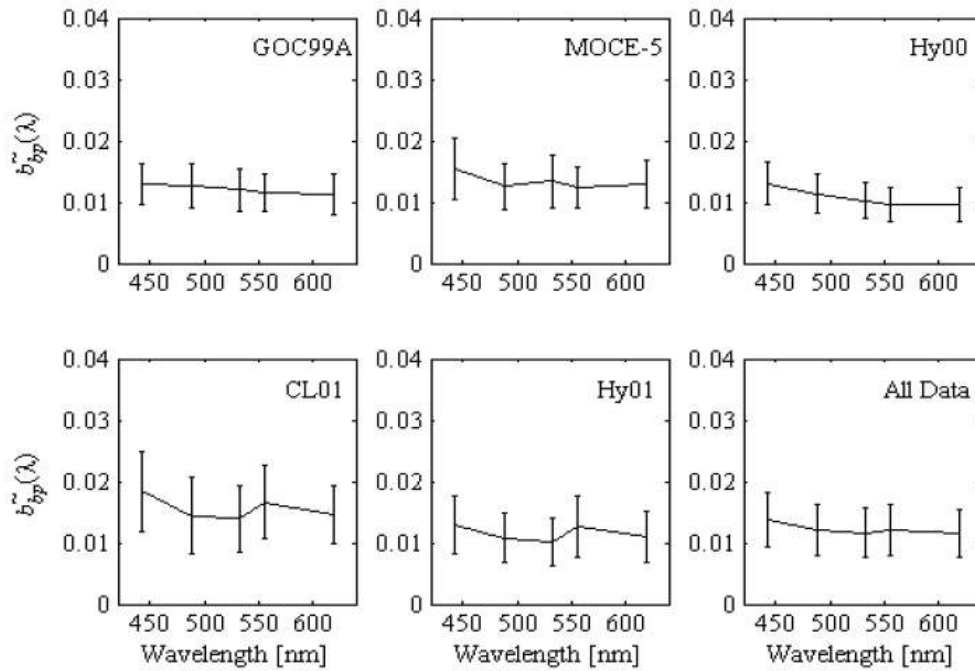


Figure 6.11: Particle backscattering fraction as a function of wavelength for the data of Fig. 6.10. The labeled panels refer to various data sets. The error bars are one standard deviation of the measured values. Reproduced from Fig. 4 of Whitmire et al. (2007) under Optica’s Fair Use License

cal components at fixed but different locations.) In a time-gated, monostatic lidar system, a pulse of light is sent out, and the light scattered by the water column returns directly back to the lidar instrument, where it is measured at a later time. As seen in Eq. (9.26) of Section 9.6, the volume scattering function at exactly (or very nearly) $\psi = 180$ deg is one of the parameters needed to predict the return signal in a monostatic configuration.

Enhanced backscatter refers to coherent backscatter when light travels back and forth along the same path through a medium. It can be caused by several different processes (Barabanenkov et al., 1991), but the end result is that, if it is present, the VSF will show a sharp peak very near $\psi = 180$ deg. In the early days of lidar remote sensing of the oceans, there was speculation that enhanced backscatter might explain a factor of 2 or 3 difference in predicted and measured lidar return signals. Existing instruments for measurement of the VSF (e.g., as used by Petzold to obtain the VSF measurements seen in the next section) did not make measurements beyond about 170 deg and therefore could not test this hypothesis. A novel instrument, called Beta Pi for its measurement of $\beta(\psi = \pi)$, was therefore developed at SRI International by the same group that developed the backscattering instruments described above (Maffione and Honey, 1992; Maffione, 1996; Maffione and Dana, 1996). The Beta Pi instrument measured the VSF from $\psi = 178.8$ to 180 deg with a resolution of 0.02 deg. Figure 6.12 shows an example of data taken by Beta Pi.

Beta Pi measurements in a variety of ocean waters showed that there was no factor-of-two enhancement of the VSF near 180 deg. There was consequently little interest in

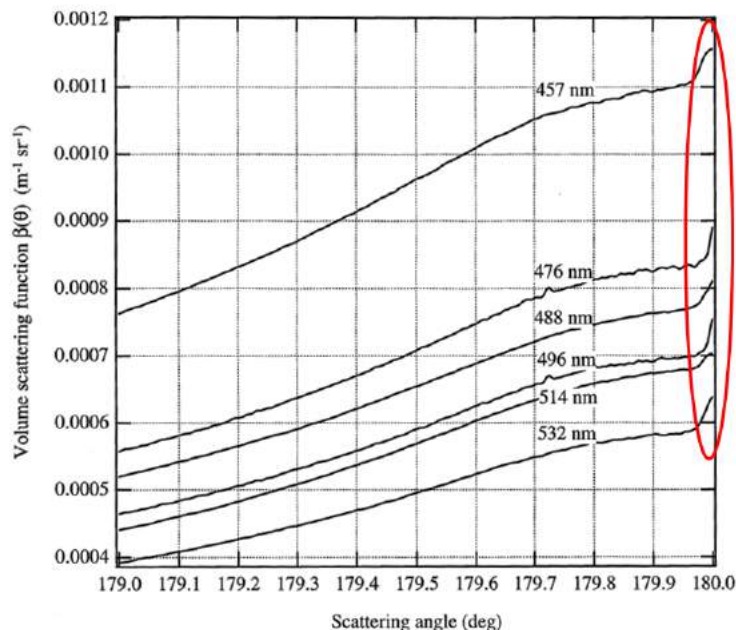


Figure 6.12: Measurements of the VSF near $\psi = 180$ deg made by the Beta Pi instrument. The red oval shows the location of the small amount of enhanced backscattering observed in ocean water. Annotated from Maffione (1996, Fig. 3.5).

further measurements, funding dried up, and the unique Beta Pi instrument, developed at great cost, was eventually taken to the scrap yard.

Enhanced backscattering remains a topic of active research in atmospheric optics, relevant for example to lidar measurements of atmospheric particles or radar detection of raindrops. The “backscattering cross section” σ_b that is a standard output of Mie codes (Section 12.4) is proportional to $\beta(\psi = \pi)$. The Mie σ_b is often called the radar cross section; see Bohren and Huffman (1983, Section 4.6) for discussion. Hu et al. (2020) studied the relation between the particle VSF $\beta_p(\psi = \pi)$ and the particle backscattering coefficient b_{bp} as a function of particle shape. The conversion factor $\chi_p(\psi) = b_{bp}/(2\pi\beta_p(\psi = \pi))$ was found to depend on particle shape and internal structure, but was almost independent of wavelength.

6.5 Petzold’s VSF Measurements

Several early researchers in optical oceanography built instruments to measure the volume scattering function (VSF) of oceanic waters. (See Jerlov (1976) for data and references for early measurements.) The most carefully made and widely cited scattering measurements are found in the classic technical report of Petzold (1972) (summarized in Petzold, 1977). He combined two instruments, one for VSF measurements at very small angles ($\psi = 0.172, 0.344,$ and 0.688 deg) and one for angles between 10 and 170 degrees, to obtain VSF measurements over almost the whole range of scattering angles. Petzold’s report describes his instruments, their calibration and validation, and tabulates data from very clear (Bahamas), productive coastal (California), and turbid harbor (San Diego, California)

waters. The Petzold VSFs and phase functions plotted here can be downloaded from the corresponding Web Book page.

Figure 6.13 shows three of Petzold's VSF curves displayed on a log-log plot to emphasize the forward scattering angles. The same data are displayed on log-linear axes in Fig. 6.14, which emphasizes large scattering angles. The instruments he used had a spectral response centered at $\lambda = 514\text{ nm}$ with a bandwidth of 75 nm (full width at half maximum). In these figures the top (red) curve was obtained in the very turbid water of San Diego Harbor, California; the center (green) curve comes from near-shore coastal water in San Pedro Channel, California; and the bottom (blue) curve is from very clear water in the Tongue of the Ocean, Bahama Islands. The striking feature of these volume scattering functions from very different waters is the similarity of their shapes.

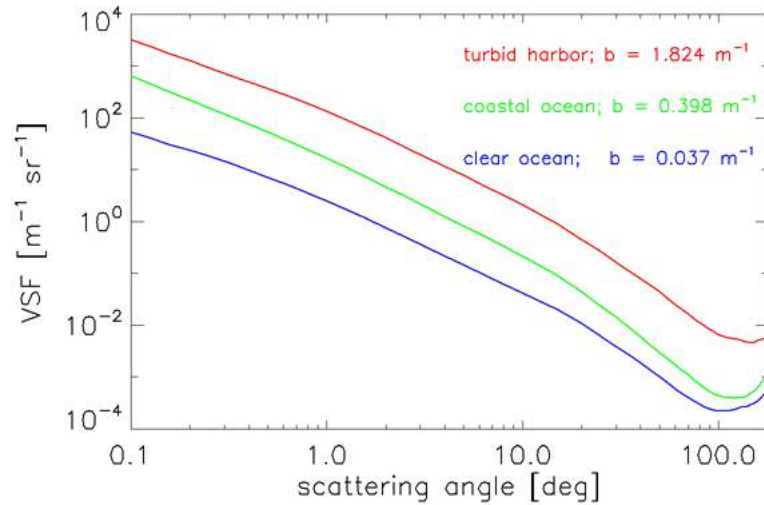


Figure 6.13: Log-log plots of Petzold's measured volume scattering functions from three different waters, as labeled.

Although the scattering coefficients b of the curves in Figs. 6.13 and 6.14 vary by a factor of 50, the uniform shapes suggest that it may be reasonable to define a "typical" particle phase function $\tilde{\beta}_p(\psi)$. This has been done with three sets of Petzold's data from waters with a high particulate load (one set being the top curve of Figs. 6.13 and 6.14), as follows (Mobley et al., 1993):

1. Subtract the pure sea water VSF at 514 nm from each curve to get three particle volume scattering functions $\beta_p^i(\psi)$, $i = 1, 2, 3$.
2. Obtain the corresponding particle scattering coefficients from $b_p^i = b^i - b_{\text{water}}$;
3. Compute three particle phase functions via $\tilde{\beta}_p^i(\psi) = \beta_p^i(\psi)/b_p^i$;
4. Average the three particle phase functions at each scattering angle to define the "average particle" phase function.

The three phase functions so obtained and the average-particle phase function are shown in Figs. 6.15 and 6.16. This average-particle phase function satisfies the normalization condition $2\pi \int_0^\pi \tilde{\beta}_p(\psi) \sin \psi d\psi = 1$ if a behavior of $\tilde{\beta}_p \propto \psi^{-m}$ is assumed for $\psi < 0.1$ deg

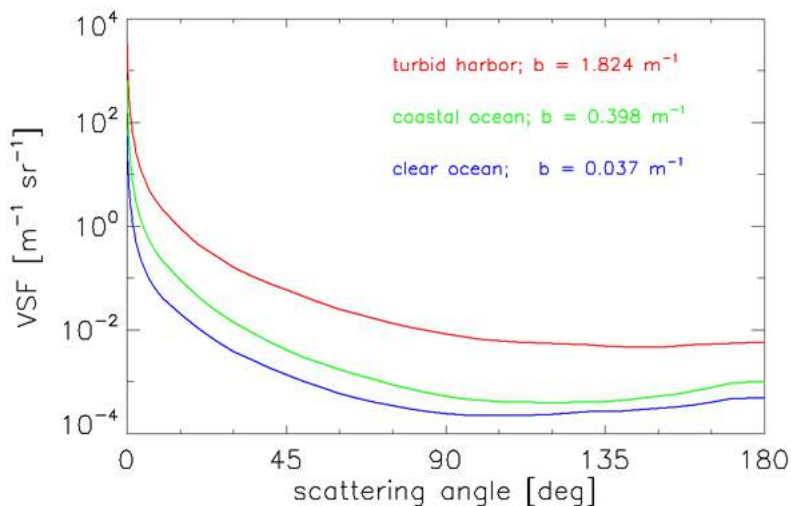


Figure 6.14: Log-linear plots of Petzold’s measured volume scattering functions from three different waters, as labeled.

and a trapezoidal-rule integration is used for $\psi \geq 0.1$ deg, with linear interpolation in $\log \tilde{\beta}_p(\psi)$ versus $\log \psi$ used between the tabulated values. Here $m = 1.346$ is the negative of the slope of $\log \tilde{\beta}_p$ versus $\log \psi$, as determined from the two smallest measured scattering angles.

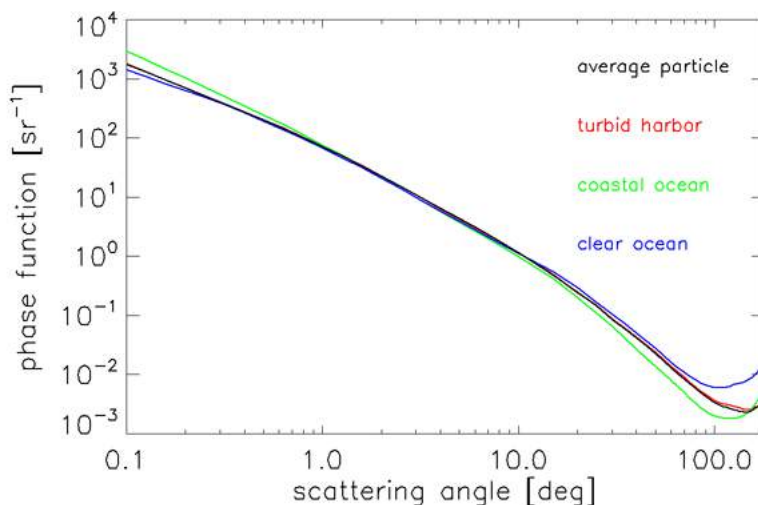


Figure 6.15: Log-log plots of the phase functions for the VSFs of Figs. 6.13 and 6.14, along with the average particle phase function.

This “Petzold average-particle” phase function has been widely used in radiative transfer calculations and is one of the standard phase functions available in HydroLight (Section 10.6). However, it must be remembered that this phase function is based on very limited data from turbid harbor waters at one wavelength and likely corresponds to a mixture of phytoplankton and mineral particles as might be found in harbor waters. This phase

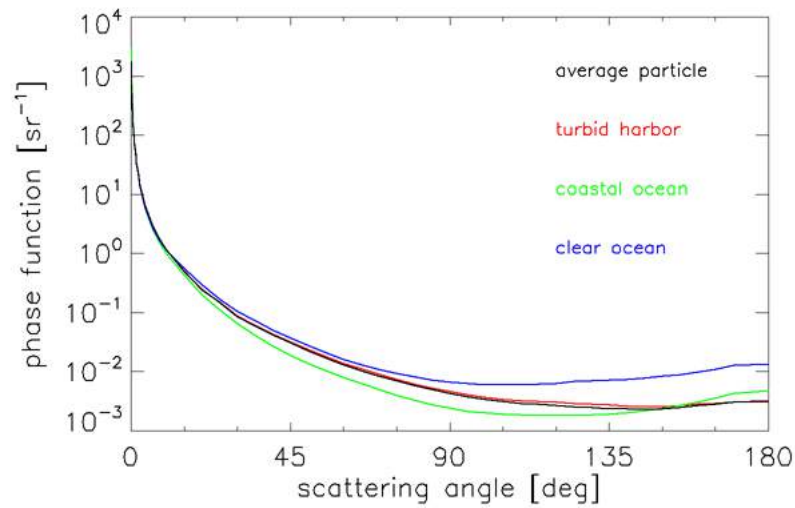


Figure 6.16: Log-linear plots of the phase functions for the VSFs of Figs. 6.13 and 6.14 , along with the average particle phase function.

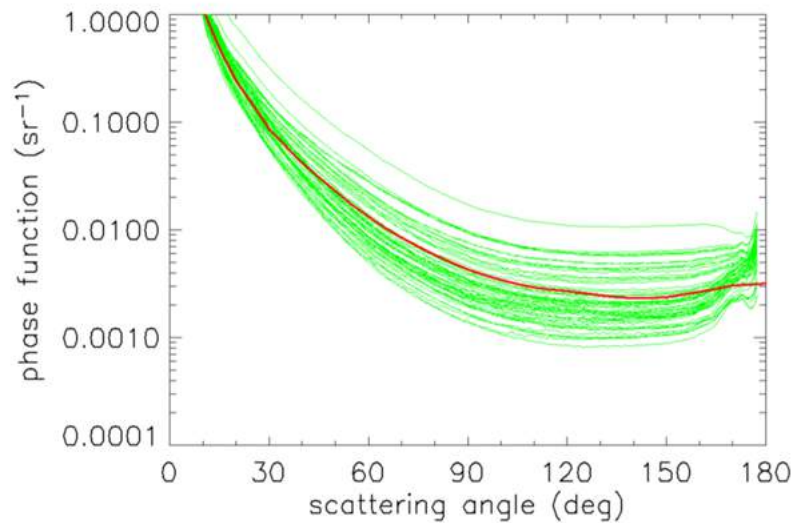


Figure 6.17: 62 phase functions measured in coastal waters at 530 nm (green curves). The red curve is the Petzold average-particle phase function of Figs. 6.15 and 6.16.

function thus corresponds closely to the turbid harbor phase function seen in Figs. 6.15 and 6.16.

Figure 6.17 compares the Petzold average-phase function with 62 phase functions measured in coastal waters at 530 nm using the more recently developed VSM (volume scattering meter) instrument of Lee and Lewis (2003). The Petzold average-particle phase function does indeed give a good *average* for these phase functions. However, it is important to note that there is an order-of-magnitude variability in the VSM phase functions at backscattering angles. The large variability in the measured phase functions of Fig. 6.17 will give corresponding variability in the remote-sensing reflectance, for example.

Thus, as is always the case with a simple model, the average-particle phase function may be satisfactory on average, but may be very wrong in a simulation of a particular water body. When attempting to model a particular water body, it is always best to use a VSF or phase function measured at the particular time and place being modeled, rather than relying on a “generic” phase function or analytic model. Examples of this are given in [Mobley et al. \(2002\)](#).

6.6 The Henyey-Greenstein Phase Function

In pre-computer days it was necessary to have a simple phase function for use in pencil-and-paper radiative transfer calculations. A number of such functions can be found in the literature, and some are still occasionally used in numerical calculations. The most important and long-lived of these phase functions is that of [Henyey and Greenstein \(1941\)](#):

$$\tilde{\beta}_{\text{HG}}(g, \psi) \triangleq \frac{1}{4\pi} \frac{1 - g^2}{(1 + g^2 - 2g \cos \psi)^{3/2}}, \quad (6.12)$$

The Henyey-Greenstein (HG) phase function was proposed (without a single word of justification) for studies of scattering by interstellar dust (Note: interstellar dust, not phytoplankton!). The parameter g can be adjusted to control the relative amounts of forward and backward scattering in $\tilde{\beta}_{\text{HG}}$: $g = 0$ corresponds to isotropic scattering, and $g \rightarrow 1$ gives highly peaked forward scattering. Note that $\tilde{\beta}_{\text{HG}}$ satisfies the normalization condition

$$2\pi \int_{-1}^1 \tilde{\beta}_{\text{HG}}(g, \psi) d \cos \psi = 1 \quad \text{for any } g.$$

The physical interpretation of g comes from noting that

$$2\pi \int_{-1}^1 \tilde{\beta}_{\text{HG}}(g, \psi) \cos \psi d \cos \psi = g.$$

Thus the Henyey-Greenstein parameter g is just the average of the cosine of the scattering angle for $\tilde{\beta}_{\text{HG}}$.

Equation (6.12) can be integrated over ψ from $\pi/2$ to π to obtain the backscatter fraction:

$$B_{\text{HG}} = \frac{1 - g}{2g} \left[\frac{1 + g}{\sqrt{1 + g^2}} - 1 \right].$$

Because of its mathematical simplicity, the HG phase function has been widely used in many fields, including oceanography. However, we can anticipate discrepancies between $\tilde{\beta}_{\text{HG}}$ and measured oceanic phase functions because oceanic particles are quite different in their physical properties (size, shape, composition) from interstellar dust.

The average of $\cos \psi$ for the Petzold average-particle phase function $\tilde{\beta}_{\text{p}}$ is 0.924. Using $g = 0.924$ in Eq. (6.12) thus gives the Henyey-Greenstein phase function corresponding to the particle phase function $\tilde{\beta}_{\text{p}}$, in the sense that each phase function then has the same average cosine. Figures 6.18 and 6.19 compare $\tilde{\beta}_{\text{HG}}(g = 0.924, \psi)$ and $\tilde{\beta}_{\text{p}}$. $\tilde{\beta}_{\text{HG}}$ is also shown for $g = 0.7$ and 0.99 . Note that the best-fit $\tilde{\beta}_{\text{HG}}$ differs noticeably from $\tilde{\beta}_{\text{p}}$ at scattering angles greater than $\psi \approx 150$ deg and less than 20 deg, and that $\tilde{\beta}_{\text{HG}}$ is much too small

at angles of less than a few degrees. The small-angle behavior of Eq. (6.12) is inherently incompatible with $\tilde{\beta}_p$ because $\tilde{\beta}_{\text{HG}}$ always levels off as $\psi \rightarrow 0$, whereas $\tilde{\beta}_p$ continues to rise. Even for $g = 0.99$, $\tilde{\beta}_{\text{HG}}$ is nearly constant for $\psi < 0.5$ deg. For $g = 0.924$, $\tilde{\beta}_{\text{HG}} = 0.0170$, compared to the backscatter fraction of 0.0183 for the Petzold phase function.

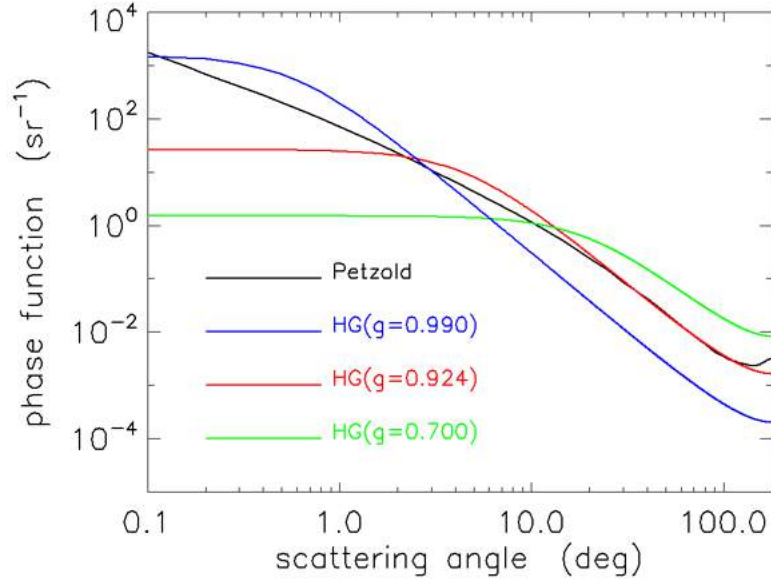


Figure 6.18: Log-log comparison of the Petzold average-particle phase function $\tilde{\beta}_p$ (black line) with $\tilde{\beta}_{\text{HG}}(g, \psi)$ of Eq. (6.12) for three values of g .

Equation (6.12) depends on only one free parameter, the average cosine g , so this equation is often called the One-Term Henyey-Greenstein (OTHG) phase function. Because of the poor fits of the OTHG phase function to measurements at small and large scattering angles, a linear combination of Henyey-Greenstein phase functions is sometimes used to improve the fit at small and large angles. The so-called two-term Henyey-Greenstein (TTHG) phase function is

$$\tilde{\beta}_{\text{TTHG}}(\gamma, g_1, g_2, \psi) \triangleq \gamma \tilde{\beta}_{\text{HG}}(g_1, \psi) + (1 - \gamma) \tilde{\beta}_{\text{HG}}(g_2, \psi). \quad (6.13)$$

Enhanced small-angle scattering is obtained by choosing g_1 near one, and enhanced backscatter is obtained by making g_2 negative; γ is a weighting factor between zero and one. Kattawar (1975) shows how to determine best-fit values of γ , g_1 , and g_2 for a given phase function. Reasonable fits can be obtained with the TTHG function for phase functions that are not highly peaked, for example atmospheric haze phase functions (see Kattawar's paper for example fits). However, the fit of $\tilde{\beta}_{\text{TTHG}}$ to highly peaked oceanic phase functions such as $\tilde{\beta}_p$ always remains unsatisfactory at very small angles, for the reason already noted.

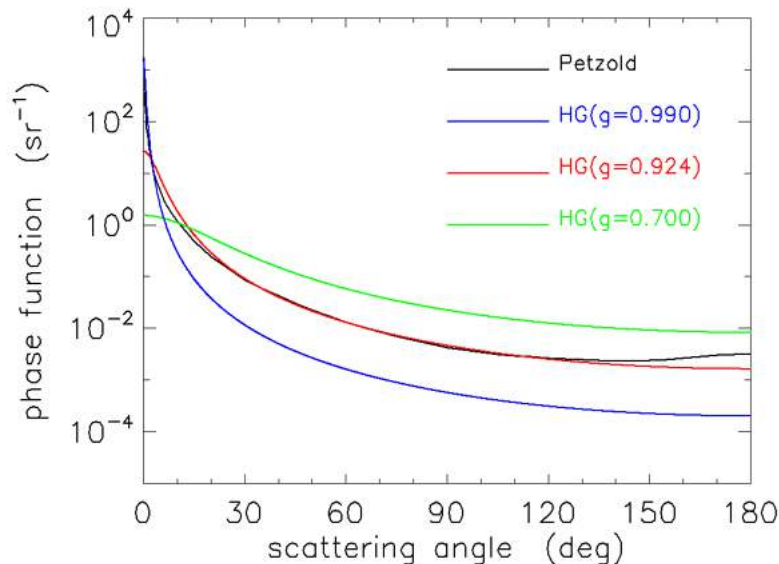


Figure 6.19: Log-linear comparison of the Petzold average-particle phase function $\tilde{\beta}_p$ (black line) with $\tilde{\beta}_{\text{HG}}(g, \psi)$ of Eq. (6.12) for three values of g .

6.6.1 The Reynolds-McCormick Phase Function

The Reynolds-McCormick phase function (Reynolds and McCormick, 1980) is a generalization of the Henyey-Greenstein phase function:

$$\tilde{\beta}_{\text{RM}}(g, \psi) \triangleq \frac{1}{\pi} \frac{\alpha g (1 - g^2)^{2\alpha}}{[(1 + g)^{2\alpha} - (1 - g)^{2\alpha}] [1 + g^2 - 2g \cos \psi]^{\alpha+1}}, \quad (6.14)$$

The RM phase function reduces to the HG phase function when $\alpha = 0.5$. The RM phase function has two parameters, α and g , which can be adjusted to give a better fit to measured phase functions than can the one-parameter OTHG. However, g no longer equals the average cosine of the scattering angle, which is now a function of both g and α (Reynolds and McCormick, 1980, Eqs. 16 and 17). Harmel et al. (2021) defined two-term Fourier-Forand (TTF) and Reynolds-McCormick (TTRM) phase functions in analogy to the two-term Henyey-Greenstein phase function (TTHG) of Eq. (6.13). Comparisons of the one- and two-term FF and RM phase functions with measurements showed that the TTRM phase function gave fits to data that in most cases were better than the FF or TTF phase functions. This is not surprising because the TTRM phase function has five free parameters, which were determined by a non-linear optimization. The disadvantage of such a phase function is that there are no simple connections between these fitting parameters and easily measured or understood quantities such as the average cosine of the scattering angle, the backscatter fraction, the particle index of refraction, or the slope of the particle size distribution. This makes it difficult to use this phase function in bio-geo-optical IOP models that define the scattering properties in terms of chlorophyll or mineral particle concentrations and physical properties such as index of refraction or size distribution.

6.7 The Fournier-Forand Phase Function

A much more realistic—and mathematically much more complicated—analytic phase function was developed by Fournier and Forand (1994). They derived an approximate analytic form for the phase function of an ensemble of particles that have a power-law (Junge-type) particle size distribution (PSD), with each particle scattering according to the anomalous diffraction approximation to the exact Mie theory. [For a power-law cumulative size distribution, the number $N(r)$ of particles per unit volume with size greater than r (volume-equivalent spherical radius) is proportional to $r^{-\mu}$, so that $-\mu$ is the slope of the size distribution when $\log N(r)$ is plotted versus $\log r$. Oceanic particle size distributions typically have μ values between 3 and 5. See Section 8.12 for a detailed discussion of particle size distributions.]

In its latest form (Fournier and Jonasz, 1999) this phase function is given by

$$\begin{aligned} \tilde{\beta}_{\text{FF}}(\psi) = & \frac{1}{4\pi(1-\delta)^2\delta^\nu} \left[\nu(1-\delta) - (1-\delta^\nu) + [\delta(1-\delta^\nu) - \nu(1-\delta)] \sin^{-2}\left(\frac{\psi}{2}\right) \right] \\ & + \frac{1-\delta_{180}^\nu}{16\pi(\delta_{180}-1)\delta_{180}^\nu} (3\cos^2\psi - 1), \end{aligned} \quad (6.15)$$

where

$$\nu = \frac{3-\mu}{2} \quad \text{and} \quad \delta = \frac{4}{3(n-1)^2} \sin^2\left(\frac{\psi}{2}\right). \quad (6.16)$$

Here n is the real index of refraction of the particles relative to the surrounding medium, μ is the slope parameter of the hyperbolic distribution, and δ_{180} is δ evaluated at $\psi = 180^\circ$. Equation (6.15) can be integrated to obtain the backscatter fraction,

$$B = \frac{b_b}{b} = 1 - \frac{1 - \delta_{90}^{\nu+1} - 0.5(1 - \delta_{90}^\nu)}{(1 - \delta_{90})\delta_{90}^\nu}, \quad (6.17)$$

where δ_{90} is δ evaluated at $\psi = 90^\circ$. Although Eq. (6.16) uses only the real part of the index of refraction, the addition of moderate amounts of absorption does not significantly change the shape of the phase functions generated by Eq. (6.15).

Figures 6.20 and 6.21 show Fournier-Forand phase functions for a wide range of B values; each curve is labeled by its backscatter fraction. Most oceanic particles have backscatter fractions between $B = 0.001$ (e.g., very large phytoplankton) and 0.1 (e.g., very small mineral particles). Table 6.1 shows the n and μ values used in Eqs. (6.15) and (6.16) to obtain these curves.

Fournier (2011) cautions that the Fournier-Forand equations should be used only for parameter values in the ranges $3 \leq \mu \leq 5$ and $1 < n \leq 1.33$. The values seen in Table 6.1 are within those ranges. In the same paper, Fournier derives an exact but rather messy formula for the mean cosine g of the Fournier-Forand phase function. Fortunately, the exact equation can be approximated as

$$g \approx 1 - \frac{23(n-1)^{5/2}}{1 - 7.5(n-1)^{5/2}}. \quad (6.18)$$

Given the value of g , the ν and δ parameters of Eq. (6.15), which are needed to generate a phase function, are given at the same level of approximation by

$$\nu = -3 \left(\frac{1-g}{23 - 7.5(1-g)} \right)^{2/5} \quad (6.19)$$

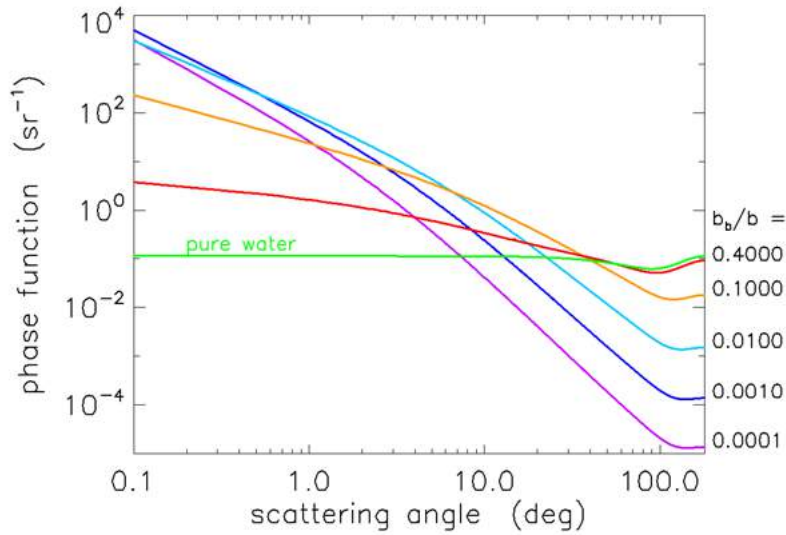


Figure 6.20: Log-log plot of Fournier-Forand phase functions for selected backscatter fractions B . The green curve is the pure sea water phase function, which has $B = 0.5$.

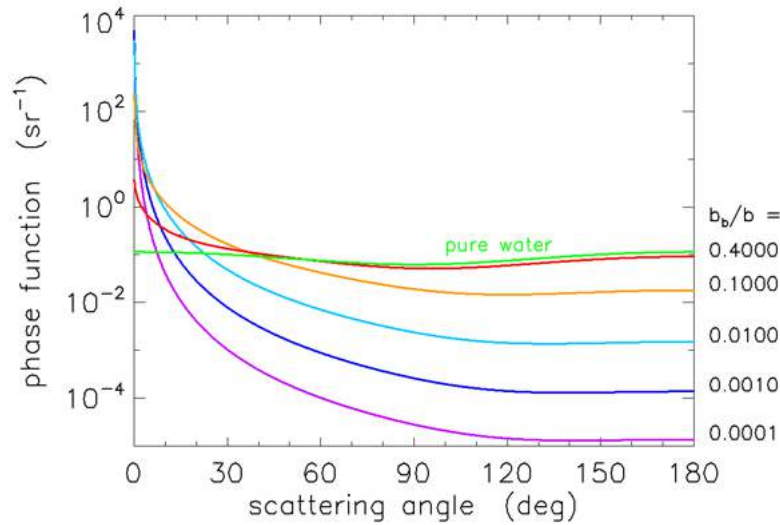


Figure 6.21: Log-linear plot of Fournier-Forand phase functions for selected backscatter fractions B . The green curve is the pure sea water phase function, which has $B = 0.5$.

$$\delta = \frac{16 \sin^2(\psi/2)}{3} \left(\frac{1-g}{23-7.5(1-g)} \right)^{-4/5}. \quad (6.20)$$

Figures 6.22 and 6.23 show how well the Fournier-Forand phase function can reproduce the shapes of measured phase functions for the types of particles found in ocean water. In those figures, the blue curves are three phase functions measured by the Volume Scattering Meter (VSM) of Lee and Lewis (2003), and the green curve is the Petzold average-particle phase function. The red curves show the Fournier-Forand phase functions with the same backscatter fractions as the measured phase functions. Each pair of measured and FF

index of refraction n	Junge slope μ	backscatter fraction B
1.021	3.0742	0.0001
1.040	3.2010	0.001
1.08	3.483	0.01
1.175	4.065	0.1
1.15	4.874	0.4

Table 6.1: The n and μ values used in Eqs. (6.15) and (6.16) to generate the Fournier-Forand phase functions of Figs. 6.20 and 6.21.

phase functions is labeled by the backscatter fraction B in Fig. 6.23.

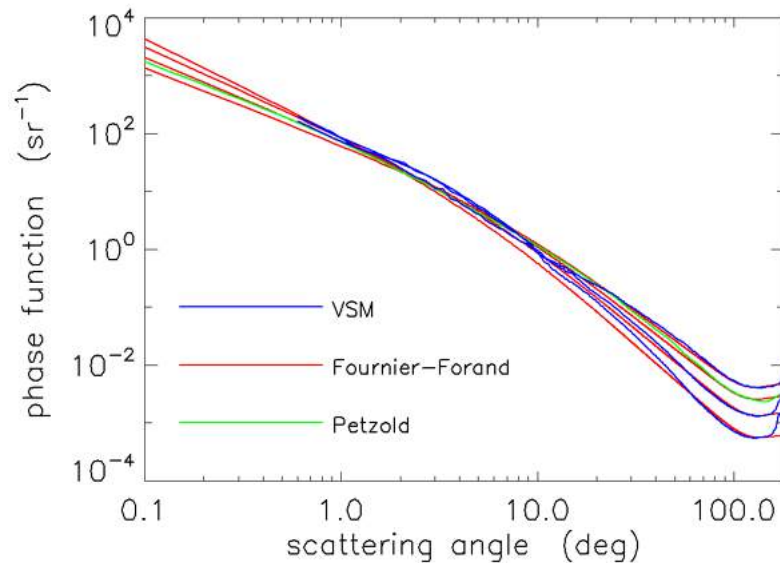


Figure 6.22: Log-log plot of 4 measured phase functions (blue and green curves), and the Fournier-Forand phase functions with the same backscatter fractions (red curves).

As seen in Figs. 6.22 and 6.23, the Fournier-Forand analytical model does a much better job of reproducing the shapes of ocean phase functions than does the historical Henyey-Greenstein phase function defined in the previous section, especially at very small and large scattering angles. There is considerable discrepancy between the measured and modeled VSM data at wavelengths near 180 deg, but it is not known how much of this may be due to instrumental error. In any case, Fournier-Forand phase functions are adequate for most radiative transfer calculations, and they are now commonly used in numerical simulations of oceanic light fields (e.g., in HydroLight). The light fields resulting from the use of Fournier-Forand and several other measured and modeled phase functions in numerical simulations are compared in Mobley et al. (2002).

The relation between n , μ , and B is not unique. Figure 6.24 shows selected backscatter

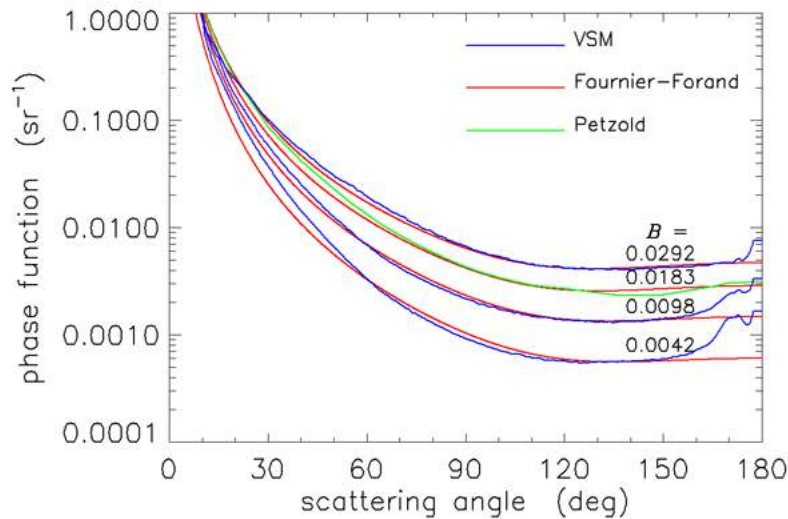


Figure 6.23: Log-linear plot of 4 measured phase functions (blue and green curves), and the Fournier-Forand phase functions with the same backscatter fractions (red curves).

values given by Eq. (6.17) as a function of n and μ . The red dot on the $B = 0.0183$ curve shows the values of $n = 1.10$ and $\mu = 3.5835$ for which Eq. (6.15) gives the best fit to the Petzold average-particle phase function, as seen in Figs. 6.22 and 6.23. However, any pair of n and μ values lying on a curve of constant B in Fig. 6.24 will generate a slightly different phase function having the given B value. The phase functions corresponding to the two red squares and the red dot on the $B = 0.0183$ contour give the phase functions shown in Figs. 6.25 and 6.26. These functions have somewhat different shapes, but any is still a reasonable approximation to the Petzold average particle phase function shown in red in those figures. Thus the exact n and μ values used to generate a Fournier-Forand phase function with a given backscattering fraction are not critical for most applications.

Given that the exact (n, μ) values used to generate a phase function are not critical, Mobley et al. (2002) proposed a way to generate Fournier-Forand phase functions given the backscattering fraction B . The green curve in Fig. 6.24 shows the line

$$n = 1.01 + 0.1542(\mu - 3). \quad (6.21)$$

This function connects the point at $(n, \mu) = (1.01, 3.0)$ with the best-fit point for the Petzold average-particle phase function (the red dot in Fig. 6.24). This function is simply a crude conceptual approximation of (n, μ) values progressing from low-index, low-backscatter biological particles to high-index, high-backscatter mineral particles. Given a desired backscatter fraction, the intersection of the green curve and the backscatter curve as seen in Fig. 6.24 determine the n and μ values to be used to generate a phase function with the given B value. (This is how the three blue diamonds in the figure were obtained for given B values.) Although no measured data were used to construct the $n(\mu)$ function, it works well in practice and has been used in many studies. This technique is used in HydroLight to generate phase functions when the backscatter value is an input.

Thus there are three ways to generate a Fournier-Forand phase function in practice: (1) choose values of n and μ for use in Eq. (6.16); (2) use a desired mean cosine g in Eqs. (6.19)

and (6.20); and (3) use a desired backscatter fraction B to obtain the the corresponding n and μ as in Fig. 6.24. The Fournier-Forand phase function clearly does a much better job of reproducing measured oceanic phase functions than does the Henyey-Greenstein phase function of the previous section (recall Figs. 6.18 and 6.19). For this reason, Fournier-Forand phase functions have replaced earlier approximations like Henyey-Greenstein in most numerical radiative transfer calculations.

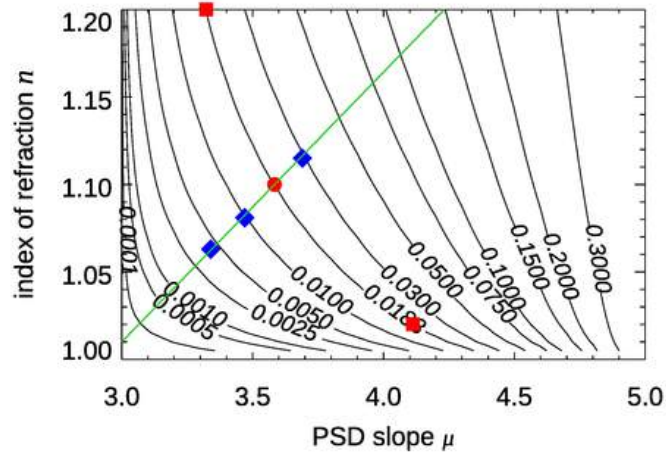


Figure 6.24: The Fournier-Forand backscattering fraction as a function of index of refraction and Junge slope. The red dot shows the value of n and μ that gives the best fit to the Petzold average-particle phase function. The red dot and the two red squares show the values used to generate the three phase functions with $B = 0.0183$ shown in Figs. 6.25 and 6.26. The red dot and blue diamonds give the n and μ values used to generate the four phase functions of Figs. 6.22 and 6.23. The green line is Eq. (6.21).

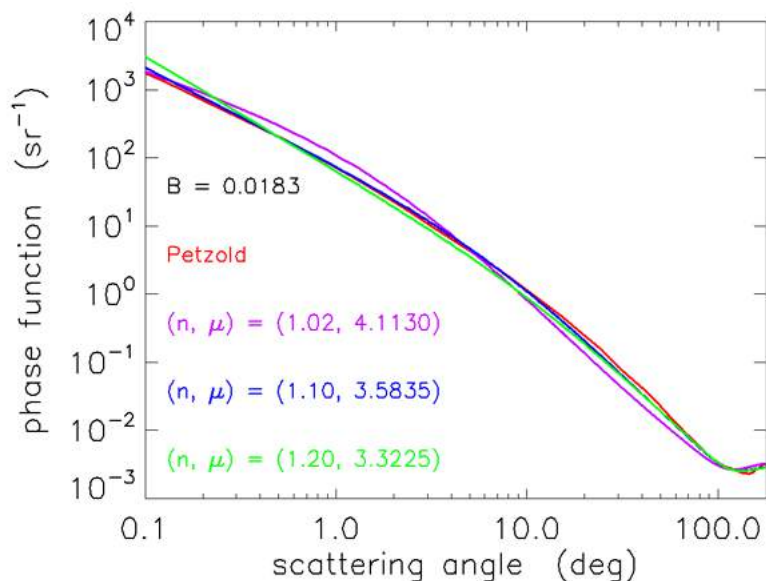


Figure 6.25: Log-log plots of Fournier-Forand phase functions having the same backscattering fraction $B = 0.0183$ as the Petzold average-particle phase function, but generated by the three n, μ pairs shown by the red symbols in Fig. 6.24. The Petzold average-particle phase function is shown in red.

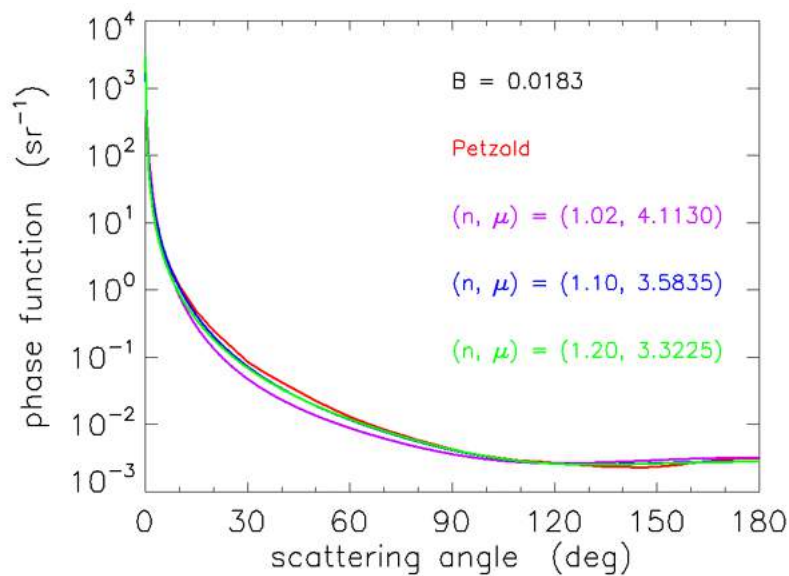


Figure 6.26: Log-linear plots of Fournier-Forand phase functions having the same backscatter fraction $B = 0.0183$ but generated by the three n, μ pairs shown by the red symbols in Fig. 6.24. The Petzold average-particle phase function is shown in red.

6.8 Effect of Particle Shape on Optical Properties

[Emmanuel Boss and Curtis Mobley contributed to this section.]

A sphere is not likely to be a good representative of the shape of the *average* aquatic particle for two main reasons: (1) the majority of marine particles are not spherical, and (2) of all the convex shapes a sphere is rather an extreme shape: for a given particle volume it has the smallest surface-area-to-volume ratio. Only a limited number of studies have examined the IOPs of non-spherical marine particles, but results indicate a strong dependence of optical properties, in particular scattering, on shape. See, for example, Kirk (1976), Voss and Fry (1984), Volten et al. (1998), Gordon and Du (2001), MacCallum et al. (2004), Quirantes and Bernard (2004), Quirantes and Bernard (2006), and Gordon (2006).

Methodical evaluation of shape effects on IOPs are presented in two non-peer-reviewed publications, Aas (1984) and Herring (2002); a short book chapter, Jonasz (1987); and in a review by Clavano et al. (2007). Considerably more research has been done on aerosol particles, which are often irregularly shaped mineral particles with high indices of refraction relative to air. Many of those results are pertinent to oceanic particles.

It has been shown in many studies that non-sphericity has its greatest effect in backscatter directions. It is these directions that contribute to the water-leaving radiance, which is the foundation of ocean-color remote sensing. Figure 6.27 shows a comparison between measured phase functions and those computed by Mie theory for a polydispersion (particle size distribution) of “equivalent” spheres of the same volume. The feature to note in the phase functions is that Mie theory gives values that differ by as much as a factor of three in backscatter directions.

Figure 6.28 shows phase functions for non-spherical particles and for spheres with the same surface areas, as computed by the T-matrix technique. The left panel compares phase functions for prolate (“football” shaped) and oblate (“pancake” shaped) spheroids with aspect ratios ranging from 1.2 to 2.4, averaged over all orientations. The particle size distributions corresponded to dust-like tropospheric aerosols with a complex index of refraction of $1.53 + 0.0085i$. The right panel shows the a similar comparison but for effective particle size parameters of 14, 16, and 18; the non-spheres are a mixture of prolate and oblate spheroids. This paper by Mishchenko et al. (1997) performed many such comparisons for other wavelengths and effective particle sizes. Some of the conclusions from their study was that spheres and non-spheres have similar phase functions only in the scattering angle range from $\psi = 0$ to 10 deg. For $90 \text{ deg} \lesssim \psi \lesssim 150 \text{ deg}$, the phase functions for the non-spheres are much greater than for spheres, and for $150 \text{ deg} \lesssim \psi \lesssim 180 \text{ deg}$, the sphere phase functions are much greater than the non-spheres. However, for the absorption and scattering cross sections, the sphere versus non-sphere values are often within 15% of each other for small size parameters, and approached the same values at large sizes (the geometric optics regime). Although these particular simulations are for atmospheric dust in air, and not for particles in water, the general conclusions are the same: non-sphericity can make factor-of-two differences in phase functions at backscatter directions, which are critical for ocean color remote sensing. Asano and Yamamoto (1975) present similar results for aspect ratios from 1 (a sphere) to 5.

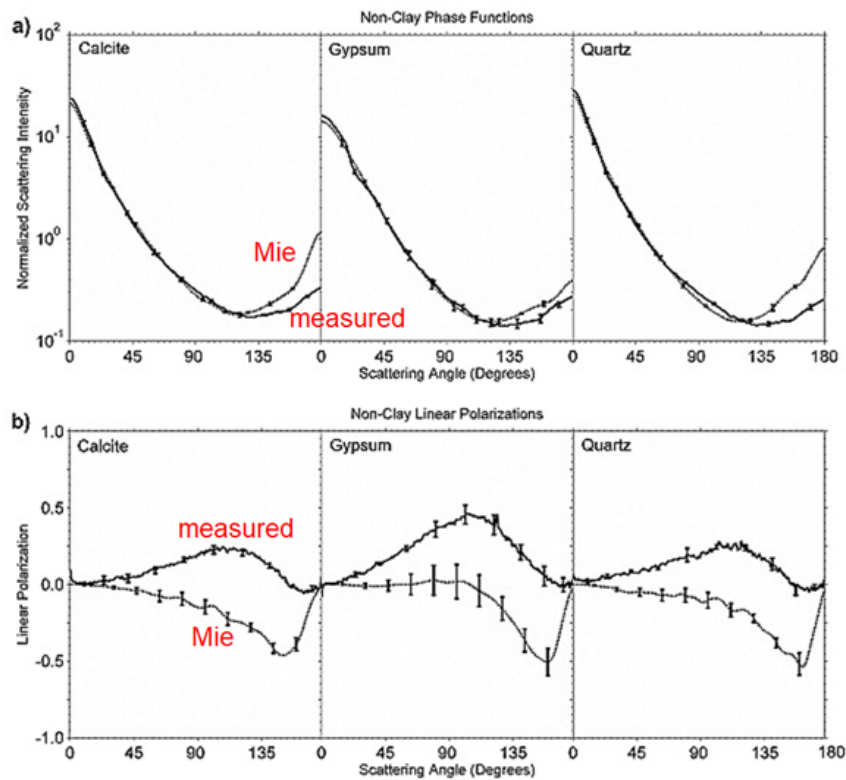


Figure 6.27: Measured and Mie-computed phase functions (top) and degree of linear polarization (bottom) for three mineral particles in air. Annotated Fig. 2 from [Curtis et al. \(2008\)](#).

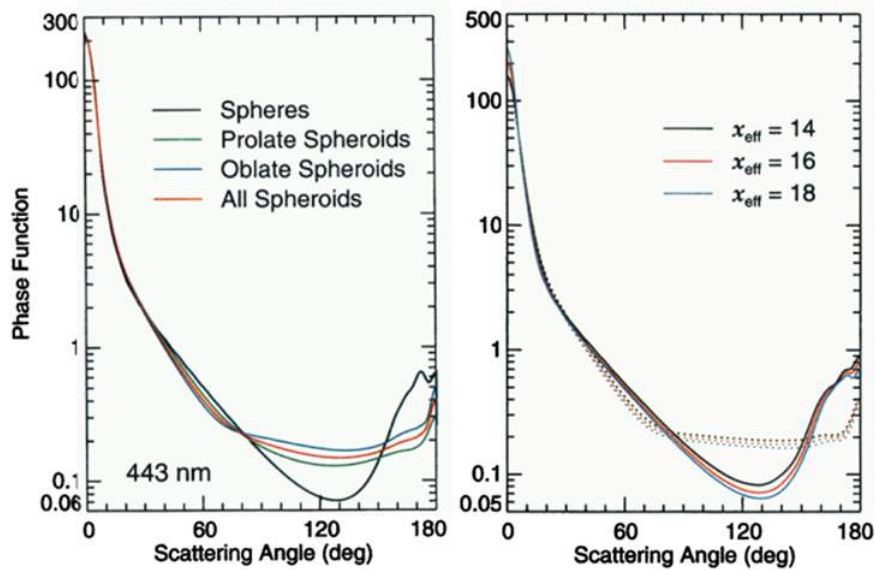


Figure 6.28: Computed phase functions for prolate and oblate spheroids versus area-equivalent spheres. The left panel is for a polydispersion of particles corresponding to tropospheric aerosols. The right panel is for three effective size parameters; spheres are solid curves, non-spheres are dotted. Figures extracted from Plates 3 and 5 of [Mishchenko et al. \(1997\)](#), which shows many more results for other wavelengths and effective sizes.

Figure 6.29 shows orientation-averaged extinction cross sections ($\overline{C_{\text{ext}}}$) normalized by the particle area computed using the radius r_v of the volume-equivalent sphere, as a function of the size parameter computed using r_v . For small particles, the shape has minimal effect, but for size parameters of order 10, there is a factor-of-two difference in the normalized extinction cross sections.

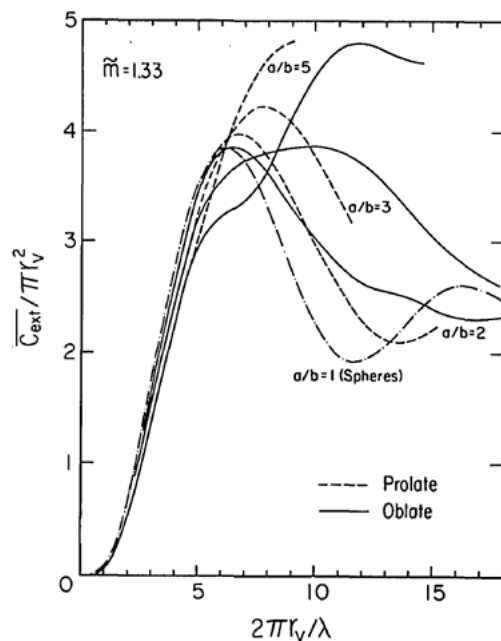


Figure 6.29: Normalized extinction cross sections for various prolate and oblate spheroids compared to volume-equivalent spheres. Redrawn from Fig. 2 of [Asano and Sato \(1980\)](#).

One of the applications of in-water optical measurements is the inversion of measured volume scattering functions (VSFs) to obtain particle size distributions. The underlying physics is that diffraction accounts for much of the light scattered through small angles, $\psi \lesssim 10$ or 20 deg. To first order, the amount of diffracted power is proportional to the cross sectional area of the particle “seen” by the incident beam and is fairly insensitive to the particle composition or index of refraction. Small particles scatter (diffract) more power into large scattering angles; large particles diffract more power into very small angles. Thus a measurement of the VSF between, $\psi \approx 0.1$ deg and 10 or 20 deg can be inverted to get an estimate of the particle *area* size distribution of the scattering particles. Converting the area PSD to a volume PSD requires an assumption about the shape of the particles in the form of a median diameter for the particles in each region of the area PSD. For details see [Agrawal and Pottsmith \(2000\)](#) or the November 1991 issue of *Applied Optics* (Vol. 30, No. 33), which was a special issue on optical particle sizing.

The bottom panel of Fig. 6.30 shows a PSD retrieved from a culture of the dinoflagellate *Ceratium longipes* using the LISST-100 instrument (Sequoia Scientific, Inc.). From just the PSD, it would be reasonable to conclude that the culture contained a mixture of small, $\sim 10 \mu\text{m}$; medium, $\sim 50 \mu\text{m}$; and large, $\sim 150 \mu\text{m}$ sized particles. The top panel of the figure shows a photograph of an actual *Ceratium longipes*. It is seen that there is a central body with long, narrow spines. The LISST was “seeing” the narrow spines, the central body, and the overall size of the phytoplankton. This is actually a quite remarkable retrieval and shows the power of the instrument for resolving objects of various sizes. This figure should be sufficient to convince you of the futility of trying to define an “equivalent spherical particle” for use in Mie theory when a particle has such a complex shape.

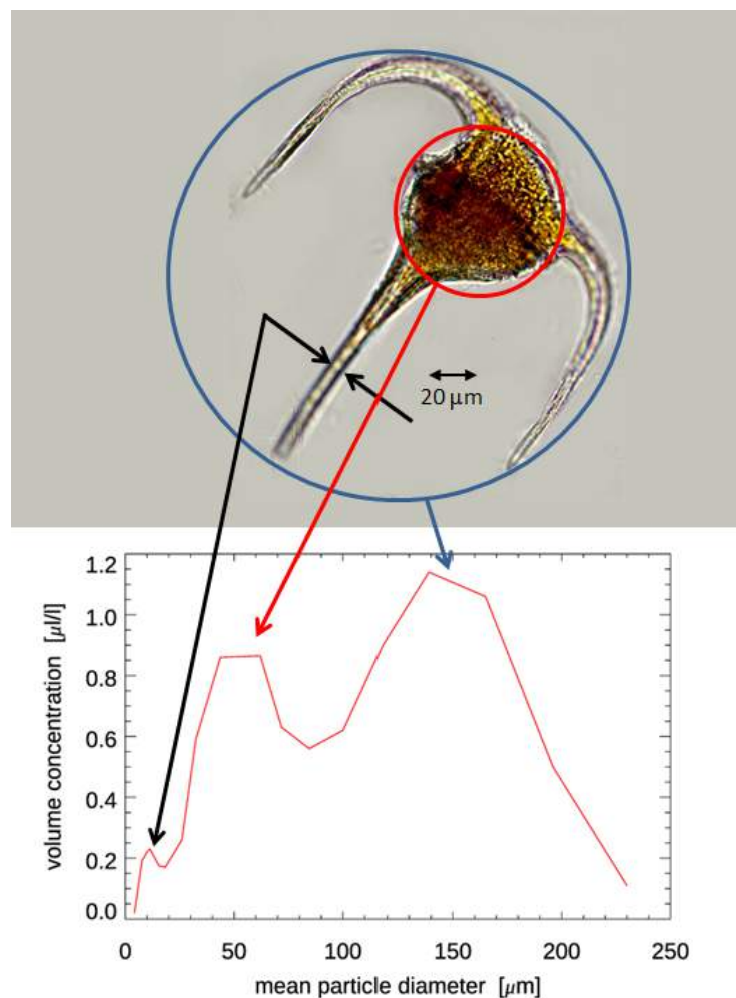


Figure 6.30: Top panel: Photograph of a *Ceratium longipes* dinoflagellate. Bottom panel: LISST-retrieved particle size distribution obtained on a culture of *Ceratium longipes*. Top image modified from [PhycoKey](#). LISST data courtesy of Lee Karp-Boss.

The preceding figures are sufficient to make the point that non-spherical particles can have much different scattering properties than spheres, and that using a volume- or area-equivalent sphere in Mie calculations can give results much different from the actual non-

spherical particles. It is difficult to generalize about the optical properties of non-spherical particles because their shapes can be so different: spheroids, long-chain diatoms, plate-like mineral liths, fractal aggregates, etc. In addition to shape effects defined by the boundary of the particle, phytoplankton have shape effects due to their non-homogeneous internal structure as defined by chloroplasts, gas vacuoles, and cell walls.

However, a few results can be summarized as follows:

- For particles much smaller than the wavelength of light, the inherent optical properties of non-spherical particles are similar to those of spheres with the same volume.
- For randomly oriented convex particles much larger than the wavelength of light (the geometric optics limit) the absorption, scattering and attenuation are similar to those of spheres with a similar average cross-sectional areas.
- In general, however, about the only safe thing to say is that non-spherical particles often have IOPs, especially scattering phase functions, that differ by a factor of two, and sometimes by a factor of ten, from what would be obtained by Mie theory using equivalent-volume or equivalent-area spheres.

There exist numerical techniques for computation of the optical properties of non-spherical particles. The two most commonly used are the T-matrix method and the Discrete Dipole Approximation (DDA). These techniques are extremely mathematical and computationally intensive, and can require extensive effort just to define the problem. The DDA, for example, divides the particle into many very small volume elements, each of which then scatters as a dipole when the incident electromagnetic radiation induces a dipole moment in the volume element. However, you must first define the size and shape of the particle. Publicly available codes for these techniques may come with a gallery of pre-defined shapes such as spheroids, disks, or cylinders, but for more irregular shapes, you must first “build in” the particle shape, which can be a laborious process. In addition, the computations may be limited in the range of size parameters that can be computed. The best source for finding such codes, in a variety of computer languages, is [SCATTERLIB](#).

Inelastic Scattering

This chapter discusses the last of the three processes—absorption, elastic scattering, and inelastic scattering—that can occur when light interacts with matter. There are various types of inelastic scattering, which are due to different physical processes. The opening section of this book mentioned Compton scattering, which is inelastic scattering of X-rays by electrons. The next section discusses Raman scattering, which is a type of inelastic scattering of light by molecules. The following sections then discuss fluorescence by chlorophyll and CDOM. The time scales of both elastic and Raman scattering are on the order of 10^{-13} s and are properly termed “scattering.” As will be seen, fluorescence results from the absorption of a photon followed by the emission of a new photon, typically on a time scale of 10^{-12} to 10^{-9} s, which is quite a long time by atomic standards. However, the time scales for each of these processes is very short on the scale at which most radiometric measurements are made, and each process can be treated by the same general mathematical formalism in the time-independent radiative transfer equation. Therefore Raman scattering and fluorescence will both be called scattering and regarded as essentially instantaneous for the purposes of this chapter.

In all of these cases, we are interested in light at some wavelength λ , and the inelastic process takes radiant energy from wavelengths $\lambda' < \lambda$ and “creates” light at wavelength λ . This is modeled in the radiative transfer equation by a source term that must be formulated for each particular process. Bioluminescence is a true source of light at λ , in which the light is created from non-thermal chemical reactions. Bioluminescence is a true emission process and has nothing whatsoever to do with scattering. However, the creation of new light by bioluminescence can be modeled by a mathematical formalism similar to that for inelastic scattering for the purpose of formulating a bioluminescence source term for inclusion in the radiative transfer equation.

7.1 General Formalism for Inelastic Scattering

The general formalism for inclusion of inelastic scattering effects into time-independent radiative transfer calculations (e.g. in HydroLight) involves the following:

- **A scattering coefficient** $b(\lambda')$ with units of m^{-1} , which quantifies how strongly the incident or excitation light at wavelength λ' interacts with the scattering substance. That is, $b(\lambda')$ tells how much of the irradiance at the excitation wavelength λ' scatters into all emission wavelengths $\lambda > \lambda'$, per unit of distance traveled by the excitation irradiance.
- **A wavelength redistribution function** $f(\lambda', \lambda)$ with units of nm^{-1} , which relates the excitation and emission wavelengths. That is, $f(\lambda', \lambda)$ specifies what wavelengths λ receive the scattered spectral irradiance for a given excitation wavelength λ' per unit of emission wavelength, or, conversely, what wavelengths λ' excite a given emission wavelength λ . The term “wavelength redistribution function” is non-standard, but descriptive, so that is what was used in *Light and Water Mobley (1994)* and again here.
- **A scattering phase function** $\tilde{\beta}(\psi)$ with units of sr^{-1} , which gives the angular distribution of the scattered or emitted light relative to the direction of the incident or exciting light.

The next sections develop the exact forms of these three quantities as needed for Raman scattering and chlorophyll and CDOM fluorescence. One each of these terms has been defined for a particular inelastic scattering process, their product gives the corresponding *volume inelastic scattering function*, $\beta_{\text{I}}(\hat{\xi}' \rightarrow \hat{\xi}; \lambda' \rightarrow \lambda)$, where $\hat{\xi}'$ and $\hat{\xi}$ represent the incident and final directions of the light:

$$\begin{aligned} \beta_{\text{I}}(\hat{\xi}' \rightarrow \hat{\xi}; \lambda' \rightarrow \lambda) &= \beta_{\text{I}}(\theta', \phi' \rightarrow \theta, \phi; \lambda' \rightarrow \lambda) \\ &= b(\lambda') f(\lambda', \lambda) \tilde{\beta}(\psi) \quad (\text{m}^{-1} \text{ nm}^{-1} \text{ sr}^{-1}). \end{aligned} \quad (7.1)$$

The scattering angle ψ is given by $\psi = \cos^{-1}(\hat{\xi}' \cdot \hat{\xi})$, which can be computed from $\theta', \phi', \theta, \phi$ via Eq. (1.9). The subscript I on β_{I} is used to distinguish the VSF for inelastic scattering, $\beta_{\text{I}}(\theta', \phi' \rightarrow \theta, \phi; \lambda' \rightarrow \lambda)$, from the VSF for elastic scattering, $\beta(\theta', \phi' \rightarrow \theta, \phi)$. The I will be replaced below by R, C, Y, or B for Raman scatter, chlorophyll fluorescence, yellow-matter (CDOM) fluorescence, or bioluminescence.

The scalar radiative transfer equation (SRTE) as developed in Sections 9.1 and 9.4 is

$$\begin{aligned} \cos \theta \frac{dL(z, \theta, \phi, \lambda)}{dz} &= -c(z, \lambda)L(z, \theta, \phi, \lambda) \\ &+ \int_0^{2\pi} \int_0^\pi \beta(z; \theta', \phi' \rightarrow \theta, \phi; \lambda)L(z, \theta', \phi', \lambda) \sin \theta' d\theta' d\phi' \\ &+ \int_0^\lambda \int_0^{2\pi} \int_0^\pi \beta_{\text{I}}(\theta', \phi' \rightarrow \theta, \phi; \lambda' \rightarrow \lambda)L(z, \theta', \phi', \lambda') \sin \theta' d\theta' d\phi' d\lambda'. \end{aligned} \quad (7.2)$$

Inelastic scattering is then incorporated into unpolarized radiative transfer calculations by using the inelastic VSF of Eq. (7.1) in the last integral of this equation. Note that this formidable equation cannot be solved at just the emission wavelength λ of interest; it must be solved at all wavelengths $\lambda' < \lambda$ that contribute inelastically scattered radiance to the radiance at λ .

7.2 Raman Scattering

First of all, the pronunciation is RAmAn, not raMAN. Upon learning of Compton’s discovery of inelastic scattering of X-rays by electrons (for which Compton received the 1928 Nobel Prize in Physics), C. V. Raman¹ thought there should be a similar effect at visible wavelengths and set out to find it. He succeeded and subsequently received the 1930 Nobel Prize in Physics for the discovery and explanation of what is now called Raman scattering or the Raman effect.

The quantum-level physics of Raman scattering is quite complicated, but it can be conceptualized as following. Incident light can excite a molecule in its ground state to a higher “virtual” energy level, which then immediately decays back to a lower level accompanied by the emission of light. If the decay returns the molecule to its initial state, the scattering is elastic and is called Rayleigh scattering. If the decay is to a molecular vibrational level above the ground state, then then emitted light has a longer wavelength (lower energy) than the incident light: this is Raman scattering. (See the Physics of Absorption Section 5.3 for a description of electronic, vibrational, and rotational energy levels.) Raman scattering is widely used in chemistry as a way to study the vibrational energy levels of molecules; this application is often called Raman spectroscopy. In those applications, the incident light usually comes from a laser. In the ocean, the molecule of interest is water, and the exciting light can come either from the Sun or a laser.

The wavelength shift for Raman scatter by water is exceptionally large, corresponding to a wavenumber (1/wavelength) shift of about 3400 cm^{-1} , which at visible wavelengths is many tens to more than a hundred nanometers.

In accordance with the general formulation for inelastic effects of the previous Section 7.1, the quantities needed to compute Raman scattering contributions to the radiance are

- the Raman scattering coefficient $b_{\text{R}}(\lambda')$, with units of m^{-1}
- the Raman wavelength redistribution function $f_{\text{R}}(\lambda', \lambda)$, with units of nm^{-1}
- the Raman scattering phase function beta $\tilde{\beta}_{\text{R}}(\psi)$, with units of sr^{-1}

The next sections discuss each of these quantities in turn.

7.2.1 The Raman Scattering Coefficient

The Raman scattering coefficient $b_{\text{R}}(\lambda')$ tells how much of the irradiance at the excitation wavelength λ' scatters into all emission wavelengths $\lambda > \lambda'$, per unit of distance traveled by the excitation irradiance. The most recently published values of $b_{\text{R}}(488\text{ nm})$ for water are $(2.7 \pm 0.2) \times 10^{-4}\text{ m}^{-1}$ (Bartlett et al., 1998) and $2.4 \times 10^{-4}\text{ m}^{-1}$ (Desiderio, 2000). (The current version 6.0 of the HydroLight radiative transfer model (Section 10.6) uses $b_{\text{R}}(488\text{ nm}) = 2.6 \times 10^{-4}\text{ m}^{-1}$ as the default value.)

Various values for the wavelength dependence of b_{R} can be found in the literature. Bartlett et al. (1998) reviewed the wavelength dependence of the Raman scattering coefficient in detail and found, based on their measurements, a wavelength dependence of

¹Raman received his bachelor’s degree in physics at 16, graduating at the head of his class, and he published his first paper at 18. The Raman family produced a number of scientists, including Raman’s nephew Subrahmanyan Chandrasekhar, who received the physics Nobel in 1983. The family seems to have had good genes for science.

$\lambda^{-4.8 \pm 0.3}$ for calculations performed in energy units (as in HydroLight). In terms of the excitation wavelength, Bartlett et al. found $b_R(\lambda') = b_R(488)(488/\lambda')^{5.5 \pm 0.4}$ for energy computations. (The current version of HydroLight uses $b_R(\lambda') = b_R(488)(488/\lambda')^{5.5}$ as the default.) For calculations in terms of photon numbers (as in a Monte Carlo simulation), Bartlett et al. found wavelength dependencies of $(\lambda')^{-5.3 \pm 0.3}$ or $(\lambda)^{-4.6 \pm 0.3}$.

7.2.2 The Raman Wavelength Redistribution Function

The Raman wavelength distribution function $f_R(\lambda', \lambda)$ relates the excitation and emission wavelengths, i.e., what wavelengths λ receive the Raman-scattered spectral irradiance for a given excitation wavelength λ' or, conversely, what wavelengths λ' excite a given emission wavelength λ . The function $f_R(\lambda', \lambda)$ is most conveniently described in terms of the corresponding *wavenumber* distribution function $f_R(\nu'')$, where ν'' is the wavenumber shift, expressed in units of cm^{-1} . This follows because the Raman-scattered light undergoes a frequency shift that is determined by the type of molecule and is independent of the incident frequency. The wavenumber ν in cm^{-1} is related to the wavelength λ in nm by $\nu = 10^7/\lambda$, and to the frequency f by $\nu = f/c$, where c is the speed of light. (The 10^7 factor converts nanometers to centimeters.)

According to Walrafen (1967), the shape of $f_R(\nu'')$ for water is given by a sum of four Gaussian functions:

$$f_R(\nu'') = \left[\left(\frac{\pi}{4 \ln 2} \right)^{\frac{1}{2}} \sum_{i=1}^4 A_i \right]^{-1} \sum_{j=1}^4 A_j \frac{1}{\Delta\nu_j} \exp \left[-4 \ln 2 \frac{(\nu'' - \nu_j)^2}{\Delta\nu_j^2} \right] \quad (\text{cm}), \quad (7.3)$$

where

- ν'' is the wavenumber shift of the Raman-scattered light, relative to the wavenumber ν' of the incident light, in cm^{-1}
- ν_j is the center of the j^{th} Gaussian function, in cm^{-1}
- $\Delta\nu_j$ is the full width at half maximum of the j^{th} Gaussian function, in cm^{-1}
- A_j is the nondimensional weight of the j^{th} Gaussian function.

The values of A_j , ν_j , and $\Delta\nu_j$ for pure water at a temperature of 25 deg C are given in Table 7.1. Figure 7.1 shows $f_R(\nu'')$ evaluated for the water parameter values of Table 7.1. The function shows a peak and a shoulder, which result from the sums of the four Gaussians seen in Eq. (7.3). For water, the wavenumber shift is roughly 3400 cm^{-1} .

Consider, for example, incident light with $\lambda' = 500 \text{ nm}$, which corresponds to $\nu' = 20000 \text{ cm}^{-1}$. Figure 7.1 shows that this light, if Raman scattered, will be shifted by roughly 3400 cm^{-1} to $\nu = 16600 \text{ cm}^{-1}$, which corresponds to $\lambda \approx 602 \text{ nm}$.

The function $f_R(\nu'')$ can be interpreted as a probability density function giving the probability that a light of any incident wavenumber $\nu' = 10^7/\lambda'$, if Raman scattered, will be scattered to a wavenumber

$$\nu = \nu' - \nu''. \quad (7.4)$$

The function $f_R(\nu'')$ satisfies the normalization condition

$$\int_0^{\nu'} f_R(\nu'') d\nu'' = 1, \quad (7.5)$$

j	A_j	$\nu_j \text{ cm}^{-1}$	$\Delta\nu_j \text{ cm}^{-1}$
1	0.41	3250	210
2	0.39	3425	175
3	0.10	3530	140
4	0.10	3625	140

Table 7.1: Parameter values for the Raman wavenumber redistribution function $f_R(\nu'')$ of Eq. (7.3), for pure water at a temperature of 25 C. Data from Walrafen (1967).

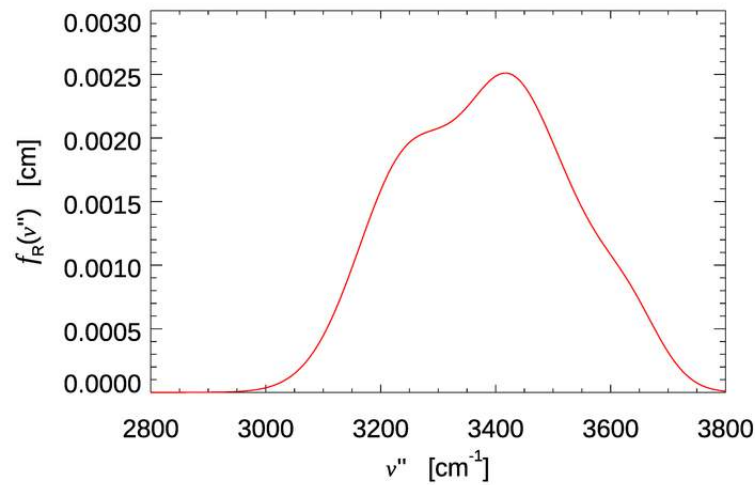


Figure 7.1: The $f_R(\nu'')$ function of Eq. (7.3) for pure water at 25 C.

as is required of any probability distribution function. The integration limits above come from observing that as $\lambda \rightarrow \infty$ then the wavenumber $\nu'' \rightarrow \nu'$, and as $\lambda \rightarrow \lambda'$ then $\nu'' \rightarrow 0$. A change of variables from ν'' to λ in Eq. (7.5) leads to the corresponding *wavelength* redistribution function $f_R(\lambda' \rightarrow \lambda)$. Thus

$$\begin{aligned} \int_0^{\nu'} f_R(\nu'') d\nu'' &= \int_{\lambda'}^{\infty} f_R\left(\frac{10^7}{\lambda''}\right) \frac{d\nu''}{d\lambda} d\lambda = \int_{\lambda'}^{\infty} f_R\left[10^7\left(\frac{1}{\lambda'} - \frac{1}{\lambda}\right)\right] \frac{10^7}{\lambda^2} d\lambda \\ &\equiv \int_{\lambda'}^{\infty} f_R(\lambda' \rightarrow \lambda) d\lambda = 1, \end{aligned}$$

where the wavelengths are in nanometers. In the last equation, we have identified the function

$$f_R(\lambda', \lambda) \equiv \begin{cases} \frac{10^7}{\lambda^2} f_R\left(\frac{10^7}{\lambda'}\right) = \frac{10^7}{\lambda^2} f_R\left[10^7\left(\frac{1}{\lambda'} - \frac{1}{\lambda}\right)\right] & \text{if } \lambda' < \lambda \\ 0 & \text{if } \lambda' \geq \lambda \end{cases} \quad (7.6)$$

as being the desired Raman wavelength redistribution function, with units of nm^{-1} .

In $f_R(\lambda', \lambda)$ we can fix the incident wavelength λ' and plot the corresponding emission wavelengths. We can also fix the emission wavelength λ and use the function to see where the light emitted at λ comes from. Both options are seen in Fig. 7.2.

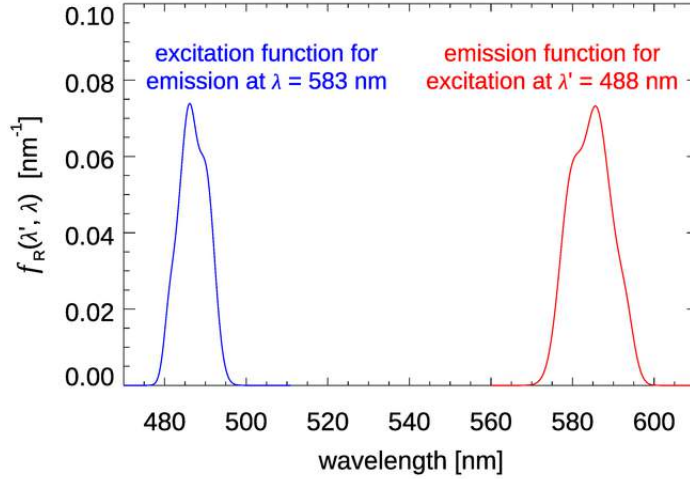


Figure 7.2: Raman emission function for excitation at 488 nm (red), and excitation function for emission at 583 nm (blue). Compare with similar figures in [Ge et al. \(1993, Fig. 3\)](#) and [Kattawar and Xu \(1992, Fig. 2\)](#).

Figure 7.3 shows $f_R(\lambda', \lambda)$ for four values of the incident wavelength λ' . These plots show that as the incident wavelength λ' increases, the emission band becomes broader and the shift from λ' to λ becomes larger. Excitation at 400 nm gives emission centered at roughly 463 nm, a shift of 63 nm, but excitation at 550 nm gives emission centered at round 677 nm, a shift of 127 nm.

Equation (7.4) can be rewritten as

$$\lambda = \frac{10^7}{\frac{10^7}{\lambda'} - 3400}$$

and used to compute the approximate center of the emission wavelength band for a given excitation wavelength. Figure 7.4 shows the result.

7.2.3 The Raman Phase Function

The Raman phase function $\tilde{\beta}_R(\psi)$ gives the angular distribution of the Raman scattered radiance. This function (averaging over all polarization states) is given by

$$\tilde{\beta}_R(\psi) = \frac{3}{16\pi} \frac{1+3\rho}{1+2\rho} \left[1 + \left(\frac{1-\rho}{1+3\rho} \right) \cos^2 \psi \right],$$

where ψ is the scattering angle between the direction of the incident and scattered radiance, and ρ is the depolarization factor. The value of ρ depends on the wavenumber shift ν'' ([Ge](#)

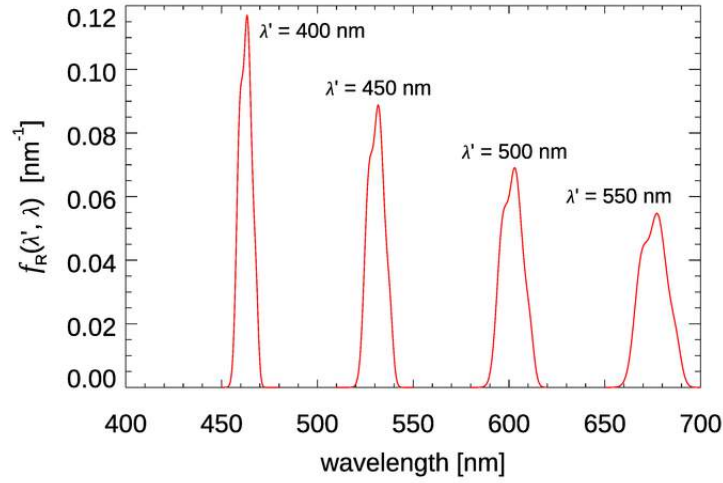


Figure 7.3: The Raman emission functions for four excitation wavelengths, as computed from Eq. (7.6).

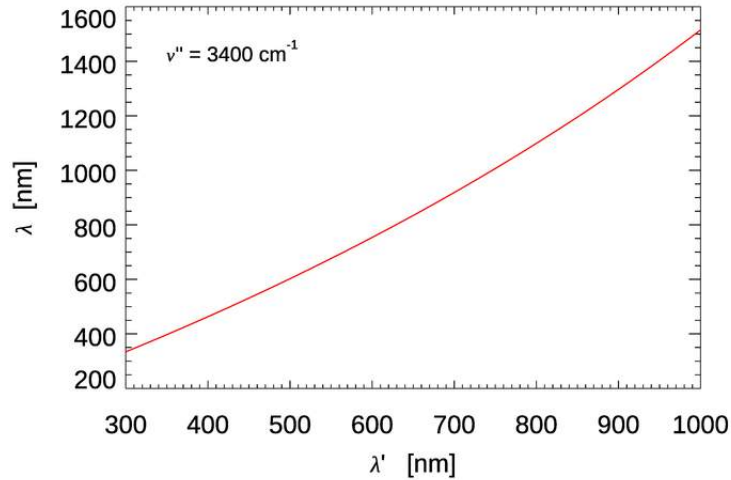


Figure 7.4: Approximate center λ of the Raman emission band for a given excitation wavelength λ' .

et al., 1993, Fig. 2). For a value of $\nu'' = 3400 \text{ cm}^{-1}$, $\rho \approx 0.18$, in which case the phase function is

$$\tilde{\beta}_R(\psi) = 0.068 (1 + 0.53 \cos^2 \psi).$$

This phase function is similar in shape to the phase function for elastic scattering by pure water.

7.2.4 Incorporation of Raman Scattering into Radiative Transfer Calculations

We now have the pieces needed to define the *volume scattering function for Raman scattering*, $\beta_{\text{R}}(\hat{\xi}' \rightarrow \hat{\xi}; \lambda' \rightarrow \lambda)$, where $\hat{\xi}'$ and $\hat{\xi}$ represent the incident and final directions of the light. This VSF specifies the strength of the Raman scattering via the Raman scattering coefficient $b_{\text{R}}(\lambda')$, the wavelength distribution of the scattered light via the wavelength redistribution function $f_{\text{R}}(\lambda', \lambda)$, and its angular distribution relative to the direction of the incident light via the Raman phase function $\tilde{\beta}_{\text{R}}(\psi)$. Thus we have a specific instance of the general form (7.1):

$$\beta_{\text{R}}(\hat{\xi}' \rightarrow \hat{\xi}; \lambda' \rightarrow \lambda) = b_{\text{R}}(\lambda') f_{\text{R}}(\lambda', \lambda) \tilde{\beta}_{\text{R}}(\psi) \quad (\text{m}^{-1} \text{nm}^{-1} \text{sr}^{-1}). \quad (7.7)$$

This inelastic VSF is then used in the radiative transfer equation (7.2).

7.2.5 Interpretation of Raman emission profiles

An interesting example of the contribution of Raman scattering to in-water radiances is seen in Fig. 7.5. This plot shows the downwelling (zenith-viewing) radiance generated from a HydroLight run using its option for simulating lidar illumination at one wavelength in an otherwise black sky. The inputs were as follows:

- The incident (lidar) irradiance was $E_{\text{d}}(\text{direct beam}) = 1 \text{ W m}^{-2}$ at 488 nm; the sky was otherwise black.
- The chlorophyll concentration was $Chl = 0.05 \text{ mg m}^{-3}$ for Case 1 water.
- The water was infinitely deep water with Eq. (7.2) solved down to 50 m.
- The output was at 560 to 610 nm at 1 nm resolution.

The curves in Fig. 7.5 are explained as follows. At depth 0, just below the sea surface, the only contribution to $L_{\text{d}}(0, \lambda)$ is upwelling radiance that is reflected back downward by the sea surface. By 5 m depth, there is now enough water above the simulated measurement instrument that the water column is generating significant downwelling radiance. Note that the shape of the emission has the same shape as the emission function seen in Fig. 7.2, namely a peak near 585 nm with a shoulder at about 580 nm. As the depth increases to 10 and then 20 m, the magnitude of $L_{\text{d}}(z, \lambda)$ increases, but the shape of the emission begins to flatten between 580 and 585. By 30 m, the magnitude of $L_{\text{d}}(z, \lambda)$ has decreased because of the decrease in the radiance penetrating to this depth from the water between the surface and 30 m, and the magnitude continues to decrease as the depth becomes greater. However, the shape of the emission at 30 m shows almost the same magnitudes at 580 and 585 nm, and at 40 and 50 m, the peak emission is actually greater at 580 than at 585 nm. This may seem strange because the shape of the Raman emission function seen in Fig. 7.2 is the same for all depths.

This “reversal” of the “peak-shoulder” shape of the emission is a consequence of the difference in absorption across the emission wavelengths. The total absorption (water plus phytoplankton) increases by a factor of three (from 0.071 to 0.221 m^{-1}) between 570 and 600 nm, and by 22% (from 0.091 to 0.111 m^{-1}) between 580 and 585 nm. These rapidly increasing absorption values change the shape of the local (at each depth) emission function

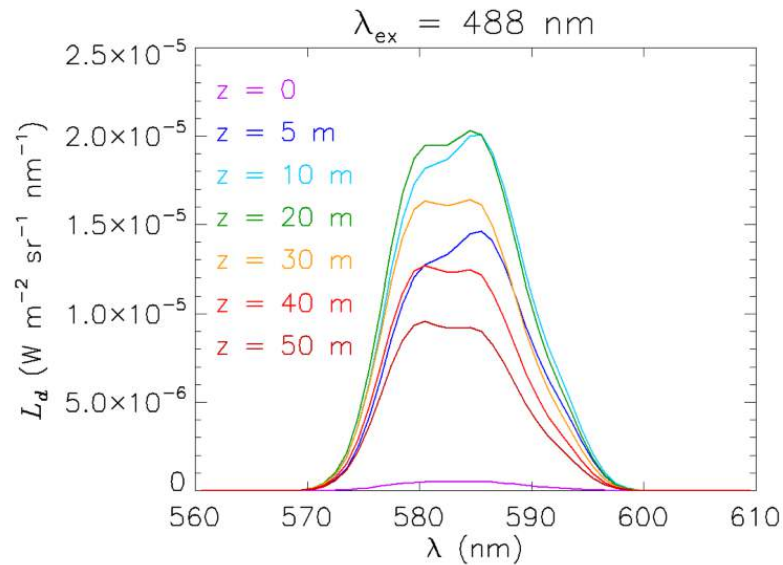


Figure 7.5: Downwelling Raman-scattered radiance $L_d(z, \lambda)$ as a function of depth and wavelength over the Raman emission band for 488 nm excitation. The wavelength shape of the $L_d(z, \lambda)$ emission band depends on depth because of wavelength-dependent absorption.

when integrated over depth to obtain the total $L_d(z, \lambda)$, which has contributions from all depths. Simply stated, the higher absorption at the 585 nm peak lets relatively less of the radiance emitted above a given depth reach the measurement depth than for the shoulder at 580 nm, so the 585 nm peak appears smaller relative to the shoulder than what is seen for the emission function of Fig. 7.2.

The claim that the change in shape with depth of the Raman emission is due to wavelength-dependent absorption can be verified as follows. An “artificial water” IOP data file was created with the IOP values between 570 and 600 nm having the values at 570 nm. The water IOPs are then the same over the entire range of emission wavelengths seen in Fig. 7.5. The resulting Raman $L_d(z, \lambda)$ spectra are seen in Fig. 7.6. Now, without the wavelength-dependent absorption, the shape of $L_d(z, \lambda)$ does not change with depth and, indeed, looks exactly like the shape of the emission function seen in Fig. 7.2. The magnitude of the $L_d(z, \lambda)$ curves is greater than before because the absorption is less. Of course, if the HydroLight run had been made at 5 or 10 nm resolution, then the shape of the emission band would not have been resolved. However, the total Raman-scattered power would have been the same, but spread over the wider bands.

These simulations show that the interpretation of Raman-scattered spectra can be complicated because of IOP effects at both the excitation and emission wavelengths, even in the simplest possible case of excitation at one wavelength in an otherwise black sky. The situation becomes even more complicated for solar-stimulated Raman scatter because many excitation wavelengths can contribute to a range of emission wavelengths, and everything blurs together in a non-obvious fashion.

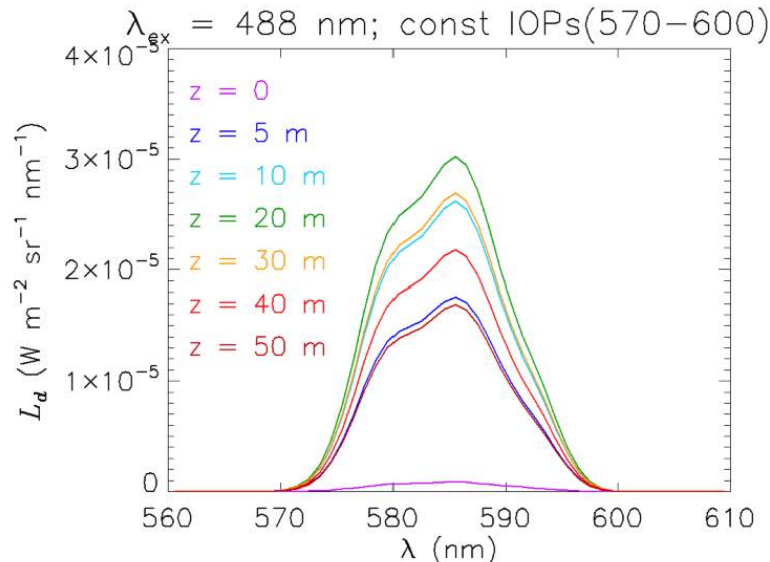


Figure 7.6: Downwelling Raman-scattered radiance as a function of depth for the case of artificial water IOPs having the same values from 570 to 600 nm. The shape of $L_d(z, \lambda)$ is now independent of depth and is the same as the Raman-emission function of Fig. 7.2.

7.2.6 Temperature and Salinity Dependence of Raman Scattering

The Walrafen data presented in Table 7.1 were determined on pure water at a temperature of 25 deg C. There is, however, a small but significant dependence on temperature and salinity of the shape of the emission curve seen in Fig. 7.1. This dependence is shown in Fig. 7.7. The upper panel shows the shape of the emission curve as a function of temperature for a salinity of 15 PSU, and the lower curve shows the dependence on salinity for a temperature of 25 deg C. Artlett and Pask (2017) have shown in laboratory measurements that these differences can be used to simultaneously determine temperature and salinity with an RMS errors of ± 0.7 deg C and ± 1.4 PSU, and they present the design for a three-channel Raman spectrometer that excites at 532 nm and measures the emission at three bands. The emission bands would be used to form band ratios, from which temperature and salinity can be extracted.

Because Raman scattering is determined by the water itself, its contributions to the light field can be computed exactly. It can therefore be used as a “known reference” for calibration a lidar system as the signal leaving the ocean is proportional to the intensity of the source; e.g. Hoge et al. (1988).

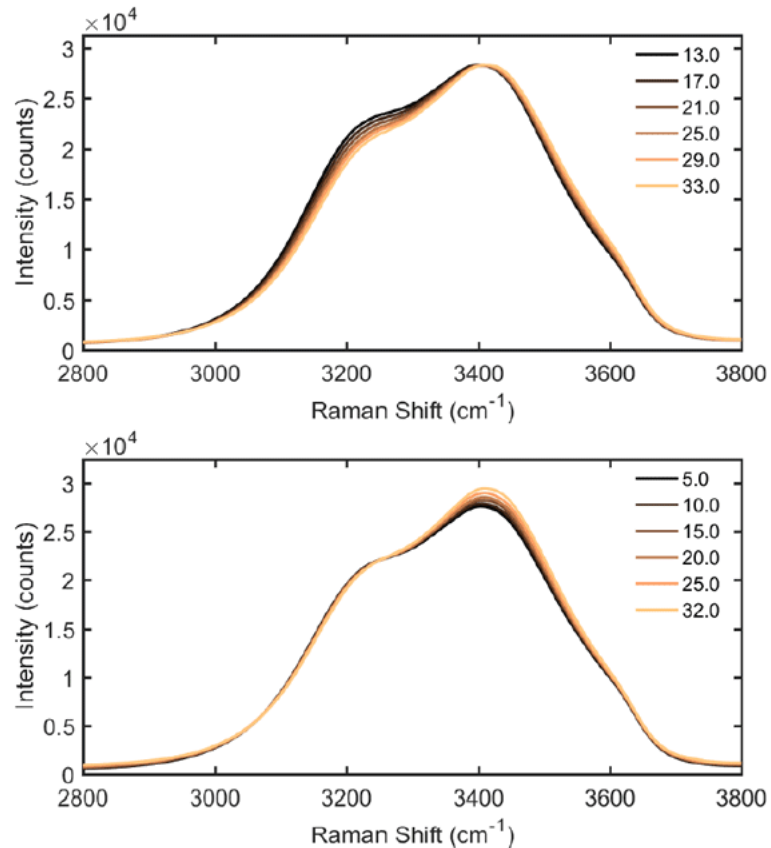


Fig. 2. (top) Raman spectra with changing temperature and fixed salinity (15 PSU). (bottom) changing salinity and fixed temperature (25 °C).

Figure 7.7: Dependence of the Raman wavelength redistribution function on temperature and salinity. (Figure 2 from [Artlett and Pask \(2017\)](#), reproduced by permission of Optica under their Fair Use policy)

7.3 Theory of Fluorescence and Phosphorescence

Luminescence refers to the emission of light by processes other than thermal emission, which is called *incandescence*. There are many types of luminescence: fluorescence, phosphorescence, chemiluminescence, triboluminescence, radioluminescence, and so on. The luminescence processes that are of primary interest in oceanography are fluorescence (by chlorophyll, CDOM, or pollutants), chemiluminescence (bioluminescence), and radioluminescence (Cherenkov radiation).

When light is absorbed by a molecule, one of three things can happen:

1. the energy is used for photochemistry, e.g. for photosynthesis in a chlorophyll molecule;
2. the energy goes into vibrational modes of the molecule, i.e. into heat; or
3. the energy is re-emitted as light via fluorescence or phosphorescence.

This section develops the general theory needed to include fluorescence by chlorophyll and colored dissolved organic matter (CDOM) in radiative transfer calculations. Specific

implementations of the general theory for chlorophyll and CDOM fluorescence are given on the next two sections.

7.3.1 The Physics of Fluorescence and Phosphorescence

The Pauli Exclusion Principle is one of the foundations of quantum mechanics. It states that no two fermions in an atom or molecule can have the same set of quantum numbers. Fermions are particles with an intrinsic angular momentum or “spin” that is a half-integer multiple of $\hbar = h/2\pi$, where h is Planck’s constant. Electrons are fermions with an angular momentum of $\hbar/2$ and are called “spin 1/2” fermions. Thus for our purposes, the Pauli Exclusion Principle says that two electrons in the same atomic or molecular orbital must have opposite spins, which are called “up” and “down.” (The labels “up” and “down” historically referred to orientations of the angular momentum vector relative the direction of an external magnetic field. “Up” and “down” correspond to spin quantum numbers of $s = +1/2$ and $-1/2$. For more on the terminology of quantum numbers, orbitals, etc. see the Physics of Absorption Section 5.3.)

A *singlet* state of a molecule is one in which all electrons are paired in up and down pairs. A *triplet* state is one in which one set of two electrons in different orbitals have the same orientation, up-up or down-down. Figure 7.8 illustrates the idea. (Again, the terminology is historical and relates to the number of lines seen in a spectrum for molecules that are not spherically symmetric or when a molecule is placed in a magnetic field, which splits the energy states. There is also a doublet state corresponding to one un-paired electron, but it does not concern us here.)

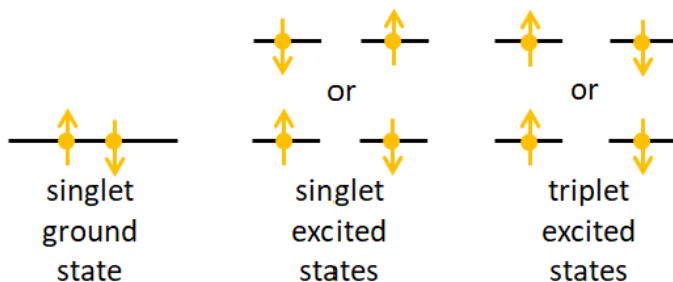


Figure 7.8: Illustration of singlet and triplet states. The black lines are energy levels and the gold dots are electrons; the arrows represent up and down angular momentum states.

Figure 7.9 is a Jablonski² diagram of energy states in a molecule. Such diagrams group the energy levels vertically and the spin states horizontally. The black lines represents the energy levels; the thick lines labeled S_0 and S_1 are the electronic levels, and the thinner lines are vibrational levels. The gold dots with arrows represent electrons in either up or down angular momentum states.

²The guy’s name is properly spelled Jabłoński. The Polish “J” is like the English “Y”; the L with a line through it is like the “W” in water, and the accented N is like the Cyrillic “H” in the Russian word “HeT,” or Nyet in English. So Jabłoński is pronounced something like YabWoNYski with English spelling; the accent is on the second syllable. He is one of those unfortunate souls who, like Khrushchev and Gengis Kahn, are forever doomed to have their names mispronounced by English speakers.

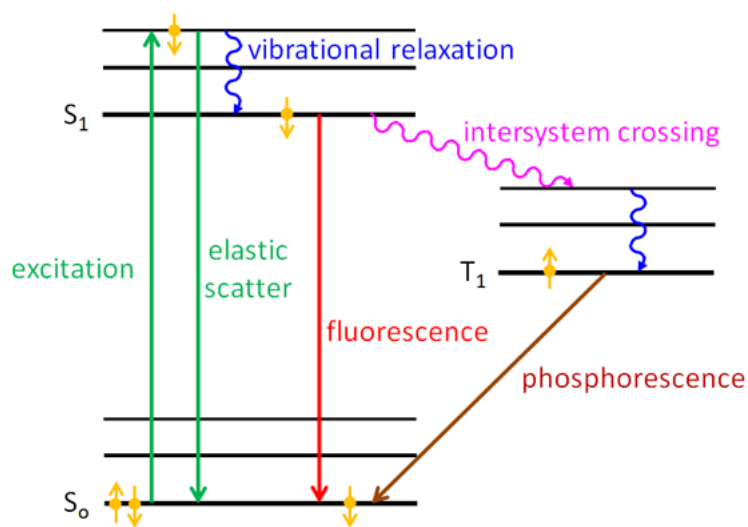


Figure 7.9: Jablonski diagram illustrating the fundamental processes of fluorescence and phosphorescence.

Now suppose that one of the electrons of an electron pair, say the down electron, in the ground state of a molecule absorbs a photon. The energy of the photon excites the electron to a higher energy level as shown by the upward green arrow. The time scale for this excitation is on the order of 10^{-15} s. The electron spin is still down in the excited energy level. One thing that can happen is that the electron almost immediately drops back to the ground state with emission of a photon with the same wavelength as the excitation photon. This is elastic scattering and is represented by the green downward arrow in the figure. This process is so fast that the excited electron “remembers” the direction of the incident photon, and the direction of the emitted/scattered photon depends on the direction of the incident photon; i.e., the scattering is not isotropic.

Another thing that can happen is that part of the energy of the excited electron is given up to other vibrational modes (i.e., to heat), so that the electron drops to a lower energy level of the excited state. This is indicated by the blue wiggly line labeled “vibrational relaxation” in Fig. 7.9. This loss of energy is called a “radiationless transition” because it does not involve the emission of a photon. The time scale for vibrational relaxation is 10^{-12} - 10^{-10} s. The electron may stay in the excited S_1 state for a while, but if the electron is still in the down state, i.e. it is in a singlet state, it can drop back to the ground level by emission of a photon of lower energy (longer wavelength) than the exciting photon. The time scale for this is of order 10^{-10} - 10^{-7} s. The ground state then once again contains an up-down pair of electrons. This is fluorescence and is shown by the red downward arrow in the figure. The processes of vibrational relaxation and fluorescence emission take so long that the electron has “forgotten” the direction of the incident photon, and the emitted photon is equally likely to be in any direction; i.e., the emission of fluoresced light is isotropic.

An electron cannot just simply “flip” from a down to an up state, which would violate the law of conservation of angular momentum. Such a flip is therefore a “forbidden tran-

sition” in quantum mechanics terminology. However, in quantum mechanics, “forbidden” does not mean “do not ever do something,” just “do not do it very often.” So a third thing that can happen is that the down electron sitting in the S_1 state can exchange some angular momentum with the orbital angular momentum of the molecule via what is called “spin-orbit coupling.” The electron can then flip to an up state. This is another radiationless transition and is called an “intersystem crossing” and is shown in purple in Fig. 7.9. The time scale for intersystem crossing is on the order of 10^{-10} - 10^{-8} s. The electron can also undergo further vibrational relaxation in the triplet state. This up electron is no longer paired with the up electron that was left in the ground state; it is in an excited triplet state, which is labeled by T_1 in figure. The energy level of the first excited triplet state T_1 is usually lower than for the first excited singlet state S_1 because the electrons are further apart, which makes their Coulomb repulsion less. This up electron cannot drop back to the ground state because the result would give two up electrons in the ground state, which violates the Pauli Exclusion Principle; this is another forbidden transition. Therefore the up electron in the T_1 state must wait for another chance for spin-orbit coupling, which can flip its spin to down. This takes a long time (on the atomic scale), so the T_1 state is called “metastable.” That down electron can then drop into the down-electron spot in the ground level, which is again paired with the up electron in the ground state. This process is phosphorescence and can take from 10^{-6} to as long as 10 s. The emitted light is again isotropic because of the long time between absorption and re-emission.

In summary, fluorescence is a transition from an excited singlet state to the ground singlet state, and phosphorescence is a transition from an excited triplet state to the ground singlet state. There are other pathways to fluorescence. For example, the original absorbed photon could excite the ground state electron to the S_2 state (second excited singlet state), which might then do a radiationless transfer of energy to an overlapping vibrational level of the S_1 orbital; this is called “internal conversion” (internal within a singlet state), and the time scale is 10^{-11} - 10^{-9} s. The electron can then vibrationally relax to the S_1 level, and then fluoresce, and so on. It should also be noted that there are many vibrational energy levels in each of the electronic levels illustrated in Fig. 7.9, and transitions can occur between any of these. This gives a spread of energies of the emitted photons, i.e., a spread of wavelengths. See the related discussion in Section 5.3.

As previously noted, time scales of elastic scattering, and of Raman Scattering, are so short, 10^{-13} s or less, that they are called “scattering.” However, as just seen, fluorescence and phosphorescence are clearly absorption followed much later by the emission of a new photon that is uncorrelated with the absorbed photon. However, for solution of the time-independent radiative transfer equation, fluorescence and phosphorescence can be regarded as “inelastic scattering” and treated with the same mathematical formalism as that used for Raman scattering. This is the case for “solar-stimulated” chlorophyll and CDOM fluorescence.

However, measuring the time dependence of fluorescence emission at time scales of 10^{-12} to 10^{-9} s gives information about the internal details of the energy transfers within the molecule. Indeed, the quantum efficiency Φ_C of chlorophyll fluorescence is determined by measuring the decay of chlorophyll fluorescence (on a time scale of order 10 ns) excited by an extremely short light pulse (or order 1 ns). Thus time-resolved measurements of chlorophyll fluorescence give information about the photosystems responsible for photosynthesis and the effect of environmental stresses (pH, light adaptation, etc) on those

systems. That requires a much more complicated modeling of fluorescence than what is discussed here. See [Falkowski et al. \(2016\)](#) for an overview of time-resolved chlorophyll fluorescence measurements.

7.3.2 Incorporation of Fluorescence into Time-independent Radiative Transfer Theory

Just as for any inelastic process, the quantities needed to compute fluorescence contributions to the radiance are

- the fluorescence scattering coefficient $b_F(z, \lambda')$, with units of m^{-1} ,
- the fluorescence wavelength redistribution function $f_F(\lambda', \lambda)$, with units of nm^{-1} , and
- the fluorescence scattering phase function $\tilde{\beta}_F(\psi)$, with units of sr^{-1} .

The subscript “F” indicates fluorescence; it will be replaced on the next sections by “C” for chlorophyll or “Y” for yellow matter (and it was “R” in the discussion of Raman scatter, which uses the same mathematical formalism). The scattering coefficient $b_F(z, \lambda')$ in general depends on the depth because *amount* of fluorescing material, e.g. the amount of chlorophyll or CDOM, depends on depth, whereas the wavelength redistribution function is determined only by the *type* of fluorescing material. The phase function is isotropic. It is thus reasonable to place the depth dependence of the fluorescence in the scattering coefficient.

Multiplied together, these functions give the *volume inelastic scattering function for fluorescence* $\beta_F(z; \theta', \phi' \rightarrow \theta, \phi; \lambda' \rightarrow \lambda) = \beta_F(z, \psi, \lambda', \lambda)$:

$$\beta_F(z, \psi, \lambda', \lambda) = b_F(z, \lambda') f_F(\lambda', \lambda) \tilde{\beta}_F(\psi) \quad [\text{m}^{-1} \text{sr}^{-1} \text{nm}^{-1}]. \quad (7.8)$$

This function is used as a source function in the scalar radiative transfer equation (7.2). This is conceptually the same set of quantities required in the previous section to describe Raman scatter. Of course, the functional forms and magnitudes of these quantities will be different for Raman scatter, chlorophyll fluorescence, and CDOM fluorescence.

It must be remembered that radiative transfer theory is formulated in terms of energy. For fluorescence, the quantity of interest in computing radiance via Eq. (7.2) is how much energy is emitted by fluorescence compared to how much energy is absorbed. However, the folks who study fluorescence usually work in quantum units, i.e., how many photons are emitted as fluorescence compared to how many photons are absorbed. Accordingly, the *spectral fluorescence quantum efficiency function* is defined as

$$\eta_F(\lambda', \lambda) = \frac{\text{the number of photons emitted at } \lambda, \text{ per unit wavelength interval}}{\text{the number of photons absorbed at } \lambda'} \quad [\text{nm}^{-1}].$$

The numerator of this function must be multiplied by hc/λ , and the denominator by hc/λ' , to convert the photon counts to energy. The result is that

$$f_F(\lambda', \lambda) = \eta_F(\lambda', \lambda) \frac{\lambda'}{\lambda} \quad [\text{nm}^{-1}]. \quad (7.9)$$

Another quantity often seen in the literature is the non-dimensional *quantum efficiency* (or *quantum yield*) of fluorescence, which is defined by

$$\Phi_F(\lambda') = \frac{\text{the number of photons emitted at all wavelengths } \lambda}{\text{the number of photons absorbed at } \lambda'}$$

$\Phi_F(\lambda')$ is obtained from $\eta_F(\lambda', \lambda)$ by

$$\Phi_F(\lambda') = \int_{\lambda'}^{\infty} \eta_F(\lambda', \lambda) d\lambda. \quad (7.10)$$

A paper on chlorophyll fluorescence may discuss other quantum efficiencies, namely a quantum efficiency for photosynthesis Φ_{ph} and a quantum efficiency for heating Φ_{H} . The quantum efficiency for photosynthesis is the ratio of photons whose energy goes into photosynthesis to the number of photons absorbed, and similarly for the number whose energy stays in vibrational modes (i.e., heat). From the opening comments of this section it follows that $\Phi_F + \Phi_{\text{ph}} + \Phi_{\text{H}} = 1$. Typical upper-ocean numbers as measured by Falkowski et al. (2016) are $\Phi_F = 0.07$, $\Phi_{\text{ph}} = 0.35$, and $\Phi_{\text{H}} = 0.58$.

7.3.3 Comments on Terminology

It might appear that $f_F(\lambda', \lambda)$ or $\eta_F(\lambda', \lambda)$ is the fluorescence “excitation-emission” (Ex-Em) function or matrix that is commonly seen in papers (e.g., Fig. 7.16 of Section 7.5). However, these are different quantities. $\eta_F(\lambda', \lambda)$ is a fluorescence *efficiency* that measures how many quanta are emitted per quanta absorbed. Ex-Em functions are measurements of the number of emitted quanta (e.g., as counts per second) without regard for how many quanta are absorbed. In addition, the factor of λ'/λ that converts $\eta_F(\lambda', \lambda)$ to $f_F(\lambda', \lambda)$ shifts the location of maxima in the functions because different wavelengths correspond to different numbers of quanta for the same energy. Thus $f_F(\lambda', \lambda)$ and Ex-Em spectra are qualitatively similar in appearance, but quantitative comparison is difficult. In addition, Ex-Em spectra are often presented as normalized values, which makes quantitative comparison impossible. Relative excitation and emission values have many uses, but the radiative transfer equation requires a calibrated function in units of 1/nm.

I have called $\eta_F(\lambda', \lambda)$ the “spectral fluorescence quantum efficiency function”, which comes from Hawes (1992). This Master’s Thesis is the *only* study I have found that measures this function (for CDOM fluorescence) in a calibrated form suitable for radiative transfer theory³. I called the energy equivalent $f_F(\lambda', \lambda)$ the “wavelength redistribution function” in *Light and Water* because it tells how energy at the excitation wavelengths λ' is redistributed to the emission wavelengths λ , but maybe the “spectral fluorescence energy efficiency function” would be a better name. Gordon (1979) calls a nondimensional function equivalent to $\eta_F(\lambda', \lambda)\Delta\lambda$ the “quantum efficiency.” Gordon calls my “volume inelastic scattering function for fluorescence” β the “volume fluorescence function,” and he calls

$$\Phi(\lambda', \lambda) = \int_{4\pi} \beta(\psi) d\Omega \quad \text{m}^{-1} \text{nm}^{-1}$$

the “coefficient of fluorescence.” I will leave it at that and define quantities as needed; I hope that my definitions are clear, even if the terminology is sometimes non-standard.

³I have talked to a couple of researchers in fluorescence about the need for calibrated excitation-emission functions in radiative transfer theory, i.e. my $f_F(\lambda', \lambda)$ or $\eta_F(\lambda', \lambda)$, but the response has been that they do not need calibrated functions for their applications, so they do not attempt to measure them.

7.4 Chlorophyll Fluorescence

This section now tailors the general fluorescence theory of the previous section to the case of fluorescence by chlorophyll in living phytoplankton. The goal is to develop the quantities needed for prediction of chlorophyll fluorescence contributions to oceanic light fields using a radiative transfer model like HydroLight. For convenience of reference, it is recalled that the quantities needed are

- the chlorophyll fluorescence scattering coefficient $b_C(z, \lambda')$, with units of m^{-1} ,
- the chlorophyll fluorescence wavelength redistribution function $f_C(\lambda', \lambda)$, with units of nm^{-1} , and
- the chlorophyll fluorescence scattering phase function $\tilde{\beta}_C(\psi)$, with units of sr^{-1} .

These quantities are then combined to create the volume inelastic scattering function for chlorophyll fluorescence,

$$\beta_C(z, \psi, \lambda', \lambda) = b_C(z, \lambda') f_C(\lambda', \lambda) \tilde{\beta}_C(\psi) \quad [\text{m}^{-1} \text{sr}^{-1} \text{nm}^{-1}]. \quad (7.11)$$

The subscript C indicates chlorophyll.

7.4.1 The Chlorophyll Fluorescence Scattering Coefficient

For chlorophyll fluorescence, the *inelastic* “scattering” coefficient in the formalism of treating the fluorescence as inelastic scattering is just the absorption coefficient for chlorophyll. In other words, what matters is how much energy is absorbed by chlorophyll at the excitation wavelength λ' , which is then available for possible re-emission at a longer wavelength λ . Note that it is only energy absorbed by the chlorophyll molecule that matters for chlorophyll fluorescence. Energy absorbed by other pigments may (or may not) fluoresce, but that is not chlorophyll fluorescence. Thus the needed chlorophyll scattering coefficient is commonly modeled as

$$b_C(z, \lambda') = Chl(z) a_{\text{Chl}}^*(\lambda') \quad [\text{m}^{-1}],$$

where $Chl(z)$ is the chlorophyll profile in mg Chl m^{-3} and $a_{\text{Chl}}^*(\lambda')$ is the chlorophyll-specific absorption spectrum in units of $\text{m}^2 (\text{mg Chl})^{-1}$. Examples of these spectra are seen in Section 8.3. (The *elastic* scattering coefficient for chlorophyll-bearing phytoplankton is often modeled as a power law, as described in the New Case I IOPs, Section 8.9.)

7.4.2 The Chlorophyll Fluorescence Wavelength Redistribution Function

Figure 7.10 shows a typical chlorophyll fluorescence emission spectrum.

Other than the general shape, the important feature of this emission spectrum is that it is independent of the excitation wavelength. The chlorophyll absorption spectrum peaks in the blue and is a minimum in the green, so a 435 nm photon is much more likely to be absorbed by a chlorophyll molecule than is a 570 nm photon. However, either photon, if absorbed, leads to the same fluorescence. Wavelengths in the range of 370 to 690 nm can, if absorbed, can lead to fluorescence. Given these observations, it is customary (e.g., Gordon, 1979) to factor the chlorophyll $f_C(\lambda', \lambda)$ into a product of functions:

$$f_C(\lambda', \lambda) = \eta_C(\lambda', \lambda) \frac{\lambda'}{\lambda} = \Phi_C g_C(\lambda') h_C(\lambda) \frac{\lambda'}{\lambda} \quad [\text{nm}^{-1}], \quad (7.12)$$

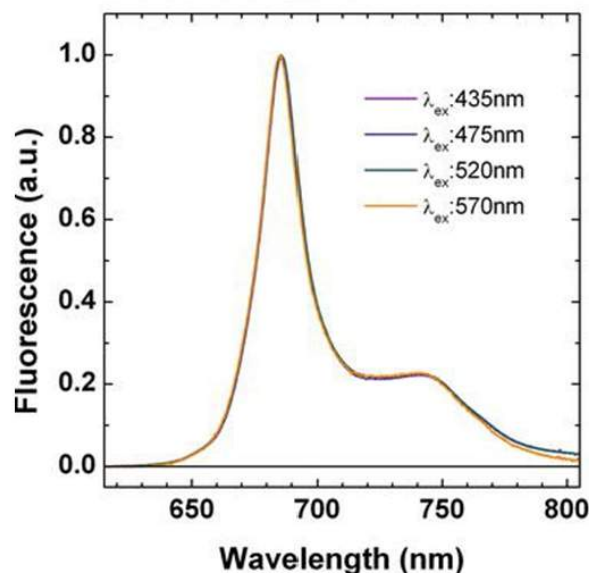


Figure 7.10: Measured chlorophyll fluorescence emission spectra for four excitation wavelengths for the chlorophyte *Chlorella sorokiniana*. Figure extracted from Fig. 3 of [Santabarbara et al. \(2020\)](#) under a Creative Commons License.

where

Φ_C is the quantum efficiency for chlorophyll fluorescence,

$g_C(\lambda')$ is a nondimensional function that specifies the interval over which light is able to excite chlorophyll fluorescence, and

$h_C(\lambda)$ is the chlorophyll fluorescence wavelength emission function, with units of nm^{-1} .

The following sections describe how each of these terms can be modeled.

7.4.3 The Quantum Efficiency of Chlorophyll Fluorescence

The first factor on the right-hand side of Eq. (7.12), the quantum efficiency Φ_C , is just a number, but it is the most difficult to model. This is because its value depends on the type and physiological state of the phytoplankton, and the physiological state is affected by the available light and nutrients, the temperature, and other factors.

Figure 7.11 shows three depth profiles of Φ_C determined as described in [Maritorena et al. \(2000\)](#). The locations were in oligotrophic areas of the equatorial Pacific where the chlorophyll values were between 0.035 and 0.29 mg Chl m^{-3} . The inset shows the fluorescence signal, which is proportional to the chlorophyll concentration and thus shows the shape of the $Chl(z)$ profiles. The arrows labeled Z_e are the depths of the euphotic zone, which was defined as the depth where the irradiance has decreased to 1% of its surface value. The irradiance decreases approximately exponentially with depth, so the ordinate axis roughly corresponds to a log-scale plot of irradiance level. In the high-irradiance, near-surface region, the quantum efficiency is between 0.005 and 0.01. However, in the

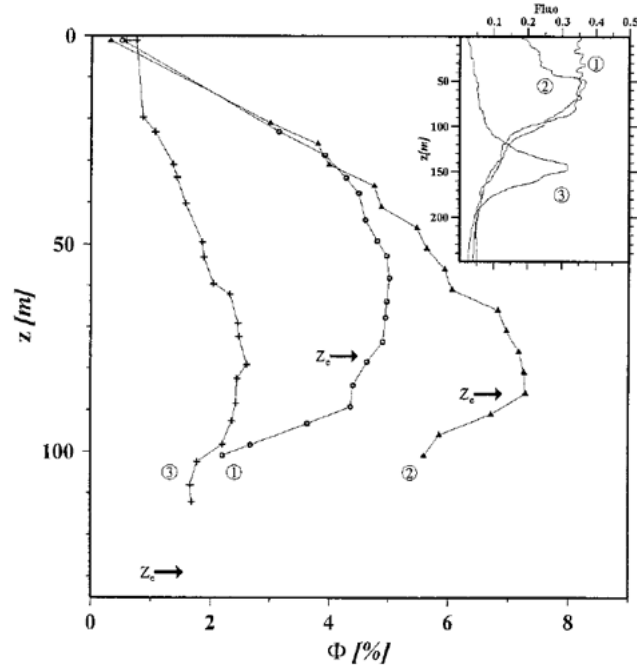


Figure 7.11: Depth profiles of Φ_C for three locations. From [Maritorena et al. \(2000, Fig. 7\)](#) and reproduced under Optica's Fair Use policy.

lower-irradiance regions below 50 m depth, Φ_C is as large as 0.07. Very similar profiles can be seen in Fig. 8 of [Morrison \(2003\)](#).

Several models have been developed to predict Φ_C as a function of the variables that affect it. Understanding these models requires a cellular-level understanding of the processes involved in photosynthesis, which is far beyond the level of this section. See, for example, [Kirk \(1994\)](#) for a general discussion of photosynthesis and [Kiefer and Reynolds \(1992\)](#) for discussion of the factors determining Φ_C . In addition to PAR, these models depend on quantities such as the fraction of open photosystem II (PSII) reaction centers. An example model is that of [Morrison \(2003, Eq. 18\)](#), which has the form

$$\Phi_C = \left[r + (1 - r) q_I e^{-PAR_o/PAR_T} \right] [\phi_{\min} A + \phi_{\max} (1 - A)], \quad (7.13)$$

where

- $r = 0.04$ is the fraction of PSII reaction centers that are unaffected by nonphotochemical quenching.
- q_I is related to nonphotochemical quenching and ranges between 0 (maximum quenching) and 1 (minimal quenching).
- PAR_o is the ambient scalar irradiance PAR in units of $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$
- $PAR_T = 350 \mu\text{mol quanta m}^{-2} \text{s}^{-1}$ is the saturation PAR value for energy dependent nonphotochemical quenching.
- $\phi_{\min} = 0.03$ and $\phi_{\max} = 0.09$.
- $PAR_k = 55 \mu\text{mol quanta m}^{-2} \text{s}^{-1}$ is the saturation PAR value for photosynthesis.

- $A = \exp(-PAR_o/PAR_k)$ is the fraction of open PSII reaction centers.

Quenching refers to any process that reduces the amount of fluorescence. These processes include the use of the absorbed energy for the chemical processes of photosynthesis (photochemical quenching) and the transfer of energy into heat (nonphotochemical quenching). Nonphotochemical quenching is common in phytoplankton as a way to protect themselves from the harmful effects of high irradiance levels. Quenching of whatever type reduces the energy available for re-emission as fluorescence and therefore reduces the quantum efficiency of fluorescence (with corresponding increases in the quantum efficiency of photosynthesis or of heating).

Figure 7.12 shows three curves for Φ_C as a function of the ambient PAR_o , for values of $q_I = 0.2, 0.4$, and 1.0 , which include the range of observed values seen in Morrison (2003, Fig. 8). Note the similarity to the profiles seen in Fig. 7.11: values of Φ_C less than 0.01 at high PAR values (i.e., near the surface), a maximum of around 0.06 at medium PAR values, and then decreasing for very low PAR values.

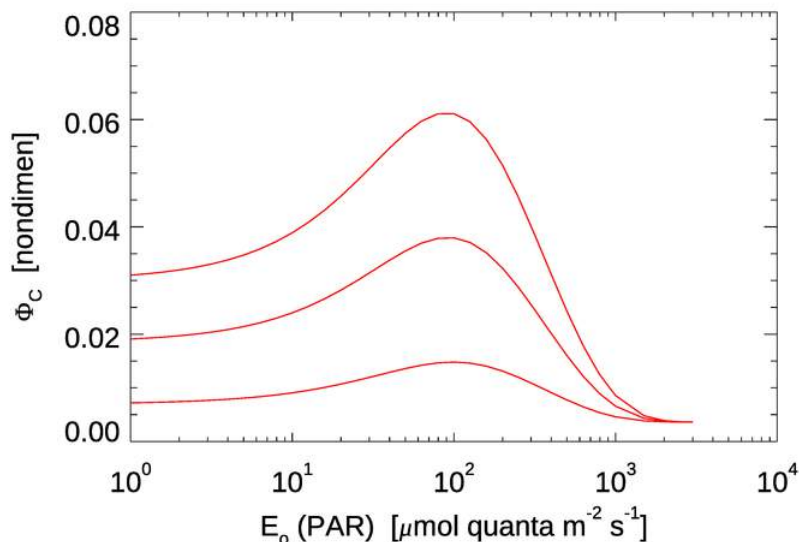


Figure 7.12: Φ_C as a function of PAR predicted by Eq. (7.13) for values of $q_I = 0.2$ (bottom curve), 0.4 , and 1.0 (top curve).

A more sophisticated model, including the effects of temperature and the surface chlorophyll concentration, is developed in Ostrovska (2012). That model gives curves qualitatively similar to those in Fig. 7.12, but with a maximum values up to 0.1 for some values of the temperature and surface chlorophyll concentration.

Thus measurements of Φ_C (e.g., Fig. 7.11) and recent models (e.g., Eq. (7.13)) are in reasonable agreement. However, a model for Φ_C in terms of PAR cannot be used in a radiative transfer model like HydroLight for the simple reason that the purpose of HydroLight is to predict the radiance and derived quantities, including PAR, by solving the radiative transfer equation (RTE), so the PAR_o values needed in Eq. 7.13 and similar models are not known until *after* the RTE has been solved. At best, HydroLight could

be run once to compute the PAR profile without fluorescence, and then run again with that PAR profile used in Eq. (7.13) to compute the chlorophyll fluorescence contribution (which conceptually leads to an iteration of solutions).

For this reason, the value of Φ_C to be used in a HydroLight simulation is left as a user input to be chosen at run time. The default value in the current version 6 is $\Phi_C = 0.02$. This is a mid-range value for moderate-irradiance, upper-ocean conditions, although Falkowski et al. (2016) report an average value of $\Phi_C = 0.07$ in surface waters for 200,000 profiles taken in a wide variety of locations.

7.4.4 The Chlorophyll Excitation Function

As previously noted, wavelengths in the range of 370 to 690 nm, if absorbed, are equally likely to excite chlorophyll fluorescence. Therefore, $g_C(\lambda')$ is modeled by

$$g_C(\lambda') = \begin{cases} 1 & \text{if } 370 \leq \lambda' \leq 690 \text{ nm,} \\ 0 & \text{otherwise.} \end{cases}$$

7.4.5 The Chlorophyll Emission Function

The emission function $h_C(\lambda)$ is commonly (e.g., Gordon, 1979) approximated as a Gaussian:

$$\begin{aligned} h_C(\lambda) &= \frac{1}{\sqrt{2\pi}\sigma_C} \exp\left[-\frac{1}{2}\left(\frac{\lambda - \lambda_C}{\sigma_C}\right)^2\right] \\ &= \sqrt{\frac{4 \ln 2}{\pi}} \frac{1}{FWHM} \exp\left[-4 \ln 2 \left(\frac{\lambda - \lambda_C}{FWHM}\right)^2\right] \quad [\text{nm}^{-1}] \end{aligned} \quad (7.14)$$

where

$\lambda_C = 685$ nm is the wavelength of maximum emission, and

$\sigma_C = 10.6$ nm is the standard deviation of the Gaussian; 10.6 nm corresponds to a full width at half maximum of $FWHM = 2\sqrt{2 \ln 2} \sigma_C = 25$ nm, as seen in the equivalent second form of the function.

It should be noted that this $h_C(\lambda)$, when used in the $\eta_C(\lambda', \lambda)$ defined in Eq. (7.12) and integrated over λ as in Eq. (7.15) of the preceding theory section,

$$\Phi_F(\lambda') = \int_{\lambda'}^{\infty} \eta_F(\lambda', \lambda) d\lambda \quad \text{for } 370 \leq \lambda' \leq 690 \text{ nm,} \quad (7.15)$$

gives the quantum efficiency Φ_C as required.

Figure 7.10 shows that this Gaussian captures only the main peak of the emission function. A better model for $h_C(\lambda)$ is a weighted sum of two Gaussians, one centered at 685 with a FWHM of 25 nm and one centered at 730 or 740 with a FWHM of 50 nm:

$$\begin{aligned} h_C(\lambda) &= W \sqrt{\frac{4 \ln 2}{\pi}} \frac{1}{25} \exp\left[-4 \ln 2 \left(\frac{\lambda - 685}{25}\right)^2\right] \\ &+ (1 - W) \sqrt{\frac{4 \ln 2}{\pi}} \frac{1}{50} \exp\left[-4 \ln 2 \left(\frac{\lambda - 730}{50}\right)^2\right] \quad [\text{nm}^{-1}], \end{aligned} \quad (7.16)$$

where W and $1 - W$ are the weights of the Gaussians at these wavelengths. These weights correspond to the fractions of the total quantum efficiency contributed by each Gaussian. Setting $W = 0.75$ gives the peak height of the second Gaussian as 0.2 of the first, consistent with Fig. 7.10. Using (7.16) in (7.12) and (7.15) then again recovers Φ_C .

The exact shape of the fluorescence emission seen in Fig. 7.10 and the corresponding best-fit parameters—heights and widths of the Gaussians and their center wavelengths—do vary somewhat with plankton species, pigment content and ratios, photoadaptation, nutrient conditions, stage of growth. and other parameters. This is, after all, why fluorescence gives information about the physiological state of phytoplankton.

7.4.6 The Chlorophyll Fluorescence Phase Function

As previously noted, fluorescence emission is isotropic. Therefore the phase function is simply

$$\tilde{\beta}_C(\psi) = \frac{1}{4\pi} \quad [\text{sr}^{-1}].$$

The models seen above give everything needed to construct the volume inelastic scattering function of Eq. (7.11) for chlorophyll fluorescence, $\beta_C(z, \psi, \lambda', \lambda)$, which is then ready for use in the radiative transfer equation as seen in Eq. (7.2).

7.5 CDOM Fluorescence

This section tailors the general fluorescence theory to the case of fluorescence by colored dissolved organic matter (CDOM). The goal is to develop the quantities needed for prediction of CDOM fluorescence contributions to oceanic light fields using a radiative transfer model. For convenience of reference, it is recalled from the theory section that the quantities needed are

- the CDOM fluorescence scattering coefficient $b_Y(z, \lambda')$, with units of m^{-1} ,
- the CDOM fluorescence wavelength redistribution function $f_Y(\lambda', \lambda)$, with units of nm^{-1} , and
- the CDOM fluorescence scattering phase function $\tilde{\beta}_Y(\psi)$, with units of sr^{-1} .

These quantities are then combined to create the volume inelastic scattering function for CDOM fluorescence

$$\beta_Y(z, \psi, \lambda', \lambda) = b_Y(z, \lambda') f_Y(\lambda', \lambda) \tilde{\beta}_Y(\psi) \quad [\text{m}^{-1} \text{sr}^{-1} \text{nm}^{-1}]. \quad (7.17)$$

The subscript Y indicates yellow matter, i.e. CDOM.

7.5.1 The CDOM Fluorescence Scattering Coefficient

For CDOM fluorescence, the *inelastic* “scattering” coefficient in the formalism of treating fluorescence as inelastic scattering is just the absorption coefficient for CDOM. In other words, what matters is how much energy is absorbed by CDOM at the excitation wavelength λ' , which is then available for possible re-emission at a longer wavelength λ . Note that it is only energy absorbed by CDOM molecules that matters for CDOM fluorescence.

Energy absorbed by chlorophyll in phytoplankton or pollutants such as oil may (or may not) fluoresce, but that is not CDOM fluorescence. Thus the needed CDOM fluorescence scattering coefficient is commonly modeled as

$$b_Y(z, \lambda') = a_Y(z, \lambda_0) \exp[-S_Y (\lambda' - \lambda_0)] \quad [\text{m}^{-1}], \quad (7.18)$$

where $a_Y(z, \lambda_0)$ is the absorption by CDOM at a reference wavelength λ_0 , which is usually taken to be 400 or 440 nm, and S_Y is a spectral slope parameter. S_Y is usually in the range of 0.016 to 0.018 nm^{-1} , but can vary from 0.007 to 0.026. See Fig. 8.16 and Section 8.4 for further discussion of the distribution of S_Y values. (As noted in Section 8.4, the elastic scattering coefficient for CDOM is usually assumed to be zero.)

7.5.2 The CDOM Fluorescence Wavelength Redistribution Function

Figure 7.13 shows a measured spectral fluorescence quantum efficiency function for CDOM, $\eta_Y(\lambda', \lambda)$, as measured by Hawes (1992) on a sample of water from the Gulf of Mexico on the West Florida Shelf (his station FA7). Other measurements in the Gulf of Mexico, Peru upwelling, and North Atlantic showed similar shapes, although with different magnitudes and some variability in the details.

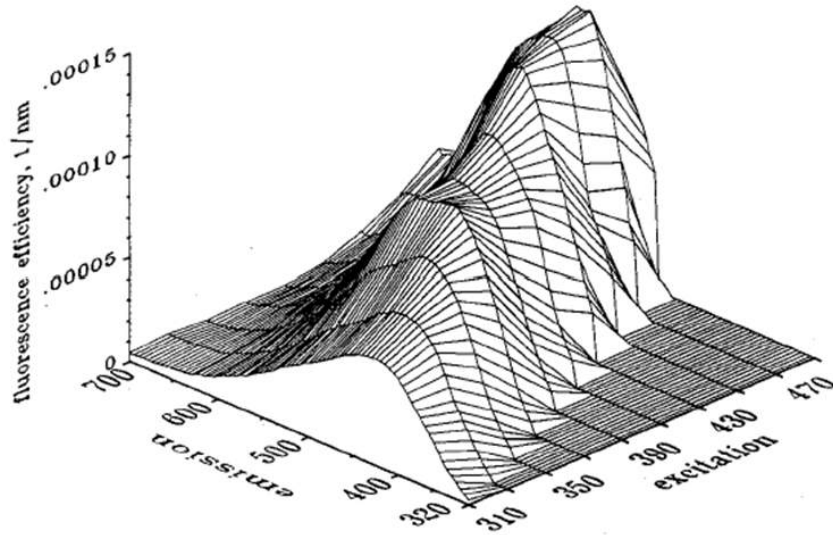


Figure 7.13: The spectral fluorescence quantum efficiency function $\eta_Y(\lambda', \lambda)$ for a sample of water from the Gulf of Mexico (Hawes, 1992, Station FA7).

Hawes was able to fit his measurements to a function of the form (his Eq. 10)

$$\eta_Y(\lambda', \lambda) = A_0(\lambda') \exp \left[- \left(\frac{\frac{1}{\lambda} - \frac{A_1}{\lambda'} - B_1}{0.6 \left(\frac{A_2}{\lambda'} + B_2 \right)} \right)^2 \right]. \quad (7.19)$$

Here $A_0(\lambda')$ has units of nm^{-1} , A_1 and A_2 are dimensionless, and B_1 and B_2 have units of nm^{-1} . The values of these model parameters are determined by a best fit of the model to the measured data. For the data of Fig. 7.13, the best-fit values are shown in Table 7.2.

λ'	A_0
310	5.81×10^{-5}
330	6.34
350	8.00
370	9.89
390	9.39
410	10.48
430	12.59
450	13.48
470	13.61
490	9.24×10^{-5}
A_1	0.470
B_1	8.077×10^{-4}
A_2	0.407
B_2	-4.57×10^{-4}

Table 7.2: Best-fit parameter values of Eq. (7.19) for the data of Fig. 7.13. All values of A_0 are times 10^{-5} . The r^2 value is 0.987. Data from Hawes (1992, Table 3).

The left panel of Fig. 7.14 shows the best-fit $A_0(\lambda')$ values of Table 7.2 as solid dots. The open dots are values extended to other wavelengths for use in HydroLight. The corresponding quantum efficiencies $\Phi_Y(\lambda')$ are obtained by integration of $\eta_Y(\lambda', \lambda)$ over λ :

$$\Phi_Y(\lambda') = \int_{\lambda'}^{\infty} \eta_Y(\lambda', \lambda) d\lambda. \quad (7.20)$$

The right panel of Fig. 7.14 shows the dependence of Φ_Y on the excitation wavelength λ' for the parameter values of Table 7.2 as extended and used in Eq. (7.19). Unlike the quantum efficiency for chlorophyll fluorescence, the quantum efficiency for CDOM fluorescence depends on the excitation wavelength.

Figure 7.15 shows $\eta_Y(\lambda', \lambda)$ as computed by Eq. (7.19) for the parameter values of Table 7.2 as extended.

As seen in Fig. 8.15 of the CDOM section and as modeled by Eq. (7.18), absorption by CDOM continues to rise rapidly in the ultraviolet (UV). As seen in Fig. 7.13, CDOM fluorescence is excited by UV wavelengths even below 300 nm, and CDOM emission occurs at wavelengths from the excitation wavelength into the blue and green. Thus CDOM is optically important both because of its strong absorption at blue and UV wavelengths, and because it can fluoresce at UV to blue and green wavelengths.

As noted on the theory section, the CDOM fluorescence wavelength redistribution function $f_Y(\lambda', \lambda)$ is obtained from $\eta_Y(\lambda', \lambda)$ via

$$f_Y(\lambda', \lambda) = \eta_Y(\lambda', \lambda) \frac{\lambda'}{\lambda} \quad [\text{nm}^{-1}]. \quad (7.21)$$

The excellent Master's Thesis by Hawes remains, three decades later, the one and only publication I can find that presents measurements and a model for *calibrated* spectral fluorescence quantum efficiency functions $\eta_Y(\lambda', \lambda)$ for CDOM (or for any other substance).

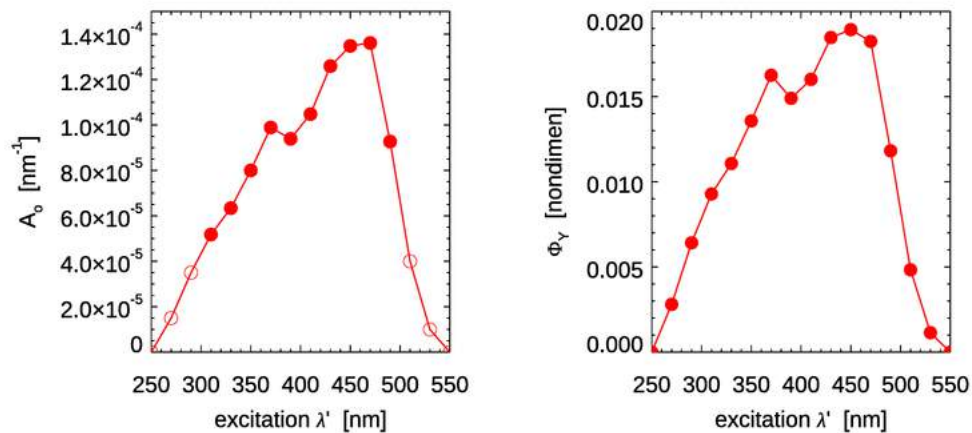


Figure 7.14: Left panel: Values of the fitting parameter $A_0(\lambda')$ as determined for the measured data of Fig. 7.13 (solid dots) and as extended to other excitation wavelengths for use in HydroLight (open circles). Right panel: the corresponding CDOM quantum efficiency $\Phi_Y(\lambda')$ as obtained from Eqs. (7.19) and (7.20).

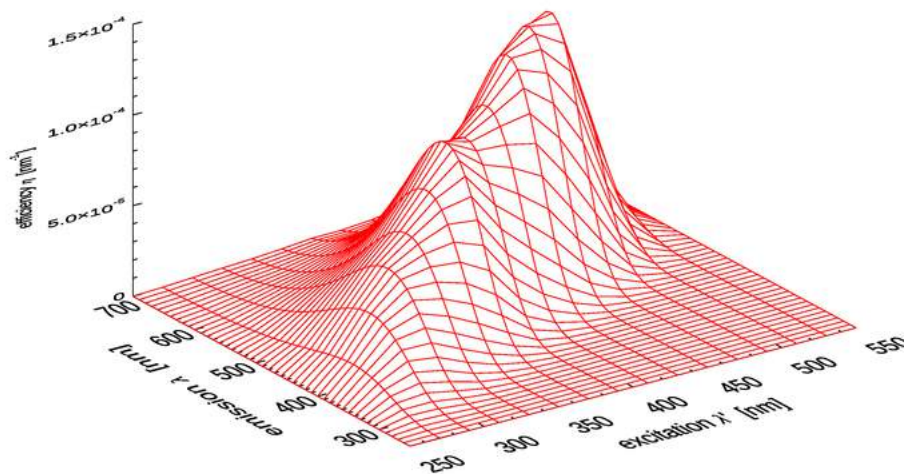


Figure 7.15: The spectral fluorescence quantum efficiency function $\eta_Y(\lambda', \lambda)$ as computed by Eq. (7.19) using the parameter values of Table 7.2 (as extended for A_0).

His functional form (7.19) is used in HydroLight to model CDOM fluorescence. The parameter values seen above are the defaults in HydroLight. Although the results of his thesis research were presented at the Ocean Optics XI conference (Hawes et al., 1992), they were never published in the refereed literature. The thesis itself cannot be found online, but a photocopy can be downloaded from the Web Book.

Figure 7.16 shows excitation-emission functions as commonly seen in the literature. Such measurements are used to identify the presence (or absence) of various types of fluorescing compounds in the water. The original figure does not comment on the measurement units, which are often counts per second or something similar. In any case, these excitation-emission plots display relative values and are not the equivalent of the calibrated spectral fluorescence quantum efficiency functions $\eta_Y(\lambda', \lambda)$ discussed above and seen in Figs. 7.13 and 7.15.

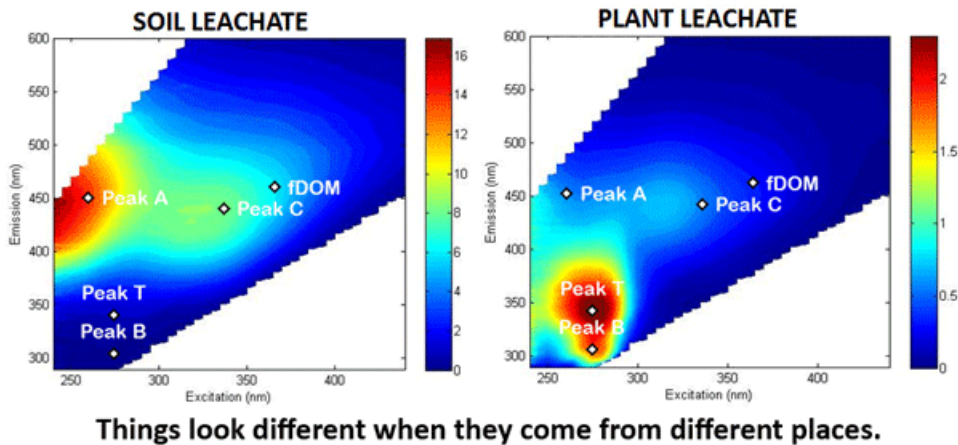


Figure 7.16: Two fluorescence excitation-emission functions from different source materials. Left, a peat soil leachate. Peaks A and C are the highest intensity, and represent high concentrations of humic acid, a primary organic component of soil. Right, a fresh plant leachate. Peaks T and B are the highest and represent high concentrations of fresh dissolved organic compounds like proteins and lignins that fluoresce in this region (Credit: Angela Hansen, U.S. Geological Survey. Public domain image from [USGS](#).)

7.5.3 The CDOM Fluorescence Phase Function

As previously noted, fluorescence emission is isotropic. Therefore the phase function is simply

$$\tilde{\beta}_Y(\psi) = \frac{1}{4\pi} \quad [\text{sr}^{-1}].$$

The models seen above give everything needed to construct the volume inelastic scattering function of Eq. (7.17) for CDOM fluorescence, $\beta_Y(z, \psi, \lambda', \lambda)$, which is then ready for use in the radiative transfer equation.

7.6 Examples of Inelastic Scattering Effects

Now that the general theory of Raman scattering and of chlorophyll and CDOM fluorescence has been developed, the inelastic scattering VSFs can be used in radiative transfer codes like HydroLight to predict the effects of various inelastic scattering processes on radiances, irradiances, reflectances, K -functions, and the like. This section presents a number of HydroLight-generated examples of light fields computed with and without inelastic scattering effects included in the simulations.

7.6.1 Effect of Raman Scattering on R_{rs}

Figure 7.17 shows the effect of Raman scattered radiance on the remote-sensing reflectance for chlorophyll values of $Chl = 0.02$ and 2 mg m^{-3} as simulated by HydroLight. The HydroLight runs used a bio-optical model for Case 1 water for which the absorption and scattering properties of the water are determined only by the chlorophyll concentration. The water was homogeneous and infinitely deep. The Sun was at a 30 deg zenith angle in a clear sky; the wind speed was 5 m s^{-1} . Identical runs were made with and without Raman scattering included in the runs. It is seen that for the very clear water with $Chl = 0.02$, R_{rs} is as much as 22% higher, but for the water with $Chl = 2$, Raman increases R_{rs} by at most 5%. The Raman effect decreases but still can be significant in higher chlorophyll waters or in turbid Case 2 waters. Although difficult to see in Fig. 7.17, the Raman effect does not “turn on” until about 340 nm even though the HydroLight run started at 300 nm. This is because the start of the emission band for excitation near 300 nm is around 340 nm (recall Fig. 7.4).

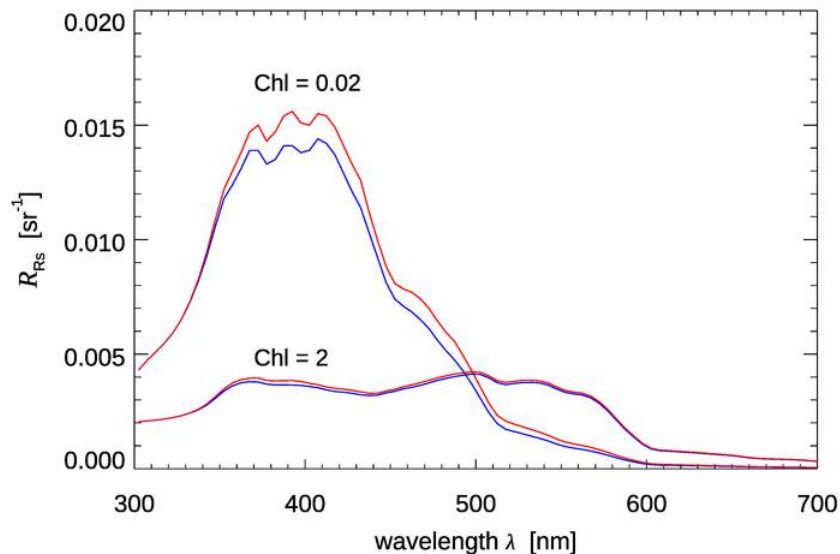


Figure 7.17: Effect of Raman scattering on remote-sensing reflectance R_{rs} . The red curves included Raman scattering and the blue curves did not.

7.6.2 Effect of Raman Scattering on Upwelling Plane Irradiance

Although Raman scattering does not have a large effect on R_{rs} except in the clearest water, it can be the dominant source of light at red wavelengths at depths where absorption by water has removed most of the incoming sunlight. This is illustrated in Fig. 7.18. HydroLight was run for Case 1 water with a chlorophyll concentration of $Chl = 0.5 \text{ mg m}^{-3}$, which is typical of open ocean water. The Sun was at a 30 deg zenith angle in a clear sky. The run started at 300 nm, so that Raman effects would be present at wavelengths greater than 340 nm. At 400 and 500 nm, the contribution by Raman scattered light to the upwelling irradiance is almost unnoticeable. This is because at these wavelengths

elastically scattered solar radiance is the main contributor to the upwelling irradiance. At 580 nm, water absorption ($a_w(580) = 0.09 \text{ m}^{-1}$) is beginning to filter out enough of the solar radiance that the Raman contribution, which comes from wavelengths around 485 nm (see Fig.7.4), where the light penetrates well to depth, is becoming the main contribution to $E_u(z, 580)$. At 600 nm, water absorption ($a_w(600) = 0.22 \text{ m}^{-1}$) has removed almost all of the solar light below about 20 m. Below 20 m, almost all of $E_u(z, 600)$ comes from Raman scattered light that originates from wavelengths around 500 nm, where sunlight penetrates well to depth. It should be noted that below about 40 m, the depth rate of decay of $E_u(z, 600)$ (i.e., $K_u(z, 600)$) is almost the same as the rate of decay the irradiance at 500 nm. This is because the light at 500 nm is the source of the light at 600 nm.

It was the measurement of unexpected upwelling irradiance $E_u(z, \lambda)$ at depths below 50 m and wavelengths greater than 520 nm that led Sugihara et al. (1984) to suggest that the unexpected upwelling irradiance came from downwelling irradiance at blue-green wavelengths being Raman scattered. A number of subsequent studies (e.g. Stavin and Weidemann, 1988; Marshall and Smith, 1990) confirmed this hypothesis. Many studies since have studied Raman effects on ocean light fields. For example, Kattawar and Xu (1992) and Ge et al. (1993) studied the filling in of Fraunhofer lines in underwater light by Raman scattering.

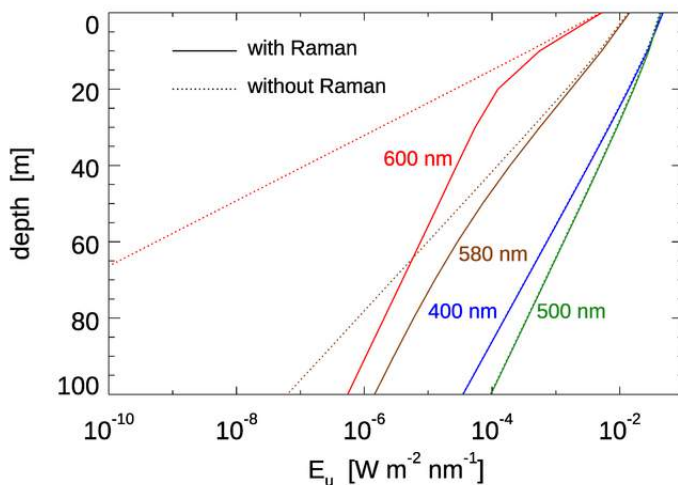


Figure 7.18: HydroLight simulations of $E_u(z, \lambda)$ with and without Raman scattering.

7.6.3 Effect of Chlorophyll Fluorescence on Downwelling Plane Irradiance

This section uses HydroLight (Section 10.6) to illustrate the effects of various chlorophyll fluorescence input parameters.

To see the effect of the shape of the chlorophyll emission function, a series of four HydroLight runs was done with the following inputs:

- A chlorophyll concentration of $Chl = 10 \text{ mg Chl m}^{-3}$ for Case 1 water (using the New Case 1 IOP model of Section 8.9 in HydroLight); the water was homogeneous and infinitely deep.
- A chlorophyll quantum efficiency of $\Phi_C = 0.06$.
- A chlorophyll emission function given by either Eq. (7.14) or (7.16) .
- Sun at a zenith angle of 30 deg in a clear sky, wind speed of 5 m s^{-1} .
- The run was from 400 to 750 nm by 5 nm.
- Output was saved at 5 m intervals from 0 to 50 m.
- Four sets of inelastic effects were simulated: (1) no inelastic effects at all, (2) Raman scatter only, (3) Raman scatter plus chlorophyll fluorescence with a single Gaussian emission function, and (4) Raman scatter plus chlorophyll fluorescence with a double Gaussian emission function.

Figure 7.19 shows the the two Gaussian chlorophyll emission functions of Eqs. (7.14) and (7.16) (upper left panel); the resulting remote-sensing reflectance R_{rs} in the region of the chlorophyll emission (upper right panel); and the depth profiles of downwelling plane irradiance at 710 nm, $E_d(710)$ (lower left); and at 730 nm, $E_d(730)$ (lower right).

Some of the features to note in Fig. 7.19 are as follows:

- The peak R_{rs} values near 685 nm are larger for the single Gaussian than for the double Gaussian. This is because both emission functions, when integrated as in Eq. (7.15), correspond to the same quantum efficiency $\Phi_C = 0.06$. Thus, as seen in the upper left panel, the 685 peak of the double Gaussian is lower than for the single Gaussian because part of the energy is going into the second Gaussian centered at 730 nm.
- There is no fluorescence contribution to R_{rs} for the single Gaussian beyond about 720 nm, but the double Gaussian gives a noticeable increase in R_{rs} even beyond 750 nm. This corresponds to the magnitudes of the two emission functions. However, magnitude of R_{rs} is quite small in the near infrared relative to the peak emission values and R_{rs} at shorter wavelengths (not shown) even for the high chlorophyll value of $Chl = 10 \text{ mg Chl m}^{-3}$ and the high efficiency of $\Phi_C = 0.06$ used here because of the high absorption by water itself beyond 700 nm; water absorption at 720 nm is $a_w(720) = 1.17 \text{ m}^{-1}$ and rises to 2.47 m^{-1} at 750 nm.
- The $E_d(z, 710)$ profiles are similar down to about 5 m. For shallower depths, solar radiance at 710 nm penetrates the water column well enough to dominate the value of $E_d(z, 710)$. Inelastic scatter contributions to the near-surface light field are minimal at 710 nm for these IOPs.
- Below about 10 m, the simulation without any inelastic scattering is greatly different from the curves with Raman or Raman plus fluorescence. Essentially the only light at 710 nm at depths below about 15 m comes from light at blue and green wavelengths, which do penetrate to depths below 15 m, that is inelastically scattered into 710 nm. As expected from the shapes of the emission functions, the double Gaussian “injects” more light into 710 nm than does the single Gaussian.

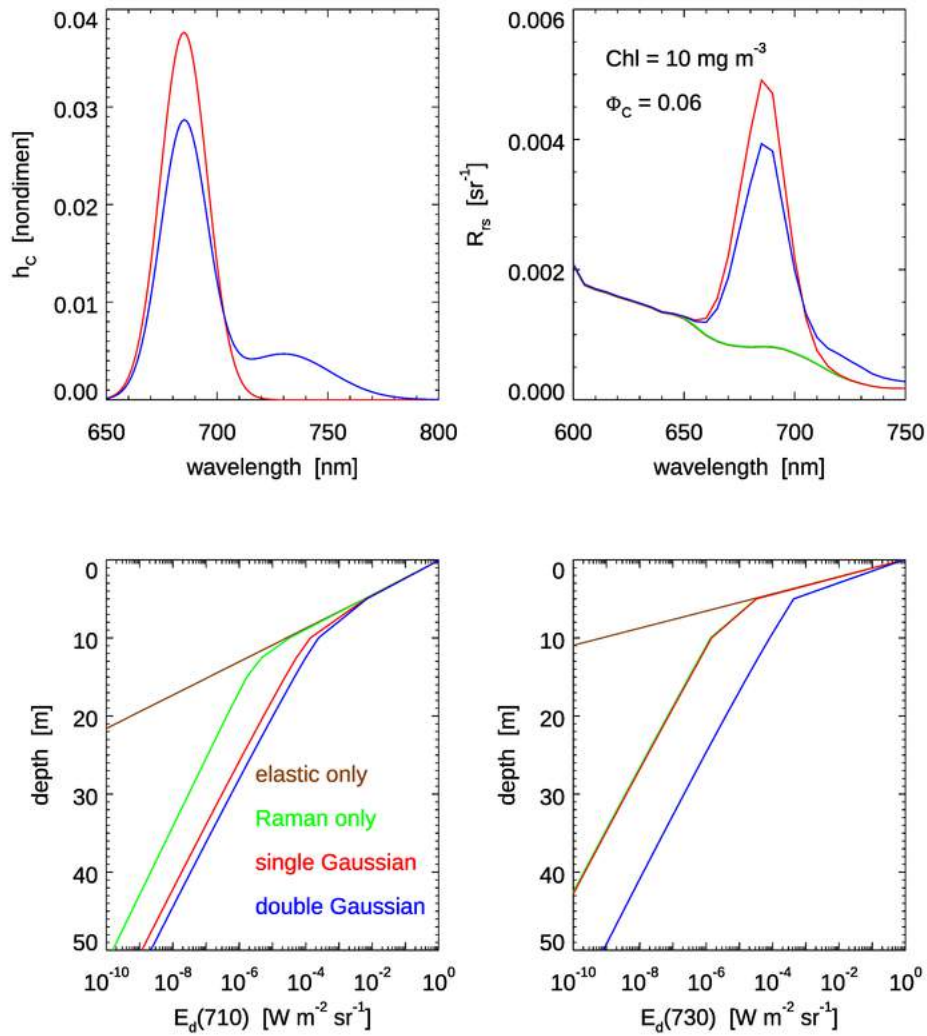


Figure 7.19: HydroLight-simulated fluorescence effects for Case 1 water with a high chlorophyll concentration of $Chl = 10 \text{ mg Chl m}^{-3}$ and a high quantum efficiency for chlorophyll fluorescence of $\Phi_C = 0.06$. See the text for discussion.

- At 730 nm, the Raman-only and Raman-plus-single-Gaussian emission are essentially identical because the single Gaussian is almost zero at 730. However, the double-Gaussian emission function still adds a significant amount of light into the deep water column.

In summary, for wavelengths greater than about 700 nm there is a significant fractional difference in R_{rs} and in the irradiances at depth for the two chlorophyll emission functions. However, these differences are likely to be unimportant for practical oceanographic problems. It is hard to imagine applications where accurate predictions of irradiances are required in the near infrared at large depths.

7.6.4 Effect of Chlorophyll Fluorescence Quantum Efficiency on R_{rs}

Figure 7.20 shows HydroLight simulations of R_{rs} in the chlorophyll fluorescence emission region for a value of $Chl = 0.5 \text{ mg Chl m}^{-3}$, typical of open-ocean waters, and for three values of the quantum efficiency. The emission function is the double Gaussian. Other run inputs were the same as for Fig. 7.19. These curves include both Raman scatter and chlorophyll fluorescence. Curves for Raman only, and for elastic scatter only are also shown. Relative to the baseline of Raman only, the chlorophyll fluorescence curves are in direction proportion to the quantum efficiency values, all else being the same, as should be expected.

On the other hand, the height of the “Raman corrected” peak of the $\Phi_C = 0.06$, double-Gaussian curve of the upper right panel of Fig. 7.19 is only about 6 times the height of the corresponding curve in Fig. 7.20, even though the chlorophyll concentration is 20 times higher for Fig. 7.19. This is because absorption and scattering do not depend linearly on the chlorophyll concentration. For an order-of-magnitude understanding of this, note that it is absorption that removes light that might otherwise contribute to R_{rs} . The new Case 1 IOP model used for these runs models phytoplankton absorption by a formula that depends on $Chl^{E(\lambda)}$. In the 400-680 nm range relevant to the chlorophyll fluorescence, $E(\lambda)$ ranges from 0.6 to 0.8. For a difference in chlorophyll values of 20, $20^{0.7} \approx 0.8$, which is close to the differences in the heights of the emission peaks. Absorption by CDOM will further reduce the light available for fluorescence.

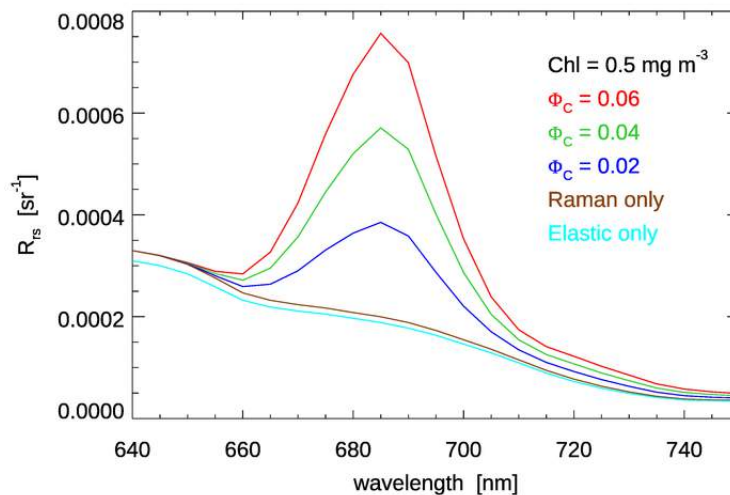


Figure 7.20: Effect of chlorophyll quantum efficiency Φ_C on the chlorophyll emission band for a chlorophyll concentration of $Chl = 0.5 \text{ mg Chl m}^{-3}$.

7.6.5 Effect of CDOM Fluorescence on R_{rs}

To see the effect of the CDOM fluorescence on the remote sensing reflectance R_{rs} , a series of HydroLight runs was done with the following inputs:

- A chlorophyll concentration of $Chl = 0.5 \text{ mg Chl m}^{-3}$ for Case 1 water (using the new Case 1 IOP model in HydroLight); the water was homogeneous and infinitely deep.
- Low and high values of CDOM absorption was included as either .
 - CDOM absorption at 440 nm was 20% of the chlorophyll absorption at 440 nm, i.e. $a_Y(440) = 0.2a_C(440)$. This is a common model for Case 1 water.
 - CDOM absorption at 440 nm was 5 times the chlorophyll absorption at 440 nm, i.e. $a_Y(440) = 5.0a_C(440)$. This represents a very a high concentration of extra CDOM as could result from river input into coastal water or by the decay of benthos such as sea grass, which gives Case II water.
- CDOM fluorescence was modeled using the parameter values shown above for the Hawes Station FA7.
- Sun at a zenith angle of 30 deg in a clear sky, wind speed of 5 m s^{-1} .
- The run was from 300 to 750 nm by 5 nm.
- Four sets of inelastic effects were simulated: (1) no inelastic effects at all, (2) Raman scatter only, (3) CDOM fluorescence only, and (4) Raman scatter plus CDOM and chlorophyll fluorescence.

Figure 7.21 shows the results of these simulations. The top panel is for the Case 1 water with $a_Y(440) = 0.2a_C(440)$. The bottom panel is for the high-CDOM Case 2 water with $a_Y(440) = 5.0a_C(440)$. For the low-CDOM, Case 1 water, the simulation with CDOM fluorescence increases R_{rs} by less than 2% over the elastic-only case. Raman scatter gives up to an 8% increase over elastic only. For the high-CDOM Case 2 water, Raman gives up to a 7% increase, but CDOM fluorescence gives up to a 27% increase in R_{rs} . Thus, it is possible for CDOM fluorescence to have a significant effect on R_{rs} , but it takes a very high CDOM concentration to do so. In low-to-medium CDOM waters, typical of the open ocean where CDOM covaries with chlorophyll, CDOM fluorescence affects R_{rs} by at most a few percent. If these R_{rs} spectra are used in band-ratio algorithms for retrieval of environmental variables such as the chlorophyll concentration, the CDOM-fluorescence enhancement to R_{rs} should have even less affect on the retrieved values. This conclusion is consistent with the findings in Hawes (1992). Note also that CDOM fluorescence has a minimal effect in the chlorophyll-fluorescence band near 685 nm.

7.6.6 Effect of Inelastic Scattering on K Functions

HydroLight was next used to simulate a homogeneous Case 1 water body with a chlorophyll concentration of $Chl = 1 \text{ mg m}^{-3}$. As for Figs. 4.3 and 4.6, the Sun was at 40 deg and the surface was level. Figure 7.22 plots several quantities as a function of wavelength at 10 m depth for this simulation. We see that between 300 and about 600 nm, the various $K(10 \text{ m}, \lambda)$ functions are very similar and proportional to the total (including water) absorption coefficient $a(10 \text{ m}, \lambda)$. However, beyond 600 nm the K functions differ from each other, and they are all much different from a . The reason for this behavior is inelastic scattering.

At the near-UV to blue to green wavelengths below 600 nm, most of the radiance at 10 m depth (for these IOPs) comes from sunlight being transmitted through the upper 10

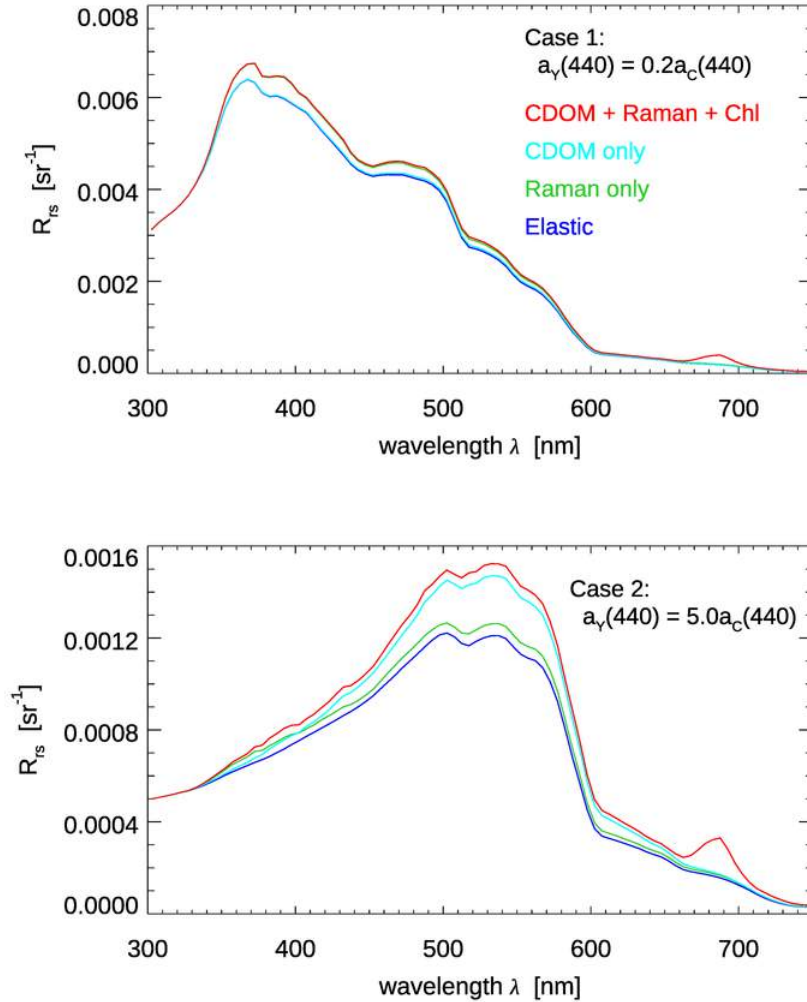


Figure 7.21: Effects of CDOM fluorescence on the remote-sensing reflectance R_{rs} . Top panel: low-CDOM water; bottom panel: high-CDOM water. See text for discussion.

m of the water column. Above about 600 nm, absorption by the water itself has removed most of the sunlight. For example, at 700 nm where $a_w = 0.65 \text{ m}^{-1}$, we expect roughly $\exp(-az) \approx 0.001$ of the surface light to reach 10 m. However, Raman scatter and CDOM fluorescence inelastically scatter light from shorter wavelengths, where sunlight is present, into the red wavelengths, and thus create additional red light at 10 m. Thus, beyond 600 nm, the various radiances and irradiances no longer decrease with depth in a simple exponential fashion. Note that K_d tracks a longer than do K_u and K_{Lu} . This is because E_d continues to collect whatever downwelling sunlight remains, and thus the inelastic contribution to K_d is not noticeable until the chlorophyll fluorescence contribution begins near 670 nm. K_u and K_{Lu} on the other hand have only a small amount of backscattered sunlight, so that the inelastic contribution becomes significant sooner, at around 600 nm.

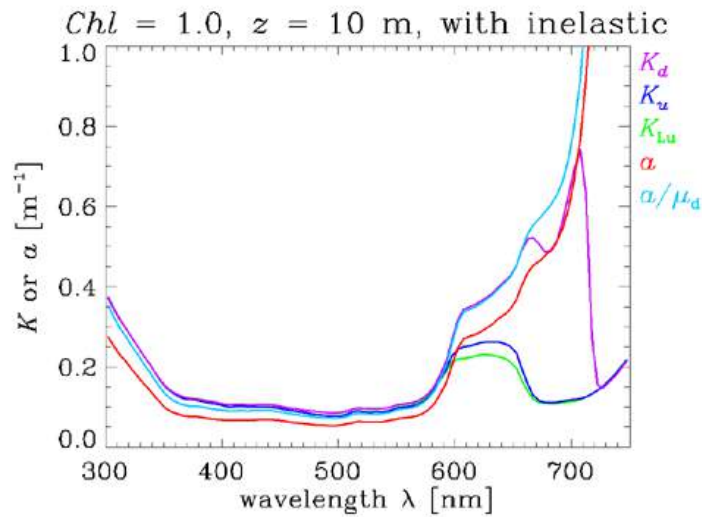


Figure 7.22: K at 10 m depth for Case 1 water with $Chl = 1 \text{ mg m}^{-3}$. The HydroLight run included Raman scattering by the water, and chlorophyll and CDOM fluorescence.

It is easy to verify that the peculiar behavior of the K functions beyond 600 nm is due to inelastic scatter. The HydroLight run was repeated with Raman scatter and CDOM and chlorophyll fluorescence “turned off.” Figure 7.23 shows the results. Now, the K functions all track the absorption nicely at all wavelengths.

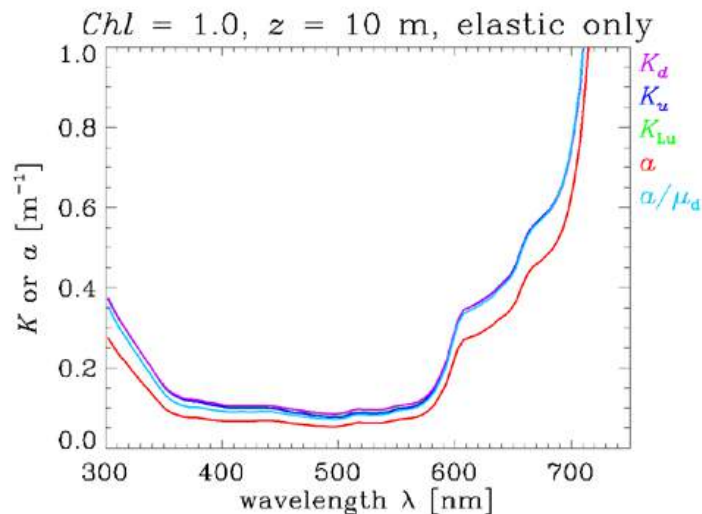


Figure 7.23: K at 10 m depth for Case 1 water with $Chl = 1 \text{ mg m}^{-3}$ as in Fig. 7.22, except that inelastic scattering was turned off in the HydroLight run.

The behavior seen in Fig. 7.22 is not just of academic interest. Figure 7.24 shows absorption coefficient spectra $a(\lambda)$ at selected depths as measured by an ac-9 instrument

and as determined by Gershun's law, Eq. (10.33),

$$a(z, \lambda) = -\frac{1}{E_o(z, \lambda)} \frac{d[E_d(z, \lambda) - E_u(z, \lambda)]}{dz}, \quad (7.22)$$

using measurements of the plane and scalar irradiances. The figure shows that for wavelengths less than about 580 nm, the directly measured and computed absorption spectra are in good agreement. For wavelengths longer than 580 nm, the irradiances are increasingly affected by inelastic scatter from shorter wavelengths as the depth increases, and the retrieved absorption spectra are much less than the measured spectra. This figure and the simulations of Figs. 7.22 and 7.23 show that care must be taken when interpreting or using in-water radiance or irradiance measurements at wavelengths longer than about 580 nm, when inelastic effects can become dominant at depth.

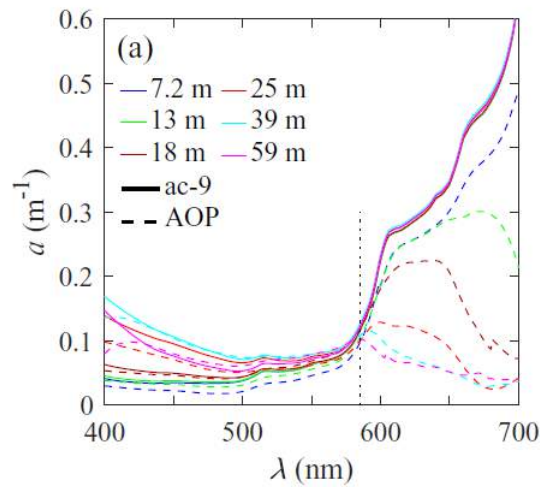


Figure 7.24: Absorption coefficient spectra $a(\lambda)$ at selected depths as measured by an ac-9 instrument and as determined by Gershun's law using measured irradiances. Figure 14(a) from Li et al. (2018), reproduced under the *Applied Sciences* Creative Commons license.

Optical Constituents of the Ocean

Now that we have laid the foundations of radiometry, defined IOPs and AOPs, and examined the processes of absorption and scattering, this chapter surveys the most important constituents of the ocean that determine its optical properties. The first few sections discuss the optical properties of specific constituents, beginning with water itself. Then commonly used models for the absorption and scattering as functions of the chlorophyll, colored dissolved organic matter, and mineral concentrations of the water body are presented. The chapter then finishes with a discussion of particle size distributions, which are a non-optical physical property of particles that is very important in determining the scattering properties of a water body.

8.1 Classifying the Optically Important Marine Particles

[Collin Roesler, Emmanuel Boss, and Curtis Mobley contributed to this section.]

The in-water constituents that affect aquatic optical properties are traditionally lumped into the following categories:

- Sea water (water + inorganic dissolved materials)
- Colored (or chromophoric) dissolved organic material (CDOM)
- Phytoplankton
- Non-phytoplankton organic particles
- Inorganic particles
- Bubbles

The constituents are identified operationally based upon how we measure their optical properties and are often are grouped by like optical properties. For example, the distinction between particulate and dissolved is operationally defined by the filter type or pore size, often a filter pore size of $0.2 \mu\text{m}$. It is essential to remember that (1) the strict chemical definition is quite different, (2) filter pore size varies from author to author, and (3) it is important to keep track of filter pore sizes to ensure closure (i.e., don't define dissolved

organic matter by the filtrate of a $0.2\ \mu\text{m}$ pore size filter and then measure particulates on a $0.7\ \mu\text{m}$ pore size GF/F filter). Similarly, often all of the non-phytoplankton particles are lumped into a single compartment because their optical properties can be quite similar. Sometimes all of the particulate material is lumped together into suspended particulate material (SPM), or part of it into particulate organic material (POM). This is often done when studying a specific bio-geochemical property using optics.

The differentiation between dissolved and particulate materials (using a filter) does not imply that the dissolved material is organic, though this is most often the assumption. For example, inorganic dissolved substances such as iron oxides (rust) could contribute in certain cases.

A more detailed classification of optically important particles is as follows:

- **Biogeochemical classification**

Particulate Organic Material (POC): Includes all living or once-living particulate matter such as phytoplankton and detritus.

Particulate Inorganic Material (PIC): Includes all lithogenic and mineralic particulate matter such as sand and calcium carbonate shells, also called tripton.

Particulate Material (PM): The total organic and inorganic, animate and inanimate particles, also called seston.

- **Biological classification**

Algae: Aquatic photosynthetic organisms that can range in size from microscopic (phytoplankton) to tens of meters (giant kelp)

Phytoplankton: Microscopic, single-celled (some form colonies or chains), free-floating, aerobic (live in an oxygenated environment), oxygenic (produce oxygen during photosynthesis) organisms that possess chlorophyll-a and photosynthesize. Some may be mixotrophic (photosynthetic and heterotrophic). Includes prokaryotes (cyanobacteria) and eukaryotes (diatoms, dinoflagellates, coccolithophorids, and many others).

Bacterioplankton: Microscopic, single-celled (some may form colonies or chains) prokaryotic (no membrane-bound organelles) organisms, can be phototrophic (cyanobacteria) or heterotrophic.

Zooplankton: Single- or multi-cellular free-floating heterotrophic organisms that may consume photosynthetic organism or other heterotrophic organisms. They range in size from microscopic to macroscopic (visible by eye). While some are locomotive, they are advected in ocean currents and therefore planktonic.

Detritus Non-living particulate organic matter, including dead bacterial, phytoplankton and zooplankton cells, fragments of cells left from zooplankton grazing, fecal pellets, shells, and marine snow aggregates.

- **Size classification**

Biological:

Picoplankton: The fraction of plankton between $0.2\ \mu\text{m}$ and $2\ \mu\text{m}$. Typically phytoplankton.

Nanoplankton: The fraction of plankton between $2\ \mu\text{m}$ and $20\ \mu\text{m}$. Typically phytoplankton and single-celled zooplankton.

Microplankton: The fraction of plankton between $20\ \mu\text{m}$ and $200\ \mu\text{m}$ ($0.2\ \text{mm}$). Typically phytoplankton and single-celled zooplankton.

Mesoplankton: The fraction of plankton between $0.2\ \text{mm}$ and $20\ \text{mm}$. Typically multicellular zooplankton, may include large single celled zooplankton and phytoplankton.

Geological:

Clay: The fraction of typically lithogenic or mineralic particles that are less than $2\ \mu\text{m}$ ($0.002\ \text{mm}$).

Silt: The fraction of typically lithogenic or mineralic particles between $2\ \mu\text{m}$ and $63\ \mu\text{m}$ ($0.002\ \text{mm}$ and $0.063\ \text{mm}$). Includes subdivisions fine, medium and coarse.

Sand: The fraction of typically lithogenic or mineralic particles between $0.063\ \text{mm}$ ($63\ \mu\text{m}$) and $2\ \text{mm}$. Includes subdivisions fine, medium and coarse.

- **Operational Classification**

Particulate: All matter captured by filtration onto a filter of defined nominal pore size. The two most common pore sizes used to identify particles are $0.2\ \mu\text{m}$ (e.g., nucleopore filters) or $0.7\ \mu\text{m}$ (e.g., glass fiber filters).

Dissolved: All matter that passes through the filter of identified nominal pore size, i.e., filtrate. May include some particulate matter such as very small picoplankton, viruses, and colloids. Optically classified as colored dissolved organic matter (CDOM). Chemists classify colloids as particles in the $1\ \text{nm}$ to $1\ \mu\text{m}$ size range. Marine viruses are typically in the $20\text{-}300\ \text{nm}$ range.

Phytoplankton: All particulate matter pigmented by phytoplankton-derived pigments, including detrital pigments. Defined by extractive methods using solvents such as acetone and methanol for most pigments and phosphate buffers for phycobilipigments.

Non-algal particles (NAP): The fraction of particulate matter that is not extractable (not a phytoplankton pigment). Includes all living and detrital organic matter such as the non-pigmented portion of phytoplankton cells, detritus, heterotrophic bacteria, and viruses. May also include inorganic mineral particles of both biogenic (e.g., calcite liths and shells) and terrestrial origin (e.g., clay, silt, and sand).

Particulate Inorganic Matter: The fraction of particulate matter that is not organic and therefore remains on the filter after combustion, also called dry ash.

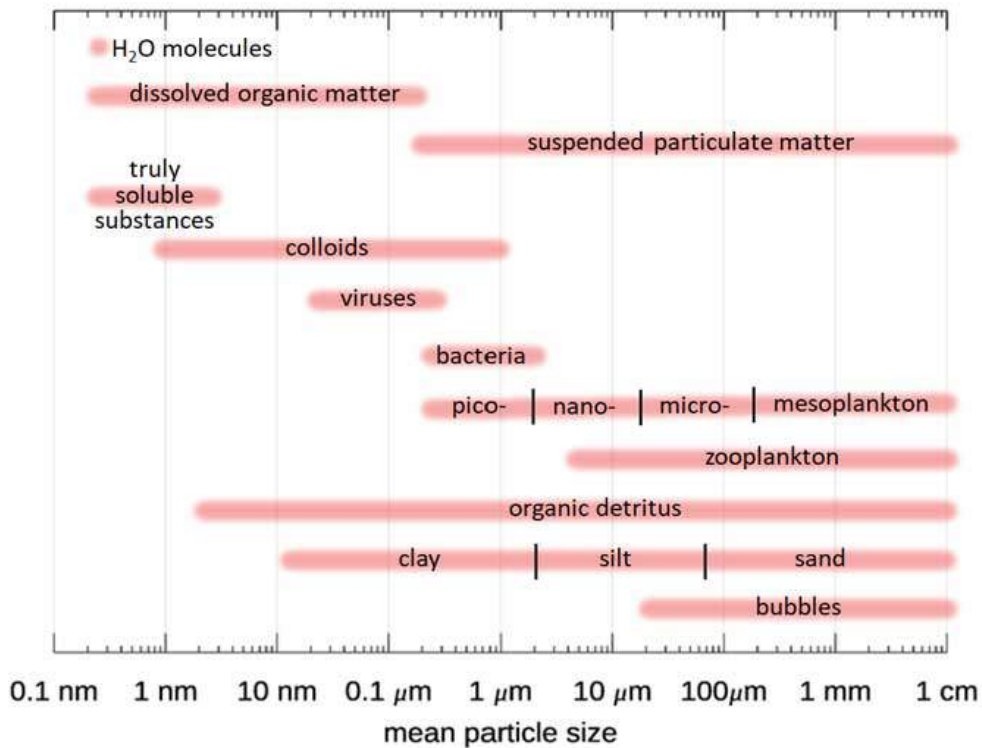


Figure 8.1: Approximate size ranges for different components of natural waters. Redrawn from [Stramski et al. \(2004a\)](#), with some additions.

Figure 8.1 shows the size ranges commonly associated with the different components of natural waters. Figure 8.2 shows how many of different types of biological particles typically occur in marine waters, expressed as a particle size distribution (PSD) $n(D)$. D is the diameter of an equivalent-volume sphere, and $n(D)$ gives the number of particles per cubic meter, per unit of size range. It must be remembered that such a size distribution may be a good average, but may be quite incorrect for a given time and location. For example, if a phytoplankton bloom is underway, there will be many more particles of a certain size range, which will put a “bump” in the PSD. In open ocean water, there may be almost no mineral particles. In coastal waters, river inputs or sediment resuspension may make minerals the optically dominant particles. The same can be true in open-ocean water if coccolithophores are shedding their calcite liths. The blue line shows a power law size distribution with an exponent of -4, i.e., $n(D) \sim D^{-S}$ with $S = 4$. A power law distribution with an exponent of -4 is also known as a Junge distribution. Figure 8.3 shows three sets of PSDs measured in coastal waters. The red lines are Junge size distributions. It is seen that the Junge distribution gives a reasonably good fit over the measured size range, but there are significant deviations from a power law distribution for some of the measurements, and in many cases a power law with a different exponent than -4 gives a better fit. Particle size distributions are discussed in detail in Section 8.12.

As just seen from their many ways of classification, marine particles are extremely complex and varied in their composition, pigmentation, shape, internal structure, and

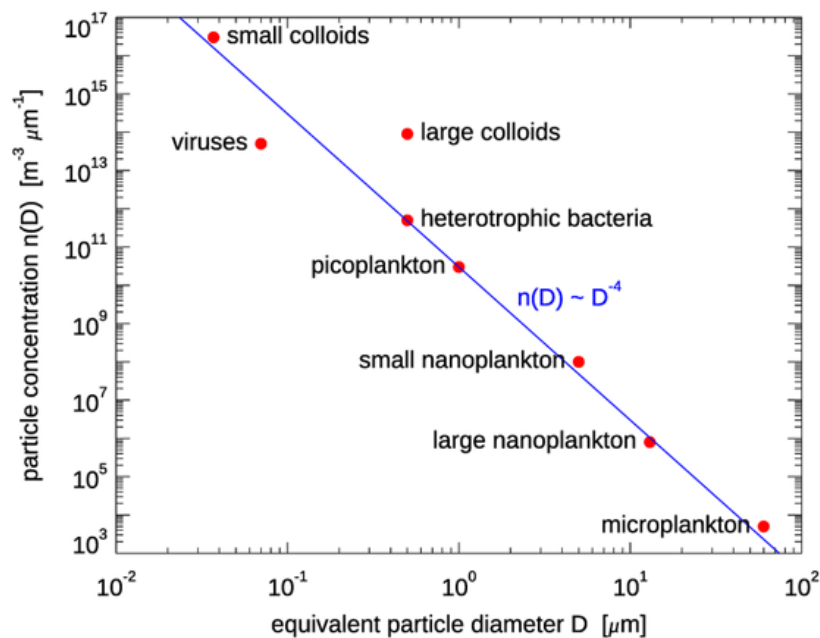


Figure 8.2: Approximate size distribution of biological particles in the ocean. The blue line is a power law distribution with an exponent of -4 centered on the picoplankton point. Data courtesy of D. Stramski.

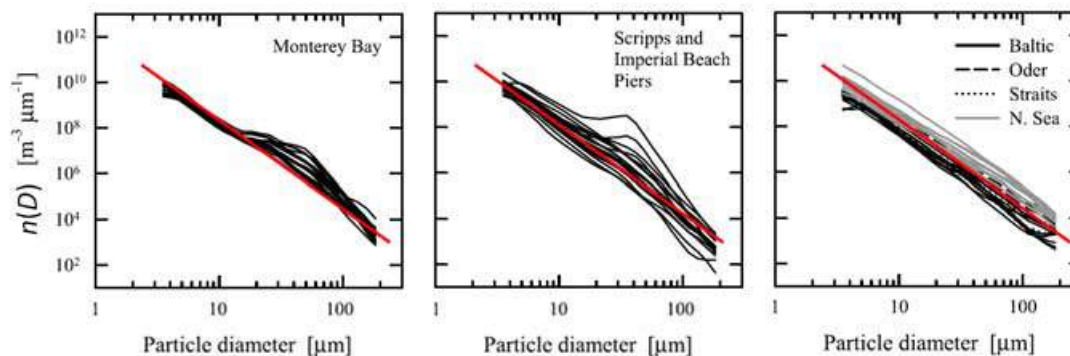


Figure 8.3: Measured particle size distributions in coastal waters. The red lines are Junge size distributions, $n(D) \sim D^{-4}$. Redrawn from Reynolds et al. (2010, with modifications).

packaging. In order to understand their interaction with light some idealizations have to be made. Because it is impractical to study or model each individual particle (there are more than a billion bacteria in one milliliter of seawater), particles are lumped into groups of particles having similar properties.

In order to model the optical properties of particles (that is, to derive analytical or numerical descriptions of these properties) we need empirical data providing the necessary inputs (size, shape, index of refraction, internal structure, and packaging). Some of these

data can be obtained from microscopy while others need to be deduced from other measurements. For example, to obtain information on the index of refraction of a sediment grain one could immerse it in oils of different known indexes and microscopically observe when the least optical contrast (scattering) is observed. Most often an inverse optical approach is used. That is, an optical model (such as Mie theory) is used to fit empirical data, and the values of the index of refraction that provide the best match with the observations are chosen to be those of the suspension.

Aas (1996, Table 1) provides a compilation for data on the real part of index of refraction of marine particles and the different methodologies used to obtain them. He also shows how the estimates for phytoplankton are consistent with a model of the index of refraction that is based on their composition. He found that the bulk real index of refraction is related to the volume fraction of water in the phytoplankton cell, V_w ; $V_w = 0.8$ means that the cell is 80% water and 20% other substances (proteins, carbohydrates, fats, pigments). His formula Aas (1996, Eq. 48),

$$n \approx 1.533 - 0.194V_w \text{ (relative to air) ,} \quad (8.1)$$

gives $n \approx 1.38 \pm 0.02$ for $V_w = 0.8 \pm 0.1$ (1.03 ± 0.02 relative to water), or 1.42 ± 0.04 (1.06 ± 0.04 relative to water) for $V_w = 0.6 \pm 0.2$. These indices are consistent with those found by others.

Specific models for the optical properties of different types particles are provided in their own sections. They vary in methodology used to obtain them; some are based strictly on observations, either in the lab or in the field. Others are based on numerical calculations (e.g. using Mie theory) with observation-based inputs. The latter are particularly useful when observations are lacking due to the complexity associated with the measurements (e.g. the volume scattering function). It is particularly satisfying when independent observations and models agree (referred to as optical closure, Section 10.9) as it provides mutual validation for the approaches.

When reading the bio-optical literature it is important to keep in mind what kind of particle is being considered. For example, Bricaud et al. (1995) develops models for the absorption by phytoplankton, a_{phi} , as a function of the chlorophyll concentration; and Bricaud et al. (1998) develops models for the absorption by particles, a_p , where particles refers to phytoplankton plus co-varying detritus.

8.2 Water

Water contributes to both absorption and scattering in natural waters. In clear ocean waters, the effect of water on ocean color in the visible cannot be neglected and must be taken into account. In the near infrared, water almost always dominates absorption. Temperature and salinity affect both absorption and scattering by water and hence need to be taken into account when the optical properties of water are computed.

8.2.1 The Index of Refraction of Water

In the study of light propagation through water using Maxwell's equations (Section 11.3), the index of refraction is usually written as a complex number,

$$m = n + ik ,$$

where n is the real part and k is the imaginary part. Both n and k are positive numbers, and both depend on wavelength or frequency. Their magnitudes and wavelength dependence depend on the material.

It is not obvious at the moment why the real the index of refraction is written as a complex number. As was discussed in the Physics of Scattering, Section 6.2, spatial variation in the real index of refraction n governs the elastic scattering of light. As will be seen in Section 11.3, the imaginary index of refraction k determines the absorption of energy in propagating light, which in oceanography is usually expressed via the absorption coefficient a of the material. These are related by (Eq. 11.43 or Bohren and Huffman (1983, Eq. 2.52))

$$a = \frac{4\pi k}{\lambda}.$$

Using a complex index of refraction will be found in Section 11.3 to be convenient when solving Maxwell's Equations for a plane wave propagating in a dielectric, where the relation between a and k will be derived. It might seem that absorption (i.e., the complex index of refraction) and scattering (i.e., the real index of refraction) should be entirely independent phenomena. However, it will be seen in Section 11.6 that absorption and scattering are intimately related.

8.2.1.1 The real part of the index of refraction

Figure 8.4 shows the real index of refraction of pure water over the wavelength range of 10 nm to 10 m. For wavelengths less than 10 nm the value remains at 1, and for wavelengths longer than 10 m the value remains at about 8.85.

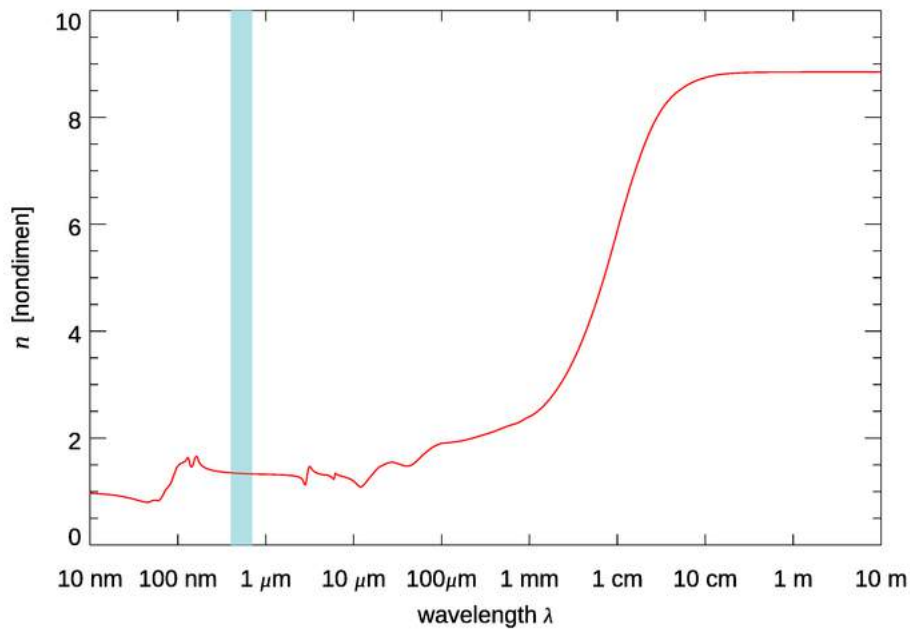


Figure 8.4: The real index of refraction of pure water at room temperature. The blue shading is the visible wavelengths. Data from Segelstein (1981).

Figure 8.4 is for pure water at room temperature. There are small temperature and salinity effects in the visible region, which are not seen in Fig. 8.4. In the 200 to 1100 nm region of interest in optical oceanography, the real part of the index of refraction of water relative to air, n , as given by Quan and Fry (1995) is

$$n(S, T, \lambda) = n_0 + (n_1 + n_2T + n_3T^2)S + n_4T^2 + \frac{n_5 + n_6S + n_7T}{\lambda} + \frac{n_8}{\lambda^2} + \frac{n_9}{\lambda^3}, \quad (8.2)$$

where T is temperature in Celsius, S is salinity in PSU, and λ is the wavelength in nanometers. [PSU is equivalent to parts per thousand mass of dissolved salts to mass of pure water, or g/kg.] Table 8.1 gives the values of the coefficients n_i , which were determined by fitting Eq. (8.2) to the measurements of Austin and Halikas (1976).

coefficient	value
n_0	1.31405
n_1	1.779×10^{-4}
n_2	-1.05×10^{-6}
n_3	1.6×10^{-8}
n_4	-2.02×10^{-6}
n_5	15.868
n_6	0.01155
n_7	-0.00423
n_8	-4382
n_9	1.1455×10^6

Table 8.1: The best-fit coefficients for use in Eq. (8.2).

For pure (fresh) water ($S = 0$), Eq. (8.2) reduces to

$$n(T, \lambda) = 1.31405 - 2.02 \times 10^{-6}T^2 + \frac{15.868 - 0.00423T}{\lambda} - \frac{4382}{\lambda^2} + \frac{1.1455 \times 10^6}{\lambda^3}. \quad (8.3)$$

Although developed from data in the 400-700 nm range, Zhang and Hu (2009) find that Eq. 8.3 fits data from 200-1100nm with an average error of 1.5×10^{-5} . Huibers (1997) reached a similar conclusion about the excellent performance of Eq. 8.3.

There is also some pressure effect on n ; however, this effect is small because water is very incompressible. Table 8.2 shows values of n for the extreme ranges of pressure, temperature, salinity, and wavelength relevant to optical oceanography. For near-surface waters, n falls in the range of 1.33 to 1.35. Even including the extreme depth values, n varies by less than 3% over the entire parameter range. For comparison, living phytoplankton typically have “low” indices of refraction, in the range 1.01 to 1.09 relative to the index of refraction of pure seawater. Detritus and inorganic particles generally have “high” indices, in the range of 1.15 to 1.20 relative to seawater.

Harvey et al. (1998) give additional tables of n at selected wavelengths, including temperatures up to 500 C and pressures up to 10^8 Pa.

p [Pa]	T [C]	S [PSU]	λ [nm]	n
1.01×10^5	0	0	400	1.344186
1.01	0	0	700	1.331084
1.01	0	35	400	1.351415
1.01	0	35	700	1.337906
1.01	30	0	400	1.342081
1.01	30	0	700	1.329128
1.01	30	35	400	1.348752
1.01	30	35	700	1.335316
1.08×10^8	0	0	400	1.360076
1.08	0	0	700	1.346604
1.08	0	35	400	1.366885
1.08	0	35	700	1.352956
1.08	30	0	400	1.356281
1.08	30	0	700	1.342958
1.08	30	35	400	1.362842
1.08	30	35	700	1.348986

Table 8.2: Index of refraction of water, n , for the extreme values of pressure p , temperature T , salinity S , and wavelength λ encountered in oceanography. A pressure of 10^5 Pa corresponds to a depth of about 10 m; 10^8 Pa corresponds to a depth of about 10,000 m. Values extracted from [Austin and Halikas \(1976\)](#).

8.2.1.2 The imaginary part of the index of refraction

The imaginary part of the index of refraction for pure water is shown as a function of wavelength in vacuo by the red line in Fig. 8.5. The corresponding absorption coefficient $a = 4\pi k/\lambda$ is shown in green. Starting at the small-wavelength (X ray) end of the spectrum, the absorption rises as Compton scattering and then the photoelectric effect (see Section 1.1) come into play, removing energy from the incident light. The peak near 100 nm is due to the collective excitation of all of the electrons in the water molecule. Going into the near-ultraviolet, there is an extremely rapid decrease in a with wavelength when the photon energy is no longer able to excite electron transitions of the water molecule. The minimum value is at blue wavelengths in the visible part of the spectrum, which is shown by the light blue band in the figure. Absorption then rises again going into the infrared. In gaseous water, near-infrared wavelengths excite rotational modes of water molecules. However, these modes are quenched in liquid water because the molecules cannot freely rotate. In liquid water, the individual peaks between the visible and about $10 \mu\text{m}$ are caused by O-H stretching modes of the water molecule being excited. Going further into the microwave region, absorption by pure water remains much higher than in the visible but decreases as the longer and longer wavelengths have less and less energy for exciting vibrational states. Finally, when the wavelength is around 10 m, far into the radio region of the spectrum, the absorption is again comparable to that at blue wavelengths.

The green dashed line in the figure for $\lambda > 10$ cm shows the absorption for sea water, which is a conductor because of the salt ions. The absorption for radio waves is then much higher than for pure water. The extrapolation of the salt-water curve to longer wavelengths is supported by theoretical considerations, which show that the absorption should be proportional to $1/\sqrt{\lambda}$ (Jackson, 1975, §7.5). If this is done, the absorption by sea water again becomes comparable to that at visible wavelengths for wavelengths of order 1000 km. This part of the spectrum is called “extremely low frequency,” or ELF, and has frequencies of 100 Hz or less. The “clarity” of seawater at ELF frequencies has not escaped the attention of the world’s navies. Several countries have built transmitters that operate at ELF frequencies to communicate with submerged submarines; see, for example, [Communication with Submarines](#) and [Project Sanguine](#). The downside of communication at ELF frequencies is that the antennas must be very large, of order 100 km in size, and the data transmission rates are as low as only a few bytes per second.

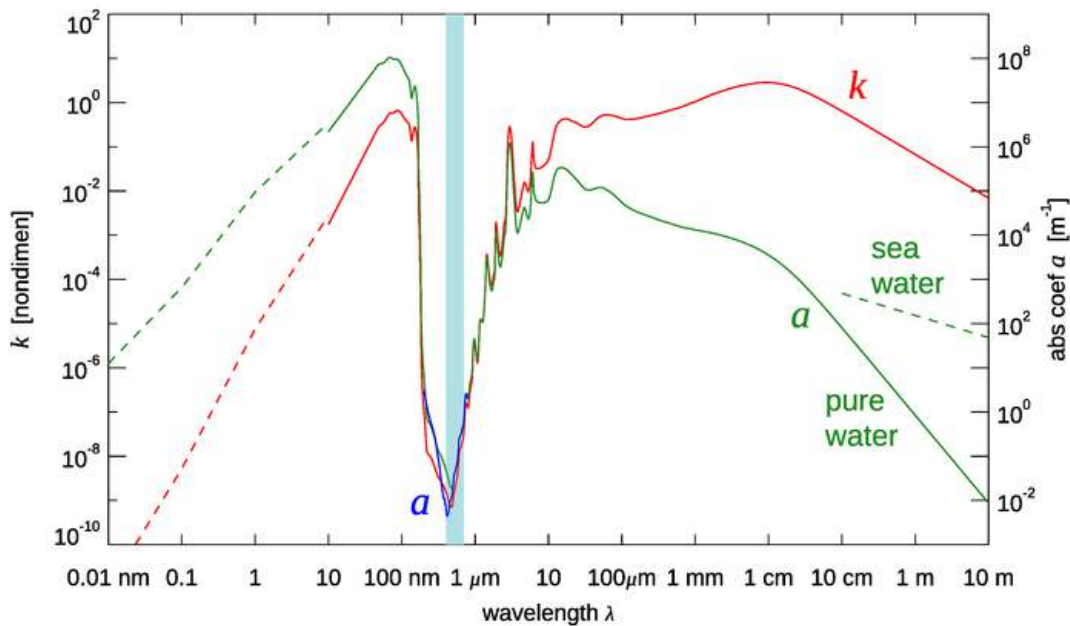


Figure 8.5: The imaginary index of refraction k of pure liquid water (red, left ordinate axis) and the corresponding absorption coefficient a (green, right ordinate axis). The light blue band is the visible wavelengths from 400 to 700 nm. The blue curve from 200 to 800 nm shows the values commonly used today. Dashed curves for $\lambda < 10$ nm are from Zoloratev and Demin (1977). Solid curves are from Segelstein (1981). The blue curve from 200 to 800 nm is from Smith and Baker (1978) and Pope and Fry (1997). The sea water curve is from Jackson (1975).

8.2.2 Absorption by Water

Absorption by water has a rich structure due to the the excitation of the different vibrational modes of the water molecule. Water absorption can be computed using empirical

functions of the water temperature T and salinity S . Figure 8.6 shows the dependence of the water absorption coefficient on T for $S = 35$ PSU, and on S for $T = 20$ deg C. These curves were computed using the semiempirical equations of [Roettgers et al. \(2014\)](#). The temperature effect on absorption is largest in the regions around 740 and 800 nm (and in several bands beyond 800 nm, not seen in this figure). The figure shows that salinity has much less effect on absorption than does temperature.

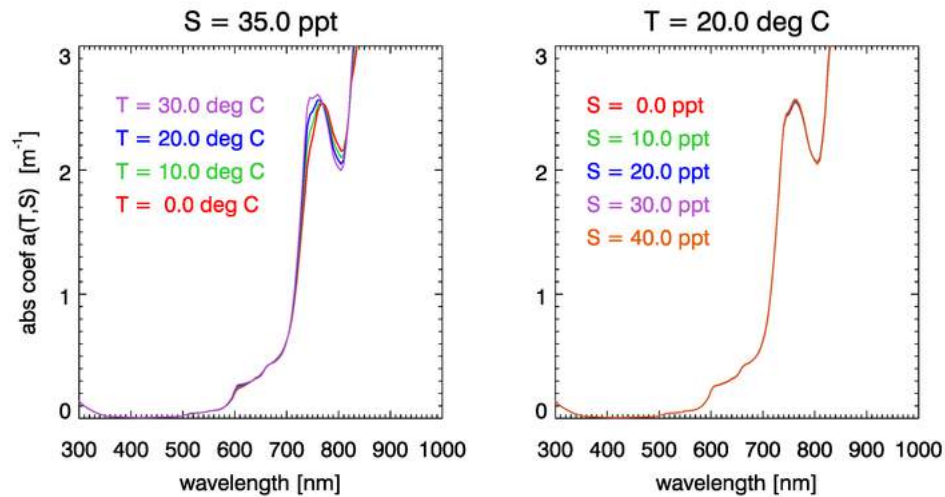


Figure 8.6: Dependence of the water absorption coefficient on temperature and salinity according to [Roettgers et al. \(2014\)](#).

The [Optical Absorption of Water Compendium](#) gives access to historical publications on water absorption going back to 1929.

8.2.3 Elastic Scattering by Water

According to Einstein-Smoluchowski theory (Section 6.3.4), the elastic scattering of light by water is due to spatial inhomogeneities in the dielectric constant caused by random motion of molecules (see [Zhang and Hu \(2009\)](#) and [Zhang et al. \(2009\)](#) for recent reviews). Elastic scattering by water has similarities in angular shape and spectral behavior to Rayleigh scattering (scattering by spherical particles that are much smaller than the wavelength of the light; Section 12.5.1), however with important differences. For example, the VSF is

$$\beta(\psi) = \beta(90) \left(1 + \frac{1 - \delta}{1 + \delta} \cos^2 \psi \right), \quad (8.4)$$

where ψ is the scattering angle and δ is the depolarization ratio. Several depolarization ratios have been suggested, with $\delta = 0.039$ providing the best fit to data ([Zhang et al., 2009](#)).

The elastic scattering coefficient of sea water depends on salinity ($\sim 30\%$ increase for range of salinities normally observed in the oceans), much less so of temperature ($\sim 4\%$ between 0 and 26 °C), and still less for pressure ($\sim 1.3\%$ for an increase in P of 100 bar). Figure 8.7 shows the dependence of the water scattering coefficient on T and S according to the model of [Zhang et al. \(2009\)](#).

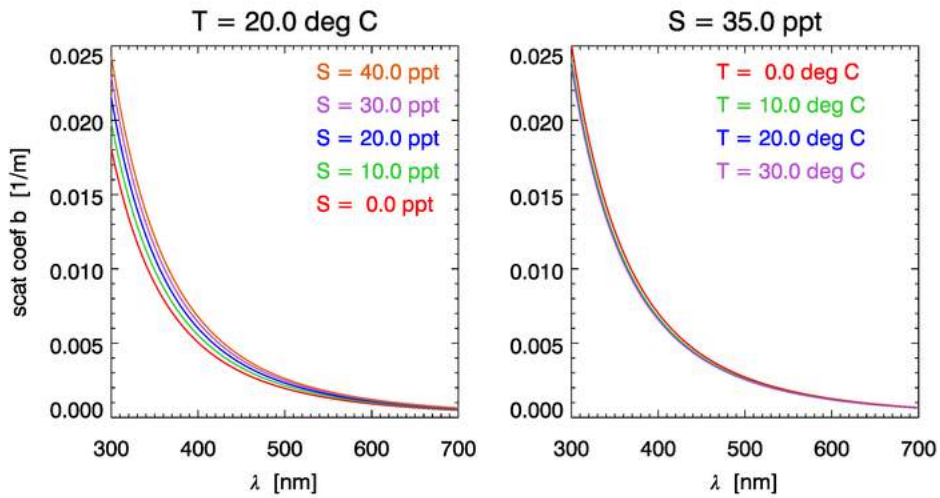


Figure 8.7: Dependence of the water scattering coefficient on temperature and salinity according to [Zhang et al. \(2009\)](#).

Rather than repeat the derivations from [Zhang and Hu \(2009\)](#) and [Zhang et al. \(2009\)](#), the corresponding [Web Book page on water](#) provides a link to a MATLAB function that provides their results and computes the scattering coefficient and the VSF of water at 90 degrees for a given wavelength, salinity and temperature. Figure 8.8 shows example plots computed using the MATLAB code of Zhang.

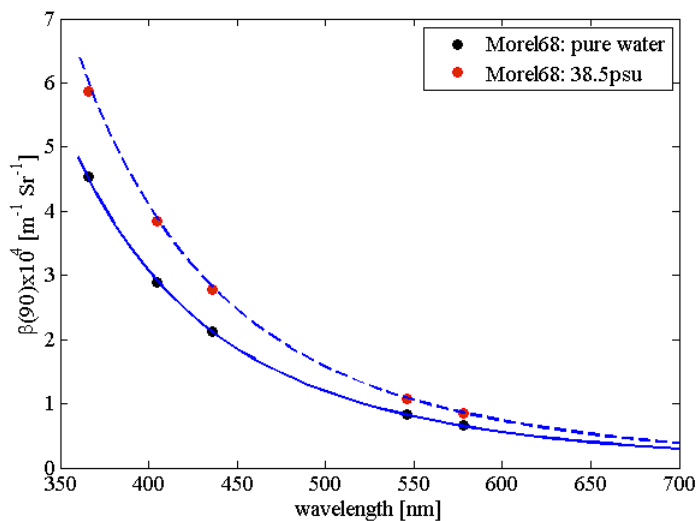


Figure 8.8: Comparison of the data of [Morel \(1968\)](#) and the model of [Zhang et al. \(2009\)](#) for scattering by pure water and pure sea water. Assumed $T = 20^\circ\text{C}$.

Raman scattering by water is discussed in detail in Section 7.2.

8.3 Phytoplankton

[Ivona Cetinić and Curtis Mobley contributed to this section.]

Phytoplankton combines the Greek words *phyto* meaning plant and *planktos* meaning wanderer or drifter, i.e., drifting plants. Phytoplankton are one of the primary reasons for studying the ocean. These microscopic, single cell, free-floating organisms have a major effect on ocean color and influence the heating of near-surface water. Additionally, they possess chlorophyll, a pigment that allows them to harvest sunlight and, through the process of photosynthesis, produce energy-rich organic material, while releasing the oxygen. That makes them the most important primary producers in the ocean, the base of the oceanic food web, and an important component of the global carbon cycle (Falkowski and Raven, 2007). Phytoplankton, through photosynthesis, contribute about one-half of the Earth’s annual oxygen production, with the other half coming from photosynthetic terrestrial plants. For all of these reasons and many more, it is of a great importance to understand phytoplankton abundance and dynamics.

There are many ways to quantify “how much” phytoplankton is in the water. You can count their number per unit volume of water using a microscope or particle sizing instruments such as a flow cytometer or diffraction-based optical instruments. You can measure their biomass or carbon content. All phytoplankton contain chlorophyll, a molecule that is the basis of photosynthesis. There are several types of chlorophyll, denoted by of chlorophyll *a*, chlorophyll *b*, chlorophyll *c*₁, etc., but all phytoplankton chlorophyll *a*. Therefore, the most common measure of phytoplankton concentration or abundance is the concentration of chlorophyll *a*, usually by *Chla* or usually just *Chl*, given in units of milligrams of chlorophyll per cubic meter (numerically equal to micrograms per liter).

Phytoplankton are a taxonomically diverse group comprising more than 10,000 species and taxa (Jeffrey and Vesk, 1997). By number, the most important groups of phytoplankton include diatoms, dinoflagellates, and coccolithophorids, although many other types of algae are part of the phytoplankton. Additionally, an important group from the realm of bacteria, cyanobacteria, is part of the phytoplankton. They come in different sizes (from 0.2 μm to $> 1000 \mu\text{m}$), shapes (spheroids, cylinders, prolate spheroids and many more), builds (cellulose, silicate or calcium-carbonate cell walls), and with many different physiologies.

Phytoplankton contain a large number of different pigments¹, which are substances that absorb specific wavelengths of light, the most important of which is chlorophyll *a*. Depending on group or taxa, varying concentrations of other pigments will be present in the cell—other chlorophylls, carotenoids, and billiproteins (Jeffrey and Vesk, 1997). All the specifics of phytoplankton make-up mentioned above impact phytoplankton optical properties, and consequently allow us to study their ecology and evolution in time by measuring their optical properties. Phytoplankton absorbing and scattering properties, and chlorophyll fluorescence, are surveyed in this section; detailed treatments are given in other sections.

¹Pigments are generally insoluble (or nearly so) in water. Dyes, on the other hand, are typically soluble in water.

8.3.1 Pigment Packaging

In order to understand absorption by phytoplankton, it is first necessary to understand pigment packaging. The *pigment packaging effect* refers to the “flattening” of mass-specific absorption spectra that occurs when the absorbing material is “packaged” into discrete particles rather than being uniformly distributed throughout a cell. This qualitative idea is as follows.

Consider an absorbing substance, e.g. chlorophyll, distributed molecule by molecule throughout a medium (water or cell cytoplasm). The absorbing material is then considered to be dissolved or in solution. When light shines into this solution, some amount will be absorbed. Moreover, each molecule has a roughly “equal chance” to receive and absorb some of the light. This is illustrated in the left panel of Fig. 8.9, which shows absorbing molecules (the blue dots) dispersed throughout a medium. Light rays (the arrows) can reach almost any molecule and be absorbed. When the absorbing molecules are grouped together into larger particles as seen in the right panel of the figure, some of the molecules are “shaded” by others and have less chance of absorbing some of the incident light. Thus less light will be absorbed when the absorbing molecules are packaged than when they are dissolved (or are in smaller packages), even though the amount of the absorbing material is the same in each case. In other words, the mass-specific absorption decreases as packaging increases.

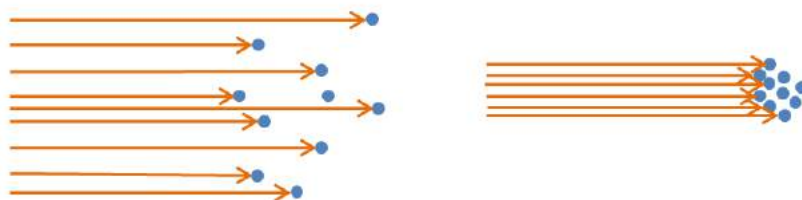


Figure 8.9: The qualitative idea of pigment packaging. The left panel shows absorbing molecules (the blue dots) dispersed throughout a medium. Light rays (the arrows) can reach almost any molecule and be absorbed. The right panel shows the absorbing molecules packaged together into larger particles. Some of the absorbing molecules are now shaded by others and therefore absorb less light.

Figure 8.10 shows the effect of pigment packaging on absorbance spectra for a particular species of phytoplankton, the flagellate *Euglena gracilis*, which is 20 to 100 μm in size. (Recall from Eq. (3.23) that the absorption coefficient a is related to the absorbance Abs by $a = 2.303Abs/R$, where R is the path length through the sample.) The red spectrum is for intact cells. The blue spectrum is for cells that have been broken up by ultrasonication so that the chlorophyll is dispersed throughout the water. The inset photo of a *Euglena gracilis* cell shows the green chloroplasts and other internal structures.

There are two ways that the effects of pigment packaging can be enhanced. First, for a given package size, increasing the absorption efficiency of the pigment molecule will give a larger packaging effect because the molecules first receiving the light will absorb more of the incident light, leaving less light for the molecules that are shaded. Second, for a given pigment molecule, making the packages larger will allow less light to penetrate to the “back” of the package and, again, the shaded molecules will absorb less.

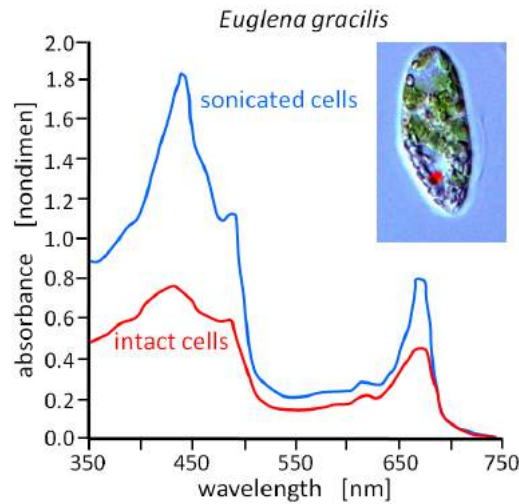


Figure 8.10: The effect of pigment packaging on absorbance spectra for in vivo (red) and sonicated (blue) cells. Spectra redrawn from Kirk (1994, Fig. 9.1). Inset photo by E. O’Neill from https://en.wikipedia.org/wiki/Euglena_gracilis, reproduced by the Creative Commons license.

The first of these situations leads to the flattening of absorption spectra. Chlorophyll absorbs more strongly at 440 nm, for example, than at 550 nm. Thus the chlorophyll-specific absorption will be relatively less at 440 than at 550, and the chlorophyll-specific absorption spectrum will have a smaller difference in the value at 440 compared to 550 as packaging increases. This is what is seen in the red curve of Fig. 8.10.

The second situation gives flatter spectra for large phytoplankton than for small ones because large cells can contain larger packages than can small cells. Large cells contain more chlorophyll than small cells, so the effect of packaging correlates with the chlorophyll concentration. This correlation between the chlorophyll concentration and spectral flattening is seen in Fig. 8.11. The left panel shows the absorption coefficient for particles (living phytoplankton plus detritus and non-algal particles) in Case 1 water. The spectra are color coded by the range of chlorophyll concentration. As expected, on average, the absorption coefficient a_p is larger when the chlorophyll concentration is larger. The right panel shows the corresponding chlorophyll-specific absorption spectra, $a_p^* = a_p/Chl$. It is seen that the highest chlorophyll values correspond to the “flattest” spectra (which are almost buried under the blue and purple spectra in the figure), and the lowest-chlorophyll values have the spectra with the most structure. This shows the effect of pigment packaging.

For a more detailed discussion of pigment packaging, see Duysens (1956), who originally recognized the importance of packaging and developed a mathematical model of the effect. Kirk (1976, Section 9.2) repeats Duysens’ development. Morel and Bricaud (1981a); Johnsen et al. (1994) and many others have built on Duysens’ original model of pigment packaging to understand and model absorption by phytoplankton.

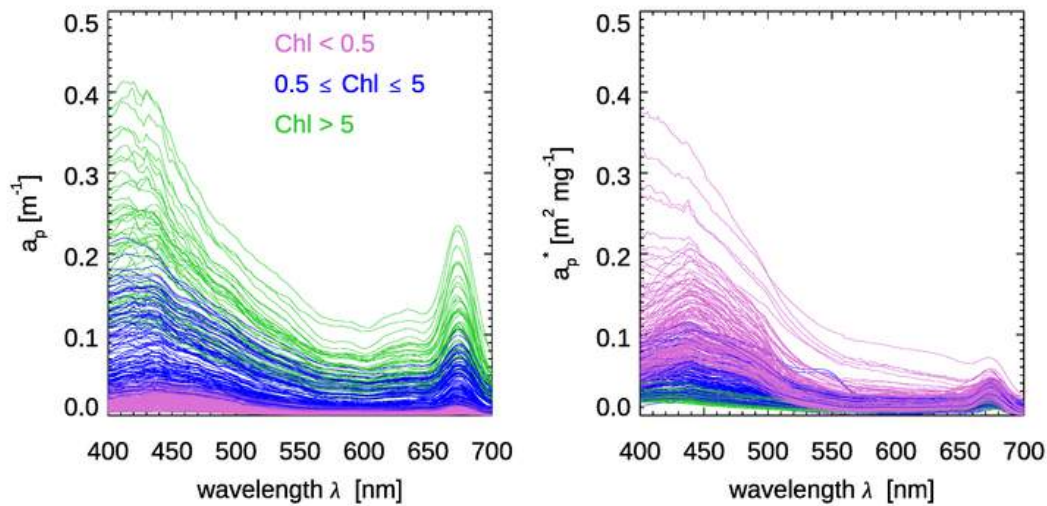


Figure 8.11: Particulate absorption spectra for Case 1 water color coded by the range of chlorophyll values. Left panel: particulate absorption spectra; right panel: the corresponding chlorophyll-specific absorption spectra. Data courtesy of A. Bricaud; for discussion of these spectra, see [Bricaud et al. \(1998\)](#).

8.3.2 Absorption by Phytoplankton

Phytoplankton absorb sunlight and use this energy to produce energy-rich organic material through the process of photosynthesis. Chlorophylls, present in all phytoplankton cells, cause two dominant peaks in absorption spectra, a primary peak in the blue (centered near 440 nm) and a secondary peak in red part of the spectrum (near 675 nm). The presence of other pigments (depending on species and taxa) causes a broadening of blue peak and the appearance of additional absorption maxima. These taxa-specific absorption peaks have been used as an optical detection tool in situ ([Kirkpatrick et al., 2000](#)) as well for development of remote sensing algorithms (e.g., [Subramaniam et al., 2002](#); [Tomlinson et al., 2009](#); [Wang and Moisan, 2021](#)). [Jeffrey et al. \(2012\)](#) give a detailed review of the pigments found in different microalgal classes.

However, absorption by phytoplankton is not a simple sum of absorption coefficients of individual pigments. Spectra of phytoplankton absorption, $a_{\text{phy}}(\lambda)$, vary in magnitude and shape due to the different cellular pigment composition and pigment packaging ([Bidigare et al., 1990](#); [Bricaud and Stramski, 1990](#); [Hoepffner and Sathyendranath, 1991](#); [Ciotti et al., 2002](#); [Bricaud et al., 2004](#)). Specific pigment-protein complexes present in the cell will cause changes in absorption spectra and magnitudes. Furthermore, the increase in cellular pigment concentration and cell size, i.e. the packaging effect, will flatten the specific absorption spectra as explained above.

For theoretical calculations of phytoplankton absorption and inversion of measured absorption spectra, we need measurements of *in vivo* pigment spectra. Figure 8.12 shows two such databases that are based on absorption in solvents, with the absorption spectra shifted to match those observed *in-vivo*.

Figure 8.13 shows examples of absorption spectra of phytoplankton normalized by chlorophyll and showcases the variability in pigments, packaging (flattening of the peak

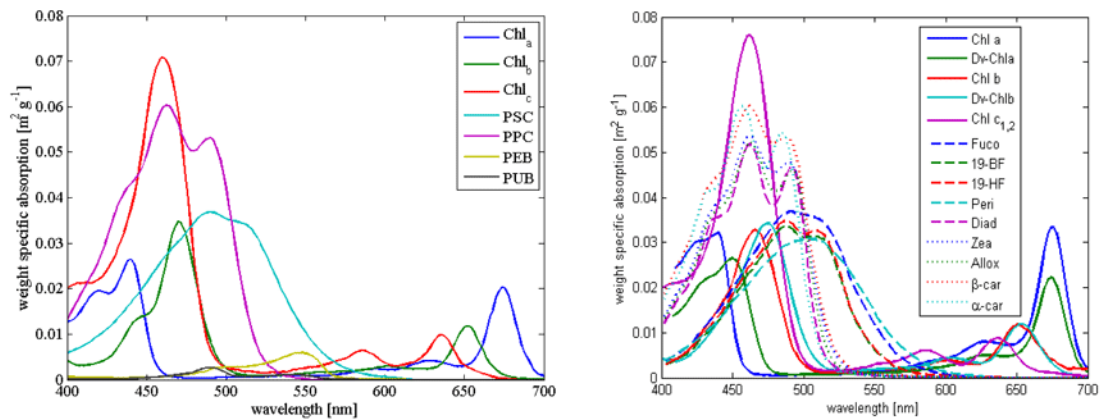


Figure 8.12: Mass-normalized absorption of phytoplankton pigments based on studies by Bidigare et al. (1990) (left) and Bricaud et al. (2004) (right). These data can be downloaded in spreadsheet format on the [Web Book page on phytoplankton](#).

where absorption is high) and differences between large and small cells.

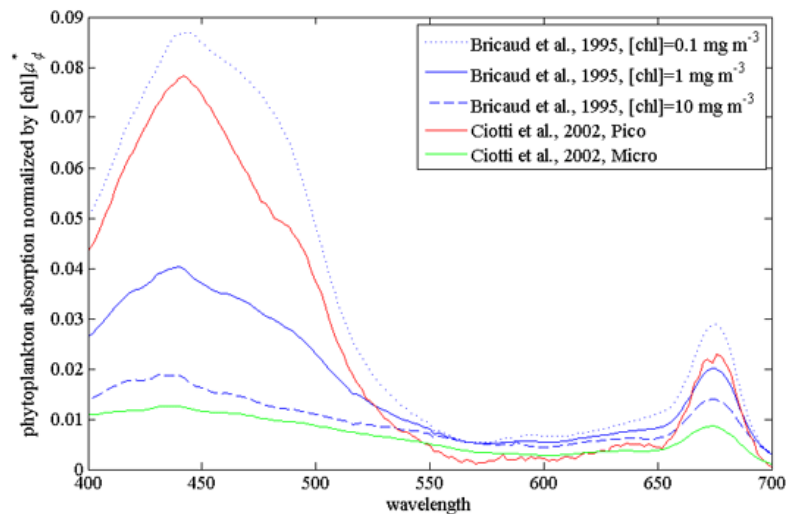


Figure 8.13: Chlorophyll-normalized phytoplankton absorption based on Bricaud et al. (1995) and Ciotti et al. (2002). These data can be downloaded in spreadsheet format on the [Web Book page on phytoplankton](#).

8.3.3 Scattering by Phytoplankton

The scattering properties of phytoplankton are important since they are directly related to remote-sensing reflectance calculations (via the backscattering to absorption ratio). Scattering and backscattering coefficients of phytoplankton as well as the volume scattering function are derived from either theoretical models (e.g., Mie theory) or direct measurements of the above mentioned properties (Bricaud et al., 1983; Volten et al., 1998;

Witkowski et al., 1998; Vaillancourt et al., 2004; Sullivan and Twardowski, 2009). They strongly depend on the size, shape and refractive index of all components of the phytoplankton cell (e.g., Volten et al., 1998; Witkowski et al., 1998; Jonasz and Fournier, 2007; Sullivan and Twardowski, 2009). Values of phytoplankton scattering coefficients, when compared to other kinds of the oceanic particles, are relatively low because of their high water content and strong absorptive properties (Aas, 1984). An exception to this rule is coccolithophores—phytoplankton that produce small calcium carbonate scales that make them very effective scatterers, which allows the detection of coccolithophorid blooms from space (Balch et al., 1996). Discriminating phytoplankton based on their angular scattering properties was suggested by Wyatt and Jackson (1989).

Our knowledge about the angular distribution of scattering for phytoplankton is scarce due to the small number of experimental datasets (Petzold, 1972; Morel and Bricaud, 1986; Volten et al., 1998; Witkowski et al., 1998; Vaillancourt et al., 2004; Lotsberg et al., 2007). In general, phytoplankton forward scattering dominates scattering in backward direction as seen in Fig. 8.14, consistent with their large size-to-wavelength ratio. Such angular distribution of scattering causes low backscattering ratios observed for phytoplankton populations in situ (Twardowski et al., 2001).

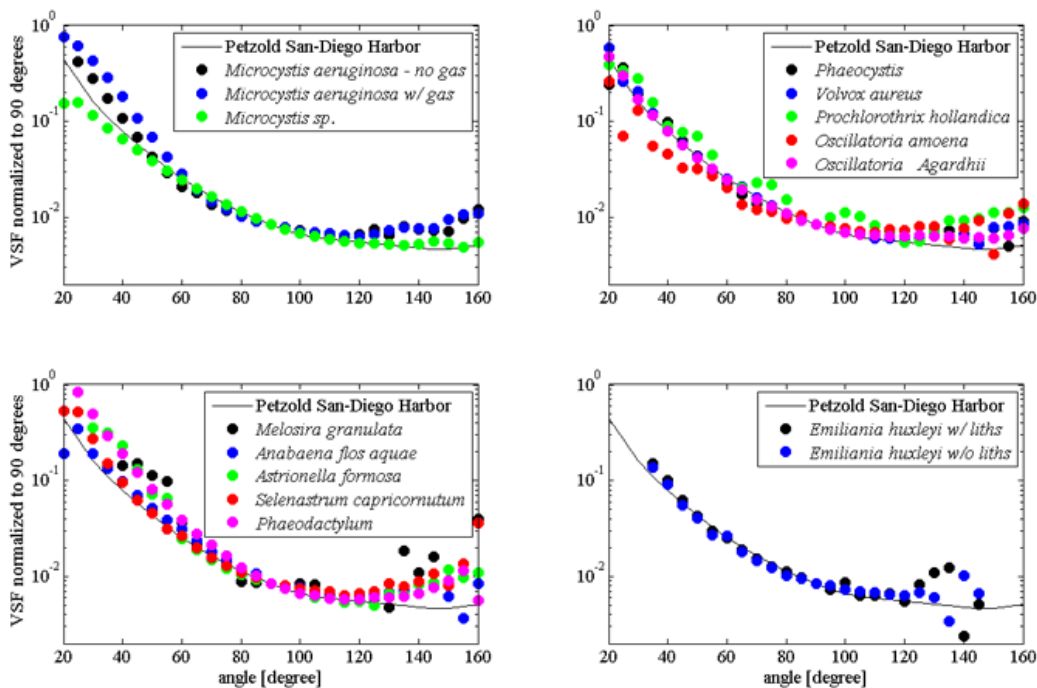


Figure 8.14: Scattering by different phytoplankton species in comparison with the San-Diego Harbor VSF of Petzold (1972). Data from Volten et al. (1998). These data and their uncertainties can be downloaded as a spreadsheet from the [Web Book page on phytoplankton](#).

Following the formula that scattering is the difference between attenuation of the light and absorption ($b = c - a$), and the rather featureless attenuation spectra of phytoplankton,

one can assume that the absorptive properties of the cell influence the shape of the spectral scattering (Bricaud et al., 1983). The scattering spectra will be, roughly speaking, an inverse picture of absorption spectra of the same species; a photon absorbed is not scattered. Similar patterns are also observed in backscattering (Stramski et al., 2001; Gordon et al., 2009). The phenomena of anomalous dispersion, which explains the shape of scattering spectra in the region of strong absorption bands, is treated in detail in Section 11.5.

8.3.4 Fluorescence by Phytoplankton

The general theory of fluorescence as used in time-independent radiative transfer theory was developed in Section 7.3 and applied to chlorophyll fluorescence in Section 7.4. However, a mathematically simpler formulation, as outlined here, is usually seen in the biological literature.

A portion of the light absorbed by a phytoplankton cell can be emitted at another, longer wavelength; a process referred to as fluorescence. Several phytoplankton pigments (chlorophylls, pheopigments, and phycobilins) fluoresce, with chlorophyll *a* fluorescence being the most significant. Although fluorescence is only a form of energy dissipation of the absorbed light, secondary to photosynthesis, it is still significant enough to be observed from space (Neville and Gower, 1977; Huot et al., 2005). Fluorescence is important for the study of ocean primary production (Yentsch, 1994).

Fluorescence from phytoplankton chlorophyll is often expressed expressed by this simplified formula (Falkowski and Kiefer, 1985; Babin, 2008):

$$F = PAR [Chla] \bar{a}_{\text{phy}}^* \phi_F . \quad (8.5)$$

Here PAR is the is the measure of light impinging on the cell [$\text{photons s}^{-1} \text{m}^{-2}$], $[Chla]$ is chlorophyll *a* concentration [mg m^{-3}], \bar{a}_{phy}^* is the wavelength-averaged chlorophyll-specific phytoplankton absorption coefficient as weighted by the scalar irradiance spectrum [$\text{m}^2 \text{mg}^{-1}$], and ϕ_F is the nondimensional quantum yield of fluorescence—the emission efficiency of the cell defined as the number of photons emitted divided by the number of photons absorbed by the chlorophyll. The resulting F is then photons emitted per second per cubic meter.

If the spectral scalar irradiance $E_o(\lambda)$ is in energy units [$\text{W m}^{-2} \text{nm}^{-1}$], then \bar{a}_{phy}^* is given by

$$\bar{a}_{\text{phy}}^* = \frac{\int_{400}^{700} a_{\text{phy}}^*(\lambda) E_o(\lambda) \frac{\lambda}{hc} d\lambda}{\int_{400}^{700} E_o(\lambda) \frac{\lambda}{hc} d\lambda} .$$

The factor of $\lambda/(hc)$ has units of photons per joule and converts energy units to quantum units (number of photons). The denominator of the previous equation is by definition PAR (Eq. 1.29). Thus Eq. (8.5) says that the number of photons fluoresced (F) is the product of how much light is available (PAR) times how much chlorophyll is present to absorb the light ($[Chla]$) times how efficiently the chlorophyll absorbs the light (\bar{a}_{phy}^*) times how efficiently the light absorbed by the chlorophyll is re-emitted as fluoresced photons. That does not sound complicated.

The complications to Eq.(8.5) lie in the fact that phytoplankton fluorescence depends (via the quantum yield of fluorescence ϕ_F and the chlorophyll-specific phytoplankton absorption spectrum $a_{\text{phy}}^*(\lambda)$) on numerous factors: taxonomic position of the algae, pigment content and ratios, photoadaptation, physiological state of the phytoplankton, nutrient conditions, and stage of growth (Prezelin and Alberte, 1978; Falkowski and Owens, 1980;

Babin, 2008; Cullen, 2008). All of that makes fluorescence a very complicated phenomenon. Measuring in-situ chlorophyll fluorescence is the most frequent method for describing the chlorophyll and phytoplankton distribution in the ocean (Cullen, 2008), but the above mentioned drivers of chlorophyll fluorescence variability make interpretation of the chlorophyll fluorescence data anything but straightforward. Refer to Babin (2008) for a comprehensive treatment of this important topic.

8.4 Colored Dissolved Organic Matter (CDOM)

Colored or Chromophoric Dissolved Organic Matter (CDOM) is an important optical constituent in water, often dominating absorption in the blue. The quantification of the amount of CDOM present in the water is based on the absorption or fluorescence by material passing through a given filter (most often with pore size of $0.2\ \mu\text{m}$). As such, it is an absorption (or fluorescence)-weighted sum of the different dissolved materials in the water. Note that most of the material comprising dissolved organic matter (DOM) does not absorb or fluoresce and that there exist inorganic dissolved materials that also absorb (e.g. iron oxides, nitrate), though it is believed that fluorescence is due solely to organic materials. From this discussion it follows that CDOM is not necessarily a good proxy for DOM, particularly in the open ocean. Nevertheless CDOM has been found to be a useful tracer of water masses, as well as an indicator of different biogeochemical processes. Coble (2007) gives a review on the link between the optical and chemical properties of DOM. Sample preparation and methodology of measurement are important in order to obtain accurate CDOM measurements. See Nelson and Coble (2009) for an analysis of methodology.

In the United States, colored dissolved organic matter is usually called CDOM. However, you will also see the terms “yellow matter,” “Gelbstoff” (German for “yellow matter”), or “gilvin.” The name gilvin, from a Latin word meaning pale yellow, was proposed by Kirk (1994) but is not widely used. Absorption by CDOM is commonly denoted by a_{CDOM} , a_g , or sometimes a_y .

8.4.1 Absorption by CDOM

Figure 8.15 shows several measurements of CDOM absorption. For wavelengths greater than 300 nm, a_{CDOM} decreases roughly exponentially with increasing wavelength. Below 300 nm there is a change in slope and a rapid increase in CDOM absorption with decreasing wavelength.

For wavelengths greater than 300 nm, which are the ones relevant to most optical oceanography, it is common to model CDOM absorption with an exponential function (e.g. Jerlov (1968)):

$$a_{\text{CDOM}}(\lambda) = a_{\text{CDOM}}(\lambda_0) \exp^{-S_{\text{CDOM}}(\lambda - \lambda_0)} \quad [\text{m}^{-1}], \quad (8.6)$$

where S_{CDOM} is referred to as the spectral slope parameter and λ_0 is a reference wavelength, which is often taken to be 440 nm. A theoretical explanation for this shape has been hypothesized by Shifrin (1988) as arising from a superposition of resonances of different molecular π -bonds in the long organic molecules comprising CDOM. Single bonds, which are most abundant, will absorb short wavelength radiation while resonances of multiple bonds, less abundant, absorb longer wavelength radiation. Since numerically there are

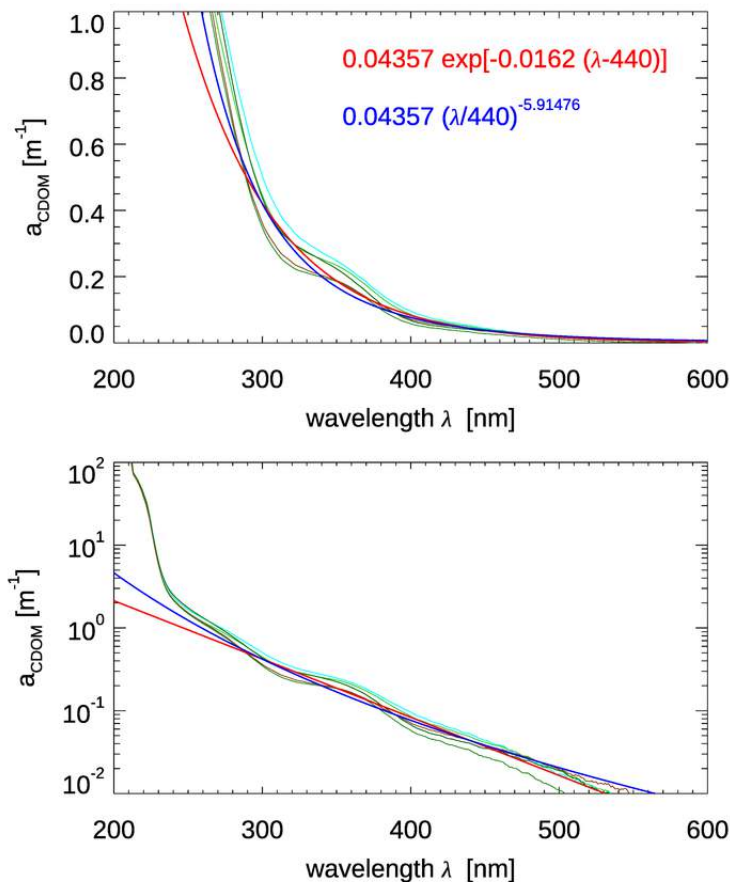


Figure 8.15: CDOM absorption spectra from five locations in Florida waters (greenish curves). The red line is Eq. (8.6) and the blue line is Eq. (8.7). For both models, the parameter values were determined from the average values of the five spectra at 300 and 440 nm. Data courtesy of L. Ayoub.

many more short bonds, the absorption is higher at shorter wavelengths. This explanation is consistent with the observation that small values of the spectral slope of CDOM, S_{CDOM} , are associated with higher molecular weight materials (e.g., Carder et al., 1989; Yacobi et al., 2003).

For visible wavelengths the most commonly used values of S_{CDOM} are near 0.014 nm^{-1} , based on measurements by Bricaud et al. (1981) and others. However, the value of S_{CDOM} varies in the visible from 0.007 to 0.026 nm^{-1} (e.g., Twardowski et al., 2004, Table 1). Figure 8.16 shows the distribution of S_{CDOM} values observed by Babin et al. (2003b). Their data show a mean of $S_{\text{CDOM}} = 0.0176 \text{ nm}^{-1}$ with a standard deviation of 0.0020 .

While (8.6) is the most frequent model of CDOM absorption, other models have been suggested and may provide a better fit to data (even when taking into account that fits improve as more free parameters are available in the fit, (e.g., Twardowski et al., 2004). In particular, a constant is often added to the exponential fit:

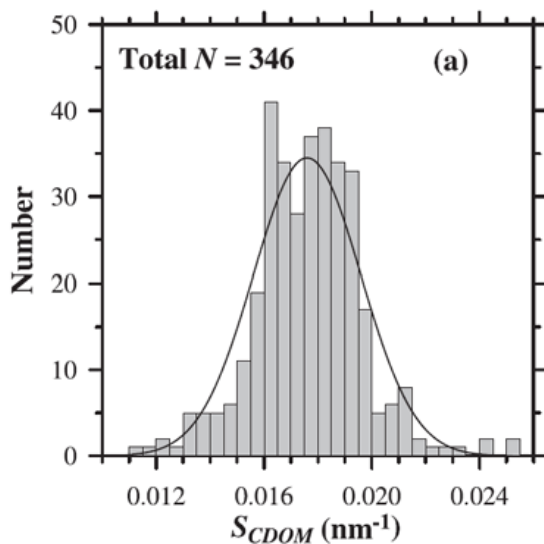


Figure 8.16: Distribution of the spectral slope parameter S_{CDOM} measured on 346 samples in European waters. From [Babin et al. \(2003b\)](#), Figure 4(a), reproduced by permission of the American Geophysical Union.

$$a_{CDOM}(\lambda) = a_{CDOM}(\lambda_0) \exp^{-S_{CDOM}(\lambda - \lambda_0)} + Constant \quad [\text{m}^{-1}].$$

What this constant represents is not clear. In some cases it is supposed to account for scattering by the dissolved component, however there is no reason to believe that such scattering would be spectrally flat (see [Bricaud et al. \(1981\)](#) for an in-depth discussion). It may account for bubbles in the sample.

Another model that has been found to work even better than the exponential model is a power-law model (e.g., [Twardowski et al., 2004](#)):

$$a_{CDOM}(\lambda) = a_{CDOM}(\lambda_0) \left(\frac{\lambda}{\lambda_0} \right)^{-S} \quad [\text{m}^{-1}]. \quad (8.7)$$

Given that molecular absorption is often a symmetric function for a given chromophore, frequency domain fits have been suggested (e.g., [Schwarz et al., 2002](#)). Those are based on Gaussian or Lorentzian functions with the visible domain being the tail end of the distribution. Trying to fit together UV and visible bands is complicated by the absorption of UV light by dissolved salts which are not part of DOC.

8.4.2 Elastic Scattering by CDOM

The CDOM contribution to scattering by seawater is somewhat controversial. By definition colloids are dissolved organic matter and, if abundant enough, could contribute significantly to scattering by sea water (particularly to backscattering, see [Stramski and Wozniak, 2005](#)). However, there is no observational evidence that CDOM contributes significantly to scattering (see [Dall'Olmo et al. \(2009\)](#) for recent measurements). Thus, currently, CDOM contribution to scattering is most often neglected.

8.4.3 Inelastic Scattering by CDOM

One of the primary methods to quantify CDOM is through fluorescence. Since not all dissolved material that absorbs also fluoresces, this material is often denoted as fluorescent dissolved organic material (FDOM). In general absorption and fluorescence covary, however their ratio can vary by orders of magnitude between different locations. The measurement of CDOM fluorescence in the field is often limited to a single excitation/emission band pair. With laboratory instrumentation, two-dimensional excitation-emission spectra can be measured and used to characterize the FDOM based on the size and presence of known excitation-emission peaks (see, for example, Coble, 2007, Table 1). This is discussed in more detail in the Section 7.5.

8.5 Non-algal Particles

As noted in the first section of this chapter, the term *non-algal particles* (NAPs) refers to particulate matter that does not have extractable (via solvents such as acetone or methanol) pigments. This includes all living and detrital organic matter such as the non-pigmented portion of phytoplankton cells, detritus, heterotrophic bacteria, and viruses. Many authors also include inorganic mineral particles of both biogenic (e.g., calcite liths and shells) and terrestrial origin (e.g., clay, silt, and sand). *Detritus* is a catch-all term for non-living particulate organic matter, including dead bacterial, phytoplankton and zooplankton cells, fragments of cells left from zooplankton grazing, fecal pellets, shells, and marine snow aggregates. Detritus both absorbs and scatters and can constitute an important fraction of the total IOPs of a water body.

8.5.1 Absorption by Non-algal Particles

There have been many measurements of the absorption spectra of detritus and NAPs. Figure 8.17 shows an example of mass-specific absorption by detritus and NAPs. These measurements include both biogenic detritus and mineral particles (including resuspended sediments). The curves show a generally exponential decay with increasing wavelength. It is thus common (e.g., Yentsch, 1962; Kirk, 1980; Roesler et al., 1989; Bricaud et al., 1998) to model absorption by detritus or NAPs via

$$a_{\text{NAP}}(\lambda) = a_{\text{NAP}}(\lambda_0) \exp^{-S_{\text{NAP}}(\lambda - \lambda_0)} \quad [\text{m}^{-1}], \quad (8.8)$$

where λ_0 is a reference wavelength and S_{NAP} the spectral slope parameter. The mean value of S_{NAP} used to model NAP absorption is 0.011 nm^{-1} (e.g., Roesler et al., 1989; Bricaud et al., 1998). However, as always with data-derived parameters, there is variability in S_{NAP} . Figure 8.18 shows the distribution of S_{NAP} values observed by Babin et al. (2003b). Their data show a mean of $S_{\text{NAP}} = 0.0123 \text{ nm}^{-1}$ with a standard deviation of 0.0013.

The absorption model of Eq. (8.8) has the same functional form as that for absorption by CDOM, although the average spectral slope parameter $S_{\text{NAP}} \approx 0.011 \text{ nm}^{-1}$ differs somewhat from the average parameter for CDOM, $S_{\text{CDOM}} \approx 0.018 \text{ nm}^{-1}$.

8.5.2 Scattering by Non-algal Particles

Although CDOM and NAPs have similar absorption spectra, they differ greatly in their scattering properties: CDOM has negligible scattering, but NAPs can be strongly scat-

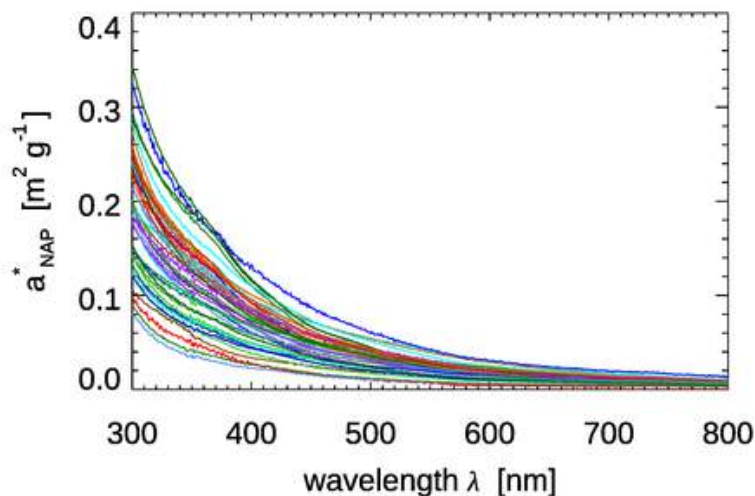


Figure 8.17: Measurements of mass-specific detritus and NAP absorption spectra from both fresh river water and coastal ocean water. Data courtesy of M. L. Estapa; see [Estapa et al. \(2012\)](#) for discussion of these data.

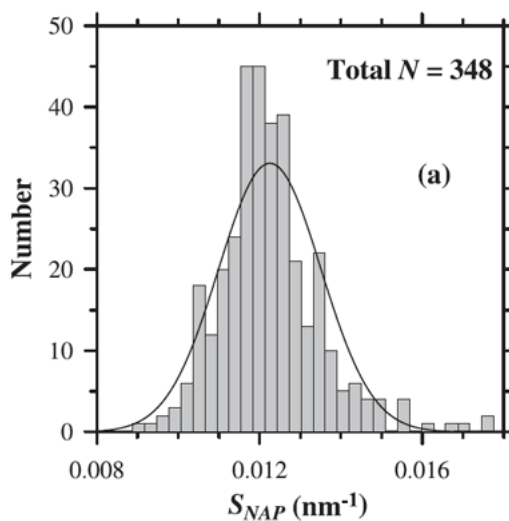


Figure 8.18: Distribution of the spectral slope parameter S_{NAP} measured on 348 samples in European waters. From [Babin et al. \(2003b, Fig. 41\(a\)\)](#), reproduced by permission of the American Geophysical Union.

tering. There is much less data on NAP scattering than for absorption, but Fig. 8.19 shows data from Great Lake (Taihu), China taken by [Sun et al. \(2010\)](#). The $b_{NAP}(\lambda)$ spectra in this dataset show a rather featureless dependence on wavelength, which is well

fit ($R^2 = 0.989$) by a power law:

$$b_{\text{NAP}}(\lambda) = b_{\text{NAP}}(550) \left(\frac{\lambda}{550} \right)^{-0.938} \quad [\text{m}^{-1}]. \quad (8.9)$$

Sun et al. (2010) show that the IOPs of this lake are dominated by NAP, with phytoplankton usually contributing less than 10% of the total scattering. (The same dataset had NAP absorption spectra that were well fit by Eq. (8.8), but with a spectral slope parameter of $S_{\text{NAP}} = 0.0052 \pm 0.0010 \text{ nm}^{-1}$, which is about one-half of typical ocean values.)

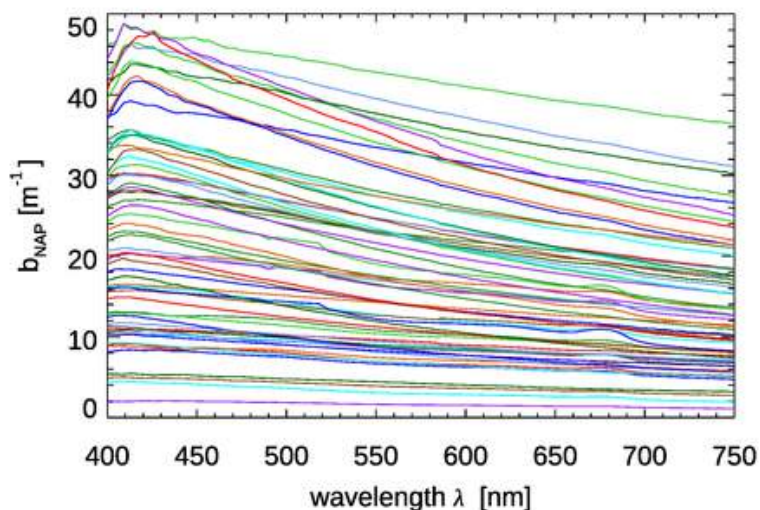


Figure 8.19: Scattering coefficient spectra $b_{\text{NAP}}(\lambda)$ from Great Lake (Taihu), China. Data from Sun et al. (2010, Fig. 1(B)) provided courtesy of D. Sun.

8.5.3 Absorption by Minerals

Mineral particles can enter the ocean from river discharge, erosion of coastal cliffs, sediment resuspension, and deposition of atmospheric dust, which is often carried long distances by winds. The optical properties of such particles are just as varied, and often just as important, as the properties of biological particles. In spite of their importance, relatively few studies have investigated the absorption and scattering properties of minerals per se. These include Ahn (1999); Stramski et al. (2004b); Babin and Stramski (2004), and Stramski et al. (2007).

Figure 8.20 shows mass-specific absorption coefficients $a^*(\lambda)$ (units of $\text{m}^2 \text{ g}^{-1}$) for four types of minerals. Figure 8.21 shows the same for several types of mineral dust suspended in sea water; some of the curves are the same mineral type but with different particle size distributions. The imaginary indices of refraction for these particles varied by an order of magnitude, from about 0.003 to 0.03, depending on wavelength and mineral type. See Stramski et al. (2007) for a full discussion. These (and other) data sets are in qualitative agreement: absorption by minerals can range from highly absorbing in the blue with a roughly exponential decrease with increasing wavelength, to almost non-absorbing and

spectrally flat. Thus, to first order, absorption by minerals can be modeled by either an exponential,

$$a_{\min}(\lambda) = a_{\min}(\lambda_o) \exp[-S_{\min}(\lambda - \lambda_o)] \quad [\text{m}^{-1}],$$

or a power law,

$$a_{\min}(\lambda) = a_{\min}(\lambda_o) \left(\frac{\lambda}{\lambda_o}\right)^{-m_{\min}} \quad [\text{m}^{-1}],$$

where λ_o is a reference wavelength. The parameters S_{\min} and m_{\min} can be adjusted as needed to fit a particular mineral type.

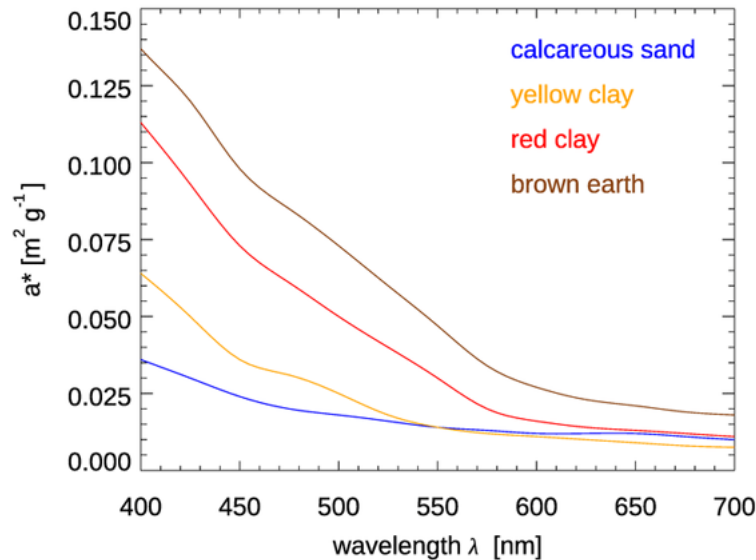


Figure 8.20: Measured mass-specific absorption coefficients $a^*(\lambda)$ for four types of minerals. From Ahn (1999, data courtesy of A. Morel).

8.5.4 Scattering by Minerals

Figures 8.22 and 8.23 show the mass-specific scattering coefficients $b^*(\lambda)$ for the same two data sets of Figs. 8.20 and 8.21. As for absorption, the scattering coefficients show a decrease from blue to red, although with a couple of exceptions that have their maximum values at green wavelengths. Many of these scattering coefficients could be modeled with an exponential or power law function. However, in all cases, use of a measured spectrum for radiative transfer calculations is preferable.

Stramski et al. (2001) used complex indices of refraction for detritus and mineral particles, a power law particle size distribution with a slope parameter of -4 (a Junge distribution), and Mie theory to compute absorption, scattering, and backscattering cross sections for detritus and minerals. Their results are shown in Table 8.3. It is interesting to note how closely the wavelength exponents for the scattering cross sections agree with the best-fit values for the NAP-dominated waters of Fig. 8.19 as seen in Eq. (8.9).

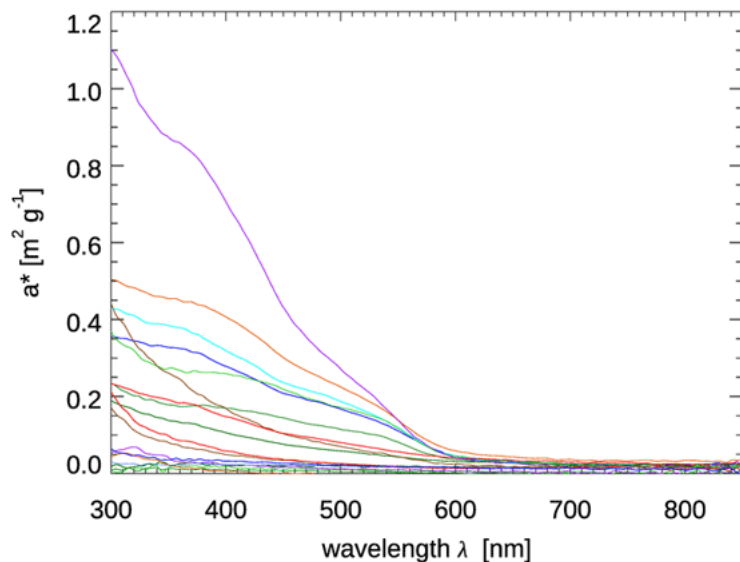


Figure 8.21: Measured mass-specific absorption coefficients $a^*(\lambda)$ for various types of mineral dust suspended in sea water. The highest magnitude curves are red soils rich in iron oxides, the lowest curves are calcite and quartz. Redrawn from [Stramski et al. \(2007, Fig. 3\)](#) with data provided courtesy of D. Stramski.

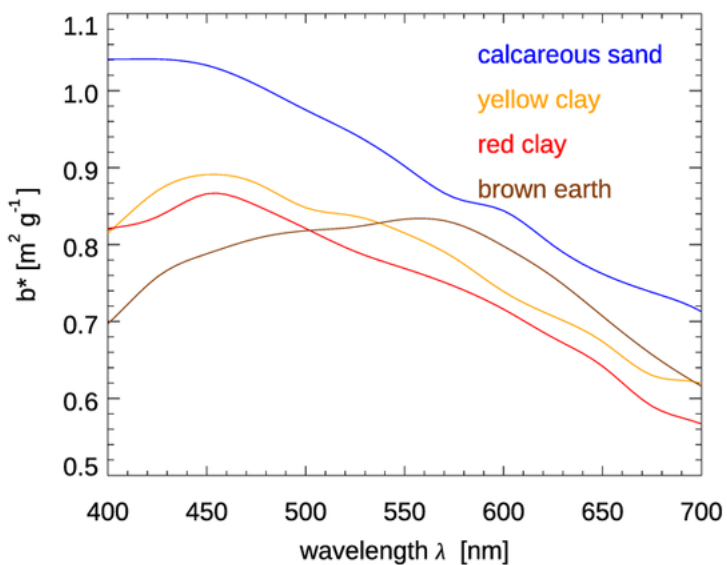


Figure 8.22: Measured mass-specific scattering coefficients $b^*(\lambda)$ for four types of minerals. From [Ahn \(1999, data courtesy of A. Morel\)](#).

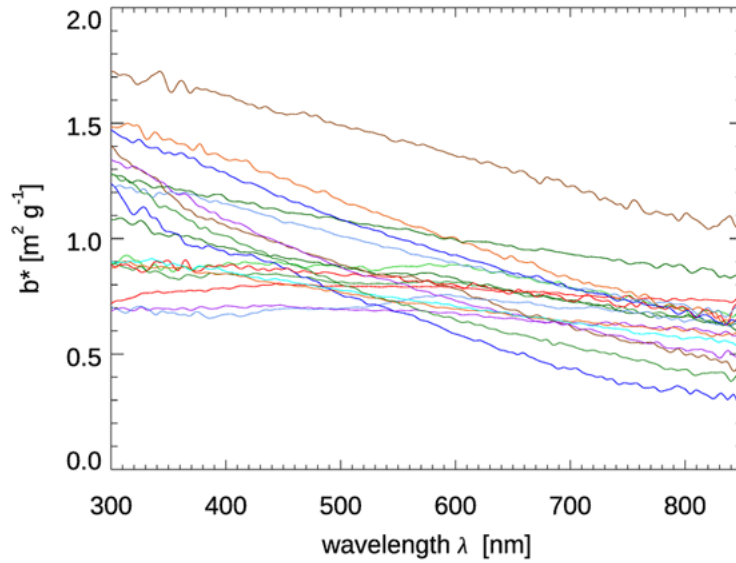


Figure 8.23: Measured mass-specific scattering coefficients $b^*(\lambda)$ for the same samples of mineral dust suspended in sea water as in Fig. 8.21. Redrawn from Stramski et al. (2007, Fig. 6) with data provided courtesy of D. Stramski.

	Detritus	Minerals
Absorption	$\sigma_{a, \text{det}} = 8.79 \times 10^{-16} e^{-0.00847\lambda}$	$\sigma_{a, \text{min}} = 1.01 \times 10^{-15} e^{-0.00846\lambda}$
Scattering	$\sigma_{b, \text{det}} = 1.43 \times 10^{-13} \lambda^{-0.9445}$	$\sigma_{b, \text{min}} = 7.71 \times 10^{-13} \lambda^{-0.9764}$
Backscattering	$\sigma_{\text{bb}, \text{det}} = 5.88 \times 10^{-16} \lambda^{-0.8997}$	$\sigma_{\text{bb}, \text{min}} = 1.79 \times 10^{-14} \lambda^{-0.9140}$

Table 8.3: Absorption, scattering, and backscattering cross sections for detritus and minerals. All have units of m^2 per particle. Multiplication by the number of particles per cubic meter gives a , b , and b_b in units of m^{-1} . Data from Stramski et al. (2001, Table 2).

8.6 Air Bubbles

[Xiaodong Zhang wrote this section.]

Bubbles in the upper ocean are generated primarily by breaking waves (Lamarre and Melville, 1991; Thorpe and Humphries, 1980). When wind speed exceeds 7 m s^{-1} , field observations have shown that a stratus layer of bubbles forms under the sea surface and persists as a result of continuous supply of bubbles by frequent wave breaking and the subsequent advection by turbulence (Crawford and Farmer, 1987; Thorpe, 1982, 1986). As the wind subsides, bubbles that have been injected will evolve under the effects of buoyancy and gas diffusion, and merge into the background population on time scales of minutes to hours (Johnson, 1986). When wind speeds are lower than 3 m s^{-1} , few waves break (Thorpe, 1982; Thorpe and Hall, 1983). In quiescent conditions, the presence of bubbles has also been

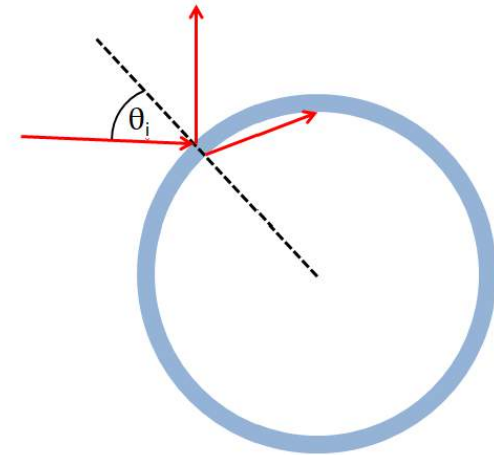


Figure 8.24: Schematic drawing of light incident on and scattered by a bubble with a coating (the shaded area).

detected (Medwin, 1977; O'Hern. et al., 1988), and the sources of these bubbles have been attributed to the pre-existing cavitation nuclei (O'Hern, 1987), remnant of bubbles injected by breaking waves (Johnson, 1986), sediment outgasing (Mulhearn, 1981), phytoplankton photosynthesis (Waaland and Branton, 1969), or zooplankton respiration (Ling and Pao, 1988).

Once formed, bubbles are coated with surfactant material almost instantaneously, and the accumulation of organic films onto their surfaces provides a stabilizing mechanism against surface tension pressure and gas diffusion (Fox and Herzfeld, 1954; Yount, 1979). These stabilized bubbles, acting as cavitation nuclei, explain the tensile strength (Apfel, 1972) and cavitation pressure (Holl, 1970) observed for natural water with magnitudes that are much lower than theoretical predictions for pure water.

The first study of optical properties of an air bubble in water probably was carried out by Davis (1955), who used the geometric optics approximation to study the angular distribution of intensity of light scattered by an air bubble in water. Later, the most intensive studies of single-bubble optics were undertaken by Marston and co-workers (Arnott and Marston, 1998, 1991, 1988; Kingsbury and Marston, 1981; Marston, 1979; Marston et al., 1988; Marston and Kingsbury, 1981; Marston et al., 1982). They examined the light scattering pattern near the critical angle (82.8°), Brewster angle (106.2°), and glory (180°). As shown in Fig. 8.24, a bubble (large compared to the wavelength of incident light) suspended in water can be regarded as a local water-to-air interface. When the incident angle θ_i is equal to or greater than 48.6° ($= \sin^{-1}(n_a/n_w)$), the light experiences total reflectance with a scattering angle of 82.8° or larger. Similarly when the incident angle θ_i is 36.9° ($= \tan^{-1}(n_a/n_w)$), the parallel polarized light has a null reflectance according to Fresnel's law, and the light at scattering angle of 106.2° is entirely perpendicularly polarized. Analogous to the optical glory of a drop but with different mechanism, light scattered from bubbles in water manifests an enhancement in the backward direction (180°). The viewable enhancement is within a circle of radius of about 2° .

For rising bubbles, their shape will become oblate due to drag, which in turn depends on the bubble size, surfactant, and water kinematic viscosity. By observing the glory pattern exhibited by freely rising bubbles, Arnott and Marston (1988) found that for bubbles with radius $< 150 \mu\text{m}$, the shape is sufficiently spherical for Mie theory to be

valid. Also photographs show that bubbles with adsorbed monolayers of surface-active material remain spherical (Johnson and Cooke, 1981). For practical purpose, bubbles in the ocean are assumed to be spherical and the errors introduced by non-spherical larger bubbles are not expected to be significant for the estimate of bulk optical properties because their number density is very small as compared to the smaller bubbles. Let m denote the relative index of a particle (including bubble) in the ocean. For practical purpose in the visible wavelengths, $m = 0.75$, which is estimated for an air bubble in pure water at 25°C at 550 nm wavelength. The variations of m with temperature, salinity, and pressure are $< 1\%$ in the visible domain and have a negligible effect on the calculation of optical properties of bubbles. The comparison of scattering and backscattering efficiencies of clean bubbles with those of soft ($m = 1.05$) and hard particles ($m = 1.20$) (Fig. 8.25a) as a function of the sizes shows that for $r < 0.5\ \mu\text{m}$, the scattering by a bubble is very similar to that of a hard particle, and both scatter about an order of magnitude more than a soft particle of equivalent size. All three achieve asymptotic value in scattering when $r > 10\ \mu\text{m}$. For $r < 0.1\ \mu\text{m}$, the bubble is the strongest in backscattering efficiency, whereas hard particles display the highest backscattering efficiency when $r > 0.5\ \mu\text{m}$. The asymptotic values for backscattering efficiency factor for bubbles, hard particles and soft particles are 0.02, 0.04 and 0.002 respectively. The backscattering ratio (Fig. 8.25b) decreases sharply as particle size $> 0.1\ \mu\text{m}$, and it diverges among various particles. The backscattering ratio for a bubble asymptotically oscillates around 0.01.

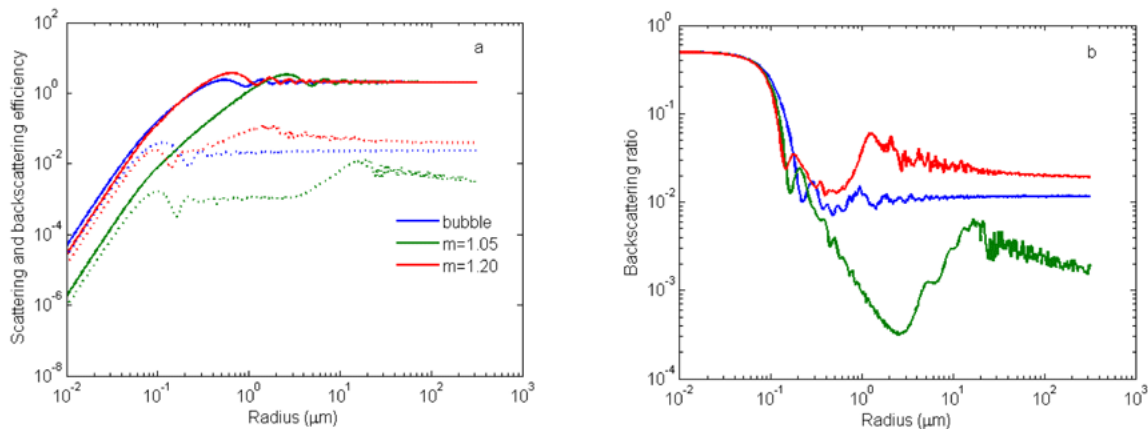


Figure 8.25: Mie scattering calculation ($\lambda = 550\text{ nm}$) of the scattering and backscattering efficiency (a) and backscattering ratio (b) for clean bubbles and for particles with relative refractive index = 1.05 and 1.20, respectively.

The coating of organic film on the surface of bubbles significantly affects the backscattering with little effect on the total scattering (Zhang et al., 1998). Scattering by bubbles coated with organic film does not change very much from those by clean bubbles of sizes $> 1\ \mu\text{m}$, but the backscattering increases with the thickness of the film, up to a factor of 4, as seen in Fig. 8.26. For bubbles of smaller sizes ($< 0.5\ \mu\text{m}$), the coating seems to reduce both scattering and backscattering efficiency as compared to clean bubbles. There are few data on the absorption properties of films coated onto the bubble surface. Calculations assuming an imaginary part of the refractive index of values 0.001 to 0.006 showed that only when the bubble sizes reach approximately $100\ \mu\text{m}$, and with an extremely high absorption

coefficient (imaginary index = 0.006), does absorption significantly reduce the scattering and backscattering from the case with no absorption. It is safe to say that the normally absorbing organic films exert a small influence on the scattering and backscattering of bubbles (Zhang et al., 1998).

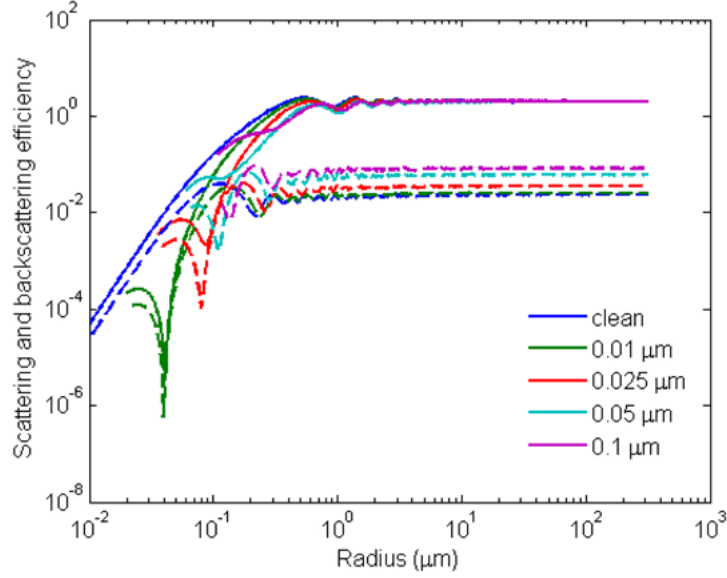


Figure 8.26: Mie scattering calculation ($\lambda = 550$ nm) of the scattering and backscattering efficiency for clean bubbles versus bubbles coated with an organic film of different thickness. The film is assumed to a protein with an index = 1.2.

The bulk optical properties of bubble populations can be easily estimated once the particle size distribution (PSD) of the bubbles is known. Let

$$\overline{Q}_j = \frac{\int_{r_{\min}}^{r_{\max}} Q_j(r) PSD(r) \pi r^2 dr}{\int_{r_{\min}}^{r_{\max}} PSD(r) \pi r^2 dr}, \quad (8.10)$$

and

$$S = \int_{r_{\min}}^{r_{\max}} PSD(r) \pi r^2 dr.$$

Then

$$j = N_0 \overline{Q}_j S \quad (8.11)$$

where \overline{Q}_j , S and N_0 are, respectively, the mean optical efficiency factors, the mean geometric cross-sectional areas, the number density of the size distribution of the bubble population; and $j = b$ or b_b , denoting the total scattering or total backscattering coefficient.

Many studies attempting to characterize the size spectra of bubble populations under various stages after wave breaking have been conducted and their results are summarized in Fig. 8.27a. Immediately after wave breaking, the newly created bubbles are typically of sizes 0.1 - 10 mm (Deane, 1997; Haines and Johnson, 1995; Phelps et al., 1997; Zhang et al., 2002). Once bubble creation processes cease, the newly formed bubble plume evolves under

the influence of turbulent diffusion, advection, buoyant degassing and dissolution, leaving behind a diffuse cloud of microbubbles (Baldy, 1988; Baldy and Bourguel, 1985; Breitz and Medwin, 1989; Cartmill and Su, 1993; Johnson and Cooke, 1979; Kolovayev, 1976; Leeuw and Cohen, 1995; Phelps and Leighton, 1998; Su and Cartmil, 1994; Su et al., 1988; Terrill et al., 2001; Vagle and Farmer, 1992, 1998; Walsh and Mulhearn, 1987). Copious amounts of bubbles have also been observed in quiescent seas (Fig. 8.27b), sometimes with a concentration even higher than in rough water (Huffman and Zveare, 1974; Ling and Pao, 1988; Medwin, 1977; O'Hern. et al., 1988; Shen et al., 1986).

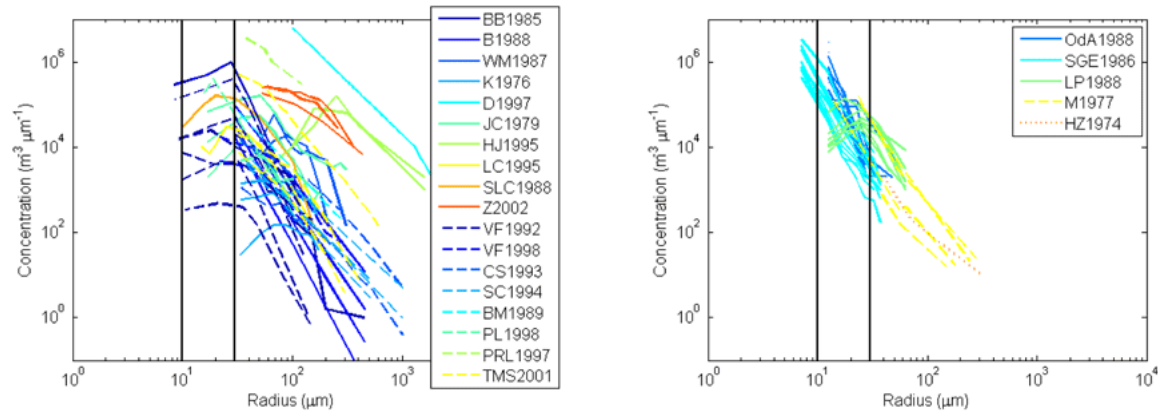


Figure 8.27: The size distributions of bubble populations measured under breaking waves (a, left) and in quiescent seas (b, right). Different colors represent different studies. Solid lines denote the methods that are based on optics and the dashed lines for acoustics. The vertical line at $30 \mu\text{m}$ denotes a size around which disagreements occur and the line at $10 \mu\text{m}$ denotes the size below which few observations are available.

Various bubble-sensing techniques have been developed and the technology can be categorized into two groups: optics and acoustics. Optics-based methods (solid lines in Fig. 8.27) include photography (Deane, 1997; Johnson and Cooke, 1979; Kolovayev, 1976; Leeuw and Cohen, 1995; Walsh and Mulhearn, 1987), light scattering and reflection (Baldy, 1988; Baldy and Bourguel, 1985; Ling and Pao, 1988; Su et al., 1988), and holography (O'Hern. et al., 1988). Bubbles pulsate under sound pressure and the intrinsic frequency is inversely proportional to the sizes. Bubbles of sizes that are in resonant with the acoustical frequency can significantly alter the sound propagation speed and exert an attenuation cross-section up to 100 times than those non-resonant bubbles or solid particles. The acoustics-based methods (dashed lines in Fig. 8.27) include resonator (Cartmill and Su, 1993; Medwin, 1970; Su and Cartmil, 1994; Terrill and Melville, 2000; Vagle and Farmer, 1998), sound speed dispersion (Huffman and Zveare, 1974), backscatter (Vagle and Farmer, 1992, 1998), and nonlinear interaction (Phelps and Leighton, 1998; Phelps et al., 1997).

Despite various techniques being developed and used, few studies have been able to detect bubbles of sizes $< 10 \mu\text{m}$ (the vertical lines at $10 \mu\text{m}$ in Fig. 1). We believe this limit, hardly of any natural origin, is artificial and imposed by technology being used. For acoustic frequencies greater than 200 KHz, which resonates with bubbles of $16 \mu\text{m}$, the off-resonant contribution from larger bubbles is very large and it is difficult to separate resonant signal from off-resonant noises. On the other hand, the size limit in photography is about $30\text{-}40 \mu\text{m}$.

For bubbles of sizes $> 10\mu\text{m}$, the mean optical efficiency factors $\overline{Q_j}$ are almost constant for both scattering and backscattering (Figs. 8.25 and 8.26); therefore the bulk scattering and backscattering coefficients (Eq. 8.11) are determined by the size distributions. Based on the bubble size spectra shown in Fig. 8.27a, the natural bubble populations account for a large portion of observed backscattering (Terrill et al., 2001; Zhang et al., 1998) or even completely dominate the scattering process in roughened seas (Stramski and Tegowski, 2001). In clear waters, the presence of bubbles shifts the color of the ocean towards green (Zhang et al., 1998).

The volume scattering functions of bubble populations were measured by a prototype volume scattering meter (Lee and Lewis, 2003) in the laboratory, and the results confirmed the elevated scattering at the critical angle and the enhanced backscattering by surfactant-coated bubbles relative to the clean bubbles (Fig. 8.28).

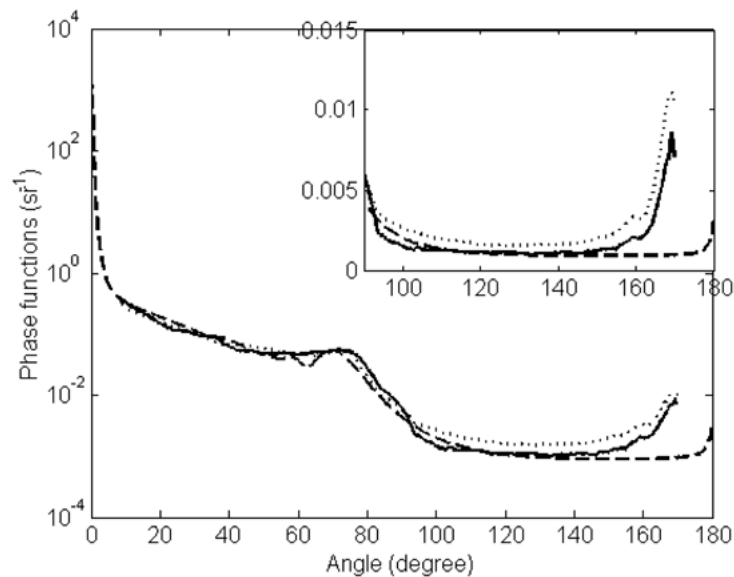


Figure 8.28: Two phase functions measured for bubble populations produced in clean seawater (solid line) and in surfactant-contaminated seawater (dotted line) are compared with the theoretical phase function calculated for clean bubbles (dashed line). The bubble distribution followed a normal distribution, with a mean radius of $25\mu\text{m}$. Because the reliable angular range of the measurement was from 10° to 170° , the measured data were scaled such that the integrations of the phase function between 10° and 170° for the measurements are the same as that of the theoretical calculation. The inset is in linear scale for angles from 90° to 170° . Copied from Zhang et al. (2002, Fig. 4) with permission.

It is impossible to directly measure the volume scattering that is due to bubble populations alone in nature. However, the scattering by oceanic bubbles can be inferred from field measurement of the total VSF (Zhang et al., 2002), even though the solution might not be unique, depending directly on how well the nature of scattering by other particles and the size distribution of all particles are represented.

Because little is known about small bubbles of sizes $< 10\mu\text{m}$ in nature, the optical properties of oceanic bubble populations in their entirety are still elusive. If the bubble concentration decreases at smaller sizes, then theoretical calculations and laboratory ob-

servations of the VSF for bubbles as shown in Fig. 8.28 are applicable to natural bubble populations; however, if the distribution is characterized by a continuous increase in the number density as size decreases (although clearly this must be bounded), then this could change the VSF in a way dependent on the size distribution of the small bubbles.

The injection of bubbles beyond the background population could significantly alter the light field both below and above the bubble layer. Being an efficient source of diffuse radiance, bubble layers contribute to a rapid transition to the diffuse asymptotic regime (Flatau et al., 1999) and enhance and modify the spectral reflectance above the surface (Flatau et al., 2000; Piskozub et al., 2009). Because of this, the typical ocean color algorithms for atmospheric correction (in particular, the black pixel assumption, Section 15.11.2) and for retrieval of the oceanic parameters (through band ratios) may no longer apply in the presence of additional bubbles (Yan et al., 2002; Zhang et al., 2004).

8.7 Aggregates

[Emmanuel Boss and Curtis Mobley wrote this section.]

Flocculation is the packaging of small particles into larger aggregated particles. The flocculation process is governed by a balance between particles colliding and sticking together and being torn apart by small-scale turbulence. The resulting aggregates (often called flocs or marine snow) are delicate aquatic particles composed of primary particles held together by polymers. They are often amorphous or “fluffy” and have sizes $> 100 \mu\text{m}$. A large fraction of an aggregate is often water ($> 90\%$ by volume) and hence its physical (e.g., settling speed) and optical properties (especially scattering) can be significantly different from the constituent particles. When mechanically disturbed (e.g., due to pumping through an instrument, sampling through the spigot of a Niskin bottle, or in the wake of a rosette) fragile aggregates may break apart, which often results in large changes in the optical properties of the water column.

Aggregation converts small particles into larger particles, which gives relatively fewer small particles and more large particles and thereby decreases the overall slope of the particle size distribution (PSD). Scattering depends strongly on the PSD (Eq. 8.10 above or Section 12.3), so aggregation can be expected to affect optical properties even though the mass remains unchanged. These changes are caused for two primary reasons. First, the packing of particles within aggregates is often dense enough that coherent interactions between scattered waves emanating from individual particles within the aggregate will cause a different scattering pattern than the simple superposition of scattering by the individual particles in suspension. Second, aggregate porosity is observed to grow with increasing aggregate size, so the cross-sectional areas of aggregates can be significantly larger than that derived by assuming that the solid mass is packed into a sphere of the same density as the component particles.

The difficulty of modeling scattering by aggregates can be understood as follows. Consider a spherical particle of diameter d . The volume of the particle is $V = (\pi/6)d^3$ and the cross-sectional area is $A = (\pi/4)d^2$. If two such particles are combined into a new spherical particle, and the density of material remains unchanged, then the new particle has twice the mass and twice the volume of an individual original particle, but the diameter increases by a factor of only $\sqrt[3]{2} = 1.26$. (This is the situation if two small raindrops coalesce into one larger drop.) The cross-sectional area thus increases by a factor of only $\sqrt[2]{\sqrt[3]{2}} = 1.58$.

Combining the two particles thus decreases their cross-sectional area compared to the total cross-sectional area of the two individual particles. Forward scattering (dominated by diffraction) is roughly proportional to the cross-sectional area, so in this case the aggregate scatters less than the two individual particles. However, the situation for actual marine aggregates is more complicated. Rather than the aggregate mass being proportional to d^3 , as for a homogeneous sphere, aggregate mass is found to be proportional to d^D , with D in the range of 1.5 to 2.5 (e.g., [Li and Logan, 1995](#), and references therein). In other words, aggregates are fractal particles ($D \neq 3$).

Consider an aggregate composed of a number N of component particles, and let F be the ratio of the gross volume of the aggregate to the total volume of the component particles. Then fractal theory shows that

$$F = N^{1-3/D} .$$

For “perfect” aggregation (e.g., the aggregation of small raindrops to make a larger one), $D = 3$ and $F = 1$. For oceanic aggregates, $D < 3$, so F decreases with increasing aggregate size, and F is typically in the range of 0.1 to 0.5 ([Latimer, 1985](#)). This complexity in aggregate structure greatly complicates modeling scattering properties (you certainly cannot use Mie theory to model scattering as an “equivalent-sphere” aggregate particle).

The role of aggregates in determining oceanic optical properties is remains poorly understood. However, aggregates have been found to contribute significantly to backscattering with little change in the backscattering-to-mass ratio as they formed in the lab or field ([Hatcher et al., 2001](#); [Flory et al., 2004](#)). [Boss et al. \(2009a\)](#) found that aggregation helps explain the relative constancy of scattering to mass or attenuation to mass found in coastal areas. These observations are inconsistent with traditional modeling of marine particles (e.g., using Mie theory), which assumes that particles are solid and which predicts a decrease in scattering efficiency per mass with size. Almost no measurements of scattering phase functions of aggregates have been made ([Hou et al., 1996](#)).

8.8 A Classic IOP Model for Case 1 Waters

The preceding sections of this chapter have surveyed the optical properties of individual constituents of natural waters. This section begins the discussion of modeling the bulk optical properties of waters that contain many types of particles and dissolved substances. Such models are often called *bio-geo-optical models* because they account for the contributions of biological (phytoplankton, CDOM, organic detritus) and geological (mineral particles) components of the water body.

This section presents an early model for the IOPs of Case 1 water. It is based on a reformulation by [Morel and Maritorena \(2001\)](#) of earlier work by [Priour and Sathyendranath \(1981\)](#) and [Gordon and Morel \(1983\)](#). This is the model seen in Eqs. (3.27) and (3.40) of *Light and Water* ([Mobley, 1994](#)). This IOP model is presented here for historical purposes and is retained in HydroLight for comparison purposes, where it is called the “Classic Case 1” IOP model in the user interface.

8.8.1 Absorption

The absorption coefficient is modeled as the sum of three components:

$$a_{\text{total}}(z, \lambda) = a_w(\lambda) + a_p(z, \lambda) + a_{\text{CDOM}}(z, \lambda), \quad (8.12)$$

where $a_w(\lambda)$ is the absorption by pure water, $a_p(z, \lambda)$ is the absorption by chlorophyll-bearing particles (phytoplankton) and co-varying detritus, and $a_{\text{CDOM}}(z, \lambda)$ is the absorption by co-varying colored dissolved organic matter (CDOM).

Absorption and scattering by water are discussed in Section 8.2.

8.8.1.1 Absorption by particles

The particle absorption in Eq. (8.12) is modeled by

$$a_p(z, \lambda) = 0.06 A_{\text{Chl}}(\lambda) \text{Chl}(z)^{0.65}, \quad (8.13)$$

where $\text{Chl}(z)$ is the chlorophyll concentration profile in mg Chl m^{-3} , and $A_{\text{Chl}}(\lambda)$ is the non-dimensional chlorophyll-specific absorption coefficient given in [Prieur and Sathyendranath \(1981, their \$a_c^*\$ \)](#) and in [Morel \(1988, Fig 10c\)](#), as extrapolated to 300 and 1000 nm. This $A_{\text{Chl}}(\lambda)$ is shown in Fig. 8.29. It should be noted that this $A_{\text{Chl}}(\lambda)$ is independent of the chlorophyll concentration. Thus only the magnitude of the particle absorption coefficient depends on Chl ; the *shape* of the a_p spectrum is the same for all Chl values. The original data covered 350-700 nm. Extrapolation of the original $A_{\text{Chl}}(\lambda)$ to 300 and 1000 nm was done by eye and is somewhat uncertain, especially for the UV wavelengths; see the discussion of absorption near 300 nm in Section 8.9.

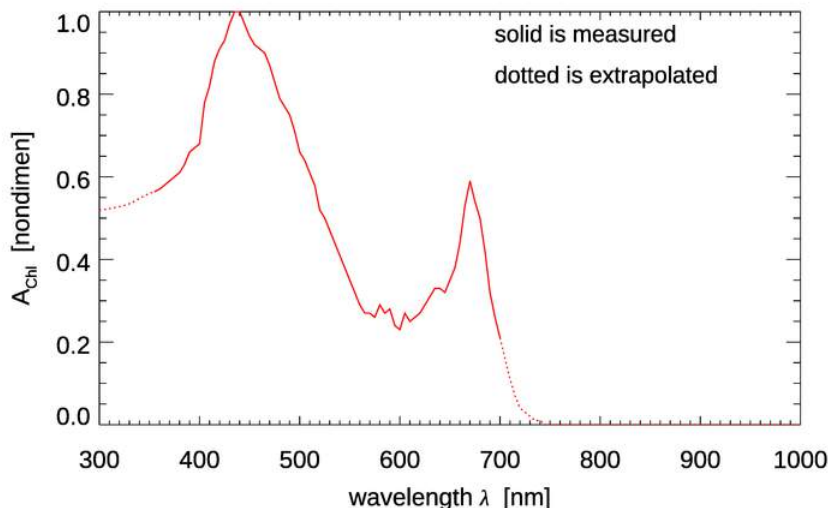


Figure 8.29: The non-dimensional chlorophyll-specific absorption coefficient $A_{\text{Chl}}(\lambda)$ used in Eq. (8.13). Data from [Prieur and Sathyendranath \(1981, for 350-700 nm\)](#).

8.8.1.2 Absorption by colored dissolved organic matter

The CDOM absorption is assumed to covary with particle absorption according to

$$a_{\text{CDOM}}(z, \lambda) = 0.2 a_{\text{p}}(z, 440) \exp[-0.014(\lambda - 440)]. \quad (8.14)$$

Thus the CDOM absorption at 440 nm is assumed to be 20% of the particulate absorption at 440 nm. This proportionality constant is an average of many measurements, as is the 0.014 wavelength slope factor. In a particular water body both of these values can be different.

Note that the absorption model of Eqs. (8.12-8.14) reduces to that of pure water if $Chl = 0$. This is slightly different than some formulations (e.g., Morel and Maritorena, 2001, Eqs. 16-18), which include a small amount of background CDOM absorption even in the absence of phytoplankton. The inclusion of a small amount of background CDOM is reasonable, but in all honesty, the residual was removed from the equations as used in HydroLight because I just got tired of explaining to HydroLight users why they didnt get exactly the same results as for pure water when they plugged in $Chl = 0$ in the Case 1 IOP model. In any case, the difference is negligible except at extremely low Chl values, in which case the model is suspect anyway.

8.8.2 Scattering

CDOM is assumed to be non-scattering.

The scattering coefficient for the particles is modeled by (Gordon and Morel, 1983)

$$b_{\text{p}}(z, \lambda) = b_{\text{o}} Chl(z)^n \left(\frac{550}{\lambda} \right)^m, \quad (8.15)$$

where the defaults are $b_{\text{o}} = 0.3$, $n = 0.62$, and $m = 1$.

The user of this IOP model must specify both the chlorophyll concentration $Chl(z)$ and the scattering phase function for the particles. One way to specify the particle phase function is to give the particle backscatter fraction, $B_{\text{p}} = b_{\text{bp}}/b_{\text{p}}$, and then use a Fournier-Forand phase function with the requested value of B_{p} , as described in Section 6.7 or Mobley et al. (2002).

Various formulas for B_{p} as a function of chlorophyll can be found in the literature. For example, Ulloa et al. (1994) give an empirical formula for B_{p} at 550 nm in Case 1 waters,

$$B_{\text{p}} = 0.01[0.78 - 0.42 \log_{10}(Chl)], \quad (8.16)$$

and Twardowski et al. (2001) present another formula,

$$B_{\text{p}} = 0.0096 Chl^{-0.253}. \quad (8.17)$$

Such formulas for B_{p} can be used for rough guidance in specifying the phase function in the absence of other information about the phase function. However, it must be remembered that scattering in general, and B_{p} in particular, correlates poorly with Chl (e.g., Fig. 6.11), and there can be order-of-magnitude variability in the measured value of B_{p} among different data sets for a given Chl . See Section 6.4 for further discussion of models for backscattering.

The particle backscatter fraction can be chosen to be the same at every wavelength, or be a function of wavelength according to the power law

$$B_p = B_p(\lambda_o) \left(\frac{\lambda_o}{\lambda} \right)^n . \quad (8.18)$$

As with all IOP models, this model may give good values on average, but can be considerably in error when applied to a specific water body. In any case, the use of the New Case 1 IOP Model described next is recommended for general use, especially near 300 nm.

8.9 A New IOP Model for Case 1 waters

This section develops an IOP model for Case 1 water based on more recent publications on absorption and scattering in Case 1 waters than those used for the classic IOP model of the previous section. The intention of this model was to have an IOP model that made use of the latest data (at the time of its creation) and was also convenient for use in the HydroLight radiative transfer model (Section 10.6), where it can be selected as the “New Case 1 IOP Model”. **In this model all IOPs are determined by the chlorophyll concentration** (after selecting low, medium, or high ultraviolet absorption). This model is presumably an improvement over previous models, although that remains to be proven by comparison with comprehensive measurements. In any case, the inherent limitations of IOP models must be remembered (in particular, see Section 4.4.3 or Mobley et al. (2004) for limitations of the “Case 1” concept). IOP models may be very good *on average*, but may or may not be (very often, *definitely are not*) correct in any particular instance. IOP models are therefore best used for “generic” studies. When modeling a particular water body, especially in a closure study, it is always best to use measured data to the greatest extent possible.

8.9.1 Absorption by Particles

It is well known that there is great variability in chlorophyll-specific absorption spectra $a^*(\lambda)$. In particular, the spectral shape of $a^*(\lambda)$ changes with the chlorophyll concentration, owing to species composition and pigment packaging effects (e.g., Bricaud et al., 1995, 1998) and Section 8.3.1. This is seen in Fig. 8.11. Another view of the chlorophyll dependence of absorption spectra is seen in Fig. 8.30. The left panels of this figure show the particulate absorption at 440 nm obtained from the spectra of Fig. 8.11. The right panels show the corresponding chlorophyll-specific absorption values, $a_p^*(440) = a_p(440)/Chl$.

Thus the next step in improving the particle absorption model is to allow the chlorophyll-specific absorption $a^*(\lambda)$ to depend on the chlorophyll concentration itself. Bricaud et al. (1998) therefore model particle absorption as

$$\begin{aligned} a_p(z, \lambda) &= a^*(Chl, \lambda) Chl(z) \\ &= A(\lambda) [Chl(z)]^{-B(\lambda)} Chl(z) \\ &= A(\lambda) [Chl(z)]^{E(\lambda)} . \end{aligned} \quad (8.19)$$

The Bricaud et al. (1998) paper gives $A(\lambda)$ and $E(\lambda)$ between 400 and 700 nm. Extending the Bricaud et al. values from 700 to 1000 nm is easy because phytoplankton

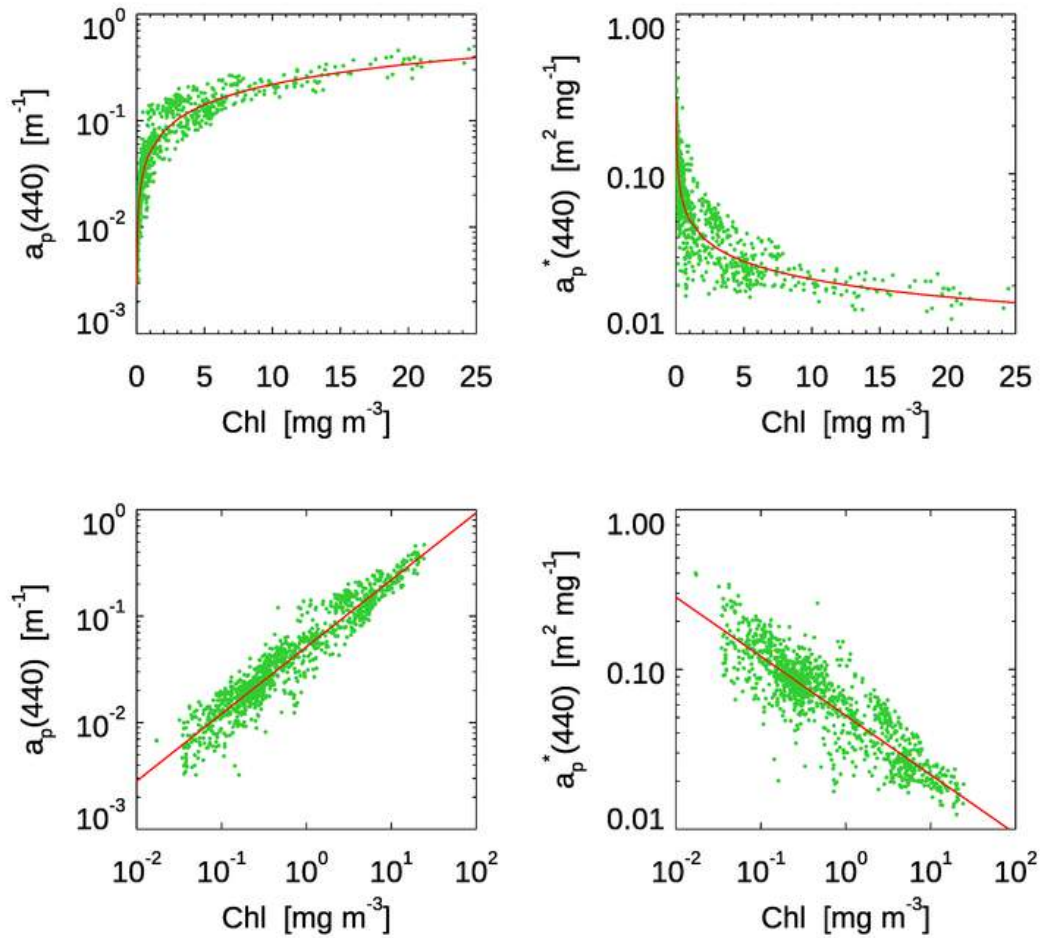


Figure 8.30: Particulate absorption at 440 nm as a function of chlorophyll for Case 1 water. Left panels: particulate absorption $a_p(440)$; right panels: the corresponding chlorophyll-specific absorption $a_p^*(440)$. The top row is chlorophyll on a linear scale; the bottom row is chlorophyll on a log scale. The red curves are the prediction of Eq. (8.19). The lower left curve reproduces Bricaud et al. (1998, Fig. 1a). The plotted values were obtained from the spectra of Fig. 8.11.

absorption is essentially zero in the infrared. However, there are very few measurements of phytoplankton absorption below 350 nm, so extending $A(\lambda)$ and $E(\lambda)$ down to 300 nm is at best an uncertain process.

Morrison and Nelson (2004, Fig. 1) show two normalized phytoplankton absorption spectra from 300 to 750 nm taken at the Bermuda Atlantic Time Series (BATS) site. The BATS Chl values ranged between 0.002 and 0.606 mg m⁻³ over the course of a year, with a mean of 0.152. Although their spectra are similar above 365 nm, they are highly variable with season and depth between 300 and 365 nm. This variability is likely due to mycosporine-like amino acids (MAAs), which strongly absorb near 320 nm. Figure 8.31 compares the Morrison and Nelson spectra (blue curves) with the Bricaud et al. a_p of Eq. (8.19) evaluated for $Chl = 0.05$ mg m⁻³ (red curve); the Morrison and Nelson spectra

are normalized to the Bricaud value of $a_p(400)$. The shapes of the Morrison and Nelson spectra are consistent with the Bricaud values for low Chl values.

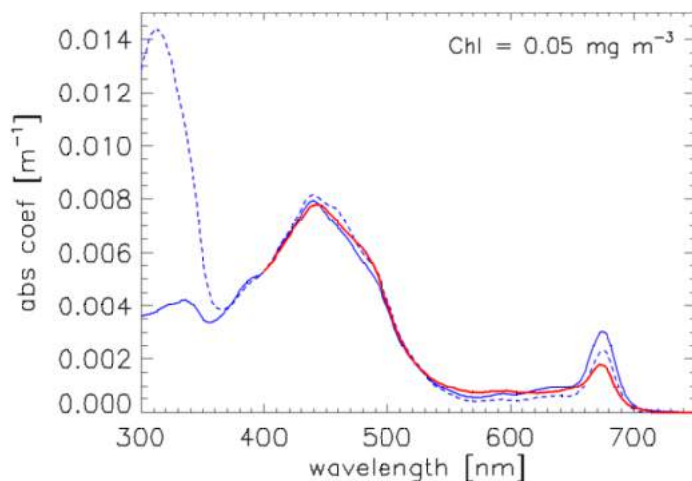


Figure 8.31: Comparison of Bricaud et al. (1998) a_p for $Chl = 0.05$ (red) with the Morrison and Nelson (2004) normalized absorption spectra (blue; dotted is summer, solid is winter).

Vasilkov et al. (2005) present spectra for $A(\lambda)$ and $B(\lambda) = 1 - E(\lambda)$ between 300 and 400 nm, as derived from absorption measurements in coastal California waters. Figure 8.32 shows their $A(\lambda)$ and $B(\lambda)$ spectra compared with those of Bricaud et al. (1998). The differences at 400 nm reflect the different databases (i.e., different waters) used to derive the coefficients.

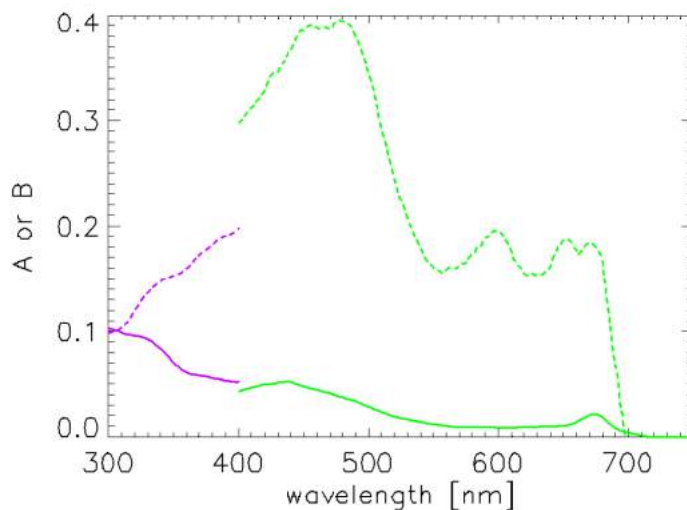


Figure 8.32: Comparison of the Vasilkov et al. (2005) $A(\lambda)$ and $B(\lambda)$ (purple curves) with those of Bricaud et al. (1998) (green curves). Solid lines are A and dashed lines are B .

Desperation is the mother of modeling (you can quote me on that), so to define $A(\lambda)$ and $B(\lambda)$ over the 300-1000 nm range, I proceeded as follows. The Bricaud et al. A and

B curves are accepted for use from 400 to 720 nm, with $A = B = 0$ between 720 and 1000 nm. The Vasilkov et al. curve for A was normalized to the Bricaud value at 400 nm, i.e., $A_V(\lambda) \leftarrow [A_V(\lambda)/A_V(400)]A_B(400)$, where subscripts V and B stand for Vasilkov et al. and Bricaud et al., respectively. The normalized $A_V(\lambda)$ was then averaged with the two normalized spectra of Morrison and Nelson (2004) seen in Fig. 8.31, *assuming* that the A spectra have the same shape as a_p . This assumption about the shapes of A and a_p is correct only if $B = 0$ or if $Chl = 1$, in which case $A = a^*$ in Eq. (8.19). The resulting average A between 300 and 400 nm then merges smoothly with the A of Bricaud at 400 nm. For B , the Vasilkov et al. curve was normalized to the Bricaud et al. curve at 400, and the result was used to extend the Bricaud et al. B down to 300 nm. The resulting A and B spectra are shown in red in Fig. 8.33, along with the three A spectra used to compute the average A between 300 and 400. These A and B give an absorption model that roughly corresponds to the mid-range of UV absorptions seen in the Morrison and Nelson data. The new Case 1 IOP model uses these A and $E = 1 - B$ as the default spectra for the last version of Eq. (8.19). The tabulated A and E spectra can be downloaded from the Web Book.

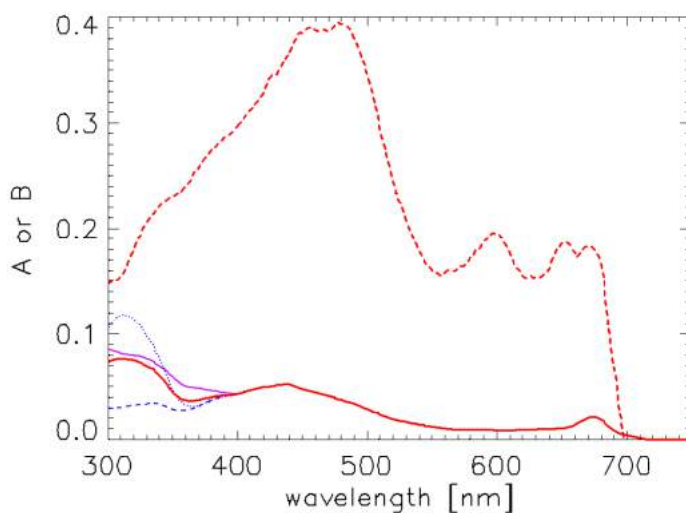


Figure 8.33: The A (red solid line) and B (dashed line) spectra used in Eq. (8.19) to define a_p for the mid-range of UV absorption. The purple and blue curves were averaged to produce A between 300 and 400 nm.

It would be interesting to compare light fields for the wide range of UV absorptions illustrated by the Morrison and Nelson (2004) spectra of Fig. 8.33. A spectra for low and high UV absorptions are therefore defined by simply using the shapes of the Morrison and Nelson spectra absorption spectra to extend the Bricaud A from 400 down to 300 nm. The B spectra were taken to be the same as for the mid-range of UV absorption just discussed. These spectra, plus those for low and high UV absorption, can be downloaded from the Web Book page on the [New IOP Model](#).

Regardless of which set of A and E spectra is chosen, the A and E spectra are used in the same manner in Eq. (8.19) to define $a_p(\lambda)$ for any Chl value. Figure 8.34 shows the resulting particle absorption spectra for low, medium, and high UV absorptions and

for $Chl = 0.01, 0.1, 1.0,$ and 10.0 mg m^{-3} . The corresponding absorption coefficients as computed by the classic Case 1 IOP model are shown for comparison, as is absorption by pure water. There are significant differences in the classic and new models, which will lead to significant differences in computed radiances, irradiances, and AOPs when used in radiative transfer calculations. Note in particular that the shape of the particle absorption spectrum now changes with the chlorophyll value. Presumably the new model gives a more realistic description, on average, of particle absorption in Case 1 waters than does the classic model for which the shape of the particle absorption spectrum is independent of the chlorophyll concentration.

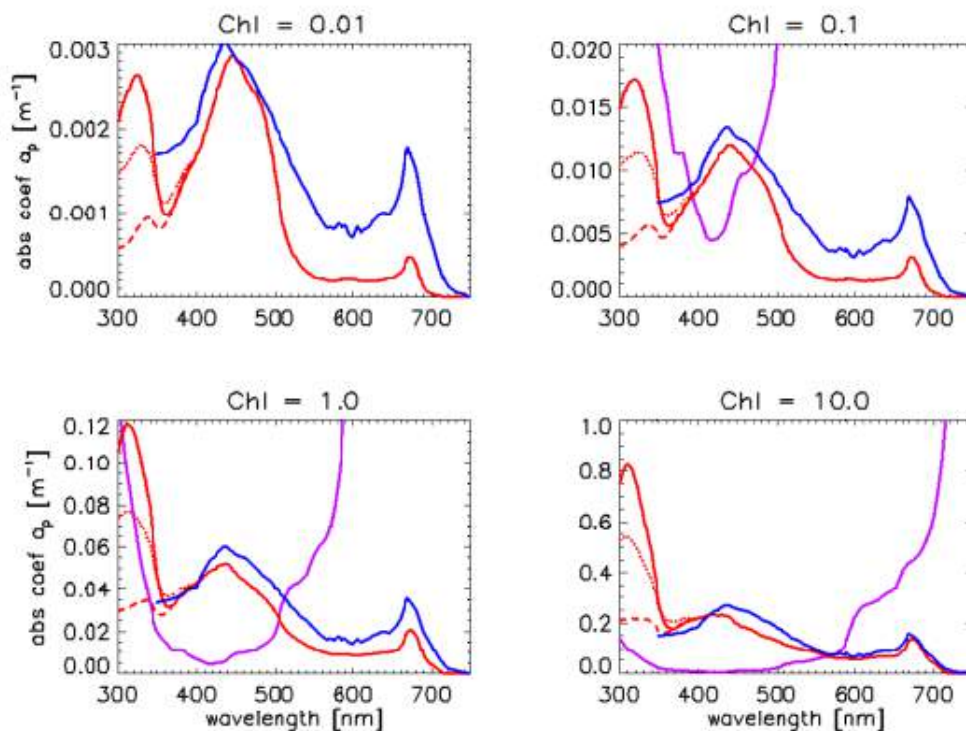


Figure 8.34: Particle absorption coefficients computed by the “New Case 1” model of Eq. (8.19) (red) and the “Classic Case 1” IOP model of Eq. (8.29) (blue). The red solid line is for the high-UV absorption, dotted is the mid-range UV absorption, and dashed is the low-range UV absorption. The purple line is absorption by pure water.

8.9.2 Absorption by CDOM

As has been seen in the CDOM section 8.4 and in Fig. 8.15, for wavelengths greater than 300 nm, CDOM absorption is acceptably well modeled by a function of the form

$$a_{\text{CDOM}}(z, \lambda) = a_{\text{CDOM}}(z, \lambda_0) \exp[-S_{\text{CDOM}}(\lambda - \lambda_0)]. \quad (8.20)$$

The functional form (8.20) underestimates CDOM absorption at wavelengths less than 300 nm and therefore should not be used below 300 nm. When incorporated into the new Case 1 IOP model, $a_{\text{CDOM}}(z, \lambda_0)$ is set to $f_{\text{CDOM}} a_{\text{p}}(z, \lambda_0)$, with default values of $f_{\text{CDOM}} = 0.2$,

$\lambda_o = 440$ nm, and $S_{\text{CDOM}} = 0.014 \text{ nm}^{-1}$ just as in the classic Case 1 model. However, the default values of f_{CDOM} , λ_o , and S_{CDOM} can be changed in HydroLight if desired.

8.9.3 Scattering

Just as for absorption, recent papers have presented improved models for particle scattering in Case 1 waters. Morel et al. (2002, Eq. 14) model the particle scattering coefficient as

$$b_p(z, \lambda) = b_o [\text{Chl}(z)]^n \left(\frac{\lambda}{550} \right)^\nu, \quad (8.21)$$

where

$$\nu = \begin{cases} 0.5 [\log_{10} \text{Chl} - 0.3] & \text{for } 0.02 < \text{Chl} < 2 \\ 0 & \text{for } \text{Chl} > 2. \end{cases} \quad (8.22)$$

Equation (8.21) with $\nu = -1$ is the “classic” scattering model proposed by Gordon and Morel (1983). The wavelength dependence of the new scattering coefficient now depends on the chlorophyll concentration. In particular, ν now lies between -1 and 0. A value of $\nu = -1$, as often used in earlier models, is known from Mie theory to be valid only for nonabsorbing particles with a Junge particle size distribution slope of -4.

Loisel and Morel (1998) studied the relationship between particle beam attenuation at 660 nm, $c_p(660)$, and the chlorophyll concentration. They found the functional form

$$c_p(z, 660) = c_o [\text{Chl}(z)]^n. \quad (8.23)$$

The values of c_o and n are different for near-surface (down to one “penetration depth,” as relevant to remote sensing) and deeper waters. Because $c_p \approx b_p$ at 660, Morel et al. (2002) adopt the coefficients for Eq. (8.23) for use in Eq. (8.21), after shifting the reference wavelength to 550 nm. Thus for near-surface waters, Morel et al. (2002) use $b_o = 0.416$ and $n = 0.766$ in Eq. (8.23).

However, a power law in wavelength of the form of Eq. (8.23) generally gives a better fit for c_p than for b_p (e.g., Voss, 1992; Boss et al., 2001). Thus it is probably better to model c_p and then obtain b_p from $b_p = c_p - a_p$ (with a_p being determined by Eq. (8.19) as described above). This is the approach taken in the new Case 1 model, which uses

$$c_p(z, \lambda) = c_o [\text{Chl}(z)]^n \left(\frac{\lambda}{660} \right)^\nu, \quad (8.24)$$

where the coefficients are the same as for b in Eq. (8.21) above and ν is given by Eq. (8.22). Thus this model uses the chlorophyll-dependence of $c_p(660)$ from Loisel and Morel (1998), Eq. (8.24), and assumes that c_p has the same chlorophyll-dependent wavelength dependence as the b_p model of Morel et al. (2002), Eq. (8.21). The default values of c_o and n , which apply to near-surface waters, are $c_o = 0.407$ and $n = 0.795$ (from Loisel and Morel, 1998, Eq. 5); other values can be used if desired.

Figure 8.35 shows example a_p for mid-range UV absorption, b_p , and c_p spectra for near-surface waters ($c_o = 0.416$ and $n = 0.766$ in Eq. 8.24), along with the “classic” b_p with $b_o = 0.3$, $n = 0.62$, and $m = 1$. The scattering coefficients are not too different at low chlorophyll values, but the new b_p has a different wavelength dependence and is

much larger in magnitude, by up to a factor of three, at high Chl values. Unlike in the classic scattering model, the wavelength dependence of b_p now depends on Chl and is more complicated. These differences in scattering will have a significant effect on computed radiances.

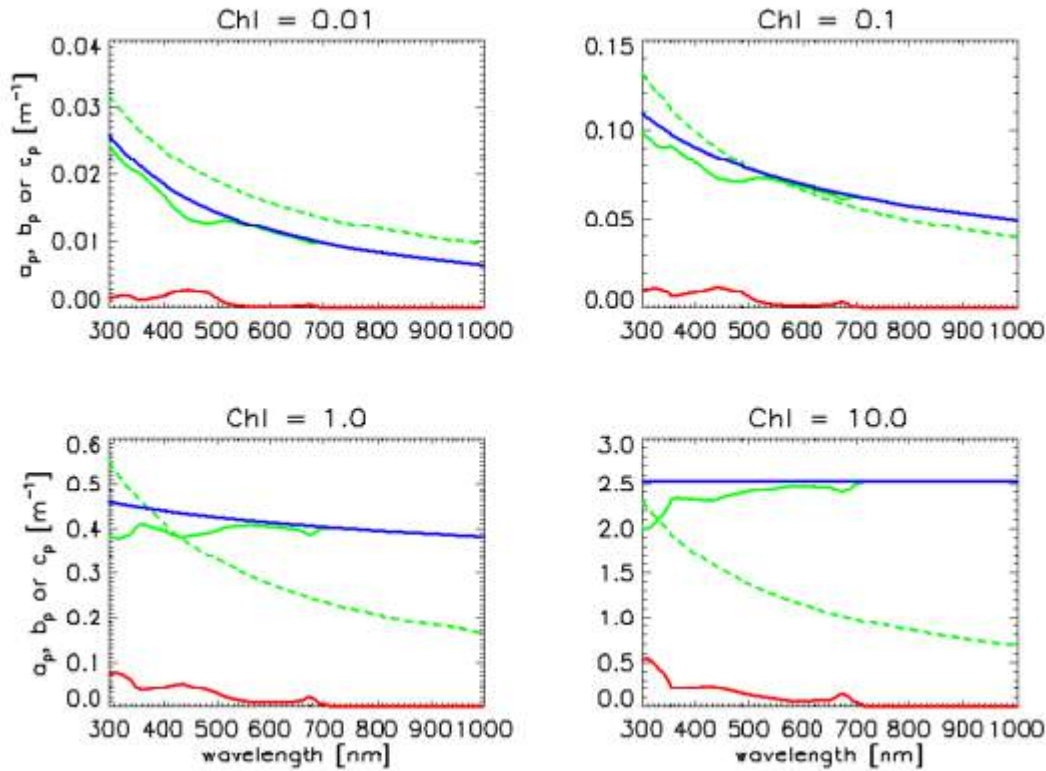


Figure 8.35: Beam attenuation c_p as determined by from Eq. (8.24) and near-surface values of $c_o = 0.407$ and $n = 0.795$ (blue). The red curves are the same a_p as the mid-UV absorptions in Fig. 8.34. The solid green curve is the new $b_p = c_p - a_p$; the dashed green curve is the “classic” b_p of Eq. (8.21) with $\nu = -1$.

8.9.4 Scattering Phase Function

Morel et al. (2002) also developed a phase function model for Case 1 water in which the phase function is a combination of “small” and “large” particle phase functions, with the fraction of each being determined by the chlorophyll concentration:

$$\tilde{\beta}_p(\psi, Chl) = \alpha_s(Chl) \tilde{\beta}_s(\psi) + \alpha_l(Chl) \tilde{\beta}_l(\psi), \quad (8.25)$$

where

$$\alpha_s(Chl) = 0.855 [0.5 - 0.25 \log_{10}(Chl)] \text{ and } \alpha_l = 1 - \alpha_s. \quad (8.26)$$

Figure 8.36 shows phase functions determined by Eqs. (8.25) and (8.26), along with the frequently used Petzold average-particle phase function 9.5.2.

It should be noted that the Morel et al. (2002) phase functions have smaller backscatter fractions ($B_p = 0.014$ for the small particles to 0.0019 for the large particles) than the

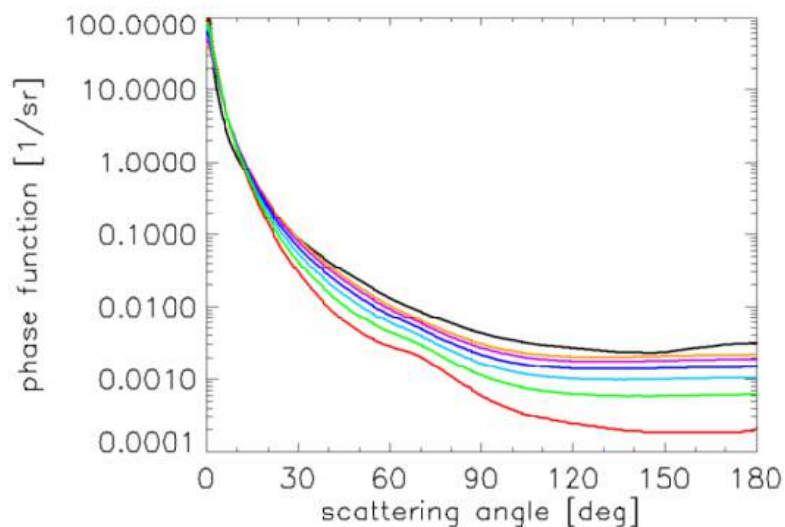


Figure 8.36: Phase functions for small (orange) and large (red) particles as given by Morel et al. (2002). Phase functions as given by Eqs. (8.25) and (8.26) for $Chl = 0.01$ (purple), 0.1 (blue), 1.0 (teal), and 10.0 (green), and the Petzold average particle phase function (black) are also shown.

Petzold phase function ($B_p = 0.018$). This is consistent with what is known about the phase functions for algal particles (e.g., Ulloa et al., 1994; Twardowski et al., 2001). The Morel phase functions of Eq. (8.25) and Fig. 8.36 are adopted for use in the new Case 1 IOP model.

To illustrate the quantitative differences (including the combined effects of absorption and scattering coefficients and the particle scattering phase function) between the classic and new Case 1 IOP models, Fig. 8.37 shows the remote-sensing reflectance $R_{rs}(\lambda)$ for homogeneous, infinitely deep waters with $Chl = 0.01, 0.1, 1,$ and 10 mg Chl m^{-3} , as computed for the classic and new Case 1 IOP models. The mid-range UV absorption model was used in the new model. The Sun was placed at a zenith angle of 30 deg in a clear sky with typical marine atmospheric parameters (sky irradiances were computed using the RADTRAN-X sky irradiance model for the annual average Earth-Sun distance). The wind speed was 6 m s^{-1} . For the classic IOP model, the particle phase function was taken to be a Fournier-Forand phase function with a backscatter fraction as given by the empirical formula

$$B_p = 0.01 [0.78 - 0.42 \log_{10}(Chl)]$$

of Ulloa et al. (1994) for B_p at 550 nm in Case 1 waters. These IOPs were then used in the HydroLight radiative transfer model, which was run from 300 to 800 nm with 5 nm bands.

Figure 8.37 shows that the computed R_{rs} spectra are very similar for $Chl = 0.01$ and 0.1, but that the differences can become very large at high chlorophyll values. The maximum difference computed as $100(\text{new} - \text{old})/\text{old}$ is less than 5% for $Chl = 0.01$ or 0.1. For $Chl = 1$, the maximum difference is less than 50% at visible wavelengths (62% at 798 nm). For $Chl = 10$, the differences are as large as 254% (more than a factor of three; at 578 nm). The larger R_{rs} for high Chl is due to the greatly increased scattering, as seen in Fig.

8.36.

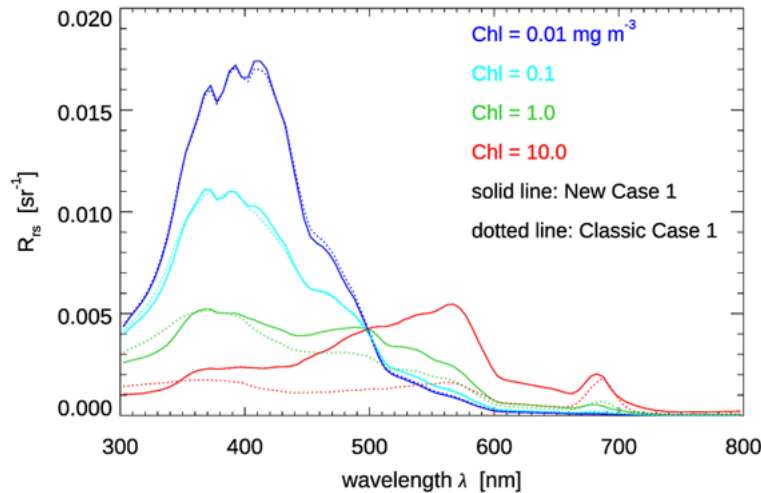


Figure 8.37: Comparison of R_{rs} as computed by the New Case 1 model with mid-range UV absorption (solid lines) and the Classic model (dotted lines), for $Chl = 0.01, 0.1, 1,$ and 10 mg Chl m^{-3} .

8.10 Commonly Used Models for IOPs and Biogeochemistry

[This section was written by Emmanuel Boss.]

This section presents a number of commonly used models relating inherent optical properties (IOPs) to the underlying biogeochemistry. Models for IOPs and AOPs are analytical expressions relating these optical variables to bio-geochemical parameters (e.g. chlorophyll, suspended matter) and/or describing their spectrum (relating their value at one wavelength with their value at another wavelength). Below is a “laundry list” of such models assembled from the literature. For a detailed discussion, see the review paper by [Sosik \(2008\)](#). The users of such models are cautioned that they were developed from specific data sets and designed with specific applications in mind, which may or not be applicable to the conditions to which the user is applying them. Also, it is important to note that the fit parameters will vary depending on the way a model is fit to the data (e.g. how uncertainties are assumed to behave) and the spectral range that is fit (e.g., [Twardowski et al., 2004](#)). Note that some of the observed variability in relationships is likely due to methodology in biogeochemical determinations (e.g. filtration) and instrumental issues, e.g., spectral filters used (narrow versus wide) and acceptance angle (e.g., [Boss et al., 2009b](#)). In addition, empirical relationships are likely to be biased to time and location of data used to derive them, and their generalization should be done with caution.

8.10.1 Models Linking CDOM Absorption to Biogeochemical Parameters

Models for CDOM absorption as a function of wavelength were discussed in Section 8.4. CDOM absorption at a specific wavelength can also be related to the concentration of dissolved organic carbon (DOC). In estuaries and coastal waters, CDOM and fluorescence by dissolved organic matter (DOM) vary in correlation with DOM (e.g., Blough and Green, 1995) via relationships of the type

$$a_{\text{CDOM}}(450) = (0.007 - 1.76)DOC \quad [\text{m}^{-1}]$$

for a variety of environmental samples as well as extracted fulvic and humic materials and where DOC has units of mg organic C L⁻¹. When restricted to whole environmental samples (and including data from Vodacek et al. (1997)), the relation is

$$a_{\text{CDOM}}(450) = (0.33 - 1.23)DOC \quad [\text{m}^{-1}].$$

Such relationships are not observed in open waters (Nelson and Siegel, 2002). However, the values of DOC observed in the open ocean (e.g. 48-68 μmol L⁻¹, (Nelson and Siegel, 2002)), are of the similar magnitude as the intercept of $a_{\text{CDOM}}-DOC$ regressions ($\sim 70 \mu\text{mol L}^{-1}$ (Vodacek et al., 1997)) and hence represent, to a large extent, the surface pool of uncolored DOC. These relationships arise from end-member mixing between terrestrial and oceanic water masses and do not hold in coastal areas not strongly affected by river inputs and where CDOM sinks (e.g. photo-oxydation) affect CDOM concentrations significantly (Blough and Vecchio, 2002).

Between rivers and estuaries $a^*(450) = a_{\text{CDOM}}(450)/DOC$ increases with increases in aromatic content and thus with lower CDOM spectral slopes (Blough and Vecchio, 2002).

Prieur and Sathyendranath (1981) suggest the following model (recall Eqs. 8.13 and 8.14)

$$a_{\text{CDOM}}(440) = 0.2 [a_w(440) + 0.06 Chl^{0.65}] .$$

That is, $a_{\text{CDOM}}(440)$ is 20% of the absorption by water plus chlorophyll at 440 nm. Babin et al. (2003b) have also found a linear relationship between CDOM and Chl for European waters.

8.10.2 Non-algal Particles

Similar to CDOM, the absorption of non-algal particles (NAP) $a_{\text{NAP}}(\lambda)$ is usually modeled with a decreasing exponential function (Yentsch, 1962; Kirk, 1980; Roesler et al., 1989; Bricaud et al., 1998) as seen in Eq. (8.8):

$$a_{\text{NAP}}(\lambda) = a_{\text{NAP}}(\lambda_o) \exp^{-S_{\text{NAP}}(\lambda-\lambda_o)} \quad [\text{m}^{-1}],$$

where λ_o is a reference wavelength and S_{NAP} the spectral slope (independent of λ_o). The mean slope (S_{NAP}) generally used to model is 0.011 nm⁻¹ (e.g., Roesler et al., 1989; Bricaud et al., 1998). However, as always with data-derived parameters, there is variability in S_{NAP} . Figure 8.18 shows the distribution of S_{NAP} values observed by Babin et al. (2003b). Their data show a mean of $S_{\text{NAP}} = 0.0123 \text{ nm}^{-1}$ with a standard deviation of 0.0013. It should be noted that the exponential function is only an *approximation* and that realistic NAP

spectra may be non-monotonic and often exhibit a “hump” in the blue (e.g. [Iturriaga and Siegel \(1989\)](#)).

Both CDOM and NAP display similar exponential absorption spectra, although with somewhat different spectral slopes. In modeling their absorption effects, CDOM and NAP are often combined and described by an exponential. Note however, that CDOM and NAP have much different scattering properties. CDOM is assumed to be non-scattering, but NAP are highly scattering.

For non-algal particles collected both in coastal and riverine waters and from mineral samples, [Babin et al. \(2003b\)](#) and [Babin and Stramski \(2004, Fig. 10\)](#) found

$$a_p(443) = (0.03 - 1.0) PM$$

where PM is the concentration of particulate matter in g m^{-3} , with the high values being associated with high iron-oxide content in the “natural assemblages of mineral particles.” Relationships with iron concentrations are considerably tighter ([Babin and Stramski, 2004](#)):

$$a_p(443) = (1 - 4) Fe,$$

where Fe is the concentration of iron in g Fe m^{-3} .

8.10.3 Particulate Organic Carbon (POC)

$$c_p(666) = (0.06 - 0.3), POC \quad [\text{m}^{-1}].$$

where POC is in mol m^{-3} .

8.10.4 Particulate Matter or Total Suspended Matter

$$b_p(555) = (0.2 - 1) PM \quad [\text{m}^{-1}].$$

where PM is in g m^{-3} .

8.10.5 Global Particulate Scattering

In open-ocean environments [Morel \(2008\)](#) found

$$b_p(550) = (0.15 - 0.45) Chl^{0.62} \quad [\text{m}^{-1}],$$

while in more turbid Case 2 waters the leading coefficient exceeds 0.45 and the exponent can be different ([Morel and Maritorena, 2001](#)); Chl is in mg m^{-3} . For the upper layer, and based on more recent measurements, [Loisel and Morel \(1998\)](#) found

$$b_p(550) = 0.4 Chl^{0.76} \quad [\text{m}^{-1}].$$

[Babin et al. \(2003b\)](#) found that

$$b_p(555) = (0.5 - 1) PM \quad [\text{m}^{-1}].$$

where PM is the particulate matter concentration in g m^{-3} . The lower values come from turbid coastal areas while the open water values are higher. This relatively tight relationship was explained as arising from the relative insensitivity to particle composition (PM is the dried mass) using theoretical calculations. [Boss et al. \(2009a\)](#) showed that the relative insensitivity of this relationship to variability in size composition may be due to aggregation.

8.10.6 Global Particulate Absorption

In open ocean environments Morel (2008) found

$$\begin{aligned} a_p(440) &= 0.052 Chl^{0.64}, \\ a_p(675) &= 0.02 Chl^{0.82} \\ a_p(550-620) &< 0.01 Chl^{0.85} \quad [\text{m}^{-1}]. \end{aligned}$$

Phytoplankton absorption $a_\phi(\lambda)$ is always smaller than particle absorption $a_p(\lambda)$ by about 30% in absorbing bands and by more than 100% in weakly absorbing bands such as in the green part of the spectrum.

8.11 Optical Properties of Shelf Seas and Estuaries

[David Bowers wrote this section.]

The majority of the literature on ocean optics has concentrated, understandably, on the majority of the ocean. For most people, however, their nearest stretch of salt water is a *shelf sea*: water lying on a continental shelf and less than 200 meters deep. Shelf seas occupy 6-8% (depending on definition) of the surface area of the world ocean but contain about 16% of the world's phytoplankton biomass. These seas are also rich in mineral particles stirred up from the sea bed by tides and waves, and in dissolved organic matter brought in by rivers. The waters are colored green or blue-green by the mixture of pigments in phytoplankton, colored dissolved organic matter (CDOM) and minerals. The optical properties of phytoplankton and their pigments are dealt with in other sections of this document and in several excellent textbooks on marine optics (e.g., Kirk, 1994) but CDOM and mineral particles have their own special features in shelf seas and deserve an account of their own.



Figure 8.38: Shelf Sea waters often appear green to the eye. The picture at the left was taken from a research ship in clear, deep, shelf sea waters and the one at the right in more energetic, tidally stirred waters in which sediment is stirred up from the bed to the sea surface. The effect of the suspended particles in coloring the water green and also making it opaque can be sensed in these pictures. Photos: D.G. Bowers

Shelf seas are considered to be optically complex because the CDOM and mineral particles can make a significant (or even the dominant) contribution to water color and brightness. Both CDOM and mineral particles absorb light most strongly in the blue part of the spectrum and so, when added to water (which absorbs mostly red light), they produce a green color which is easily confused with the effect of phytoplankton pigments. For this reason, satellite remote sensing algorithms for chlorophyll regularly fail in shelf seas in the sense that they over-estimate the chlorophyll content. In remote sensing terms, these waters are classified as Case 2, (Morel and Prieur, 1977) and Section 4.4.3, as distinct from Case 1 open ocean waters. The mineral particles suspended near the sea surface are also excellent scatterers of light. Because of this, shelf seas appear bright (highly reflective) when viewed from space.

8.11.1 CDOM in Shelf Seas

Colored dissolved organic matter is produced when organic material decays. In the open ocean, the main source of CDOM is from the decay of phytoplankton, and there is sometimes a correlation between the concentration of chlorophyll and the concentration of CDOM, which is helpful in remote sensing chlorophyll in these waters. Shelf seas in temperate latitudes, however, receive significant fresh water input from the land and this brings with it CDOM produced by the decay of terrestrial plants. This can be the main source of CDOM in shelf seas, particularly near the coast, dominating the marine production of CDOM. In these circumstances it is not reasonable to expect there to be a correlation between CDOM and chlorophyll concentrations in shelf seas. As the fresh water runoff mixes with sea water, the land-produced CDOM is diluted. This leads to a negative correlation between CDOM and salinity of the form

$$g = g_s \left(1 - \frac{S}{S_0} \right) \quad (8.27)$$

where g is the CDOM concentration at a point where the salinity is S , g_s is the concentration in the source water (where the salinity is zero) and g_s/S_0 is the gradient of a plot of g against S . Since CDOM is often measured optically, the “concentration” of CDOM, g and g_s may be expressed as its absorption coefficient of a filtered water sample at a chosen wavelength (Kirk, 1994). An example of the relationship between CDOM absorption and salinity in an estuary is shown in Fig. 8.39.

8.11.2 Effect of Variations in Source Concentration

A linear relationship between CDOM and salinity (obeying Eq. (8.27)) is observed surprisingly often (although not always) in shelf seas and estuaries. This is taken to indicate “conservative mixing”, that is the CDOM mixes passively with water in the same way as salt. This is surprising, since it is known that CDOM is consumed in water (by bacteria, and it is also destroyed by sunlight) and it is produced by the decay of organic matter at sea. The observation of a linear relationship is also surprising because the concentration in the source water can be expected to change with time. The intercept of a plot of CDOM against salinity at zero salinity is not fixed but moves up and down in response to seasonal and other changes in CDOM production at the source.

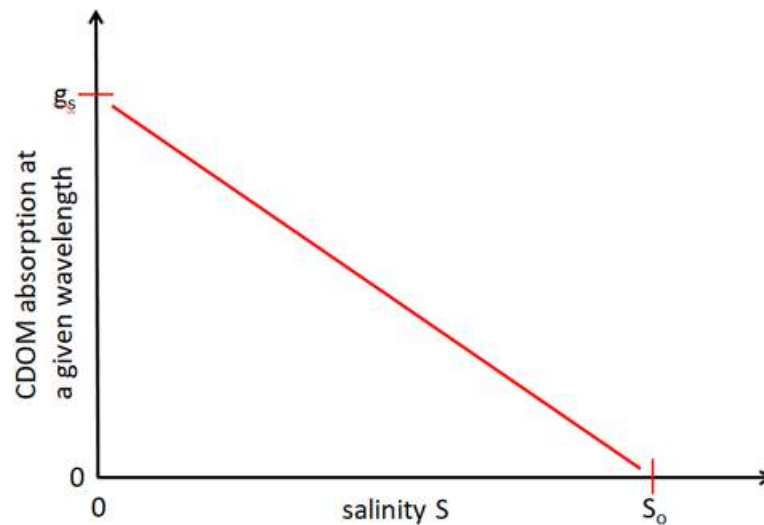


Figure 8.39: Typical variation of CDOM concentration with salinity in an estuary. CDOM concentration is expressed as the absorption coefficient of filtered seawater at a chosen wavelength. The concentration of CDOM in the source river water is g_0 .

Bowers and Brett (2008) have analysed the effect of variations in source concentration on the CDOM-salinity relationship in shelf seas and estuaries using a simple box model. They found that in a water body with a short flushing time compared to the timescale of the source variation (for example, a small body of water such as an estuary), the concentration of CDOM can “keep up” with the fluctuations in the source and so the linear relationship between salinity and CDOM is maintained, although the slope changes with time (as illustrated in Fig. 8.39). In larger water bodies, such as a bay or a gulf, the slow response time of the water body buffers it against rapid changes in the source concentration, so again there is a linear relationship between salinity and CDOM, this time with a fixed slope set by the long term mean concentration of CDOM in the source.

8.11.3 Using CDOM to Trace Water Masses in Shelf Waters

A useful technique in classical physical oceanography is water mass analysis, which uses temperature-salinity diagrams to identify the source of water in the ocean and mixing between different sources. This technique does not work in shelf seas (nor in the surface waters of the ocean) because temperature is affected too much by the heat flux through the sea surface. Instead, Stedmon et al. (2010) have proposed using CDOM instead of temperature to trace water masses in shelf seas. Water samples are placed on an x-y plot which has CDOM concentration (i.e. absorption at a chosen wavelength) on one axis and salinity on the other, that is a diagram like Fig. 8.39. If sources can be identified by their characteristic salinity-CDOM value (and for a large water body, the long term mean values can be used), then the proportion of water in a given sample from either two or three sources can be established. Stedmon and co-workers applied their technique to the Baltic-North Sea transition zone. The three end-member water masses were the German Bight, the Baltic outflow and the central North Sea. The method gave believable results

for the distribution of these 3 water masses in the study area. The method assumes that CDOM behaves conservatively. Stedmon et al. (2010) tested this assumption by using a third variable: the spectral slope of CDOM absorption, and concluded that the CDOM was behaving conservatively.

8.11.4 Remote Sensing of CDOM and Salinity

CDOM in coastal waters acts as a dye (see Fig. 8.40) producing a coloration which can be seen from space. In water in which CDOM is the principal optical component this means that it is possible to measure CDOM concentration using ocean color and, because of the relationship between CDOM and salinity, this technique can also be used to map salinity in coastal water bodies. A number of empirical relationships between water color and CDOM have been proposed (Kutser et al., 2005; Del Castillo and Miller, 2008) but it is possible to arrive at a theoretically sound relationship using the fact that CDOM is a strong absorber in the blue and a weak one in the red.



Figure 8.40: A flask of clear water (left) and one containing CDOM (right). Flask on the right was prepared by filtering coastal water through $0.2\mu m$ filters to remove particles. The coloration produced by CDOM in shelf seas and estuaries is sufficient to remotely sense its concentration using satellite ocean color data. Photo: B. D. Moate

The reflection coefficient at a given wavelength just below the water surface is proportional to the ratio of the backscattering to absorption coefficient at that wavelength (Morel and Prieur, 1977). Assuming that (i) absorption in the red is dominated by water, (ii) absorption in the blue is due to CDOM and water and (iii) the ratio of backscattering coefficients in the red and the blue is constant, it is possible to write an expected relationship between CDOM absorption in the blue and a ratio of reflection coefficients of the form:

$$g_{440} = k_1 \frac{R_R}{R_X} + k_2$$

Where g_{440} is the absorption coefficient of CDOM at 440nm (a proxy for its concentration), R_R is the reflection coefficient in the red and R_X the reflection coefficient at another (shorter) wavelength. k_1 and k_2 are constants provided CDOM is the predominant influence on water color. Binding and Bowers (2003) found a good ($R^2 = 0.94$) fit for this

expression against observations for $R_X = R(490)$, $R_R = R(665)$, $k_1 = 0.635$, $k_2 = 0.103$. They used this expression to map surface CDOM concentrations in the Clyde Sea in Scotland and to derive salinity using the known relationship between CDOM and salinity in the Clyde.

The application of water color to derive CDOM concentration and salinity remotely is most appropriate in estuaries where the strongest signal is to be found. Unfortunately, the color of the water in an estuary is also likely to be influenced by suspended particles (including mineral particles), although it may be possible to correct for this if the optical properties of the particles are known.

8.11.5 Suspended Mineral Particles in Shelf Seas

The combination of shallow water and high energy input (through waves and tides) in shelf seas means that mineral particles are lifted off the sea bed and mixed throughout the water column. The particles collide and stick together to form *flocs* (Section 8.7). This process is particularly important during and after the spring bloom when there is plenty of organic material in the water to help the flocculation process (the organic material acts as a “glue” to hold the flocs together). Since flocs tend to have a higher settling speed than the constituent particles this flocculation process after the spring bloom is important in making surface waters in shelf seas becoming clearer.

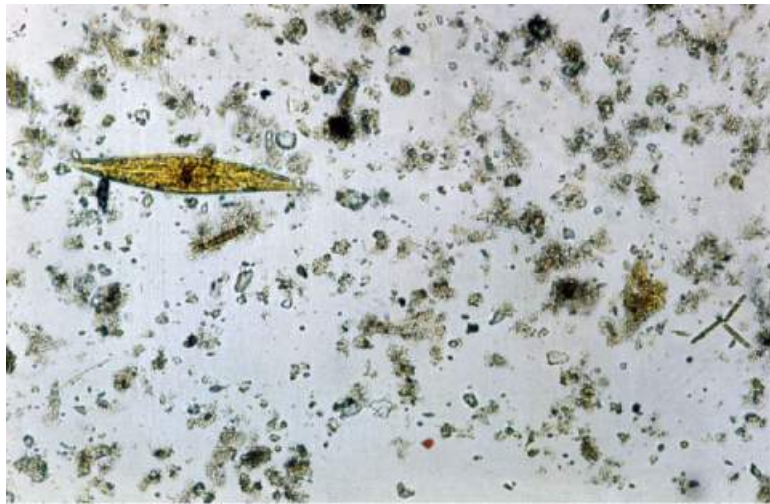


Figure 8.41: Particles in a drop of shelf sea water photographed through a microscope. The largest particle on view (a phytoplankton cell) is about $100\ \mu\text{m}$ long. Individual pieces of mineral material can be seen as well as flocs of mineral material joined together with organic “glue”. Photo: P. S. D. Smith.

Flocs are considered to be fairly delicate objects, easily broken up in the turbulent shear involved in collecting a water sample for instance. For this reason, the most reliable measurements of floc properties are made in situ using instruments which do not disturb the flocs. Because flocs cannot withstand turbulence it has been suggested that the largest flocs will have the same size as the smallest turbulent eddies, a size known as the *Kolmogorov microscale*. It is implied that other properties of the particle size distribution, such as the mean particle size, will change as the turbulent microscale changes and the largest flocs

adjust to follow this. The Kolmogorov microscale depends on the rate of dissipation of turbulent kinetic energy. In a tidal sea this will fluctuate with the tide, being largest at times of slack water. There is growing evidence that the particle size in shelf seas responds to these changes and that the mean particle size at slack water can be several times greater than that at maximum tidal flow. This will have interesting implications for the optical properties of shelf seas and estuaries.

How exactly is it best to imagine the passage of light through a flocculation? Is it best to think of the flocculation as a single particle with uniform properties (an average of the particles composing the flocculation), or is the flocculation better described as a number of separate mineral pieces (with relatively high refractive index) embedded in a matrix of low refractive index organic material? [Boss et al. \(2009a\)](#) have described experimental and theoretical results that suggest that an optical model allowing for the complexity of the flocculation performs better than one that assumes that flocculations are solid homogenous spheres, but the exact nature of the best model to use remains an open question.

8.11.6 Measuring the Mass Concentration of Mineral Particles

The concentration of mineral particles is measured by filtering a known volume of sea water through pre-weighed and pre-combusted GF/F filters (nominal pore size $0.7 \mu\text{m}$). The filters are then combusted in an oven at 500 deg C for 3 hours to remove organic material, cooled and weighed again.



Figure 8.42: GF/F filters after filtration of water samples on Sofala bank, Mozambique. The filters have been laid on a chart to show their sampling location. Photo: D.G. Bowers.

8.11.7 Light Absorption by Mineral Particles

The absorption coefficient of mineral particles suspended in seawater can be measured directly using filters such as those shown in Fig. 8.42. The filters are placed, while wet, on

a microscope slide and the slide is placed over the exit port of a spectrophotometer (the *filter pad* method). The optical density (OD) of the filter plus particles is recorded as a function of wavelength and the OD of a blank wet filter is subtracted. In calculating the absorption coefficient of the particles, allowance is made for the increase in the pathlength of the light through the particles and the filter produced by deviation of the light as it passes through the particles. The absorption coefficient of particles can also be measured in situ by measuring total absorption and subtracting the component due to water, organic material and CDOM.

The general shape of the absorption spectrum of mineral particles is one of exponential decay with increasing wavelength, often with a small “bump” near 500 nm. It appears that the absorption coefficient may not necessarily tend to zero in the near infra-red. The specific absorption spectrum of mineral flocs can therefore be represented by the following equation which includes the possibility of a non-zero absorption c_1 at long wavelength:

$$a_{\min}^* = c_1 + c_2 \exp(-S_{\min}(\lambda - \lambda_o)) \quad (8.28)$$

Here, a_{\min}^* is the concentration-specific absorption coefficient of mineral particles (i.e. absorption per unit mass concentration, units $\text{m}^2 \text{g}^{-1}$), λ is wavelength and λ_o is a reference wavelength. Bowers and Binding (2006) suggested mean values of $c_1 = 0.020 \text{ m}^{-1}$, $c_2 = 0.042 \text{ m}^{-1}$ (for a reference wavelength of 443 nm) and $S_{\min} = 0.012 \text{ nm}^{-1}$ based on their own measurements and those in the literature. There is no reason to suppose that the absorption spectrum of mineral particles should be the same from place to place and indeed variations in these coefficients are observed, but the slope parameter S_{\min} , in particular, appears to be fairly tightly constrained.

8.11.8 Light Scattering by Mineral Particles

There is no equivalent to the filter pad method for measuring light scattering by mineral particles. Scattering coefficients are instead measured at sea using instruments specifically designed for this purpose. Scattering coefficients measured in this way generally increase with the concentration of mineral particles, the rate of increase is the concentration-specific scattering coefficient for mineral particles b_{\min}^* . Specific scattering coefficients (units $\text{m}^2 \text{g}^{-1}$) are observed to be rather flat spectrally and to vary over at least an order of magnitude in the range: $0.1 < b_{\min}^* < 1$.

There is some evidence that b_{\min}^* increases as the water becomes more oceanic and less coastal. Babin et al. (2003a) have suggested this is because of changes in particle density. Inshore particles tend to be more mineral in content and of higher density. They therefore can be expected to have a smaller cross sectional area per unit mass. A full appreciation of the variation of the specific scattering coefficient includes consideration of particle size as well as density and we turn to this in the next section. Even at the lower end of the range of b_{\min}^* , mineral particles are more likely to scatter light than absorb it. A typical value of a_{\min}^* at 555 nm is $0.03 \text{ m}^2 \text{g}^{-1}$, therefore light of this wavelength will be scattered at least 3 (and up to 30) times by mineral particles before it is absorbed by a mineral particle.

8.11.9 Theoretical Ideas About the Optical Properties of Mineral Particles

The probability that light will interact with a particle in suspension is proportional to the cross sectional area of the particle. For a suspension of identical particles, the absorption (or scattering) *per unit* cross sectional area is known as the absorption (or scattering) efficiency. For a suspension of spherical particles with a range of diameters (from D_1 to D_2) we can calculate the scattering (or absorption) coefficient by integrating the product of scattering (or absorption) efficiency and particle area over the range of sizes. For example, in the case of scattering by spherical particles with a range of sizes, the scattering coefficient can be written:

$$b = \frac{\pi}{4} \int_{D_1}^{D_2} Q_b(D)n(D)D^2dD \quad (8.29)$$

and there is an equivalent expression for the absorption coefficient. Here $n(D)dD$ is the number of particles per unit volume in the size range D to $D + dD$ and $Q_b(D)$ is the scattering efficiency. Values of the scattering efficiency for spherical particles of given size and refractive index (both real and imaginary) are available from Mie theory or the simpler anomalous diffraction theory of [van de Hulst \(1957\)](#). It is often assumed that the size distribution of particles obeys a power law (or Junge) distribution, of the form

$$n(D) = KD^{-J} \quad (8.30)$$

where K and J are constants for a particular size distribution. A feature of this equation is that the number of particles rises rapidly as the size decreases. As a result, calculations based on Eqs. (8.29) and (8.30) lead to the conclusion that optical properties are dominated by the large number of very small particles. For example, in the case of the scattering coefficient, it has been estimated that 50% of light scattering by minerals is produced by particles smaller than $1 \mu\text{m}$ ([Babin et al., 2003a](#)). If this is the case, it would mean that much of the signal seen in visible band satellite images of shelf seas is produced by very small particles with low settling speeds. This would be an important conclusion for the interpretation of these images. However, there is no direct evidence that sub-micron sized particles exist in the numbers predicted by equation (8.30). Currently, measurements of particle numbers are limited to particles greater than a few microns in size. Furthermore, the flocculation process will tend to remove small particles and incorporate them in larger flocs ([Flory et al., 2004](#); [Boss et al., 2009a](#)).

8.11.10 Relationship Between Scattering and Mean Particle Size and Density

For *any* size distribution and shape of particle, Eq. (8.29) can be written as

$$b = Q'_b A \quad (8.31)$$

where A is the cross sectional area of particles in suspension in unit volume of water (units m^{-1}) and Q'_b is an effective scattering efficiency defined by

$$Q'_b \triangleq \frac{\int_{D_1}^{D_2} Q_b n(D) D^2 dD}{\int_{D_1}^{D_2} n(D) D^2 dD}. \quad (8.32)$$

Equation (8.31) can then be written in terms of the bulk, measurable, properties of the particles as

$$b = Q'_b \frac{C}{\rho D_A} \quad (8.33)$$

where C is the mass concentration of particles (dry mass per unit volume of water), ρ is the particle density (dry mass of particle per unit volume of particle in situ) and D_A is the Sauter diameter (volume of particles in situ per unit cross sectional area of particle in situ). It is possible to measure the volume of particles of different sizes in situ using laser diffraction instruments and from these measurements, the Sauter diameter (Section 8.12.5) can be calculated. Combining these observations with the mass concentration of particles weighed on a filter gives the particle density and therefore the validity of Eq. (8.33) can be tested. Bowers et al. (2010) found that 86 % of the variance in the scattering coefficient at stations along the west coast of Britain was explained by Eq. (8.33) with a constant $Q'_b = 1.96$. Since the concentration specific scattering coefficient of mineral particles $b^*_{\min} = b/C$ this means that most of the variation in b^*_{\min} is explained by changes in both particle size and density. In fact, for the flocculated particles in the study by Bowers et al. (2010), the Sauter diameter changed by a factor of 3 and the density by a factor of 13, so it is mostly changes in particle density that are responsible for the variation in the specific scattering coefficient of mineral particles.

8.11.11 Relationship Between Absorption and Mean Particle Size and Density

We can write a similar equation to (8.33) for the absorption coefficient:

$$a = Q'_a \frac{C}{\rho D_A}, \quad (8.34)$$

where Q'_a is an effective absorption efficiency and the other terms are as before. Unlike the scattering efficiency, Q'_a for semi-transparent particles can be expected to vary with the particle size, since it is harder for light to pass through large particles than small ones. A simple analysis of this problem proceeds as follows. The proportion of light that pass through the center of a particle of diameter D composed of material with uniform absorption coefficient a_{\min} is $\exp(-a_{\min}D)$. The proportion of light that is absorbed within the particle is therefore $(1 - \exp(-a_{\min}D))$. For small D , $(1 - \exp(-a_{\min}D)) \approx a_{\min}D$ and the proportion of light absorbed within the particle increases linearly with the size of the particle.

For a suspension of particles of different sizes, Bowers et al. (2010) found that the effective absorption efficiency of particles in north-west European seas does increase with a measure of the mean size of the suspension—namely the Sauter diameter, D_A . That is, it is possible to write

$$Q'_a = k_1 + k_2 D_A, \quad (8.35)$$

where k_1 and k_2 are constants. At 555 nm a reasonable fit to the data at stations where $C > 5 \text{ mg l}^{-1}$ ($N = 28$, $R^2 = 0.75$) was given by the expression

$$Q'_a = 0.02 + 0.0039 D_A \quad (8.36)$$

in which D_A is expressed in micrometers. In the case of light absorption by mineral flocs, therefore, the absorption coefficient depends on the cross sectional area of particles in suspension (which depends on their size and density) and, in addition, the absorption efficiency depends on the size of the particles.

8.11.12 Remote Sensing of Mineral Particles in Shelf Seas

Mineral particles suspended near the sea surface color the water and increase the reflection coefficient. Figure 8.43 shows a false color SeaWiFS image of the north-west European shelf. The regions of bright, colored water in the Irish Sea, the English Channel and the southern North Sea are mostly caused by mineral suspended sediments near the sea surface.

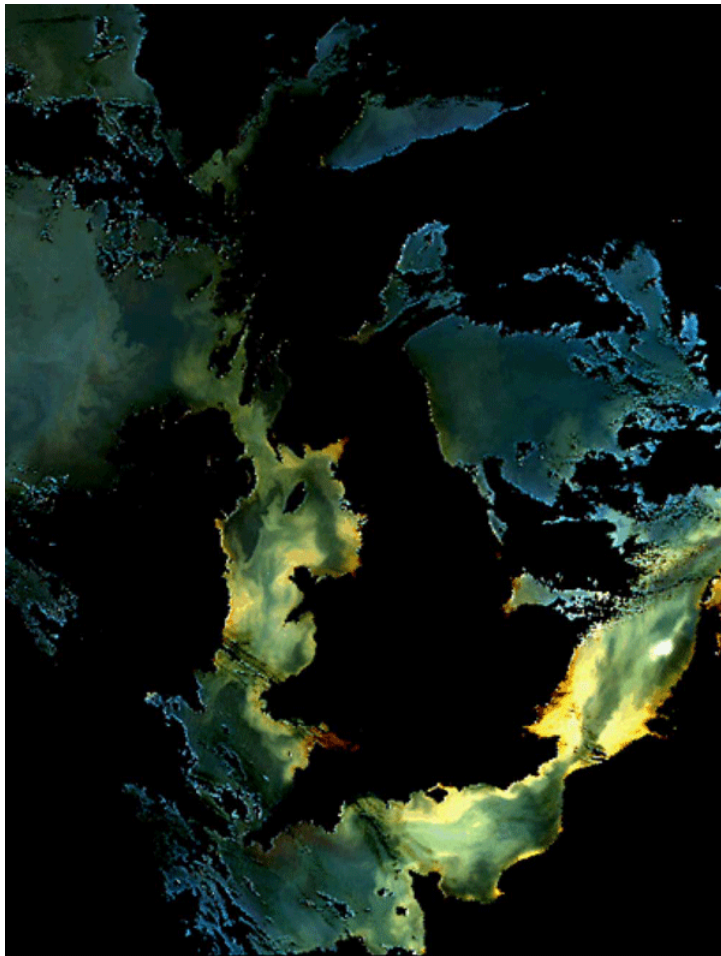


Figure 8.43: SeaWiFS false color image of the shelf seas of north west Europe. Image courtesy NERC NEODAAS.

Quantitative interpretation of images such as Fig. 8.43 is based on the scattering and absorption properties of the mineral particles and of water itself. The most successful algorithms for predicting the concentration of mineral suspended sediments use the reflection coefficients in the red (Binding et al., 2005) or, in the most turbid water, near infra-red parts of the spectrum (Doxaran et al., 2009). Analysis of the vertical flux of light in an ab-

sorbing ocean allows us to write the sub surface reflection coefficient as (Morel and Prieur, 1977)

$$R = f \frac{b_b}{a},$$

where f is a factor that depends on solar elevation and sky conditions, b_b is the backscattering coefficient and a is the absorption coefficient. In water where the reflectance is mostly produced by mineral particles, this equation can be expanded as

$$R = f \frac{\gamma b_{\min}^*}{a_w + a^* C_{\min} + a_X}, \quad (8.37)$$

where γ is the backscattering fraction (b_b/b), C_{\min} is the concentration of mineral particles, b^* and a^* are the concentration-specific scattering and absorption coefficients, respectively, a_w is absorption by water and a_X is absorption by materials other than water and mineral particles. In the red and infra-red parts of the spectrum, the spectrum, absorption by water tends to dominate the denominator of Eq. 8.37 and, for low to moderate concentrations of mineral particles, the reflectance increases in proportion to C_{\min} , with a slope of $f\gamma b^*/a_w$. At higher values of C_{\min} , absorption by mineral particles becomes more important in the denominator and the reflectance tends towards an asymptotic value of $f\gamma b^*/a^*$. Choosing a wavelength at the red end of the spectrum where absorption by water is high extends the region where reflectance is proportional to concentration. Equation (8.37) forms the basis of most general quantitative algorithms for suspended sediments in shelf seas (Stumpf and Pennock, 1989; Nechad et al., 2010).

As we have seen, a^* and b^* depend on the density and size of particles in suspension and changes in these properties, as well as of concentration, will produce changes in reflectance. What we are really seeing in images such as Fig. 8.43 is the cross sectional area of the particles in suspension. In order that remote sensing can be used to test and verify models of suspended particles (which predict mass concentration) we therefore need a better understanding of the relationship between particle mass and cross sectional area, and that is a challenge for the near future.

8.12 Particle Size Distributions

Figure 8.1 of this chapter shows the size *ranges* commonly associated with different components of natural waters. This section discusses the theory of how to quantify the number of particles as a function of their size. The next section, Creating Particle Size Distributions from Data, discusses the mechanics of creating particle size distributions from measured data.

8.12.1 Non-uniqueness of Particle Size

The first problem is to decide what is meant by particle “size,” especially when speaking of statistical measures like the “average size” of a collection of particles.

Suppose we have three spherical particles of radii $r = 1, 2,$ and 3 . What is the “average size” of these three spheres?

If the average size is based on the radius, then the average (radius) sphere has a radius of $(1 + 2 + 3)/3 = 2$.

If the average is based on the surface area of the spheres, $A = 4\pi r^2$, then the average surface area is $4\pi(1^2 + 2^2 + 3^2)/3$ which corresponds to a radius of $r = 2.16$. The same holds true for the cross-sectional area of the spheres.

If the average is based on the volume of the spheres, $V = \frac{4}{3}\pi r^3$, then the average volume is $\frac{4}{3}\pi(1^3 + 2^3 + 3^3)/3$ which corresponds to a radius of $r = 2.29$.

So there is ambiguity in how to define the “average particle size” even for a collection of spherical particles.

The situation is ever worse for non-spherical particles, which do not have a single parameter like radius that can be used to specify the size of the particle. The “size” of a non-spherical particle is sometimes taken to be

- the largest dimension of the particle,
- the arithmetic mean of the largest dimension of the particle, D_L , and the smallest dimension, D_S : $\frac{1}{2}(D_L + D_S)$,
- the geometric mean of the largest and the smallest dimensions: $\sqrt{D_L D_S}$,
- the diameter of a sphere with the same surface area as the particle (an “area-equivalent” sphere), or
- the diameter of a sphere with the same volume as the particle (a “volume-equivalent” sphere),

and there are also other measures of the “size” of a non-spherical particle. Note that the arithmetic and geometric means are equal for spherical particles, but are in general not equal.

All of these measures of particle size are valid, but one measure may be optimum for a particular application. For example, a Coulter Counter measures the change in electrical conductivity when a particle passes through the sensor. This change is proportional to the amount of material in the particle, i.e. to the particle volume. Thus a Coulter Counter measures the size of a volume-equivalent sphere. Laser diffraction instruments measure diffraction of light caused by a particle in the light beam. The angular shape of the diffraction pattern is determined by the cross-sectional (projected) area of the particle as seen by the beam. Thus laser diffraction measures the size of an area-equivalent sphere.

The diameters of area-equivalent and volume-equivalent spheres can be considerably different. Consider a cubical particle of length ℓ on a side. The volume is $V_c = \ell^3$ and the equivalent-volume sphere ($V_s = \pi D^3/6$) has a diameter

$$D_{\text{vol}} = \left(\frac{6}{\pi}\right)^{1/3} \ell = 1.24\ell.$$

The surface area of the cube is $A_c = 6\ell^2$ and of a sphere is $A_s = \pi D^2$, which leads to

$$D_{\text{area}} = \left(\frac{6}{\pi}\right)^{1/2} \ell = 1.38\ell.$$

If cubes of this size are seen randomly oriented in a laser beam, the average projected area is one-fourth of the surface area². Thus the laser sees, on average, particles of area

²Cauchy’s Average Projected Area Theorem shows that for a convex polyhedron, the average projected area over all orientations, i.e. the average cross section, is one-fourth the surface area of the polyhedron.

$6\ell^2/4$, and the equivalent-projected-area sphere has diameter $\pi D^2/4$. This again leads to $D_{\text{area}} = \sqrt{6/\pi}\ell$.

Thus for a cube the diameter of the equivalent-area sphere is 1.11 times the diameter of the equivalent-volume sphere. Coulter-principle and laser-diffraction instruments will therefore report somewhat different equivalent-sphere sizes for the same particle. Clearly, it is important to understand exactly what an instrument measures and what instrument was used to measure particle sizes. Unfortunately, publications sometimes fail to state exactly what measure they are using for the size of non-spherical particles.

8.12.2 Cumulative and Particle Number Size Distributions

First decide on some measure of particle size; call it the diameter D . If the particles are known to be spherical (e.g., fog droplets or small bubbles in water), then D is the diameter of the sphere. For non-spherical particles, D is most often taken to be the diameter of the volume-equivalent sphere. The next step is to quantify how many particles there are of each size D . There are two equivalent ways to do this.

The *cumulative number size distribution* (CSD_n) $N(D)$ is usually defined (in the particle-sizing literature) as the number of particles per unit volume *larger* than size D . $N(D)$ is usually expressed in units of m^{-3} . In principle, this function can be measured simply by counting the numbers of particles larger than a given size D , as D in principle ranges from 0 to ∞ , but in practice D ranges from some minimum D_{min} to some maximum D_{max} . These minimum and maximum sizes are usually determined by the instrument used to do the counting.

The *particle number size distribution* (PSD_n) $n(D)$ is a function defined so that $n(D)dD$ is the number of particles per unit volume between size D and $D+dD$ (or in the size range $D \pm D/2$). The units of $n(D)$ are usually number of particles per cubic meter per micrometer of size range, i.e. $\text{m}^{-3} \mu\text{m}^{-1}$.

The CSD_n , $N(D)$, as defined above decreases as D increases, so $dN(D)/dD$ is negative. The corresponding PSD is then the negative of the derivative of the CSD, and is usually written as

$$n(D) = \left| \frac{dN(D)}{dD} \right| \quad (\text{m}^{-3} \mu\text{m}^{-1}). \quad (8.38)$$

Because of the large range of values for both $N(D)$ and D , it is customary to plot these functions on log-log scales and to work with logarithmic size intervals. You then see (e.g., [Junge, 1953](#); [Liley, 1992](#)) $n(D)$ defined in terms of a log derivative, which is related to $dN(D)/dD$ by

$$\frac{dN(D)}{d \log D} = \frac{D}{\log e} \frac{dN(D)}{dD}. \quad (8.39)$$

The $\log e$ comes from a change of base from base 10 logarithms, which are convenient for working with data, to base e natural logarithms, which are convenient for mathematics. Thus PSDs defined by logarithmic and linear derivatives will differ by a factor of $D/\log e = 2.30D$:

$$n'(D) = \left| \frac{dN(D)}{d \log D} \right| = \frac{D}{\log e} \left| \frac{dN(D)}{dD} \right| = \frac{D}{\log e} n(D).$$

The converse of Eq. (8.38) is

$$N(D) = \int_D^\infty n(D)dD \quad (\text{m}^{-3}).$$

The total number of particles per unit volume is given by

$$N_t = N(0) = \int_0^\infty n(D)dD \quad (\text{m}^{-3}).$$

In the particle sizing literature, it is customary to use a diameter D as the measure of particle size. However, Mie Theory (Chapter 12) usually expresses its size parameter in terms of the particle radius $r = D/2$. If using a PSD(D) in Mie calculations as described in Section 12.3, it is necessary to use $PSD(r) = 2PSD(D)$.

8.12.3 Area and Volume Size Distributions

The previous definition (8.38) of the PSD was for the *number* of particles in a unit volume, hence the notation PSD_n or $n(D)$, with the subscript n indicating number of particles, and the name number size distribution. We can also define size distributions for the area and volume of the same particles. The surface area of a sphere is πD^2 , so the PSD for area (subscript a) is

$$PSD_a(D) = \pi D^2 PSD_n(D) = \pi D^2 n(D) \quad (\mu\text{m}^2 \text{m}^{-3} \mu\text{m}^{-1}). \quad (8.40)$$

Note that the units of the area size distribution $PSD_a(D)$ are particle surface area, usually in μm^2 , per unit volume per unit size interval. The size distribution for the cross-sectional area of the particles is $\frac{\pi}{4}D^2 PSD_n(D)$, with the same units. The PSD for particle volume, $\frac{\pi}{6}D^3$, is then (subscript v)

$$PSD_v(D) = \frac{\pi}{6}D^3 PSD_n(D) = \frac{\pi}{6}D^3 n(D) \quad (\mu\text{m}^3 \text{m}^{-3} \mu\text{m}^{-1}). \quad (8.41)$$

Now the units of the volume size distribution are particle volume, usually in units of μm^3 , per unit volume per unit size interval. For particle volume measured in cubic micrometers, $PSD_v(D)$ thus gives a size distribution for particle volume in parts per million, per unit size interval. Of course, these formulas are based on the assumed validity of area- and volume-equivalent spheres.

The total surface area of particles per unit volume is

$$A_t = N_a(0) = \int_0^\infty PSD_a(D)dD = \pi \int_0^\infty D^2 n(D)dD \quad (\mu\text{m}^2 \text{m}^{-3}).$$

The total volume of particles per unit volume is

$$V_t = N_v(0) = \int_0^\infty PSD_v(D)dD = \frac{\pi}{6} \int_0^\infty D^3 n(D)dD \quad (\mu\text{m}^3 \text{m}^{-3}).$$

If the quantity of interest is the number of particles (e.g., the number of phytoplankton of various sizes), then the number size distribution $n(D)$ is the relevant distribution. However, some processes depend more on the surface area of the particles than on their number. For example, chemical reactions such as combustion (e.g., burning of coal dust in a power plant) depend strongly on the surface area, and laser diffraction is governed primarily by particle cross-sectional area. For such problems, PSD_a is the distribution of interest. For processes that depend on particle volume or mass (e.g., the mass of sediment

material resuspended in a bottom boundary layer or particle buoyancy in a fluid), the volume distribution PSD_v is the distribution of interest.

Many commercial particle sizing instruments allow the user to select the output as a number, area, or volume PSD. Keep in mind that the conversions between one form of PSD and another are being made on the basis of an assumed equivalent spherical particle, even though the underlying measurement may be based on particle area (laser diffraction) or volume (Coulter counter). Thus a laser diffraction instrument may give a good measurement for the area PSD_a even for nonspherical particles, but errors can be introduced when converting the area PSD to number or volume PSDs. A Coulter counter may give a good measurement of a volume distribution PSD_v even for nonspherical particles, but errors can be introduced with converting the volume distribution to area or number distributions.

8.12.4 Models for Particle Size Distributions

Many models have been proposed for particle size distributions. These differ according to the environment (atmosphere, ocean, industrial process) and particular data set being analyzed. The classic studies by Junge (1953) and Junge (1955) were concerned with aerosol size distributions. He found that a number size distribution of the form

$$\frac{dN(D)}{d \log D} \propto D^{-3} \quad (8.42)$$

gave a good fit to aerosols down to about $1 \mu\text{m}$ in diameter. Note that according to Eq. (8.39), the D^{-3} dependence of Eq. (8.42) corresponds to a D^{-4} dependence when written as

$$n(D) = \left| \frac{dN(D)}{dD} \right| \propto D^{-4} \quad (8.43)$$

Thus you usually see the statement that a Junge distribution has a slope of -4 when $n(D)$ and D are plotted on logarithmic axes.

Figure 8.2 showed one instance of the number PSD for various size classes of marine biological particles; D is the diameter of a volume-equivalent sphere. The Junge distribution is shown as a blue line in that figure, with the magnitude fixed by the value for picoplankton.

A more general number size distribution model is the *power law distribution*, which is usually written as

$$n(D) = n(D_o) \left(\frac{D}{D_o} \right)^{-S}. \quad (8.44)$$

The exponent S is usually called the slope parameter. $n(D_o)$ is the value of the distribution in $\text{m}^{-3} \mu\text{m}^{-1}$ at a reference diameter D_o . The Junge distribution (8.43) is a special case of a power law distribution for $S = 4$. (Some authors call the power law distribution a Junge distribution, but others are careful to use the term Junge distribution only for $S = 4$ because that is the only slope Junge considered in his papers.)

Figure 8.44 shows 168 number size distributions $n(D)$ measured over the equivalent-sphere size range $0.8 \leq D \leq 120 \mu\text{m}$ in Arctic waters. When fit to the power law model of Eq. (8.44), the best-fit values of $-S$ have the distribution of values shown in Fig. 8.45. The average of the 168 slope parameters is $S = 3.6 \pm 0.37$.

Figure 8.46 shows two of the PSDs of Fig. 8.44 along with the corresponding area and volume distributions, and the cumulative distribution functions. Although the fit of the

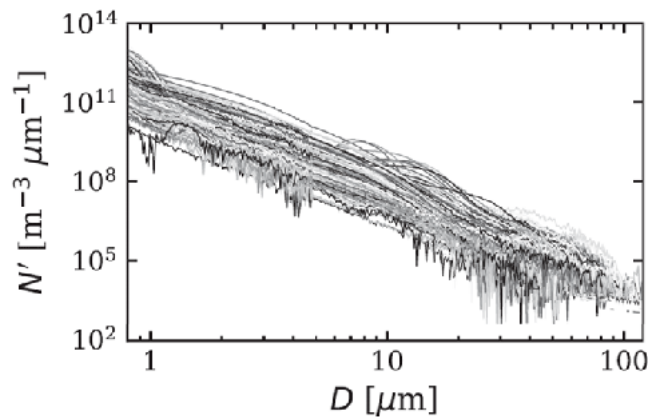


Figure 8.44: 168 particle number size distributions $N' = n(D)$ measured in Arctic waters. From Runyan et al. (2020, Fig. 2), reproduced by permission under a Creative Commons license.

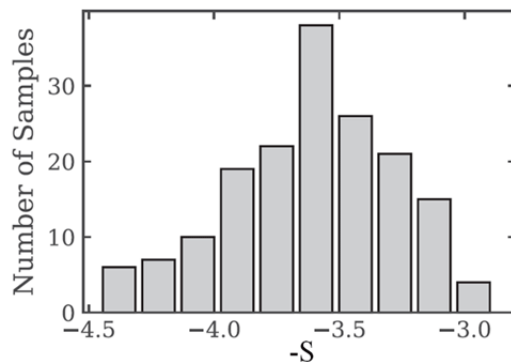


Figure 8.45: Distribution of the best-fit exponents of the power-law model of Eq. (8.44) when fit to the 168 PSDs of Fig. 8.44. From Runyan et al. (2020, Fig. 3). Reproduced by permission under a Creative Commons license.

power-law model (8.44) to these number PSDs appears quite good on a log-log plot, it must be noted that there are order-of-magnitude deviations from the best-fit curves, namely near $40\text{--}50\ \mu\text{m}$ in the upper left figure, and near $1.5\ \mu\text{m}$ in the lower-left figure. Such deviations are common in fits such as these and can result, for example, from a bloom of a particular species of phytoplankton. It should be noted from the CDFs in the right panels that the total number of particles comes mostly from particles less than about $3\ \mu\text{m}$, but the total volume of particles comes mostly from the larger particles greater than about $30\ \mu\text{m}$. This is simply because one particle with $D = 10\ \mu\text{m}$ has as much volume as 1000 particles of size $D = 1\ \mu\text{m}$, for example.

It sometimes happens that the small size and large size ends of the distribution have different slopes. A better fit is then obtained by separate fits of a power law for the two size regions, but other distributions have also been used.

It is important to recognize that a power-law distribution implies that the size range extends from $D = 0$ to ∞ . This is never the case in nature. The smallest phytoplankton

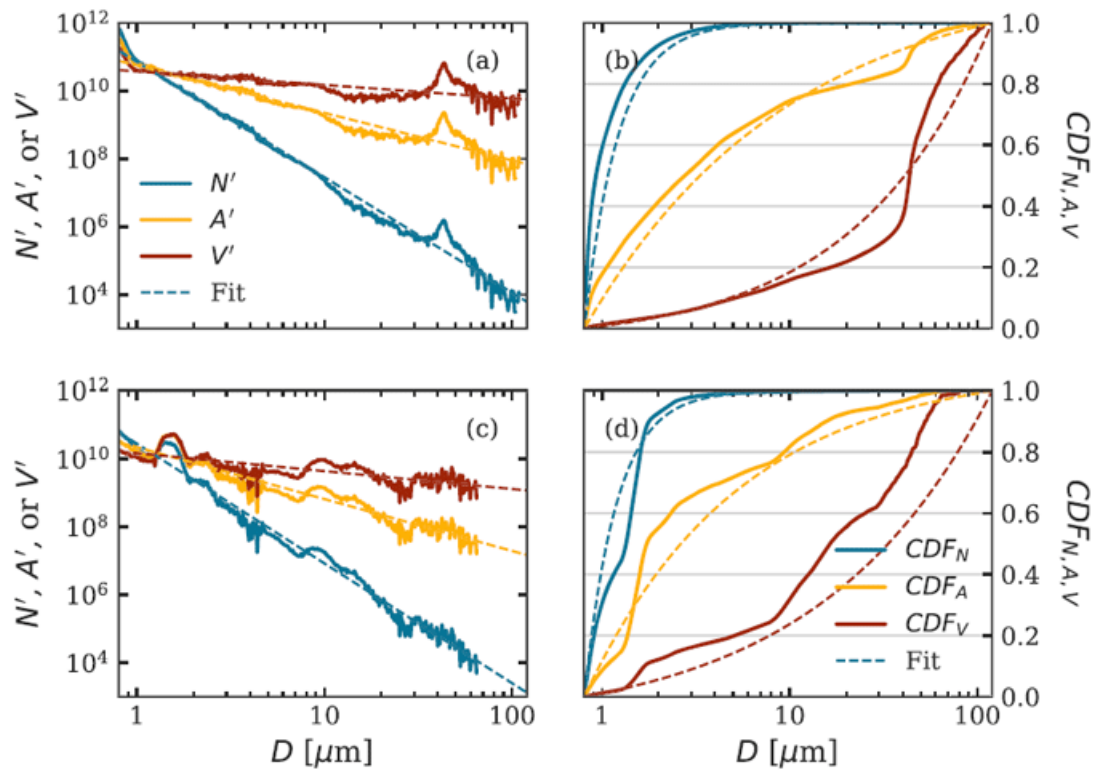


Figure 8.46: Two sets of PSDs from the deep chlorophyll maximum at two locations (top and bottom rows) in the Arctic. The left panels show the number ($N' = n(D)$), area ($A' = PSD_a(D)$), and volume ($V' = PSD_v(D)$) distributions. The right panels show the corresponding cumulative distribution functions for number, area, and volume. The dashed lines show the power law distributions, based on the best-fit to the number distributions. From Runyan et al. (2020, Fig. 5), reproduced by permission under a Creative Commons license.

have a diameter of about $0.2 \mu\text{m}$ and the largest are about $200 \mu\text{m}$. Thus a PSD for phytoplankton, whatever its mathematical form, should be applied only for the $D = 0.2$ to $200 \mu\text{m}$ size range, and perhaps for an even smaller range. Particular species or size classes of phytoplankton have much smaller size ranges. The upper panel of Fig. 8.47 shows measured number PSDs for 18 classes and species of phytoplankton (VIRU is viruses, HBAC is heterotrophic bacteria, PROC is *prochlorococcus*, ..., MICA is *Prorocentrum micans*; see Stramski et al. (2001) for the full listing and description of these microbes). The bottom panel of the figure shows the sum of the individual PSDs, with the concentrations of each class chosen within the ranges of typical values so that the total (except for viruses) obeys a Junge distribution ($S = 4$ in Eq. (8.44)). Modeled distributions for detritus, minerals, and bubbles are also shown. These PSDs were used to model the inherent optical properties of oligotrophic waters; the total chlorophyll of the sum is $0.18 \text{ mg Chl m}^{-3}$.

The PSDs of individual species or classes of particles as seen in the upper panel of Fig. 8.47 are often modeled by a log-normal distribution. In this distribution, the numbers are normally distributed when D is plotted on a logarithmic scale as in the preceding figures,

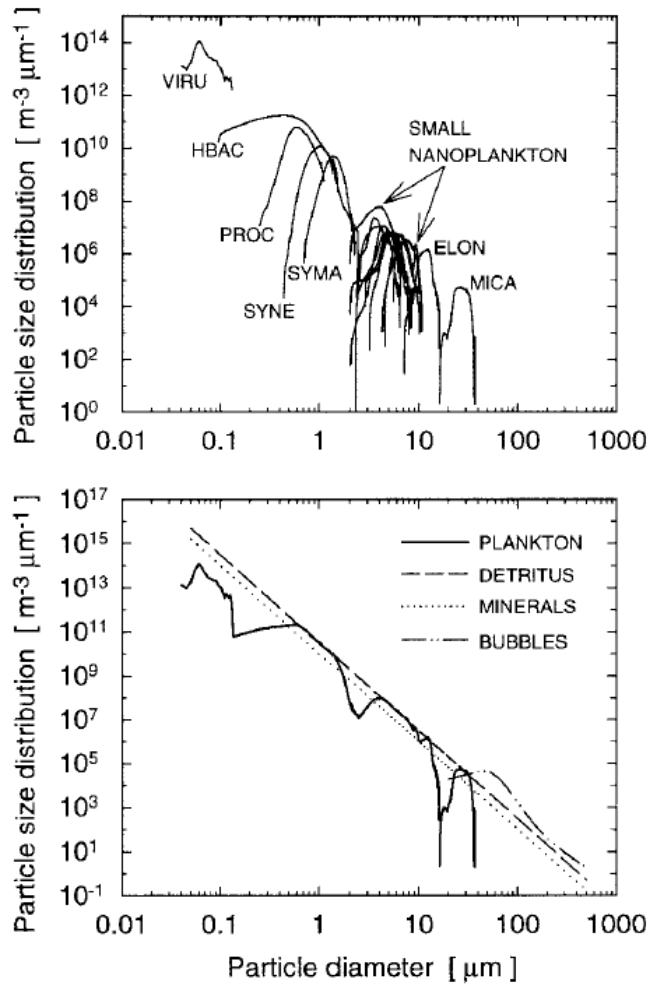


Figure 8.47: Top panel: measured number size distributions for 18 classes or species of microbes. Bottom panel: the sums of the microbial PSDs (solid line), and modeled PSDs for detritus, minerals, and bubbles. Reproduced from [Stramski et al. \(2001, Fig. 9\)](#) under Optica's Fair Use policy.

or equivalently, using a normal distribution in $\ln D$ rather than in D . That is, $\ln D$ is distributed as a Gaussian probability distribution function:

$$pdf(\ln D) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln D - \mu)^2}{2\sigma^2}\right], \quad (8.45)$$

where μ and σ are the mean and standard deviation of $\ln D$, respectively. Note that this $pdf(\ln D)$ satisfies

$$\int_{-\infty}^{\infty} pdf(\ln D) d \ln D = 1,$$

as is required of any probability distribution function. Noting that $pdf(D)dD = pdf(\ln D)d \ln D$

represents the same probabilities gives

$$pdf(D) = pdf(\ln D) \frac{d \ln D}{dD} = \frac{1}{D} pdf(\ln D),$$

which is the form seen in [Campbell \(1995, Eq. A1\)](#), and is equivalent to [Ahmad et al. \(2010a, Eq. 2\)](#). To express the log-normal distribution in terms of base 10 logarithms, the same procedure plus the observation that $d \log D / dD = \log e / D$ gives

$$pdf(D) = \frac{\log e}{D} pdf(\log D) = \frac{1}{(\ln 10)D} pdf(\log D),$$

which is the form seen in [Shettle and Fenn \(1979, Eq. 1\)](#). Now μ and σ are the mean and variance of $\log D$. Note that $pdf(D)$ is defined for $0 < D < \infty$, which corresponds to $-\infty < \ln D < \infty$ in Eq. (8.45). Multiplying these probability distribution functions by magnitudes (e.g., the total number of particles per cubic meter) scales them for use as PSDs.

Log-normal distributions are commonly used to describe atmospheric aerosol size distributions (e.g., [Shettle and Fenn, 1979](#)). The current NASA atmospheric correction algorithm models aerosols as a sum of two log-normal distributions: one for small, continental (dust) aerosols and one for large, marine (sea salt) aerosols ([Ahmad et al., 2010a](#)):

$$n(r) = \frac{1}{r} \frac{dN(r)}{d \ln r} = \sum_{i=1}^2 \frac{N_{oi}}{r \sigma_i \sqrt{2\pi}} \exp \left[-\frac{(\ln r - \mu_i)^2}{2\sigma_i^2} \right]. \quad (8.46)$$

The parameters N_{oi} , μ_i , and σ_i , $i = 1, 2$ are adjusted to give the relative amounts of continental and marine aerosols, and the parameters also depend on the relative humidity. Figure 8.48 illustrates these aerosol distributions for a typical open-ocean mix of aerosol types and for a relative humidity of 50%. The left panel plots the number PSD $n(r) = dN(r)/dr$ and the right panel plots the same information as $dN(r)/d \log r$. In the left panel, the dashed line shows the -4 slope of a Junge distribution (Eq. 8.43). In the right panel, the Junge distribution is represented by a slope of -3 on the $\log r$ abscissa scale (Eq. 8.42). The aerosol model of Eq. (8.46) uses particle radius r rather than diameter D because these particle size distributions are used as input to Mie calculations of aerosol optical properties.

The log-normal distribution finds many applications beyond particle size distributions. [Campbell \(1995\)](#) shows examples of log-normal distributions of chlorophyll concentrations, normalized water-leaving radiances, the diffuse attenuation coefficient $K_d(490)$, normalized photosynthetic yields, and several other biological variables. In fields other than oceanography, the log-normal distribution has been found to describe phenomena as diverse as the distribution of income (excluding billionaires), daily rainfall amounts, the populations of cities, and much more. Qualitative understanding of these observations can be obtained as follows. If a random variable is the *sum* of other random variables, then the distribution of the sum will be normal, regardless of the distribution of the individual variables. This is known as the Central Limit Theorem. (See Section E.5 on error estimation in Monte Carlo calculations for numerical examples.) If the total results from a *product* of random processes, then the product will obey a log-normal distribution. This is because the logarithm of a product is the sum of the logarithms, and the sum of the individual logarithms

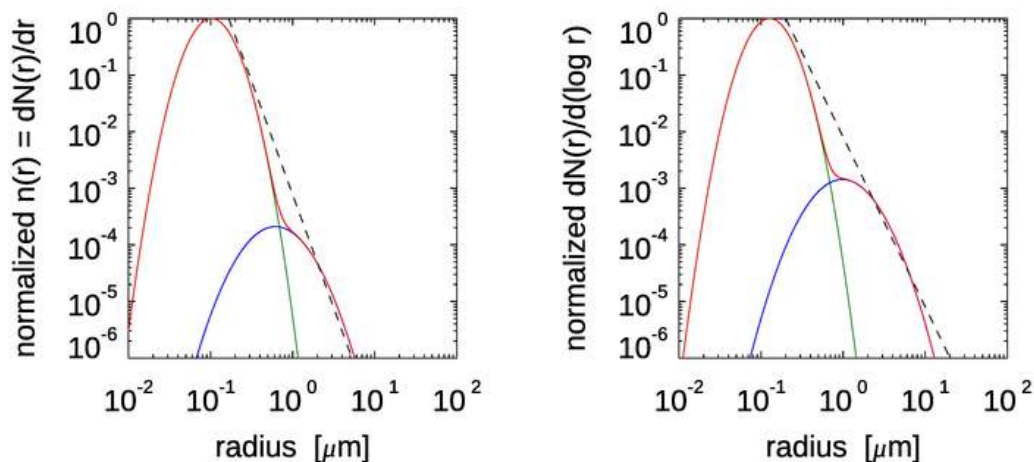


Figure 8.48: Typical open-ocean aerosol distributions plotted as $dN(r)/dr$ in the left panel and as $dN(r)/d \log r$ in the right panel. The green curve is the contribution by continental aerosols and the blue curve is the marine aerosols; the red curve is the total. The black dashed lines show the slope of a Junge distribution.

is then normally distributed. Thus, for example, if the total chlorophyll concentration is the result of a repeated daily fractional increase (daily growth rate) (e.g., the total Chl is the value at day 1 times 1.1 to get the value at day 2, time 1.1 again to get the value at day 3, ...), the total chlorophyll concentration will be log-normally distributed.

8.12.5 Single-parameter Measures of Particle Size

A PSD contains the full information about the sizes of particles in a sample. However, a PSD is a function of size, and there is a human tendency to want to have a single number to describe the particle sizes. Using a single number to describe a distribution may sometimes be useful, but it often can be misleading or downright dangerous. As shown in the introduction, it is not even possible to define a unique “size” for more than one sphere, let along a distribution of non-spherical particles. A power-law distribution diverges as the size goes to 0, so such a distribution can be integrated only over a finite size range D_1 to D_2 . Numbers like the mean or median particle size can be computed for a distribution, but the “mean” size will be different when based on a number distribution versus a volume distribution, for example.

Perhaps the most commonly used single-number size parameter is the Sauter mean diameter (SMD), which is defined by

$$SMD = \frac{\int_{D_1}^{D_2} D^3 n(D) dD}{\int_{D_1}^{D_2} D^2 n(D) dD}.$$

Recalling Eqs. 8.40 and 8.41, the SMD can be written as $SMD = 6V_t/A_t$, where V_t and A_t are the total volume and area of the particles in the distribution. For a single spherical particle of volume $\pi D^3/6$ and area πD^2 , the SMD reduces to the diameter of the sphere.

Thus in words, the SMD is the diameter of a sphere that has the same total volume to total area ratio as the particles in the distribution.

Figure 8.49 shows the differences in the mode, median, mean, and Sauter mean diameter for a log-normal probability distribution function,

$$pdf(x) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right],$$

with parameters $\mu = 1.0$ and $\sigma = 0.5$. This illustrates that these measures of particle “size” are in general all different. The Sauter mean diameter is larger than the others and is determined most strongly by the largest particles of a distribution. If you are interested in total volume or mass, then the SMD is a useful quantity. This is why the SMD is used in sediment transport studies. On the other hand, if you are interested in the smallest particles, then the mode or median is probably a better statistic to use. Other single-number measures of a distribution have been developed for specific problems. Clearly, which single measure of a size distribution might be most useful depends on the problem at hand.

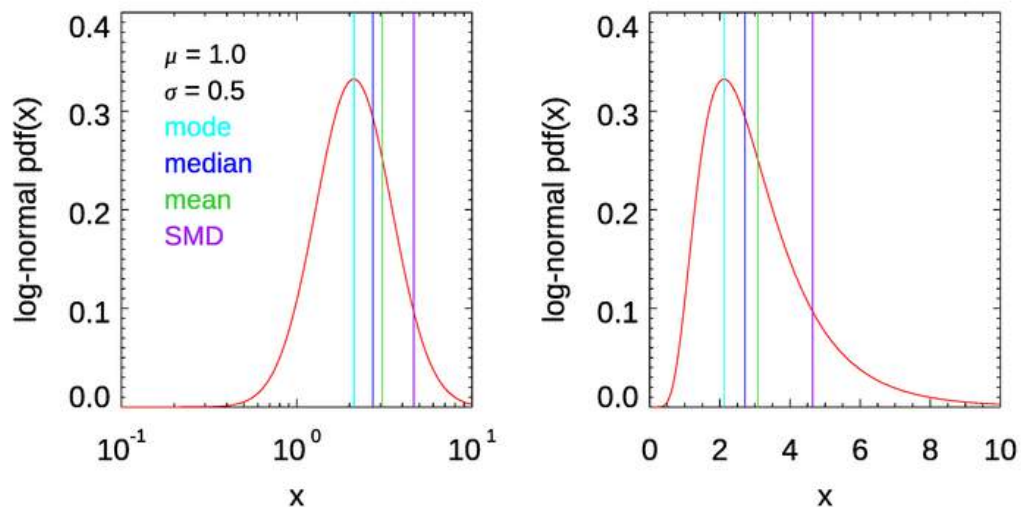


Figure 8.49: Differences in the mode, median, mean, and Sauter mean diameter for a log-normal distribution with parameters $\mu = 1.0$ and $\sigma = 0.5$.

8.12.6 Comments on Terminology

As is so often the case, different authors sometimes use different terminology. For example, [Bader \(1970\)](#) calls $N(D)$ the cumulative number distribution function and $n(D)$ the number distribution function. [Jonasz \(1983\)](#) calls $N(D)$ the cumulative size distribution and $n(D)$ the particle size distribution. I use the term cumulative number size distribution (CSD_n) for $N(D)$ and particle number size distribution (PSD_n) for $n(D)$ because that seems to be the most common in the literature (subscript n for number).

However, in probability theory, a cumulative distribution function $CDF(D)$ refers to the how much of a total is *less* than D , rather than greater than D as used for $N(D)$. Runyan et al. (2020) compute a true $CDF(D)$ by dividing the $PSD(D)$ by the total number of particles over all sizes, N_t , to get a true $PDF(D)$, and then computing a (nondimensional) CDF as $CDF(D) = \int_0^D [n(D')/N_t] dD'$. They also employ a carefully designed nomenclature with N' , A' , and V' denoting the number, area, and volume PDFs, which they refer to as density functions. This is somewhat in analogy to a spectral density function, which shows how much of something there is per unit interval of distance or frequency. Their true CDFs are denoted by $CDF_n(D)$, $CDF_a(D)$, and $CDF_v(D)$ for the number, area, and volume CDFs computed from the respective density functions.

Any of the functions $CSD_n(D) = N(D)$, $PSD_n(D) = n(D)$, $PSD_a(D)$, $PSD_v(D)$, $CDF_n(D)$, $CDF_a(D)$, $CDF_v(D)$ etc. is properly termed a particle size distribution. They are all different ways of describing how some feature of particles (number, area, volume) change with size, and one can be computed from and of the others (under certain assumptions like spherical particle shapes).

I suspect that the “backwards” definition of the cumulative size distribution as used in particle analysis traces back to measurement of size distributions of soil or other particles using a succession of wire mesh sieves. You first sieve out the large particles (gravel, for example) and count those. Then a smaller mesh collects the next smaller size (sand, for example), and so on down to the finest particles (e.g., silt). The fine meshes are very delicate, so you have to remove the large particles first. Thus as the sieving process works from large to small particles sizes, you are accumulating the number of particles larger than the current sieve size, and building up a $CSD(D)$ that shows how many particles are *larger* than D . Given a normalized probability distribution function $PDF(D)$ (e.g., $PDF_n(D) = n(D)/N_t$), it is convenient to integrate (or sum) the PDF from small to large values of D , and thus build up a $CDF(D)$ that shows the fraction of the total that is *less* than D .

8.12.7 Comments on Units

Suppose you have some data $Y(x)$ that you wish to model with a normal or Gaussian distribution. Y might be the number, area, or volume of particles, and x is some measure of their size. The mean and standard deviation of the data are m and s , respectively, and Y_t is the total “amount” of Y for all sizes (e.g., the total number of particles or their total volume). Then the Gaussian model of the data would be

$$Y(x) = \frac{Y_t}{\sqrt{2\pi}s} \exp \left[-\frac{(x-m)^2}{2s^2} \right]. \quad (8.47)$$

If x has units, say μm , then m and s have the same units. The argument of the exponential is nondimensional, and

$$\int_{-\infty}^{\infty} Y(x) dx = Y_t$$

since a Gaussian distribution integrates to 1.

Now suppose you wish to model the same data using a log-normal distribution. The idea is that

$$Y(\log x) = \frac{Y_t}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(\log x - \mu)^2}{2\sigma^2} \right], \quad (8.48)$$

where now $\mu = \log m$ and $\sigma = \log s$ are the parameters of the distribution. However, *this equation is not correct because you cannot take the logarithm of a dimensional quantity.* (Nor can you compute e^x or $\sin(x)$ or any other such function unless x is a nondimensional number.) However, you can compute $\log(ax)$ if a has units of 1 over the units of x , so that ax is a nondimensional number. Thus, before using a log-normal distribution, you must non-dimensionalize the size x . If x is in μm , then this can be done by dividing all values of x by $1 \mu\text{m}$, i.e. setting $a = 1/1 \mu\text{m}$. However, you could just as well normalize the x values by dividing by $17.3 \mu\text{m}$. Note that converting the x data values from, say, micrometers to millimeters, or even to inches, also amounts to multiplying all x values by a scale factor a . If $x \rightarrow ax$, then $m \rightarrow am$ and $s \rightarrow as$, and the exponential of the log-normal distribution does not change. The log-normal distribution is then properly written as

$$Y[\log(ax)] = \frac{Y_t}{\sqrt{2\pi}\sigma} \exp \left[-\frac{(\log(ax) - \mu)^2}{2\sigma^2} \right], \quad (8.49)$$

where now $\mu = \log(am)$ and $\sigma = \log(as)$. Of course, if x is measured in micrometers and $a = 1/1 \mu\text{m}$, then nothing changes numerically in Eq. (8.48) versus (8.49), but it should be kept in mind when using equations like (8.45) or (8.46) that the size measure D or r must be non-dimensional. This subtlety seems never to be mentioned in the particle sizing literature, although the mathematicians have commented on this; see [Matta et al. \(2011\)](#) and [Finney \(1977\)](#).

8.13 Creating Particle Size Distributions from Data

[Emmanuel Boss and Nils Haëntjens contributed to this section.]

Particle size distributions (PSDs), the descriptions of how particle number, area, or volume depend on particle size, are useful tools in oceanography used in applications such as characterizing the ecosystem ([Cermeno and Figueiras, 2008](#)) or computing the carbon flux to depth ([Guidi et al., 2009](#)). While PSDs have been extensively used in the literature, surprisingly little has been written about how they are actually derived from measured data.

8.13.1 Empirical PSDs

How one constructs a PSD depends on the tool used to size the assembly of particles. Single-particle analyzers, such as the Coulter Counter or cytometers, provide information on the size of each individual particle passing through the instrument (typically based on volume or cross-sectional area and assuming an equivalent sphere). Other methods, such as the Laser In Situ Scattering and Transmission meter (LISST, [Agrawal and Pottsmith \(2000\)](#)), provide a PSD of the bulk assembly of particles by inverting a bulk measurement (near-forward angular scattering in the case of the LISST) to obtain the most likely underlying PSD.

The process of building a PSD from the size information of individual particles is as follows. Choose a number of bins (M) and denote their boundaries b_1, b_2, \dots, b_{M+1} . Place a particle in the particular bin for which its diameter (D) obeys $b_j \leq D < b_{j+1}$ (by “placing it” is meant that the number of particles in that bin is incremented by one). Ideally, the size characterizing each bin is based on the mean size of the particles in that bin. Typically, however, that size is based on the boundaries of the bin (e.g. the arithmetic or

geometric mean of the bin boundaries). Thus the “discrete” PSD, denoted by $N(D_j)$, gives the number of particles with mean diameter D_j (units of number per volume of water). To obtain a continuous PSD, $n(D_j)$, (for example for the purpose of comparison between different instruments each having different bin sizes), one divides the discrete PSD by the bin width: $n(D_j) = N(D_j)/(b_{j+1} - b_j)$ (units of number per volume per size). To obtain a volume or area size distribution, the number of particles in each bin is multiplied by the average volume or area of a particle in that bin, respectively. When using imaging cytometers such as the Imaging Flow Cytobot (Olson and Sosik, 2007) the cross-sectional area is measured directly for each cell, from which the volume can be estimated (Moberg and Sosik, 2012). Hence another option to derive the PSD is to sum the volume or area of each individual cells instead of the average cross-sectional area or volume of the bin times the number of particles.

Continuous particle number size distributions in the surface ocean are often approximated by a power-law distribution (e.g. Jackson et al. (1997)):

$$n(D) = AD^{-\xi} \quad [\text{number cm}^{-3} \mu\text{m}^{-1}]. \quad (8.50)$$

Power-law differential distributions are observed to have an exponent (ξ above) varying between 3 and 4 in the surface ocean (Jackson et al., 1997; Sheldon et al., 1972). An exponent of 4 implies that volume is constant within bins increasing in size by a power-law rule (e.g. Sheldon et al. (1972)). A problem in the above equation is that, in principle, we should never exponentiate with a fraction a quantity that has physical units. Often, (8.50) will be written instead as function of a nondimensional ratio, e.g.

$$n(D) = A(D/D_0)^{-\xi},$$

with D_0 being a reference diameter. In what follows we assume, without loss of generality, that $D_0 = 1 \mu\text{m}$, and D is reported in μm and hence this normalization is implicitly assumed.

Before the PSD can be built certain decisions need to be made. The upper and lower bounds for particle size range, the number of bins, and the rule according to which bins are allocated. Traditionally, due to the rapid decrease in particle concentration with size, bin sizes have been chosen to follow a power-law scaling (Sheldon et al., 1972; Jackson et al., 1997; Agrawal and Pottsmith, 2000). That is to say, a subsequent bin is q times larger than the previous bin (Section 8.13.5 shows other possible choices for bin sizing). This choice has the advantage that bins are of equal size on a logarithmic D (abscissa) axis and that oceanic volume distributions are nearly flat (Sheldon et al., 1972), which provides a quick check on the data. The downside is that over a decade in size the number of particles per bin still decreases rapidly. For example, for a choice of ten bins over a decade, the number of particles between the first and last bin fall by a factor of ~ 1000 . This means that to reduce counting errors to 10% at the largest bin (counting errors scale like \sqrt{N}) for such a choice, more than 100,000 particles are required per sample, which is typically unrealistically large for cytometry.

8.13.2 Parametric Description of a PSD

Assume we want to produce a size distribution for oceanic plankton (e.g. Lombard et al. (2019, Fig. 2)). Denote the boundaries of the bins by b_1, b_2, \dots, b_{M+1} . Assuming that the

bins they bound grow following a power-law:

$$q = \frac{b_3 - b_2}{b_2 - b_1} = \frac{b_4 - b_3}{b_3 - b_2} = \dots = \frac{b_{M+1} - b_M}{b_M - b_{M-1}},$$

which is satisfied if for any j ,

$$b_j = b_1 q^{j-1} \implies b_{M+1} = b_1 q^M \implies q = \sqrt[M]{\frac{b_{M+1}}{b_1}}.$$

Thus, if we have the lowest and largest boundaries of the PSD (b_1 and b_{M+1}) and the number of bins in the PSD (M), we can compute the boundaries of all other bins. The volume of material associated with spherical particles distributed as a power-law with a differential power law exponent of 4 (the ‘‘canonical’’ value of [Sheldon et al. \(1972\)](#)) is

$$V(b_j < D < b_{j+1}) = \frac{\pi A}{6} \int_{b_j}^{b_{j+1}} D^{-1} dD = \frac{\pi A}{6} \ln q,$$

which, as discussed above, is the same for all bins.

The average size of a particle associated with each bin for such a PSD is

$$\begin{aligned} \bar{D}(b_j < D < b_{j+1}) &= \frac{\int_{b_j}^{b_{j+1}} n(D) D dD}{\int_{b_j}^{b_{j+1}} n(D) dD} = \frac{\int_{b_j}^{b_{j+1}} D^{-3} dD}{\int_{b_j}^{b_{j+1}} D^{-4} dD} \\ &= \frac{3(b_{j+1}^{-2} - b_j^{-2})}{2(b_{j+1}^{-3} - b_j^{-3})} = \frac{3b_{j+1}(1+q)}{2(1+q+q^2)} = \frac{3b_1 q^j (1+q)}{2(1+q+q^2)}. \end{aligned} \quad (8.51)$$

This, however, is different from the typical size chosen to represent bins in the literature. Typically the size associated with PSD bins is computed as the geometric mean of the bin boundaries:

$$\bar{D}(b_j < D < b_{j+1}) = \sqrt{b_{j+1} b_j} = b_j \sqrt{q} = b_1 q^{j-1} \sqrt{q},$$

which is larger than the size computed in (8.51). While the two may be close (their ratio is constant and depends on q), they are not identical. Choosing the arithmetic mean to represent the bin is even more biased.

This issue is of importance because the mean size of the bin is used to convert between number, area, and volume size distributions. For example the LISST particle size output is a volume distribution (in parts per million, ppm). Converting it to a size distribution requires a division of the volume in each bin by the volume of the average particle in that bin. The choice of the average diameter, because it is cubed, can result in a significant bias.

In coastal areas the differential PSD power-law slope of particles has an exponent closer to $\xi = 3$ (e.g., [Jackson et al., 1997](#)). In such cases, where we expect the power-law exponent not to be 4, the mean size representing a bin changes from that of (8.51) to ([Boss et al., 2001](#)):

$$\bar{D}(b_j < D < b_{j+1}) = \frac{\int_{b_j}^{b_{j+1}} n(D) D dD}{\int_{b_j}^{b_{j+1}} n(D) dD} = \frac{\int_{b_j}^{b_{j+1}} D^{1-\xi} dD}{\int_{b_j}^{b_{j+1}} D^{-\xi} dD} = \frac{(1-\xi)(b_{j+1}^{2-\xi} - b_j^{2-\xi})}{(2-\xi)(b_{j+1}^{1-\xi} - b_j^{1-\xi})}.$$

Normalizing a PSD by the total number of particles between the lower and upper boundaries provides a probability distribution for a particle to be within a specific bin. For a power-law PSD with exponent ξ :

$$p(b_j < D < b_{j+1}) = \frac{\int_{b_j}^{b_{j+1}} n(D)dD}{\int_{b_1}^{b_{M+1}} n(D)dD} = \frac{b_{j+1}^{1-\xi} - b_j^{1-\xi}}{b_{M+1}^{1-\xi} - b_1^{1-\xi}}.$$

The cumulative distribution, $P(D < D_c) = \int_{b_1}^{D_c} p(D)dD$ is therefore

$$P(b_1 < D < D_c) = \frac{\int_{b_1}^{D_c} n(D)dD}{\int_{b_1}^{b_{M+1}} n(D)dD} = \frac{D_c^{1-\xi} - b_1^{1-\xi}}{b_{M+1}^{1-\xi} - b_1^{1-\xi}},$$

which is useful to answer questions such as how numerically abundant are certain planktonic species compared to all other groups. A similar calculus is used to derive the probability distribution for particle area or volume.

[Jackson et al. \(1997\)](#) further discuss the case of computing the PSD for the solid fraction of particles if the particles are fractals (as is the case for oceanic aggregates). This requires assumptions regarding the change of fractal dimension with size and hence is not discussed further here (see, for example, [Maggi \(2007\)](#) and [Khelifa and Hill \(2006\)](#)).

8.13.3 Ecological Size Spectra

In ecology, size spectra are typically represented by an abundance or biomass size spectrum with the size axis often represented as mass or volume (e.g., the review by [Blanchard et al. \(2017\)](#)). A power-law function is often fit to the spectrum whose value is interpreted based on ecological theory. In such a case the number distribution is represented as

$$n(V) = AV^{-\xi_2} \quad [\text{particles cm}^{-3} \mu\text{m}^{-1}],$$

and we expect that $\xi_2 = \xi/3$, which is less steep than is a function of size as seen in (8.50). Because it is less steep, using the *arithmetic* mean for bin size is reasonable for the canonical distribution, more so than the *geometric* mean. In this case,

$$\bar{D}(b_j < V < b_{j+1}) = \frac{\int_{b_j}^{b_{j+1}} n(V)DdV}{\int_{b_j}^{b_{j+1}} n(V)dV} = \frac{\int_{b_j}^{b_{j+1}} DV^{-\xi_2}dD}{\int_{b_j}^{b_{j+1}} V^{-\xi_2}dD} = \frac{(1 - \xi_2)(b_{j+1}^{4/3-\xi_2} - b_j^{4/3-\xi_2})}{(4/3 - \xi_2)(b_{j+1}^{1-\xi_2} - b_j^{1-\xi_2})}.$$

8.13.4 Correctly Fitting a Power Law to a PSD

Suppose one is interested in fitting a power-law model to a PSD, for example as a simple descriptor of a relative contribution of large versus small particles in a given sample. What is the correct way to fit the PSD such that the exponent will be appropriately computed whether it is computed from a number, area, or volume distribution?

The answer is as follows. The number distribution $n(D_i)$, where D_i is the size representing the i^{th} bin, has an uncertainty $\delta(D_i)$. For example, if the uncertainty is due to counting alone, $\delta(D_i) = \sqrt{n(D_i)}$ (another source of uncertainty may be instrument sensitivity, particularly at the small end of the PSD). If the metric for fitting is to minimize

the root-mean-square error, we want to find the fit parameters A and ξ that minimize the cost function

$$\chi = \sum_{i=1}^M \left[\frac{n(D_i) - AD_i^\xi}{\delta(D_i)} \right]^2.$$

A more robust fit that reduces the weight of outliers may be found by minimizing

$$\chi = \sum_{i=1}^M \frac{|n(D_i) - AD_i^\xi|}{\delta(D_i)}.$$

If the appropriate uncertainties are used for each type of size distribution (be it number, area, or volume), the exponent found should be the same as we would expect, e.g. the exponent for the differential volume distribution equals that of the differential number distribution plus three.

On the other hand, if one simply fits a type-I regression line to $\log[n(D_i)]$ versus $\log[D_i]$, the implicit assumption is that the relative uncertainty in $n(D_i)$ is constant and the exponent obtained will not be consistent between the different size distributions.

8.13.5 Building PSDs with a Different Rule for the Bin Size

In principle, one could use a different bin size convention from the canonical use of bins increasing as a power law. For example, if one desires a PSD where, for any ξ , the volume in each bin is the same, the process is as follows. The total volume (assuming spheres) is

$$\int_{b_1}^{b_{M+1}} A \frac{\pi D^3}{6} D^{-\xi} dD = \frac{A\pi(b_1^{4-\xi} - b_{M+1}^{4-\xi})}{6(4-\xi)} = V_0.$$

For every bin to have the same volume we have

$$\int_{b_j}^{b_{j+1}} A \frac{\pi D^3}{6} D^{-\xi} dD = \frac{A\pi(b_j^{4-\xi} - b_{j+1}^{4-\xi})}{6(4-\xi)} = \frac{V_0}{M},$$

which implies that

$$b_1^{4-\xi} - b_2^{4-\xi} = b_2^{4-\xi} - b_3^{4-\xi} = \dots = b_M^{4-\xi} - b_{M+1}^{4-\xi} = \frac{b_1^{4-\xi} - b_{M+1}^{4-\xi}}{M}.$$

With a choice of b_1 , b_{M+1} , and M , we can compute all the other bin boundaries.

On the other hand, if one wanted to have bins of constant numbers of particles (for example to have similar counting errors in all bins), we would require the same number of particles in each bin. Assuming we have N_0 particles in M bins spanning from b_1 to b_{M+1} gives

$$\int_{b_1}^{b_{M+1}} AD^{-\xi} dD = \frac{A(b_1^{1-\xi} - b_{M+1}^{1-\xi})}{1-\xi} = N_0.$$

For every bin to have the same number of particles requires that

$$\int_{b_j}^{b_{j+1}} AD^{-\xi} dD = \frac{A(b_j^{1-\xi} - b_{j+1}^{1-\xi})}{1-\xi} = \frac{N_0}{M},$$

which implies that

$$b_1^{1-\xi} - b_2^{1-\xi} = b_2^{1-\xi} - b_3^{1-\xi} = \dots = b_M^{1-\xi} - b_{M+1}^{1-\xi} = \frac{b_1^{1-\xi} - b_{M+1}^{1-\xi}}{M}$$

or

$$b_2^{1-\xi} = b_1^{1-\xi} - \frac{b_1^{1-\xi} - b_{M+1}^{1-\xi}}{M}, \dots, b_{j+1}^{1-\xi} = b_j^{1-\xi} - \frac{b_1^{1-\xi} - b_{M+1}^{1-\xi}}{M}.$$

Again, with a choice of b_1 , b_{M+1} , and M , we can compute all of the other bin boundaries.

Radiative Transfer Theory

The Light and Radiometry Chapter 1 showed how various physical and mathematical quantities such as energy and solid angle can be combined to describe light in terms of radiance. The Inherent Optical Properties Chapter 3 showed how various IOPs are used to describe the absorbing and scattering properties of the medium through which light propagates. If the radiance is known, then all irradiances and all apparent optical properties (AOPs) can be computed. As was seen in the organization chart of Fig. 1, the radiative transfer equation provides the connection between the IOPs and boundary conditions (such as the incident sky radiance) and the radiance.

The goal of this chapter is to develop a family of *radiative transfer equations* (RTEs), which can be used to predict the radiance within a medium, given the IOPs of the medium and suitable boundary conditions. These equations provide the theoretical framework for all of optical oceanography and ocean-color remote sensing. The first section gives a qualitative derivation of the scalar (unpolarized) radiative transfer equation (SRTE). Although not physically rigorous, this derivation is seen in many textbooks and is provided here both for historical reasons and to point out the inadequacies of that derivation. However, there is a hierarchy of RTEs. At the top is a very general vector radiative transfer equation (VRTE, Eq. 9.14) capable of describing polarized light propagation in matter that is directionally non-isotropic, that can absorb light differently for different states of polarization, and that contains scattering particles of any shape and random or non-random orientation. Although very general in its ability to simulate any situation encountered in optical oceanography, the full set of IOP inputs to this equation is never measured in the oceanographic setting.

The IOP inputs to the most general VRTE become considerably simpler if the medium has mirror symmetry. The resulting VRTE (Eq. 9.18) is suitable for computation of polarized radiative transfer in the oceanic setting, especially after restriction to a plane-parallel geometry.

The vector-level equations can be further simplified to obtain, in a rigorous fashion, the SRTE (Eq. 9.21) shown in Fig. 1. That equation for the total radiance is only approximate, but the inputs are relatively simple to measure and model, so this equation finds wide use

in oceanography¹.

The chapter finishes with sections on several related topics in radiative transfer theory, namely beam and point spread functions, and the lidar equation.

Solution of the vector and scalar RTEs must be done numerically except for a few trivial cases such as non-scattering media. Approximate analytical solutions for the radiance can be obtained under very restrictive (and unphysical) conditions such as the Sun in a black sky and only single scattering within the water. These solutions are developed in the next chapter. Chapter 11 on Maxwell's equations discusses light from the viewpoint of the classical theory of electromagnetism.

9.1 The Scalar Radiative Transfer Equation; Heuristic Derivation

This section presents a heuristic or qualitative derivation of the the scalar radiative transfer equation (SRTE). The end results (Eqs. (9.9), (9.10), or (9.12)) are correct but, strictly speaking, the steps of the derivation are physically incorrect because polarization is ignored. In addition, the derivation of this section gives no way to estimate the errors that result from ignoring polarization. Nevertheless, this form of the derivation of the SRTE is found in almost every textbook and it does serve as a useful way to remember the various physical processes that contribute to light propagation in an absorbing and scattering medium like the ocean or atmosphere. This derivation is also of historical interest because it shows how the founding fathers of radiative transfer theory proceeded in order to obtain a governing equation (e.g., [Preisendorfer, 1965](#), page 65) before the link between fundamental physics and radiative transfer theory was firmly established.

The development beginning in section 9.2 on the General Vector Radiative Transfer Equation outlines the steps of a physically rigorous derivation of various levels of RTEs, ending with the SRTE (9.21) and including an error estimate.

9.1.1 Radiative Processes

To the extent that polarization can be ignored, the SRTE expresses conservation of energy written for a collimated beam of radiance traveling through an absorbing, scattering and emitting medium. We thus begin by considering the various processes that can occur when light interacts with an atom or molecule.

The light (electromagnetic radiation) may be annihilated, leaving the atom or molecule in an excited state with higher internal (electronic, vibrational, or rotational) energy. All or part of the absorbed radiant energy may be subsequently converted into thermal (kinetic) or chemical energy (manifested, for example, in the formation of new chemical compounds during photosynthesis). If the molecule almost immediately (on a femtosecond (10^{-15} s) or shorter time scale) returns to its original internal energy state by re-emitting radiation of the same energy as the absorbed radiation (but probably traveling in a different direction

¹The SRTE yields further equations involving the irradiances. One example is the two-flow equations for the plane irradiances E_d and E_u seen in *Light and Water* ([Mobley, 1994](#), Section 5.11). However, those equations cannot be solved for the irradiances unless *ad hoc* assumptions are made about the angular shape of the radiance distribution. Although once important, the two-flow equations have little value now that computers allow for very accurate numerical solutions of both the vector and scalar radiative transfer equations. The two-flow equations are therefore not discussed in this book.

from the original radiation), the process is called *elastic scattering*. Because of the extremely short time required for these events, elastic scattering can reasonably be thought of as the light interacting with the molecule and simply “changing direction” without an exchange of energy with the scattering molecule.

The excited molecule also may emit radiation of lower energy (longer wavelength) than the incident radiation. The molecule thus remains in an intermediate excited state and may at a later time emit new radiation and return to its original state, or the retained energy may be converted to thermal or chemical energy. Indeed, if the molecule is initially in an excited state, it may absorb the incident light and then emit light of greater energy (shorter wavelength) than the absorbed light, thereby returning to a lower energy state. In either case the scattered (emitted) radiation has a different wavelength than the incident (absorbed) radiation, and the process is called *inelastic scattering*. One important example of this process in the ocean is Raman scattering by water molecules (Section 7.2). Fluorescence (Section 7.3) is an absorption and re-emission process that occurs on a time scale of 10^{-11} to 10^{-8} sec. If the re-emission requires longer than about 10^{-8} sec, the process is usually called phosphorescence. The physical and chemical processes that lead to the vastly different times scales of Raman scattering versus fluorescence versus phosphorescence are much different. The distinctions between the very short time scale of Raman “scattering” versus the longer time scale of fluorescence “absorption and re-emission” do not concern us in the derivation of the time-independent RTE. However, the terminology has evolved somewhat differently, e.g., Raman scattering usually refers to “incident” and “scattered” wavelengths, whereas fluorescence usually refers to “excitation” and “emission” wavelengths.

The reverse process to absorption is also possible, as when chemical energy is converted into light; this process is called *emission*. An example of this is bioluminescence (Section 2.6), in which an organism converts part of the energy from a chemical reaction into light.

In order to formulate the RTE, it is convenient to imagine the total light field as many beams of electromagnetic radiation of various wavelengths coursing in all directions through each point of a water body. We then consider a single one of these beams, which is traveling in some direction (θ, ϕ) and has wavelength λ . This beam and the processes affecting it are illustrated in Fig. 9.1.

Now think of all the ways in which that beam’s energy can be decreased or increased. Bearing in mind the preceding comments, the following six processes are both necessary and sufficient to write down an energy balance equation for a beam of light on the phenomenological level:

- Process 1:** loss of energy from the beam through annihilation of the light and conversion of radiant energy to nonradiant energy (absorption)
- Process 2:** loss of energy from the beam through scattering to other directions without change in wavelength (elastic scattering)
- Process 3:** loss of energy from the beam through scattering (perhaps to other directions) with change in wavelength (inelastic scattering)
- Process 4:** gain of energy by the beam through scattering from other directions without change in wavelength (elastic scattering)
- Process 5:** gain of energy by the beam through scattering (perhaps from other directions) with a change in wavelength (inelastic scattering)

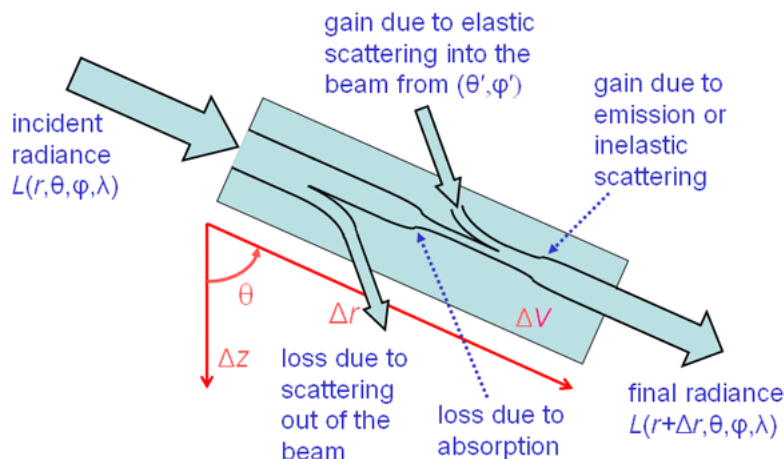


Figure 9.1: Illustration of a single beam of radiance and the processes that affect it as it propagates a distance Δr .

Process 6: gain of energy by the beam through creation of light by conversion of non-radiant energy into radiant energy (emission)

Next we must mathematically express how these six processes change the radiance as the beam travels a short distance Δr in passing through a small volume ΔV of water, which is represented by the blue rectangle of Fig. 9.1.

Processes 1 and 3. It is reasonable to assume that the *change* in radiance while traveling distance Δr due to absorption is proportional to the incident radiance, i.e., the more incident radiance there is, the more is lost to absorption. Thus we can write

$$\frac{L(r + \Delta r, \theta, \phi, \lambda) - L(r, \theta, \phi, \lambda)}{\Delta r} = \frac{\Delta L(r + \Delta r, \theta, \phi, \lambda)}{\Delta r} = -a(r, \lambda)L(r, \theta, \phi, \lambda). \quad (9.1)$$

Here $\Delta L(r + \Delta r, \theta, \phi, \lambda)$ denotes the change in L between r and $r + \Delta r$. The minus sign is necessary because the radiance decreases (energy is disappearing, so ΔL is negative) along Δr . Referring back to Eq. (3.4) of the Inherent Optical Properties section, it is easy to see that the present Eq. (9.1) is just the definition of the absorption coefficient written as a change in radiance over distance Δr , rather than as a change in absorptance. Thus the proportionality constant $a(r, \lambda)$ in Eq. (9.1) is just the absorption coefficient as defined in the IOP section. Note that absorption at the wavelength λ of interest accounts both for energy converted to non-radiant form (absorption) and for energy that disappears from wavelength λ and re-appears at a different wavelength (inelastic scattering). Either process leads to a loss of energy from the beam at wavelength λ .

Process 2. In a similar fashion, the loss due to elastic scattering out of the (θ, ϕ) beam direction into all other directions can be written as

$$\frac{\Delta L(r + \Delta r, \theta, \phi, \lambda)}{\Delta r} = -b(r, \lambda)L(r, \theta, \phi, \lambda), \quad (9.2)$$

where $b(r, \lambda)$ is the scattering coefficient as defined in the IOP Chapter 3.

Process 4. This process accounts for elastic scattering from all other directions into the beam direction (θ, ϕ) . Figure 9.2 shows Fig. 9.1 redrawn to illustrate scattering along path

length Δr from one direction (θ', ϕ') into the direction (θ, ϕ) of interest. These incident and final directions correspond to scattering angle ψ as shown in Fig. 9.2.

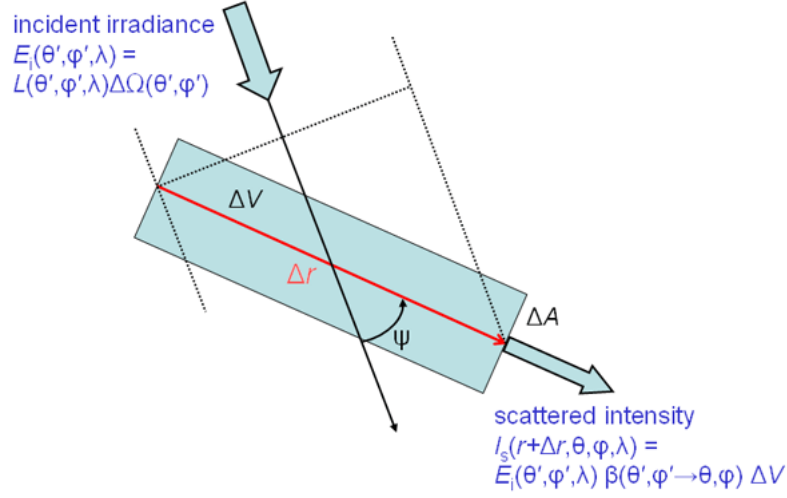


Figure 9.2: Illustration of a beam of radiance in direction (θ', ϕ') generating radiance in the direction of interest (θ, ϕ) by elastic scattering.

Recalling from Eq. (3.9) that one definition of the volume scattering function β is scattered intensity per unit incident irradiance per unit volume, we can write

$$I_s(r + \Delta r, \theta, \phi, \lambda) = E_i(\theta', \phi', \lambda) \beta(\theta', \phi' \rightarrow \theta, \phi; \lambda) \Delta V . \quad (9.3)$$

Here $I_s(r + \Delta r, \theta, \phi, \lambda)$ is the intensity exiting the scattering volume at location $r + \Delta r$ in direction (θ, ϕ) . All of this intensity is created along Δr by scattering from direction (θ', ϕ') into (θ, ϕ) , so $\Delta I_s(r + \Delta r, \theta, \phi, \lambda) = I_s(r + \Delta r, \theta, \phi, \lambda)$. The incident irradiance E_i is computed on a surface normal to the incident beam direction, as illustrated by the dotted lines in Fig. 9.2. We can rewrite E_i as the incident radiance times the solid angle of the incident beam:

$$E_i(\theta', \phi', \lambda) = L(\theta', \phi', \lambda) \Delta \Omega(\theta', \phi') . \quad (9.4)$$

Next recall from the Geometrical Radiometry section that intensity is radiance times area. Thus the intensity created by scattering along pathlength Δr and exiting the scattering volume over an area ΔA can be written as

$$\Delta I_s(r + \Delta r, \theta, \phi, \lambda) = \Delta L(r + \Delta r, \theta, \phi, \lambda) \Delta A , \quad (9.5)$$

where $\Delta L(r + \Delta r, \theta, \phi, \lambda)$ is the radiance created by scattering along Δr and exiting the scattering volume over a surface area ΔA . Using Eqs. (9.4) and (9.5) in (9.3) and writing the scattering volume as $\Delta V = \Delta r \Delta A$ gives

$$\frac{\Delta L(r + \Delta r, \theta, \phi, \lambda)}{\Delta r} = L(\theta', \phi', \lambda) \beta(\theta', \phi' \rightarrow \theta, \phi; \lambda) \Delta \Omega(\theta', \phi') . \quad (9.6)$$

This equation gives the contribution to $\Delta L(r + \Delta r, \theta, \phi, \lambda) / \Delta r$ by scattering from one particular direction (θ', ϕ') . However, ambient radiance may be passing through the scattering

volume from all directions. We can sum up the contributions to $\Delta L(r + \Delta r, \theta, \phi, \lambda)/\Delta r$ from all directions by integrating the right hand side of Eq. (9.6) over all directions,

$$\frac{\Delta L(r + \Delta r, \theta, \phi, \lambda)}{\Delta r} = \int_0^{2\pi} \int_0^\pi L(\theta', \phi', \lambda) \beta(\theta', \phi' \rightarrow \theta, \phi; \lambda) \sin \theta' d\theta' d\phi' , \quad (9.7)$$

where we have written the element of solid angle in terms the angles using Eq. (1.12).

Processes 5 and 6. Process 5 accounts for radiance created along pathlength Δr in direction (θ, ϕ) at wavelength λ by inelastic scattering from other all other wavelengths $\lambda' \neq \lambda$. Each such process, such as Raman scattering by water molecules or fluorescence by chlorophyll or CDOM molecules, requires a separate mathematical formulation to specify how radiance is absorbed from an incident beam at wavelength λ' and converted to the wavelength λ of interest.

Process 6 accounts for radiance created *de novo* by emission, e.g., by bioluminescence, and each emission process again requires a separate formulation to define how the light is emitted as a function of location, direction, and wavelength. For the moment, we can simply include a generic *source function* that represents creation of radiance along pathlength Δr in direction (θ, ϕ) at wavelength λ by any inelastic scattering or emission process. Thus we write just

$$\frac{\Delta L(r + \Delta r, \theta, \phi, \lambda)}{\Delta r} = S(r, \theta, \phi, \lambda) , \quad (9.8)$$

without specifying the mathematical form of the source function S . Previous sections have developed the source functions for bioluminescence (Section 2.6.1) , Raman scattering (Section 7.2), and fluorescence by chlorophyll (Section 7.4) and CDOM (Section 7.5).

We can now sum of the various contributions to the changes in L along Δr . We can also take the conceptual limit of $\Delta r \rightarrow 0$ and write

$$\frac{dL(r, \theta, \phi, \lambda)}{dr} = \lim_{\Delta r \rightarrow 0} \frac{\Delta L(r + \Delta r, \theta, \phi, \lambda)}{\Delta r} .$$

9.1.2 Standard Forms of the RTE

The net change in radiance due to all six radiative processes is the sum of the right hand sides of Eqs. (9.1), (9.2) , (9.7), and (9.8). We thus obtain an equation relating the changes in radiance with distance along a given beam direction to the optical properties of the medium and the ambient radiance in other directions:

$$\begin{aligned} \frac{dL(r, \theta, \phi, \lambda)}{dr} &= - [a(r, \lambda) + b(r, \lambda)] L(r, \theta, \phi, \lambda) \\ &+ \int_0^{2\pi} \int_0^\pi L(r, \theta', \phi', \lambda) \beta(r; \theta', \phi' \rightarrow \theta, \phi; \lambda) \sin \theta' d\theta' d\phi' \\ &+ S(r, \theta, \phi, \lambda) \quad (\text{W m}^{-3} \text{ sr}^{-1} \text{ nm}^{-1}) . \end{aligned} \quad (9.9)$$

This is one form of the SRTE, written for changes in radiance along the beam path.

In oceanography, it is usually convenient to use a coordinate system with the depth z being normal to the mean sea surface and positive downward. Thus depth z is a more convenient spatial coordinate than location r along the beam path. Changes in r are related

to changes in z as shown in Fig. 9.1: $dr = dz/\cos\theta$. Using this in Eq. (9.9), assuming that the ocean is horizontally homogeneous, and recalling that $a + b = c$, we get

$$\begin{aligned} \cos\theta \frac{dL(z, \theta, \phi, \lambda)}{dz} &= -c(z, \lambda)L(z, \theta, \phi, \lambda) \\ &+ \int_0^{2\pi} \int_0^\pi L(z, \theta', \phi', \lambda)\beta(z; \theta', \phi' \rightarrow \theta, \phi; \lambda) \sin\theta' d\theta' d\phi' \\ &+ S(z, \theta, \phi, \lambda) . \end{aligned} \quad (9.10)$$

This equation expresses location as geometric depth z and the IOPs in terms of the beam attenuation c and the volume scattering function β .

Other forms of the RTE are often used. The nondimensional *optical depth* ζ is defined by

$$d\zeta = c(z, \lambda)dz . \quad (9.11)$$

Dividing Eq. (9.10) by $c(z, \lambda)$ and using (9.11) gives the SRTE written in terms of optical depth. It is also common to use $\mu = \cos\theta$ as the polar angle variable. Recalling Eq. (3.14) of the Volume Scattering Function section, we can factor the volume scattering function β into the scattering coefficient b times the scattering phase function $\tilde{\beta}$. Finally, recalling the definition of the albedo of single scattering $\omega_o = b/c$, we can re-write Eq. (9.10) as

$$\begin{aligned} \mu \frac{dL(\zeta, \mu, \phi, \lambda)}{d\zeta} &= -L(\zeta, \mu, \phi, \lambda) \\ &+ \omega_o(\zeta, \lambda) \int_0^{2\pi} \int_{-1}^1 L(\zeta, \mu', \phi', \lambda)\tilde{\beta}(\zeta; \mu', \phi' \rightarrow \mu, \phi; \lambda) d\mu' d\phi' \\ &+ \frac{1}{c(\zeta, \lambda)} S(\zeta, \mu, \phi, \lambda) . \end{aligned} \quad (9.12)$$

This equation now shows all quantities as a function of optical depth.

Any of Eqs. (9.9), (9.10), or (9.12) is called the monochromatic (1 wavelength), one-dimensional (the depth is the only spatial variable), time-independent SRTE.

Form (9.12) of the SRTE yields an important observation: *In source-free ($S = 0$) waters, any two water bodies having the same single-scattering albedo ω_o , phase function $\tilde{\beta}$, and boundary conditions (including incident radiances) will have the same radiance distribution L at a given optical depth.* This is why optical depth, albedo of single scattering, and phase function are often the preferred variables in radiative transfer theory. Note, for example, that doubling the absorption and scattering coefficients a and b leaves ω_o unchanged, so that the radiance remains the same for a given optical depth. However, the geometric depth corresponding to a given optical depth will be different after such a change in the IOPs.

We can convert geometric depth to optical depth, or vice versa, by integrating Eq. (9.11):

$$\zeta = \int_0^z c(z', \lambda)dz' \quad \text{or} \quad z = \int_0^\zeta \frac{d\zeta'}{c(\zeta', \lambda)} , \quad (9.13)$$

Note that the optical depth ζ corresponding to a given geometric depth z is different for different wavelengths because the beam attenuation c depends on wavelength. This is inconvenient for oceanographic work, so Eq. (9.10) is usually the preferred form of the SRTE for oceanography.

We have now derived the SRTE in a form adequate for much oceanographic work. Technically, the SRTE is a linear integrodifferential equation because it involves both an integral and a derivative of the unknown radiance. This makes solving the equation for given IOPs and boundary conditions quite difficult. Fortunately, the radiance appears only to the first power. Nevertheless, there are almost no analytic (i.e., pencil and paper) solutions of the SRTE except for trivial special cases, such as non-scattering waters. Sophisticated numerical methods therefore must be employed to solve the SRTE for realistic oceanic conditions.

As mentioned at the start, the development on this section has been for unpolarized light. In a sense, this is always incorrect because polarization is an inherent property of electromagnetic waves (light in particular). Even if the incident beam is unpolarized, scattering (either by particles within the water or by reflection and transmission by the air-water surface) induces polarization. Thus underwater radiance in a particular direction is usually at least partially polarized. Nevertheless, the unpolarized, or scalar RTE (SRTE), derived here gives sufficiently accurate solutions for many (but not all) oceanographic applications. There are three main reasons for the utility of the SRTE in underwater optics:

- The particles responsible for scattering within the ocean are usually much larger than the wavelength of light. Polarization by scattering is greatest for particles much smaller than the wavelength (for example, Rayleigh scattering by water molecules). The larger the particle, the less polarization is induced by scattering of unpolarized sunlight.
- Multiple scattering is almost always significant underwater. Multiple scattering tends to depolarized the radiance, i.e., reduce the overall degree of polarization induced by single scattering events.
- Irradiances are often the radiometric quantity of interest. Irradiances are computed from integration of the radiance over direction, which tends to average out different polarizations in different directions.

However, if very accurate results are needed (e.g., radiance with an error of less than 10% in a given direction), or if the state of polarization itself is of interest, then a polarized or vector RTE (VRTE), must be used. The VRTE is more complicated than the SRTE developed here. The development of a general VRTE suitable of oceanography begins in the next section.

9.2 The General Vector Radiative Transfer Equation

After the introductory comments of the previous section of this chapter, it is now time to outline a rigorous derivation of the family of radiative transfer equations. The goal of this section and the next two is to obtain the vector and scalar radiative transfer equations in the forms that are commonly used in optical oceanography. The steps to these equations are outlined starting from the most fundamental physics. This sequence makes explicit what assumptions must be made along the way to get from an exact but highly abstract and mathematical theory to equations that are only approximate but which can be solved to get results that are accurate enough for many applications. In particular, the presentation

shows what is given up at each step in exchange for increased simplicity in the resulting physics and mathematics. The final result will be the scalar radiative transfer equation (SRTE), which was seen above (Eq. 9.10) and is found in many textbooks, and which is solved by the widely used the HydroLight radiative transfer software (Section 10.6).

The phenomenological nature radiative transfer theory as developed over the decades made its connections to fundamental physics unclear because of its heuristic derivations and (sometimes physically indefensible) assumptions (e.g. [Preisendorfer, 1965](#); [Mishchenko, 2013, 2014](#)). Theoreticians worked for decades to construct “an analytical bridge between the mainland of physics and the island of radiative transfer theory”, as R. W. Preisendorfer worded it ([Preisendorfer, 1965](#), page 389). Most recently, M. I. Mishchenko expended enormous effort to develop a physically and mathematically rigorous connection between Maxwell’s equations and the vector radiative transfer equation (VRTE), from which the SRTE is obtained after further simplifications. Thus Preisendorfer’s bridge has now been completed with a level of rigor worthy of any branch of physics ([Mishchenko, 2008a](#); [Mishchenko et al., 2016](#)). This section outlines how that bridge is constructed.

The path from fundamental physics to radiative transfer theory works its way through the following stages:

1. Quantum Electrodynamics
2. Maxwell’s Equations
3. The general vector radiative transfer equation (VRTE)
4. The VRTE for particles with mirror symmetry
5. The SRTE for the first component of the Stokes vector

9.2.1 Quantum Electrodynamics

Quantum electrodynamics (QED) is the fundamental physical theory that explains with total accuracy (as far as we know) the interactions of light and matter, or of electrically charged particles. In this theory, the electromagnetic field itself is quantized, and the photon is the quantum of the field. The electromagnetic force between charged particles is described by the exchange of so-called virtual photons (called “virtual” because they are not detectable in the role of transferring forces between charged particles). The photon is said to mediate or carry the electromagnetic force.

As an example of the accuracy of QED, the most recently measured value of a quantity called the electron spin g-factor is ([Hanneke et al., 2011](#))

$$g/2 = 1.001\,159\,652\,180\,73(28)$$

where the last two digits in parentheses give the uncertainty in the preceding two digits. The value predicted by the most recent QED calculations is ([Aoyama et al., 2015](#))

$$g/2 = 1.001\,159\,652\,181\,643(763)$$

This is an agreement between theory and experiment of around 1 part in 10^{12} , which is an accuracy equivalent to measuring the circumference of the Earth to within 0.04 millimeters. Similar results are obtained for other QED predictions. Because of its exceptional success

in explaining nature at its most fundamental level of elementary particle interactions, QED is often regarded as the most successful theory in physics.

Unfortunately, QED is exceptionally abstract and mathematical, and the calculations can be carried out only for simple interactions, such as the scattering of one elementary charged particle by another (although the calculations for even the simplest of interactions can be exceedingly complex according to the rules of QED). QED is the conceptual starting point for much of modern physics, but it isn't even remotely practical for everyday calculations. If you want to start learning about how the theory is formulated, there is no better place to begin than Feynman's classic popular exposition *QED: The Strange Story of Light and Matter* (Feynman, 1985). After that, if you want to get a bit more serious and see how the calculations are actually done, the best text I've found is *Introduction to Elementary Particles, Second Edition* (Griffiths, 2008). Learning about QED a fascinating journey, but this is not a road that oceanographers are required to travel.

9.2.2 Maxwell's Equations

QED is a quantum field theory that describes light and electromagnetism at the level of individual photons, which in QED are viewed as the quantized vibrational modes of the electromagnetic field. It is possible to take a "classical physics limit" of QED to get a classical field theory, in which the electromagnetic field is not quantized. The result is Maxwell's equations, which describe electric and magnetic fields as continuous functions of space and time. One way to think of this limiting process was stated by Sakurai (1967) as "The classical limit of the quantum theory of radiation is achieved when the number of photons becomes so large that the occupation number may as well be regarded as a continuous variable. The space-time development of the classical electromagnetic wave approximates the dynamical behavior of trillions of photons." Not surprisingly, getting from QED to Maxwell requires a high level of mathematical and physical sophistication.

Maxwell's equations are a set of four equations that govern electric and magnetic fields. Maxwell built upon earlier work by Coulomb, Gauss, Ampere, Faraday, and others; and the equations developed by these earlier geniuses are contained within his equations. These equations are applicable to the every-day needs of science and engineering. They govern matters as seemingly diverse as the generation, propagation, and detection of radio waves; the generation of electrical power for your house; the generation of Earth's magnetic field, the refraction of light at an air-water surface, and the scattering of light by phytoplankton. However, the greatest achievement of Maxwell and his equations is this: In a vacuum (as viewed by classical physics), a bit of vector calculus shows (Section 11.1.4) that each of the x , y , and z components of the electric and magnetic field vectors satisfies an equation of the form

$$\nabla^2 f = \mu_o \epsilon_o \frac{\partial^2 f}{\partial t^2}.$$

Here $\epsilon_o = 8.85 \times 10^{-12} \text{ A}^2 \text{ s}^4 \text{ kg}^{-1} \text{ m}^{-3}$ or $\text{coul}^2 \text{ N}^{-1} \text{ m}^{-2}$) is the electric permittivity of free space, and $\mu_o = 4\pi \times 10^{-7} \text{ kg m s}^{-2} \text{ A}^{-2}$ or N A^{-2}) is the magnetic permeability of free space. These are physical constants that describe the ability of electric and magnetic fields to penetrate a vacuum.

This equation describes a wave propagating with speed $c = 1/\sqrt{\mu_o \epsilon_o}$. Plugging in the numbers for ϵ_o and μ_o gives $c = 3 \cdot 10^8 \text{ m s}^{-1}$. As Maxwell observed in 1862, "This velocity is so nearly that of light that it seems we have strong reason to conclude that light itself

(including radiant heat and other radiations) is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws.” *This result is one of the greatest intellectual achievements of all time: not only were electric and magnetic fields tied together in the first unified field theory in physics, but light was shown to be an electromagnetic phenomenon*².

Although Maxwell’s equations are very accurate for predicting the phenomena of classical physics, they do fail when applied to atomic-scale processes, extremely strong fields, and the realm of “quantum optics,” which deals with individual photons. In particular, classical physics failed to describe blackbody radiation and the photoelectric effect, which is why Planck and Einstein had to make radical new assumptions in order to explain the observations. Those assumptions were the start of quantum mechanics³.

For solution in material media, Maxwell’s equations must be augmented with information about the electrical and magnetic properties of the material and with boundary conditions describing the specific problem of interest, such as sunlight passing through the sea surface and propagating through an ocean full of absorbing and scattering particles. The equations become somewhat more complicated in material media, although the general form remains the same. These equations do find occasional use in oceanography. For example, they have been used for measurement of oceanic currents via the electric fields set up by a conductor (salt water) moving through the earth’s magnetic field (e.g., [Larsen, 1973, 1992](#)). They are also the foundation for computations of sea ice emissivity and reflection at microwave frequencies (e.g., [Tan et al., 2016](#)). However, those applications are either non-optical or at wavelengths much longer than the visible spectrum. Direct numerical solutions of Maxwell’s equations are reviewed in [Mishchenko et al. \(2016, Section 7\)](#).

Maxwell’s equations are discussed in detail in Chapter 11.

For geometry as complex as a wind-blown sea surface and a water column filled with spatially varying absorbing and scattering matter, Maxwell’s equations and their boundary conditions are just too complex to solve for typical optical oceanographic problems. So, elegant and useful as they are for some problems, optical oceanographers must seek still further simplifications.

9.2.3 The General Vector Radiative Transfer Equation

The next simplifying step is to go from the world of electric and magnetic fields to the world of radiance. At optical wavelengths, the frequency of electromagnetic waves (light) is of order 10^{15} Hz. This is far higher than can be directly measured for a time-dependent propagating electric field. In practice, time-averaged irradiances are measured, in conjunction with various combinations of polarizing filters, over times long compared to the wave

²Maxwell developed his theory of electromagnetism in a series of papers in the 1860s, culminating with “A Dynamical Theory of the Electromagnetic Field,” published in 1865. If you do an internet search on “important events in world history in 1865” or check the Wikipedia page on 1865, the results will be dominated by events of the U.S. Civil War, plus mention of the first ascent of the Matterhorn, the arrival of Jumbo the elephant at the London zoo, and endless trivia about various kings and political shenanigans. You will not find a single mention of the publication of Maxwell’s paper, which is a sad commentary on what most historians consider important.

³To be precise, benefiting from the insights of a century of quantum mechanics, the failure of classical physics to describe the photoelectric effect and blackbody radiation was not due to the radiation field not being quantized (Maxwell’s equations), but due to the discrete nature of atomic absorption and emission processes, which in turn lead to discrete frequencies of the associated radiation. But that’s another story.

period. This time-averaging destroys the phase information contained in the instantaneous vector fields, but preserves the directional information about the plane of oscillation of the electric (and magnetic) field, i.e., it preserves information about the state of polarization of the light.

As described in the Polarization: Stokes Vectors Section 1.6, the state of polarization of a light field is specified by the four-component Stokes vector, whose elements are related to the complex amplitudes of the electric field vector \mathbf{E} resolved into directions that are parallel (E_{\parallel}) and perpendicular (E_{\perp}) to a conveniently chosen reference plane. In the oceanographic setting, this reference plane is usually the meridian plane, which is defined by the normal to the mean sea surface, $\hat{\mathbf{z}}$, and the azimuthal direction of the propagating light; the meridian plane is thus perpendicular to the mean sea surface. In a polar coordinate system (r, θ, ϕ) , $E_{\parallel} = E_{\theta}$ is the polar angle component of the electric field and $E_{\perp} = E_{\phi}$ is the azimuthal angle component. Figure 1.20 shows the meridian plane as commonly used for oceanography radiative transfer calculations.

As discussed in Section 1.7, there are two versions of the Stokes vector, which have different units and refer to different physical quantities. The *coherent* Stokes vector describes a quasi-monochromatic plane wave propagating in one exact direction, and the vector components have units of power per unit area (i.e., irradiance) on a surface perpendicular to the direction of propagation. The *diffuse* Stokes vector is defined as in Eq. (1.35) but describes light propagating in a small set of directions surrounding a particular direction and has units of power per unit area per unit solid angle (i.e., radiance). It is the diffuse Stokes vector that appears in the radiative transfer equations as developed here. The differences in coherent and diffuse Stokes vectors are rigorously presented in Mishchenko et al. (2002).

The transition from electric and magnetic fields to Stokes vectors yields a general 3-D vector radiative transfer equation. Let $\mathbb{S}(\mathbf{x}, \hat{\boldsymbol{\xi}})$ denote the diffuse Stokes vector at spatial location $\mathbf{x} = (x, y, z) = (r, \theta, \phi)$ for light propagating in the $\hat{\boldsymbol{\xi}} = (\theta, \phi)$ direction: $\mathbb{S}(\mathbf{x}, \hat{\boldsymbol{\xi}}) = \underline{\mathbb{S}}(\mathbf{x})\hat{\boldsymbol{\xi}}$. Then the VRTE has the form (Mishchenko (2008a), with slightly different notation and with the addition of an internal source term)

$$\hat{\boldsymbol{\xi}} \cdot \nabla \mathbb{S}(\mathbf{x}, \hat{\boldsymbol{\xi}}) = -\mathbb{K}(\mathbf{x}, \hat{\boldsymbol{\xi}}) \mathbb{S}(\mathbf{x}, \hat{\boldsymbol{\xi}}) + \iint_{4\pi} \mathbb{Z}(\mathbf{x}, \hat{\boldsymbol{\xi}}' \rightarrow \hat{\boldsymbol{\xi}}) \mathbb{S}(\mathbf{x}, \hat{\boldsymbol{\xi}}') d\Omega(\hat{\boldsymbol{\xi}}') + \mathbb{Q}(\mathbf{x}, \hat{\boldsymbol{\xi}}). \quad (9.14)$$

In this equation,

- $\mathbb{K}(\mathbf{x}, \hat{\boldsymbol{\xi}})$ is a 4×4 *extinction matrix*, which describes the attenuation (by the background medium and any particles imbedded in the medium) of the light propagating in direction $\hat{\boldsymbol{\xi}}$.
- $\mathbb{Z}(\mathbf{x}, \hat{\boldsymbol{\xi}}' \rightarrow \hat{\boldsymbol{\xi}})$ is a 4×4 *phase matrix*, which describes how light in an initial state of polarization and direction $\hat{\boldsymbol{\xi}}'$ in the incident meridian plane is scattered to a different state of polarization and direction $\hat{\boldsymbol{\xi}}$ in the final meridian plane.
- $\mathbb{Q}(\mathbf{x}, \hat{\boldsymbol{\xi}})$ is a 4×1 internal source term, which specifies the Stokes vector of any emitted light such as bioluminescence or light at the wavelength of interest that comes from other wavelengths via inelastic scattering.

In general, all 16 elements of \mathbb{K} and \mathbb{Z} are nonzero, and they depend on both location and direction (and wavelength; not shown). Equation (9.14) can describe polarized light propagation in matter that is directionally non-isotropic (e.g., in a crystal), that can absorb

light differently for different states of polarization (linear or circular dichroism), and that contains scattering particles of any shape and random or non-random orientation. (See [Mishchenko et al. \(2002\)](#) or [Mishchenko et al. \(2016\)](#) for a rigorous discussion of how these matrix elements are defined and computed. The $\mathbb{K}(\mathbf{x}, \hat{\xi})$ of Eq. 9.14 is a bulk extinction matrix. In Mishchenko’s papers this is written as $n_o \langle \mathbb{K} \rangle_{\hat{\xi}}$, where $n_o = N/V$ is the number of particles N in the volume of interest V , and $\langle \mathbb{K} \rangle_{\hat{\xi}}$ is a single-particle extinction matrix averaged over all N particles. A similar comment holds for \mathbb{Z} .)

By definition the phase matrix $\mathbb{Z}(\hat{\xi}' \rightarrow \hat{\xi})$ scatters light from one meridian plane (defined by $\hat{\mathbf{z}}$ and $\hat{\xi}'$) to another (defined by $\hat{\mathbf{z}}$ and $\hat{\xi}$). It is customary to write $\mathbb{Z}(\hat{\xi}' \rightarrow \hat{\xi})$ as a product of three matrices that

1. Transform the initial (unscattered) Stokes vector from the incident meridian plane to the scattering plane, which is the plane containing the incident ($\hat{\xi}'$) and scattered ($\hat{\xi}$) directions,
2. Scatter the Stokes vector from direction $\hat{\xi}'$ to $\hat{\xi}$, with calculations performed in the scattering plane, and
3. Transform the final Stokes vector from the scattering plane to the final meridian plane.

When this is done, the phase matrix is written as

$$\mathbb{Z}(\hat{\xi}' \rightarrow \hat{\xi}) = \mathbb{R}(\alpha) \mathbb{M}(\hat{\xi}' \rightarrow \hat{\xi}) \mathbb{R}(\alpha'). \quad (9.15)$$

Here $\mathbb{R}(\alpha')$ is a 4×4 matrix that transforms (“rotates” through an angle α') the incident Stokes vector into the scattering plane; $\mathbb{M}(\hat{\xi}' \rightarrow \hat{\xi})$ is a 4×4 matrix, the *scattering matrix*, which by definition scatters the incident Stokes vector to the final Stokes vector, with both expressed in the scattering plane; and $\mathbb{R}(\alpha)$ is a 4×4 matrix that rotates the final Stokes vector from the scattering plane to the final meridian plane. (In general, \mathbb{M} is called the scattering matrix. In the laboratory setting, \mathbb{M} is usually called the Mueller matrix.) This factoring of \mathbb{Z} separates the physics of the scattering process (\mathbb{M}) from the geometrical bookkeeping related to the coordinate systems (the rotation matrices). For the choice of a positive rotation being counterclockwise when looking into the beam, the Stokes vector rotation matrix is given by Eq. (1.51) (e.g., [Mishchenko et al., 2002](#), page 25). The rotation angles α' and α are shown in Fig. 1.21. These angles determined by spherical trigonometry as described in Section 1.7.2.

Equation (9.14) is almost never applied to oceanic problems. The reason is less for mathematical reasons than because there are no comprehensive data or models for the needed inputs \mathbb{K} and \mathbb{M} . To be useful, all elements of these matrices must be available for a wide range of oceanic waters. Further simplification is required.

9.3 The VRTE for Mirror-symmetric Media

A great simplification to the general VRTE of Eq. (9.14) results if the scattering particles are assumed to be (1) randomly oriented and (2) mirror symmetric. The concept of mirror (or bilateral) symmetry is illustrated in Fig. 9.3. Panel A represents a particle. The dashed line represents a mirror, and particle B is the mirror image of A. This B does not “look

like” or overlay A. However, if B is rotated by 180 deg about an axis normal to the plane of the figure, as illustrated by the green arrow, B becomes C, which looks exactly like A (is congruent with A). This means that A is a mirror-symmetric particle. Particle D has mirror image E. Rotating D by 180 deg about a vertical axis gives F, which (even without the “head”) cannot be made to overlay D. This means that the shape of D is not mirror symmetric. In general, a particle is mirror symmetric if a translation and/or rotation of the mirror-reflected particle can make it congruent with the original particle.

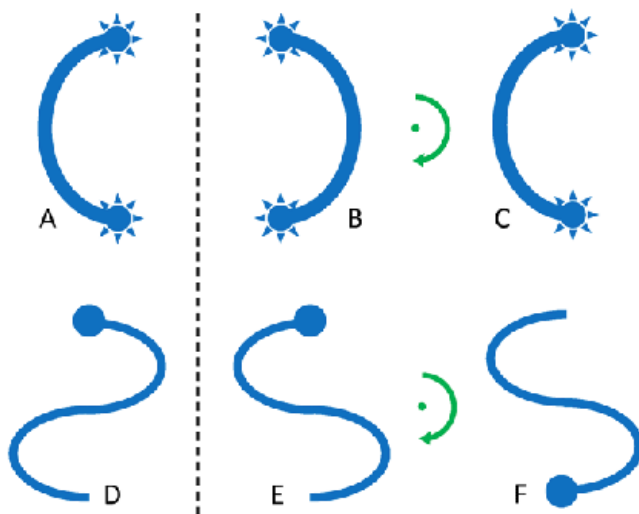


Figure 9.3: Figures A and D represent two particles. The dashed line is a mirror. The second column in the mirror image of the first column. The last column is the mirror image rotated by 180 deg about an axis normal to the figure, as illustrated by the green arrows. Top row: a mirror-symmetric particle; bottom row: a particle that is not mirror-symmetric.

Suppose that the scattering medium consists of randomly oriented, mirror-symmetric particles. The bulk medium is then directionally isotropic and mirror symmetric. In this case, the extinction matrix becomes diagonal, with each element equal to K_{11} . There is thus a common extinction coefficient for all states of polarization and directions of propagation:

$$\mathbb{K}(\mathbf{x}, \hat{\xi}) = c(\mathbf{x}) \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (9.16)$$

Here $c = K_{11}$ is the oceanographers’ “beam attenuation coefficient.” (In Mishchenko’s notation, $c = 2k'' + n_o \langle K_{11} \rangle$, which explicitly shows the contributions of the background medium (the water; the k'' term) and the imbedded particles.)

In addition, the scattering matrix becomes block diagonal and the scattering depends only on the included angle between directions $\hat{\xi}'$ and $\hat{\xi}$, and not on the directions $\hat{\xi}'$ and $\hat{\xi}$ themselves. This scattering angle ψ is given by (1.9), one form of which is

$$\cos \psi = \hat{\xi}' \cdot \hat{\xi} = \cos \theta' \cos \theta + \sin \theta' \sin \theta \cos(\phi - \phi').$$

The form of the scattering matrix is then (Mishchenko et al., 2002)

$$\mathbb{M} = \begin{bmatrix} M_{11}(\psi) & M_{12}(\psi) & 0 & 0 \\ M_{12}(\psi) & M_{22}(\psi) & 0 & 0 \\ 0 & 0 & M_{33}(\psi) & M_{34}(\psi) \\ 0 & 0 & -M_{34}(\psi) & M_{44}(\psi) \end{bmatrix}. \quad (9.17)$$

The restriction to an isotropic, mirror-symmetric medium gives a scattering matrix with only six independent elements.

The obvious question is now, “How realistic is the assumption that oceanic particles are mirror-symmetric?” Figure 9.4 shows a collection of oceanic diatoms (arranged for artistic purposes). Many of these are clearly not spherical (contrary to what is often assumed by modelers who cannot wean themselves away from Mie theory), but they all appear to be mirror-symmetric to a good approximation. The same holds true for many other species of phytoplankton which, if not roughly spherical, at least have bilateral symmetry. Likewise, atmospheric particles such as fog droplets, snowflakes, and ice crystals are often mirror symmetric. Use of Eqs. (9.16) and (9.17) is then justified.



Figure 9.4: Diatom shapes. Photo Kreispraeparat-25-G from <http://www.mikroskopie-ph.de>

Of course, this is not always the case. Figure 9.5 shows a photo of a chain-forming diatom *Chaetoceros debilis*. This cute little microbe appears to be a left-handed helix (assuming that the original image has not been flipped). Chiral (left- or right-handed) particles are not mirror symmetric. If the ocean contains lots of such particles (e.g., during a bloom), and all have the same handedness, then you have solve Eq. (9.14) of the previous section (perhaps with the simplification to one spatial dimension). Of course, no one ever does that, which is one more reason that numerical predictions of radiances may not agree

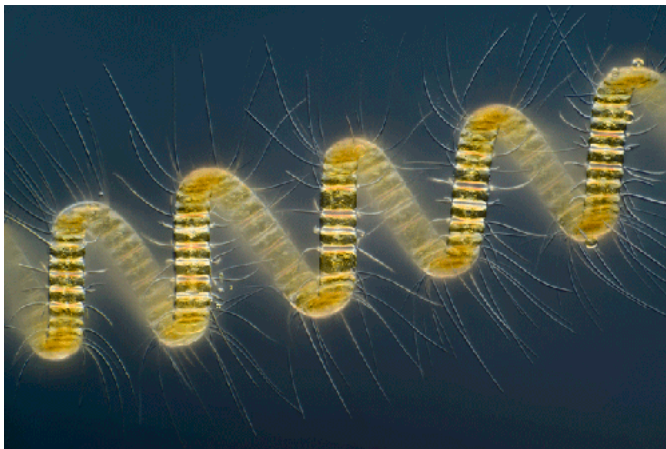


Figure 9.5: Photograph of *Chaetoceros debilis*; magnification of 250x. Photo 01-25002-Egmond-Chaetoceros-debilis from <https://www.wired.com/2013/10/nikon-small-world-2013/> ©Wim van Egmond.

perfectly with measurements. However, if there are equal numbers of randomly oriented left- and right-handed helical particles, then the bulk medium itself is mirror symmetric even though the individual particles are not, and the simplified VRTE is still applicable.

Consider next the calcium carbonate plates that are at times shed by coccolithophores. The material itself, CaCO_3 , is birefringent (meaning that the real index of refraction depends on the polarization), so in general would require the full 16-element \mathbb{K} matrix to describe the polarization-dependent extinction. However, the shapes of the liths (little platelets of CaCO_3) are mirror symmetric. So if the water is filled with randomly oriented liths, and the spatial distribution of the refractive index is also randomly oriented, then the bulk medium is mirror symmetric and Eqs. (9.16) and (9.17) are applicable. Likewise, the medium might contain spherical beads of dichroic glass (dichroic means that the imaginary index of refraction depends on the polarization). The spherical shape is mirror symmetric. However, if the spatial pattern of the dichroism is not random, then the full VRTE must be used. If the beads are randomly oriented in the pattern of dichroism, then the bulk medium is again mirror symmetric and Eqs. (9.16) and (9.17) are applicable.

There are very few measurements of scattering matrices for ocean waters. However, those that have been made indicate that the form of \mathbb{M} seen in Eq. (9.17) is a reasonable approximation. The reduced scattering matrix $\tilde{\mathbb{M}}(\psi)$ is defined as the scattering matrix with each element normalized by the $M_{11}(\psi)$ element, which is the volume scattering function for scattering of unpolarized light into unpolarized light. That is, $\tilde{M}_{ij}(\psi) = M_{ij}(\psi)/M_{11}(\psi)$.

Figure 9.6 shows measurements of the reduced scattering matrix for a location in the Atlantic Ocean. To a good approximation, the matrix elements shown in this figure obey the symmetries shown in Eq. (9.17). The measured \mathbb{M} is block symmetric with $M_{21} \approx M_{12}$ and $M_{43} \approx -M_{34}$ (although both of those elements are zero to within a bit of noise). In addition, $M_{33} \approx M_{44}$, which is characteristic of rotationally symmetric particles. For spherical particles, in addition, $M_{11} = M_{22}$; these are noticeably, but not greatly, different in the measured matrix elements. Similar results were obtained for about 200 measurements in Atlantic, Pacific, and Gulf of Mexico waters. Thus observation supports the use of

scattering matrices of the form of Eq. (9.17), corresponding to the assumption of non-spherical but mirror-symmetric particles, except for unusual circumstances such as a bloom of helical phytoplankton like those of Fig. 9.5.

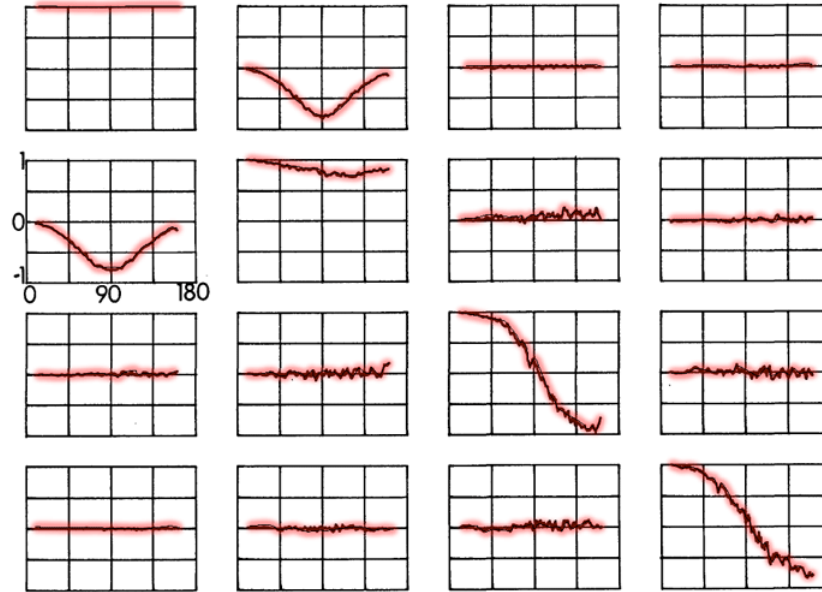


Figure 9.6: Measurements of an oceanic scattering matrix. Each element is normalized by the M_{11} element, which is not shown. Redrawn from Voss and Fry (1984, Fig. 3(a)).

In summary, the assumption of randomly oriented, mirror-symmetric particles gives an isotropic, mirror-symmetric medium with two very important simplifications to the general VRTE:

1. the attenuation coefficient does not depend on direction or state of polarization, and
2. the scattering matrix has only six independent elements, which depend only on the scattering angle (and of course on location and wavelength).

Figure 9.4 should convince you that the additional assumption of an ocean dominated by spherical phytoplankton may not be valid except during bloom conditions by a species that is nearly spherical.

If we further assume that the optical properties of the ocean depend only on depth z , and that the boundary conditions are horizontally homogeneous, then the left hand side of the three-dimensional VRTE reduces to a one-dimensional, ordinary (in space) derivative,

$$\hat{\xi} \cdot \nabla \mathbb{S}(\mathbf{x}, \hat{\xi}) = \cos \theta \frac{d}{dz} \mathbb{S}(z, \theta, \phi).$$

Likewise, the other location arguments simplify to depth only. The resulting one-dimensional (1-D) VRTE then reads

$$\cos \theta \frac{d}{dz} \mathbb{S}(z, \theta, \phi) = -c(z) \mathbb{S}(z, \theta, \phi) + \iint_{4\pi} \mathbb{R}(\alpha) \mathbb{M}(z, \psi) \mathbb{R}(\alpha') \mathbb{S}(z, \theta', \phi') d\Omega(\theta', \phi') + \mathbb{Q}(z, \theta, \phi). \quad (9.18)$$

Here $\mathbb{M}(\psi)$ has the form shown in Eq. (9.17) and the rotation matrices have the form seen in Eq. (1.51). In polar coordinates, the element of solid angle is given by $d\Omega(\theta', \phi') = \sin \theta' d\theta' d\phi'$. The integration is over all 4π steradians of the incident directions.

Equation (9.18) along with Eq. (9.17) is as far as we should go in simplifying the VRTE if we are interested in computing underwater polarized radiance distributions.

Solving Eq. (9.18) requires knowledge of the $M_{11}, M_{12}, M_{22}, M_{33}, M_{34}$, and M_{44} elements of the phase matrix. Several modern instruments have been developed for in situ measurement of the VSF, $M_{11}(\psi)$, over a wide range of scattering angles (e.g., Lee and Lewis (2003), Harmel et al. (2016), Li et al. (2012), Tan et al. (2013)). The POLVSM instrument described in Chami et al. (2014) can measure the nine elements of \mathbb{M} that do not involve circular polarization, i.e. $M_{ij}, i = 1, 2, 3$ and $j = 1, 2, 3$. The MASCOT instrument (Twardowski et al., 2012) measures the top row of \mathbb{M} , i.e. $M_{1j}, j = 1, 2, 3, 4$. The LISST-VSF (Slade et al., 2013) measures M_{11}, M_{12} and M_{22} . Only the LISST-VSF is commercially available. Not even the VSF is routinely measured during oceanographic field work. Therefore, when the VRTE is solved, usually some model is used to create the needed scattering matrix elements. This model is often Mie theory, which assumes spherical particles.

The three-dimensional VRTE is most commonly solved by Monte Carlo techniques. The assumption of a one-dimensional geometry gives a VRTE that is amenable to a variety of other numerical solution techniques, which have been employed in many studies of atmospheric and oceanic light fields.

9.4 The Scalar Radiative Transfer Equation; Rigorous Derivation

In many cases we are interested only in the total radiance, without regard to its state of polarization. Most instruments used in optical oceanography are by design not sensitive to polarization. Furthermore, it is commonly assumed that many processes of interest, such as absorption of light by phytoplankton to drive photosynthesis, do not depend on the polarization state⁴. However, underwater light fields are partially linearly polarized by transmission through the sea surface and by scattering within the water column, even though the Sun's direct beam is unpolarized. Thus there will be a contribution to the total radiance through conversion of this linearly polarized light to unpolarized by the (1,2) element of the scattering matrix.

9.4.1 From the VRTE for Total Radiance to the SRTE

Now let $c = \cos 2\alpha$, $s' = \sin 2\alpha'$, etc. for the rotation angles seen in Eq. (9.15) and in the rotation matrix of Eq. (1.51); let $M_{ij} = M_{i,j}(\psi)$; and use the form of \mathbb{M} seen in Eq. (9.17)

⁴However, an internet search on “polarization-dependent phytoplankton absorption” or similar key words results in exactly zero papers on this topic.

of the previous section. The phase matrix after matrix multiplications then becomes

$$\mathbb{Z} = \begin{bmatrix} M_{11} & c' M_{12} & -s' M_{12} & 0 \\ c M_{12} & c' c M_{22} - s' s M_{33} & -s' c M_{22} - c' s M_{33} & 0 \\ s M_{12} & c' s M_{22} + s' c M_{33} & -s' s M_{22} + c' c M_{33} & 0 \\ 0 & 0 & 0 & M_{44} \end{bmatrix}. \quad (9.19)$$

This equation shows that linear polarization (the Q and U Stokes parameters) contributes to the total I component via the M_{12} matrix element and the rotation angles that carry the incident linear polarization into the scattering plane. I does not depend on the rotation angle from the scattering plane to the final meridian plane because total radiance is independent of the coordinate system. Scattering by the volume of the water body cannot convert circular polarization (V) into unpolarized light because $Z_{14} = 0$, nor can circular polarization be converted to linear because $Z_{24} = Z_{34} = 0$. Similarly, unpolarized and linearly polarized light cannot be converted to circular because $Z_{41} = Z_{42} = Z_{43} = 0$. (However, total internal reflection can convert linear into circular and vice versa, as seen in Section 13.3.)

Writing Eq. (9.18) of the previous section for just the first element of the Stokes vector then gives

$$\begin{aligned} \cos \theta \frac{d}{dz} I(z, \theta, \phi) &= -c(z) I(z, \theta, \phi) \\ &+ \iint_{4\pi} M_{11}(z, \psi) I(z, \theta', \phi') d\Omega(\theta', \phi') + Q_I(z, \theta, \phi) \\ &+ \iint_{4\pi} \cos \alpha' M_{12}(z, \psi) Q(z, \theta', \phi') d\Omega(\theta', \phi') \\ &- \iint_{4\pi} \sin \alpha' M_{12}(z, \psi) U(z, \theta', \phi') d\Omega(\theta', \phi'). \end{aligned} \quad (9.20)$$

Here Q_I is the first element of the source term $\mathbb{Q} = [Q_I, Q_Q, Q_U, Q_V]^T$, which contributes to the first element of the Stokes vector $\mathbb{S} = [I, Q, U, V]^T$.

Equation (9.20) is the correct 1-D radiative transfer equation for the total radiance $I(z, \theta, \phi)$, under the simplifications described above. Unfortunately, this equation cannot be solved as written because Q and U are not known unless the full VRTE (9.18) is solved.

As already noted, routine solution of the VRTE is hindered not so much by the mathematics as by the lack of the needed inputs for the scattering matrices for various water-column constituents (phytoplankton, mineral particles, microbubbles, etc.). In response to this situation, modelers often raise their hands in surrender and simply drop the two integrals involving M_{12} in Eq. (9.20). This amounts to an ad hoc assumption—almost always incorrect—that underwater radiance distributions are unpolarized. However, the result is a scalar RTE for the total radiance that requires only beam attenuation and the VSF as input. In particular, this is the SRTE that is solved by the HydroLight radiative transfer software. There are much more data and models available for the VSF than for the other elements of the scattering matrix, so the trade off is decreased accuracy in exchange for requiring less input and doing simpler mathematics (i.e., faster computer programs).

In optical oceanography it is common to write the total radiance as $L(z, \theta, \phi, \lambda)$ rather than $I(z, \theta, \phi, \lambda)$. The volume scattering function is usually written as either $\beta(z, \psi, \lambda) =$

$M_{11}(z, \psi, \lambda)$, which emphasizes the scattering angle, or as $\beta(z; \theta', \phi' \rightarrow \theta, \phi; \lambda)$, which emphasizes the initial and final directions of the scattered light. With these changes in notation, the 1-D SRTE as commonly seen in textbooks (e.g., [Mobley, 1994](#), Eq. 5.23) is

$$\begin{aligned} \cos \theta \frac{dL(z, \theta, \phi, \lambda)}{dz} = & -c(z, \lambda)L(z, \theta, \phi, \lambda) \\ & + \int_0^{2\pi} \int_0^\pi L(z, \theta', \phi', \lambda) \beta(z; \theta', \phi' \rightarrow \theta, \phi; \lambda) \sin \theta' d\theta' d\phi' \\ & + S(z, \theta, \phi, \lambda) . \end{aligned} \quad (9.21)$$

Following *Light and Water* ([Mobley, 1994](#)), the source term in the SRTE is denoted by S , which is not to be confused with the Stokes vector \mathbb{S} . This equation expresses location as geometric depth z and the IOPs in terms of the beam attenuation c and the volume scattering function β . This form is convenient for studies at multiple wavelengths because the geometric depth z is independent of the IOPs.

The beam attenuation coefficient is the sum of the absorption and scattering coefficients: $c = a + b$. The VSF is often written as the product of the scattering coefficient and the scattering phase function: $\beta(z, \psi, \lambda) = b(z, \lambda) \tilde{\beta}(\psi)$. Dividing the SRTE by c , defining the optical depth by $d\zeta = c dz$, and the albedo of single scattering by $\omega_o = b/c$, gives

$$\begin{aligned} \cos \theta \frac{dL(\zeta, \theta, \phi, \lambda)}{d\zeta} = & -L(\zeta, \theta, \phi, \lambda) \\ & + \omega(\zeta, \lambda) \int_0^{2\pi} \int_0^\pi L(\zeta, \theta', \phi', \lambda) \tilde{\beta}(\zeta; \theta', \phi' \rightarrow \theta, \phi; \lambda) \sin \theta' d\theta' d\phi' \\ & + \frac{1}{c(\zeta, \lambda)} S(\zeta, \theta, \phi, \lambda) . \end{aligned} \quad (9.22)$$

This equation expresses location as nondimensional optical depth ζ and the IOPs in terms of the albedo of single scattering and the scattering phase function. This form is convenient for theoretical studies at one wavelength. It is difficult to use for multiple wavelengths because optical depth depends on the inherent optical properties. Thus the same geometric depth in meters corresponds to different optical depths at different wavelengths.

Equations (9.21) and (9.22) are the same as were obtained from the heuristic derivation in Section 9.1 (Eqs. 9.10 and 9.12). Now, however, Eq. (9.20) allows us to go one step further and estimate the errors that occur by ignoring polarization.

The magnitude of the error in computed values of the total radiance induced by ignoring the M_{12} terms in Eq. (9.20) can be crudely estimated as follows. Measurements show that underwater light fields are partially linearly polarized (e.g., [Cronin and Shashar, 2001](#); [?; You et al., 2011](#)). The degree of linear polarization is seldom more than 0.5, and is typically in the 0.1 to 0.3 range. Figure 9.6 shows that the maximum magnitude of $M_{12}(\psi)/M_{11}(\psi)$ is about 0.8 at $\psi = 90$ deg. For $Q/I = 0.3$ and a 90 deg scattering angle, the ratio of $M_{12}Q$ (or $M_{12}U$) to the $M_{11}I$ term is then

$$\frac{M_{12}Q}{M_{11}I} = 0.8 \times 0.3 \approx 0.25.$$

Thus for a viewing direction at right angles to the unscattered direction (often the Sun’s direct beam in the water), single scattering gives an error of order 25%. However, for near-forward and near-backward scattering angles, $M_{12} \approx 0$, and the error is much smaller. In addition, multiple scattering tends to reduce the degree of polarization. It thus seems reasonable to say that neglecting the M_{12} terms in Eq. (9.20) can cause errors in the total radiance as large as a few tens of percent in certain directions, but in other directions the errors will be of order 10% or even smaller. This error can be positive or negative, depending on the sign of Q and U . That is, in some directions the radiance value will be too large, and in other directions too small. These errors tend to cancel when computing irradiances, which are integrals of the radiance over direction. The historical justification for dropping the M_{12} terms in Eq. (9.20) has been that, although the errors in the total computed radiances are of order 10% (and possibly greater), the errors in the corresponding irradiances are only of order 1%. That is considered an acceptable trade off for the increased simplicity of solving Eq. (9.21) compared to the VRTE of the previous section.

9.4.2 Summary

This development started with QED and part-per-trillion accuracy and ended up with the SRTE and order of 10% accuracy in radiances but only order of 1% percent error in irradiances. This loss of accuracy is the price to be paid for an equation that is simple enough to solve on a routine basis for situations of practical oceanographic interest.

The development outlined here accomplishes two things:

1. It shows how Preisendorfer’s “analytical bridge between the mainland of physics and the island of radiative transfer theory” was constructed, although the actual construction requires an exceptionally high level of physics and mathematics.
2. The physically rigorous development makes clear all of the assumptions needed to arrive at an equation for the total radiance without regard for the state of polarization—the scalar radiative transfer equation, or SRTE. Moreover, it enables an estimate of the errors that may occur if polarization is ignored.

Neither of these accomplishments can be obtained from the heuristic derivation of the SRTE seen in Section 9.1.

The largest errors in computing underwater radiances usually come not from the approximations made to the general VRTE, but from the inputs to the simplified VRTE or SRTE. This is especially true if generic bio-geo-optical models for the absorption and scattering coefficients are used to convert chlorophyll and/or mineral particle concentrations to IOPs. Even if correct on average, such models can be off by tens of percent to an order of magnitude in any particular situation. A 10% error in the absorption coefficient can easily result in an order-of-magnitude error in the computed radiance or irradiance at depth. Guessing or modeling the phase function can give a factor-of-two or more error in backscatter directions, which gives an error of the same magnitude in the water-leaving radiance. It is certainly possible to get very good agreement between measurements and predictions based on the SRTE, but heroic efforts are required to measure all of the needed inputs. Examples of such comparisons can be seen in [Tzortziou et al. \(2006\)](#) and [Tonizzo et al. \(2017\)](#).

9.5 Beam and Point Spread Functions

This section defines two equivalent quantities, the beam and point spread functions, which describe light propagation in absorbing and scattering media. The beam spread function will be used in Section 9.6 to understand the return of a lidar signal when used for detection of in-water scattering layers or of bathymetric measurements. The point spread function is fundamental to image analysis, as will be seen in Appendix G.

9.5.1 The Beam Spread Function (BSF)

Consider a collimated source emitting spectral power P (units of W nm^{-1}) in direction $\psi = 0$ as shown in Fig. 9.7. As the beam passes through the medium, scattering will spread out the beam as illustrated by the green arrows in the figure, and absorption will reduce the beam power. The combined effects of scattering and absorption give some spectral plane irradiance $E(R, \psi)$ (units of $\text{W m}^{-2} \text{nm}^{-1}$) on the surface of a sphere of radius R at an off-axis angle ψ relative to the direction of the emitted beam. Distance R is often called the range. The irradiance sensor used to measure $E(R, \psi)$ has a cosine response for angles relative to the normal to the detector surface. It is assumed here that the emitted light is unpolarized and that the medium is isotropic, so that there is no azimuthal dependence of the detected irradiance. The green arrows in the figure represent one possible path for a light ray between the source and the detector.

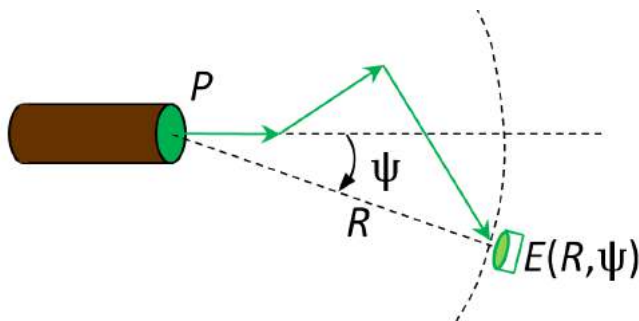


Figure 9.7: Geometry for definition of the Beam Spread Function.

The beam spread function (BSF) is then defined as the detected irradiance normalized by the emitted power:

$$BSF(R, \psi) \triangleq \frac{E(R, \psi)}{P} \left[\frac{\text{W m}^{-2} \text{nm}^{-1}}{\text{W nm}^{-1}} = \text{m}^{-2} \right]. \quad (9.23)$$

Recall that the Volume Scattering Function (VSF) defined in Section 3.1 describes a *single* scattering event. Two such scatterings are shown for the green arrows in Fig. 9.7. The BSF on the other hand describes the cumulative effects on the emitted beam of *all* of the scattering and absorption events between the source and the detector. The BSF thus depends both on the IOPs and on the distance R between the source and the detector, whereas the VSF depends only on the optical properties of the medium and is independent of the locations of the source and detector.

9.5.2 The Point Spread Function (PSF)

Now suppose that there is a source at the location of the detector in Fig. 9.7 and that this source is emitting spectral intensity $I(\gamma)$ (units of $\text{W sr}^{-1} \text{nm}^{-1}$) with an angular pattern

given by

$$I(\gamma) \triangleq \begin{cases} \frac{P}{\pi} \cos(\gamma) & 0 \leq \gamma \leq \pi/2 \\ 0 & \pi/2 < \gamma \leq \pi. \end{cases} \quad [\text{W sr}^{-1} \text{ nm}^{-1}]$$

For this intensity pattern the total emitted power is

$$\int_{4\pi} I(\gamma) d\Omega = 2\pi \int_0^{\pi/2} \frac{P}{\pi} \cos(\gamma) \sin(\gamma) d\gamma = P.$$

The short green arrows in Fig. 9.8 illustrate the emission of light by this cosine source. As illustrated by the long green arrows in this figure, the emitted intensity will give rise to a radiance $L(R, \psi)$, where R is the distance from the source and ψ is direction measured from the $\gamma = 0$ axis of the emitted intensity. The light emitted by this source is then detected by a well collimated radiance sensor that can scan past the source as shown in Fig. 9.8.

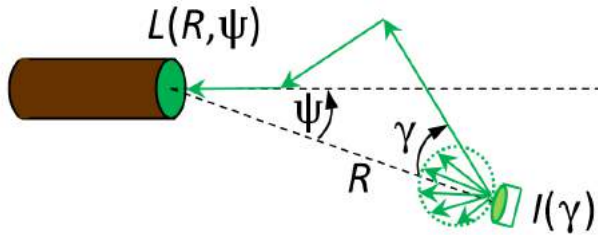


Figure 9.8: Geometry for definition of the Point Spread Function.

The Point Spread Function (PSF) is then defined as the detected radiance normalized by the maximum of the emitted intensity:

$$PSF(R, \psi) \triangleq \frac{L(R, \psi)}{P/\pi} \quad \left[\frac{\text{W m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}}{\text{W sr}^{-1} \text{ nm}^{-1}} = \text{m}^{-2} \right].$$

The PSF can be visualized as the “glow” of light around distant street light seen through a foggy atmosphere as seen in Fig. 9.9. Although a street light is not a cosine-emitting point source, the angular pattern of the glow seen around the light gives a qualitative feeling for the PSF, which is just the pattern of the radiance distribution centered on the source.

Figure 9.10 shows PSFs for a homogeneous water body with a Petzold average-particle phase function and a single-scattering albedo (3.7) of $\omega_o = 0.8$. These PSFs were computed by Monte Carlo ray tracing simulations as described in Appendix E. The left panel shows the PSF for the first 15 deg of ψ and for nondimensional optical distances between the source and detector of $\tau = cR = 0.1, 1, 5, 10,$ and 30 (c is the beam attenuation coefficient). These curves show that the magnitude of the PSF decreases as τ increases because of absorption. The colored dots are the centers of the angular bins used to tally the light rays in the Monte Carlo simulations.

The right panel of the figure shows the PSF values normalized to 1 at $\psi = 0$. These curves show that the shape of the PSF starts out very highly peaked near $\psi = 0$ for small τ and eventually becomes relatively flat in ψ as the optical distance increases. This progression of PSF shapes can be understood as follows.

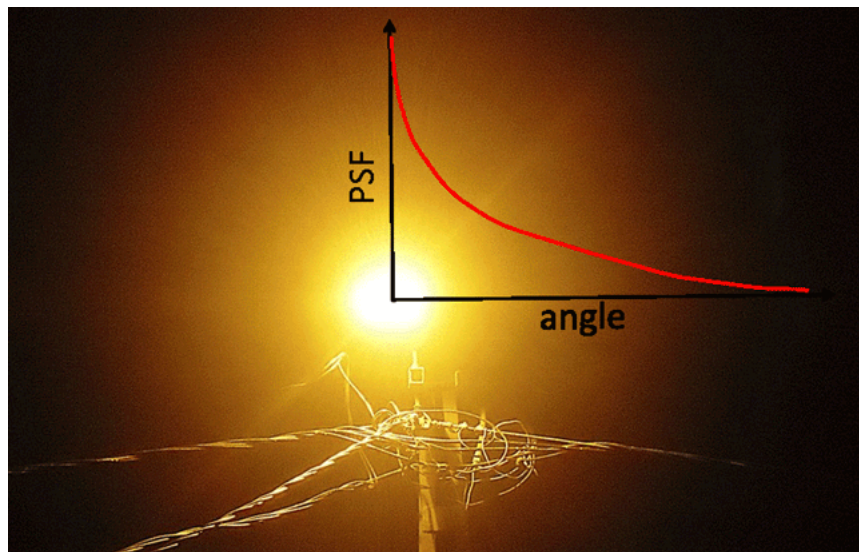


Figure 9.9: The glow of light around a distance street light seen through fog is essentially the PSF for the given distance and atmospheric properties.

As τ goes to zero, there is almost no chance for light to scatter. As the radiometer seen in Fig. 9.8 scans past the point source, it sees either almost nothing, or it sees the point source at $\psi = 0$. The PSF then approaches a Dirac delta function in ψ as $\tau \rightarrow 0$. As the optical distance increases, there is more and more chance for scattering until many rays have been scattered once. The PSF then begins to look similar to the scattering phase function, which describes the redistribution of radiance by single scattering. The red curve lying close to the curve for $\tau = 5$ is the normalized Petzold phase function used in these simulations. (The PSF would not in general ever have exactly the same shape as the phase function. Even if, at some distance, most light rays have been scattered once, others will not yet be scattered, and others will have been scattered more than once. Thus a PSF never describes just single scattering.)

Finally, as τ becomes very large, all rays have been scattered many times, and the resulting radiance distribution, hence the PSF, approaches the shape of the asymptotic radiance distribution L_∞ , which is described in Section 10.4. The black curve near the $\tau = 30$ PSF shows L_∞ as computed by HydroLight for the IOPs of these simulations. The $\tau = 30$ curve is still noticeably different from L_∞ . It is computationally expensive to trace enough rays to large optical distances to reproduce the asymptotic distribution. A computation for $\tau = 50$ required emitting 10^9 initial rays, of which about 0.05% reached $\tau = 50$ for $\psi < 15$ deg, but the resulting PSF (not shown) is much closer to L_∞ . Similarly, statistical noise can become large for large ψ angles if the angular bins are small. However, it is the small ψ values that are most important for image analysis (you are usually looking generally toward an object, not away from it), so noise at large angles is seldom a problem in practical applications.

Because of the importance of the PSF in image analysis (to be seen in Appendix G), considerable effort has been expended to develop models of the PSF as a function of water IOPs, e.g., Mertens and Replogle (1977), Voss and Chapin (1990), Voss (1991), McLean and Voss (1991), Gordon (1994b), McLean et al. (1998), Sanchez and McCormick (2002),

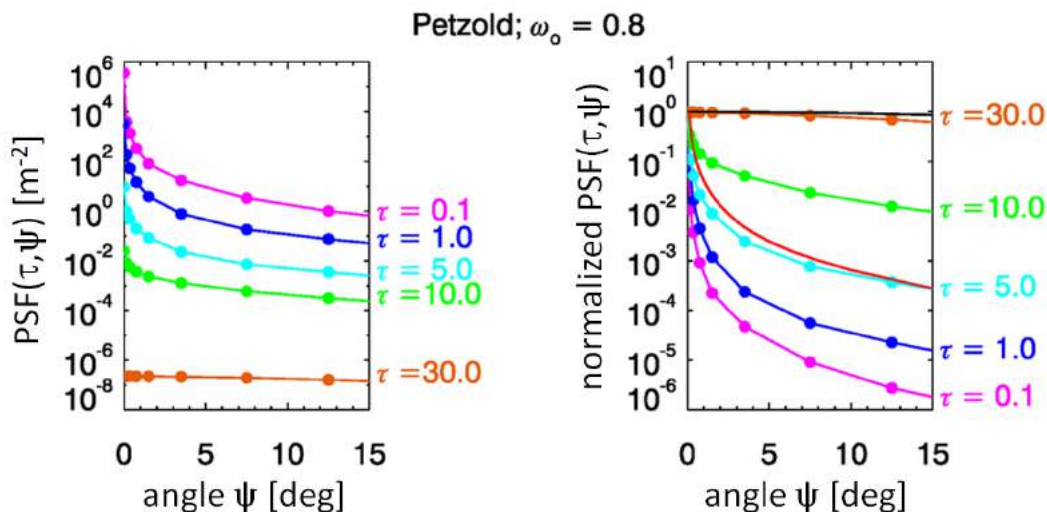


Figure 9.10: Example point spread functions. Left panel: PSF as a function of optical distance for $\psi = 0$ to 15 deg. Right panel: the same curves normalized to 1 at $\psi = 0$. In the right panel, the red line near the $\tau = 5$ curve is the normalized Petzold phase function used in the Monte Carlo simulations. The black line at the top is the asymptotic radiance distribution.

and [Xu and Yue \(2015\)](#).

9.5.3 Equivalence of the BSF and PSF

In Maxwell's Equations for the propagation of electromagnetic energy, if you change time to minus time, nothing changes except that the direction of propagation is reversed. This means that if light is propagating from the source to the receiver along the path shown in [Fig. 9.7](#), then light emitted from the receiver location in the same direction as the detected beam will propagate back along the same path to the source, as shown by the green arrows in [Fig. 9.8](#). This is known as the principle of electromagnetic reciprocity, or Helmholtz reciprocity, or the “if I can see you, you can see me” theorem. [Figures 9.7 and 9.8](#) were drawn to highlight the symmetry between the BSF and the PSF.

The BSF and PSF are superficially different: a collimated source and a cosine detector vs a cosine source and a collimated detector. However, reciprocity suggests that the BSF and PSF contain equivalent information. Indeed, the BSF and PSF are numerically the same. [Figure 9.11](#) shows PSF and BSF measurements from two very similar water bodies obtained by [Mertens and Replogle \(1977\)](#). The instruments were mounted on a 20 m long underwater frame, which allowed a maximum range R between source and detector of approximately 19 m. Although the measurements in the figure were made on different days and the water IOPs were slightly different, the closeness of the PSF and BSF over several orders of magnitude suggests that the BSF and PSF are numerically equal, as they stated without proof. [Gordon \(1994b\)](#) started with the radiative transfer equation and used reciprocity to show that the BSF and PSF are indeed numerically equal. It is thus customary refer to just the PSF, even if the geometry of a problem corresponds to that

of the BSF. (This is the case in the derivation of the lidar equation the next section, for example.)

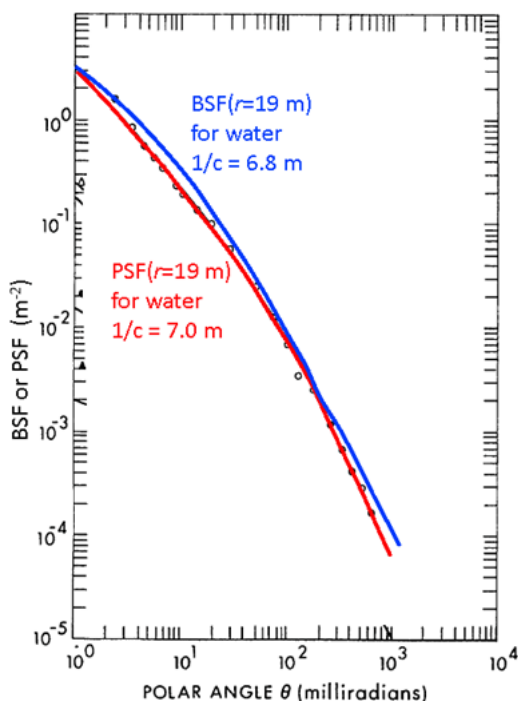


Figure 9.11: The BSF and PSF at 488 nm measured on two different days in very clear Bahamas water. The range was approximately 19 m. The BSF was measured when the water beam attenuation length was $1/c(488) = 6.8$ m . The PSF was measured on the following day when $1/c(488) = 7.0$ m . Curves extracted from Figs. 4 and 6 of [Mertens and Replogle \(1977\)](#).

9.6 A Lidar Equation for Oceanography

The purpose of a “lidar⁵ equation” is to compute the power returned to a receiver for given transmitted laser power, optical properties of the medium through which the lidar beam passes, and target properties. There are, however, many versions of lidar equations, each of which is tailored to a particular application. For example, [Measures \(1992\)](#) develops lidar equations for elastic and inelastic backscattering scattering by the medium, for fluorescent targets, for topographical targets, for long-path absorption, for broad-band lasers, and so on.

The lidar equation developed here⁶ applies to the detection of a scattering layer or in-water target as seen by a narrow-beam laser imaging the ocean from an airborne platform.

⁵Is it lidar, or Lidar, or LIDAR, or something else? [Deering and Stoker \(2014\)](#) argue that the acronym has now reached same status as radar and sonar, so lidar is the preferred form.

⁶Acknowledgment: I learned this derivation from Richard C. Honey, one of the pioneering geniuses of optical oceanography, and of many other fields ranging from antenna design to laser eye surgery. Dick Honey is unfortunately little known to the general community because he spent much of his career doing classified work.

This equation explicitly shows the effects of atmospheric and sea-surface transmission, the water volume scattering function and beam spread function, water-column diffuse attenuation, and transmitter and receiver optics.

9.6.1 Preliminaries

Table 9.1 lists for reference the variables involved with the derivation of the present form of the lidar equation.

variable	definition	units
H	Height of the airplane above the sea surface	m
z	Depth of the water layer being imaged	m
Δz	Thickness of the water layer being imaged	m
P_t	Power transmitted by the laser	W
ΔP_r	Power detected from water layer Δz	W
T_a	Transmittance by the air	nondimen
T_s	Transmittance by the water surface	nondimen
A_r	Receiver aperture area	m ²
Ω_{FOV}	Receiver field of view solid angle	sr
A_z	Area at depth z seen by the receiver	m ²
Ω	Solid angle of the receiver aperture as seen from depth z	sr
β_π	Water VSF for 180-deg backscatter	m ⁻¹ sr ⁻¹
BSF	Water beam spread function	m ⁻²
E_i	Irradiance incident (downward) onto the water layer at z	W m ⁻²
ΔE_r	Irradiance reflected (upward) by the water layer Δz	W m ⁻²
\overline{K}_{up}	Depth-averaged (over 0 to z) attenuation coefficient for upwelling (returning) irradiance ΔE_r	m ⁻¹

Table 9.1: Variables occurring in the lidar equation.

Figure 9.12 shows the geometry of the lidar system for detection of a scattering layer.

Recall from Eq. (3.24) of the section of measuring the volume scattering function that the VSF is operationally defined by

$$\beta(\psi) = \frac{\Delta \Phi_s(\psi)}{\Phi_i \Delta r \Delta \Omega},$$

where Φ_i is the power incident onto an element of volume defined by a surface area A and thickness Δr , and $\Delta \Phi_s(\psi)$ is the power scattered through angle ψ into solid angle $\Delta \Omega$. These quantities are illustrated in Fig. 3.2. In this derivation, quantities directly proportional to the layer thickness Δz ($= \Delta r$ in Fig. 3.2) will be labeled with a Δ . Thus

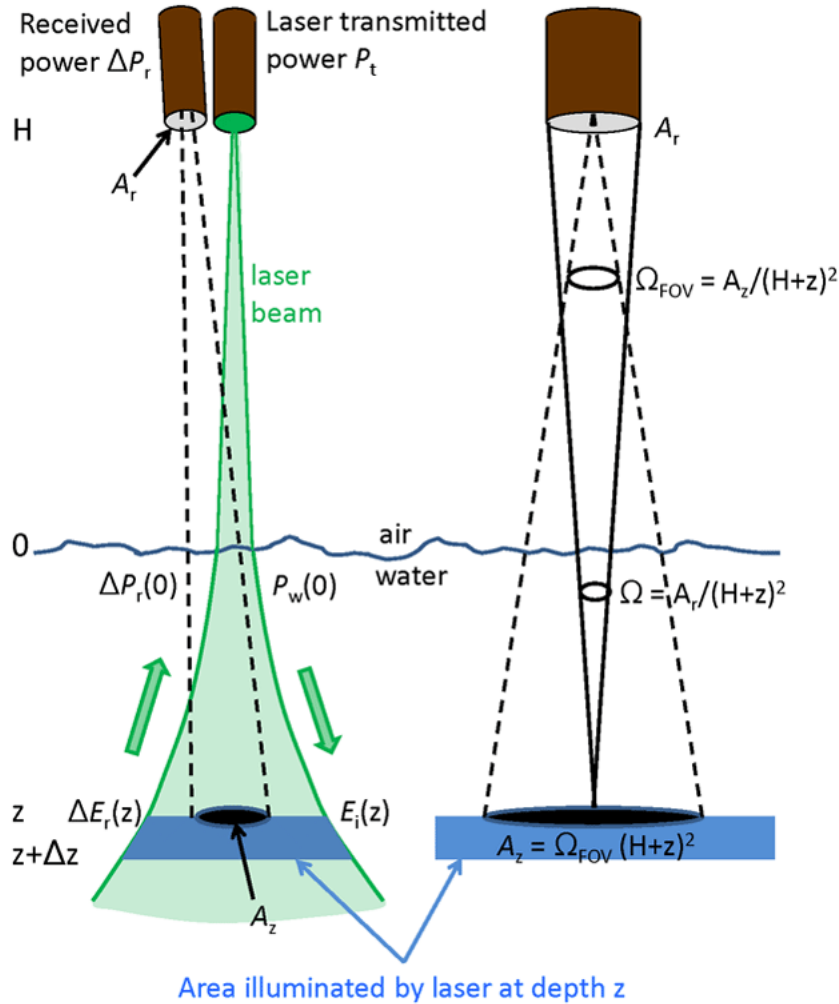


Figure 9.12: Geometry of the lidar system for detection of a scattering layer. The left panel illustrates the physical variables. The right panel illustrates the system optics; this panel ignores refraction at the sea surface.

$\Delta\Phi_s(\psi)$ is the power scattered by the water layer of thickness Δz . The incident power Φ_i onto area A gives an incident irradiance $E_i = \Phi_i/A$.

Now consider exact backscatter, which is a scattering angle of $\psi = \pi$, or $\psi = 180$ deg. The backscattered power $\Delta\Phi_s(\pi)$ exits the scattering volume through the same area A , so the backscattered irradiance is $\Delta E_s(\pi) = \Delta\Phi_s(\pi)/A$. The VSF for exact backscatter can then be written as

$$\beta_\pi \triangleq \beta(\psi = \pi) = \frac{\Delta\Phi_s(\pi)/A}{(\Phi_i/A) \Delta z \Omega} = \frac{\Delta E_s(\pi)}{E_i \Delta z \Omega}.$$

This gives

$$\Delta E_s(\pi) = E_i \beta_\pi \Omega \Delta z. \quad (9.24)$$

(Do not confuse this backscattered or reflected irradiance with $E_u = R E_d$, which is defined for any incident radiance distribution and for an arbitrarily thick layer of water. Here E_i is the irradiance for a collimated incident laser beam and a thin layer of water.)

Now suppose a collimated beam is emitting power P in direction $\psi = 0$. Then scattering and absorption in the medium will give some irradiance $E(R, \psi)$ on the surface of a sphere of radius R at an angle of ψ relative to the direction of the emitted beam, as illustrated by the green arrow in Fig. 9.7. Recall that the beam spread function (BSF) is then defined by Eq. (9.23), which is

$$BSF(R, \psi) \triangleq \frac{E(R, \psi)}{P} \quad (\text{m}^{-2}). \quad (9.25)$$

9.6.2 Derivation of the Lidar Equation

The derivation of the lidar equation for the stated application now proceeds via the following steps:

1. As illustrated in Fig. 9.12, the laser pulse has transmitted power P_t to start. After transmission through the atmosphere and sea surface, the pulse has power

$$P_w(0) = P_t T_a T_s$$

just below the water surface.

2. The laser pulse is still a narrow beam when it enters the water, but it then begins to spread out because of scattering, and it is attenuated by absorption. This process is quantified by the beam spread function. We are interested in the “on-axis” irradiance incident onto a layer of water at depth z . From Eq. (9.25), this is given by

$$E_i(z) = P_w(0) BSF(R = z, \psi = 0).$$

3. The irradiance that is reflected by a layer of thickness Δz at depth z is given by Eq. (9.24):

$$\Delta E_r(z) = E_i(z) \beta_\pi(z) \Omega \Delta z.$$

The solid angle Ω is determined by the aperture size of the receiver and the distance from the receiver to the layer. This is discussed in step 7 below.

4. The irradiance $\Delta E_r(z)$ heading upward from the layer Δz will be attenuated by some diffuse attenuation function \bar{K}_{up} as ΔE_r propagates back to the sea surface:

$$\Delta E_r(0) = \Delta E_r(z) \exp(-\bar{K}_{\text{up}} z).$$

The receiver sees only an area A_z at depth z . It is assumed that A_z is less than the total area illuminated by the spreading laser beam at depth z . Thus we can multiply $\Delta E_r(z)$ by A_z to get the power leaving the illuminated area that is seen by the receiver. The fraction of the total power reaching the surface, which is seen by the detector, is then

$$\Delta P_r(0) = \Delta E_r(z) A_z \exp(-\bar{K}_{\text{up}} z).$$

A_z is determined by the receiver field of view and the distance from the receiver to the layer.

5. The power just below the surface is now transmitted through the water surface and through the atmosphere to the receiver. The power detected is then

$$\Delta P_r = \Delta P_r(0) T_s T_a .$$

6. Combining the above results gives

$$\Delta P_r = P_t T_a^2 T_s^2 A_z \Omega B S F(z, 0) \beta_\pi \exp(-\bar{K}_{\text{up}} z) \Delta z .$$

7. A_z and Ω can be written in terms of known parameters. We are interested only in power that is reflected by the layer Δz into the solid angle Ω that will take the light into the receiver. As shown in the right panel of Fig. 9.12, the receiver aperture A_r and the range $H + z$ determine $\Omega = A_r / (H + z)^2$. Only power heading into Ω from area A_z , which is seen by the receiver, reaches the receiver. A_z is determined by the receiver field of view solid angle Ω_{FOV} and the range: $A_z = \Omega_{\text{FOV}} (H + z)^2$. Thus the $A_z \Omega$ factor in the preceding equation can be rewritten as

$$A_z \Omega = \Omega_{\text{FOV}} (H + z)^2 \frac{A_r}{(H + z)^2} = A_r \Omega_{\text{FOV}} .$$

This equation shows how the receiver optics affects the detected power. This is an application of the “ $A\Omega$ ” theorem of optical engineering. $A\Omega$ is also called the throughput or étendue of the system. Strictly speaking, the in-air solid angle Ω_{FOV} decreases by a factor of $1/n^2$ upon entering the water, where n is the water real index of refraction. However, the in-water Ω increases by a factor of n^2 when entering the air. For a round trip, air to water to air, these n^2 factors cancel. They are thus ignored here and in Fig. 9.12.

8. Collecting the above results gives the desired lidar equation:

$$\Delta P_r = P_t T_a^2 T_s^2 A_r \Omega_{\text{FOV}} B S F(z, 0) \beta_\pi \exp(-\bar{K}_{\text{up}} z) \Delta z . \quad (9.26)$$

Equation (9.26) nicely shows the effects of the transmitted power (P_t); atmospheric and surface transmission ($T_a^2 T_s^2$); receiver-optics ($A_r \Omega_{\text{FOV}}$); water-column optical properties ($B S F(z, 0) \beta_\pi \exp(-\bar{K}_{\text{up}} z)$); and layer thickness (Δz). The take-home message from this equation is that in order to understand lidar data, the water inherent optical properties you need to know are the beam spread function and β_π . This observation was in part the incentive for the work of [Mertens and Replogle \(1977\)](#), [Voss and Chapin \(1990\)](#), [Voss \(1991\)](#), [McLean and Voss \(1991\)](#), [Maffione and Honey \(1992\)](#), [Gordon \(1994b\)](#), [McLean et al. \(1998\)](#), [Sanchez and McCormick \(2002\)](#), [Dolin \(2013\)](#), [Xu and Yue \(2015\)](#), and others. These papers present several models for beam/point spread functions in terms of the water absorption and scattering properties. Several of those models are reviewed in [Hou et al. \(2008\)](#).

There are various arguments about what to use for \bar{K}_{up} , which depends on both the water optical properties and on the imaging system details. It is intuitively expected that \bar{K}_{up} , which is defined for a horizontally small patch of upwelling irradiance, will be greater than the diffuse attenuation coefficient for upwelling irradiance, K_u , which is defined for a horizontally infinite light field. Likewise, we expect that \bar{K}_{up} will be less than c , the beam

attenuation coefficient. Thus $K_u < \bar{K}_{up} < c$. Because \bar{K}_{up} is an attenuation function for a finite patch of reflected irradiance, computing its value is inherently a three-dimensional radiative transfer problem. To pin down the value of \bar{K}_{up} more accurately thus requires either actual measurements for a particular system and water body, or three-dimensional radiative transfer simulations (usually Monte Carlo simulations) tailored to a given system and water properties.

9.6.3 Example Lidar Calculation

To develop some intuition about Eq. (9.26), consider the following example application. Suppose a 532 nm laser is being used to look for objects in the water that have an area of 1 m^2 . The receiver FOV must be small enough that the object can be distinguished from its surroundings. For $H = 100 \text{ m}$ and $z = 10 \text{ m}$, this requires that

$$\Omega_{\text{FOV}} \leq \frac{1 \text{ m}^2}{(100 + 10) \text{ m}^2} \approx 8 \times 10^{-5} \text{ sr}.$$

For a 15 cm radius receiving telescope, $A_r = \pi(0.15 \text{ m})^2 \approx 0.07 \text{ m}^2$. For normal incidence at the sea surface, $T_s \approx 0.97$, and for a clear atmosphere, $T_a \approx 0.98$. Suppose the water is Jerlov type 1 coastal water for which $K_d(532) \approx 0.15 \text{ m}^{-1}$ (Fig. 4.24 or Mobley (1994, page 130)), and assume that $\bar{K}_{up} = K_d$. Further assume that $BSF \approx \exp(-0.2z)$, since the lidar beam attenuation will be more “beam like” than diffuse attenuation, and $c > K_d$. Finally, take $\beta_\pi = 10^{-3} \text{ m}^{-1} \text{ sr}^{-1}$ (Fig. 6.13 or 6.13 for “coastal ocean” water). Then for $\Delta z = 1 \text{ m}$ at a depth of 10 m, the water returns the fraction

$$\begin{aligned} \frac{\Delta P_r}{P_t} &= (0.98)^2 (0.97)^2 (0.07) (8 \times 10^{-5}) e^{-(0.2)10} (10^{-3}) e^{-(0.15)10} (1) \\ &\approx 1.5 \times 10^{-10} \end{aligned}$$

of the transmitted power.

Now suppose that the laser beam hits an object at $z = 10 \text{ m}$ whose surface is a Lambertian reflector of 2% (irradiance) reflectance. Then 0.02 of E_i is reflected into $2\pi \text{ sr}$. The layer backscatter $\beta_\pi \Delta z = 0.001 \text{ sr}^{-1}$ is then replaced by

$$\frac{0.02}{2\pi \text{ sr}} \approx 0.003 \text{ sr}^{-1}.$$

If all other terms remain the same, the object would return about three times the signal as the water itself.

Radiative Transfer Solutions

With the exception of the conceptually simple and widely used Monte Carlo ray tracing techniques (Appendices E and F), this book does not discuss numerical methods for solving radiative transfer equations. There are many such methods, all highly mathematical and not of interest to most optical oceanographers. (One such technique, invariant imbedding theory, briefly mentioned here, is used in the commercial HydroLight software.) However, there are a few approximate analytical methods that are frequently cited in the oceanographic and remote-sensing literature, namely the successive-order-of-scattering (SOS) series, and the resulting single-scattering (SSA) and quasi-single-scattering (QSSA) approximations. These warrant discussion, even if just for historical reasons. There is also a solution for the SRTE at the irradiance level, namely Gershun’s law, that is fundamental to light calculations in many aquatic ecosystem models. The chapter finishes with the derivation of Gershun’s law, a discussion of the behavior of light at large depths in homogeneous water (the asymptotic radiance distribution), a qualitative description of HydroLight, and discussions of energy conservation and closure in optical oceanography.

10.1 Solving the Radiative Transfer Equation

The 1-D, time-independent RTEs are *linear two-point boundary value problems*. That is to say, there are boundary conditions describing the radiance at the top and bottom of the atmosphere or ocean (the two spatial points), and the propagation of radiance within the medium—between the boundaries—is governed by the linear integro-differential RTE. In the oceanographic setting, the upper boundary condition specifies the sky radiance incident onto the sea surface. The lower boundary condition specifies how the sea bottom reflects the downwelling radiance. The “sea bottom” can be a physical bottom, or the deepest depth in the water column at which the RTE needs to be solved to obtain the radiances above that depth. In the vector-level equations, the sky radiance specification is in terms of Stokes vectors; for the scalar RTE the sky input is specified by the unpolarized sky radiance.

Two-point boundary value problems in general do not have solutions. Consider a simple example:

$$\frac{dy}{dx} = x^2$$

with the two-point boundary conditions

$$y = 0 \quad \text{at} \quad x = 0 \quad \text{and} \quad y = 2 \quad \text{at} \quad x = 1.$$

What is the function $y(x)$ that satisfies this differential equation and the two boundary conditions? Integration gives $y = x^3/3$, which satisfies the first boundary condition but not the second. Thus the problem as stated has no solution; i.e., there is no such function $y(x)$. This equation however does have a solution for the two boundary conditions

$$y = 0 \quad \text{at} \quad x = 0 \quad \text{and} \quad y = \frac{1}{3} \quad \text{at} \quad x = 1.$$

In the case of the radiative transfer equation and boundary conditions on the radiance at the sea surface and sea bottom (when properly formulated), the equation does have a solution. We know a solution must exist simply because light propagates in the ocean! However, finding that solution is quite difficult. Even the scalar radiative transfer equation quite difficult to solve.

10.1.1 Exact Analytical Solutions

Exact analytical (i.e., pencil and paper) solutions of the SRTE can be obtained only for very simple situations, such as no scattering. There is no function (that anyone has ever found) of the form

$$L(z, \theta, \phi, \lambda) = f(a, VSF, \text{Sun angle, bottom reflectance, etc.}),$$

where f is some function into which you can “plug in” the absorption coefficient, the VSF, and other parameters and get back the radiance. This is true even for very simple situations such as homogeneous water with isotropic scattering. Even the conceptually simple geometry of an isotropically emitting point light source in an infinite homogeneous ocean is unsolved. (A very complicated solution for the the scalar irradiance $E_o(r)$ around an isotropically emitting point source with isotropic scattering does exist ([Davison and Sykes, 1957](#), Eq 5.25) or ([Mobley, 1996](#), Section 9.2). This may seem surprising because other point-source problems, such as the electric field around a point charge or the gravitational field around a point mass, are often very simple. The difference with optics lies in the complications caused by absorption and scattering within the medium surrounding the point source, which do not exist for problems like the gravitational field around a point mass.

10.1.2 The Bouguer-Lambert-Beer Law

If there is no scattering, the SRTE ([9.9](#)) reduces to just

$$\frac{dL(r, \theta, \phi, \lambda)}{dr} = -a(r, \lambda)L(r, \theta, \phi, \lambda) + S(r, \theta, \phi, \lambda). \quad (10.1)$$

This is a linear, first order, ordinary differential equation, which is easily solved (see any text on differential equations, e.g., [Rainville, 1964](#), page 36). If the medium is homogeneous,

so that the absorption coefficient and source function do not depend on distance r , the solution of Eq. (10.1) is (dropping the wavelength and direction arguments for brevity)

$$L(r) = L(0)e^{-ar} + \frac{S}{a}[1 - e^{-ar}] , \quad (10.2)$$

where $L(0)$ is the initial radiance at distance $r = 0$. In source-free water, $S = 0$ and the solution is a simple exponential decay of the initial radiance with distance:

$$L(r) = L(0)e^{-ar} , \quad (10.3)$$

This result is known as Beer's Law, Lambert's Law, Bouguer's Law, or some hyphenated combination of these names¹. An equivalent result was derived in another way in the discussion of measuring IOPs (Eq. 3.21).

Note that for great distances, $r \rightarrow \infty$, the radiance depends only on the source function and the absorption coefficient: $L \rightarrow S/a$. This result for the asymptotic behavior of the radiance with distance also holds when scattering is present, as shown in Section 10.5 on an analytical asymptotic solution for internal sources.

10.1.3 Approximate Analytical Solutions

A number of approximate analytical solutions to the SRTE can be derived after simplifying the SRTE in various ways. One of these approximate solutions is the Single-scattering Approximation (SSA), which will be obtained in Section 10.2. This solution is developed by assuming that the water is homogeneous, the sea surface is level, the Sun is a point source in a black sky, there are no internal sources, and only single-scattering of light rays is considered. The Quasi-single-scattering Approximation (QSSA) to be seen in Section 10.3 is a further simplification of the SSA. These solutions are derived below and in Gordon (1994a). A much more complicated solution including second-order scattering (i.e., including rays that have been scattered twice) is developed in Walker (1994, Section 2-6).

These approximate solutions of the SRTE are useful for isolating the main factors influencing underwater radiances. However, the solutions depend on various simplifying assumptions and the predicted radiances are typically accurate to a few tens of percent at best, and can be off by an order of magnitude.

10.1.4 Numerical Solutions

If accurate solutions of the vector or scalar RTE are to be obtained for realistic oceanic conditions, numerical methods must be used. Many such methods have been developed.

¹The exponential decay of light through a medium was first reported by Bouguer (1729, Fig. 4) based on observations of candles made by eye. Lambert (1760) placed Bouguer's graphical result in the mathematical form of an exponential decay with distance. Beer (1852) found that for a fixed distance, the transmitted light decreased as an exponential function of the concentration of salts in solution. It seems that Bouguer should be credited with the original understanding of the exponential decrease of light when traveling through a material medium, and Lambert gets credit for putting Bouguer's results into the mathematical form seen in Eq. (10.3). It is then "Lambert's law" if you use Eq. (10.3) to predict the attenuation of light as a function of distance for a given substance. It is "Beer's law" if you use Eq. (10.3) to predict the attenuation of light as a function of concentration for a given distance. To give full credit to everyone, it should be called the "Bouguer-Lambert-Beer law." It seems that "Beer's law" or the "Lambert-Beer law" seems to be most common in optical oceanography.

Some of these solution techniques have been tailored to specific environments, such as stellar or planetary atmospheres, and are not commonly used in oceanography. The numerical methods most commonly employed in oceanographic radiative transfer, and their salient characteristics, can be summarized as follows:

Monte Carlo

- based on conceptually simple physics that mimics how nature absorbs and scatters idealized light rays
- completely general; can solve time-dependent and 3-D problems with arbitrary geometry
- relatively easy to incorporate polarization
- easy to program
- computed radiances have statistical errors, which can be reduced by tracing more initial rays (requiring longer computer times)
- computer run times can be extremely slow for some problems (e.g., solving the RTE to large optical depths; run times increase exponentially with optical depth)
- sophisticated mathematical “tricks” can sometimes be used to speed up the calculations

Invariant Imbedding

- computed radiances do not have statistical errors
- includes all orders of multiple scattering
- can solve only 1-D problems (the one dimension being the depth in optical oceanography)
- highly mathematical
- difficult to program
- is extremely fast (run times increase linearly with optical depth)

Discrete Ordinates

- computed radiances do not have statistical errors
- does not handle highly peaked scattering phase functions well
- highly mathematical
- difficult to program
- models the medium as a stack of homogeneous layers
- is fast for irradiance calculations and homogeneous water, but can be slow for radiances or if many layers are needed to resolve depth-dependent IOPs (run time is proportional to the number of layers)

Because of their simplicity and generality, Monte Carlo methods are widely used to solve RTEs in fields as diverse as oceanography, atmospheric sciences, astronomy, medical physics, and nuclear engineering. The trade-off for their simplicity and generality is long

computer run times for many problems. Invariant imbedding is the solution technique used in the HydroLight numerical model. The scattering phase functions for atmospheric aerosols are not as highly peaked at very small scattering angles as are those for oceanic particles. Discrete ordinates can handle aerosol phase functions well and is often used in atmospheric optics, but is seldom used for underwater calculations because of the need to resolve highly peaked phase functions and to have many layers if the IOPs vary strongly with depth. Each of these techniques gives the same answer for the same inputs and boundary conditions for the RTE, for problems where all three techniques are applicable, as can be seen in the model comparison study of Mobley et al. (1993). They differ only in their internal mathematics and the resulting computer run times, and well-debugged and validated computer programs exist for each. In this sense, solving the RTE in the oceanographic setting can be considered a “solved problem.”

10.2 The Single-Scattering Approximation (SSA)

As previously noted 10.1, exact analytical solutions of the RTE exist only for a few idealized and unphysical situations such as no scattering. There are, however, a few approximate analytic solutions. In pre-computer days these were useful computational tools. These approximate solutions are no longer needed for numerical computation, but they are still useful for isolating the most important processes governing light propagation in the ocean and can provide guidance in interpretation of radiometric data. This section develops one such solution: the single-scattering approximation (SSA). The next section discusses the related quasi-single scattering approximation (QSSA).

10.2.1 The Successive-Order-of-Scattering (SOS) Solution Technique

Recall Eq. (9.22), the optical depth form of the time-independent, 1-D (plane parallel geometry) RTE:

$$\begin{aligned} \mu \frac{dL(\zeta, \mu, \phi, \lambda)}{d\zeta} &= -L(\zeta, \mu, \phi, \lambda) \\ &+ \omega_o(\zeta, \lambda) \int_0^{2\pi} \int_{-1}^1 L(\zeta, \mu', \phi', \lambda) \tilde{\beta}(\zeta; \mu', \phi' \rightarrow \mu, \phi; \lambda) d\mu' d\phi' \\ &+ \frac{1}{c(\zeta, \lambda)} \Sigma(\zeta, \mu, \phi, \lambda) . \end{aligned}$$

We next make a number of simplifications by assuming that

- The water is homogeneous, so that the IOPs do not depend on depth;
- The water is infinitely deep;
- The sea surface is level (zero wind speed);
- The Sun is a point source in a black sky, so that the incident radiance onto the sea surface is collimated;
- There are no internal sources or inelastic scattering.

The RTE then becomes, for a given wavelength λ , which we henceforth drop for brevity,

$$\begin{aligned} \mu \frac{dL(\zeta, \mu, \phi)}{d\zeta} &= -L(\zeta, \mu, \phi) \\ &+ \omega_o \int_0^{2\pi} \int_{-1}^1 L(\zeta, \mu', \phi') \tilde{\beta}(\mu', \phi' \rightarrow \mu, \phi) d\mu' d\phi' . \end{aligned} \quad (10.4)$$

A powerful technique for solving differential equations is to attempt a power series solution in which higher order terms of the series are weighted by a powers of a parameter whose magnitude is less than 1. The higher order terms then contribute less and less to the sum that represents the solution. The albedo of single scattering, ω_o , meets the requirement for an expansion parameter. We therefore attempt a solution of Eq. (10.4) of the form

$$\begin{aligned} L(\zeta, \mu, \phi) &= \sum_{k=0}^{\infty} \omega_o^k L^{(k)}(\zeta, \mu, \phi) \\ &= L^{(0)}(\zeta, \mu, \phi) + \omega_o L^{(1)}(\zeta, \mu, \phi) + \omega_o^2 L^{(2)}(\zeta, \mu, \phi) + \dots \end{aligned} \quad (10.5)$$

The notation $L^{(0)}$ denotes radiance that is unscattered, $L^{(1)}$ is radiance from rays that have been scattered once, $L^{(2)}$ is radiance from rays that have been scattered twice, and so on. This is consistent with the interpretation of ω_o as the probability of ray survival in an interaction with matter, i.e., the probability that a ray will be scattered and not absorbed.

We now substitute Eq. (10.5) for the radiance into Eq. (10.4) to obtain

$$\begin{aligned} &\mu \left[\frac{dL^{(0)}}{d\zeta} + \omega_o \frac{dL^{(1)}}{d\zeta} + \omega_o^2 \frac{dL^{(2)}}{d\zeta} + \dots \right] \\ &= - [L^{(0)} + \omega_o L^{(1)} + \omega_o^2 L^{(2)} + \dots] \\ &+ \omega_o \int_0^{2\pi} \int_{-1}^1 [L^{(0)} + \omega_o L^{(1)} + \omega_o^2 L^{(2)} + \dots] \tilde{\beta}(\mu', \phi' \rightarrow \mu, \phi) d\mu' d\phi' . \end{aligned} \quad (10.6)$$

We next group terms that have the same power of ω_o :

$$\begin{aligned} &\left[\mu \frac{dL^{(0)}}{d\zeta} + L^{(0)} \right] \\ &+ \omega_o \left[\mu \frac{dL^{(1)}}{d\zeta} + L^{(1)} - \int_0^{2\pi} \int_{-1}^1 L^{(0)} \tilde{\beta}(\mu', \phi' \rightarrow \mu, \phi) d\mu' d\phi' \right] \\ &+ \omega_o^2 \left[\mu \frac{dL^{(2)}}{d\zeta} + L^{(2)} - \int_0^{2\pi} \int_{-1}^1 L^{(1)} \tilde{\beta}(\mu', \phi' \rightarrow \mu, \phi) d\mu' d\phi' \right] \\ &+ \dots = 0 . \end{aligned}$$

This equation must hold true for any value of $0 \leq \omega_o < 1$. Setting $\omega_o = 0$ would leave only the first line of the equation, whose terms must sum to 0. Similarly, when $\omega_o \neq 0$, each group of terms multiplying a given power of ω_o must equal zero in order for the entire left side of the equation to sum to zero. We can therefore equate to zero the groups of terms in brackets multiplying each power of ω_o . This gives a sequence of equations:

$$\mu \frac{dL^{(0)}}{d\zeta} = -L^{(0)} \quad (S0)$$

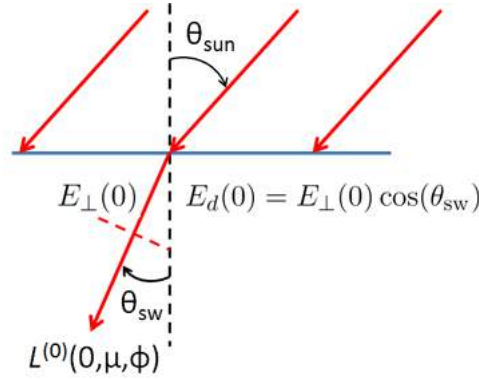


Figure 10.1: The collimated incident radiance onto the sea surface and the refracted radiance just below the surface.

$$\mu \frac{dL^{(1)}}{d\zeta} = -L^{(1)} + \int_0^{2\pi} \int_{-1}^1 L^{(0)} \tilde{\beta}(\mu', \phi' \rightarrow \mu, \phi) d\mu' d\phi' \quad (\text{S1})$$

$$\mu \frac{dL^{(2)}}{d\zeta} = -L^{(2)} + \int_0^{2\pi} \int_{-1}^1 L^{(1)} \tilde{\beta}(\mu', \phi' \rightarrow \mu, \phi) d\mu' d\phi' \quad (\text{S2})$$

and so on. Note that because ω_o multiplies the path integral term in Eq. (10.6), the path integrals in this sequence of equations always involve the radiance at one order of scattering less than the derivative term. We first solve Eq. (S0), which governs the unscattered radiance. The solution for $L^{(0)}$ then can be used in Eq. (S1) to evaluate the path integral, which becomes a source function for singly scattered radiance. After solving Eq. (S1) for singly-scattered radiance, $L^{(1)}$ can be used to evaluate the path function in Eq. (S2), and so on. *This process constitutes the successive-order-of-scattering (SOS) solution technique.*

10.2.1.1 Solution for the unscattered radiance

To solve Eq. (S0) we need boundary conditions at the sea surface and bottom. Figure 10.1 reminds us that the incident unscattered radiance onto the sea surface, and transmitted into the water, is perfectly collimated because we have assumed that the Sun is a point source in a black sky and the surface is level. In that figure, $E_{\perp}(0)$ denotes the irradiance measured just below the sea surface on a plane that is perpendicular to the direction of photon travel (denoted by the red dashed line), and θ_{sw} is the Sun's zenith angle in the water after refraction by the level surface.

Recalling the Dirac delta function 1.4.3, we can write the unscattered radiance just below the surface as

$$L^{(0)}(0, \mu, \phi) = E_{\perp}(0) \delta(\mu - \mu_{sw}) \delta(\phi - \phi_{sw}), \quad (\text{BC1})$$

where (μ_{sw}, ϕ_{sw}) is the direction of the Sun's beam in the water. The two delta functions, which together have units of sr^{-1} , “pick out” the direction of the Sun's beam; the unscattered radiance is zero in all other directions. Note that integrating this radiance over all downward directions to compute the downwelling plane irradiance gives

$$\begin{aligned} E_d(0) &= \int_0^1 \int_0^{2\pi} E_{\perp}(0) \delta(\mu - \mu_{sw}) \delta(\phi - \phi_{sw}) \mu d\mu d\phi \\ &= E_{\perp}(0) \mu_{sw} \end{aligned}$$

as expected.

It is assumed that the incident solar irradiance is given, so Eq. (BC1) is the boundary condition on $L^{(0)}(\zeta, \mu, \phi)$ at the sea surface (i.e., in the water at depth $\zeta = 0$). We are assuming that the water is infinitely deep and source free, so the radiance must approach 0 at great depth. The boundary condition at the bottom is thus

$$L^{(0)}(\zeta, \mu, \phi) \rightarrow 0 \quad \text{as } \zeta \rightarrow \infty. \quad (\text{BC2})$$

We can now solve Eq. (S0) subject to boundary conditions (BC1) and (BC2). Rewriting (S0) as

$$\frac{dL^{(0)}(\zeta)}{L^{(0)}(\zeta)} = -\frac{d\zeta}{\mu}$$

and integrating from depth 0 to ζ , corresponding to radiances $L^{(0)}(0)$ and $L^{(0)}(\zeta)$ respectively, gives

$$\ln L^{(0)} \Big|_{L^{(0)}(0)}^{L^{(0)}(\zeta)} = -\frac{\zeta}{\mu} \Big|_0^{\zeta}$$

or

$$L^{(0)}(\zeta, \mu, \phi) = L^{(0)}(0, \mu, \phi) e^{-\zeta/\mu} \quad (10.7)$$

$$= E_{\perp}(0) \delta(\mu - \mu_{\text{sw}}) \delta(\phi - \phi_{\text{sw}}) e^{-\zeta/\mu}. \quad (10.8)$$

Solution (10.7) is simply the Lambert-Beer law: the initial unscattered radiance decays exponentially with optical depth. Using (BC1) to rewrite the radiance at the surface gives (10.8), which will be the convenient form for solution of (S1) below. Equation (10.8) also shows explicitly that the unscattered radiance is 0 except in direction $(\mu_{\text{sw}}, \phi_{\text{sw}})$. The exponential forces the radiance to 0 as the depth increases, so that (BC2) is satisfied. Thus our solution satisfies both the surface and bottom boundary conditions and thus constitutes a complete solution of the two-point boundary value problem for unscattered radiance. This solution gives the contribution of unscattered radiance to the total radiance.

10.2.1.2 Solution for the singly-scattered radiance

The first step in solving Eq. (S1) for the singly-scattered radiance is to evaluate the scattering term using the solution for $L^{(0)}$. To do this we use (10.8) to get

$$\begin{aligned} & \int_0^{2\pi} \int_{-1}^1 L^{(0)}(\zeta, \mu', \phi') \tilde{\beta}(\mu', \phi' \rightarrow \mu, \phi) d\mu' d\phi' \\ &= \int_0^{2\pi} \int_{-1}^1 E_{\perp}(0) \delta(\mu' - \mu_{\text{sw}}) \delta(\phi' - \phi_{\text{sw}}) e^{-\zeta/\mu'} \tilde{\beta}(\mu', \phi' \rightarrow \mu, \phi) d\mu' d\phi' \\ &= E_{\perp}(0) e^{-\zeta/\mu_{\text{sw}}} \tilde{\beta}(\mu_{\text{sw}}, \phi_{\text{sw}} \rightarrow \mu, \phi). \end{aligned} \quad (10.9)$$

This result shows how much of the unscattered radiance reaches depth ζ and then gets scattered into the direction of interest (μ, ϕ) . In other words, *the unscattered radiance is a local (at depth ζ) source term for singly scattered radiance.*

All quantities on the right hand side of Eq. (10.9) are known from the given IOPs and surface boundary condition. We can therefore proceed with the solution of (S1) for the

singly scattered radiance $L^{(1)}$. The equation to be solved is

$$\mu \frac{dL^{(1)}}{d\zeta} + L^{(1)} = E_{\perp}(0) e^{-\zeta/\mu_{sw}} \tilde{\beta}(\mu_{sw}, \phi_{sw} \rightarrow \mu, \phi), \quad (10.10)$$

where the right hand side is now a known function of depth. There is no incident scattered radiance from the sky because the Sun's collimated beam is all unscattered light. Thus the boundary conditions for Eq. (10.10) are

$$L^{(1)}(0, \mu, \phi) = 0 \quad \text{and} \quad L^{(1)}(\zeta, \mu, \phi) \rightarrow 0 \quad \text{as} \quad \zeta \rightarrow \infty. \quad (10.11)$$

Figure 10.2 shows that the singly scattered downwelling radiance at depth ζ comes only from above depth ζ , and that the upwelling radiance at ζ comes only from depths below ζ . We can thus consider the downwelling, $L_d^{(1)}(\zeta, \mu, \phi)$, and upwelling, $L_u^{(1)}(\zeta, \mu, \phi)$, radiances separately. We can integrate from the surface down to ζ to compute $L_d^{(1)}$, and we can integrate from ζ to ∞ to compute $L_u^{(1)}$.

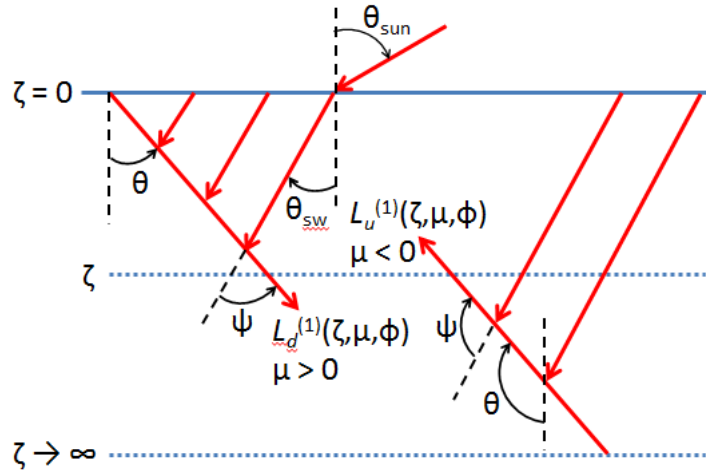


Figure 10.2: Single-scattering contributions to $L_d^{(1)}$ and $L_u^{(1)}$.

If you were paying attention in your undergraduate differential equations class, you recognize Eq. (10.10) as an ordinary differential equation with constant coefficients, which can be solved by means of an integrating factor. Multiplying Eq. (10.10) for downwelling radiance by $\frac{1}{\mu} e^{\zeta/\mu}$ (the integrating factor) gives

$$\begin{aligned} \frac{1}{\mu} e^{\zeta/\mu} \left[\mu \frac{dL_d^{(1)}(\zeta)}{d\zeta} + L_d^{(1)}(\zeta) \right] &= \frac{1}{\mu} e^{\zeta/\mu} \left[E_{\perp} \tilde{\beta} e^{-\zeta/\mu_{sw}} \right] \\ \frac{d}{d\zeta} \left[L_d^{(1)}(\zeta) e^{\zeta/\mu} \right] &= \frac{E_{\perp} \tilde{\beta}}{\mu} \exp \left[\left(\frac{1}{\mu} - \frac{1}{\mu_{sw}} \right) \zeta \right]. \end{aligned} \quad (10.12)$$

Now integrating from depth 0 to ζ , where the radiances are $L_d^{(1)}(0)$ and $L_d^{(1)}(\zeta)$, respectively, and recalling that $L_d^{(1)}(0) = 0$ by the upper boundary condition (10.11) gives

$$L_d^{(1)}(\zeta) e^{\zeta/\mu} = \frac{E_{\perp} \tilde{\beta}}{\mu} \frac{1}{\left(\frac{1}{\mu} - \frac{1}{\mu_{sw}} \right)} \left[\exp \left(\frac{1}{\mu} - \frac{1}{\mu_{sw}} \right) \zeta - 1 \right],$$

provided that $\mu \neq \mu_{\text{sw}}$. Recalling that $E_d(0) = E_{\perp}(0) \mu_{\text{sw}}$, the preceding equation can be rewritten as

$$L_d^{(1)}(\zeta, \mu, \phi) = E_d(0) \tilde{\beta}(\mu_{\text{sw}}, \phi_{\text{sw}} \rightarrow \mu, \phi) \frac{1}{\mu_{\text{sw}} - \mu} \left[e^{-\zeta/\mu_{\text{sw}}} - e^{-\zeta/\mu} \right]. \quad (10.13)$$

For the special case of $\mu = \mu_{\text{sw}}$ but $\phi \neq \phi_{\text{sw}}$, so that the scattering angle is nonzero, Eq. (10.12) reduces to

$$\frac{d}{d\zeta} \left[L_d^{(1)}(\zeta) e^{\zeta/\mu_{\text{sw}}} \right] = \frac{E_{\perp} \tilde{\beta}}{\mu_{\text{sw}}}$$

which integrates to

$$\begin{aligned} L_d^{(1)}(\zeta, \mu_{\text{sw}}, \phi) &= E_{\perp} \tilde{\beta}(\mu_{\text{sw}}, \phi_{\text{sw}} \rightarrow \mu_{\text{sw}}, \phi) \frac{\zeta}{\mu_{\text{sw}}} e^{-\zeta/\mu_{\text{sw}}} \\ &= E_d(0) \tilde{\beta}(\mu_{\text{sw}}, \phi_{\text{sw}} \rightarrow \mu_{\text{sw}}, \phi) \frac{\zeta}{\mu_{\text{sw}}^2} e^{-\zeta/\mu_{\text{sw}}}. \end{aligned} \quad (10.14)$$

The second form results from $E_d(0) = E_{\perp}(0) \mu_{\text{sw}}$, which was derived above.

The direction of $\mu = \mu_{\text{sw}}$ and $\phi = \phi_{\text{sw}}$ is the case of no scattering, so there is no singly scattered radiance.

We next compute the upwelling radiance at ζ by integrating Eq. (10.10) from ζ to ∞ , keeping in mind that now $\mu = \cos \theta < 0$ since θ is measured from 0 in the nadir direction. The integration gives (writing $\mu = -|\mu|$ to emphasize the negativity of μ)

$$\begin{aligned} &\left[L_u^{(1)}(\zeta') e^{-\zeta'/|\mu|} \right]_{\zeta' \rightarrow \infty} - L_u^{(1)}(\zeta) e^{-\zeta/|\mu|} = \\ &\frac{E_{\perp} \tilde{\beta}}{-|\mu|} \frac{1}{\left(\frac{1}{-|\mu|} - \frac{1}{\mu_{\text{sw}}} \right)} \left\{ \left[\exp \left(\frac{1}{-|\mu|} - \frac{1}{\mu_{\text{sw}}} \right) \zeta' \right]_{\zeta' \rightarrow \infty} - \exp \left(\frac{1}{-|\mu|} - \frac{1}{\mu_{\text{sw}}} \right) \zeta \right\} \end{aligned}$$

Both limits as $\zeta' \rightarrow \infty$ are zero. The result can be rewritten as

$$L_u^{(1)}(\zeta) = E_d(0) \tilde{\beta}(\mu_{\text{sw}}, \phi_{\text{sw}} \rightarrow \mu, \phi) \frac{1}{\mu_{\text{sw}} - \mu} e^{-\zeta/\mu_{\text{sw}}}. \quad (10.15)$$

10.2.2 Assembling the SSA solution

Recalling from Eq. (10.5) that the SSA is given by

$$L^{(\text{SSA})}(\zeta, \mu, \phi) = L^{(0)}(\zeta, \mu, \phi) + \omega_o L^{(1)}(\zeta, \mu, \phi),$$

we can assemble $L^{(\text{SSA})}$ from the pieces computed in Eqs. (10.7) and (10.13-10.15):

$$\begin{aligned} L_d^{(\text{SSA})}(\zeta, \mu, \phi) &= L^{(0)}(0, \mu_{\text{sw}}, \phi_{\text{sw}}) e^{-\zeta/\mu_{\text{sw}}} \\ &\text{if } \mu = \mu_{\text{sw}} \text{ and } \phi = \phi_{\text{sw}} \end{aligned} \quad (10.16)$$

$$\begin{aligned} L_d^{(\text{SSA})}(\zeta, \mu, \phi) &= \omega_o E_d(0) \tilde{\beta}(\mu_{\text{sw}}, \phi_{\text{sw}} \rightarrow \mu_{\text{sw}}, \phi) \frac{\zeta}{\mu_{\text{sw}}^2} e^{-\zeta/\mu_{\text{sw}}} \\ &\text{if } \mu = \mu_{\text{sw}} \text{ but } \phi \neq \phi_{\text{sw}} \end{aligned} \quad (10.17)$$

$$L_d^{(\text{SSA})}(\zeta, \mu, \phi) = \omega_o E_d(0) \tilde{\beta}(\mu_{\text{sw}}, \phi_{\text{sw}} \rightarrow \mu, \phi) \frac{1}{\mu_{\text{sw}} - \mu} \left[e^{-\zeta/\mu_{\text{sw}}} - e^{-\zeta/\mu} \right] \quad (10.18)$$

$$\begin{aligned}
& \text{if } \mu > 0 \text{ and } \mu \neq \mu_{\text{sw}} \\
L_u^{(\text{SSA})}(\zeta, \mu, \phi) &= \omega_o E_d(0) \tilde{\beta}(\mu_{\text{sw}}, \phi_{\text{sw}} \rightarrow \mu, \phi) \frac{1}{\mu_{\text{sw}} - \mu} e^{-\zeta/\mu_{\text{sw}}} \\
& \text{if } \mu \leq 0
\end{aligned} \tag{10.19}$$

Equations (10.16)-(10.19) constitute the SSA solution to the RTE. This solution is seen, for example, in Gordon (1994a), where it is presented without derivation.

It is easy to show that

$$\lim_{\mu \rightarrow \mu_{\text{sw}}} \frac{1}{\mu_{\text{sw}} - \mu} \left[e^{-\zeta/\mu_{\text{sw}}} - e^{-\zeta/\mu} \right] = \frac{\zeta}{\mu_{\text{sw}}^2} e^{-\zeta/\mu_{\text{sw}}}, \tag{10.20}$$

in which case Eq. (10.18) reduces to Eq. (10.17), which was derived independently as a special case of the depth integration.

It must be remembered that the SSA rests upon a number of simplifying assumptions. In particular, the input sky radiance was collimated. The delta functions in direction then made evaluation of the scattering path function in Eq. (10.9) easy. This would not be the case for any other sky radiance distribution, or for a non-level sea surface. Likewise, the assumption of infinitely deep water removed any bottom effect.

The SSA will be a good approximation to actual radiances only if the higher order terms in the Eq. (10.5) are negligible. This means that ω_o must be sufficiently small, but how small? Figure 10.3 compares $L_u^{(\text{SSA})}$ and $L_d^{(\text{SSA})}$ with radiances computed by HydroLight for nadir- and zenith-viewing radiances. The Sun was at 42 deg, which gives the in-water solar zenith angle of $\theta_{\text{sw}} = 30$ deg or $\mu_{\text{sw}} = 0.866$. This gives a scattering angle of $\psi = 30$ deg for $L_d^{(\text{SSA})}$ and $\psi = 150$ deg for $L_u^{(\text{SSA})}$. The Petzold “average-particle” phase function 9.5.2 was used, for which $\tilde{\beta}(\psi = 30) = 0.08609$ and $\tilde{\beta}(\psi = 150) = 0.002365 \text{ sr}^{-1}$. HydroLight includes all orders of multiple scattering, so comparison of its radiances with the SSA values shows the importance of multiple scattering. The HydroLight runs modeled the SSA conditions as closely as possible, the difference being that the SSA is for one exact direction and HydroLight computes nadir and zenith radiances as averages over polar caps with a 5 deg half angle, and the Sun’s direct beam in water is spread out over a quad from $\theta = 25$ to 35 deg. The HydroLight runs set $E_d(\text{in air}) = 1.028 \text{ W m}^{-2} \text{ sr}^{-1}$ so that $E_d(0) = 1.0 \text{ W m}^{-2} \text{ sr}^{-1}$.

Figure 10.3 shows that for $\omega_o = 0.01$ the agreement between the SSA and HydroLight is very good. The HydroLight values are slightly higher than the SSA values because there is still a small multiple scattering contribution to the total radiance even at very small ω_o values. For $\omega_o = 0.1$ the SSA still gives good results near the sea surface but differs from the multiple scattering solution by a factor of 3 at 10 optical depths. For $\omega_o = 0.85$, which is typical of blue and green wavelengths in ocean waters, the SSA upwelling radiance is a factor of five too small even at the surface, and the SSA radiances are off by orders of magnitude at large optical depths. Thus, as expected, we see that the SSA is of little use in optical oceanography because multiple scattering almost always dominates underwater radiance distributions at visible wavelengths.

We end the SSA discussion by noting that Walker (1994) has carried the SOS solution through second order scattering (i.e., an analytical solution of Eq. (S2) above). His development requires a good bit of mathematical masochism and results in a much more complicated set of equations, which can be seen in his Section 2-6. There is little need

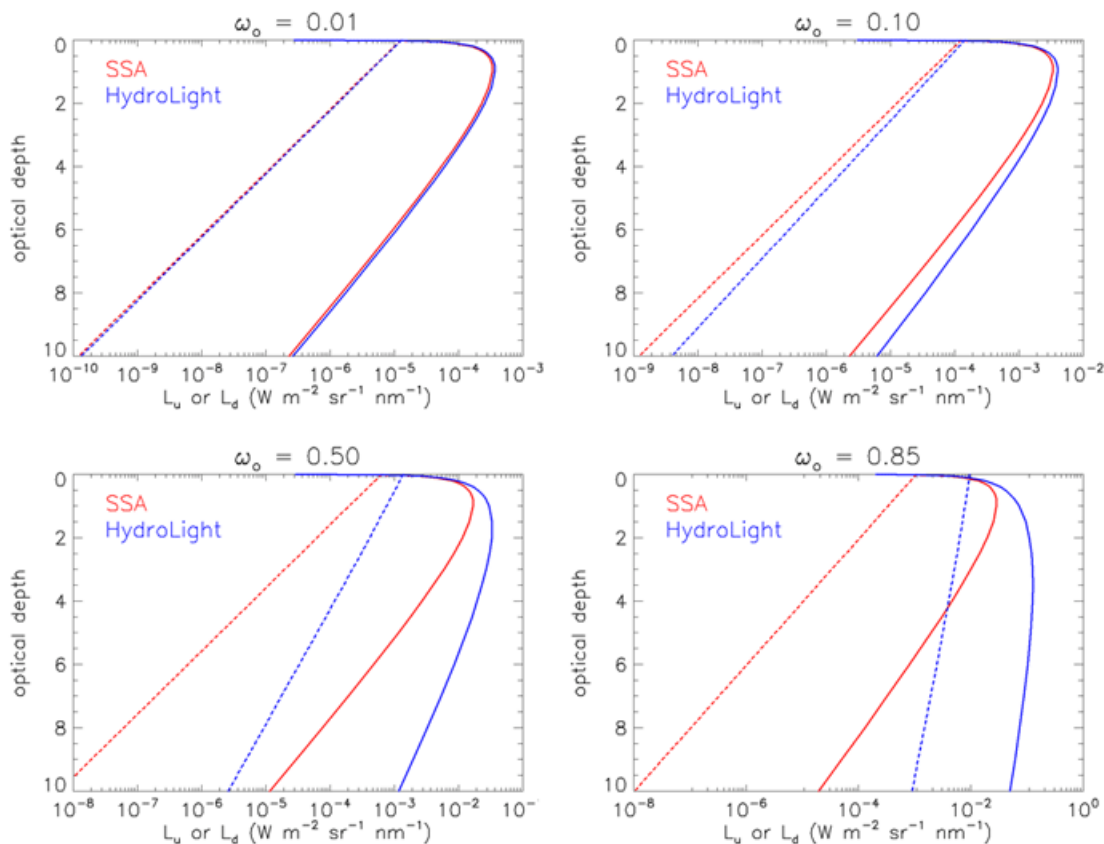


Figure 10.3: Comparison of the SSA and HydroLight solutions for nadir-viewing (L_u ; dashed lines) and zenith-viewing (L_d ; solid lines) radiances for four ω_o values; $E_d(0) = 1 \text{ W m}^{-2} \text{ sr}^{-1}$.

for such approximations given the ease of numerical solution of the RTE to include all (in HydroLight) or at least many (in Monte Carlo models) orders of multiple scattering, and without any of the assumptions required for the analytic evaluation of the path integrals in the SSA. Perhaps the greatest value of the SSA solution is that it can be used to check numerical models when ω_o is small.

10.3 The Quasi-Single-Scattering Approximation (QSSA)

For highly peaked phase functions such as those typical of ocean waters, most of the scattering is at very small scattering angles ψ . For some purposes, scattering through a small scattering angle is almost the same as no scattering at all. The quasi-single-scattering approximation (QSSA) exploits this observation by assuming that the forward-scattering part of the phase function can be represented by a Dirac delta function at scattering angle $\psi = 0$, with no scattering at all for $0 < \psi < 90$ deg. At first glance this seems like a terribly inaccurate approximation of reality for phase functions like the Petzold average-particle phase function shown in the left panel of Fig. 10.4. However, when plotted on linear axes as in the right panel of the figure, the approximation looks more reasonable. In practice,

it can yield surprisingly accurate results for quantities that depend mostly on absorption and/or backscatter (such as K_d and R_{rs}).

The QSSA can be traced back to Hansen (1971), who used it in studies of reflection by planetary atmospheres. Gordon (1973) introduced it to oceanography for ocean color remote sensing of the ocean.

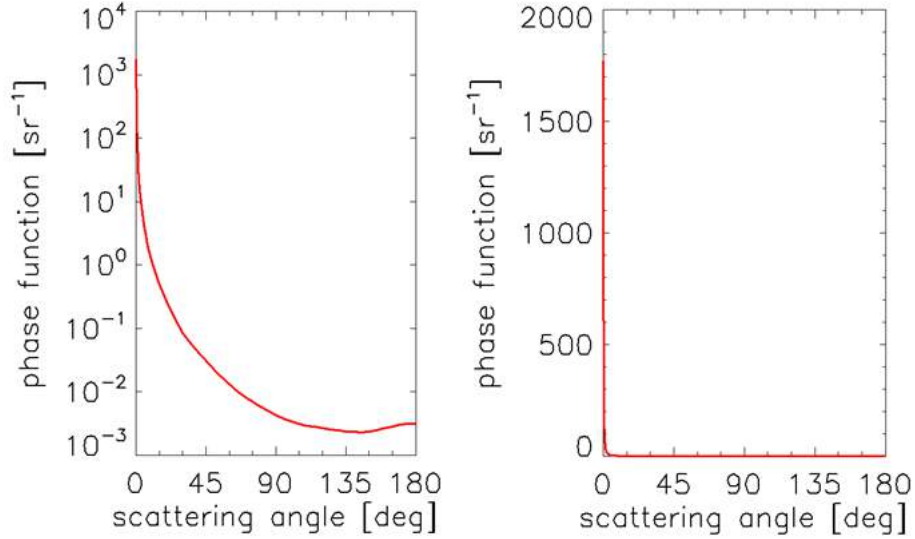


Figure 10.4: The Petzold average particle phase function plotted on log-linear and linear-linear axes.

The QSSA uses the formulas of the SSA, but treats forward scattering as no scattering at all. With this approximation, the beam attenuation coefficient c becomes

$$c = a + b = a + b_f + b_b \approx a + b_b \triangleq c^* .$$

With this approximation for c , the single scattering albedo ω_o and the optical depth ζ become

$$\omega_o = \frac{b}{c} \approx \frac{b}{a + b_b} = \frac{b_b}{a + b_b} \frac{1}{B} \triangleq \omega_o^* ,$$

and

$$\zeta = cz \approx (a + b_b)z \triangleq \zeta^* .$$

where B is the backscatter fraction. The QSSA thus replaces c , ω_o , and ζ with c^* , ω_o^* , and ζ^* respectively. This is an example of a *similarity transformation* in which the solution of one problem is rescaled to obtain the solution to a different problem.

The QSSA was developed for reflectance calculations, so let us compute the QSSA approximation for the remote-sensing reflectance R_{rs} . The SSA solution for $L_u^{(SSA)}$ (Eq. 10.19),

$$L_u^{(SSA)}(\zeta, \mu, \phi) = \omega_o E_d(0) \tilde{\beta}(\mu_{sw}, \phi_{sw} \rightarrow \mu, \phi) \frac{1}{\mu_{sw} - \mu} e^{-\zeta/\mu_{sw}} , \quad (10.21)$$

with the approximations of the QSSA becomes

$$\frac{L_u^{(\text{QSSA})}(\zeta^*, \mu, \phi)}{E_d(0)} = \frac{b_b}{a + b_b} \frac{\tilde{\beta}(\mu_{\text{sw}}, \phi_{\text{sw}} \rightarrow \mu, \phi)}{B} \frac{1}{\mu_{\text{sw}} - \mu} e^{-\zeta^*/\mu_{\text{sw}}}$$

When evaluated just below the sea surface at $\zeta^* = 0$, this quantity is related to R_{rs} by

$$R_{\text{rs}} = \frac{t}{n^2} \frac{E_d(0)}{E_d(\text{in air})} \frac{L_u^{(\text{QSSA})}(0, \mu, \phi)}{E_d(0)}. \quad (10.22)$$

Here t is the transmittance of radiance from water to air. For nadir-viewing radiance, $\mu = -1$ and $t \approx 0.98$. $n \approx 1.34$ is the index of refraction of water, and the ratio of irradiances (close to 1 for solar zenith angles away from the horizon) converts the underwater irradiance to the above-water value used in the definition of R_{rs} . Thus for nadir-viewing R_{rs} we have

$$R_{\text{rs}} = \frac{b_b}{a + b_b} \frac{\tilde{\beta}(\mu_{\text{sw}}, \phi_{\text{sw}} \rightarrow \mu = -1, \phi = 0)}{B} \frac{1}{\mu_{\text{sw}} + 1}, \quad (10.23)$$

The factor of $\tilde{\beta}/B$ is determined by the shape of the total phase function, which in turn is determined by the type of particles in the ocean. The μ_{sw} factor is determined by solar angle. The remaining factor,

$$G = \frac{b_b}{a + b_b},$$

shows that, to first order, R_{rs} depends on the IOPs via $b_b/(a + b_b)$, where both a and b_b are functions of depth and wavelength. G is sometimes called the Gordon parameter in recognition of his use of this quantity in numerous ocean remote-sensing studies.

Figure 10.5 compares the accuracy of the QSSA as a function of ω_o with the SSA and with HydroLight computations that include all orders of multiple scattering. The HydroLight run used the Petzold phase function of Fig. 10.4, for which the backscatter fraction is $B = 0.0183$. The Sun was placed at 30 deg in a black sky, and the sea surface was level. This figure shows that the SSA does well only for $\omega_o < 0.3$. The SSA is almost a factor of four too small at $\omega_o = 0.7$, whereas the QSSA is only 16% too small. Even at $\omega_o = 0.85$ the QSSA is only 35% too small, whereas the SSA is only one tenth of the correct value.

The blue curve of Fig. 10.6 shows an R_{rs} spectrum computed by HydroLight for homogeneous Case 1 water with a chlorophyll concentration of 2 mg m^{-3} . The phytoplankton were modeled with a Petzold average particle phase function as used above. The Sun was at 30 deg in a clear sky, and the wind was 5 m s^{-1} . Those sky and surface conditions violate the assumptions of collimated incident radiance and a level sea surface that underlie the SSA and QSSA. Nevertheless, the QSSA, evaluated using the IOPs generated by the HydroLight bio-optical model, gives an amazingly close prediction of the exact spectrum even though ω_o was in the range of 0.80 to 0.86 between 350 and 575 nm. The QSSA of course does not capture the chlorophyll fluorescence peak near 685 nm. The SSA fails badly until near 750 nm, where ω_o is less than 0.15.

The excellent performance of the QSSA seems counterintuitive given the crudeness of its phase function approximation. Why, in particular, does it do better than the SSA, since the SSA uses the full phase function, which is certainly a better description of nature than a delta function for forward scattering? The answer lies in how backscattering is

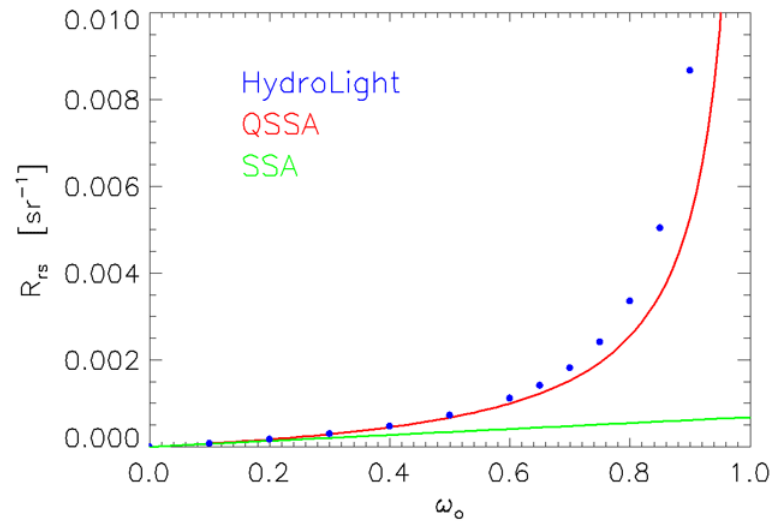


Figure 10.5: Predictions of R_{rs} for the SSA, QSSA, and HydroLight as a function of ω_o .

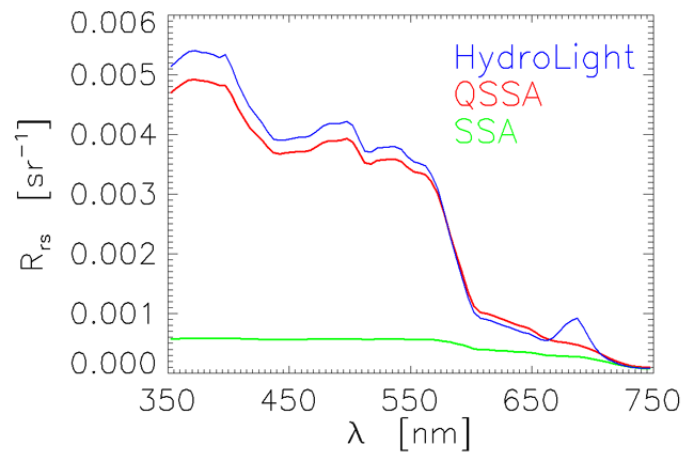


Figure 10.6: Predictions of R_{rs} for the SSA, QSSA, and HydroLight as a function of wavelength for Case 1 water with a chlorophyll concentration of 2 mg m^{-3} .

parameterized in the QSSA. Note from Eq. (10.21) that in the SSA the backscatter part of the phase function is weighted by (multiplied by) ω_o . For a typical water body with a scattering coefficient four times the absorption coefficient, $b = 4a$, and a Petzold phase function with a backscatter fraction of $B = 0.018$, this gives

$$\omega_o \tilde{\beta} = \frac{b}{a+b} \tilde{\beta} = 0.80 \tilde{\beta}.$$

Equation (10.23) shows that the QSSA for the same water weights $\tilde{\beta}$ by

$$\frac{b_b}{a+b_b} \frac{1}{B} \tilde{\beta} = \frac{bB}{a+bB} \frac{1}{B} \tilde{\beta} = 3.73 \tilde{\beta}.$$

Thus, for these IOPs, the QSSA weights the backscatter part of the phase function by a factor 4.66 times greater than that of the SSA. This corresponds precisely to the increase in R_{rs} for the QSSA compared to the SSA as seen in Fig. 10.5 for $\omega_o = 0.8$. In effect, the QSSA accounts for the “missing” multiple scattering in the single-scattering formulation by increasing the amount of backscattering. In other words, the QSSA parameterizes multiple scattering within the single-scattering mathematical framework by artificially increasing the backscattering.

The main utility of the SSA and QSSA is in the insight they give to the dependence of various AOPs on the IOPs. We saw above that R_{rs} depends primarily on $b_b/(a+b_b)$. Likewise, the SSA and QSSA both show that near the sea surface K_d depends primarily on $a+b_b$. Gordon (1994a) shows comparisons between the SSA, QSSA and numerical (Monte Carlo) computations of other quantities such as the near-surface irradiance reflectance R and K_d/c .

Two more comments must be made for completeness. As Gordon noted in his original paper, the desired partitioning of scattered radiance is into upward versus downward components. His development and the one above truncated the phase function out to a scattering of $\psi = 90$ deg. This separation of backward and forward scattering corresponds to radiance scattered into upwelling and downwelling directions, respectively, only if the Sun is at the zenith. If the Sun is not at the zenith, then the phase function used in the QSSA should include only the angles that contribute to upward scattering, in which case B in Eq. (10.23) is not exactly the backscatter fraction of the original phase function. However, as we have seen in the example calculations with the Sun at a 30 deg zenith angle, the QSSA still works well for non-zenith solar angles even when evaluated with B equal to the phase function backscatter fraction. Finally, note that in the limit of no scattering, i.e., $\omega_o \rightarrow 0$, the QSSA reduces to the SSA.

10.4 The Asymptotic Radiance Distribution

Deep in *homogeneous, source-free* waters the radiance distribution $L(z, \theta, \phi)$ approaches a *shape* $L_\infty(\theta)$ that depends only on the IOPs. Moreover, the radiance distribution at great depth decays in magnitude *exactly exponentially* with a decay rate K_∞ that, once again, depends only on the IOPs. The shape $L_\infty(\theta)$ is called the *asymptotic radiance distribution*, and K_∞ is called the *asymptotic decay rate* or the *asymptotic K function*. $L_\infty(\theta)$ and K_∞ depend on the wavelength λ via the wavelength dependence of the IOPs; the wavelength is omitted here for brevity. An asymptotic radiance distribution exists only

if the IOPs do not depend on depth (homogeneous water) and if there is no inelastic scatter or bioluminescence contributing to the radiance (source-free water). Preisendorfer (1976, Vol. 5, page 212) and Højerslev and Zaneveld (1977) give rigorous mathematical proofs that $L_\infty(\theta)$ and K_∞ exist for any physically realistic phase function and single-scattering albedo.

These statements imply that *the directional and depth dependencies of the radiance distribution decouple at great depths*. That is to say,

$$L(z, \theta, \phi) \xrightarrow{z \rightarrow \infty} L_\infty(\theta) \exp(-K_\infty z) . \quad (10.24)$$

This in turn implies that *all irradiances decay in the asymptotic regime at the same rate as the radiance*. For example,

$$\begin{aligned} \lim_{z \rightarrow \infty} E_d(z) &= \int_0^{2\pi} \int_0^{\pi/2} L_\infty(\theta) \exp(-K_\infty z) \cos \theta \sin \theta d\theta d\phi \\ &= \left[2\pi \int_0^{\pi/2} L_\infty(\theta) \cos \theta \sin \theta d\theta \right] \exp(-K_\infty z) \\ &\triangleq E_d(\infty) \exp(-K_\infty z) . \end{aligned}$$

We can compute corresponding values for E_u , E_{od} , and E_{ou} . Clearly, each of these irradiances has the same asymptotic K function. Using these asymptotic irradiances, we can compute asymptotic values for any apparent optical property. For example, we have

$$R_\infty \triangleq \frac{E_u(\infty)}{E_d(\infty)} .$$

Note that any normalization factor in $L_\infty(\theta)$ divides out when computing AOPs.

Because the asymptotic radiance $L_\infty(\theta)$ is determined solely by the IOPs, it follows that any quantity computed from $L_\infty(\theta)$ is also in IOP. Therefore *all apparent optical properties become inherent optical properties in the asymptotic regime*. The K 's, $\bar{\mu}$'s, R 's, and their ilk, which are influenced by boundary conditions near the water surface, all approach values at depth that are independent of the boundary conditions. This was seen in Figs. 4.3- 4.6, 4.16, and 4.21.

10.4.1 An Integral Equation for the Asymptotic Radiance Distribution

The obvious question is, “How do you compute the asymptotic radiance distribution and the asymptotic decay rate, given the IOPs?” One way is to recall the SRTE for homogeneous, source-free water

$$\begin{aligned} \cos \theta \frac{dL(z, \theta, \phi)}{dz} &= -c L(z, \theta, \phi) \\ &+ \int_0^{2\pi} \int_0^\pi L(z, \theta', \phi') \beta(\theta', \phi' \rightarrow \theta, \phi) \sin \theta' d\theta' d\phi' , \end{aligned}$$

and assume that the radiance has the form seen in Eq. (10.24). This gives an *integral equation for the shape $L_\infty(\theta)$ and decay rate K_∞ of the asymptotic radiance distribution*:

$$(c - K_\infty \cos \theta) L_\infty(\theta) = \int_0^{2\pi} \int_0^\pi L_\infty(\theta') \beta(\theta', \phi' \rightarrow \theta, \phi) \sin \theta' d\theta' d\phi' . \quad (10.25)$$

(Note that since the scattering angle ψ depends only on $\cos(\phi - \phi')$, we can set $\phi = 0$ in this equation, so that the result is a function only of θ .) This equation is often written in terms of $\mu = \cos \theta$ and a nondimensional asymptotic decay rate $\kappa_\infty = K_\infty/c$ as

$$(1 - \kappa_\infty \mu) L_\infty(\mu) = \omega_o \int_0^{2\pi} \int_{-1}^1 L_\infty(\mu') \tilde{\beta}(\mu', \phi' \rightarrow \mu, \phi) d\mu' d\phi'. \quad (10.26)$$

Given the IOPs c and β for Eq. (10.25) or ω_o and $\tilde{\beta}$ for Eq. (10.26), we can solve either of these equations for the corresponding L_∞ and K_∞ or κ_∞ . Another form of Eq. (10.26), often seen in the literature, is

$$(1 - \kappa_\infty \mu) L_\infty(\mu) = 2\pi\omega_o \int_{-1}^1 L_\infty(\mu') h(\mu', \mu) d\mu',$$

where $h(\mu', \mu)$ is the azimuthally averaged phase function

$$h(\mu', \mu) \triangleq \frac{1}{2\pi} \int_0^{2\pi} \tilde{\beta}(\mu', \phi' \rightarrow \mu, \phi) d\phi'.$$

For the idealized case of *isotropic scattering*, $\tilde{\beta} = 1/4\pi$ and the solution of Eq. (10.26) has the simple form

$$L_\infty(\mu) = \frac{1 - \kappa_\infty}{1 - \kappa_\infty \mu}, \quad (10.27)$$

where $L_\infty(\mu)$ is normalized to 1 at $\mu = 1$ (or at the nadir direction $\theta = 0$). This $L_\infty(\mu)$ has the shape of an ellipse whose major axis is oriented vertically. The corresponding value of κ_∞ is the solution of the transcendental equation

$$1 = \frac{\omega_o}{2\kappa_\infty} \ln \left(\frac{1 + \kappa_\infty}{1 - \kappa_\infty} \right),$$

as can be seen by substitution of Eq. (10.27) into Eq. (10.26).

[Kattawar and Plass \(1976\)](#) obtained an analytic solution of Eq. (10.26) for the Rayleigh phase function $\tilde{\beta} = (3/16\pi)(1 + \cos^2 \psi)$. However, for other phase functions, in particular for those characteristic of oceanic waters, the solution of Eq. (10.25) or (10.26) must be obtained numerically.

Solving the integral Eq. (10.25) is mathematically equivalent to solving a certain eigenmatrix equation² for its eigenfunctions (which give L_∞) and eigenvalues (which give K_∞). The eigenmatrix approach is described in *Light and Water* ([Mobley, 1994](#), Section 9.6).

The asymptotic radiance distribution including polarization has been studied by [Kattawar and Plass \(1976\)](#) for selected phase matrices and by [Sun et al. \(2016\)](#) for arbitrary phase matrices. In the asymptotic regime the Stokes parameters (I, Q, U, V) (Section 1.6) are independent of the azimuthal angle and U and V are zero. The degree of polarization Q/I in the asymptotic regime depends on the phase matrix and on the albedo of single scattering, ω_o . As $\omega_o \rightarrow 1$ the radiance in the asymptotic regime becomes unpolarized. However, as ω_o decreases, the asymptotic radiance becomes increasingly polarized and can approach that for singly scattered radiance for the given phase matrix.

²HydroLight uses the eigenmatrix approach to obtain its asymptotic values. The excellent agreement between the asymptotic values computed by the eigenmatrix method and the approach with depth to those values, as obtained by independent numerical solution of the SRTE, is an excellent check on the correctness of the HydroLight code.

10.4.2 Dependence of Asymptotic Values on Inherent Optical Properties

As just seen, the two asymptotic properties L_∞ and K_∞ are determined solely by the IOPs c , ω_o and $\tilde{\beta}$. Figure 10.7 shows how the nondimensional asymptotic decay rate $\kappa_\infty = K_\infty/c$ depends on the albedo of single scattering ω_o for three phase functions. The dotted line is for the pure water phase function $\tilde{\beta}_w$; the dashed line is for a Henyey-Greenstein phase function $\tilde{\beta}_{HG}$ with an asymmetry parameter $g = 0.7$; and the solid line is for a Petzold “average-particle” phase function $\tilde{\beta}_p$, which is typical of phase functions for oceanic particles. The squares show experimental data taken in laboratory suspensions containing milk (Timofeeva and Gorobetz, 1967). The fat globules in milk are large ($\gg \lambda$), efficient scatterers, which explains the similarity between the milk solution and the $\tilde{\beta}_p$ phase function, which is typical of particle-laden natural waters.

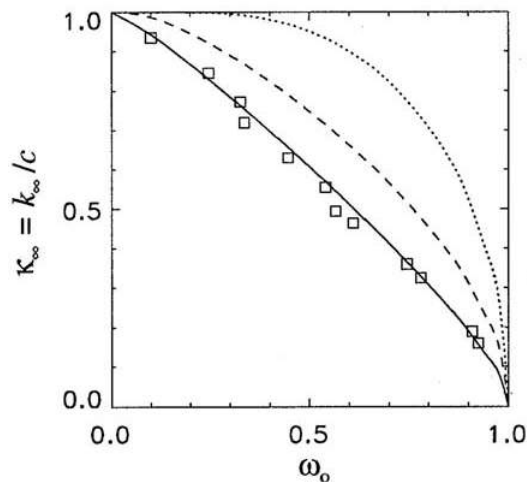


Figure 10.7: Dependence of κ_∞ on ω_o for selected phase functions. The solid line is for $\tilde{\beta}_p$, the dashed line is for $\tilde{\beta}_{HG}$, and the dotted line is for $\tilde{\beta}_w$. The squares are the data of Timofeeva and Gorobetz (1967).

Figure 10.8 shows the shape of $L_\infty(\theta)$ as a function of ω_o for the average-particle phase function $\tilde{\beta}_p$. Since it is the shape that is determined by the IOPs, it is customary to normalize $L_\infty(\theta)$ to one for the nadir radiance direction (looking upward in the zenith direction). The viewing angle θ_v as plotted is the angle in which an underwater observer would look in order to see radiance traveling in direction $\theta = 180^\circ - \theta_v$; θ_v and θ are both measured from the $+\hat{z}$, or nadir, direction. Thus, $\theta_v = 180^\circ$ corresponds to looking toward the zenith and seeing radiance heading straight down ($\theta = 0$). As we would expect, in highly scattering water (large ω_o) the upwelling radiance is relatively much greater than in weakly scattering water (small ω_o). Corresponding curves for the Rayleigh phase function can be seen in Kattawar and Plass (1976). Priour and Morel (1971) show such curves as a function of the relative contributions by molecular and particle scattering, i.e., for phase functions that are in between $\tilde{\beta}_w$ and $\tilde{\beta}_p$.

Figure 10.9 shows the asymptotic mean cosines and irradiance reflectance for the same phase functions used in Fig. 10.7.

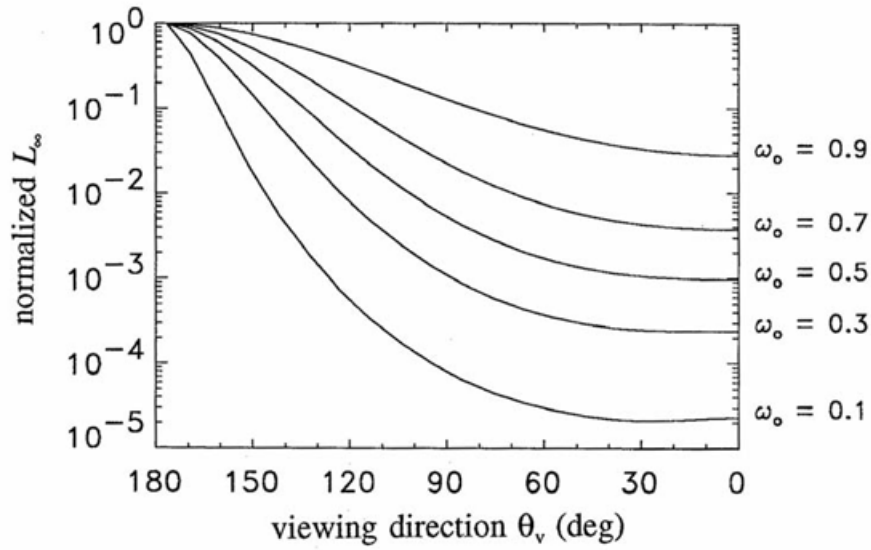


Figure 10.8: Shape of the asymptotic radiance distribution $L_\infty(\theta)$ as a function of ω_o , for the particle phase function $\tilde{\beta}_p$. The viewing angle θ_v is $180^\circ - \theta$, as discussed in the text.

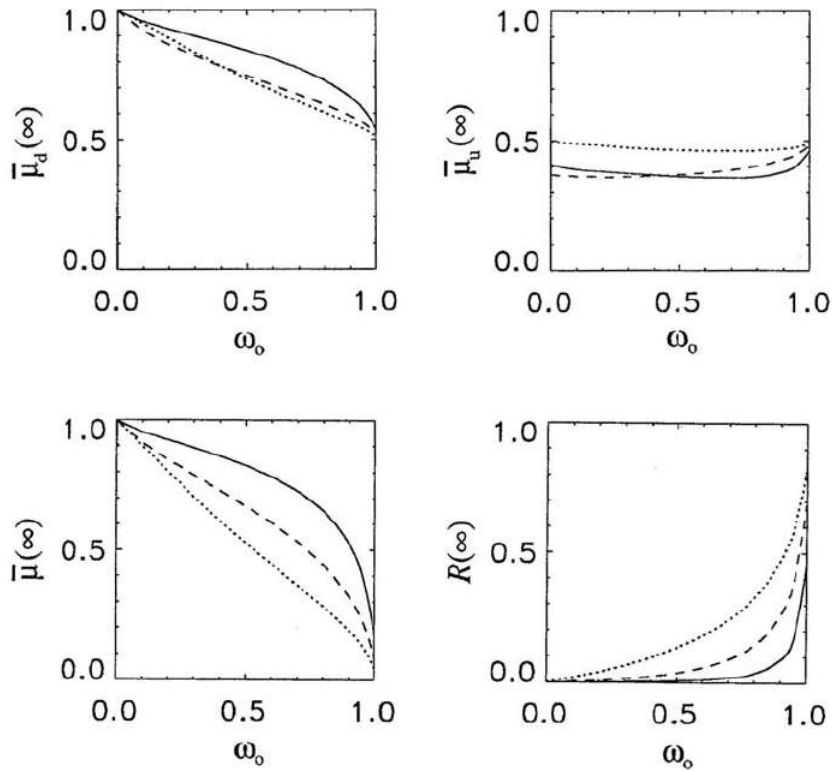


Figure 10.9: Asymptotic values of the mean cosines and of the irradiance reflectance, as a function of ω_o , for the Petzold average-particle phase function. The solid, dashed, and dotted lines correspond to those of Fig. 10.7.

10.4.3 Rate of Approach to Asymptotic Values

The asymptotic values are determined solely by the IOPs of a homogeneous water body. However, *how quickly* a given quantity approaches its asymptotic value depends on both the IOPs and the boundary conditions. We have already seen this in the discussion of K functions 4.1, but a few more examples will be instructive. HydroLight was run for Case 1 water with a chlorophyll concentration of $Chl = 1.0 \text{ mg m}^{-3}$. The wavelength was 443 nm. The corresponding IOPs (including water) were $a = 0.0680 \text{ m}^{-1}$, $b = 0.4346 \text{ m}^{-1}$, so that $c = 0.5026 \text{ m}^{-1}$. Thus one meter of geometric depth is about 0.5 optical depths. The albedo of single scattering is then $\omega_o = b/c = 0.8648$, and the total backscatter fraction was 0.1253. Figure 10.10 shows K_d at 443 nm for four different surface boundary conditions:

- the Sun is at the zenith in a clear sky ($\theta_{\text{sun}} = 0$) and the sea surface is level (wind speed $U = 0$)
- the Sun is at a 50 deg zenith angle in a clear sky ($\theta_{\text{sun}} = 50$) and the sea surface is level ($U = 0$)
- the Sun is at a 50 deg zenith angle in a clear sky ($\theta_{\text{sun}} = 50$) and the sea surface is wind blown, with a wind speed of 10 m/s ($U = 10$)
- the sky has a cardioidal radiance distribution, $L(\theta, \phi) = L_o(1 + 2 \cos \theta)$, $0 \leq \theta \leq \pi/2$, which is similar to a heavy overcast through which the location of the Sun cannot be determined, and the sea surface is wind blown, with a wind speed of 10 m/s ($U = 10$)

The curves of Fig. 10.10 show the boundary effects on the rate of approach to the asymptotic value of $K_\infty = 0.1075 \text{ m}^{-1}$, which depends only on the IOPs. The curve for the heavily overcast sky approaches K_∞ the quickest. The physical reason is that the overcast sky radiance is already a diffuse radiance distribution, so that less scattering (i.e., less propagation to depth) is required to redirect the initial photon directions towards the asymptotic angular distribution within the water. The other cases with the Sun in a clear sky have a strongly collimated incident radiance distribution, which requires more scattering (a deeper depth) to “erase the memory” of where the Sun is in the sky and achieve the asymptotic shape of $L_\infty(\theta)$. For the 50 deg Sun zenith angle, the surface roughness makes a noticeable but minor difference.

Figure 10.11 shows the corresponding results for the approach of the irradiance reflectance $R = E_u/E_d$ to its asymptotic value of $R_\infty = 0.0406$.

Figure 10.12 shows the average cosines $\bar{\mu}_d$, $\bar{\mu}_u$, and $\bar{\mu}$.

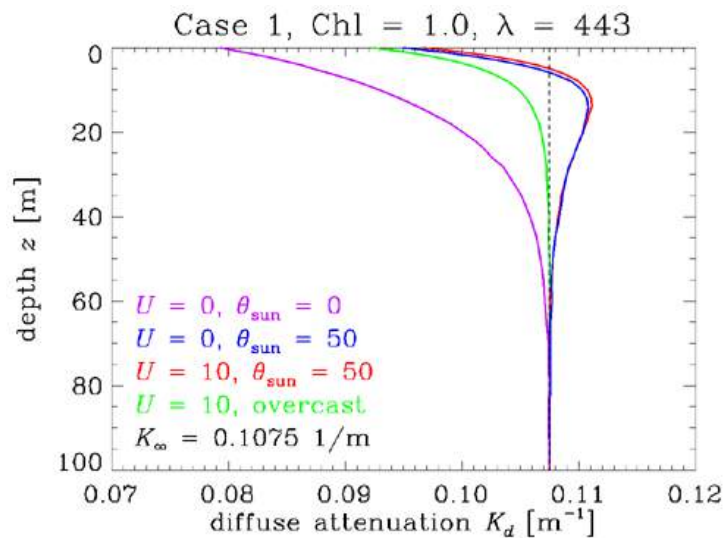


Figure 10.10: Approach of K_d to K_{∞} for one set of IOPs and different sea-surface boundary conditions, as described in the text.

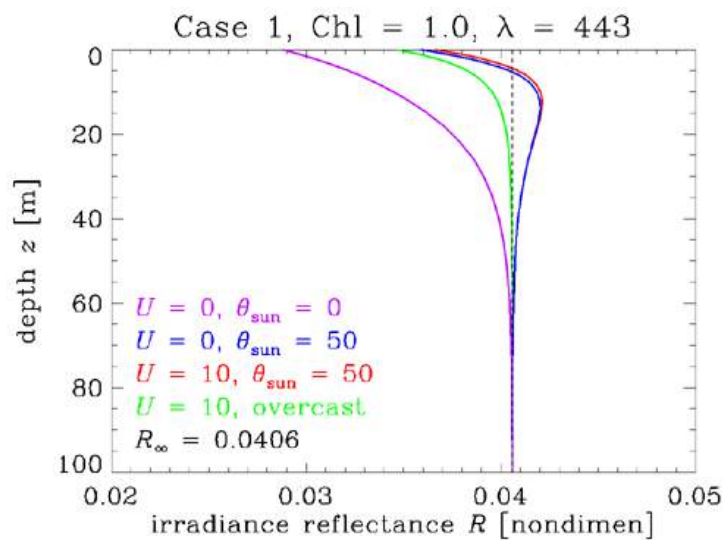


Figure 10.11: Approach of R to R_{∞} for one set of IOPs and different sea-surface boundary conditions, as described in the text. The color-coded curves correspond to those of Fig. 10.10.

Finally, Fig. 10.13 shows the approach of measured and modeled radiances to the asymptotic shape. The dots are radiances measured in the azimuthal plane of the Sun by Tyler (1960) at the depths indicated. The blue curves are the corresponding HydroLight simulation. The measured data were published only as relative values. Therefore, the radiances are normalized for plotting to a value of 1 in the nadir-viewing direction ($\theta_v = 0$) at depth 4.2 m. The $\phi_v = 0$ direction is looking toward the Sun, and $\phi_v = 180$ is looking away from the Sun. The red curve shows the shape of L_{∞} , normalized to the nadir-

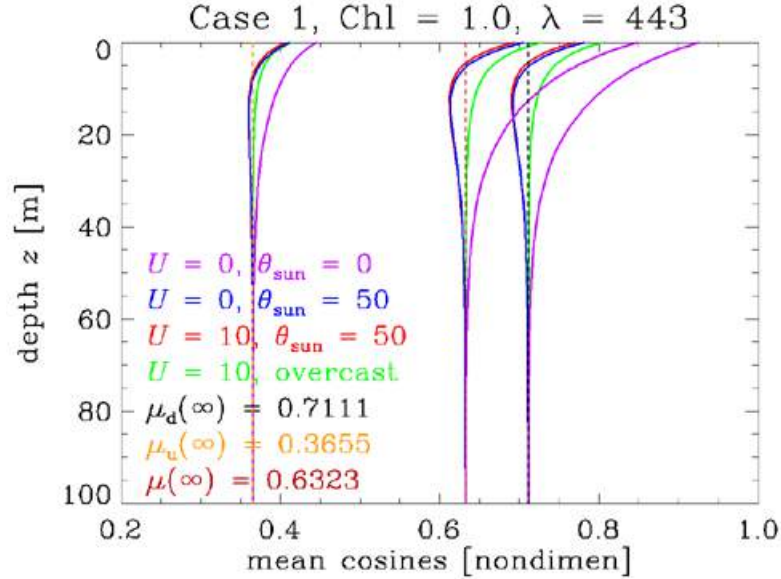


Figure 10.12: Approach of the mean cosines $\bar{\mu}_d$, $\bar{\mu}_u$, and $\bar{\mu}$ to their respective asymptotic values. The color-coded curves correspond to those of Fig. 10.10.

viewing measured value at 66.1 m. The data are described in detail in Tyler’s report, and the modeling is described in Mobley (1994, Section 11.1). Given the uncertainties in the measured data and the educated guesses that had to be made about unmeasured inputs needed by HydroLight, the overall agreement between data and model predictions is quite good.

Near the surface, the Sun’s location is obvious and the unscattered direct beam gives a large spike in the radiance. At 29.0 m, the Sun’s azimuthal direction can still be discerned, but the large spike of the direct beam has been removed by scattering. By 66.1 m, there is only a slight asymmetry remaining to indicate the Sun’s azimuthal direction. Clearly, both the measured and modeled radiances are close to the asymptotic shape at 66.1 m, which was about 26 optical depths.

10.5 An Analytical Asymptotic Solution for Internal Sources

As already noted, exact pencil-and-paper solutions of the radiative transfer equation are extremely rare, but there are a few. Here is an interesting one related to bioluminescence.

Consider an infinitely deep ocean with homogeneous IOPs and a uniform distribution of bioluminescing material. The material is isotropically emitting with a source function $S(z, \theta, \phi, \lambda) = S_o \text{ W m}^{-3} \text{ sr}^{-1} \text{ nm}^{-1}$ at some wavelength. What is the spectral radiance deep within this ocean?

The RTE for this problem is (dropping the wavelength λ)

$$\begin{aligned} \cos \theta \frac{L(z, \theta, \phi)}{dz} &= -cL(z, \theta, \phi) \\ &+ \int_0^{2\pi} \int_0^\pi L(z, \theta', \phi') \beta(\theta', \phi' \rightarrow \theta, \phi) \sin \theta' d\theta' d\phi' + S_o \end{aligned}$$

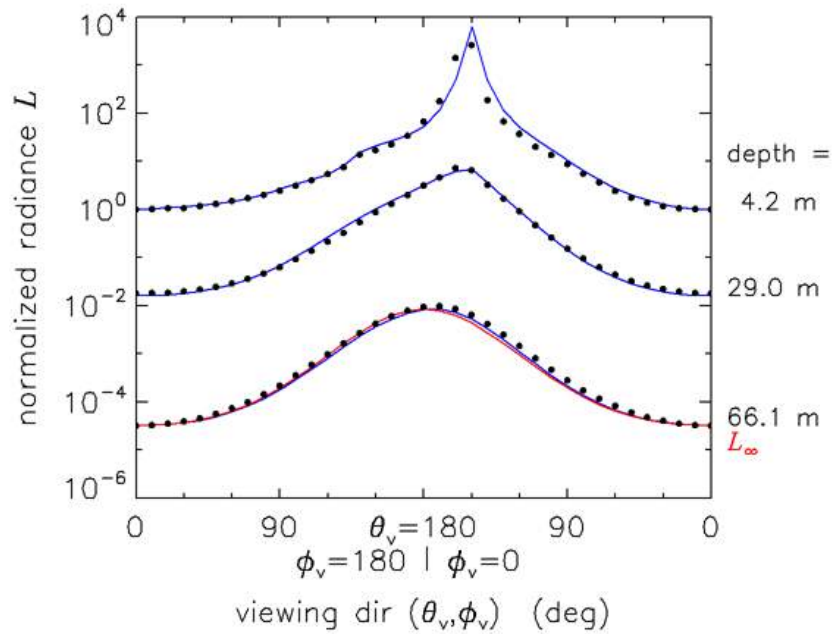


Figure 10.13: Approach of measured and modeled radiances to L_∞ . The dots are radiances measured at the depths indicated. The blue curves are the HydroLight simulation, and the red curve is L_∞ . Measured and modeled radiances are normalized to 1 for the nadir-viewing direction at 4.2 m depth. L_∞ is normalized to the nadir-viewing radiance at 66.1 m.

Because the source is directionally isotropic, the radiance also should be independent of direction when the depth is sufficiently far from the surface boundary. Because the source is constant with depth, the radiance should not depend on the depth when far from the surface boundary. Thus $L(z, \theta, \phi) \rightarrow L_o$, a constant value, at great depth. The left side of the RTE is then zero, and the radiance can be taken out of the path integration. The RTE then becomes

$$0 = -cL_o + L_o \int_0^{2\pi} \int_0^\pi \beta(\theta', \phi' \rightarrow \theta, \phi) \sin \theta' d\theta' d\phi' + S_o$$

The integral of the volume scattering function over all directions (all scattering angles) is by definition the scattering coefficient b . Thus the RTE reduces to just

$$0 = -cL_o + L_o b + S_o ,$$

which gives the radiance as

$$L_o = \frac{S_o}{a} ,$$

where $a = c - b$ is the absorption coefficient. This is the same asymptotic solution as was derived with the assumption of no scattering in the discussion of exact analytical solutions of the SRTE (Eq. 10.2). Here we see that this simple solution holds for *any* volume scattering function.

This analytic solution is easily tested with HydroLight. A run was done using $a = 1 \text{ m}^{-1}$, $b = 4 \text{ m}^{-1}$, and a Petzold average-particle phase function for the water IOPs. The Sun was placed at a 50 deg zenith angle in an otherwise black sky; the solar irradiance onto the level sea surface was $E_d = 1 \text{ W m}^{-2} \text{ nm}^{-1}$. The source magnitude was $S_o = 0.1 \text{ W m}^{-3} \text{ sr}^{-1} \text{ nm}^{-1}$. Figure 10.14 shows the radiance as a function of depth and direction looking upward to see the downwelling radiance L_d (green), looking downward at the upwelling radiance L_u (purple), and looking horizontally toward (blue), at perpendicular to (red), and away from (orange) the Sun's azimuthal direction.

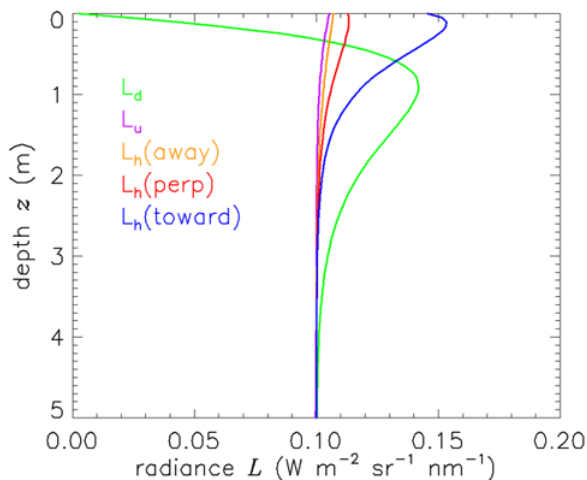


Figure 10.14: Radiances in selected directions for an average-particle scattering phase function and other conditions as described in the text.

Figure 10.15 shows the radiances when an isotropic phase function is used, and all other simulation conditions remain the same. The near-surface radiances are now different. In particular, the three horizontal directions are all identical because of the isotropic scattering. However, the radiance at depth remains the same.

Thus the radiance does depend on depth and direction near the sea surface, where transmitted sky radiance contributes significantly to the total. However, within a few meters (about 20 optical depths for these IOPs) of this boundary, all radiances converge to the predicted value of $L_o = S_o/a = 0.1 \text{ W m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$. Although the *rate of approach* to the final radiance value depends on the scattering phase function, Sun and sky conditions, and sea state, the solution far from the surface boundary does not.

Depending on your point of view, this simulation can be taken as a numerical confirmation that the analytic solution is correct, or as a validation of the HydroLight code by comparison with a known solution of the RTE.

10.6 HydroLight

Full disclosure: HydroLight was developed by Curtis Mobley and is now a commercial software product of [Numerical Optics, Ltd.](#) Because HydroLight is used to generate many of the figures in this book, and because it is widely by the oceanic optics community, some discussion of the model is warranted.

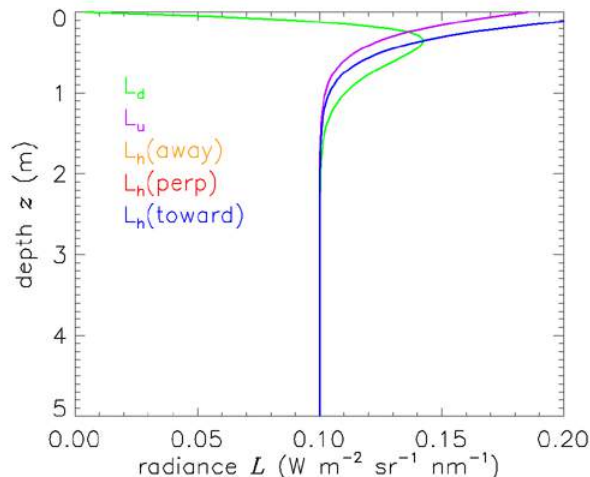


Figure 10.15: Radiances in selected directions for an isotropic scattering phase function and other conditions the same as for Fig. 10.14.

HydroLight is a radiative transfer numerical model that computes radiance distributions and derived quantities (irradiances, reflectances, K functions, etc.) for natural water bodies. It is designed to solve a wide range of problems in optical oceanography and ocean color remote sensing. Many of the figures of this book show HydroLight-computed results.

In brief, HydroLight solves the time-independent, depth-dependent, scalar radiative transfer equation (9.21) to compute the radiance distribution within and leaving any plane-parallel water body. The spectral radiance distribution is computed as a function of depth, direction, and wavelength within the water. The upwelling radiance just above the sea surface includes both the water-leaving radiance and that part of the incident direct and diffuse sky radiance that is reflected upward by the wind-blown sea surface. The water-leaving and reflected-sky radiances are computed separately in order to isolate the water-leaving radiance, which is the quantity of interest in most remote sensing applications. Input to the model consists of the absorbing and scattering properties of the water body, the nature of the wind-blown sea surface and of the bottom of the water column, and the Sun and sky radiance incident on the sea surface. Output consists both of archival printout, Excel spreadsheets, and of files of digital data from which graphical or other analyses can be performed.

The input absorbing and scattering properties of the water body can vary arbitrarily with depth and wavelength. These IOPs can be obtained from actual measurements or from analytical models, which can build up the IOPs from contributions by any number of components. The software comes with various bio-optical models for Case 1 and 2 waters, which are based on historical and recent publications on absorption and scattering by various water constituents. The most general case 2 IOP model can have any number of components (such as different phytoplankton functional groups, different types of mineral particles, dissolved substances, bubbles, etc.). The user can also write subroutines to define the IOPs in any chosen way.

The input sky radiance distribution can be completely arbitrary in the directional and wavelength distribution of the solar and diffuse sky light. HydroLight does not solve the RTE for the atmosphere to obtain the radiance incident onto the sea surface. However, it

does include default sky radiance and irradiance models based on published atmospheric radiative transfer models.

In its most general solution mode, HydroLight includes the effects of inelastic scatter by chlorophyll fluorescence, by colored dissolved organic matter (CDOM) fluorescence, and by Raman scattering by the water itself. The model also can simulate internal layers of bioluminescing microorganisms (but not point sources of bioluminescence, which give a 3-D radiative transfer problem).

HydroLight employs mathematically sophisticated invariant imbedding techniques to solve the radiative transfer equation. Details of this solution method are given in [Mobley \(1994\)](#). When computing the full radiance distribution, invariant imbedding is computationally extremely fast compared to other solution methods such as discrete ordinates and Monte Carlo simulation. Computation time is almost independent of the depth variability of the inherent optical properties (whereas a discrete ordinates model, which resolves the depth structure as N homogeneous layers, takes N times as long to run for stratified water as for homogeneous water). Computation time depends linearly on the depth to which the radiance is desired (whereas Monte Carlo computation times increase exponentially with depth). All radiance directions are computed with equal accuracy. There is no statistical noise caused by the invariant imbedding in-water RTE solution (although there can be a small amount of statistical noise resulting from HydroLight's Monte Carlo treatment of wind-blown water surfaces). Monte Carlo models suffer from statistical noise, and quantities such as radiance contain more statistical noise than quantities such as irradiance, because the simulated rays must be partitioned into smaller directional bins when computing radiances. The water-leaving radiance—the fundamental quantity in remote sensing studies—is very time consuming to compute with Monte Carlo simulations because so few incident rays are backscattered into upward directions.

HydroLight has been under continuous development for over 30 years, with its first published description in [Mobley \(1989\)](#). It has been extensively compared with independent numerical models (e.g., [Mobley et al., 1993](#)) (wherein HydroLight version 3.0 is referred to as “Invariant Imbedding”). The literature contains many comparisons between HydroLight predictions and measured data; representative examples are seen in [Mobley et al. \(2002\)](#), [Chang et al. \(2003\)](#), [Tzortziou et al. \(2006\)](#), and [Tonizzo et al. \(2017\)](#). Although several researchers have developed excellent numerical codes for solving the RTE in the oceanographic setting, their codes are not readily available. HydroLight is commercially available, thoroughly validated, and therefore is widely used and trusted.

10.6.1 The HydroLight Physical Model

The version of the RTE solved by HydroLight is describes the following physical conditions:

- time-independent
- horizontally homogeneous IOPs and boundary conditions
- arbitrary depth dependence of IOPs
- wavelengths between 300 and 1000 nm (for the default underlying databases)
- sea surfaces are modeled either by wave variance spectra or by Cox-Munk sea-surface slope statistics
- finite or infinitely deep (non-Lambertian) water-column bottom

- can optionally include Raman scatter by water
- can optionally include fluorescence by chlorophyll and CDOM
- can optionally include horizontally homogeneous internal sources such as bioluminescing layers
- includes all orders of multiple scattering
- does not include polarization
- does not include whitecaps

These conditions are appropriate for many (but not all) oceanographic simulations. HydroLight cannot, for example, simulate time-dependent wave focusing by surface waves because its sea surface treatment describes the spatially or temporally averaged effects of surface waves. It cannot be used for pulsed lidar bathymetry simulation, which is an inherently time-dependent problem. It cannot simulate sloping bottoms or the radiance reflected by an object in the water, which are inherently 3-D problems. Probably the most limiting simplification of the physics of HydroLight is that it solves the scalar, or unpolarized, RTE. It thus cannot be used for studies where the state of polarization is of interest.

10.6.2 The HydroLight Computational Model

Any radiance sensor actually measures an average of $L(z, \theta, \phi, \lambda)$ taken over some finite solid angle $\Delta\Omega$, which is determined by the field of view of the instrument, and over some finite bandwidth $\Delta\lambda$, which is determined by the wavelength response of the instrument. Likewise, in order to solve the RTE numerically, it must be discretized (or otherwise simplified) by averaging over direction and wavelength to obtain a finite number of values that must be computed. In HydroLight, this directional averaging is performed by first partitioning the set of all directions (θ, ϕ) , $0 \leq \theta \leq 180$ deg, $0 \leq \phi < 360$ deg, into regions bounded by lines of constant θ (like lines of constant latitude) and constant ϕ (constant longitude), plus two polar caps. These quadrilateral regions and polar caps are collectively called “quads.” The individual quads Q_{uv} are labeled by discrete indices $u = 1, 2, \dots, M$ and $v = 1, 2, \dots, N$ to show their θ and ϕ positions, respectively. The standard (default) quad layout is shown in Figure 10.16. In this layout, which has $M = 20$ and $N = 24$, the polar caps have a 5 deg half angle and the θ boundaries lie at 5, 15, 25, ..., 75, 85, 90, 95, 105, ..., 175 deg. For mathematical reasons there is no quad centered on the “equator” at $\theta = 90$ deg. However, the radiances computed for the 85-90 and 90-95 deg quads can be averaged to get the “horizontal” radiance at a nominal angle of $\theta = 90$ deg. Thus the HydroLight standard quad layout essentially gives 10 deg resolution in θ and 15 deg in ϕ . This is adequate for most oceanographic simulations.

Similarly, the wavelength region of interest is partitioned into a number of contiguous wavelength bands of width $\Delta\lambda_j$, $j = 1, 2, \dots, J$. The $\Delta\lambda_j$ need not be the same size for different j values.

The fundamental quantities computed by HydroLight are then the quad- and band-averaged radiances at any selected set of depths z_k , $k = 1, 2, \dots, K$:

$$L(k, u, v, j) = \frac{1}{\Delta\Omega_{uv} \Delta\lambda_j} \int_{\Delta\lambda_j} \int_{Q_{uv}} L(z_k, \theta, \phi, \lambda) d\Omega(\theta, \phi) d\lambda .$$

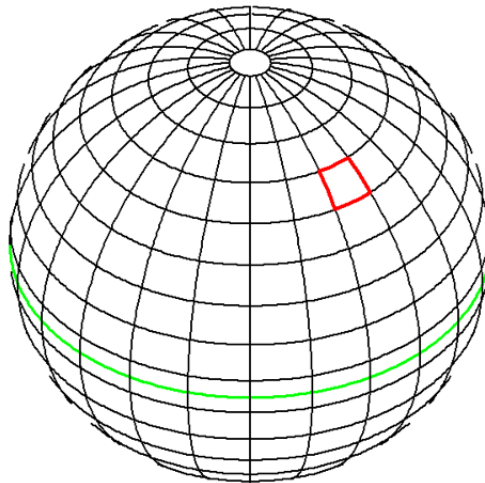


Figure 10.16: The HydroLight standard quad layout, which has a nominal resolution of $\theta = 10$ deg and $\phi = 15$ deg. The red box indicates a particular 10×15 deg quad. The green line is the “equator” between the upper and lower hemispheres of directions.

The quads “homogenize” or average the radiance within each quad, just like a diffuser does in an instrument. Thus, in the quad layout of Fig. 10.16, it is not possible to resolve the difference in the radiance for polar angles $\theta = 26$ deg and $\theta = 34$ deg, because they both lie in the same quad extending from $\theta = 25$ deg and $\theta = 35$ deg. However, there is a difference in $\theta = 34$ deg and $\theta = 36$ deg, because those angles lie in different quads and thus are represented by (probably) different quad-averaged radiances. This same sort of directional averaging of radiances occurs in Monte Carlo models, which collect rays in directional “bins.” If it is necessary to have greater angular resolution in the radiance distribution, a different quad layout can be created. However, the computer storage and run time are proportional to the square of the number of quads, so increasing the angular resolution comes with increased computational cost, just as for other solution techniques.

10.6.3 Ways in Which HydroLight Can Be Used

HydroLight has been used in numerous published studies on topics as diverse as biological primary production, ecosystem modeling, remote sensing, underwater visibility, mixed-layer thermodynamics, and generation of large synthetic data sets needed for neural network training, generation of spectrum-matching libraries, and design of ocean color satellite sensors and retrieval algorithms. These studies have used HydroLight in various ways:

- HydroLight can be run with modeled input values to generate in-water scalar irradiances, which in turn become the input to models of primary productivity or mixed-layer thermodynamics. Accurate light calculations are fundamental to the coupling of physical, biological, and optical ecosystem models.
- HydroLight can be run with the IOPs of different water types to simulate in-water light fields for the purpose of selecting or designing instruments for use in various

water types. Such information aids in the planning of field experiments.

- HydroLight can be run with assumed water inherent optical properties as input, in order to obtain estimates of the signals that would be received by various types or configurations of remote sensors, when flown over different water bodies and under different environmental conditions. Such information can guide the planning of specific operations.
- HydroLight can be used to isolate and remove unwanted contributions to remotely sensed signatures. Consider the common remote-sensing problem of extracting information about a water body from a downward-looking imaging spectrometer. The detected radiance contains both the water-leaving radiance (the signal, which contains information about the water body itself) and sky radiance reflected upward by the sea surface (the noise). HydroLight separately computes each of these contributions to the radiance heading upward from the sea surface and thus provides the information necessary to correct the detected signature for surface-reflection effects.
- When analyzing experimental data, HydroLight can be run repeatedly with different water optical properties and boundary conditions, to see how particular features of the data are related to various physical processes or features in the water body, to substance concentrations, or to boundary or other external environmental effects. Such simulations can be valuable in formulating hypotheses about the causes of various features in the data.
- HydroLight can be used to simulate optical signatures for the purpose of evaluating proposed remote-sensing algorithms for their applicability to different environments or for examining the sensitivity of algorithms to simulated noise in the signature.
- HydroLight can be used to characterize the background environment in an image. When attempting to extract information about an object in the scene, all of the radiance from the natural environment may be considered noise, with the radiance from the object being the signal. HydroLight can be used to compute and remove the environmental contribution to the image.
- HydroLight can be run with historical (climatological) or modeled input data to provide estimates about the marine optical environment during times when remotely or in-situ sensed data are not available.

10.6.4 Inputs to HydroLight

In order to run HydroLight to predict the spectral radiance distribution within and leaving a particular body of water, during particular environmental (sky and surface wave) conditions, the user supplies the core model with the following information (via built-in submodels, or user-supplied subroutines or data files):

- The inherent optical properties of the water body. These optical properties are the absorption and scattering coefficients and the scattering phase function. These properties must be specified as functions of depth and wavelength.
- The state of the wind-blown sea surface. HydroLight can model the sea surface using the Cox-Munk capillary-gravity wave slope statistics (Section 13.4), which adequately describe the optical reflection and transmission properties of the sea surface

for moderate wind speeds and solar angles away from the horizon. In this case, only the wind speed needs to be specified. HydroLight can also use wave variance spectra (Appendices B and C), which resolve both wave height and wave slope.

- The sky spectral radiance distribution. This radiance distribution (including background sky, clouds, and the Sun) can be obtained from semi-empirical models that are built into HydroLight, from observation, or from a separate user-supplied atmospheric radiative transfer model (such as MODTRAN).
- The nature of the bottom boundary. The bottom boundary is specified via its bidirectional reflectance distribution function (BRDF, Section 13.6). If the bottom is a Lambertian reflecting surface at a finite depth, the BRDF is defined in terms of the irradiance reflectance of the bottom. For infinitely deep water, the inherent optical properties of the water body below the region of interest are given, from which HydroLight computes the needed (non-Lambertian) BRDF describing the infinitely deep layer of water below the greatest depth of interest.

The absorption and scattering properties of the water body can be provided to HydroLight in various ways. For example, if actual measurements of the total absorption and scattering are available at selected depths and wavelengths, then these values can be read from files provided at run time. Interpolation is used to define values for those depths and wavelengths not contained in the data set. In the absence of actual measurements, the IOPs of the water body can be modeled in terms of contributions by any number of components. Thus the total absorption can be built up as the absorption by water itself, plus the absorption by chlorophyll-bearing microbial particles, plus that by CDOM, by detritus, by mineral particles, and so on. In order to specify the absorption by chlorophyll-bearing particles, for example, the user can specify the chlorophyll profile of the water column and then use a bio-optical model to convert the chlorophyll concentration to the needed absorption coefficient. The chlorophyll profile also provides information needed for the computation of chlorophyll fluorescence effects. Each such absorption component has its own depth and wavelength dependence. Similar modeling can be used for scattering.

Phase function information can be provided by selecting (from a built-in library) a phase function for each IOP component, e.g., using a Rayleigh-like phase function for scattering by the water itself, by using a Petzold type phase function for scattering by particles, and by assuming that dissolved substances like CDOM do not scatter. HydroLight can also generate phase functions that have a specified backscatter fraction. For example, if the user has both measured scattering coefficients $b(z, \lambda)$ (e.g., from a WETLabs ac-s instrument) and measured backscatter coefficients $b_b(z, \lambda)$ (e.g., from a WETLabs bb-9 or HOBILabs HydroScat-6 instrument), then HydroLight can use the ratio $b_b(z, \lambda)/b(z, \lambda)$ to generate a Fournier-Forand phase function (Section 6.7) that has the same backscatter fraction at each depth and wavelength. The individual-component phase functions are weighted by the respective scattering coefficients and summed in order to obtain the total phase function.

HydroLight does not carry out radiative transfer calculations for the atmosphere. The sky radiance for either cloud-free or overcast skies can be obtained from simple analytical models or from a combination of semi-empirical models. Such models are included in the HydroLight code. Alternatively, if the sky radiance is measured, that data can be used as input to HydroLight via a user-supplied data file. It is also possible to run an independent atmospheric radiative transfer model (such as MODTRAN) in order to generate the sky

radiance coming from each part of the sky hemisphere, and then give the model-generated values to HydroLight as input.

The bottom boundary condition is applied at the deepest depth of interest in the simulation at hand. For a remote sensing simulation concerned only with the water-leaving radiance, it is usually sufficient to solve the radiative transfer equation only for the upper two optical depths, because almost all light leaving the water surface comes from this near-surface region. In this case, the bottom boundary condition can be taken to describe an optically infinitely deep layer of water below two optical depths. In a biological study of primary productivity, it might be necessary to solve for the radiance down to five (or more) optical depths to reach the bottom of the euphotic zone, in which case the bottom boundary condition would be applied at that depth. In such cases, HydroLight computes the needed bottom boundary BRDF from the inherent optical properties at the deepest depth of interest. The bottom boundary condition also can describe a physical bottom at a given geometric depth. In that case, irradiance reflectance of the bottom must be specified (for a Lambertian bottom). In general, this reflectance is a function of wavelength and depends on the type of bottom—mud, sand, sea grass, etc. The user can also write a subroutine to define a non-Lambertian bottom BRDF.

10.6.5 Outputs from HydroLight

HydroLight generates files of archival “printout,” which are convenient for a quick examination of the results, and larger files of digital data. The digital files include Excel spreadsheets and files of data (including the full radiance distribution) formatted for input into graphics packages such as IDL. (The software package contains example routines that use IDL to read and plot HydroLight output. Those routines created many of the HydroLight-computed figures in this book.). The default printout gives a moderate amount of information to document the input to the run and to show selected results of interest to most oceanographers (such as various irradiances, reflectances, mean cosines, K -functions, and radiances in selected directions). This output is easily tailored to the user’s requirements. A file of digital data contains the complete input and output for the run, including the full radiance distribution. This file is generally used as input to plotting routines that give graphical output of various quantities as functions of depth, direction, or wavelength. All input and output files are in ASCII (text file) or Excel spreadsheet format to enable easy transfer between different computer systems.

10.6.6 Documentation

The invariant imbedding algorithms used within HydroLight are described in detail in *Light and Water* (Mobley, 1994), in particular Chapters 4 and 8. The source code is extensively documented with comments referencing equations in *Light and Water* and other publications. There is a *Users’ Guide* (Hedley and Mobley, 2019b) that describes how to run the code, and *Technical Documentation* (Hedley and Mobley, 2019a) that gives information about the included models for IOPs, bottom reflectances, sky radiances, and such. The latest versions of these documents can be downloaded from the references page of the Web Book.

The software itself comes in native versions for the Microsoft Windows, Apple OSX, and Linux operating systems. The mathematical source code is written in Fortran 95.

The user interface is written in C++ and uses the Qt library for graphical elements. There are related specialized versions of HydroLight³ called EcoLight and EcoLight-S(ubroutine). For further details, see [Numerical Optics, Ltd.](#)

10.7 Gershun's Law

The radiative transfer equation is a statement of energy conservation in the sense that it accounts for all the losses and gains to a beam of light moving through the water along a path in a fixed direction. We now derive a useful conservation statement that holds at a fixed point in the water, through which light is moving in all directions.

The desired result is obtained by integrating the 1-D, time-independent, source-free RTE

$$\begin{aligned} \cos \theta \frac{dL(z, \theta, \phi, \lambda)}{dz} = & -c(z, \lambda) L(z, \theta, \phi, \lambda) \\ & + \int_0^{2\pi} \int_0^\pi L(z, \theta', \phi', \lambda) \beta(z, \theta', \phi' \rightarrow \theta, \phi, \lambda) \sin \theta' d\theta' d\phi' \end{aligned} \quad (10.28)$$

over all directions. Dropping the wavelength argument for brevity and writing the differential element of solid angle $\sin \theta d\theta d\phi$ as $d\Omega(\theta, \phi)$, the left hand side of Eq. (10.28) yields

$$\begin{aligned} \int_0^{2\pi} \int_0^\pi \cos \theta \frac{dL(z, \theta, \phi)}{dz} d\Omega(\theta, \phi) &= \frac{d}{dz} \int_0^{2\pi} \int_0^\pi L(z, \theta, \phi) \cos \theta d\Omega(\theta, \phi) \\ &= \frac{d}{dz} [E_d(z) - E_u(z)] \end{aligned} \quad (10.29)$$

after noting that $\cos \theta < 0$ for $\pi/2 < \theta \leq \pi$ and recalling the definitions (1.22) and (1.23) of the upwelling and downwelling plane irradiances as integrals of the radiance. The $-cL$ term becomes

$$\begin{aligned} \iint -c(z)L(z, \theta, \phi)d\Omega(\theta, \phi) &= -c(z) \iint L(z, \theta, \phi)d\Omega(\theta, \phi) \\ &= -c(z)E_o(z), \end{aligned} \quad (10.30)$$

where the double integration over all directions is the same as shown in Eq. (10.29), and $E_o(z)$ is the scalar irradiance. The elastic scatter path function gives

$$\begin{aligned} & \iint \left[\iint L(z, \theta', \phi') \beta(z, \theta', \phi' \rightarrow \theta, \phi) d\Omega(\theta', \phi') \right] d\Omega(\theta, \phi) \\ &= \iint L(z, \theta', \phi') \left[\iint \beta(z, \theta', \phi' \rightarrow \theta, \phi) d\Omega(\theta, \phi) \right] d\Omega(\theta', \phi') \\ &= b(z) \iint L(z, \theta', \phi') d\Omega(\theta', \phi') = b(z)E_o(z). \end{aligned} \quad (10.31)$$

³*Caveat Emptor:* There are many ‘‘HydroLight’’ products on the market, including hydrogen powered lighting systems, hydroelectric power systems, underwater dive lights, lighting for irrigation systems, lighting for growing recreational plants in your basement, and even skin care lotions, bicycles, sports drinks, and a toothbrush. However, none of those other HydroLights can solve the radiative transfer equation.

Here we recall that the integral of the volume scattering function over all directions is the scattering coefficient.

Collecting terms (10.29)-(10.31) resulting from the directional integration of the RTE, we have

$$\frac{d}{dz} [E_d - E_u] = -cE_o + bE_o ,$$

or

$$\frac{d}{dz} [E_d(z, \lambda) - E_u(z, \lambda)] = -a(z, \lambda)E_o(z, \lambda) \quad (\text{W m}^{-3} \text{ nm}^{-1}), \quad (10.32)$$

which is the desired result. This equation is known as Gershun's law (Gershun, 1936) and (Gershun, 1939).

The physical significance of Eq. (10.32) is that it relates the depth rate of change of the net irradiance $E_d - E_u$ to the absorption coefficient a and the scalar irradiance E_o . If inelastic contributions (fluorescence and Raman scattering) and internal sources (such as bioluminescence) are negligible at the wavelength of interest, then Eq. (10.32) can be used to obtain the absorption coefficient a from *in situ* measurements of the irradiance triplet E_d, E_u , and E_o :

$$a(z, \lambda) = -\frac{1}{E_o(z, \lambda)} \frac{d[E_d(z, \lambda) - E_u(z, \lambda)]}{dz} . \quad (10.33)$$

This is an example of an *explicit inverse model*—a model that retrieves an inherent optical property from measurements of the light field. Voss (1989) used Eq. (10.33) to recover a values to within an estimated error of order 20%. Inelastic scattering and internal source effects were reasonably assumed to be negligible in his study. The needed irradiances were all computed from a measured radiance distribution, so that no intercalibration of instruments was required. Li et al. (2018) made extensive measurements of the irradiance quartet E_d, E_u, E_{od} , and E_{ou} and used those to retrieve absorption, which was compared with measurements made by an ac9. Some of their results are seen in Fig. 7.24.

A more general development can be made to account for internal sources or inelastic scatter and for 3-D and time-dependent light fields, as shown in *Light and Water* (Mobley, 1994, Section 5.10). The result is known as the *divergence law for irradiance*:

$$\frac{1}{v} \frac{\partial E_o}{\partial t} + \nabla \cdot \vec{E} = -aE_o + E_o^S . \quad (10.34)$$

Here v is the speed of light in the water, \vec{E} is the vector irradiance, and E_o^S is a source term. For time-independent, 1-D, source-free water, this equation reduces to Eq. (10.32).

Maffione et al. (1993) determined absorption values by writing the source-free form of the 3-D divergence law in spherical coordinates and applying the result to irradiance measurements made using an underwater, artificial, *isotropic* light source. The artificial light source allowed measurements to be made at night, thus there was no inelastic scattering from other wavelengths. Their instrument did not require absolute radiometric calibration.

Note, however, that Gershun's law will give *incorrect* absorption values if naively applied to waters and wavelengths where inelastic processes such as Raman scattering or fluorescence are significant; see Li et al. (2018) for examples. For this reason, and because of calibration difficulties if different instruments are used to measure E_d, E_u , and E_o , Gershun's law is seldom used as a way to measure absorption. Nevertheless, it is sometimes a useful check on the internal consistency of numerical models or measured data, and it leads to a convenient way of calculating radiant heating rates.

10.7.1 Heating Rates in the Upper Ocean

Gershun’s law has a very important application in the computation of heating rates in the upper ocean. The rate of heating of water depends on how much scalar irradiance is available and on the total absorption coefficient of the water (pure water plus all other constituents). Combining the First Law of Thermodynamics (i.e., conservation of energy) with Eq. (10.32) gives

$$\frac{\partial T}{\partial t}(z, \text{at } \lambda) = \frac{1}{c_v \rho} a(z, \lambda) E_o(z, \lambda) = -\frac{1}{c_v \rho} \frac{\partial [E_d(z, \lambda) - E_u(z, \lambda)]}{\partial z} \left[\frac{\text{deg C}}{\text{sec}} \right].$$

In this equation,

- T is the temperature in deg C
- t is the time in seconds
- $\frac{\partial T}{\partial t}(z, \text{at } \lambda)$ denotes the rate of change of temperature at depth z due to energy absorbed at wavelength λ . It does not mean that temperature is a function of wavelength.
- $c_v = 3900 \text{ J (kg deg C)}^{-1}$ is the specific heat of sea water at constant volume
- $\rho = 1025 \text{ kg m}^{-3}$ is the density of sea water

This equation is usually applied with wavelength-integrated irradiances. In ocean modeling, “short-wave” radiation is usually taken to be the range of 400–1000 nm. Letting

$$E_{d,u}(z) = \int_{400}^{1000} E_{d,u}(z, \lambda) d\lambda$$

gives

$$\frac{\partial T}{\partial t}(z, \text{short-wave}) = -\frac{1}{c_v \rho} \frac{\partial [E_d(z) - E_u(z)]}{\partial z}. \quad (10.35)$$

In optically deep water, $E_u(z) \ll E_d(z)$, and the upwelling irradiance is usually dropped. The downwelling irradiance is then modeled, e.g. in terms of the chlorophyll concentration in Case 1 water. The EcoLight-S radiative transfer code (Mobley, 2011) was developed to provide extremely fast solutions of the radiative transfer equation to obtain $E_d(z)$ and $E_u(z)$ for any water body (deep or shallow, Case 1 or Case 2) for use in Eq. (10.35) in coupled physical-biological-optical ecosystem models. (EcoLight-S also computes other quantities such as PAR and remote-sensing reflectance, which are inputs to photosynthesis calculations or are useful for model validation.) For an example application of such a model, see Mobley et al. (2015).

10.8 Energy Conservation

When making or computing radiance or irradiance measurements in the ocean, it is sometimes hard to verify that energy is being conserved. This is equally true when looking at the output from a numerical model such as HydroLight. Confusion most often arises when comparing measurements made just above and below the sea surface and observing that the measured (or predicted) radiances do not seem to “add up” correctly. To resolve such confusion, first note that it’s the Law of Conservation of *Energy*, not the law of conservation of radiance or irradiance. We must therefore express the law of conservation of energy in terms of radiance or irradiance as appropriate to the measurements being made.

10.8.1 Energy Conservation Within the Water

Within the water column, if there are no internal sources such as bioluminescence or inelastic scatter contributing to the wavelength of interest, conservation of energy is expressed in terms of irradiances by Gershun’s law (10.33). This can be illustrated using numerical results generated by HydroLight. A run was made with the Sun placed at a 60 deg zenith angle in a clear sky; the wind speed was 5 m s⁻¹, and the water was infinitely deep. The absorption coefficient was set to $a = 0.1 \text{ m}^{-1}$, the scattering coefficient to $b = 0.4 \text{ m}^{-1}$, and a Petzold “average-particle” phase function was chosen. (The run used HydroLight’s “Constant IOPs” submodel for these inputs.) No internal sources were included in the run. Table 10.1 shows the computed irradiances at depths of 5.00 and 5.01 m. Closely spaced depths are used to evaluate Eq. (10.33) so that the depth derivative can be accurately approximated as a finite difference and the scalar irradiance can be taken as the average of the values at 5.00 and 5.01 m.

depth (m)	E_o	E_u	E_d
5.00	0.51833	0.013151	0.32613
5.01	0.51749	0.013129	0.32559

Table 10.1: Irradiances computed within the water body as described in the text. The irradiances have units of $\text{W m}^{-2} \text{ nm}^{-1}$.

Using these values in the finite difference approximation of Eq. (10.33) gives

$$\begin{aligned}
 a &= - \frac{1}{0.5(0.51833 + 0.51749)} \frac{(0.32559 - 0.013129) - (0.32613 - 0.013151)}{5.01 - 5.00} \\
 &= 0.1 \text{ m}^{-1}.
 \end{aligned}$$

The absorption coefficient as computed from the irradiances via Gershun’s law is the same as the one used as input to the run, so HydroLight is conserving energy in its in-water solution of the radiative transfer equation. Such accuracy cannot be expected with real data because of imperfect instrument calibration and other sources of error such as Raman scatter (an internal source not accounted for in Eq. (10.32), but which will always be present to some degree during daylight measurements⁴). Nevertheless, Gershun’s law can be a useful way to retrieve absorption coefficients from in-water radiance or irradiance measurements, as seen in papers by Voss (1989) and Maffione et al. (1993).

10.8.2 Energy Conservation Across the Air-Water Surface

The situation is more complicated when considering energy conservation across the air-water surface. Consider a collimated beam of light incident onto a level sea surface. Part will be reflected and part will be transmitted into the water. Figure 10.17 shows this geometry and various quantities needed to express conservation of energy at the surface.

⁴HydroLight also conserves energy if internal sources are present. Demonstration of that requires adding an appropriate source term to Gershun’s equation; see Eq. (10.34).

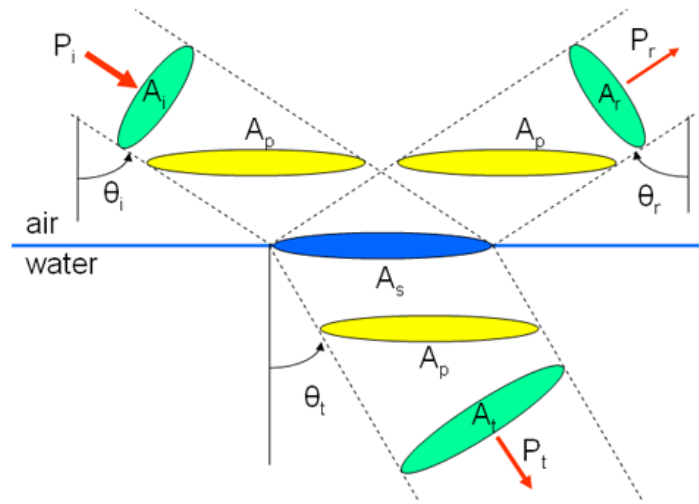


Figure 10.17: Geometry needed for expressing the law of conservation of energy in terms of irradiances across a level sea surface.

Conservation of energy (or equivalently, power) across the surface requires that

$$P_i = P_r + P_t, \quad (10.36)$$

where P_i is the power in an incident collimated beam of cross section A_i , P_r is the power in the reflected beam of cross section A_r , and P_t is the power in transmitted beam of cross section A_t . These beam cross sectional areas are shown in green in Fig. 10.17. The beam of incident power illuminates a horizontal area A_s of the level sea surface (shown in blue).

10.8.2.1 Energy conservation in terms of plane irradiances

Conservation of energy can be expressed in terms of incident (downwelling), reflected (upwelling), and transmitted (downwelling) plane irradiances using the horizontal projections of the beam cross section areas. Recalling that plane irradiance is power per unit horizontal area, Eq. (10.36) becomes

$$A_p E_d(\text{incident, in air}) = A_p E_u(\text{reflected, in air}) + A_p E_d(\text{transmitted, in water}),$$

where the horizontally projected areas A_p corresponding to each beam are shown in yellow in Fig. 10.17. Since each beam has the same horizontally projected area, this equation reduces to just

$$E_d(\text{incident, in air}) = E_u(\text{reflected, in air}) + E_d(\text{transmitted, in water}). \quad (10.37)$$

Equation (10.37) is conservation of energy across the surface expressed in terms of plane irradiances for a collimated incident beam.

10.8.2.2 Energy conservation in terms of scalar irradiances

To express Eq. (10.36) in terms of the scalar irradiances, first recall from (1.24) how E_{od} is obtained from the radiance:

$$E_{\text{od}} = \iint_{2\pi_d} L(\hat{\xi}') d\Omega(\hat{\xi}') .$$

For a collimated incident beam in direction $\hat{\xi}_o$, we can write the radiance in terms of the beam irradiance and a Dirac delta function to get

$$E_{\text{od}} = \iint_{2\pi_d} E_{\perp} \delta(\hat{\xi}' - \hat{\xi}_o) d\Omega(\hat{\xi}') = E_{\perp} ,$$

where E_{\perp} is the incident irradiance on the beam cross sectional area; $E_{\perp} = P_i/A_i$. Similar equations hold for the reflected and transmitted scalar irradiances. These beam cross sectional areas are shown in green in Fig. 10.17. Thus Eq. (10.36) becomes

$$\begin{aligned} A_i E_{\text{od}}(\text{incident, in air}) &= A_r E_{\text{ou}}(\text{reflected, in air}) \\ &+ A_t E_{\text{od}}(\text{transmitted, in water}) , \end{aligned}$$

or

$$\begin{aligned} E_{\text{od}}(\text{incident, in air}) &= E_{\text{ou}}(\text{reflected, in air}) \\ &+ A_t E_{\text{od}}(\text{transmitted, in water}) \frac{A_t}{A_i} , \end{aligned} \quad (10.38)$$

after noting that $A_r = A_i$ by the law of reflection. By geometry, the ratio of areas A_t/A_i is just $\cos \theta_t / \cos \theta_i$. θ_t can be written in terms of θ_i using Snell's law, $\sin \theta_i = n \sin \theta_t$, where n is the real index of refraction of the water. Equation (10.38) then becomes

$$\begin{aligned} E_{\text{od}}(\text{incident, in air}) &= E_{\text{ou}}(\text{reflected, in air}) \\ &+ E_{\text{od}}(\text{transmitted, in water}) \frac{1}{\cos \theta_i} \sqrt{1 - \frac{\sin^2 \theta_i}{n^2}} . \end{aligned} \quad (10.39)$$

Equation (10.39) is conservation of energy expressed in terms of scalar irradiance for a collimated incident beam.

To illustrate Eqs. (10.37) and (10.39) with HydroLight output, a run was made with the Sun at a zenith angle of $\theta_i = 60$ deg in a black sky, in order to obtain a collimated incident beam. The incident E_d onto the surface was set to 1. The sea surface was level (zero wind speed), and the water index of refraction was $n = 1.34$. The scattering coefficient within in the water was set to zero, so that upwelling light from the water column would not confuse the illustration of energy conservation across the sea surface itself. The absorption coefficient was set to $a = 0.1 \text{ m}^{-1}$. (These runs used HydroLight IOP model ‘‘Constant IOPs’’ and the ‘‘idealized sky model.’’)

The computed irradiances just above and below the surface are shown in Table 10.2. The upwelling irradiances in the water are zero because there was no scattering by the water and the bottom was infinitely deep. These computed irradiances satisfy Equations (10.37) and (10.39), showing that HydroLight conserves energy across the surface.

location	E_{ou}	E_{od}	E_u	E_d
in air	0.12690	2.0077	0.063209	1.0000
in water at depth 0	0.00000	1.2277	0.0000	0.93693

Table 10.2: Plane and scalar irradiances for a collimated incident beam and no scattering within the water.

location	E_{ou}	E_{od}	E_u	E_d
in air	0.18863	1.7764	0.050456	1.0000
in water at depth 0	0.00000	1.1430	0.0000	0.94954

Table 10.3: Plane and scalar irradiances for a diffuse incident sky radiance and no scattering within the water.

A run was then done with a heavily overcast sky to generate a diffuse sky radiance pattern with the Sun’s location not discernible. The wind speed was 10 m s^{-1} to generate a non-level sea surface. The resulting irradiances are shown in Table 10.3

As before, these plane irradiances satisfy Eq. (10.37). However, the scalar irradiances do not satisfy Eq. (10.39). The reason is that Eq. (10.39) must be applied for each particular incident direction θ_i of the diffuse sky radiance, in order to correctly compute the angle factor seen in Eq. (10.39). If this is done one direction at a time (with the Sun in a black sky), each individual incident direction satisfies Eq. (10.39) just as in the previous example. Thus energy is conserved across the surface “beam by beam,” hence *in toto*, even if it is not obvious from the total scalar irradiances.

A further complication occurs when the water body has scattering. To illustrate this, the scattering coefficient was set to $b = 0.4 \text{ m}^{-1}$ and a Petzold “average-particle” phase function was used in the previous simulation. The computed irradiances were then

location	E_{ou}	E_{od}	E_u	E_d
in air	0.21264	1.7764	0.062372	1.0000
in water at depth 0	0.07922	1.2046	0.031115	0.96873

Table 10.4: Plane and scalar irradiances for a diffuse incident radiance, with scattering in the water.

Now even the plane irradiances appear to violate conservation of energy, Eq. (10.37). The reason is that E_{ou} and E_u in air contain both surface-reflected radiance (as before) and radiance transmitted upward through the sea surface. Likewise, E_{od} and E_d in the water contain both downward transmitted irradiance (as before) as well as upwelling radiance that has been reflected back downward by the surface. This makes it appear at first glance that the irradiances do not add up correctly, even though energy actually is conserved

across the air-water surface. The effects of both the surface itself and the water body can be accounted for simply by tallying the plane irradiances incident onto and leaving the surface from above and below:

$$\begin{aligned} E_d(\text{downwelling, in air}) - E_u(\text{upwelling, in air}) = \\ E_d(\text{downwelling, in water}) - E_u(\text{upwelling, in water}) . \end{aligned} \quad (10.40)$$

Equation (10.40) expresses conservation of energy for the surface plus water body, for any conditions of incident lighting and water IOPs. The irradiances seen in Table (10.4) satisfy this equation.

10.8.2.3 Energy conservation in terms of radiance

Another complication occurs when expressing conservation of energy across the surface in terms of radiance. Now it is necessary to account for changes in both the beam cross sectional area and in solid angle when going across the surface. When interpreting HydroLight output, it must be remembered that the computed radiances are averages over “quads” or solid angles of finite size, as discussed on the HydroLight section. Such quads are represented by the green areas in Fig. (10.18).

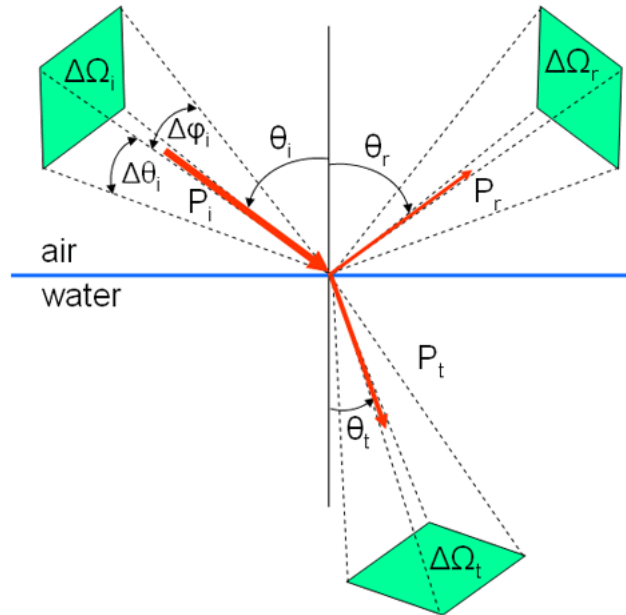


Figure 10.18: Illustration of the HydroLight radiance quads as needed to express energy conservation for a collimated beam across a level sea surface.

Remembering that radiance is power per unit area per unit solid angle, Eq. (10.36) can be written as

$$L_i A_i \Delta \omega_i = L_r A_r \Delta \omega_r + L_t A_t \Delta \omega_t ,$$

or

$$L_i = L_r + L_t \frac{A_t \Delta \omega_t}{A_i, \Delta \omega_i} ,$$

after noting that $A_r = A_i$ and $\Delta\omega_r = \Delta\omega_i$ by the law of reflection. As noted before, $A_t/A_i = \cos\theta_t/\cos\theta_i$. The solid angles are given by $\Delta\Omega = \Delta(\cos\theta)\Delta\phi$, where the angle ranges are determined by the θ and ϕ boundaries of the quads. Thus the last equation becomes

$$L_i = L_r + L_t \frac{\cos\theta_t}{\cos\theta_i} \frac{\Delta(\cos\theta_t)}{\Delta(\cos\theta_i)}, \quad (10.41)$$

since $\Delta\phi_t = \Delta\phi_i$ for the HydroLight quad layout. These angle factors must be evaluated for the particular quads associated with the incident and transmitted beams.

Consider again the previous case of the Sun in a black sky at $\theta_i = 60$ deg, a level water surface, and no scattering. The incident and reflected HydroLight quads are centered at 60 deg and extend from 55 to 65 deg. Snell's law gives the transmitted direction as $\theta_t = \sin^{-1}[(\sin\theta_i)/n] = 40.26$ deg. Thus the transmitted radiance is contained in the HydroLight quad centered at 40 deg and extending from 35 to 45 deg. For the transmitted quad, for example, $\Delta(\cos\theta_t) = \cos 35 - \cos 45 = 0.11205$. Evaluating the angle factors in Eq. (10.41) for these quads gives

$$L_i = L_r + L_t \frac{0.76604}{0.50000} \frac{0.11205}{0.15096}. \quad (10.42)$$

The corresponding computed quad-averaged radiances for this run were $L_i = 50.7996$, $L_r = 3.21094$, and $L_t = 41.8544 \text{ W m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$. These radiances satisfy Eq. (10.42). Thus energy expressed as radiance is conserved beam-by-beam across the water surface.

We have seen that conservation of energy across the sea surface can be expressed in various ways in terms of radiance and irradiances. However, those expressions are often written for collimated incident beams and consider only the surface itself (Eqs. 10.37, 10.39, and 10.41). It is therefore not always obvious that energy is conserved for the entire surface plus water column when looking at data or simulations for realistic conditions of sky radiances and water IOPs. In general, Eq. (10.40) should be used to check conservation of energy across the sea surface with measured or modeled plane irradiances.

10.9 Closure

We have now assembled all of the pieces needed for optical oceanography: radiometric variables, inherent and apparent optical properties, and a family radiative transfer equations. There remains one more topic: closure. "Closure" is just a fancy word for "getting everything to agree," and it comes in various forms:

Scale Closure: Do measurements made on one spatial scale agree with measurements made on a different spatial scale? An ac-s instrument ([Sea-Bird Scientific](#)) measures the absorption coefficient on a few cubic centimeters of water, as does a spectrophotometer (Section 5.5). However, it is also possible to extract the absorption coefficient from Gershun's law (Eq. 10.33) or from remote-sensing measurements, and both of these techniques employ optical measurements (irradiances or remote-sensing reflectance) that are determined by large volumes of the ocean. In a given set of measurements, do these small-volume and large-volume measurements give the same absorption coefficient? If they do, this provides a good check on the quality of the data (and perhaps on the spatial homogeneity of the water body being observed). If they do not, the discrepancy needs to be explained.

Instrument Closure: Do different instruments that supposedly measure the same quantity give the same answers? Do the just-mentioned ac-s and spectrophotometer give the same absorption coefficient spectrum for the same water sample? Both the Coulter counter ([Beckman Instruments](#)) and the LISST ([Sequoia Scientific, Inc.](#)) measure particle size distributions, but using different physical principles to detect the particles. Do they give the same size distributions for a given water body? If so, this gives reassurance that the data are good. If they disagree, what is the reason? Is the difference expected because their different physical principles make them “see” different particle properties, or is the difference because one of the instruments was badly calibrated?

Model Closure: Do different models give the same outputs for the same inputs? Both HydroLight (Section 10.6) and Monte Carlo models (Appendix E) can compute radiometric variables in one-dimensional geometries, but using very different numerical methods to solve the SRTE. Do these very different techniques give the same answers to a given problem? If the answers agree, both models are likely correct. But if not, is the difference due to explainable differences in the physical assumptions of the models, or to different accuracies of the solution algorithms (e.g., statistical noise in the Monte Carlo solution), or perhaps to a bug on one of the codes? There are many bio-geo-optical models that compute various IOPs given the concentrations of the water column constituents such as chlorophyll, CDOM, and mineral particles. Do different IOP models give the same predictions for absorption and scattering coefficients given the same water composition? If not, why not?

Model-Data Closure: Do model predictions agree with measurements? This is perhaps the most difficult closure to achieve. A field experiment may have instruments measuring chlorophyll and other component concentrations, which are used in a bio-optical model to predict the IOPs. Those IOPs are then used in HydroLight to predict in-water $E_d(z, \lambda)$ and $L_u(z, \lambda)$. There may also be a profiling instrument such as a Hyper Pro ([Sea-Bird Scientific](#)) measuring $E_d(z, \lambda)$ and $L_u(z, \lambda)$ directly. Do the measurements and the HydroLight predictions agree? If not, is the disagreement due to bad input to HydroLight (e.g., from the bio-optical model that converted the concentrations to IOPs), or due to bad radiometer calibration, or due to physics not included in HydroLight (e.g., polarization or the effects of 3-D geometry)?

Figure 10.19 show absorption coefficients $a_{\text{total}} - a_w$ measured five different ways in a coastal lagoon. Four instruments were used: two versions of reflective tube absorption meters, the ac9 and acs ([www.seabird.com](#)); and two designs of integrating cavity absorption meters, ICAM ([www.turnerdesigns.com](#)) and PSICAM ([unstonesci.com](#)). (For a detailed description of these instruments and the protocols for their use, see [Neeley and Maninno \(2018\)](#).) Comparison of these instruments gives an example of instrument closure. In addition, Gershun’s law (10.33) was used to retrieve the absorption coefficient from measurements of the scalar and plane irradiances. The four instruments make measurements on small volumes of water inside the instruments, whereas Gershun’s law uses in situ measurements of irradiances that are determined by a large water volume. Comparing the instruments with Gershun’s law thus gives a test of scale closure. As seen in the figure, sometimes the measurements differ by tens of percent (their Station 4, left panel), and sometimes they are within a few percent (Station 9, right panel). This level of disagreement among different instruments and methods for absorption measurements is

typical. The paper by [Kostakis et al. \(2021\)](#) shows an additional ten stations and many other comparisons, including the effects of different scattering corrections for the ac9 and acs instruments.

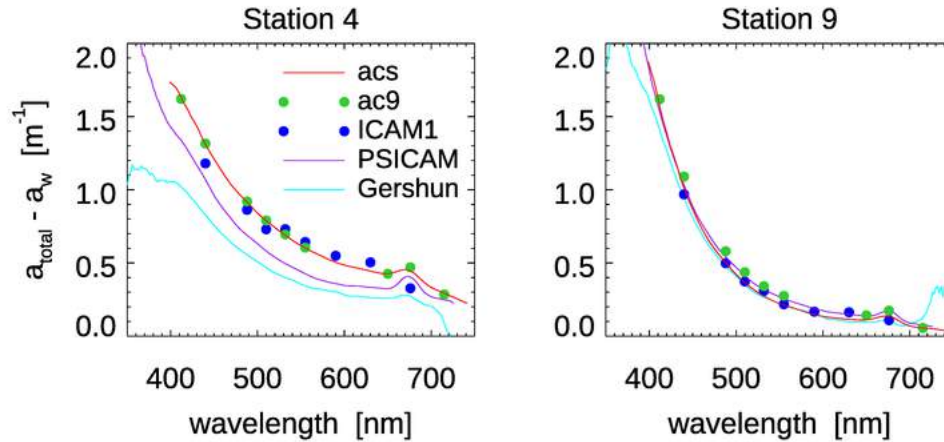


Figure 10.19: Comparison of non-water absorption coefficients measured five different ways in coastal waters. Data from [Kostakis et al. \(2021, Fig 5\)](#) provided courtesy of I. Kostakis.

Figure 10.20 shows measurements of the particle backscatter coefficient b_{bp} made by four different instruments on a water sample containing *E. huxleyi* coccolithophores. As with the absorption measurements just discussed, in principle all four instruments should give the same value at a given wavelength. In reality, there are differences of 30% or more at some wavelengths. Note that the supposed error bars for the individual measurements do not overlap, so the actual errors are greater than the error bars imply.

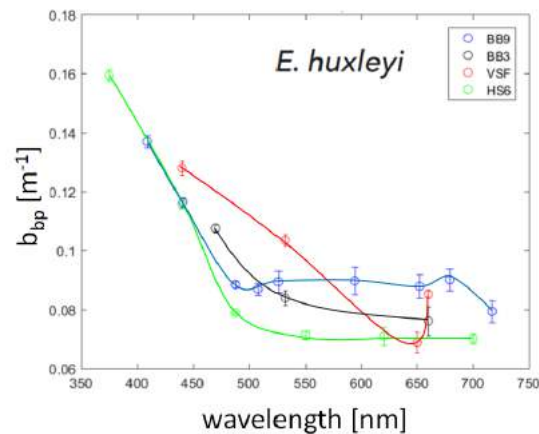


Figure 10.20: Comparison of backscattering coefficients measured by four different instruments on a sample containing *E. huxleyi* coccolithophores. Modified from [Mannino et al. \(2019\)](#).

These two figures show the typical spread of values seen for the same IOP for oceanographic measurements. Optical oceanography is not a field where you can simply buy

an instrument and get six-figure accuracy in the output. Such comparisons give valuable information about the true uncertainties in data values, which then can be used in error analyses of final products such as the remote-sensing reflectance. To first order, the remote-sensing reflectance R_{rs} is proportional to backscatter over absorption, so if absorption and backscatter coefficients with errors of tens of percent are used as input to HydroLight, the HydroLight-predicted R_{rs} values can easily differ by tens of percent (depending on whether the errors in a and b_b have the same or different signs, with all else being the same).

An example of model closure is shown in Fig. 10.21. In that study (Mobley et al., 1993), numerical models using invariant imbedding (HydroLight before it was named), discrete ordinates, and Monte Carlo solution methods for the SRTE were compared for a “canonical” set of problems. The differences in the computed in-water radiances are due to different angular resolutions in the model outputs and to Monte Carlo statistical noise. This paper convinced the community that all of the compared models were giving correct outputs for a given set of inputs and can be trusted in routine use. Thirty years of use of these models has established that when a model like HydroLight gives a “wrong answer,” the cause is almost always due to “bad” IOP inputs from either inaccurate IOP measurements or the use of bio-geo-optical models that are not applicable for the water body being studied.

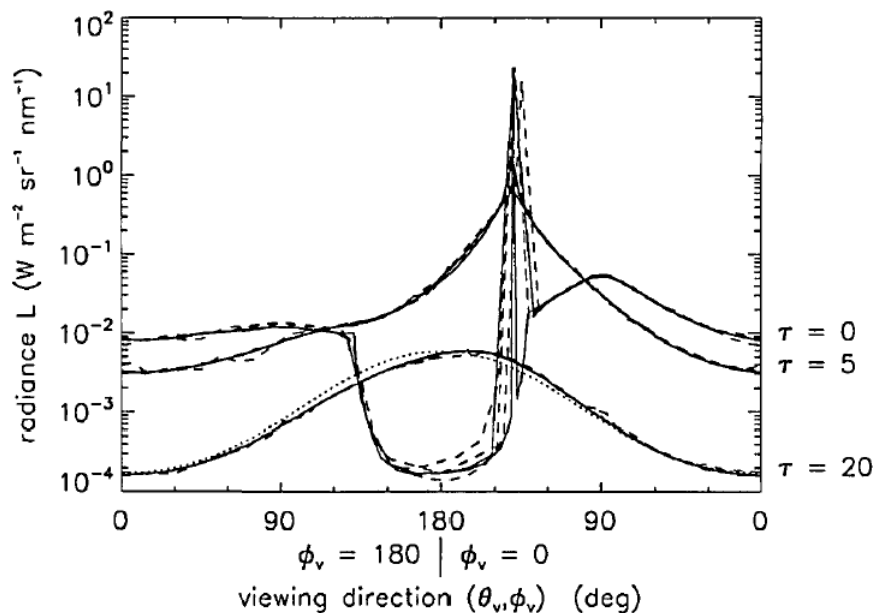


Figure 10.21: Comparison of radiance distributions $L(\tau, \theta_v, \phi_v)$ in the azimuthal plane of the Sun as computed by six different numerical models. (In this figure τ is the optical depth; θ_v and ϕ_v are viewing directions.) The solid lines are computed by invariant imbedding (HydroLight) and discrete ordinates; the dash curves are four different Monte Carlo codes. The dotted line is the asymptotic radiance distribution $L_\infty(\theta_v)$ normalized to the value of $L(\tau = 20, \theta_v = 180)$. Figure 8 of Mobley et al. (1993), reproduced under Optica’s Fair Use policy.

As a final example of closure, Fig. 10.22 shows an example of model-data closure from Tzortziou et al. (2006). This outstanding paper is a gold standard for closure studies and is

highly recommended to see what is involved in taking good data. As seen in Fig. 10.22 and several other figures in the paper, it is possible to achieve closure between measurement and model predictions, but only with great care in the instrument calibration and data processing. In the present case, it was a full PhD project.

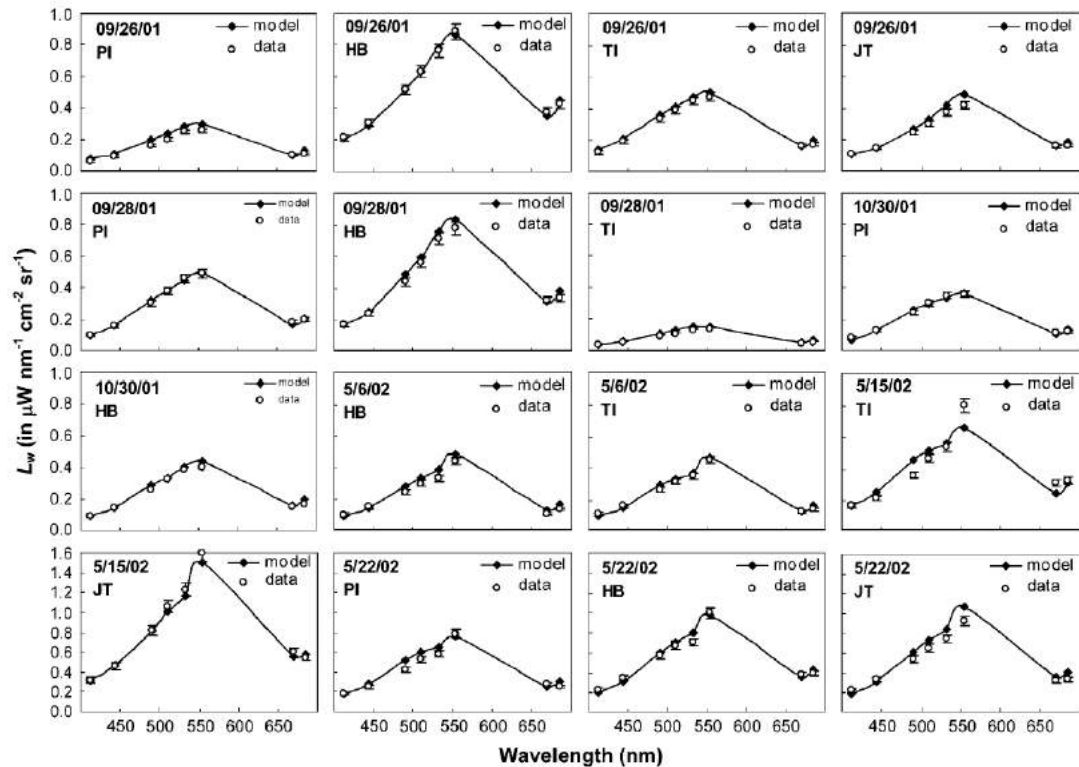


Figure 10.22: Comparison of HydroLight-computed (model) and measured (data) water-leaving radiances L_w at various times and locations in Chesapeake Bay, USA. Figure 7 of Tzortziou et al. (2006), reproduced by paid permission.

Another excellent model-data closure paper is Tonizzo et al. (2017). They were able to get agreement between measured and modeled (HydroLight) R_{rs} spectra to within 20%. They found that about one-half of the difference in model and measurement came from uncertainties in the measurement of R_{rs} , and one-half came from the combined uncertainties in IOPs measurements and radiative transfer modeling (HydroLight does not include polarization).

The examples are sufficient to illustrate the concept of optical closure and the necessity of comparing as many different instruments or data processing methods as possible in order to assess the errors in the measurements or predictions. When working with data and models, it is always necessary to perform as many closure tests as possible. This is the basis of convincing yourself that your models and data are correct, and you can learn a lot from any disagreements. Redundancy is always good, whether in instruments, models, or algorithms.

CHAPTER 11

Maxwell's Equations

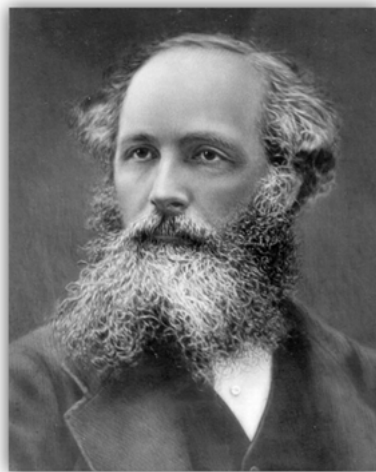


Figure 11.1:
James Clerk Maxwell (1831-1879)
“War es ein Gott, der diese Zeichen schrieb?”
 (“Was it a God who wrote these symbols?”)
—Ludwig Boltzmann, commenting on Maxwell’s equations
 (and recycling a quote from Goethe’s *Faust*)

As seen Section 9.2, radiative transfer theory rests on the theory of electromagnetism as expressed by Maxwell’s equations. In spite of their fundamental importance to everything in optical oceanography and ocean remote sensing, most oceanography texts mention Maxwell’s equations only in passing: a dozen or so pages in [Apel \(1987\)](#), two pages in [Shifrin \(1988\)](#), one paragraph in [Walker \(1994\)](#), and no mention at all in [Neumann and Pierson, Jr. \(1966\)](#) or [Jerlov \(1976\)](#). Specialized, advanced texts like [Bohren and Huffman \(1983\)](#) and [Mishchenko et al. \(2002\)](#) make extensive use of Maxwell’s equations, but with the assumption that the reader is already familiar with them.

A physics major will learn a lot about Maxwell's equations; a biology or chemistry major may learn almost nothing about them. It thus seems fitting to present an overview of these equations at a level in between these two extremes. The goal of this chapter is to discuss Maxwell's equations in a way that minimizes the mathematics but still gives a non-physicist reader some appreciation of these equations and their implications.

The first section of this chapter introduces Maxwell's equation in a vacuum. The next section reformulates them for use in material media such as water. They are then used to study the propagation of electromagnetic waves, i.e. light, in dielectrics like water. Two sections on dispersion follow—the first on the basic concepts of dispersion and the differences in phase and group speeds, and the second on anomalous dispersion. These concepts have been used, for example, to gain a fundamental understanding of absorption by phytoplankton. The chapter finishes with a short section on the Kramers-Kronig relations, which establish a fundamental connection between absorption and the real index of refraction.

The solution of Maxwell's equations for a plane electromagnetic wave incident onto a homogeneous dielectric sphere, the so-called Mie theory, is discussed in the next chapter.

11.1 Maxwell's Equations in Vacuo

This section begins a qualitative overview of Maxwell's equations as they govern electric and magnetic fields in a vacuum. Entire books have been written about these equations, so a few sections are not going to teach you much. The goal here is to present the fundamental ideas and, hopefully, inspire you to continue to study these equations in the references provided. The discussion presumes a knowledge of basic physics (concepts such as electric charge and current, and electric and magnetic fields). Knowledge of vector calculus (divergence and curl in particular) is needed to understand the equations, but you can understand the basic ideas even without the math. If you are unfamiliar with the basic physics and math of electric and magnetic fields, or need a good review, an excellent place to start is *A Student's Guide to Maxwell's Equations* by Fleisch (2008). That tutorial spends 130 pages covering what is presented here.

11.1.1 Physical Preliminaries: Electric and Magnetic Fields

The Lorentz equation for the force \mathbf{F} exerted on an electric charge q moving with velocity \mathbf{v} through an electric field \mathbf{E} and a magnetic field \mathbf{B} is (in SI units)

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}).$$

In this discussion, vectors in 3-D space are indicated by bold-faced symbols. The \times indicates the vector cross product. The Lorentz equation gives us the units for electric and magnetic fields. The force on the charge due to the electric field is $\mathbf{F} = q\mathbf{E}$, so the units of electric field must be

$$[E] = \frac{[F]}{[q]} = \frac{\text{newton}}{\text{coulomb}},$$

where [...] denotes “units of ...”. Similarly, magnetic fields have units of

$$[B] = \frac{[F]}{[qv]} = \frac{\text{newton}}{\text{coulomb meters per second}}.$$

You will see equivalent forms for these units. A newton per coulomb is the same as a volt per meter. An ampere is a current of a coulomb per second, so we can write $[B] = \text{N}/(\text{A m})$, which is called a Tesla (T). Table 11.1 summarizes for reference the quantities seen in Maxwell's equations.

The first two quantities in Table 11.1 are worthy of comment. The electric constant or permittivity of free space, ϵ_0 , is an empirical constant that measures an electric field's ability to "penetrate" a vacuum. In other words, it sets the strength of the force between two electric charges separated by some distance in a vacuum. This is seen if you write Coulomb's law as

$$F = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2},$$

where F is the magnitude of the force (in newtons) between charges q_1 and q_2 (in coulombs) separated by a distance r (in meters) in a vacuum. The value of ϵ_0 is not derived from fundamental physics; it must be measured. This can be done by measuring the force between two charges, but the measurement is more accurately made with a parallel plate capacitor. Similarly, the magnetic constant or permeability of free space, μ_0 , measures a magnetic field's ability to penetrate a vacuum. This is seen in Ampere's law, which when used to compute the strength of the magnetic field B (in Tesla) at distance r from a long, straight wire carrying a current I (amperes) gives

$$B(r) = \frac{\mu_0 I}{2\pi r}.$$

μ_0 also must be measured, e.g. by measuring the strength of the magnetic force between two current-carrying wires separated by some distance in a vacuum. Why do these two fundamental constants have the particular values shown in Table 11.1? This is a question like "why does an electron have the charge it has, and not some other value?" All that can be said is that these values are what they are because that is just how the universe works.

By the way, an electric field of 1 V/m is a very weak field: just think of a large parallel plate capacitor with the plates separated by 1 m and connected by a 1 V battery. The electric field between a thundercloud and the ground is of order 10^5 V/m just before a lightning discharge. On the other hand, a 1 T magnetic field is really strong¹. The Earth's magnetic field at the surface is about $5 \cdot 10^{-5}$ T.

11.1.2 Mathematical Preliminaries: Divergence and Curl

In order to enjoy Maxwell's equations, it is necessary to understand the mathematical notation. For the benefit of readers who are not familiar with vector calculus, the needed operations are as follows.

A *scalar field* $S(x, y, z, t)$ associates a number with each point in space and time. An example is the temperature in room. A *vector field* $\mathbf{V}(x, y, z, t) = \mathbf{V}(\mathbf{x}, t)$ associates a vector (a magnitude and a direction)

$$\mathbf{V}(x, y, z, t) = V_x(x, y, z, t)\hat{\mathbf{x}} + V_y(x, y, z, t)\hat{\mathbf{y}} + V_z(x, y, z, t)\hat{\mathbf{z}}$$

with each point in space and time. An example is the wind blowing outside your home.

¹Important research has shown that a 16 T magnetic field is so strong that it can overcome the force of gravity and levitate a living frog (Berry and Geim, 1997).

Physical quantity	Symbol	SI Units	Comment
Electric constant	ϵ_0	$\approx 8.8542 \cdot 10^{-12}$ $\text{A}^2 \text{s}^4 \text{kg}^{-1} \text{m}^{-3}$ (or $\text{C}^2 \text{N}^{-1} \text{m}^{-2}$)	measures a vacuum's ability to support an electric field; also called the permittivity of free space
Magnetic constant	μ_0	$\approx 1.2566 \cdot 10^{-6}$ $\text{kg m s}^{-2} \text{A}^{-2}$ (or N A^{-2})	measures a vacuum's ability to support a magnetic field; also called the permeability of free space
Electric charge	q	coulomb (C)	a fundamental physical quantity
Charge density	ρ	C m^{-3}	charge per unit volume
Electric current	\mathbf{I}	ampere ($\text{A} = \text{C/s}$)	measures flow of electric charge per unit time
Current density	\mathbf{J}	A m^{-2}	current per unit area
Electric field	\mathbf{E}	$\text{N/C} = \text{V/m}$	a vector field set up by stationary electric charges or time varying magnetic fields; acts on stationary electric charges
Magnetic field	\mathbf{B}	$\text{N}/(\text{A m}) = \text{T}$	a vector field set up by moving electric charges (currents) or by time-varying electric fields; acts on moving electric charges
Electric dipole moment	\mathbf{p}	C m	measures charge separation; direction is from negative to positive charge
Polarization	\mathbf{P}	$\text{C m}/\text{m}^3$	electric dipole moment per unit volume
Magnetic dipole moment	\mathbf{m}	A m^2	measures the magnetic field set up by a loop of current; direction is by a right-hand rule or from south pole to north
Magnetization	\mathbf{M}	$(\text{A m}^2)/\text{m}^3$	magnetic dipole moment per unit volume
Electric displacement	\mathbf{D}	C/m^2	$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P}$
Magnetic intensity	\mathbf{H}	A/m	$\mathbf{H} = \mathbf{B}/\mu_0 - \mathbf{M}$

Table 11.1: Quantities involved in Maxwell's equations. Other than the physical constants ϵ_0 and μ_0 , all quantities are functions of time and space, e.g., $\mathbf{E} = \mathbf{E}(\mathbf{x}, t) = \mathbf{E}(x, y, z, t)$.

The “del” operator ∇ (sometimes also called “nabla”) can be thought of as a vector whose elements are partial derivatives defined (in Cartesian coordinates) as

$$\nabla = \hat{\mathbf{x}} \frac{\partial}{\partial x} + \hat{\mathbf{y}} \frac{\partial}{\partial y} + \hat{\mathbf{z}} \frac{\partial}{\partial z}.$$

Applying the del operator to a scalar gives a vector, called the *gradient* of the scalar field:

$$\nabla S = \hat{\mathbf{x}} \frac{\partial S}{\partial x} + \hat{\mathbf{y}} \frac{\partial S}{\partial y} + \hat{\mathbf{z}} \frac{\partial S}{\partial z}.$$

Just like any vector, we can take the dot product of ∇ with a vector, and the result is a scalar. Taking the dot product of the del operator with a gradient gives a scalar:

$$\nabla \cdot \nabla S = \nabla^2 S = \frac{\partial^2 S}{\partial x^2} + \frac{\partial^2 S}{\partial y^2} + \frac{\partial^2 S}{\partial z^2}.$$

This is usually called the Laplacian of S , and ∇^2 is called the Laplace operator.

The *divergence* of a vector field is defined as

$$\nabla \cdot \mathbf{V} = \frac{\partial V_x}{\partial x} + \frac{\partial V_y}{\partial y} + \frac{\partial V_z}{\partial z}.$$

The cross product of two vectors $\mathbf{a} = a_x \hat{\mathbf{x}} + a_y \hat{\mathbf{y}} + a_z \hat{\mathbf{z}}$ and $\mathbf{b} = b_x \hat{\mathbf{x}} + b_y \hat{\mathbf{y}} + b_z \hat{\mathbf{z}}$ is

$$\mathbf{a} \times \mathbf{b} = (a_y b_z - a_z b_y) \hat{\mathbf{x}} + (a_z b_x - a_x b_z) \hat{\mathbf{y}} + (a_x b_y - a_y b_x) \hat{\mathbf{z}}.$$

In the same fashion we get the *curl* of a vector field, which is the cross product of ∇ with the vector field and yields a vector:

$$\nabla \times \mathbf{V} = \left(\frac{\partial V_z}{\partial y} - \frac{\partial V_y}{\partial z} \right) \hat{\mathbf{x}} + \left(\frac{\partial V_x}{\partial z} - \frac{\partial V_z}{\partial x} \right) \hat{\mathbf{y}} + \left(\frac{\partial V_y}{\partial x} - \frac{\partial V_x}{\partial y} \right) \hat{\mathbf{z}}.$$

There is a useful trick for remembering the order of the vector components and derivatives in the curl if you know how to expand the determinant of a 3×3 matrix. Write the unit direction vectors in the first row of the determinant, the partial derivatives in the second row, and the vector components in the third row:

$$\nabla \times \mathbf{V} = \begin{vmatrix} \hat{\mathbf{x}} & \hat{\mathbf{y}} & \hat{\mathbf{z}} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ V_x & V_y & V_z \end{vmatrix}.$$

Then expand the determinant just as though the elements were ordinary numbers, and let the derivatives operate on the vector elements.

Thus the divergence and curl are just certain combinations of the spatial derivatives of a vector field. Each has a physical interpretation when the vector field is a physical variable such as the velocity or an electric field. However, just knowing the definitions is sufficient for our level of presentation of Maxwell's equations.

11.1.3 Maxwell's Equations in a Vacuum

Without further ado, Maxwell's equations for the electric field $\mathbf{E}(\mathbf{x}, t)$ and magnetic field $\mathbf{B}(\mathbf{x}, t)$ in a vacuum are (in differential form, in SI units)

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0} \rho \quad (11.1)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (11.2)$$

$$\nabla \times \mathbf{E} = - \frac{\partial \mathbf{B}}{\partial t} \quad (11.3)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \quad (11.4)$$

Note that “in a vacuum” means that the electric and magnetic fields are in empty space. There can still be electric charges located here and there in space (the ρ term), and the same for currents (\mathbf{J}), which give rise to the fields in the region of interest.

These equations can be described as follows:

Eq.(11.1) This equation is called Gauss’s law for electric fields. It shows how electric charges (via the charge density ρ) create electric fields. This equation is the equivalent of Coulomb’s law for a point charge.

Eq.(11.2) This equation is sometimes called Gauss’s law for magnetic fields. It says that there are no magnetic charges corresponding to electric charges.

Eq.(11.3) This is Faraday’s law. It shows that a time-varying magnetic field creates an electric field.

Eq.(11.4) This is Ampere’s law as modified by Maxwell. The first term on the right, deduced by Ampere, shows that electric currents create magnetic fields. The second term on the right, added by Maxwell, shows that a time-varying electric field also creates a magnetic field.

Thus there are two ways to create electric fields: electric charges create them, and time-dependent magnetic fields create them. One might suppose that the electric fields resulting from these two entirely different creation mechanisms could some way be different, but they are not. An electric field is an electric field, no matter how it is created. That’s just the way the universe works. (Pondering this equivalence of electric fields, no matter how created, was one of the things that lead Einstein to the development of special relativity.) The same situation holds for magnetic fields. They can be created by electric currents or by time-dependent electric fields, but the nature of the magnetic field is the same in either case.

Simply stating Maxwell’s equations is really no different than simply stating Newton’s law of gravity for the magnitude of the force of attraction between two spherical masses M_1 and M_2 separated by a distance r :

$$F = G \frac{M_1 M_2}{r^2}. \quad (11.5)$$

Newton did not *derive* his law of gravity from more fundamental principles; it *is* the fundamental principle. Newton found that if he *assumed* Eq. (11.5) to be true, then he could derive Kepler’s laws of planetary motion, explain the motion of the moon, and (to first order) predict the ocean tides. The same can be said of Maxwell’s equations. They are

based on decades of observational work by Coulomb, Gauss, Faraday, Ampere and others, but we can view them as the mathematical statement of the fundamental laws governing electric and magnetic fields. We can simply accept these equations as given and get on with the business of applying them to problems of interest. (Of course, “fundamental” laws of nature may turn out to be imperfect in the light of new data. That happened to Newton’s law of gravity, which was replaced by, and can be derived from, Einstein’s theory of general relativity. Likewise, Maxwell’s equations can now be derived from the more fundamental laws of quantum electrodynamics developed by Feynman and others.)

It may at first glance seem that Maxwell’s equations are over-determined. That is, there are four equations but only two unknowns, \mathbf{E} and \mathbf{B} . This would be true for algebraic equations, in which case we could solve two linearly independent equations for two unknowns. However, for vector fields, Helmholtz’s theorem (also known as “the fundamental theorem of vector calculus”) says that an arbitrary vector field in 3 dimensions can be uniquely decomposed into a divergence part (with zero curl) and a curl part (with zero divergence) (under a few conditions, namely vector functions that are sufficiently smooth and that decay to zero at infinity). Conversely, knowing the divergence and curl of a vector field determines the vector field. That is the case here for both \mathbf{E} and \mathbf{B} . Given the charge density ρ and current density \mathbf{J} , the four Maxwell equations uniquely determine the electric and magnetic fields via their divergences and curls. (To be rigorous, a vector field is determined from its divergence and curl to within an additive term. This is somewhat like saying that knowing a derivative $df(x)/dx$ determines f to within an additive constant. Adding a boundary condition $f(x_o) = f_o$ then fixes the value of the constant.)

11.1.4 Light as an Electromagnetic Phenomenon

Starting with equations (11.1) to (11.4), Maxwell derived what is probably the most elegant and important result in the history of physics. Consider a region of space where there are no charges ($\rho = 0$) or currents ($\mathbf{J} = 0$). Equations (11.1)-(11.4) then become

$$\nabla \cdot \mathbf{E} = 0 \quad (11.6)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (11.7)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (11.8)$$

$$\nabla \times \mathbf{B} = \mu_o \epsilon_o \frac{\partial \mathbf{E}}{\partial t} \quad (11.9)$$

Now take the curl of Eq. (11.8), use the vector calculus identity $\nabla \times (\nabla \times \mathbf{E}) = \nabla(\nabla \cdot \mathbf{E}) - \nabla^2 \mathbf{E}$, use Eq. (11.6) to eliminate the $\nabla(\nabla \cdot \mathbf{E})$ term, and use Eq. (11.9) to rewrite the $\partial(\nabla \times \mathbf{B})/\partial t$ term. The result is

$$\nabla^2 \mathbf{E} = \mu_o \epsilon_o \frac{\partial^2 \mathbf{E}}{\partial t^2}.$$

The same process starting with the curl of Eq. (11.9) gives an equation of the same form for \mathbf{B} . Equations of the form

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = \frac{1}{v^2} \frac{\partial^2 f}{\partial t^2}$$

describe a wave propagating with speed v . Thus each component of \mathbf{E} and \mathbf{B} satisfies a wave equation with a speed of propagation

$$v = \frac{1}{\sqrt{\mu_0 \epsilon_0}}. \quad (11.10)$$

Inserting the experimentally determined values of μ_0 and ϵ_0 given in Table 11.1 gives $v = 3 \cdot 10^8 \text{ m s}^{-1}$. As Maxwell observed (Maxwell, 1864, page 466), “This velocity is so nearly that of light that it seems we have strong reason to conclude that light itself (including radiant heat and other radiations) is an electromagnetic disturbance in the form of waves propagated through the electromagnetic field according to electromagnetic laws.” *This conclusion is one of the greatest intellectual achievements of all time: not only were electric and magnetic fields tied together in Maxwell’s equations, but light itself was shown to be an electromagnetic phenomenon.* This is the first example of a “unified field theory,” in which seeming independent phenomena—here electric fields, magnetic fields, and light—were shown to be related and governed by the same underlying equations.

11.2 Maxwell’s Equations in Matter

This section considers electric and magnetic fields inside dielectric materials.

11.2.1 Dielectrics

We begin with the effects of electric fields on dielectrics. Dielectrics are materials that do not easily allow the flow of electric charge, so they are also called insulators. Dielectrics include materials like water, glass, wood, or plastic, but not metals, which easily conduct electricity.

The molecules making up many dielectrics have the center of the negative electric charge (due to the electrons surrounding the nuclei) offset slightly from the center of the positive charge (due to the protons in the nuclei). Such molecules are called polar molecules. (The molecule overall is of course electrically neutral.) This offset gives the molecule a dipole moment \mathbf{p} whose magnitude is defined as the product of the positive charge times the distance between the charge centers. By convention, the direction of the dipole moment vector points from the negative to the positive charge. For example, in the asymmetric water molecule, the electrons tend to cluster around the oxygen atom, leaving the center of the positive charge a bit toward the point between the two hydrogen atoms. A water molecule has a dipole moment of about $6 \cdot 10^{-30} \text{ C m}$.

For macroscopic volumes of matter, the combined effect of the molecular dipole moments is described by the net dipole moment per unit volume \mathbf{P} , which is called the polarization and has units of $(\text{C m})/\text{m}^3$. (Note that this use of the term “polarization” has nothing to do with the polarization of light.) If the molecules are randomly oriented as illustrated in Fig. 11.2(a), the molecular dipole moments in the different directions cancel out so that the net dipole moment of the substance is zero.

However, if the dielectric is placed in an external electric field, that field can cause the dipole moments to align so that the substance has a net dipole moment, or non-zero polarization \mathbf{P} , as illustrated in Fig 11.2(b). In this figure, the green symbols with

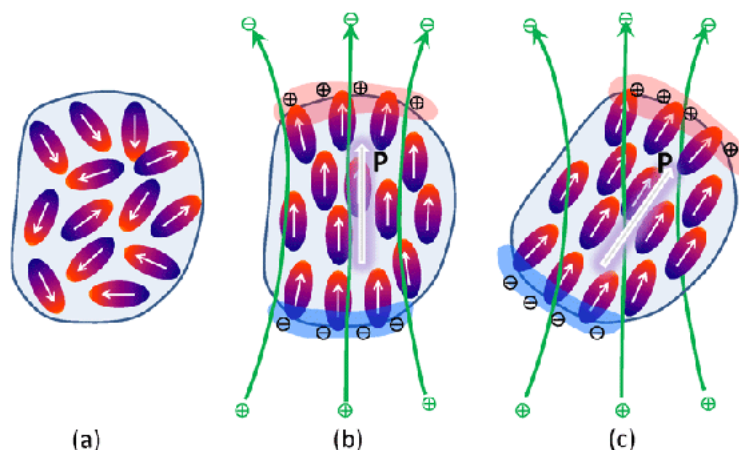


Figure 11.2: Concepts of electric polarization in dielectrics. The small red-blue ovals represent polar molecules, with the negative charge in blue and the positive charge in red. The small white arrows illustrate the molecular dipole moments \mathbf{p} . The large white arrow is the polarization \mathbf{P} . The green arrows represent an applied electric field. Panel (a) represents unpolarized matter. Panel (b) shows polarization induced by an applied electric field. Panel (c) represents an electret.

plus and minus signs represent positive and negative charges creating the external electric field, which is illustrated by the green arrows. The negative ends of the polar molecules are attracted to the positive charges creating the external field, and the positive ends to the negative external charges, so the molecules align as shown. (In practice, this is a complicated business. The applied electric field tends to align the polar molecules, but random thermal motions tend to randomize the directions. Thus, for a given material, \mathbf{P} depends on temperature. It takes time for the molecules to rotate into alignment, so if the applied field is not constant, \mathbf{P} depends on the frequency of the applied field. These details need not concern us here, but this is the origin of the frequency (wavelength) dependence of the index of refraction, for example.)

Note that as the molecules align in response to the external field, there is a net accumulation of positive charge on the surface of the material nearest the negative external charges, as illustrated in Fig 11.2(b) by the reddish area and the black symbols with plus signs. In essence, the positive “heads” of the molecules are sticking out of the volume of material. Likewise, there is an accumulation of negative charge on the opposite side of the material caused by the negative “tails” of the molecules, as illustrated by the bluish area and the black symbols with minus signs. In the interior of the matter, the net charge remains zero because nearby positive heads and negative tails cancel each other out. The external charges (the green + and - symbols in Fig. 11.2(b)) creating the applied field are called “free” charges because I am free to place them as desired in order to create the external electric field. The charges on the surface of the material are called “bound” charges because they are fixed to the material and cannot be moved around as desired.

The bound charges on the surface of the material create an electric field directed opposite to the applied field. Recall the bookkeeping: electric fields are by convention directed from positive to negative charges, but dipole moments, hence \mathbf{P} , are directed from negative

to positive charges. Thus we expect that the total electric field inside the material will be less than the externally applied electric field. This is why capacitors are filled with dielectric material. By decreasing the electric field between the plates of the capacitor, the capacitor can store more charge before a spark jumps from one plate to the other. The field set up by the bound charges will also distort the field of the free charges in the region outside the dielectric.

Even if the molecules or atoms of the dielectric are symmetric and have no dipole moment in the absence of an electric field, they will become polarized in the presence of an applied field. This is because the applied field tends to pull the negative electron cloud in one direction and the positive nucleus in the other direction, creating a dipole moment.

The situation just discussed is called induced polarization, because the polarization is induced by the applied external electric field. The direction of the polarization vector \mathbf{P} is then parallel to the applied field \mathbf{E} . This is not always the case. Some crystals have a permanent polarization due to their molecular structure. These substances are called “electrets” and are the electrical equivalent of permanent bar magnets. (Indeed, the name comes from “*electric magnets*”.) Examples are quartz crystals and barium titanate, BaTiO_3 . In this case, illustrated in Fig. 11.2(c), the polarization is not induced by the applied field and can be in any direction relative to the applied field.

11.2.2 Maxwell's Equations Inside Matter

To continue the development, recall for reference the forms of Maxwell's equations in vacuo:

$$\nabla \cdot \mathbf{E} = \frac{1}{\epsilon_0} \rho \quad (11.11)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (11.12)$$

$$\nabla \times \mathbf{E} = - \frac{\partial \mathbf{B}}{\partial t} \quad (11.13)$$

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J} + \mu_0 \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \quad (11.14)$$

The total electric field is equal to the field of the free charges plus the field of the bound charges. In practice, we control the free charges and can place them as desired, but not the bound charges, which are stuck to the material. In the most general case, there may even be bound charges induced within the material if the material is inhomogeneous so that \mathbf{P} varies with location. A general expression for the bound charge density is then (e.g., Griffiths, 1981, Section 4.2)

$$\rho_b = -\nabla \cdot \mathbf{P}.$$

We can now write the total charge density as the sum of the free charge density and the bound charge density:

$$\rho = \rho_f + \rho_b = \rho_f - \nabla \cdot \mathbf{P}.$$

Inserting this into Eq. (11.11) gives

$$\epsilon_0 \nabla \cdot \mathbf{E} = \rho_f - \nabla \cdot \mathbf{P}.$$

Defining the electric *displacement* \mathbf{D} as

$$\mathbf{D} = \epsilon_0 \mathbf{E} + \mathbf{P} \quad (11.15)$$

allows Eq. (11.11) to be rewritten as

$$\nabla \cdot \mathbf{D} = \rho_f. \quad (11.16)$$

The significance of this development warrants discussion. In Gauss's law, $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$, the charge density ρ includes all charges, both free and bound (if any), and the electric field \mathbf{E} is the total electric field caused by all charge density. This \mathbf{E} is the electric field either outside or inside a dielectric. It is \mathbf{E} that we normally want to know. Although we can define the free charge density ρ_f as desired, we may not know what the bound charge density is because that depends on the dielectric material. Determining ρ_b for a given material and applied electric field can be very difficult (this is what solid state physicists get paid to do). However, we can compute the displacement field \mathbf{D} given only the free charge density ρ_f , which we do know, without any knowledge of the nature of the dielectric material or its bound charge density. This gives us something to start with, using only what we know. Then, if we can determine the polarization \mathbf{P} (which may or may not be easy), we can obtain the electric field \mathbf{E} from Eq. (11.15).

In summary, we can avoid having to know the details of the dielectric properties by solving for something, the displacement field \mathbf{D} , that is not really what we want, but which is related to the electric field. Note that the displacement \mathbf{D} does not have the same units as \mathbf{E} , and may not even be in the same direction as \mathbf{E} (the situation of Fig. 11.2(c)). (There are other subtleties associated with \mathbf{D} . For example, there is no equivalent of Coulomb's law for \mathbf{D} , but those discussions are left to the textbooks.) Note also that the remaining three Maxwell's equations (11.12)-(11.14) remain unchanged because they do not involve the charge density.

We can repeat the above process for conducting materials such as metals. Now, instead of molecules or atoms having electric dipole moments, either inherent or induced, we consider molecules or atoms having magnetic dipole moments \mathbf{m} . Magnetic dipole moments are generated by current loops enclosing some area and have units of A m^2 . The small electric dipoles shown in Fig 11.2 are now replaced by small current loops, and instead of bound charges we have bound currents. The sum of these magnetic dipole moments is expressed by the *magnetization* \mathbf{M} , which is the magnetic dipole moment per unit volume, with units of $(\text{A m}^2)/\text{m}^3 = \text{A/m}$. As was seen above, the induced electrical polarization \mathbf{P} aligns with the applied electric field. However, the induced magnetization \mathbf{M} can align either with the applied magnetic field (paramagnetic materials) or opposite the applied field (diamagnetic materials). Materials with permanent magnetization are called ferromagnetic (an example is a common bar magnet).

The Ampere's law part of (11.14) shows how currents generate magnetic fields.

$$\nabla \times \mathbf{B} = \mu_0 \mathbf{J}. \quad (11.17)$$

Here the current density \mathbf{J} refers to all currents, and \mathbf{B} is the magnetic field either outside or inside matter. In addition to free currents \mathbf{J}_f and bound currents \mathbf{J}_b , in analogy with free and bound charges, there will also be a current associated with a time-dependent polarization. A change in the electrical polarization of a material implies moving charges around, which gives a "polarization" current. We can thus partition the total current density into three parts:

$$\mathbf{J} = \mathbf{J}_f + \mathbf{J}_b + \mathbf{J}_p.$$

The bound currents can be either on the surface on an object, or within the object if the magnetization is non uniform. In general (e.g., Griffiths, 1981, Section 6.2) the bound current and magnetization are related by

$$\mathbf{J}_b = \nabla \times \mathbf{M}.$$

The polarization current is given by

$$\mathbf{J}_p = \frac{\partial \mathbf{P}}{\partial t}.$$

The Ampere-Maxwell law (11.14) can now be rewritten as

$$\begin{aligned} \frac{1}{\mu_0} \nabla \times \mathbf{B} &= \mathbf{J} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \\ &= \mathbf{J}_f + \mathbf{J}_b + \mathbf{J}_p + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t} \\ &= \mathbf{J}_f + \nabla \times \mathbf{M} + \frac{\partial \mathbf{P}}{\partial t} + \epsilon_0 \frac{\partial \mathbf{E}}{\partial t}. \end{aligned}$$

Defining the *magnetic intensity* \mathbf{H} by

$$\mathbf{H} = \frac{1}{\mu_0} \mathbf{B} - \mathbf{M} \quad (11.18)$$

and recalling the definition (11.15) of \mathbf{D} then gives

$$\nabla \times \mathbf{H} = \mathbf{J}_f + \frac{\partial \mathbf{D}}{\partial t}.$$

In this equation, we have hidden our ignorance about the magnetic properties of the material inside the \mathbf{H} term. Note that the magnetic intensity \mathbf{H} does not have the same units as the magnetic field \mathbf{B} , and in general the two are not even in the same direction. However, we can solve for \mathbf{H} given only the free current density \mathbf{J}_f and the displacement \mathbf{D} , which depends only on the free charge density ρ_f . Then, if we are smart enough to figure out \mathbf{M} for the material at hand, we can extract the desired \mathbf{B} from Eq. (11.18).

We now have Maxwell's equations in the form usually seen in discussions of arbitrary materials:

$$\nabla \cdot \mathbf{D} = \rho_f \quad (11.19)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (11.20)$$

$$\nabla \times \mathbf{E} = - \frac{\partial \mathbf{B}}{\partial t} \quad (11.21)$$

$$\nabla \times \mathbf{H} = \mathbf{J}_f + \frac{\partial \mathbf{D}}{\partial t} \quad (11.22)$$

These equations are sometimes called the “macroscopic” Maxwell equations, because they hold inside and outside of macroscopic amounts of matter large enough for the spatial averages underlying \mathbf{D} and \mathbf{H} to be defined. The original equations (11.11)-(11.14) are then called the “microscopic” Maxwell equations because they hold at even the smallest spatial scales, even between the atoms within materials. Equations (11.15) and (11.18) are called *constitutive* equations.

It may seem that we now have four equations in four unknowns, but Eqns. (11.15) and (11.18) make clear that all we have done is rewrite the original four equations with their two unknowns. In order to solve these equations, we must specify the free charge density ρ_f and the free current density \mathbf{J}_f along with boundary conditions on the fields.

11.2.3 Comments on Terminology

A word of warning is needed regarding terms and units for magnetic fields. I grew up with \mathbf{B} called the magnetic field and \mathbf{H} the magnetic intensity, and I'm too old to change. However, the venerable Jackson (1962), the standard graduate-level text on electromagnetic theory for over 60 years, calls \mathbf{H} the magnetic field and \mathbf{B} the magnetic induction. The equally competent and respected Griffiths (1981) calls \mathbf{B} the magnetic field and \mathbf{H} the auxiliary field. Griffiths (page 232) says calling \mathbf{B} the magnetic induction is “an absurd choice” because it leads to confusion with electromagnetic induction, which is something totally different. He also says, “ \mathbf{H} has no sensible name; just call it ‘ \mathbf{H} ’”. The great Arnold Sommerfeld, doctoral or postdoctoral adviser to seven students who later won Nobel prizes, says (Sommerfeld, 1952, page 45) that “The unhappy term ‘magnetic field’ for \mathbf{H} should be avoided as far as possible.” The thing to keep in mind is that \mathbf{E} and \mathbf{B} are the fundamental quantities; \mathbf{D} and \mathbf{H} arise from rewriting the fundamental equations (11.6) to (11.9) in forms convenient for material media. To make matters even worse, the units of \mathbf{B} and \mathbf{H} , whatever you call them, depend on the system of fundamental units chosen. Here I have used SI units (previously called “rationalized mks” units). \mathbf{B} and \mathbf{H} then have different units, as we have seen. However, in the cgs (centimeter-gram-second) system, \mathbf{H} is defined by $\mathbf{H} = \mathbf{B} - \mathbf{M}/(\epsilon_0 c^2)$, where c is the speed of light. This is convenient because \mathbf{B} and \mathbf{H} then have the same units of Tesla, but then Maxwell's equations are slightly different. In cgs, the Ampere-Maxwell equation (11.22) reads $\epsilon_0 c^2 \nabla \times \mathbf{H} = \mathbf{J}_f + \partial \mathbf{D} / \partial t$. This whole business is a confusing mess; see the discussion in Feynman et al. (1964, Section 36.2).

This qualitative discussion is sufficient for most oceanographers. If nothing else, you can now hold your own at a cocktail party full of undergraduate physics majors. But in all seriousness, it is worth keeping in mind the importance of Maxwell's synthesis. He tied together apparently unrelated results describing electric and magnetic fields, finished off Ampere's law, and showed that light of any frequency is a propagating electromagnetic wave. Everything in the modern world that has to do with electricity or magnetism or optics rests upon these equations—generation of electrical current in hydroelectric plants, the use of that current in electric motors, radios, televisions, mobile phones, computers, microwave ovens, lasers, and so on *ad infinitum*. In oceanography, the well known Fresnel equations (Section 13.3) for the reflection and transmission of light at the sea surface arises from solving Eqs. (11.19)-(11.22) for a plane wave incident onto an interface between two dielectric media. The widely used Mie Theory (Chapter 12) for the scattering of light by spherical particles is the solution of these equations for a plane wave incident onto a spherical dielectric.

You would have to return to the early 1800s—when lighting was by candles and whale oil lamps, communication was by handwritten letters delivered by riders on horses, and calculations were done with a quill pen and paper—to live in a world without applications of Maxwell's equations.

In addition to his work on electromagnetism, Maxwell made major contributions to thermodynamics (the Maxwell relations connecting temperature, entropy, pressure, and volume) and statistical mechanics (the Maxwell-Boltzmann distribution). He developed an explanation for the dynamics of Saturn's rings that is still used today. He even made the first color photographs. Although less known to the general public, Maxwell stands with Newton, Einstein, and Darwin as one of the most profound intellects the world has ever

known.

11.3 Plane Wave Solutions of Maxwell's Equations

Now that we have been introduced to Maxwell's Equations in the previous two sections, we can attempt to solve them. Textbooks on classical electrodynamics (e.g. Jackson, 1962; Griffiths, 1981; Bohren and Huffman, 1983) take a general approach of assuming little and letting Maxwell's equations force upon you conclusions about what functional forms of solutions are possible for propagating waves. This is not a physics text, so I will propose simple forms for propagating electric and magnetic fields and show that they satisfy Maxwell's equations. That approach is sufficient to show with a minimum of math and physics how waves can propagate in dielectrics like water. In so doing, we will discover the relation between the absorption coefficient and the imaginary parts of the index of refraction and the wavenumber. The discussion of plane waves in this section sets the stage for a deeper investigation of wave propagation in the section on dispersion.

Before proceeding, we must again face the never-ending problem of notation. In Section 8.2.1, I wrote the complex index of refraction as $m = n + ik$, where n was the real index of refraction and k was called the imaginary part of the index of refraction. It is also common to use k as the angular wavenumber: $k = 2\pi/\lambda$ (Section 1.5.9). We will also need to introduce a complex wavenumber. Every author seems to have a slightly different way to avoid using the same k symbol for two different quantities. Bohren and Huffman (1983) use a Roman k for the complex wavenumber, which they write as $k = k' + ik''$, and they use an italic k for the imaginary part of the index of refraction. They also use N rather than m for the complex index of refraction. Thus my $m = n + ik$ is Bohren and Huffman's $N = n + ik$. (Griffiths, 1981) uses κ_+ for the real part of the complex wavenumber and κ_- for the complex part; he writes the complex wavenumber as $\kappa = \kappa_+ + i\kappa_-$. Mishchenko et al. (2002) write $k = k_R + ik_I$ and $m = m_R + im_I$. For this section, I have chosen to use k for the imaginary part of the index of refraction and $\tilde{\kappa}$ for the complex wavenumber, with a single prime on the real part and a double prime on the imaginary part: $\tilde{\kappa} = \kappa' + i\kappa''$. You will similarly see the complex index of refraction written as $n = n' + in''$. To further complicate matters, some authors write the time dependence in Eq. (11.23) as $+\omega t$, in which case the complex index of refraction (in my notation) becomes $m = n - ik$ and the complex wavenumber is $\tilde{\kappa} = \kappa' - i\kappa''$. The whole notation business is a mess, and you just have to figure out each author's preferences.

11.3.1 Plane Waves in a Dielectric

Other than electromagnetic waves propagating in a vacuum, the simplest solution of Maxwell's equations for wave propagation is for a plane wave in a dielectric material. The term "plane wave" refers to an electromagnetic wave (i.e., light) that is propagating in some direction, and which has the same properties at all points of a plane perpendicular to the direction of propagation. As will be seen, if we know the electric field of the wave, we can find the magnetic field from Maxwell's equations (or vice versa). Thus it is customary to consider only the electric field of the wave.

Consider first an electromagnetic wave propagating in a vacuum. We are free to choose a convenient coordinate system, so let the wave propagate in the $+\hat{\mathbf{x}}$ direction, and let the electric field \mathbf{E} oscillate in the $\pm\hat{\mathbf{y}}$ direction (i.e., the electric field is linearly polarized in

the y direction). The associated magnetic field then oscillates in the $\pm\hat{z}$ direction. This situation is illustrated in Fig. 11.3.

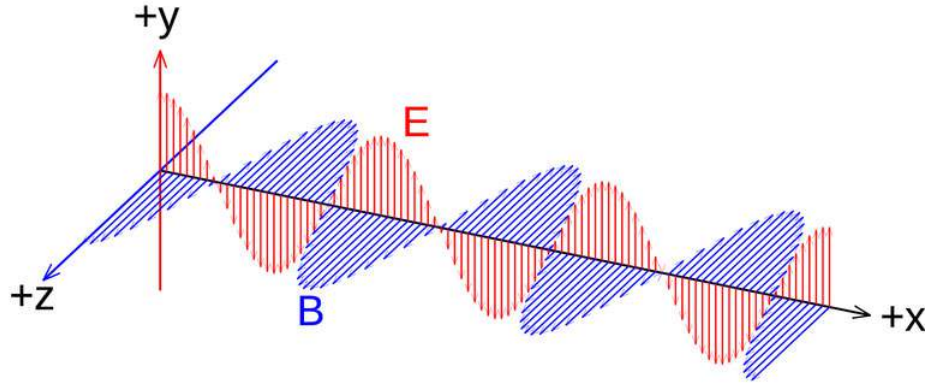


Figure 11.3: Illustration of sinusoidal electric and magnetic plane waves in a vacuum. The red arrows represent the electric field and the blue arrows are the magnetic field.

We can write the magnitude of the electric field as

$$E(x, t) = E_o \cos(\kappa'x - \omega t + \phi), \quad (11.23)$$

where E_o sets the magnitude of the electric field, $\kappa' = 2\pi/\lambda$ is the angular wavenumber, and $\omega = 2\pi/T$ is the angular frequency; λ is the wavelength and T is the period of the oscillation. For propagation in a vacuum, $\lambda = cT$; i.e., a wave propagating at the speed of light c travels a distance λ in one period of the oscillation. It is common to use k rather than κ' for the angular wavenumber (recall Section 1.5.9); the reason for my choice of κ' was explained above.

Electric fields, wavenumbers, and frequencies are of course real quantities. However, it will prove to be convenient to write the electric field of Eq. (11.23) as the real part of a complex quantity:

$$E(x, t) = \Re\{E_o \exp i(\kappa'x - \omega t + \phi)\},$$

where $\Re\{\dots\}$ stands for the real part of the argument, and we recall that $e^{i\theta} = \cos\theta + i\sin\theta$; $i = \sqrt{-1}$. Some authors use a subscript c to indicate a complex field, and some use a tilde, and some leave it to the reader to figure out from the context which variables are real and which are complex. I will use a tilde and write

$$\begin{aligned} \tilde{E}(x, t) &= E_o \exp i(\kappa'x - \omega t + \phi) \\ &= \tilde{E}_o \exp i(\kappa'x - \omega t) \end{aligned} \quad (11.24)$$

In the second equation, the phase angle has been incorporated into the amplitude so that $\tilde{E}_o = E_o \exp(i\phi)$. Now $E = \Re\{\tilde{E}\}$, and so on. It is customary to omit writing the \Re symbol, in which case it is understood that at the end of any calculation involving complex numbers we must take the real part to get back to a real physical quantity. Some authors call \tilde{E} the “complex electric field”, but a better terminology is to say that \tilde{E} is the complex representation of the real electric field E .

We can rewrite (11.24) as

$$\tilde{E}(x, t) = \tilde{E}_o \exp i\kappa'(x - \frac{\omega}{\kappa'}t).$$

It is easily shown by substitution that any function of the form $f[s(x - vt)]$, where s is a constant scale factor, satisfies the 1-D wave equation

$$\frac{\partial^2 f(x, t)}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 f(x, t)}{\partial t^2}.$$

Thus any function of the form $f[s(x - vt)]$ describes a function with shape $f(sx)$ at time $t = 0$ propagating in the $+x$ direction with speed v . We can thus identify the speed

$$v_p = \frac{\omega}{\kappa'} \quad (11.25)$$

as the speed of the electromagnetic wave. This speed is known as the *phase speed* of the sinusoidal wave because it shows how fast a point with a given phase of the sinusoidal wave, say the wave crest, travels.

A function of the form $\tilde{E}_o \exp i(\kappa'x - \omega t)$ propagates with an amplitude \tilde{E}_o that is independent of time and location. That is correct for propagation in a vacuum. However, in absorbing matter, the Bouguer-Lambert-Beer law says that the irradiance attenuates exponentially with distance according to $\exp(-ax)$, where a is the absorption coefficient. We can build this attenuation into the electric field in an ad hoc manner by replacing the constant amplitude of the electric field with a function of x with the form

$$\tilde{E}_o \leftarrow \tilde{E}_o \exp(-\kappa''x),$$

where κ'' is a positive constant. (The \leftarrow means “is replaced by”.) Equation (11.24) is then replaced by

$$\begin{aligned} \tilde{E}(x, t) &= \tilde{E}_o \exp(-\kappa''x) \exp i(\kappa'x - \omega t) \\ &= \tilde{E}_o \exp i[(\kappa' + i\kappa'')x - \omega t] = \tilde{E}_o \exp i(\tilde{\kappa}x - \omega t). \end{aligned} \quad (11.26)$$

Here $\tilde{\kappa} = \kappa' + i\kappa''$ is called the *complex wavenumber*.

Now let us turn the electric and magnetic field magnitudes into vectors so we can insert them into Maxwell's equations. In Fig. 11.3 the electric field oscillates in the y plane, and the magnetic field lies in the z plane. So for these fields we can write

$$\tilde{\mathbf{E}}(x, t) = \tilde{E}_o \exp i(\tilde{\kappa}x - \omega t) \hat{\mathbf{y}} \quad (11.27)$$

$$\tilde{\mathbf{B}}(x, t) = \tilde{B}_o \exp i(\tilde{\kappa}x - \omega t) \hat{\mathbf{z}} \quad (11.28)$$

At this point I have proposed two traveling waves for the electric and magnetic fields. Now we can ask, “under what conditions do waves of the form of Eqs. (11.27) and (11.28) satisfy Maxwell's equations?”

For reference, Maxwell's equations in matter are

$$\nabla \cdot \mathbf{D} = \rho_f \quad (11.29)$$

$$\nabla \cdot \mathbf{B} = 0 \quad (11.30)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \quad (11.31)$$

$$\nabla \times \mathbf{H} = \mathbf{J}_f + \frac{\partial \mathbf{D}}{\partial t} \quad (11.32)$$

with the constitutive relations

$$\mathbf{D} = \epsilon_o \mathbf{E} + \mathbf{P} \quad (11.33)$$

$$\mathbf{H} = \frac{1}{\mu_o} \mathbf{B} - \mathbf{M}. \quad (11.34)$$

For a linear medium, \mathbf{P} is proportional to \mathbf{E} and is commonly written $\mathbf{P} = \epsilon_o \chi \mathbf{E}$, where χ is the electric susceptibility (i.e., χ measures how susceptible the material is to being polarized by an electric field). χ is assumed to be independent of \mathbf{E} and is independent of location and direction in a homogeneous, isotropic medium. Thus $\mathbf{D} = \epsilon_o(1 + \chi)\mathbf{E}$. The quantity $\epsilon = \epsilon_o(1 + \chi)$ is the permittivity of the medium. If we assume that there are no free electric charges in the medium ($\rho_f = 0$), the first of Maxwell's equations (Eq. 11.29) is $\nabla \cdot \mathbf{D} = 0$. Inserting the complex representation $\tilde{\mathbf{D}} = \epsilon \tilde{\mathbf{E}}$ into $\nabla \cdot \tilde{\mathbf{D}}$ gives

$$\begin{aligned} \nabla \cdot \tilde{\mathbf{D}} &= \left\{ \frac{\partial}{\partial x} \hat{\mathbf{x}} + \frac{\partial}{\partial y} \hat{\mathbf{y}} + \frac{\partial}{\partial z} \hat{\mathbf{z}} \right\} \cdot \epsilon \tilde{E}_o \exp i(\tilde{\kappa}x - \omega t) \hat{\mathbf{y}} \\ &= \epsilon \left[i\tilde{\kappa} \tilde{E}_o \exp i(\tilde{\kappa}x - \omega t) \hat{\mathbf{x}} \cdot \hat{\mathbf{y}} + 0 + 0 \right] = 0 \end{aligned}$$

after noting that $\hat{\mathbf{x}} \cdot \hat{\mathbf{y}} = 0$ and that the y and z derivatives are 0 because the wave varies only in x . Thus the form of Eq. (11.27) for $\tilde{\mathbf{E}}(x, t)$ satisfies $\nabla \cdot \tilde{\mathbf{D}} = 0$. Likewise, the form of Eq. (11.28) for $\tilde{\mathbf{B}}(x, t)$ satisfies $\nabla \cdot \tilde{\mathbf{B}} = 0$.

Next insert the proposed $\tilde{\mathbf{E}}(x, t)$ and $\tilde{\mathbf{B}}(x, t)$ into Eq. (11.31). The result is

$$i\tilde{\kappa} \tilde{E}_o \exp i(\tilde{\kappa}x - \omega t) \hat{\mathbf{z}} = i\omega \tilde{B}_o \exp i(\tilde{\kappa}x - \omega t) \hat{\mathbf{z}}$$

This equation is satisfied only if

$$\tilde{B}_o = \frac{\tilde{\kappa}}{\omega} \tilde{E}_o. \quad (11.35)$$

Thus, given the electric field, we can compute the magnetic field. Taking the real part of this equation gives

$$B_o = \frac{\kappa'}{\omega} E_o = \frac{1}{v_p} E_o = \frac{n}{c} E_o \quad (11.36)$$

after recalling the definition of the phase speed from Eq. (11.25), and recalling that the speed of light in a medium with real index of refraction n is the speed in vacuo divided by n .

For a dielectric, the magnetization \mathbf{M} is zero, so $\mathbf{H} = \frac{1}{\mu} \mathbf{B}$, where μ is the permeability of the medium. (In general one can write $\mu = \mu_o(1 + \chi_m)$, where χ_m is the magnetic susceptibility of the medium. In a dielectric, μ is very nearly equal to μ_o , the permeability

of a vacuum.) If there are no free currents ($\mathbf{J}_f = 0$), the remaining Maxwell equation 11.32 reduces to

$$\nabla \times \frac{1}{\mu} \mathbf{B} = \frac{\partial}{\partial t} \epsilon \mathbf{E}$$

Inserting the complex representations into this equation gives

$$-i \frac{\tilde{\kappa}}{\mu} \tilde{B}_o \hat{\mathbf{y}} = -i \omega \epsilon \tilde{E}_o \hat{\mathbf{y}}$$

or

$$\tilde{B}_o = \frac{\omega}{\tilde{\kappa}} \mu \epsilon \tilde{E}_o$$

This result combined with that of Eq. (11.35) implies that

$$\tilde{\kappa}^2 = \omega^2 \mu \epsilon$$

or

$$\tilde{\kappa} = \omega \sqrt{\mu \epsilon}.$$

When the wave equation is derived starting with Maxwell's equations in matter, the wave speed is $v = 1/\sqrt{\mu \epsilon}$, rather than $v = c = 1/\sqrt{\mu_o \epsilon_o}$ as was seen in Eq. (11.10) for the vacuum case. The speed of light in a medium with real index of refraction n is c/n . Given that $\tilde{\kappa}$ in the last equation is complex (which implies that μ and/or ϵ must also be complex, as will be seen in the discussion of dispersion), we can introduce the complex index of refraction $m = n + ik$ and write

$$\tilde{\kappa} = \omega \sqrt{\mu \epsilon} = \frac{\omega m}{c} = \frac{2\pi m}{\lambda}, \quad (11.37)$$

after recalling that ω/c is the free-space wavenumber $2\pi/\lambda$.

We can now rewrite the complex representation of the electric field as

$$\tilde{\mathbf{E}}(x, t) = \tilde{E}_o \exp i \left(\frac{2\pi m x}{\lambda} - \omega t \right) \hat{\mathbf{y}} \quad (11.38)$$

$$= \tilde{E}_o \exp \left(\frac{-2\pi k x}{\lambda} \right) \exp i \left(\frac{2\pi n x}{\lambda} - \omega t \right) \hat{\mathbf{y}}, \quad (11.39)$$

Taking the real part gets us back to the real electric field:

$$\mathbf{E}(x, t) = E_o \exp \left(\frac{-2\pi k x}{\lambda} \right) \cos \left(\frac{2\pi n x}{\lambda} - \omega t + \phi \right) \hat{\mathbf{y}}. \quad (11.40)$$

The corresponding equation for $\tilde{\mathbf{B}}$ is

$$\mathbf{B}(x, t) = \frac{1}{v_p} E_o \exp \left(\frac{-2\pi k x}{\lambda} \right) \cos \left(\frac{2\pi n x}{\lambda} - \omega t + \phi \right) \hat{\mathbf{z}}, \quad (11.41)$$

or an equivalent based on Eq. (11.36).

We have now shown that sinusoidal plane waves can propagate through a dielectric provided that the magnitude of the magnetic field is proportional to the magnitude of the electric field according to Eq. (11.36), and that the wavenumber is related to the angular frequency according to Eq. (11.37).

11.3.2 The Absorption Coefficient

The equations above are for electric and magnetic fields, which are what physicists like to play with. Oceanographers, however, almost always work with irradiance. For the moment, let \mathcal{E} be irradiance, rather than the usual E , to avoid confusion with the electric field.

The Poynting vector is defined by

$$\mathbf{S} \triangleq \frac{1}{\mu} \mathbf{E} \times \mathbf{B}.$$

This vector points in the direction of wave propagation, and it has units of $\text{Js}^{-1}\text{m}^{-2}$, i.e. of irradiance. The Poynting vector thus describes the irradiance of the propagating electromagnetic wave in the medium. Inserting \mathbf{E} and \mathbf{B} from Eqs. (11.40) and (11.41) gives

$$\mathbf{S}(x, t) = \sqrt{\frac{\epsilon}{\mu}} E_o^2 \exp\left(\frac{-4\pi kx}{\lambda}\right) \cos^2\left(\frac{2\pi nx}{\lambda} - \omega t + \phi\right) \hat{\mathbf{x}},$$

after using Eq. (11.36), $v = 1/\sqrt{\mu\epsilon}$, and $\hat{\mathbf{y}} \times \hat{\mathbf{z}} = \hat{\mathbf{x}}$. This is the instantaneous irradiance of the wave at time t , which cannot be measured by an instrument at optical frequencies of order 10^{14} Hz. What is measured is the time-average of $\mathbf{S}(x, t)$ over many wave cycles. Recalling that the average of the cosine squared over a wave period is $1/2$ gives

$$\mathcal{E}(x) = \langle \mathbf{S}(x) \rangle = \frac{1}{2} \sqrt{\frac{\epsilon}{\mu}} E_o^2 \exp\left(\frac{-4\pi kx}{\lambda}\right) \hat{\mathbf{x}}. \quad (11.42)$$

Here $\mathcal{E}(x) = \mathcal{E}(x)\hat{\mathbf{x}}$. The thing to note in this equation is that the irradiance is proportional to the square of the electric field amplitude, and that its magnitude it damps out exponentially as

$$\mathcal{E}(x) = \mathcal{E}(0) \exp\left(\frac{-4\pi kx}{\lambda}\right) = \mathcal{E}(0) \exp(-ax).$$

This is the Bouguer-Lambert-Beer law as derived from Maxwell's equations². In the last equation we have identified the usual absorption coefficient a as

$$a = \frac{4\pi k}{\lambda}. \quad (11.43)$$

You will also see the absorption coefficient written as

$$a = 2\kappa'', \quad (11.44)$$

which follows from Eq. (11.26).

11.3.3 Generalizations

I have made a number of simplifications in the preceding development, which should be noted for completeness.

²This may look like circular reasoning because the exponential decay was built into the solution in Eq. (11.26) when we allowed for in imaginary part of the wavenumber. What was not seen in that equation was that the oceanographers' absorption coefficient is related to the complex index of refraction k by Eq. (11.43) and to the imaginary part of the wavenumber, κ'' , by Eq. (11.44)

I assumed that the electric field was in the $\hat{\mathbf{y}}$ plane and the magnetic field was in the $\hat{\mathbf{z}}$ plane, and that both were perpendicular to the direction of propagation and to each other. In a more general treatment, you find that Maxwell's equations can be satisfied for a plane wave only if \mathbf{E} and \mathbf{B} are perpendicular to the direction of propagation, i.e., electromagnetic plane waves are transverse waves. Furthermore, in a vacuum or in a homogeneous, isotropic, linear medium like water or glass, \mathbf{E} and \mathbf{B} are also perpendicular to each other. (A linear medium means that \mathbf{D} and \mathbf{H} are proportional to \mathbf{E} and \mathbf{B} , respectively.) The choice of \mathbf{E} in the $\hat{\mathbf{y}}$ plane was arbitrary. In general, you can write

$$\tilde{\mathbf{E}} = \tilde{\mathbf{E}}_0 \exp i(\tilde{\boldsymbol{\kappa}} \cdot \mathbf{x} - \omega t),$$

where \mathbf{x} and $\tilde{\boldsymbol{\kappa}}$ are arbitrary directions in space, and the direction of the electric field is contained in the amplitude vector $\tilde{\mathbf{E}}_0$. It is then found that

$$\tilde{\boldsymbol{\kappa}} \cdot \tilde{\mathbf{E}}_0 = \tilde{\boldsymbol{\kappa}} \cdot \tilde{\mathbf{B}}_0 = \tilde{\mathbf{E}}_0 \cdot \tilde{\mathbf{B}}_0 = 0.$$

That is, these three vectors are mutually perpendicular. The magnetic field can then be written as

$$\tilde{\mathbf{B}} = \frac{1}{v_p} \tilde{\boldsymbol{\kappa}} \times \tilde{\mathbf{E}}$$

My example was just a special case of these equations with $\tilde{\boldsymbol{\kappa}} = \tilde{\kappa} \hat{\mathbf{x}}$ and $\mathbf{x} = \hat{\mathbf{x}}$.

The electric and magnetic fields are perpendicular for plane waves propagating in a dielectric. However, they are not always perpendicular, for example in the 3-D fields created by scattering by a particle or radiated by an antenna. Similarly, although the electric and magnetic fields are in phase in a dielectric, they are out of phase by as much as 45 deg in a conducting medium. When decomposing electric and magnetic fields using Fourier transforms, the \mathbf{E} and \mathbf{B} fields of each Fourier mode are orthogonal, however the total fields resulting from a sum of modes may not be orthogonal. That is to say, if \mathbf{E}_1 and \mathbf{B}_1 are orthogonal (i.e., $\mathbf{E}_1 \cdot \mathbf{B}_1 = 0$), and \mathbf{E}_2 and \mathbf{B}_2 are orthogonal, it does not follow that $(\mathbf{E}_1 + \mathbf{E}_2) \cdot (\mathbf{B}_1 + \mathbf{B}_2) = 0$.

The complex representation of the Poynting vector is

$$\tilde{\mathbf{S}} = \frac{1}{\mu} \tilde{\mathbf{E}} \times \tilde{\mathbf{B}} = \tilde{\mathbf{E}} \times \tilde{\mathbf{H}}.$$

To get back to the real version, the formula is

$$\mathbf{S} = \Re\{\tilde{\mathbf{E}}\} \times \Re\{\tilde{\mathbf{H}}\} = \Re\{\tilde{\mathbf{E}} \times \tilde{\mathbf{H}}^*\},$$

where $\tilde{\mathbf{H}}^*$ is the complex conjugate of $\tilde{\mathbf{H}}$. Note that these formulas are not the same as $\Re\{\tilde{\mathbf{E}} \times \tilde{\mathbf{H}}\}$.

11.4 Dispersion: Phase and Group Speeds

The preceding section discussed wave propagation for a sinusoidal wave with given spatial (κ') and temporal (ω) angular frequencies. In that case, there is a single speed of propagation, the phase speed $v_p = \omega/\kappa'$. I will now revert to using the customary k rather than κ' for the real wavenumber, $k = 2\pi/\lambda$. However a true sinusoidal wave cannot exist in nature because real waves must start and stop at finite times. This in turn means that the finite

length “wave packet” must contain more than one frequency. (This is easily seen using Fourier transforms. The Fourier transform of a finite-length spatial function contains all frequencies.) It is then possible that the overall wave packet or group of waves propagates at a different speed, called the group speed, than the phase speeds of the individual waves comprising the wave packet. This results in some waves “outrunning” others, and the original group of waves spreads out, or disperses.

This section examines dispersion in detail. The basic phenomenon of dispersion and different phase and group speeds is illustrated first by the sum of just two sinusoids. This is then generalized to wave packets containing all frequencies. This allows for an understanding of “normal” and “anomalous” dispersion. A specific example of phase and group speeds is then given for light wave propagation in pure water, which leads to a surprise.

11.4.1 Phase and Group Speeds for a Two-component System

Let ψ be the magnitude of a propagating disturbance. ψ might be the electric field of a light wave, or the height of the sea surface relative to the mean sea surface, or the fluctuating pressure of a sound wave. Suppose that ψ is composed of two cosines of equal amplitudes but differing wavenumbers and temporal frequencies. That is to say

$$\psi(x, t) = A \cos(k_1 x - \omega_1 t) + A \cos(k_2 x - \omega_2 t). \quad (11.45)$$

The trigonometric identity

$$\cos \alpha + \cos \beta = 2 \cos \left(\frac{\alpha - \beta}{2} \right) \cos \left(\frac{\alpha + \beta}{2} \right)$$

can be used to rewrite ψ as

$$\psi(x, t) = 2A \cos\left[\frac{1}{2}(\Delta k x - \Delta \omega t)\right] \cos(\bar{k} x - \bar{\omega} t),$$

where $\Delta k = k_2 - k_1$, $\Delta \omega = \omega_2 - \omega_1$, $\bar{k} = \frac{1}{2}(k_1 + k_2)$, and $\bar{\omega} = \frac{1}{2}(\omega_1 + \omega_2)$. This can be rewritten as

$$\psi(x, t) = 2A \cos\left[\frac{1}{2}\Delta k \left(x - \frac{\Delta \omega}{\Delta k} t\right)\right] \cos\left[\bar{k} \left(x - \frac{\bar{\omega}}{\bar{k}} t\right)\right]$$

Recalling that a function of the form $f[s(x - vt)]$ propagates with speed v , we see that the two cosines describe waves that propagate with speeds

$$= 2A \cos\left[\frac{1}{2}\Delta k (x - v_g t)\right] \cos\left[\bar{k} (x - v_p t)\right],$$

where we identify the two speeds as

$$v_p = \frac{\bar{\omega}}{\bar{k}}, \quad \text{and} \quad (11.46)$$

$$v_g = \frac{\Delta \omega}{\Delta k}. \quad (11.47)$$

v_p has the form of the phase speed of the previous section, computed now from the average wavenumber and average angular frequency. v_g is called the *group speed* and is computed as the ratio of the differences in the two wavenumbers and frequencies.

The implications of these equations are illustrated in Figs. 11.4–11.6. In the first figure the wavenumbers and frequencies are chosen so that both the individual waves (the

red curves) and the group envelope (the blue dots) propagate in the same direction and $v_p > v_g$. In the second figure the wavenumbers and frequencies are chosen so that both the waves and the group envelope propagate in the same direction but with $v_p < v_g$. In the third figure the wavenumbers and frequencies are chosen so that the waves propagate in one direction but the group envelope propagates in the opposite direction.

11.4.2 Dispersion Relations

We have seen that the phase speed is the ratio of the angular frequency to the angular wavenumber. It is easy to see from the previous example that if the two waves have nearly the same frequencies and wavenumbers that $\Delta\omega/\Delta k$ is a good approximation to the derivative $d\omega/dk$. In the general development below, the group speed will be defined as

$$v_g = \frac{d\omega}{dk}. \quad (11.48)$$

Clearly, knowing how ω depends on k allows the computation of both the phase speed and the group speed. This relation is known as a *dispersion relation*.

Do the situations illustrated in Figs. 11.4–11.6 actually occur in nature? Indeed they do. Consider waves propagating on a water surface. The dispersion relation for small amplitude³, free-surface water waves is (e.g. [Apel, 1987](#))

$$\omega^2 = \left(gk + \frac{\sigma}{\rho} k^3 \right) \tanh(kh), \quad (11.49)$$

where $g = 9.8 \text{ m s}^{-2}$ is the acceleration of gravity, $\sigma = 0.074 \text{ N m}^{-1}$ is the surface tension of water, $\rho = 1000 \text{ kg m}^{-3}$ is the density of water, and h is the depth of the water. This dispersion relation has three limiting cases of interest.

For waves whose wavelengths are small compared to the water depth, so-called deep-water waves, $\tanh(kh) = \tanh(2\pi h/\lambda) \approx 1$. Then

$$\omega^2 = \left(gk + \frac{\sigma}{\rho} k^3 \right).$$

For long-wavelength “gravity waves”, $gk \gg \sigma k^3/\rho$ and the surface-tension term is negligible (i.e., surface tension is a negligible restoring force compared to gravity). Then

$$\omega^2 \approx gk.$$

From this we can compute $v_p = \omega/k = \sqrt{g/k}$ and $v_g = d\omega/dk = \frac{1}{2}\sqrt{g/k}$. Thus the group speed for deep-water gravity waves is one-half the phase speed. For very short-wavelength capillary waves, $\sigma k^3/\rho \gg gk$ and the surface-tension term dominates. Then

$$\omega^2 \approx \frac{\sigma}{\rho} k^3,$$

from which we find $v_p = \sqrt{\sigma k/\rho}$ and $v_g = \frac{3}{2}\sqrt{\sigma k/\rho}$. Thus the group speed is $\frac{3}{2}$ times the phase speed. For wavelengths that are large compared to the water depth, so-called

³The assumption that the waves have a small amplitude allows the equations of fluid motion to be linearized, which results in the dispersion relation shown here.

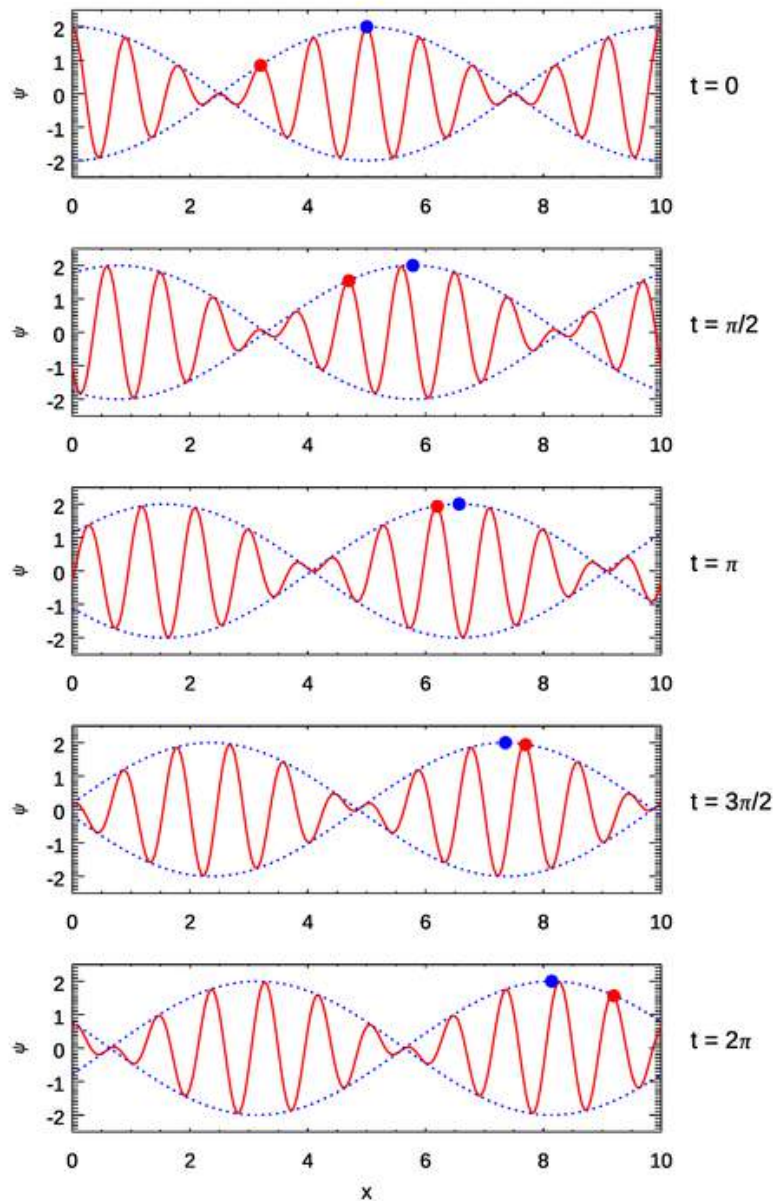


Figure 11.4: Illustration of phase and group speeds for the two-wave system giving $v_p > v_g$. The blue dots track the crest of the wave envelope. The red dots track the crest of a particular wave phase that forms at the rear of an envelope and moves forward, growing in amplitude and eventually passing the crest of the envelope, after which it decreases in amplitude as it propagates further and eventually dies out. The parameter values used in Eq. (11.45) are $A = 1$, $\omega_1 = 2\pi$, $\omega_2 = 1.1\omega_1$, and $k_1 = 2\pi$, $k_2 = 1.2k_1$. The resulting phase and group speeds are $v_p = 0.9545$ and $v_g = 0.50$, so both waves propagate to the right in the figure but the individual waves outrun the wave packet.

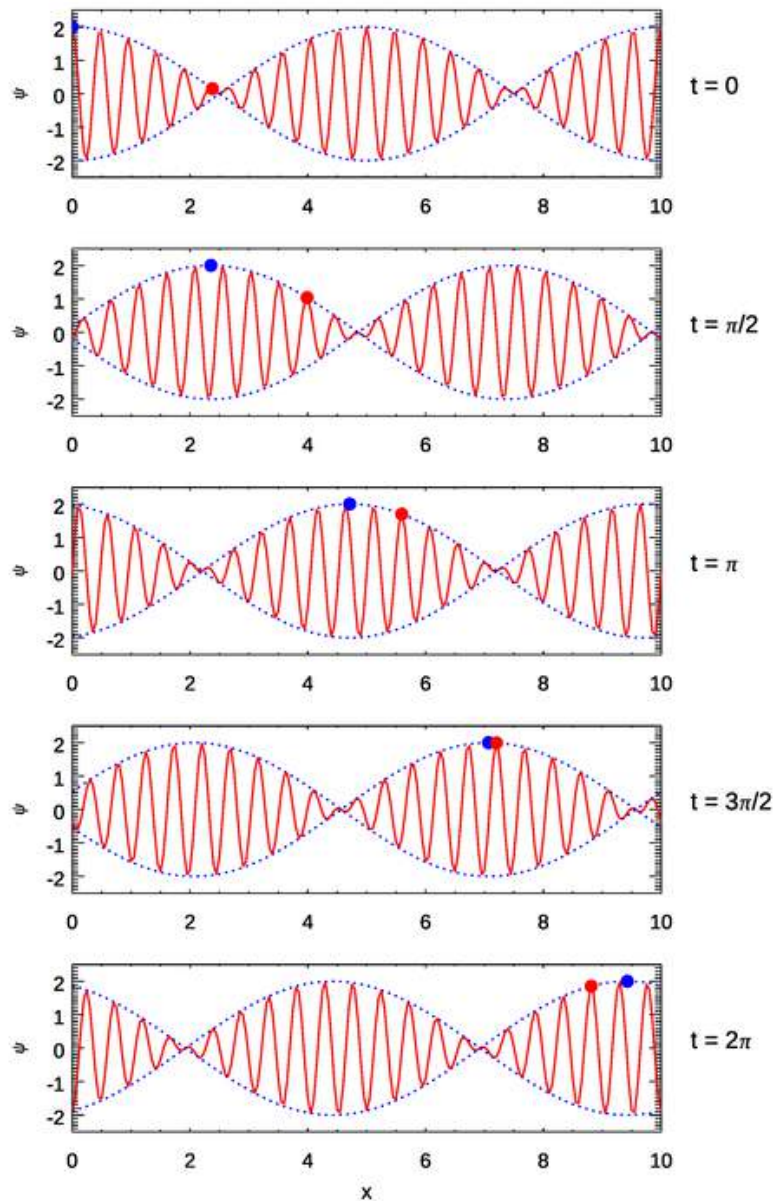


Figure 11.5: Illustration of phase and group speeds for two waves giving $v_p < v_g$. As before, the blue dots track the crest of the wave envelope. The red dots track the crest of a particular wave phase that forms at the front of an envelope. Although the red dot moves to the right ($v_p > 0$), from the viewpoint of the wave packet, the wave forms at the front of the packet and moves to the rear of the packet, where it dies out. The parameter values used in Eq. (11.45) are $A = 1$, $\omega_1 = 4\pi$, $\omega_2 = 1.15\omega_1$, and $k_1 = 4\pi$, $k_2 = 1.1k_1$. The resulting phase and group speeds are $v_p = 1.024$ and $v_g = 1.50$, so now the individual waves propagate slower than the wave packet.

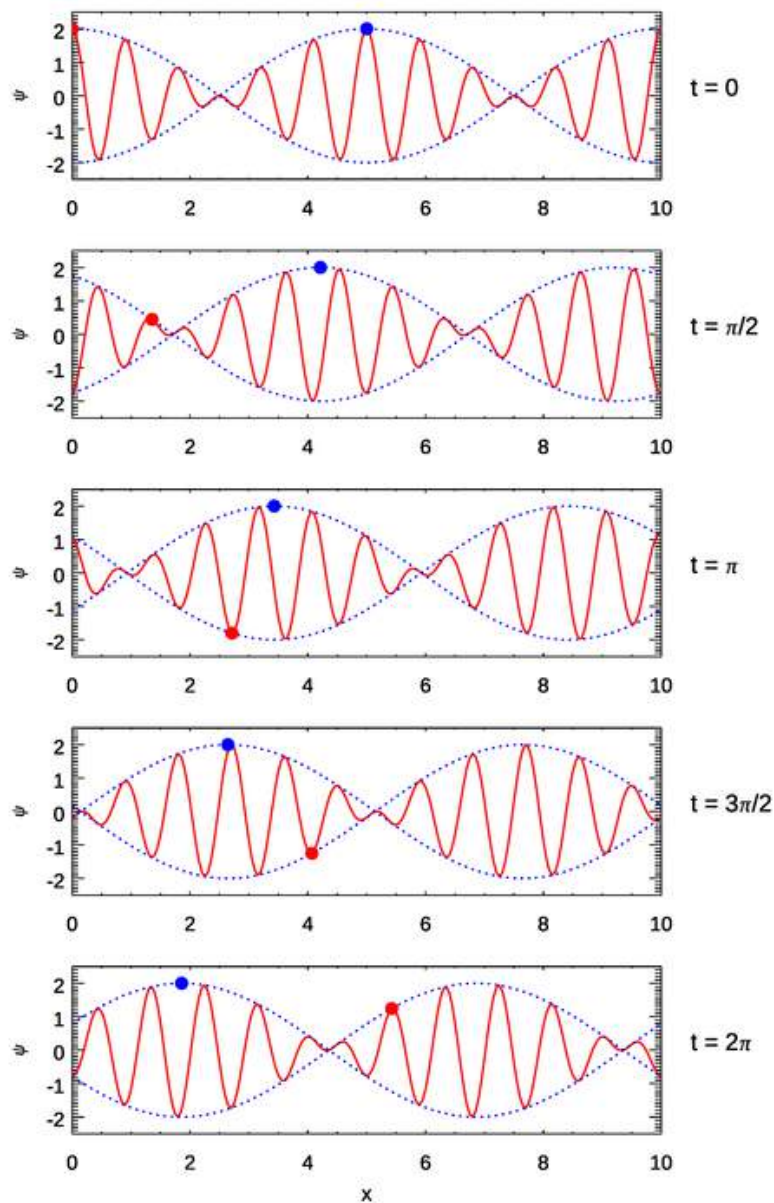


Figure 11.6: Illustration of phase and group speeds for two waves giving $v_p > 0$ and $v_g < 0$. The blue dots track the crest of the wave envelope and the red dots track the crest of a particular wave phase. The parameter values used in Eq. (11.45) are $A = 1$, $\omega_1 = 2\pi$, $\omega_2 = 0.9\omega_1$, and $k_1 = 2\pi$, $k_2 = 1.2k_1$. The resulting phase and group speeds are $v_p = 0.8636$ and $v_g = -0.50$. Now the individual waves propagate to the right, but the group envelope propagates to the left.

shallow-water waves, $\tanh kh \approx kh$ and, again, the surface tension term is negligible. The dispersion relation is then

$$\omega^2 \approx ghk^2,$$

from which we get $v_p = v_g = \sqrt{gh}$. Thus for shallow-water waves the phase and group speeds are equal.

The deep-water case of $v_g = \frac{1}{2}v_p$ is qualitatively like Fig. 11.4. The individual waves travel faster than the group envelope. The individual waves appear to form at the back of the envelope, move forward to the crest of the envelope while growing in amplitude, and then move past the crest of the envelope and decrease in amplitude until they disappear at the front of the group. I have watched this happen many times while sea kayaking on a nearly calm water body. Then a ship goes by some distance away and leaves a large wake composed of a string of smaller waves. As the wake reaches my kayak, smaller waves form at the rear of the wake group and propagate to the front of the wake, where they disappear. I never cease to be amazed as I watch waves appear from nowhere, grow in size as they propagate forward, and then simply disappear when they reach the front of the wake, exactly as predicted by the deep-water dispersion formula.

The capillary-wave case of $v_g = \frac{3}{2}v_p$ is similar to Fig. 11.5. The very short-wavelength capillary waves form at the front of the wave packet and propagate to the rear, or rather, the wave packet envelope outruns them. An example of a positive phase speed and a negative group speed will be seen below.

11.4.3 The General Development

This section merely outlines the general development of wave propagation in dispersive media; a more complete discussion is given in Jackson (1962, Section 7.3) and Towne (1967, Sections 16-4 and 16-5). We have learned that a single sinusoidal wave of given frequency ω and wavenumber k satisfies the wave equation and has a wave propagation speed of ω/k . Because the wave equation is linear, a linear combination of sinusoids also satisfies the wave equation. Thus a general wave function $\psi(x, t)$ can be built up as a sum of sinusoids of various amplitudes, frequencies, and wavenumbers. Rather than deal with real sinusoids as seen in Eq. (11.45), it is convenient (indeed, almost necessary) to use the complex representation of propagating waves. We thus write

$$\tilde{\psi}(x, t) = \int_{-\infty}^{\infty} \tilde{A}(k) \exp i[kx - \omega(k)t] dk, \quad (11.50)$$

where the tilde reminds us that the quantity is a complex function whose real value must be taken at the end of the development. Positive and negative k values represent waves with the same physical wavenumber but traveling in opposite directions. This equation thus adds up waves of any amplitude and wavenumber to create a general waveform. Note that the angular frequency is a function of the wavenumber according to the dispersion relation for the medium under study. It is noted that at $t = 0$ this equation is, to within a factor of 2π , precisely the inverse Fourier transform of $\tilde{A}(k)$ as defined in Eq. (A.4), although that connection will not be used here.

The dispersion relation can be expanded in a power series about any value of k_o :

$$\omega(k) = \omega_o + \left. \frac{d\omega}{dk} \right|_o (k - k_o) + \text{higher order terms}. \quad (11.51)$$

If the distribution of $\tilde{A}(k)$ values is peaked around the value k_o , or if the dependence of ω on k is weak, then the higher order terms can be neglected. Use of this approximation in

Eq. (11.50) eventually leads to the identification of $v_g = d\omega/dk$ as shown in (11.48) as the speed at which a wave pulse or packet travels. But note: as will be seen, neglecting the higher order terms in this expansion is not always a good approximation.

For light waves in a dielectric medium like water or glass the dispersion relation is

$$\omega(k) = \frac{c k}{n(k)}, \quad (11.52)$$

where c is the speed of light in a vacuum and $n(k)$ is the real index of refraction expressed as a function of k . This gives the phase speed as

$$v_p = \frac{\omega}{k} = \frac{c}{n(k)}. \quad (11.53)$$

This makes clear that “the speed of light in a medium” as presented in freshman physics is the phase speed. The corresponding group speed is

$$v_g = \frac{d\omega}{dk} = \frac{d}{dk} \left[\frac{ck}{n(k)} \right] = \frac{c}{n(k)} - \frac{ck}{n(k)^2} \frac{dn(k)}{dk}.$$

It is more convenient to view the index of refraction as a function of frequency than of wavenumber, so using

$$\frac{dn}{dk} = \frac{dn}{d\omega} \frac{d\omega}{dk} = \left(\frac{c}{n} - \frac{ck}{n^2} \right) \frac{dn}{d\omega}$$

in the previous equation leads to

$$v_g = \frac{c}{n(\omega) + \omega \frac{dn(\omega)}{d\omega}}.$$

This is the form usually seen in advanced physics texts. The corresponding equation with n viewed as a function of wavelength (as oceanographers prefer to do) is

$$v_g = \frac{c}{n(\lambda)} \left(1 - \frac{\lambda}{n(\lambda)} \frac{dn(\lambda)}{d\lambda} \right) = v_p \left(1 - \frac{\lambda}{n(\lambda)} \frac{dn(\lambda)}{d\lambda} \right). \quad (11.54)$$

Note that if n is independent of wavelength, $v_g = v_p$. This is approximately the case for water at wavelengths near 1000 nm, as seen in the top panel of Fig. 11.7. If $n = 1$, $v_g = v_p = c$. The only “substance” with $n = 1$ is a vacuum.

It is very instructive to examine these phase and group speeds when applied to the propagation of light in pure water. The top panel of Fig. 11.7 shows the real index of refraction of pure water for $\lambda = 0.1\text{--}10 \mu\text{m}$. The bottom panel shows the corresponding phase and group speeds computed from Eqs. (11.53) and (11.54), respectively. There are a number of features to note in these curves. Below 72 nm, the real index of refraction is less than 1, which makes the phase speed $v_p = c/n$ greater than c . This is not a violation of special relativity. The “universal speed limit” of c applies to material objects and the speed at which energy or signals can propagate; phase speeds can have any value. Between about 200 and 2000 nm, which includes the region of interest to optical oceanographers, $n(\lambda)$ decreases smoothly with increasing wavelength. The group speed is a bit greater than the phase speed, and both are about three-fourths the speed of light in a vacuum. However, there are regions where the group speed displays a seemingly bizarre behavior—it can be

greater than the speed of light, and it can be negative, and there are rapid fluctuations between these two extremes. The conventional wisdom of undergraduate physics says that energy propagates at the group speed, which therefore should not exceed the speed of light in vacuo.

What is happening here is that the concept of the group speed of a wave packet has broken down. Note in the figure that the large fluctuations in v_g occur near the wavelengths where $n(\lambda)$ is changing rapidly. Rapid changes in the index of refraction give rapid changes in the dispersion relation of Eq. 11.52. This in turn means that the higher order terms in the expansion of $\omega(k)$ seen in Eq. (11.51) cannot be neglected. When that is the case, the simple concept of a group speed as defined by Eq. (11.48) is inadequate to describe the propagation of the wave pulse envelope. As the venerable Jackson (1962, page 211) points out, a value of $v_g > c$ “...is no cause for alarm that our ideas of special relativity are here violated; group velocity is no longer a meaningful concept.” and “The behavior of the pulse is much more involved.” Towne (1967, Section 6-5) and Bohren and Huffman (1983, Section 9.1.3) discuss the physical processes leading to $n < 1$. It is true for all substances that at high frequencies (short wavelengths) n approaches 1 from values less than 1, just as is seen for water in Fig. 11.7.

11.5 Anomalous Dispersion: The Lorentz Model

When the real index of refraction decreases with increasing wavelength (i.e., $dn/d\lambda < 0$, or $dn/d\omega > 0$) the associated dispersion is called “normal” dispersion. This is the situation for water or glass at visible wavelengths. According to Snell’s law, red light will deviate less than blue when passing from air into a water drop. The result is that the red band of color (longer wavelengths) is at the outside of a rainbow, and the violet (shorter wavelengths) is at the inside. Regions of the spectrum where n increases with wavelength ($dn/d\lambda > 0$, or $dn/d\omega < 0$) would give the reverse order of wavelengths: shorter wavelengths would be at the outside of the rainbow and longer wavelengths at the inside. This reversal of “colors” from that seen in normal life at visible wavelengths is called “anomalous dispersion.” There is really nothing normal or abnormal about either situation, you simply get a particular order of refracted wavelengths depending on the sign of $dn/d\lambda$.

Anomalous dispersion occurs near strong absorption bands, which is how it enters into the understanding of the optical properties of phytoplankton (e.g., Morel and Bricaud, 1981a,b; Bricaud et al., 1983; Bricaud and Morel, 1986; Stramski et al., 1986). Because of this important application, anomalous dispersion warrants discussion. To understand anomalous dispersion, we must construct a model for light propagation in a dielectric. A simple, but very fruitful, model was constructed by H. A. Lorentz⁴ around 1900. I will merely outline the development of the Lorentz model—the details are a standard topic in texts on classical electrodynamics; see, for example, Griffiths (1981, Section 8.4.2). The arguments are as follows.

⁴This is the same Lorentz of the Lorentz transformation in special relativity and of many other results that still bear his name. He shared the 1902 Nobel Prize with P. Zeeman for their discovery and explanation of the Zeeman effect, which is the splitting of certain spectral lines in the presence of a magnetic field. His model of light propagation in dielectrics grew out of his efforts to understand the physical basis for the index of refraction. His work was one of the first couplings of Maxwell’s equations with the new “electron theory of matter” after the discovery of the electron in 1897.

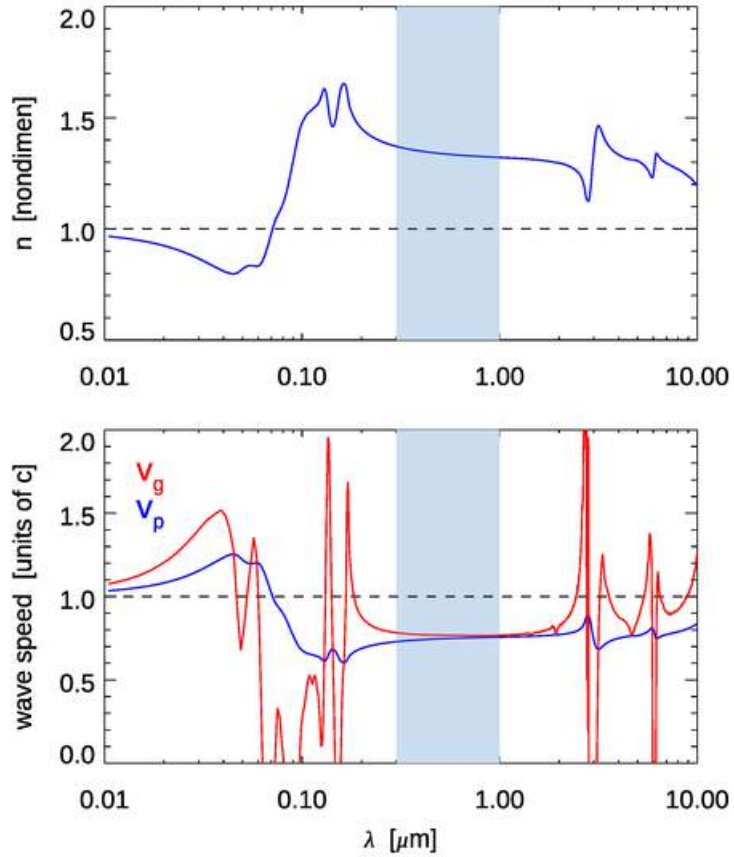


Figure 11.7: Phase and group speeds for water. The top panel shows the real index of refraction of pure water (data of Segelstein (1981) redrawn from Fig. 8.4). The bottom panel shows the phase speed v_p computed using Eq. (11.53) (blue curve) and the group speed v_g computed from Eq. (11.54) (red curve). The black dashed line is the speed of light. The region of interest to optical oceanography, 300-1000 nm, is shaded in blue.

Electrons in a dielectric are bound to specific molecules by electrostatic forces from the nuclei. When an electromagnetic wave of frequency ω passes by, the oscillating electric field causes the electrons to oscillate at the same frequency as the wave. However, the electrons also have resonant or “natural” frequencies ω_o . If the incident wave frequency is near to resonant frequency, the light-molecule interaction is particularly strong.

Let y be the displacement of an electron of charge q and mass m_e from its equilibrium position in the molecule. When the electron is pulled away from its equilibrium position at $y = 0$ there is a binding or restoring force that pulls it back towards its equilibrium position:

$$F_{\text{binding}} = -m_e\omega_o^2y.$$

If the sign of the force pulling the electron away from equilibrium is taken to be the positive y direction, then the binding force pulling it back is in the negative y direction. If the electron is set in motion by a passing electromagnetic wave, it will oscillate forever

unless there is some damping force. That force is taken to be proportional to the speed of the oscillating electron:

$$F_{\text{damping}} = -m_e \gamma \frac{dy}{dt},$$

where γ is a measure of the strength of the damping. The nature of the damping need not concern us here, but one mechanism is “radiation damping,” which arises because oscillating (i.e., accelerated) charges radiate energy. The driving force is the electromagnetic field:

$$F_{\text{driving}} = qE_o \cos(\omega t).$$

The total force on the electron is then

$$F_{\text{total}} = m_e \frac{d^2 y}{dt^2} = F_{\text{binding}} + F_{\text{damping}} + F_{\text{driving}},$$

which gives the equation

$$\frac{d^2 y}{dt^2} + \gamma \frac{dy}{dt} + \omega_o^2 y = \frac{q}{m_e} E_o \cos(\omega t).$$

This is the equation for a damped harmonic oscillator. Thus the electrons are modeled in classical physics terms as though they are charged masses tied to molecules like a masses on springs driven by an applied sinusoidal force. This equation is linear, so we can replace the terms with complex representations, just as was done in the previous section on wave propagation:

$$\frac{d^2 \tilde{y}}{dt^2} + \gamma \frac{d\tilde{y}}{dt} + \omega_o^2 \tilde{y} = \frac{q}{m_e} E_o e^{-i\omega t}.$$

In steady state conditions, the electron oscillates at the driving frequency:

$$\tilde{y}(t) = \tilde{y}_o e^{-i\omega t}.$$

Inserting this into the previous equation gives

$$\tilde{y}_o = \frac{q}{m_e} \frac{1}{(\omega_o^2 - \omega^2) - i\gamma\omega} E_o.$$

The dipole moment induced by the electric field is the product of the charge and the distance from equilibrium:

$$\tilde{p}(t) = q\tilde{y}(t) = \frac{q^2}{m_e} \frac{1}{(\omega_o^2 - \omega^2) - i\gamma\omega} E_o e^{-i\omega t}.$$

If N is the number of electrons per cubic meter in the material, then the bulk polarization is $\tilde{P} = N\tilde{p}$. Moreover, there are many electrons in a molecule, each of which has its own resonant frequency, denoted ω_{oj} for the j^{th} electron. Each electron contributes to the bulk polarization, so we sum over all electrons to obtain

$$\tilde{P}(t) = \frac{Nq^2}{m_e} \left(\sum_j \frac{f_j}{(\omega_{oj}^2 - \omega^2) - i\gamma_j\omega} \right) \tilde{E}(t),$$

where f_j is the number of electrons in each molecule with resonant frequency ω_{oj} and damping parameter γ_j .

Now recall from the previous section that the polarization vector was written as $\mathbf{P} = \epsilon_o \chi \mathbf{E}$. We now introduce a complex susceptibility χ and write the magnitudes as $\tilde{P} = \epsilon_o \chi \tilde{E}$. Recall also that the proportionality between the the displacement field \mathbf{D} and the electric field \mathbf{E} is $\mathbf{D} = \epsilon \mathbf{E} = \epsilon_o(1 + \chi)\mathbf{E}$, so we now introduce a complex permittivity $\epsilon = \epsilon_o(1 + \chi)$. Assembling these pieces gives the Lorentz model for the permittivity:

$$\epsilon = \epsilon_o \left(1 + \frac{Nq^2}{m_e \epsilon_o} \sum_j \frac{f_j}{(\omega_{oj}^2 - \omega^2) - i\gamma_j \omega} \right). \quad (11.55)$$

We also recall from Eq. (11.37) that the complex wavenumber was written as

$$\tilde{\kappa} = \omega \sqrt{\mu \epsilon} = \frac{\omega m}{c} = \kappa' + i\kappa'',$$

where m is the complex index of refraction.

To simplify the math, let us assume that the χ term is small compared to 1, and that $\mu = \mu_o$, so that we can write

$$\tilde{\kappa} = \omega \sqrt{\mu \epsilon} = \omega \sqrt{\mu_o \epsilon_o} \sqrt{1 + \chi} \approx \frac{\omega}{c} \left(1 + \frac{1}{2} \chi \right).$$

We now have

$$\tilde{\kappa} = \kappa' + i\kappa'' = \frac{\omega}{c} \left(1 + \frac{Nq^2}{2m_e \epsilon_o} \sum_j \frac{f_j}{(\omega_{oj}^2 - \omega^2) - i\gamma_j \omega} \right). \quad (11.56)$$

The summands in this equation are complex numbers of the form $a/(b - ic)$, which can be written as

$$\frac{a}{b - ic} = \frac{a}{b - ic} \frac{b + ic}{b + ic} = \frac{ab + iac}{b^2 + c^2} = \frac{ab}{b^2 + c^2} + i \frac{ac}{b^2 + c^2}.$$

This make it easy to extract the real and imaginary parts of 11.56. Recalling that the real index of refraction is $n = c\kappa'/\omega$ and that the absorption coefficient is $a = 2\kappa''$ gives

$$n \approx 1 + \frac{Nq^2}{2m_e \epsilon_o} \sum_j \frac{f_j(\omega_{oj}^2 - \omega^2)}{(\omega_{oj}^2 - \omega^2)^2 + \gamma_j^2 \omega^2} \quad (11.57)$$

and

$$a \approx \frac{Nq^2 \omega^2}{m_e \epsilon_o c} \sum_j \frac{f_j \gamma_j}{(\omega_{oj}^2 - \omega^2)^2 + \gamma_j^2 \omega^2}. \quad (11.58)$$

Equation (11.57) is what Lorentz was after: a model linking parameters of the substance, e.g., the density of electrons and their resonant frequencies, to the index of refraction. His development also gave a model for the absorption coefficient with no additional effort. The shape of the absorption curve defined by Eq. (11.58) is called a Lorentzian line shape. (There are other line shapes that describe effects in gases such as pressure broadening and Doppler shifting, but these do not concern oceanographers.)

The limiting cases of the Lorentz model are worth noting. At very low frequencies ($\omega \rightarrow 0$), the permittivity seen in Eq. (11.55) becomes real and of the form $\epsilon = \epsilon_0(1 + \chi)$, where χ is now a constant that depends on the particular substance. Water is a highly polar molecule, which means that the bulk polarization $\mathbf{P} = \epsilon_0\chi\mathbf{E}$ achieves a large magnitude as the water molecules align themselves with the slowly varying electric field. This gives a large value for χ and thus for ϵ . For water, $\epsilon \approx 78$ at low frequencies. This gives a real index of refraction $n = \sqrt{\epsilon} \approx 8.8$, which is seen in Fig. 8.4 for $\lambda \gtrsim 10\text{cm}$ ($\omega \lesssim 2 \cdot 10^{10}\text{rad s}^{-1}$). For very high frequencies, $\omega \rightarrow \infty$, the $-\omega^2$ term in the denominator of Eq. (11.55) dominates because the applied frequency is greater than any resonant frequency of the molecule. The permittivity then has the form $\epsilon \approx \epsilon_0(1 - \text{constant}/\omega^2)$. Thus the index of refraction approaches 1 from values less than 1 as $\omega \rightarrow \infty$. This is true for any substance, and is seen for water in Figs. 8.4 and 11.7 for $\lambda \lesssim 0.04\ \mu\text{m}$ ($\omega \gtrsim 5 \cdot 10^{16}\text{rad s}^{-1}$).

Figure 11.8 shows the shapes of the functions for the index of refraction and the absorption coefficient near a single resonant frequency. (That is to say, the n plot shows the frequency-dependent function without the factor $Nq^2/2m_e\epsilon_0$ and similarly for the a plot. A value of $\gamma = 0.4$ was used.) For most of the frequency range, $dn/d\omega > 0$ (or $dn/d\lambda < 0$). This is the case of “normal” dispersion. However, near the absorption line, the derivative of n is reversed. This is the case of “anomalous dispersion.” In materials like water there are many absorption lines due to electronic transitions (in the UV), vibrational modes (in the near IR), and rotational modes (in the far IR and longer wavelengths). These lines are often closely spaced, so they tend to overlap and blur out the features seen here for a single isolated line, and n can remain above 1. However, near very strong absorption features, the anomalous dispersion effect on the index of refraction can be seen. This is the case near $\lambda = 4$ and $6\ \mu\text{m}$ in Fig. 11.7, where the shape of the $n(\lambda)$ curve looks qualitatively like that of the bottom panel of Fig. 11.8. These features are due to strong vibrational modes of the water molecule (O-H stretching modes). The structure of $n(\lambda)$ in at wavelengths from 0.115 to $0.180\ \mu\text{m}$ is due to a number of electronic transitions.

To finish the discussion of the Lorentz model, it is first noted that the assumption that χ was small compared to 1 is fairly good for gases. However, for liquids and solids, where one molecule interacts with its neighbors and not just with the sinusoidal electric field, the equations seen here require modification to allow for the “packing” of the molecules in space. That leads to a result known as the Clausius-Mosotti, or the Lorentz-Lorenz, equation. However, that modification does not change the basic idea developed here and needs not be pursued. Of course, real electrons are not attached to molecules as though they were on springs. Somewhat surprisingly, a proper quantum mechanical treatment of the problem leads to a result with exactly the same functional form as seen in Eq. (11.55). However, the various terms are interpreted differently. The resonant frequencies are replaced by the frequencies corresponding to the energy differences between the quantized energy levels in the atom or molecule, and so on.

11.6 Kramers-Kronig Relations

At first glance it would seem that the real index of refraction and the absorption coefficient should be unrelated optical properties of a material. As emphasized in Section 4.2, spatial changes in the index are responsible for scattering. Absorption, on the other hand, describes how energy is removed from a beam of light. However, as seen in the development of the

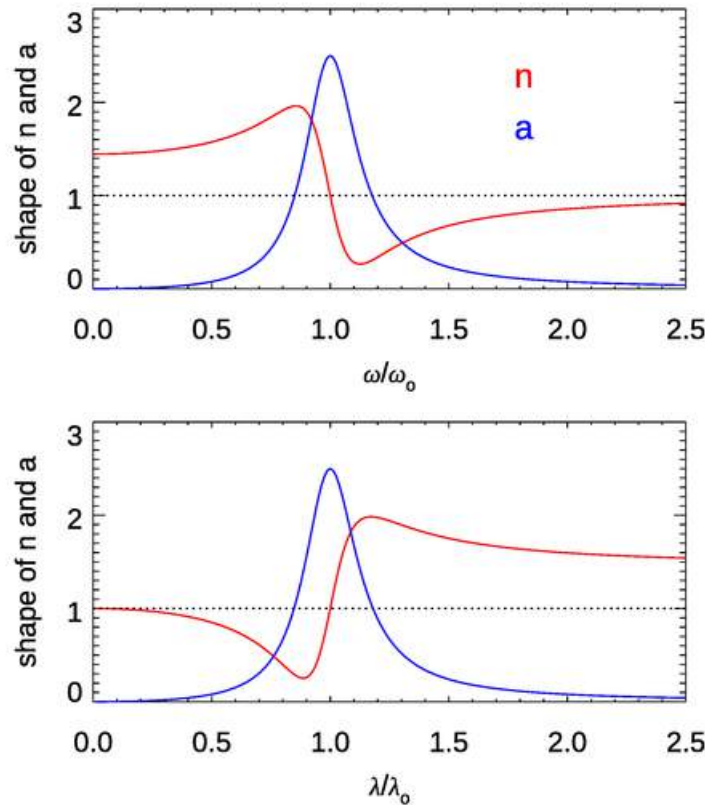


Figure 11.8: Lorentz line shapes for the real index of refraction n and the absorption coefficient a . These are plots of the summands in Eqs. (11.57) and (11.58) with the scale factors set to 1. Top panel: n and a as functions of frequency; bottom panel: n and a as functions of wavelength.

Lorentz model, equations for both the index of refraction and the absorption coefficient were obtained from the same analysis of how matter responds to a sinusoidal electromagnetic wave, i.e. to light. The frequency-dependent functions for $n(\omega)$ and $a(\omega)$ are very similar in form. This hints at a deeper connection between $n(\omega)$ and $a(\omega)$.

There is indeed a profound connection between the real index of refraction and the absorption coefficient. These two quantities are in fact so closely related that if you know one at all wavelengths, you can compute the other at all wavelengths. This relation was discovered independently by R. de L. Kronig in 1926 and by H. A. Kramers in 1927. The derivation of these equations is mathematically difficult and will not be given here; see [Bohren and Huffman \(1983, Section 2.3.2\)](#) for a discussion.

Kramers-Kronig equations relate the real and imaginary parts of what are called analytic functions. Without going into the details, there are many physical functions that satisfy the requirements to be analytic. Examples relevant to optics are the frequency dependent complex index of refraction $m(\omega)$ and related quantities such as the complex dielectric function $\epsilon(\omega)$ and the electric susceptibility $\chi(\omega)$. When stated for the complex index of refraction $m(\omega) = n(\omega) + ik(\omega)$, the Kramers-Kronig relations are ([Bohren and](#)

Huffman, 1983, Eqs. 2.49 and 2.50)

$$n(\omega) - 1 = \frac{2}{\pi} \wp \int_0^\infty \frac{\Omega k(\Omega)}{\Omega^2 - \omega^2} d\Omega \quad (11.59)$$

$$k(\omega) = -\frac{2\omega}{\pi} \wp \int_0^\infty \frac{n(\Omega) - 1}{\Omega^2 - \omega^2} d\Omega. \quad (11.60)$$

If these equations do not scare you, they should. Note that as the frequency is integrated from 0 to ∞ , somewhere along the way the integration variable Ω will equal ω and the denominator of the integrand will equal 0, while the numerator is non-zero. Thus the integrand becomes infinite and the integral diverges. The \wp symbol in front of the integrals indicates the *Cauchy principle value* of the integrals. The Cauchy principle value is a way to assign a finite value to some divergent integrals by computing a contour integral in the complex plane, “going around” the singular point, and then taking suitable limits⁵. The mathematical details are not needed for the present discussion, which will take as given that the integrals can be evaluated.

The absorption coefficient is determined by the imaginary part of the index of refraction via Eq. (11.43): $a = 4\pi k/\lambda$. Using this and the relation $\omega/c = 2\pi/\lambda$ allows the previous form of the Kramers-Kronig relations to be converted into the corresponding form for $n(\omega)$ and $a(\omega)$:

$$n(\omega) - 1 = \frac{c}{\pi} \wp \int_0^\infty \frac{a(\Omega)}{\Omega^2 - \omega^2} d\Omega \quad (11.61)$$

$$a(\omega) = -\frac{4\omega^2}{\pi c} \wp \int_0^\infty \frac{n(\Omega) - 1}{\Omega^2 - \omega^2} d\Omega. \quad (11.62)$$

These equations are not just of academic interest. For example, it is often easier to measure the absorption coefficient than the index of refraction. Then a measurement of absorption allows the determination of the index of refraction via Eq. (11.61). This is what [Zoloratev and Demin \(1977\)](#) did, although they do not give the details of their numerical calculations. [Segelstein \(1981\)](#) did essentially the same thing, although he did his numerical calculations using a fast Fourier transform technique derived from the Kramers-Kronig relation (11.59).

It is important to note that you cannot measure absorption at one frequency (or wavelength) and then determine the index of refraction at that frequency. You have to measure $a(\omega)$ over *all* frequencies, and then you can determine $n(\omega)$ over all frequencies by repeated evaluations of Eq. (11.61). In practice you never have measurements over all frequencies, but you have to have measurements over a wide-enough range of frequencies to enable an accurate approximate evaluation of the integrals. Of course, the integrations must be performed numerically, which is not trivial because of the singularity at $\Omega = \omega$. There is considerable literature on this; [Fitzgerald \(2020\)](#) gives a listing of MATLAB code to carry out the calculations.

The purpose of this brief discussion of Kramer-Kronig relations is to show that absorption and the real index of refraction are closely related. You might have a need for a material with particular absorption and refractive properties. So you mix together just

⁵To learn how that is done, you need to take a class in complex analysis, where you will learn about wonderful things like poles, residues, branch points, contour integration of complex functions and, of course, how to evaluate such integrals. The standard text on this topic was published by R. V. Churchill in 1948. It is still in press as an 8th edition ([Brown and Churchill, 2009](#)).

the right combination of dyes to give the absorption spectrum you want. However, at that point you have no freedom to somehow define the real index of refraction; it has now been fixed by Eq. (11.61).

There is much more to be said about Kramers-Kronig relations, which occur throughout physics and engineering. They are much more general than just the forms seen here for optical variables. At the deepest level, they are a necessary and sufficient for causality, which means that an effect cannot occur before its cause. This also means that no signal can propagate faster than the speed of light in a vacuum. A necessary and sufficient condition that a signal speed in a medium be less than the speed of light c is that the real and imaginary parts of the medium's refractive index satisfy Eqs. (11.59) and (11.60).

CHAPTER 12

Mie Theory

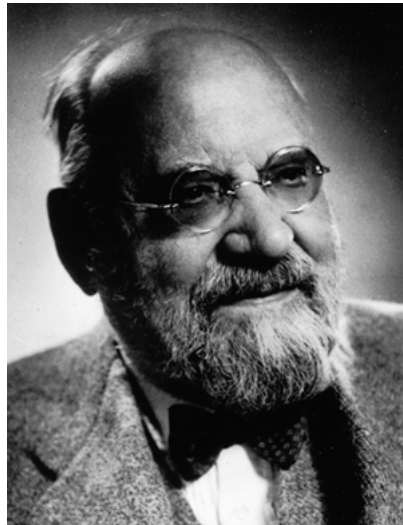


Figure 12.1: Gustav Adolf Feodor Wilhelm Ludwig Mie, later in life

As discussed in the Physics of Scattering, Section 6.2, one way to change the real index of refraction and thereby cause elastic scattering is to imbed a particle of some index of refraction within a medium with a different index of refraction. If the imbedded particle is a homogeneous sphere (of any radius), the solution of Maxwell's equations for a plane wave incident onto the sphere is now called Mie theory.

Gustav Mie (1868-1957) began his career in mathematics and mineralogy. One of the mysteries of the late 1800s was why colloidal suspensions of metallic particles displayed a rainbow of colors. Figure 12.2 shows an example of red to violet colors in suspensions of gold particles. The difference in colors is due to the different sizes of the gold particles, which are smaller than the wavelength of visible light. Understanding the optical effects of small concentrations of very small particles had important industrial applications because adding metallic nanoparticles to molten glass was (and still is) a common way to make glass of different colors.

Mie realized that particles made of the same material might absorb and scatter light much differently just because of differences in their sizes. He approached this problem by

working out the solution to scattering of light by spheres, starting with Maxwell's equations. His approach is all the more remarkable because, at the time, the importance of Maxwell's equations was not yet recognized by all physicists. His classic paper, [Mie \(1908\)](#), is titled "Beiträge zur Optik trüber Medien, speziell kolloidaler Metallösungen," or "Contributions to the optics of turbid media, particularly colloidal metal solutions." Mie used his solution equations to explain how particle size and absorption properties can explain the different colors. After that success, he moved on to other problems and never published another paper on the scattering of light.

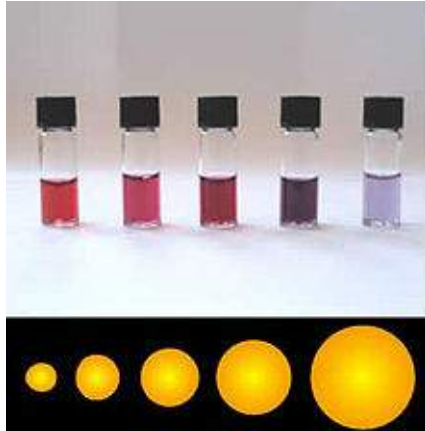


Figure 12.2: Suspensions of gold nanoparticles of various sizes showing a range of colors. Photo by Aleksandar Kondinski from en.wikipedia.org/wiki/Colloidal_gold.

The fundamental importance of Mie's paper went unrecognized for the next 50 years, apparently even by Mie himself. He does not even mention this paper in his autobiographical notes of 1948. This is perhaps understandable because his solution equations are so complicated that they cannot be evaluated except by modern computers. Mie gets credit for the complete solution of this scattering problem, but many of the most famous physicists of the late 1800s also worked on a solution; [Kerker \(1969\)](#) gives a detailed review of who did exactly what and when. There are several short biographies of Mie, e.g., [Lilienfeld \(1991\)](#), [Stout and Bonod \(2020\)](#), and [Mishchenko and Travis \(2008\)](#).

12.1 Statement of the Problem

Mie's problem is formulated as follows.

- We have given a single, homogeneous sphere of radius ρ , whose material has a complex index of refraction $m_s = n_s + ik_s$. Here n_s is the real index of refraction, and k_s is the complex index of refraction. The complex index is related to the absorption coefficient a_s of the sphere material by Eq. (11.43): $a_s(\lambda) = 4\pi k_s(\lambda)/\lambda$, where λ is the wavelength in vacuo corresponding to the frequency of an electromagnetic wave.
- The sphere is imbedded in a non-absorbing, homogeneous, infinite medium whose index of refraction is $m_m = n_m$.
- A plane electromagnetic wave of frequency f is incident onto the sphere. The wavelength on the incident light in the medium is thus $\lambda_m = c/(n_m f) = \lambda/n_m$, which corresponds to a wavelength in vacuo of $\lambda = c/f$.

- We wish to find the electric field within the sphere and throughout the surrounding medium. That is, we wish to determine how the incident light is absorbed and scattered by the sphere, including the angular distribution of the scattered light and its state of polarization.

The solution of this geometrically simple problem is exceptionally difficult. Indeed, this is one of the classic problems of applied mathematics, and its solution was attempted (and partially achieved in various forms) by many of the most illustrious figures of nineteenth-century physics. For historical reasons, Mie usually gets credit for the first complete solution of the problem, and his solution of Maxwell's equations is commonly called Mie theory. Mie's paper, (Mie, 1908), is 69 pages of dense equations, and I doubt that more than a handful of people have actually read the entire paper, although it has been cited in tens of thousands of papers. Bohren and Huffman (1983, page 93) say that someone who works through the details of Mie's solution will have "acquired virtue through suffering." I second that. Mie's paper is full of scary equations (see Fig. 12.4) connected by phrases like "It is easily shown that...", "Symmetry shows that...", and "You can convince yourself that..."

The details of Mie's solution are given (along with much needed extra explanation and modern notation) in the texts by van de Hulst (1957) and by Bohren and Huffman (1983). The purpose of the present section is to state the problem and outline its solution, so that you will understand the inputs to and outputs from computer programs that implement Mie's equations, and also have a qualitative idea of what is happening deep inside those programs. The chapter closes with examples of Mie-computed quantities.

12.1.1 Geometry

Figure 12.3 shows the geometry of Mie theory. An incident electromagnetic plane wave (i.e., a collimated beam of light) of frequency f (cycles per second) is incident onto a homogeneous spherical particle at the origin of a coordinate system. The coordinate system is chosen so that the wave is propagating in the $+z$ direction, and the origin of the coordinate system is chosen so that the wave is a cosine at time 0. The incident electric field in the medium of real index of refraction n_m then can be written as

$$\mathbf{E}_i(z, t) = \mathbf{E}_{oi} \cos(kz - \omega t),$$

where $k = 2\pi/\lambda_m = 2\pi n_m/\lambda$ is the wavenumber (cycles per meter) in the medium, and $\omega = 2\pi f$ is the angular frequency (radians per second). \mathbf{E}_{oi} is the amplitude of the incident electric field vector, and the direction of propagation is $\hat{\mathbf{e}}_z$. Life will be mathematically easier later on if we write the incident wave as a complex variable,

$$\mathbf{E}_i(z, t) = \mathbf{E}_{oi} e^{i(kz - \omega t)},$$

and keep in mind that we're interested in only the real part of the complex variable $\mathbf{E}_i(z, t)$. We're dealing with Maxwell's equations, which involve both electric and magnetic fields. However, if you know one, then you can get the other, so it suffices to discuss just the electric field.

The incident wave \mathbf{E}_i will interact with the particle at the origin of the coordinate system and generate a scattered wave \mathbf{E}_s traveling in direction $\hat{\mathbf{e}}_r$, which is at polar and

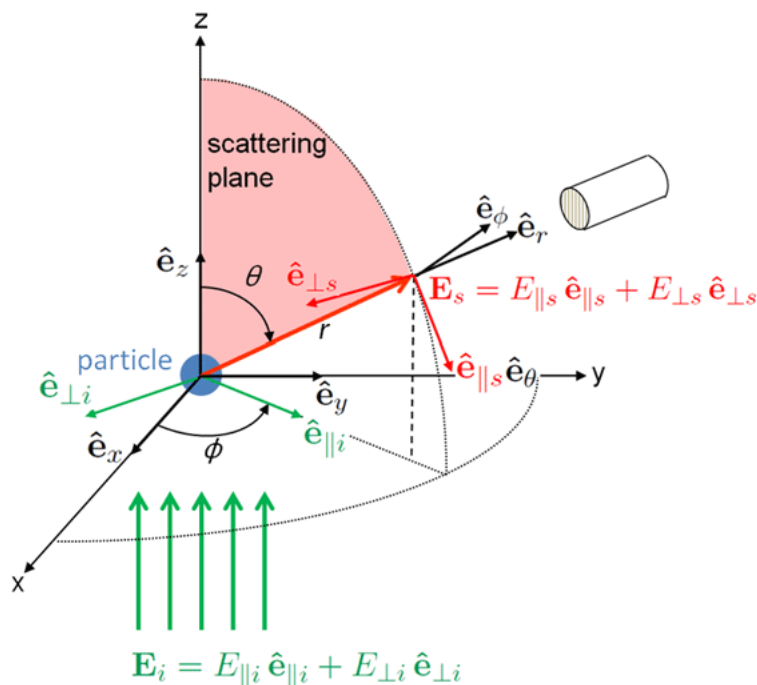


Figure 12.3: Geometry for Mie theory. Unit direction vectors are indicated by hatted letters, $\hat{\mathbf{e}}_z$, $\hat{\mathbf{e}}_{\perp i}$, etc. The thick green arrows represent the incident plane wave, and the thick red arrow represents the scattered wave. The scattering particle is the blue sphere at the origin.

azimuthal angles (θ, ϕ) as seen in Fig. 12.3. The incident direction $\hat{\mathbf{e}}_z$ and the scattered direction $\hat{\mathbf{e}}_r$ define the *scattering plane*, part of which is shaded in pink in the figure.

The incident plane wave is a transverse electromagnetic wave, which means that the electric and magnetic fields are perpendicular to the direction of travel. The incident wave is also arbitrarily polarized. An arbitrary state of polarization of $\mathbf{E}(z, t)$ can be written as a combination of two components, which are orthogonal to the direction of propagation. We choose these two directions to be parallel and perpendicular to the scattering plane. Thus we can write the incident electric field as

$$\mathbf{E}_i = E_{\parallel i} \hat{\mathbf{e}}_{\parallel i} + E_{\perp i} \hat{\mathbf{e}}_{\perp i}, \quad (12.1)$$

where the parallel ($\hat{\mathbf{e}}_{\parallel i}$) and perpendicular ($\hat{\mathbf{e}}_{\perp i}$) directions are shown by the thin green arrows in Fig. 12.3. Note that $\hat{\mathbf{e}}_{\perp i} \times \hat{\mathbf{e}}_{\parallel i} = \hat{\mathbf{e}}_z$. At large distances from the particle (the so-called “far field”), the scattered field becomes transverse and can also be written as a combination of components in directions parallel and perpendicular to the scattering plane: $\mathbf{E}_s = E_{\parallel s} \hat{\mathbf{e}}_{\parallel s} + E_{\perp s} \hat{\mathbf{e}}_{\perp s}$. The directions $\hat{\mathbf{e}}_{\parallel s}$ and $\hat{\mathbf{e}}_{\perp s}$ are parallel and perpendicular to the scattering plane at the point (r, θ, ϕ) where the scattered light is being measured by an instrument looking towards the particle at the origin. As seen in Fig. 12.3, $\hat{\mathbf{e}}_{\perp s} = \hat{\mathbf{e}}_{\perp i}$ but $\hat{\mathbf{e}}_{\parallel s} \neq \hat{\mathbf{e}}_{\parallel i}$. In particular,

$$\hat{\mathbf{e}}_{\parallel s} = \hat{\mathbf{e}}_{\theta}, \quad \hat{\mathbf{e}}_{\perp s} = -\hat{\mathbf{e}}_{\phi}, \quad \text{and} \quad \hat{\mathbf{e}}_{\perp s} \times \hat{\mathbf{e}}_{\parallel s} = \hat{\mathbf{e}}_r, \quad (12.2)$$

where $\hat{\mathbf{e}}_r, \hat{\mathbf{e}}_{\theta}, \hat{\mathbf{e}}_{\phi}$ give the directions of increasing r, θ, ϕ in the spherical coordinate system.

12.2 The Solution

For an arbitrary (i.e., non-spherical and/or inhomogeneous) particle at the origin, the scattered wave can be written as (e.g., van de Hulst, 1981, Section 4.41), (Bohren and Huffman, 1983, Section 3.2)

$$\begin{bmatrix} E_{\parallel s} \\ E_{\perp s} \end{bmatrix} = \frac{e^{ik(r-z)}}{-ikr} \begin{bmatrix} S_2 & S_3 \\ S_4 & S_1 \end{bmatrix} \begin{bmatrix} E_{\parallel i} \\ E_{\perp i} \end{bmatrix}. \quad (12.3)$$

The four S_j , $j = 1, 2, 3, 4$, are the elements of the *amplitude scattering matrix*. These functions transform the amplitudes of the incident electric field into the amplitudes of the scattered field. For an arbitrary particle, all four elements of the amplitude scattering matrix are non-zero and depend on both the polar θ and azimuthal ϕ scattering angles. These functions of course depend on the particle size, shape, and composition, as well as on the wavelength of the incident light, and it is that dependence that we wish to determine.

This present discussion accepts the form of Eq. (12.3) as given to us by the physicists, but it is worth a comment. When working with 3-D waves, it is common to seek a solution that separates the radial (r) and angular (θ, ϕ) variables. Here the S_j depend only on (θ, ϕ) . As shown in Eq. 11.42, the *irradiance* of an electromagnetic wave is proportional to square of the amplitude of the electromagnetic field. Squaring Eq. (12.3) gives a factor of

$$\left| \frac{e^{ik(r-z)}}{-ikr} \right|^2 = \frac{1}{k^2 r^2}$$

for the radial dependence of the scattered irradiance. We are considering scattering by a single particle, so the farther away we are from the particle, the less the irradiance detected by a sensor looking at the particle will be by a factor of $1/r^2$. This result is known as the “ r^2 law for irradiance.” We see here how the form of (12.3) for the scattered electric field has the r^2 law for irradiance built into the radial dependence of the electric field amplitudes. In particular, we are interested in the “far field” of the scattered light, which means that $kr \gg 1$. Note also that since kr is non-dimensional, so must be the S_j matrix elements.

For a homogeneous spherical particle, $S_3 = S_4 = 0$ and the amplitude scattering matrix reduces to

$$\begin{bmatrix} E_{\parallel s} \\ E_{\perp s} \end{bmatrix} = \frac{e^{ik(r-z)}}{-ikr} \begin{bmatrix} S_2 & 0 \\ 0 & S_1 \end{bmatrix} \begin{bmatrix} E_{\parallel i} \\ E_{\perp i} \end{bmatrix}. \quad (12.4)$$

Now comes the hard part: how to compute S_1 and S_2 given the particle radius ρ , the complex index of refraction of the spherical particle, $m_s = n_s + ik_s$, and the real index of refraction of the medium, $m_m = n_m$.

Comment on notation. It is common in Mie theory papers to use a as the radius of the spherical particle. However, in applications of Mie theory to optical oceanography, that leads to confusion with the absorption coefficient. Mie used ρ , and that’s good enough for me (pun intended). There is also confusion between the common use of k as wavenumber and k as the complex part of the index of refraction; Bohren and Huffman use a Roman k for wavenumber and an italic k for the imaginary part of the index of refraction. I avoid that subtlety by using k for wavenumber and k_s for the imaginary part of the index of refraction of the sphere, but then I’m using a subscript s for both “sphere” and “scattered,” although

context should keep things clear. Bohren and Huffman use ρ for the phase shift parameter $2x(n_s/n_m - 1)$. Unfortunately, choosing good notation is a never-ending problem.

The incident and scattered electric fields must satisfy *both* Maxwell's equations *and* boundary conditions for the behavior of the electric field at the surface of the sphere and at infinity. It is these boundary conditions that determine exactly which of all possible electric fields that satisfy Maxwell's equations is the one particular field that describes scattering by a particular sphere.

Figure 12.4 shows a couple of the pages of Mie's 1908 paper. This figure should be sufficient to convince you that we should skip the mathematical details and jump straight to the answer.

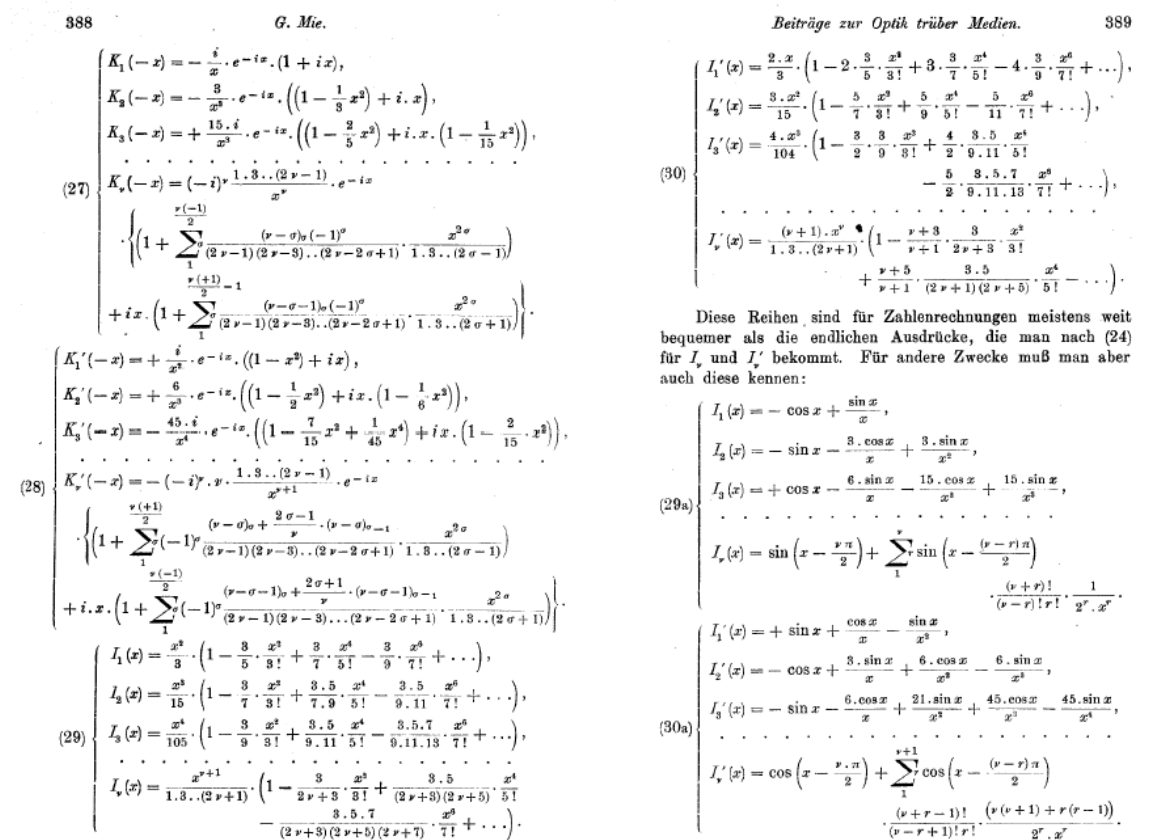


Figure 12.4: Two of the pages of Mie's 1908 paper.

Mie's solution is in the form of infinite series of very complicated mathematical functions. The terms in these series depend on a size parameter x ,

$$x = \frac{2\pi\rho}{\lambda_m} = \frac{2\pi\rho n_m}{\lambda}, \tag{12.5}$$

and the refractive index of the sphere relative to that of the surrounding medium,

$$m = \frac{n_s}{n_m} + i \frac{k_s}{n_m}. \tag{12.6}$$

The size parameter x is a measure of the sphere's size relative to the wavelength of the incident light in the surrounding medium. This parameter shows why oceanographers tend to use wavelength rather than frequency as the measure of light's oscillations: it is particle size relative to wavelength that is important for scattering (whether or not the particle is spherical). Note that the real part of the relative refractive index m can be less than 1, for example if the spherical particle is an air bubble ($n_s \approx 1$) in water ($n_m \approx 1.33$).

Mie's solution (in modern notation) is

$$\begin{aligned} S_1 &= \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} (a_n \pi_n + b_n \tau_n) \\ S_2 &= \sum_{n=1}^{\infty} \frac{2n+1}{n(n+1)} (a_n \tau_n + b_n \pi_n) \end{aligned} \quad (12.7)$$

where

$$\begin{aligned} a_n &= \frac{m \psi_n(mx) \psi'_n(x) - \psi_n(x) \psi'_n(mx)}{m \psi_n(mx) \xi'_n(x) - \xi_n(x) \psi'_n(mx)} \\ b_n &= \frac{\psi_n(mx) \psi'_n(x) - m \psi_n(x) \psi'_n(mx)}{\psi_n(mx) \xi'_n(x) - m \xi_n(x) \psi'_n(mx)} \end{aligned} \quad (12.8)$$

The a_n and b_n are often called the ‘‘Mie coefficients.’’ These functions describe multipole expansions of the electric (a_n) and magnetic (b_n) fields of the scattered wave: $n = 1$ is the dipole term, $n = 2$ is the quadrupole term, and so on. The ψ_n and ξ_n are Riccati-Bessel functions; the prime denotes derivatives of these functions with respect to the argument of the function (either x or mx). Riccati-Bessel functions are obtained from something called spherical Bessel and spherical Hankel functions, which in turn are obtained from something called Bessel functions of the first and second kind, which are themselves.... You get the idea. You eventually get down to something normal people can understand, likes sines and cosines. The π_n and τ_n are angle-dependent functions obtained by recursion relations:

$$\begin{aligned} \pi_n &= \left(\frac{2n-1}{n-1} \cos \theta \right) \pi_{n-1} - \frac{n}{n-1} \pi_{n-2} \\ \tau_n &= (n \cos \theta) \pi_n - (n+1) \pi_{n-1} \end{aligned} \quad (12.9)$$

starting with $\pi_0 = 0$ and $\pi_1 = 1$. The first 8 of these functions are plotted in Fig. 12.8 of the next section.

Thus the amplitude functions S_1 and S_2 depend on the particle size and index of refraction via the x and m in the a_n and b_n , and on scattering angle via the $\cos \theta$ factors in π_n and τ_n . For the geometry of Fig. 12.3, the polar angle θ is the scattering angle (ψ is my preferred symbol for scattering angle). It should be noted that if $m = 1 + i0$, i.e., if the sphere has the same index of refraction as the surrounding medium, then $a_n = 0$ and $b_n = 0$ for all n . That is to say, there is no scattering. This observation highlights that scattering is caused by differences in index of refraction.

If the incident light is unpolarized, then the scattered light is independent of the azimuthal scattering angle ϕ because of the symmetry of the sphere. However, the original problem was formulated for arbitrary polarization of the incident light, as shown in Fig. 12.3. If the incident light is plane polarized, then the plane of the polarization defines

an azimuthal direction that carries through as a ϕ dependence of the scattered light. For convenience, let the plane of polarization in Fig. 12.3 be the x - z plane, so that angle ϕ in the figure is the azimuthal angle from the plane of polarization to the scattering plane. The *far field* of the scattered light is the region far away from the scattering particle, usually measured in terms of λ/r . In the far-field limit $r \rightarrow \infty$, the scattered wave becomes a transverse electromagnetic plane wave and the scattered energy flows radially outward. Mie's solution for the θ and ϕ components of the electric field in the far field are (Kerker, 1969, Section 3.3.3), (van de Hulst, 1981, page 124) (and recalling Eq. 12.2)

$$E_\phi = -E_{\perp s} = \frac{i}{kr} \exp(-ikr - i\omega t) \sin \phi S_1(\theta) \quad (12.10)$$

$$E_\theta = E_{\parallel s} = \frac{-i}{kr} \exp(-ikr - i\omega t) \cos \phi S_2(\theta). \quad (12.11)$$

The irradiance is proportional to the square of the electric field (Eq. 11.42), so the corresponding scattered irradiance components are given by

$$\mathcal{E}_\phi = \frac{1}{k^2 r^2} \sin^2 \phi |S_1(\theta)|^2 \quad (12.12)$$

$$\mathcal{E}_\theta = \frac{1}{k^2 r^2} \cos^2 \phi |S_2(\theta)|^2. \quad (12.13)$$

Here \mathcal{E} is used for irradiance to prevent confusion with E for the electric field. As indicated in Eqs. (12.10) and (12.11), these components are respectively perpendicular and parallel to the scattering plane. As described by Kerker (1969, page 47), “Each of these components of the scattered light can be thought of as arising from that component of the incident beam polarized in the same sense, i.e., \mathcal{E}_ϕ arises from an incident beam of irradiance (proportional to) $\sin^2 \phi$ polarized perpendicularly to the scattering plane, and \mathcal{E}_θ from a beam of irradiance (proportional to) $\cos^2 \phi$ polarized parallel to the scattering plane¹.” This is made clear by Fig. 12.5, which is a view of the coordinate system of Fig. 12.3 as seen “from above,” i.e., looking in the $-\hat{z}$ direction. The solid green arrow is the incident beam polarized in the x - z plane, i.e., $\mathbf{E}_i = E_i \hat{e}_x$. This incident field can be resolved into components parallel ($E_{\parallel i}$) and perpendicular ($E_{\perp i}$) to the scattering plane, which is the red arrow at an angle ϕ to the incident polarization plane. If $\phi = 0$, then $E_\phi = 0$ because there is no component of the incident wave in the $\hat{e}_\phi = \hat{e}_y$ direction. If $\phi = 90$, then $E_\theta = 0$ because there is no component of the incident wave in the $\hat{e}_\theta = \hat{e}_y$ direction.

It must be remembered that the total field outside the scattering particle, which is what an instrument measures, is the vector sum of the scattered field just discussed and the incident (unscattered) field. This addition of fields must take proper account of the phase differences in the waves, which were induced by the incident wave passing through the particle.

Mie codes such as those used in Section 12.4 output only the two amplitude matrix elements $S_i(\theta)$; if needed, the ϕ dependence of the scattered light can be constructed from the preceding equations. However, most applications of Mie theory use only the efficiencies or the phase function (defined below), which do not depend on the ϕ dependence of the scattered electric field.

For completeness, it must be mentioned that in the *near field* of the sphere, i.e., within a few wavelengths of the surface of the sphere, the electromagnetic field is extremely complicated. In particular, there are also radial components of the electric and magnetic fields,

¹Kerker uses I_ϕ and I_θ can call them intensity rather than irradiance.

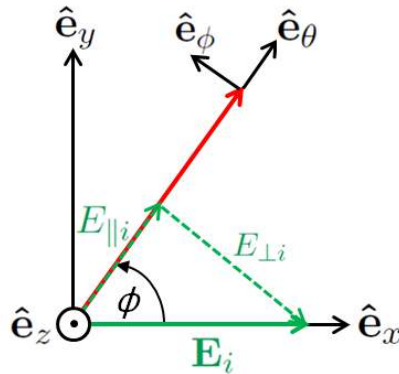


Figure 12.5: Geometry for interpreting azimuthal dependence of Mie scattering. This is Fig. 12.3 as seen “from above.” Unit direction vectors are indicated by hatted letters, \hat{e}_z , $\hat{e}_{\perp i}$, etc. The thick green arrows represent the incident plane wave polarized in the x - z plane, i.e., $\mathbf{E}_i = E_i \hat{e}_x$. The thick red arrow represents the scattered wave.

so that the scattered light is not a transverse electromagnetic wave, in spite of what you were taught in freshman physics. However, the radial electric fields decay with distance according to $(\lambda/r)^2$ (or $(\lambda/r)^4$ for the radial irradiances) or faster, so these radial components die away very quickly with distance from the scattering particle and are of no interest for oceanographic applications of Mie theory.

Mie’s solution is exact and valid for all sizes of homogeneous spheres, indices of refraction, and wavelengths.

12.2.1 Computational Issues and Derived Quantities

Unfortunately, it is not possible to compute the sums of the infinite series (12.7) analytically. The sums must be approximated numerically by adding only a finite number of terms in the series. If the size parameter x is less than roughly 10, e.g. typically when $\rho < \lambda$, only the first few terms are needed to get an accurate approximation for S_1 and S_2 . However, for large size parameters, e.g. when $\rho \gg \lambda$, many terms must be computed and convergence of the series is very slow. A commonly used rule is that the number of terms that must be computed is the integer closest to

$$N_{\max} = x + 4x^{1/3} + 2.$$

Suppose we want to compute scattering by a spherical phytoplankton of radius $\rho = 0.5 \mu\text{m}$, real index of refraction $n_s = 1.4$, in water with $n_m = 1.33$, at a wavelength of $\lambda = 500 \text{ nm}$. The size parameter is then $x = 8.36$ and $N_{\max} = 18$. That is no problem for a computer. But suppose we want to compute the scattering for a rain drop of size $\rho = 1 \text{ mm}$, $n_s = 1.33$, in air with $n_m = 1.0$, and for $\lambda = 500 \text{ nm}$. Then $x = 12,566$ and $N_{\max} = 12,661$. That can take a while.

The BHMIE code used to generate the examples on the next section is restricted to $x < 2 \cdot 10^4$ for reasons of computational accuracy. However, specialized codes have been designed for use with size parameters larger than 10^7 , but you are then getting into the world of quadruple-precision arithmetic and supercomputers. In Mie’s original paper, he

had to do the calculations by hand. He was able to compute only the first three terms of the infinite series, which limited his applications to particles less than 200 nm in size for visible wavelengths (x of order 1). However, that was sufficient to explain the optical effects of scattering by the colloidal particles that prompted his study.

It must be remembered that S_1 and S_2 are complex variables that transform incident complex electric fields into scattered complex electric fields. In oceanography, we are interested in scattered energy, which can be detected and turned into radiances. We are also interested in the shape of scattering phase function as a function of scattering angle, and in other quantities like absorption and scattering coefficients. The energy in an electric field is proportional to its amplitude squared so, not surprisingly, the quantities of real interest are obtained from various functions of the absolute values squared of S_1 and S_2 . The $|S_j|^2$ are real functions that are proportional to the scattered power (i.e., to scattered irradiance).

Suppose that the incident light is polarized parallel to the scattering plane, i.e. $E_{\perp i} = 0$ in Eq. (12.1). Then Eq. 12.3 shows that for an arbitrary particle, this incident light can be scattered into light that has components that are both parallel and perpendicular to the scattering plane. But for a spherical particle, Eq. 12.4 shows that incident light polarized parallel (perpendicular) to the scattering plane is scattered into light that remains polarized parallel (perpendicular) to the scattering plane.

Thus the angular pattern (ignoring normalization factors) of the incident parallel-polarized light that is scattered parallel to the scattering plane is given by

$$I_{\parallel s} = |S_2|^2 = S_2 S_2^*,$$

where S^* denotes complex conjugation. Likewise the perpendicular-incident to perpendicular-scattered pattern is given by $I_{\perp s} = |S_1|^2$. It is therefore common to plot $|S_1|^2$ and $|S_2|^2$ as functions of the scattering angle to see how these two polarization states are scattered.

$I_{\parallel s}$ and $I_{\perp s}$ can be thought of as unnormalized scattering phase functions for particular states of polarization. The oceanographer's scattering phase function for unpolarized light is, to within a normalization factor, given by

$$\tilde{\beta} = \frac{1}{2} (|S_1|^2 + |S_2|^2) = \frac{1}{2} (S_1 S_1^* + S_2 S_2^*). \quad (12.14)$$

A word of warning here: Mie codes return $S_1(\psi)$ and $S_2(\psi)$ for the set of scattering angles requested by the user (e.g., ψ from 0 to 180 deg by 0.1 deg). You can then use Eq. 12.14 to compute the phase function, but you can be guaranteed that it will be unnormalized. For example, if you study Bohren and Huffman (and you should), you will see many places where they say something like “where we have omitted the factor $1/k^2 r^2$ ” (page 113). You need to integrate the $\tilde{\beta}$ obtained from 12.14 to determine the needed normalization factor. A phase function used in a radiative transfer code such as HydroLight must satisfy the normalization $2\pi \int \tilde{\beta}(\psi) \sin \psi \, d\psi = 1$.

Additional output of Mie codes is usually given as various absorption and scattering *efficiencies*. The absorption efficiency Q_a , for example, gives the fraction of radiant energy incident on the sphere that is absorbed by the sphere. The term “energy incident on the sphere,” means the energy of the incident plane wave passing through an area equal to the cross-sectional (projected, or “shadow”) area of the sphere, $A_s = \pi \rho^2$. Likewise, the total scattering efficiency Q_b gives the fraction of incident energy that is scattered into all

directions. Other efficiencies can be defined: $Q_c = Q_a + Q_b$ for total attenuation, Q_{bb} for backscattering, and so on.

Mie solutions may also be presented in terms of absorption and scattering *cross sections*. The physical interpretation of these cross sections is simple. The absorption cross section σ_a , for example, is the cross sectional area of the incident plane wave that has energy equal to the energy absorbed by the sphere. The absorption and scattering cross sections are therefore related to the corresponding efficiencies by the geometrical cross section of the sphere. Thus

$$\sigma_a = Q_a A_s = Q_a \pi \rho^2 \quad (\text{m}^2).$$

Likewise, $\sigma_b = Q_b A_s$, and so on for σ_c, σ_{bb} , etc.

For the record, these cross sections are obtained within the Mie code from the a_n and b_n functions of Eq.(12.8):

$$\begin{aligned} \sigma_b &= \frac{2\pi}{k^2} \sum_{n=1}^{\infty} (2n+1) (|a_n|^2 + |b_n|^2) \\ \sigma_c &= \frac{2\pi}{k^2} \sum_{n=1}^{\infty} (2n+1) \Re(a_n + b_n) \end{aligned}$$

where $\Re(\dots)$ denotes the real part of the argument; $2\pi/k^2$ is equivalent to $2\pi\rho^2/x^2$ or $\lambda^2/(2\pi n_m^2)$. The absorption cross section can be obtained from $\sigma_a = \sigma_c - \sigma_b$.

Warning: Mie computer codes usually output what they call a backscattering efficiency or cross section, often denoted by Q_{back} or σ_{back} , which is given in terms of the Mie coefficients by

$$\sigma_{\text{back}} = \frac{\pi}{k^2} \left| \sum_{n=1}^{\infty} (2n+1)(-1)^n (a_n - b_n) \right|^2.$$

This quantity is also called the “radar cross section” or “radar backscattering.” The radar cross section is the hypothetical area required to intercept incident power onto the particle such that if the total intercepted power were re-radiated isotropically with a scattering amplitude equal to the amplitude for exact backscattering (at 180 deg), the power actually observed at the receiver is produced. See [Bohren and Huffman \(1983, Section 4.6\)](#) for further discussion. Do not confuse this quantity with the backscatter cross section σ_{bb} , which corresponds to scattering over the backward hemisphere of the phase function without any assumption of isotropic scattering pinned to the phase function value at 180 deg. The oceanographers’ backscattering coefficient can be computed by integrating the Mie-computed phase function from 90 to 180 deg².

12.3 Mie Theory with Particle Size Distributions

The cross sections obtained from Mie theory are for a single particle and have units of m² per particle, for the given particle properties. In oceanography, we are usually interested

²I have never found even one paper in the Mie literature that mentions backscattering over 90-180 deg, let alone seen a formula for computing the 90-180 deg backscattering cross section or efficiency from sums of the the Mie coefficients. It might be possible to obtain such a formula, but the derivation looks really ugly.

in a water body containing a huge number of particles per cubic meter. If there are \mathcal{N} particles per cubic meter corresponding to particle radius ρ (for given other conditions of indices of refraction and wavelength, which determine the size parameter x and relative index of refraction m), then the oceanographers' scattering coefficient corresponding to a collection of identical particles is

$$b(\rho) = \mathcal{N}(\rho) \sigma_b(\rho) \quad (\text{m}^{-1}).$$

Now, of course, the ocean does not contain just one size and type of particle. The range of particle sizes is described by the particle number size distribution $PSD(\rho)$, where $PSD(\rho)$ is a function such that the number of particles with radii between ρ and $\rho + d\rho$ is $PSD(\rho)d\rho$ (Section 8.12). The units of PSD are particles per cubic meter per size interval, which is usually written as $1/(\text{m}^3\mu\text{m})$ because particle sizes are usually measured in micrometers. Particle size distributions are often modeled as a power law of the form $PSD(\rho) = K\rho^{-s}$, where K sets the scale and s is in the range of 4 or 5. So the total scattering coefficient due to all particles of a given type is then

$$b(\text{all sizes}) = \int_0^\infty \sigma_b(\rho) PSD(\rho) d\rho.$$

In practice, the integration over all ρ will be approximated as a summation from some minimum size ρ_{\min} to some maximum size ρ_{\max} for which the $\sigma_b PSD$ term makes a significant contribution to the summation.

There will also be different types of particles for a given size ρ , which gives different x and m parameters in the Mie equations, hence different cross sections for the different particle types: $\sigma_b(\rho, i)$, where $i = 1, \dots, M$ labels the type of particle (cyanobacteria, quartz sand, etc.). Each particle type can have its own size distribution, $PSD(\rho, i)$. Then the total scattering coefficient for all sizes of all particle types is

$$b(\text{all sizes, all types}) = \sum_{i=1}^M \sum_{j=\rho_{\min}}^{\rho_{\max}} \sigma_b(\rho_j, i) PSD(\rho_j, i) \Delta\rho_j. \quad (12.15)$$

Thus there must be many evaluations of the Mie equations to obtain the cross sections for a realistic range of particle types and size bins. Keep in mind that if Mie theory is used to obtain the cross sections, then it is being assumed that the particles are homogeneous spheres, which is almost never the case in the ocean. However, Eq. (12.15) shows how Mie calculations can be used to compute the quantities used in optical oceanography if the particles can be approximated as homogeneous spheres.

12.4 Mie Theory Examples

This section shows examples computed from the equations discussed on the preceding Mie Theory Overview section. These examples were generated with the IDL version of the Bohren and Huffman Mie code (BHMIE) downloaded from [SCATTERLIB](#).

For ease of reference, recall the physical inputs to the Mie calculations:

- The radius ρ of the homogeneous sphere
- The complex index of refraction of the sphere, $m_s = n_s + ik_s$, where n_s is the real index of refraction, and k_s is the complex index of refraction. The complex index is related to the absorption coefficient a_s of the sphere material by $a_s(\lambda) = 4\pi k_s(\lambda)/\lambda$, where λ is the wavelength in vacuo corresponding to the frequency ν of the incident electromagnetic wave.
- The real index of refraction n_m of the non-absorbing, homogeneous, infinite medium
- The wavelength in vacuo, λ , of the incident plane electromagnetic wave. This corresponds to a frequency $\nu = c/\lambda$ and to a wavelength in the medium of $\lambda_m = \lambda/n_m$

These inputs are recast into the quantities actually used in the Mie calculations:

- The size parameter

$$x = \frac{2\pi\rho n_m}{\lambda} = \frac{2\pi\rho}{\lambda_m}$$

- The complex index of the particle relative to the medium,

$$m = \frac{n_s}{n_m} + i \frac{k_s}{n_m}$$

Most Mie codes then return the following outputs:

- The *unnormalized* complex amplitude matrix elements S_1 and S_2 as a function of scattering angle, $0 \leq \psi \leq 180$ deg, from which can be computed:
 - The scattering phase function for incident perpendicular polarization to scattered perpendicular polarization, $I_{\perp} = |S_1|^2$
 - The scattering phase function for incident parallel polarization to scattered parallel polarization, $I_{\parallel} = |S_2|^2$
 - The scattering phase function for unpolarized light, $\tilde{\beta} = \frac{1}{2}(|S_1|^2 + |S_2|^2)$
- Various efficiencies
 - The attenuation efficiency Q_c
 - The scattering efficiency Q_b
 - The absorption efficiency $Q_a = Q_c - Q_b$
- or the equivalent cross sections
 - The attenuation cross section σ_c
 - The scattering cross section σ_b
 - The absorption efficiency $\sigma_a = \sigma_c - \sigma_b$
- The average cosine of the scattering angle, g

The radar cross section may also be returned, but it is not of interest to oceanographers.

12.4.1 Example 1: Bohren and Huffman Fig. 4.9(b)

As a first example, let us reproduce one of the figures from [Bohren and Huffman \(1983\)](#). They considered a water droplet in air with $\rho = 0.263 \mu\text{m}$, an index of refraction of the particle relative to the air of $m = 1.33 + i10^{-8}$, and a wavelength of $\lambda = 550 \text{ nm}$. These values give a size parameter of $x = 3$. Figure 12.6 shows the curves for $I_{\perp} = |S_1|^2$, $I_{\parallel} = |S_2|^2$, and $\tilde{\beta} = \frac{1}{2}(I_{\perp} + I_{\parallel})$. The I_{\perp} and I_{\parallel} curves exactly reproduce the curves in Fig. 4.9(b) of Bohren and Huffman. This is a check that the Mie code is working correctly.

The I_{\perp} and I_{\parallel} curves can be viewed as unnormalized phase functions for scattering of incident light that is polarized perpendicular (parallel) to the scattering plane into light that is polarized perpendicular (parallel) to the scattering plane. The red curve shows the phase function for scattering of unpolarized incident light into a sensor that is not polarization sensitive. This is the phase function usually used by oceanographers. Recall, however, the warning of the previous section about Mie codes outputting unnormalized amplitude matrix elements. Integrating the red curve of Fig 12.6 gives

$$2\pi \int_0^{\pi} \tilde{\beta}(\psi) \sin(\psi) d\psi = 51.70.$$

Thus the red curve in the figure must be divided by 51.70 to obtain a properly normalized phase function that could be used, for example, in HydroLight.

The figure also shows the scattering, absorption, and attenuation efficiencies, and the mean cosine of the scattering angle. The scattering efficiency Q_b is 1.759, which means that the particle is scattering more than would be expected from its geometric cross section $\pi\rho^2$. We will return to this peculiar result below.

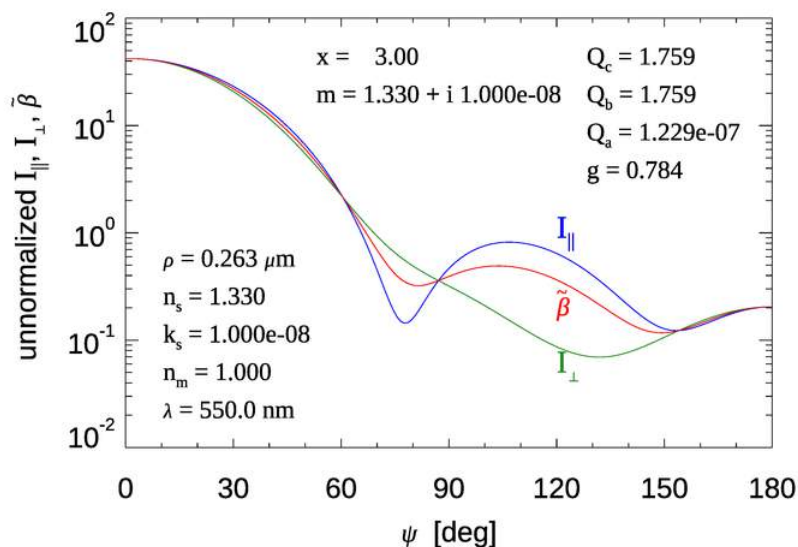


Figure 12.6: Figure 4.9(b) from [Bohren and Huffman \(1983\)](#), with additional information.

12.4.2 Example 2: Oceanic Particles

Let us next consider a particle that might be an approximation to a small spherical phytoplankton like *Synechococcus*. For physical parameters we use

- $\rho = 0.5 \mu\text{m}$
- $n_s = 1.37$
- $k_s = 0.015$ (corresponding to $a_s = 3.77 \cdot 10^5 \text{ m}^{-1}$)
- $n_m = 1.33$
- $\lambda = 500 \text{ nm}$

Remember that the absorption coefficient a_s is the absorption coefficient of the plankton material. These values are typically in the range of 10^4 - 10^5 m^{-1} at visible wavelengths. Figure 12.7 shows the resulting normalized phase function for unpolarized light.

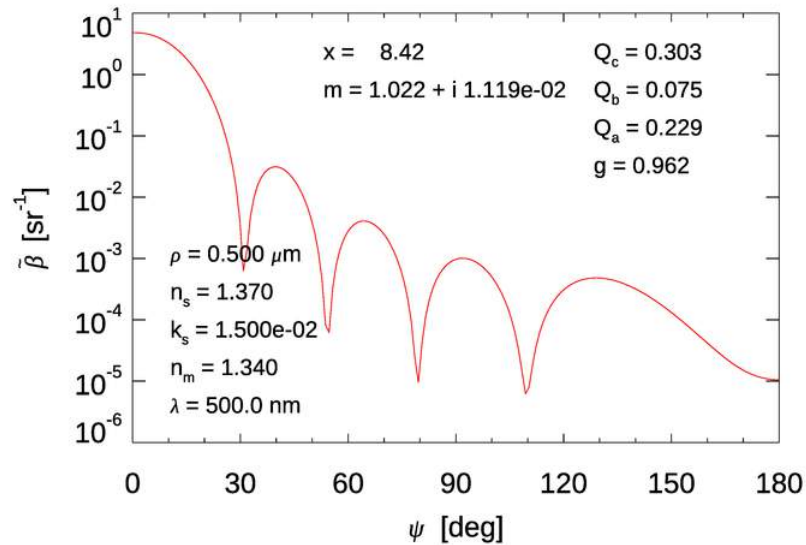


Figure 12.7: The normalized phase function for unpolarized light for the simulated phytoplankton with $\rho = 0.5 \mu\text{m}$.

As seen in this figure, a single particle scatters light in a very complex angular pattern. The peaks and valleys of the phase function result from constructive and destructive interference between the incident electric field and the scattered field that arises in the region of the sphere. You can think of this as being the 3-D version of the diffraction pattern of bright and dark lines seen when a plane wave is incident onto a slit in screen (recall Fig. 1.2). You can also think of this pattern as resulting from the sum of the infinite series of multipole modes of the scattered field, that is, the sum of a dipole electric field, plus a quadrupole field, plus an octupole field, plus....

Recall from Eq. (12.7) that the amplitude matrix elements are given by infinite sums of particle-dependent coefficients, the a_n and b_n of Eq. (12.8), times angle-dependent functions $\pi_n(\psi)$ and $\tau_n(\psi)$ given by Eq. (12.9). Figure 12.8 shows the first 8 of the angle

functions. The most important feature to note in these curves is that both π_n and τ_n are always positive as $\psi \rightarrow 0$. Therefore, as more and more terms are added in the amplitude matrix sums, the small-angle amplitudes (hence the associated phase function) become more and more peaked. The oscillations at larger scattering angles combine to create the peaks and valleys of seen in the phase functions of Figs. 12.6 and 12.7 and in 12.10 below.

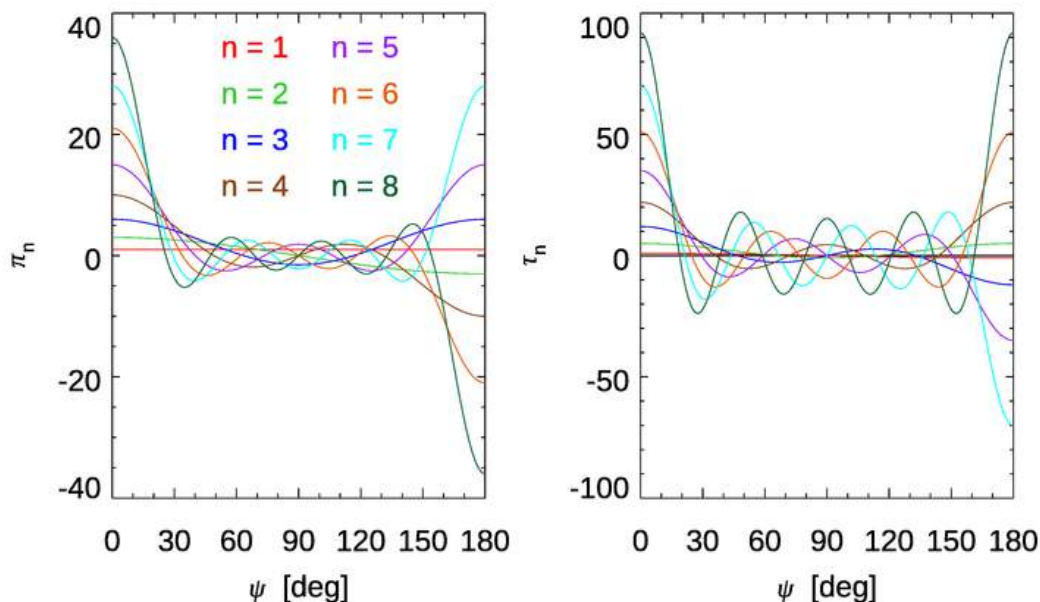


Figure 12.8: The first 8 of the Mie angular-dependent functions π_n and τ_n defined by Eq. (12.9).

The shapes of the π_n and τ_n functions trace back to mathematical functions called spherical harmonics, which are buried deep inside the Mie equations. Expanding a 3-D function of (θ, ϕ) as a sum of spherical harmonics is analogous to expanding a 1-D function as a sum of sines and cosines (a Fourier series). Figure 12.9 shows a graphical representation of the first few spherical harmonics. (If you think these patterns look suspiciously like the shapes of the orbitals seen in Fig. 5.6 of the Physics of Absorption section, you would not be wrong. The underlying physics is different—quantum mechanics vs light scattering—but the same mathematical functions occur in both cases.)

The locations of the peaks and valleys of the scattering pattern depend on the particle's physical properties—its size, relative index of refraction, and the wavelength. Figure 12.10 shows three normalized phase functions for the same particle type (a simulated phytoplankton) as in Fig. 12.7, but for particles of radii $\rho = 0.5, 1.0, 2.0 \mu\text{m}$. These particles have size parameters of $x = 8.42, 16.84,$ and 33.68 , respectively; each particle has $m = 1.02239 + i0.01119$. It is seen that the peaks and valleys of the phase functions are at different scattering angles. In general, the larger the size parameter x , the more features there are in the phase function. Note also that as the particle size increases, the phase function becomes more peaked at small scattering angles. The scattering cross section increases rapidly with particle size; that is, a large particle scatters more strongly than a small particle.

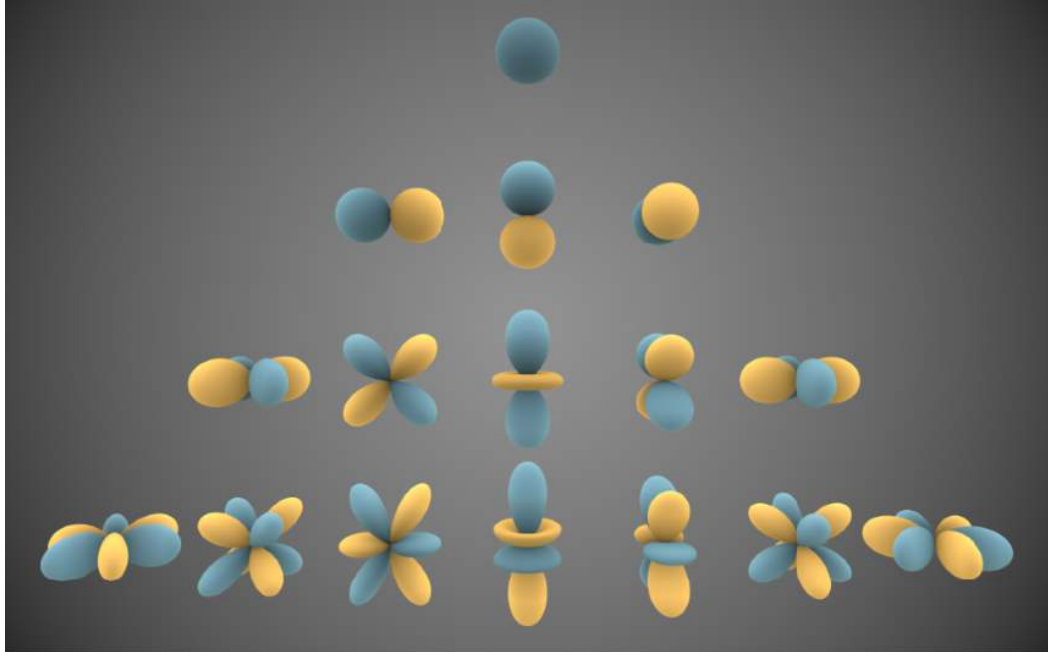


Figure 12.9: Graphical representation of the first few spherical harmonics. The blue color shows where the function is positive, and the yellow is negative values; the distance from the origin represents the magnitude for a given (θ, ϕ) direction. Creative Commons image created by Inigo Quilez, downloaded from [Wikipedia–Spherical Harmonics](#).

In the ocean, there are particles of many sizes and many indices of refraction, all occurring with different numbers of particles per cubic meter. When the phase functions for many different particle sizes and compositions are added together, the peaks and valleys of the individual phase functions tend to cancel out, leaving a much smoother phase function for the mixture of particles corresponding to what is measured on a sample of ocean water.

To combine phase functions for different particle types or sizes, remember that IOPs, including the volume scattering function, are additive. Thus the total phase function for the three particles used to generate Fig. 12.10 would be combined as follows. Let $\mathcal{N}_i, i = 1, 2, 3$ be the number density (particles per cubic meter) of each size of particle. The particle scattering cross sections obtained from Mie theory are σ_{si} ; the values are shown in Fig. 12.10. Then

$$\begin{aligned} VSF_{\text{total}} &= VSF_1 + VSF_2 + VSF_3 \\ b_{\text{total}} \tilde{\beta}_{\text{total}} &= b_1 \tilde{\beta}_1 + b_2 \tilde{\beta}_2 + b_3 \tilde{\beta}_3 \\ \tilde{\beta}_{\text{total}} &= \frac{\mathcal{N}_1 \sigma_{s1} \tilde{\beta}_1 + \mathcal{N}_2 \sigma_{s2} \tilde{\beta}_2 + \mathcal{N}_3 \sigma_{s3} \tilde{\beta}_3}{\mathcal{N}_1 \sigma_{s1} + \mathcal{N}_2 \sigma_{s2} + \mathcal{N}_3 \sigma_{s3}}. \end{aligned}$$

Suppose, just for the sake of illustration, that there are one-fifth as many particles of radius $\rho = 1 \mu\text{m}$ as there are particles of radius $\rho = 0.5 \mu\text{m}$, and one-fifth as many of radius 2 as of radius 1. Then the three individual-particle phase functions seen in the red, green, and blue curves of Fig. 12.10 would combine to give the total phase function shown by the black curve in the figure. This shows that the highly peaked features of the single-particle phase functions are starting to average out to leave a smoother total phase function. When

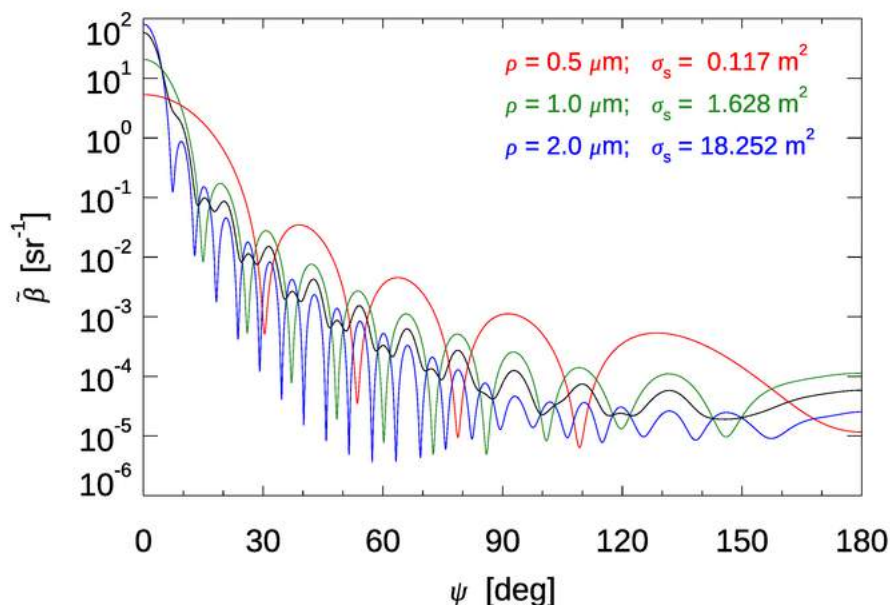


Figure 12.10: The normalized phase functions for unpolarized light for simulated phytoplankton with $\rho = 0.5, 1.0, 2.0 \mu\text{m}$. The black curve is the total phase function for a particular size distribution of these three particle sizes, as discussed in the text.

the same process is carried out for many sizes of particles, say from 0.1 to $100 \mu\text{m}$, and for many different kinds of particles (living phytoplankton of various types, detritus, sediment particles, etc.), a much smoother, typical ocean phase function can result.

Adding the results for individual particles together in the manner just shown assumes that each scattering particle is in the “far field” of its neighbors, so that the scattered field set up by one particle is not affected by nearby particles. This means that the particles should be separated by many wavelengths of the light. If there are, say, 10^{12} particles per cubic meter (a typical value for small phytoplankton) and the wavelength is 500 nm , then each particle is separated by 200 wavelengths, on average. This more than satisfies the far-field approximation.

Note, however, that the relative contributions of different particles is highly dependent not just on particle size and index of refraction, but on how many particles there are. Very small particles generally occur in high numbers, but they are individually weak scatterers and thus may contribute little to the total because of their small cross sections. Very large particles are strong scatterers, but they occur in very small numbers, and thus also may contribute little because of their small numbers. It is often the medium-sized particles, say radii from 0.5 to $5 \mu\text{m}$, that contribute the bulk of the total scattering in typical waters.

12.4.3 The Extinction Paradox

Recall from Fig. 12.6 that the scattering efficiency was $Q_b = 1.759$ for a small water droplet in air at 550 nm . This says that the particle is scattering more light than just the

light that encounters the cross-sectional area of the particle. This seems counterintuitive from everyday experience. If light hits a baseball, some will be absorbed and some will be reflected (scattered) by the ball, but light that misses the ball travels onward—or so it seems.

Figure 12.11 shows the absorption (Q_a), scattering (Q_b), and extinction ($Q_c = Q_a + Q_b$) efficiencies for highly absorbing soot particles, which are generated by incomplete combustion and are a common component of air pollution caused by coal-fired power plants, diesel exhaust, or forest fires. Soot has a real index of refraction of about 1.5, and an imaginary index of about 0.05 (Adler et al., 2010). The wavelength of the incident light is 532 nm.

For small size parameters, starting from $x = 0$, the scattering curve rises rapidly and then displays broad but damped oscillations with increasing x . These oscillations are caused by constructive and destructive interference between the incident and scattered light waves. Thus very small particles can be very efficient scatterers if their size matches the wavelength of the light in just the right way. For these soot particles in air, the maximum in scattering efficiency near $x = 4$ corresponds to a particle radius of about 340 nm. There is also a fine “ripple structure,” which requires a more complicated explanation (Bohren and Huffman, 1983). Thus these very small particles are scattering much more energy than would be expected from their physical cross section size. This behavior of large extinction for small x is seen in measurements of soot extinction efficiency in Fig. 12.12. The agreement between measurements and Mie theory is surprisingly good given that real soot particles are far from spherical.

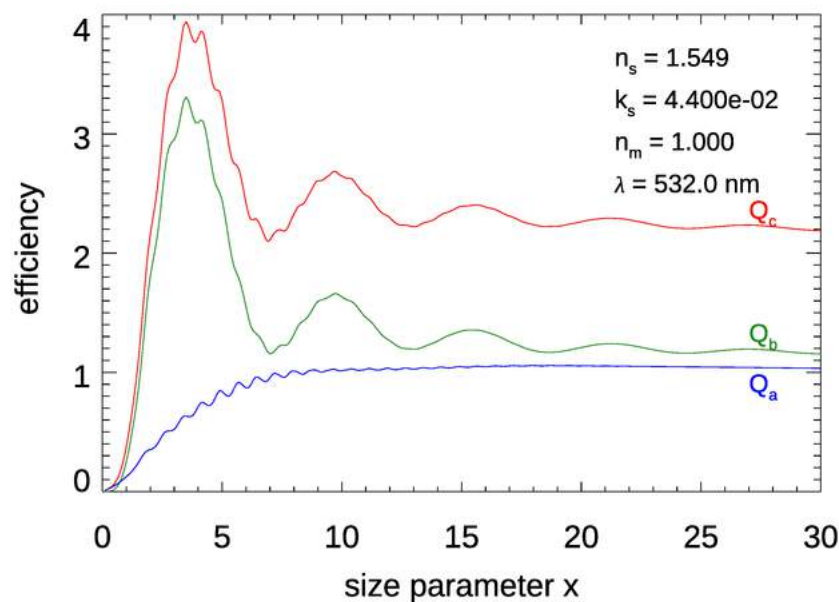


Figure 12.11: Efficiencies as a function of the size parameter for highly absorbing soot particles with $m = 1.549 + i0.044$. Compare with panel A of Fig. 12.12.

Figure 12.11 shows that, for soot in air, as x becomes large, Q_a is close to 1, but so

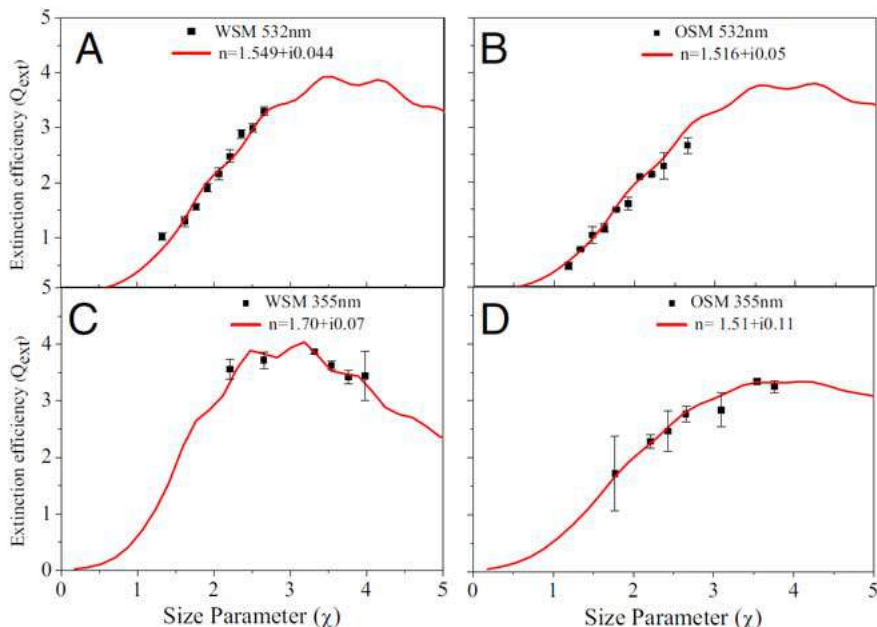


Fig. 2. The measured extinction efficiency (Q_{ext}) as a function of size parameter (x) of WSM (a) and OSM (b) at 532 nm (Upper) and 355 nm (Lower). The solid curve represents the Mie fit to the experimental data.

Figure 12.12: Measurements and Mie predictions of the extinction efficiency of diesel-fuel soot particles. Panel A corresponds to the Mie simulations of Fig. 12.11. (WSM refers to water-soluble material and OSM is organic soluble material. Figure 2 from Adler et al. (2010).

is Q_b , so Q_c is close to 2. These values remain similar as x continues to increase. For $x = 10,000$ (a particle size of $\rho = 0.85 \text{ nm}$ for $\lambda = 532 \text{ nm}$), $Q_a = 0.9085$, $Q_b = 1.096$, and $Q_c = 2.004$. The Q_a value makes sense: the soot is highly, but not totally absorbing, so a 91% absorption efficiency is plausible. However, if the light incident onto the particle is almost all absorbed, then it would seem that the particle should not be scattering as much or more light as it absorbs.

Geometric optics is a model of light propagation using rays and is often used to model light scattering by objects that are much, much larger than the wavelength of light. Geometric optics corresponds to large size parameters x in Mie theory. Geometric optics predicts that the maximum of Q_c should be 1, that is, a particle can absorb and/or scatter at most the energy that is incident onto the particle. The asymptotic approach of Q_c to 2 is known as the “extinction paradox.” This very general result is called a paradox because geometric optics predicts $Q_c = 1$. An equivalent statement is that the asymptotic value of the extinction cross section σ_c is twice the particle’s geometric cross section $\pi\rho^2$.

The standard explanation for the extinction paradox is that the “extra” scattering is due to diffraction by the particle. The electric field of light passing near to, but not intersecting, the particle is perturbed by the presence of the particle, which causes the light to change direction, even if only slightly. Diffraction is not easily observed for everyday objects like baseballs because the angle of deviation of the diffracted light from the direction of the

incident light is extremely small. Indeed, it was Newton's inability to observe the bending of light around large objects, and their apparently sharp shadow edges, that led him to conclude that light consisted of particles and not waves. Nevertheless, diffraction occurs for all sizes of objects, and any deviation, no matter how small, of light from its initial direction counts as scattering and contributes to Q_b . However, the diffraction explanation can be shown to be incomplete, and papers are still being written about the fundamental cause of the $Q_c = 2$ limit (e.g., Berg et al., 2011). The full explanation requires a deep understanding of how the incident and scattered waves interact within and surrounding the particle. Regardless of the full explanation, the effect is very real and occurs for all particles.

Another example of the extinction paradox is seen in Fig. 12.13. These simulations are for quartz particles in water. Quartz has a real index of refraction of about 1.54 and is modeled here as completely non-absorbing. The wavelength is taken to be 500 nm. Since the particles are non-absorbing, the absorption efficiency is identically 0, and the total extinction is due to scattering ($Q_c = Q_b$). Both the interference structure and the ripple structure are clearly seen. The ripple structure is much less noticeable for the soot particles in Fig. 12.11 because the ripples are damped out by absorption. The quartz particles also display an oscillating, asymptotic approach of Q_c to a value of 2, just as do the soot particles.

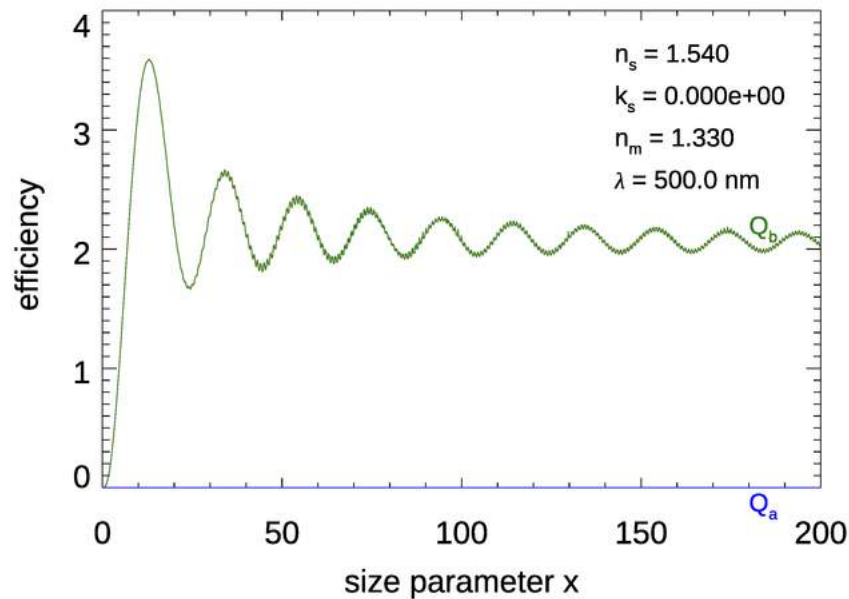


Figure 12.13: Efficiencies as a function of the size parameter for non-absorbing quartz particles in water at 500 nm.

Figure 12.14 shows the efficiencies for the phytoplankton model used to generate the phase function of Fig. 12.7. For this relative index of refraction, the scattering efficiency shows only one broad maximum near $x = 50$, which corresponds to a particle radius of about $\rho = 3 \mu\text{m}$. By $x = 150$, the extinction efficiency $Q_c = 2.02$, so within 1% of its

asymptotic value.

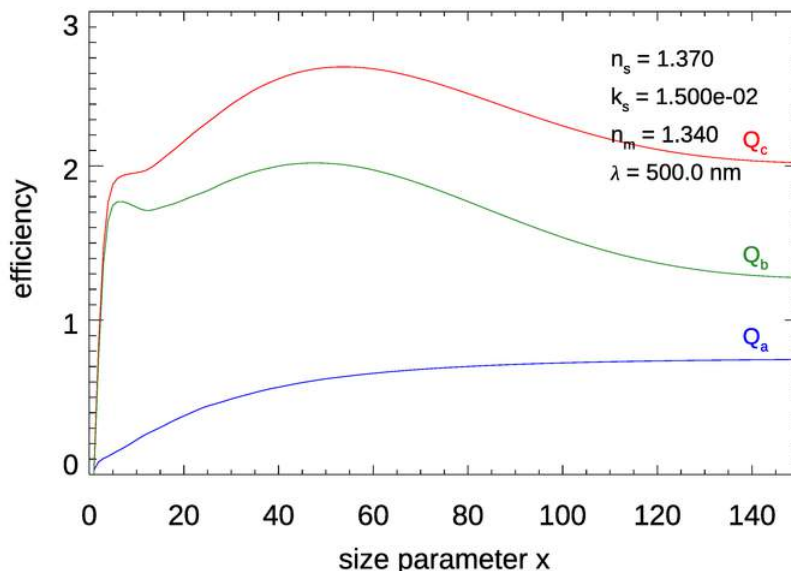


Figure 12.14: Efficiencies as a function of the size parameter x for the phytoplankton model used to generate the phase function of Fig. 12.7.

A fine way to spend a rainy Saturday is to download a Mie code and do a few thousand runs to see the effects of real and imaginary indices of refraction, particle size, and wavelength on the various phase functions and efficiencies. However, regarding the wavelength effect, keep in mind that as wavelength changes so do the indices of refraction of the particle and medium. Suppose that, for a given particle radius ρ you want to generate a figure like Figs. 12.14 and 12.13, but showing the efficiencies as a function of wavelength. That can be done, but you have to change the particle and medium indices of refraction for each wavelength. The absorption coefficient of water changes from $a_w \approx 0.015 \text{ m}^{-1}$ at 440 nm to 0.65 m^{-1} at 700 nm. If modeling a water droplet in air, this wavelength dependence of a_w determines the water imaginary index of refraction via $k_s(\lambda) = a_w(\lambda)\lambda/(4\pi)$, and thus affects the size parameter $x(\lambda) = 2\pi\rho k_s(\lambda) n_m(\lambda)/\lambda = \rho a_w(\lambda) n_m(\lambda)/2$. If modeling a phytoplankton in water, the needed phytoplankton $k_s(\lambda)$ might be constructed from a phytoplankton chlorophyll-specific absorption spectrum, and so on.

Finally, remember that Mie theory is valid only for homogeneous spherical particles, even though it often, but not always, gives reasonable approximate results for non-spherical particles.

12.5 Mie Theory Approximations

As discussed on the previous two sections, Mie theory is exact for homogeneous spheres of any size, but it can be computationally expensive for spheres that are large relative to the wavelength of the light incident on them. There are, however, analytical approximations to the exact theory that can be useful in limited situations. Loosely speaking, these might

be called approximations for “really small particles,” “weakly scattering particles,” and “really large particles.” This section presents these three approximations and illustrates the limits of each.

12.5.1 Small Particles, $x \ll 1$: Rayleigh’s Approximation

John William Strutt had the good fortune to be born a wealthy British aristocrat at a time when that still meant something. Unlike some of his peers, he did not spend his life in the idle dissipation of an inherited fortune. On the contrary, he became one of Britain’s greatest scientists. He worked in many areas including optics, acoustics, and fluid mechanics, publishing 446 papers. He received many honors, including the Nobel Prize in Physics in 1904 for “for his investigations of the densities of the most important gases and for his discovery of argon in connection with these studies.” Upon the death of his father in 1873, he inherited the title of Baron Rayleigh, and was henceforth known as Lord Rayleigh. [Masters \(2009\)](#) gives a short biography of his scientific life.

In a series of three papers published in 1871, he developed equations to describe scattering by non-absorbing particles that are small compared to the wavelength of the incident light. The first of these papers, [Strutt \(Lord Rayleigh\) \(1871\)](#), titled “On the Light from the Sky, its Polarization and Colour,” begins “It is now, I believe, generally admitted that the light which we receive from the clear sky is due in one way or another to small suspended particles which divert the light from its regular course.” He first used dimensional analysis to conclude that the scattering must be proportional to the inverse fourth power of the wavelength.

Rayleigh’s dimensional analysis argument was as follows. Incident irradiance E_i will be scattered by a particle to give some amount of scattered irradiance E_s . How much irradiance is scattered and observed at a distance r from the particle should depend on the amount of material doing the scattering, i.e., on the volume V of the particle, the indices of refraction of the particle (n_p) and of the surrounding medium (n_m), the distance r , the wavelength λ , and perhaps on the speed of light c . Thus we can write

$$E_s = f(V, n_p, n_m, r, \lambda, c) E_i,$$

where f is a nondimensional function that converts incident irradiance into scattered irradiance. For a single particle viewed from a distance r , the scattered irradiance must decrease as $1/r^2$; this is the r^2 law for irradiance. The speed of light c is the only parameter that depends on time, and the steady-state scattering is independent of time, so c cannot appear in f . The scattered electric field will be proportional to the volume of the particle, and the irradiance is proportional to the square of the electric field, so the scattered irradiance must be proportional to V^2 . The indices of refraction are dimensionless. So the only way to keep

$$f(V, n_p, n_m, r, \lambda, c) = f'(n_p, n_m, \lambda) \frac{V^2}{r^2}$$

dimensionless is for the wavelength to appear as $1/\lambda^4$, because V^2 has dimensions of length to the sixth power. Thus it must be that

$$E_s \propto \frac{V^2}{r^2 \lambda^4} E_i.$$

After reaching this conclusion, he proceeded to derive the exact mathematical form of the scattering.

Rayleigh found, under the assumption that the particle is much smaller than the wavelength of the incident light, that the single-particle volume scattering function for unpolarized light (to use modern terminology and notation) is

$$\beta(\psi) = \frac{8\pi^4 \rho^6}{\lambda^4} \left(\frac{m^2 - 1}{m^2 + 2} \right)^2 (1 + \cos^2 \psi), \quad (12.16)$$

where ρ is the particle radius, λ is the wavelength, m is the real index of refraction of the particle relative to that of the surrounding medium, and ψ is the scattering angle. This result can be written as the product of a single-particle scattering cross section σ_b and a scattering phase function $\tilde{\beta}$, $\beta = \sigma_b \tilde{\beta}$, where

$$\sigma_b = \frac{\pi^5 \rho^6}{96\lambda^4} \left(\frac{m^2 - 1}{m^2 + 2} \right)^2 \quad (12.17)$$

$$\tilde{\beta} = \frac{3}{16\pi} (1 + \cos^2 \psi). \quad (12.18)$$

This phase function satisfies the normalization condition $2\pi \int_0^\pi \tilde{\beta}(\psi) \sin \psi d\psi = 1$. Note that σ_b has units of m^2 . After multiplication by \mathcal{N} particles per cubic meter, the result is a scattering coefficient $b = \mathcal{N}\sigma_b$ with the customary units of inverse meters. $\tilde{\beta}$ describes the angular scattering per steradian, so the bulk VSF then has units of $\text{m}^{-1} \text{sr}^{-1}$. Dividing σ_b by the particle cross section $\pi\rho^2$ and rewriting in terms of the Mie theory size parameter $x = 2\pi\rho/\lambda$ (for a medium index of refraction of 1) gives the scattering efficiency

$$Q_b = \frac{8}{3} x^4 \left(\frac{m^2 - 1}{m^2 + 2} \right)^2. \quad (12.19)$$

Rayleigh used the λ^{-4} dependence of his equations to explain the blue sky as wavelength-dependent scattering by the “small suspended particles” of his first papers. However, in [Rayleigh \(1899\)](#) he returned to “...the interesting question whether the light from the sky can be explained by diffraction from the molecules of air themselves, or whether it is necessary to appeal to suspended particles composed of foreign matter, solid or liquid.” and concluded that “...even in the absence of foreign particles we should still have a blue sky.”

12.5.2 Rayleigh’s Approximation Obtained from Mie Theory

Rayleigh’s result (12.16) can be obtained from and extended by Mie theory, which came 36 years later. Recall from Eqs. (12.5 – 12.9) in the Mie Theory Overview section that Mie’s solution for scattering by a sphere is in the form of infinite series, the terms of which depend on powers of the size parameter x . Expanding the series solution and keeping terms through x^4 eventually leads to the efficiency factors (see [Bohren and Huffman \(1983, Section 5.1\)](#) for the math)

$$Q_c \approx 4x\Im \left\{ \frac{m^2 - 1}{m^2 + 2} \left[1 + \frac{x^2}{15} \left(\frac{m^2 - 1}{m^2 + 2} \right) \frac{m^3 + 27m^2 + 38}{2m^2 + 3} \right] \right\} + \frac{8}{3} x^4 \Re \left\{ \left(\frac{m^2 - 1}{m^2 + 2} \right)^2 \right\} \quad (12.20)$$

and

$$Q_b \approx \frac{8}{3}x^4 \left| \frac{m^2 - 1}{m^2 + 2} \right|^2, \quad (12.21)$$

where now the index of refraction m can be complex, i.e. the sphere can be absorbing. $\Re\{\dots\}$ and $\Im\{\dots\}$ indicate the real and imaginary parts of the quantities in braces, and $|\dots|^2$ indicates the absolute value squared of the complex quantity. If the particle is non-absorbing, m is real. The first term in Eq. (12.20) is then zero, and the second term is then the same as Eq. (12.21) and Rayleigh's Q_b seen in Eq. (12.19). The Rayleigh scattering coefficient (in the form of either b , σ_b , or Q_b) thus falls out of the first terms of the Mie solution. If x is small enough that terms of order $|m|x$ and higher can be ignored, then the absorption efficiency $Q_a = Q_c - Q_b$ reduces to just

$$Q_a \approx 4x\Im \left\{ \frac{m^2 - 1}{m^2 + 2} \right\}. \quad (12.22)$$

Remembering that $x = 2\pi\rho/\lambda$, then if the $(m^2 - 1)/(m^2 + 2)$ factor is almost independent of wavelength over some wavelength interval, then $Q_a \propto 1/\lambda$ and $Q_b \propto 1/\lambda^4$ over that interval.

12.5.3 Applicability of Rayleigh's Approximation

Rayleigh's scattering result (12.19) was derived for very small, non-absorbing particles. The question remains as to how small is small enough for the Rayleigh formulas to be accurate within some error compared to the exact Mie theory. In particular, can Rayleigh's equations be used to compute scattering by phytoplankton or other oceanic particles? At visible wavelengths, phytoplankton typically have real indices of refraction in the range of 1.02 to 1.1, relative to water (e.g., Ackleson and Spinrad, 1988). The complex index of refraction is in the region of 0.001 at 500 nm up to 0.005 in an absorption band (e.g., Bricaud et al., 1983, Table 1). Figure 12.15 compares the Mie and Rayleigh scattering efficiencies for a typical phytoplankton index of refraction and a wavelength of 500 nm. Suppose we accept a 10% error in Q_b as an acceptable trade-off for the ease of computation. For a size parameter of $x = 0.5$, the Rayleigh Q_b is about 9% too large. Figure 12.16 shows the corresponding difference in phase functions for $x = 0.5$. Again, the maximum difference is about 10% (at $\psi = 0$ and 180 deg). So we could use the Rayleigh formulas for size parameters up to 0.5 for phytoplankton. The problem for oceanography is that a size parameter of $x = 2\pi\rho/\lambda = 0.5$ for $\lambda = 500$ nm gives a particle radius of $\rho = 0.04$ μm , which is an order of magnitude smaller than bacteria or the smallest phytoplankton. Thus the Rayleigh scattering formulas are not useful for computing the scattering coefficients or phase functions for phytoplankton or other oceanographic particles, which are usually of size $\rho = 0.5$ μm or larger.

Comment: Rayleigh was not the first to recognize that scattering by very small particles was a primary contributor to the blue of the sky; he was the first to work out the physics and math. His work captured much, but not all, of the physics of Earth's sky color. We now understand that the "small suspended particles" assumed by Rayleigh are mainly the nitrogen and oxygen molecules that comprise most of the atmosphere. To really understand sky color, in addition to the $1/\lambda^4$ scattering law, one must also take into account the wavelength dependences of the solar spectrum and the response of the human eye. The

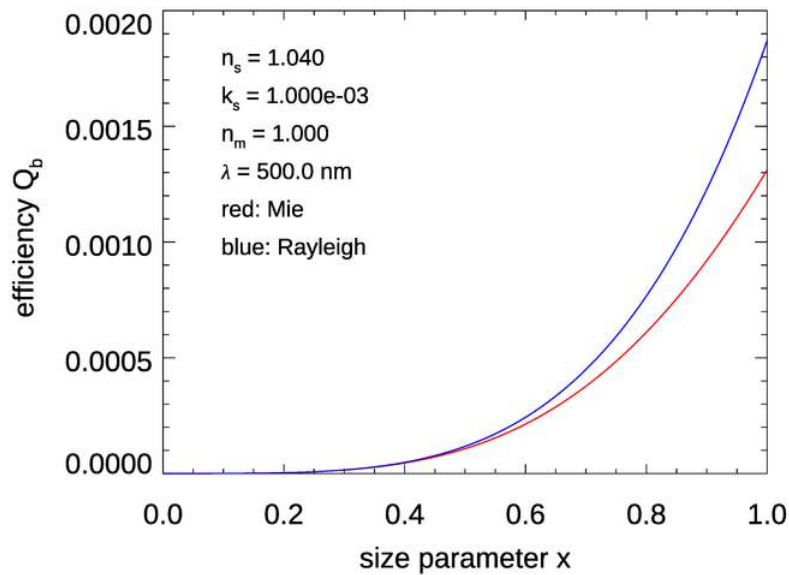


Figure 12.15: Comparison of exact Mie and approximate Rayleigh scattering efficiencies. The red Mie curve is computed numerically from the BHMIE code; the blue Rayleigh curve is from Eq. (12.21), which allows for an absorbing particle. However, Eq. (12.19) for a non-absorbing particle gives essentially the same result (0.01% difference at $x = 1$).

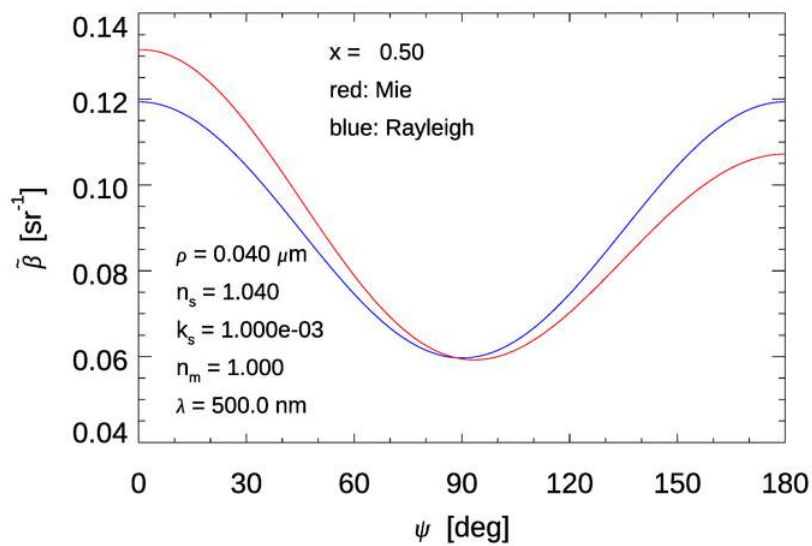


Figure 12.16: Comparison of exact Mie and approximate Rayleigh phase functions. The red Mie curve is computed numerically from a Mie code; the blue Rayleigh curve is from Eq. (12.18).

excellent text by [Bohren and Clothiaux \(2006\)](#) devotes much a chapter to the blue-sky problem, including absorption by ozone at green to red wavelengths. See also [Bohren and Fraser \(1985\)](#) for a non-mathematical summary. That Rayleigh did not totally account for every last contribution to the color of the sky in no way detracts from the brilliance of his work at age 29, in an era when light was still supposed to propagate through a “luminiferous aether” and the very existence of atoms and molecules was disputed by many eminent scientists³.

12.5.4 Weakly Scattering or “Soft” Particles: The Rayleigh-Gans Approximation

A second approximation to Mie theory is available for particles whose complex index of refraction relative to the surrounding medium, m , and size parameter x satisfy two conditions:

$$|m - 1| \ll 1 \tag{12.23}$$

$$\delta = 2x|m - 1| \ll 1. \tag{12.24}$$

These are two independent requirements. $|m - 1| \ll 1$ says that the particle scatters only weakly; there would be no scattering if the real part of the index of refraction were exactly 1. $2x|m - 1| \ll 1$ says that the particle is small enough that there is only a small change in the phase and amplitude of the incident electromagnetic wave as it passes through the particle. That is, the electric field inside the particle is almost the same as that of the incident wave. δ is usually called the *phase shift parameter*. Particles that satisfy Eqs. (12.23) and (12.24) are usually called “optically soft” particles. Recall that Rayleigh’s equations require that the particle size parameter x be small. The Rayleigh-Gans simplification allows x to be large so long as the index of refraction is small enough to satisfy $2x|m - 1| \ll 1$.

To develop the solution, the volume of the particle is divided into small volume elements. The particle does not need to be spherical. Each volume element receives essentially the same incident electromagnetic wave (condition 12.24), which it then scatters according to the Rayleigh approximations for $x \ll 1$. However, there will be phase differences for the scattered waves from different volume elements, which lead to interference effects. These are accounted for via an integration of the phase differences over the volume of the particle. The result of that integration is a non-dimensional *form factor* $G(\psi, \alpha)$, which depends on both the polar (ψ) and azimuthal (α) scattering angles if the particle is non-spherical. The form factor contains all of the information about the shape of the particle.

The scattering-matrix elements have the same general form as those obtained from Mie theory (see Eq.12.4), except for extra factors of G^2 ; see [Bohren and Huffman \(1983, Section 6.1\)](#) for the details. These matrix elements eventually lead to a phase function for scattering of unpolarized light of the form

$$\tilde{\beta}(\psi) = KV^2 |m - 1|^2 G^2(\psi, \alpha) (1 + \cos^2 \psi), \tag{12.25}$$

³The debate about whether atoms and molecules actually exist, or whether they are just convenient mathematical artifices, was quite acrimonious. The great Ludwig Boltzmann became so depressed trying to convince people that his statistical mechanics proved that molecules are real entities that he committed suicide in 1906.

where V is the volume of the particle, and K is the proportionality constant that normalizes the phase function; this is easily computed by numerical integration after the rest of the calculations are performed. This equation holds for any shape of particle.

As noted, $G(\psi, \alpha)$ must be computed by an integration over the volume of the particle. For a homogeneous spherical particle, G depends only on the polar scattering angle, and the integration can be done analytically with the result

$$G(\psi) = \frac{3}{u^3}(\sin u - u \cos u) \quad \text{where} \quad u = 2x \sin \frac{\psi}{2}. \quad (12.26)$$

Although the phase function of Eq. (12.25) contains a $(1 + \cos^2 \psi)$ factor like that of the Rayleigh phase function of Eq. (12.18), the form factor G makes the Rayleigh-Gans phase function much more peaked at small scattering angles.

Let the index of refraction of the particle relative to the surrounding medium be written as $m = n + ik$ and define a parameter ξ by

$$\tan \xi = \frac{k}{n - 1}. \quad (12.27)$$

Note that ξ ranges from 0 (for non-absorbing particles) to ∞ (for absorbing particles with n approaching 1). Then the Rayleigh-Gans extinction and absorption efficiency factors for spherical particles are functions of ξ and the phase shift parameter δ :

$$Q_c = 2 - 4 \exp(-\delta \tan \xi) \frac{\cos \xi}{\delta} \left[\sin(\delta - \xi) + \frac{\cos \xi}{\delta} \cos(\delta - 2\xi) \right] + 4 \left(\frac{\cos \xi}{\delta} \right)^2 \cos(2\xi) \quad (12.28)$$

$$Q_a = 1 + \frac{\exp(-2\delta \tan \xi)(2\delta \tan \xi + 1) - 1}{2\delta^2 \tan^2 \xi}. \quad (12.29)$$

For nonspherical particles, the Rayleigh-Gans scattering efficiency factor depends on the polarization state of the incident light. For a spherical particle, the scattering efficiency $Q_b = Q_c - Q_a$ can be obtained from the preceding two equations. For non-absorbing particles, $\tan \xi = 0$, $\xi = 0$, and Eq. (12.28) reduces to just

$$Q_c = 2 - \frac{4}{\delta} \sin \delta + \frac{4}{\delta^2} (1 - \cos \delta). \quad (12.30)$$

12.5.5 Applicability of the Rayleigh-Gans Approximation

To give an idea of its applicability in oceanography, Fig. 12.17 compares the Rayleigh-Gans efficiency factors with those of the exact Mie numerical calculations for typical phytoplankton IOPs. The Rayleigh scattering or extinction efficiency of Eq. (12.19) is also shown, as is the equivalent Rayleigh absorption efficiency of Eq. (12.22), obtained from the lowest-order term of the Mie series expansion. The figure displays the results as a function of the phase shift parameter $\delta = 2x|m - 1|$. Not surprisingly, the Rayleigh scattering efficiency blows up for $\delta \gtrsim 0.1$, corresponding to $x \gtrsim 1$. However, the Rayleigh-Gans equations do very well for much larger δ values. (For these IOPs, $\delta = 20$ corresponds to a size parameter of $x = 200$.) This is quite unexpected and remarkable given that Rayleigh-Gans theory was developed on the assumption that $\delta \ll 1$. Indeed, [van de Hulst \(1981, page 176\)](#) comments on Eq. (12.30) that “This is one of the most useful formulae in the whole domain of the

Mie theory, because it describes the salient features of the extinction curve not only for m close to 1 but even for values of m as large as two.” The same holds for Eqs. (12.28) and (12.29).

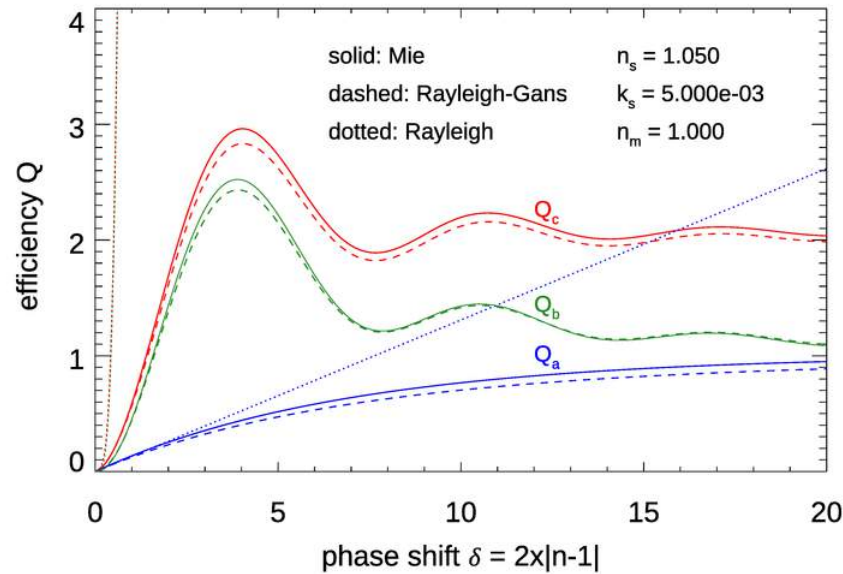


Figure 12.17: Comparison of exact Mie, Rayleigh-Gans, and Rayleigh efficiencies for phytoplankton IOPs.

Figure 12.18 shows the efficiencies for the IOPs typical of soot; recall Fig. 12.11. Even for this high-index-of-refraction, highly absorbing particle, Rayleigh-Gans does amazingly well. (For these IOPs, $\delta = 20$ corresponds to a size parameter of $x = 18.45$.)

Because of their wide range of validity, the Rayleigh-Gans efficiency formulas have been used to gain insight into phytoplankton optical properties. Examples are Bricaud et al. (1983) and Gordon (2007).

Now consider how well Rayleigh-Gans phase functions computed by Eqs. (12.25) and (12.26) compare with the exact Mie phase functions. Picoplankton have diameters on the order of $1\mu\text{m}$, or $\rho = 0.5\mu\text{m}$; nanoplankton have diameters of order $10\mu\text{m}$. For a wavelength of $\lambda = 500\text{ nm}$, these give size parameters of $x = 6.28$ and 62.8 , respectively. Figure 12.19 compares exact Mie and Rayleigh-Gans phase functions for these size parameters and indices of refraction that are very near 1, $m = 1.01 + i0.0001$; typical of phytoplankton in water, $m = 1.05 + i0.005$; and typical of soot in air, $m = 1.54 + i0.044$. For the very low index (top two figures), the envelopes of the maximum values of the Rayleigh-Gans phase function are close to those of the Mie phase functions. For the large particle (upper right plot), the peaks are somewhat out of phase for large scattering angles, but this is probably acceptable for most applications because a disperse range of sizes would cause the individual interference features to average out, leaving only the upper bound of the curves (recall Fig. 12.10). By the same argument, the Rayleigh-Gans phase function for the small size, typical index particle (middle left figure) would be acceptable for many applications. However, the large-phytoplankton phase functions (middle right figure) differ by an order of magnitude over almost the entire range of scattering angles. Averaging over a range

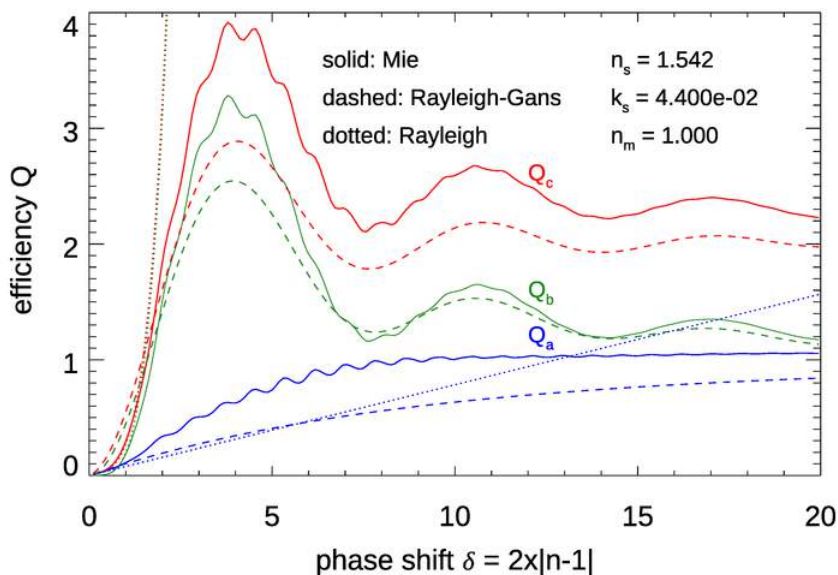


Figure 12.18: Comparison of exact Mie, Rayleigh-Gans, and Rayleigh efficiencies for soot IOPs.

of sizes will not bring those curves together. Similarly, for the high-index soot particles (bottom row), Mie and Rayleigh-Gans differ by one or two orders of magnitude for most scattering angles. Thus it seems that, although Rayleigh-Gans efficiency factors perform well beyond the range of their derivation, the Rayleigh-Gans phase functions cannot be pushed as far.

12.5.6 Large Particles, $x \gg 1$: Geometric Optics

At the other end of the size spectrum are particles that are much larger than the wavelength of the light incident onto them. This is the realm of geometrical optics and ray tracing. The needed tools are simply Snell's law, Fresnel's law, and a computer. Geometric ray tracing can compute the optical properties, phase functions in particular, for any shape of particle, but at the expense of missing any effects due to diffraction or interference. For visible wavelengths, diffraction can be significant for particles in the size range of most phytoplankton, $\rho \lesssim 10 \mu\text{m}$ or $x \lesssim 100$. For diffraction to appear insignificant compared to reflection and refraction, particle sizes need to be of order 0.05 mm or larger, which gives a size parameter of $x \gtrsim 1000$. Thus ray tracing is seldom used of computing the phase functions of oceanic particles. However, ray tracing is commonly used to compute the reflectance and transmission properties of wind-blown sea surfaces (e.g., Mobley, 2015) or of underwater objects or surfaces (e.g., Mobley, 2018).

In geometric optics ray tracing, a ray that misses a particle by even the smallest distance continues onward unperturbed. Consider a particle that is so highly absorbing that it absorbs all of the light that hits it. Then no light will be scattered. The absorption and scattering efficiencies are then $Q_a = 1$ and $Q_b = 0$, so that the extinction efficiency $Q_c = Q_a + Q_b = 1$. Conversely, if the particle is non-absorbing, all light that hits it will

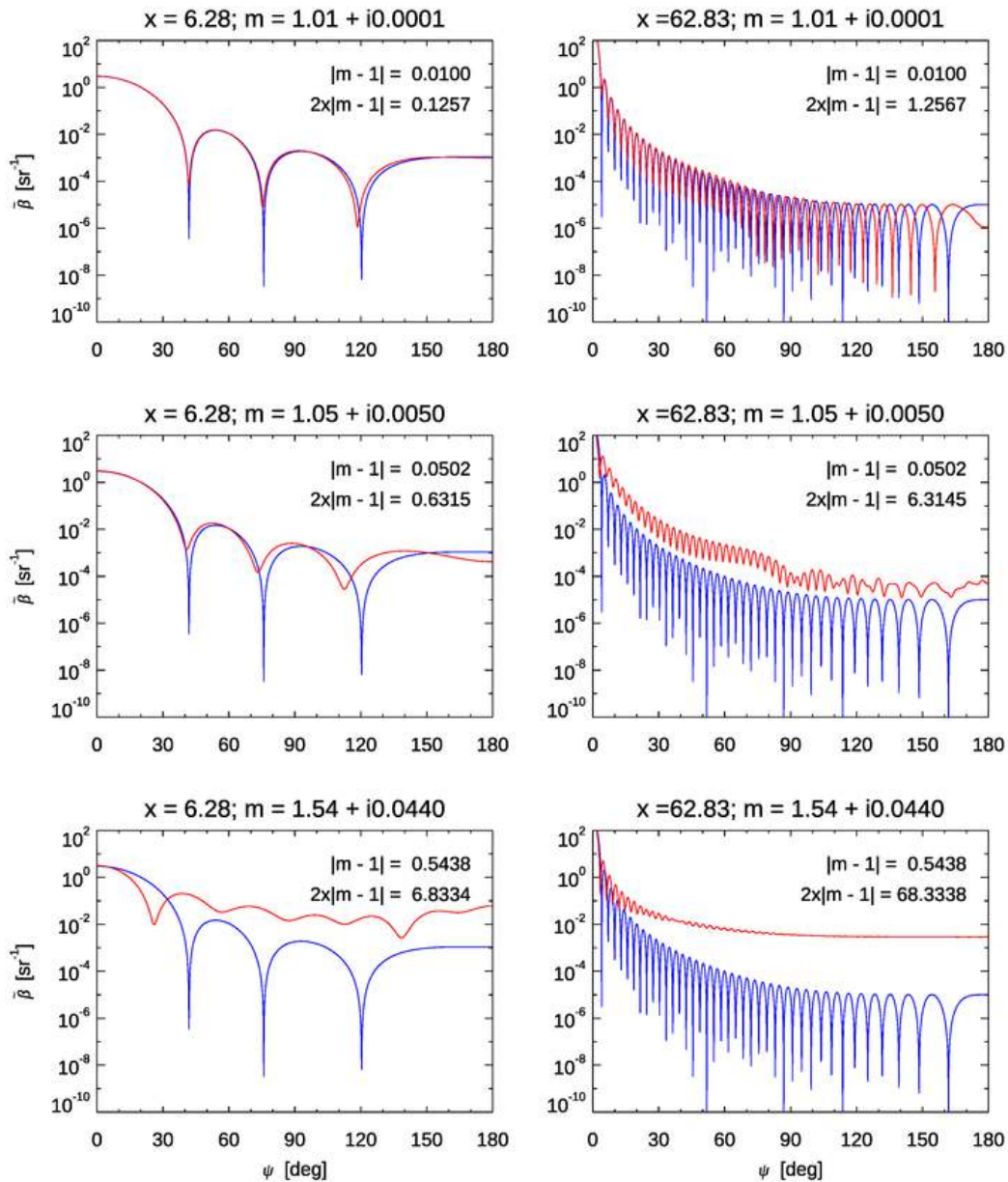


Figure 12.19: Comparison of exact Mie (red) and Rayleigh-Gans (blue) phase functions. Left column: a picoplankton-sized particle; right column: a nanoplankton-sized particle. Top row: a very low index of refraction, $|m| = 1.01$; middle row, typical phytoplankton index of refraction, $|m| = 1.05$; bottom row, soot index of refraction, $|m| = 1.54$.

be scattered. Then $Q_a = 0$, $Q_b = 1$, and $Q_c = 1$. This is the origin of the “extinction paradox,” which was discussed on the preceding section. There we saw that Mie theory leads an asymptotic large-particle value of $Q_c = 2$, not 1. The difference is that Mie theory fully accounts for diffraction and interference effects, and geometric optics does not.

Incidentally, if you want to convert a geometric optics efficiency Q into a cross section σ , there is a wonderful result, Cauchy's Average Projected Area Theorem, which shows that for a convex polyhedron, the average projected area over all orientations, i.e. the average cross section, is one-fourth the surface area of the polyhedron. (This result is obvious only for a sphere, whose surface area is $4\pi r^2$ and whose cross section is πr^2 .) If $\langle A \rangle$ is the average area of the particle as seen from all orientations, the the cross section is given by $\sigma = Q\langle A \rangle$, regardless of the particle shape (as long as it is convex, i.e., with any "indentations" in its surface).

12.5.7 A Geometric Optics Ray Tracing Example

The use of ray tracing to explain rainbows goes back to Descartes in the early 1600s. Another area of geophysical optics where ray tracing has proved very useful is the computation of phase functions for atmospheric ice crystals in cirrus clouds. These clouds are important for modeling the Earth's radiation balance. Ice has a hexagonal crystal structure, so cloud ice crystals commonly form as hexagonal plates or solid or hollow columns, although more complex shapes (e.g., snowflakes and columns with pyramidal caps or indentations) can form, depending on the temperature and humidity during formation.

Figure 12.20 illustrates possible ray paths through a solid hexagonal column of ice, which is very common in cirrus clouds. Such columns are on the order of 0.05 to 0.5 mm in size. When ray tracing through such columns, millions of rays are traced for randomly oriented columns. The reflected and refracted rays for different directions are then used to build up, ray by ray, the shape of the scattering phase function. Certain directions, such as those shown by the red and green arrows in Fig. 12.20, tend to generate caustics, or "collections" of rays near particular directions (just as happens for spherical water drops in the formation of rainbows).

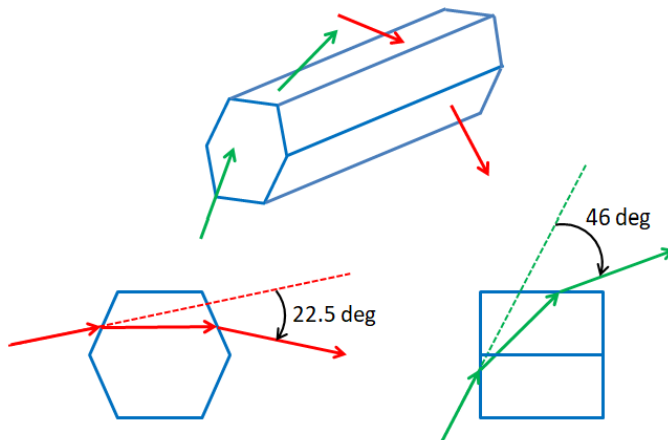


Figure 12.20: Illustration of rays passing through the sides and ends of a solid hexagonal ice crystal. Rays generating the 22.5 and 46 degree cirrus cloud halos are shown.

Figure 12.21 shows the scattering phase function for cirrus clouds containing hexagonal columns like those of Fig. 12.20. The prominent peaks near 22 and 46 degrees give halos or "rings around the Sun" (or Moon) at those angles. The 22 deg halo is common. The 46 deg halo, resulting from rays passing through the flat ends of hexagonal column, is seen

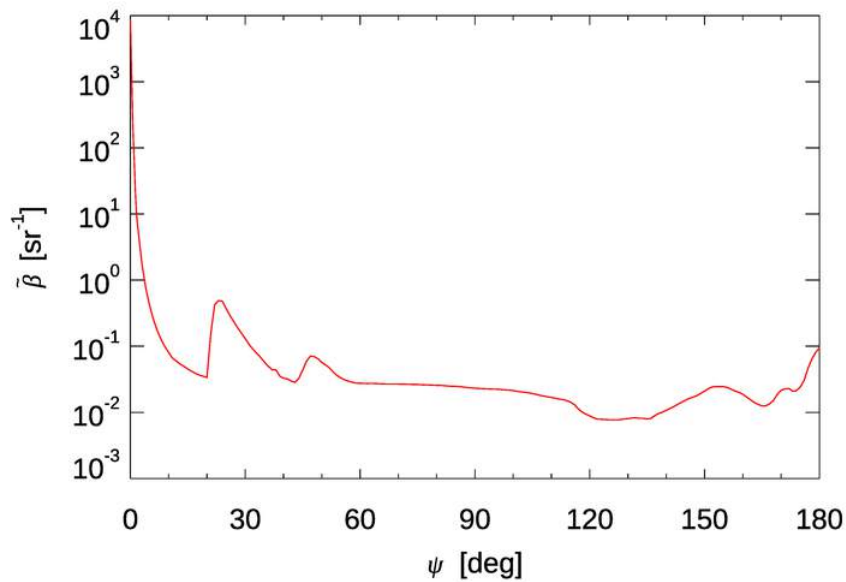


Figure 12.21: Scattering phase function $\tilde{\beta}(\psi)$ for hexagonal column cirrus cloud ice crystals at $\lambda = 550$ nm. Data of [Takano and Liou \(1989, Table 1\)](#) divided by 4π for normalization.

less often. The peak near 160 deg results from two internal reflections in the crystal. The left panel of Fig. 6.1 shows a spectacular 22 deg halo. Note how well the phase function of Fig. 12.21 matches the halo: There is a very sharp transition from darker to brighter on the inner side of the halo, where the phase function rises very rapidly between $\psi = 20$ and 22 deg. Then there is a slower decrease in brightness on the outer side of the halo, corresponding to the slower decay of the phase function between 23 and 44 degrees. There is no 46 deg halo in this image. That will be the case if the hexagonal crystals do not have the flat ends needed to create the 46 deg peak in the phase function, as illustrated by the green rays in Fig. 12.20. The color in the halo results from the small difference in the ice index of refraction as a function of wavelength.

A myriad of other halos can be generated by ice crystals of other shapes. Some of these crystal shapes rarely form and the resulting halos are almost “once-in-a-lifetime” events. An excellent book showing photographs of many types of halos, rainbows, glories, and other atmospheric phenomena, along with explanations of their causes, is [Greenler \(2020\)](#). There are free ray tracing codes for halo simulation now available, e.g., [HaloSim 3](#) and [HaloPoint 2](#).

12.5.8 Closing Thoughts

Rayleigh developed his scattering theory in the late 1800s, and it worked quite well to solve the problem of interest to him. Rayleigh-Gans theory arose in the late 1800’s and early 1900’s. (Gans worked out the theory for homogeneous spheres in 1925.) In those pre-computer days, use of the exact Mie theory was not possible, so analytical approximations were the only option for computation of scattering properties of particles. As we have seen, Rayleigh scattering theory is not applicable to scattering by oceanic particles like

phytoplankton or minerals, which are too large. Rayleigh-Gans theory does have some usefulness, but it too has its limits. Geometric optics can be useful for atmospheric particles like large ice crystals, but oceanic particles tend to be too small for geometric optics. Thus the approximations surveyed in this section are of interest, but they do not find frequent application in oceanic optics. There are additional analytical approximations for spherical particles, but these have found little if any application in optical oceanography. Those approximations are surveyed in [Kokhanovsky and Zege \(1997\)](#). If you have spherical particles, it is usually feasible with modern computers to do numerical Mie calculations without the need for approximations. For non-spherical particles, there are other numerical methods that can be used (T-matrix theory or the discrete dipole approximation).

It should be noted that this section refers to “Rayleigh’s approximation” and “the Rayleigh-Gans approximation,” rather than to “Rayleigh scattering,” or “Rayleigh-Gans scattering,” as is often seen. This is to emphasize that “Rayleigh scattering,” “Rayleigh-Gans scattering,” and Geometric optics are not physically different types of scattering. They are *approximate mathematical models* for computing scattering quantities such as cross-sections and phase functions in particular size and index-of-refraction domains. As was explained in the Physics of Scattering, Section [6.2](#), all scattering is caused by a change in the real index of refraction and in that sense all scattering is fundamentally the same.

Surfaces

Chapter 9 on radiative transfer theory developed the equations that govern light propagation within a water body. This chapter now develops the equations that describe how light is reflected and transmitted by the surfaces that bound the water body. These surfaces include the wind-blown sea surface and, in shallow water, an opaque sea floor of sediment or vegetation. There may also be objects within the water column that reflect light.

The discussion begins with the basics of reflection and transmission of unpolarized light by a level or flat water surface. The case of polarized light is then considered. Although the ocean is rarely glassy smooth, reflection and transmission by rough, wind-blown surfaces are modeled using the Fresnel equations for a flat surface applied to each small patch of sea surface, which although tilted from the normal to the mean sea surface can be assumed to be locally flat. The next sections then introduce the bidirectional reflectance distribution function or BRDF. The BRDF is the fundamental quantity for specifying how an opaque surface reflects light. The chapter closes with a section on the various measures of surface reflectance that are encountered in the literature.

Appendices A-D discuss the advanced topics of how random sea surfaces can be described in terms of wave energy spectra or autocovariance functions and, conversely, how random sea surfaces can be generated starting with wave energy spectra or autocovariances.

13.1 The Level Sea Surface

The wavelength of visible light is much, much less than the millimeter and larger spatial wavelengths of the waves on wind-blown sea surfaces. Therefore, the laws of geometrical optics and the idealization of a narrow ray of collimated light give a good description of the relevant physical processes.

13.1.1 Geometric Relations

Figure 13.1 illustrates a level surface with light incident onto the surface from the air side (panel a), and from the water side (panel b). The real index of refraction of the air is n_a , which is taken to be one. n_w is the real index of refraction of the water, which is weakly wavelength dependent at visible wavelengths (Section 8.2.1.1). However, for many

calculations n_w can be taken to be approximately 1.34 at visible wavelengths. $\hat{\mathbf{n}}$ is a unit vector normal to the surface. $\hat{\boldsymbol{\xi}}_i$ is a unit vector in the incident direction; $\hat{\boldsymbol{\xi}}_r$ and $\hat{\boldsymbol{\xi}}_t$ are respectively the directions of the reflected and transmitted rays. $\theta_i = \cos^{-1}(|\hat{\boldsymbol{\xi}}_i \cdot \hat{\mathbf{n}}|)$ is the acute angle between the incident direction and the normal, and θ_r and θ_t are the angles of the reflected and transmitted rays relative to the normal.

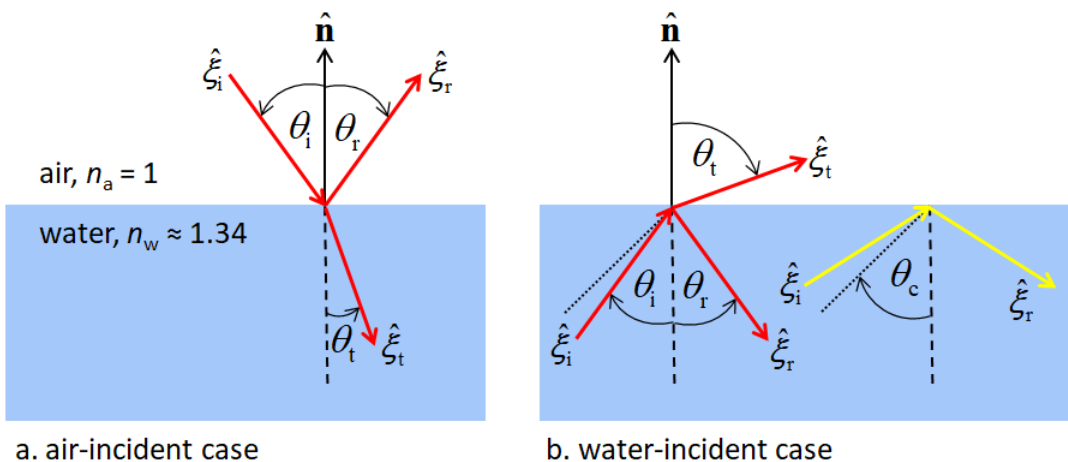


Figure 13.1: Illustration of incident, reflected, and transmitted rays for air- and water-incident light.

The incident, reflected, and refracted directions all lie in the plane defined by $\hat{\boldsymbol{\xi}}_i$ and $\hat{\mathbf{n}}$. The reflected angle is always equal to the incident angle: $\theta_r = \theta_i$, which is known as the *Law of Reflection*. The incident and transmitted angles are related by

$$n_1 \sin \theta_1 = n_2 \sin \theta_2, \quad (13.1)$$

where subscripts 1 and 2 refer to any two media. This equation is usually called Snell's law, although more properly it should be Snel's law (see the footnote on page 174).

Figure 13.2 gives a visual representation of the relations between the various unit vectors, angles, and indices of refraction.

For air-incident light, $n_a = 1$ and Snell's law reads $\sin \theta_i = n_w \sin \theta_t$. Then the angle of transmission is given by

$$\theta_t = \sin^{-1} \left(\frac{1}{n_w} \sin \theta_i \right). \quad (13.2)$$

The relations between the unit vectors are given by the following equations (with $n_a = 1$):

$$\hat{\boldsymbol{\xi}}_r = \hat{\boldsymbol{\xi}}_i - 2(\hat{\boldsymbol{\xi}}_i \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}, \quad (13.3)$$

and

$$\hat{\boldsymbol{\xi}}_t = \frac{1}{n_w}(\hat{\boldsymbol{\xi}}_i - c\hat{\mathbf{n}}), \quad (13.4)$$

where

$$c = \hat{\boldsymbol{\xi}}_i \cdot \hat{\mathbf{n}} + \sqrt{(\hat{\boldsymbol{\xi}}_i \cdot \hat{\mathbf{n}})^2 + n_w^2 - 1}. \quad (13.5)$$

It should be noted from Eq. (13.2) that if the incident ray is nearly parallel to the water surface, $\theta_i \approx 90$ deg, then the angle of the transmitted ray is $\theta_t \approx \sin^{-1}(1/n_w)$. For

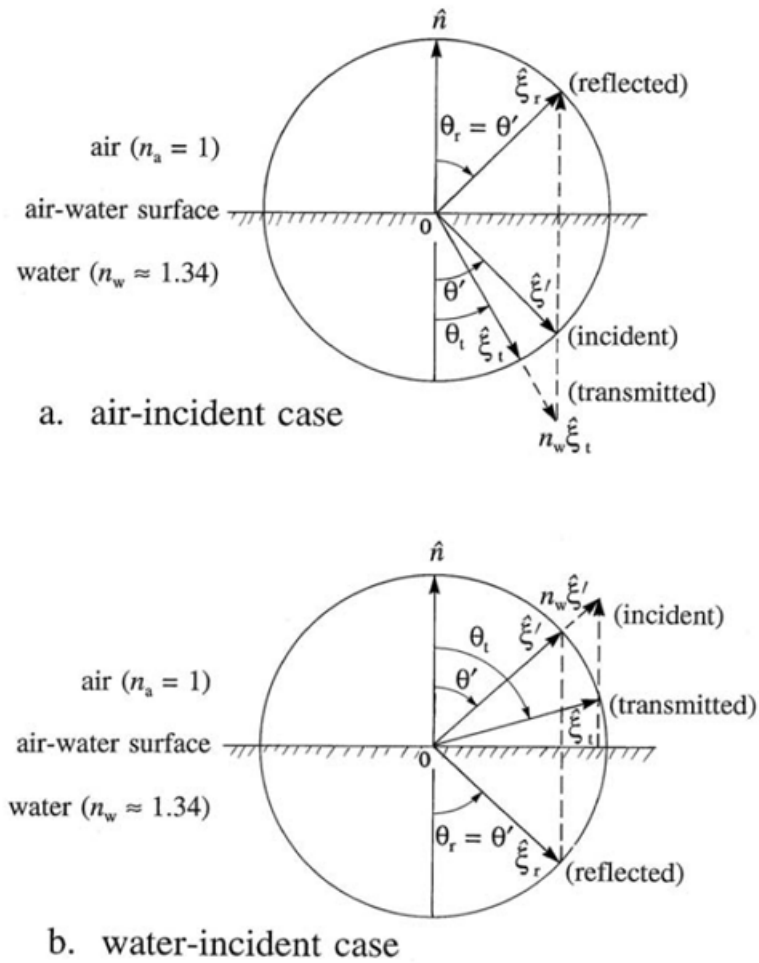


Figure 13.2: Graphical representations of relations among incident, reflected, and transmitted directions. Reproduced from Fig. 4.2 of Mobley (1994), where $\hat{\xi}_i = \hat{\xi}'_i$.

$n_w = 1.34$, this gives $\theta_t \approx 48$ deg. Thus the entire hemisphere of sky directions is mapped into an underwater cone of half angle 48 deg, as illustrated in Fig. 13.3. This phenomenon is known as “Snell’s cone,” or the “optical manhole.”

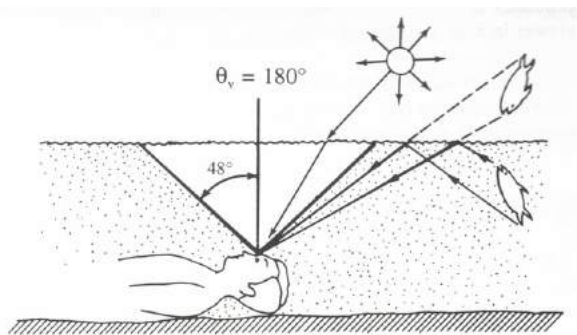


Figure 13.3: Illustration of Snell’s cone. Based on Preisendorfer (1976, Fig. 1.18)

For the water-incident case, Snell's law reads $n_w \sin \theta_i = \sin \theta_t$, in which case the angle of transmission is given by

$$\theta_t = \sin^{-1}(n_w \sin \theta_i). \quad (13.6)$$

The relations between the unit vectors are then

$$\hat{\xi}_r = \hat{\xi}_i - 2(\hat{\xi}_i \cdot \hat{n})\hat{n}, \quad (13.7)$$

and

$$\hat{\xi}_t = n_w \hat{\xi}_i - c\hat{n}, \quad (13.8)$$

where

$$c = n_w \hat{\xi}_i \cdot \hat{n} - \sqrt{(n_w \hat{\xi}_i \cdot \hat{n})^2 - n_w^2 + 1}. \quad (13.9)$$

If $n_w \sin \theta_i < 1$, Eq. (13.6) gives a real value for θ_t and light is transmitted from the water to the air. However, if θ_i is greater than the *critical angle for total internal reflection*

$$\theta_c = \sin^{-1}(1/n_w), \quad (13.10)$$

then there is no real solution for the inverse sine. In this case, all light incident onto the water side of the air-water surface is reflected back into the water. This is called *total internal reflection*. The dotted line in the right panel of Fig. 13.1 represents the critical angle. The red unit vectors illustrate the case $\theta_i < \theta_c$ with both reflected and transmitted light, and the yellow vectors represent the case of $\theta_i > \theta_c$ and total internal reflection. As shown in Fig. 13.3, total internal reflection by the underside of a water surface makes the image of an underwater object appear inverted when seen reflected from the water side of the sea surface.

13.1.2 The n^2 Law for Radiance

Snell's law yields an important result governing how unpolarized radiance changes when going from one medium to another, e.g., when crossing an air-water surface. Figure 13.4 shows two beams of radiance, one incident onto an interface and one transmitted. Let L_1 be the incident radiance in medium 1 defined by power Φ_1 passing through an area ΔA_1 normal to the direction of photon travel and contained in a solid angle $\Delta\Omega_1 = \sin \theta_1 \Delta\theta_1 \Delta\phi_1$, where θ_1 is polar angle measured relative to the normal to the surface and $\Delta\phi_1$ is the width of the solid angle in the azimuthal direction. Likewise, L_2 is the transmitted radiance in medium 2 defined by the corresponding quantities as illustrated. The azimuthal angle does not change when crossing the surface, so $\Delta\Omega_2 = \sin \theta_2 \Delta\theta_2 \Delta\phi_1$. The incident and transmitted power passes through a common area ΔA at the interface.

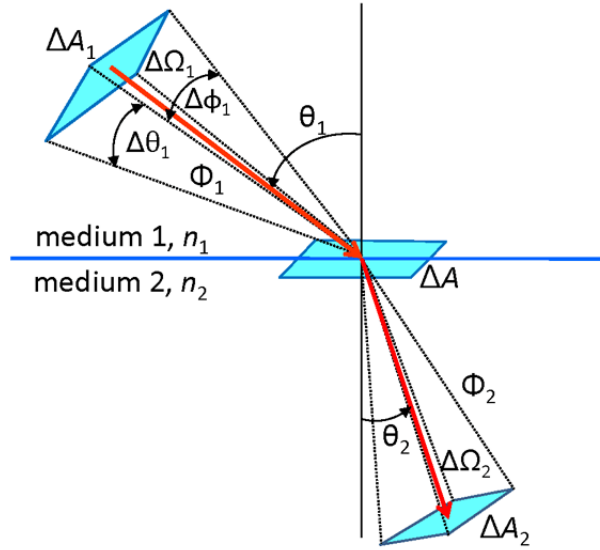
The indices of refraction n_1 and n_2 are fixed, but the polar angle θ changes when crossing the interface. Squaring Eq. (13.1) and differentiating gives

$$n_1^2 \sin \theta_1 \cos \theta_1 \Delta\theta_1 = n_2^2 \sin \theta_2 \cos \theta_2 \Delta\theta_2.$$

Multiplying each side of this equation by the common value of $\Delta\phi$ and rewriting in terms of solid angles gives

$$n_1^2 \cos \theta_1 \Delta\Omega_1 = n_2^2 \cos \theta_2 \Delta\Omega_2,$$

which is known as *Straubel's invariant*.

Figure 13.4: Geometry for deriving the n^2 law for radiance.

The radiances are defined by

$$L_1 = \frac{\Delta\Phi_1}{\Delta A_1 \Delta\Omega_1} \quad \text{and} \quad L_2 = \frac{\Delta\Phi_2}{\Delta A_2 \Delta\Omega_2}.$$

Fresnel's equation (13.11) gives the transmitted power as $\Delta\Phi_2 = [1 - R_F(\theta_1)]\Delta\Phi_1 = T_F\Delta\Phi_1$. The areas are related by $\Delta A_1 = \Delta A \cos \theta_1$ and $\Delta A_2 = \Delta A \cos \theta_2$. Thus the ratios of the incident and transmitted radiances can be written as

$$\begin{aligned} \frac{L_2}{L_1} &= \frac{\Delta\Phi_2}{\Delta\Phi_1} \frac{\Delta A_1 \Delta\Omega_1}{\Delta A_2 \Delta\Omega_2} \\ &= T_F \frac{\cos \theta_1 \Delta\Omega_1}{\cos \theta_2 \Delta\Omega_2} \\ &= T_F \frac{n_2^2}{n_1^2} \end{aligned}$$

or

$$\frac{L_2}{n_2^2} = T_F \frac{L_1}{n_1^2}.$$

This result is called the *n-squared law for radiance*. The quantity L/n^2 is sometimes called the *reduced radiance* or the *basic radiance*.

Although energy is conserved when crossing a boundary, the radiance changes by a factor proportional to the change in the index of refraction squared. This is a simple consequence of the change in solid angle resulting from the change in θ when crossing the boundary. Note that for normal incidence and $n_w = 1.34$, $T_F \approx 0.979$ and the radiance just below a water surface is $0.979(1.34)^2 \approx 1.76$ times the radiance in the air. Conversely, when going from water to air, the in-water radiance is reduced by a factor of 1.76.

To the extent that losses to absorption and scattering out of the beam can be ignored (sometimes a good approximation for atmospheric transmission, but almost never the case in water), the radiance divided by the square of the index of refraction is constant along any path. This result has even been called the *Fundamental Theorem of Radiometry*, which is perhaps a bit grandiose given that real beams always lose at least some radiance due to absorption and can lose or gain radiance due to scattering.

Finally, note that the n^2 law applies only to radiance transmission. When tracing rays in a Monte Carlo simulation, from which the radiance can be estimated by appropriate binning of the transmitted rays, no n^2 factor is applied to the energy of the transmitted rays or to the radiance estimated from the detected rays. This is because the n^2 effect is automatically built into the radiance estimate ray by ray as the directions of the individual rays are computed by Snell's law.

13.2 Fresnel's Equations for Unpolarized Light

The equations of the previous section show the relations between the angles and directions of the incident and the reflected and transmitted light. However, they do not show how much energy is reflected or transmitted. That information is given by Fresnel's equations.

Consider first a collimated beam of unpolarized incident light, which has some irradiance measured on a surface normal to the direction $\hat{\xi}_i$ of propagation. The fraction of this incident irradiance that is reflected by the air-water surface is (e.g., [Hapke \(1993, Section 4.C\)](#) or [Hecht \(1989, Section 4.3\)](#))

$$R_F(\theta_i) = \frac{1}{2} \left\{ \left[\frac{\sin(\theta_i - \theta_t)}{\sin(\theta_i + \theta_t)} \right]^2 + \left[\frac{\tan(\theta_i - \theta_t)}{\tan(\theta_i + \theta_t)} \right]^2 \right\}, \quad (13.11)$$

which holds for $\theta_i \neq 0$. For normally incident light, $\theta_i = 0$, the reflectance is¹

$$R_F(\theta_i = 0) = \left(\frac{n_w - 1}{n_w + 1} \right)^2. \quad (13.12)$$

Equations (13.11) and (13.12) hold for both air- and water-incident light. Given the incident angle θ_i , the transmitted angle θ_t is computed using either Eq. (13.2) or (13.6), and then Eq. (13.11) (or 13.12) can be evaluated. For water-incident light and $\theta_i \geq \theta_c$, $R_F = 1$. Figure 13.5 shows the Fresnel reflectance for the range of water indices of refraction at visible wavelengths.

To be completely general, the Fresnel equations should use the complex index of refraction $m = n + ik$, where n is the real index of refraction seen above and $k(\lambda) = \lambda a(\lambda)/2\pi$ is the complex part (a is the absorption coefficient). Thus Eq. (13.12) should be

$$R_F(\theta_i = 0) = \left| \frac{m - 1}{m + 1} \right|^2 = \frac{(n_w - 1)^2 + k^2}{(n_w + 1)^2 + k^2}.$$

However, for water at near-UV to near-IR wavelengths, $k < 10^{-6}$ and the difference is negligible. However k can be of order 0.1 to 1 at some UV and far-IR wavelengths, in which case the complex index of refraction must be used.

¹In general, $R_F(\theta_i = 0) = \left(\frac{n_1 - n_2}{n_1 + n_2} \right)^2$.

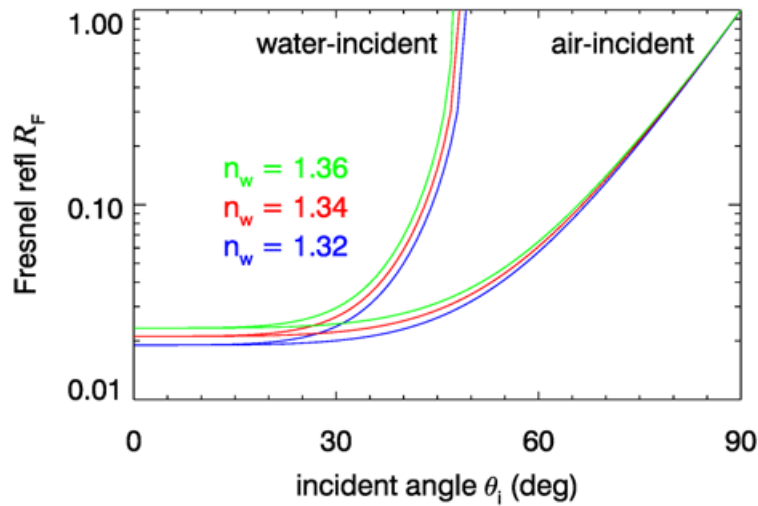


Figure 13.5: Fresnel reflectance for unpolarized light and selected water indices of refraction.

Conservation of energy requires that the sum of the reflected and transmitted energy equal the incident energy. Thus the fraction of the incident energy that is transmitted is $T_F = 1 - R_F$. It can be confusing to see that energy is conserved when different quantities such as plane irradiance, scalar irradiance, or radiance are used to describe the light, or when the incident light is not a single collimated beam. This is discussed in detail Section 10.8 on energy conservation.

13.3 Fresnel's Equations for Polarized Light

Reflection and transmission are much more complicated when the incident light is polarized. The geometry is the same as for the preceding discussion of Fresnel reflectance and transmittance of unpolarized light by a level sea surface. Now, however, the state of polarization of the incident light is described by a four-component Stokes vector, as described in Section 1.6. Consequently, reflection and transmission by the surface are described by 4×4 matrices.

The state of polarization of a light field is specified by the four-component Stokes vector, whose elements are related to the complex amplitudes of the electric field vector \mathbf{E} resolved into components that are parallel (E_{\parallel}) and perpendicular (E_{\perp}) to a conveniently chosen reference plane. As explained in Section 1.7, there are two versions of the Stokes vector, and these two versions have different units and refer to different physical quantities. The *coherent Stokes vector* describes a quasi-monochromatic plane wave propagating in one exact direction, and the vector components have units of power per unit area (i.e., irradiance) on a small surface element perpendicular to the direction of propagation. The *diffuse Stokes vector* describes light propagating in a small set of directions surrounding a particular direction and has units of power per unit area per unit solid angle (i.e., radiance). It is the diffuse Stokes vector that appears in the vector radiative transfer equation. The differences in coherent and diffuse Stokes vectors are rigorously discussed in [Mishchenko](#)

(2008b).

For either air- or water-incident light, \mathbb{S}_i denotes the diffuse Stokes vector of the incident light, \mathbb{S}_r is the reflected light, and \mathbb{S}_t is the transmitted light. Angles θ_i , θ_r , and θ_t are the incident, reflected, and transmitted directions of the light propagation measured relative to the normal to the surface. For a level surface, \mathbb{S}_i , \mathbb{S}_r , and \mathbb{S}_t all lie in the same plane as seen in Fig. 13.1.

There are four matrices to describe reflection and transmission: \mathbb{R}_{aw} describes how air-incident light is reflected by the water surface back to the air, \mathbb{T}_{aw} describes how air-incident light is transmitted through the surface into the water, \mathbb{R}_{wa} reflects water-incident light back to the water, and \mathbb{T}_{wa} transmits light from the water into the air. However, because \mathbb{S}_i , \mathbb{S}_r , and \mathbb{S}_t are coplanar, scattering by the level surface does not involve rotation matrices as does scattering within the water body. (Or, from another viewpoint, the incident and final meridian planes and the scattering plane are all the same, the rotation angles between meridian and scattering planes are 0, and the rotation matrices reduce to identity matrices.)

The reflection and (especially) transmission of polarized light by a dielectric surface such as a level water surface are rather complicated processes, and the literature contains a number of different (and, indeed, sometimes incorrect) mathematical formulations of the equations. The formulas given in Garcia (2012) are used here. Note, however, that although the equations in Garcia (2012) are correct, some of his derivations and interpretations are incorrect, as explained by Zhai et al. (2012). Both papers must be used to understand the equations now presented. The equations in Garcia will be referenced by (G21) and so on; the corresponding equations in Zhai et al. (2012) will be referenced as (Z5), etc.

The reflectance and transmittance matrices have a general formulation for the interface between any two dielectric media a and b . Let n_a be the index of refraction of medium a and n_b be that of medium b . In general n_a and n_b are complex numbers, but for the air-water surface we take $n_{air} = 1$ and $n_{water} \approx 1.34$ to be real indices of refraction. For reflection, the reflected angle θ_r equals the incident angle θ_i . For transmission from a to b , the transmitted angle is given by Snell's law, $n_a \sin \theta_a = n_b \sin \theta_b$, or

$$\theta_b = \sin^{-1} \left(\frac{n_a \sin \theta_i}{n_b} \right). \quad (13.13)$$

For water-incident light, $n_a = n_{water}$ and $n_b = n_{air}$, in which case the transmitted angle becomes undefined beyond the *critical angle for total internal reflection*, which for water is $\theta_c = \sin^{-1}(1/n_{water}) \approx 48$ deg. For water-incident angles greater than θ_c the incident light is totally reflected back to the water and no light is transmitted to the air.

Let \mathbb{R}_{ab} denote the reflectance matrix for light incident from medium a and reflected back by medium b . \mathbb{R}_{ab} thus represents either \mathbb{R}_{aw} or \mathbb{R}_{wa} . Likewise, let \mathbb{T}_{ab} denote the reflectance matrix for light incident from medium a and transmitted through the surface into medium b . \mathbb{T}_{ab} thus represents either \mathbb{T}_{aw} or \mathbb{T}_{wa} .

With these preliminaries, the reflectance matrix \mathbb{R}_{ab} is (G10)

$$\mathbb{R}_{ab} = \begin{bmatrix} \frac{1}{2}(R_{\parallel}R_{\parallel}^* + R_{\perp}R_{\perp}^*) & \frac{1}{2}(R_{\parallel}R_{\parallel}^* - R_{\perp}R_{\perp}^*) & 0 & 0 \\ \frac{1}{2}(R_{\parallel}R_{\parallel}^* - R_{\perp}R_{\perp}^*) & \frac{1}{2}(R_{\parallel}R_{\parallel}^* + R_{\perp}R_{\perp}^*) & 0 & 0 \\ 0 & 0 & \Re\{R_{\parallel}R_{\perp}^*\} & \Im\{R_{\parallel}R_{\perp}^*\} \\ 0 & 0 & -\Im\{R_{\parallel}R_{\perp}^*\} & \Re\{R_{\parallel}R_{\perp}^*\} \end{bmatrix}. \quad (13.14)$$

Here $\Re\{R_{\parallel}R_{\perp}^*\}$ denotes the real part of $R_{\parallel}R_{\perp}^*$ and $\Im\{R_{\parallel}R_{\perp}^*\}$ is the imaginary part.

The transmission matrix \mathbb{T}_{ab} is (G11 or Z3)

$$\mathbb{T}_{ab} = f_{\text{T}} \begin{bmatrix} \frac{1}{2}(T_{\parallel}T_{\parallel}^* + T_{\perp}R_{\perp}^*) & \frac{1}{2}(T_{\parallel}T_{\parallel}^* - T_{\perp}T_{\perp}^*) & 0 & 0 \\ \frac{1}{2}(T_{\parallel}T_{\parallel}^* - T_{\perp}R_{\perp}^*) & \frac{1}{2}(T_{\parallel}T_{\parallel}^* + T_{\perp}T_{\perp}^*) & 0 & 0 \\ 0 & 0 & \Re\{T_{\parallel}T_{\perp}^*\} & \Im\{T_{\parallel}T_{\perp}^*\} \\ 0 & 0 & -\Im\{T_{\parallel}T_{\perp}^*\} & \Re\{T_{\parallel}T_{\perp}^*\} \end{bmatrix}. \quad (13.15)$$

The components of these equations are given by (G7):

$$R_{\parallel} = \frac{n_b \cos \theta_a - n_a \cos \theta_b}{n_b \cos \theta_a + n_a \cos \theta_b} \quad (13.16)$$

$$R_{\perp} = \frac{n_a \cos \theta_a - n_b \cos \theta_b}{n_a \cos \theta_a + n_b \cos \theta_b} \quad (13.17)$$

$$T_{\parallel} = \frac{2n_a \cos \theta_a}{n_b \cos \theta_a + n_a \cos \theta_b} \quad (13.18)$$

$$T_{\perp} = \frac{2n_a \cos \theta_a}{n_a \cos \theta_a + n_b \cos \theta_b}. \quad (13.19)$$

The factor f_{T} is defined below in Eq. (13.35). In general, the indices of refraction are complex numbers and these equations must be used. However, for real indices of refraction, the matrix elements can be simplified at the expense of having a special case for water-incident angles greater than the critical angle.

Define

$$n_{ab} = \frac{n_a}{n_b} \quad \text{and} \quad n_{ba} = \frac{n_b}{n_a}. \quad (13.20)$$

Then for the case of air-incident light, i.e., $n_a \leq n_b$, or water-incident light with the incident angle less than the critical angle, i.e., $n_a > n_b$ and $\theta_a < \theta_c$, the equations yield the real forms (G14 and G15)

$$R_{\parallel}R_{\parallel}^* = \left(\frac{\cos \theta_a - n_{ab} \cos \theta_b}{\cos \theta_a + n_{ab} \cos \theta_b} \right)^2 \quad (13.21)$$

$$R_{\perp}R_{\perp}^* = \left(\frac{n_{ab} \cos \theta_a - \cos \theta_b}{n_{ab} \cos \theta_a + \cos \theta_b} \right)^2 \quad (13.22)$$

$$\Re\{R_{\parallel}R_{\perp}^*\} = \left(\frac{\cos \theta_a - n_{ab} \cos \theta_b}{\cos \theta_a + n_{ab} \cos \theta_b} \right) \left(\frac{n_{ab} \cos \theta_a - \cos \theta_b}{n_{ab} \cos \theta_a + \cos \theta_b} \right) \quad (13.23)$$

$$\Im\{R_{\parallel}R_{\perp}^*\} = 0 \quad (13.24)$$

$$T_{\parallel}T_{\parallel}^* = \left(\frac{2n_{ab} \cos \theta_a}{\cos \theta_a + n_{ab} \cos \theta_b} \right)^2 \quad (13.25)$$

$$T_{\perp}T_{\perp}^* = \left(\frac{2n_{ab} \cos \theta_a}{n_{ab} \cos \theta_a + \cos \theta_b} \right)^2 \quad (13.26)$$

$$\Re\{T_{\parallel}T_{\perp}^*\} = \frac{4n_{ab}^2 \cos^2 \theta_a}{(\cos \theta_a + n_{ab} \cos \theta_b)(n_{ab} \cos \theta_a + \cos \theta_b)} \quad (13.27)$$

$$\Im\{T_{\parallel}T_{\perp}^*\} = 0. \quad (13.28)$$

It should be noted that for the case of normal incidence, $\theta_i = 0$, both $R_{\parallel}R_{\parallel}^*$ and $R_{\perp}R_{\perp}^*$ reduce to

$$R_{\parallel}R_{\parallel}^* = R_{\perp}R_{\perp}^* = \left(\frac{n_b - n_a}{n_b + n_a} \right)^2. \quad (13.29)$$

This gives a reflectance of $R_{ab}(\theta_i = 0) = 0.021$ for $n_{\text{water}} = 1.34$, for both air- and water-incident light.

For the case of total internal reflection, i.e., $n_a > n_b$ and $\theta_a \geq \theta_c$, the following equations are to be used (G22):

$$R_{\parallel}R_{\parallel}^* = 1 \quad (13.30)$$

$$R_{\perp}R_{\perp}^* = 1 \quad (13.31)$$

$$\Re\{R_{\parallel}R_{\perp}^*\} = \frac{2 \sin^4 \theta_a}{1 - (1 + n_{ba}^2) \cos^2 \theta_a} - 1 \quad (13.32)$$

$$\Im\{R_{\parallel}R_{\perp}^*\} = - \frac{2 \cos \theta_a \sin^2 \theta_a \sqrt{\sin^2 \theta_a - n_{ba}^2}}{1 - (1 + n_{ba}^2) \cos^2 \theta_a} \quad (13.33)$$

and all elements of the transmission matrix elements are 0:

$$\mathbb{T}_{ab} = \mathbb{T}_{wa} = \mathbb{O}. \quad (13.34)$$

(\mathbb{O} is a 4×4 matrix of zeros.) Finally, the all-important transmission factor f_T in Eq. (13.15) is given by

$$f_T = n_{ba}^3 \left(\frac{\cos \theta_b}{\cos \theta_a} \right) \quad (\text{for diffuse Stokes vectors}), \quad (13.35)$$

when computing the transmittance for diffuse Stokes vectors. These equations give everything needed to describe reflection and transmission of polarized light by a level sea surface.

Figure 13.6 shows the \mathbb{R}_{aw} and \mathbb{T}_{aw} matrices as a function of incident angle θ_i for $n_{air} = 1$ and $n_{\text{water}} = 1.34$. The (1,1) matrix elements are shown in the upper-left plot, and the (4,4) elements are in the lower-right plot. The red curves are $\mathbb{R}_{aw}(\theta_i)$ and the blue curves are $\mathbb{T}_{aw}(\theta_i)$. The reflectance curve for $R_{aw}(1,1)$ is the Fresnel reflectance for unpolarized light as given in the previous section on Fresnel formulas for unpolarized light: it starts at 0.021 for normal incidence and $n_{\text{water}} = 1.34$, and rises to 1 at grazing incidence. The transmission curve for $T_{aw}(1,1)$ on the other hand may look incorrect because it has values greater than one. Its maximum value at normal incidence is

$$T_{aw}(1,1) = \frac{4n_b^3}{(1+n_b)^2} = 1.758 \quad (13.36)$$

However, this value is indeed correct and is a consequence of the fact that we are now dealing with a *diffuse* Stokes vector with units of radiance, and the n^2 law for radiance (Section 13.1.2) applies. The curves in Fig.(13.6) agree exactly with the corresponding plots in Garcia (2012, Figs. 1-3).

If we were dealing with *coherent* Stokes vectors with units of irradiance, then the f_T factor of Eq. (13.35) would be

$$f_T = n_{ba} \frac{\cos \theta_b}{\cos \theta_a} \quad (\text{for coherent Stokes vectors}). \quad (13.37)$$

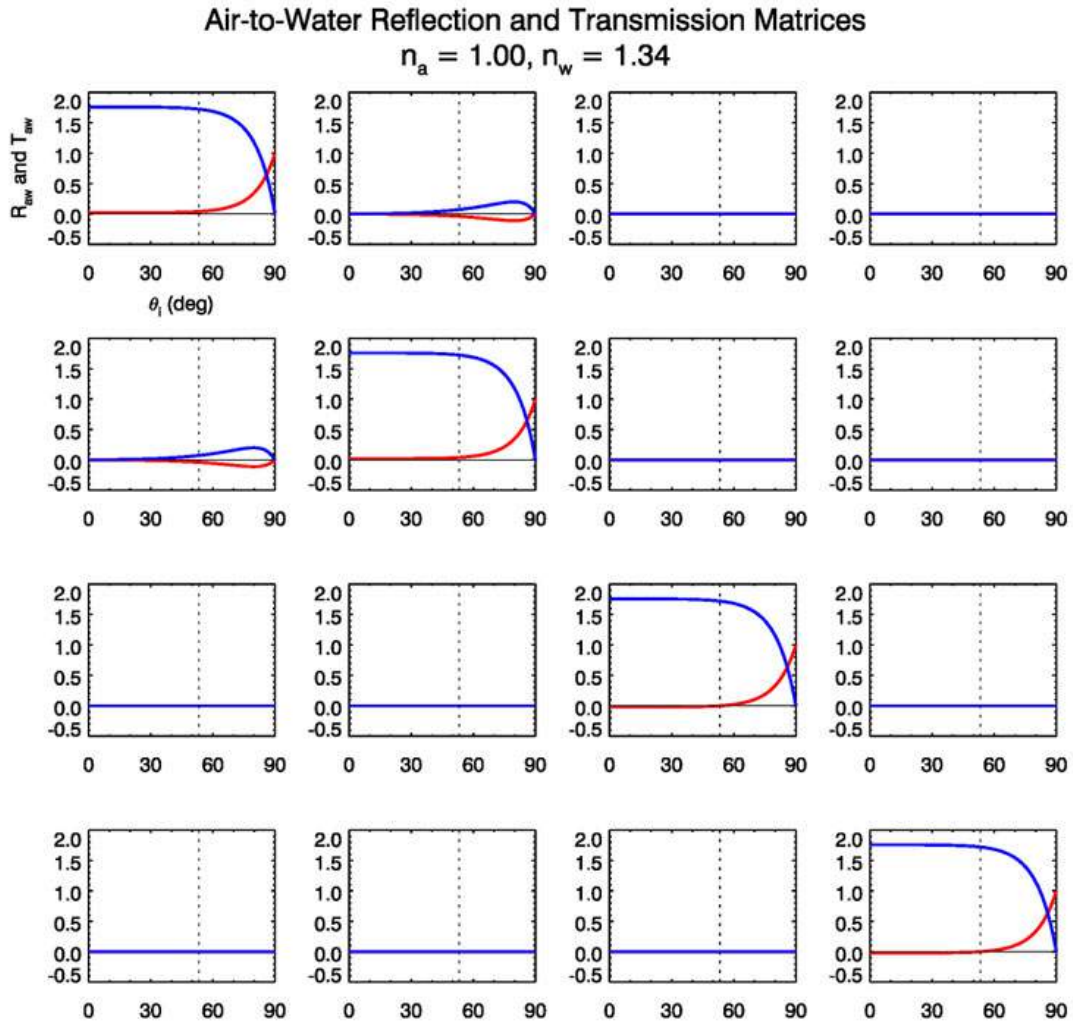


Figure 13.6: Reflectance and transmittance matrices as functions of the incident angle θ_i for air-incident radiance. \mathbb{R}_{aw} is in red and \mathbb{T}_{aw} is in blue. The vertical dotted line at $\theta_i = 53.3$ deg is Brewster's angle.

The transmittance for normal incidence then would be $(4n_b)/(1+n_b)^2 = 0.979$, which with the reflectance sums to one (and also sums to one for all other incident angles). As noted elsewhere, it is the law of conservation of energy, not the law of conservation of radiance. This business of diffuse and coherent Stokes vectors and their different f_T values may seem like a confusing bit of trivia, that is until you write a computer program and cannot figure out why your numbers do not agree with observation or with someone else's computations.

The vertical dotted line in Fig. (13.6) shows the location of Brewster's angle,

$$\theta_{Brew} = \arctan(n_b) \quad (13.38)$$

which is $\arctan(1.34) = 53.3$ deg in the present case. At this angle, $R_{aw}(1, 2) = R_{aw}(2, 1) = -R_{aw}(1, 1)$, and $R_{aw}(3, 3) = R_{aw}(4, 4) = 0$. In the present case $R_{aw}(1, 1) \approx 0.04$ at θ_{Brew} ,

and the reflection process $\mathbb{S}_r = \mathbb{R}_{aw}(\theta_i = \theta_{\text{Brew}})\mathbb{S}_i$ becomes

$$\mathbb{S}_r = \begin{bmatrix} 0.04 & -0.04 & 0 & 0 \\ -0.04 & 0.04 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} I \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0.04I \\ -0.04I \\ 0 \\ 0 \end{bmatrix}. \quad (13.39)$$

Thus, at Brewster's angle, unpolarized incident radiance is totally horizontally polarized upon reflection.

It should also be noted that the non-zero $T_{aw}(2, 1)$ means that unpolarized radiance becomes partly horizontally polarized upon transmission through the surface.

Figure (13.7) shows \mathbb{R}_{aw} and \mathbb{T}_{aw} as reduced scattering matrices, i.e. after dividing each element by its (1,1) component. These plots show more clearly the behavior of the \mathbb{R}_{aw} matrix elements at Brewster's angle. These curves agree exactly with the corresponding plots in Kattawar and Adams (1989, Fig. 4).

Figure (13.8) shows \mathbb{R}_{wa} and \mathbb{T}_{wa} . The vertical dotted line is at the critical angle for total internal reflection, which in the present case is $\theta_c = 48.3$ deg. For angles less than the critical angle, the transmission is never more than about 0.54. This again shows the n-squared law for radiance. In going from water to air, the in-water radiance is decreased by a factor of $1/n_{\text{water}}^2$ when crossing the surface because the solid angle in air is greater than that in water by a factor of n_{water}^2 . The (1,1) elements show that beyond the critical angle there is no transmission and total reflection. These curves agree with the corresponding plots in Garcia (2012, Figs. 4-6).

Figure (13.9) shows the reduced water-to-air matrices. These curves agree with the corresponding plots in Kattawar and Adams (1989, Fig. 5) (The signs of the $\mathbb{T}_{wa}(3, 4)$ and $\mathbb{T}_{wa}(4, 3)$ elements are reversed in the original Fig. 5, which had a sign error.).

The non-zero matrix elements of course depend on incident angle as seen above, but also depend weakly on the wavelength via the wavelength dependence of n_{water} .

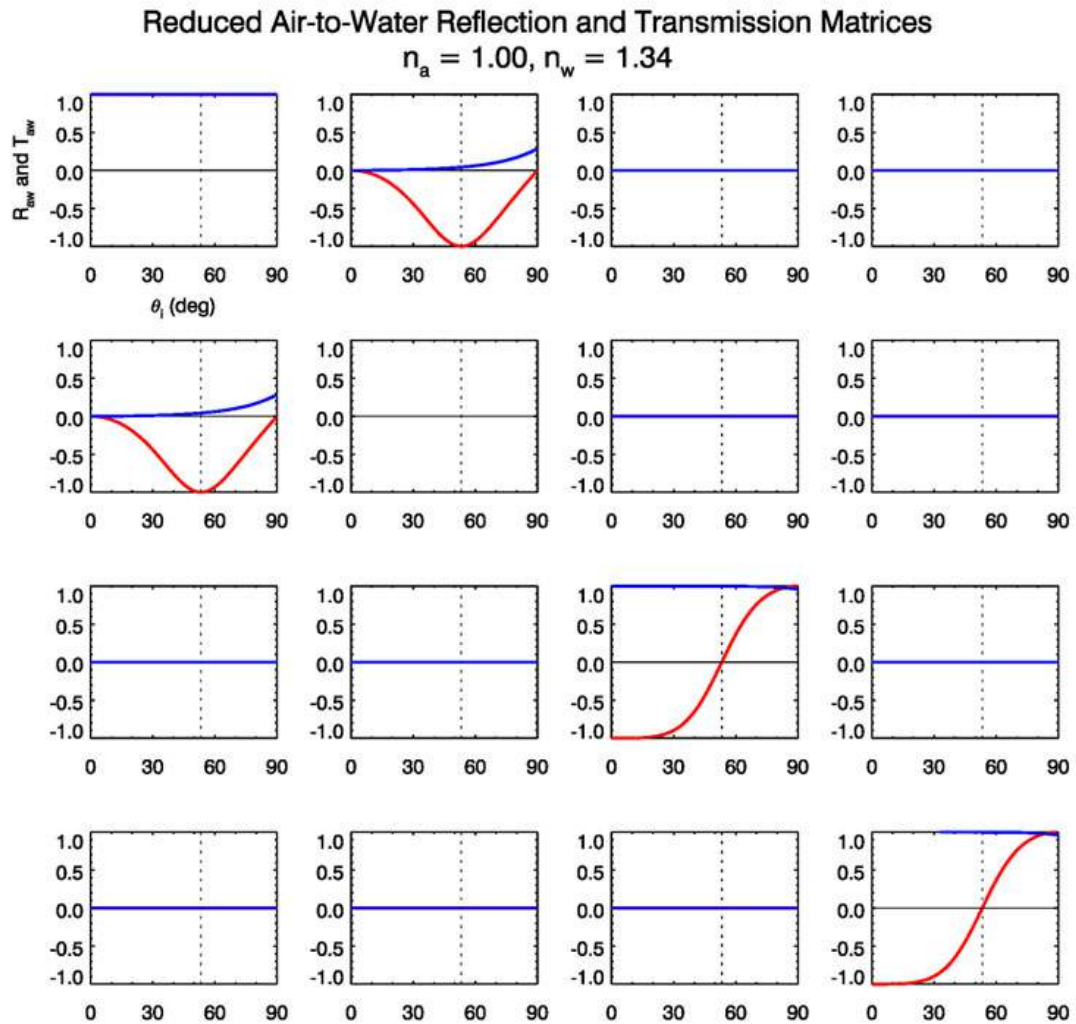


Figure 13.7: Reduced reflectance and transmittance matrices for air-incident radiance [the reflectance and transmittance matrices of Fig.(13.6) normalized by their (1,1) elements]. The vertical dotted line is Brewster's angle.

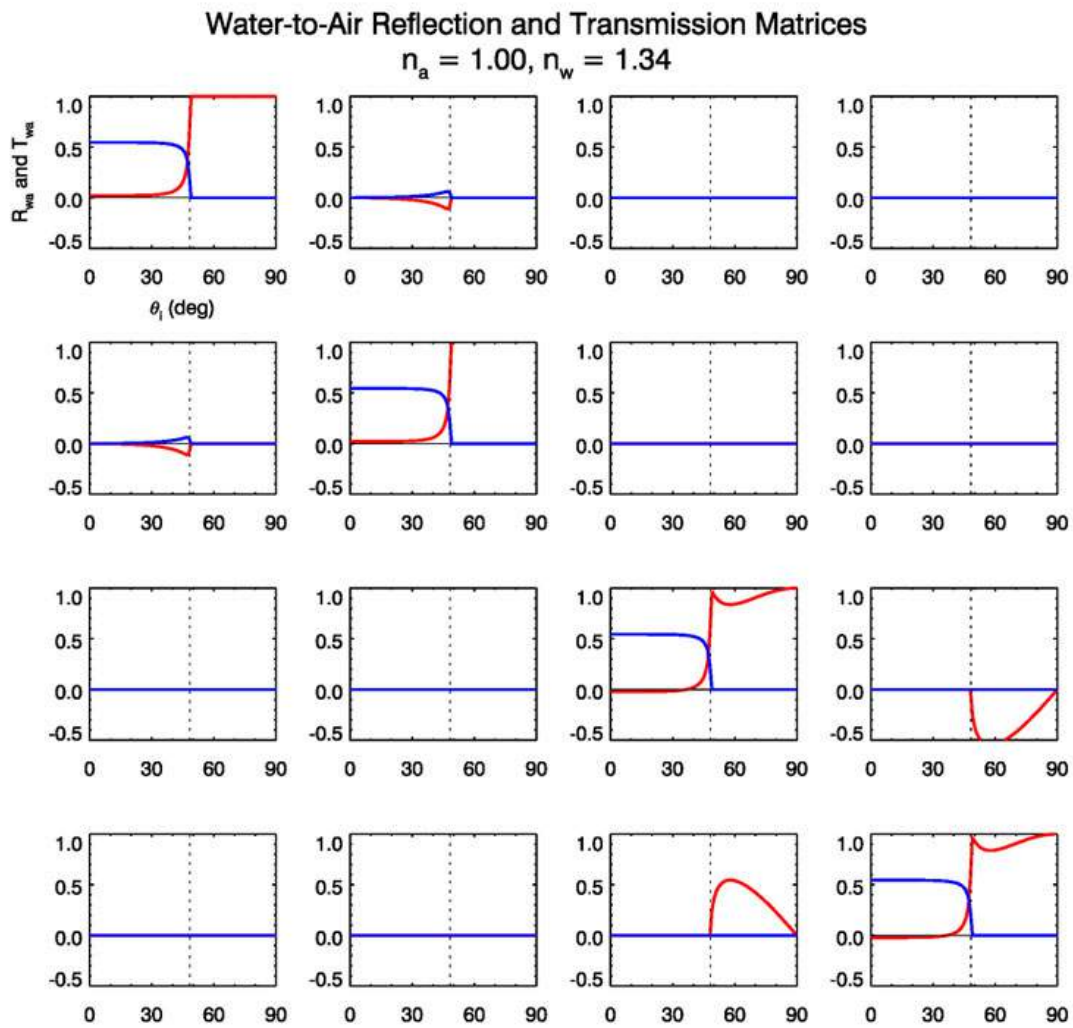


Figure 13.8: Reflectance and transmittance matrices as functions of the incident angle θ_i for water-incident radiance. \mathbb{R}_{wa} is in red and \mathbb{T}_{wa} is in blue. The vertical dotted line is the critical angle for total internal reflection.

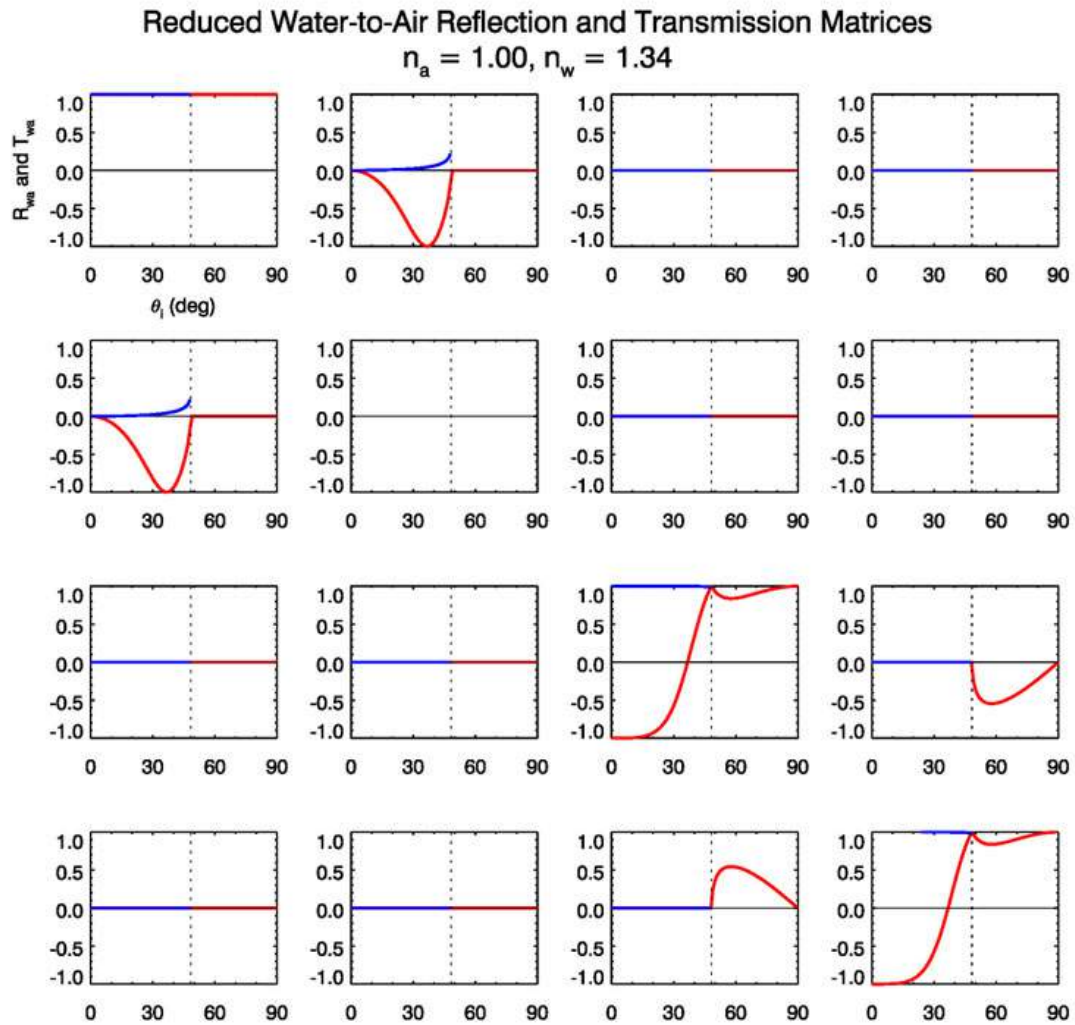


Figure 13.9: Reduced reflectance and transmittance matrices for air-incident radiance [the reflectance and transmittance matrices of Fig.(13.8) normalized by their (1,1) elements]. The 34 and 43 elements are the reverse of Fig. 5 in [Kattawar and Adams \(1989\)](#) due to a sign error in the original paper.

13.4 The Cox-Munk Sea Surface Slope Statistics

Cox and Munk (1954a,b) analyzed aerial photographs of the Sun's glitter patterns on wind-blown sea surfaces, from which they were able to deduce the slope statistics of the sea surface as a function of the wind speed (see also Cox and Munk, 1955). Figure 13.10 shows one such photograph. The wind speed, measured from a ship within the area photographed, was 4.6 m/s.

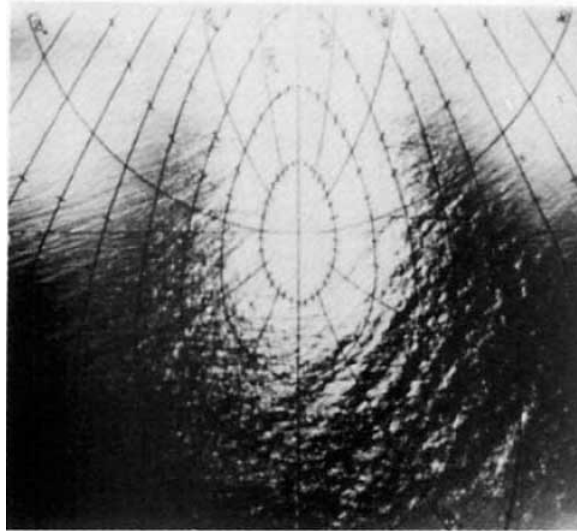


Figure 13.10: A glitter pattern photograph for a wind speed of 4.6 m/s. The view is looking azimuthally toward the Sun and downward at the sea surface from an altitude of 2,000 feet (610 m). The Sun's glitter pattern is the central bright area. The upper left and right areas are glint from the background sky or perhaps from clouds. The superimposed grid shows lines of constant wave facet tilt β and azimuth α as described in the text and in Fig. 13.11.

If the surface were perfectly flat (i.e., zero wind speed and no swell) and the sky were black, the glitter pattern would be a single bright spot at the Sun's specular reflection direction, which is at the center of the photograph. However, wind ruffles the sea surface so that the Sun's direct beam can be reflected from a wider area of the sea surface into the observer's direction. In order for a wave facet to reflect a solar ray towards the observer, the facet must be tilted in just the right way so that the tilted facet can reflect an incident ray from the Sun into the direction of the observer. This is illustrated in Fig. 13.11. The blue triangle represents a tilted wave facet that is reflecting an incident solar ray $\hat{\xi}'$ into the observer's direction $\hat{\xi}$. The $(\hat{x}, \hat{y}, \hat{z})$ coordinate system shown by the green arrows is a Sun-centered system with $-\hat{x}$ pointing horizontally towards the Sun, \hat{z} vertically upward (normal to the mean sea surface), and $\hat{y} = \hat{z} \times \hat{x}$. The Sun's incident ray $\hat{\xi}' = (\xi'_x, \xi'_y, \xi'_z)$ thus has $\xi'_x > 0$, $\xi'_y = 0$, and $\xi'_z < 0$. \hat{n} is the normal to the tilted facet. Polar angle $\beta = \cos^{-1}(\hat{n} \cdot \hat{z})$ measures the tilt of the facet from the normal to the mean sea surface. Azimuthal angle α measures the orientation of the facet relative to the \hat{x} axis, with α being measured clockwise from \hat{x} as shown.

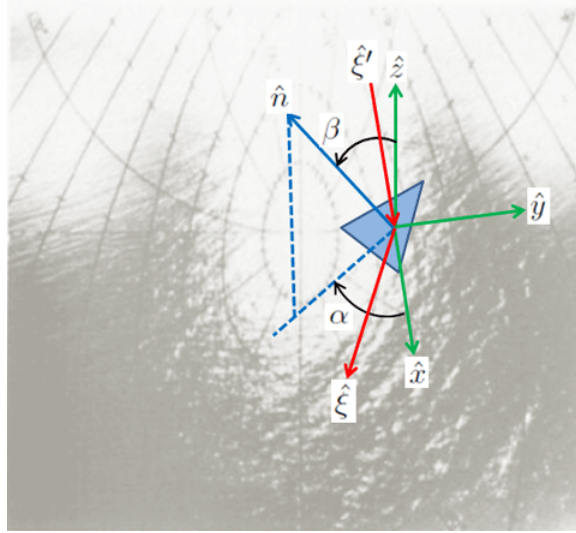


Figure 13.11: The image of Fig. 13.10 with an illustration of a tilted wave facet (blue triangle) reflecting an incident solar ray $\hat{\xi}'$ into the direction $\hat{\xi}$ of an observer.

Let $\eta(x_a, x_c)$ be the sea surface elevation in a wind-centered coordinate system². where x_a is in the along-wind direction (with x_a positive in the downwind or \hat{x} direction) and x_c is in the cross-wind direction (with x_c positive in the \hat{y} direction). Then

$$\eta_a = \frac{\partial \eta}{\partial x_a} \quad \text{and} \quad \eta_c = \frac{\partial \eta}{\partial x_c}$$

are respectively the sea surface slopes in the along-wind and cross-wind directions. After laborious analysis of numerous photographs for different solar zenith angles and wind speeds, Cox and Munk found that the statistical distribution of the random sea surface slopes η_a and η_c is, to a good approximation, a bivariate Gaussian:

$$p(\eta_a, \eta_c) = \frac{1}{2\pi\sigma_a\sigma_c} \exp \left[-\frac{1}{2} \left(\frac{\eta_a^2}{\sigma_a^2} + \frac{\eta_c^2}{\sigma_c^2} \right) \right] \quad (13.40)$$

where σ_a^2 and σ_c^2 are respectively the variances of the slopes in the along-wind and cross-wind directions. Note that $p(\eta_a, \eta_c)$ is normalized as required for a probability distribution function, i.e.,

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(\eta_a, \eta_c) d\eta_a d\eta_c = 1.$$

²The analysis of the glint photographs requires several coordinate systems. One is Sun-centered, as seen in Fig. 13.11. Another system is used to define the image plane of the camera recording the glitter patterns. Use of the surface statistics to be presented below for generation of random realizations of wind-blown surfaces requires a wind-centered system $(\hat{x}, \hat{y}, \hat{z})$, with \hat{x} pointing downwind, \hat{y} cross-wind, and \hat{z} upward. Figure 13.11 shows the $(\hat{x}, \hat{y}, \hat{z})$ system as used by Preisendorfer and Mobley. The Sun system used in the Cox and Munk papers has \hat{y} pointing horizontally toward the Sun, with α measured from \hat{y} “to the right of the Sun.” Conversion from one system to the other is straightforward but tedious trigonometry; the details are given in Section 7(a) of Preisendorfer and Mobley (1985) and in Preisendorfer and Mobley (1986). Fortunately, these details do not concern us here.

These slope variances were found to be related to the wind speed U in meters per second at “mast height” (41 feet or 12.5 m) by

$$\sigma_a^2 = 0.000 + 3.16 \cdot 10^{-3}U \pm 0.004, \quad r = 0.945 \quad (13.41a)$$

$$\sigma_c^2 = 0.003 + 1.92 \cdot 10^{-3}U \pm 0.002, \quad r = 0.956 \quad (13.41b)$$

$$\sigma^2 = \sigma_a^2 + \sigma_c^2 = 0.003 + 5.12 \cdot 10^{-3}U \pm 0.004, \quad r = 0.986 \quad (13.41c)$$

Equations (13.40) and (13.41) are the celebrated Cox-Munk wind speed-wave slope statistics. The non-zero value of σ_c^2 at zero wind speed results from a residual amount of slope that is not attributable to the local wind. This contribution to σ_c^2 is often ignored when modeling sea surfaces, so that a wind speed of zero corresponds to an exactly flat sea surface.

Duntley (1954) used closely spaced vertical wires to measure surface elevations at the two wires, from which the slope could be obtained. His measurements were consistent with the values obtained by Cox and Munk. Several later studies have found some dependence on the air-sea temperature difference, i.e., on the atmospheric stability, although sometimes with conflicting conclusions, perhaps because the sea states were not in a mature state for the given wind speed. The slope variances are also sensitive to the presence of films (e.g., from oil) that tend to dampen waves, especially at the smallest spatial scales. Cox and Munk themselves did measurements within areas where they had poured a mixture of oils onto the sea surface³; those values are found in the papers cited. Regardless of potential improvements to the Cox and Munk values, the original values of the slope variances are widely used, e.g. they are one option for surface generation in the HydroLight radiative transfer code (Section 10.6), and they are used by NASA for atmospheric correction (Section 15.9). The numerous successful applications of the Cox-Munk equations have proven that their values are sufficiently accurate for a wide range of conditions.

It should be noted that the Cox-Munk statistics are based on observations, so they describe the slope effects of whatever waves were on the sea surface at the times the photographs were taken. They thus describe the full range of long-wavelength gravity to short-wavelength capillary waves for the sea states at the time of observation. Note, indeed, the obvious presence of long-wave swell at the lower right of Fig. 13.10. It is sometimes stated (e.g., in Preisendorfer and Mobley (1985) and in Section 4.3 of Mobley (1994)) that these slope statistics refer to capillary waves, which is only partially correct. Capillary waves are responsible for much of the slope variance, and they are included in the Cox-Munk statistics, but the effects of gravity waves on the surface slopes are also included.

13.5 Sea Surface Simulations

The numerical modeling of random water surfaces is an exceedingly complex business, and there are many techniques for simulating random sea surfaces (Tessendorf, 2004). At one extreme are hydrodynamics-based methods that solve the equations of fluid motion (the

³As described in Cox and Munk (1954b), they pumped a mixture of “40 percent used crankcase oil, 40 percent Diesel oil, and 20 percent fish oil” in the water to create a slick approximately 2000 by 200 feet in size (approximately 600 by 60 meters). In 1954 that was good science; today they probably would be thrown in jail for such an action.

Navier-Stokes and related equations as seen in [Kinsman \(1965\)](#). At the other extreme are methods that generate a sea surface that visually looks good enough for use in a video game or movie, but often does not obey even the simplest of physical laws such as conservation of energy. What technique you use depends on your problem. Do you want to compute the energy transferred to the waves in the wake of passing ship, do you want to compute the energy transferred from the wind blowing over the ocean to the waves and currents, or do you want to make a video game that runs in real time? For the purposes of optical oceanography, we need to simulate water surfaces that can be used to compute the reflectance and transmittance of light incident onto the water surface. That problem is intermediate in difficulty between the hydrodynamics and visual-appearance methods.

[Preisendorfer and Mobley \(1985\)](#) and [Preisendorfer and Mobley \(1986\)](#) showed how to use the Cox-Munk statistics to generate random realizations of wind-blown sea surfaces (see also Section 4.3 of [Mobley, 1994](#)). The mathematical details need not be repeated here, but the procedure is as follows. First, they divide a hexagonal patch of sea surface into triangular wave facets. Then the Cox-Munk along-wind and cross-wind variances are used in conjunction with a random number generator to define, for a given wind speed, the relative surface elevations at the vertices of the triangular wave facets. The slopes of the resulting surface facets then reproduce, on average, the slope statistics of a real sea surface. Figure 13.12 shows an example of such a surface realization for a wind speed of 10 m/s. These Cox-Munk surfaces are “scale independent.” That is, only the slopes matter, not the actual physical size of the patch of surface. Thus no units are shown for the axes. Note that, in this figure, the vertical scale (the surface elevations) is greatly expanded relative to the horizontal scales (the x axis is the along-wind direction, and the y axis is the cross-wind direction).

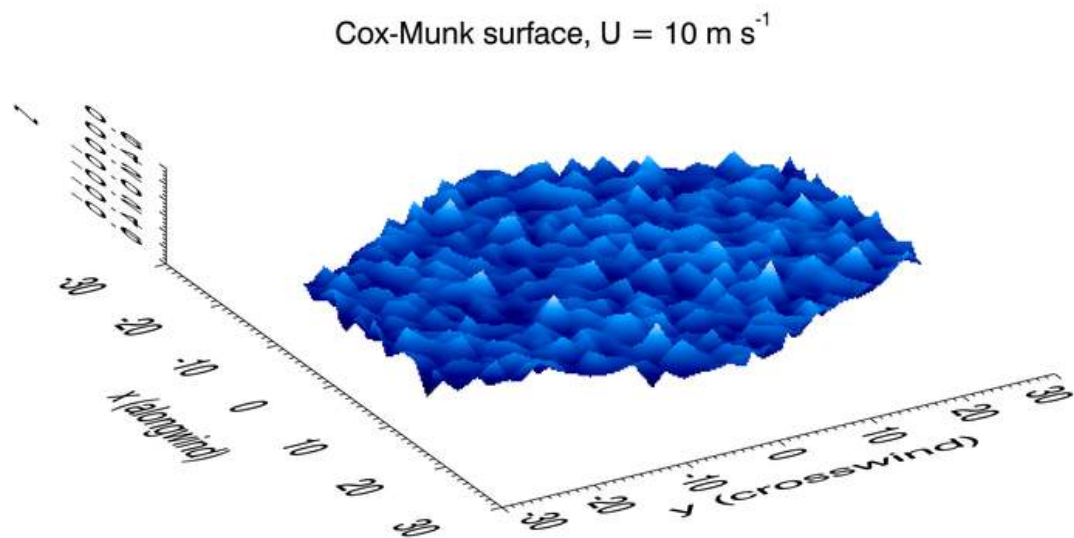


Figure 13.12: Example of a Cox-Munk sea surface for a wind speed of $U = 10 \text{ m s}^{-1}$. High surface elevations (above the mean sea level of zero) are light blue; low elevations (below mean sea level) are dark blue.

This technique does not reproduce the sea surface *elevation* statistics of a real sea

surface because the Cox-Munk equations describe the statistics of only the sea surface *slopes*. Although the surface realization seen in Fig. 13.12 correctly reproduces the Cox-Munk slope variances, it simply does not look like a real sea surface. In particular, there is no spatial correlation from one point on the surface to another nearby point, as occurs with real water waves. However, both the elevation and slope statistics can be reproduced using more advanced techniques, which are described in the appendices. Those techniques are based on sea surface elevation variance spectra (Appendix B) and use fast Fourier transforms (FFTs; Appendix A). Here, surfaces generated by these techniques (as described in Appendix C) will be called FFT surfaces for brevity. Figure 13.13 shows an example surface constructed using the techniques of Appendix C. The inset shows the mean square slopes in the along-wind (mss_x) and cross-wind (mss_y) directions; these values correspond to σ_a^2 and σ_c^2 , respectively, in the Cox-Munk equations. For a wind speed of 10 m s^{-1} , the Cox-Munk equations give $\sigma_a^2 = 0.0316$ and $\sigma_c^2 = 0.0222$, which agree well with the values shown in the figure for this particular FFT surface realization. The value $H_{1/3} = 2.14 \text{ m}$ is the significant wave height, which is in agreement with the wave heights for a mature sea at this wind speed. Note that the figure axes are now in meters, because an actual 100×100 meter patch of sea surface is being simulated, although the scale of the vertical axis is still exaggerated compared to the horizontal scales.

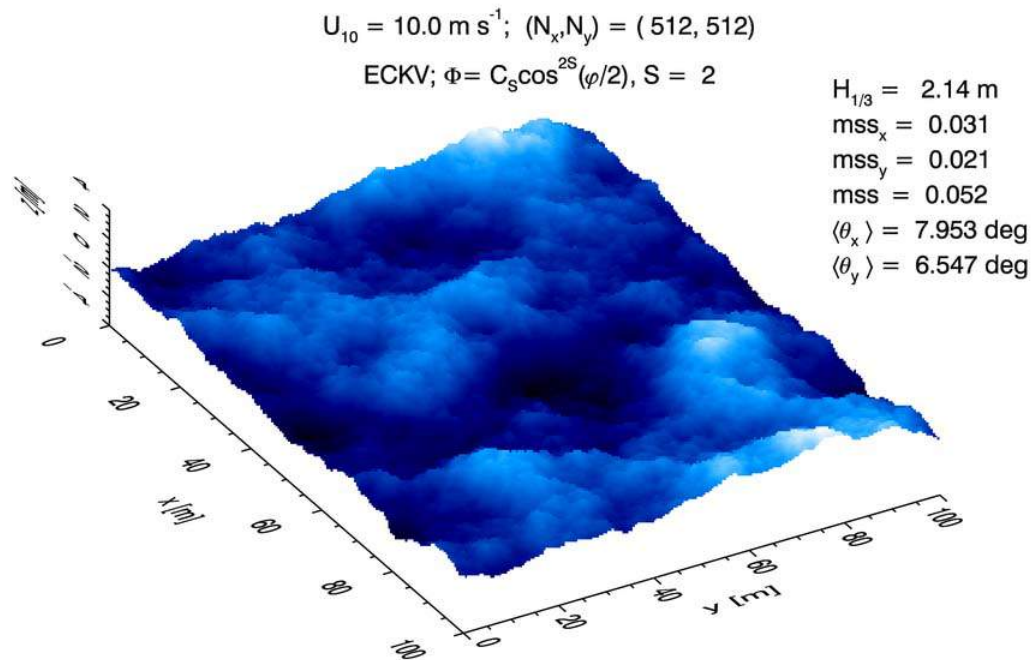


Figure 13.13: Example realization of a 10 m s^{-1} sea surface constructed using the techniques of Appendix C. This surface reproduces both the surface elevation statistics and the slope statistics. Reproduced from Fig. 3.4 of Mobley (2016).

13.5.1 Optical Differences in Sea Surfaces

The obvious next question is this: How much do optical quantities such as the sea surface reflectance or water-leaving radiance differ for a Cox-Munk surface like that of Fig. 13.12 versus a more realistic FFT surface like that of Fig. 13.13? This can be answered by Monte Carlo simulation, which is described in Appendix E. In such simulations, a large number (often millions) of random sea surface realizations are generated. For each surface, rays simulating the Sun and sky incident radiance are sent toward the surface, where they are reflected and transmitted by the surface wave facets according to the Fresnel equations applied to the point where a ray intersects a wave facet. In this manner, the reflected and transmitted radiances are built up ray by ray. The mathematical details of these computations are rather ugly; see Preisendorfer and Mobley (1985) or Mobley (2014, Appendix B) for a detailed description of a ray tracing algorithm that fully accounts for the possibility of multiple scattering between surface wave facets.

Figure 13.14 shows an example of a simulated solar glitter pattern created by ray tracing and a Cox-Munk surface. In this figure, the final direction of each ray is plotted as a dot where the ray intersects the image plane of a camera photographing the glitter pattern from the air. The pattern of this simulated glitter pattern should be compared with central glitter pattern of Fig. 13.10. The two patterns are in qualitative agreement.

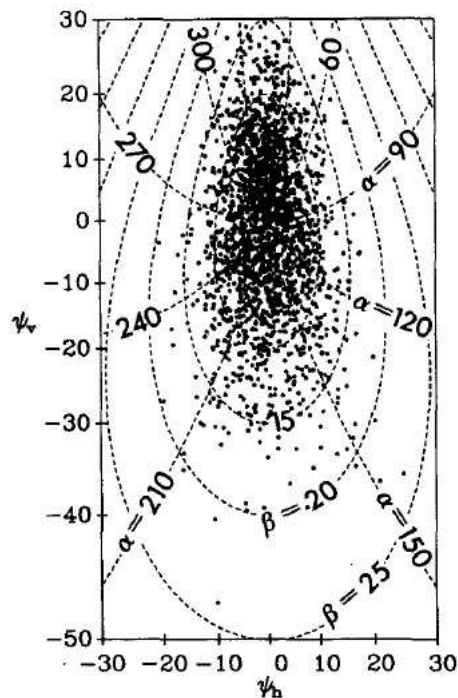


Figure 13.14: Monte Carlo simulation of a glitter pattern for the Sun at a 60 deg zenith angle and a wind of 5 m s^{-1} . The ψ_h and ψ_v axes refer to the coordinate system used to describe the viewing direction in the image plane of the observer's camera. In this figure, the glitter pattern is being viewed ± 30 degrees in the horizontal from the specular point, and from 30 deg above to 50 deg below the specular point. Reproduced from Preisendorfer and Mobley (1986).

A feeling for the optical differences of various sea surfaces can be obtained by comparison of their surface-reflected radiances, i.e., their glitter patterns. The left panel of Fig. 13.15 shows the surface-reflected radiance for a Cox-Munk surface and a wind speed of 5 m s^{-1} , as generated by HydroLight. The Sun was at a zenith angle of 50 deg in a clear sky. The Sun's azimuthal angle was in the down-wind direction (the arrow at the middle of the plot indicates the wind direction). The water IOPs were based on an albedo of single scattering of 0.8 and the sky conditions were for a wavelength of 550 nm . The ray tracing that underlies these calculations used $250,000$ sea surface realizations; there is a negligible amount of Monte Carlo noise in these results. The colors display contours of the radiance, with the lightest color being the highest radiance; the contour spacing is not linear but was chosen for visual effect. The light colored area at the right of the polar plot is the glitter pattern as would be seen by an observer looking downward at the sea surface and facing the Sun at an azimuthal viewing angle of $\varphi_v = 0$. The concentric circles show off-nadir viewing angles of 30 , 60 , and 90 deg (the horizon). The radiance is largest near the horizon, rather than near the specular direction, because of the large increase in the Fresnel reflectance for angles of reflection greater than about 60 deg . The right panel of the figure shows the glitter pattern for an FFT surface, with all else being the same.

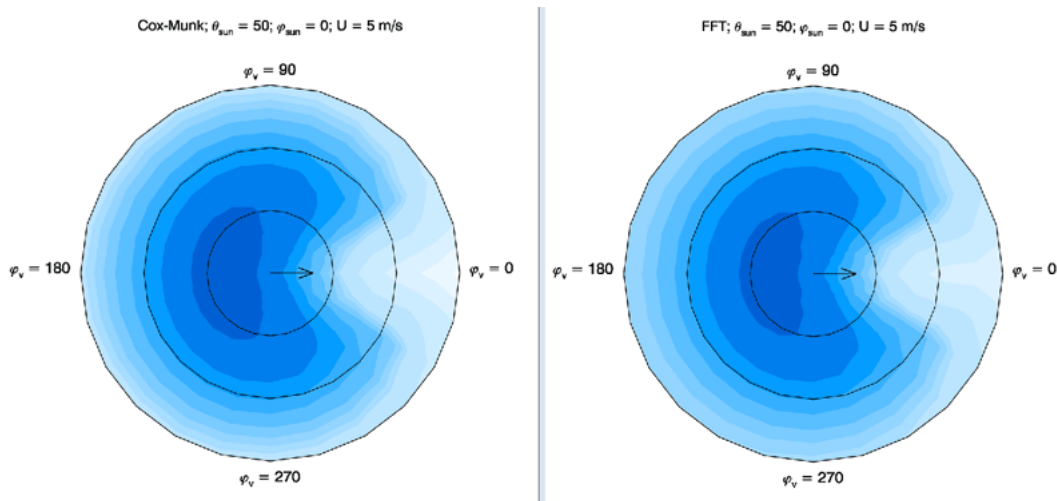


Figure 13.15: HydroLight-computed surface-reflected radiances for a solar zenith angle of 50 deg and a wind speed of 5 m s^{-1} . Left panel: a Cox-Munk surface realizations; right panel: FFT surfaces.

Some differences in the glitter patterns can be seen in the contour plots of Fig. 13.15, but a more quantitative comparison can be made by plotting the radiances as a function of polar angle in the plane of the Sun. This is done in Fig. 13.16, using the data of Fig. 13.15. This plot shows that for viewing directions out to about 60 deg in the azimuthal direction of the Sun, there is only a few percent difference in the surface-reflected radiances. For viewing directions near the horizon, the difference increases to several tens of percent, with the Cox-Munk surface having the “brighter” glitter pattern near the horizon.

Figure 13.17 shows the corresponding water-leaving radiances, which determine the remote-sensing reflectance. Again, there is very little difference (at most a few percent) for off-nadir viewing directions less than about 50 deg , which covers the range of most ocean-color remote sensing. However, for viewing directions near the horizon, the difference in

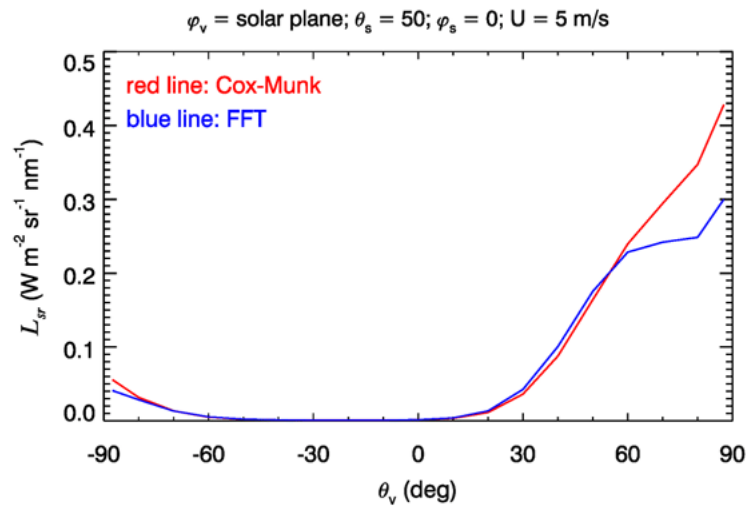


Figure 13.16: Surface-reflected radiances in the plane of the Sun. Negative off-nadir viewing directions θ_v correspond to looking away from the Sun (in the $\varphi_v = 180$ direction in Fig. 13.15; positive θ_v values correspond to looking toward the Sun.

the Cox-Munk and the FFT surface is again a few tens of percent.

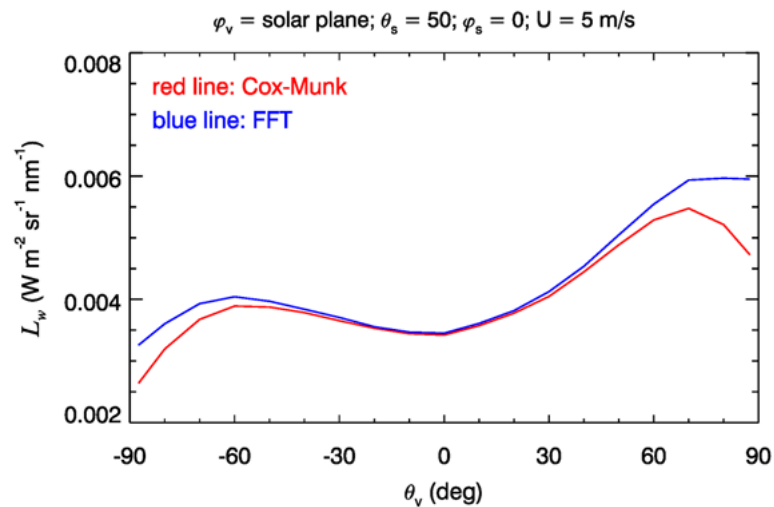


Figure 13.17: Water-leaving radiances in the plane of the Sun, as in Fig. 13.16.

The details of these comparisons will be different for different Sun zenith angles and different wind speeds. However, it is generally true that the sea surface affects the water-leaving radiance by only a few percent for the near-nadir viewing directions relevant to most remote sensing. Surface effects are most prominent in the glitter patterns themselves. There are also differences when the Sun is in the along-wind versus the cross-wind direction. However, a full discussion of such matters is a topic for elsewhere.

13.6 The Bidirectional Reflectance Distribution Function (BRDF)

The preceding sections have discussed the optical effects of air-water surfaces. This section considers how light is reflected by opaque surfaces, such as a sandy sea bottom or a submarine within the water. Scientists in many different fields including astronomy, geology, agronomy, the paint industry, camouflage technology, and remote sensing have studied how surfaces reflect light. Unfortunately, different fields often use different measures of “reflectance,” and they all have their own terminology and notation even when they are measuring the same physical quantity. (The next section discusses other measures of reflectance.) There are many opportunities for losing factors of π and cosines of angles, and it is sometimes nearly impossible to figure out *exactly* what is being discussed when reading a paper. This section gives an overview of the definitions, terminology, and notation as needed for optical oceanography and remote sensing.

For the most part, the definitions and terminology used here are given in Hapke (1993), which is a good introductory textbook on reflectance, and in Nicodemus et al. (1977) (referenced here as NBS160). NBS160 is a National Bureau of Standards⁴ document that discusses the measurement of reflectance in great detail and is the authoritative document on the subject. However, we have changed some notation to correspond to what is commonly used in optical oceanography. Table 13.1 at the end of this section compares the notation used in these books.

For convenience, let the “surface” reflecting the light be a horizontal plane. This can be a physical surface such as a sandy ocean bottom, or it can be simply a particular depth in the water column, say at 1 m above a sea grass bed or at 100 m in optically deep mid-ocean water. To conform to NBS160, let subscript i denote *incident* and r to denote *reflected*. In the oceanographic setting of a horizontal bottom, the light incident onto the surface is traveling downward, and the light reflected by the surface is traveling upward. Thus we sometimes use subscript d for *downward* (incident) and u for *upward* (reflected) when necessary to conform to common oceanographic usage.

In nature, light is usually incident onto a surface from all directions, and some of the incident light gets reflected by the surface into all directions. Therefore, *to completely understand the optical properties of a surface, it is necessary to know how the surface reflects radiance incident from any incident direction into any reflected direction.*

Figure 13.18 shows the geometry used to describe reflectance from a surface. A Cartesian $(\hat{x}, \hat{y}, \hat{z})$ coordinate system is chosen with the surface lying in the x-y plane and with the z axis normal to the surface (upward in our case), an element of which is shown in aqua. There is a *collimated light source*, which provides the incident light coming from direction (θ_i, ϕ_i) ; and there is a detector, which receives the reflected light traveling toward the viewing direction (θ_r, ϕ_r) . Surface optical properties usually depend on the wavelength λ , so the complete description of the reflectance properties of a surface will be a function (the BRDF) of five variables: $\theta_i, \phi_i, \theta_r, \phi_r$, and λ . To make our equations as simple as possible, we drop the λ , but keep in mind that everything discussed below depends on wavelength.

For oceanography, it is often reasonable to assume that the surface is *azimuthally isotropic*, which means that its reflectance properties depend on the *difference* of ϕ_i and

⁴The NBS is now NIST; see the footnote on page 15

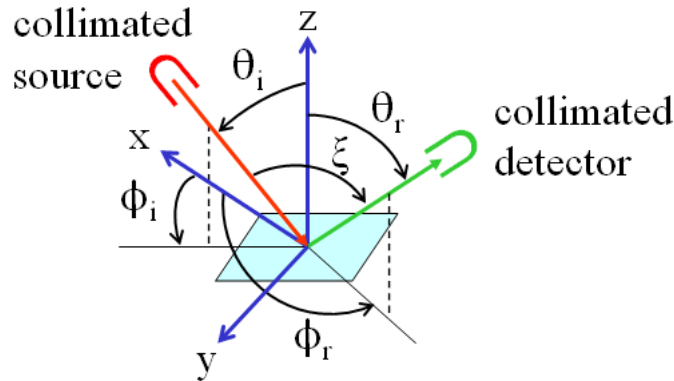


Figure 13.18: Geometry for discussion of surface reflectance. The surface is in aqua, the incident light is red, and the reflected light is green.

ϕ_r . (This would not be the case for long parallel ripples on a sandy bottom, for example.) The *specular direction* is the direction that a level mirror surface would reflect light: $(\theta_r, \phi_r) = (\theta_i, \phi_i + 180^\circ)$. The *retroreflection direction* is the direction of exact backscatter: $(\theta_r, \phi_r) = (\theta_i, \phi_i)$. The angle ξ between the source and detector is called the *phase angle*⁵; it is computed from

$$\cos \xi = \cos \theta_i \cos \theta_r + \sin \theta_i \sin \theta_r \cos(\phi_i - \phi_r). \quad (13.42)$$

Great care and precise language must be used when talking about reflectance. In particular, “reflectance” always should be preceded by two adjectives: the first describes the source and the second the detector. Thus we have

The *directional-hemispherical reflectance* tells how much light is reflected from a particular incident (e.g., downward) direction (θ_i, ϕ_i) into the hemisphere of all reflected (e.g., upward) directions

The *hemispherical-directional reflectance* tells how much light is reflected from all incident directions into a particular reflected direction (θ_r, ϕ_r) . The remote-sensing reflectance $R_{rs} = L_w/E_d$ used in optical oceanography is a hemispherical-directional reflectance.

The *bi-hemispherical (i.e., hemispherical-hemispherical) reflectance* tells how much light is reflected from all incident directions into all reflected directions. The irradiance reflectance $R = E_u/E_d$ used in optical oceanography is a bi-hemispherical reflectance.

The *bi-directional (i.e., directional-directional) reflectance* tells how light is reflected from a particular incident direction into a particular reflected direction.

We now define the *bi-directional reflectance distribution function* (BRDF), which tells us *everything we need to know about how a surface reflects light*. The following discussion is based on NBS160, which treats these matters in great detail.

⁵If the source is the Sun and the surface is the Moon and the Earth is the detector, then the phase angle determines the phase of the Moon as seen from the Earth. This is the historical origin of the term “phase function” for the function that describes the angular pattern of scattered light. The scattering angle ψ as used in radiative transfer theory is the complement of the phase angle: $\psi = 180 - \xi$.

Conceptually, we think about a light beam traveling in a particular direction (θ_i, ϕ_i) being reflected into another particular direction (θ_r, ϕ_r) . But since any source has some finite divergence, and any detector has some finite field of view, we can associate small solid angles $d\Omega_i$ and $d\Omega_r$ with the incident and reflected beams, respectively. The radiance of the incident beam is $L_i(\theta_i, \phi_i)$, and $L_r(\theta_r, \phi_r)$ is the reflected radiance. These quantities are shown in Fig. 13.19, which is a redrawn version of Fig. 13.18.

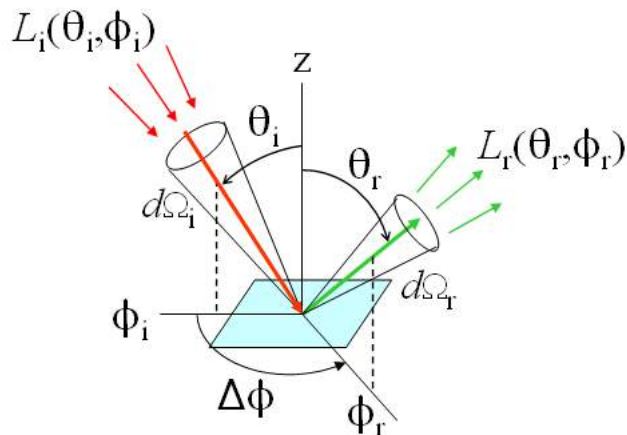


Figure 13.19: Quantities used in the definition of the BRDF.

Our goal is to define an inherent optical property that tells us how the reflective properties of the surface vary with incident and reflected directions (and wavelength). Therefore, consider a measurement in which we hold the direction of the detector in Fig. 13.19 constant while we vary the direction of the source. The BRDF is then defined as

$$BRDF(\theta_i, \phi_i, \theta_r, \phi_r) \triangleq \frac{dL_r(\theta_r, \phi_r)}{L_i(\theta_i, \phi_i) \cos \theta_i d\Omega_i(\theta_i, \phi_i)} \quad [\text{sr}^{-1}]. \quad (13.43)$$

Note that if only the magnitude of the incident radiance changes, the reflected radiance will change proportionately, and the BRDF will remain unchanged. However, if the direction of the incident or reflected beams changes while holding all else constant, the BRDF will in general change.

Equation (13.43) allows an easy transition to radiative transfer theory. Suppose we want to compute the total radiance heading upward in direction (θ_r, ϕ_r) owing to light incident onto the surface from all directions. We then rewrite (13.43) as

$$dL_r(\theta_r, \phi_r) = BRDF(\theta_i, \phi_i, \theta_r, \phi_r) L_i(\theta_i, \phi_i) \cos \theta_i d\Omega_i$$

and then integrate over all incident directions to get the total reflected radiance in direction (θ_r, ϕ_r) :

$$L_r(\theta_r, \phi_r) = \int_{2\pi_i} L_i(\theta_i, \phi_i) BRDF(\theta_i, \phi_i, \theta_r, \phi_r) \cos \theta_i d\Omega_i \quad (13.44)$$

$$\triangleq \int_{2\pi_i} L_i(\theta_i, \phi_i) r(\theta_i, \phi_i, \theta_r, \phi_r) d\Omega_i. \quad (13.45)$$

This last equation is exactly what is seen (with slightly different notation) in [Mobley \(1994, Eq. 4.3\)](#), where $r(\theta_i, \phi_i, \theta_r, \phi_r)$ is called the *radiance reflectance function*. Clearly,

$$r(\theta_i, \phi_i, \theta_r, \phi_r) = BRDF(\theta_i, \phi_i, \theta_r, \phi_r) \cos \theta_i,$$

and the two functions are equivalent ways of describing a surface. In radiative transfer theory irradiances are measured on surfaces normal to the direction of light propagation, whereas actual irradiance measurements are made on the surface of interest. The $\cos \theta_i$ factor in Eq. (13.44) just projects the incident beam irradiance onto the horizontal surface. This is one of those places where it is easy to lose a cosine factor when comparing an observational paper and a theory paper. Also, some investigators add a factor of π sr to the numerator of Eq. (13.43) and define the BRDF as a nondimensional quantity, although this is non-standard⁶. Finally, note that the BRDF is a reflectance per unit solid angle; it can have any non-negative value. It is only when the BRDF is integrated over solid angle to get, for example, an irradiance reflectance that the resulting irradiance reflectance is bounded by one.

It is emphasized that the BRDF completely describes the net effect of everything that happens on or below the surface where it is measured. For example, if the BRDF is measured in the water column 1 m above a sea grass bed, then all the effects of the light interacting with the grass, sediments, and water below the 1 m level are accounted for in this BRDF. Knowing the BRDF on this imaginary surface would, for example, allow HydroLight to compute the radiance distribution in the region *above* the depth where the BRDF was measured⁷. *Predicting or modeling* the BRDF of the grass and sediments from first principles is, however, very difficult and requires understanding and modeling all of the extremely complicated interactions of light with the grass and sediment particles.

All other reflectances can be computed if the BRDF is known. For example, the irradiance reflectance $R = E_u/E_d$ is given by (here the d and u subscripts to identify the downwelling and upwelling irradiances, respectively, as is common in optical oceanography, which correspond to the i and r subscripts in the integrals):

$$\begin{aligned} R &= \frac{E_u}{E_d} = \frac{\iint_{2\pi_r} L_r(\theta_r, \phi_r) |\cos \theta_r| d\Omega_r}{\iint_{2\pi_i} L_i(\theta_i, \phi_i) |\cos \theta_i| d\Omega_i} \\ &= \frac{\iint_{2\pi_r} \left[\iint_{2\pi_i} L_i(\theta_i, \phi_i) BRDF(\theta_i, \phi_i, \theta_r, \phi_r) |\cos \theta_i| d\Omega_i \right] |\cos \theta_r| d\Omega_r}{\iint_{2\pi_i} L_i(\theta_i, \phi_i) |\cos \theta_i| d\Omega_i} \end{aligned} \quad (13.46)$$

The first equation here is just the definitions of the plane irradiances in terms of the incident (downwelling) and reflected (upwelling) radiances. The differential of solid angle is $d\Omega = \sin \theta d\theta d\phi$, and the double integrals are over the corresponding hemispheres of

⁶This is how the MODTRAN atmospheric radiative transfer model defines its BRDFs for various types of surfaces that form the bottom boundary of the atmosphere.

⁷This is indeed how HydroLight models infinitely deep, homogeneous water without actually solving the radiative transfer equation to extreme depth. The BRDF of an infinitely deep, homogeneous layer of water with known inherent optical properties can be found analytically ([Mobley, 1994, Section 9.5](#)). Thus, when HydroLight simulates infinitely deep water, it first computes the BRDF of the infinitely deep water below the maximum depth of interest, z_{\max} , and it then uses that BRDF at z_{\max} just as though there were an actual physical bottom at depth z_{\max} .

2π sr. In going to the second equation, Eq. (13.44) has been used to write the upwelling radiance reflected from the surface in terms of the downwelling radiance onto the surface and the BRDF of the surface. Equation (13.46) is completely general and is the fundamental equation for computing the irradiance reflectance of any surface, given the BRDF of the surface and the incident radiance onto the surface. Note that in general the irradiance reflectance depends both on the surface (the BRDF) and on the incident lighting (the incident radiance $L_i(\theta_i, \phi_i)$). As will be seen in Section 13.7, if the surface is Lambertian, R becomes independent of the incident radiance distribution.

The equations of this section show the BRDF as used in radiative transfer theory. The use of the BRDF as a probability distribution function in Monte Carlo simulations is discussed in Appendix Section F.8.

Finally, there is an important reciprocity theorem about what happens if the positions of the source and detector are interchanged. It states simply that

$$BRDF(\theta_i, \phi_i, \theta_r, \phi_r) = BRDF(\theta_r, \phi_r, \theta_i, \phi_i) . \quad (13.47)$$

If you measure or define a BRDF that does not obey Eq. (13.47), then it is simply wrong.

Table 13.1 compares the notation used in several texts.

Quantity	This Book	Light and Water	Hapke	NBS160
radiance	L	L	I	L
irradiance	E	E	J	E
single-scattering albedo	ω_o	ω_o	w	—
scattering angle	ψ	ψ	θ	—
mean cosine of scattering angle	g	g	ξ	—
phase angle	ξ	—	g	—
incident polar angle	θ_i	θ'	i	θ_i
reflected polar angle	θ_r	θ	e	θ_r
incident azimuthal angle	ϕ_i	ϕ'	set to 0	ϕ_i
reflected azimuthal angle	ϕ_r	ϕ	ψ	ϕ_r
solid angle	Ω	Ω	Ω	ω
BRDF	BRDF	$r / \cos \theta_i$	BRDF	f_r
irradiance reflectance	R	R	r	ρ

Table 13.1: Comparison of the notation used here with that used in *Light and Water* (Mobley, 1994), Hapke (1993), and NBS 160 (Nicodemus et al., 1977).

13.7 Lambertian Surfaces

Actual measurements of BRDFs of ocean bottom materials like sand or sea grass canopies have rarely been made, although some measurements of sediments do exist (e.g. Zhang et al., 2003). Because of the lack of measurements and models of the BRDF for actual ocean bottom materials, it is usually assumed that a bottom is a Lambertian reflecting surface.

A *Lambertian surface* by definition reflects radiance equally into all directions. Its BRDF is simply

$$BRDF_{Lamb}(\theta_i, \phi_i, \theta_r, \phi_r) = \frac{\rho}{\pi}, \quad (13.48)$$

where ρ is called the *reflectivity* of the surface. The reflectivity varies from zero for a completely absorbing (“black”) surface, to one for a completely reflecting (“white”) surface. There are no Lambertian surfaces in nature, but matte paper is good approximation except at grazing angles (θ_i and θ_r near 90 degrees), where the surface begins to look “shiny.”

There is sometimes confusion as to how Lambertian surfaces reflect light. You will sometimes see a statement like

- A Lambertian surface reflects light equally into all directions. Lambertian surfaces therefore are also called isotropic/uniform/perfectly diffuse reflectors.

You can also see statements like

- A Lambertian surface reflects light with a cosine angular distribution. Lambertian surfaces therefore are also called cosine reflectors.

These statements conjure up different mental images, as shown in Fig. 13.20, and appear to be contradictory.

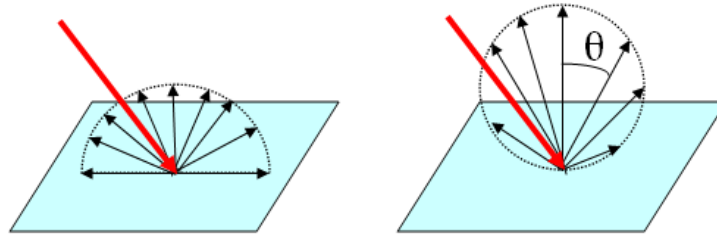


Figure 13.20: The mental images corresponding to the two descriptions of Lambertian surfaces.

However, either statement can be correct or incorrect, depending on which measure of “light” is used. Each *point* of a Lambertian surface reflects *intensity* in a cosine pattern, as in the right-hand panel of Fig. 13.20, which is the correct form of the second statement. However, when the surface is viewed with a radiometer with a finite field of view, the area of the surface seen by the radiometer is proportional to $1/\cos\theta$. The radiance measured comes from the the intensity reflected from each point ($\propto \cos\theta$) times the number of points seen ($\propto 1/\cos\theta$), and is therefore independent of θ . This is illustrated in Fig. 13.21. Thus the measured *reflected radiance is independent of the viewing direction*, which corresponds to the first statement.

In addition to their mathematical simplicity, Lambertian surfaces have an extremely important property. To see what it is, compute the irradiance reflectance $R = E_r/E_i$ of a Lambertian surface. Recall the general equation (13.46) for the irradiance reflectance:

$$R = \frac{\iint_{2\pi_r} \left[\iint_{2\pi_i} L_i(\theta_i, \phi_i) BRDF(\theta_i, \phi_i, \theta_r, \phi_r) |\cos\theta_i| d\Omega_i \right] |\cos\theta_r| d\Omega_r}{\int \int_{2\pi_i} L_i(\theta_i, \phi_i) |\cos\theta_i| d\Omega_i}$$

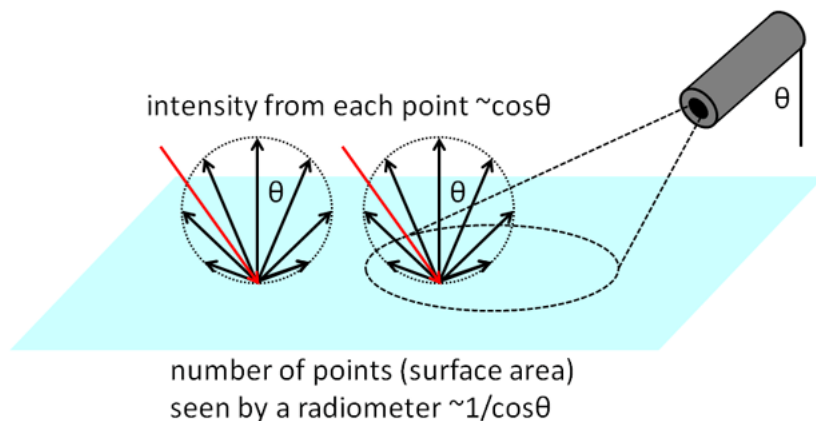


Figure 13.21: Resolution of the paradoxical statements about how Lambertian surfaces reflect light.

Substituting the Lambertian BRDF of Eq. (13.48) into this equation and rearranging gives

$$R = \frac{\frac{\rho}{\pi} \left[\iint_{2\pi_i} L_i(\theta_i, \phi_i) |\cos \theta_i| d\Omega_i \right] \iint_{2\pi_r} |\cos \theta_r| d\Omega_r}{\iint_{2\pi_i} L_i(\theta_i, \phi_i) |\cos \theta_i| d\Omega_i} = \rho ,$$

since the integrals over $2\pi_i$ cancel, and the integral over $2\pi_r$ equals π . ***A Lambertian surface thus has the property that its irradiance reflectance R equals its reflectivity ρ and, furthermore, its irradiance reflectance R is independent of the incident radiance.*** Both of these results are true *only* for Lambertian surfaces. For non-Lambertian surfaces, R generally depends both on the surface and on the incident lighting, as seen in Eq. (13.46).

Fortunately, the use of a Lambertian BRDF for ocean bottom materials in radiative transfer calculations usually results in errors of less than 10 per cent (and often less, see [Mobley et al., 2003](#)) in predicted upwelling radiances as viewed for near-nadir directions (the geometry for most remote sensing), so long as the reflectivity ρ of the assumed Lambertian surface equals the irradiance reflectance R of the actual non-Lambertian surface. Thus the crucial measurement to make is the irradiance reflectance of the bottom for the incident lighting conditions of interest.

13.8 Other Measures of Reflectance

The bi-directional reflectance distribution function (BRDF) just discussed tells us everything we need to know about how a surface reflects light. However, it in general depends on four angles ($\theta_i, \phi_i, \theta_r, \phi_r$) plus the wavelength λ (and in general on the state of polarization, which is ignored for the moment), which makes it a complicated function. The BRDF is consequently difficult to measure or model. Measuring even a partial BRDF for a few values of $\theta_i, \phi_i, \theta_r, \phi_r, \lambda$ is a difficult and tedious task in the laboratory and is almost never attempted in the ocean. Of course, everyone wants to have some easily made (compared to a BRDF) measure of surface reflectance that, with appropriate assumptions, can be used

to describe the optical properties of the surface. (The surface can be the water surface itself, or the water surface plus the water column, or the surface of an object.) This leads us to various other reflectances and quantities derived from the BRDF.

13.8.1 Albedos

There are many definitions of “albedo.” [Hapke \(1993\)](#) defines bolometric, Bond, geometric, hemispherical, normal, physical, plane, single-scattering, and spherical albedos, as well as an albedo factor. Not a single one of these albedos corresponds to how albedo is defined in [Mobley \(1994, page 193\)](#) and commonly used in optical oceanography. Fortunately, oceanographers do not have to deal with all of these albedos, but we should clarify a point that can cause confusion when reading papers on reflectance, which sometimes just say, “...and the albedo is...” without telling you which one they are using.

Optical oceanographers generally think of the albedo as being the ratio of the upwelling plane irradiance to the downwelling plane irradiance, *for whatever conditions of incident lighting you have in nature at the time of measurement*. (This is how the albedo is defined in [Mobley \(1994\)](#).) This is what you need to know to compute an energy balance in the ocean, for example. Thus the oceanographers’ albedo is the same as the irradiance reflectance $R = E_u/E_d$.

Many surface-reflectance scientists (e.g., in the paint industry) like to define their albedos and reflectances in terms of *isotropic illumination of the surface*, i.e., the incident radiance $L_i(\theta_i, \phi_i)$ is a constant independent of (θ_i, ϕ_i) . For isotropic incident radiance, the general equation (13.46) for the irradiance reflectance R ,

$$R = \frac{\iint_{2\pi_r} \left[\iint_{2\pi_i} L_i(\theta_i, \phi_i) BRDF(\theta_i, \phi_i, \theta_r, \phi_r) |\cos \theta_i| d\Omega_i \right] |\cos \theta_r| d\Omega_r}{\iint_{2\pi_i} L_i(\theta_i, \phi_i) |\cos \theta_i| d\Omega_i}, \quad (13.49)$$

reduces to just

$$A \triangleq \frac{1}{\pi} \iint_{2\pi_r} \left[\iint_{2\pi_i} BRDF(\theta_i, \phi_i, \theta_r, \phi_r) |\cos \theta_i| d\Omega_i \right] |\cos \theta_r| d\Omega_r. \quad (13.50)$$

This quantity is called the Bond or spherical albedo, or the spherical or bi-hemispherical reflectance, or just the albedo, depending on the author’s preference. Note that this A is not equal to $R = E_u/E_d$ unless the incident lighting is isotropic (which never occurs in nature) or unless the surface is Lambertian (which never occurs in nature). For a Lambertian surface, $A = \rho$, where ρ is the reflectivity of the surface.

The same convention of assuming isotropic illumination of the surface is often used when defining other reflectances, e.g., the hemispherical-directional reflectance. Note that the convention of using isotropic illumination when defining albedos and reflectances is not necessarily bad: it removes a complicating factor—variable incident lighting—from the discussion of surface properties. However, oceanographers cannot control their incident lighting; they have to live with whatever incident radiance nature gives them.

13.8.2 The Irradiance Reflectance versus The Bi-Hemispherical Reflectance

As already noted, the oceanographers’ albedo or irradiance reflectance $R = E_u/E_d$ as given by Eq. (13.46) is a bi-hemispherical reflectance, but it is not the bi-hemispherical

reflectance as defined in books such as Hapke's, because the oceanographer's R uses the actual incident radiance distribution in Eq. (13.46) rather than an isotropic incident radiance.

13.8.3 The Remote-Sensing Reflectance

The oceanographers' remote-sensing reflectance

$$R_{\text{rs}}(\theta_{\text{r}}, \phi_{\text{r}}) \triangleq \frac{L_{\text{w}}(\theta_{\text{r}}, \phi_{\text{r}})}{E_{\text{d}}} \quad (\text{sr}^{-1}),$$

has the same units as the BRDF, but they are not the same thing. Note in particular that R_{rs} uses the downwelling radiance from all directions (as contained in E_{d}), whereas the incident radiance in the definition of the BRDF is in a collimated beam. The water-leaving radiance L_{w} is the total upward radiance L_{u} minus the radiance L_{sr} reflected by the surface itself: $L_{\text{w}} = L_{\text{u}} - L_{\text{sr}}$. Determining L_{w} from a measurement of L_{u} is always problematic, so for practical reasons some people approximate R_{rs} as $L_{\text{u}}/E_{\text{d}}$. If the measurement is made underwater at depth z , the ratio of upwelling radiance to downwelling irradiance is often called the "remote-sensing ratio", RSR :

$$RSR(z, \theta_{\text{r}}, \phi_{\text{r}}) \triangleq \frac{L_{\text{u}}(z, \theta_{\text{r}}, \phi_{\text{r}})}{E_{\text{d}}(z)} \quad (\text{sr}^{-1}).$$

13.8.4 The Reflectance Factor and the Radiance Factor

The *reflectance factor* $REFF$ (also called the *reflectance coefficient*) is defined as the ratio of the BRDF of a surface to that of a perfectly diffuse surface under the same conditions of illumination and observation. "Perfectly diffuse" means a Lambertian surface with $\rho = 1$. Thus

$$REFF(\theta_{\text{i}}, \phi_{\text{i}}, \theta_{\text{r}}, \phi_{\text{r}}) \triangleq \frac{BRDF(\theta_{\text{i}}, \phi_{\text{i}}, \theta_{\text{r}}, \phi_{\text{r}})}{BRDF_{\text{Lamb}}(\text{with } \rho = 1)} = \pi BRDF(\theta_{\text{i}}, \phi_{\text{i}}, \theta_{\text{r}}, \phi_{\text{r}}). \quad (13.51)$$

The *radiance factor* $RADF$ is defined as the reflectance factor for normal illumination, i.e., for $\theta_{\text{i}} = 0$. Thus

$$RADF(\theta_{\text{r}}, \phi_{\text{r}}) \triangleq \pi BRDF(0, 0, \theta_{\text{r}}, \phi_{\text{r}}). \quad (13.52)$$

13.8.5 The Fresnel Reflectance

The Fresnel reflectance R_{F} is the reflectance of a perfectly smooth surface between two media of different indices of refraction n . R_{F} is discussed in detail in Section 13.2 for unpolarized light and in Section 13.3 for polarized light.

Note that the Fresnel reflectance describes the reflectance of the surface itself. The irradiance and remote-sensing reflectances, R and R_{rs} , when measured just above the sea surface describe the reflectance of the sea surface plus the water beneath the surface.

R_{F} can be combined with Dirac delta functions (Section 1.4.3) to create a BRDF. Consider the BRDF

$$BRDF(\theta_{\text{i}}, \phi_{\text{i}}, \theta_{\text{r}}, \phi_{\text{r}}) = 2R_{\text{F}} \delta(\sin^2 \theta_{\text{r}} - \sin^2 \theta_{\text{i}}) \delta(\phi_{\text{r}} - \phi_{\text{i}} \pm \pi).$$

Inserting this BRDF into the general equation (13.45) for reflected radiance,

$$L_r(\theta_r, \phi_r) = \iint_{2\pi_i} L_i(\theta_i, \phi_i) BRDF(\theta_i, \phi_i, \theta_r, \phi_r) \cos \theta_i d\Omega_i ,$$

gives

$$\begin{aligned} L_r(\theta_r, \phi_r) &= R_F \iint_{2\pi_i} L_i(\theta_i, \phi_i) \delta(\sin^2 \theta_r - \sin^2 \theta_i) \delta(\phi_r - \phi_i \pm \pi) 2 \cos \theta_i \sin \theta_i d\theta_i d\phi_i \\ &= R_F \int_0^1 L_i(\theta_i, \phi_i = \phi_r \pm \pi) \delta(\sin^2 \theta_r - \sin^2 \theta_i) d \sin^2 \theta_i \\ &= R_F L_i(\theta_i = \theta_r, \phi_i = \phi_r \pm \pi) . \end{aligned}$$

This last equation is the form usually seen in the definition of the Fresnel reflectance as being the ratio of reflected to incident (ir)radiances for angles related by the law of reflection.

Ocean Color Remote Sensing

Remote sensing of the ocean uses electromagnetic signals from the near UV (wavelengths from ~ 300 to 400 nm) to various radar bands (wavelengths from ~ 1 cm to ~ 1 m). This chapter introduces *ocean color radiometry*, commonly called “ocean color remote sensing,” which typically uses visible (400 to 700 nm) and near-IR (wavelengths from 700 to less than 2000 nm) light. Ocean color remote sensing uses data obtained from aircraft- or satellite-borne instruments to obtain information about the constituents of natural waters, the corresponding IOPs, the bottom depth and type, or the surface wave state.

The applications of ocean color remote sensing are extensive, varied, and fundamental to understanding and monitoring the global ecosystem. The current applications of ocean color data include

- Mapping of chlorophyll concentrations
- Measurement of inherent optical properties such as absorption and backscatter
- Determination of phytoplankton physiology, phenology, and functional groups
- Studies of ocean carbon fixation and cycling
- Monitoring of short- and long-term ecosystem changes
- Fisheries management
- Mapping of coral reefs, sea grass beds, and kelp forests
- Mapping of shallow-water bathymetry and bottom type
- Monitoring of water quality
- Detection of harmful algal blooms and pollution events

The International Ocean Colour Coordinating Group (IOCCG) has a lengthy report, *Why Ocean Colour? The Societal Benefits of Ocean-Colour Technology* (Platt et al., 2008) describing the many applications and societal benefits of ocean color remote sensing. A National Research Council study *Assessing the Requirements for Sustained Ocean Color Research and Operations* (National Research Council, 2011) also shows the diversity of ocean color applications. These reports give many additional applications and details

about each. The definitive treatise on the physical and mathematical foundations of ocean color remote sensing is *Physical Principles of Ocean Color Remote Sensing* (Gordon, 2019), which can be downloaded from the [IOCCG website](#).

Ocean color remote sensing from satellites began with the the Coastal Zone Color Scanner (CZCS), which was launched in 1978. CZCS was a *multi-spectral* sensor, meaning that it had only a few wavelength bands with bandwidths of 10 nm or more. After the phenomenal success of that “proof of principle” sensor, numerous other multispectral sensors have been developed and launched. Those later sensors generally had a few more bands with narrower bandwidths. Thus the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) added a band near 412 nm to improve the detection of CDOM. The near-IR bands are used for atmospheric correction. There is today much interest in the use of *hyperspectral* sensors, which typically have 100 or more bands with nominal bandwidths of 5 nm or less. Figure 14.1 shows the wavelength bands for a few representative sensors. The MODIS (MODerate resolution Imaging Spectroradiometer) sensor has additional bands in the 400-900 nm range, not shown in Fig. 14.1, which are used for detection of clouds, aerosols, and atmospheric water vapor. The bands shown are the ones used for remote sensing of water bodies. The launch dates of the three satellite sensors are shown. Acker (2015) gives a very nice history of these sensors and NASA’s long involvement with ocean color remote sensing. The [Compact Airborne Hyperspectral Imager](#) (CASI) is a commercially available hyperspectral sensor that is widely used in airborne remote sensing of coastal waters. It has 228 slightly overlapping bands, each with a nominal 1.9 nm bandwidth and covering the 400 - 1000 nm range. CASI users often select a subset of these bands as needed for a particular application. Lidar bathymetry systems typically use either 488 nm in “blue” water or 532 nm in “green” water. Those wavelengths can be obtained from high-power lasers and give close to optimum water penetration for the respective water types.

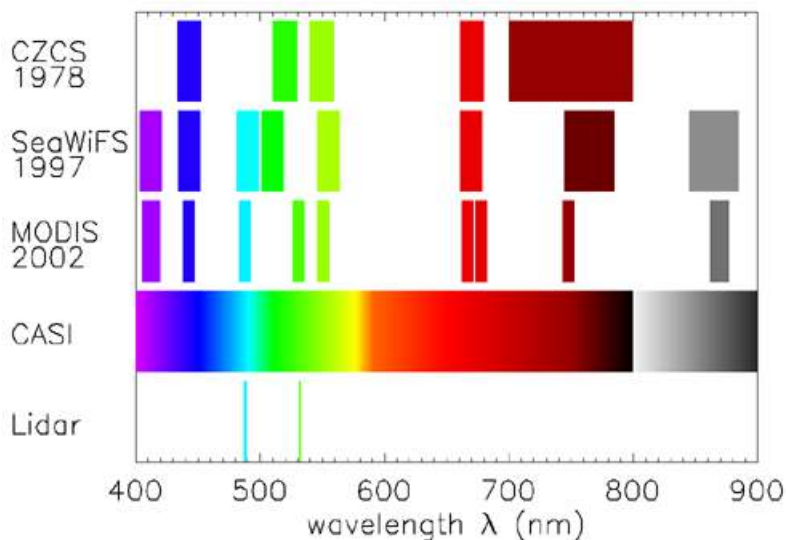


Figure 14.1: Wavelength bands used by various ocean color remote sensors.

Ocean color remote sensing usually obtains information for one spatial point at a time, most applications combine measurements from many points to build up an image, i.e. a 2-D spatial map of the ocean displaying the desired information at a give time. Imagery

acquired at different times then gives temporal information. Satellite systems typically have spatial resolution (the size of one image pixel at the ocean surface) of 250 m to 1 km. Those systems are useful for regional to global scale studies. Airborne systems can have resolutions as small as 1 m, as required for applications such as mapping coral reefs.

Ocean-color remote sensing is conceptually simple. Sunlight, whose spectral properties are known, enters the water body. The spectral character of the sunlight is then altered, depending on the absorption and scattering properties of the water body, which of course depend on the types and concentrations of the various constituents of the particular water body. Part of the altered sunlight eventually makes its way back out of the water and is detected by the sensor on board an aircraft or satellite. If we know how different substances alter sunlight, for example by wavelength-dependent absorption, scattering, or fluorescence, then we can hope to deduce from the altered sunlight what substances must have been present in the water, and in what concentrations. However, this process of “working backwards” from the sensor to the ocean is an inverse problem that is fraught with difficulties, as seen on the next pages. Nevertheless, these difficulties can be overcome, and ocean color remote sensing has completely revolutionized our understanding of the oceans at local to global spatial scales and daily to decadal temporal scales.

This chapter introduces ocean color remote sensing. Terminology is defined in the next section. Remote sensing is then described as a type of radiative transfer inverse problem. Constraints on satellite sensors are then illustrated by counting the number of photons reaching the top of the atmosphere (TOA). Finally, differences in ocean and terrestrial remote sensing via thematic mapping are discussed. The topic of atmospheric correction for ocean-color remote sensors is treated in the next chapter.

14.1 Remote-sensing Terminology

Remote sensing, like any field of science, has specialized terminology, which we introduce here.

14.1.1 Active versus Passive Remote Sensing

Remote sensing can be *active* or *passive*.

Active remote sensing means that a signal of known characteristics is sent from the sensor platform—an aircraft or satellite—to the ocean, and the return signal is then detected after a time delay determined by the distance from the platform to the ocean and by the speed of light. One example of active remote sensing at visible wavelengths is the use of laser-induced fluorescence to detect chlorophyll, yellow matter, or pollutants. In laser fluorosensing, a pulse of UV light is sent to the ocean surface, and the spectral character and strength of the induced fluorescence at UV and visible wavelengths gives information about the location, type and concentration of fluorescing substances in the water body. Another example of active remote sensing is lidar bathymetry. This refers to the use of pulsed lasers to send a beam of short duration, typically about a nanosecond, toward the ocean. The laser light reflected from the sea surface and then slightly later from the bottom is used to deduce the bottom depth. The depth is simply $0.5(c/n)\Delta t$, where c is the speed of light in vacuo, n is the water index of refraction, Δt is the time between the arrival of the surface-reflected light and the light reflected by the bottom, and the 0.5

accounts for the light traveling from the surface to the bottom and back to the surface. Laser fluorosensing and lidar bathymetry are discussed in detail in [Measures \(1992\)](#).

Passive remote sensing simply observes the light that is naturally emitted or reflected by the water body. The night-time detection of bioluminescence from aircraft is an example of the use of emitted light at visible wavelengths. The most common example of passive remote sensing, and the one primarily discussed in this chapter, is the use of sunlight that has been scattered upward within the water and returned to the sensor. This light can be used to deduce the concentrations of chlorophyll, CDOM, or mineral particles within the near-surface water; the bottom depth and type in shallow waters; and other ecosystem information such as net primary production, phytoplankton functional groups, or phytoplankton physiological state.

14.1.2 Data Resolution

The quality of remote sensing data is determined by the spatial, spectral, radiometric and temporal resolutions.

- **Spatial resolution** refers to the “ground” size of an image pixel, which may be as small as 1 m for airborne systems to more than 1000 meters for satellite systems.
- **Spectral resolution** refers to the number, spacing, and width of the different wavelength bands recorded. This can range from one broad band covering the visible spectrum to several hundred bands, each a few nanometers wide. The spectral resolution is quantified by the
 - **Sampling interval**, which is the spectral distance between the centers or peaks of adjacent spectral channels along a spectrum
 - **Band width** (or band pass), which is the full width at half maximum (FWHM) of a spectral channel response to monochromatic light

Figure 14.2 illustrates the difference in sampling interval and band width. The three blue curves indicate a sampling interval that is greater than the band width, as is common for multispectral sensors. The red curves illustrate a sampling interval that is less than the band width, which is often the case for hyperspectral sensors.

Spectral resolution can be further described the sensor type:

- **Monochromatic** refers to a sensor with 1 very narrow wavelength band, e.g. at a laser wavelength.
- **Panchromatic** refers to 1 very broad wavelength band, usually over the visible range, e.g. a black and white photograph.
- **Multispectral** sensors have several (typically 5-10) wavelength bands, each typically 10-20 nm wide.
- **Hyperspectral** sensors have 30 or more bands with 10 nm or better resolution. Typical hyperspectral sensors have more than 100 bands, each less than 5 nm wide.

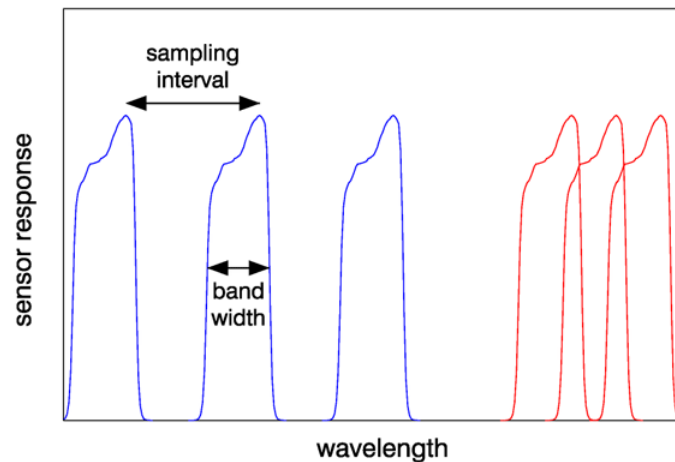


Figure 14.2: Illustration of the difference in sampling interval and band width.

- **Radiometric resolution** refers to the number of different intensities of light the sensor is able to distinguish, often specified as the number of recorded bits. An n -bit sensor can record 2^n levels of light intensity between the smallest or darkest expected value (usually zero, or no light) and the brightest expected value. Typically sensors range from 8 to 14 bits, corresponding to $2^8 = 256$ to $2^{14} = 16,384$ levels or “shades” of color in each band. The radiometric resolution determines how small of a change in the measured quantity can be recorded by the instrument. Usable resolution depends on the instrument noise.
- **Temporal resolution** refers to the frequency of flyovers by the sensor. This is relevant for time-series studies, or if cloud cover over a given area makes it necessary to repeat the data collection.

14.1.3 Data Processing Levels

Processing remotely sensed imagery involves many steps to convert the radiance measured by the sensor to the information desired by the user. These processing steps result in different “levels” of processed data, which are often described as follows:

- **Level 0** refers to unprocessed instrument data at full resolution. Data are in “engineering” units such as volts or digital counts.
- **Level 1a** is unprocessed instrument data at full resolution, but with information such as radiometric and geometric calibration coefficients and georeferencing parameters appended, but not yet applied, to the Level 0 data.
- **Level 1b** data are Level 1a data that have been processed to sensor units (e.g., radiance units) via application of the calibration coefficients. Level 0 data are not recoverable from level 1b data. Science starts with Level 1b data.

Atmospheric correction converts the top-of-the-atmosphere radiance of Level 1b to the normalized water-leaving reflectance $[\rho]_{\text{N}}^{\text{ex}}$ of Level 2. $[\rho]_{\text{N}}^{\text{ex}}$ is defined in Section

15.3, and the Atmospheric Correction Chapter 15 describes in detail the steps needed to get from Level 1b to Level 2.

- **Level 2** refers to normalized reflectance $[\rho]_N^{\text{ex}}$ and derived geophysical variables (e.g., chlorophyll concentration or bottom depth) at the same resolution and location as Level 1 data.
- **Level 3** are variables mapped onto uniform space-time grids, often with missing points interpolated or masked, and with regions mosaiced together from multiple orbits to create large-scale maps, e.g. of the entire earth.
- **Level 4** refers to results obtained from a combination of satellite data and model output (e.g., the output from an ocean ecosystem model), or results from analyses of lower level data (i.e., variables that are not measured by the instruments but instead are derived from those measurements).

14.1.4 Validation

It is always of interest to compare remotely sensed values with “ground truth” or known values, usually measured in situ, of the quantity being determined by remote sensing. This leads to various ways of describing the difference in remotely sensed and in situ values of the same quantity.

- **Reliability** refers to the certainty with which the value of a remotely sensed quantity is actually a measure of the quantity of interest. For example, if chlorophyll concentration is the quantity of interest, it is desirable that the value obtained be a measure only of chlorophyll and not, perhaps, a false value caused by high concentrations of mineral particles or dissolved substances that can change a spectrum so as to give an incorrect chlorophyll value in a retrieval algorithm. A reliable chlorophyll retrieval algorithm would give the chlorophyll value regardless of what other substances are in the water. (Note that in some fields, reliability is defined as the ability to reproduce a measurement; this is different than the meaning used here.)
- **Reference value** is the quantity (often measured in situ) being used for comparison with a remotely sensed value. This is often thought of as the “true” value of the quantity, but it must be remembered that the “truth” is seldom known. A value measured in situ is also subject to errors determined by the instrument and methodology used, and thus also may not be the true value of the quantity being measured.
- **Error** is the difference between a measurement and a reference value. Errors can be systematic or random.
 - **Random errors** are differences between a measurement and a reference value that are determined by random physical processes such as electronic noise. Although the statistical properties of the errors can be determined, the value of the error in any particular measurement cannot be predicted. The statistical distribution of random errors is often assumed to be Gaussian, but this is often just a mathematical convenience rather than a correct description of the underlying physical process. The effects of random errors can be reduced by making repeated measurements of the same quantity and then averaging the results.

- **Systematic errors** are biases or offsets between a measurement and reference value that are caused by imperfect instruments, methodologies, or algorithms. Examples are additive offsets caused by failure to remove dark current values or multiplicative errors caused by imperfect instrument calibration. The purpose of instrument calibration is the removal of systematic errors from the measurements. The systematic error is quantified as the difference in the reference value and the average of repeated measurements. Systematic errors cannot be reduced by making additional measurements and averaging the results.
- **Precision** is the reliability with which an instrument will give the same value when repeated measurements of the same quantity are made. Precision is determined by repeated measurements, without regard for any reference value. It can be quantified by the standard deviation of the values measured.

Figure 14.3 illustrates the differences in accuracy and precision, and systematic and random errors. The red curve represents the distribution of values obtained by many repeated measurements; the red dot represents any one of those measurements. The arrow showing the precision is drawn as the mean value ± 1.96 standard deviations of the measured values; this means that 95% of the measurements lie within that range of values. In this example, there is a large systematic error, so that the measured quantity is not very accurate, but there is a relatively small spread of measured values, so that the precision is high compared to the systematic error. Ideally, measurements have a small systematic error and high precision. In practice, if repeated measurements can be made, it is better to have a small systematic error even if the precision is low, because averaging repeated measurements reduces the effects of random errors and leads to an accurate average value.

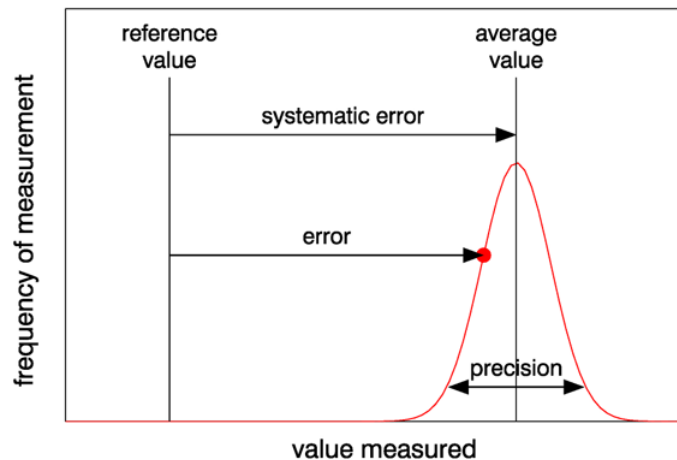


Figure 14.3: Illustration of terms used in validating measurements.

Finally, when developing a mathematical model or algorithm, there are two potential sources of error. The model may leave out some of the physical processes essential to describing the phenomenon of interest. Even if the physics is correct, there may be errors in the computer programming of the model equations. The model must be checked for both of these types of errors.

- **Verification** refers to making sure that the model equations have been programmed correctly.
- **Validation** then refers to checking the model output against reference values to make sure that both the model physics and the computer programming are correct.

14.2 Inverse Problems

The development of the radiative transfer equation and its solution techniques in Chapters 9 and 10 and Appendices E and F is concerned with the *forward* or *direct problem* of radiative transfer theory. The rules of the game are simple: *Given the inherent optical properties of the water and the properties of the boundaries, find the radiance distribution throughout and leaving the water.* This problem has a unique solution, which means that a given set of IOPs and boundary conditions yields a unique radiance distribution. The only limits on the accuracy of computed radiances are the accuracy with which we specify the IOPs and the boundary conditions, and the amount of computer time we wish to devote to the numerical solutions. In this sense *the direct problem of computing radiances can be regarded as solved.*

Figure 14.4 shows the conceptual process of solving the forward problem. In principle, we can start with the fundamental physical properties of the particles and dissolved substances in the ocean and derive the water IOPs from the physical properties (e.g., using Mie theory to compute the VSF from the particle properties and size distribution). In optical oceanography, we often begin with direct measurements of the IOPs. We then apply suitable boundary conditions and solve the very complicated RTE to obtain the radiance distribution. Any other quantities of interest, such as irradiances or AOPs can then be computed from the radiances.

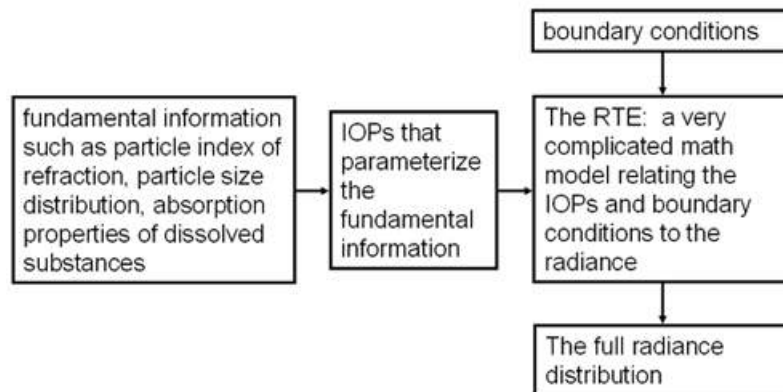


Figure 14.4: The conceptual process involved in solving a forward radiative transfer problem.

The inverse problem of radiative transfer theory can be stated as follows: given radiometric measurements of underwater or water-leaving light fields, determine the inherent optical properties of the water. This is very much an *unsolved problem*. Both conceptual and practical limits are encountered in inverse problems. Unfortunately, *remote sensing is an inverse problem.*

The first problem we encounter is *uniqueness of the solution*. Consider the following situation. A body of water with a particular set of IOPs and boundary conditions has an underwater radiance distribution $L_1(z, \theta, \phi, \lambda)$. If the boundary conditions now change, perhaps because the Sun moves, there will be a different radiance distribution $L_2(z, \theta, \phi, \lambda)$ within the water, even though the IOPs remain unchanged. Can we correctly recover the same set of IOPs from the two different light fields? Can we distinguish between $L_1 \neq L_2$ because of a change in boundary conditions, as opposed to $L_1 \neq L_2$ because of a change in IOPs? Because the same set of IOPs can yield different radiance distributions for different boundary conditions, as we just saw, we are led to ask if two different sets of IOPs and boundary conditions can lead to the same radiance distribution. In other words, *is there even in principle a unique solution to the inverse problem stated above?*

Another problem often encountered with inverse solutions is the *stability of the solution*, or its *sensitivity to errors in the measured radiometric variables*. In direct problems we often find that a small error (say 5%) in the IOPs or boundary conditions leads to a correspondingly small error in the computed radiance. With inverse problems we often find that small errors in the measured radiometric quantities lead to large errors, or even unphysical results, in the retrieved IOPs. Sensitivity of the inversion scheme to small errors in the input data often renders inversion algorithms useless in practice, even though they appear in principle to be quite elegant and satisfactory.

It can be shown that if the full radiance distribution is measured with perfect accuracy, there is in principle a unique inverse solution to the RTE to obtain the full set of IOPs. But from a practical standpoint, if we have to measure the entire radiance distribution throughout the water body with high accuracy to obtain the IOPs, we could measure the IOPs themselves just as easily. An inverse method is useful only when it saves us time, money, or effort. What is desired is a recovery of at least some of the IOPs from a limited set of imperfect radiometric measurements. We have seen one example of this in Gershun's law, Eq. (10.33), which allows us to recover the absorption coefficient from measured values of the plane and scalar irradiances, if there are no internal sources present.

In remote sensing, we have a very limited set of imperfect light field measurements, namely just the water leaving radiance or remote-sensing reflectance, from which we want to retrieve as much information as possible about the water body. Our input measurements fall far short of measuring the full radiance distribution, and the measurements we do have may contain substantial errors due to poor atmospheric correction or inaccurate radiometer calibration. Thus we expect *a priori* that we will not be able to recover a full set of water IOPs, and that what is recovered may contain large errors. Oceanic remote sensing is thus a very difficult inverse problem. The various inversion algorithms discussed on the following pages show the wide range of techniques developed over the years to address the inherent difficulties of inverting remote sensing measurements to obtain information about the ocean. Each of these techniques has its strengths and weaknesses, and each is imperfect, but each still has demonstrated great value to oceanographers.

Inversions are always based on an assumed model that relates what is known to what is desired. The inversion is then effected by using the known quantities as inputs to the model, whose output is an estimate of the desired quantities. In some cases the model is simple. For example, if historical data relating the chlorophyll concentration to the ratio of remote sensing reflectance at two wavelengths are used to find a best-fit function of the form $Chl = f[R_{rs}(\lambda_1)/R_{rs}(\lambda_2)]$, then the model is that function. Inserting a newly measured

reflectance at λ_1 and λ_2 then gives an estimate of the chlorophyll concentration. The accuracy of that estimate will depend on the scatter in the original data and on whether the water body being studied is similar to the one used to determine the function. In other cases the model is complicated. A neural network with many layers is a complex model in which it is often not obvious how a particular input is related to a particular output. The accuracy of a neural network inversion depends on how well the neural network represents nature and on the data used to train the network.

Because of the limited measurements available for remote sensing, inversion algorithms usually require *constraints* to limit the possible solutions obtained from a given remote sensing reflectance. Constraints can be “built in” for example as simplifications to the RTE. They can also be external, typically as additional required measurements (such as a measurement of the water leaving radiance or bottom depth at one point in an image). They can also be implicit constraints, such as a limitation of retrieved values to the range of values found in the data set used to predetermine certain parameters in the inverse model.

Figure 14.5 summarizes the conceptual issues involved with inverting remotely sensed data to obtain estimates of oceanic properties.

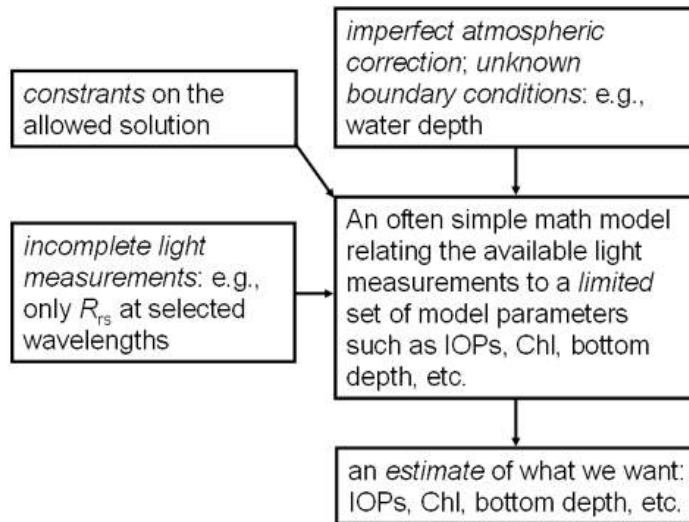


Figure 14.5: The conceptual process involved in solving a remote-sensing inverse radiative transfer problem.

14.2.1 Classification of Inverse Problems

There are many kinds of inverse problems. For example, there are *medium characterization problems*, for which the goal is to obtain information about the IOPs of the medium, which in our case is the water body with all of its constituents. This is the type of problem considered in this chapter. There are also *hidden-object characterization problems*, for which the goal is to detect or obtain information about an object imbedded within the medium, for example a submerged submarine. This book does not discuss this type of problem. Inverse problems may use optical measurements made *in situ*, as with the use of Gershun’s equation to obtain the absorption coefficient. *Remote sensing* uses measurements made outside the medium, typically from a satellite or aircraft.

Another type of inverse problem seeks to determine the properties of individual particles from light scattered by *single* particles. Such problems usually start with considerable knowledge about the particles (for example, the particles are spherical and have a known radius) and then seek to determine another specific bit of information (such as the particle index of refraction). The associated inversion algorithms usually assume that the detected light has been singly scattered. Even these highly constrained problems can be very difficult. These “individual-particle” inverse problems are not discussed in this book. In the ocean, there is no escaping multiple scattering, which greatly complicates the inverse problem, and we do not generally have the requisite *a priori* knowledge of the individual particle properties needed to constrain the inversion.

Solution techniques to inverse problems fall into two categories: explicit and implicit. *Explicit solutions* are formulas that give the desired IOPs as functions of measured radiometric quantities. A simple example is Gershun’s law when solved for the absorption in terms of the irradiances. *Implicit solutions* are obtained by *solving a sequence of direct or forward problems*. In crude form, we can imagine having a measured remote-sensing reflectance (or set of underwater radiance or irradiance measurements). We then solve direct problems to predict the reflectance for each of many different sets of IOPs. Each predicted reflectance is compared with the measured value. The IOPs associated with the predicted reflectance that most closely matches the measured reflectance are then taken to be the solution of the inverse problem. Such a plan of attack can be efficient if we have a rational way of changing the IOPs from one direct solution to the next, so that the sequence of direct solutions converges to the measured reflectance or radiance.

14.3 Counting Photons for Remote Sensing

[Bryan Monosmith, Jeremy Werdell, and Curtis Mobley contributed to this section.]

The foundation of ocean color remote sensing is sunlight that has entered the ocean, been transformed through absorption and scattering by the myriad constituents of the water body, and then been scattered out of the ocean and into a detector. It is worthwhile to consider how many photons going through this process are actually available for detection by a satellite sensor. An order-of-magnitude calculation suffices to identify many of the engineering constraints on the design of an ocean color sensor.

14.3.1 Radiance from the Sea Surface

Figure 2.18 shows that the downwelling spectral plane irradiance E_d onto the sea surface on a clear day is of order $1 \text{ W m}^{-2} \text{ nm}^{-1}$ at visible wavelengths. Most (90-98%, depending on Sun zenith angle and wind speed) of this irradiance enters the ocean. In-water irradiance reflectances $R = E_u/E_d$ are typically in the 0.01 to 0.05 range, depending on the water constituents and wavelength (R can reach 0.1 in very turbid, highly scattering waters).

Suppose that $R = 0.03$ of downwelling plane irradiance of magnitude $1 \text{ W m}^{-2} \text{ nm}^{-1}$ entering the ocean is backscattered into upward directions. If this upwardly scattered light is isotropically scattered into the 2π sr of the upward hemisphere, then the upwelling isotropic radiance just below the sea surface would be

$$L_u = \frac{0.03}{2\pi \text{ sr}} (1 \text{ W m}^{-2} \text{ nm}^{-1}).$$

For ocean waters the upwelling radiance is not isotropic. Just below the sea surface, upwelling radiance within 30 deg of the zenith (the angles relevant to remote sensing) is less than the radiance in more nearly horizontal directions by a factor of two or three, as shown in Fig. 14.6. Thus the above value for isotropic scattering can be reduced by a factor of roughly $\frac{1}{2}$ to estimate the underwater radiance in near-zenith directions.

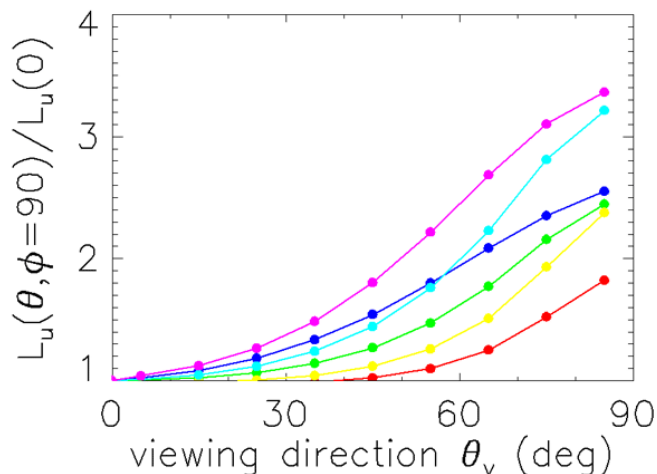


Figure 14.6: Example upwelling radiances at 90 deg to the Sun's azimuthal direction, normalized to the zenith (nadir-viewing) radiance. The six curves are for chlorophyll values of 0.05, 0.5 and 5 mg m^{-3} in Case 1 water and Sun zenith angles of 30 and 60 deg.

A beam of this radiance just below the sea surface will be reduced by a factor of t/n_w^2 when passing through the sea surface. Here t is the water-to-air radiance transmittance, which is close to 1 for directions relevant to remote sensing, namely directions within a few tens of degrees of the zenith. $n_w \approx 1.34$ is the water index of refraction. The water-leaving radiance is then roughly

$$L_w = \frac{0.03}{2n_w^2 2\pi} \approx 0.0013 \quad [\text{W m}^{-2} \text{sr}^{-1} \text{nm}^{-1}]$$

Atmospheric transmittance is 0.7 to 0.95 at visible wavelengths, so most of the water-leaving radiance will be transmitted to the top of the atmosphere (TOA), where it can be detected by a satellite. However, along the way, atmospheric scattering of solar radiation into the beam will add typically 10 to 20 times as much radiance to the beam. The TOA radiance seen by a satellite would then be of order $10 \times L_w \approx 0.01 \text{ W m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$. Typical open-ocean TOA radiances for the MODIS sensor are in the range of 0.01 (red wavelengths) to 0.08 (blue wavelengths) $\text{W m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$.

14.3.2 Photons Detected at the Top of the Atmosphere

Now that we have the radiance detected at the satellite, we can compute the numbers of photons collected and consider the related engineering matters. We start by computing the number of detected photons that come from a 1 m^2 patch of sea surface in 1 second of observation time. For specific numbers, the MODIS sensors have the following characteristics:

Physical quantity	value
altitude	705 km = 7.1×10^5 m
off-nadir viewing angle	20 deg
slant range	750 km = 7.5×10^5 m
sensor aperture radius	89 mm = 0.09 m
band width	$\Delta\lambda = 10$ nm
optical efficiency	OE = 0.6
quantum efficiency	QE = 0.9

Table 14.1: MODIS sensor characteristics.

The slant range is the distance from the satellite to the observed point on the ocean surface along the line of sight, which is taken here to be 20 deg off nadir. The optical efficiency is the fraction of light incident onto the sensor fore-optics that eventually reaches sensor material itself; the losses are due to reflections from lens surfaces or diffraction gratings, absorption by filters, etc.

The quantum efficiency is the fraction of photons reaching the sensor material that actually results in a signal, e.g., by the generation of a photo-electron. Figure 14.7 shows the quantum efficiencies of typical sensors. The “standard CCD” curve is for CCDs like those used in consumer-grade video cameras. “Scientific CCD” is for a much more expensive “science grade” CCD. The MODIS curve shows what is achievable if your budget is almost unlimited. The bottom curves are for the human eye. The human eye has a maximum QE of only a few percent for color (photopic) vision and 10% for night-time (scotopic) vision¹.

The solid angle of the sensor as seen from the Earth’s surface is

$$\Omega_{\text{aperture}} = \frac{\pi(\text{aperture radius})^2}{(\text{slant range})^2} = 4.5 \times 10^{-14} \text{ sr} .$$

The power detected by the sensor coming from a square meter of the ocean surface is

$$\begin{aligned} P_{\text{detector}} &= L \Omega_{\text{aperture}} \text{Area}_{\text{surface}} \text{OE} \Delta\lambda \\ &= (0.01 \text{ W m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}) (4.5 \times 10^{-14} \text{ sr}) (1 \text{ m}^2) (0.6) (10 \text{ nm}) \\ &\approx 2.7 \times 10^{-15} \text{ W} \end{aligned}$$

This equation has been written for viewing the satellite from the sea surface. We could also take the viewpoint of the satellite viewing the earth. In that case, the relevant solid angle would be that of the 1 m^2 pixel as seen from the satellite, and the area factor would be that of the sensor aperture. The *throughput*, or *etendue*, of an optical system is

$$\Omega_{\text{aperture}} \text{Area}_{\text{surface}} = \Omega_{\text{surface}} \text{Area}_{\text{aperture}} ,$$

which shows that these viewpoints are equivalent.

¹It thus seems that evolution has given us eyes that are adequate for finding food and avoiding tigers, and even for reading this book, but which are shockingly inefficient from an optical engineering standpoint. This figure gives another refutation of “intelligent design” for the human eye. The intelligent design in this figure was done by the physicists and optical engineers working on the CCD and MODIS sensors.

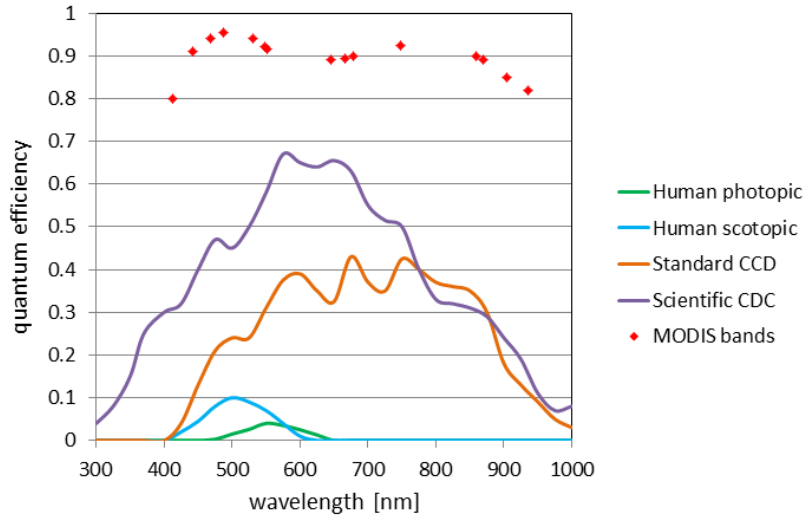


Figure 14.7: Quantum efficiencies for various light detectors. (Human eye and CCD data redrawn from learn.hamamatsu.com/articles/quantumefficiency.html.)

If we assume a wavelength of 550 nm, the corresponding number of photo-electrons released in the detector in time $t = 1$ s is

$$\begin{aligned}
 N_{\text{electrons}} &= P_{\text{detector}} t QE \frac{\lambda}{hc} \\
 &= (2.7 \times 10^{-15} \text{ J s}^{-1})(1 \text{ s})(0.9) \frac{550 \times 10^{-9} \text{ m}}{(6.63 \times 10^{-34} \text{ J s})(3 \times 10^8 \text{ m s}^{-1})} \\
 &\approx 6800
 \end{aligned}$$

where h is Planck's constant and c is the speed of light.

These 6800 electrons are for the total TOA radiance in this green band. As previously noted, the TOA radiance is typically 90% atmospheric path radiance, in which case only 680 of these electrons correspond to the water-leaving radiance from a square meter of the sea surface.

Kepler's third law of planetary motion and Newton's law of gravity give the relation between a satellite's orbital period T and the radius r of its orbit:

$$T^2 = \frac{4\pi^2 r^3}{GM},$$

where $G = 6.67 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$ is Newton's gravitational constant and $M = 5.97 \times 10^{24} \text{ kg}$ is the mass of the earth. For a satellite at altitude 705 km above the earth (whose mean radius is 6731 km), this gives a period of 6385 s. This corresponds to a speed of $v = (2\pi r)/T = 7318 \text{ m s}^{-1}$ relative to the ground. The time required for the satellite to travel 1 m is then $1.37 \times 10^{-4} \text{ s}$. For 1 m spatial resolution, this short exposure time reduces the number of detected photons by a factor of 1.37×10^{-4} compared to the 1 s collection time computed above. Thus the number of water-leaving photons collected during the time the satellite passes over the 1 m^2 area is only 0.093. Alas, the actual situation is even worse because the sensor is not collecting light from just one pixel, but from perhaps 1000 pixels

as the sensor either scans back and forth or rotates to observe a wide swath to either side of the satellite nadir point. Thus, in the 10^{-4} s time required for the satellite to move forward by 1 m, the sensor must collect photons from 1000 pixels, reducing the number for each to roughly 0.0001 photoelectrons per pixel. Collecting only 0.0001 photoelectrons per pixel would not yield a very good image.

These physical and orbital constraints show one reason why orbiting satellites do not obtain meter-scale ocean color imagery: There simply are not enough photons leaving the ocean surface from a square meter of area to form an image. Many more photons must be collected, and there are several ways to do this:

- View a larger surface area, which both increases the number of photons leaving the surface and allows for longer integration times.
- View the surface area for a longer time, e.g., from a geostationary satellite that can stare at the same point for very long times (but a geostationary satellite has an altitude of 36,000 km, which makes the solid angle much smaller).
- Get closer to the surface, e.g. by using an airborne sensor flying at a few kilometers above the sea surface. This greatly increases the solid angle of the sensor and allows for longer integration times for a slowly flying aircraft.
- Increase the bandwidth.
- Increase the aperture of the receiving optics.
- Use multiple detector elements to observe the same ground pixel nearly simultaneously, either on the same or successive scans, and then combine the photons collected from the different sensors.

Suppose, as is typical of ocean color sensors such as MODIS, that we image a 1 km^2 area of ocean surface. This increases the number of photons leaving the imaged pixel by a factor of 10^6 and the integration time by a factor of 10^3 (the increased time for the satellite to travel 1 km rather than 1 m). The sensor now collects 93,000 photoelectrons from the water-leaving radiance. The total number of TOA photoelectrons, including the atmospheric path radiance, would be roughly ten times as large, about 10^6 . In practice, this number will be less because of the duty-cycle time of the sensor: a rotating sensor will be viewing the ocean only about one-third of the time, and a scanning sensor will require time to stop and start each scan. Nevertheless, we can still collect roughly 10^6 photoelectrons for each pixel, of which 10^5 correspond to water-leaving radiance.

The signal to noise ratio, SNR, is given in general by

$$SNR = \frac{N_{PE}}{\sqrt{\mathcal{N}_{PE}^2 + \mathcal{N}_{DC}^2 + \mathcal{N}_{RO}^2 + \mathcal{N}_{QN}^2}},$$

where the signal N_{PE} is the number of photoelectrons counted. The terms in the denominator represent the noise terms for the photoelectrons (PE), dark current (DC), sensor read-out (RO), and quantization (QN). Dark current noise results from the spontaneous emission of photoelectrons within the sensor, read-out noise comes from the sensor's analog front-end electronics when the collected photoelectrons are read from the sensor, and quantization noise comes from uncertainty when the analog signal is digitized. (Writing the total noise as the square root of the sum of the individual noise terms squared assumes that

the individual noise processes are uncorrelated.) The number of photoelectrons counted during a given time interval is described by a Poisson probability distribution, for which the noise is the standard deviation of the distribution, which in turn equals the number N_{PE} of photoelectrons counted. The emission of dark current photoelectrons is also a Poisson process. The SNR can thus be written

$$SNR = \frac{N_{\text{PE}}}{\sqrt{N_{\text{PE}} + N_{\text{DC}} + \mathcal{N}_{\text{RO}}^2 + \mathcal{N}_{\text{QN}}^2}}.$$

Assuming no dark current or other noise, this gives an SNR of

$$SNR = \frac{10^6}{\sqrt{10^6}} = 1000.$$

for the number of photoelectrons estimated above. The actual SNR will be somewhat less due to dark current, read out, and quantization noise. The MODIS sensor has SNR requirements of 750 to 1000 for various bands, so we have achieved the approximate number of photons needed at the sensor.

These simple estimates illustrate the severe constraints on the design of any ocean color sensor. In practice, the engineering of a satellite ocean color sensor requires great sophistication to achieve the needed SNR.

14.4 Thematic Mapping

Thematic mapping refers to the determination and display of a particular type of information (the theme). In terrestrial and oceanic remote sensing, a common theme is the type of surface material. On land, a thematic map might display the land areas covered by forest, grassland, water, crops, bare soil, pavement, etc. In shallow waters, the thematic map might distinguish bottom areas covered by mud, sand, rock, sea grass, coral, etc. Much work has been done recently on mapping bathymetry, bottom type, and water IOPs as extracted from hyperspectral imagery. This page compares the supervised classification technique used for terrestrial thematic mapping with spectrum-matching techniques (Mobley et al., 2005; Dekker et al., 2011, e.g.) for shallow-water mapping of bottom type.

The simultaneous retrieval of bathymetry, bottom classification, and water IOPs is a much more difficult task than traditional thematic mapping to determine land surface type, as used in terrestrial remote sensing. In terrestrial thematic mapping, only the type of land surface must be deduced from an atmospherically corrected image spectrum; there are no confounding influences by water IOPs and depth. We will see that *terrestrial techniques for supervised classification are **not** well suited to the oceanic problem because of the additional complications of bottom depth and water optical properties, neither of which are present in terrestrial remote sensing.*

14.4.1 Supervised Classification

In *supervised classification* the object is to associate a given image spectrum with one of several pre-determined classes of spectra. In terrestrial remote sensing these classes are typically defined as soil, grass, trees, water, pavement, etc. A thematic map of earth

surface features is then generated by classifying the spectrum from each image pixel into one of the pre-determined classes.

One approach to supervised classification is to compute the mean spectrum for each class and a corresponding covariance matrix that defines the “size” of each class of spectra about its mean. The image spectrum is then compared only with the mean spectrum and size for each class, and the image spectrum is statistically associated with the class it is most likely to belong to according to some metric for distance between the image and mean spectra and user-specified assumptions about the statistical properties of the class members.

This section considers the terrestrial and oceanic problems in more detail and shows that *the standard terrestrial thematic mapping methodology based on supervised classification is not easily applied to the ocean remote sensing problem.*

14.4.2 Covariance and Correlation Matrices

Consider a collection of N remote sensing reflectance spectra R_{rs} , each with K wavelengths, which we denote by $R_n(\lambda_k)$, $n = 1, \dots, N$ and $k = 1, \dots, K$ (dropping the rs subscript on R_{rs} for convenience). The spectra can be regarded as column vectors:

$$\mathbf{R}_n = \begin{pmatrix} R_n(\lambda_1) \\ R_n(\lambda_2) \\ \vdots \\ R_n(\lambda_K) \end{pmatrix} = [R_n(\lambda_1), R_n(\lambda_2), \dots, R_n(\lambda_K)]^T, \quad (14.1)$$

where bold type indicates a vector or matrix, and superscript T indicates transpose. In the spectrum matching technique described previously, these spectra are the database spectra, N is usually 10^5 or more, and K would be 75 for spectra from 380 to 750 nm with 5 nm resolution. Let

$$\mathbf{I} = [R(\lambda_1), R(\lambda_2), \dots, R(\lambda_K)]^T$$

be the image spectrum that is to be classified.

Now consider subsets of the entire database that define various classes of spectra. To be specific in the illustrative computations below, we chose four classes of spectra: R_{rs} for 10 sand and sediment spectra seen through 0.01 m of water, 10 coral spectra seen through 0.01 m of water, and the same sand and coral spectra seen through 10 m of the same water. The water IOPs were based on measurements of the very clear water in the Bahamas. The sand and sediment spectra range from clean ooid sand to heavily biofilmed, darker sand. The coral spectra are different species of corals. Figure 14.8 shows the individual spectra in these four classes. To minimize the array sizes for the printout of Table 1 below, we subsampled the spectra to wavelengths of 400, 450, ..., 650, 700 nm, so that $K = 7$. The subsampled spectra are shown in Fig. 14.9.

These spectra are obviously correlated in wavelength. The amount of correlation between one wavelength and another is quantified by the covariance and correlation matrices, which are computed as follows. Let $m = 1, \dots, M$ label the class, with M being the total number of classes (here 4). Class m contains N_m spectra (here, $N_m = 10$ for each class).

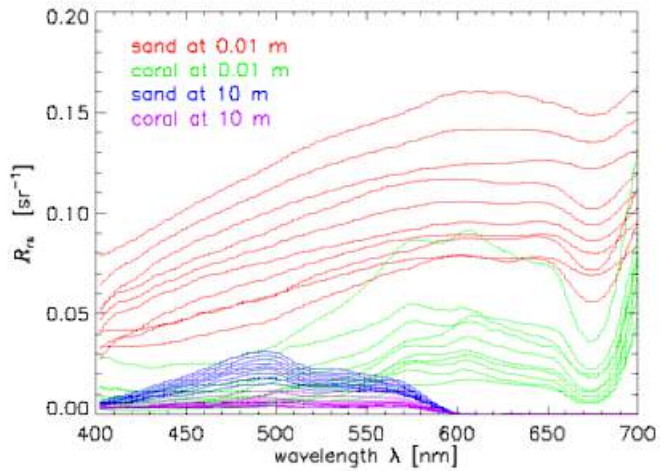


Figure 14.8: Example database R_{rs} spectra defining the four thematic classes; each class has 10 spectra.

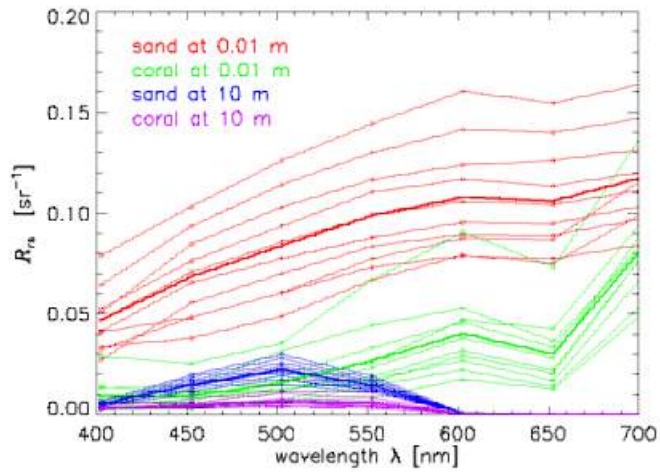


Figure 14.9: The spectra of Fig. 14.8 as re-sampled at 50 nm intervals for use in the illustrative computations (light lines). The heavy lines are the class mean spectra.

Then the mean or average spectrum for each class is defined by

$$\bar{R}_m(\lambda_i) = \frac{1}{N_m} \sum_{n=1}^{N_m} R_n(\lambda_i) , \quad (14.2)$$

where the sum is over the spectra belonging to class m . In vector notation this is

$$\bar{\mathbf{R}}_m = \frac{1}{N_m} \sum_{n=1}^{N_m} \mathbf{R}_n . \quad (14.3)$$

The mean spectra for the example four classes are shown by the heavy lines in Fig. 14.9.

The elements of the $K \times K$ class covariance matrices Σ_m are defined by

$$\Sigma_m(i, j) = \frac{1}{N_m - 1} \sum_{n=1}^{N_m} [R_n(\lambda_i) - \bar{R}_m(\lambda_i)][R_n(\lambda_j) - \bar{R}_m(\lambda_j)] . \quad (14.4)$$

$\Sigma_m(i, j)$ expresses the covariance of the class spectra at wavelength λ_i with λ_j ; $\Sigma_m(i, i)$ is the variance of the class spectra at λ_i . For remote-sensing reflectance spectra R_{rs} with units of sr^{-1} , the units of $\Sigma_m(i, j)$ are sr^{-2} . If we arrange the spectrum column vectors for class m in a $K \times N_m$ matrix with the class mean removed,

$$\mathbf{R}_{(m)} = \begin{pmatrix} R_1(\lambda_1) - \bar{R}_m(\lambda_1) & \cdots & R_{N_m}(\lambda_1) - \bar{R}_m(\lambda_1) \\ \vdots & \cdots & \vdots \\ R_1(\lambda_K) - \bar{R}_m(\lambda_K) & \cdots & R_{N_m}(\lambda_K) - \bar{R}_m(\lambda_K) \end{pmatrix} , \quad (14.5)$$

then the covariance matrix for class m can be compactly written as

$$\Sigma_m = \frac{1}{N_m - 1} \mathbf{R}_{(m)} \mathbf{R}_{(m)}^T . \quad (14.6)$$

The elements of the $K \times K$ correlation matrix ρ_m for class m are defined from the class covariance matrix Σ_m by

$$\rho_m(i, j) = \frac{\Sigma_m(i, j)}{\sqrt{\Sigma_m(i, i)\Sigma_m(j, j)}} . \quad (14.7)$$

The following arrays show the class covariance and correlation matrices computed by these equations for the four classes of spectra shown in Fig. 14.9.

$$\Sigma_{(\text{sand at } 0.1 \text{ m})} = \begin{pmatrix} 2.462e-4 & 3.049e-4 & 3.612e-4 & 3.797e-4 & 4.141e-4 & 3.994e-4 & 3.625e-4 \\ 3.049e-4 & 4.547e-4 & 5.361e-4 & 5.447e-4 & 5.712e-4 & 5.639e-4 & 4.569e-4 \\ 3.612e-4 & 5.361e-4 & 6.338e-4 & 6.462e-4 & 6.787e-4 & 6.692e-4 & 5.449e-4 \\ 3.797e-4 & 5.447e-4 & 6.462e-4 & 6.658e-4 & 7.046e-4 & 6.912e-4 & 5.780e-4 \\ 4.141e-4 & 5.712e-4 & 6.787e-4 & 7.046e-4 & 7.546e-4 & 7.364e-4 & 6.340e-4 \\ 3.994e-4 & 5.639e-4 & 6.692e-4 & 6.912e-4 & 7.364e-4 & 7.230e-4 & 6.179e-4 \\ 3.625e-4 & 4.569e-4 & 5.449e-4 & 5.780e-4 & 6.340e-4 & 6.179e-4 & 5.856e-4 \end{pmatrix}$$

$$\rho_{(\text{sand at 0.1 m})} = \begin{pmatrix} 1.000 & 0.911 & 0.914 & 0.938 & 0.961 & 0.947 & 0.955 \\ 0.911 & 1.000 & 0.999 & 0.990 & 0.975 & 0.983 & 0.885 \\ 0.914 & 0.999 & 1.000 & 0.995 & 0.981 & 0.989 & 0.894 \\ 0.938 & 0.990 & 0.995 & 1.000 & 0.994 & 0.996 & 0.926 \\ 0.961 & 0.975 & 0.981 & 0.994 & 1.000 & 0.997 & 0.954 \\ 0.947 & 0.983 & 0.989 & 0.996 & 0.997 & 1.000 & 0.950 \\ 0.955 & 0.885 & 0.894 & 0.926 & 0.954 & 0.950 & 1.000 \end{pmatrix}$$

$$\Sigma_{(\text{coral at 0.1 m})} = \begin{pmatrix} 5.793e-5 & 4.952e-5 & 6.378e-5 & 1.143e-4 & 1.389e-4 & 1.129e-4 & 1.552e-4 \\ 4.952e-5 & 4.442e-5 & 6.222e-5 & 1.080e-4 & 1.299e-4 & 1.059e-4 & 1.417e-4 \\ 6.378e-5 & 6.222e-5 & 1.070e-4 & 1.712e-4 & 2.007e-4 & 1.655e-4 & 2.057e-4 \\ 1.143e-4 & 1.080e-4 & 1.712e-4 & 3.010e-4 & 3.578e-4 & 2.961e-4 & 3.775e-4 \\ 1.389e-4 & 1.299e-4 & 2.007e-4 & 3.578e-4 & 4.432e-4 & 3.774e-4 & 4.950e-4 \\ 1.129e-4 & 1.059e-4 & 1.655e-4 & 2.961e-4 & 3.774e-4 & 3.300e-4 & 4.306e-4 \\ 1.552e-4 & 1.417e-4 & 2.057e-4 & 3.775e-4 & 4.950e-4 & 4.306e-4 & 6.132e-4 \end{pmatrix}$$

$$\rho_{(\text{coral at 0.1 m})} = \begin{pmatrix} 1.000 & 0.976 & 0.810 & 0.865 & 0.867 & 0.817 & 0.823 \\ 0.976 & 1.000 & 0.903 & 0.934 & 0.926 & 0.875 & 0.859 \\ 0.810 & 0.903 & 1.000 & 0.954 & 0.922 & 0.881 & 0.803 \\ 0.865 & 0.934 & 0.954 & 1.000 & 0.980 & 0.940 & 0.879 \\ 0.867 & 0.926 & 0.922 & 0.980 & 1.000 & 0.987 & 0.949 \\ 0.817 & 0.875 & 0.881 & 0.940 & 0.987 & 1.000 & 0.957 \\ 0.823 & 0.859 & 0.803 & 0.879 & 0.949 & 0.957 & 1.000 \end{pmatrix}$$

$$\Sigma_{(\text{sand at 10 m})} = \begin{pmatrix} 5.792e-7 & 2.174e-6 & 3.341e-6 & 1.880e-6 & 4.235e-8 & 2.783e-9 & 2.886e-12 \\ 2.174e-6 & 1.006e-5 & 1.542e-5 & 8.357e-6 & 1.805e-7 & 1.214e-8 & 1.047e-11 \\ 3.341e-6 & 1.542e-5 & 2.369e-5 & 1.288e-5 & 2.785e-7 & 1.872e-8 & 1.645e-11 \\ 1.880e-6 & 8.357e-6 & 1.288e-5 & 7.086e-6 & 1.545e-7 & 1.033e-8 & 9.641e-12 \\ 4.235e-8 & 1.805e-7 & 2.785e-7 & 1.545e-7 & 3.409e-9 & 2.267e-10 & 2.151e-13 \\ 2.783e-9 & 1.214e-8 & 1.872e-8 & 1.033e-8 & 2.267e-10 & 1.517e-11 & 1.457e-14 \\ 2.886e-12 & 1.047e-11 & 1.645e-11 & 9.641e-12 & 2.151e-13 & 1.457e-14 & 2.776e-17 \end{pmatrix}$$

$$\rho_{(\text{sand at 10 m})} = \begin{pmatrix} 1.000 & 0.900 & 0.902 & 0.928 & 0.953 & 0.939 & 0.720 \\ 0.900 & 1.000 & 0.999 & 0.990 & 0.975 & 0.983 & 0.626 \\ 0.902 & 0.999 & 1.000 & 0.994 & 0.980 & 0.987 & 0.641 \\ 0.928 & 0.990 & 0.994 & 1.000 & 0.994 & 0.996 & 0.687 \\ 0.953 & 0.975 & 0.980 & 0.994 & 1.000 & 0.997 & 0.699 \\ 0.939 & 0.983 & 0.987 & 0.996 & 0.997 & 1.000 & 0.710 \\ 0.720 & 0.626 & 0.641 & 0.687 & 0.699 & 0.710 & 1.000 \end{pmatrix}$$

$$\Sigma_{(\text{coral at 10 m})} = \begin{pmatrix} 1.956e-7 & 5.520e-7 & 9.484e-7 & 9.105e-7 & 2.255e-8 & 1.305e-9 & 1.714e-12 \\ 5.520e-7 & 1.638e-6 & 3.068e-6 & 2.868e-6 & 7.055e-8 & 4.100e-9 & 4.858e-12 \\ 9.484e-7 & 3.068e-6 & 7.054e-6 & 6.135e-6 & 1.484e-7 & 8.731e-9 & 8.282e-12 \\ 9.105e-7 & 2.868e-6 & 6.135e-6 & 5.845e-6 & 1.424e-7 & 8.393e-9 & 9.001e-12 \\ 2.255e-8 & 7.055e-8 & 1.484e-7 & 1.424e-7 & 3.620e-9 & 2.197e-10 & 2.322e-13 \\ 1.305e-9 & 4.100e-9 & 8.731e-9 & 8.393e-9 & 2.197e-10 & 1.372e-11 & 1.359e-14 \\ 1.714e-12 & 4.858e-12 & 8.282e-12 & 9.001e-12 & 2.322e-13 & 1.359e-14 & 2.221e-17 \end{pmatrix}$$

$$\rho_{(\text{coral at 10 m})} = \begin{pmatrix} 1.000 & 0.975 & 0.807 & 0.852 & 0.847 & 0.797 & 0.823 \\ 0.975 & 1.000 & 0.903 & 0.927 & 0.916 & 0.865 & 0.805 \\ 0.807 & 0.903 & 1.000 & 0.955 & 0.928 & 0.888 & 0.662 \\ 0.852 & 0.927 & 0.955 & 1.000 & 0.979 & 0.937 & 0.790 \\ 0.847 & 0.916 & 0.928 & 0.979 & 1.000 & 0.986 & 0.819 \\ 0.797 & 0.865 & 0.888 & 0.937 & 0.986 & 1.000 & 0.778 \\ 0.823 & 0.805 & 0.662 & 0.790 & 0.819 & 0.778 & 1.000 \end{pmatrix}$$

These specific examples make it clear that

- For a given class, R_{rs} at one wavelength is highly correlated with R_{rs} at another wavelength, as expected.
- The covariance and correlation matrices are different for each class. These matrices depend not only on bottom type (sand vs coral) but also on bottom depth (and water IOPs, not explicitly shown here). In other words, *the wavelength covariances carry information about both bottom type and water depth and IOPs.*

14.4.3 Spectrum Matching versus Statistical Classification

One metric for comparing two spectra is the simple Euclidean metric, which measures the squared distance (in units of sr^{-2}) between an image spectrum \mathbf{I} and each \mathbf{R}_m in the database:

$$D_{\text{E}}^2(m) = \sum_{i=1}^K [I(\lambda_i) - R_m(\lambda_i)]^2 = [\mathbf{I} - \mathbf{R}_m]^T [\mathbf{I} - \mathbf{R}_m] \quad (14.8)$$

The spectrum \mathbf{R}_m giving the minimum distance $D_{\text{E}}^2(m)$ of all N database spectra determines the closest match to the image spectrum \mathbf{I} . *Note that this is not a statistical estimate in the sense that no probability model is involved. Note also that the image spectrum is being compared with every spectrum in the database, not just with pre-defined class mean spectra.*

In traditional thematic classification, an image spectrum \mathbf{I} is compared only with the mean spectrum and “size” for each class, as expressed by the class mean $\bar{\mathbf{R}}_m$ and covariance $\mathbf{\Sigma}_m$. Here “size” is used in the sense that the variances and covariances in $\mathbf{\Sigma}_m$ are larger when the spread of R_{rs} spectra is greater. Inspect, for example, the elements of $\mathbf{\Sigma}_m$ for the class of sand at 0.01 m compared to sand at 10 m, for which the spectra are all much closer together (especially at blue and red wavelengths) and thus have smaller covariances. The class covariance matrix defines the size of the “swarm of points” surrounding the centroid (mean class spectrum) representing the class in K -dimensional R_{rs} space. The image spectrum is assigned to a particular class according to a statistical model (often based on the assumption of a multivariate normal distribution of the swarm of points) that determines the probability that the image spectrum belongs to a particular the swarm of points defining a given class. The class spectra (K -dimensional swarms of points) generally

overlap, so that an unambiguous, non-probabilistic association of \mathbf{I} with a given class is not possible.

In maximum likelihood estimation (MLE; see Richards and Jia (1996) for an excellent discussion of this whole business), the distance metric is

$$D_{\text{MLE}}^2(m) = \ln |\boldsymbol{\Sigma}_m| + [\mathbf{I} - \bar{\mathbf{R}}_m]^T \boldsymbol{\Sigma}_m^{-1} [\mathbf{I} - \bar{\mathbf{R}}_m] , \quad (14.9)$$

where $|\boldsymbol{\Sigma}_m|$ denotes the determinant of $\boldsymbol{\Sigma}_m$ and $\boldsymbol{\Sigma}_m^{-1}$ denotes the inverse. $|\boldsymbol{\Sigma}_m|$ and $\boldsymbol{\Sigma}_m^{-1}$ are of course pre-computed for each class before doing the spectrum matching. The image spectrum \mathbf{I} is assigned to the class m having the smallest value of $D_{\text{MLE}}^2(m)$. *Note that now the image spectrum is compared only with the class mean spectra $\bar{\mathbf{R}}_m$.* The assignment of the image spectrum to a particular class is based on its closeness to the class mean and the spread of the “swarm of points” surrounding the mean, as described by the covariance matrix. This metric involves matrix multiplication, which is computationally expensive, but the number of classes is generally small, so in practice this may not be a problem.

It is often said that the incorporation of $\boldsymbol{\Sigma}_m$ into the distance metric “removes the effect of correlations between wavelengths.” This interpretation of the effect of $\boldsymbol{\Sigma}_m$ relates to the fact that covariance matrices are the foundation of principle component analysis (PCA; see Preisendorfer (1988)). In PCA the original independent, physical variables (here, the wavelengths) are transformed to obtain a new set of (generally unphysical) independent variables for which the data are uncorrelated. This transformation can be viewed as a rotation of the axes of the original, physical data space (here the wavelength axes used for plots in K -dimensional space) to generate new (generally unphysical) axes for which the data are uncorrelated.

If the class covariances are equal (or assumed to be equal), then $\ln |\boldsymbol{\Sigma}_m|$ is the same for each class and can be ignored. The MLE metric then reduces to the Mahalanobis distance metric,

$$D_{\text{M}}^2(m) = [\mathbf{I} - \bar{\mathbf{R}}_m]^T \boldsymbol{\Sigma}^{-1} [\mathbf{I} - \bar{\mathbf{R}}_m] , \quad (14.10)$$

where $\boldsymbol{\Sigma}$ is the common value of $\boldsymbol{\Sigma}_m$. The image spectrum \mathbf{I} is then assigned to the class m having the smallest value of $D_{\text{M}}^2(m)$.

We have seen by the specific examples of Fig. 14.9 and Table 1 that the covariance matrices are different for different classes of the sort that are relevant for ocean-bottom remote sensing. Indeed, Table 1 shows that the elements of the $\boldsymbol{\Sigma}_m$ can change by orders of magnitude as a function of water depth. This inequality of the $\boldsymbol{\Sigma}_m$ for different classes precludes use of the Mahalanobis metric for classes as defined here. For the retrievals needed for shallow-water mapping of bottom type, MLE (or something else) would have to be used with a different covariance matrix for each class.

However, it is not at all clear how meaningful classes should be defined for simultaneous retrievals of bottom type, water column IOPs, and bottom depth. Should one class be “sand spectra at 5.25 m depth with a particular set of water absorption, scattering, and backscatter spectra,” and another class be “sand spectra at 5.25 m depth with the same absorption and scattering spectra but a different backscatter fraction,” and another class be “sand spectra at 5.50 m with the first set of IOPs,” and then another class be “sea grass spectra at 7.50 m with yet another set of IOPs,” and so on? If so, then the number of classes quickly becomes as large as the number of depths, IOP sets, and pre-chosen classes of bottom type (sand, coral, sea grass, etc.). A database generated as previously described easily could have hundreds or thousands of classes (a database often has 50-100

bottom depths, several dozen to several hundred sets of IOPs, and more than 100 bottom reflectance spectra). With such a large number of classes, the validity of doing traditional thematic mapping becomes uncertain, not to mention the additional computational costs involved with the matrix multiplications.

Clearly spectrum matching for shallow-water applications addresses a much more complicated problem than classic terrestrial thematic mapping, which corresponds to retrieval of bottom type if there were no water present, i.e. no simultaneous retrieval of depth and IOPs. Because of the greater complexity of the oceanographic retrieval problem, and because of the difficulty in defining meaningful classes, shallow-water spectrum matching does not use statistical classification techniques such as MLE. The spectrum matching approach of [Mobley et al. \(2005\)](#) does not compare an image spectrum to a class mean spectrum. In that technique, an image spectrum is compared to every spectrum in a database to find the closest match by the chosen (Euclidean or some other) metric, which is appropriate in this case. In a manner of speaking, each database R_{rs} spectrum is a separate class corresponding to a particular depth, bottom reflectance spectrum, and set of IOPs. In such a situation (only one member in each class) the covariance matrix is undefined.

Moreover, for the present problem it is not even desirable to remove the effects of wavelength correlations, as can be done with the MLE or Mahalanobis metrics, because the wavelength correlations carry information that is critical to separating depth and IOPs effects from bottom type effects.

The spectrum-matching approach of [Mobley et al. \(2005\)](#) for shallow-water benthic mapping therefore avoids defining predetermined classes and finds the closest match from the entire database. This gives the highest possible resolution (in depth, bottom type, and water IOPs) of retrievals. This approach retrieves a particular bottom reflectance spectrum (which represents a particular bottom type), not just a generic bottom type such as sand or coral. If the user later wishes to group the particular spectra for the retrieved bottom types into broader classes such as corals versus sediments, or to group the retrieved IOPs into low, medium, and high absorption bins, for example, then that is easily done from the full-resolution retrieval.

Atmospheric Correction

[The development of this chapter was a collaborative effort by Curtis Mobley, Jeremy Werdell, Bryan Franz, Ziauddin Ahmad, Sean Bailey, Howard Gordon and David Antoine.]

The topic of atmospheric correction perhaps belongs in the preceding Remote Sensing chapter. However, this is such an important and complicated topic that it warrants a chapter of its own.

The initial sections of this chapter discuss the generic atmospheric correction problem. The first section shows via numerical simulations how the atmosphere can affect top-of-the-atmosphere (TOA) radiances. The following section then formulates the atmospheric correction problem in terms of the various contributions to the TOA radiance measured by a satellite-borne sensor. Those contributions come from solar radiance scattered by atmospheric molecules and aerosols, Sun and sky radiance reflected by the sea surface (either by the water surface itself or by foam from whitecaps), and finally from water-leaving radiance. The details of how normalized water-leaving radiances and normalized reflectances are defined and how they should be interpreted are then given. Sections on vicarious calibration and the computation of diffuse transmittances then follow.

Several sections then describe one by one the specific algorithms used (as of the date of this writing) by the NASA Ocean Biology Processing Group (OBPG) to effect the various steps of the atmospheric correction process, namely the corrections for absorption and scattering by gases and aerosols, Sun and sky reflectance by the sea surface, whitecap reflectance, and finally corrections for sensor out-of-band response and polarization effects. The end result is an estimate of the “exact normalized water-leaving radiance,” or its equivalent reflectance, which carries the information about the water-column itself.

Once obtained, the normalized reflectance is the input to algorithms for retrieval of various quantities of scientific interest. These ocean-color products include—among others—the Chlorophyll *a* concentration, the water-column diffuse attenuation for downwelling plane irradiance at 490 nm (K_d490 , which is a proxy for water transparency), water-column absorption and backscatter coefficients, and particulate organic and inorganic carbon. The algorithms for retrieval of specific products, given the normalized reflectance, are given in a series of reports found at [NASA Algorithms](#). Those retrieval algorithms are under continual refinement are not discussed here.

There are many other sources with additional information about atmospheric correction. The [NASA ocean color web site](#) contains a wealth of information about how NASA collects, processes, calibrates, validates, archives and distributes ocean color data from a variety of satellite sensors. That web site has many pages with links to various technical memos and other information about ocean color, and many of the data files underlying the atmospheric correction process can be downloaded there.

There are also many non-NASA sources of information on atmospheric correction. The [Maine In-situ Sound and Color Lab](#) website links to PowerPoint presentations of lectures given at the University of Maine summer courses, and to videos of the 2015 and 2017 lectures. The International Ocean Color Coordinating Group (IOCCG) has hosted summer lecture series during which the lectures were videoed. The IOCCG lectures delivered by Menghua Wang in 2012 and 2014 cover much of the material presented here; they can be found at [IOCCG 2012 Lectures](#) and [IOCCG 2014 Lectures](#). [IOCCG Report 10](#) compares the SeaWiFS-MODIS versus MERIS versus OCTS-GLI versus POLDER atmospheric correction algorithms, but assumes that the reader is already familiar with the general process.

The OBPG-specific material of this chapter is available as a NASA Technical Memorandum, [Atmospheric Correction for Satellite Ocean Color Radiometry](#) (Mobley et al., 2016).

The chapter finishes with two sections on methods for atmospheric correction that are sometimes used for airborne imagery or in situations for which the OBPG algorithms are not applicable.

15.1 The Atmospheric Correction Problem

An instrument viewing the ocean from a satellite or aircraft measures upwelling radiances that include contributions by the atmosphere, the water surface, and the water column. The atmospheric contribution L_a comes from solar radiance that is scattered one or more times by atmospheric gases and aerosols into the direction of the sensor. The surface-reflected radiance (Sun and sky glint) L_r is downwelling solar radiance that is reflected toward the sensor by the water surface. The water-leaving radiance L_w comes from light that is transmitted through the surface into the ocean, L_t , where it is changed by the absorbing and scattering components in the water, and is then scattered into an upward direction, L_u , and eventually leaves the sea surface in the sensor direction. Figure 15.1 shows this conceptually.

Radiance reflected by the sea surface contains information about the wave state of the surface, which may be of interest in itself or which, for example, may be useful for detection of surface oil slicks. Radiance scattered by the atmosphere along the path between the sea surface and the sensor contains information about atmospheric aerosols and other atmospheric parameters. However, only the water-leaving radiance carries information about the water column (and, in optically shallow water, about the sea bottom). A sensor looking downward measures the total radiance $L_u = L_a + L_r + L_w$ and cannot separate the various contributions to the total. *Atmospheric correction*¹ refers to the process of

¹The term “atmospheric correction” offends some in the atmospheric optics community who rightly say that the atmosphere does not need correcting because it has not done anything wrong. They prefer to call the process “atmospheric compensation.” They make a valid point, but “atmospheric correction” is commonly used by those doing ocean color remote sensing, so that is what is used here.

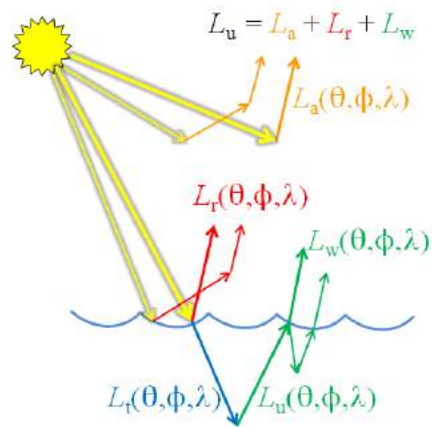


Figure 15.1: Contributions to the total upwelling radiance above the sea surface, L_u . Yellow arrows are the Sun’s unscattered beam; orange arrows are atmospheric path radiance L_a ; red is surface-reflected radiance L_r ; green is water-leaving radiance L_w . Thick arrows represent single-scattering contributions; thin arrows illustrate multiple scattering contributions.

removing the contributions by surface glint and atmospheric scattering from the measured total to obtain the water-leaving radiance.

This section uses numerical atmosphere-ocean radiative transfer simulations to show examples of the nature and magnitude of the atmospheric correction problem. These figures were generated by a coupled MODTRAN-HydroLight radiative transfer code. The MODTRAN atmospheric code (Acharya et al., 1998) was used to propagate Sunlight from the top of the atmosphere to the sea surface. The atmospheric radiance incident onto the sea surface was then used to initialize the HydroLight in-water code (Section 10.6). HydroLight transmitted the radiance through the sea surface, computed its interaction with the water constituents, and eventually returned the water-leaving radiance back to MODTRAN. MODTRAN then propagated the water-leaving radiance, plus the glint radiance and atmospheric path radiance, to the sensor. Both of these codes are for unpolarized radiance calculations. The HydroLight-MODTRAN code can separate the Rayleigh versus [aerosol + aerosol-Rayleigh] contributions, but cannot separate the pure aerosol from the aerosol-Rayleigh contributions. Likewise, it does not normally separate Sun glint and sky glint contributions, although that separation can be effected with some extra effort. The partitioning of atmospheric radiance contributions in the model simulations is thus not exactly the same as is done operationally, but the model simulations can still illustrate the various contributions to the TOA radiance.

Figure 15.2 shows examples of at-sensor total radiances L_u for different sensor altitudes. The inputs to MODTRAN were for

- cloudless mid-latitude summer atmosphere
- marine aerosols
- relative humidity of 76% at sea level
- solar zenith angle of 50 deg

- surface wind speed of 6 m s^{-1}
- nadir-viewing sensor

These atmospheric conditions give excellent remote-sensing conditions with a horizontal visibility of 63 km at sea level. The HydroLight inputs were for

- homogeneous water
- Case 1 water with a chlorophyll concentration of 1 mg m^{-3}
- infinitely deep water

The runs were from 300 to 1000 nm with 10 nm bandwidths. The Sun and viewing geometry gives almost no Sun glint radiance. However, there is always surface-reflected sky radiance, which shows up in the surface (zero altitude) spectrum (black curve) as non-zero L_u radiance beyond 750 nm, where the water-leaving radiance is very close to zero. The differences in these curves are due solely to atmospheric contributions along the different path lengths from the sea surface to the sensor because all other conditions were the same for each curve. This figure shows that there is a noticeable atmospheric contribution to the total radiance even for sensors just a few hundred meters above the surface in very clear atmospheres. Airborne sensors typically fly at altitudes from 3,000 to 10,000 m, and even at lower altitudes the atmospheric contribution is greater than the water-leaving radiance. At 30,000 m, a sensor is effectively above the top of the atmosphere (TOA) and atmospheric path radiance is typically 90-95% of the total.

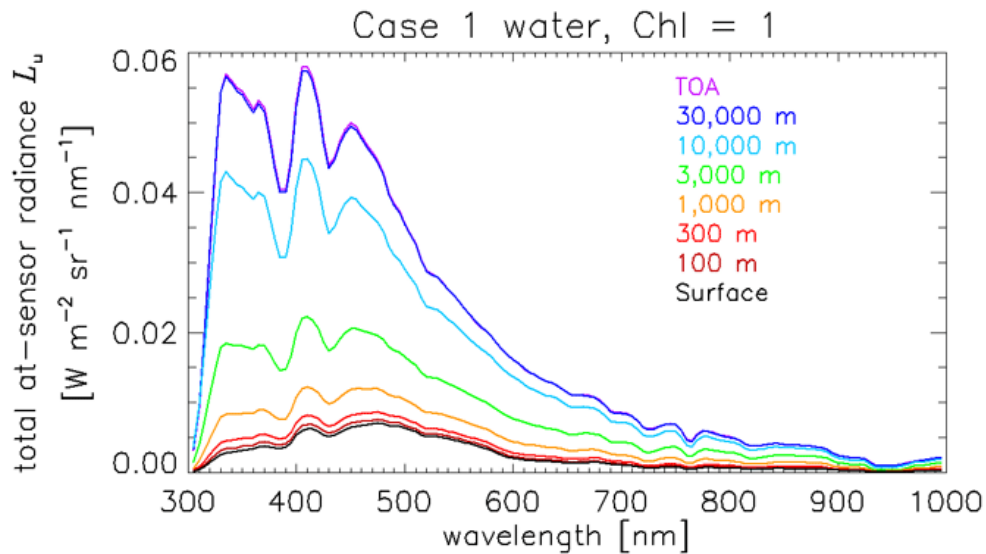


Figure 15.2: Example at-sensor radiances L_u for different sensor altitudes. The water-leaving radiance and surface-reflected radiance (not shown) are the same in all cases.

Figure 15.3 shows the Fig. 15.2 at-sensor radiance at 3,000 m partitioned into the contributions by water-leaving, surface-reflected, and atmospheric path radiances; the individual contributions can be separated in the numerical models. This altitude is typical for aircraft acquiring high spatial resolution hyperspectral imagery.

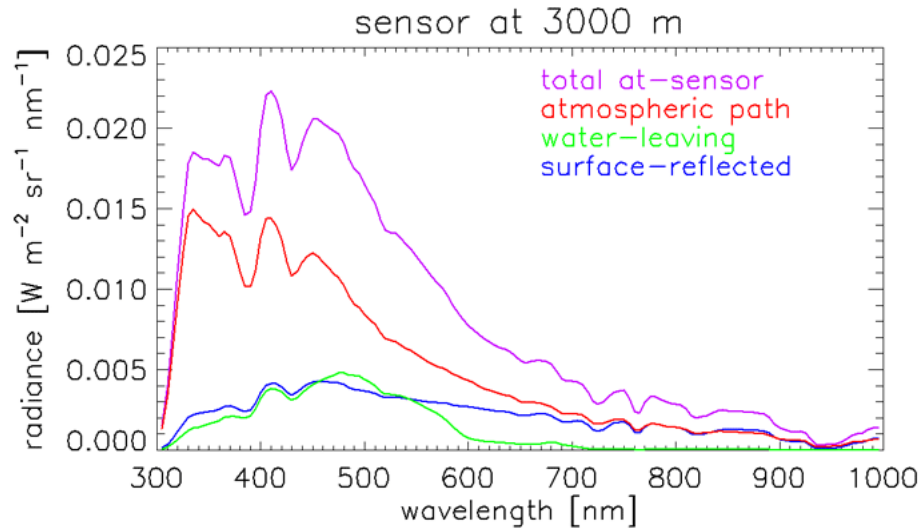


Figure 15.3: The total radiance curve of Fig. 15.2 at 3000 m sensor altitude partitioned into the contributions by water-leaving radiance, surface reflectance, and atmospheric path radiance.

The atmospheric correction problem becomes even more intimidating when effects of Sun and sensor viewing directions, atmospheric conditions, and surface wave state are considered. The following figures show a few examples of simulations performed in support of the design of the proposed NASA *HypIRI* (Hyperspectral InfraRed Imager) mission. That sensor will measure radiances at the top of the atmosphere at 10 nm resolution from 380 to 2500 nm with 60 m ground resolution. The *HypIRI* orbital characteristics were used to obtain the Sun zenith and azimuthal angles at the time the sensor would fly over a point at (latitude, longitude) = (28.75 N, 158.00 W) on June 21. This point is north of the island of Oahu in Hawaii and is known as Station ALOHA (A Long-term Oligotrophic Habitat Assessment). Figure 15.4 shows the relevant angles needed for the simulation.

The simulations shown below all have a chlorophyll value of 0.05 mg m^{-3} characteristic of the very clear Case 1 ocean water near the Hawaiian Islands. The atmospheric conditions are for either a very clear tropical atmosphere (horizontal visibility of 100 km at sea level) or one with considerable haze (horizontal visibility of 10 km). The wind speed was 0 (a level sea surface) or 10 m s^{-1} . The sensor viewing direction was 30 deg east of nadir, nadir, or 30 deg west of nadir. The Sun zenith angle was 18 deg, which corresponds to the time of the satellite overpass on 21 June at Station Aloha, near Hawaii.

A simulation was done for the following environmental conditions:

- The water was homogeneous and infinitely deep.
- The water IOPs were simulated using a chlorophyll concentration of $Chl = 0.05 \text{ mg m}^{-3}$ in the “new Case 1” IOP model (Section 8.9) in *HydroLight*.
- The Sun zenith angle was $\theta_{\text{Sun}} = 17.99 \text{ deg}$ and the Sun’s azimuthal angle was east of the nadir point at 84.34 deg from true north.
- The off-nadir viewing angle was $\theta_v = 30 \text{ deg}$, $\phi_v = 281.12 \text{ deg}$, which is at right angles to the satellite’s orbital direction and looking to the west side of the orbit, away from

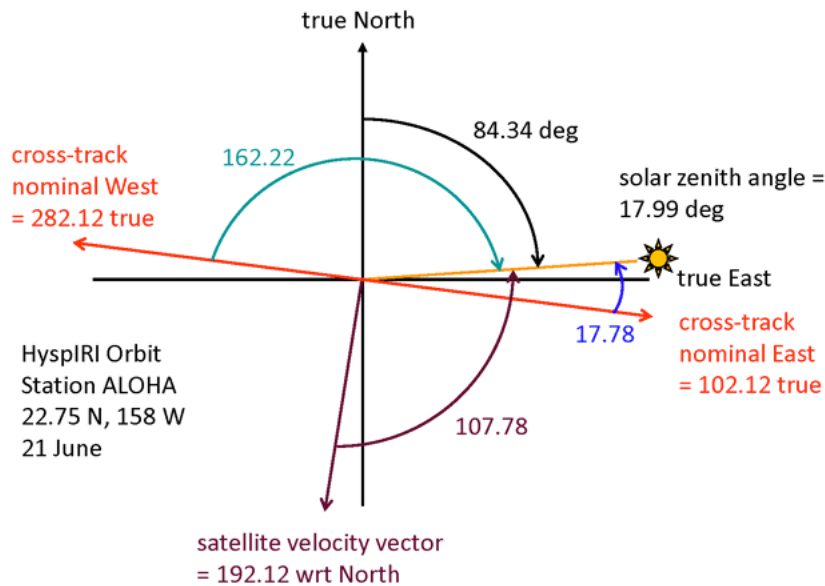


Figure 15.4: Sun and viewing geometry for a mid-day HypsIRI pass over Station ALOHA on 21 June.

the Sun’s direction.

- The atmospheric conditions (temperature profile, water vapor, ozone, etc.) were typical of a tropical marine atmosphere (defined via MODTRAN’s “Tropical Atmosphere” option). The sky conditions were clear. The aerosols were for an open-ocean marine atmosphere.
- The wind speed was 10 m s^{-1} .
- The wavelength resolution was 10 nm from 350 to 1500 nm .

Figure 15.5 shows various radiances and irradiances obtained from this simulation. The solid curves are values at the TOA, and the dotted curves are the corresponding quantities just above the sea surface. The E_d TOA curve (the solid purple line) is the extraterrestrial solar irradiance (averaged over 10 nm bands) on a surface parallel to the mean sea surface. The dips in the curve below 700 nm are due to absorption by various elements in the Sun’s photosphere; these are Fraunhofer lines averaged over the 10 nm bands of this simulation. Above 700 nm , the Sun’s irradiance is close to a blackbody spectrum (Section 2.4.2). The purple dotted line shows how much of the TOA solar irradiance reaches the sea surface. There are large dips in the TOA irradiance that reached the sea surface in the regions around 940 and 1130 ; these are due to absorption by water vapor, as are the smaller dips near 720 and 820 nm . The large opaque region between 1350 and 1450 nm is due to water vapor and carbon dioxide. The dip at 760 nm is due to absorption by atmospheric oxygen. These absorption features of the Earth’s atmosphere will affect any radiation passing through the atmosphere.

The solid blue line shows the TOA radiance L_t that would be measured by a satellite looking in the direction 30 deg West of the nadir point. The orange curve shows how much

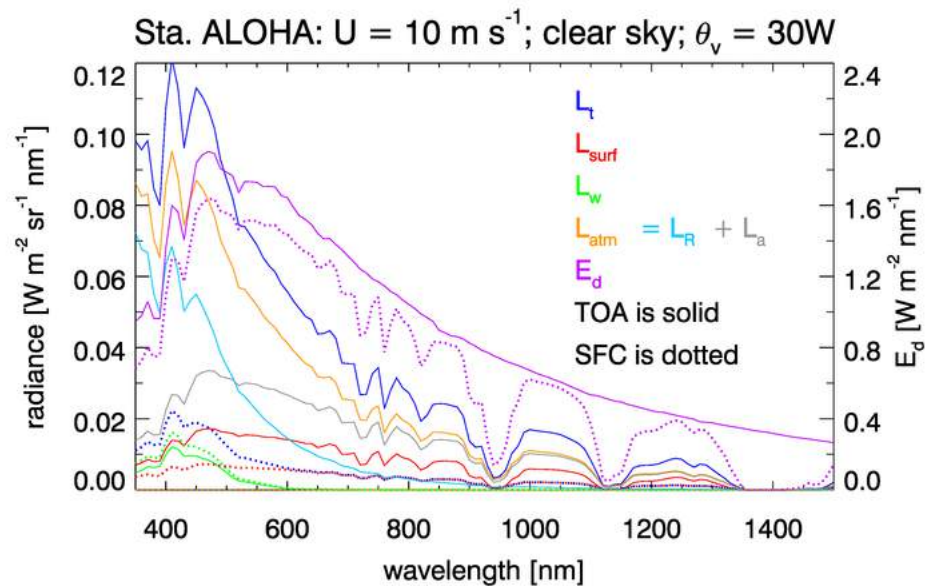


Figure 15.5: Example radiances contributing to the TOA radiance. Solid lines are radiances at the TOA; dotted lines are at the sea surface (SFC). The geometric, atmospheric, and water conditions are described in the text.

of the total is atmospheric path radiance, L_{atm} . The aqua and gray curves respectively show how much of the path radiance is due to Rayleigh scattering by atmospheric gases and by aerosols (including aerosol-gas interactions). The green curves in Fig. 15.5 show that the water-leaving radiance at the TOA (the solid curve) is less than the water-leaving radiance just above the sea surface (the dotted curve). This makes intuitive sense, because part of the water-leaving radiance would be lost to atmospheric absorption or scattering into other directions before that radiance reaches the TOA.

The red curves show the total radiances due to surface reflectance, i.e., the sum of the background sky reflectance and the direct Sun glint. However, the red curves show that the surface-reflectance contribution is *greater* at the TOA than at the surface. This seems counterintuitive and requires explanation. Figure 15.6 gives a more detailed view of the contributions to the TOA radiance. In this figure, the arrow labeled L_g represents Sun glint due to the occasional wave facet that is tilted in just the right direction to create glint that is seen by the sensor. The arrow labeled L_{gs} represents the very bright glint in the Sun's specular direction; the sensor is looking in the direction away from the Sun's azimuthal direction in order to avoid viewing this specular glint. However, the specular glint gives a strong reflected radiance, some of which is being scattered by the atmosphere into the sensor viewing direction; this is illustrated by the $L_{\text{gs}2}$ arrow in Fig. 15.6. The surface contribution in Fig. 15.5 is the sum of the L_{sky} , L_g , and $L_{\text{gs}2}$ contributions. If the ocean is viewed from just above the sea surface (the red dotted line in Fig. 15.5), the surface-reflected radiance comes only from reflected sky radiance L_{sky} and a small amount of direct Sun glint L_g from wave facets that are tilted in just the right way to reflect the Sun's direct beam into the direction of the sensor. (This direct Sun glint L_g is minimal because

of the choice of viewing direction.) These surface-reflected sky and Sun radiances decrease between the surface and the TOA, just as does the water-leaving radiance. However, the total TOA radiance that arises from sea surface reflectance, as partitioned in the simulation, comes from the sum of the surface reflectance in the viewing direction (decreasing with altitude) and the contribution by atmospheric scattering of specular Sun glint into the viewing direction (L_{gs2} , increasing with altitude). The specular radiance L_{gs2} is large in magnitude, so the atmospheric scattering of this radiance into the viewing direction can be significant. If the present simulation is run with a level sea surface, for which there is no L_g Sun glint into the sensor direction, this behavior is still present and, indeed, is even somewhat greater in magnitude.

To fully isolate the effect of specular Sun glint being atmospherically scattered into the sensor, a run was made in which any ray from the Sun's unscattered beam that was reflected by the sea surface was killed. That is, all Sun glint was forced to be zero. Only rays from the background sky that were reflected by the sea surface were allowed to contribute to the TOA surface-reflected radiance. In that case, the surface-reflected radiance behaves the same as the water-leaving radiance: less sky glint reaches the TOA than leaves the sea surface.

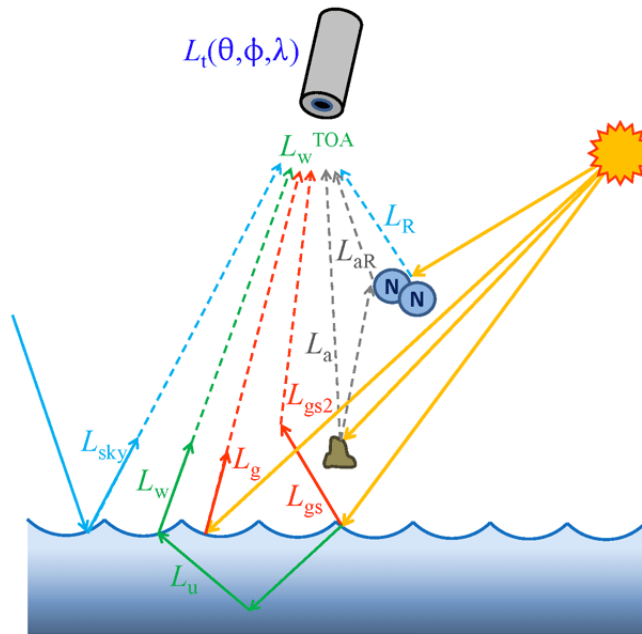


Figure 15.6: Qualitative illustration of the various processes contributing to the total TOA radiance. The notation corresponds to Table 15.1 in the next section. The blue N-N represents a nitrogen (N_2) molecule, or any other atmospheric gas molecule; the brown blob represents an aerosol particle.

Figure 15.7 shows the fractional contributions by these various processes to the total TOA radiance L_t . For this particular case, the water-leaving radiance is at most 10% of the total TOA radiance. The atmospheric path radiance contributes 70 to 90% of L_t , with the rest coming from surface glint. Below 500 nm, Rayleigh scattering by atmospheric gases

is the largest contributor to the TOA radiance. This is also true in the band from 1350 to 1400 nm, where the atmosphere is essentially opaque, because the aerosols are located mostly near the sea surface. Aerosols are the greatest contributor between 500 and 1350 nm.

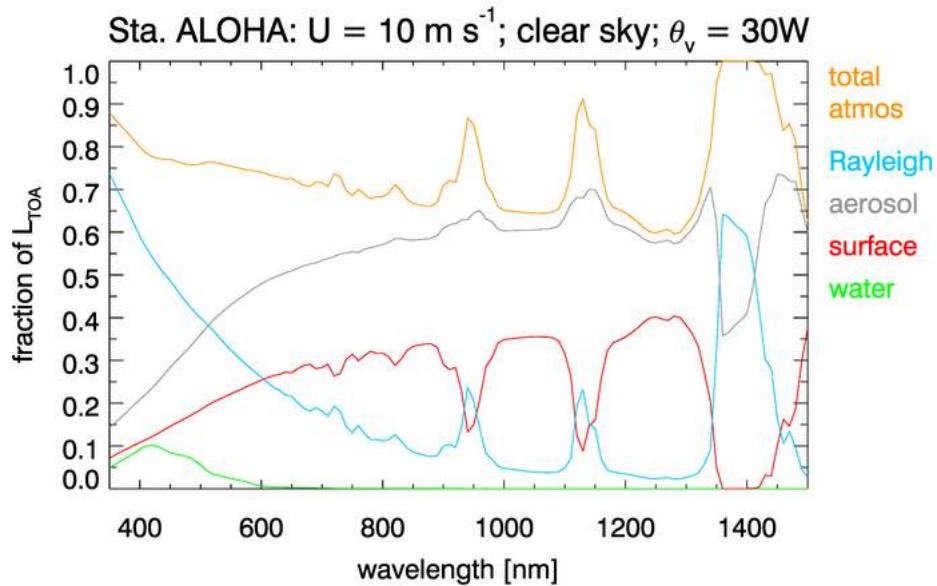


Figure 15.7: The fraction of L_t due to various processes, for the particular simulation of Fig. 15.5.

Figure 15.8 shows the L_t of Fig. 15.5, plus the corresponding L_t radiances for wind speeds of 0 and 10 m s^{-1} , clear and hazy atmospheric conditions, and viewing directions of 0 (nadir) and 30 deg East or West of the flight direction. Each of these 12 much-different TOA radiances has essentially the same water-leaving radiance L_w , which is shown in green.

The atmospheric correction problem can be visually summarized as follows: Given any of these TOA radiance spectra seen in Fig. 15.7 and the geometry (Sun location and viewing direction), recover the water-leaving radiance spectrum L_w . This is clearly a difficult problem because the needed atmospheric conditions (aerosol type and concentration in particular) are not obtained as part of the L_t measurement.

Atmospheric correction is the central problem of ocean color remote sensing. However, the problem is not intractable—if it were, we could not do ocean color remote sensing! There are many inversion algorithms that can retrieve ocean properties such as chlorophyll, mineral, or CDOM concentrations, or bottom depth and reflectance, IF an accurate water-leaving radiance (or, usually, its equivalent reflectance, to be developed in Section 15.3) is available. The hard part of the game is to get from radiance measured at the TOA to water-leaving radiance at the sea surface.

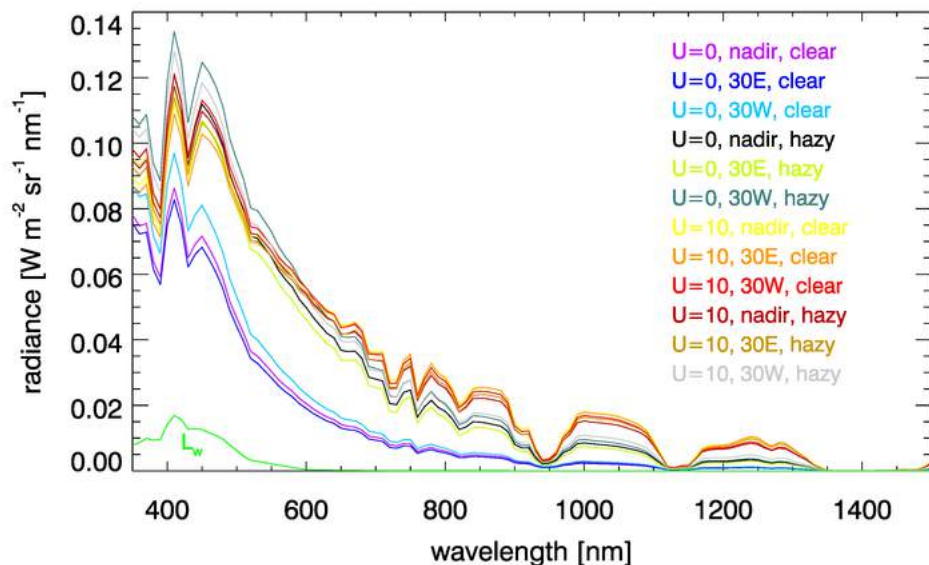


Figure 15.8: Upper curves: 12 TOA radiances for various environmental and viewing conditions as labeled. Each TOA radiance has the same water-leaving radiance L_w , which is shown by the green curve.

15.2 Problem Formulation

The introduction to atmospheric correction given in Section 15.1 illustrated the nature of the problem. It is now time to define the problem quantitatively.

The total radiance L_t measured by a satellite-borne sensor at the top of the atmosphere (TOA) comes from contributions by atmospheric scattering, L_{atm} ; Sun and sky radiance reflected back upward by the sea-surface and reaching the TOA, $L_{\text{surf}}^{\text{TOA}}$; and water-leaving radiance that reaches the TOA, L_w^{TOA} :

$$L_t = L_{\text{atm}} + L_{\text{surf}}^{\text{TOA}} + L_w^{\text{TOA}}. \quad (15.1)$$

For brevity, the viewing direction (θ_v, ϕ_v) and wavelength λ are not shown. Expanding this equation into further levels of detail requires the definition of many different radiances, and precise notation is needed to minimize confusion. The atmospheric contribution L_{atm} is always considered to be at the TOA. However, the surface-reflected radiance and water-leaving radiance can be formulated either at the sea surface or at the TOA. For these radiances, a superscript TOA will be used to specify the TOA value. Thus L_w will denote the water-leaving radiance just above the sea surface, and L_w^{TOA} will denote how much of L_w reaches the TOA. Table 15.1 summarizes the various radiances introduced in this chapter and used throughout this report.

The atmospheric contribution in Eq. (15.1), usually called atmospheric path radiance, comes from scattering by atmospheric gases and aerosols, including multiple scattering between gases and aerosols. The path radiance that comes solely from scattering by atmospheric gas molecules is usually called the Rayleigh radiance, L_R , because scattering by

Table 15.1: Radiance notation. The notation here corresponds to that seen in Fig. 15.6. Spectral radiance L has SI units of $\text{W m}^{-2} \text{nm}^{-1} \text{sr}^{-1}$; in practice $\text{mW cm}^{-2} \mu\text{m}^{-1} \text{sr}^{-1}$ is often used.

Symbol	Definition
L_t	total upwelling radiance at the top of the atmosphere
L_{atm}	total contribution of atmospheric scattering to the TOA radiance
$L_{\text{surf}}^{\text{TOA}}$	total contribution of surface-reflected radiance to the TOA radiance
L_R	total Rayleigh radiance at the TOA
L_r	“standardized” Rayleigh radiance at the TOA
L_a	TOA radiance due to scattering by aerosols only
L_{aR}	TOA radiance due to aerosol-molecule scattering
L_A	$L_a + L_{\text{aR}}$; total aerosol radiance at the TOA
L_w	water-leaving radiance just above the sea surface
L_w^{TOA}	the part of the water-leaving radiance L_w that reaches the TOA
L_g	direct Sun glint radiance just above the sea surface
L_g^{TOA}	the part of the direct Sun glint radiance L_g that reaches the TOA
L_{sky}	surface-reflected background sky radiance at the sea surface
$L_{\text{sky}}^{\text{TOA}}$	the part of the surface-reflected background sky radiance L_{sky} that reaches the TOA
L_{wc}	radiance due to whitecaps and foam just above the sea surface
$L_{\text{wc}}^{\text{TOA}}$	the part of the whitecap radiance L_{wc} that reaches the TOA
L_u	upwelling underwater radiance just beneath the sea surface

molecules is well described by the Rayleigh mathematical model of scattering by particles that are much smaller than the wavelength of light. In the absence of any aerosols, the atmospheric path radiance would equal the Rayleigh radiance. Let L_a denote the aerosol contribution, which is the path radiance that would occur if the atmosphere consisted only of aerosol particles. Let L_{aR} denote the contribution resulting from multiple scattering between aerosols and gases. The total surface reflectance can be separated into a contribution due to direct Sun glint from the water surface, L_g^{TOA} ; by background sky radiance reflected by the water surface, $L_{\text{sky}}^{\text{TOA}}$; and by Sun and sky radiance reflected by whitecaps and foam, $L_{\text{wc}}^{\text{TOA}}$. Thus Eq. (15.1) can be further partitioned into

$$L_t = L_R + [L_a + L_{\text{aR}}] + L_g^{\text{TOA}} + L_{\text{sky}}^{\text{TOA}} + L_{\text{wc}}^{\text{TOA}} + L_w^{\text{TOA}}. \quad (15.2)$$

In practice, the aerosol and aerosol-gas contributions are usually grouped together and

treated as one contribution, sometimes denoted $L_A = L_a + L_{aR}$ and often called just the aerosol contribution. The sky reflectance term is accounted for as part of the Rayleigh correction, which incorporates reflectance by the sea surface. For some sensors that were specifically optimized for ocean color (e.g., CZCS and SeaWiFS), the strongest part of the Sun glint (the Sun’s glitter pattern) is avoided by pointing the sensor in a direction away from the Sun so that almost no direct glint is present in the image. However, there is still a correction for residual amounts of Sun glint. Figure 15.6 illustrates these contributions to the TOA radiance.

Most papers (e.g., Wang and Bailey, 2001; Wang, 2002) rewrite Eq. (15.2) as

$$L_t = L_R + [L_a + L_{Ra}] + TL_g + tL_{wc} + tL_w, \quad (15.3)$$

or something very similar. Now, however, L_g , L_{wc} , and L_w are all measured at sea level. T is the direct transmittance between the sea surface and the TOA along the viewing direction, and t is diffuse transmittance in the viewing direction. These transmittances are discussed in Section 15.4.

Yet a third formulation can be found in the literature (e.g., Franz et al., 2007, Eq. 1):

$$L_t = \left(L_r + [L_a + L_{ra}] + t_{dv}L_{wc} + t_{dv}L_w \right) t_{gv}t_{gs}f_p. \quad (15.4)$$

Here t_{dv} is the diffuse transmittance along the viewing path of the sensor. t_{gv} is the transmittance by atmospheric gases in the viewing direction, and t_{gs} is the transmittance by atmospheric gases in the Sun’s direction; these transmittances are usually called gaseous transmittances. f_p is a known instrument polarization-correction factor. Comparison of Eqs. (15.3) and (15.4) shows, for example, that

$$L_R = L_r t_{gv} t_{gs} f_p.$$

Thus the total TOA Rayleigh contribution L_R has been factored into a product of terms involving a Rayleigh term times gaseous transmittances and a polarization-correction factor. The difference between Eqs. (15.3) and (15.4) is primarily a matter of simplification for presentation purposes. The f_p term enters because MODIS has large polarization sensitivity and this requires correction. Earlier papers by Gordon and Wang often ignored the gaseous transmission terms because they were only considering ozone, which could be “taken off the top,” so to speak, with the remaining problem being effectively formulated below the ozone layer. The L_r term is computed using a standard atmosphere and only non-absorbing gases N_2 and O_2 . This allows “standard” Rayleigh radiances L_r to be computed as a function of Sun and viewing geometry. The gaseous transmittances are computed by use of absorption coefficients, computed path lengths, and gas concentrations for the various gases. The f_p term is computed for each image pixel as a function of atmosphere and surface polarization states (modeled Rayleigh and glint Stokes vectors) and the sensor-specific polarization sensitivity with viewing direction.

All of Eqs. (15.2), (15.3), and (15.4) can be found in the literature. They all give the same TOA total radiance L_t . Which form is used in a particular instance is determined by convenience. Forms (15.2), (15.3) are often convenient for discussions of theory, whereas form (15.4) is convenient for operational atmospheric correction algorithms. Table 15.2 summarizes for later reference the various transmittances seen in the literature and in later sections of this chapter.

Table 15.2: Transmittances used in Eqs. (15.3) and (15.4).

Symbol	Definition
T	the direct transmittance between the sea surface and the TOA along the viewing direction
t_{dv}	the diffuse transmittance along the viewing path of the sensor
t_{gv}	the diffuse transmittance by atmospheric gases in the viewing direction
t_{gs}	the diffuse transmittance by atmospheric gases in the solar direction
f_p	the sensor-specific polarization-correction factor

The goal of atmospheric correction is to convert a measured top-of-the-atmosphere radiance L_t into the corresponding sea-level water-leaving radiance L_w . Since only L_t is measured, this requires estimation of the various atmospheric and surface-reflectance terms seen in Eqs. (15.3) or (15.4) so that they can be subtracted from L_t in order to arrive at L_w . How this is done is discussed in the following sections.

15.3 Normalized Reflectances

Ocean-color remote sensing algorithms usually work with remote-sensing reflectances or normalized water-leaving reflectances. The calculation and interpretation of those quantities are discussed in detail in this section.

As seen in Section 4.2.2, the ratio of water-leaving radiance L_w to incident sky irradiance E_d is an apparent optical property (AOP) that has only weak dependence on external parameters such as solar zenith angle and sky conditions, but which is strongly correlated with water-column inherent optical properties (IOPs). However, the remote-sensing reflectance $R_{rs} \triangleq L_w/E_d$ still does depend somewhat on the atmospheric and other conditions (recall the discussion of Section 4.2.3) at the time of measurement and thus, strictly speaking, is tied to the particular time and location of the observation.

It would be desirable to have an AOP that completely removes the effects of solar zenith angle, viewing direction, atmospheric conditions, and sea state, while retaining a strong dependence on the water IOPs. It would then be possible to compare this AOP for measurements made at different times and/or locations, and thereby extract information about the differences in the water columns for the different measurements. Even for measurements made at the same time and location, normalization to a common set of conditions is needed, e.g., when comparing in situ measurements having different viewing directions. Such an AOP is obtained via the concept of the normalized water-leaving reflectance.

15.3.1 Normalized Radiances and Reflectances

Let $L_u(z, \theta_s, \theta_v, \phi)$ be the in-water, upwelling radiance at depth z , for a Sun zenith angle of θ_s and a viewing direction of θ_v, ϕ . Polar viewing direction $\theta_v = 0$ indicates a direction looking at the nadir, detecting radiance traveling toward the zenith. The azimuthal

angle ϕ is measured relative to the Sun’s azimuthal direction. $L_w(\theta_s, \theta_v, \phi)$ denotes the corresponding water-leaving radiance, which is measured just above the sea surface. These radiances depend strongly on wavelength, which is not shown. In practice, L_u is measured by instruments in the water. L_w can be obtained by atmospheric correction of a radiance measured at the top of the atmosphere, from an above-surface measurement at sea level after correction for surface reflectance, or from an in-water measurement of L_u extrapolated upward through the sea surface.

One goal of normalization is to transform a satellite-based measurement of top-of-the-atmosphere radiance L_t into something that can be compared with a standard measurement made in situ, in the ocean, for whatever Sun zenith angle, viewing direction, atmospheric conditions, and wave state occurred at the time of the satellite measurement. Let this standard in situ measurement be the nadir-viewing radiance measured just below the sea surface, $L_u(0^-, \theta_v = 0)$. Depth $z = 0^-$ refers to a location in the water just below the sea surface; 0^+ refers to a location in the air just above the sea surface. Dividing $L_u(0^-, \theta_v = 0)$ by the downwelling plane irradiance within the water, $E_d(0^-)$, gives the in-water Remote Sensing Ratio *RSR*:

$$RSR \triangleq \frac{L_u(0^-, \theta_v = 0)}{E_d(0^-)}. \quad (15.5)$$

The division of $L_u(0^-, \theta_v = 0)$ by $E_d(0^-)$ removes the “zeroth order” effect of solar zenith angle θ_s and the “first order” atmospheric effects (including aerosol effects) on the magnitude of $L_u(0^-, \theta_v = 0)$. We now want to transform L_t into something comparable.

In the early days of satellite remote sensing, it was sometimes assumed that the upwelling underwater radiance distribution is isotropic. Under that assumption, *RSR* is approximately what you would get if the Sun were at the zenith and there were no atmosphere (i.e., the sky were black). This was the origin of statements that the normalized water-leaving radiance is the water-leaving radiance “which would exit the sea surface if the Sun were at the zenith and if the atmosphere were absent” (Gordon et al., 1988, page 10,910).

Fig. 15.9 compares upwelling and water-leaving radiances for “no atmosphere” or “black sky” versus realistic sky conditions. The curves of this figure were generated using HydroLight with the Sun placed at the zenith. The sky was either black (a collimated incident sky radiance) or had a diffuse radiance angular distribution typical of a clear sky. The water IOPs were determined using the same “new Case 1” bio-optical model (Section 8.9) as for the simulations of the previous section. Runs were made for chlorophyll values of 0.05, 0.5 and 5 mg m⁻³. The runs were at a wavelength of 550 nm and the sea surface was level. Each radiance is normalized by its value at the nadir-viewing direction to isolate the differences in the shapes of the curves. The upper set of curves shows the shape of the upwelling radiances L_u just below the sea surface as functions of the in-water, off-nadir viewing angle θ'_v at right angles ($\phi = 90$ deg) to the solar plane. The lower set of curves (those curving downward in the figure) shows the corresponding water-leaving radiances L_w as functions of the in-air, off-nadir viewing angle θ_v . For a level sea surface, these in-water and in-air angles are related by Snell’s law $\sin \theta_v = n_w \sin \theta'_v$, where $n_w \approx 1.34$ is the water index of refraction. Although the differences in the black-sky and real-sky radiances are only a few percent over the range of angles relevant to most remote sensing ($\theta_v \lesssim 60$ deg), differences of this magnitude are significant given the high accuracy requirements for retrieved water-leaving radiances in ocean remote sensing.

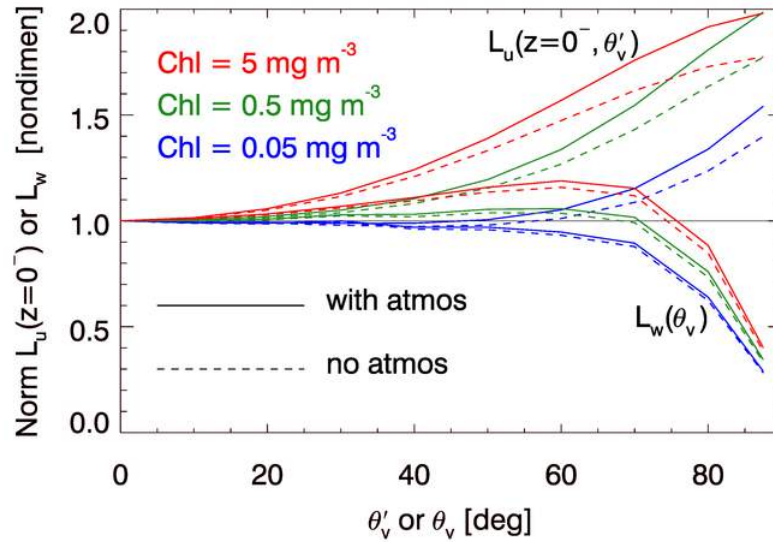


Figure 15.9: Comparisons of $L_u(z = 0^-, \theta_s = 0, \theta'_v, \phi = 90)$ and $L_w(\theta_s = 0, \theta_v, \phi = 90)$ for a zenith Sun in a black sky (no atmosphere; dashed lines) versus a zenith Sun in a typical atmosphere (solid lines). The colors identify the chlorophyll concentrations. The black line at an ordinate value of 1 corresponds to an isotropic radiance distribution.

The observation that the upwelling radiance distribution depends on the sky radiance distribution indicates that the idea of removing the atmosphere is too extreme. [Morel and Gentili \(1996, page 4852\)](#) therefore revised the definition of normalized water-leaving radiance to be “...the radiance that could be measured by a nadir-viewing instrument, if the Sun were at the zenith in the absence of any atmospheric loss, and when the Earth is at its mean distance from the Sun.” *The distinction between “atmosphere were absent” (i.e., a vacuum) and “absence of any atmospheric loss” (i.e., no attenuation by the atmosphere) is fundamental to understanding the developments of this section.*

The normalization proceeds as follows. The first step is to account for solar zenith angle and atmospheric attenuation effects on $L_w(\theta_s, \theta_v, \phi)$ via (e.g., [Gordon and Clark \(1981, page 10,910\)](#) or [Gordon and Wang \(1994b, Eq. 4\)](#)):

$$[L_w(\theta_v, \phi)]_N \triangleq \frac{L_w(\theta_s, \theta_v, \phi)}{\cos \theta_s t(\theta_s)},$$

where $t(\theta_s)$ is the atmospheric diffuse transmittance for irradiance in the Sun’s direction for the given atmospheric conditions. Recent papers include an explicit factor to correct L_w for the Earth-Sun distance at the time of measurement:

$$[L_w(\theta_v, \phi)]_N \triangleq \left(\frac{R}{R_o} \right)^2 \frac{L_w(\theta_s, \theta_v, \phi)}{\cos \theta_s t(\theta_s)}. \quad (15.6)$$

Here R is the Earth-Sun distance at the time of measurement, and R_o is the mean Earth-Sun distance. The $(R/R_o)^2$ factor corrects the L_w measurement to what it would be at

the mean Earth-Sun distance. (The solar irradiance at the TOA varies by about 8% over the course of a year due to the Earth's elliptical orbit.)

$[L_w(\theta_v, \phi)]_N$ is called the *normalized water-leaving radiance*. It is the water-leaving radiance that would occur if the earth were at the mean Earth-Sun distance, the Sun were at the zenith, and the atmosphere were non-attenuating. Note that although the factors of $(R/R_o)^2$, $\cos \theta_s$, and $t(\theta_s)$ largely remove the effects of Earth-Sun distance, solar zenith angle, and atmospheric attenuation, respectively, on the measured L_w , the normalized water-leaving radiance still refers to a particular viewing direction and depends on the sky angular radiance distribution at the time of observation.

Multiplying the $[L_w(\theta_v, \phi)]_N$ of Eq. (15.6) by a factor of π/F_o , where π has units of steradian and F_o is the extraterrestrial solar irradiance at the mean Earth-Sun distance, gives the *nondimensional normalized water-leaving reflectance*² $[\rho_w]_N$ (Gordon and Wang, 1994b, page 7756):

$$[\rho_w(\theta_v, \phi)]_N \triangleq \frac{\pi}{F_o} [L_w(\theta_v, \phi)]_N = \pi \frac{\left(\frac{R}{R_o}\right)^2 L_w(\theta_s, \theta_v, \phi)}{F_o \cos \theta_s t(\theta_s)}. \quad (15.7)$$

The *remote-sensing reflectance* R_{rs} is defined as

$$R_{rs}(\theta_s, \theta_v, \phi) \triangleq \frac{L_w(\theta_s, \theta_v, \phi)}{E_d(0^+, \theta_s)}. \quad (15.8)$$

In this definition, both L_w and E_d are values for the Earth-Sun distance at the time of measurement. However, this R_{rs} is numerically the same as what would be obtained if both L_w and E_d were corrected to the mean Earth-Sun distance by $(R/R_o)^2$ factors applied to each, because the correction factors on L_w and E_d cancel out. That is to say, the irradiance at the sea surface for Earth-Sun distance R is

$$E_d(0^+, \theta_s) = F_o \left(\frac{R_o}{R}\right)^2 \cos \theta_s t(\theta_s). \quad (15.9)$$

It thus follows that

$$[\rho_w(\theta_v, \phi)]_N = \pi \frac{\left(\frac{R}{R_o}\right)^2 L_w(\theta_s, \theta_v, \phi)}{F_o \cos \theta_s t(\theta_s)} = \pi \frac{L_w(\theta_s, \theta_v, \phi)}{E_d(\theta_s)} = \pi R_{rs}(\theta_v, \phi). \quad (15.10)$$

Another way to view $[\rho_w]_N$ is to think of it as the bidirectional reflectance distribution function (BRDF) of the ocean normalized by the BRDF of a perfectly reflecting Lambertian surface. The BRDF of a surface as measured in the laboratory is the radiance reflected by the surface divided by the incident plane irradiance onto the surface. The BRDF of a

²Radiant energy is a physical quantity that propagates through space and that can leave the water. Radiance is a derived physical quantity that likewise can leave the water, so it makes sense to speak of the “water-leaving radiance.” Reflectance, however, is a property of a surface. Reflectance is not a physical quantity that can leave a surface and propagate through space, so it does not make sense to speak of the “water-leaving reflectance.” However, the term “water-leaving reflectance” is well established shorthand for “reflectance based on the water-leaving radiance and the incident irradiance,” so I will use the term even if it is physically improper.

Lambertian surface whose irradiance reflectance is R is R/π , with units of inverse steradian. For a perfect Lambertian reflector, $R = 1$, and

$$[\rho_w]_N = \frac{BRDF_{\text{ocean}}}{BRDF_{\text{Lamb}}} = \frac{L_w/E_d}{1/\pi} = \pi R_{\text{rs}}.$$

This makes it clear that the π carries units of solid angle, so that $[\rho_w]_N$ is nondimensional.

15.3.2 The BRDF Effect

As noted above, the normalizations contained in $[\rho_w(\theta_v, \phi)]_N$ or $R_{\text{rs}}(\theta_v, \phi)$ remove the effects of solar zenith angle, atmospheric attenuation, and Earth-Sun distance on a measured radiance L_w . However, $[\rho_w(\theta_v, \phi)]_N$ still refers to a particular viewing direction (θ_v, ϕ) . This dependence ties $[\rho_w(\theta_v, \phi)]_N$ to the angular distribution of the upwelling underwater radiance and to the transmittance through the sea surface from water to air (which depends on the wave state, i.e., on the wind speed). The upwelling underwater radiance in turn depends on the angular distribution of the incident sky radiance, surface transmittance from air to water, and on the absorbing and scattering properties of the water body (the scattering phase function in particular). The dependence of the upwelling radiance distribution on the sky radiance distribution, viewing geometry, and water optical properties is commonly called the BRDF effect. The final step is to remove the BRDF effect to the greatest extent possible.

The BRDF effect was studied by Morel and colleagues in a series of papers, [Morel and Gentili \(1991\)](#), [Morel and Gentili \(1993\)](#), and [Morel and Gentili \(1996\)](#), culminating in [Morel et al. \(2002\)](#). They used numerical radiative transfer models to compute correction factors that would transform a measurement made for a particular Sun zenith angle, viewing direction, wind speed, atmospheric conditions, and water IOPs into a measurement that corresponds to a zenith Sun and nadir viewing direction *for a typical marine atmosphere and for Case 1 water with a given chlorophyll value*. The correction involves three separate factors, \mathfrak{R} , f , and Q , as follows.

Let $\mathfrak{R}(\theta'_v, U)$ be a nondimensional factor that accounts for all transmission and reflection effects by the wind-blown sea surface when $E_d(0^+)$ is transmitted downward through the surface to give $E_d(0^-)$, and $L_u(0^-, \theta'_v, \phi)$ is transmitted upward through the surface to give $L_w(0^+, \theta_v, \phi)$. Polar angle θ'_v (measured from the nadir) is the underwater angle that is refracted by the surface into the above-surface viewing direction θ_v of the water-leaving radiance $L_w(\theta_v, \phi_v)$. U is the wind speed. $\mathfrak{R}(\theta'_v, U)$ depends on the wind speed (i.e., the surface wave state) and the water index of refraction via the Snell's law mapping of θ'_v to θ_v . However, [Gordon \(2005\)](#) showed that the dependence of $\mathfrak{R}(\theta'_v, U)$ on wind speed is very weak, and usually \mathfrak{R} can be computed with adequate accuracy over a wide range of viewing angles using $U = 0$. The detailed derivation of $\mathfrak{R}(\theta'_v, U)$ in [Morel and Gentili \(1996, Eq. 5 and Appendix D\)](#) shows that, strictly speaking, \mathfrak{R} also depends on the solar zenith angle and the angular distribution of the incident atmospheric radiance, which affect how much incident irradiance is transmitted through the sea surface. Likewise, \mathfrak{R} depends on the water IOPs via the in-water irradiance reflectance $R(0^+) = E_u(0^+)/E_d(0^+)$. However, these dependencies are weak compared to the directional (θ'_v) dependency and so, for compactness and consistency with Morel's notation, are not shown as arguments.

Let $\mathfrak{R}_o(U)$ be the reference value of $\mathfrak{R}(\theta'_v, U)$ corresponding to transmission normal to the mean sea surface ($\theta_v = \theta'_v = 0$). Multiplying the $[L_w(\theta_v)]_N$ of Eq. (15.6) by

$\mathfrak{R}_o(U)/\mathfrak{R}(\theta'_v, U)$ corrects for surface-transmission effects for the actual viewing direction θ'_v and wind speed U . For $\theta'_v \approx 0$, $\mathfrak{R}_o \approx 0.53$; \mathfrak{R}_o is greater than 0.52 out to angles of 50 deg. (Figure 4 of [Morel et al. \(2002\)](#), which shows ratios of $\mathfrak{R}_o(U)/\mathfrak{R}(\theta'_v, U)$, was incorrectly computed; see the discussion and revised figures in [Gordon \(2005\)](#).)

The dimensionless factor f is defined by $E_u(0^-)/E_d(0^-) \triangleq f \times (b_b/a)$, where a and b_b are the water absorption and backscatter coefficients, respectively, which are assumed to be independent of depth for the present discussion. This factor parameterizes how downwelling irradiance in the water is converted to upwelling irradiance by backscatter and reduced by absorption. That is to say, f relates the irradiance reflectance within the water to the most relevant IOPs. f values are in the range of 0.3 to 0.5 ([Morel and Gentili, 1996](#), Fig. 2).

The factor $Q \triangleq E_u(0^-)/L_u(0^-)$ (units of steradian) describes the angular distribution of the upwelling radiance. $Q = \pi$ sr for an isotropic upwelling radiance distribution; actual in-water radiance distributions typically have Q values in the range of 3 to 6 sr ([Morel and Gentili, 1996](#), Fig. 3).

In practice, f and Q are combined to give a term that has less variability than the individual factors. The combined factor $f/Q = L_u(0^-, \theta'_v, \phi)/[E_d(0^-)(b_b/a)]$ describes how the downwelling irradiance just beneath the sea surface is reflected back upward as upwelling radiance in the direction (θ'_v, ϕ) . The f/Q term thus describes both the efficiency of conversion of downwelling irradiance into upwelling radiance, and the angular (non-isotropic) distribution of the upwelling underwater radiance that generates the water-leaving radiance. f/Q values are typically in the range to 0.07 to 0.15 ([Morel and Gentili, 1996](#), Fig. 6). Let f_o/Q_o refer to the ratio for the nadir viewing direction and Sun at the zenith.

Multiplying $[L_w(\theta_v, \phi)]_N$ by $(f_o/Q_o)/(f/Q)$ corrects for the difference of the actual angular distribution of the upwelling radiance and what that distribution would be for the Sun at the zenith, nadir viewing, *for the particular atmospheric and oceanic conditions used to compute f and Q* .

Applying these BRDF corrections to the $[L_w(\theta_v, \phi)]_N$ of Eq. (15.6) gives ([Morel et al., 2002](#), Eq. 13):

$$[L_w]_N^{\text{ex}} \triangleq [L_w(\theta_v, \phi)]_N \frac{\mathfrak{R}_o(U)}{\mathfrak{R}(\theta'_v, U)} \frac{f_o(\text{ATM}, U, \text{IOP})}{Q_o(\text{ATM}, U, \text{IOP})} \left[\frac{f(\theta_s, \text{ATM}, U, \text{IOP})}{Q(\theta_s, \theta'_v, \phi, \text{ATM}, U, \text{IOP})} \right]^{-1}. \quad (15.11)$$

Morel et al. call $[L_w(\lambda)]_N^{\text{ex}}$ the “exact normalized water-leaving radiance.” The arguments “ATM” and “IOP” refer to the specific set of atmospheric conditions and water inherent optical properties used to compute f and Q . (As previously noted, these arguments are omitted from the \mathfrak{R} terms because the ATM and IOP dependencies are small for \mathfrak{R} .) This $[L_w(\lambda)]_N^{\text{ex}}$ is equivalent to the normalization seen in [Franz et al. \(2007, Eq. 2\)](#), except for a sensor-specific correction factor for out-of-band wavelength response. It should be noted that the f_o/Q_o factor has arguments of (ATM, U , IOP) because those values correspond to the same atmospheric and oceanic conditions as the f/Q factor; the difference is that f_o/Q_o corresponds to $\theta_s = 0$ and $\theta'_v = 0$. Values of the $(\mathfrak{R}_o/\mathfrak{R})(f_o/Q_o)/(f/Q)$ product are typically in the 0.6 to 1.2 range, depending on the IOPs, solar zenith angle, atmospheric conditions, wind speed, and wavelength.

It is noted that the wind speed at the time of observation, U , is shown in both the \mathfrak{R}_o and (f_o/Q_o) terms in Eq. (15.11). It can be argued that these terms should use a reference value of $U = 0$, so that all quantities are referred to a level sea surface. However,

as presently implemented by the OBPB, the same wind speed U is used in the reference terms.

The atmospheric conditions used to compute the f/Q factors were typical of marine atmospheres and are summarized as follows (Morel and Gentili, 1996; Morel et al., 2002):

- The atmosphere was modeled as 50 layers, each 1 km thick
- The aerosol optical thickness was $\tau_a = 0.2$ at 550 nm.
- The tropospheric aerosols had a relative humidity of 70% and were distributed in the upper 45 atmospheric layers
- The marine aerosols had a relative humidity of 90% and were distributed in the lower 5 atmospheric layers
- The aerosol phase functions were modeled using Mie theory and the data of Shettle and Fenn (1979).
- The sea surface was modeled using a Gaussian distribution of wave slopes based on the empirical wind-speed, wave-slope data of Cox and Munk (1954b, page 847) for the given wind speed U . U was taken to be 0 for the Q calculations, although there are still some residual waves according to the empirical slope data of Cox and Munk.

The water optical properties were obtained from bio-optical models for Case 1 water, for which the IOPs can all be parameterized by the chlorophyll concentration Chl . In particular, the scattering phase function was computed as a chlorophyll-weighted sum of phase functions for “small” and “large” particles, which themselves were computed using T-matrix theory for non-spherical particles with different size distributions. The details of the IOP bio-optical models are given in Morel et al. (2002).

The radiative transfer calculations were carried out using a Monte Carlo code for calculation of the \mathfrak{R} factors, and using HydroLight for the f/Q calculations. The \mathfrak{R} are tabulated for an exactly level sea surface and for wind speeds of $U = 0, 4, 10$ and 16 m s^{-1} ; the $U = 0$ values include a small amount of residual waves because the Cox-Munk mean square sea surface slopes are not exactly zero for a zero wind speed. The f/Q calculations were done both with and without Raman scattering by water. None of the calculations included polarization. These codes were run for

- 7 wavelengths (412.5, 442.5, 490, 510, 560, 620, and 660 nm)
- 6 chlorophyll values ($Chl = 0.03, 0.1, 0.3, 1.0, 3.0$ and 10 mg m^{-3}). The water was homogeneous.
- 6 solar zenith angles ($\theta_s = 0, 15, 30, 45, 60$ and 75 deg)
- 13 azimuthal angles ($\phi = 0$ to 180 deg by steps of 15 deg)
- 17 nadir angles ($\theta'_v = 1.078, 3.411, 6.289, 9.278, 12.300, 15.330, 18.370, 21.410, 24.450, 27.500, 30.540, 33.590, 36.640, 39.690, 42.730, 45.780,$ and 48.830 deg)

These runs give a total of $7 \times 6 \times 6 \times 13 \times 17 = 55,629$ f/Q values, which are organized into $7 \times 6 \times 6 = 252$ tables, each with 13 columns and 17 rows. A separate table gives values of $\mathfrak{R}(\theta'_v, U)$ at θ'_v increments of 1 deg and for wind speeds of $U = 0, 4, 10$ and 16 m s^{-2} . The tables including Raman effects are available at [BRDF tables](#).

Use of these tables requires the chlorophyll concentration, which is not *a priori* known. In the initial study of Morel and Gentili (1996), a band-ratio algorithm was used to obtain

an initial guess for Chl . That value was then used in the tables to obtain the BRDF correction factors. Those factors give a new estimate of $[L_w(\lambda)]_N^{\text{ex}}$, which can be used to obtain a new value of Chl , and so on. However, the tables are not tied to how the Chl value is obtained, so in practice any algorithm can be used to obtain Chl values from the reflectances in an iterative process.

Finally, the exact normalized water-leaving radiance of Eq. (15.11) is used in Eq. (15.7) to obtain the *exact normalized water-leaving reflectance*:

$$\begin{aligned}
 [\rho_w]_N^{\text{ex}} &\triangleq \frac{\pi}{F_o} [L_w]_N^{\text{ex}} & (15.12) \\
 &= \left\{ \frac{\left(\frac{R}{R_o}\right)^2 \pi \mathfrak{R}_o(U) f_o(\text{ATM}, U, \text{IOP})}{F_o \cos \theta_s t(\theta_s) \mathfrak{R}(\theta'_v, U) Q_o(\text{ATM}, U, \text{IOP})} \left[\frac{f(\theta_s, \text{ATM}, U, \text{IOP})}{Q(\theta_s, \theta'_v, \phi, \text{ATM}, U, \text{IOP})} \right]^{-1} \right\} \\
 &\times L_w(\theta_s, \theta_v, \phi). & (15.13)
 \end{aligned}$$

The quantities in brackets can all be obtained from pre-computed look-up-tables given the Sun and viewing geometry, wavelength, atmospheric conditions used to obtain $L_w(\theta_s, \theta_v, \phi)$ from $L_t(\theta_s, \theta_v, \phi)$ (used to determine $t(\theta_s)$), and the chlorophyll concentration. The chlorophyll concentration determines the IOPs according to the bio-optical models for Case 1 water used in the Morel et al. calculations.

Although Morel and others call $[L_w]_N^{\text{ex}}$ the “exact” normalized water-leaving radiance, and $[\rho_w]_N^{\text{ex}}$ the “exact” normalized water-leaving reflectance, it must be remembered that these quantities are exact only if the atmosphere and ocean have exactly the same absorbing and scattering properties as used in the model simulations upon which the BRDF correction factors are based. That will of course not in general be the case. In Case 1 waters, the differences between the Morel et al. Case 1 IOP model and the actual ocean IOPs are often small enough that the BRDF-corrected quantities are sufficiently accurate for remote sensing. However, the differences can become large in Case 2 waters. Research therefore continues on ways to improve the BRDF correction, both to extend its validity to Case 2 water and to remove the need to estimate the chlorophyll concentration in order to use the look-up tables (Lee et al., 2011; Fan et al., 2016). Although further improvements can be anticipated, the BRDF correction as described above remains the OBPG operational algorithm at the time of this writing.

The OBPG works with radiance to get to $[L_w]_N^{\text{ex}}$. However, when doing atmospheric correction on TOA radiances, the various look-up-tables used for Rayleigh correction, etc. are in terms of reflectances $[\rho_w]_N^{\text{ex}}$. Equation (15.12) allows easy conversion from one to the other, depending on which quantity is most convenient for a given step of the atmospheric correction process.

It is to be noted that the “remote-sensing reflectance” reported by the NASA OBPG as a standard Level 2 product of the NASA ocean color satellites such as MODIS and VIIRS is the exact normalized water-leaving reflectance of Eq. (15.13), divided by π :

$$R_{\text{rs}}(\text{NASA}) = \frac{[\rho_w]_N^{\text{ex}}}{\pi} = \frac{[L_w]_N^{\text{ex}}}{F_o}. \quad (15.14)$$

However, the “remote-sensing reflectance” computed by HydroLight is $R_{\text{rs}}(\theta_s, \theta_v, \phi)$ as defined by Eq. (15.8). A given HydroLight run computes $R_{\text{rs}}(\theta_s, \theta_v, \phi)$ for all viewing directions θ_v, ϕ for the given θ_s and other conditions of wind speed, IOPs, and atmospheric

radiance distribution. Thus the HydroLight R_{rs} outputs for various viewing directions incorporate the BRDF effects. (Indeed, comparison of HydroLight-computed $R_{\text{rs}}(\theta_{\text{s}}, \theta_{\text{v}}, \phi)$ values with HydroLight values for a zenith Sun and nadir viewing direction is how the f/Q BRDF factors were determined in the Morel et al. studies.) Thus there is no need for an explicit BRDF correction to the HydroLight-computed R_{rs} . *If a HydroLight run has the Sun at the zenith, then the nadir-viewing HydroLight R_{rs} times π corresponds to $[\rho_{\text{w}}]_{\text{N}}^{\text{ex}}$:*

$$[\rho_{\text{w}}]_{\text{N}}^{\text{ex}} = \pi R_{\text{rs}}(\text{HydroLight}; \theta_{\text{s}} = 0, \theta_{\text{v}} = 0). \quad (15.15)$$

Recalling Eq. (15.14), this gives

$$R_{\text{rs}}(\text{NASA}) = R_{\text{rs}}(\text{HydroLight}; \theta_{\text{s}} = 0, \theta_{\text{v}} = 0). \quad (15.16)$$

There should be only a very small difference in HydroLight's πR_{rs} and $[\rho_{\text{w}}]_{\text{N}}^{\text{ex}}$, attributable to any differences in the angular distributions of the sky radiances used to compute the Morel BRDF factors and as used in HydroLight. Morel et al. (2002, page 6295) notes that these differences are negligible.

It should be noted that HydroLight R_{rs} values are valid for whatever IOPs were used in the run; there is no restriction to homogeneous Case 1 water and no need to estimate a chlorophyll concentration as must be done when applying the Morel BRDF factors. Thus *HydroLight can give a very general $[\rho_{\text{w}}]_{\text{N}}^{\text{ex}}$ via Eq. (15.15), without the assumptions made by Morel et al. when developing the BRDF correction factors seen in Eq. (15.13).*

Figure 15.10 illustrates the magnitudes of the corrections to $\pi R_{\text{rs}}(\theta_{\text{s}}, \theta_{\text{v}}, \phi)$ values. HydroLight was first run to generate remote-sensing reflectances for a Sun zenith angle of $\theta_{\text{s}} = 50$ deg, with the Sun in a clear sky. The water IOPs and chlorophyll values were the same as those used for Fig. 15.9. The runs included Raman scatter by water and fluorescence by chlorophyll and CDOM. The dashed lines in Fig. 15.10 show the values of $\pi R_{\text{rs}}(\theta_{\text{s}} = 50, \theta_{\text{v}} = 30, \phi = 90)$. These viewing angles correspond to a sensor viewing the ocean at an off-nadir angle of 30 deg at right angles to the solar plane. The HydroLight runs were then repeated with the Sun at the zenith. The solid lines in the figures show the resulting values of $[\rho_{\text{w}}]_{\text{N}}^{\text{ex}}$ as determined by Eq. (15.15). Depending on the water IOPs, Sun zenith angle, viewing direction, and wavelength, the normalization can change a spectrum by tens of percent, or almost not at all. The changes tend to be greatest in high-chlorophyll waters, at large solar zenith angles, and at large off-nadir viewing angles.

When formulated in terms of reflectances, the partitioning of Eq. (15.3) of the Problem Formulation section 15.2 becomes (e.g., Gordon and Wang, 1994a)

$$\rho_{\text{t}} = \rho_{\text{R}} + [\rho_{\text{a}} + \rho_{\text{Ra}}] + T\rho_{\text{g}} + t\rho_{\text{wc}} + t\rho_{\text{w}}, \quad (15.17)$$

where the terms correspond to those of Eq. (15.3). A similar equation applies to the reflectance form of Eq. 15.4.

Figure 15.11 shows the radiances of Fig. 15.5 recast as normalized reflectances. It should be noted in Fig. 15.11 that the solar-spectrum features (most noticeable below 600 nm) seen in the TOA E_{d} spectrum of Fig. 15.5 are removed by the normalization process. However, the atmospheric absorption features (most noticeable beyond 600 nm) remain in the TOA reflectances, but are not present in the surface reflectances. Thus the dotted curves for ρ_{t} , ρ_{g} , and ρ_{w} , are very smooth functions of wavelength. The surface glint reflectance spectrum ρ_{g} is almost independent of wavelength, whereas the surface

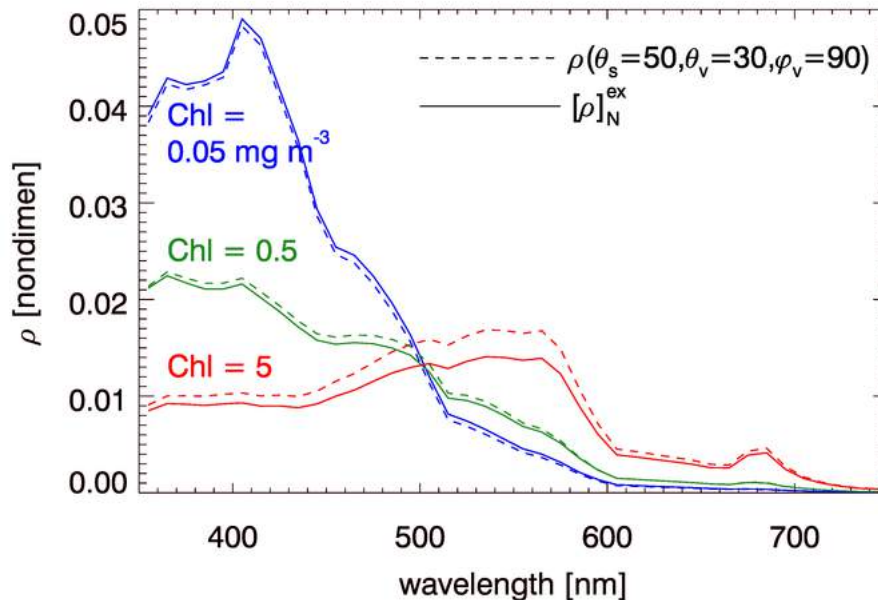


Figure 15.10: Comparison of exact normalized water-leaving reflectances $[\rho_w]_N^{\text{ex}}$ (solid lines) with unnormalized reflectances (dashed lines) for $\theta_s = 50, \theta_v = 30, \phi = 90$. The color identifies the chlorophyll values of 0.05, 0.5, and 5 mg m^{-3} .

glint radiance seen in Fig. 15.5 depends of wavelength in the same manner as the surface irradiance spectrum³.

15.3.3 Summary

The *interpretation* of $[\rho_w]_N^{\text{ex}}$ as developed above can be summarized as follows:

- Start with a measured $L_w(\theta_s, \theta_v, \phi)$. Then
- The division by $\cos \theta_s$ in Eq. (15.6) moves the Sun to the zenith.
- The division by $t(\theta_s)$ in Eq. (15.6) rescales the radiance magnitude to account for atmospheric attenuation, but the angular distribution of the radiance is unchanged. This diffuse transmission is for the actual atmosphere at the time of observation.
- The $(R/R_o)^2$ factor in Eq. (15.6) corrects for the Earth-Sun distance.
- The BRDF factors in Eq. (15.11) normalize L_w as measured for the actual atmospheric and in-water radiance distributions to *what L_w would be for a reference atmospheric radiance distribution and for Case 1 water with the given chlorophyll value.*

The proposed atmospheric retrieval accuracy requirements for the PACE (Pre-Aerosol, Clouds, and ocean Ecosystem) mission are (Franz, 2014)

³Note that the glint reflectance ρ_g considered here is not the same as the surface radiance reflectance factor ρ of Mobley (1999, Eq. 4) and Mobley (2015), which is a ratio of reflected to incident sky radiances for the given wind speed, and Sun and viewing directions.

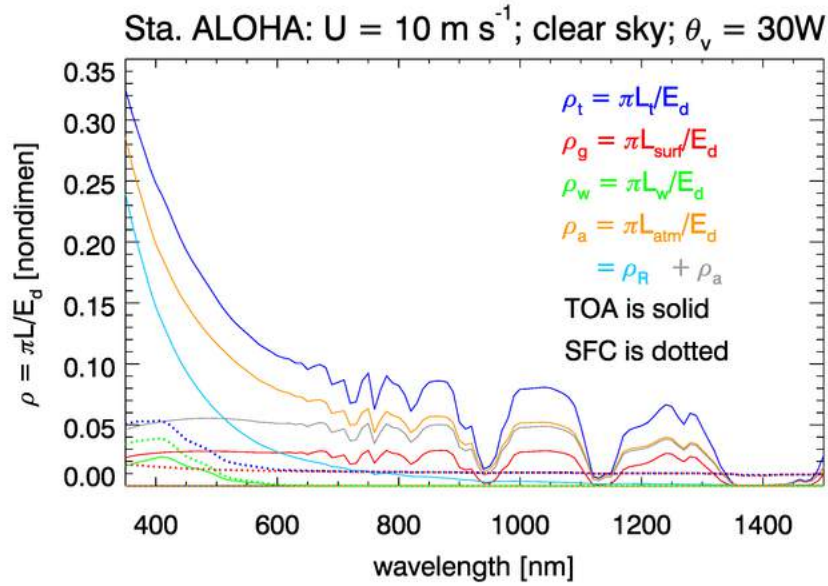


Figure 15.11: Normalized reflectances corresponding to the radiances of Fig. 15.5. These reflectances are $[\rho]_N$, not $[\rho]_N^{\text{ex}}$.

- for $[\rho]_N^{\text{ex}}$ in the 350-400 nm range, a maximum error of 0.002 or 10%
- for $[\rho]_N^{\text{ex}}$ in the 400-710 nm range, a maximum error of 0.001 or 5%

In terms of $R_{\text{rs}} = [\rho]_N^{\text{ex}}/\pi$, this gives

- for R_{rs} in the 350-400 nm range, a maximum error of $6 \times 10^{-4} \text{ sr}^{-1}$
- for R_{rs} in the 400-710 nm range, a maximum error of $3 \times 10^{-4} \text{ sr}^{-1}$

Given that the water-leaving radiance L_w is at most 10% of the TOA radiance (e.g., Fig. 15.7), a requirement for a 5% maximum error in L_w (expressed as the same percentage error in $[\rho]_N^{\text{ex}}$ or R_{rs}) implies a maximum error of roughly 0.5% in the measured TOA radiance L_t . Such a small error in the TOA radiance cannot be achieved by pre-launch sensor radiometric calibration alone. Thus ocean color sensors require post-launch “vicarious calibration.” This term refers to the process of forcing measured TOA radiances (or the equivalent reflectances) to agree with measured water-leaving radiances (or reflectances) as propagated from the sea surface to the TOA. This agreement is obtained via an empirically determined set of sensor calibration correction factors to be applied to the measured TOA radiances so that they lead, after atmospheric correction, to the proper sea-level water-leaving radiances and associated reflectances. Vicarious calibration is described in more detail in Section 15.5.

The Q part of the Morel et al. correction has been validated against empirical data for Case 1 waters by Voss et al. (2007) and found to give good agreement with measured radiance distributions. As concluded there, “...the bidirectional corrections based on the lookup tables generated from the model, and presently applied to ocean color imagery, are sound and amply validated for Case 1 waters...” However, it should be remembered that

the above BRDF correction is based on a particular atmospheric model and on particular bio-optical models for Case 1 water. For different atmospheric conditions, the f/Q correction would be different, although the difference would probably be small. However, for different water IOPs, in particular for Case 2 waters with high mineral particle loads or high concentrations of CDOM, the differences in the water-column scattering and absorption properties could have a significant effect on the f/Q correction. That variability has not yet been studied, and the above correction remains the current state of the science and is implemented by the OBPG for operational ocean color image processing.

15.4 Atmospheric Transmittances

Equations such as (15.3) of the Problem Formulation section and (15.17) of the Normalized Reflectances section involve various direct (T) and diffuse (t) atmospheric transmission terms, which require discussion.

15.4.1 Direct Transmittance

Direct atmospheric transmission is used if only one particular path, or a narrow bundle of nearly colinear paths, connects the source and the observer. This is the case for specular reflection, as illustrated in the left panel of Fig. 15.12. When the sea surface is viewed by the sensor, only a small patch of sea surface is seen as the Sun's specular reflection or direct glint. Other points of the sea surface would be seen as localized patches of Sun glint for other viewing directions, but not by the sensor viewing direction as shown. In each case, the reflected radiance is traveling in a very narrow set of directions determined by the Sun's location and the law of reflection ($\theta_r = \theta_s$ for a level surface). The narrow beams of Sun glint for other viewing directions are represented by the dotted and dash lines in the left panel of the figure. These beams can influence the direction of interest only via two scatterings: the first out of the reflected beam and the second into the direction of interest. Such a two-scatter path is shown by the light dashed line. Two scatterings makes the contributions of the unseen beams of specular reflection to the direction of interest very small.

Let τ be the atmospheric optical depth along a vertical path (the nadir viewing direction for the sensor). This τ includes all effects of atmospheric absorption and attenuation by all atmospheric constituents. For an off-nadir viewing direction θ_v , the direct transmittance is then simply

$$T(\theta_v) = \exp(-\tau / \cos \theta_v). \quad (15.18)$$

This geometry is analogous to the Lambert-Beer law for radiance propagation of a beam through a homogeneous medium: $L(r) = L(0) \exp(-cr)$, where c is the beam attenuation coefficient and r is the distance traveled. In the present case, $\tau = cr_a$, where r_a is the distance though the atmosphere on a vertical (nadir-viewing) path, and $r = r_a / \cos \theta_v$ is the atmospheric path length along the viewing direction.

15.4.2 Diffuse Transmittance

For water-leaving radiance L_w , every point of the sea surface is emitting an upward distribution of radiance $L_w(\theta, \phi)$, as illustrated in the right panel of Fig. 15.12. Radiance from

all locations and various directions can be scattered into the direction of interest via only one scattering, as illustrated by the dotted line in the right panel of Fig. 15.12. Radiance scattered out of the beam along the viewing direction can thus be replaced via a single scattering from the radiance emitted by a neighboring point propagating in a different direction. The diffuse transmittance therefore depends not just on the atmospheric properties and viewing direction, but also on the angular distribution of L_w , which in general is unknown. This situation is analogous to the diffuse attenuation K_d of downwelling irradiance within the water column of a plane parallel ocean that is illuminated at all points of the surface. Because radiance absorbed or scattered out of the path between the source and sensor can be replaced by radiance from other sea surface points and directions, the diffuse transmittance $t(\theta_v)$ will be greater than the direct transmittance $T(\theta_v)$ (just as the the diffuse transmittance $\exp(-K_d z)$ is always greater than than the beam transmittance $\exp(-cz)$). Accounting for the angular distribution of L_w and for the scattering processes makes the computation of diffuse transmittance more complicated than for direct transmittance.

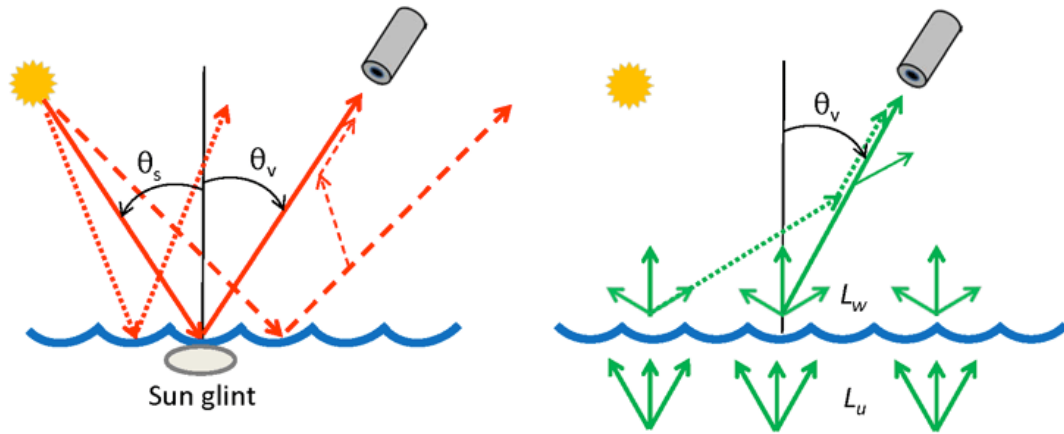


Figure 15.12: Left panel: Illustration of Sun glint as seen from the TOA, which is described by a direct transmittance. Right panel: Illustration of water-leaving radiance as seen from the TOA, which is described by a diffuse transmittance.

The diffuse transmittance of the water-leaving radiance along a particular viewing direction (θ_v, ϕ_v) is by definition

$$t(\theta_v, \phi_v, \lambda) \triangleq \frac{L_w^{\text{TOA}}(\theta_v, \phi_v, \lambda)}{L_w(\theta_v, \phi_v, \lambda)}, \quad (15.19)$$

where L_w is the water-leaving radiance at the sea surface and L_w^{TOA} is the water-leaving radiance that reaches the TOA. Henceforth, all quantities are assumed to be functions of wavelength, and the argument λ will be omitted. One way to compute $t(\theta_v, \phi_v)$ is to run a coupled ocean-atmosphere radiative transfer model to compute the needed $L_w(\theta_v, \phi_v)$ and $L_w^{\text{TOA}}(\theta_v, \phi_v)$ for a wide range of atmospheric and oceanic conditions, Sun and viewing geometries, and wavelengths. The values of $t(\theta_v, \phi_v)$ would then be obtained via Eq. (15.19) and tabulated for later use. Such tables would need to be constructed for all possible aerosol types (i.e, aerosol phase function shapes), aerosol optical thicknesses, water IOPs

(or at least for chlorophyll concentrations in Case 1 waters), and viewing geometries. The tables would necessarily be large because of the large number of parameters that can affect $t(\theta_v, \phi_v)$.

Yang and Gordon (1997) examined the computation of diffuse transmittance and errors therein on the retrieved water-leaving radiance. Using a combination of radiative transfer numerical modeling of the ocean and atmosphere and reciprocity principles, they compared diffuse transmittances computed using realistic modeled and measured $L_u(\theta', \phi')$ distributions versus diffuse transmittances computed on the assumption that L_u is isotropic. Here $L_u(\theta', \phi')$ denotes the upwelling underwater radiance distribution just beneath the sea surface. For a level sea surface, $L_w(\theta_v, \phi_v) = T_F(\theta')L_u(\theta', \phi')/n_w^2$, where n_w is the index of refraction of the water, the in-water θ' is related to the in-air θ_v by Snell's law $\sin \theta_v = n_w \sin \theta'$, and $T_F(\theta')$ is the Fresnel transmittance of the surface from water to air. When L_u is isotropic, the diffuse transmittance is independent of the azimuthal angle and is denoted by $t^*(\theta_v)$.

Yang and Gordon (1997, Eq. 3) show via a clever use of reciprocity that the diffuse transmittance of radiance t^* along an atmospheric path in the direction of the Sun at solar zenith angle θ_o is numerically equal to the diffuse transmittance of irradiance from the TOA to a depth just beneath the sea surface, on the assumption that there is no upwelling radiance within the water. That is,

$$t^*(\theta_o) = \frac{E_R(\theta_o)}{F_o \cos(\theta_o) T_F(\theta_o)}, \quad (15.20)$$

where $E_R(\theta_o)$ is the downwelling plane irradiance just beneath the sea surface for an extraterrestrial solar irradiance F_o incident onto the TOA at angle θ_o , and $T_F(\theta_o)$ is the Fresnel downward transmission of the sea surface for radiance incident at angle θ_o from the normal. Since the upwelling radiance $L_u(\theta', \phi')$ used to obtain this result is assumed to be isotropic, the azimuthal dependence of t^* is irrelevant, and the desired diffuse attenuation for radiance at viewing direction $\theta_v = \theta_o$ is equal to the value of the irradiance transmission at the same polar angle. The great virtue of Eq. (15.20) is that it allows the efficient numerical computation of t^* using backward Monte Carlo simulation (Sections E.7 and F.7) of downwelling irradiance for a given aerosol type and optical thickness.

The retrieved water-leaving radiance is, by Eq. (15.19),

$$L_w(\theta_v, \phi_v) = \frac{L_w^{\text{TOA}}(\theta_v, \phi_v)}{t(\theta_v, \phi_v)}. \quad (15.21)$$

Let L_w^* denote the retrieval when t^* rather than t is used in Eq. (15.21). The error in the retrieved water-leaving radiance due to using t^* rather than the exact t is

$$\frac{\Delta L_w}{L_w} = \frac{L_w^* - L_w}{L_w} = \frac{t - t^*}{t^*}. \quad (15.22)$$

Yang and Gordon (1997) found that for viewing angles ϕ_v perpendicular to the principle plane (the plane of the Sun), the errors in the retrieved L_w are no more than 4% for viewing angles $\theta_v \leq 45$ deg, Sun zenith angles $\theta_o \leq 60$ deg, and aerosol optical thicknesses typical of clear atmospheres. The errors in band-ratio algorithms were less; e.g., the error in $L_w(443)/L_w(555)$ (used to retrieve chlorophyll concentration) is less than 2% except for very clear water and some viewing directions, for which the error in the ratio is about

3%. These parameter ranges covered most of the needs for SeaWiFS and errors of this magnitude were deemed acceptable compared to other errors in the retrieval process (i.e., removal of atmospheric path radiance). In that case, the diffuse attenuation can be obtained by pre-computed functions of the form

$$t^*(\theta_v) = A(\theta_v) \exp[-B(\theta_v)\tau_a]. \quad (15.23)$$

where $A(\theta_v)$ and $B(\theta_v)$ are tabulated for each aerosol type. Thus, for N angles θ_v , only $2N$ numbers $A(\theta_v)$ and $B(\theta_v)$ need to be tabulated for each aerosol type. Moreover, the aerosol type and optical thickness are determined as part of the process to remove the aerosol contribution to the total path radiance. This enables selection of the appropriate A , B , and τ_a , and evaluation of Eq. (15.21) is operationally feasible. If aerosols are ignored, Eq. (15.23) reduces to $t^*(\theta_v) = \exp[-\frac{1}{2}\tau_R/\cos\theta_v]$, where τ_R is the Rayleigh optical thickness. This is the formula used in the early days for CZCS atmospheric correction. The virtue of Eq. (15.23) is that it allows the aerosol optical thickness τ_a to be incorporated into the diffuse transmittance calculations via a simple exponential and pre-computed A and B values.

However, for larger off-nadir viewing angles θ_v , for azimuthal viewing directions near $\phi_v = 0$ or 180 deg, and for very clear water, the errors $\Delta L_w/L_w$ can be as much as 6%. This could be significant for the MODIS Aqua sensor, which views a wide range of ϕ_v directions. Gordon and Franz (2008) therefore re-examined the model of Yang and Gordon (1997) and developed a correction term to t^* so that Eq. (15.21) becomes

$$L_w(\theta_v, \phi_v) = \frac{L_w^{\text{TOA}}(\theta_v, \phi_v)}{t^*(\theta_v)[1 + \delta(\theta_v, \phi_v)]}. \quad (15.24)$$

The $\delta(\theta_v, \phi_v)$ factor corrects for the bi-directional effects resulting from the use of an isotropic L_u in the computation of t^* , rather than the exact, non-isotropic $L_u(\theta', \phi')$. The $L_u(\theta', \phi')$ needed for computation of δ was obtained from $L_u = (f/Q)(b_b/a)$ and chlorophyll-based models for f/Q and b_b/a . Again, the δ correction term can be pre-computed and tabulated for various aerosol types and water IOPs. Evaluation of the impact of the δ correction applied to both SeaWiFS and MODIS Aqua data showed that retrieved water-leaving radiances will be in error by no more than $\sim 1\%$ if $\theta_v < 60$ deg and the δ correction is omitted. For $\theta_v > 60$ deg, i.e. near the edges of scan lines, the use of the δ correction is warranted.

In current operational practice, the tabulated A and B functions are applied for both downwelling (solar irradiance) and upwelling L_w paths. The $\delta(\theta_v)$ correction of Gordon and Franz (2008) is not applied because it is an added complication with no significant impact in most instances.

15.5 Vicarious Calibration

The atmospheric correction techniques described in the following sections remove most of the effects of atmospheric scattering and absorption and sea surface reflectance from a measured TOA radiance. The end result is an estimate of the water-leaving radiance, which can be converted to a normalized reflectance as shown in Section 15.3. However, neither the sensor radiometric calibration nor the atmospheric correction are perfect, so

a satellite-derived water-leaving radiance usually will not exactly match a water-leaving radiance measured at the sea surface. It is therefore necessary to make comparisons between satellite-derived radiances and radiances measured at the sea surface in order to determine the gain or correction factor needed to convert a best estimate of a radiance into one that agrees with the radiance measured at the sea surface. This process is called *vicarious calibration*.

As described in [Antoine et al. \(2006\)](#) and at the [BOUSSOLE website](#), there are two main vicarious calibration paths that can produce ocean color products of the desired accuracy, i.e., water-leaving radiances within an uncertainty of about 5% in the blue for an oligotrophic ocean. The first path is usually referred to as vicarious calibration, and consists in forcing satellite-derived water-leaving radiances to agree with a set of in situ water-leaving radiances (so-called match-up analyses). A set of vicarious calibration coefficients is obtained, which is applied to the top-of-atmosphere (TOA) total radiances measured by the sensor. The second procedure, which is also an indirect (vicarious) calibration and is sometimes referred to as a vicarious radiometric calibration, consists in simulating the TOA signal that the sensor should measure under certain conditions, and to compare it to the measured signal.

One of the difficulties of the first type of vicarious calibration is that it is dependent upon the procedure used for the atmospheric correction of the TOA observations. Even if it is admittedly less dependent upon the selected set of in situ water-leaving radiances, these measurements also contribute to the final accuracy. The advantage of this technique, besides the fact that atmospheric measurements are not needed, is that the marine signals delivered by several sensors that use different atmospheric correction algorithms can be cross-calibrated provided that the same set of in situ water-leaving radiances is used to perform the vicarious calibration. This is the case, for instance, for the SeaWiFS and OCTS sensors.

Inconveniences of the vicarious radiometric calibration are that it requires a set of in situ measurements that is usually difficult to collect, among other things because a high accuracy is needed. In addition to the in-water measurements of the water-leaving radiances, this data set includes sea state and atmospheric pressure, ozone concentration, aerosol optical thickness, aerosol type, and even aerosol vertical profile if the aerosols are found to be absorbing. If this data set is successfully assembled, the advantage of the vicarious radiometric calibration is that it is independent of the atmospheric correction algorithms, so that the TOA signals of various sensors can be cross-calibrated. Then it is up to any user to apply their preferred atmospheric correction to these TOA signals. The marine signals in that case might be inconsistent if significant differences exist in the various atmospheric corrections.

The greatest difficulty of the vicarious radiometric calibration lies in the estimation of the aerosol optical thickness, phase function, and single scattering albedo. These parameters are accessible through the inversion of Sun photometer measurements, yet uncertainties inevitably occur when applying such methods, for instance because of multiple scattering, perturbations from the ground reflectance, and uncertainties in the photometer calibration. Assembling all data needed for these vicarious calibration experiments is often compromised because the aerosol parameters are not accurate enough.

The current practice, including at NASA's OBPG, is to take the atmospheric path radiance from the atmospheric correction lookup tables. The reasoning is that the vicarious

calibration is then consistent with the atmospheric correction, which, in some ways, is a good point, but as just explained, also can be considered an issue.

The vicarious calibration methodology employed by the OBPB is described in Franz et al. (2007). Their procedure simultaneously corrects for residual errors in both sensor radiometric calibration and in atmospheric correction. The correction factors are therefore specific to a particular sensor and atmospheric correction procedure, but they are independent of the how the water-leaving radiance is obtained. It is assumed that the sensor has been calibrated as well as possible, so that only residual calibration errors need be corrected by the vicarious calibration process.

Operational data processing starts with a measured TOA total radiance L_t and derives the corresponding water-leaving radiance L_w . The process is based on Eq. (15.4) of the Problem Formulation section:

$$L_t = \left(L_r + L_A + t_{dv}L_{wc} + t_{dv}L_w \right) t_{gv}t_{gs}f_p, \quad (15.25)$$

where the radiances are defined in Table 15.1 and the transmittances are defined in Table 15.2. These quantities are all functions of wavelength. No specular-reflection term is included in Eq. (15.25) because it is assumed that pixels containing a detectable amount of specular reflection are omitted from consideration. In this equation, the aerosol radiance L_A and the water-leaving radiance L_w are the two primary unknowns. The aerosol-dependent diffuse transmittance t_{dv} can be computed for a given aerosol type as described in the Atmospheric Transmittances section. The Rayleigh and whitecap radiances and the gaseous transmittances can be computed for given atmospheric conditions as described in Sections 15.7 on non-absorbing gases, 15.8 on absorbing gases, and 15.10 on whitecaps. The determination of the aerosol contribution is the crux of the atmospheric correction process and is described in the Aerosols Section 15.11. These computations constitute the atmospheric correction process. Once these terms have been computed, Eq. (15.25) can be solved for a measured L_t to obtain the corresponding L_w .

To develop the correction factors, this process is reversed. Let L_w^{tar} denote the known or “target” water-leaving radiance that is to be matched by the satellite-derived value. In most instances, the target water-leaving radiance is obtained from in-situ measurement, but can in principle be the value retrieved by another satellite sensor or predicted by a model. In any case, this value can be propagated to the TOA under various assumptions about the atmospheric conditions to obtain the corresponding target TOA radiance L_t^{tar} which, ideally, would match the satellite-measured TOA radiance. The ratio

$$g(\lambda) = \frac{L_t^{\text{tar}}(\lambda)}{L_t(\lambda)} \quad (15.26)$$

is then the correction or gain factor that, when multiplied by a measured L_t gives an adjusted TOA radiance L_t^{tar} that, when atmospherically corrected, yields the correct water-leaving radiance L_w^{tar} . Note that the gain factor is different for each wavelength. The gain factors are created via a series of “match-up” comparisons of satellite and in situ data and then, once determined, are routinely applied as part of the operational reduction of satellite-measured TOA radiances to water-leaving radiances.

Now consider the details of the computation of the gain factors. A satellite-derived L_w is converted to an exact normalized water-leaving radiance as described in Section 15.3.

This process can be summarized as

$$[L_w]_N^{\text{ex}} = \frac{L_w}{\mu_s f_s t_{\text{ds}} f_b f_\lambda}, \quad (15.27)$$

where

- μ_s is the cosine of the solar zenith angle,
- f_s is the Earth-Sun distance correction factor,
- t_{ds} is the Rayleigh-aerosol diffuse transmittance in the Sun's direction,
- f_b is the BRDF correction factor, and
- f_λ is a band-pass adjustment factor.

A target water-leaving radiance is converted to an exact normalized water-leaving radiance in the same manner:

$$[L_w^{\text{tar}}]_N^{\text{ex}} = \frac{L_w^{\text{tar}}}{\mu_s^{\text{tar}} f_s^{\text{tar}} t_{\text{ds}}^{\text{tar}} t_{\text{gs}}^{\text{tar}} f_b^{\text{tar}} f_\lambda^{\text{tar}}}, \quad (15.28)$$

where now the superscript tar on the terms in Eq. (15.28) indicates that these terms are evaluated for the Sun and viewing geometry at the time of the measurement of L_w^{tar} , which may be different from the geometry at the time of the satellite observation leading to L_w . A factor of t_{gs} is included in Eq. (15.28) to account for the diffuse transmittance due to absorption by gases in the Sun's direction at the time of measurement of L_w^{tar} . This factor does not appear in Eq. (15.27) because that correction to the total measured radiance L_t^{tar} is accounted for in Eq. (15.25). The radiances in Eq. (15.25) are computed for the full spectral response of each sensor band. The f_λ factor converts these full-band values to nominal band-center wavelengths to remove residual out-of-band response effects. Since the satellite and in situ instruments usually have different spectral responses, this factor adjusts the satellite and in situ values to a common wavelength dependence.

Writing L_w in terms of $[L_w]_N^{\text{ex}}$ via Eq. (15.27) and then replacing $[L_w]_N^{\text{ex}}$ by the target value $[L_w^{\text{tar}}]_N^{\text{ex}}$ gives an equation for the target value of the TOA radiance:

$$L_t^{\text{tar}} = \left(t_{\text{dv}} [L_w^{\text{tar}}]_N^{\text{ex}} (\mu_s f_s t_{\text{ds}} f_b f_\lambda) + L_r + L_A^{\text{tar}} + t_{\text{dv}} L_{\text{wc}} \right) t_{\text{gv}} t_{\text{gs}} f_p. \quad (15.29)$$

The total transmittance along the Sun's path is the product of the diffuse transmittance for the Rayleigh and aerosol scattering and the diffuse transmittance for gaseous absorption. The $t_{\text{ds}}^{\text{tar}}$ term depends on the aerosols and is thus an unknown for the calibration target. The total transmittance for the target could be obtained from auxiliary measurements (e.g., from a Sun photometer) made at the time of the target radiance measurement. However, such measurements are not generally available and, even if available, any error in those measurements would be an additional source of error in the target radiance. Therefore, the satellite-retrieved atmospheric and aerosol properties are used to evaluate the total transmittance for the target measurement via

$$t_{\text{ds}}^{\text{tar}} t_{\text{gs}}^{\text{tar}} = \exp \left[\ln(t_{\text{ds}} t_{\text{gs}}) \frac{\mu_s}{\mu_s^{\text{tar}}} \right]. \quad (15.30)$$

The total transmittance for the target is therefore the total transmittance for the satellite with a correction for the difference in the solar zenith angles. Other terms in Eq. (15.29)

such as the Rayleigh radiance and gaseous transmittances along the viewing direction are evaluated for the atmospheric conditions of the satellite retrieval as described in sections 15.7 and 15.8. Thus the t_{dv} that multiplies L_{wc} is determined by Rayleigh-scattering calculations based on the sea-level pressure. The whitecap radiance is modeled as a function of wavelength and wind speed as described in section 15.10. These choices reference both the target and the satellite radiances to a common atmosphere, which is desirable for the development of the gain factors.

Finally, the BRDF correction factors f_b and f_b^{tar} must be evaluated. As discussed in the BRDF Effect, Section 15.3.2, the IOPs needed for evaluation of the BRDF correction are parameterized in terms of the chlorophyll concentration Chl . If a measurement of Chl is made in conjunction with the target measurement, then that value of Chl can be used to evaluate the BRDF correction. However, chlorophyll measurements are not usually available. Operationally, the chlorophyll concentration is obtained via insertion of the satellite-derived $[L_w]_N^{ex}$ (or the corresponding reflectance $[\rho_w]_N^{ex}$) into a chlorophyll-retrieval algorithm. That is an iterative process because $[L_w]_N^{ex}$ is required to determine Chl , and Chl is required to determine $[L_w]_N^{ex}$. During the determination of gain factors, the target radiance can be used as input to the operational chlorophyll-retrieval algorithm and no iteration is required.

The final issue is the determination of the aerosol properties. This is a two-step process based on the “black-pixel” assumption to be described in Section 15.11.2. As seen in the Table 15.4 in that section, satellite sensors use two wavelengths in the near-infrared (NIR) for aerosol retrievals. Call the longer of these wavelengths λ_L and the shorter λ_S . (For VIIRS, the NIR bands are at the nominal wavelengths $\lambda_L = 862$ nm and $\lambda_S = 745$ nm.) During determination of the gain factors, it is first assumed that the water-leaving radiance at these two wavelengths is zero (the black-pixel assumption). This is usually a good approximation for the mid-ocean, oligotrophic waters used for vicarious calibration. It is further assumed that the instrument calibration is perfect for the λ_L band, in which case the gain factor for the longer NIR band is $g(\lambda_L) = 1$. The black-pixel assumption means that Eqs. (15.25) and (15.29) reduce to

$$L_t(\text{NIR}) = (L_r + L_A + t_{dv}L_{wc}) t_{gv}t_{gs}f_p \quad (15.31)$$

$$L_t^{tar}(\text{NIR}) = (L_r + L_A^{tar} + t_{dv}L_{wc}) t_{gv}t_{gs}f_p, \quad (15.32)$$

respectively, at the two NIR bands. Given the satellite-measured TOA radiances at the two NIR bands, Eq. (15.31) can be solved for L_A at the two NIR bands. The assumption that $g(\lambda_L) = 1$ means that $L_A^{tar}(\lambda_L) = L_A(\lambda_L)$. Thus $L_t^{tar}(\lambda_L)$ is determined via Eq. (15.32) evaluated at λ_L . The locations for match-ups are purposely chosen at times and locations where it is reasonable to assume that the aerosol type is stable and predictable over the image area, e.g. mid-ocean areas where the aerosols are predominately sea salt and water droplets. The aerosol model derived from the satellite measurements as described in the Black Pixels Section 15.11.2 can then be used along with the value of $L_t^{tar}(\lambda_L)$ to determine $L_t^{tar}(\lambda_S)$. Both L_t and L_t^{tar} are then known at the two NIR wavelengths, and the NIR gain factor at $g(\lambda_S)$ can be determined by Eq. (15.26).

Values of $g(\lambda_S)$ are computed for various times and locations during the lifetime of the mission. These values are averaged to determine the mean gain $\bar{g}(\lambda_S)$. The criteria for selection of suitable images for gain determination are quite rigorous and most candidate

pixels are eliminated from consideration because of glint, inhomogeneous water at the target location, or non-ideal atmospheric conditions. The details of the selection criteria and statistical determination of the mean gain factors are given in Franz et al. (2007). Experience shows that 20 to 40 match-ups are required for the determination of $\bar{g}(\lambda_S)$ values that are stable to within 0.1% of their long-term values. Once the NIR gains $g(\lambda_L) = 1$ and $\bar{g}(\lambda_S)$ have been determined for the given sensor, the extrapolation algorithm described in the Black Pixels section can be used to determine $L_A^{\text{tar}}(\lambda)$ at all wavelengths. Equation (15.32) then gives $L_t^{\text{tar}}(\lambda)$, and the gains at the remaining visible wavelengths are obtained from Eq. (15.26). Once determined, the gains are held fixed and applied as part of the operational atmospheric correction process.

For the SeaWiFS sensor the gains ranged from 1.0377 at 412 nm to 0.972 at 765 nm. A correction of 3 or 4% to the TOA radiance can correspond to a 30 or 40% correction to the water-leaving radiance because the water-leaving radiance is typically about 10% of the TOA radiance. Thus the determination of accurate gain factors is critical to the overall retrieval process. It must be remembered that a set of gains must be determined for each sensor and atmospheric correction algorithm. As improvements are made to the atmospheric correction algorithms described in this chapter, the gains must be recomputed. However, these recomputations can be made using the original target radiances. Gain recalculation is a part of the standard reprocessing of data sets.

15.6 Algorithm Introduction

The following seven sections discuss, one by one, the various corrections made to the TOA radiance during the atmospheric correction process as implemented (as of 2021) by the NASA Ocean Biology Processing Group for processing satellite imagery. The philosophy of these sections is simply to show “Here is what is done.” See the references for the historical development and scientific basis of the algorithms.

The first two sections show how to account for absorption and scattering by atmospheric gases. Sun glint and whitecap reflectances are then discussed, followed by the various aspects of correction for the effects of atmospheric aerosols. Two sections then discuss the sensor-specific corrections for spectral out-of-band response and for polarization.

For satellite imagery, the entire sequence of data processing beginning with the measurement of a TOA radiance and ending with the output of a geophysical product such as a global map of chlorophyll concentration is divided into a number of processing levels, which were defined in Section 14.1.3. The OBPG atmospheric correction process described here takes the data from Level 1b, TOA radiances that have been processed to radiance units, to Level 2, normalized water-leaving radiance $[L_w(\lambda)]_N^{\text{ex}}$ (or the equivalent reflectance $[\rho_w(\lambda)]_N^{\text{ex}}$). Figure 15.13 shows the sequence in which the various corrections are applied during the overall process.

It is important to keep in mind that there are severe computational constraints on how atmospheric correction is performed on an operational basis. The MODIS-Aqua sensor, for example, collects about 1.4 terabytes of data per day. The requirement to routinely process this amount of data (along with data from many other sensors) requires that various approximations be made in order to speed up the calculations. Some of the corrections require ancillary information such as sea level pressure, wind speed, and ozone concentration, which are not collected by ocean color sensors themselves. These ancillary data may

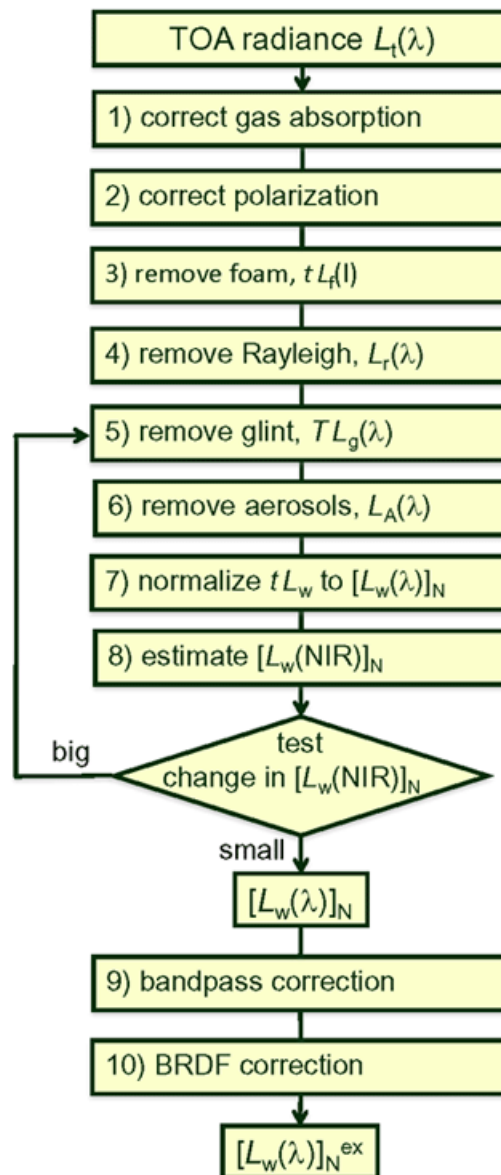


Figure 15.13: Flowchart of the atmospheric correction process.

be inaccurate or missing, in which case climatological values must be used. The quality of the ancillary information impacts the accuracy of the atmospheric correction. Table 15.3 shows some of the ancillary data and its sources as used by the various OBPG atmospheric correction algorithms.

The sea surface temperature and salinity are used to compute the water index of refraction and water backscattering coefficient. Operationally, pixels are masked before atmospheric correction for only a few conditions, namely the presence of land or clouds, and saturation of the measured radiance. An attempt is made to process all other pixels. A

Table 15.3: Ancillary data needed for atmospheric correction. Abbreviations: NCEP, National Centers for Environmental Prediction; OMI, Ozone Monitoring Instrument; TOMS, Total Ozone Mapping Spectrometer; SCIAMACHY, SCanning Imaging Absorption spectroMeter for Atmospheric CHartography; GOME, Global Ozone Monitoring Experiment; Reynolds/NCDC, Reynolds analysis, National Climate Data Center; NCEI, NOAA National Centers for Environmental Information (previously NODC); NSIDC, National Snow and Ice Data Center.

Data	Source	Use
atmospheric pressure	NCEP	Rayleigh correction
water vapor	NCEP	transmittance
wind speed	NCEP	Rayleigh, Sun glint, white caps
ozone concentration	OMI/TOMS	transmittance
NO ₂ concentration	SCIAMACHY/OMI/ GOME	transmittance
sea surface temperature	Reynolds/NCDC	seawater index of refraction and backscattering
sea surface salinity	NCEI World Ocean Atlas, Salinity Climatology	seawater index of refraction and backscattering
sea ice coverage	NSIDC	masking

separate mask is applied during atmospheric correction to pixels with too much Sun glint. Various flags are incorporated into Level 2 and 3 data after the atmospheric correction process described below. These identify pixels that may have various problems such as sea ice contamination, turbid water, bottom effects, or failed atmospheric correction. These flags are listed at [NASA Processing Flags](#). Other information on flags is given in [Patt et al. \(2003, Chapter 6\)](#). With the exception of Sun glint, applying masks and flags is not a part of the atmospheric correction process per se, so this topic is not discussed here.

15.7 Non-absorbing Gases

This section describes the Rayleigh corrections made for non-absorbing gases. The next section describes the more complicated problem of absorbing gases.

A radiative transfer numerical model is used to solve the vector (polarized) radiative transfer equation for non-absorbing atmospheric gases only. The radiative transfer model includes atmospheric multiple scattering, polarization, and sea surface roughness modeled analytically by a Cox-Munk slope distribution (Section 13.4) with an added analytical wave shadowing function. The Cox-Munk slope distribution as used is azimuthally isotropic (no dependence on wind direction); therefore only the relative angle between Sun and viewing direction matters. This greatly simplifies the Fourier decomposition seen in [Wang \(2002\)](#),

Eq. 3). The radiative transfer model is described in [Ahmad and Fraser \(1982\)](#).

15.7.1 Wind Speed and Surface Reflectance Effects

Background sky reflectance by the rough sea surface is accounted for as part of the Rayleigh correction. Some sensors (CZCS, SeaWiFS, OCTS) can be tilted to avoid looking at glint near the Sun's specular direction. Other sensors (MODIS, VIIRS, MERIS, OLI) do not tilt and therefore must account for specular reflection.

Wind speed and surface glint corrections are computed as described in [Wang \(2002\)](#): Run the numerical model to compute a look-up table (LUT) of TOA Fourier components $L_r^m(\lambda, \theta_s, \theta_v, \tau_{Ro}, U)$ (his Eq. 3) for the following conditions:

- The sensor wavelength bands (e.g., bands centered at 412, 443, 490, 510, 555, 670, 765, 865 for SeaWiFS). The radiative transfer model is run using band-averaged optical thicknesses (rather than running at high-wavelength resolution, and then averaging the L_r values over the band response functions to get the nominal band values of L_r for a particular sensor).
- 45 Sun zenith angles θ_s from 0 to 88 deg by 2 deg
- 41 viewing nadir angles θ_v from 0 to 84 deg by roughly 2 deg
- Rayleigh optical thickness τ_{Ro} for standard sea level atmospheric pressure $P_o = 1013.25$ hPa (1013.25 millibar).
- 8 wind speeds $U = 0, 1.9, 4.2, 7.5, 11.7, 16.9, 22.9$ and 30.0 m/s. (These wind speeds correspond to convenient spacing in the mean square slopes of the sea surface according to the Cox-Munk equation $mss = 0.00512U$: $mss = 0.0, 0.01, 0.02, 0.04, 0.06, \dots$ for $U = 0, 1.9, 4.2, 7.5, 11.7, \dots$) Linear interpolation is used for values between these wind speeds.

The Rayleigh optical thickness $\tau_{Ro}(\lambda)$ at 1 atmosphere of pressure, P_o (1013.25 hPa), temperature of 288.15K, and a CO₂ concentration of 360 ppm is given by [Bodhaine et al. \(1999, Eq. 30\)](#):

$$\tau_{Ro}(\lambda) = 0.0021520 \left(\frac{1.0455996 - 341.29061\lambda^{-2} - 0.90230850\lambda^2}{1.0 + 0.0027059889\lambda^{-2} - 85.968563\lambda^2} \right), \quad (15.33)$$

where λ is in micrometers. These values and the corresponding Rayleigh depolarization ratio $\rho(\lambda)$ are shown in Fig. 15.14. (At the scale of this figure, the Bodhaine values are almost indistinguishable from the values given by the formula of [Hansen and Travis \(1974\)](#) [Bodhaine et al.](#) (see 1999, Eq. 15), which was used in earlier calculations.)

The Rayleigh LUTs for L_r contain the I, Q and U Stokes vector components in reflectance units, as a function of wind speed and geometry. The Stokes vector V component for circular polarization is assumed to be zero. There is a separate LUT for each wavelength. During image correction, the wind speed U for a given pixel comes from NCEP 1 deg gridded data, interpolated to the image pixel.

15.7.2 Pressure Effects

The Rayleigh optical thickness at the time of the observation depends on the number of atmospheric gas molecules between the sea surface and the top of the atmosphere. The

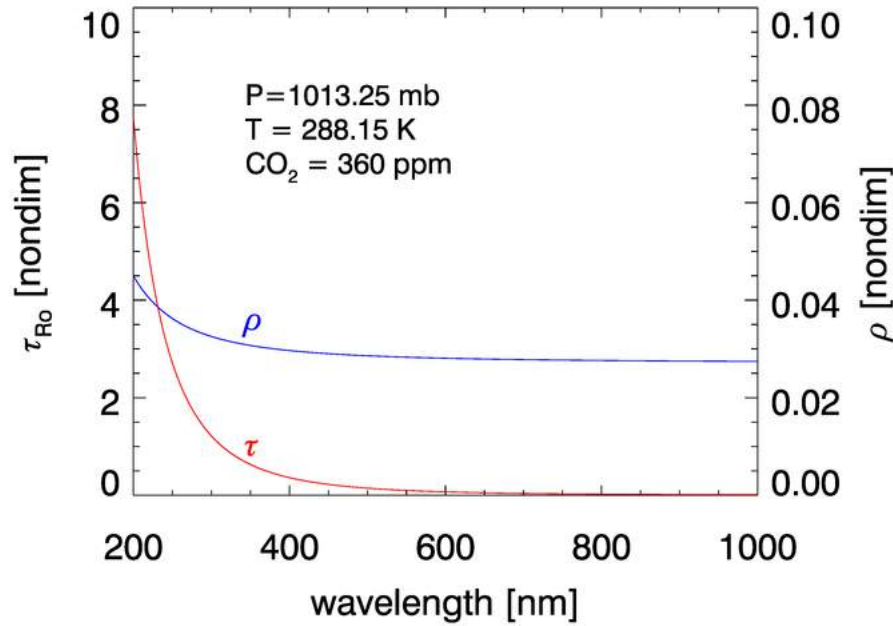


Figure 15.14: The Rayleigh optical thickness τ_a and depolarization ratio ρ based on Bodhaine et al. (1999). Plotted from NASA data file rayleigh_bodhaine.txt

number of molecules is directly proportional to the sea-level pressure P . Thus the Rayleigh optical thickness at any pressure P is given by

$$\tau_R(P, \lambda) = \frac{P}{P_o} \tau_{Ro}(P_o, \lambda).$$

The TOA L_R is then computed by Wang (2005, Eq. 5) and subsequent equations:

$$L_R[\tau_R(P, \lambda)] = L_R[\tau_R(P_o, \lambda)] \frac{1 - \exp[-C(\lambda, M)\tau_R(P, \lambda)M]}{1 - \exp[-C(\lambda, M)\tau_R(P_o, \lambda)M]}, \quad (15.34)$$

where

$$M = \frac{1}{\cos \theta_s} + \frac{1}{\cos \theta_v} \quad (15.35)$$

is the geometric air mass factor for the total path through the atmosphere. $C(\lambda, M)$ is a coefficient that is determined so that Eq. (15.34) gives the best fit to $L_R[\tau_R(P, \lambda)]$ as computed by an extremely accurate atmospheric radiative transfer model when run for values of sea level pressure $P \neq P_o$. Numerical simulations show that this coefficient can be modeled as

$$\begin{aligned} C(\lambda, M) &= a(\lambda) + b(\lambda) \ln(M) \\ a(\lambda) &= -0.6543 + 1.608\tau_R(P_o, \lambda) \\ b(\lambda) &= 0.8192 - 1.2541\tau_R(P_o, \lambda). \end{aligned}$$

15.8 Absorbing Gases

CO, N₂O, CH₄, and CO₂ have negligible absorption at the visible and NIR wavelengths relevant to ocean color remote sensing. However, O₂, O₃, NO₂, and H₂O have absorption bands in the visible and NIR. The O₂ and H₂O bands can be avoided by judicious choice of sensor bands, as shown in Fig. 15.15 for the MODIS bands. However, as seen in Figs. 15.16 and 15.17, O₃ and NO₂ have broad, concentration-dependent absorption bands that cannot be avoided. It is therefore necessary to account for absorption by these two gases.

The concentrations of absorbing gases are usually measured as column concentrations, i.e., the number of molecules per unit area, or as the equivalent in Dobson units. One Dobson unit refers to a layer of gas that would be 10 μm thick at standard temperature and pressure, or about 2.69×10^{16} molecules cm⁻². 1000 DU = 1 atm-cm; that is, 1000 DU is the number of molecules that would give a layer of gas 1 cm thick at a pressure of one atmosphere.

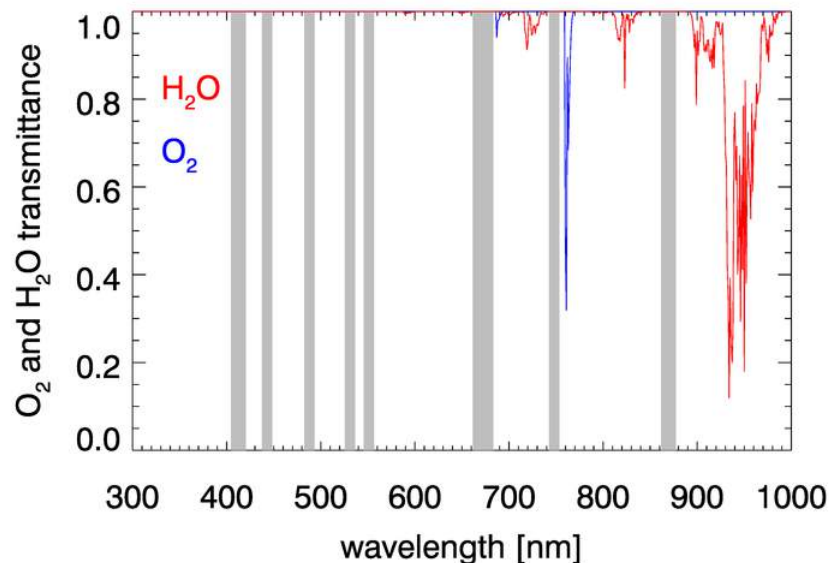


Figure 15.15: Transmittance by O₂ and H₂O for a moist tropical atmosphere. The resolution is 1 nm. The MODIS sensor bands are shaded in gray.

For optically thin absorbing gases that are high in the atmosphere (O₃ in particular), it is possible to correct for absorption using just the geometric air mass factor M computed by (15.35) because scattering is not significant. However, for gases near the surface (NO₂ in particular), multiple scattering by dense gases and aerosols is significant and increases the optical path length, hence the absorption. Thus M is not a good approximation for the total optical path length through an absorbing gas near sea level.

Note that an absorbing gas reduces the TOA radiance because light is lost to absorption. Correcting for this loss will increase the TOA radiance or reflectance, with the effect being greatest at blue wavelengths where multiple scattering is greatest.

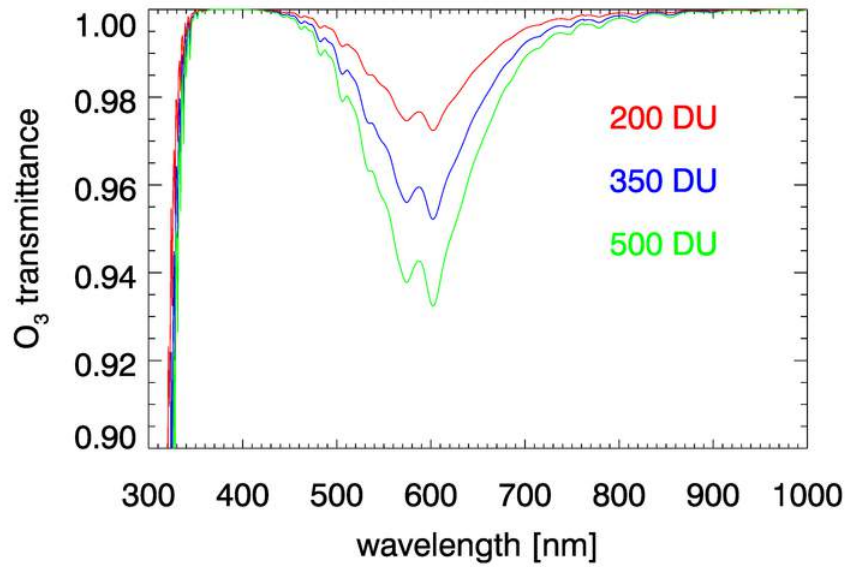


Figure 15.16: Transmittance by O_3 for 200, 350, and 500 Dobson units and a vertical path through the atmosphere.

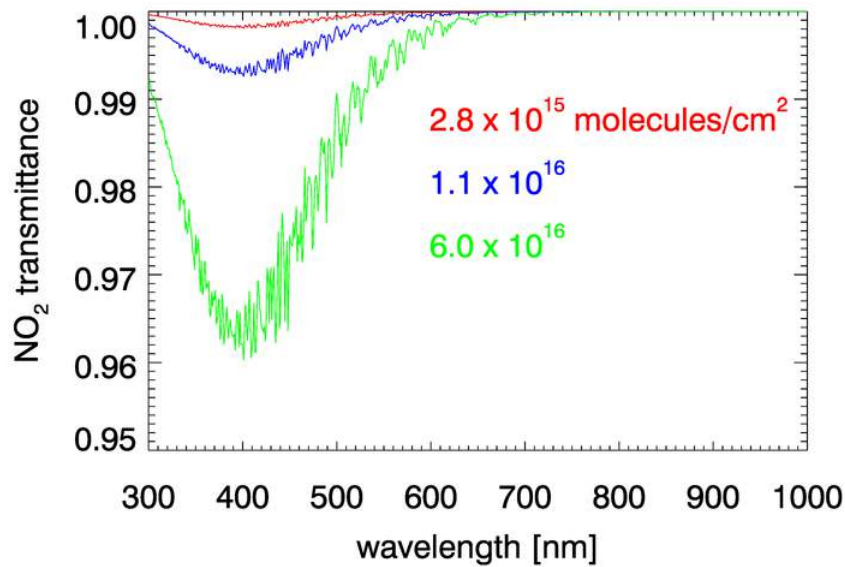


Figure 15.17: Transmittance by NO_2 for low (2.8×10^{15}), typical (1.1×10^{16}), and high (6.0×10^{16} molecules cm⁻²) concentrations of NO_2 and a vertical path through the atmosphere.

15.8.1 Absorption by Ozone

The diffuse transmission by ozone can be written as

$$\begin{aligned} t_{\text{O}_3} &= \exp \left[-\tau_{\text{O}_3} \left(\frac{1}{\cos \theta_s} + \frac{1}{\cos \theta_v} \right) \right] \\ &= \exp [-\tau_{\text{O}_3} M], \end{aligned} \quad (15.36)$$

where M is the geometric air mass factor defined in Eq. (15.35), and τ_{O_3} is the optical thickness of the ozone for a vertical path through the atmosphere. Scattering by ozone is negligible, but absorption is significant at some wavelengths. Thus τ_{O_3} is the optical thickness for absorption by ozone, which is given by

$$\tau_{\text{O}_3}(\lambda) = [\text{O}_3] k_{\text{O}_3}(\lambda), \quad (15.37)$$

where $[\text{O}_3]$ is the ozone concentration (column amount in molecules cm^{-2}), and k_{O_3} is the absorption cross section (in $\text{cm}^2 \text{molecule}^{-1}$). The ozone concentration $[\text{O}_3]$ for a given image pixel is obtained from the NASA OMI or TOMS sensors (Ozone Monitoring Instrument; Total Ozone Mapping Spectrometer, now replaced by OMI).

As was seen in for transmittance in Figs. 15.16 and 15.17, the absorption cross sections for gases like O_3 and, especially, NO_2 can vary with wavelength on a nanometer scale. To fully resolve the effects of such wavelength dependence on sensor signals, radiative transfer calculations would require computationally intensive “line-by-line” calculations followed by integration over the sensor bands. To avoid that computational expense, band-averaged values of the Rayleigh optical depth and absorption cross sections k_{O_3} and k_{NO_2} are computed for each sensor and tabulated. Radiative transfer calculations then use the band-averaged values with just one radiative transfer calculation done for each sensor band. These band-averaged values depend on the sensor even for the same nominal wavelength band (e.g, the 412 nm blue band) because of different band widths about the nominal center wavelength and different sensor response functions within a band. Figure 15.18 shows example band-averaged values for the VIIRS and MODIS Aqua sensors.

15.8.2 Absorption by Nitrogen Dioxide

Nitrogen dioxide NO_2 occurs both in the stratosphere and near the Earth’s surface. NO_2 in the lower atmosphere is generated primarily by human activities (automobiles, industry, fires), and the highest concentrations are near the earth’s surface in industrial areas. Numerical simulations show that failure to correct for absorption by NO_2 can give errors of approximately 1% in TOA radiances at blue wavelengths, which result in $\sim 10\%$ errors in retrieved water-leaving radiances (Ahmad et al., 2007).

The geometric air mass factor M of Eq. (15.35) and a simple atmospheric transmittance function like that of Eq. (15.36) are valid for ozone absorption corrections because scattering by ozone in the upper atmosphere is negligible. However, such functions may be not adequate for NO_2 correction calculations because of multiple scattering in the dense lower atmosphere. Further guidance for the form of the NO_2 correction comes from the observation that the water-leaving radiance L_w sees all NO_2 in the atmosphere, so the total NO_2 concentration N must be used to correct L_w for NO_2 absorption. However, the upwelling atmospheric path radiance L_{atm} , which is generated throughout the atmosphere,

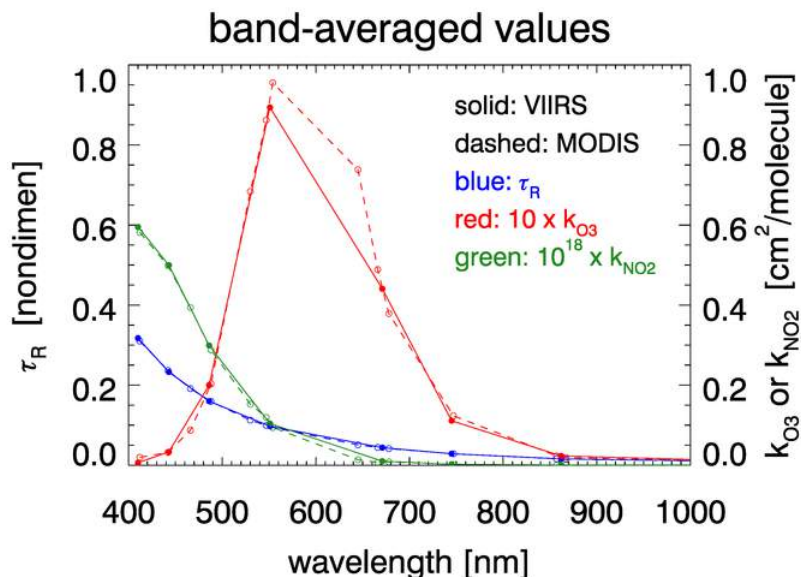


Figure 15.18: Band-averaged Rayleigh optical depth τ_R , ozone absorption cross section k_{O_3} , and NO_2 absorption cross section k_{NO_2} for the VIIRS and MODIS Aqua sensors. (Data from [NASA ocean color documents](#))

is not strongly influenced by the absorbing gas very near the surface. Extremely accurate numerical simulations show that in that case, N' , the NO_2 concentration between an altitude of 200 m and the TOA, can be used as a satisfactory measure of NO_2 concentration. This result leads to different corrections for the measured TOA path radiance and for the water-leaving radiance.

Let ρ_{obs} be the uncorrected, observed (measured) TOA reflectance, and let ρ_{corr} be the TOA reflectance corrected for NO_2 absorption effects. Comparison of numerical simulations and analytical approximations justifies a correction of the form

$$\rho_{\text{corr}} = \rho_{\text{obs}} \exp \left[\alpha N' \left(\frac{1}{\cos \theta_s} + \frac{1}{\cos \theta_v} \right) \right],$$

where N' is the NO_2 concentration between an altitude of 200 m and the TOA. This simple formula gives ρ_{corr} values that are within 0.15% of the values obtained by exact numerical simulations that account for the total column NO_2 concentration and multiple scattering.

The correction for water-leaving radiance proceeds as follows. The reflectances can be written as ([Ahmad et al., 2007](#), Eqs. 1 and 7)

$$\rho_t(\theta_s, \theta_v) = \rho_{\text{path}}(\theta_s, \theta_v) + t_3(\theta_v)t_d(\theta_s)\rho_w(\theta_s, \theta_v) + \text{glint and whitecap terms}$$

(omitting the arguments for wavelength, azimuthal angle, and wind speed). Here t_3 is the diffuse transmission along the viewing direction from the sea surface to the sensor, t_d is the diffuse transmission of downwelling solar irradiance, and ρ_w is the water-leaving reflectance at the sea surface. Consider now only the path and water-leaving terms, and omit the directional arguments for brevity. Then multiplying this equation by the exponential

correction factor for the observed TOA reflectance gives

$$\rho_t \exp [\alpha N' (\sec \theta_s + \sec \theta_v)] = \rho_{\text{path}} \exp [\alpha N' (\sec \theta_s + \sec \theta_v)] \\ + [t_3 \exp(\alpha N' \sec \theta)] [t_d \rho_w \exp(\alpha N' \sec \theta_s)].$$

Numerical simulations show that both the path reflectance and the diffuse transmission term $[t_3 \exp(\alpha N' \sec \theta)]$ are accurate to within 0.2% with this correction. However, the error in the water-leaving factor, $[t_d \rho_w \exp(\alpha N' \sec \theta_s)]$ is in error by 0.5 to 1.5%, which is unacceptably large. The reason for the greater error in this term is that it depends on the downwelling irradiance, which passes through the entire atmosphere and thus sees the total concentration N , not just the reduced concentration N' that is adequate for correction of the TOA path term. However, the error in this term also decreases to $\sim 0.2\%$ if N' is replaced by N . This error can be reduced still further by the following empirical procedure.

For bands where NO_2 absorption is significant (e.g., at 412 or 443 nm), the atmospheric correction is determined as always (without NO_2 correction) using the NIR bands. However, rather than subtract these terms from the corrected TOA reflectance, the atmospheric correction terms (including the Rayleigh reflectance) are reduced for NO_2 absorption by applying a factor of $\exp[-\alpha N' (\sec \theta_s + \sec \theta_v)]$. The computed path reflectance for NO_2 is then subtracted from the observed TOA reflectance to obtain $\Delta \rho_{\text{obs}} = t_3 t_d \rho_w$, which is the TOA value for water-leaving reflectance in the presence of NO_2 . The NO_2 -corrected value of the water-leaving reflectance is then obtained by multiplying this $\Delta \rho_{\text{obs}}$ by $\exp(\alpha N' \sec \theta) \exp(\alpha N \sec \theta_s)$, which gives

$$\exp(\alpha N' \sec \theta) \exp(\alpha N \sec \theta_s) \Delta \rho_{\text{obs}} = [t_3 \exp(\alpha N' \sec \theta)] [t_d \rho_w \exp(\alpha N \sec \theta_s)].$$

Note that $t_3 \exp(\alpha N' \sec \theta)$ is the NO_2 -corrected transmission term, and $\rho'_w = t_d \rho_w \exp(\alpha N \sec \theta_s)$ is the desired NO_2 -corrected water-leaving reflectance. This equation is then solved to obtain ρ'_w :

$$\rho'_w = \frac{\exp(\alpha N' \sec \theta) \exp(\alpha N \sec \theta_s) \Delta \rho_{\text{obs}}}{t_3 \exp(\alpha N' \sec \theta)}. \quad (15.38)$$

Note that the exponentials are increasing the magnitude of the water-leaving radiance compared to the no- NO_2 case, which accounts for the loss due to NO_2 absorption along the paths of the Sun's direct beam and the viewing direction. The term for the Sun's direct beam uses the full column NO_2 concentration N , whereas the viewing-path term uses the reduced concentration N' . This is an artifice that brings the analytical correction of Eq. (15.38) into close agreement with the exact numerical calculations. The absorption cross section α is a function of wavelength. Computations are done for α at 18 deg C, and then a temperature correction is made⁴.

15.9 Sun Glint

Both direct Sun glint and background sky reflectance are included in the radiative transfer calculations. However, the Sun glint contribution is removed from the look-up-tables (LUTs) so as to allow the historical approach of [Gordon and Wang \(1994a\)](#) to be used.

⁴[Ahmad et al. \(2007, Table 1\)](#) gives the band-averaged NO_2 absorption cross sections for SeaWiFS and MODIS bands, which are called σ in that table. This σ is same quantity as α in his Eq. (4) and in the equations of this chapter.

The current Rayleigh LUTs are wind-speed dependent, but the aerosol LUTs assume a flat ocean. Thus the LUTs include diffuse sky reflectance but not specular reflection. There is thus an explicit correction for Sun glint, but not for diffuse sky reflectance, which is accounted for as part of the Rayleigh correction.

Even for sensors (such as SeaWiFS) that are designed with tilt capabilities allowing them to be oriented so that they do not look at the Sun's glitter pattern, there can still be significant residual glint radiance reaching the sensor, especially near the edges of the obvious glint area. This is corrected as follows.

Recall again (15.3) of the Problem Formulation section:

$$L_t = L_R + [L_a + L_{Ra}] + TL_g + tL_{wc} + tL_w ,$$

where L_g is the direct Sun glint radiance. L_g is computed using the analytical Cox-Munk wind speed-wave slope distribution and the Sun and viewing geometry. Wang and Bailey (2001, Eq. 2) write the Sun glint radiance L_g in terms of a normalized Sun glint L_{GN} , which is defined by

$$L_g(\lambda) \triangleq F_o(\lambda)T(\theta_s, \lambda)L_{GN} .$$

L_{GN} is computed using an azimuthally symmetric analytical form of the Cox-Munk wind speed-wave slope distribution for the given Sun and viewing directions, and an incident irradiance of magnitude $F_o(\lambda) = 1 \text{ W m}^{-2} \text{ nm}^{-1}$. L_{GN} thus has the angular distribution of the surface-reflected radiance, but its units are 1/steradian. Note that L_{GN} is independent of wavelength.

During image processing, pixels with a value of $L_{GN} > 0.005 \text{ sr}^{-1}$ are masked out as having too much glint to be useful. Pixels with $L_{GN} \leq 0.005 \text{ sr}^{-1}$ have a glint correction applied before use.

For the glint correction, atmospheric attenuation occurs first along the path of the Sun's direct solar beam as the Sun's beam travels from the TOA to the sea surface; the associated transmittance is $T(\theta_s, \lambda)$. Attenuation then occurs along the viewing direction from the sea surface back to the TOA; this transmittance is $T(\theta_v, \lambda)$. These are both direct beam transmittances because only one particular path connects the Sun with a point on the sea surface that reflects the direct beam into the sensor (recall the left panel of Fig. 15.12). The total "two-path" transmittance is the product of the transmittances. The glint radiance correction, which is subtracted from L_t , is then (Wang and Bailey, 2001, Eqs. 4 and 5)

$$T(\theta_v, \lambda)L_g(\theta_v, \lambda) = F_o(\lambda)T(\theta_s, \lambda)T(\theta_v, \lambda)L_{GN} ,$$

where

$$T(\theta_s, \lambda)T(\theta_v, \lambda) = \exp \left\{ -[\tau_R(\lambda) + \tau_a(\lambda)] \left(\frac{1}{\cos \theta_s} + \frac{1}{\cos \theta_v} \right) \right\} , \quad (15.39)$$

and where $\tau_R(\lambda)$ and $\tau_a(\lambda)$ are the Rayleigh and aerosol optical thicknesses, respectively.

Wang and Bailey (2001, page 4792, left column) comment that the effects of ozone absorption have already been accounted for before this state of processing. This is now also true for NO_2 absorption.

Note that the glint correction cannot be computed unless the aerosol optical thickness (AOT) τ_a is known. The AOT is obtained in a two-step process. First, the measured $L_t(\lambda)$ and the wind speed U are used to get a first estimate $\tau_a^{(1)}(\lambda)$ of the AOT. This estimate

is obtained using the algorithms described on the aerosols section 15.11. This estimate is then used in Eq. (15.39), and the glint-corrected TOA radiance is then computed as

$$L'_t(\lambda) = L_t(\lambda) - F_o(\lambda)T(\theta_s, \lambda)T(\theta_v, \lambda)L_{GN}. \quad (15.40)$$

This gives the initial estimate $L_t^{(1)'}$ of $L'_t(\lambda)$. This value is then used again in the AOT algorithm to obtain the second estimate $\tau_a^{(2)}$ for the AOT. The second AOT estimate is then used again in Eqs. (15.39) and (15.40) to obtain an improved estimate $L_t^{(2)'}$. In practice, only two iterations give a satisfactory final estimate for the AOT, $\tau_a(\lambda) = \tau_a^{(2)}$, and thus for the glint-corrected TOA radiance. This final $\tau_a(\lambda)$ is then used to compute the aerosol contribution to the TOA radiance, as described on the aerosols section below.

15.10 Whitecaps

The contribution of white caps and foam to the TOA radiance depends on two factors: the reflectance of whitecaps per se and the fraction of the sea surface that is covered by whitecaps.

Following [Gordon and Wang \(1994b\)](#), the contribution of whitecaps and foam at the TOA is

$$t(\theta_v, \lambda)\rho_{wc}(\lambda) = [\rho_{wc}(\lambda)]_N t(\theta_s, \lambda) t(\theta_v, \lambda),$$

where $t(\theta_v, \lambda)$ is the diffuse atmospheric transmission in the viewing direction, $t(\theta_s, \lambda)$ is the diffuse transmission in the Sun's direction, and $[\rho_{wc}(\lambda)]_N$ is the non-dimensional normalized whitecap reflectance. $[\rho_{wc}(\lambda)]_N$ is defined in the same manner as was the normalized water-leaving reflectance $[\rho_w(\lambda)]_N$ in Eq. (15.7), namely

$$[\rho_{wc}]_N \triangleq \frac{\pi}{F_o} [L_{wc}]_N = \pi \frac{\left(\frac{R}{R_o}\right)^2 L_{wc}(\theta_s)}{F_o \cos \theta_s t(\theta_s)}, \quad (15.41)$$

where L_{wc} is the whitecap radiance. It is assumed that the whitecaps are Lambertian reflectors, so that (unlike for L_w) L_{wc} does not depend on direction θ_v, ϕ . This gives the interpretation ([Gordon and Wang, 1994b](#), page 7754) that “ ρ is the reflectance—the reflected irradiance divided by the incident irradiance—that a Lambertian target held horizontally at the TOA would have to have to produce the radiance L .” $[\rho_{wc}]_N$ can be interpreted as the average reflectance of the sea surface that results from whitecaps in the absence of atmospheric attenuation.

The effective whitecap irradiance reflectance is taken from [Koepke \(1984\)](#) to be 0.22 (albeit with $\pm 50\%$ error bars). This reflectance is independent of wavelength. This gives $[\rho_{wc}]_N = 0.22F_{wc}$, where F_{wc} is the fraction of the sea surface that is covered by whitecaps. The fractional coverage is taken from [Stramska and Petelski \(2003\)](#), who give two models for for F_{wc} :

$$F_{wc} = 5.0 \times 10^{-5} (U_{10} - 4.47)^3 \quad \text{for developed seas} \quad (15.42)$$

$$F_{wc} = 8.75 \times 10^{-5} (U_{10} - 6.33)^3 \quad \text{for undeveloped seas} \quad (15.43)$$

where U is the wind speed in m s^{-1} at 10 m. Formula (15.43) for undeveloped seas is used on the assumption that if the seas are well developed it is probably stormy, hence

cloudy, so that remote sensing is not possible. The blue curve in Fig. (15.44) shows F_{wc} for undeveloped seas.

The final model for $[\rho_{\text{wc}}]_{\text{N}}$ is then taken to be

$$\begin{aligned} [\rho_{\text{wc}}]_{\text{N}}(\lambda) &= a_{\text{wc}}(\lambda) \times 0.22 \times F_{\text{wc}} \\ &= a_{\text{wc}}(\lambda) \times 1.925 \times 10^{-5} (U_{10} - 6.33)^3. \end{aligned} \quad (15.44)$$

A whitecap correction is applied for wind speeds in the range $6.33 \leq U_{10} \leq 12 \text{ m s}^{-1}$. The factor $a_{\text{wc}}(\lambda)$ is a normalized whitecap reflectance that describes the decrease in reflectance at red and NIR wavelengths. This factor is taken from Frouin et al. (1996, Figs. 3 and 4); the values are

$\lambda =$	412	443	490	510	555	670	765	865
$a_{\text{wc}} =$	1.0	1.0	1.0	1.0	1.0	0.889	0.760	0.645

Linear interpolation is used as needed between these values. Figure 15.19 shows the whitecap reflectance as given by Eq. (15.44) when $a_{\text{wc}} = 1$.

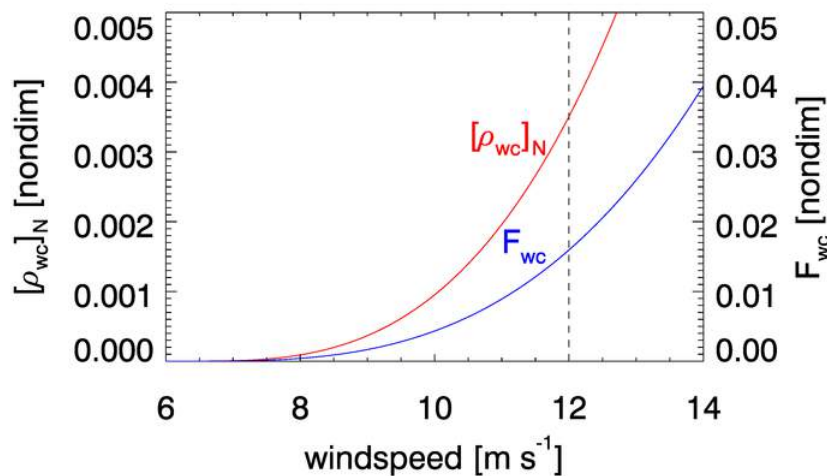


Figure 15.19: Whitecap normalized reflectance as given by Eq. (15.44) with $a_{\text{wc}} = 1$, and the fraction of sea surface whitecap coverage F_{wc} as given by Eq. (15.43).

15.11 Aerosols

15.11.1 Aerosol Properties

Aerosols are solid or liquid particles that are much larger than gas molecules but small enough to remain suspended in the atmosphere for periods of hours to days or longer. Typical sizes are 0.1 to 10 μm . An aerosol's optical properties are determined by its composition, usually parameterized via its complex index of refraction, and its particle size distribution (PSD).

For the purposes of atmospheric correction, aerosol particle size distributions are modeled as a sum of “fine” (small; radii less than roughly $1\ \mu\text{m}$) and “coarse” (large; radii greater than roughly $1\ \mu\text{m}$) particles, with a log-normal distribution for each. The log-normal distribution (8.45) is discussed in Section 8.12.4. That section also discusses the relations between particle number, area, and volume size distributions. The cumulative volume distribution is then (Ahmad et al., 2010a)

$$\frac{dV(r)}{d\ln r} = \sum_{i=1}^2 \frac{V_{oi}}{\sqrt{2\pi}\sigma_i} \exp \left[- \left(\frac{\ln r - \ln r_{voi}}{\sqrt{2}\sigma_i} \right)^2 \right].$$

Here $V(r)$ is the volume of particles per volume of space with size less than or equal to r ; $V(r)$ is typically specified as $\mu\text{m}^3\ \text{cm}^{-3}$. r_{voi} is the volume geometric mean radius, and σ_i is geometric standard deviation for class i . The integral of $dV(r)/d\ln r$ over all sizes $r = 0$ to ∞ (i.e., $\ln r$ from $-\infty$ to ∞) gives $V(\infty) = V_{oi}$. Thus V_{oi} is the total volume of particles of class i per volume of space.

A similar equation holds for the cumulative number distribution $dN(r)/d\ln r$, where $N(r)$ is the number of particles per volume of space with size less than or equal to r . The corresponding parameters r_{noi} and N_{oi} can be obtained from r_{voi} and V_{oi} ; see the equations in Ahmad et al. (2010a). The particle size distribution (PSD) is given by

$$n(r) = \frac{dN(r)}{dr} = \frac{1}{r} \frac{dN(r)}{d\ln r},$$

where $n(r)dr$ is the number of particles per unit volume in size range r to $r + dr$. The units for $n(r)$ are usually expressed as particles $\text{m}^{-3}\ \mu\text{m}^{-1}$.

Figure 15.20 illustrates shapes of the volume $V(r)$, number $N(r)$, and particle size $n(r)$ distributions for an open ocean aerosol, computed using the parameter values of Ahmad et al. (2010a, Table 2). The distributions of the fine aerosols are given by the green lines. The blue lines are the coarse aerosols, and the red lines are the sums. The two roughly comparable distributions of left panel of the figure show that the fine and coarse aerosols each contribute a significant amount of the total particle volume. In the present example the fines are 25.7% of the total volume and the coarse particles are 74.3% of the total volume. The middle panel shows that the fines dominate the number of particles; in the present case there are 477 times as many fine particles as coarse. The right panel shows the PSD. The black dashed line shows the -4 slope of a Junge distribution for comparison.

The radius parameter r_{voi} and index of refraction both depend on the aerosol type (dust, sea salt, soot, etc.) and on the relative humidity Rh . The index of refraction generally depends on wavelength. Figure 15.21 shows the effect of relative humidity on cumulative volume and particle size distributions for an open-ocean aerosol (parameter values from Ahmad et al. (2010a, Table 4)). Note that as Rh increases, the particles absorb more water and increase in size, so the distributions shift to the right. The shape of the distribution also changes with Rh .

As modeled in Ahmad et al. (2010a), the fine particles are generally of “continental” origin and include both dust and soot. The fine particles are sometimes absorbing. The coarse particles are of “oceanic” origin and are assumed to be non-absorbing sea salts. The tables⁵ in Ahmad et al. (2010a) give the PSD parameters and indices of refraction for different aerosol types (dust, sea salt, soot, etc.) and relative humidities.

⁵Table 1 of Ahmad et al. (2010a) has errors. The corrected table is given in Ahmad et al. (2010b)

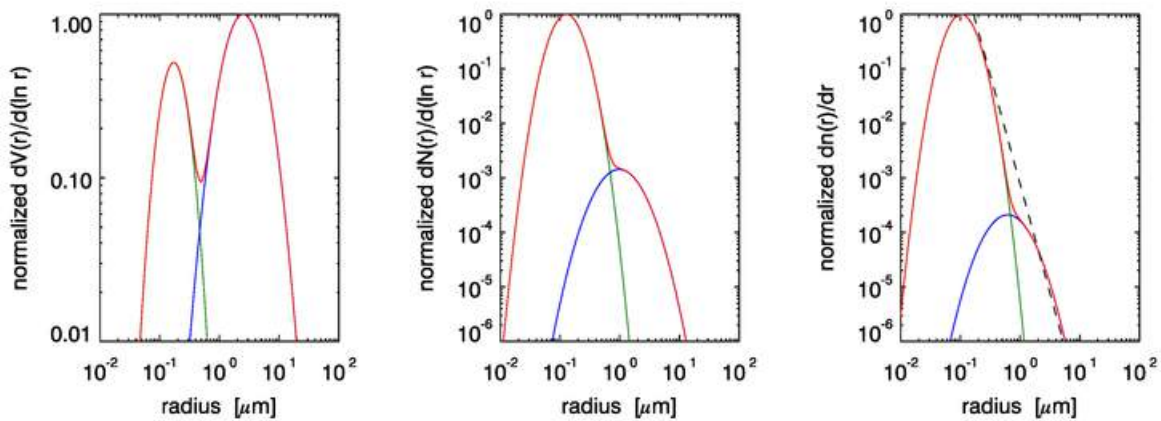


Figure 15.20: Illustration of aerosol volume, number, and particle size distributions. The parameter values are for an open-ocean aerosol. The green curve is fine particles; blue is coarse particles; red is the sum of fine and coarse particles. Each total curve is normalized to 1 at its maximum value.

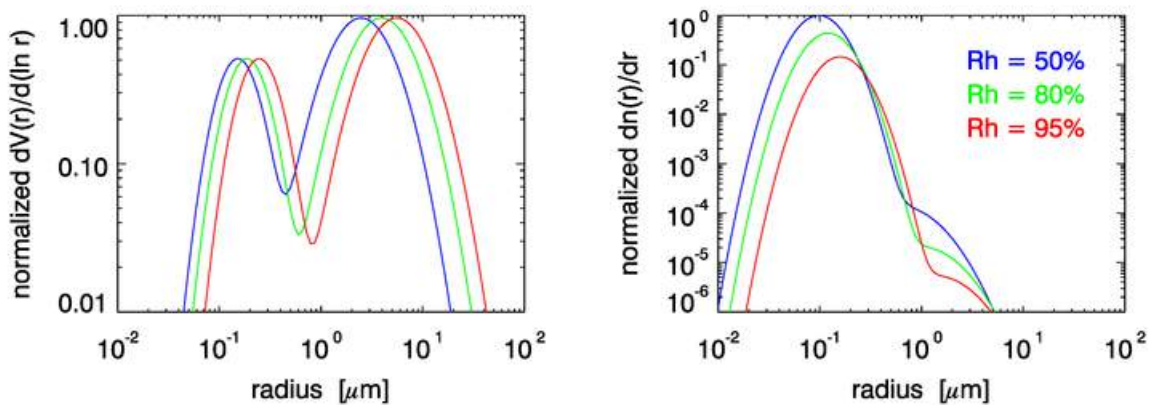


Figure 15.21: Effect of relative humidity Rh on the cumulative volume and particle size distributions for a typical open-ocean aerosol. The curves in each panel are normalized to 1 at the maximum value of the three curves.

An aerosol's physical properties determine its optical properties, namely its mass-, number-, or volume-specific absorption $a^*(\lambda)$ and scattering $b^*(\lambda)$ coefficients and scattering phase function $\tilde{\beta}(\psi, \lambda)$, where ψ is the scattering angle. If the particles are homogeneous spheres, Mie theory can be used to compute the optical properties from the physical properties. This is often done, although the assumption of homogeneous spherical particles may or may not be valid in a given situation. In any case, once the $a^*(\lambda)$ and scattering $b^*(\lambda)$ coefficients are known, then given the concentration profile $Conc(z)$ as a function of altitude z , the extinction coefficient $c(z, \lambda) = Conc(z)[a^*(\lambda) + b^*(\lambda)]$ can be computed.

The aerosol optical thickness or aerosol optical depth is then given by

$$\tau_a(\lambda) = \int_{z_0}^{TOA} c(z, \lambda) dz,$$

where z_0 is the surface elevation. (Generally $z_0 = 0$ for mean sea level, but may also be the elevation of a lake, for example.)

For all else held fixed, the aerosol optical thickness at wavelength λ is approximately related to the value at a reference wavelength λ_o by

$$\frac{\tau_a(\lambda)}{\tau_a(\lambda_o)} = \left(\frac{\lambda_o}{\lambda} \right)^\alpha \quad (15.45)$$

The parameter α is known as the Ångström exponent or Ångström coefficient. Smaller (larger) particles generally have a larger (smaller) Ångström exponent.

The single scattering albedo ω_o defined by

$$\omega_o(\lambda) = \frac{b^*(\lambda)}{c^*(\lambda)}$$

is also of use in modeling the optical effects of aerosols on the radiance distribution.

Ahmad et al. (2010a) constructed look-up tables (LUTs) for 10 aerosol types and 8 relative humidities, for a total of 80 aerosol tables. The fine fraction was a mixture of 99.5% dustlike and 0.5% soot particles (not modeled by Shettle and Fenn) for all 10 aerosol types, which gives good agreement on average with AERONET measurements of aerosol optical properties. The ten aerosol models have different weights of fine and coarse particles, but the effective radius

$$r_{\text{eff}} = \frac{\int r^3 n(r) dr}{\int r^2 n(r) dr}$$

and mean radius

$$\mu_o = \frac{\int r n(r) dr}{\int n(r) dr}$$

are the same. The aerosol types were then defined by letting the fine-to-coarse fraction vary from 0 to 1. For each aerosol type, relative humidities of $Rh = 30, 50, 70, 75, 80, 85, 90,$ and 95% were used. The actual aerosol LUTs contain the components of the model from which to derive single-scattering aerosol reflectance ratios in each band (relative to any reference band), plus a set of quadratic coefficients relating single to multiple scattering (as in Gordon and Wang (1994a), plus a separate table of Rayleigh-aerosol diffuse transmittance coefficients of the form $t = A \exp(-B\tau_a)$ (recall Eq. (15.23) of the Atmospheric Transmittances section). Mie is theory used to compute the aerosol phase functions for use in the radiative transfer model.

15.11.2 Black-pixel Calculations

The algorithm developed by Gordon and Wang (1994a) is used, although the original aerosol models and LUTs have been updated as described in Ahmad et al. (2010a), and model selection is now partitioned by relative humidity (Rh). (Gordon and Wang ignored the glint term since they were considering SeaWiFS, whose viewing direction was chosen to avoid direct Sun glint.)

Beginning with Eq. (15.17) of the Normalized Reflectances section,

$$\rho_t = \rho_R + [\rho_a + \rho_{Ra}] + T\rho_g + t\rho_{wc} + t\rho_w, \quad (15.46)$$

The basic theory in Gordon and Wang (1994a) is developed using single-scattering theory, in which case the ρ_{Ra} term is zero because there is no multiple scattering. Multiple-scattering effects are then added via numerical models using the guidance of the single-scattering theory.

Assume that the corrections for Rayleigh, whitecaps, O₃, NO₂, and Sun glint have all been made. Then the left hand side of

$$\rho_t - \rho_R - T\rho_g - t\rho_{wc} = [\rho_a + \rho_{Ra}] + t\rho_w \quad (15.47)$$

or

$$\rho_{Aw} \triangleq \rho_A + t\rho_w$$

is known. Here ρ_{Aw} is just convenient shorthand for the measured TOA reflectance ρ_t with the Rayleigh and other effects removed. The next task is to compute ρ_A , the combined aerosol and aerosol-Rayleigh reflectance, and move it to the left hand side, after which the desired ρ_w will be known.

Low-chlorophyll, Case 1 waters have negligible water-leaving radiance at near-infrared (NIR) wavelengths, i.e. beyond roughly 700 nm. For such waters, it can be assumed that $\rho_w(\lambda > 700 \text{ nm}) = 0$, which is known as the “black-pixel” assumption. Let λ_1 and λ_2 be two NIR wavelengths, with $\lambda_1 < \lambda_2$. At these two wavelengths, the TOA normalized reflectance (corrected as shown in Eq. (15.47)) is due entirely to atmospheric path radiance: $\rho_{Aw}(\lambda_i) = \rho_A(\lambda_i), i = 1, 2$. Table 15.4 shows the λ_1 and λ_2 bands for several sensors.

Band Label	Wavelengths [nm]	Nominal Wavelength [nm]
SeaWiFS		
7	745-785	$\lambda_1 = 765$
8	845-855	$\lambda_2 = 865$
MODIS		
15	743-753	$\lambda_1 = 748$
16	862-877	$\lambda_2 = 869$
VIIRS		
M6	739-754	$\lambda_1 = 745$
M7	846-885	$\lambda_2 = 862$

Table 15.4: NIR bands used for aerosol correction.

Now consider the ratio

$$\epsilon(\lambda_1, \lambda_2) \triangleq \frac{\rho_A(\lambda_1)}{\rho_A(\lambda_2)} \stackrel{\text{Black Pixel}}{=} \frac{\rho_{Aw}(\lambda_1)}{\rho_{Aw}(\lambda_2)} \quad (15.48)$$

The quantity $\epsilon(\lambda_1, \lambda_2)$, and more generally the quantity $\epsilon(\lambda, \lambda_2)$ for any $\lambda < \lambda_2$, depends on the aerosol type, which is determined by the particle type, PSD, and relative humidity. Figure 15.22 shows the dependence of $\epsilon(\lambda, \lambda_2 = 865)$ for the ten aerosol models, for one particular set of solar zenith angle, viewing direction, and relative humidity.

For the given image pixel being corrected, the process is as follows:

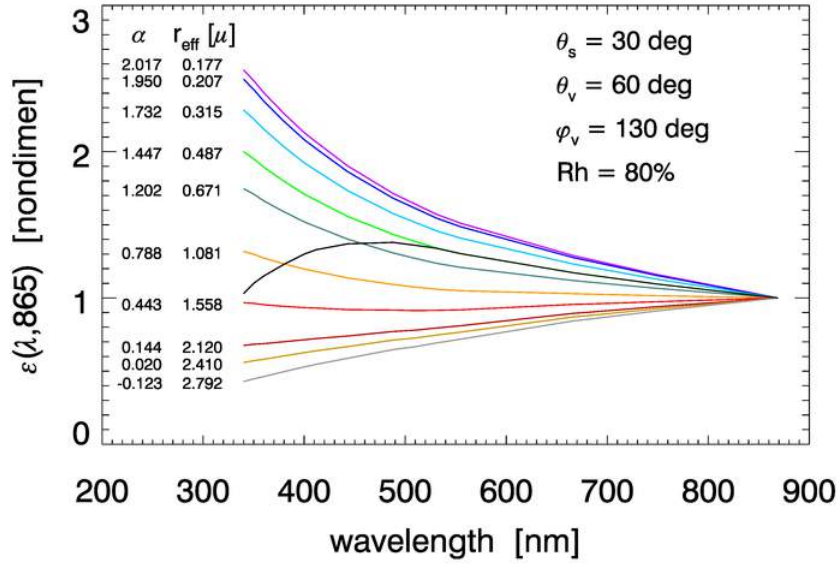


Figure 15.22: Dependence of $\epsilon(\lambda, \lambda_2 = 865)$ on aerosol model. The aerosol model is determined by the value of the Ångström exponent α and the effective particle size r_{eff} . All curves are for a particular Sun zenith angle θ_s , viewing direction θ_v, ϕ_v , and relative humidity Rh , as shown. The black curve shows the qualitative behavior of $\epsilon(\lambda, \lambda_2 = 865)$ for a blue-absorbing aerosol.

- The relative humidity Rh is taken from NCEP.
- For each bounding Rh value in the family of 8 Rh values in the database, the corresponding family of 10 aerosol types in the database is then searched to find the two aerosol types whose precomputed values of $\epsilon(\lambda_1, \lambda_2)$, call them ϵ_{low} and ϵ_{high} , bracket the measured value of $\epsilon(\lambda_1, \lambda_2)$ for the given Sun and viewing geometry.
- This selects two of the curves like those of Fig. 15.22, for which the corresponding $\epsilon(\lambda, \lambda_2)$ and τ_a values have been precomputed and stored in the aerosol LUT.
- It is assumed that the difference in the precomputed $\epsilon(\lambda, \lambda_2)$ is at all wavelengths in the same proportion as the measured $\epsilon(\lambda_1, \lambda_2)$ is to the bracketing values at the NIR reference wavelengths. Thus let

$$\Delta = \frac{\epsilon(\lambda_1, \lambda_2) - \epsilon_{\text{low}}(\lambda_1, \lambda_2)}{\epsilon_{\text{high}}(\lambda_1, \lambda_2) - \epsilon_{\text{low}}(\lambda_1, \lambda_2)} \quad (15.49)$$

- The aerosol reflectance at all wavelengths is then computed from the measured $\rho_{\text{Aw}}(\lambda_2)$ and the tabulated ϵ_{low} and ϵ_{high} values using

$$\rho_A(\lambda) = [(1 - \Delta)\epsilon_{\text{low}}(\lambda, \lambda_2) + \Delta\epsilon_{\text{high}}(\lambda, \lambda_2)] \rho_{\text{Aw}}(\lambda_2) \quad (15.50)$$

Now that $\rho_A(\lambda)$ is known, Eq. (15.47) gives

$$t\rho_w = \rho_{\text{Aw}} - \rho_A = \rho_t - \rho_R - T\rho_g - t\rho_{wc} - [\rho_a + \rho_{Ra}] \quad (15.51)$$

Recall that t is the diffuse transmission in the viewing direction, and that $t\rho_w$ is the contribution of water-leaving radiance (in reflectance form) at the TOA. The desired ρ_w at the sea surface is thus obtained from

$$\rho_w = \frac{\rho_{Aw} - \rho_A}{t}. \quad (15.52)$$

The aerosol optical depth is computed at all wavelengths using the values of $\tau(\lambda_2)$ and the Ångström exponent in Eq. (15.45).

This technique rests on two main assumptions:

- The water-leaving radiance is negligible at the NIR reference wavelengths. This is valid only for optically deep, Case 1 waters, with a chlorophyll concentration of 0.3 mg m^{-3} or less. Waters containing higher chlorophyll concentrations or mineral particles will violate this assumption. Figure 15.23 shows an example of very turbid water for which the remote-sensing reflectances at NIR wavelengths is not negligible.
- The aerosols are not strongly absorbing. Some mineral aerosols absorb strongly at blue wavelengths but not in the NIR. Their $\epsilon(\lambda, \lambda_2)$ functions look like the black curve in Fig. 15.22. Thus their presence cannot be detected from the NIR TOA radiances.

Note that these are unrelated assumptions: the water can have non-zero NIR reflectance and the atmosphere can have non-absorbing aerosols, or there can be zero NIR reflectance but absorbing aerosols. If the water-leaving radiance is not zero at the λ_2 reference wavelength, then the ρ_w contribution to ρ_{Aw} will be interpreted as a larger aerosol concentration. This leads to over-correction for the aerosol, i.e., subtracting too much ρ_A from ρ_{TOA} . The resulting ρ_w is then too small, and can even be negative at blue wavelengths. Likewise, if the aerosol is blue-absorbing, over-correction occurs at blue wavelengths and, again, ρ_w is too small or even negative at blue wavelengths.

15.11.3 Non-black-pixel Calculations

Many ways to treat non-black pixels have been developed; the method currently implemented by the OBPG is described in Bailey et al. (2010). This algorithm works as follows.

It is necessary to estimate $R_{rs}(\lambda)$ (or equivalently $\rho_w(\lambda)$) at the NIR reference wavelengths so that the non-zero water-leaving radiance can be removed from the TOA signal, leaving only the aerosol reflectance as the contribution to ρ_{Aw} , from which the aerosol type (i.e., $\epsilon(\lambda, \lambda_2)$) can be determined. However, $R_{rs}(\text{NIR})$ can't be estimated until the aerosol contribution is removed. Thus an iterative solution must be used to obtain both $R_{rs}(\text{NIR})$ and the aerosol type.

The remote-sensing reflectance can be written

$$R_{rs}(\lambda) = \frac{f(\lambda)}{Q(\lambda)} \frac{b_b(\lambda)}{a(\lambda) + b_b(\lambda)}. \quad (15.53)$$

As discussed in detail on the Normalized Reflectances section, the f/Q factor describes the angular distribution of the water-leaving radiance, i.e., the BRDF of the ocean. This factor depends on the Sun and sky radiance distribution (parameterized by the solar zenith angle θ_s and AOT τ_a), water-column IOPs (parameterized by the chlorophyll concentration Chl

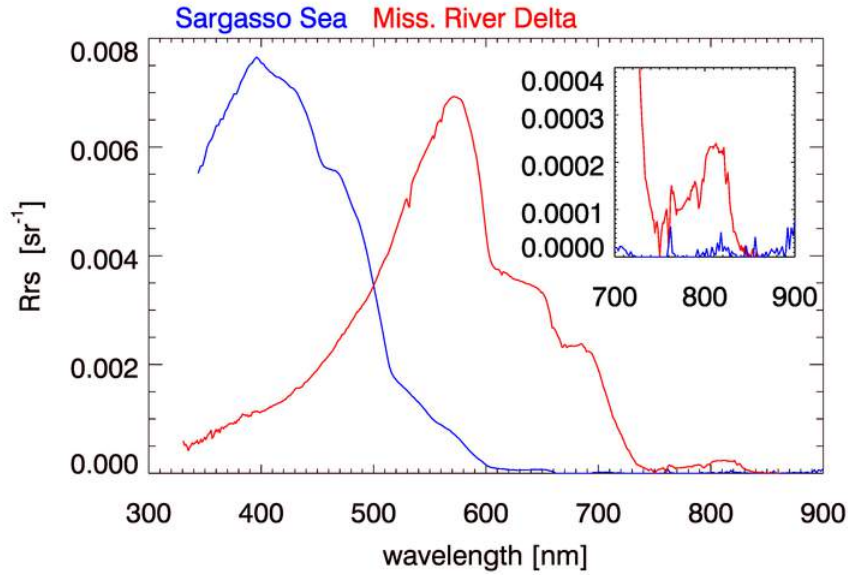


Figure 15.23: Examples of R_{rs} for Case 1 and Case 2 waters. R_{rs} is not zero in the NIR for the Mississippi River Delta water.

in Case 1 waters), sea state (wind speed U), viewing direction (θ_v, ϕ_v), and wavelength. It has been extensively studied and numerically modeled by Morel et al. (2002), who present tabulated values as functions of $\theta_s, \tau_a, \theta_v, \phi_v, Chl, U$, and λ . Although the Morel et al. (2002) f/Q table was generated for Case 1 waters, analysis shows (Bailey et al., 2010) that it is often adequate for Case 2 waters as well. The f/Q factor is thus considered known for the present calculations.

The iterative correction for waters where the black-pixel assumption cannot be made has the following steps:

1. Assume that $R_{rs}(765)$ and $R_{rs}(865)$ are both 0, i.e. make the black-pixel assumption for both NIR reference bands.
2. Complete the atmospheric correction process as described for black pixels. This gives the initial estimate of $\rho_w(\lambda)$, or equivalently $R_{rs}(\lambda)$.
3. Use $R_{rs}(443)$ and $R_{rs}(555)$ from the initial estimate of $R_{rs}(\lambda)$ to get η by the empirical relationship (Lee et al., 2010, Eq. 8), (Bailey et al., 2010, Eq. 3)

$$\eta = 2 \left[1 - 1.2 \exp \left(0.9 \frac{R_{rs}(443)}{R_{rs}(555)} \right) \right]. \quad (15.54)$$

4. Use the initial $R_{rs}(\lambda)$ to get an initial estimate of the chlorophyll concentration Chl . The particular algorithm used to obtain Chl from $R_{rs}(\lambda)$ depends on the sensor.
5. Use this Chl to obtain $a(670)$ via the empirical relationship (Bailey et al., 2010, Eq. 4)

$$a(670) = \exp[0.9389 \ln(Chl) - 3.7589] + a_w(670). \quad (15.55)$$

where the $a_w(670) = 0.439 \text{ m}^{-1}$ is the absorption by pure water.

6. Use $a(670)$ and $R_{rs}(670)$ in Eq. (15.53) to solve for $b_b(670) = b_{bw}(670) + b_{bp}(670)$, where $b_{bw}(670) = 4.26 \times 10^{-4} \text{ m}^{-1}$ is the backscatter coefficient for pure sea water.
7. Use η from Eq. (15.54) and (Bailey et al., 2010, Eqs. 2b and 3) to compute $b_b(765)$:

$$b_b(765) = b_{bw}(765) + b_{pb}(670) \left(\frac{670}{765} \right)^\eta \quad (15.56)$$

where $b_{bw}(765) = 2.38 \times 10^{-4} \text{ m}^{-1}$. $b_b(865)$ is computed in the same manner using $b_{bw}(865) = 1.41 \times 10^{-4} \text{ m}^{-1}$.

8. Use this $b_b(765)$ and $a(765) = a_w(765) = 2.85 \text{ m}^{-1}$ to get $R_{rs}(765)$ from Eq. (15.53). Similarly, compute $R_{rs}(865)$ using $b_b(865)$ and $a_w(865) = 4.61 \text{ m}^{-1}$.
9. Use the new, non-zero value of $R_{rs}(765)$ (i.e. $\rho_w(765)$) to remove the non-zero $\rho_w(765)$ contribution to $\rho_t(765)$. Do the same calculation for 865 nm.
10. Return to Step 2 and repeat the atmospheric correction using the black-pixel algorithm. This will give a new (hopefully better) estimate of $R_{rs}(\lambda)$, thus an new estimate of the other parameters, and finally new estimates of $R_{rs}(765)$ and $R_{rs}(865)$ at Step 8. After using the new values of $\rho_w(765)$ and $\rho_w(865)$ to correct for the non-zero water contribution to $\rho_t(765)$ and $\rho_t(865)$, return to Step 2 for a new iteration. Continue iterating until the change in $R_{rs}(765)$ from one iteration to the next is less than 2%, which typically takes 2-4 iterations, or when 10 iterations have been made.

If this iteration process fails to converge within 10 cycles, then re-initialize with $\rho_a(\text{NIR}) = 0$, i.e., set the NIR aerosol contribution to zero. This implies that all of the NIR reflectance is due to the water (after correction for Rayleigh and Sun glint). Repeat the iteration until convergence is reached. If convergence is still not reached, do one more calculation with $\rho_a(\text{NIR}) = 0$ and flag the pixel as “atmospheric correction warning.” However, even in this case, the retrieval may still be useful.

The above iteration is not done if the initial *Chl* estimate is less than 0.3 mg m^{-3} , and it is always done if the initial *Chl* estimate is greater than 0.7 mg m^{-3} . To prevent discontinuities in the final results, the $R_{rs}(765)$ estimate is linearly weighted from 0 to 1 for $0.3 \leq \text{Chl} \leq 0.7$. Figure 15.24 shows the regions of the ocean where the non-black-pixel algorithm is likely to be applied.

Finally, rather than the exact NIR reference wavelengths of $\lambda_1 = 765$ and $\lambda_2 = 865$ shown above, in practice band-averaged IOP values are used for the particular sensor. Thus for VIIRS-NPP, the reference bands are centered at 745 and 862 nm, with the band-averaged $a_w(745) = 2.806 \text{ m}^{-1}$, and so on. Band averaged IOPs for different sensors are given at [NASA ocean color documents](#).

15.11.4 Strongly Absorbing Aerosols

The aerosol models of Ahmad et al. (2010a) discussed in Section 15.11.2 include a small fraction of absorbing soot for the fine part of the aerosol size distribution. These aerosol models are used for routine correction for aerosols. These models, however, cannot account for the presence of strongly absorbing aerosols, which frequently occur in coastal regions downwind from continents, and even in mid-ocean regions because winds can transport fine dust particles long distances.

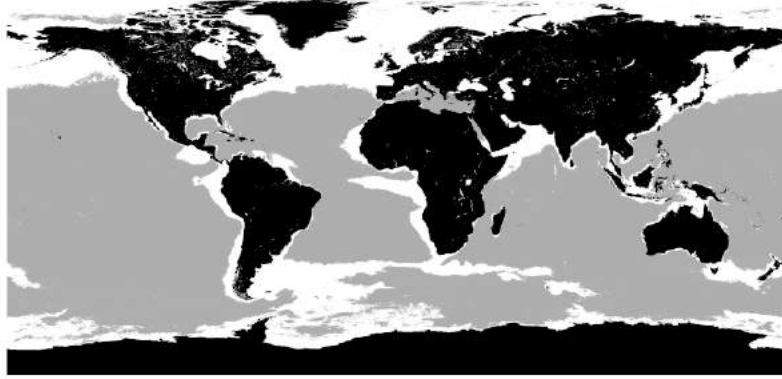


Figure 15.24: Regions of the ocean where the non-black-pixel correction is likely to be applied, based on SeaWiFS climatology. Land is black, gray is $Chl < 0.3 \text{ mg m}^{-3}$ (where the correction is not applied), and white is $Chl > 0.3 \text{ mg m}^{-3}$ (where the correction is applied). (Bailey et al., 2010, Fig. 3), used by permission.

Strongly absorbing aerosols (especially at blue wavelengths) have been a topic of much research over the years (e.g., Gordon et al., 1997b). However, as was illustrated in Fig. 15.22, there is no reliable way to detect the presence of absorbing aerosols from the NIR bands of Table 15.4 during the atmospheric correction process. Therefore, all pixels are processed using the algorithms for non- or weakly absorbing aerosols as described in the previous sections.

15.12 Out-of-band Response

Figure 15.25 shows the sensor response functions (SRF) of the MODIS-Aqua bands used for ocean color remote sensing; each response function is normalized to 1 at its maximum. Visually, in this plot with a linear ordinate axis, the bands appear to be well defined and narrow, with full-width, half-maximum (FWHM; the wavelengths at which the function is one-half of its maximum) widths of 10 to 15 nm. However, when plotted with a logarithmic ordinate as in Fig. 15.26, it is seen that there is significant “out-of-band” (OOB) sensitivity; i.e., a non-zero response outside the nominal band width. In each plot, the black curve represents a TOA radiance with a wavelength dependence of λ^{-4} .

Rayleigh scattering with a λ^{-4} dependence dominates the total TOA radiance. If a radiance with such a wavelength dependence is measured by sensors having the response functions shown in Figs. 15.25 and 15.26, the total measured radiance in the i^{th} band over the 380-1100 nm range shown in the figures will be (Gordon, 1995, Eq. 8)

$$L_i(\text{total}) = \frac{\int_{380}^{1100} L_{\text{TOA}}(\lambda) SRF_i(\lambda) d\lambda}{\int_{380}^{1100} SRF_i(\lambda) d\lambda}.$$

Define the “in-band” part of the total signal to be the part detected between chosen lower (λ_{low}) and upper (λ_{up}) wavelengths. The OBPG uses the wavelengths at which the SRF drops to 0.1% of its maximum value to define the lower and upper boundaries of the in-band region. For the nominal 488 nm band, for example, $\lambda_{\text{low}} = 460 \text{ nm}$ and $\lambda_{\text{up}} = 503 \text{ nm}$. (The

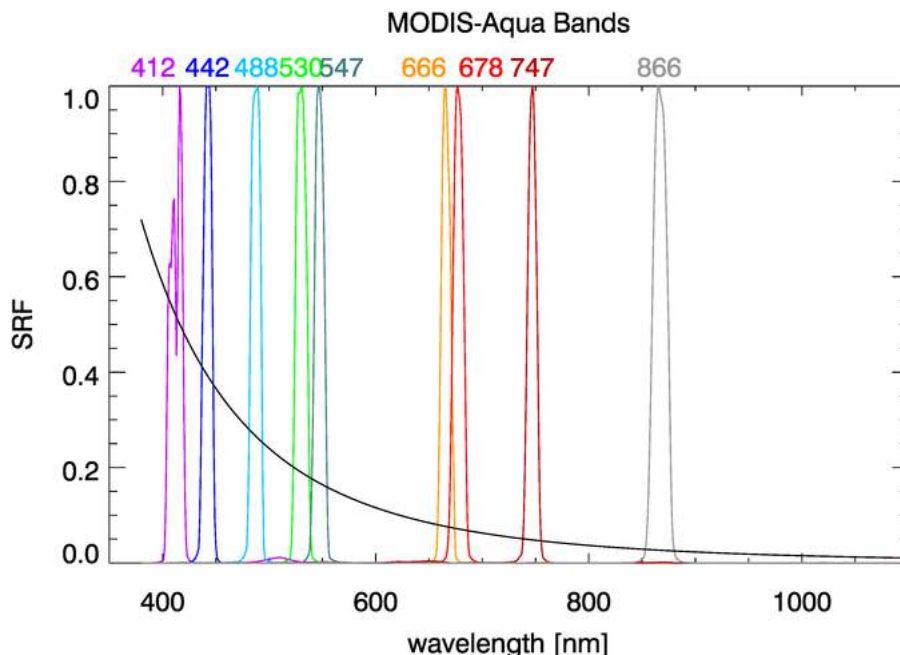


Figure 15.25: Relative sensor response functions (SRF) for the MODIS-Aqua bands used for ocean color remote sensing. The black curve represents a TOA radiance proportional to λ^{-4} . (Data from from [NASA ocean color documents](#))

FWHM boundaries for the 488 nm band are $\text{FWHM}_{\text{low}} = 482\text{nm}$ and $\text{FWHM}_{\text{up}} = 493\text{nm}$.) The part of the measured radiance that comes from in-band wavelengths is then

$$L_i(\text{in band}) = \frac{\int_{\lambda_{\text{low}}}^{\lambda_{\text{up}}} L_t(\lambda) \text{SRF}_i(\lambda) d\lambda}{\int_{380}^{1100} \text{SRF}_i(\lambda) d\lambda}, \quad (15.57)$$

with similar equations for the out-of-band contributions at wavelengths less than λ_{low} and greater than λ_{up} . Numerical integration shows that for a λ^{-4} TOA radiance and the nominal 488 nm band, 99.24% comes from the in-band wavelengths, 0.55% comes from out-of-band response at wavelengths less than $\lambda_{\text{low}} = 460$ nm, and 0.20% comes from out-of-band response at wavelengths greater than $\lambda_{\text{up}} = 503$ nm. For the 866 nm band, the corresponding numbers are 99.22% in-band, 0.63% from wavelengths less than $\lambda_{\text{low}} = 843$ nm, and 0.16% comes from wavelengths greater than $\lambda_{\text{up}} = 891$ nm. Thus, for the MODIS-Aqua bands, almost 1% of the TOA radiance attributed to a nominal bandwidth actually comes from outside that band. This magnitude of misattribution of radiances between bands is significant and requires correction for proper interpretation of measured data.

[Gordon \(1995\)](#) points out that the OOB corrections must be applied separately to the individual components of the TOA radiance because the OOB response depends on the spectral shape of the radiance. That is, separate corrections must be applied to the Rayleigh, aerosol, and water-leaving radiances. Those corrections are built into the sensor-specific Rayleigh and aerosol look-up tables described above. This section describes how

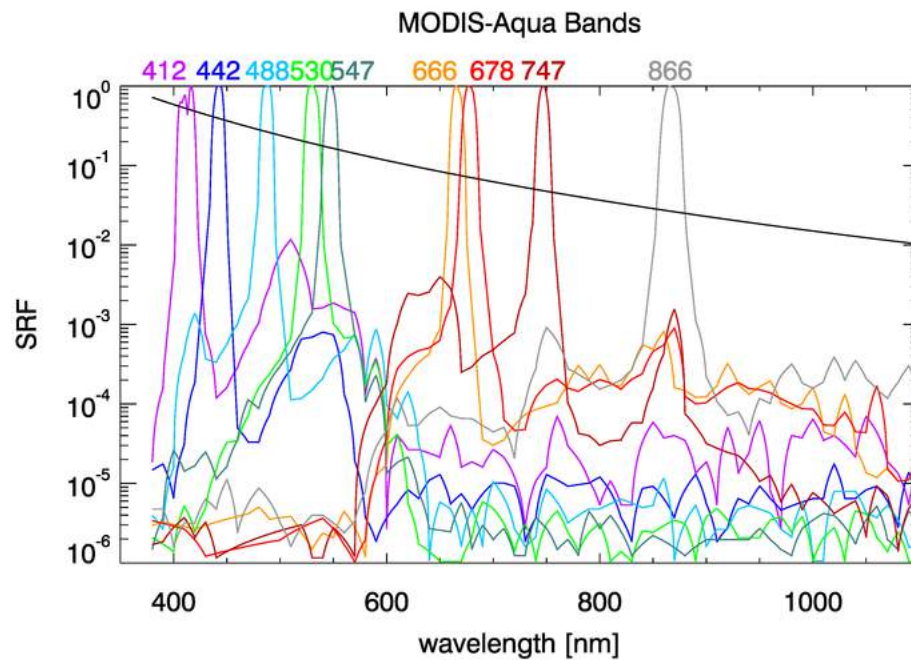


Figure 15.26: Same as Fig. 15.25 except that the ordinate axis is logarithmic to show the out-of-band response.

the OOB correction is applied to the remote-sensing reflectance R_{rs} after the preceding steps of the atmospheric correction process have been carried out.

It should be noted that OOB corrections are also required when comparing measurements made by sensors having different spectral responses. This happens, for example, when comparing a nominal MODIS 412 nm band value with an in-situ measurement made by a multispectral radiometer having a nominal 412 band. The filters in the MODIS and in situ radiometer instruments will not have exactly the same spectral responses or nominal bandwidths. In general, it is desirable to reference any measurement to what would be obtained by an ideal sensor with a perfect response function defined by the nominal FWHM. This is illustrated in Fig. 15.27 for the MODIS-Aqua 412 band and a sensor with a perfect “top hat” response for $407 \leq \lambda \leq 412$ nm. For a radiance with a λ^{-4} wavelength dependence, 18.1% of the MODIS nominal 412 band response comes from $\lambda < 407$, 60.5% comes from within the nominal 10 nm bandwidth of the perfect sensor, and 21.4% comes from $\lambda > 412$ nm.

Figure 15.28 shows $R_{rs}(\lambda)$ as computed for Case 1 water using a model of the type developed in Morel and Maritorena (2001) for $K_d(\lambda)$ and $R(\lambda)$.

For each nominal sensor band labeled by $\lambda_i, i = 1, \dots, N_{\text{bands}}$, and for each chlorophyll value $Chl_j, j = 1, \dots, N_{\text{Chl}}$, the $R_{rs}(\lambda)$ spectra of Fig. 15.28 are used in equations of the form of (15.57) with appropriate integration limits to compute:

- $R_{rs}^{11}(\lambda_i, Chl_j)$ = The mean R_{rs} over idealized 11-nm bandwidths (center wavelength ± 5 nm) corresponding to the nominal satellite bands (as illustrated by the gray 412 nm band in Fig. 15.27)

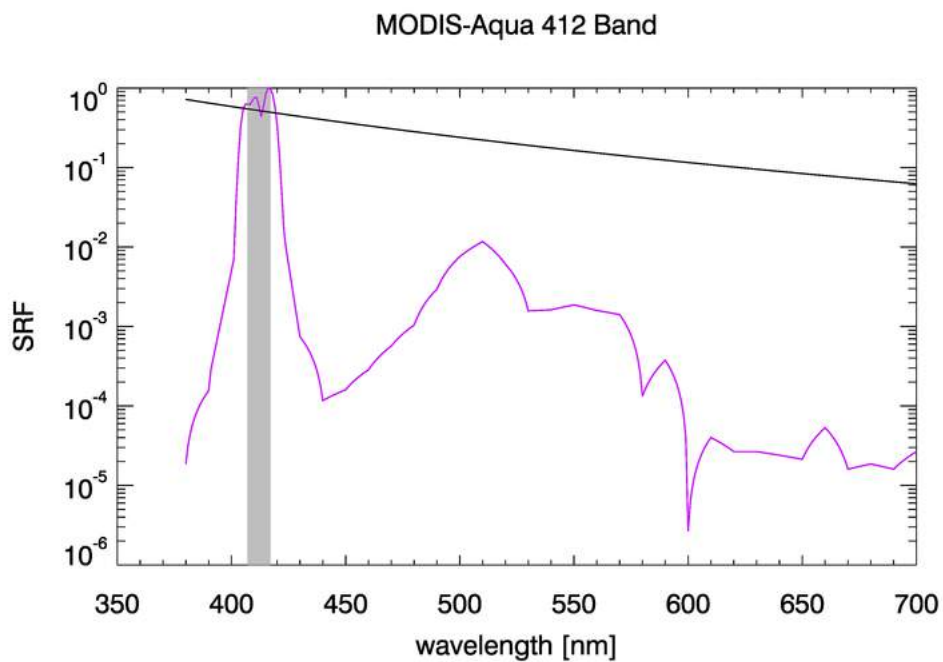


Figure 15.27: The MODIS-Aqua relative sensor response (purple) and a perfect sensor with a 10-nm FWHM (gray shading). The black line represents a radiance proportional to λ^{-4} .

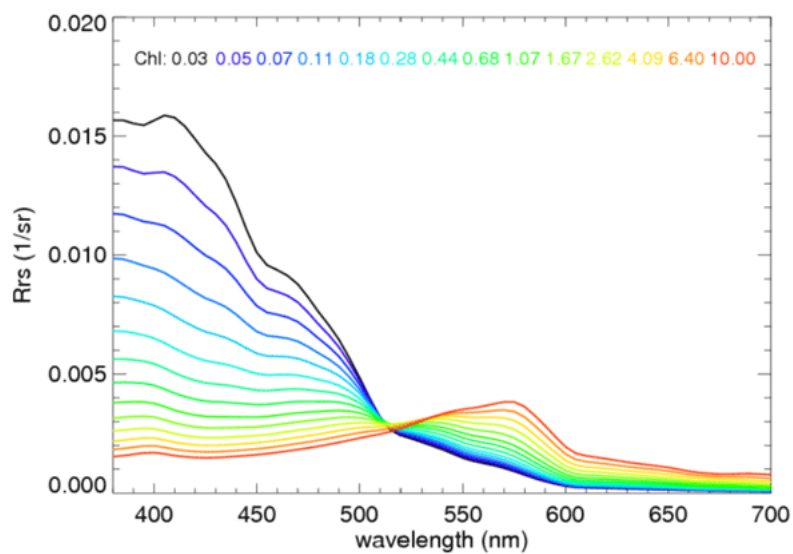


Figure 15.28: $R_{rs}(\lambda)$ spectra as functions of the chlorophyll concentration.

- $R_{rs}^{\text{full}}(\lambda_i, Chl_j) = R_{rs}$ computed using the full spectral response function for the i^{th} sensor band.

- The ratio

$$r(\lambda_i, Chl_j) = \frac{R_{rs}^{11}(\lambda_i, Chl_j)}{R_{rs}^{full}(\lambda_i, Chl_j)}. \quad (15.58)$$

Figure 15.29 illustrates the results of these calculations for the SeaWiFS nominal 555 nm band. This figure shows chlorophyll values only for $Chl \leq 3 \text{ mg m}^{-3}$, which was felt to be the upper limit of reliability of the chlorophyll-based $R_{rs}(\lambda)$ model of Fig. 15.28. A similar figure can be drawn for each sensor band. If the chlorophyll concentration were known, a rearrangement of Eq. (15.58) and ratio curves like that of Fig. 15.28 could be used to compute the correction to the measured R_{rs}^{full} for the i^{th} band.

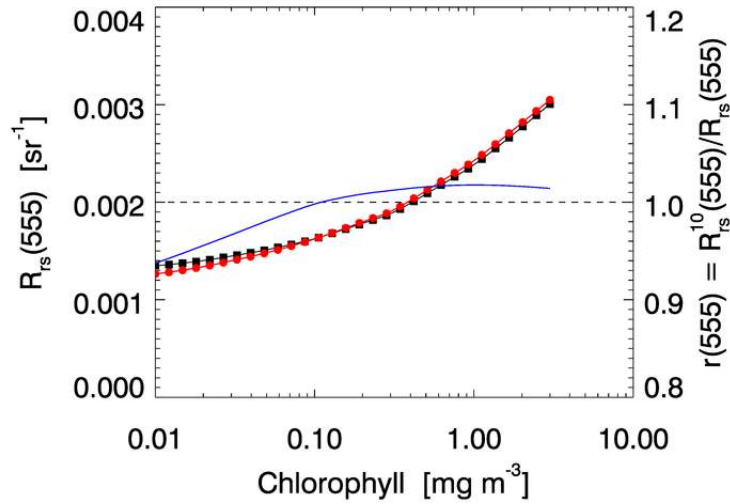


Figure 15.29: $R_{rs}^{11}(\lambda_{555}, Chl_j)$ (red dots), $R_{rs}^{full}(\lambda_{555}, Chl_j)$ (black squares), and the ratio $r(\lambda_{555}, Chl_j)$ (blue line, right ordinate axis) for the SeaWiFS nominal 555 nm band.

However, the chlorophyll concentration is not yet known. To proceed, the $r(\lambda_i, Chl_j)$ curves like the one shown by the blue line of Fig. 15.29 are used to compute correction factors $r(\lambda_i)$ as functions of the ratio of the uncorrected $R_{rs}(490 \text{ nm})$ to the uncorrected $R_{rs}(555 \text{ nm})$. Each chlorophyll value shown in figures like 15.29 for the various bands gives a point like those shown in Fig. 15.30 for $r(555)$ versus the uncorrected $R_{rs}(490)/R_{rs}(555)$. A corresponding set of $r(\lambda_i)$ points is computed for each band, but in each case as a function of $R_{rs}(490)/R_{rs}(555)$. (This choice of the ratio of 490 to 555 nm for the independent variable traces back to SeaWiFS, for which these were the most trustworthy bands.) A best-fit function to the set of points so generated is then found for each of the $r(\lambda_i)$ versus $R_{rs}(490)/R_{rs}(555)$ functions.

These functions are then used to apply the OOB correction to the measured $R_{rs}(555)$ values as follows. Given the measured full-band (uncorrected) $R_{rs}(i)$, the value of $R_{rs}(490)/R_{rs}(555)$ is used to evaluate the functional fit to the points of Fig. 15.30 in order to obtain the correction to be applied to the $R_{rs}(555)$ value. The corresponding functions for the other bands are used to correct those bands. For example, the fit to $r(412)$ versus $R_{rs}(490)/R_{rs}(555)$ is used to correct the 412 nm band, and so on. If the $R_{rs}(490)/R_{rs}(555)$ value is outside the range of the points used for the data fit as illustrated in Fig. 15.30,

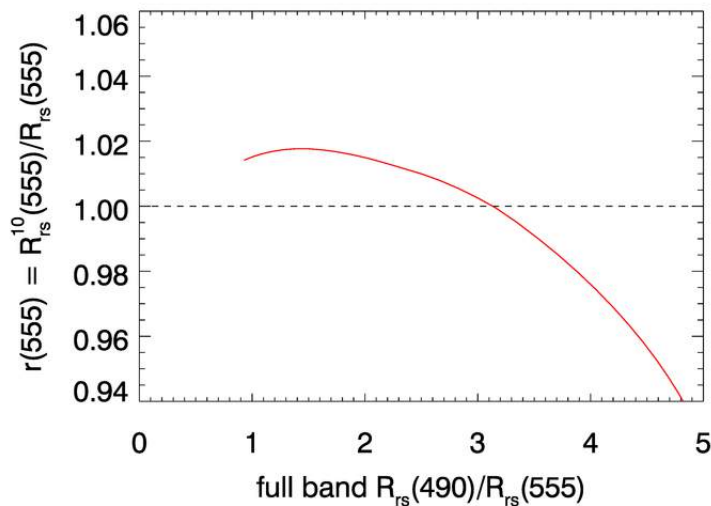


Figure 15.30: Out-of-band correction factor for the SeaWiFS 555 nm band as a function of the ratio of the uncorrected $R_{rs}(490)$ to $R_{rs}(555)$.

the value of the nearest point is used, rather than extrapolating with the fitting function beyond the range of the underlying data.

The $R_{rs}(i)$ functions were developed using a Case 1 model for $R_{rs}(\lambda)$ as shown in Fig. 15.28. Case 2 waters can have much different $R_{rs}(\lambda)$ spectra and therefore should have different OOB corrections. However, in practice, Case 2 waters have the same correction applied as for Case 1 waters.

A final comment is warranted regarding the use of this out-of-band correction when comparing satellite-derived and in-situ measurements of R_{rs} (e.g, when doing vicarious calibration):

- If comparing a multispectral satellite band R_{rs} with in an in situ multispectral measurement, perform the OOB adjustment to the satellite data. (Of course, an adjustment should also be made to the in situ values based on the relative spectral responses $SRF(\lambda)$ of the in situ radiometer.)
- If comparing a multispectral satellite band R_{rs} with in an situ hyperspectral measurement that has been filtered with a 10 nm bandpass filter, perform the adjustment.
- If comparing a multispectral satellite band R_{rs} with an in situ hyperspectral measurement that has not been filtered, do not perform the OOB adjustment to the satellite data. However, the hyperspectral in situ spectrum should processed using the satellite sensor $SRF(\lambda)$ spectra. That is, replace the spectrum used in integrals of the form of Eq. (15.57) with the hyperspectral $R_{rs}(\lambda)$.

Performing the OOB adjustment is the default for processing imagery at OBPG. Therefore, if a user wants to compare satellite data with unfiltered hyperspectral data as in the third bullet above, the standard Level 2 files cannot be used. The user would need to begin with the Level 1b TOA radiances, disable the OOB correction in the atmospheric correction software, and reprocess the TOA radiances to Level 2.

15.13 Polarization

Radiance leaving the top of the atmosphere can be strongly polarized even though the Sunlight incident onto the TOA is unpolarized. This is because scattering by atmospheric constituents, reflection by the sea surface, and scattering within the water all can generate various states of polarization from unpolarized radiance. Although remote sensing as considered here is based on the total TOA radiance without regard to its state of polarization, many instruments are sensitive to polarization. Therefore, the total TOA radiance they measure may depend on the state of polarization of the TOA radiance and the orientation of the instrument relative to the plane of linear polarization. Correction for these effects is required so that instruments give consistent measurements of the total TOA radiance.

The MODIS sensors are polarization sensitive. MODIS radiance measurements vary by up to $\pm 5.4\%$ for totally linearly polarized radiance, depending of the orientation of the sensor relative to the plane of polarization. This amounts to about $\pm 3\%$ differences in measured TOA radiances for typical values of atmospheric polarization (Meister et al., 2005). These effects must be accounted for in order to achieve the desired 0.5% accuracy in measured TOA radiance. Similarly, the VIIRS instrument polarization sensitivity is 1-2% and requires a polarization correction. SeaWiFS by design was not very sensitive to polarization ($< 0.25\%$), and no polarization correction was applied.

As explained in Section 1.6.2, the state of polarization is described by the four-component Stokes vector $[I, Q, U, V]^T$, where superscript T denotes transpose, I is the total radiance without regard for its state of polarization, Q specifies the linear polarization resolved in planes parallel and perpendicular to a conveniently chosen reference plane, U specifies the polarization resolved in planes oriented ± 45 deg to the reference plane, and V specifies the right or left circular polarization. The choice of the reference plane for specification of the Q and U components is arbitrary and can be made for convenience. The direction of propagation of the radiance is given by a unit vector $\hat{\mathbf{i}}$, so that the Stokes vector can be written as $\mathbf{I} = [I, Q, U, V]^T \hat{\mathbf{i}}$ when it is desired to indicate both its components and direction. Direction $\hat{\mathbf{i}}$ can be specified by the polar (θ) and azimuthal (ϕ) directions in a spherical coordinate system as shown in Fig. 15.31. In that figure, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ are unit vectors specified by the directions of increasing θ ($\theta = 0$ at the pole or $\hat{\mathbf{z}}$ direction in a Cartesian coordinate system) and increasing ϕ ($\phi = 0$ in a conveniently chosen azimuthal direction such as $\hat{\mathbf{x}}$ pointing east or toward the Sun).

In the geophysical setting, it is customary to define the Stoke vector components Q and U with reference to a plane defined by the normal to the sea surface and the direction of propagation of the radiance. This plane is known as the meridional plane and is partly shaded in light blue in Fig. 15.31. Following the notation and choices of Gordon (1997), let $\hat{\mathbf{r}}_t = \hat{\boldsymbol{\phi}}$ be the reference direction perpendicular to the meridional plane, and let $\hat{\mathbf{l}}_t = -\hat{\boldsymbol{\theta}}$ be the reference direction parallel to the meridional plane. The direction of propagation of the radiance to be measured is then $\hat{\mathbf{i}} = \hat{\mathbf{r}}_t \times \hat{\mathbf{l}}_t$. The total TOA radiance resolved in these directions is denoted $\mathbf{I}_t = [I_t, Q_t, U_t, V_t]^T \hat{\mathbf{i}}$.

This radiance is being measured by a sensor illustrated by the red rectangle in Fig. 15.31. The Stokes vector measured by that sensor has its Q and U components resolved along perpendicular ($\hat{\mathbf{r}}$) and parallel ($\hat{\mathbf{l}}$) directions chosen for convenience relative to the orientation of the sensor. The sensor will measure the TOA radiance as a Stokes vector $\mathbf{I}_m = [I_m, Q_m, U_m, V_m]^T \hat{\mathbf{i}}$. For an incident radiance \mathbf{I} , the optical system comprising the sensor itself and any associated optical components (mirrors, lenses, etc.) will convert

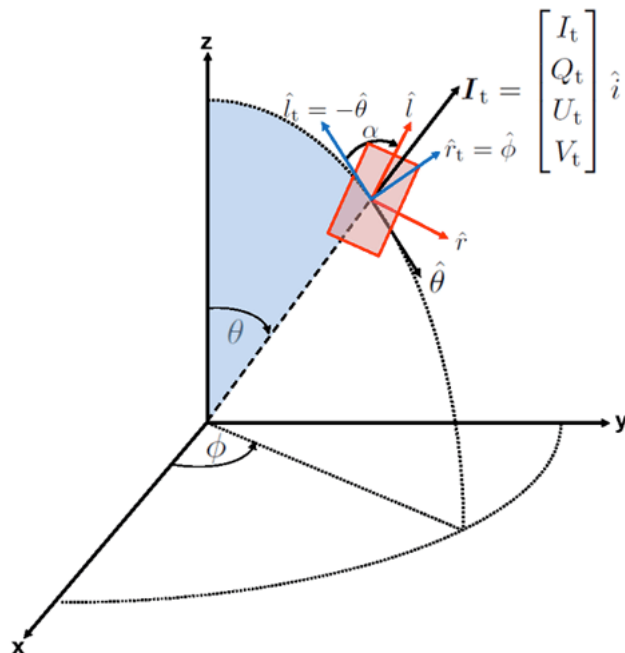


Figure 15.31: Angles and directions used in specification of Stokes vectors. The blue-shaded wedge indicates the meridional plane used to define the TOA Stokes vector \mathbf{I}_t defined via the blue unit vectors; the red-shaded rectangle represents the sensor that measures \mathbf{I}_m defined via the red unit vectors.

the incident radiance into a measured value given by $\mathbf{I}_m = \mathbf{M}\mathbf{I}$, where \mathbf{M} is the 4×4 Mueller matrix that describes the optical properties of the sensor optical system. \mathbf{M} is defined relative to the sensor reference directions $\hat{\mathbf{r}}$ and $\hat{\mathbf{l}}$. In order for \mathbf{M} to operate on the TOA radiance \mathbf{I}_t , which is defined with reference directions $\hat{\mathbf{r}}_t$ and $\hat{\mathbf{l}}_t$, \mathbf{I}_t must first be transformed (rotated) from the meridional-based $\hat{\mathbf{r}}_t, \hat{\mathbf{l}}_t$ system to the sensor-based $\hat{\mathbf{r}}, \hat{\mathbf{l}}$ system.

Let $\alpha = \cos^{-1}(\hat{\mathbf{l}}_t \cdot \hat{\mathbf{l}})$ be the angle between the parallel reference directions for \mathbf{I}_t and for the sensor. With the choice of α being positive for clockwise rotations from $\hat{\mathbf{l}}_t$ to $\hat{\mathbf{l}}$ as seen looking “into the beam” (looking in the $-\hat{\mathbf{i}}$ direction), the transformation is given by the 4×4 rotation matrix

$$\mathbf{R}(\alpha) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(2\alpha) & \sin(2\alpha) & 0 \\ 0 & -\sin(2\alpha) & \cos(2\alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (15.59)$$

Thus, the radiance as measured by the sensor is

$$\mathbf{I}_m = \mathbf{M}\mathbf{R}(\alpha)\mathbf{I}_t. \quad (15.60)$$

It is important to note that the radiance measured by the sensor, \mathbf{I}_m depends both on the “true” TOA radiance \mathbf{I}_t , the sensor optical properties via \mathbf{M} , and the orientation α of the

sensor relative to the local meridional plane. \mathbf{M} is fixed for a given sensor, but \mathbf{I}_t and α change from moment to moment as the sensor orbits and views the TOA radiance in different locations and directions. (To be exact, there are long-term changes to \mathbf{M} caused by degradation of the sensor optical surfaces. These changes are monitored on-orbit and corrected by a cross-calibration technique.)

The quantity of interest here is the measured TOA radiance magnitude, which is given by the first element of the Stokes vector. Using (15.59) in Eq. (15.60) gives this to be

$$I_m = M_{11}I_t + M_{12}[\cos(2\alpha)Q_t + \sin(2\alpha)U_t] \\ + M_{13}[-\sin(2\alpha)Q_t + \cos(2\alpha)U_t] + M_{14}V_t. \quad (15.61)$$

Clearly, if $M_{11} = 1$ and all elements of \mathbf{M} other than the M_{11} element are zero, then the sensor is not sensitive to the state of polarization and $I_m = I_t$. Physical arguments and numerical simulation show the circular polarization of the TOA radiance is very small: $|V_t| \leq 10^{-3}I_t$. This term is therefore neglected in the present correction algorithm. It is customary to define the elements of the *reduced* Mueller matrix by $m_{ij} \triangleq M_{ij}/M_{11}$. Similarly defining reduced Stokes vector element by $q_t \triangleq Q_t/I_t$ and $u_t \triangleq U_t/I_t$, Eq. (15.61) becomes

$$I_m = I_t \{1 + m_{12}[\cos(2\alpha)q_t + \sin(2\alpha)u_t] \\ + m_{13}[-\sin(2\alpha)q_t + \cos(2\alpha)u_t]\}. \quad (15.62)$$

Gordon et al. (1997a) give the general procedure for measuring m_{12} and m_{13} in the laboratory. Meister et al. (2005) give the details of these measurements for the MODIS sensors. These quantities, which specify the polarization sensitivity of the instrument, are determined before the instrument is launched; they are thus known. Angle α is determined by the orbit and pointing geometry of the sensor. It remains to determine the elements of \mathbf{I}_t .

Following Eq. (15.3) of the Problem Formulation section, the total TOA polarized radiance can be decomposed as

$$\mathbf{I}_t = \mathbf{I}_R + \mathbf{I}_a + \mathbf{I}_{Ra} + T\mathbf{I}_g + t\mathbf{I}_{wc} + t\mathbf{I}_w. \quad (15.63)$$

Here, as before, the Rayleigh (R), aerosol (a), and Rayleigh-aerosol (Ra) radiances are at the TOA; the glint (g), whitecap (wc), and water-leaving (w) radiances are at the sea surface. The surface values are transmitted to the TOA via the appropriate direct (T) and diffuse (t) transmittances. According to Eq. (15.60), each of these radiances must be known in order to predict what the sensor will measure for a given TOA radiance and, thereby, to determine the correction needed to account for sensor polarization effects.

The surface-glint and atmospheric polarization contributions to the TOA signal are computed separately. Sea-surface glint can be highly polarized. This glint contribution to the TOA signal is computed using a vector radiative transfer code assuming a Rayleigh-scattering atmosphere above a rough Fresnel-reflecting ocean surface (Gordon and Wang, 1992; Wang, 2002), meister2005. The water-leaving radiance is at most 10% of the TOA total, and the whitecap contribution is generally even less. These two terms are therefore ignored in the present development. The effect of aerosols and Rayleigh-aerosol interactions depends on the particle size distribution and concentration of the aerosols, which are unknown during atmospheric correction. Fortunately, numerical simulations show that the

Rayleigh contribution to the TOA polarization is usually much greater than the aerosol-related contributions. Therefore, the aerosol contributions are also ignored and the polarization correction is based on the TOA Rayleigh radiance. The total TOA Stokes vector is then modeled as the sum of the glint and Rayleigh contributions.

Meister et al. (2005, Eq. 15) defines the polarization correction via $p_c \triangleq I_m/I_t$. Equations (15.61) and (15.62) allow this to be written as

$$p_c = \frac{1}{1 - m_{12}[\cos(2\alpha)Q_t + \sin(2\alpha)U_t]/I_m - m_{13}[-\sin(2\alpha)Q_t + \cos(2\alpha)U_t]/I_m}.$$

As applied during atmospheric correction, the unknown total TOA radiance components Q_t and U_t are replaced by the corresponding TOA Rayleigh components Q_R and U_R . The Rayleigh components are precomputed and tabulated for use during the first step of the atmospheric correction process, namely the removal of the Rayleigh contribution as described in Section 15.7. The end result is that the actual measured value I_m is used along with the Rayleigh radiance for the given atmospheric conditions and viewing geometry to obtain an estimate of the TOA radiance via

$$I_t = I_m - m_{12}[\cos(2\alpha)Q_R + \sin(2\alpha)U_R] - m_{13}[-\sin(2\alpha)Q_R + \cos(2\alpha)U_R]. \quad (15.64)$$

Gordon et al. (1997a) show that this approximate polarization correction is acceptably accurate (errors $\Delta I_t/I_t < 0.01$ so long as the error has the same sign throughout the spectrum) when m_{12} is independent of wavelength and less than about 0.1 in magnitude. If m_{12} depends on wavelength, the approximation does not perform well for m_{12} as small as 0.02. Application of this polarization correction to MODIS Aqua imagery shows (Meister et al., 2005) that the polarization correction p_c is largest at blue wavelengths (the MODIS 412 nm band), where p_c lies in the range of 0.978 to 1.032.

[This section ends the discussion of the atmospheric correction algorithms used by the NASA Ocean Biology Processing Group for processing of satellite imagery.]

15.14 Empirical Line Fits

[Contributors to this section include Richard Zimmerman and Victoria Hill of Old Dominion University, who made the field measurements and developed the ELFs; Paul Bissett (then at WeoGeo, Inc.), who provided the WV2 image; and Curtis Mobley.]

The “black-pixel” technique discussed in the Aerosols Section 15.11.2 works well in many open-ocean situations, but it is not applicable if the water-leaving radiance is not very close to zero at the near-IR wavelengths used to model the aerosol scattering. This can occur for two reasons. First, if the water contains mineral particles, these highly scattering particles can give a significant amount of water-leaving radiance even out to 1000 nm and beyond. Such particles are common in coastal waters because of inputs by rivers or sediment resuspension by strong currents. Second, if the water is clear and less than a meter deep, there can be significant water-leaving radiance due to bottom reflectance, especially for bright sand bottoms.

Case 2 and shallow waters are of great interest for reasons such as ecosystem management, recreation, and military operations, and such waters are often observed from aircraft

with hyperspectral imaging sensors. New techniques have been developed for atmospheric correction of such imagery.

In general, we need an atmospheric correction technique that

- works for any water body (Case 1 or 2, deep or shallow),
- works for any atmosphere (including absorbing aerosols), and
- does not require zero water-leaving radiance at particular wavelengths (no black-pixel assumption).

Two basic atmospheric correction methods have been developed in response to these needs. The first of these is a correlational technique called empirical line fitting, which is discussed in this section. The second, discussed in the next section, uses atmospheric radiative transfer calculations.

The essence of the empirical line fit (ELF) technique is as follows:

- Make field measurements of the remote-sensing reflectance $R_{rs}(\lambda)$ [or water-leaving radiance $L_w(\lambda)$, or non-dimensional reflectance $\rho_w(\lambda)$, or whatever is needed by your retrieval algorithms] *at the same time as the image acquisition and at various points within the imaged area.*
- The $R_{rs}(\lambda)$ measurements at various points in the imaged area are then correlated with the at-sensor measurements for the image pixels viewing the stations where $R_{rs}(\lambda)$ was measured. The at-sensor spectra can be in any units, e.g. radiance or digital counts. The correlation functions that convert at-sensor spectra to sea-surface $R_{rs}(\lambda)$ values are the empirical line fits. A different ELF is obtained for each wavelength.
- Assume that the atmospheric conditions, surface waves, and illumination are the same for every pixel in the image.
- The ELFs, which were developed from a few image pixels, are then used to convert the at-sensor measurements to sea-level $R_{rs}(\lambda)$ spectra for every pixel in the image

To illustrate how this process works, ELFs were developed for a DigitalGlobe WorldView-2 image of shallow Case 2 waters in St. Joseph's Bay, Florida, USA. WorldView-2 (WV2) is a commercial satellite that provides high spatial resolution (approximately 2 m pixel size), 8-band multispectral imagery. Figure 15.32 shows an RGB image of this area, created from WV2 bands 5 (656 nm), 3 (546 nm) and 2 (478 nm) for the red, green, and blue values. The area includes some dry land (at the lower left of the image), areas of dense bottom vegetation (reddish color in this image), clean sand bottom (white to green, depending on depth), and an optically deep channel (darkest area). These Case 2 waters are optically deep (the bottom cannot be seen) for depths greater than about 3 m.

Hyperspectral $R_{rs}(\lambda)$ measurements were made from a small boat at 10 stations in St. Joseph's Bay. Those spectra were then weighted by the WV2 spectral response functions for each of the 8 bands to obtain multispectral $R_{rs}(\lambda)$ spectra that correspond to the 8 WV2 bands. Figure 15.33 shows the relative spectral response of the 8 WV2 bands. Thus $R_{rs}(j)$, the equivalent of $R_{rs}(\lambda)$ corresponding to WV2 band j , is obtained from

$$R_{rs}(j) = \frac{\int_{400}^{1100} w_j(\lambda) R_{rs}(\lambda) d\lambda}{\int_{400}^{1100} w_j(\lambda) d\lambda},$$

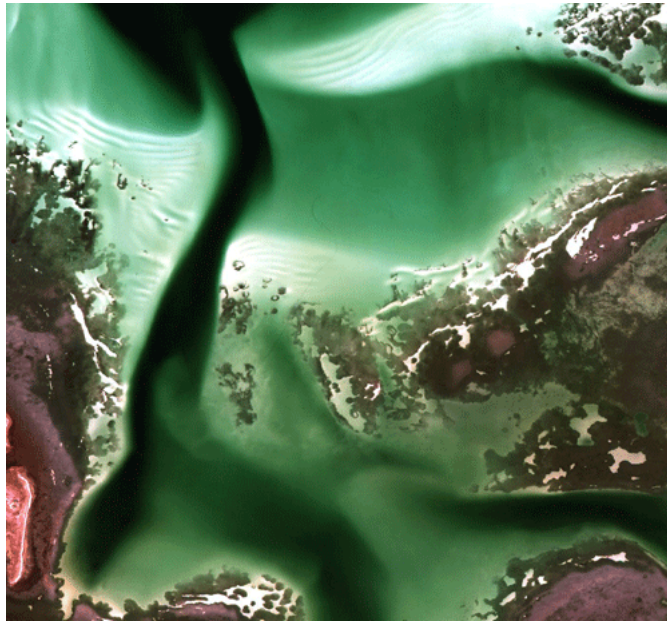


Figure 15.32: RGB image created from the WorldView-2 image of the south end of St. Joseph's Bay, Florida. This image covers about 6.3 km^2 and is 1400×1297 pixels.

where $w_j(\lambda)$, $j = 1, \dots, 8$ is the spectral response function for band j as seen in Fig. 15.33. Figure 15.34 shows one of the measured hyperspectral $R_{\text{rs}}(\lambda)$ spectra and the corresponding values for the WV2 bands.

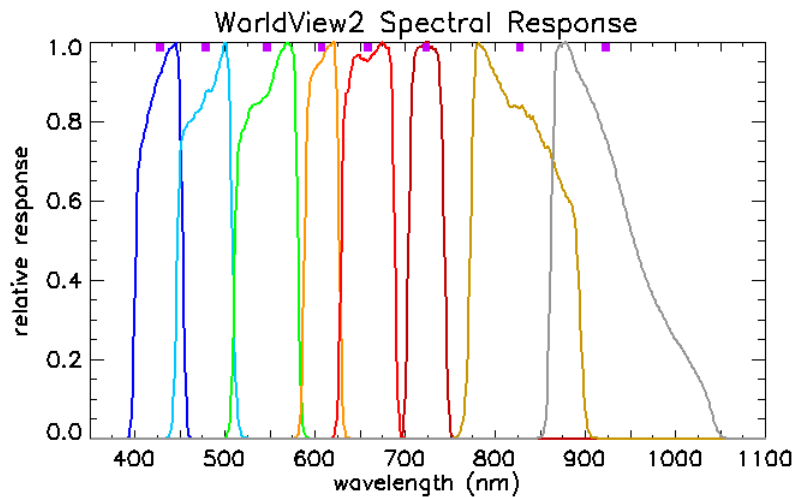


Figure 15.33: Normalized WV2 spectral response functions.

Each of the WV2 image pixels “looking” at the 10 ground stations was then used to correlate the WV2 top-of-the-atmosphere (TOA) band values in digital counts (DC) with the sea-level $R_{\text{rs}}(j)$ values for each of the 8 bands. Figure 15.35 shows the results for WV2 band 3, which is centered at 546 nm. *The best-fit line to these 10 points is the ELF*

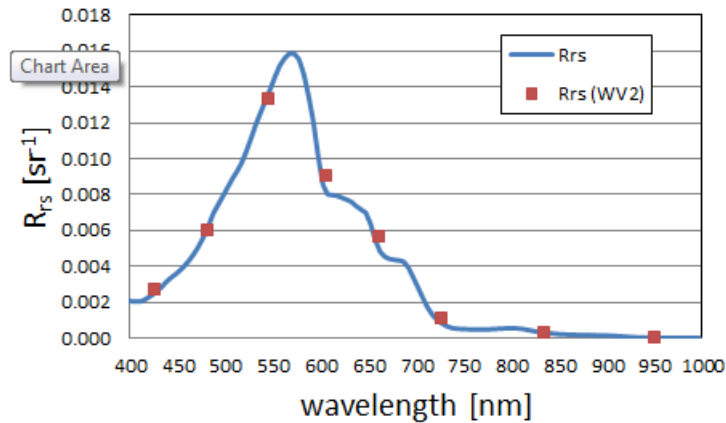


Figure 15.34: Example of a measured hyperspectral $R_{rs}(\lambda)$ spectrum and the 8 values corresponding to the WV2 multispectral bands.

for this wavelength band. In this example, the ELF's convert the TOA measurements in digital counts to sea-level $R_{rs}(j)$ values in units of sr^{-1} . These ELF's, obtained from only 10 points, are then applied to every pixel in the image. Note that the measured $R_{rs}(\lambda)$ spectra correspond to different water IOPs, bottom depths, and bottom types, but the atmospheric conditions and viewing geometry (and resulting atmospheric path radiance) are assumed to be the same at each ground station.

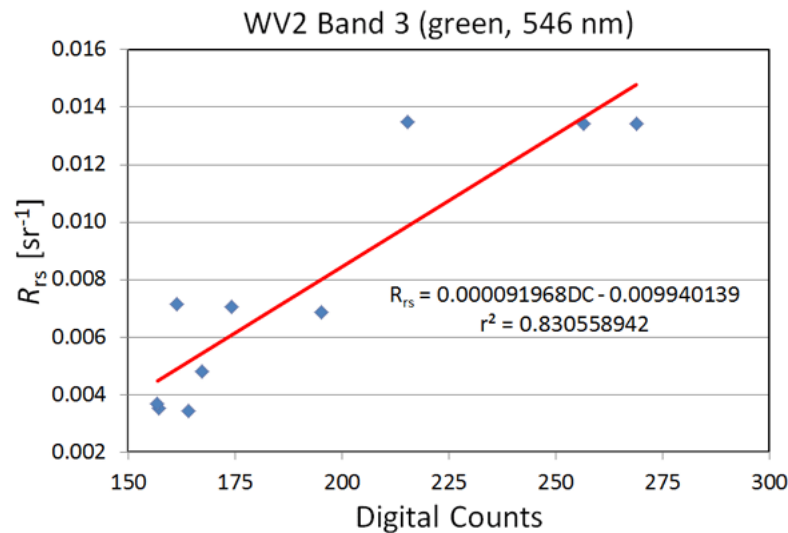


Figure 15.35: $R_{rs}(j = 3)$ values for the 10 ground stations, and the corresponding TOA radiances in digital counts. The red line is the ELF for this wavelength band.

Figure 15.36 shows an example WV2 TOA spectrum (excluding the two bands in the IR) and the corresponding sea-level R_{rs} spectrum obtained from the 6 ELF's. Note for band 3, for example, that the TOA value of 170 DC is converted to an R_{rs} value of 0.0057 sr^{-1} , in accordance with the ELF of Fig. 15.35.

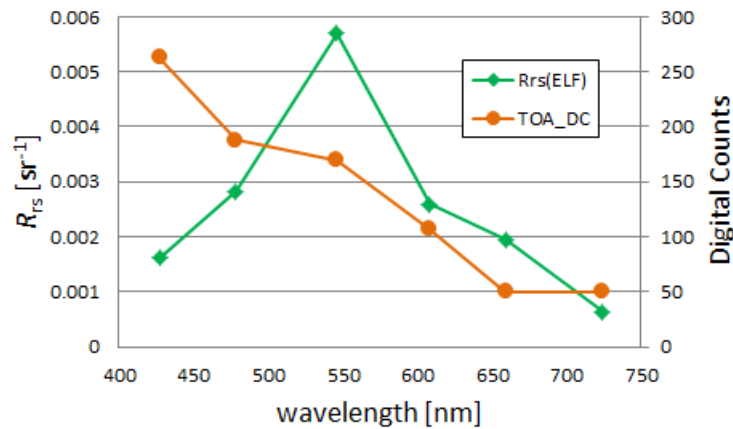


Figure 15.36: Example of conversion of a TOA spectrum in digital counts (DC) to a sea-level R_{rs} in sr^{-1} .

The advantages of the ELF technique are

- The ELFs account for atmospheric path radiance for any atmospheric conditions, without the need to know what these conditions are. *No atmospheric measurements are needed.*
- The technique works for shallow or Case 2 waters, for which $L_w(\lambda)$ is not zero.
- The technique works for any Sun and sensor geometry, or sensor altitude (airborne or satellite sensors).

The disadvantages of the ELF technique are

- *Field measurements of $R_{rs}(\lambda)$ must be made at the time of image acquisition*, which is labor intensive and often impossible.
- *A set of ELFs is valid only for the one image used for their development.* ELFs for one image cannot be applied to a different image of the same area, or to a different area, because the atmospheric conditions, or Sun and viewing geometry, will differ for other locations and times.
- The field measurements always contain errors, which introduces an unknown amount of error into the ELFs, hence into the final $R_{rs}(\lambda)$ spectra.
- The same ELFs are applied to all image pixels, even though the atmospheric and water conditions and viewing geometry may vary from one part of the image to another. (For airborne sensors, the viewing geometry and atmospheric path radiances can vary greatly from one part of an image to another.)

15.15 Radiative Transfer Techniques

As just seen, the empirical line fit technique has the virtues that it works in principle for any atmosphere and no atmospheric measurements are needed. However, it has the disadvantage that it requires field measurements of $R_{rs}(\lambda)$ to use in developing the ELF

functions. These are often impossible to obtain and certainly are not practicable for routine observation of large areas. The ELF technique also does the same atmospheric correction for every image point, even though for airborne imaging the viewing geometry may be significantly different for different image pixels.

Radiative transfer techniques take the sophisticated approach of numerically computing the atmospheric path radiance for each pixel in an image. The virtue of this approach is that every image pixel can get a different atmospheric correction, which can account for differences in viewing geometry or even for differences in atmospheric conditions in various parts of the image. The disadvantage of this approach is that the *atmospheric conditions must be measured in the field at the time of image acquisition*, or perhaps obtained from atmospheric forecast models, in order to obtain the inputs needed to compute the atmospheric path radiance for each pixel.

A pioneering radiative transfer model developed for atmospheric correction of airborne hyperspectral images over water is the TAFKAA code developed by the U.S. Naval Research Laboratory (Gao et al., 2000; Montes and Gao, 2004). TAFKAA is based on the earlier ATREM⁶ (ATmospheric REMoval) code of Gao and Goetz (1990). Other such codes exist (e.g., ACORN and FLAASH; see the comparison in San and Suzen (2010)), but they are used primarily for terrestrial remote sensing. TAFKAA is discussed here because one of its versions is designed for the particular problems of oceanic hyperspectral remote sensing. (One TAFKAA version is designed for terrestrial imagery and does not remove surface reflectance; the other version is designed for ocean imagery and can remove sea-surface reflectance.)

When using a radiative transfer code such as TAFKAA, the Sun and viewing geometry for each pixel are known from the location, time, sensor altitude and heading, and pixel location in a georectified image. If atmospheric conditions such as sea-level pressure; aerosol type, altitude, and optical thickness; and humidity are known, the path radiance can be computed for each path from pixel to sensor. Knowing the wind speed allows for an estimate of the background sky reflectance by the sea surface to be computed. These path radiance calculations are performed with an atmospheric radiative transfer model, usually including polarization. Such calculations are computationally intensive, so one set of calculations is performed for a wide range of conditions to create a look-up table of path radiances (and other factors, such as atmospheric transmittances). Given the viewing geometry and atmospheric conditions for a pixel, interpolation in the look-up table is used to obtain the appropriate wavelength-dependent path radiance and surface reflectance to subtract from the at-sensor radiance or reflectance to obtain the water-leaving radiance or reflectance at the sea surface.

The original TAFKAA look-up table included path radiances and atmospheric transmittances for the following grid of values:

- 5 aerosol types (maritime, 2 coastal, tropospheric, and urban)
- 5 relative humidities (50, 70, 80, 90, and 98%)
- 10 aerosol altitudes from 0 to 84 km
- 10 aerosol optical depths from 0.0 to 2.0 (at 550 nm)
- 9 solar zenith angles from 1.5 to 72 deg.

⁶Indeed, TAFKAA stands for The Algorithm Formerly Known As ATREM.

- 17 off-nadir viewing angles from 0 to 88 deg
- 17 azimuthal viewing angles from 0 to 180 deg, relative to the Sun
- 17 wavelengths from 390 nm to 2.25 μm
- 3 wind speeds of 2, 6, and 10 m s^{-1}

This grid of inputs required about 332 million solutions of the vector radiative transfer equation. These calculations required many months of computer time, but they needed to be done only once.

When processing an image, the user inputs the image information (location, time, aircraft or satellite altitude and heading, georectification information) and atmospheric conditions. TAFKAA can then look up the appropriate value to subtract from each at-sensor spectrum, interpolating as necessary in the look-up table. If no aerosol information is available, then TAFKAA defaults to making a “black-pixel” assumption and estimating the aerosol type from wavelengths of 750 and 865 nm (if no wavelengths greater than 1000 nm are available in the image), as previously described. TAFKAA allows the user to input a file with the sensor wavelength responses for the different wavelength bands, so that TAFKAA output matches a particular sensor’s wavelength response as closely as possible. Inputs to TAFKAA are made via an ENVI-format image header file. Applications of TAFKAA can be seen at [Montes et al. \(2003\)](#) and [Goodman et al. \(2003\)](#).

There are assumptions in the TAFKAA calculations that limit its applicability or accuracy. The water-leaving radiance is assumed to be Lambertian; this excludes modeling multiple scattering effects between the water-leaving radiance and the atmosphere. It is assumed that the viewing geometry avoids direct Sun glint; thus TAFKAA cannot correct Sun glint in an image. The surface reflectance calculations for background sky radiance cannot correct for swell effects. The code does not interpolate in wind speed. Nevertheless, given the needed atmospheric information, TAFKAA provides adequate atmospheric corrections for a wide range of imagery and environmental conditions. However, as with any such model, if the inputs do not describe the imaged environment, the TAFKAA corrections can be poor.

Photometry and Visibility

In general the well defined and objectively measured quantities of geometrical radiometry are used in quantitative studies of optical oceanography, remote sensing, and radiative transfer. However, there are times when, either unavoidably or by choice, the human eye becomes one of our instruments. Such is the case when we enjoy the beautiful colors of nature or when someone observes the appearance of a distant object as a semi-quantitative measure of the clarity or “visible range” of the atmosphere or ocean. In other instances, the eye-brain system may be the preferred instrument, as in visual searches for underwater objects. We therefore must understand how the human visual system responds to radiant energy. This takes us into the domain of *photometry*, which, for our present purpose, is defined as the study of the human visual response to the quantities of geometrical radiometry. Photometry provides the theoretical structure for understanding visibility, in particular for predicting whether or not an object will be visible under given environmental conditions.

Perhaps the first person interested in underwater visibility was Alexander the Great around 330 BCE. Legend says that he had constructed “a very fine barrel made entirely of white glass” in which he had himself lowered into the sea as shown in Fig. 16.1. The first underwater photograph was made in 1893, and the first underwater color photograph was made in 1923. The first movie filmed in part underwater was a 1916 adaptation of Jules Verne’s *20,000 Leagues Under the Sea*.¹ Mertens (1970) gives a short review of this history. Today underwater visibility and imaging have a wide variety of applications such as visual searches for objects in marine archaeology and recreational scuba diving and, of course, military applications.

This chapter first covers the basic ideas of human visual response and visibility, beginning with an overview of how the eye senses light and color. The next topic discussed is the photopic luminosity function, which describes the spectral response of the average human eye, and luminance, which corresponds to the visual sensation of brightness. Luminance is an inherently broad-band (all visible wavelengths) correspondent of monochromatic radiance. An equation for luminance transfer is then developed from the monochromatic

¹You can watch the entire movie at <https://www.youtube.com/watch?v=QsrXuyjci7U> There is a 20 minute segment of underwater scenes starting at 39:10. Although murky, these images were wildly popular with audiences of a century ago, who had never seen the underwater world.



Figure 16.1: Alexander the Great in his glass diving bell. From an illuminated manuscript circa 1400, now in the Bodleian Library at Oxford University (MS. Bodl. 264).

radiative transfer equation. This requires the development of photopic (broad-band) replacements for wavelength-dependent inherent optical properties. The quantitative specification of color via CIE chromaticity coordinates is then developed. The introductory material closes with a discussion of the relation between CIE chromaticity coordinates and the red-green-blue RGB values used by computer monitors.

The mathematical formulation of classical visibility theory is then developed. A luminance transfer equation, corresponding to the radiative transfer equation for monochromatic light, is first developed. Whether or not an object can be seen depends on the visual contrast between the object and its background. This is quantified by the inherent and apparent contrast, and a law of contrast reduction is developed. These tools give us what is needed to understand the theory of the Secchi disk and the prediction of the Secchi depth.

Appendix G develops a rigorous mathematical framework for image propagation through absorbing and scattering media.

16.1 Human Color Vision

Writing this section is something of a no-win undertaking. How the eye works and how humans perceive color are so complicated that anything I say will be a great oversimplification and therefore open to criticism (justly deserved) that what I'm saying isn't quite correct. Indeed, half a dozen Nobel Prizes in Physiology or Medicine have been awarded for studies of how eyes work. Nevertheless, I'll try to give a qualitative summary of human color vision that will be sufficient for the needs of the rest of this chapter.

The quantitative study of color began in the 1600s with Isaac Newton, who showed that white light can be decomposed into various colors, which could then be recombined to create white light. He also showed that it was possible to add certain “primary” colors together to obtain a different color, e.g., combining red and green to get yellow. (Primary colors are colors such that combining any two of them cannot produce the third. These are usually taken to be red, green, and blue; but other combinations such as blue, yellow

and red could also be used as the primaries. Then blue plus yellow gives green.) In the early 1800s Thomas Young proposed the idea that color vision is based on combinations of three different colors and suggested that the eye must have three types of color receptors; this is called trichromatic color vision. James Clerk Maxwell and Hermann von Helmholtz furthered the study of trichromatic color vision in the late 1800s.

However, not all visual phenomena can be explained by the trichromatic hypothesis. Consider, in particular, “afterimage” color reversals as illustrated in Fig.16.2. First look at the black dot in the white area at the right of the yellow patch. You see a black dot against a white background. Now stare at the black dot in the center of the yellow area for at least 30 seconds. Then shift your focus to the black dot in the white area to the right. When staring at the black dot on the white background, you now will perceive the surrounding white area as a light bluish shade. It is difficult to explain the perceived change of the white background to blue in terms of a simple sum of red, green, and blue inputs. This and other visual phenomena led Ewald Hering to develop in the late 1800s the opponent-process theory of color vision, in which opposing responses generated by the color opposites blue-yellow and red-green, along with black-white, generate color vision. These opposite or complementary color pairs are based on the observation that you might describe a color as bluish-green or greenish-yellow, but never as bluish-yellow, so blue and yellow are complementary colors. Likewise no one ever describes a color as reddish-green.

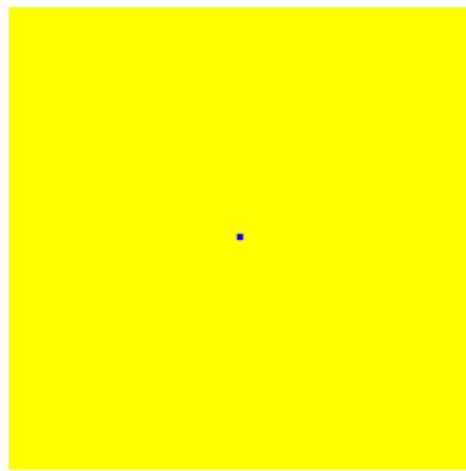


Figure 16.2: Color patches for illustration of afterimage colors. See the text for instructions.

We now know that the retinas of normal human eyes contain two types of cells that detect light. One type, called rods, contain only one type of pigment and are achromatic, i.e. they “see in black and white.” The rods have a slow response (longer integration time) and operate at low light levels. These cells give us “night vision” (scotopic vision), at the expense of spatial resolution. The other cell type, called cones, have three types of pigments. The cones have faster responses (shorter integration times) and therefore operate at higher light levels. These cells give us daytime color vision (photopic vision), with good spatial resolution. The three types of cones (for the three types of pigments) are sensitive to different but somewhat overlapping wavelength ranges, which are usually called short-wavelength (S), middle-wavelength (M), and long-wavelength (L) cones. Figure 16.3 shows the normalized sensitivities of the S, M, and L cones. These curves are derived from

measurements of the absorption spectra of the three different pigments found in cone cells.

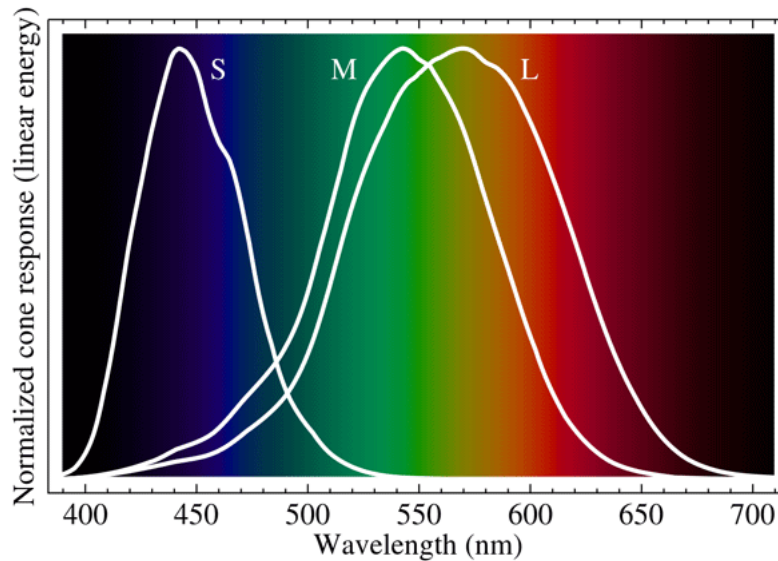


Figure 16.3: Sensitivities of the three different types of cone cells. Figure created by [BenRG](#)

The three types of cones give us trichromatic color vision, i.e. color vision that is based on the responses of the three different S, M, and L cone cell receptors. This is often called red (R), green (G), and blue (B) color vision although, as the overlapping spectra of Fig. 16.3 show, it is not correct to say that our eyes simply detect red, green, and blue light. The color perceived depends on the relative levels of excitation of the S, M, and L cones *and* on how those signals are processed. Thus the total signal ($S + M + L$) gives an achromatic measure of overall brightness. Then $L + M - S$ gives yellow or blue, depending on the excitation strengths, and $L - M + S$ leads to red or green. Thus the trichromatic and opponent-process models are complementary and both are necessary to explain the complexities of color vision. The trichromatic model describes the detection of light at the receptor level, and the opponent-process model is needed to explain how the detected RGB inputs are processed by the eye-brain system, i.e. the opponent-process model applies at the signal-processing level.

It must be understood that color is not a property of a material, or even of the light reflected from or emitted by an object. Color is a *perceived sensation*, not a physical property of light. Using well designed and calibrated radiometers, we can all agree on the spectrum of light received by an eye. However, different people may perceive that light differently. There is some natural variability in the chemical composition of the pigments that give the S, M, and L responses. Thus my “red” may not be quite the same as your “red” when we view the same spectrum. If you are red-green color blind (deuteranopia), you are missing the M cones, which alters how a spectrum of detected light is perceived. Figure 16.4 shows a simulation of how a spectrum appears to a normal person and to one with deuteranopia. Deuteranopia occurs in about 6% of males, but in less than 0.5% of females (which illustrates another of the many disadvantages of having a Y chromosome).

There are many other abnormalities in color vision. The most extreme is achromatopsia, in which all three types of cones are missing. Such people see only in shades of gray via their rod cells, which also makes them generally hypersensitive to bright light and degrades

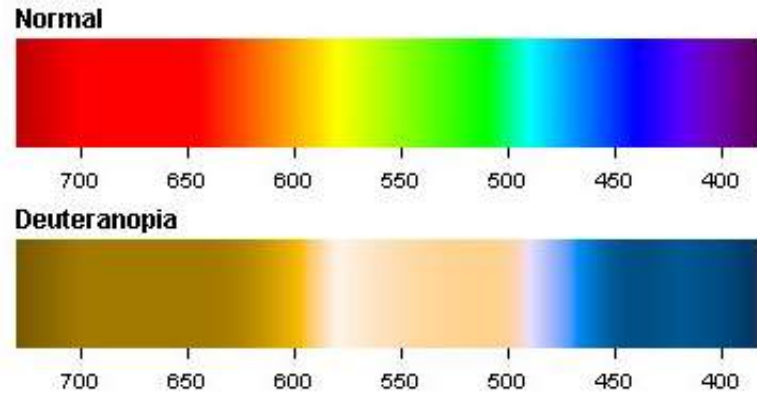


Figure 16.4: Visual appearances of spectra for a normal person (top) and a deuteranopia (bottom). Figure from [Colblindor](#)

their visual acuity. This is a rare condition affecting only about 1 in 30,000 people, except on the small island of Pingelap Atoll in Micronesia, where about 10% of the population is achromatopic².

The perceived color also depends on the magnitude (“brightness”) of the detected spectrum. For example, brown is just red that is not very bright. This is shown in Fig. 16.5. The color patches were generated on an RGB computer monitor for which the minimum intensity is 0 (the pixel is turned off), and the maximum intensity is 255. The upper left patch has the red pixels at maximum intensity, with the green and blue pixels turned off. The remaining patches have the red pixels at lower and lower intensities. As the magnitude of the red spectrum decreases, the color is perceived as going from bright red to dark brown.

I will close this section with an interesting connection between linguistics and color vision. Every European language spoken from Portugal to Russia and Iceland to Greece has (as far as I know) separate words for green and blue (and has words for finer distinctions of each—just visit any paint store), as do many other languages around the world. However, there are also many languages that do not have separate words for green and blue. There are so many of these languages that linguists have a term for them: the *grue* languages, where *grue* comes from combining *green* and *blue*. Most (but not all) *grue* languages are minor languages spoken by people living in either desert or tropical areas. For example, in the deserts of the southwestern United States and northwestern Mexico, Cahuilla, O’odham, Seri, Guarajío, and Western Tarahumara are *grue* languages. So the question arises: Do speakers of *grue* languages not see the difference in green and blue, or do they see the difference in green and blue, but just don’t consider the difference important enough to warrant having different words? I don’t know the answer. It seems hard (for me at least) to believe that someone would think the color difference between green leaves and blue sky is unworthy of distinction with different words. However, it’s equally hard to imagine that diverse populations from tropical areas around the world all have a similar genetic difference that gives them some sort of green-blue color blindness. More research is needed, as they say.

²That’s an interesting genetics story; see *The Island of the Colorblind* by Sachs (1996).

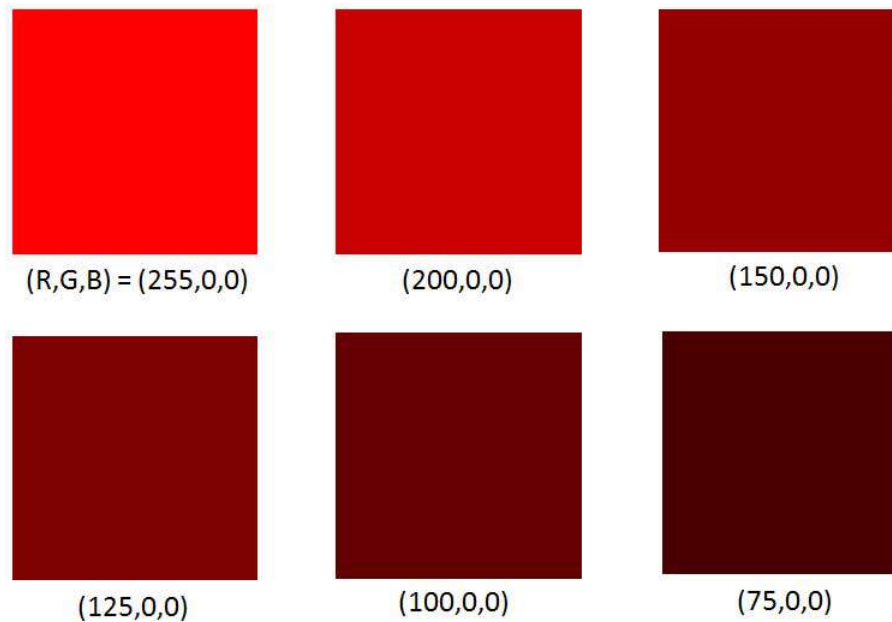


Figure 16.5: Change in the appearance of “red” as the intensity of the signal decreases from the maximum possible value on a computer monitor. See the additional discussion of these colors in Section 16.4.

The above discussion shows only a small part of the complexity of color vision. Just do an internet search on “optical illusions” and you can entertain yourself for hours with visual phenomena that show the incredibly complex ways that the eye-brain system can interpret images and colors. An excellent site is that of [Michael Bach](#). Likewise, investigation of how different cultures around the world describe colors is another good way to spend rainy winter days when you can’t go outside and play. There is a short [history of color vision](#) at this Smithsonian website.

16.2 Luminosity Functions

If the human eye is the sensor, then all visible wavelengths are seen simultaneously, and monochromatic radiometric variables, IOPs, and AOPs must be replaced by broad-band values that depend on both the wavelength dependence of the ambient radiance and on the relative sensitivity of the eye to different wavelengths. This is effected by replacing radiometric quantities by their photopic equivalents.

Not all wavelengths of light evoke the same sensation of brightness in the human eye-brain system. For example, suppose a person with “normal” eyesight is exposed to monochromatic radiance of wavelength 550 nm and magnitude of $10^3 \text{ W m}^{-2} \text{ sr}^{-1} \text{ nm}^{-1}$. (This is comparable in magnitude to the Sun’s spectral radiance at this wavelength when seen through a hazy atmosphere or at a large solar zenith angle.) The person will “see” a bright yellowish-green light. However, if the person is exposed to light of the same radiance magnitude, but of wavelength 300 nm, the person will not “see” anything because the eye is not sensitive to this ultraviolet wavelength. However, if the exposure lasts long enough,

permanent and severe damage will be done to the eye by the ultraviolet radiant energy.

The relative ability of radiant energy of different wavelengths to evoke differing sensations of brightness in the human observer is described by luminosity functions. The cone cells of the human eye are responsible for color vision at daylight levels of the ambient illumination. These cells have a sensitivity described, when averaged over many individuals, by the *photopic luminosity function*. The rod cells are responsible for vision in very dim light, such as at night. These cells are more efficient at seeing blue wavelengths and less efficient at red wavelengths than are the cones. The eye sensitivity at low-light conditions is given by the *scotopic luminosity function*.

The photopic and scotopic luminosity functions are plotted in Fig. 16.6. These functions are empirically derived averages based on visual response studies of numerous humans. In these studies a colored light is viewed next to a reference light. The observer adjusts the power of the colored light until it subjectively appears to have the same brightness as the reference light. The reciprocal of the measured radiance of the colored light is then plotted at the wavelength of the colored light. This process of “brightness matching” is subjective and there is considerable variance among observers, so the resulting average over many observers has somewhat the same statistical validity as the “average American male, age 30.” Nevertheless, the functions serve as reasonable reference standards for human eye response. Suppose, for example, that monochromatic radiance $L(\lambda = 500 \text{ nm})$ (blue-green light) of some given magnitude (in $\text{W m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$) evokes a certain qualitative sensation of brightness in the eye. Then from Figure 16.6 we see that in order to produce the same sensation of brightness with red light of wavelength 650 nm requires about three times the radiance, i.e. $L(\lambda = 650) \approx 3L(\lambda = 500)$. The left panel of this figure shows for comparison the relative spectra response of a low-light video camera³. That camera outputs a gray-scale image, and its spectral response gathers as much light as possible over a broad spectral range (even out to 1000 nm).

The normalized photopic luminosity function is denoted by $\bar{y}(\lambda)$. The spectral radiance $L(\lambda)$, weighted by $\bar{y}(\lambda)$ and integrated over all wavelengths (in practice, usually from 380 to 720 nm or even just 400 to 700 nm) gives the *luminance* L_v , which is the photopic, or vision, equivalent of radiance:

$$L_v \triangleq K_{\text{cd}} \int_0^\infty L(\lambda) \bar{y}(\lambda) d\lambda \quad [\text{lm m}^{-2} \text{sr}^{-1}]. \quad (16.1)$$

Here K_{cd} is a fundamental physical constant called the *luminous efficacy* that by definition has the value $K_{\text{cd}} = 683 \text{ lumen W}^{-1}$ exactly (recall the discussion of Section 1.3 and Table 1.1). This quantity converts radiance from energy units (Watts or Joules per second) to the visual unit of lumens (abbreviated lm). The numerical value of K_{cd} traces back to the idea of the visual brightness of a “standard candle.” The modern definition of a lumen is that the surface of melting platinum (at a temperature of 2042 K) emits luminance of $6 \times 10^5 \text{ lm m}^{-2} \text{sr}^{-1}$. The luminance L_v corresponds to the visual impression of brightness. In photometry, the subscript v (for visual) on a radiometric quantity flags it as the corresponding photometric quantity.

As seen in the units of K_{cd} , the lumen is the visual correspondent of radiometric power in watts. The SI base unit for photometric variables in the candela, abbreviated cd (Table

³This is the sort of camera that is used for television nature programs showing tigers running around in the jungle at night.

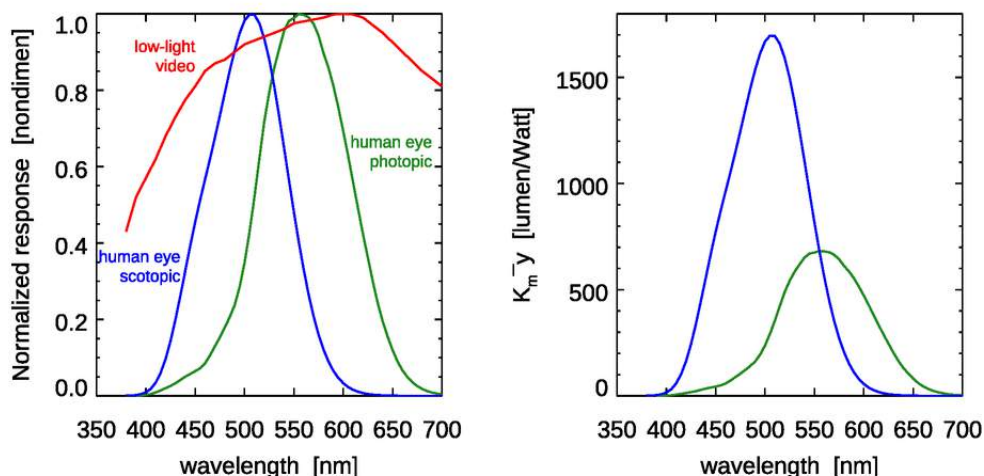


Figure 16.6: Luminosity functions for bright and dim light. Left panel: normalized functions; right panel: functions in lumen per watt. (The photopic function is the 1931 CIE 2-degree function in energy units; see www.cvrl.org/lumindex.htm; see also Mobley (1994, Tab. 2.1).)

1.2). The definition of the candela is “The candela is the luminous intensity, in a given direction, of a source that emits monochromatic radiation of frequency 540×10^{12} hertz and that has a radiant intensity in that direction of $1/683$ watt per steradian.” (This frequency corresponds to $\lambda = 555$ nm for light in a vacuum.) Recall that radiometric intensity is power per unit solid angle. The candela is correspondingly luminous power per unit solid angle, i.e. $1 \text{ cd} = 1 \text{ lm, sr}^{-1}$. The lumen is then a derived quantity, which by definition is $1 \text{ lm} \triangleq 1 \text{ cd sr}$. Unsuccessful attempts have been made to have the lumen adopted as the SI base unit; the proposed definition being “the lumen is the luminous power of monochromatic radiant energy whose radiant power is $1/683$ W and whose frequency is 540×10^{12} Hz.” Neither of these definitions is particularly enlightening, so for intuitive purposes, think of an ordinary candle as having a luminous intensity of about 1 cd, or 1 lm sr^{-1} . The candela (Latin for candle) indeed traces back to the historical use of a “standard candle” as the unit of brightness⁴.

For conversion of night-time, or dim-light, radiances to luminance, an equation of the same form as (16.1) is used but with the scotopic luminosity function, denoted $\bar{y}'(\lambda)$. The conversion factor for the scotopic luminosity function is $K'_{\text{cd}} = 1700 \text{ lm W}^{-1}$. Thus the rods are more efficient at converting radiant energy into visible light, but at the trade off of giving only gray-scale images.

As an example, we can compute the luminance of the Sun. The Sun’s output can be approximated as that of a blackbody at a temperature of $T = 5782$ K. The radiance of a blackbody is given by Eq. 2.6 in the blackbody radiation section:

$$L_{\text{BB}}(\lambda) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{hc/\lambda kT} - 1}. \quad (16.2)$$

⁴One of the historical definitions of a standard candle was the “light produced by a pure spermaceti candle weighing one sixth of a pound and burning at a rate of 120 grains per hour.” A grain is an obsolescent unit of weight equal to approximately 0.065 g.

Using this radiance in Eq. (16.1) gives

$$L_v^{\text{Sun}} = K_{\text{cd}} \int_0^\infty L_{\text{BB}}(T = 5782 \text{ K}, \lambda) \bar{y}(\lambda) d\lambda = 1.86 \times 10^9 \text{ lm m}^{-2} \text{ sr}^{-1}.$$

This is the value in the first line of Table 16.1.

Source	Luminance (lm m ⁻² sr ⁻¹ = cd m ⁻²)
solar disk, above the atmosphere	2 × 10 ⁹
solar disk, at Earth's surface, Sun near the zenith	1 × 10 ⁹
melting platinum at 2042 K	≜ 6 × 10 ⁵
60 W frosted light bulb	1 × 10 ⁵
sunlit snow surface	1 × 10 ⁴
full Moon's disk	6 × 10 ³
clear blue sky, directions away from the Sun	3 × 10 ³
heavy overcast, zenith direction	1 × 10 ³
twilight sky	3
clear sky, moonlit night	3 × 10 ⁻²
overcast sky, moonless night	3 × 10 ⁻⁵

Table 16.1: Typical luminances L_v .

All radiometric quantities have a photometric equivalent obtained by an equation of the form of (16.1). Thus the photometric equivalent of the downwelling plane irradiance $E_d(\lambda)$, which is called the downwelling plane illuminance, is given by

$$E_{\text{dv}} \triangleq K_{\text{cd}} \int_0^\infty E_d(\lambda) \bar{y}(\lambda) d\lambda \quad [\text{lm m}^{-2}],$$

and so on. Just as irradiances can be computed from radiances, illuminances can be computed from luminances by equations of the same form as for radiometric variables. For example,

$$E_{\text{dv}} = \int_{2\pi_d} L_v(\theta, \phi) |\cos \theta| \sin \theta d\theta d\phi.$$

Table 16.2 shows typical illuminances E_{dv} . In illumination engineering, one lumen per square meter is called a “lux.” Thus the instruments used to measure brightness in rooms are often called lux meters (photographers usually call them light meters).

Photometric equivalents of apparent optical properties are obtained from the photometric variables just as AOPs are obtained from radiometric variables. Thus the illuminance reflectance is given by

$$R_v = \frac{E_{\text{uv}}}{E_{\text{dv}}},$$

and the photometric diffuse attenuation function for E_{dv} is given by

$$K_{\text{dv}}(z) = -\frac{d \ln E_{\text{dv}}(z)}{dz},$$

and so on.

Source	Illuminance ($\text{lm m}^{-2} = \text{lux}$)
Sun at the zenith, clear sky	1×10^5
Sun at 60 deg zenith angle, clear sky	5×10^4
overcast day	1000
well-lit room	300-500
very dark, heavily overcast day	100
full Moon at 60 deg zenith angle, clear sky	0.2
starlight, moonless night, clear sky	4×10^{-3}
moonless night, heavy overcast, in a thick forest	10^{-4}

Table 16.2: Typical illuminances E_{dv} .

16.3 Chromaticity

This section develops a way to specify color “coordinates” such that if I give you the coordinates of a color I see, you can reproduce that color.

16.3.1 Color Matching

“Color-matching” experiments were performed in the 1850s by James Clerk Maxwell as he worked on his theory of trichromatic color vision. Further experiments in the late 1920s by W. D. Wright and J. Guild formed the foundation of modern color matching. The design of the Wright and Guild experiment is shown in Fig. 16.7. A box is divided into two sections by a partition. On one side of the partition, a test light shines light of any color onto a white viewing screen. On the other side of the partition, three adjustable lights can shine primary colors onto the viewing screen. These primary colors are usually taken to be red, green, and blue. The viewer can adjust the brightnesses of the individual R, G, and B lights, but not their wavelengths. The viewer then adjusts the brightnesses of the R, G, and B lights until their combined light matches the test color. The geometry restricts the observer’s vision to a 2 deg field of view. This means that the light from the viewing screen is collected at the fovea area of the observer’s retina, where the cone cells are most dense and where there is a minimum of rod cells.

The purpose of the Wright and Guild experiments was not to show that red, green, and blue lights can be combined to create another color. That was well known. Their intention was to get the data needed for developing a quantitative and unambiguous way to describe all colors visible to the human eye. That is to say, if I give you the “coordinates” of a color I see, you can reproduce that color. Wright and Guild of course had no idea in the 1920s that their work would find application decades later in color television, computer monitors, and digital cameras.

In their experiments it was found that a combination of monochromatic red, green, and blue lights (say, at wavelengths of 650, 550, and 450 nm) could match many, but not all, test colors. For example, if the test light is 500 nm, then no combination of 650, 550, and 450 nm light will give a perfect match to the visual color of light with a wavelength of 500 nm. To make a long story short, it was eventually determined that

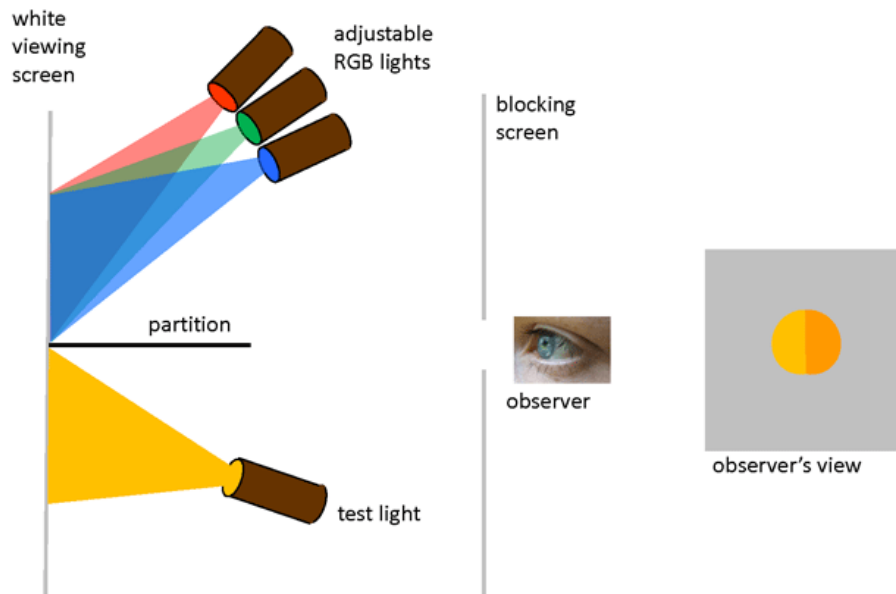


Figure 16.7: The experimental layout of a color-matching experiment.

any color visible to humans with normal vision can be matched by mixing the light from three *non-monochromatic* “color-matching” spectra. These reference spectra are the CIE *tristimulus functions*, which are often called the CIE color-matching functions or spectra. [Service \(2016\)](#) has an excellent explanation of the path from the Wright-Guild experiments to the CIE tristimulus functions.

16.3.2 CIE Chromaticity Coordinates

The [Commission Internationale de l’Eclairage](#) (CIE, or International Commission on Illumination) develops and publishes standards for all matters relating to photometry, including lighting, vision, and colorimetry. In 1931, based on the Wright and Guild data, the CIE published three color-matching, or tristimulus, functions. Although other similar functions have been developed over the years, e.g, for a 10 deg observer field of view, which includes more rod cells, the 1931 CIE two-degree tristimulus functions remain the most commonly used starting point for specification of color. These functions are shown in Fig. [16.8](#).

The tristimulus functions are used as follows. Given a spectrum $\Lambda(\lambda)$, which could be radiance or irradiance, first compute the integrals

$$X = K_{\text{cd}} \int_0^{\infty} \Lambda(\lambda) \bar{x}(\lambda) d\lambda, \quad (16.3a)$$

$$Y = K_{\text{cd}} \int_0^{\infty} \Lambda(\lambda) \bar{y}(\lambda) d\lambda, \quad (16.3b)$$

$$Z = K_{\text{cd}} \int_0^{\infty} \Lambda(\lambda) \bar{z}(\lambda) d\lambda. \quad (16.3c)$$

The tristimulus functions have equal areas under the three curves, and each integrates to a value of $106.9K_{\text{cd}} = 73000$ (to three place accuracy). The $\bar{x}(\lambda)$, $\bar{y}(\lambda)$ and $\bar{z}(\lambda)$ spectra do

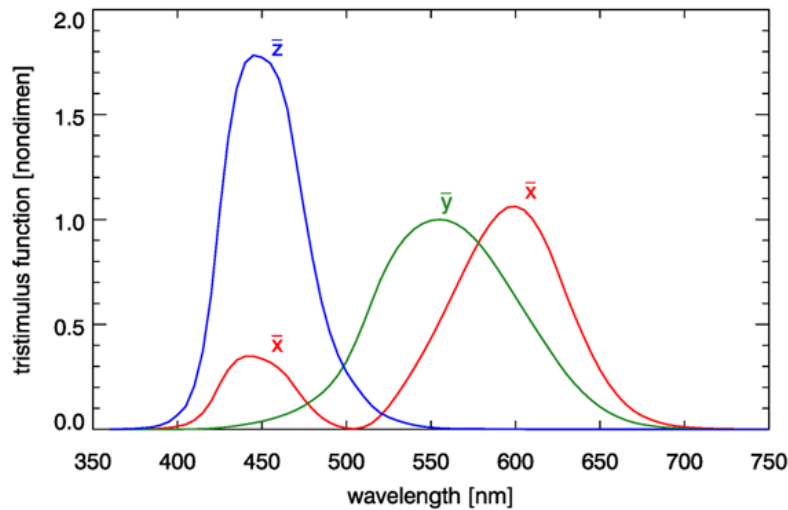


Figure 16.8: The CIE 1931 two-degree tristimulus functions. (The data can be downloaded from the corresponding Web Book page.)

not represent actual light sources as might be used in a color-matching experiment, and they are certainly not red, green, and blue primary colors, because there is considerable overlap of the three spectra. The functions are just mathematical constructions that allow the unique specification of the visual color of a given spectrum $\Lambda(\lambda)$ via the values of X , Y , and Z . The magnitudes of X , Y , and Z depend on the magnitude of the $\Lambda(\lambda)$ spectrum, and the units of X , Y , and Z depend on whether $\Lambda(\lambda)$ is a radiance (in which case X , Y , and Z are in lumens per square meter per steradian) or an irradiance (in which case X , Y , and Z are in lumens per square meter).

To remove the dependence of the magnitude of $\Lambda(\lambda)$ from the computation of its color, next form

$$(x, y, z) = \left(\frac{X}{X + Y + Z}, \frac{Y}{X + Y + Z}, \frac{Z}{X + Y + Z} \right). \quad (16.4)$$

The quantities (x, y, z) are called the *chromaticity coordinates* of the spectrum $\Lambda(\lambda)$. Note that $x + y + z = 1$, so only two of these quantities are independent. By convention, x and y are used to specify the color.

The $\bar{y}(\lambda)$ tristimulus function is identical to the photopic luminosity function discussed previously (and $\bar{y}(\lambda)$ is almost identical to the response of the M cones). Thus the Y integral of Eq. (16.3b) is the same as the luminance integral seen previously. The values of (x, y) are used to specify the color, independent of the magnitude of the spectrum, and Y is used to specify the brightness, which depends on both the shape of the spectrum and its magnitude.

Consider a spectrum that has unit magnitude at all wavelengths; this represents pure white light. This is known as CIE reference illuminant E, with the “E” standing for Equal energy at each wavelength. Inserting this spectrum into Eqs. (16.3) gives $X = Y = Z = 73000$, so $(x, y) = (0.333, 0.333)$. This is known as the *white point*.

Now consider a spectrum that has unit magnitude at 500 nm and is zero at all other wavelengths: $\Lambda(\lambda) = \delta(\lambda - 500)$, where δ is the Dirac delta function of Section 1.4.3. This spectrum picks off the values of the tristimulus functions at 500 nm, which are $\bar{x}(500) = 0.0049$, $\bar{y}(500) = 0.3230$, and $\bar{z}(500) = 0.2720$. From equations 16.3, the values of (X, Y, Z) are then $(16.7, 1103, 929)$ e.g., $X = 0.0049K_{cd}\Delta\lambda$, with $\Delta\lambda = 5$ nm for integration (summation) at 5 nm resolution. These values give $(x, y) = (0.0082, 0.5384)$. If all pure wavelengths between 360 and 830 nm are processed in the same way, this generates a sequence of (x, y) values known as the *spectrum locus*. The spectrum locus gives the color coordinates of monochromatic spectra, and the white point gives the color of a spectrum containing equal energy at all wavelengths.

The spectrum locus defines a region of (x, y) space in which any color visible to the human eye has a corresponding point. This plot is known as the CIE *chromaticity diagram* and is shown in Fig. 16.9. The region between 360 and 830 nm (often taken to be 400 and 700 nm since the eye is very insensitive outside 400-700 nm) is called the purple line. Purple is not a spectral color; it is the visual sensation generated by a mixture of red and blue light. Figure 16.10 shows the chromaticity diagram with the different regions labeled by their common color names.

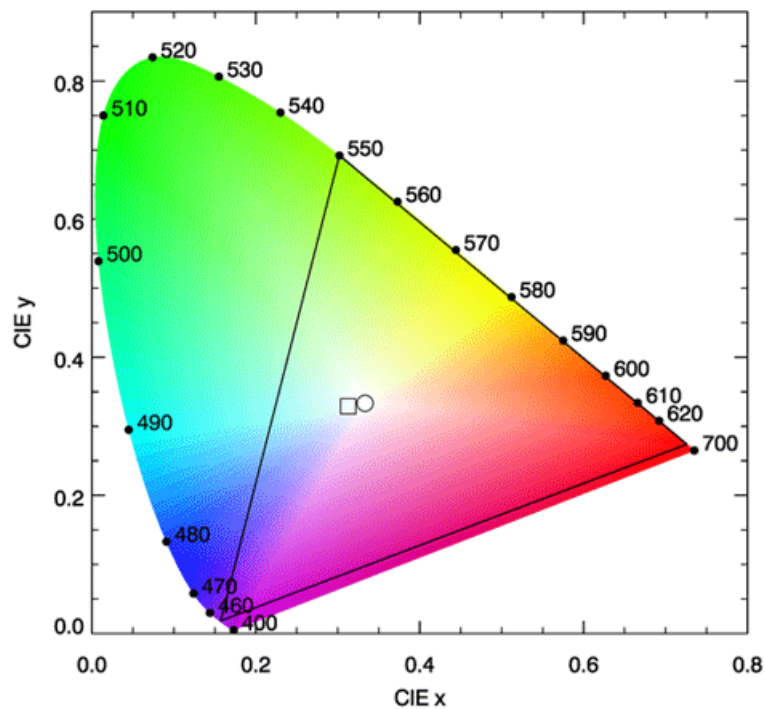


Figure 16.9: The CIE 1931 two-degree chromaticity diagram. The white point is indicated by the circle at $(x, y) = (0.333, 0.333)$. The square at $(x, y) = (0.313, 0.329)$ is the color of the D65 spectrum shown in Fig. 16.11. The large black triangle shows the gamut of colors that can be represented by a linear combination of pure red, green, and blue colors at 650, 550, and 450 nm. The colors seen here are only a rough approximation of what the eye would actually see.

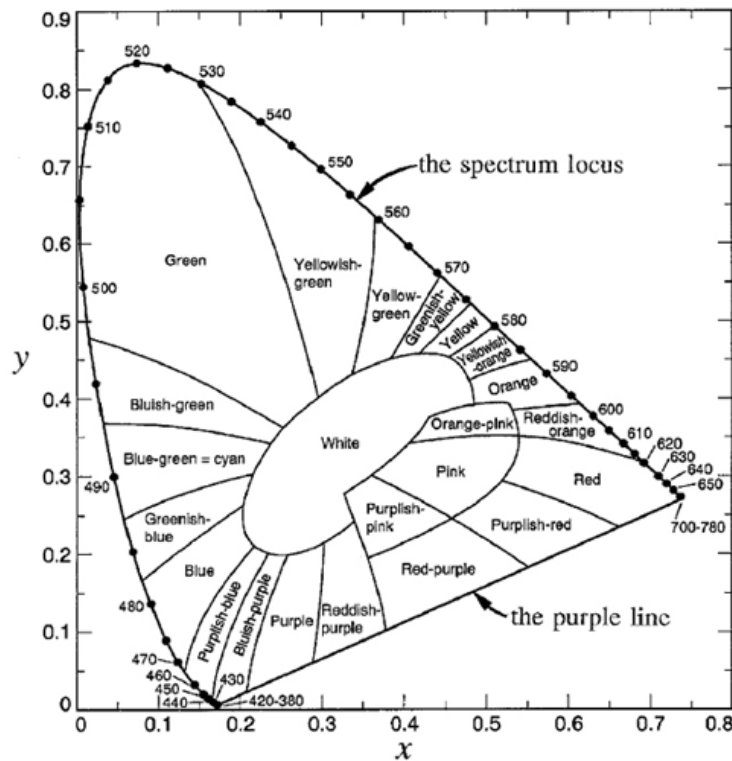


Figure 16.10: Color names for different regions of the CIE chromaticity diagram as defined by Kelly (1943).

16.3.3 Metameric Spectra

Every spectrum $\Lambda(\lambda)$ (that is not zero at all visible wavelengths) has a unique associated color (x, y) , which is obtained from Eqs. (16.3) and (16.4). However, the converse is not true: every color (x, y) corresponds to many different spectra. Different spectra $\Lambda_i(\lambda)$ and $\Lambda_j(\lambda)$, $i \neq j$, that have the same color (x, y) are called *metameric* spectra or *metamers*. (Strictly speaking, two different spectra are metameric if they have the same color coordinates for a given reference illuminant and reference observer.)

CIE reference illuminant D65 is a commonly used spectrum that simulates outdoor light at mid-day on a sunny, clear day. This spectrum has color coordinates $(x, y) = (0.3128, 0.3290)$, which lies somewhat on the blue side of the white point. This shift towards the blue simulates the contribution by blue sky radiance to the whiter light of the Sun's direct beam. The "D" stands for daylight, and the "65" indicates that it has a "correlated color temperature" of 6,500 K. (The CIE has published many other reference spectra for tungsten-filament lights, fluorescent lights, LED lights, etc.). The color of the D65 spectrum is shown by the square symbol in Fig. 16.9.

Figure 16.11 shows the D65 spectrum and three spectra that are metameric to it. To create the first of these spectra, I started with a pure white spectrum and added more energy at 490 nm until the chromaticity coordinates of the resulting spectrum, shown by the cyan curve in Fig. 16.11, was the same as that of the D65 spectrum. Another

metameric spectrum was created by starting with white and removing one half of the energy at wavelengths of 590 to 610 nm. Removing red from white leaves a spectrum with a bluish color. Since the D65 color point lies within the gamut of the 650, 550, and 450 nm pure colors, I then adjusted the amounts of monochromatic 650, 550, and 450 nm energy until the resulting three-wavelength spectrum had the same (x, y) as the original D65 spectrum (0.4503 parts of 450 nm, 0.6311 parts of 550 nm, and 1.000 parts of 650 nm).

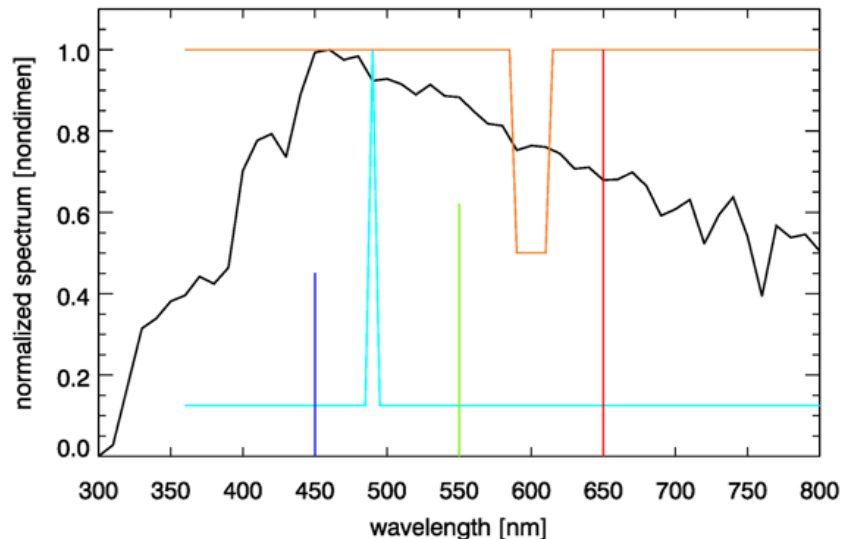


Figure 16.11: The black curve is the CIE D65 reference spectrum. The cyan curve is a white spectrum plus extra energy at 490 nm. The orange curve is white with one-half of the energy removed for wavelengths of 590-610 nm. The blue, green, and red spikes at 450, 550, and 650 nm show the relative amounts of energy at those wavelengths for an RGB spectrum. All spectra are normalized to 1 at their maximum values. Each of these metameric spectra has the same color coordinates $(x, y) = (0.313, 0.329)$.

The four spectra seen in Fig. 16.11 are entirely different spectra from the viewpoint of radiometry, but they are identical from the viewpoint of photometry. That is, a radiometer would measure the much different spectra seen in the figure, but the eye sees each of these as having exactly the same color. It may seem peculiar that the eye-brain system cannot distinguish these four spectra, but the reason goes back to the three types of pigments in cone cells as discussed in Section 16.1. You cannot do better spectrum discrimination with just three overlapping color sensors. Our simian ancestors no doubt found it advantageous to be able to distinguish ripe blueberries and ripe red strawberries from unripe green ones, so that is the ability that evolution has given us. There is no need for us to see a full spectrum. One might imagine our eyes eventually evolving into radiometers, in which case we might see tables of numbers or plots of spectra, instead of colors, when we look at things. That might give us the ability to mentally extract the chlorophyll concentration when looking at the ocean, but there was no evolutionary value of that ability on the plains of Africa 5 million years ago.

16.4 From XYZ to RGB

Computer monitors and TVs have light sources made of many groups of three small pixels. One pixel in each group generates visually red light, one green, and one blue. Each group can generate various levels of brightness for each color. Computer monitors thus use a triplet of red, green, and blue (RGB) values to define how much of each primary color is used to generate a desired color. In most monitors, each value of R, G, or B ranges from 0 (no light of that color emitted; the pixel is turned off) to $2^8 - 1 = 255$ (maximum brightness). This is usually called “8 bits per channel” or “24 bit,” i.e. 3×8 , color. These 256 brightness levels for each primary color give a total of $256^3 = 16,777,216$ different colors and levels of brightness that can be displayed, even though the human eye can perceive differences in far fewer colors and brightness levels. Digital cameras likewise store images in terms of RGB values.

As was seen in the preceding section, converting a spectrum $\Lambda(\lambda)$ into CIE (X, Y, Z) and then into (x, y) chromaticity values is easy. Converting between (X, Y, Z) and (R, G, B) values is more complicated.

The personal computer (PC) industry (Microsoft and Hewlett Packard in particular, see [Stokes et al. \(1996\)](#)) defined an RGB color model or color space, i.e. a gamut of colors, for use with PC monitors. This color model, called sRGB, is the gamut of colors that can be generated on most PC computer monitors. (Not surprisingly, Apple has its own color model, as does Adobe, and there are at least a dozen others.)

The conversion from CIE XYZ to sRGB coordinates is given by

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 3.2404542 & -1.5371385 & -0.4985314 \\ -0.9692660 & 1.8760108 & 0.0415560 \\ 0.0556434 & -0.2040259 & 1.0572252 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} \quad (16.5)$$

In this equation, the (X, Y, Z) values are assumed to be scaled from 0 to 1. The conversion matrix seen here comes from the excellent web site maintained by Bruce Lindbloom. The [Lindbloom web site](#) gives all of the details about color spaces and conversions from one color space to another, including online calculators and downloadable spreadsheets for doing various calculations.

Now consider CIE reference illuminant E, which is pure white, with equal energy at each wavelength. Setting $\Lambda(\lambda) = 1$ for all λ gives a values of $7.30 \cdot 10^4$ for each of X, Y , and Z . Rescaling these values to 1 and using them in Eq. (16.5) gives $(R, G, B) = (1.205, 0.948, 0.909)$, which can be rescaled to $(255, 201, 192)$. This result is not $(255, 255, 255)$ as might be expected for a pure white color. The computed RGB color is actually a pale red or pink. The reason for this color mismatch is that sRGB is based on the D65 reference illuminant, whereas the CIE 1931 chromaticity diagram uses reference illuminant E to define “white.”

If we let $\Lambda(\lambda)$ be the normalized D65 spectrum seen in Fig. 16.11, then $(X, Y, Z) = (58228.0, 61262.5, 66713.0)$. Rescaling these values to a maximum value of 1 and inserting them into Eq. (16.5) then returns each of R, G , and B as 0.918. Multiplying these values by 255 rescales them to $(234, 234, 234)$, which is a very light gray. Keep in mind that gray is just white that is not very bright, so these values can be rescaled to $(255, 255, 255)$ without changing the color. Thus the sRGB color model returns the D65 illuminant as the white

color, and it returns the equal-energy spectrum E as a pink color.

The conversion from sRGB values to XYZ is given by the inverse of Eq. (16.5):

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124564 & 0.3575761 & 0.1804375 \\ 0.2126729 & 0.7151522 & 0.0721750 \\ 0.0193339 & 0.1191920 & 0.9503041 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (16.6)$$

The RGB values in Eq. (16.6) are considered to be between 0 and 1. If we rescale $(R, G, B) = (255, 255, 255)$ to $(1, 1, 1)$ and insert into this equation, it returns $(X, Y, Z) = (0.950, 1.000, 1.089)$, which gives the CIE chromaticity coordinates $(x, y) = (0.313, 0.329)$. This is the slightly bluish color of the D65 spectrum shown by the box symbol in Fig. 16.12.

Using Eq. (16.6) to convert $(R, G, B) = (1, 0, 0)$ gives $(X, Y, Z) = (0.4125, 0.2127, 0.01933)$, which gives $(x, y) = (0.640, 0.330)$. Converting $(0, 1, 0)$ gives $(x, y) = (0.30, 0.60)$, and $(0, 0, 1)$ gives $(x, y) = (0.15, 0.06)$. These three (x, y) values define the vertices of the sRGB gamut triangle seen in Fig. 16.12. It is clear that this gamut cannot accurately reproduce many of the colors visible to the eye (nor can any of the other gamuts used in computer monitors and TVs). Indeed, the sRGB color space represents only 35% of the entire CIE (x, y) color space. In other words, only 35% of the colors as seen by the human eye can be accurately reproduced on a computer monitor using the sRGB color model.

It was seen in Fig. 16.5 that “brown is just red that is not very bright.” In that figure the perceived color went from bright red for $(R, G, B) = (255, 0, 0)$ to a dark reddish brown for $(R, G, B) = (125, 0, 0)$. Those two color patches are reproduced in Fig. 16.13. If the RGB values for these two color patches are rescaled to $(1, 0, 0)$ and $(0.49, 0, 0)$, respectively, and inserted into Eq. (16.6), the resulting CIE chromaticities are the same for each: $(x, y) = (0.640, 0.330)$. What is different is the luminance: $Y = 0.2127$ for the bright red patch, and $Y = 0.1043$ for the dark brown patch. It is likely that most people would call these patches different colors, but in the language of color science, they are the same color but different brightnesses (same chromaticity but different luminances).

There are additional complications to mention in the conversion from XYZ to RGB color specifications.

First, the eye does not respond in a linear fashion to differences in brightness. In particular, the eye is more sensitive to differences in darker color tones than it is to differences in brighter tones. Therefore, a linear spacing of RGB values allocates more of the 0-255 range to high values, which the eye cannot distinguish well, than to low values where the eye sees greater differences for a the same difference in RGB values. Therefore, after using Eq. (16.5) to obtain “linear” RGB values, is it customary to apply a *gamma correction* to the result. A gamma correction is a non-linear “stretching” of the RGB values so as to make better use of the limited range of values (usually 0 to 255) available for storing the color information. The gamma correction for the sRGB color model is given by

$$\begin{aligned} &\text{if } R \leq 0.0031308, \text{ then reset } R \text{ to } 12.92R \\ &\text{if } R > 0.0031308, \text{ then reset } R \text{ to } 1.055R^{1/2.4} - 0.055. \end{aligned}$$

The same transformation is applied to the G and B values. Here the gamma value is $\gamma = 2.4$. After this gamma correction, an input of $(X, Y, Z) = (1, 1, 1)$ results in $(R, G, B) =$

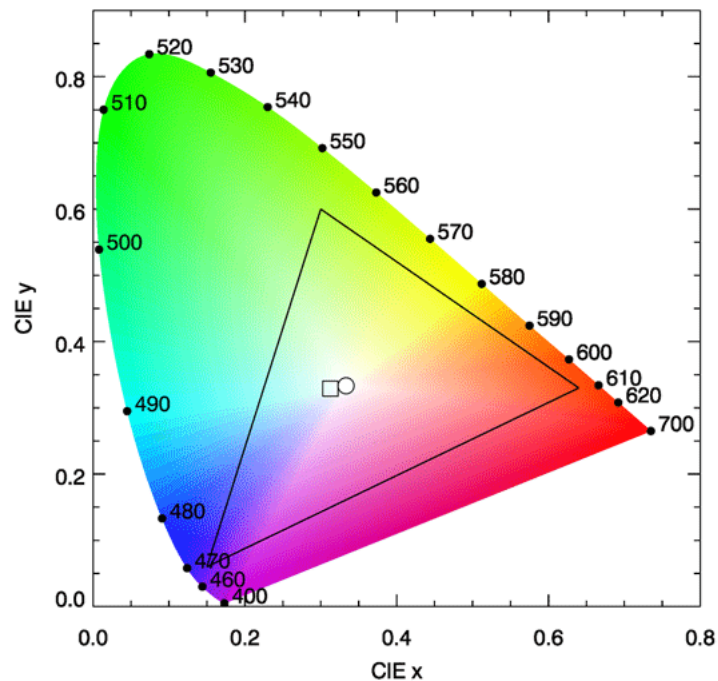


Figure 16.12: The CIE chromaticity diagram with the sRGB gamut shown by the triangle. The circle is the CIE illuminant E white point, and the box is the sRGB D65 illuminant white point.

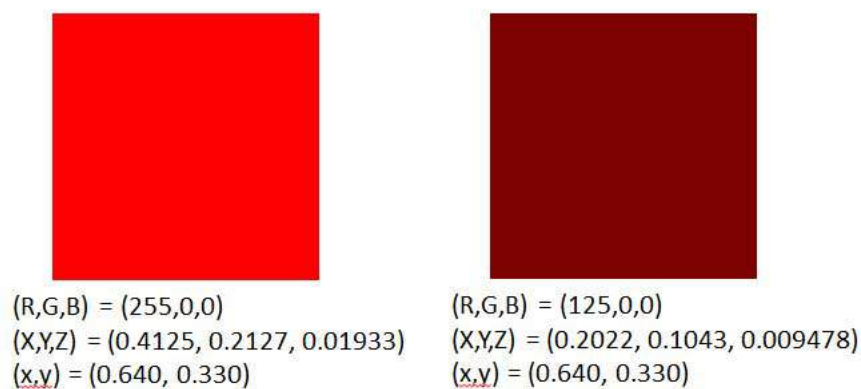


Figure 16.13: Two color patches with the same chromaticities $(x, y) = (0.64, 0.33)$ but with different luminances $Y = 0.2127$ and $Y = 0.1043$.

(255, 230, 225) rather than the values of (255, 201, 192) found above using only Eq.(16.5). Conversely, if gamma-corrected RGB values are to be converted to XYZ values, the gamma correction needs to be undone via

$$\begin{aligned} &\text{if } R \leq 0.04045, \text{ then reset } R \text{ to } R/12.92 \\ &\text{if } R > 0.04045, \text{ then reset } R \text{ to } \frac{(R + 0.055)^{2.4}}{1.055}. \end{aligned}$$

The same transformation is applied to the gamma-corrected G and B values. The resulting “linear” RGB values can then be used as Eq. 16.6.

Second, you often have to make a subjective decision on how to rescale the RGB values to get the right brightness when displayed. For example, [Dierssen et al. \(2006\)](#) found that scaling (X, Y, Z) so that $Y = 0.4$ gave RGB values in the mid-range of brightness and reasonable screen colors from RGB values computed by Eq. (16.5). The value of 0.4 was an ad hoc choice determined by trial and error for visual appearance of ocean radiance spectra colors. As was seen in Fig. 16.13, the brightness affects the perceived color.

Third, note that the 3x3 transformation matrix in Eq. (16.6) has positive elements. Thus any (R, G, B) triplet of positive numbers generates an (X, Y, Z) triplet of positive numbers, i.e. a valid point (x, y) in the CIE chromaticity diagram. That is to say, any RGB color can be represented on the CIE diagram. However, some of the matrix elements in Eq. (16.5) are negative. This means that not every valid (X, Y, Z) or (x, y) can be represented by valid (R, G, B) values; some of the $R, G,$ and B values may be negative. When that happens, about all you can do is set the negative value to 0 and hope that the displayed color is not too different from what your eye would see in nature.

Fourth, color reproduction gets even more complicated when printing. Computer monitors are emitting light, and various colors are generated by adding together different amounts of red, green, and blue light. Thus yellow is generated by turning on the red and green pixels. The RGB system is an additive color system. When viewing colors on a printed page, you are seeing reflected light. The dyes in the ink absorb, or subtract, various colors from the incident (usually white) light. The printing industry thus deals with subtractive colors, e.g., green is generated by subtracting (absorbing) blue and red from white. Color printers thus use the CMYK (cyan, magenta, yellow, black) subtractive color system, which specifies how much ink of various colors to put onto the page, rather than how much light to generate. Some printers have additional ink colors. Thus the color space or gamut for a printer, when plotted on a CIE diagram, will have more than three vertices if the printer uses more than three colors of ink. This can give better coverage of the human vision color space than can a three-color RGB system. Needless to say, converting a computer screen image from RGB to CMYK values is a messy business, and full account of the reference illuminants and color models for the monitor and the printer must be taken into account. This is what is done in PhotoShop, for example, when you select “save as CMYK” on an RGB jpeg file. The resulting printed image may or may not be satisfactory, depending on whether or not the underlying color models are consistent with the printer to be used.

The HydroLight radiative transfer model (Section 10.6) computes the (x, y, Y) values for $L_w, L_u, E_d,$ and E_u whenever the run covers at least the 400-700 nm range. These quantities are uniquely determined for each spectrum. However, HydroLight does not attempt to compute corresponding RGB values because those values are device dependent and may require subjective scaling to get good color reproduction. If you want to turn your spectra into RGB or CMYK values, you can start with the HydroLight-computed

(X, Y, Z) values, but beyond that you are on your own. Note that given (x, y, Y) , you can recover (X, Y, Z) from

$$\begin{aligned} X &= \frac{xY}{y} \\ Y &= Y \\ Z &= \frac{(1-x-y)Y}{y} \end{aligned}$$

where Y is in the range $[0, 1]$.

This is enough to say about colors. This section barely gets you started into the very complicated business of color management and rendering on specific monitors or printers. The previously mentioned web site maintained by Bruce Lindstrom is an excellent place to start if you wish to dig deeper into these topics.

16.5 The Luminance Transfer Equation

The radiative transfer chapter derived the scalar radiative transfer equation (SRTE), Eq. (9.21):

$$\begin{aligned} \cos \theta \frac{dL(z, \theta, \phi, \lambda)}{dz} &= -c(z, \lambda)L(z, \theta, \phi, \lambda) \\ &+ \int_0^{2\pi} \int_0^\pi \beta(z; \theta', \phi' \rightarrow \theta, \phi; \lambda) L(z, \theta', \phi', \lambda) \sin \theta' d\theta' d\phi' \\ &+ S(z, \theta, \phi, \lambda). \end{aligned} \quad (16.7)$$

This equation governs the propagation of unpolarized monochromatic radiance at a particular wavelength λ in a one-dimensional absorbing and scattering medium.

The question now arises: Is there a similar equation for the propagation of luminance? It is to be anticipated that a luminance transfer equation may be more complicated than the SRTE because it of necessity must involve all visible wavelengths and the response of the human eye.

To develop a luminance transfer equation, multiply Eq. (16.7) by the photopic luminosity function $K_{\text{cd}} \bar{y}(\lambda)$ and integrate over all visible wavelengths. Let Λ denote the range of wavelengths for which $\bar{y}(\lambda) > 0$. The term on the left hand side of the SRTE then becomes

$$K_{\text{cd}} \int_{\Lambda} \left\{ \cos \theta \frac{dL(z, \theta, \phi, \lambda)}{dz} \right\} \bar{y}(\lambda) d\lambda = \cos \theta \frac{dL_v(z, \theta, \phi)}{dz},$$

where the luminance L_v is defined by Eq. (16.1):

$$L_v \triangleq K_{\text{cd}} \int_{\Lambda} L(\lambda) \bar{y}(\lambda) d\lambda.$$

The first term on the right hand side of the SRTE becomes

$$K_{\text{cd}} \int_{\Lambda} \{-c(z, \lambda) L(z, \theta, \phi, \lambda)\} \bar{y}(\lambda) d\lambda.$$

This term does not give a product of an integral over wavelength of the beam attenuation coefficient times the luminance. However, we can rewrite this term as

$$\left\{ \frac{K_{\text{cd}} \int_{\Lambda} -c(z, \lambda) L(z, \theta, \phi, \lambda) \bar{y}(\lambda) d\lambda}{K_{\text{cd}} \int_{\Lambda} L(\lambda) \bar{y}(\lambda) d\lambda} \right\} K_{\text{cd}} \int_{\Lambda} L(\lambda) \bar{y}(\lambda) d\lambda.$$

The term in braces is a radiance-weighted integral of the beam attenuation coefficient times the photopic luminosity function \bar{y} . If we define the photopic beam attenuation coefficient c_v as

$$c_v(z, \theta, \phi) \triangleq \frac{K_{\text{cd}} \int_{\Lambda} c(z, \lambda) L(z, \theta, \phi, \lambda) \bar{y}(\lambda) d\lambda}{K_{\text{cd}} \int_{\Lambda} L(\lambda) \bar{y}(\lambda) d\lambda}, \quad (16.8)$$

then the $-cL$ term of the SRTE maintains the same form, $-c_v L_v$, in the luminance transfer equation.

A similar treatment of the path radiance term of the SRTE leads to a definition for the photopic volume scattering function:

$$\beta_v(z, \theta', \phi' \rightarrow \theta, \phi) \triangleq \frac{K_{\text{cd}} \int_{\Lambda} \beta(z, \theta', \phi' \rightarrow \theta, \phi, \lambda) L(z, \theta', \phi', \lambda) \bar{y}(\lambda) d\lambda}{K_{\text{cd}} \int_{\Lambda} L(z, \theta', \phi', \lambda) \bar{y}(\lambda) d\lambda}. \quad (16.9)$$

The source term in the SRTE leads to a photopic source term:

$$S_v(z, \theta, \phi) \triangleq K_{\text{cd}} \int_{\Lambda} S(z, \theta, \phi, \lambda) \bar{y}(\lambda) d\lambda.$$

Collecting the above results gives the desired luminance transfer equation:

$$\begin{aligned} \cos \theta \frac{dL_v(z, \theta, \phi)}{dz} &= -c_v(z, \theta, \phi) L_v(z, \theta, \phi) \\ &+ \int_0^{2\pi} \int_0^{\pi} \beta_v(z; \theta', \phi' \rightarrow \theta, \phi) L_v(z, \theta', \phi') \sin \theta' d\theta' d\phi' \\ &+ S_v(z, \theta, \phi). \end{aligned} \quad (16.10)$$

This equation governs the propagation of broadband luminance as seen by the human eye in a one-dimensional absorbing and scattering medium. Equation (16.10) is the basis for the classical definition of contrast as used in visibility studies.

16.5.1 Dependence of c_v on the Ambient Radiance

It is important to note that the photopic beam attenuation coefficient as defined in Eq. (16.8) depends on the ambient radiance distribution, hence on direction (θ, ϕ) , even though the beam attenuation $c(z, \lambda)$ is an inherent optical property (IOP) that does not depend on the ambient radiance or direction. Moreover, c_v meets the definition of an apparent optical property as defined in Chapter 4: it depends on the IOPs of the medium (here the beam attenuation c) and on the ambient radiance distribution, and it is insensitive to external conditions (e.g., rescaling L by a multiplicative factor does not change the value of c_v). The same holds true for the photopic volume scattering function defined in Eq. (16.9) and for any other IOP. Thus, *in going from a monochromatic radiative transfer equation to a luminance transfer equation, inherent optical properties become apparent optical properties.*

This is the penalty to be paid for going from an equation for monochromatic radiance as measured by instruments to an equation for luminance observed by a human eye.

However, in practice, there seems to be very little dependence of c_v on the ambient radiance (as would be expected for a “good” AOP). The left panel of Fig. 16.14 shows the beam attenuation $c(\lambda)$ for a simulation of homogeneous Case 1 water with a chlorophyll concentration of 0.5 mg m^{-3} (obtained using the new Case 1 IOP model (Section 8.9) in HydroLight). The Sun was at a zenith angle of $\theta_{\text{sun}} = 40 \text{ deg}$ in a clear sky, which gives a transmitted solar beam of about 29 deg in the water; that beam will lie in the HydroLight quad centered at $\theta = 30 \text{ deg}$. The right panel of Fig. 16.14 shows the radiance at 10 meters depth looking in four directions: looking upward into the Sun’s transmitted beam, looking in the nadir and zenith directions, and looking horizontally at right angles to the solar plane.

The spectra in this figure were used to compute the photopic beam attenuation c_v via Eq. (16.8). The values are all close to 0.31 m^{-1} , which is close to the beam attenuation at the peak of the photopic luminosity function: $c(555 \text{ nm}) = 0.313 \text{ m}^{-1}$.

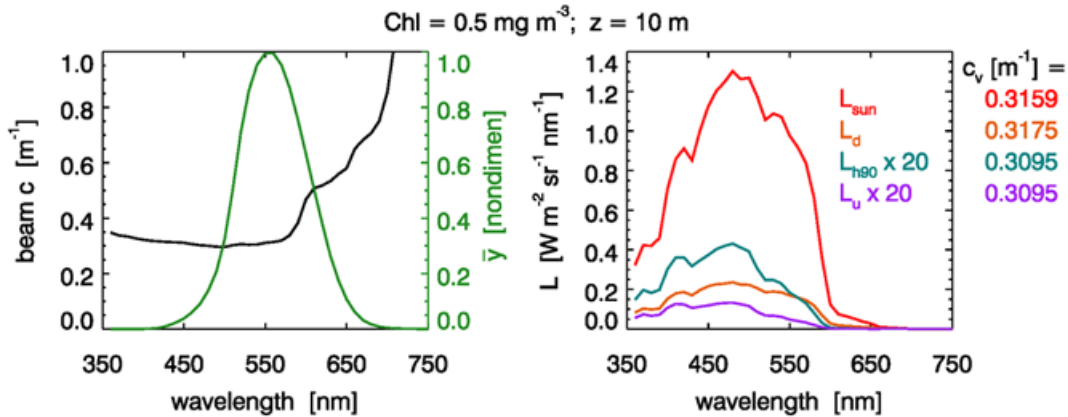


Figure 16.14: Left panel: Total (including water) beam attenuation $c(\lambda)$ for a chlorophyll concentration of 0.5 mg m^{-3} (black curve), and the photopic luminosity function \bar{y} (green). Right panel: Radiances at a depth of 10 m for the Sun at a zenith angle of 40 deg in a clear sky. L_{sun} (red curve) is looking upward into the Sun’s refracted beam. L_{u} (purple) is the upwelling (nadir-viewing) radiance; L_{d} (orange) is the downwelling (zenith-viewing) radiance; and L_{h90} (green) is the horizontal radiance in the direction perpendicular (azimuthal angle of $\phi = 90 \text{ deg}$) to the solar plane. L_{u} and L_{h90} have been multiplied by 20 for better visibility in the plot. Numbers at the right show the photopic beam attenuation c_v for the four radiance spectra.

Figure 16.15 shows the corresponding results for a chlorophyll concentration of 10 mg m^{-3} and a 5 m depth. Again, the four different radiances give the same c_v to within a fraction of a percent, and these c_v values are within one percent of the beam attenuation value $c(555 \text{ nm}) = 2.573 \text{ m}^{-1}$.

Figure 16.16 shows a chlorophyll profile consisting of a background value of 0.5 mg m^{-3} plus a Gaussian that gives a maximum value of 5.5 mg m^{-3} at 10 m depth. For this profile, an observer at 5 m depth looking upward would be looking into low-chlorophyll water, and looking downward would be looking into high-chlorophyll water. An observer at 10

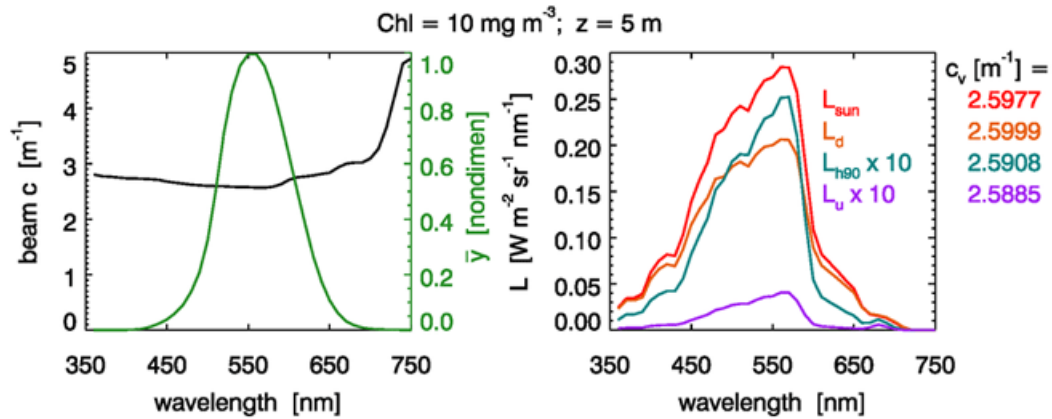


Figure 16.15: Same as for Fig. 16.14, but for a chlorophyll concentration of 10 mg m⁻³ and a 5 m depth.

m depth looking horizontally would be looking into high-chlorophyll water, but looking upward or downward would be looking into lower chlorophyll, clearer water. It might be supposed that the different IOPs ($c(z, \lambda)$ values in particular) would give radiances that might give significantly different c_v values for the different viewing directions at a given depth.

However, this is not the case. Figure 16.17 shows the radiances seen by an observer at 15 m depth. Again, the c_v values differ by only about one percent from the value of $c(15 \text{ m}, 555 \text{ nm}) = 0.719 \text{ m}^{-1}$. The same holds true at other depths (not shown).

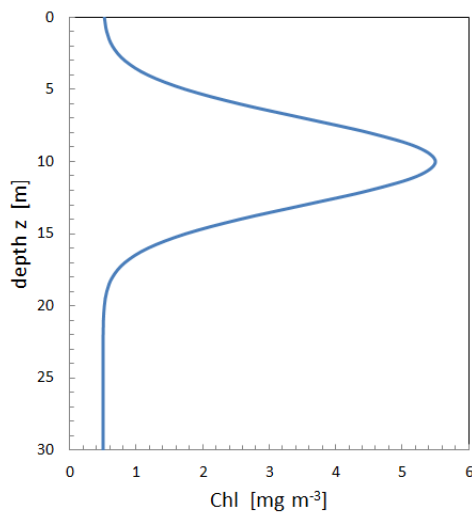


Figure 16.16: The chlorophyll profile used in the simulations of Fig. 16.17.

Exhaustive simulations have not been made, so it might be possible to create a contrived water body for which the photopic beam attenuation would be significantly different for different viewing directions, and for which $c_v(z)$ would be significantly different from $c(z, 555 \text{ nm})$. However, the above simulations indicate that in many situations of practical interest, there is little dependence of c_v on viewing direction, and that c_v is within a percent or so of the beam attenuation at the 555 nm wavelength of the maximum of the photopic

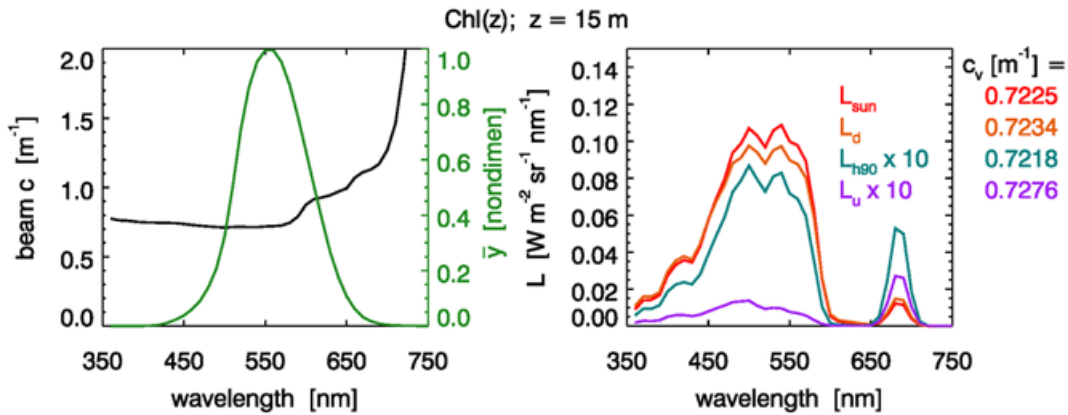


Figure 16.17: Same as for Fig. 16.14, but for the depth-dependent chlorophyll profile seen in Fig. 16.16 and a 15 m depth.

luminosity function.

These simulations are consistent with the results of Zaneveld and Pegau (2003), who found that the beam attenuation coefficient at 532 nm (excluding the water contribution) is a good proxy for the photopic beam attenuation.

16.6 Contrast

The luminance transfer equation developed in the previous section provides the theoretical structure needed to begin discussing the topic of visibility. The development here will be in terms of luminance, as is appropriate for a human observer.

16.6.1 The Luminance Difference Law

Figure 16.18 illustrates the quantities involved with viewing an object through an absorbing and scattering medium under ambient daytime illumination. The object, usually called the target, is being viewed against a background. The luminance leaving the surface of the target and heading in the direction of the observer is $L_{vT}(0, \hat{\xi}_T)$ (subscript “v” for visual and “T” for target). The luminance of the background immediately adjacent to the target is $L_{vB}(0, \hat{\xi}_B)$, where “B” is for background. These two luminances eventually reach the observer, where they are focussed onto different points of the retina (or, in an instrument, perhaps onto adjacent pixels of a CCD array).

Now assume that

1. The object is small and is illuminated by ambient daylight.
2. The luminance leaving the target does not significantly affect the ambient luminance that is present in the absence of the target, and
3. The two directions $\hat{\xi}_T$ and $\hat{\xi}_B$ are almost parallel, i.e. $\hat{\xi}_T \approx \hat{\xi}_B = \hat{\xi}$, so that the absorbing and scattering losses and additions to each beam are the same between the target area and the imaging system. Likewise, any internal sources are the same along each path.

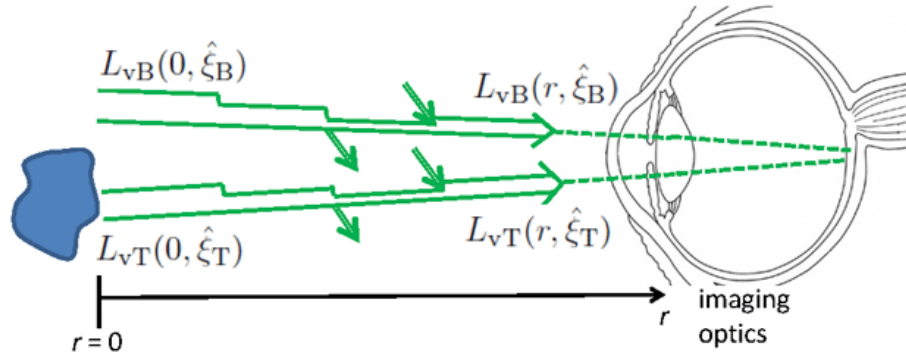


Figure 16.18: Geometry for developing the law of contrast reduction.

Equation (16.10) gives the equation governing propagation of luminance at any point r along either path:

$$\begin{aligned} \frac{dL_v(r, \theta, \phi)}{dr} &= -c_v(r, \theta, \phi) L_v(r, \theta, \phi) \\ &+ \int_0^{2\pi} \int_0^\pi \beta_v(r; \theta', \phi' \rightarrow \theta, \phi) L_v(r, \theta', \phi') \sin \theta' d\theta' d\phi' \\ &+ S_v(r, \theta, \phi) \\ &\triangleq -c_v(r) L(r, \hat{\xi}) + L_v^*(r, \hat{\xi}) + S_v(r, \hat{\xi}). \end{aligned} \quad (16.11)$$

Because the ambient luminance distribution, the IOPs, and any sources are assumed to be the same for each path, the path luminance term L_v^* , which describes broad-band scattering into the beam, and the source term S_v , will be the same for both background and target luminances. Thus the two luminances obey

$$\frac{dL_{vT}(r, \hat{\xi})}{dr} = -c_v(r) L_{vT}(r, \hat{\xi}) + L_v^*(r, \hat{\xi}) + S_v(r, \hat{\xi}) \quad (16.12)$$

$$\frac{dL_{vB}(r, \hat{\xi})}{dr} = -c_v(r) L_{vB}(r, \hat{\xi}) + L_v^*(r, \hat{\xi}) + S_v(r, \hat{\xi}) \quad (16.13)$$

Subtracting Eq. (16.13) from Eq. (16.12) gives

$$\frac{d[L_{vT}(r, \hat{\xi}) - L_{vB}(r, \hat{\xi})]}{dr} = -c_v(r) [L_{vT}(r, \hat{\xi}) - L_{vB}(r, \hat{\xi})].$$

This equation has the solution

$$[L_{vT}(r, \hat{\xi}) - L_{vB}(r, \hat{\xi})] = [L_{vT}(0, \hat{\xi}) - L_{vB}(0, \hat{\xi})] \exp \left[- \int_0^r c_v(r') dr' \right].$$

In most practical situations, $c_v(r)$ can be assumed constant along the viewing path (in the ocean, a few tens of meters at most), in which case this solution reduces to

$$[L_{vT}(r, \hat{\xi}) - L_{vB}(r, \hat{\xi})] = [L_{vT}(0, \hat{\xi}) - L_{vB}(0, \hat{\xi})] \exp[-c_v r]. \quad (16.14)$$

Even though each radiance $L_{vT}(r, \hat{\xi})$ and $L_{vB}(r, \hat{\xi})$ individually depends on all photopic IOPs, their *difference* depends only on the photopic beam attenuation c_v . This is known as the *Luminance Difference Law* (e.g., Preisendorfer, 1986).

16.6.2 The Law of Contrast Reduction

Now define the *apparent visual contrast* between the target and background as seen at distance r as the difference in the target and background luminances normalized by the background luminance:

$$C(r, \hat{\xi}) \triangleq \frac{L_{vT}(r, \hat{\xi}) - L_{vB}(r, \hat{\xi})}{L_{vB}(r, \hat{\xi})}. \quad (16.15)$$

From Eq. (16.14), this can be rewritten as

$$C(r, \hat{\xi}) = \frac{L_{vT}(0, \hat{\xi}) - L_{vB}(0, \hat{\xi})}{L_{vB}(0, \hat{\xi})} \frac{L_{vB}(0, \hat{\xi})}{L_{vB}(r, \hat{\xi})} \exp[-c_v(\hat{\xi}) r] \quad (16.16)$$

$$= C(0, \hat{\xi}) \frac{L_{vB}(0, \hat{\xi})}{L_{vB}(r, \hat{\xi})} \exp[-c_v(\hat{\xi}) r]. \quad (16.17)$$

This result holds for any viewing direction (for the assumptions of the derivation).

$$C(0, \hat{\xi}) \triangleq \frac{L_{vT}(0, \hat{\xi}) - L_{vB}(0, \hat{\xi})}{L_{vB}(0, \hat{\xi})}$$

is the *inherent visual contrast*, i.e., the contrast of the target against the background as seen from zero distance. If the target is darker (lighter) than the background, $C(0)$ is negative (positive). The apparent contrast keeps the same sign as $C(0)$ as r increases and $C(r)$ approaches zero. For a black target, $L_T(0) = 0$, and $C(0) = -1$. For a black background, $L_{vB}(0) = 0$, and the inherent contrast becomes infinite. This situation (such as viewing a distant light at night) violates the present assumption of viewing an object in ambient daylight, and a different analysis is required.

Looking horizontally, the background radiance is independent of the viewing distance if the water is horizontally homogeneous. In this case, the Eq. (16.17) reduces to

$$C(r, \hat{\xi}) = C(0, \hat{\xi}) \exp[-c_v(\hat{\xi}) r]. \quad (16.18)$$

Other viewing directions may also satisfy $L_{vB}(0, \hat{\xi}) \approx L_{vB}(r, \hat{\xi})$. An exception to $L_{vB}(0, \hat{\xi}) \approx L_{vB}(r, \hat{\xi})$ would be looking upward at a target near the sea surface from a larger depth. The downwelling radiance could then be much brighter at the target depth than at the viewing depth.

Either of equations (16.17) or (16.18) is called the *Law of Contrast Reduction* or *Law of Contrast Transmittance*.

For horizontal viewing and a black target, experience shows that the target can be detected at a *visual range* VR when $C(r = VR) \approx 0.02$, although you can find recommended minimum contrast values from 0.01 to 0.05 in the literature. Equation (16.18) can be solved for this minimum contrast to get the corresponding visual range:

$$VR(\hat{\xi}) = -\frac{\ln |0.02|}{c_v(\hat{\xi})} \approx \frac{3.9}{c_v(\hat{\xi})}. \quad (16.19)$$

This result, originally developed for viewing dark targets in the atmosphere, is known as Koschmieder's law. In very clear water ($Chl = 0.01 \text{ mg m}^{-3}$), $c_v \approx 0.04 \text{ m}^{-1}$, which limits

underwater visibility to less than 100 m. In a very clear atmosphere, visibility can be as much as 200 km.

The limiting value of 0.02 for the minimum contrast depends somewhat on factors such as the angular size of the target, whether the target is fading from view as r increases or appearing out of the background as r decreases, and on the observer's visual acuity. In any case, this result is in rough agreement with the experiments of [Zaneveld and Pegau \(2003\)](#) showing that the visibility range of a black Secchi disk 0.2 m in diameter, when viewed horizontally against the background ocean, is about

$$r \approx \frac{4.8}{c_v}.$$

This range corresponds to a minimum detectable contrast of about 0.01, rather than 0.02. Airports use a value 0.05, which gives an added safety factor when judging how far a pilot can see.

The contrast defined by Eq. (16.15) is called the Weber contrast. This is a suitable measure of contrast when viewing a small object against a background. If the target has a pattern of bright and dark features with roughly equal areas of each (like the stripes on a zebra), then a more suitable measure of contrast is the Michelson contrast:

$$C \triangleq \frac{L_{vT}(\max) - L_{vT}(\min)}{L_{vT}(\max) + L_{vT}(\min)}, \quad (16.20)$$

where $L_{vT}(\max)$ and $L_{vT}(\min)$ are respectively the maximum and minimum luminances for the bright and dark areas of the target.

16.7 The Secchi Disk

Father Angelo Secchi, S. J., seen in Fig. 16.19, was primarily an astronomer and spectroscopist, but he also made significant contributions to meteorology and several other fields. Among his many accomplishments, he pioneered the use of spectroscopy as a way to classify stars by their spectral type, he was the first to realize that the Sun is a star, and he was the first to use maps of atmospheric pressure as an aid to weather forecasting. He had only one publication in oceanography, in 1865, but it was a good one. [Pitarch \(2020\)](#) gives a nice overview of Secchi's oceanographic work.

A Secchi disk is a white disk, typically 30 cm in diameter (in oceanographic applications), weighted and attached to a cord marked with the distance from the disk. The disk is lowered into the water and observed as it goes deeper and deeper, as seen in Fig. 16.20. The depth at which it disappears from view is the Secchi depth, z_{SD} . The Secchi depth gives an easily obtained measure of water transparency. However, there are many sources of variability in this measurement. Sky light reflected by the water surface can reduce the visibility (hence a dependence on solar zenith angle and sky conditions), as do waves on the surface, which tend to break up the image of the disk as seen from above the surface. The disk will be easier to see when well illuminated on a bright day than near twilight. Finally, this is one of the few measurements still made in science where the human eye is an integral part of the measurement. Thus if you have better vision than I have, you may be able to see the disk deeper than I can. In spite of all of these uncertainties, the Secchi depth is surprisingly reproducible for different environmental conditions and observers, so

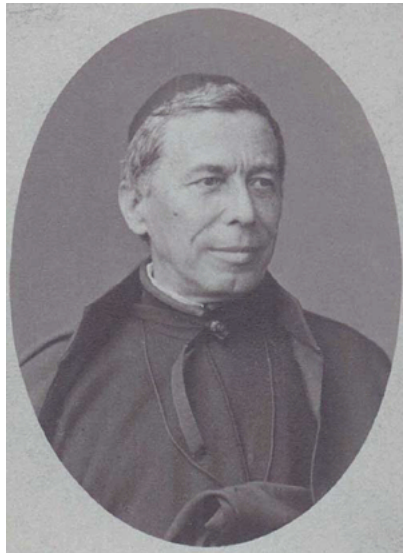


Figure 16.19: Angelo Secchi, 1818-1878.



Figure 16.20: A Secchi disk being lowered into greenish water. Photo by R. Kirby from [The Secchi Disk Study](#).

it is useful for some purposes. There is also a large historical database of Secchi disk readings going back to the days of sailing ships (e.g., the [World Ocean Database](#)), so proper interpretation of Secchi depths may even be useful for studies of long-term changes in ocean waters.

The previous sections on photometry give the background needed to derive the maximum depth at which a Secchi disk can be seen.

16.7.1 The Classical Secchi Depth Model of Preisendorfer

Consider only the case of looking straight down, and drop the direction arguments in luminances and contrasts, e.g. $L_{vB}(z, \hat{\xi}) = L_{vB}(z)$. The underlying idea is that a disk at some depth z is illuminated by the downwelling plane illuminance $E_{dV}(z)$. The luminance reflected by the disk then propagates upward to the observer as a narrow beam of luminance. The development then proceeds as follows.

The downwelling plane illuminance at depth z is given by

$$E_{\text{dv}}(z) = E_{\text{dv}}(0) \exp[-\langle K_{\text{dv}} \rangle_z z], \quad (16.21)$$

where $\langle \dots \rangle_z$ denotes the average over 0 to z .

The target is assumed to be a Lambertian reflector (Section 13.7) with an illuminance reflectance of R_{vT} . The luminance reflected by the target is then

$$L_{\text{vT}}(z) = E_{\text{dv}}(z) R_{\text{vT}}/\pi. \quad (16.22)$$

The background water is also assumed to be a Lambertian reflector, so that

$$R_{\text{vB}}(z) = \frac{E_{\text{uv}}(z)}{E_{\text{dv}}(z)}. \quad (16.23)$$

The luminance of the background water is then

$$L_{\text{vB}}(z) = E_{\text{dv}}(z) R_{\text{vB}}(z)/\pi. \quad (16.24)$$

The inherent contrast at depth z is

$$\begin{aligned} C_{\text{in}}(z) &= \frac{L_{\text{vT}}(z) - L_{\text{vB}}(z)}{L_{\text{vB}}(z)} \\ &= \frac{R_{\text{vT}} - R_{\text{vB}}(z)}{R_{\text{vB}}(z)} \end{aligned} \quad (16.25)$$

where the last equation follows from (16.22) and (16.24).

The apparent contrast of the Secchi disk as seen from just below the sea surface is

$$C_{\text{ap}}(0) = \frac{L_{\text{vT}}(0) - L_{\text{vB}}(0)}{L_{\text{vB}}(0)} \quad (16.26)$$

(Note that in this development the argument 0 refers to depth z , not to the distance from the target, which is z .)

The luminance difference law

$$L_{\text{vT}}(0) - L_{\text{vB}}(0) = [L_{\text{vT}}(z) - L_{\text{vB}}(z)] \exp[-\langle c_{\text{v}} \rangle_z z] \quad (16.27)$$

allows the apparent contrast to be written as

$$C_{\text{ap}}(0) = \frac{[L_{\text{vT}}(z) - L_{\text{vB}}(z)]}{L_{\text{vB}}(0)} \exp[-\langle c_{\text{v}} \rangle_z z] \quad (16.28)$$

by (16.27) into (16.26).

Inserting (16.22) and (16.24) into (16.28) then gives

$$C_{\text{ap}}(0) = \frac{R_{\text{vT}} - R_{\text{vB}}(z)}{R_{\text{vB}}(0)} \frac{E_{\text{dv}}(z)}{E_{\text{dv}}(0)} \exp[-\langle c_{\text{v}} \rangle_z z] \quad (16.29)$$

Assuming that $R_{\text{vB}}(0) = R_{\text{vB}}(z)$ and using (16.21) and (16.25) gives

$$C_{\text{ap}}(0) = C_{\text{in}}(z) \exp[-(\langle K_{\text{dv}} \rangle_z + \langle c_{\text{v}} \rangle_z) z] \quad (16.30)$$

This equation gives the apparent contrast of the Secchi disk as seen from just below the water surface. For viewing from above the surface, we must account for loss of contrast caused by the water surface. This loss is due both to refraction by waves and to surface-reflected sky light. Thus

$$C_{\text{ap}}(\text{air}) = \mathcal{T}C_{\text{ap}}(0) = \mathcal{T}C_{\text{in}}(z) \exp[-(\langle K_{\text{dv}} \rangle_z + \langle c_v \rangle_z) z]$$

where \mathcal{T} denotes the transmission of *contrast*, not of luminance or illuminance.

The Secchi depth z_{SD} is the depth at which the apparent contrast in air falls below a threshold contrast C_T . Solving for z_{SD} when $C_{\text{ap}}(\text{air}) = C_T$ gives

$$z_{\text{SD}} = \frac{\ln \left[\frac{\mathcal{T}C_{\text{in}}(z)}{C_T} \right]}{\langle K_{\text{dv}} \rangle_{z_{\text{SD}}} + \langle c_v \rangle_{z_{\text{SD}}}} \quad (16.31)$$

$$\triangleq \frac{\Gamma}{\langle K_{\text{dv}} \rangle_{z_{\text{SD}}} + \langle c_v \rangle_{z_{\text{SD}}}}. \quad (16.32)$$

Studies with human observers show that C_T depends on the angular subtense of the disk and on the ambient luminance (e.g., [Preisendorfer, 1986](#), Table 1). The values of Γ vary from about 6 to 9 for a disk with $R_{\text{vT}} = 0.85$, depending on the water reflectance R_{vB} (which is 0.015 to 0.1; see [Preisendorfer \(1986, Table 2\)](#)). HydroLight uses $\Gamma = 8$ as its default.

Note that Eq. (16.32) must be solved iteratively because $\langle K_{\text{dv}} \rangle_{z_{\text{SD}}}$ and $\langle c_v \rangle_{z_{\text{SD}}}$ are averages over the (unknown) Secchi depth z_{SD} . This is easily done after solution of the radiative transfer equation to some depth greater than z_{SD} over the visible wavelengths. The photopic $K_{\text{dv}}(z)$ and $c_v(z)$ can then be computed from $E_{\text{d}}(z, \lambda)$ and $c(z, \lambda)$. The values of K_{dv} and c_v just below the water surface (at depth 0) are then used to get an initial estimate of z_{SD} , which is then used to compute an improved estimate of the depth-averaged K_{dv} and c_v , and so on. Convergence is obtained within a few iterations.

16.7.2 The Secchi Depth Model of Lee et al.

Preisendorfer's analysis does not consider variability in z_{SD} due to factors like the disk diameter or the solar zenith angle. Therefore, [Lee et al. \(2015\)](#) re-examined the classic theory of the Secchi disk. They assumed that

- The disk needs not be angularly small and can perturb the ambient light field seen near the edge of the disk.
- Visibility is not based on target vs background *luminance* differences at the sharp edge of the disk, but on differences in target and background *reflectances*.
- Visibility is determined by the wavelength where the disk is most visible (which can change with depth and between water bodies), rather than on broadband photopic variables.

They argue that the classic analysis should

- Replace the photopic $K_{\text{dv}}(z)$ with $K_{\text{d}}(z, \lambda_{\text{min}})$, where λ_{min} is the wavelength at which $K_{\text{d}}(z, \lambda)$ is a minimum; and
- Replace the photopic $c_v(z)$ with $1.5K_{\text{d}}(z, \lambda_{\text{min}})$.

One end result of their analysis is a formula of the form (Eq. 28 of their paper)

$$z_{SD} = \frac{\gamma}{2.5K_d(z, \lambda_{\min})}, \quad (16.33)$$

where γ depends on a difference in reflectances, rather than on contrasts as seen in Eq. (16.31). This formula has the great virtue that $K_d(z, \lambda_{\min})$ can be estimated from multi- or hyperspectral satellite imagery.

Comparison of z_{SD} measured and computed by Eq. (16.33) gives reasonable agreement (see Fig. 6 of their paper). However, comparison of Lee et al. z_{SD} predictions with those of the classic theory have not yet been made.

16.7.3 Black-and-White Secchi Disks

Secchi himself, and oceanographers ever since, used an all-white disk, usually 30 cm in diameter (Secchi also measured the depths at which colored disks disappeared). However, limnologists almost always use a black-and-white disk, usually 20 cm in diameter. An example is seen in Fig. 16.21. The use of a disk with black and white quadrants traces back to a civil engineer, G. C. Whipple, who described such a disk in a book published in 1899 (Whipple, 1899, page 115). This “Whipple disk” then became the standard for work in fresh water. There seems to be no good reason to pick an all-white versus a black-and-white disk. A qualitative argument has been made that the black quadrants give a “standard reference” for comparison with the white quadrants when used in shallow water that may be affected by bottom reflectance (but if bottom reflectance is detectable, the Secchi depth is probably deeper than the bottom and a measurement cannot be made). Although Hou et al. (2007) give a rigorous analysis of why a Secchi disk disappears based on spatial frequencies of the imaged disk, they treat only the all-white disk.

Oceanographers and limnologists generally live in different worlds and never speak to each other, so I have never seen a comparison of Secchi depth measured by all-white and black-and-white disks employed at the same time in the same water.



Figure 16.21: A black-and-white Secchi disk as used in limnology.

APPENDIX A

Fourier Transforms

Jean-Baptiste Joseph Fourier (1768-1830) proved in the early 1800s that, under very general mathematical conditions, an arbitrary function can be written as a sum of sines and cosines of different amplitudes and wavelengths¹. Fourier Transforms are today a powerful mathematical tool used in many areas of science to “decompose” functions of time or space into functions depending on temporal or spatial frequency. In this text, they will be used both to decompose random sea surfaces into sums of sinusoids of different spatial or temporal frequencies and to simulate random sea surfaces using models for the wave energy contained surface waves of different frequencies. They will also be fundamental to the prediction of underwater visibility and imaging. In other fields, Fourier Transforms are ubiquitous in electrical engineering (e.g., analysis of time-dependent currents and voltages), economics (analysis of times series of economic indicators), and medical imaging (analysis of 2-D imagery such as CAT scans and MRIs).

There are many texts on Fourier transforms. [Bracewell \(1986\)](#) is a standard reference, and excellent sets of lecture notes and videos of lectures can be found on the web. You can take a class in Fourier transforms and prove many beautiful theorems about their properties. However, this introduction gives all of the mathematical tools needed to describe wind-blown sea surfaces or to analyze images. The results needed for the applications in this book will be stated without proof.

A.1 Notation for Fourier Transforms

This Appendix on Fourier Transforms is rather mathematical, and precise notation is needed to keep everything straight. [Table A.1](#) therefore collects for reference the notation used throughout the discussion of Fourier transforms and their application to the generation

¹Fourier was interested in computing the temperature $T(x, t)$ of a metal bar as a function of distance x along the bar and time t , if the end of the bar at $x = 0$ is heated by an arbitrary function of time $T(0, t)$. He found a solution if the applied temperature is a sinusoid, $T(0, t) = T_o \sin(\omega t)$. He then reasoned that if an arbitrary function $T(0, t)$ can be written as a sum of sinusoids, each with its own solution, and if the solutions could be linearly added, then the solution to the arbitrary problem could be written as a sum of the solutions for sinusoids of different frequencies. This let him to study how arbitrary functions can be written as sums of sinusoids.

of random sea surfaces in Appendices B, C, and D.

Symbol	Definition	Units
1-D Geometry, Continuous Transforms		
x	continuous spatial variable	m
$z(x)$	sea surface elevation at point x	m
ν	continuous spatial frequency	m^{-1}
$k = 2\pi\nu$	continuous angular spatial frequency	rad m^{-1}
$\hat{z}(\nu)$	continuous Fourier transform of $z(x)$	m^2
1-D Geometry, Discrete Transforms		
L	finite length of a sea surface region	m
N	number of grid points in spatial samples	nondimen
$\Delta x = L/N$	grid point spacing	m
$r = 0, 1, \dots, N - 1$	spatial index for discrete samples	nondimen
$x_r = r\Delta x$	spatial locations of discrete samples	m
$z(x_r) \triangleq z(r)$	sea surface elevation at sample point x_r	m
$u = 0, 1, \dots, N - 1$	frequency index for discrete samples	nondimen
$\Delta\nu = 1/L$	fundamental discrete frequency interval	m^{-1}
$\nu_u = u\Delta\nu$	discrete spatial frequencies	m^{-1}
$k_u = 2\pi\nu_u$	discrete angular spatial frequency	rad m^{-1}
$\hat{z}(\nu_u) \triangleq \hat{z}(u)$	discrete Fourier transform of $z(x_r)$	m
2-D Geometry, Discrete Transforms		
L_x, L_y	dimensions of a 2-D sea surface region	m
N_x, N_y	number of grid points in (x, y) directions	nondimen
$(\Delta x, \Delta y) = (L_x/N_x, L_y/N_y)$	grid point spacing in x and y directions	m
$r = 0, 1, \dots, N_x - 1; s = 0, 1, \dots, N_y - 1$	spatial indices for discrete samples	nondimen
$(x_r = r\Delta x, y_s = s\Delta y)$	spatial locations of discrete samples	m
$z(x_r, y_s) \triangleq z(r, s)$	sea surface elevation at sample point (x_r, y_s)	m
$u = 0, 1, \dots, N_x - 1; v = 0, 1, \dots, N_y - 1$	frequency indices for discrete samples	nondimen
$(\nu_u, \nu_v) = (u/L_x, v/L_y)$	spatial frequencies in the (x, y) directions	m^{-1}
$(k_u, k_v) = 2\pi(\nu_u, \nu_v)$	discrete angular spatial frequencies	rad m^{-1}
$\hat{z}(\nu_u, \nu_v) \triangleq \hat{z}(u, v)$	discrete Fourier transform of $z(x_r, y_s)$	m

Table A.1: Notation for Fourier transforms. See also page 644 and Table B.1.

A.2 Continuous Fourier Transforms

Given a real function $f(x)$ of a continuous variable x , the Fourier transform $\hat{f}(\nu)$ of $f(x)$ is defined as

$$\hat{f}(\nu) \triangleq \mathcal{F}\{f(x)\} \triangleq \int_{-\infty}^{\infty} f(x)e^{-i2\pi\nu x} dx. \quad (\text{A.1})$$

The inverse Fourier transform is given by

$$f(x) \triangleq \mathcal{F}^{-1}\{\hat{f}(\nu)\} \triangleq \int_{-\infty}^{\infty} \hat{f}(\nu)e^{+i2\pi\nu x} d\nu. \quad (\text{A.2})$$

It can be shown that if we insert the $\hat{f}(\nu)$ integral of Eq. (A.1) into Eq. (A.2), we recover $f(x)$. This is known as Fourier's integral theorem, the proof of which is not trivial. Equations (A.1) and (A.2) are termed a Fourier transform pair.

Understanding the units of a Fourier transform is important. Suppose that $f(x)$ is sea surface elevation and x is horizontal position, and that both f and x have units of meters. Equation (A.1) shows that $\hat{f}(\nu)$ thus has units of m^2 . The variance of f also has units of m^2 , which gives the first hint at a profound connection between the variance of a physical quantity and its Fourier transform. The units of m^2 in the Fourier transform can be rewritten as $\text{m}/(1/\text{m})$, which is units of z divided by units of spatial frequency ν . Indeed, the transform $\hat{f}(\nu)$ can be interpreted as showing “how much of f there is per unit frequency interval.” The inverse transform then has units of (f over frequency) times frequency, which returns the original units of f .

A Fourier transform is a *spectral density function*. The integral of a spectral density function over a given frequency interval gives the variance in the physical quantity contributed by the frequencies in the integration interval. Density functions are rather peculiar mathematical creatures compared to point functions, which simply give the value of something at a given value of the independent variable (e.g. the temperature at location x and time t). The blackbody radiation function is another example of a spectral density function. The blackbody radiation function discussed in Section 2.3 shows how much energy is emitted (at a given temperature) per unit frequency interval of the emitted electromagnetic radiation. If you are not familiar with the distinction between point and density functions, especially regarding how to change variables in density functions, you should take a look at Sections 2.5 and 2.3 before continuing with the present discussion.

The Fourier transform definitions above with the 2π in the exponents are those of the “Stanford school” of Bracewell (1986) and Goodman (1996). You will see others in the literature. For example, if we use $k = 2\pi\nu$ as the frequency variable, then Eqs. (A.1) and (A.2) become

$$\hat{f}(k) \triangleq \mathcal{F}\{f(x)\} \triangleq \int_{-\infty}^{\infty} f(x)e^{-ikx} dx \quad (\text{A.3})$$

and

$$f(x) \triangleq \mathcal{F}^{-1}\{\hat{f}(k)\} \triangleq \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(k)e^{+ikx} dk. \quad (\text{A.4})$$

This reappearance of the 2π in the second equation is required so that the inverse transform of the transform gets you back to where you started. In the $e^{\pm ikx}$ version, some people put the $1/2\pi$ in front of the other integral, and some put a $1/\sqrt{2\pi}$ in front of each integral. Some authors, e.g. Press et al. (1992), put the $+i$ in Eq. (A.1) and the $-i$ in Eq. (A.2).

The choice of which sign to use on the i and where to put the 2π is almost a religion—most people stay with what they first learned, are convinced of the superiority of their definition, and are willing to die rather than change. Fortunately, it doesn't matter which definitions you use, so long as you are consistent in how the transform pair is defined so that you get back to where you started if you inverse transform a transform, or vice versa.

In our work, $f(x)$ is usually the sea surface elevation, which is a real number. However, even though $f(x)$ is real, $\hat{f}(\nu)$ (or $\hat{f}(k)$) is complex. Expanding the complex exponential in Eq. (A.1) as the sum of a cosine and a sine, it is easy to see that $\hat{f}^*(\nu) = \hat{f}(-\nu)$, where the $*$ denotes the complex conjugate². Such functions are called *Hermitian*. *A real function has a Hermitian Fourier transform. Conversely, if a function is Hermitian, it has a real inverse Fourier transform.* The Hermitian property is an important constraint in the generation of random realizations of sea surfaces by computing the inverse Fourier transform of a complex function: we will have to conjure up a Hermitian function \hat{f} so that we end up with a real sea surface.

A.3 Discrete Fourier Transforms

Now suppose that we have sampled the sea surface $f(x)$ at a set of N evenly spaced points $x_r, r = 0, 1, \dots, N - 1$, in a region of size L ; $x_r = r\Delta x = rL/N$. We want to describe this sampled sea surface $f(x_r)$ as a sum of sinusoids. In general, these N values can be represented as a sum of a constant term, $N/2$ cosine terms, and $N/2 - 1$ sine terms (there is no two-point sine term; see Section B.1.1):

$$f(x_r) = \frac{a_0}{2} + \sum_{u=1}^{N/2} [a_u \cos(k_u x_r) + b_u \sin(k_u x_r)], \quad (\text{A.5})$$

with $b_{N/2} \triangleq 0$. Note that the sum runs from the fundamental frequency ($u = 1, k_1 = 2\pi/L$) through the Nyquist frequency ($u = N/2, k_{N/2} = 2\pi/(2\Delta x)$), with only a cosine term for the two-point wave. This sum is equivalent to

$$f(x_r) = \sum_{u=-N/2+1}^{N/2} c_u e^{ik_u x_r}, \quad (\text{A.6})$$

which also contains a total of N independent real and imaginary parts of the c_n coefficients. (Equation B.5 of the next appendix shows that $c_{-k} = c_k^*$, so these coefficients are not independent for $+k$ and $-k$ pairs.) These equations are the discrete-variable forms of Eqs. (A.1) and (A.2) of the previous section.

Comment on notation: It is common to use the letters i, j, k for dummy summation indices. However, we've already used i for $\sqrt{-1}$ and k for angular wavenumber, so the preceding equations would be hopelessly confusing if we reused i and k for summation indices. We will therefore use r and s for indices on spatial variables, e.g., (x_r, y_s) , and u and v for indices on frequency variables, e.g., k_u or ν_v . n and m also will be used as needed for dummy indices.

²Complex conjugation means replace i by $-i$. Thus the complex conjugate of $x \pm iy$ is $x \mp iy$.

We now have a finite number N of discrete samples $f(x_r)$ of the sea surface, so we need a discrete form of the Fourier transform. The discrete Fourier transform (DFT) of $f(x_r)$ is defined as

$$\hat{f}(\nu_u) \triangleq \mathcal{D}\{f(x_r)\} \triangleq \frac{1}{N} \sum_{r=0}^{N-1} f(x_r) e^{-i2\pi\nu_u x_r} .$$

for $u = 0, 1, \dots, N - 1$. Recalling that $\nu_u = u/L = u/(N\Delta x)$ and $x_r = r\Delta x = rL/N$ gives $\nu_u x_r = ru/N$. It is also common to write $f(x_r) = f(r)$ and $\hat{f}(\nu_u) = \hat{f}(u)$, in which case the previous equation becomes

$$\hat{f}(u) \triangleq \mathcal{D}\{f(r)\} \triangleq \frac{1}{N} \sum_{r=0}^{N-1} f(r) e^{-i2\pi ru/N} . \quad (\text{A.7})$$

The corresponding inverse discrete Fourier transform is given by

$$f(r) \triangleq \mathcal{D}^{-1}\{\hat{f}(u)\} \triangleq \sum_{u=0}^{N-1} \hat{f}(u) e^{+i2\pi ru/N} . \quad (\text{A.8})$$

It was emphasized above that the continuous function $\hat{f}(\nu)$ defined by the continuous Fourier transform (A.1) is a density function with units of $f(x)$ per spatial frequency interval, e.g., $\text{m}/(1/\text{m})$ if z is sea surface height in meters. However, the discrete function $\hat{f}(u)$ defined by the discrete Fourier transform (A.7) has the same units as $f(r)$. The discrete Fourier transform is a point function that shows how much of $f(r)$ is contained in a *finite* frequency interval $\Delta\nu = 1/L$ centered at frequency $\nu_u = u/L$. Discrete Fourier transforms convert point functions $f(r)$ to point functions $\hat{f}(u)$.

The \mathcal{F} and \mathcal{D} notations will be used to distinguish continuous versus discrete Fourier transforms. As just seen, the continuous and discrete transforms are different mathematical constructs with different units and interpretations; they must not be confused. Likewise, if necessary, a subscript can be appended to show the frequency variable, e.g., $\mathcal{F}_\nu\{f(x)\}$ as in Eq. (A.1) or $\mathcal{F}_k\{f(x)\}$ as in Eq. (A.3). As always, there are competing definitions. Equations (A.7) and (A.8) are used in Bracewell and the IDL computer language. *Numerical Recipes* (Press et al., 1992) interchanges the i and $-i$. MATLAB puts the $1/N$ factor on the inverse transform. Also, MATLAB does not support array indices of 0, so the summation indices are shifted from 0 to $N - 1$ to 1 to N , with a corresponding shift from ru to $(r - 1)(u - 1)$ in the exponentials. The devil is in the details, and details like where to put the $1/N$ factor and differences in array indexing in different computer languages can cause great misery when it comes time to actually write computer programs or to compare results computed by different canned subroutines.

The sums in the last two equations require computing complex exponentials (i.e., sines and cosines), multiplying by the corresponding values of $f(r)$ or $\hat{f}(u)$ and adding up the results. These equations can be evaluated for any value of N . The number of computations required to do this is of order N^2 . The computation time thus increases very rapidly for large N .

However, a classic paper by Cooley and Tukey (1965) showed how these sums can be computed in order $N \log_2 N$ computations, if N is a power of 2. Their technique is now called the *Fast Fourier Transform* or FFT. The difference in computer time becomes enormous for large N . For example, if $N = 2^{12} = 4096$, then $N \log_2 N = 4096 \times 12$, and

the difference in computation time is a factor of $N^2/(N \log_2 N) \approx 341$. Thus in the case of $N = 4096$, a roughly 6 minute computer run becomes a 1 second run. The development (or, perhaps, reinvention, since the basic idea goes all the way back to Gauss) of the FFT was a major advance in numerical analysis, which enables the computations on the following sections to be performed extremely efficiently. Subroutines for computing FFTs and inverse FFTs are available in all computer languages commonly used in science (Fortran, C, IDL, MATLAB, Python, etc). Fortunately we do not need to concern ourselves here with the details of how the FFT algorithm actually works, any more than we need to worry about how a canned subroutine actually computes the cosine of a number. If you are interested in how the FFT works, a web search will yield many detailed explanations. It is important to remember that the FFT is not another type of transform; the FFT is a numerically efficient way to evaluate the DFT if the number of data values is a power of two.

The one-dimensional (1-D) equations seen above are easily extended to two or more dimensions. For two dimensions (x, y) , we can sample a region of size L_x by L_y meters over N_x points in the x direction and N_y points in the y direction, with N_x and N_y both powers of 2 so we can use FFTs. Equations (A.5) and (A.6) then become

$$\begin{aligned} f(x_r, y_s) &= \frac{a_0}{2} + \sum_{u=1}^{N_x/2} \sum_{v=1}^{N_y/2} [a_{u,v} \cos(k_u x_r + k_v y_s) + b_{u,v} \sin(k_u x_r + k_v y_s)] \\ &= \sum_{u=-N_x/2+1}^{N_x/2} \sum_{v=-N_y/2+1}^{N_y/2} c_{u,v} e^{i(k_u x_r + k_v y_s)}. \end{aligned} \quad (\text{A.9})$$

The corresponding 2-D DFT pair is

$$\hat{f}(u, v) \triangleq \mathcal{D}\{f(r, s)\} \triangleq \frac{1}{N_x N_y} \sum_{r=0}^{N_x-1} \sum_{s=0}^{N_y-1} f(r, s) e^{-i2\pi(ru/N_x + sv/N_y)} \quad (\text{A.10})$$

and

$$f(r, s) \triangleq \mathcal{D}^{-1}\{\hat{f}(u, v)\} \triangleq \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} \hat{f}(u, v) e^{+i2\pi(ru/N_x + sv/N_y)}. \quad (\text{A.11})$$

It will be important below to keep notational track of continuous versus discrete versions of various functions. For any variable \mathcal{S} ,

- $\mathcal{S}(k)$ will denote a continuous function of k ,
- $\mathcal{S}(k = k_u)$ will denote the continuous function evaluated at the discrete value k_u , and
- $\mathcal{S}(k_u) = \mathcal{S}(u)$ will denote a discrete point function.

Keep in mind that the density function $\mathcal{S}(k = k_u)$ and the point function $\mathcal{S}(k_u)$ have different units. In the next appendix, $\mathcal{S}(k)$ and $\mathcal{S}(k_u)$ will denote the continuous and discrete versions, respectively, of wave elevation variance spectra.

The differences in units between continuous and discrete Fourier amplitudes makes it tricky to make the transition between discrete and continuous versions of the same quantity. In particular, it will be necessary to explicitly include the frequency intervals in some of the later calculations that involve both continuous and discrete variables. For example, if

we have a continuous density function and we need to convert to a corresponding discrete function, we must multiply the continuous function by the finite frequency interval, e.g.

$$\hat{f}(u) = \hat{f}(\nu = \nu_u) \Delta\nu. \quad (\text{A.12})$$

Conversely, if we have discrete amplitudes $\hat{f}(u)$ and we wish to estimate the continuous spectral density $\hat{f}(\nu)$, then we must divide by the frequency interval:

$$\hat{f}(\nu) = \hat{f}(u) / \Delta\nu. \quad (\text{A.13})$$

If you are an optical oceanographer familiar with the scattering phase function, you can find an analogous situation in the estimation of the scattering phase function from measurements of scattered light. The scattering phase function is a measure of how much light energy is scattered from an incident direction into a particular direction, per unit solid angle; it therefore has units of 1/steradian. If you measure the scattered light using an instrument with a finite solid angle $\Delta\Omega$, then you get the total amount of energy scattered into the solid angle $\Delta\Omega$. To estimate the phase function from this measurement, you must divide the measured value by the solid angle of the instrument; this gets you back to units of 1/steradian.

A.3.1 Frequency Ordering

There is a peculiarity to most (perhaps all) FFT subroutines. The discrete FFT of Eq. (A.7) returns an array of N complex numbers $\hat{f}(u)$, which are associated with N discrete spatial frequencies. What is peculiar is the order in which the elements of the $\hat{f}(u)$ array are returned by an FFT subroutine.

Let Δ_f represent the fundamental frequency. If wavenumber ν is the frequency variable, then $\Delta_f = 1/L$. If angular spatial frequency k is the frequency variable, then $\Delta_f = 2\pi/L$; for temporal angular frequency ω , $\Delta_f = 2\pi/T$. In any case, the discrete frequencies associated with the discrete Fourier amplitudes are equally spaced at intervals of Δ_f and are in the negative-to-positive order

$$\left\{ -\frac{N}{2} + 1, -\frac{N}{2} + 2, \dots, -1, 0, 1, \dots, \frac{N}{2} - 1, \frac{N}{2} \right\} \Delta_f. \quad (\text{A.14})$$

I'll call this the "math frequency order" because this is the natural order of arranging values in mathematics. This frequency order is convenient for plotting all of the amplitudes. Plots showing both negative and positive frequencies are called "two-sided," and examples will be seen in the discussion of modeling sea surfaces.

However, FFT routines generally return their amplitudes corresponding to frequencies in the order of 0 first, then the positive frequencies from the smallest to the Nyquist frequency (Section B.1.1), then the negative frequencies in reverse order:

$$\left\{ 0, 1, \dots, \frac{N}{2} - 1, \frac{N}{2}, -\frac{N}{2} + 1, -\frac{N}{2} + 2, \dots, -1, \right\} \Delta_f. \quad (\text{A.15})$$

I'll call this the "FFT frequency order." Given the Hermitian symmetry of the amplitudes about the 0 frequency, the FFT order is convenient for plotting amplitudes only for the positive frequencies, with the negative-positive symmetry of \hat{f} understood. Such plots are called "one-sided" or "folded."

Either frequency order can be obtained from the other by a repeated *circular shift*, which moves an array element off of one end of an array and copies it to the other end of the array, shifting all elements right or left by one position in the process. The detail to watch is that different computer languages define a circular shift in different ways. For example, the IDL routine SHIFT (and the MATLAB routine CIRCSHIFT) moves the array elements to the *right* for a “positive” shift (a negative shift moves elements to the left), whereas the Fortran 95 CSHIFT routine moves the array elements to the *left* for a positive shift (a negative shift moves elements to the right). Thus

In IDL:

```
math order = SHIFT(FFT order, N/2-1)
FFT order = SHIFT(math order, -(N/2-1))
```

In Fortran 95:

```
math order = CSHIFT(FFT order, -(N/2-1))
FFT order = CSHIFT(math order, N/2-1)
```

Some FFT routines allow the user to set a flag specifying which frequency order is to be returned. In any case, sorting out the frequency order of the amplitudes returned by a particular FFT routine, and figuring out how to shift between math and FFT frequency orders in a particular computer language, can drive you to tears.

Note finally that in Eq. (A.7) you are simply providing an array of $f(r)$ values and getting back the same number of $\tilde{f}(u)$ values. What $x(r)$ values correspond to the $f(r)$ values is irrelevant. That is, $x(r)$, $r = 0, \dots, N - 1$ might correspond to the spatial range from $x = 0$ to L , or from $x = -L/2$ to $L/2$, or to any other x range. It is only the number of samples N and the corresponding $f(u)$ values that matters. In other words, the amplitudes coming out of the Fourier transform are independent of the origin of the spatial coordinate system. The frequencies depend on both the number of points N and the physical size of the sampled region via the presence of L (or time interval T) in the fundamental frequency Δ_f . Thus the size of the region sampled and the number of samples, along with the sample values themselves, fully define the associated Fourier transforms.

A.3.2 Interpretation of Negative Frequencies

The appearance of negative frequencies in Fourier transforms may seem somewhat mysterious. Frequency, after all, is a physical quantity that is inherently a positive number, e.g., the number of wavelengths in a given distance. However, one way to interpret mathematically negative frequencies is that they are simply the mathematical price we pay for the convenience of using complex numbers. Consider, for example, the representation of the cosine as a sum of complex exponentials for the u^{th} frequency:

$$\begin{aligned} \cos(2\pi\nu_u x_r) &= \frac{1}{2} \left(e^{i2\pi\nu_u x_r} + e^{-i2\pi\nu_u x_r} \right) \\ &= \frac{1}{2} \left(e^{i2\pi(+\nu_u)x_r} + e^{i2\pi(-\nu_u)x_r} \right). \end{aligned}$$

We can interpret the complex representation of the real cosine as having one term with a positive frequency $+\nu_u$ and one term with a negative frequency $-\nu_u$. A similar equation

holds for the complex representation of $\sin(2\pi\nu_u x_r)$. The Fourier transform of a real function always contains both negative and positive frequencies, which arise from the complex exponentials in the definition of the transform.

Additional comments about negative frequencies will be made in Appendix C, where it will be seen that positive frequencies can be associated with waves propagating in the $+x$ direction, and negative frequencies correspond to waves propagating in the opposite, $-x$, direction.

A.4 Parseval's Relations

The physical and spectral variables of a continuous Fourier transform pair satisfy

$$\int_{-\infty}^{\infty} |f(x)|^2 dx = \int_{-\infty}^{\infty} |\hat{f}(\nu)|^2 d\nu, \quad (\text{A.16})$$

which is known as Parseval's relation. For complex amplitudes, $|f|^2 = f\hat{f}^*$. The corresponding equation for the discrete Fourier transform pair defined by Eqs (A.7) and (A.8) is

$$\sum_{r=0}^{N-1} |f(r)|^2 = N \sum_{u=0}^{N-1} |\hat{f}(u)|^2. \quad (\text{A.17})$$

The extension to the two-dimensional case is straightforward:

$$\sum_{r=0}^{N_x-1} \sum_{s=0}^{N_y-1} |f(r, s)|^2 = N_x N_y \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} |\hat{f}(u, v)|^2. \quad (\text{A.18})$$

The discrete forms of Parseval's relations provide important checks on numerical calculations. For example, it is easy to misplace factors of N , which appear in different places depending on the exact form used for the definition of the discrete transforms.

A.5 Discrete Convolution Theorems

The results of this section will be needed in Appendix G on image prediction. I include this derivation here two reasons: (1) it is exceptionally important, and (2) it is almost never discussed in texts. For example, Bracewell (1986, page 367) merely says "The theorem is" and writes the equivalent of Eq. (A.21) below, without any further discussion. He does not present the 2-D equivalent of Eq. (A.22), nor is there any discussion of the importance of this equation for applications such as image prediction. For simplicity, the derivation in this section is given for a function of one variable. The extension to two variables as needed for 2-D images is straightforward.

A.5.1 The 1-D Convolution Theorem

Let $f(x_r) = f(r)$ and $g(x_r) = g(r)$, $r = 0, 1, \dots, N-1$ be discrete functions of spatial location x_r with $x_0 = 0$ and $x_{N-1} = L$. The spacing of the discrete spatial points is $\Delta x = L/N$, so

$x_r = r\Delta x = rL/N$. Furthermore, let $H(\ell)$ be the *Heaviside step function* defined by

$$H(\ell) \equiv \begin{cases} 0 & \text{if } \ell \leq 0 \\ 1 & \text{if } \ell > 0 \end{cases}$$

Fourier transforms imply functions that are periodic in space. In one dimension, i.e. $f(r)$, $r = 0, 1, \dots, N-1$, this means that $f(N) = f(0)$, $f(-1) = f(N-1)$, $f(N+2) = f(2)$, and so on. That is, each function “wraps around” via a *cyclic shift* for values outside the defined index range of 0 to $N-1$. For 2-D functions such as images, the x and y directions wrap around independently. The Heaviside step function is used to wrap the indices of g whenever $r - r'$ is outside the range of 0 to $N-1$. For example, if $r = 2$ and $r' = 4$, then $r - r' = -2$, but $g(-2)$ is not defined. The value of $2 - 4 + NH(4 - 2) = -2 + N$, so the value of $g(N - 2)$, which is defined, will be used.

The Discrete Fourier Transform (DFT) of f , denoted by $\mathcal{D}\{f(x_r)\}$ or $\hat{f}(\nu_u)$ defined by Eq. (A.7) is

$$\hat{f}(u) = \mathcal{D}\{f(r)\} = \frac{1}{N} \sum_{r=0}^{N-1} f(r) e^{-i2\pi ur/N}. \quad (\text{A.19})$$

As discussed in Sections G.3.1 and G.3.2, for discrete variables of finite length, there are two kinds of convolution: linear or serial and circular or cyclic. We next develop the DFT of the cyclic convolution, which is the basis for fast computations of image degradation using the FFT. The 1-D form of the 2-D cyclic convolution of Eq. (G.6) is

$$f(r) \circledast g(r) \triangleq \sum_{r'=0}^{N-1} f(r') g(r - r' + N_x H(r' - r)), \quad (\text{A.20})$$

which will be used for the present derivation.

The DFT of a 1-D cyclic convolution is, by Eqs. (A.19) and (G.15),

$$\mathcal{D}\{f(r) \circledast g(r)\} = \frac{1}{N} \sum_{r=0}^{N-1} \left\{ \sum_{k=0}^{N-1} f(k) g(r - k) \right\} e^{-i2\pi ur/N},$$

where the cyclic indexing of $g(r - k)$ is understood. Now use

$$1 = e^{-i2\pi ku/N + i2\pi ku/N} = e^{-i2\pi ku/N} e^{i2\pi ku/N}$$

and interchange the summation order to write the last equation as

$$\mathcal{D}\{f(r) \circledast g(r)\} = \frac{1}{N} \sum_{k=0}^{N-1} f(k) \left\{ \sum_{r=0}^{N-1} g(r - k) e^{-i2\pi(r-k)u/N} \right\} e^{-i2\pi ku/N}.$$

Next let $\ell = r - k$ so that the term in braces becomes

$$\mathcal{D}\{f(r) \circledast g(r)\} = \frac{1}{N} \sum_{k=0}^{N-1} f(k) \left\{ \sum_{\ell=-k}^{N-1-k} g(\ell) e^{-i2\pi \ell u/N} \right\} e^{-i2\pi ku/N}.$$

Because $g(\ell)$ is periodic, $g(-k) = g(N - 1 - k)$ and so on, so that the dummy summation index ℓ includes exactly the same terms (just in a different order) as if the sum runs from $\ell = 0$ to $N - 1$. Thus the last equation can be rewritten as

$$\mathcal{D}\{f(r) \circledast g(r)\} = N \left\{ \frac{1}{N} \sum_{k=0}^{N-1} f(k) e^{-i2\pi k u / N} \right\} \left\{ \frac{1}{N} \sum_{\ell=0}^{N-1} g(\ell) e^{-i2\pi \ell u / N} \right\}.$$

The two terms in braces are the definitions of $\hat{f}(u)$ and $\hat{g}(u)$. Thus we have the *convolution theorem for cyclic convolution of discrete functions of length N* :

$$\mathcal{D}\{f(r) \circledast g(r)\} = N \hat{f}(u) \hat{g}(u). \quad (\text{A.21})$$

The factor of N in the last equation does not occur with continuous variables. That is, for a function $f(x)$ of a continuous variable x , the Fourier transform $\hat{f}(\nu)$ of $f(x)$ is defined as

$$\hat{f}(\nu) \triangleq \mathcal{F}\{f(x)\} \triangleq \int_{-\infty}^{\infty} f(x) e^{-i2\pi \nu x} dx.$$

For continuous functions $f(x)$ and $g(x)$, convolution is defined by

$$f(x) * g(x) = \int_{-\infty}^{\infty} f(x') g(x - x') dx'.$$

The convolution theorem for continuous functions is (see [Bracewell \(1986, page 110\)](#) for the derivation)

$$\mathcal{F}\{f(x) * g(x)\} = \hat{f}(\nu) \hat{g}(\nu).$$

Based on this result, you often see that statement that “the Fourier transform of the convolution of f and g is the product of the transforms of f and g .” That is true for continuous functions, but not for discrete, which has the extra array length factor N .

A.5.2 The 2-D Convolution Theorem

Let $f(x_r, y_s) = f(r, s)$ be a discrete function defined at a set of N_x evenly spaced points in the x direction, $x_r, r = 0, 1, \dots, N_x - 1$, and a set of N_y evenly spaced points in the y direction, $y_s, s = 0, 1, \dots, N_y - 1$. Then the 2-D DFT of $f(r, s)$ is defined by Eq. (A.10), and the 2-D cyclic convolution is given by Eq. (G.6) of Section G.3.2. These 2-D equations can be followed through in parallel to the 1-D equations above to derive the 2-D discrete convolution theorem. The result is

$$\mathcal{D}\{f(r, s) \circledast g(r, s)\} = N_x N_y \hat{f}(u, v) \odot \hat{g}(u, v). \quad (\text{A.22})$$

As for the 1-D case, the discrete convolution theorem has factors of $N_x N_y$, which do not occur in the continuous case. These factors are crucial when doing numerical image prediction calculations on a digital image of size $N_x \times N_y$.

The functions $\hat{f}(u, v)$ and $\hat{g}(u, v)$ are 2-D arrays of size $N_x \times N_y$, i.e. matrices. However, the product seen in Eq. (A.22) is an element-by-element multiplication, not matrix multiplication. This is denoted by the \odot symbol and is sometimes called the Hadamard product of the arrays. This equation is the key to the use of the fast Fourier transform for computation of image degradation. In that application $f(r, s)$ is the image, and $g(r, s)$ is obtained from the point spread function.

A.6 Orthogonality of Sines and Cosines

In general, two functions $f(x)$ and $g(x)$ are orthogonal over the interval $a \leq x \leq b$ if

$$\int_a^b f(x)g(x)dx = 0.$$

Although not obvious from the presentation above, the orthogonality of sines and cosines is fundamental to the theory of Fourier series and transforms. (That is, these relations are needed to prove the results that were stated above without proof.) For completeness, these orthogonality results are given here.

When $f(x)$ and $g(x)$ are continuous sines and cosines, the following set of orthogonality relations holds:

$$\begin{aligned} \int_c^{c+2\pi} \sin(mx) \cos(nx) dx &= 0 \quad \text{for all } m, n \\ \int_c^{c+2\pi} \sin(mx) \sin(nx) dx &= \begin{cases} 0 & \text{if } m \neq n \\ \pi & \text{if } m = n \end{cases} \\ \int_c^{c+2\pi} \cos(mx) \cos(nx) dx &= \begin{cases} 0 & \text{if } m \neq n \\ \pi & \text{if } m = n \end{cases} \end{aligned}$$

A similar set of relations can be developed for discrete functions. Let $\Delta\phi = 2\pi/N$, where N is an even number, and let $\phi_v = v\Delta\phi$, $v = 0, 1, \dots, N-1$ be a set of points on the interval $0 \leq \phi_v \leq 2\pi - \Delta\phi$. Furthermore, let δ_k be the Kronecker delta function defined by

$$\delta_k \triangleq \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{if } k \neq 0 \end{cases}$$

or equivalently

$$\delta_{k-\ell} \triangleq \begin{cases} 1 & \text{if } k = \ell \\ 0 & \text{if } k \neq \ell \end{cases}$$

where k and ℓ are integers. The discrete orthogonality relations for sines and cosines can then be written

$$\begin{aligned} \sum_{v=0}^{N-1} \cos(m\phi_v) \sin(n\phi_v) &= 0 \quad \text{for all } m \text{ and } n. \\ \sum_{v=0}^{N-1} \sin(m\phi_v) \sin(n\phi_v) &= \begin{cases} 0 & \text{if } m \neq n \\ N/2 & \text{if } m = n = 1, 2, \dots, N/2 - 1, N/2 + 1, \dots, N - 1, N + 1, \dots \\ 0 & \text{if } m = n = 0, N/2, N, 3N/2, 2N, \dots \end{cases} \\ &= \frac{N}{2}(\delta_{m-n} - \delta_{m+n} - \delta_{m+n-N}) \end{aligned}$$

$$\sum_{v=0}^{N-1} \cos(m\phi_v) \cos(n\phi_v) = \begin{cases} 0 & \text{if } m \neq n \\ N/2 & \text{if } m = n = 1, 2, \dots, N/2 - 1, N/2 + 1, \dots, N - 1, N + 1, \dots \\ N & \text{if } m = n = 0, N/2, N, 3N/2, 2N, \dots \end{cases}$$

$$= \frac{N}{2}(\delta_{m+n} + \delta_{m-n} + \delta_{m+n-N})$$

It is difficult to find the discrete forms in textbooks even though they are widely used in numerical calculations. For example, the discrete forms are used to decompose the azimuthal dependence of the radiative transfer equation deep inside HydroLight (Moblely, 1994, Sections 8.3-8.5).

Wave Variance Spectra

This Appendix discusses the fundamentals of wave variance spectra, which will be needed for sea surface generation. This is considered “well known” material—well known, that is, to those physical oceanographers who already know it. However, the entire business of wave spectra can be exceedingly confusing, and journal articles always assume that the reader already understands the underlying ideas and mathematics. It is therefore worthwhile to review the results that will be needed. A good reference for the development of wave spectra is [Holthuijsen \(2007\)](#), who is very careful in his notation and terminology. [Banner \(1990\)](#) and [Massel \(2011\)](#) provide review articles. The notation used below is chosen to conform to that used in [Elfouhaily et al. \(1997\)](#).

The discussion begins with an overview of the representation of waves as sums of sinusoids and the associated issues of sampling a wave at a discrete number of points. The theory of wave variance spectra is then introduced, followed by numerical examples using two of the most commonly used models for wave variance spectra. This material is the foundation of modeling sea surfaces as discussed in Appendix C.

B.1 Sinusoidal Wave Representations

This section first defines the sinusoidal functions that can be used to decompose waves on the sea surface into sums of sines and cosines. Sampling of sea surfaces at a finite number of points is then discussed.

The sea surface can be described using a Cartesian coordinate system. Let $\hat{\mathbf{x}}$ be a unit vector chosen to point in the downwind direction, and let $\hat{\mathbf{z}}$ be directed upward (from the water to the air) normal to the mean sea surface at $z = 0$. Then $\hat{\mathbf{y}} = \hat{\mathbf{z}} \times \hat{\mathbf{x}}$ is in the cross-wind direction. $z(x, y, t)$ is the surface elevation in meters at spatial location (x, y) and time t .

As shown in Appendix A, an arbitrary function of space or time can be written as a sum of sines and cosines of different amplitudes and wavelengths. The sea surface elevation $z(x, y, t)$ is usually such a function. An exception is a sea surface with waves curling over and breaking on a beach. There can then be multiple air-water surfaces for a given (x, y, t) : the point on the surface that sees the full sky, the point above a surfer’s head as the wave curls over her, and the point on the water that supports the surf board. However, for a

sea surface without breaking waves, there is only one air-water surface for a given location and time, which can be described as a sum of sinusoids.

The terms and notation needed to describe propagating sinusoidal waves are defined as follows:

- A is the *amplitude* (in units of meters) of the wave. This is one-half the distance from the trough to the crest of a sinusoidal wave. Oceanographers often talk about the *wave height*, H , which is the distance from a wave trough to a wave crest. Wave height is what most people visualize when they talk about how “big” a wave is. For a sinusoidal wave, $H = 2A$.
- T is the *temporal period* (seconds), usually called just the period. This is the time it takes for one wave crest to pass by a fixed point, i.e. for the argument of a sinusoid to go through 2π radians.
- $f = 1/T$ is the *temporal frequency* (1/seconds), usually called just the frequency. This is how many waves pass by a fixed point per second.
- $\omega = 2\pi/T = 2\pi f$ is the *angular temporal frequency* (radians/second), usually called the angular frequency. This is the number of waves passing a fixed point in 2π seconds.
- Λ is the *wavelength*, or spatial period, (meters). This is the distance from one wave crest to the next.
- $\nu = 1/\Lambda$ is the *spatial frequency* or *wavenumber* (1/meters). This is the number of wavelengths per meter.
- $k = 2\pi/\Lambda = 2\pi\nu$ is the *angular spatial frequency* or *angular wavenumber* (radians/meter). This is the number of wavelengths in 2π meters.
- ϕ is the phase (nondimensional). This fixes the location of a wave crest relative to the origin of a coordinate system.

As was discussed in Section 1.5.9, there is no uniformity of notation in the literature. People often refer to “the frequency” without specifying whether they mean f or ω , and “wavenumber” or “spatial frequency” may mean either ν or k . People use $\nu, \tilde{\nu}$, or σ for the wavenumber, some use σ for ω , and so on. You just have to figure it out on a case-by-case basis. For pedagogic purposes, we will start with wavenumber ν as the measure of spatial frequency, and then switch to the more common angular spatial frequency k . This presentation will use ω as the measure of temporal frequency.

Now consider a one-dimensional sea surface with elevation $z(x, t)$ over a spatial region of length L , e.g. $0 \leq x \leq L$ or $-L/2 \leq x \leq L/2$. According to Fourier, $z(x, t)$ can be written as a sum of sinusoids, the n^{th} one of which is

$$z_n(x, t) = A_n \cos\left(\frac{2\pi nx}{L} + \phi_n \pm \frac{2\pi nt}{T}\right), \quad (\text{B.1})$$

where $n = 0, 1, 2, \dots$ is simply an index for which sinusoid is being considered. (Note that we could just as well write a sine here, which corresponds to a different value for the phase.) For $n = 0$ this reduces to a constant offset $z_0(x, t) = A_0 \cos(\phi_0)$, which is usually taken to be mean sea level and is set to a reference value of 0 via setting $A_0 = 0$. For the moment, take $t = 0$, i.e., we have a snapshot of the ocean surface at time zero. This sinusoid has

a wavelength of $\Lambda_n = L/n$, i.e., the cosine returns to the same value after a distance of $x = L/n$. It is a pure cosine in the chosen coordinate system if the phase is 0 or an even integer multiple of $\pm\pi$. If the phase is an odd integer multiple of $\pm\pi/2$ it is a sine. If time is included, the wave is propagating in the $+x$ direction if the time term is $-2\pi nt/T$; the cosine returns to its initial value after a time of $t = T/n$. A wave propagating in the $-x$ direction is described by a $+2\pi nt/T$ term. The physical angular frequency $\omega_n \triangleq 2\pi n/T$ is always positive, but mathematically we can write just $+\omega_n t$ in the equation and then view a wave propagating in the $+x$ direction as having a negative temporal frequency.

With this interpretation of $\omega_n t$, Eq. B.1 is conveniently written as

$$z_n(x, t) = A_n \cos(k_n x + \phi_n + \omega_n t), \quad (\text{B.2})$$

where $k_n = 2\pi n/L = 2\pi/\Lambda_n$ is the angular wavenumber of the n^{th} wave of wavelength $\Lambda_n = L/n$. Likewise, $\omega_n = 2\pi n/T = 2\pi/T_n$ is the angular frequency of the n^{th} wave with period $T_n = T/n$. It may be intuitively easier to think of wavelengths per meter than of wavelengths per 2π meters, but the convenience of writing k rather than $2\pi/\Lambda$ or $2\pi\nu$ makes k rather than ν the spatial frequency variable used in most publications. A similar comment holds for ω versus $2\pi/T = 2\pi f$, so ω rather than f is the common choice for the temporal frequency variable.

If we take $t = 0$, or just combine the time term with the phase, expanding the cosine in Eq. (B.2) gives an equivalent form

$$z_n(x) = a_n \cos(k_n x) + b_n \sin(k_n x), \quad (\text{B.3})$$

where $a_n = A_n \cos \phi_n$ and $b_n = -A_n \sin \phi_n$, $n = 1, 2, \dots$, $a_0 = A_0$, and $b_0 = 0$. This equation can be written in terms of complex numbers:

$$z_n(x) = c_{+n} e^{ik_n x} + c_{-n} e^{-ik_n x}, \quad (\text{B.4})$$

where $i = \sqrt{-1}$ and

$$c_{+n} = (a_n - ib_n)/2, \quad c_{-n} = (a_n + ib_n)/2, \quad \text{and} \quad c_0 = a_0/2. \quad (\text{B.5})$$

Recalling that $e^{\pm i\theta} = \cos \theta \pm i \sin \theta$, we see that the imaginary parts of Eq. (B.4) add to zero, so that $z_n(x)$ is still a real variable even though it is written in complex form. Also note that $c_{+n}^* = c_{-n}$, where c^* denotes the complex conjugate. (Complex conjugation means replace i by $-i$ in all terms.) Although two complex numbers in general contain four independent real numbers, these c_{+n} and c_{-n} pairs contain only two independent numbers, a_n and b_n .

We thus have three ways to describe a sinusoidal wave: (1) The cosine of Eq. (B.2) defined by an amplitude and a phase; (2) the cosine and sine of Eq. (B.3) defined by two amplitudes, or (3) the complex exponentials of Eq. (B.4) defined by two complex numbers containing two independent real numbers. These equations all give the same $z_n(x)$, so we are free to choose whichever form is mathematically convenient for the problem at hand. For visualization of sea surfaces, the real forms (B.2) or (B.3) are convenient, but the calculations on the next sections are most conveniently carried out using complex exponentials.

Returning now to Fourier, linear wave theory says that we can write the sea surface elevation as a Fourier series

$$z(x) = \sum_{n=0}^{\infty} z_n(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(k_n x) + b_n \sin(k_n x)] , \quad (\text{B.6})$$

which in conjunction with Eq. (B.4) can be rewritten as

$$z(x) = \sum_{n=-\infty}^{\infty} c_n e^{ik_n x} . \quad (\text{B.7})$$

As previously noted, the a_0 or c_0 term is usually set to 0. This equation is the mathematical heart of our subsequent descriptions sea surfaces.

Although the physics that leads to a given sea surface in general involves non-linear interactions between waves of different wavelengths and periods, the shape of the resulting sea surface can be written as a linear sum of sinusoids of different frequencies. Owing to the orthogonality of sines and cosines (Section A.6) for different n values as defined above, these wave components $z_n(x)$ are independent of each other. Thus a description of a time-dependent surface based on an expansion like that of Eq. (B.6) (including the time-dependent terms as seen in Eq. (B.2)) cannot be used to predict the nonlinear evolution of a sea surface from a given initial condition. To do that, you must return to the world of hydrodynamics in all of its nonlinear glory. For our present purposes, we only want to model the shapes of sea surfaces, not predict their hydrodynamic development from an initial state, so linear wave representations are adequate.

B.1.1 Sampling

Next consider the implications of *sampling* the sea surface at a finite number of discrete locations or times. It is usually the case when we want to measure the sea surface elevation that we either make measurements at discrete locations $x_r, r = 0, 1, \dots, N - 1$ at a given time, or at discrete times $t_r, r = 0, 1, \dots, N - 1$ at a given location.

For a specific example, consider a region of sea surface $L = 10$ m long where we take $N = 8$ evenly spaced samples. This gives a measurement every $\Delta x = L/N = 1.25$ m at a given time. We are going to describe this surface as a sum of sinusoids. Figure B.1 illustrates the first few sinusoids shown as cosines. The cosines in the figure all have the same amplitude and each is offset vertically for ease of viewing. The dots show the sampled values for each cosine component.

The longest wave that can fit into a region of length L has a wavelength of $\Lambda = L$. This is called the fundamental wave or first harmonic. This wave is shown by the blue $n = 1$ curve in the figure. Note that the $n = 4$ cosine in the figure has a period of $2\Delta x$. The smallest wave that can be captured correctly in a sampling scheme has a period of twice the sampling interval. This called the two-point wave or two-point oscillation. Note that a sine term with period $2\Delta x$ would be sampled at arguments of $0, \pi, 2\pi, \dots$ where the sine is zero and is thus not detectable. The spatial frequency $1/(2\Delta x)$ (or temporal frequency $1/(2\Delta t)$ if sampling in time at a given location) is called the spatial (or temporal) *Nyquist frequency*. Waves with higher frequencies than the Nyquist frequency are still sampled, as illustrated by the dots on the wave with $n = 10$. Note, however, that the pattern of sampled points for the $n = 10$ wave is exactly the same as for the $n = 2$ wave. If all we have

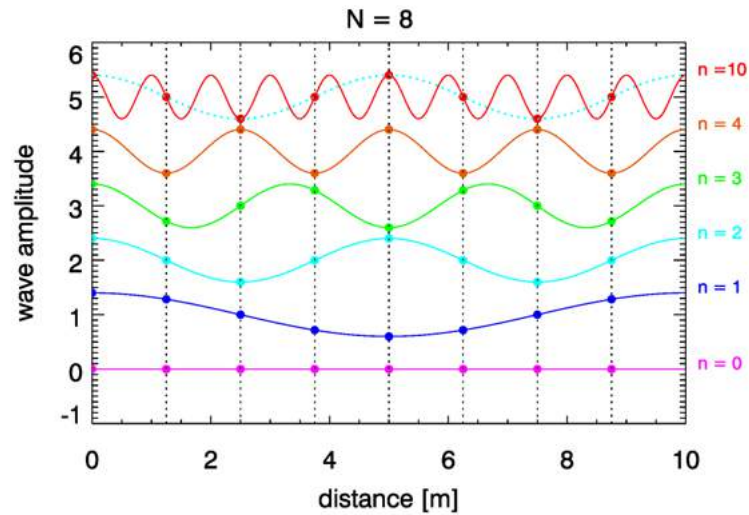


Figure B.1: Illustration of sampling the harmonics of a given wave and of aliasing. The vertical dotted lines show the sampling locations, and the dots show the sampled values for the various cosine components.

are the measured points, we cannot distinguish between the $n = 10$ and the $n = 2$ waves in this example. This illustrates the phenomenon of *aliasing*. In general, a wave with a frequency greater than the Nyquist frequency (i.e., a wavelength or period less than twice the sampling interval) is still sampled, but it appears as though the high frequency wave is a wave with a frequency less than the Nyquist frequency. The information about the high frequency (short wavelength) wave is added to or aliased into a lower frequency (longer wavelength) wave, thereby giving incorrect information about the lower frequency wave. Aliasing places a severe constraint on any sampling scheme. We can correctly sample only waves with wavelengths that lie between the size of the spatial region, L , and twice the size of the sampling interval. (In the temporal setting, the limits are the length of sampling time and twice the temporal sampling interval.)

The relation between a sampled frequency f_s greater than the Nyquist frequency, the sampling rate or Nyquist frequency f_{Ny} , and the frequency f_a receiving the aliased signal is

$$f_a = |mf_{Ny} - f_s|$$

where m is the closest integer multiple of the sampling rate to the signal being sampled. In the example of Fig. B.1, $f_s = f_{n=10} = 10$ waves per 10 meters = 1 m^{-1} , $f_{Ny} = 0.4$, and either $m = 2$ or $m = 3$ gives $f_a = |2 \times 0.4 - 1| = 0.2$ (or $f_a = |3 \times 0.4 - 1| = 0.2$). Thus the $n = 10$ wave is aliased into the $n = 2$ wave, which has a frequency of 2 waves per 10 meters, or 0.2 m^{-1} , just as seen in the figure. Note that many other frequencies are also aliased into this frequency. For example, a wave with $f_s = 1.4$, corresponding to $n = 14$, also gives $f_a = 0.2$ with $m = 4$, and so on.

B.2 Theory of Wave Variance Spectra

The figures of the following sections are illustrated using two commonly used wave variance spectra. The one-dimensional surfaces of Section D.3 use the Pierson-Moskowitz spectrum (Pierson and Moskowitz, 1964), and the two-dimensional surfaces Sections C.4 and C.7 use the spectrum of Elfouhaily et al. (1997). After the introductory overview on this section, the next section presents these two specific wave spectra for later reference.

B.2.1 Wave Energy

Consider a sinusoidal wave of amplitude A and wavenumber ν . It can be shown (e.g., Preisendorfer, 1976, Volume VI, page 72) that the total energy per unit horizontal area of sea surface of this wave, averaged over a wavelength, is

$$\frac{\text{energy}}{\text{area}} = \frac{1}{2}\rho g A^2 + \frac{1}{4}\tau \nu^2 A^2, \quad (\text{B.8})$$

where

$\rho \approx 10^3 \text{ kg m}^{-3}$ is the density of water,

$g \approx 9.8 \text{ m s}^{-2}$ is the acceleration of gravity,

$\tau \approx 0.072 \text{ N m}^{-1}$ (at 25 deg C) is the surface tension of water.

The $\frac{1}{2}\rho g A^2$ term is the sum of the kinetic and gravitational potential energy of the wave. (The potential energy is relative to the mean sea surface at $z = 0$.) The $\frac{1}{4}\tau \nu^2 A^2$ term is the energy required to stretch the level surface into a sinusoid, working against surface tension. It is easy to see that the units of these terms are J m^{-2} .

The variance of a sinusoidal surface is the average over one wavelength Λ of the surface elevation squared (assuming that the mean surface is at zero):

$$\text{var}\{z\} \triangleq \frac{1}{\Lambda} \int_0^\Lambda \left[A \sin\left(\frac{2\pi x}{\Lambda}\right) \right]^2 dx = \frac{1}{2} A^2. \quad (\text{B.9})$$

Thus the energy per unit area of a sinusoidal wave also can be written in terms of its variance:

$$\frac{\text{energy}}{\text{area}} = \left(\rho g + \frac{1}{2}\tau \nu^2 \right) \text{var}\{z\}. \quad (\text{B.10})$$

B.2.2 Wave Elevation Variance Spectra

The fundamental quantity of interest for the description of sea surfaces is a spectral density function that tells how much of the surface wave elevation variance is contained in particular frequencies. Other quantities of interest, in particular a spectral density function that tells how much sea surface slope variance is contained in different frequencies, can be derived from the elevation variance spectral density.

As just seen, the variance of the surface elevation is proportional to the amplitude squared. Equation (B.8) shows that the energy per unit horizontal area of a sinusoidal wave of a given amplitude and spatial frequency is also proportional to the amplitude squared, or to the variance of the surface elevation. For a sea surface containing waves of

many different amplitudes and spatial frequencies, the total variance of the sea surface is the sum of the variances of in the individual waves (the variance of a sum of independent random variables is is the sum of the variances of the individual variables). Likewise, the total energy of the waves is the sum of the energies of the individual waves.

We thus have the following line of reasoning:

- For a discrete sample $z(r)$ of a zero-mean sea surface, the variance is

$$\text{var}\{z\} = \frac{1}{N} \sum_{r=0}^{N-1} |z(r)|^2.$$

(Division by N rather than $N - 1$ in computing the sample variance is correct in this case because it is assumed that the mean surface elevation is known and equal to 0.)

- Parseval's identity (A.17),

$$\sum_{r=0}^{N-1} |z(r)|^2 = N \sum_{u=0}^{N-1} |\hat{z}(u)|^2,$$

gives

$$\text{var}\{z\} = \sum_{u=0}^{N-1} |\hat{z}(u)|^2.$$

- We therefore identify $\mathcal{S}(u) \triangleq |\hat{z}(u)|^2$ as the *discrete variance spectrum*, with units of m^2 .

Such reasoning led Arthur Schuster (Schuster, 1897, 1898) to the seminal observation that *Fourier transforms can be used to decompose the total variance contained in a signal into the variance contained in each frequency*. Schuster's original interest was in searching for what he call "hidden periodicities" in weather phenomena (as opposed to obvious periodicities such as daily or seasonal cycles). He soon applied the technique to looking for periodicities in earthquakes, sunspots, and other phenomena. Today, the computational speed of the FFT allows his method of analysis to be widely applied in science, engineering, economics and other fields where it is desired to know the energy or power of a signal as a function of spatial or temporal frequency. Regardless of what physical quantity is under consideration, the essence of a spectrum is that it gives the distribution of variance in that quantity as a function of frequency.

Because of the proportionality (B.10) of elevation variance to wave energy, $\mathcal{S}(u)$ is often loosely referred to as the discrete energy (or power) spectrum. To be precise, $\mathcal{S}(u)$ is not an energy variance spectrum unless it is multiplied by the factor $\rho g + \frac{1}{2}\tau\nu^2$ to convert elevation units to energy units. (In other applications, the variance of some quantity such as voltage or current in an electrical circuit can be multiplied by an appropriate conversion factor to get the spectrum of energy or power.)

Care is required to formulate wave spectra for continuous variables. As in Eqs. (B.2) and (B.6), we can write the continuous surface as a sum of sinusoids, e.g.

$$z(x) = \sum_{n=0}^{\infty} z_n(x) = \sum_{n=0}^{\infty} A_n \cos(k_n x + \phi_n). \quad (\text{B.11})$$

As we saw in the Wave Energy section above, the variance of the sinusoid with frequency k_n is $\frac{1}{2}A_n^2$. The waves of different frequencies are independent, so the total variance of the surface can be written as the sum of the variances of the individual waves:

$$\text{var}\{z\} = \sum_{n=0}^{\infty} \frac{1}{2}A_n^2. \quad (\text{B.12})$$

Now let Δk_n be a frequency interval centered on frequency k_n , whose sinusoid has amplitude A_n . We then define

$$\mathcal{S}(k = k_n) \triangleq \lim_{\Delta k_n \rightarrow 0} \frac{\frac{1}{2}A_n^2}{\Delta k_n}. \quad (\text{B.13})$$

In this definition, keep in mind that each A_n is associated with a particular frequency k_n , and that the limit operation holds for each value of n . We are thus defining a function of the spatial frequency, which becomes a continuous function of k as the bandwidth Δk_n goes to zero.

The continuous function $\mathcal{S}(k)$ is called the *omnidirectional elevation variance spectrum*. “Omnidirectional” means that there is no reference direction (e.g., a direction of wave propagation relative to the wind direction) included in the quantity. This is the case if a wave record is made at a single point as a function of time: the waves go past and their elevations are recorded, but no information is obtained about the direction the waves are traveling. $\mathcal{S}(k)$ is also called the omnidirectional elevation spectrum for obvious reasons. As is to be expected, there is no uniformity of notation for this spectrum, but \mathcal{S} seems to be the most common symbol—and what is used in both [Pierson and Moskowitz \(1964\)](#) and [Elfouhaily et al. \(1997\)](#), to be discussed in the next sections—so that is what is used here. (\mathcal{E} seems to be the second-most-common symbol and is used in [Holthuijsen \(2007\)](#).) Equation (B.13) shows that the units of $\mathcal{S}(k)$ are clearly $\text{m}^2/(\text{rad}/\text{m})$. Equations (B.12) and (B.13) show that integrating the omnidirectional variance spectrum over all frequencies gives the total elevation variance:

$$\text{var}\{z\} = \langle z^2 \rangle = \int_0^{\infty} \mathcal{S}(k) dk.$$

(The equations above are written in terms of spatial angular frequency k , as used for surface generation, but the reasoning and functional form of Eq. (B.13) are the same for any other spatial or temporal frequency variable.)

We can repeat this process for two dimensions, starting with

$$z(x, y) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} A_{n,m} \cos[k_n x + \ell_m y + \phi_{n,m}]. \quad (\text{B.14})$$

Here k is associated with spatial frequencies in the x direction, and ℓ corresponds to frequencies in the y direction. This leads to a function

$$\Psi(k = k_n, \ell = \ell_m) \triangleq \lim_{\Delta k_n \rightarrow 0} \lim_{\Delta \ell_m \rightarrow 0} \frac{\frac{1}{2}A_{n,m}^2}{\Delta k_n \Delta \ell_m}. \quad (\text{B.15})$$

The frequency intervals in the x and y directions, Δk_n and $\Delta \ell_m$, do not depend on the respective frequency indices n and m . That is, the frequency intervals are equally spaced, but they need not be the same in the x and y directions.

The notation in the last two equations is non-standard. The convention is to use k_x for frequencies in the x direction (k above), and k_y for frequencies in the y direction (ℓ above). The $\Psi(k, \ell)$ of Eq. (B.15) is then written as $\Psi(k_x, k_y)$. This leads to confusion in the subscripts, which can denote either frequency variables k_x, k_y , or specific discrete values k_n, k_m . However, the $\Psi(k_x, k_y)$ notation is standard in the literature, so that is what is used below.

$\Psi(k_x, k_y)$ (i.e., $\Psi(k, \ell)$) is the *directional variance spectrum in Cartesian coordinates*. Its units are clearly $\text{m}^2/(\text{rad/m})^2$. This spectrum is often called the “two-dimensional wavenumber spectrum,” and its arguments are often written in vector form, $\Psi(\mathbf{k})$, where $\mathbf{k} = (k_x, k_y)$ denotes the location of the frequency point in the 2-D frequency plane.

Equation (B.15) is the conceptual definition of $\Psi(\mathbf{k})$. In practice, if we have discrete measurements of the two-dimensional sea surface elevation at a given time, $z(x_r, y_s) = z(r, s) = z(\mathbf{x}_{rs})$, then the two-dimensional discrete Fourier transform of $z(r, s)$ gives the two-dimensional amplitudes

$$\hat{z}(\mathbf{k}_{uv}) = \hat{z}[k_x(u), k_y(v)] = \hat{z}(u, v) = \mathcal{D}\{z(r, s)\}.$$

Dividing by the discrete frequency bandwidths gives an *estimate* (called a 2-D periodogram) of the two-dimensional elevation variance spectral density

$$\mathcal{P}(\mathbf{k}_{uv}) \triangleq \frac{|\hat{z}(\mathbf{k}_{uv})|^2}{\Delta k_u \Delta k_v} \triangleq \frac{\Psi(\mathbf{k}_{uv})}{\Delta k_u \Delta k_v}. \tag{B.16}$$

The arguments of $z(\mathbf{x}_{rs})$, $\hat{z}(\mathbf{k}_{uv})$, and $\Psi(\mathbf{k}_{uv})$ show that these are discrete functions, whereas $\Psi(\mathbf{k})$ denotes a spectral density function of the continuous variable \mathbf{k} . The 2-D periodogram is an approximation of the 2-D variance spectral density, $\mathcal{P}(\mathbf{k}_{uv}) \doteq \Psi(\mathbf{k})$, where the symbol “ \doteq ” is used to denote “is an estimate of.” As Eq. (B.16) shows, the discrete function $\Psi(\mathbf{k}_{uv})$ has units of m^2 , whereas $\mathcal{P}(\mathbf{k}_{uv})$ and $\Psi(\mathbf{k})$ have units of $\text{m}^2/(\text{rad/m})^2$. A single periodogram contains statistical noise because it is computed from a single realization of a random sea surface. However, if many sets of observations are made, and the respective periodograms are averaged, then the noise tends to average out, and the average of the periodograms approaches the conceptual limit of the spectrum $\Psi(\mathbf{k})$. Questions such as how many periodograms must be averaged to obtain a spectrum with a given level of uncertainty lie in the domain of *spectrum estimation*, which need not concern us here.

A 2-D spectrum depends on direction, i.e., on the direction of the (k_x, k_y) point in a two-dimensional frequency plane. Usually, the $+x$ direction is chosen to be pointing downwind and, correspondingly, $+k_x$ represents the spatial frequencies of the waves propagating downwind. In this case, the angle $\varphi = \tan^{-1}(k_y/k_x)$ gives the direction relative to the wind direction. As Eq. (B.15) shows, the integral of $\Psi(k_x, k_y)$ over all frequencies gives the variance of the two-dimensional surface:

$$\text{var}\{z\} = \langle z^2 \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi(k_x, k_y) dk_x dk_y.$$

It is also common to define a directional spectrum in terms of polar coordinates given by the magnitude k and direction φ of the vector \mathbf{k} . These are related to k_x, k_y by

$$k = \sqrt{k_x^2 + k_y^2}$$

$$\varphi = \tan^{-1} \left(\frac{k_y}{k_x} \right)$$

and inversely by

$$k_x = k \cos \varphi$$

$$k_y = k \sin \varphi .$$

In this case we define

$$\tilde{\Psi}(k, \varphi) \triangleq \lim_{\Delta k \rightarrow 0} \lim_{\Delta \varphi \rightarrow 0} \frac{1}{2} \frac{A_{n,m}^2}{\Delta k \Delta \varphi} . \quad (\text{B.17})$$

This spectrum has units of $\text{m}^2 / [(\text{rad}/\text{m}) \text{rad}]$. (The tilde notation is used here to distinguish this spectrum from the $\Psi(k, \varphi)$ spectrum of Elfouhaily, et al. which is defined in the next section. (Some authors reserve the name “directional spectrum” for $\tilde{\Psi}(k, \varphi)$ and refer to $\Psi(k_x, k_y)$ as the wavenumber spectrum.) As before, definition (B.17) shows that integrating $\tilde{\Psi}(k, \varphi)$ over k and φ gives the variance:

$$\text{var}\{z\} = \int_0^\infty \int_0^{2\pi} \tilde{\Psi}(k, \varphi) dk d\varphi .$$

The ECKV directional spectrum given in the next section is specified in terms of polar coordinates k, φ . However, in the next sections we will need a spectrum in terms of Cartesian coordinates k_x, k_y for use in a rectangular FFT grid. The change of variables from polar to Cartesian coordinates is effected by the Jacobian

$$\begin{aligned} \Psi(k_x, k_y) &= \tilde{\Psi}(k, \varphi) \left| \frac{\partial(k, \varphi)}{\partial(k_x, k_y)} \right| \\ &= \tilde{\Psi}(k, \varphi) \begin{vmatrix} \frac{\partial k}{\partial k_x} & \frac{\partial k}{\partial k_y} \\ \frac{\partial \varphi}{\partial k_x} & \frac{\partial \varphi}{\partial k_y} \end{vmatrix} \\ &= \tilde{\Psi}(k, \varphi) \frac{1}{k} . \end{aligned} \quad (\text{B.18})$$

Note that the $1/k$ factor converts the units of $\tilde{\Psi}(k, \varphi)$ into the units of $\Psi(k_x, k_y)$.

In Eq. (B.27) below, this last equation is partitioned as

$$\Psi(k_x, k_y) = \frac{1}{k} \mathcal{S}(k) \Phi(k, \varphi) \triangleq \Psi(k, \varphi) , \quad [\text{ECKV 45}] \quad (\text{B.19})$$

where $\mathcal{S}(k)$ is an omnidirectional spectrum and $\Phi(k, \varphi)$ is a nondimensional *spreading function*, which shows how waves of different frequencies propagate (or “spread out”) relative to the downwind direction at $\varphi = 0$. Labels such as [ECKV 45] refer to the corresponding equations in [Elfouhaily et al. \(1997\)](#). The spreading function by definition satisfies

$$\int_0^{2\pi} \Phi(k, \varphi) d\varphi = 1 \quad (\text{B.20})$$

for all k .

Equation (B.19) shows that to obtain the ECKV variance spectrum in Cartesian coordinates we need only evaluate the ECKV $\Psi(k, \varphi)$ spectrum for the corresponding values of k and φ , i.e.

$$\Psi(k_x, k_y) = \Psi\left(k = \sqrt{k_x^2 + k_y^2}, \varphi = \tan^{-1}(k_y/k_x)\right). \quad (\text{B.21})$$

Note in particular that there is no “extra” k factor involved in the conversion of $\Psi(k, \varphi)$ to $\Psi(k_x, k_y)$; both quantities have the same units. (The k factor seen in the differentials $dk_x dk_y = k dk d\varphi$ is canceled by the $1/k$ coming from the Jacobian as seen in Eq. B.18.)

Integration of Eq. (B.19) over the respective (k_x, k_y) and (k, φ) frequency planes gives the variance as

$$\begin{aligned} \langle z^2 \rangle &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi(k_x, k_y) dk_x dk_y \\ &= \int_0^{\infty} \int_0^{2\pi} \frac{1}{k} \mathcal{S}(k) \Phi(k, \varphi) k dk d\varphi \\ &= \int_0^{\infty} \mathcal{S}(k) dk, \quad [\text{ECKV A2}] \end{aligned} \quad (\text{B.22})$$

after noting the normalization of Eq. (B.20). Thus the variance of the sea surface is still contained in the omnidirectional spectrum, even in the two-dimensional case. The omnidirectional spectrum $\mathcal{S}(k)$ is obtained from $\Psi(k, \varphi)$ via

$$\mathcal{S}(k) = \int_{-\pi}^{\pi} \Psi(k, \varphi) k d\varphi. \quad [\text{ECKV A3}]$$

Unfortunately, making measurements of 2-D sea surfaces is extremely difficult. There are very few such data sets—obtained, for example, by laser reflectance measurements (Huang et al., 2000)—and these do not cover the full range of spatial scales. Given the paucity of empirical 2-D wave data from which to develop 2-D variance spectra, the common procedure is to start with a 1-D or omnidirectional spectrum $\mathcal{S}(k)$ and add an angular spreading function $\Phi(k, \varphi)$ to distribute the wave energy over different directions relative to the downwind direction. In nature, most waves travel more or less downwind, a small amount of energy (i.e., variance) is contained in waves propagating in nearly cross-wind directions, and almost no energy is contained in waves that by some chance (such as the breaking of a larger wave generating smaller waves in all directions) might be propagating in upwind directions. The spreading function must capture this behavior. Although omnidirectional wave spectra are well grounded in observations, the choice of a spreading function is still something of a black art.

B.2.3 Wave Slope Variance Spectra

Now return to Eq. (B.11) and take the derivative to get the slope of the sea surface for the n^{th} wave:

$$\frac{dz_n(x)}{dx} = -A_n k_n \sin(k_n x + \phi_n).$$

As in Eq. (B.9), the variance of this slope is

$$\text{var} \left\{ \frac{dz_n}{dx} \right\} \triangleq \frac{1}{\Lambda_n} \int_0^{\Lambda_n} [A_n k_n \sin(k_n x + \phi_n)]^2 dx = \frac{1}{2} A_n^2 k_n^2.$$

A limit operation corresponding to Eq. (B.13) gives

$$\lim_{\Delta k_n \rightarrow 0} \frac{\frac{1}{2} A_n^2 k_n^2}{\Delta k_n} = k^2 \mathcal{S}(k). \quad (\text{B.23})$$

The quantity $k^2 \mathcal{S}(k)$ is the *omnidirectional slope variance spectrum*, usually called just the slope spectrum. The units of $k^2 \mathcal{S}(k)$ are m rad. Integrating the slope spectrum over all frequencies gives the total variance σ^2 of the sea surface slope:

$$\sigma^2 \triangleq \text{var} \left\{ \frac{dz}{dx} \right\} = \left\langle \left(\frac{dz}{dx} \right)^2 \right\rangle = \int_0^\infty k^2 \mathcal{S}(k) dk.$$

This variance is usually called the *mean square slope* or mss. The units of the mss are rad². Radians are nondimensional numbers, but the label of rad² reminds us that we can think of the slope as an angle from the horizontal measured in radians.

Comment: There is a subtle inconsistency in the units of mean square slopes as seen in the literature. As obtained above from the slope variance spectrum, the mss has units of radians squared. However, as defined using a finite difference of a sea surface elevation sample $z(x_r)$, the slope of the surface between two sample points x_r and x_{r+1} is

$$\text{slope} = \frac{\Delta z}{\Delta x} = \frac{z(x_{r+1}) - z(x_r)}{x_{r+1} - x_r}.$$

This finite difference is a non-dimensional slope as defined in analytic geometry, and the corresponding mss is obtained by averaging the squares of the finite differences over all of the sample points. However, a mss computed from finite differences $(\Delta z / \Delta x)^2$ is not the same as a mss with units of rad² as computed from a slope spectrum. For example, a slope of $\Delta z / \Delta x = 0.1$ corresponds to a slope angle of $\tan^{-1}(0.1) = 0.09966865$ rad. This is a negligible numerical difference for this slope, which is typical of actual sea surfaces, but there is a philosophical difference in a nondimensional slope as defined in analytic geometry and a slope defined as an angle with units of radians. The difference would not be negligible for large slopes: e.g., a slope of $\Delta z / \Delta x = 1$ corresponds to an angle of 0.78 rad (45 deg), not 1 rad. I have never seen any reference to this distinction in the literature, which seems to apply “mean square slope” to both forms of the surface slope. (Perhaps a reader of this section can inform me of how this issue is resolved in the wave spectrum community when comparing theoretical mean square slopes with measured ones.)

There is another way to view slope spectra. As we know from Eq. A.4 of Appendix A on Fourier transforms, the 1-D surface elevation $z(x)$ is related to the Fourier amplitude $\hat{z}(k)$ by

$$z(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{z}(k) e^{+ikx} dk.$$

Differentiating this equation with respect to x gives the 1-D *slope* of the sea surface as

$$\sigma(x) \triangleq \frac{dz(x)}{dx} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{z}(k) i k e^{+ikx} dk.$$

This leads us to identify $ik\hat{z}(k)$ as the Fourier amplitude corresponding to the sea surface slope. This gives us two ways to study the slope statistics of random sea surfaces, given

the Fourier amplitude $\hat{z}(k)$ (which we will learn to create from wave variance spectra in the following sections). The first way is to take the inverse Fourier transform of $\hat{z}(k)$ to obtain $z(x)$, and then to differentiate $z(x)$ to get the slope. The second way is to take the inverse transform of $ik\hat{z}(k)$ to get the slope directly, without ever creating the surface $z(x)$ itself. These two processes will in general give different realizations of the surface slopes, but the slope *statistics*, e.g. the mean square slopes σ^2 , will be the same.

The corresponding relations for two dimensions are derived in the same fashion and lead to similar results. Assuming that the wind is blowing in the $+x$ direction, the mean-square slope in the along-wind direction is given by either of

$$\begin{aligned} \left\langle \left(\frac{\partial z(x, y)}{\partial x} \right)^2 \right\rangle &\triangleq \sigma_x^2 \triangleq mss_x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k_x^2 \Psi(k_x, k_y) dk_x dk_y \\ &= \int_{-\infty}^{\infty} \int_{-\pi}^{\pi} k^2 \cos^2 \varphi \Psi(k, \varphi) k dk d\varphi. \end{aligned} \quad [\text{ECKV A4}]$$

The corresponding equation for the cross-wind direction is

$$\begin{aligned} \left\langle \left(\frac{\partial z(x, y)}{\partial y} \right)^2 \right\rangle &\triangleq \sigma_y^2 \triangleq mss_y = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} k_y^2 \Psi(k_x, k_y) dk_x dk_y \\ &= \int_{-\infty}^{\infty} \int_{-\pi}^{\pi} k^2 \sin^2 \varphi \Psi(k, \varphi) k dk d\varphi. \end{aligned} \quad [\text{ECKV A5}]$$

Recalling that variances of random variables add to get the total variance due to all variables gives the total mean square slope

$$\begin{aligned} mss &= mss_x + mss_y = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (k_x^2 + k_y^2) \Psi(k_x, k_y) dk_x dk_y \\ &= \int_{-\infty}^{\infty} \int_{-\pi}^{\pi} k^2 \Psi(k, \varphi) k dk d\varphi \\ &= \int_{-\infty}^{\infty} k^2 \mathcal{S}(k) dk. \end{aligned} \quad [\text{ECKV A6}]$$

Thus, even in the 2-D case, the total slope variance can be obtained from the omnidirectional slope spectrum.

Table B.1 summarizes the spectral quantities used on the following sections.

B.3 Examples of Wave Variance Spectra

This section gives two examples of wave elevation variance spectra. The Pierson-Moskowitz omnidirectional spectrum (Pierson and Moskowitz, 1964) describes gravity waves in a “fully developed” sea. A fully developed sea is an idealization of the statistically steady-state wave field resulting from a steady wind blowing for an infinitely long time over an infinite fetch. In practice, the duration and fetch required to achieve something close to a fully developed sea depend on the wind speed. A steady wind of 5 m s^{-1} blowing for 10 hours over a fetch of 60 km might be adequate; for hurricane winds of 35 m s^{-1} , a fetch of a few thousand kilometers and a duration of several days would be required. Thus it is much easier to approach a fully developed sea at low wind speeds than at high.

Spectrum Name	Symbols	Units
1-D or omnidirectional		
variance or elevation	$\mathcal{S}(k)$	$\text{m}^2/(\text{rad}/\text{m})$
slope	$k^2\mathcal{S}(k)$	m rad
2-D or directional		
variance or elevation	$\Psi(k_x, k_y), \Psi(k, \varphi)$	$\text{m}^2/(\text{rad}/\text{m})^2$
alongwind slope	$k_x^2 \Psi(k_x, k_y), k^2 \cos^2 \varphi \Psi(k, \varphi)$	rad^2
crosswind slope	$k_y^2 \Psi(k_x, k_y), k^2 \sin^2 \varphi \Psi(k, \varphi)$	rad^2
total slope	$(k_x^2 + k_y^2) \Psi(k_x, k_y), k^2 \Psi(k, \varphi)$	rad^2

Table B.1: Summary of wave variance spectral quantities.

The directional spectrum of [Elfouhaily et al. \(1997\)](#) includes both gravity and capillary wave scales. Moreover, it has a parameter that describes the wave age, so that any sea state from young (the wind has just started blowing) to fully developed can be simulated. The Elfouhaily et al. spectrum will be used to generate the two-dimensional sea surface examples in Section [C.4](#).

B.3.1 The Pierson-Moskowitz Omnidirectional Gravity Wave Spectrum

The omnidirectional [Pierson and Moskowitz \(1964\)](#) spectrum, formulated in terms of angular spatial frequency k , is

$$\mathcal{S}_{\text{PM}}(k) = \frac{\alpha}{2k^3} \exp \left[-\beta \left(\frac{g}{k} \right)^2 \frac{1}{U_{19}^4} \right] \quad [\text{m}^2/(\text{rad}/\text{m})], \quad (\text{B.24})$$

where

$$\alpha = 0.0081,$$

$$\beta = 0.74,$$

$g = 9.82 \text{ m s}^{-2}$ is the acceleration of gravity,

U_{19} is the wind speed in m s^{-1} at 19.5 m above the sea surface, and

k is the angular spatial frequency in rad m^{-1} .

The wind speed at 19.5 m can be obtained from the more commonly used wind at 10 m above the sea surface by the approximate formula

$$U_{19} \approx 1.026U_{10}.$$

As has already been noted, it is often of interest to express a variance spectrum in terms of other variables, such as the wavenumber ν or the temporal angular frequency ω . To change variables in a spectral density function, the key is to recall that a variance density function by definition expresses the variance per unit frequency interval. The

variance contained in some interval dk of the spatial angular frequency equals the variance contained in the corresponding interval $d\nu$ of the wavenumber or the interval $d\omega$ of the temporal frequency. Thus we have

$$\mathcal{S}_{\text{PM}}(k)dk = \mathcal{S}_{\text{PM}}(\nu)d\nu = \mathcal{S}_{\text{PM}}(\omega)d\omega.$$

To change the variable from $k = 2\pi\nu$ to ν , the previous equation gives

$$\mathcal{S}_{\text{PM}}(\nu) = \mathcal{S}_{\text{PM}}(k)\frac{dk}{d\nu} = \mathcal{S}_{\text{PM}}(k = 2\pi\nu)2\pi,$$

which results in

$$\mathcal{S}_{\text{PM}}(\nu) = \frac{\alpha}{8\pi^2\nu^3} \exp\left[-\beta\left(\frac{g}{2\pi\nu}\right)^2 \frac{1}{U_{19}^4}\right] \quad [\text{m}^2/(1/\text{m})]. \quad (\text{B.25})$$

To change variables from spatial angular frequency k to temporal angular frequency ω , we use the dispersion relation for gravity waves in deep water,

$$\omega^2 = gk,$$

to evaluate

$$\frac{dk}{d\omega} = \frac{2\omega}{g},$$

which leads to

$$\mathcal{S}_{\text{PM}}(\omega) = \frac{\alpha g^2}{\omega^5} \exp\left[-\beta\left(\frac{g}{\omega U_{19}}\right)^4\right] \quad [\text{m}^2/(\text{rad/s})]. \quad (\text{B.26})$$

All of these formulas have units of meters squared over the appropriate frequency. (The quantities $dk/d\nu$ and $dk/d\omega$ seen in the conversions are the Jacobians for the one-dimensional changes of variables.) Figure B.2 shows the Pierson-Moskowitz spectrum as functions of k and ω for wind speeds of $U_{10} = 5, 10$, and 15 m s^{-1} .

This spectrum has withstood the test of time fairly well (Alves and Banner, 2003) as a description of gravity waves in fully developed seas. However, it should not be used for high-frequency (short-wavelength) gravity waves, and certainly not for capillary waves. Likewise, it does not describe young seas, which have not had the time or fetch needed to approach the state of a well developed sea.

Figure B.3 shows the Pierson-Moskowitz slope spectra for three wind speeds. Note that the slope spectrum falls off much more slowly for high frequencies than does the elevation spectrum. That means that the higher frequencies contribute much more to the total slope variance than they do to the total elevation variance.

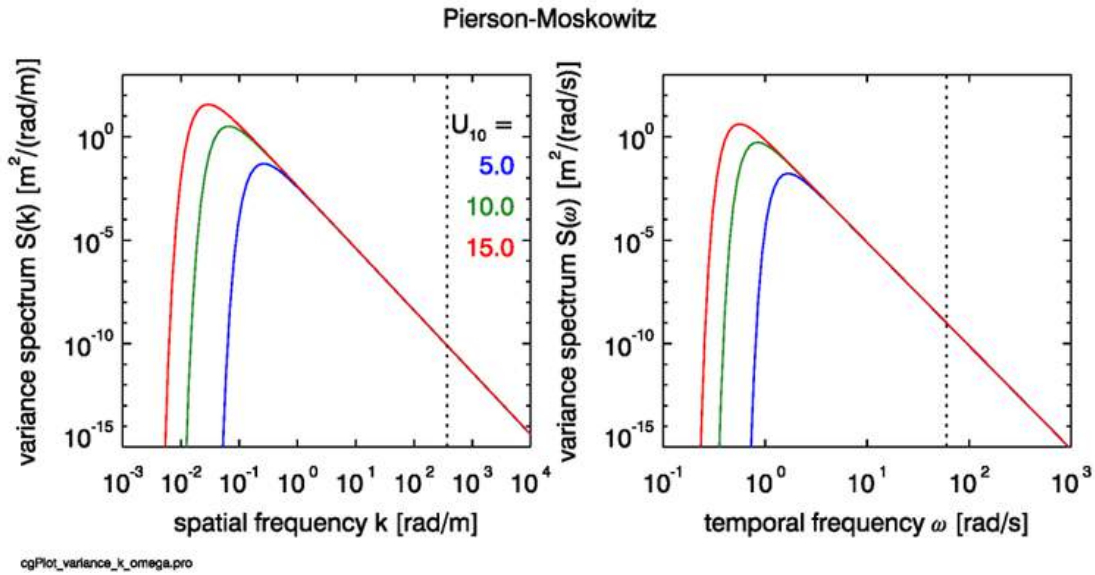


Figure B.2: The Pierson-Moskowitz variance spectrum as functions of k and ω for wind speeds of $U_{10} = 5, 10,$ and 15 m s^{-1} . The vertical dotted lines at $k = 370 \text{ rad/m}$ and $\omega = 60.3 \text{ rad/s}$ show the boundary between gravity and capillary waves.

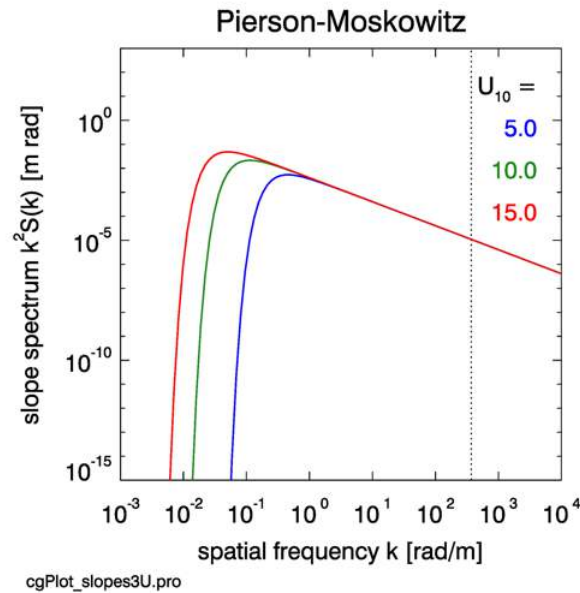


Figure B.3: The Pierson-Moskowitz slope spectrum for wind speeds of $U_{10} = 5, 10, 15 \text{ m s}^{-1}$. The plot uses the same ordinate scale as used for the elevation spectrum in the left panel of Fig. B.2 in order to highlight the slow falloff of the slope spectrum compared to the elevation spectrum. The vertical dotted line is the boundary between gravity and capillary waves.

B.3.2 The ECKV Directional Gravity-Capillary Wave Spectrum

In order to model two-dimensional sea surfaces $z(x, y)$, we need a 2-D wave elevation variance spectrum that describes the distribution of wave variance for waves propagating in difference directions. The numerical examples of 2-D sea surfaces to be seen in Appendix C use the 2-D elevation variance spectrum of [Elfouhaily et al. \(1997\)](#), which is described here for later reference. For brevity, this paper and their model are denoted here by “ECKV,” taken from the initials of the authors’ last names.

The ECKV spectrum has an omnidirectional variance spectrum that explicitly includes both the gravity and capillary wave scales. The boundary between gravity and capillary waves is taken to be $k = \sqrt{\rho g / \tau} = 370$ rad/m, the angular spatial frequency at which the restoring forces (which tend to bring a wave surface back to an unperturbed level surface) of gravity and surface tension are equal. The corresponding wavelength is $\Lambda = 2\pi/370 = 0.017$ m. ECKV then combine their omnidirectional spectrum with a spreading function to obtain their one-sided, directional variance spectrum. Using their notation, the 2-D ECKV spectrum has the form

$$\Psi(k, \varphi) = \frac{1}{k} \mathcal{S}(k) \Phi(k, \varphi) \quad [\text{ECKV 45}]. \quad (\text{B.27})$$

Here $\mathcal{S}(k)$ is the 1-D omnidirectional spectrum with units of $\text{m}^2/(\text{rad}/\text{m})$, and $\Phi(k, \varphi)$ is a non-dimensional spreading function. $\Psi(k, \varphi)$ thus has units of $\text{m}^2/(\text{rad}/\text{m})^2$. Equation labels such as [ECKV 45] give for reference the corresponding equation in the ECKV paper.

The ECKV omnidirectional spectrum is

$$\mathcal{S}(k) = \frac{B_l + B_h}{k^3} \quad [\text{ECKV 30}], \quad (\text{B.28})$$

where B_l is the low-frequency (long gravity wave) contribution to the variance, and B_h is the high-frequency (short gravity wave to capillary wave) contribution. (The quantity $k^3 \mathcal{S}(k)$ is called the *curvature* or *saturation* spectrum and is of interest in physical oceanography because it is related to the rate of variance dissipation of the waves. Thus ECKV refer to B_l and B_h as the low and high frequency curvature spectra.) The components of the omnidirectional spectrum are given by

$$L_{PM} = \exp[-1.25(k_p/k)^2] \quad [\text{ECKV 2}]$$

$$\Gamma = \exp\left\{-\frac{1}{2\sigma^2}[(\sqrt{k/k_p} - 1)^2]\right\} \quad [\text{below ECKV 3}]$$

$$J_p = \gamma^\Gamma \quad [\text{ECKV 3}]$$

$$F_p = L_{PM} J_p \exp\{-0.3162 \Omega_c (\sqrt{k/k_p} - 1)\} \quad [\text{ECKV 32}]$$

$$F_m = L_{PM} J_p \exp[-0.25(k/k_m - 1)^2] \quad [\text{ECKV 41}]$$

(Note : A typo in ECKV Eq. 41 omitted the $L_{PM} J_p$ factor in F_m)

$$B_l = 0.5 \alpha_p (c_p/c) F_p \quad [\text{ECKV 31}]$$

$$B_h = 0.5 \alpha_m (c_m/c) F_m \quad [\text{ECKV 40}]$$

where

$$\alpha = 0.0081,$$

$$\beta = 1.25,$$

$g = 9.82 \text{ m s}^{-2}$ is the acceleration of gravity,

U_{10} is the wind speed in m s^{-1} at 10 m above the sea surface

k is the angular spatial frequency in rad m^{-1}

Ω_c is defines the age of the waves for the given wind speed:

= 0.84 for a fully developed sea (corresponds to Pierson-Moskowitz)

= 1 for a “mature” sea [used in ECKV Fig 8a]

= 2 to 5 for a “young” sea; the maximum allowed value is 5

$Cd_{10N} = 0.00144$ is a drag coefficient [value deduced from ECKV Fig 11]

$u^* = \sqrt{Cd_{10N}U_{10}}$ is the friction velocity [using ECKV 61]

$$a_o = 0.1733 \text{ (ECKV 59)}$$

$$a_p = 4.0$$

$$k_m = 370.0 \text{ rad/m}$$

$c_m = 0.23 \text{ m/s}$ is the phase speed of the wave with spatial frequency k_m

$$a_m = 0.13u^*/c_m \text{ [ECKV 59]}$$

$$\gamma = 1.7 \text{ if } \Omega_c \leq 1 \text{ else } \gamma = 1.7 + 6 \log_{10}(\Omega_c)$$

$$\sigma = 0.08(1 + 4\Omega_c^{-3})$$

$$\alpha_p = 0.006\Omega_c^{0.55} \text{ [ECKV Eq. 34]}$$

$$\alpha_m = 0.01[1 + \ln(u^*/c_m)] \text{ if } u^* \leq c_m \text{ else } \alpha_m = 0.01[1 + 3 \ln(u^*/c_m)] \text{ [ECKV 44]}$$

$$k_o = g/U_{10}^2$$

$k_p = k_o\Omega_c^2$ is the spatial frequency of the maximum of the spectrum

$c_p = \sqrt{g/k_p}$ is the phase speed of the wave with spatial frequency k_p

$c = \sqrt{(g/k)(1 + (k/k_m)^2)}$ is the phase speed of the wave

At the lower frequencies, the ECKV spectrum is essentially the Pierson-Moskowitz spectrum (the L_{PM} term above) with an enhancement (the J_p term) that adds more energy to the lower frequencies. The highest frequencies have a cutoff due to viscous damping of the smallest capillary waves. The ECKV omnidirectional elevation and slope spectra are illustrated in Fig. B.4 for the case of a fully developed sea and three wind speeds. Figure B.5 shows the spectra as a function of wave age for a wind speed of $U_{10} = 10 \text{ m s}^{-1}$.

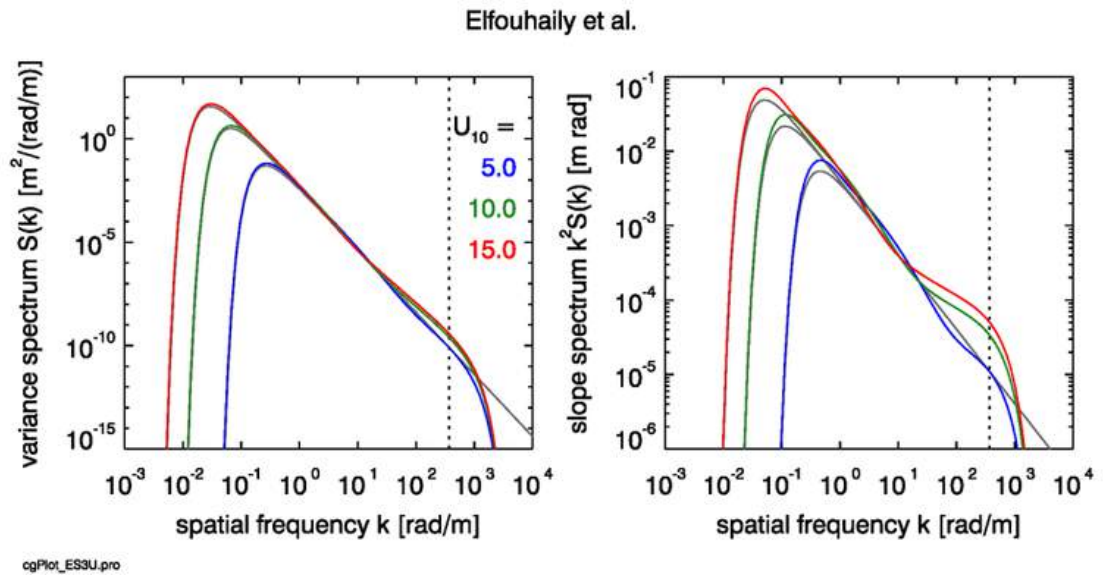


Figure B.4: The omnidirectional part \mathcal{S} of the ECKV elevation variance spectrum (left panel) and slope spectrum $k^2\mathcal{S}$ (right panel) for fully developed seas and wind speeds of $U_{10} = 5, 10, 15 \text{ m s}^{-1}$. The gray lines show the corresponding Pierson-Moskowitz spectra from Fig. B.3.

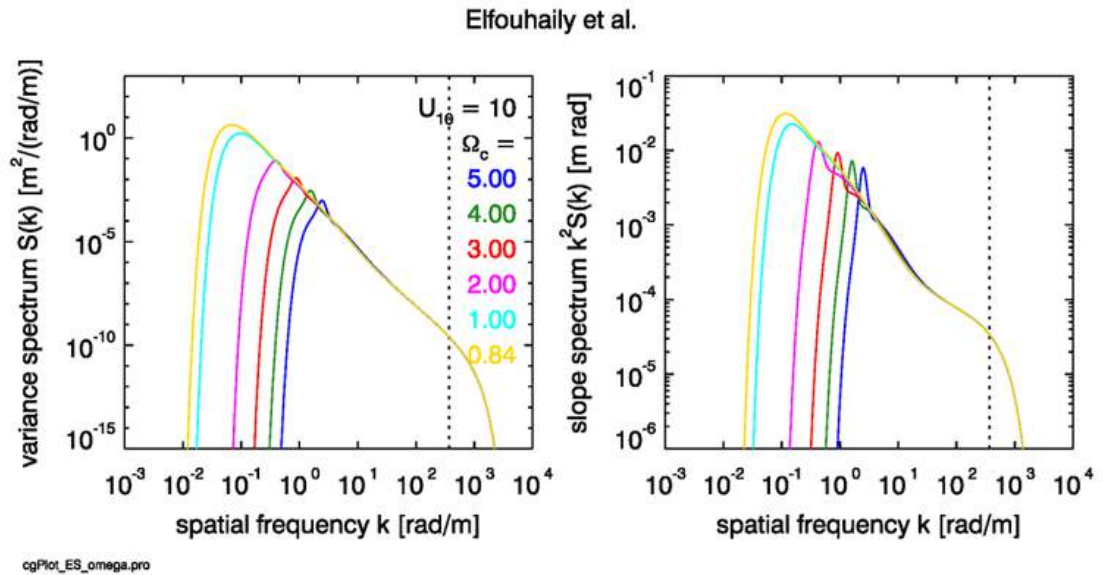


Figure B.5: The omnidirectional part of the ECKV elevation spectrum (left panel) and slope spectrum (right panel) for a wind speed of $U_{10} = 10 \text{ m s}^{-1}$ and wave ages from very young ($\Omega_c = 5$) to mature ($\Omega_c = 1$) to fully developed ($\Omega_c = 0.84$). Compare with Fig. B.4.

B.3.3 Spreading Functions

The ECKV spreading function is given by

$$\begin{aligned}\Phi(k, \varphi) &= \frac{1}{2\pi} [1 + \Delta(k) \cos(2\varphi)] \\ &= \frac{1}{2\pi} \{1 + \tanh [a_o + a_p(c/c_p)^{2.5} + a_m(c_m/c)^{2.5}] \cos(2\varphi)\}\end{aligned}\quad (\text{B.29})$$

Note that this function is symmetric about $\varphi = \pi/2$; i.e., the function has symmetric spreading downwind and upwind. This is consistent with a symmetric variance spectrum $\Psi(-\mathbf{k}) = \Psi(\mathbf{k})$ as would be obtained from the Fourier transform of a snapshot of a sea surface. This symmetry will be explained on the following sections.

A commonly used family of alternate spreading functions is given by the “cosine-2s” functions of [Longuet-Higgins et al. \(1963\)](#), which have the form

$$\Phi(k, \varphi) = C_s \cos^{2s}(\varphi/2), \quad (\text{B.30})$$

where the normalizing coefficient is

$$C_s = \frac{1}{2\sqrt{\pi}} \frac{\Gamma(s+1)}{\Gamma(s+1/2)},$$

and s is a spreading parameter that in general depends on k , U_{10} , and wave age. In this equation Γ is the customary gamma function defined by $\Gamma(p) \triangleq \int_0^\infty x^{p-1} e^{-x} dx$ where $p > 0$. The cosine-2s functions are asymmetric, with much stronger downwind than upwind propagation.

The ECKV and cosine-2s spreading functions are illustrated in [Fig. B.6](#). Both of these functions satisfy the normalization condition [\(B.20\)](#). Both spreading functions transition from strongly forward peaked at low spatial frequencies (long gravity waves; the red curves) to curves with significant propagation at right angles to the wind at high frequencies (capillary waves; the blue curves). The cosine-2s curves are asymmetric in $\pm\mathbf{k}$ and have at least a small amount of upwind propagation at all frequencies (except at exactly upwind, $\varphi = \pi$). Not surprisingly, the real ocean is more complicated than either of these models. In particular, observations of long-wave gravity waves tend to show a bimodal spreading about the downwind direction, which transitions to a more isotropic, unimodal spreading at shorter wavelengths ([Heron, 2006](#)). However, the simple models of [Eqs. \(B.29\) and \(B.30\)](#) are adequate for the present purpose of illustrating surface-generation techniques. The effect of the choice of spreading function on the generated waves is illustrated in [Section C.5](#).

The next appendix introduces the important distinction between “one-sided” or “folded” spectra and the associated “two-sided” spectra. The ECKV spectrum as given above is a one-sided spectrum.

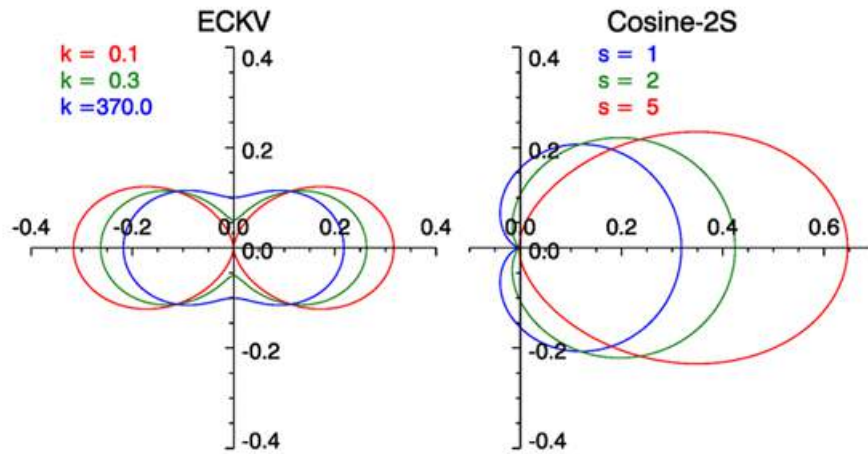


Figure B.6: Example spreading functions according to the ECKV model (left) and the cosine-2s model (right) for a wind speed of 10 m s^{-1} . Small s values correspond to large spatial frequencies k . Downwind is to the right, upwind is to the left.

Modeling Sea Surfaces: Wave Variance Spectra Techniques

Appendices [A](#) on Fourier transforms and [B](#) on wave variance spectra give the tools needed for a very sophisticated and computationally efficient analysis and prediction of random, wind-blown sea surfaces. The next sections walk step by step through the processes of how to compute a wave elevation variance spectrum from sea surface elevation measurements and, conversely, of how to generate a random realization of a sea surface given an elevation variance spectrum. The algorithms are given first for a time-independent, one-dimensional (1-D) sea surface for mathematical simplicity. Two-dimensional (2-D) time-independent surfaces are then presented. It is then shown how to generate a time-dependent sequence of random sea surfaces that can be used, for example, to create a movie loop of propagating waves on the ocean surface. Such techniques are commonly employed in the movie industry. The end result will be the ability to generate random sea surfaces that properly model both the elevation and slope statistics of real sea surfaces. Such surfaces can then be used in Monte Carlo ray-tracing simulations to compute the reflectance and transmission properties (including polarization) of wind-blown sea surfaces, which is described in [Appendix E](#).

C.1 Surfaces to Spectra: 1-D

This section uses a contrived example to illustrate the basic theory for a one-dimensional (1-D) surface. Detailed comments on this simple example emphasize the mathematical subtleties and physical characteristics of Fourier transforms and wave variance spectra derived from surface elevations. A thorough understanding of this example takes us much of the way to understanding the case of a real ocean surface.

An ad hoc, one-dimensional wave profile is constructed using the formula

$$z(r) = z(x_r) = \sum_{j=0}^{N/2} A(j) \cos[2\pi j x_r \nu_f + \phi(j)] \quad r = 0, \dots, N-1. \quad (\text{C.1})$$

The x_r locations are given by $r\Delta x = rL/N$, where L is the length of the sea surface region being sampled and N is the number of samples. $\nu_f = 1/L$ is the fundamental spatial

frequency, that is, the spatial frequency or wavenumber of the wave with a wavelength of L . The amplitude of the wave at the j^{th} frequency, $j = 1, 2, \dots, N/2$, is chosen to be

$$A(j) = 0.1 \exp(-3j/N),$$

and $A(0) = 0$. The phase of the j^{th} wave component is randomly distributed over $[0, 2\pi)$ using

$$\phi(j) = 2\pi\mathcal{U}$$

where \mathcal{U} is a uniform $[0, 1)$ random number. A different random number is drawn for each j value.

The upper left panel of Fig. C.1 shows the surface generated in this manner for $L = 10$ m, $N = 16$, and a particular set of random phases. Note that N is a power of 2 as will be needed for the FFT. The thin colored lines show the $N/2 + 1 = 9$ waves for each of the frequencies. The blue line is the wave for the fundamental frequency $\nu_f = 1/L = 0.1 \text{ m}^{-1}$; the thin black line is the two-point wave at the Nyquist frequency $\nu_{\text{Ny}} = 1/(2\Delta x) = 0.8 \text{ m}^{-1}$; the purple line is the constant $j = 0$ wave, which is set to $z = 0$ for the mean sea surface. The black dots connected by the thick black line show the sum of the individual waves. These points represent a discrete sampling of the continuous sea surface elevation.

In this example, the sampled region of the sea surface is $L = 10$ m in length, but the $N = 16$ samples do not include the point at $x = 10$ m. This is because the surface elevation at $x = L$ is always the same as at $x = 0$ when using Fourier techniques. Resolving the surface as a sum of sinusoids that are harmonics of the fundamental frequency $\nu_f = 1/L$ gives sinusoids that always return to their initial value after distance L . Real sea surfaces are of course not periodic, but we do not know the true value at L because it was not measured by the present sampling scheme. Likewise, we do not know the true surface elevations at points in between the sampled locations. *When we use Fourier techniques to generate random surface realization, we are always generating a sea surface that is a periodic tiling*; the tile dimension is L . This periodicity is useful if we want to generate a visual rendering of a large region of sea surface from a smaller computed region; the edges of the small tiles will match perfectly and the larger surface will often look reasonable, if you don't look too closely. An example of a tiled two-dimensional surface can be seen in Mobley (2016, Fig. 3.9).

We now take the sequence of the $N = 16$ real wave elevations $z(x_r) = z(r)$ seen in Fig. C.1 and feed them into an FFT routine. We soon get back 16 complex numbers, the $\hat{z}(\nu_u) = \hat{z}(u)$ Fourier amplitudes, at a set of 16 corresponding frequencies ν_u . The upper right panel of Fig. C.1 plots the real part of the $\hat{z}(u)$ complex numbers, and the lower-left panel plots the imaginary part¹.

Note first that the FFT routine returned both negative and positive spatial frequencies:

$$\nu = \left(-\frac{N}{2} + 1\right)\Delta\nu = -0.7 \text{ m}^{-1}, \dots, -0.1, 0, 0.1, \dots, \frac{N}{2}\Delta\nu = 0.8 \text{ m}^{-1},$$

for a total of $N = 16$ discrete spatial frequencies. Note that the frequency spacing $\Delta\nu$ equals the fundamental frequency $\nu_f = 1/L$. Section A.3.2 discusses the interpretation of negative frequencies. (The order of the frequencies as returned by the FFT routine

¹Here \hat{z} denotes a Fourier amplitude, which is not to be confused with the boldface $\hat{\mathbf{z}}$, which is a unit vector in the $+z$ direction.

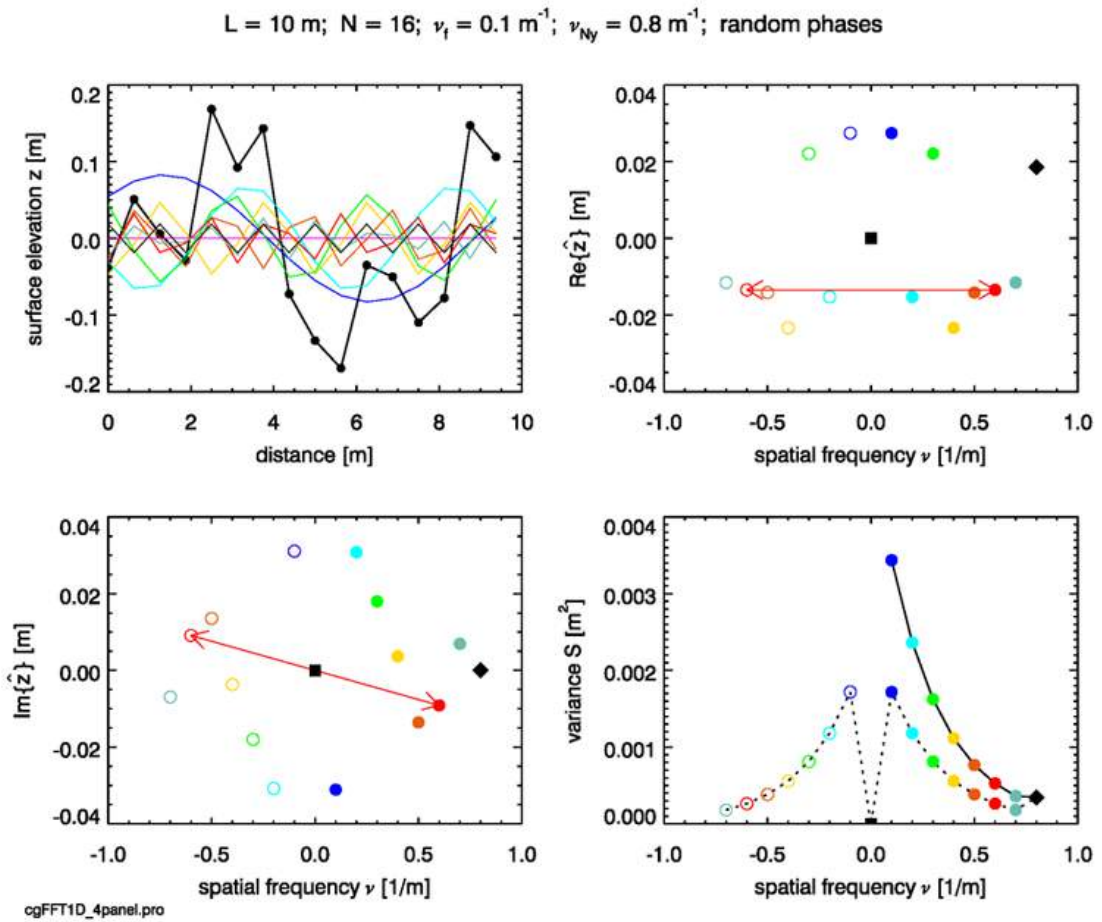


Figure C.1: A one-dimensional random sea surface and its Fourier transform and variance spectrum. The black dots in the upper-left panel show the points $z(x_r)$ of the sampled sea surface. The light colored lines show the sinusoidal components used to create the surface. The upper-right panel shows the real part of the complex amplitude $\hat{z}(\nu_u)$ and the lower-left panel shows the imaginary part. The lower right panel shows the two-sided and one-sided discrete variance spectra. Compare with Figs. C.2 and C.3.

was the “FFT order” discussed in Section A.3.1. The frequencies, and the corresponding amplitudes, were reordered to get the “math order” used for plotting.)

Note next that the real parts of the complex amplitudes $\hat{z}(u)$ are even functions of frequency: $Re\{\hat{z}(-\nu_u)\} = Re\{\hat{z}(+\nu_u)\}$. The imaginary parts are odd functions of frequency: $Im\{\hat{z}(-\nu_u)\} = -Im\{\hat{z}(+\nu_u)\}$. In Fig. C.1 the positive and negative frequencies are connected by a red arrow for one particular frequency pair, $\nu_u = \pm 0.6 \text{ m}^{-1}$. If a complex function $c(\nu) = a(\nu) + ib(\nu)$ has an even real part $a(\nu)$ and an odd imaginary part $b(\nu)$, then $c^*(-\nu) = c(\nu)$, i.e. the function is Hermitian. Thus the plots verify that the amplitudes are Hermitian, as is always the case for the Fourier transform of a real function. The Hermitian character of the complex amplitudes means that these $N = 16$ complex numbers contain only 16 independent real and imaginary numbers, not 32 as would be the case for 16 arbitrary complex numbers. In general the FFT of N real numbers (e.g., N spatial

samples of a sea surface) gives back N independent numbers, so that the “information content” of the physical and Fourier representations is the same.

The positive frequency at $\frac{N}{2}\Delta\nu = 0.8 \text{ m}^{-1}$ is the Nyquist frequency (Section B.1.1). There is, however, no value for the negative of the Nyquist frequency. Note also in the lower left panel that the imaginary part of the amplitude is identically zero at the Nyquist frequency. We will explain these values below.

The lower-right panel of the figure shows the values of $|\hat{z}(\nu_u)|^2$. The values at the negative to positive frequencies are connected by the black dotted line. These points constitute the *two-sided discrete variance spectrum*,

$$\mathcal{S}_{2s}(\nu_u) = |\hat{z}(\nu_u)|^2 \quad \text{for } u = -\frac{N}{2} + 1, \dots, \frac{N}{2}. \quad (\text{C.2})$$

“Two-sided”, denoted by the subscript 2s, refers to spectra showing both the negative and positive frequencies. The variance at zero frequency is the variance contained in the constant mean sea level. This value is zero because we have set the mean sea level to zero.

Oceanographers are often concerned only with the variance at a given magnitude of the spatial frequency, and not with whether the frequency is negative or positive. Nor is there any reason to plot the point at zero frequency, which is usually zero by choice of zero for the mean sea level. It is therefore customary to define the *one-sided* variance spectrum

$$\mathcal{S}_{1s}(\nu_u) = \mathcal{S}_{2s}(-\nu_u) + \mathcal{S}_{2s}(\nu_u), \quad (\text{C.3})$$

for $u = 1, 2, \dots, \frac{N}{2} - 1$, and $\mathcal{S}_{1s}(\nu_{\text{Ny}}) = \mathcal{S}_{2s}(\nu_{\text{Ny}})$. Then only the positive frequencies are plotted. The points connected by the solid black line in the lower-right panel of Fig. C.1 comprise the one-sided variance spectrum. In the present simulation, the two-sided spectrum is symmetric for positive and negative frequencies (except for the Nyquist frequency, which does not have a negative counterpart and is always a special case), and the one-sided function is simply twice the value of the two-sided function for the positive frequencies, except for the Nyquist frequency. When you read a paper and it refers to or plots “the variance (or energy or power) spectrum” without further comment, it is always the one-sided spectrum. However, on the next sections we will have to use two-sided spectra, in which case we will have to account for the magnitude difference in one- and two-sided spectra.

There is an important detail to note in the computation and plotting of $\mathcal{S}(u)$, as in the lower-right panel of Fig. C.1. The values of $\mathcal{S}(u)$ were obtained by the discrete Fourier transform of Eq. (A.7), and $\mathcal{S}(u)$ gives the *variance contained in a finite frequency interval* $\Delta\nu = 1/L$ at the discrete frequency ν_u . $\Delta\nu$ equals the fundamental frequency and is the frequency interval used in the calculations and the plot. As noted in the discussion of the discrete transform, $\mathcal{S}(u)$ is a point function. As was seen in Eq. (A.13) of the Fourier Transforms appendix, if we wish to convert the discrete $\mathcal{S}(u)$ to an estimate of the continuous variance spectral *density* $\mathcal{S}(\nu)$, we must divide by the discrete function by the frequency interval: $\mathcal{S}(\nu) = \mathcal{S}(u)/\Delta\nu$. The units of $\mathcal{S}(\nu)$ are then $m^2/(1/m)$, as expected for a spectral density function of spatial frequency. It is important to distinguish between a discrete variance point function and a continuous variance spectral density.

Now return to Eq. (C.1) and set all of the phases $\phi(j)$ to zero. We are then adding together cosines to create the surface wave profile, which is seen in the upper left panel of Fig. C.2. The FFT of this profile gives the real part of $\hat{z}(\nu_u)$ as positive numbers except for the 0 frequency, and the imaginary part is identically zero for all frequencies.

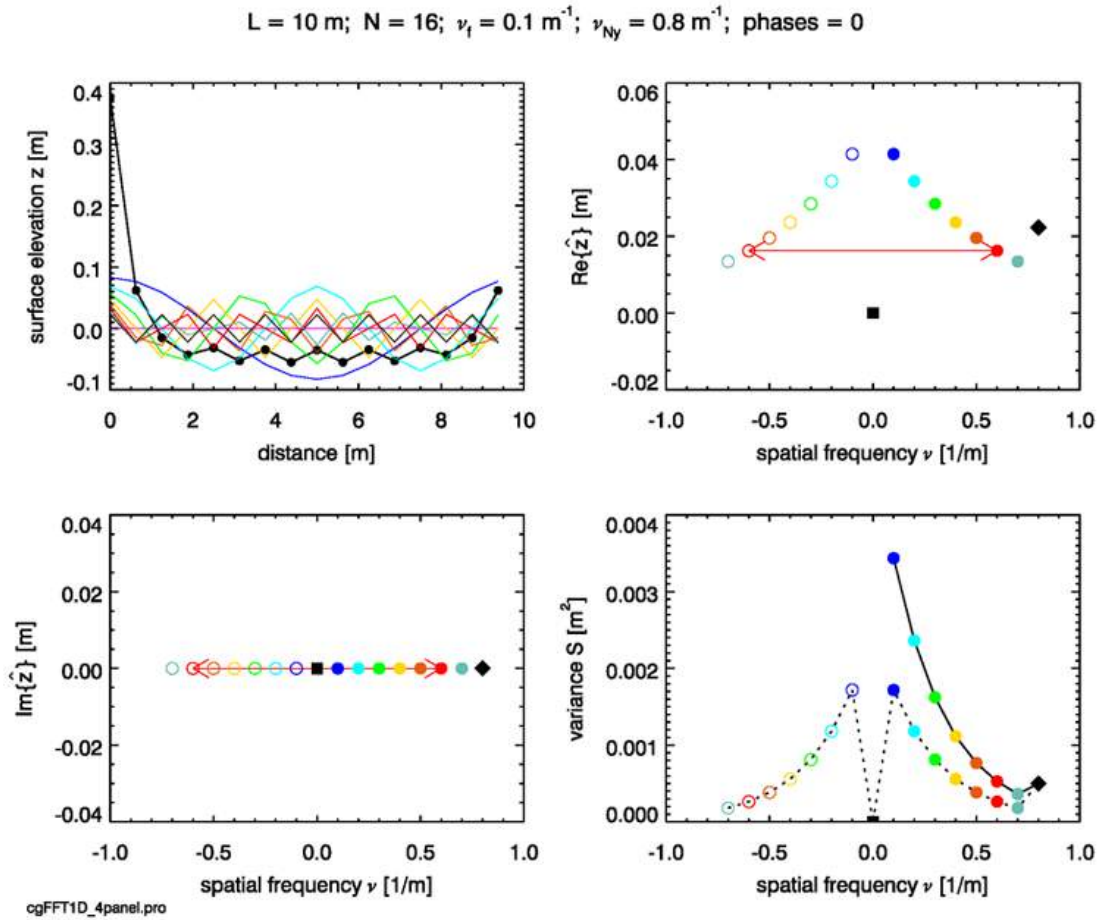


Figure C.2: A one-dimensional surface composed of cosine waves. Compare with Figs. C.1 and C.3.

If we set all of the phases $\phi(j)$ to $\pi/2$, we are then adding together sines to create the surface wave profile, which is seen in the upper left panel of Fig. C.3. The FFT of this profile gives the real part of $\hat{z}(\nu_u)$ identically zero and the imaginary part is nonzero except for the 0 and Nyquist frequencies.

These two figures show that the real part of the complex amplitude $\hat{z}(\nu_u)$ tells us how much of $z(x_r)$ is composed of cosine waves, and the imaginary part shows how much of $z(x_r)$ is composed of sine waves. This explains why the imaginary part of the amplitude is zero at the Nyquist frequency. The two-point wave at the Nyquist frequency is inherently a cosine wave because, as noted previously, a two-point sine wave is sampled only at its zero values. The general case of a wave component with a phase that is neither 0 (nor a multiple of 2π) nor $\pi/2$ (nor an odd integer multiple of $\pi/2$) can be written as a sum of cosine and sine waves, as in Eqs. (B.2) and (B.6) of the wave representations section. In that case, both the real and imaginary parts of the amplitudes are nonzero (except for the special cases of the 0 and Nyquist frequencies).

Note that in each of these three simulations, which differ only by the phases of the component sinusoidal waves, the variance spectrum is exactly the same (except at the

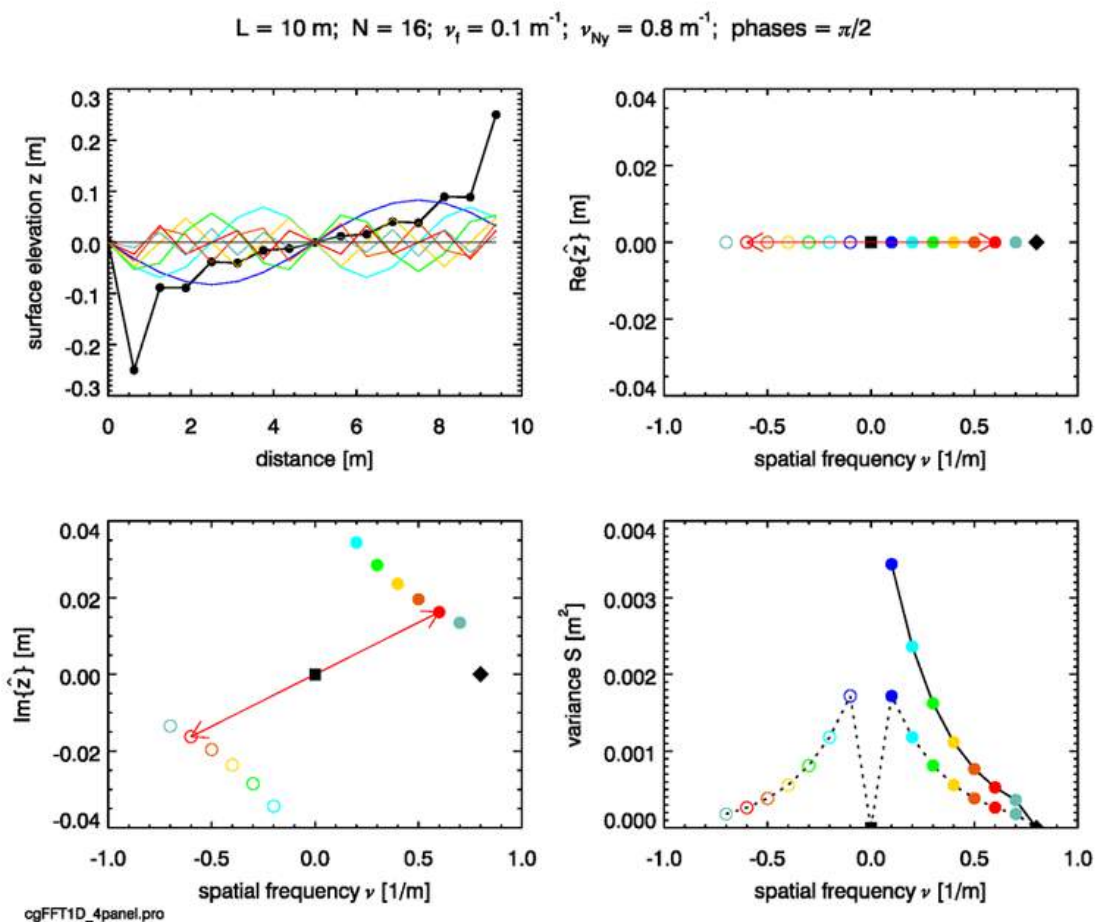


Figure C.3: A one-dimensional surface composed of sine waves. Compare with Figs. C.1 and C.2.

Nyquist frequency), as seen in the lower right panels of Figs. C.1-C.3. That is to say, the variance contained in a wave does not depend on the reference coordinate system used to describe it, even though the Fourier amplitudes $\hat{z}(u)$ do depend on the coordinate system. The variance depends only on the amplitude of the wave. The variance at the Nyquist frequency is largest when cosine waves are added and is zero when sine waves are added. In the first case, we have the maximum possible amplitude of the two-point cosine wave, and in the latter case there is no two-point wave.

C.2 Spectra to Surfaces: 1-D

The previous section showed the major features of the Fourier analysis of a sea-surface elevation record. We started with a sample of a random sea surface and ended with the corresponding discrete variance spectrum (or estimate of the variance spectral density after division by the frequency interval). This section shows how to go in the reverse direction: start with a variance spectrum and generate a random realization of the corresponding sea surface. We first outline the theory, and then show a specific example.

C.2.1 Theory for 1-D Surfaces

To create a one-dimensional (1-D) slice through a sea surface, the essence of the process is as follows:

1. **Choose the domain size.** To create a sea surface over a spatial domain at a given time, we pick the length L of the region $[0, L]$. To generate a time series at a given point, we pick the length of the time series.
2. **Choose the number of points for the DFT.** This number N is the number of frequencies at which we will sample the variance spectrum, and equals the number of samples of the sea surface that will be generated. In normal usage, pick N to be a power of 2 so that an inverse FFT routine can be used to evaluate the inverse DFT.
3. **Choose the frequency variable.** To generate a sea surface over a spatial domain at a given time, we can use either wavenumber ν or angular spatial frequency k . To generate a time series at a given point, we can use either f or ω .
4. **Choose a variance spectrum.** The variance spectrum must be expressed in terms of the chosen frequency variable.
5. **Choose the wind speed.** Pick a wind speed, and perhaps other physical parameters such as the age of the waves to be generated if required by the chosen variance spectrum.
6. **Create random Hermitian amplitudes.** This is the tricky part. We must create an array of randomized discrete Hermitian Fourier amplitudes $\hat{z}(u)$, starting with the chosen continuous variance spectrum.
7. **Take the inverse DFT of the random amplitudes.** The inverse DFT converts the Fourier amplitudes to the physical wave heights.
8. **Extract the sea surface heights.** The inverse DFT of the complex amplitudes returns a complex array. The real part of this array is the random realization of the sea surface heights, and the imaginary part is zero.
9. **Check your results.** This is extremely important during code development. For example, take the DFT of the generated surface heights to see if you get back to the Fourier amplitudes and variance spectrum you started with.

We now proceed through these steps and discuss them in detail for a specific example.

Steps 1 and 2: Let us generate a sea surface over the region from $x = 0$ to $x = L = 100$ m. The longest wavelength that can be resolved is then 100 m. We use $N = 1024$, which gives a spatial grid resolution of $\Delta x = L/N = 0.0977$ m. This means that the shortest wavelength that can be resolved, the two-point wave or Nyquist wavelength, is $2\Delta x = 0.1954$ m.

Step 3: We used wavenumber ν in the previous sections because of its easy interpretation. Now let's use angular spatial frequency k , which is more commonly used. The fundamental frequency is then $k_f = 2\pi/L = 0.0628$ rad m⁻¹. The highest frequency sampled, the Nyquist frequency, is $k_{Ny} = (N/2)k_f = 32.15$ rad m⁻¹. Note that $(2\pi)/k_{Ny} = 0.1954$ m which, as noted above, is the wavelength of the two-point wave.

Step 4: For this example, we use the Pierson-Moskowitz variance spectrum in terms of angular spatial frequency k , which is given by Eq.(C.20). *Note that this is a one-sided spectrum*, which is defined for positive k values.

Step 5: The wind speed at 10 m elevation is $U_{10} = 5 \text{ m s}^{-1}$. The wind speed is the only input to the Pierson-Moskowitz spectrum.

Step 6: We now discuss in detail how to generate a set of random discrete Fourier amplitudes that are physically consistent with the chosen variance spectrum. These amplitudes must be defined for both positive and negative frequencies, and the amplitudes must be Hermitian. We first define

$$\hat{z}_o(k_u) \triangleq \frac{1}{\sqrt{2}} [\rho(k_u) + i\sigma(k_u)] \sqrt{\frac{\mathcal{S}_{1s}(k = k_u)}{2}} \Delta k. \quad (\text{C.4})$$

Here $\mathcal{S}_{1s}(k = k_u)$ denotes the continuous spectral density $\mathcal{S}_{1s}(k)$ evaluated at $k = k_u$. $\Delta k = k_f$ is the spatial frequency sampling interval. $\hat{z}_o(k_u)$ must be defined for both positive and negative discrete frequencies in order to create the Hermitian amplitudes for use in the inverse DFT. As was mentioned on the previous section, we must convert the one-sided continuous spectrum $\mathcal{S}_{1s}(k)$ into a two-sided discrete variance function by

1. dividing its magnitude by 2, assuming that $\mathcal{S}_{2s}(-k) = \mathcal{S}_{2s}(k)$;
2. multiplying the continuous spectral density by the fundamental frequency interval Δk , which gives the variance contained in a finite frequency interval at each frequency k_u .

To emphasize the discrete vs continuous functions, and for brevity of notation, let us write the frequency index u for the frequency k_u . Then Eq. (C.4) becomes

$$\hat{z}_o(u) = \frac{1}{\sqrt{2}} [\rho(u) + i\sigma(u)] \sqrt{\mathcal{S}_{2s}(u)}, \quad (\text{C.5})$$

where $\mathcal{S}_{2s}(u)$ denotes the *two-sided* discrete variance spectrum at frequency k_u . $\hat{z}_o(u)$ can now be evaluated for both positive and negative k_u . The 0 and Nyquist frequencies are always special cases: set $\mathcal{S}_{2s}(0) = 0$ and $\mathcal{S}_{2s}(k_{Ny}) = \mathcal{S}_{1s}(k_{Ny})$. $\rho(k_u) \triangleq \rho(u)$ and $\sigma(k_u) \triangleq \sigma(u)$ are independent random numbers drawn from a normal distribution with zero mean and unit variance, denoted $\rho, \sigma \sim \mathcal{N}(0, 1)$. A different pair is drawn for each u value.

$\hat{z}_o(u)$ is a random variable. Let $\langle \dots \rangle$ denote the expectation of the enclosed variable. The expected value of $|\hat{z}_o(u)|^2$, $\langle \hat{z}_o(u) \hat{z}_o^*(u) \rangle$, gives back whatever variance function is used for $\mathcal{S}_{2s}(u)$:

$$\begin{aligned} \langle \hat{z}_o(u) \hat{z}_o^*(u) \rangle &= \left\langle \left\{ \frac{1}{\sqrt{2}} [\rho(u) + i\sigma(u)] \sqrt{\mathcal{S}_{2s}(u)} \right\} \left\{ \frac{1}{\sqrt{2}} [\rho(u) - i\sigma(u)] \sqrt{\mathcal{S}_{2s}(u)} \right\} \right\rangle \\ &= \frac{\mathcal{S}_{2s}(u)}{2} [\langle \rho^2 \rangle + \langle \sigma^2 \rangle] = \mathcal{S}_{2s}(u) \end{aligned}$$

because $\langle \rho\sigma \rangle = 0$ and $\langle \rho^2 \rangle = \langle \sigma^2 \rangle = 1$ for $\mathcal{N}(0, 1)$ random variables. Thus $\hat{z}_o(u)$ is consistent with the chosen variance spectrum. However, $\hat{z}_o(u)$ is not Hermitian, so the inverse DFT would not give a real sea surface.

Next define $\hat{z}(u)$ as

$$\hat{z}(u) \triangleq \frac{1}{\sqrt{2}} [\hat{z}_o(u) + \hat{z}_o^*(-u)]. \quad (\text{C.6})$$

This function is clearly Hermitian, so the inverse DFT applied to $\hat{z}(u)$ will give a real-valued $z(x_r) \triangleq z(r)$. Moreover, this $\hat{z}(u)$ is consistent with the variance spectrum:

$$\begin{aligned}
\langle |\hat{z}(u)|^2 \rangle &= \langle \hat{z}(u)\hat{z}^*(u) \rangle \\
&= \left\langle \frac{1}{\sqrt{2}} \left[\frac{1}{\sqrt{2}}[\rho(u) + i\sigma(u)]\sqrt{\mathcal{S}_{2s}(u)} + \frac{1}{\sqrt{2}}[\rho(-u) - i\sigma(-u)]\sqrt{\mathcal{S}_{2s}(-u)} \right] \times \right. \\
&\quad \left. \frac{1}{\sqrt{2}} \left[\frac{1}{\sqrt{2}}[\rho(u) - i\sigma(u)]\sqrt{\mathcal{S}_{2s}(u)} + \frac{1}{\sqrt{2}}[\rho(-u) + i\sigma(-u)]\sqrt{\mathcal{S}_{2s}(-u)} \right] \right\rangle \\
&= \frac{1}{4} \left\langle [\rho^2(u) - i\rho(u)\sigma(u) + i\sigma(u)\rho(u) + \sigma^2(u)] \mathcal{S}_{2s}(u) + \right. \\
&\quad [\rho(u)\rho(-u) + i\rho(u)\sigma(-u) + i\sigma(u)\rho(-u) - \sigma(u)\sigma(-u)] \sqrt{\mathcal{S}_{2s}(u)}\sqrt{\mathcal{S}_{2s}(-u)} + \\
&\quad [\rho(-u)\rho(u) - i\rho(-u)\sigma(u) - i\sigma(-u)\rho(u) - \sigma(-u)\sigma(u)] \sqrt{\mathcal{S}_{2s}(-u)}\sqrt{\mathcal{S}_{2s}(u)} + \\
&\quad \left. [\rho^2(-u) + i\rho(-u)\sigma(-u) - i\sigma(-u)\rho(-u) + \sigma^2(-u)] \mathcal{S}_{2s}(-u) \right\rangle \\
&= \frac{1}{2} [\mathcal{S}_{2s}(u) + \mathcal{S}_{2s}(-u)] = \mathcal{S}_{2s}(u).
\end{aligned}$$

Here we have noted that $\langle \rho(u)\rho(-u) \rangle = 0$, etc., because the random variables are uncorrelated for different u values.

Equations (C.4) and (C.6) are the key to generating random sea surfaces from variance spectra. $\hat{z}(u)$ defined by these equations contains random noise, which leads to a sea surface with random amplitudes and phases for the component waves of different frequencies. Any one of these surfaces has a variance spectrum that looks like the chosen spectrum plus random noise. However, on average over many realizations, the noise in these spectra will average out, leaving the variance spectrum. Figure C.4 below illustrates these important ideas, but first we must complete the surface generation.

Step 7: Compute the inverse DFT of the $\hat{z}(u)$ of Eq. (C.6). The result is a complex function $Z(x_r)$:

$$Z(x_r) \triangleq Z(r) = \mathcal{D}^{-1}\{\hat{z}(u)\}.$$

A crucial warning to this step is that the $u = 0, \dots, N-1$ elements of the $\hat{z}(u)$ array *must* be in the FFT frequency order given by Eq. (A.15) of the Fourier Transforms appendix when using an FFT routine to evaluate the DFT. $Z(r)$ is returned with x_r values in the order from $x_0 = 0$ to $x_{N-1} = (N-1)\Delta x$.

Step 8: Extract the surface. The inverse DFT returns a complex array $Z(x_r)$ whose real part is the surface elevations $z(x_r)$ and whose imaginary part is 0. The surface elevations are extracted as the real part of $Z(x_r)$:

$$z(x_r) = \text{Re}\{Z(x_r)\}.$$

Step 9: Check the results! There are many places along the way to lose a $\sqrt{2}$ or to mess up array indexing. At the minimum, it is worthwhile to check that the mean of the generated surface is zero, and that the imaginary part of $Z(x_r) = 0$ (to within a small amount of numerical roundoff error).

When developing computer code, or when first learning this material, it is also a good idea to take the forward DFT of $Z(x_r)$ to make sure that the input Fourier amplitudes $\hat{z}(u)$ are recovered, and that the variance spectrum corresponding to $z(x_r)$ is consistent with the one chosen in Step 4. Indeed, it was the failure of this check in surfaces I was

generating using equations from the literature that led me to develop the [Mobley \(2016\)](#) tutorial and this appendix.

Equations (C.4) and (C.6) are, with minor changes in notation, Eqs. (42) and (43), respectively, of [Tessendorf \(2004\)](#). However, Tessendorf’s version of Eq. (C.4) appears to use a one-sided variance spectrum (his example used the one-sided Phillips spectrum of his Eq. (40)) without the division by 2 seen in Eq. (C.4), which is needed to convert the one-sided spectrum to a two-sided spectrum. Nor does he show the Δk factor needed to convert a continuous spectral density to a discrete function. His version of Eq. (C.6) does not contain the overall factor of $1/\sqrt{2}$ seen above. These missing factors mean that in a round-trip calculation

$$\text{variance spectrum} \rightarrow \text{DFT}^{-1} \rightarrow \text{sea surface} \rightarrow \text{DFT} \rightarrow \text{variance spectrum},$$

you do not get back to the original variance spectrum. In other words, *the Tessendorf equations do not conserve wave variance (i.e., wave energy)*. Even if he included the Δk factor in his actual computations, the missing factors of $1/\sqrt{2}$ in his versions of our Eqs. (C.4) and (C.6) give an overall factor one-half on the amplitudes, which corresponds to a factor of four error in the variance. That is, waves generated using the Tessendorf equations have amplitudes that are too large.

[Tessendorf \(2004\)](#) discusses much more than just Fourier transform techniques, and his notes have been very influential in the computer graphics industry. In 2008 he deservedly received an Academy Award for Technical Achievement for showing the movie industry how to generate and render sea surfaces, as well as for his many other pioneering accomplishments in efficiently computing and rendering fluid motions into visually appealing images. (The first movie to use his techniques was *Waterworld*, followed by dozens of others including *Titanic*.) When I checked with him about the missing numerical factors, he readily acknowledged that Eqs. (C.4) and (C.6) are the correct ones, but pointed out that “Hollywood doesn’t care about conservation of energy.” I suppose that should be no surprise, since movies seem to have no problem with rockets going faster than the speed of light, sound propagating through the vacuum of outer space, or time travel that violates causality. Tessendorf’s equations are widely cited (especially in the computer graphics literature), always without comment about the missing scale factors. Even if Tessendorf had included the needed numerical factors in his equations, graphics artists would distort the resulting images to make them look “better,” e.g., to make the ocean waves look bigger than nature would allow. That may be acceptable in a fantasy world, but such laxness is not permissible if we wish to use numerically generated waves to compute sea surface optical properties.

C.2.2 Example: A Roundtrip Calculation

Figure C.4 shows an example of 1-D surface waves generated using the Pierson-Moskowitz spectrum for a wind speed of 5 m s^{-1} , and the recovery of the variance spectrum from the generated surface. The blue curve in the upper-left panel shows the Pierson-Moskowitz spectrum as defined by Eq. (C.20) of Section B.3.1. The red dots show the frequencies at which the continuous spectrum is sampled. Those dots blur together at the higher frequencies because of the log scale, but the k_u points are equally spaced at intervals of the fundamental frequency $\Delta k = k_f = 2\pi/L = 0.0628 \text{ rad/m}$. The last sampled frequency

is $k_{Ny} = 32.17$ rad/m. The bottom panel shows the sea surface elevations $z(x_r)$ generated for a particular sequence of random numbers $\rho(u), \sigma(u)$.

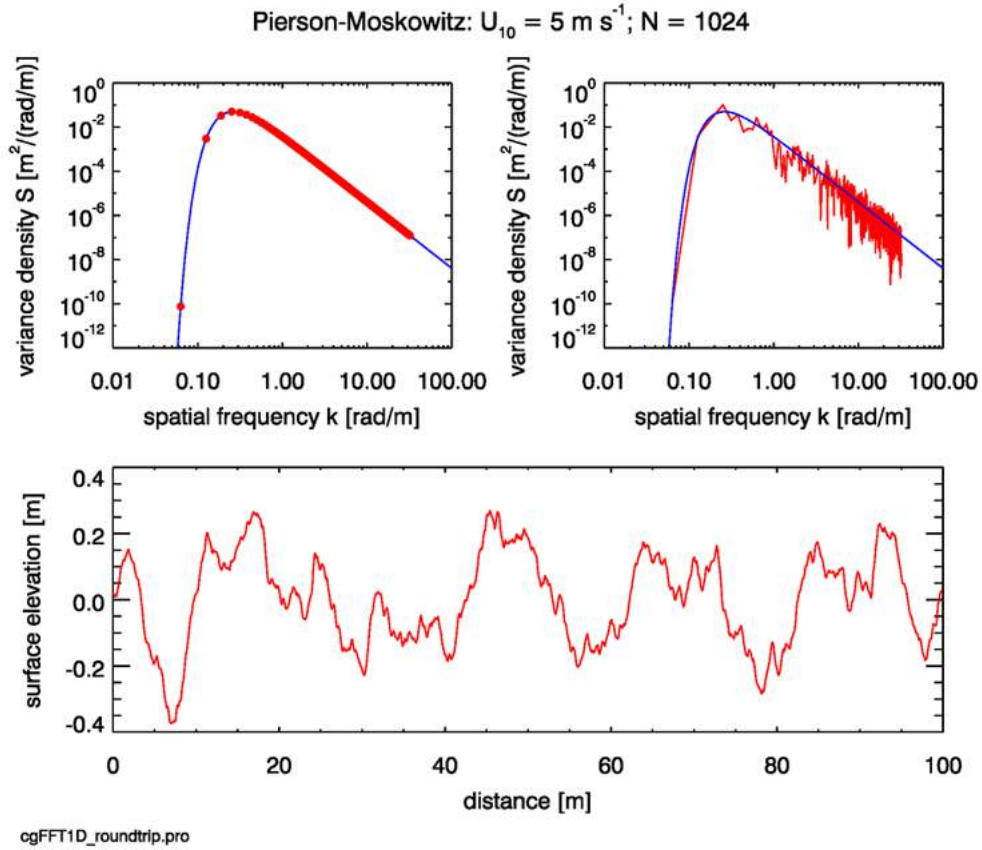


Figure C.4: Example of a 1-D random sea surface generated from the Pierson-Moskowitz spectrum for $U_{10} = 5 \text{ m s}^{-1}$ (the wind speed at 10 m above mean sea level). The longest resolvable wave has wavelength $L = 100\text{m}$. For $N = 1024$, the two-point wave has wavelength $2L/N = 0.195 \text{ m}$. This is already less than the smallest wavelength (highest frequency) for which the Pierson-Moskowitz spectrum should be used.

The red line in the upper-right panel shows the function

$$\mathcal{P}(u) \triangleq \frac{|\hat{z}(u)|^2}{\Delta k} = \frac{1}{\Delta k} |\mathcal{D}\{z(r)\}|^2. \quad (\text{C.7})$$

$\mathcal{P}(u)$ is the discrete variance function for this particular $z(x_r)$ surface. Schuster (1898) called $\mathcal{P}(u)$ a *periodogram*. The periodogram $\mathcal{P}(u)$ contains random noise because $z(x_r)$ is a random realization of the sea surface, which was generated by applying random noise to the variance spectrum. This particular $z(x_r)$ is analogous to a particular measurement of the sea surface. Had we drawn a difference sequence of random numbers for use in Eq. (C.5), we would have generated a different sea surface, and a different $\mathcal{P}(u)$. However, we can expect that if we average together many different $\mathcal{P}(u)$, corresponding to many different sets of $z(x_r)$, the noise would average out and we would be left with a curve close to the variance spectrum we started with, which is shown in blue. Numerical experimentation shows that averaging 100 $\mathcal{P}(u)$ generated from 100 independent sea surface realizations

gives an average $\mathcal{P}(u)$ that is almost indistinguishable from the blue curve at the scale of this plot. Thus $\mathcal{P}(u) \triangleq \mathcal{P}(k_u)$ is an approximation of the variance density spectrum $\mathcal{S}(k)$, denoted $\mathcal{P}(k_u) \doteq \mathcal{S}(k)$. This averaging process leads to the topic of *spectrum estimation*, which considers such problems as how many sets of measurements of $z(x_r)$ are needed to estimate the variance spectrum to within certain error bounds. Fortunately, we need not pursue that here. (The noise in the upper right panel is Gaussian distributed about the theoretical spectrum. However, the log axis makes it look asymmetric about the blue curve.)

At the minimum, you should always check to see that Parseval's relation, Eq. (A.17) of the Fourier Transforms appendix, is satisfied. For the simulation of Fig. C.4, we have

$$\sum_{r=0}^{N-1} |z(r)|^2 = N \sum_{u=0}^{N-1} |\hat{z}(u)|^2 = 19.395 \text{ m}^2.$$

There are sometimes other checks that can be made. For example, the Pierson-Moskowitz spectrum is simple enough that it can be analytically integrated over all frequencies. This gives

$$\langle z^2 \rangle = \int_0^\infty \mathcal{S}_{\text{PM}}(k) dk = 3.04 \cdot 10^{-3} \frac{U_{10}^4}{g^2}, \quad (\text{C.8})$$

where we have recalled that the variance spectral density is related to the variance of the sea surface. The variance of the generated zero-mean sea surface can be computed from

$$\text{var}(z) = \frac{1}{N} \sum_{r=0}^N z^2(x_r) \quad (\text{C.9})$$

and compared with the analytical expectation. For the surface seen in Fig. C.4, Eq. (C.9) gives $\text{var}\{z\} = 0.0189 \text{ m}^2$ versus the theoretical value of $\langle z^2 \rangle = 0.0197 \text{ m}^2$ from Eq. (C.8). This agreement to within a small amount of random noise indicates that all is probably well with the calculations. Indeed, the average $\text{var}(z) \pm$ one standard deviation for 100 independent simulations is 0.020 ± 0.007 , which agrees well with the theoretical value.

The significant wave height $H_{1/3}$ is by definition the height (trough-to-crest distance) of the highest one-third of the waves. To a good approximation, this is related to the expectation of the variance by

$$H_{1/3} = 4\sqrt{\langle z^2 \rangle}.$$

In the present example, this formula gives $H_{1/3} = 0.55 \text{ m}$. The average significant wave height for 100 simulations is 0.56 ± 0.09 (average \pm one standard deviation). If you spend enough time in a sea kayak to develop intuition about wave heights as a function of wind speed, a half-meter height for the largest waves is about right for a 5 m s^{-1} or 10 knot wind speed.

To summarize this section: we have learned how to start with a wave variance spectral density function and generate random discrete Fourier amplitudes. The inverse DFT of those amplitudes gives a random realization of a sea surface that is physically consistent with the chosen variance spectrum. That is all that we can ask of the Fourier transform technique.

C.3 Surfaces to Spectra: 2-D

The preceding sections have given a detailed look at one-dimensional surfaces. We now consider the more useful case of two-dimensional sea surfaces. The extension to two dimensions is mathematically straight forward.

Let $z(\mathbf{x}, t) = z(x, y, t)$ be the sea surface elevation in meters at point $\mathbf{x} = (x, y)$ at time t . The spatial extent of the sea surface is $0 \leq x < L_x$ and $0 \leq y < L_y$. This surface is sampled on a rectangular grid of N_x by N_y points, where both N_x and N_y are powers of 2 for the FFT. The spatial sampling points are then

$$x(r) = [0, 1, 2, \dots, N_x - 1] \frac{L_x}{N_x} = r\Delta x, r = 0, \dots, N_x - 1$$

$$y(s) = [0, 1, 2, \dots, N_y - 1] \frac{L_y}{N_y} = s\Delta y, s = 0, \dots, N_y - 1.$$

This spatial sampling frequency gives N_x and N_y spatial frequencies k_x and k_y (in math frequency order)

$$k_x(u) = [-(N_x/2 - 1), \dots, -1, 0, 1, \dots, N_x/2] \frac{2\pi}{L_x} = u\Delta k_x, u = -(N_x/2 - 1), \dots, N_x/2$$

$$k_y(v) = [-(N_y/2 - 1), \dots, -1, 0, 1, \dots, N_y/2] \frac{2\pi}{L_y} = v\Delta k_y, v = -(N_y/2 - 1), \dots, N_y/2.$$

As before, the x-dimension 0 frequency is at array element $u = N_x/2 - 1$ and the Nyquist frequency is at element $u = N_x - 1$. Thus the positive and negative pairs of k_x values are related by $k_x(u) = -k_x(N_x - 2 - u)$, $u = 0, \dots, N_x - 2$, with the Nyquist frequency always being a special case because there is only a positive Nyquist frequency. Corresponding relations hold for the y direction.

Let $\mathbf{k} = (k_x, k_y)$ denote a spatial frequency vector, where k_x and k_y are frequencies in the x and y directions, respectively. For discrete values, we write $\mathbf{k}_{uv} = (k_x(u), k_y(v))$. The magnitude of \mathbf{k} is $k = \sqrt{k_x^2 + k_y^2}$. In our (x, y) coordinate system, let the wind blow in the $+x$ direction. The $-x$ direction is then upwind, and the $\pm y$ directions are the cross-wind directions. With this choice, $k_x > 0$ indicates frequencies of waves propagating more or less downwind, and $k_x < 0$ for waves propagating against the wind. $k_x = 0$ and $k_y \neq 0$ indicates a wave propagating at exactly a cross-wind $\pm y$ direction. The angle of wave propagation relative to the downwind direction for a wave of frequency $(k_x(u), k_y(v))$ is given by

$$\varphi(\mathbf{k}_{uv}) = \varphi(u, v) = \tan^{-1} \left[\frac{k_y(v)}{k_x(u)} \right]. \quad (\text{C.10})$$

We can thus write the 2-D surface as $z(\mathbf{x})$, its Fourier amplitude as $\hat{z}(\mathbf{k})$, and the associated variance spectrum as $\Psi(\mathbf{k})$. The discrete variance spectrum $\Psi(\mathbf{k}_{uv}) = \Psi(u, v)$ gives the variance of the wave with wavelength $2\pi/k_{uv}$ propagating in direction $\varphi(\mathbf{k}_{uv})$ relative to the downwind direction.

Even though the mathematical transition from one to two dimensions causes no problems, it is again educational to take a careful look at a couple of contrived examples.

C.3.1 Example: A Random Sea Surface

For the first example, a sea surface area of size $L_x \times L_y = 10 \times 10$ m is sampled using $N_x = 16$ and $N_y = 8$ points in the x and y directions. Sea surface elevations $z(x, y)$ were created by drawing a $\mathcal{N}(0, 1)$ random number at each (x, y) value.

The upper-left panel of Fig. C.5 shows a contour plot of this surface for a particular sequence of random numbers. The spatial periodicity of the Fourier representation is used to extend the contour plot to the full (L_x, L_y) range of the tile, which gives a good visual appearance. Thus the elevations at $x = L_x$ are the same as those at $x = 0$, those at $y = L_y$ equal those at $y = 0$, and the elevation at (L_x, L_y) duplicates the elevation at $(0, 0)$. The $N_x \times N_y$ grid of sample points is shown by the solid silver dots. The points at $x = L_x$ and $y = L_y$ obtained by periodicity are shown by open silver circles at the right and top of the surface plot.

The Fourier amplitudes $\hat{z}(\mathbf{k}_{uv}) = \hat{z}(k_x(u), k_y(v))$ are obtained from the 2-D DFT of $z(\mathbf{x}_{rs}) = z(x(r), y(s))$:

$$\hat{z}(\mathbf{k}_{uv}) = \mathcal{D}\{z(\mathbf{x}_{rs})\}.$$

The usual warning on FFT frequency order applies here. The 2-D FFT gives back the amplitudes $\hat{z}(\mathbf{k}_{uv})$ with the array elements corresponding to the FFT frequency order of Eq. (A.15) of the Fourier Transforms appendix. Before plotting, the $\hat{z}(\mathbf{k}_{uv})$ array elements must be shifted into the math frequency order (A.14) in both the k_x and k_y array directions using the appropriate shift function for the computer language used in the plotting. For the IDL routine used to generate Fig. C.5, the 2-D circular shift is given by the IDL command

$$\text{realzhatplot} = \text{SHIFT}(\text{REAL_PART}(\text{zhat}), N_x/2 - 1, N_y/2 - 1),$$

where `zhat` is the complex 2-D array returned by the FFT routine, and `realzhatplot` is the 2-D array contoured in the upper right panel of Fig. C.5.

The upper right panel of Fig. C.5 plots the real part of $\hat{z}(\mathbf{k}_{uv})$, and the lower left panel plots the imaginary part. The Nyquist frequency $k_x^{\text{Ny}} = 2\pi/(2\Delta x) = 5.03$ rad/m lies along the right side of the contour plot. The Nyquist frequency $k_y^{\text{Ny}} = 2\pi/(2\Delta y) = 2.51$ rad/m lies along the top of the contour plot. The white space at the left and bottom highlights that there are no negative Nyquist frequencies. In each amplitude plot a particular pair of $\pm\mathbf{k}_{uv}$ values is indicated by the black arrows; the $(0, 0)$ frequency is shown by a black dot. Note that in the plot of the real part, $\text{Re}\{\hat{z}(-\mathbf{k}_{uv})\} = \text{Re}\{\hat{z}(\mathbf{k}_{uv})\}$, whereas in the plot of the imaginary part, $\text{Im}\{\hat{z}(-\mathbf{k}_{uv})\} = -\text{Im}\{\hat{z}(\mathbf{k}_{uv})\}$. The contouring is rather low quality for so few points, but it is easy to see in the digital output that when the k_x is the Nyquist frequency (the points along the right column of points in the plot), the symmetries are given by $\text{Re}\{\hat{z}(k_x^{\text{Ny}}, -ky)\} = \text{Re}\{\hat{z}(k_x^{\text{Ny}}, +ky)\}$ and $\text{Im}\{\hat{z}(k_x^{\text{Ny}}, -ky)\} = -\text{Im}\{\hat{z}(k_x^{\text{Ny}}, +ky)\}$. A corresponding relation holds for $\pm k_x$ when $k_y = k_y^{\text{Ny}}$ (the points along the top row of the plot). The point at the upper right of the plot corresponds to both k_x and k_y being at their respective Nyquist frequencies. As always, the array elements at the Nyquist frequencies must be treated as special cases when writing computer programs. These symmetries show that the 2-D amplitudes are Hermitian: $\hat{z}^*(-\mathbf{k}_{uv}) = \hat{z}(\mathbf{k}_{uv})$. The discrete 2-D variance spectrum $\Psi(\mathbf{k}_{uv}) = \Psi(u, v) = \hat{z}(u, v)\hat{z}^*(u, v)$ is contoured in the lower right panel of the figure. Note the $\Psi(-\mathbf{k}_{uv}) = \Psi(\mathbf{k}_{uv})$ symmetry.

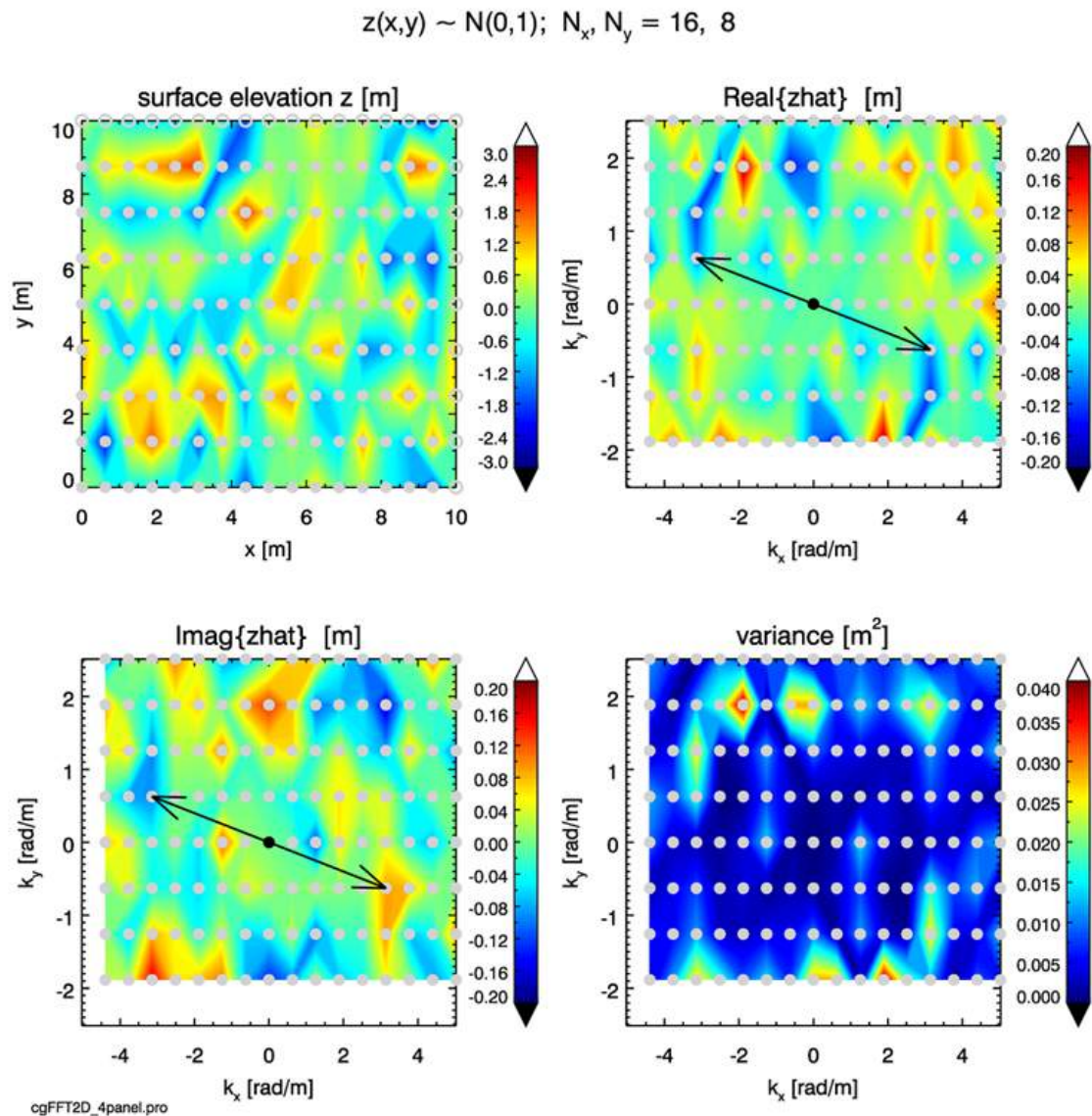


Figure C.5: A two-dimensional random sea surface and its Fourier transform and two-sided variance spectrum. The surface elevations in the upper left panel were randomly drawn from a $\mathcal{N}(0,1)$ distribution. The black arrows highlight a particular $\pm \mathbf{k}_{uv}$ frequency pair. The blue-blue and red-red symmetry of the real part of the Fourier amplitudes, and the red-blue symmetry of the imaginary part, shows the Hermitian nature of the amplitudes. The gray dots show the locations of the discrete values that were contoured to create the figures.

For this example, the standard check on the 2-D discrete Parseval's relation (A.18) gives

$$\begin{aligned}
\sum_{r=0}^{N_x-1} \sum_{s=0}^{N_y-1} |z(r, s)|^2 &= N_x N_y \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} |\hat{z}(u, v)|^2 \\
&= N_x N_y \sum_{u=0}^{N_x-1} \sum_{v=0}^{N_y-1} \Psi(u, v) = 130.90 \text{ m}^2.
\end{aligned}$$

In all of these plots, it should be remembered that the discrete values are known only at the locations of the silver dots. The contouring routine simply interpolates between these points to create a visually appealing figure. The continuous color in the plots does not imply that the values are continuous and known in between the discrete points.

C.3.2 Example: A Sea Surface of Crossing Sinusoids

For a second example, define a surface from a pair of crossing sinusoids as follows:

$$z(x_r, y_s) = A_1 \cos(k_{x1}x_r + k_{y1}y_s + \phi_1) + A_2 \cos(k_{x2}x_r + k_{y2}y_s + \phi_2), \quad (\text{C.11})$$

where as usual $r = 0, \dots, N_x - 1$; $s = 0, \dots, N_y - 1$; and where

$A_1 = 1.0 \text{ m}$ is the amplitude of the first wave

$N_{x1} = 2$ is the number of wave lengths in the x direction in $[0, L_x]$ for the first wave

$k_{x1} = 2.0\pi N_{x1}/L_x = 1.257 \text{ rad/m}^2$ is the k_x frequency of the first wave

$N_{y1} = 1$ is the number of wave lengths in the y direction in $[0, L_y]$ for the first wave

$k_{y1} = 2.0\pi N_{y1}/L_y = 0.628 \text{ rad/m}^2$ is the k_y frequency of the first wave

$\phi_1 = 0$ is the phase of the first wave; 0 gives a cosine wave

$A_2 = 0.5 \text{ m}$ is the amplitude of the second wave

$N_{x2} = 4$ is the number of wave lengths in the x direction in $[0, L_x]$ for the second wave

$k_{x2} = 2.0\pi N_{x2}/L_x = 2.513 \text{ rad/m}^2$ is the k_x frequency of the second wave

$N_{y2} = 3$ is the number of wave lengths in the y direction in $[0, L_y]$ for the second wave

$k_{y2} = -2.0\pi N_{y2}/L_y = -1.885 \text{ rad/m}^2$ is the k_y frequency of the first wave

$\phi_2 = \pi/2$ is the phase of the second wave; $\pi/2$ gives a sine wave

The wavelength of the first wave is $\Lambda_1 = 2\pi/\sqrt{k_{x1}^2 + k_{y1}^2} = 4.47 \text{ m}$, and that of the second wave is $\Lambda_2 = 2.00 \text{ m}$. The direction of propagation of the first wave relative to the $+x$ axis is

$$\varphi_1 = \tan^{-1} \left(\frac{k_{y1}}{k_{x1}} \right) = 26.57 \text{ deg},$$

and the direction of propagation of the second wave relative to the $+x$ axis is

$$\varphi_2 = \tan^{-1} \left(\frac{k_{y2}}{k_{x2}} \right) = -36.87 \text{ deg}.$$

The upper left panel of Fig. C.6 shows this surface elevation pattern when Eq. (C.11) is sampled with $N_x = N_y = 16$. The dominant red-blue pattern shows the first wave

oriented with the direction of propagation along either the $+\mathbf{k}_1$ direction at $\varphi_1 = 26.57$ deg or the $-\mathbf{k}_1$ direction at $\varphi_1 = 26.57 + 180 = 206.57$ deg. We cannot of course determine the actual direction $+\mathbf{k}_1$ or $-\mathbf{k}_1$ of propagation from sea surface elevations at a single time. The dominant wave pattern is modulated by the second wave, which has one half the amplitude of the first wave.

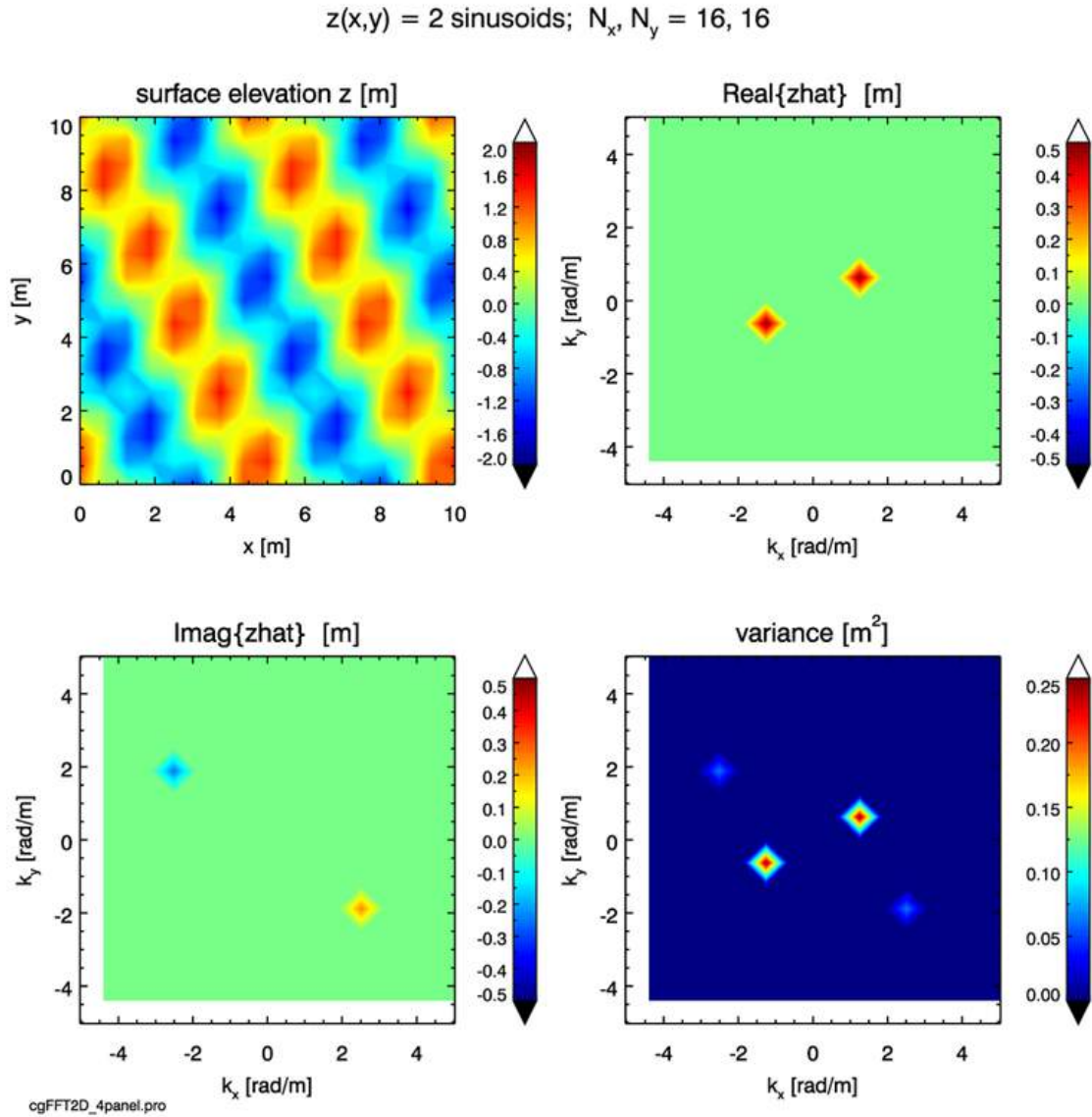


Figure C.6: A two-dimensional sea surface composed of two crossing sinusoids, and the resulting Fourier amplitudes and variance.

The choice above of $\phi_1 = 0$ in Eq. (C.11) makes the dominate wave a cosine in our coordinate system, and $\phi_2 = \pi/2$ makes the second wave a sine. As we saw in the 1-D examples, variance associated with cosine waves appears in the real part of the Fourier amplitudes, and the variance in sine waves appears in the imaginary parts. We see this again here for the first (cosine) and second (sine) waves of the surface. Since each wave

pattern has only one frequency, there is only one pair of points at $\pm \mathbf{k}_1$ in the plot of the real part, and one pair at $\pm \mathbf{k}_2$ in the plot of the imaginary part. The amplitudes at all other frequencies are zero. The symmetries of these points again show the Hermitian nature of the amplitudes.

The lower right panel shows the two-sided variance function $\Psi_{2s}(\mathbf{k})$ for this surface. Note that the first wave has four times the variance of the second wave because the amplitude of the first wave is twice that of the second wave. $\Psi_{2s}(-\mathbf{k}) = \Psi_{2s}(\mathbf{k})$. Note also that you can look at a 2-D variance spectrum and see how much energy is propagating in a given $\pm \mathbf{k}$ direction.

For this simple example involving just four frequencies it is easy (from the digital output) to hand check that the right-hand side of the 2-D discrete Parseval's relation, Eq. (A.18), is

$$N_x N_y \sum_u \sum_v |\hat{z}(u, v)|^2 = 16 \cdot 16 [(0.5)^2 + (0.5)^2 + (0.25)^2 + (-0.25)^2] = 160 \text{ m}^2.$$

This value agrees exactly with the corresponding sums of the surface elevations, $\sum_r \sum_s [z(r, s)]^2$, and variance values, $N_x N_y \sum_u \sum_v \Psi(u, v)$.

C.4 Spectra to Surfaces: 2-D

Now, again, we face the reverse process: start with a two-dimensional, one-sided variance density spectrum and generate a random 2-D realization of a sea surface. The first task is to properly formulate a two-dimensional variance density spectrum, which requires some effort.

C.4.1 Theory

Let $\Psi(\mathbf{k}) = \Psi(k_x, k_y)$ denote a 2-D elevation variance spectrum as defined in Section B.2 on the theory of wave variance spectra. This spectrum has units of $\text{m}^2/(\text{rad}/\text{m})^2$. By definition, the integral of $\Psi(k_x, k_y)$ over all frequencies gives the elevation variance:

$$\text{var}\{z\} = \langle z^2 \rangle = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \Psi(k_x, k_y) dk_x dk_y.$$

As in the 1-D case in Section D.3, $\langle \dots \rangle$ indicates expectation or ensemble average over many measurements of the sea surface.

In the discrete case, the \mathbf{k}_{uv} values coming out of $\hat{z} = \mathcal{D}\{z\}$ point both “downwind” (positive $k_x(u)$ values) and “upwind” (negative $k_x(u)$ values), i.e. there are both positive and negative frequencies represented in the “two-sided” variance spectrum, denoted as $\Psi_{2s}[u, v]$, $u = -(N_x/2 - 1), \dots, N_y/2$, $v = -(N_y/2 - 1), \dots, N_y/2$ as above. In general, $\Psi_{2s}(-\mathbf{k}_{uv}) \neq \Psi_{2s}(\mathbf{k}_{uv})$ because more energy propagates downwind than upwind at a given frequency. $k_x(N_x/2)$ is the Nyquist frequency for waves propagating in the x direction; $k_y(N_y/2)$ is the Nyquist frequency for waves propagating in the y direction. $\Psi_{2s}(0, 0)$ is the variance at zero frequencies, i.e., the variance of the mean sea surface; this term is normally set to 0 so that the mean sea surface is at height 0.

The 2-D equivalents of Eqs. (C.4) and (C.5) in Section D.3 are

$$\hat{z}_o(\mathbf{k}_{uv}) \triangleq \frac{1}{\sqrt{2}} [\rho(\mathbf{k}_{uv}) + i\sigma(\mathbf{k}_{uv})] \sqrt{\frac{\Psi_{1s}(\mathbf{k} = \mathbf{k}_{uv})}{2} \Delta k_x \Delta k_y} \quad (\text{C.12})$$

$$= \frac{1}{\sqrt{2}} [\rho(\mathbf{k}_{uv}) + i\sigma(\mathbf{k}_{uv})] \sqrt{\Psi_{2s}(\mathbf{k}_{uv})}. \quad (\text{C.13})$$

Here, as before, $\rho(\mathbf{k}_{uv})$ and $\sigma(\mathbf{k}_{uv})$ are independent $\mathcal{N}(0, 1)$ random variables, with a different random variable drawn for each \mathbf{k}_{uv} value. As in the 1-D case, the expected value of $\hat{z}_o(\mathbf{k}_{uv})$ gives back whatever variance spectrum is used for $\Psi_{2s}(\mathbf{k}_{uv})$, but is not Hermitian. As before, the notation in these equations distinguishes between the value of the continuous spectral density function evaluated at a discrete frequency value, $\Psi_{1s}(\mathbf{k} = \mathbf{k}_{uv})$, and the discrete variance point function, $\Psi_{2s}(\mathbf{k}_{uv})$.

We must define random Hermitian amplitudes for use in the inverse Fourier transform. Looking ahead to the section on time dependent surfaces (Section C.7), which extends the results of this section to time-dependent surfaces, define the time-dependent spectral amplitude

$$\hat{z}(\mathbf{k}_{uv}, t) \triangleq \frac{1}{\sqrt{2}} [\hat{z}_o(\mathbf{k}_{uv}) e^{-i\omega_{uv}t} + \hat{z}_o^*(-\mathbf{k}_{uv}) e^{+i\omega_{uv}t}]. \quad (\text{C.14})$$

This function is clearly Hermitian, so the inverse DFT applied to $\hat{z}(\mathbf{k}_{uv}, t)$ will give a real-valued $z(\mathbf{x}_{rs}, t)$. Recall that a function of the form $\cos(kx - \omega t)$ gives a wave propagating in the $+x$ direction. The corresponding $\hat{z}_o(\mathbf{k}_{uv}) e^{-i\omega t}$ in this equation gives a wave propagating in the $+\mathbf{k}$ direction, which is in the downwind half plane of all directions. In general the temporal angular frequency ω_{uv} is a function of the spatial frequency k_{uv} . For example, for deep-water gravity waves, $\omega_{uv}^2 = gk_{uv}$.

For simplicity of notation, let us momentarily drop the rs and uv subscripts on the discrete variables. The $\hat{z}(\mathbf{k}, t)$ of Eq. (C.14) is also consistent with the variance spectrum:

$$\begin{aligned} \langle |\hat{z}(\mathbf{k}, t)|^2 \rangle &= \langle \hat{z}(\mathbf{k}, t) \hat{z}^*(\mathbf{k}, t) \rangle \\ &= \left\langle \frac{1}{\sqrt{2}} \left[\frac{1}{\sqrt{2}} [\rho(\mathbf{k}) + i\sigma(\mathbf{k})] \sqrt{\Psi_{2s}(\mathbf{k})} e^{i\omega t} + \frac{1}{\sqrt{2}} [\rho(-\mathbf{k}) - i\sigma(-\mathbf{k})] \sqrt{\Psi_{2s}(-\mathbf{k})} e^{-i\omega t} \right] \times \right. \\ &\quad \left. \frac{1}{\sqrt{2}} \left[\frac{1}{\sqrt{2}} [\rho(\mathbf{k}) - i\sigma(\mathbf{k})] \sqrt{\Psi_{2s}(\mathbf{k})} e^{-i\omega t} + \frac{1}{\sqrt{2}} [\rho(-\mathbf{k}) + i\sigma(-\mathbf{k})] \sqrt{\Psi_{2s}(-\mathbf{k})} e^{i\omega t} \right] \right\rangle \\ &= \frac{1}{4} \left\langle [\rho^2(\mathbf{k}) - i\rho(\mathbf{k})\sigma(\mathbf{k}) + i\sigma(\mathbf{k})\rho(\mathbf{k}) + \sigma^2(\mathbf{k})] \Psi_{2s}(\mathbf{k}) + \right. \\ &\quad [\rho(\mathbf{k})\rho(-\mathbf{k}) + i\rho(\mathbf{k})\sigma(-\mathbf{k}) + i\sigma(\mathbf{k})\rho(-\mathbf{k}) - \sigma(\mathbf{k})\sigma(-\mathbf{k})] \sqrt{\Psi_{2s}(\mathbf{k})} \sqrt{\Psi_{2s}(-\mathbf{k})} e^{i2\omega t} + \\ &\quad [\rho(-\mathbf{k})\rho(\mathbf{k}) - i\rho(-\mathbf{k})\sigma(\mathbf{k}) - i\sigma(-\mathbf{k})\rho(\mathbf{k}) - \sigma(-\mathbf{k})\sigma(\mathbf{k})] \sqrt{\Psi_{2s}(-\mathbf{k})} \sqrt{\Psi_{2s}(\mathbf{k})} e^{-i2\omega t} + \\ &\quad \left. [\rho^2(-\mathbf{k}) + i\rho(-\mathbf{k})\sigma(-\mathbf{k}) - i\sigma(-\mathbf{k})\rho(-\mathbf{k}) + \sigma^2(-\mathbf{k})] \Psi_{2s}(-\mathbf{k}) \right\rangle \\ &= \frac{1}{2} [\Psi_{2s}(\mathbf{k}) + \Psi_{2s}(-\mathbf{k})]. \end{aligned}$$

Here we have noted that all terms like $\langle \rho(\mathbf{k})\rho(-\mathbf{k}) \rangle$ are zero because of the independence of the random variables for different \mathbf{k} values, as are terms like $\langle \rho(\mathbf{k})\sigma(\mathbf{k}) \rangle$. The remaining term is the total variance associated with waves propagating in the downwind and upwind directions at the spatial frequency of magnitude k . It should be noted that this term is independent of time even though the waves $z(x, y, t)$ depend on time. This is because the

total variance (or energy) of the wave field is the same at all times, even though the exact shape of the sea surface varies with time.

If only a “snapshot” of the sea surface at one time is available, it is not possible to resolve how much of the total variance is associated with waves propagating in direction \mathbf{k} compared to the opposite direction $-\mathbf{k}$. The forward DFT, $\hat{z}(\mathbf{k}, t) = \mathcal{D}\{z(x, y, t)\}$, and $\hat{z}(\mathbf{k}, t)\hat{z}^*(\mathbf{k}, t)$ then gives $\Psi_{2s}(-\mathbf{k}) = \Psi_{2s}(\mathbf{k})$, in which case the last equation reduces to

$$\langle \hat{z}(\mathbf{k}, t)\hat{z}^*(\mathbf{k}, t) \rangle = \Psi_{2s}(\mathbf{k}).$$

In any case, the amplitudes defined by Eq. (C.14) are Hermitian, so that the real part of the inverse Fourier transform $Z(\mathbf{x}, t) = \mathcal{D}^{-1}\{\hat{z}(\mathbf{k}, t)\}$ gives a real-valued sea surface $z(\mathbf{x}, t)$. That sea surface is consistent with the variance spectrum $\Psi_{2s}(\mathbf{k})$ at every time t . Wave variance (energy) is thus conserved in a round-trip calculation from variance spectrum to sea surface and back to variance spectrum. In the time-dependent case, if $\Psi_{2s}(\mathbf{k}) > \Psi_{2s}(-\mathbf{k})$, then more variance will be contained in waves propagating in the $+\mathbf{k}$ direction than in the $-\mathbf{k}$ direction. This is all that we can ask from Fourier transform techniques.

Although Eqs. (C.13) and (C.14) are compact representations of the random spectral amplitudes, the actual evaluation of these equations in a computer program warrants further examination. In particular, the Nyquist frequencies are always special cases because there is only a positive Nyquist frequency, $k_x^{Ny} = \frac{N_x}{2} \frac{2\pi}{L_x}$ at array element $N_x - 1$, with a corresponding equation for the Nyquist frequency in the y direction. Writing $e^{\pm i\omega t} = \cos[\omega(k)t] \pm i \sin[\omega(k)t]$ in (C.14) and expanding Eq. (C.14) gives

$$\begin{aligned} 2\hat{z}(\mathbf{k}, t) = & \left[\rho(\mathbf{k})\sqrt{\Psi_{2s}(\mathbf{k})} + \rho(-\mathbf{k})\sqrt{\Psi_{2s}(-\mathbf{k})} \right] \cos[\omega(k)t] \\ & - \left[\sigma(\mathbf{k})\sqrt{\Psi_{2s}(\mathbf{k})} + \sigma(-\mathbf{k})\sqrt{\Psi_{2s}(-\mathbf{k})} \right] \sin[\omega(k)t] \\ & + i \left\{ \left[\rho(\mathbf{k})\sqrt{\Psi_{2s}(\mathbf{k})} - \rho(-\mathbf{k})\sqrt{\Psi_{2s}(-\mathbf{k})} \right] \sin[\omega(k)t] \right. \\ & \left. + \left[\sigma(\mathbf{k})\sqrt{\Psi_{2s}(\mathbf{k})} - \sigma(-\mathbf{k})\sqrt{\Psi_{2s}(-\mathbf{k})} \right] \cos[\omega(k)t] \right\}. \end{aligned} \quad (\text{C.15})$$

These terms are all $N_x \times N_y$ arrays, but note that terms like $\rho(\mathbf{k})\sqrt{\Psi_{2s}(\mathbf{k})}$ represent element-by-element multiplications, not matrix multiplications.

For a particular array element $\hat{z}(k_x(u), k_y(v), t) = \hat{z}(u, v, t)$, and using the indexing relation $k(u) = -k(N - 2 - u)$, $u = 0, \dots, N - 2$ for frequencies written in math order, we can write

$$\begin{aligned}
& 2 \hat{z}(u, v, t) \\
&= \left[\rho(u, v) \sqrt{\Psi_{2s}(u, v)} + \rho(N_x - 2 - u, N_y - 2 - v) \sqrt{\Psi_{2s}(N_x - 2 - u, N_y - 2 - v)} \right] \cos[\omega(k(u, v))t] \\
&- \left[\sigma(u, v) \sqrt{\Psi_{2s}(u, v)} + \sigma(N_x - 2 - u, N_y - 2 - v) \sqrt{\Psi_{2s}(N_x - 2 - u, N_y - 2 - v)} \right] \sin[\omega(k(u, v))t] \\
&+ i \left\{ \left[\rho(u, v) \sqrt{\Psi_{2s}(u, v)} - \rho(N_x - 2 - u, N_y - 2 - v) \sqrt{\Psi_{2s}(N_x - 2 - u, N_y - 2 - v)} \right] \sin[\omega(k(u, v))t] \right. \\
&\left. + \left[\sigma(u, v) \sqrt{\Psi_{2s}(u, v)} - \sigma(N_x - 2 - u, N_y - 2 - v) \sqrt{\Psi_{2s}(N_x - 2 - u, N_y - 2 - v)} \right] \cos[\omega(k(u, v))t] \right\}.
\end{aligned} \tag{C.16}$$

This equation allows for efficient computation within loops over array elements. In particular, the code can loop over the non-positive $k_x(u)$ values, $u = 0, \dots, N_x/2$, and over all $k_y(v)$ values, $v = 0, \dots, N_y - 2$ to evaluate the amplitudes for all non-Nyquist frequencies. The positive $k_x(u)$ values are then obtained from the negative $k_x(u)$ values by Hermitian symmetry. The Nyquist frequencies are evaluated by an equation of the same form, but with one or the other index held fixed (e.g., $u = N_x - 1$ while $v = 0, \dots, N_y - 2$). The amplitude $\hat{z}(k_x = 0, k_y = 0)$ at array element $(u, v) = (N_x/2 - 1, N_y/2 - 1)$ is usually set to zero, corresponding to the mean sea level being set to 0. For generation of time-independent surfaces, we can set $t = 0$ so that the cosines are one and the sines are zero, which cuts the number of terms to be evaluated in half.

If the frequencies are in the FFT order, then the last equation has the same general form, but the indexing that expresses Hermitian symmetry is different: $k(u) = -k(N - u)$, $u = 1, \dots, N - 1$, with $k = 0$ being a special case. The equation corresponding to Eq. (C.16) is then

$$\begin{aligned}
& 2 \hat{z}(u, v, t) \\
&= \left[\rho(u, v) \sqrt{\Psi_{2s}(u, v)} + \rho(N_x - u, N_y - v) \sqrt{\Psi_{2s}(N_x - u, N_y - v)} \right] \cos[\omega(k(u, v))t] \\
&- \left[\sigma(u, v) \sqrt{\Psi_{2s}(u, v)} + \sigma(N_x - u, N_y - v) \sqrt{\Psi_{2s}(N_x - u, N_y - v)} \right] \sin[\omega(k(u, v))t] \\
&+ i \left\{ \left[\rho(u, v) \sqrt{\Psi_{2s}(u, v)} - \rho(N_x - u, N_y - v) \sqrt{\Psi_{2s}(N_x - u, N_y - v)} \right] \sin[\omega(k(u, v))t] \right. \\
&\left. + \left[\sigma(u, v) \sqrt{\Psi_{2s}(u, v)} - \sigma(N_x - u, N_y - v) \sqrt{\Psi_{2s}(N_x - u, N_y - v)} \right] \cos[\omega(k(u, v))t] \right\}.
\end{aligned} \tag{C.17}$$

However, this equation does not explicitly show the special cases. Let $\text{zhat}[u, v]$ be the array of $\hat{z}(\mathbf{k}, t) = \hat{z}(k_x(u), k_y(v), t)$ values at a particular time t . $\text{r}[u, v]$ and $\text{s}[u, v]$ are the arrays of random numbers $\rho(u, v)$ and $\sigma(u, v)$, respectively. $\text{Psiroot}[u, v]$ is $\frac{1}{2} \sqrt{\Psi_{2s}(k_x(u), k_y(v))}$ (incorporating the 2 seen on the left-hand side of Eq. (C.17)). With other obvious definitions, Fig. C.7 shows the pseudo code to evaluate Eq. (C.17) at a particular time t .

```

LOOP OVER u=1,Nx-1
  LOOP OVER v=1,Ny-1
    zhat[u,v] = COMPLEX(
      ( r[u,v]*Psiroot[u,v]
        + r[Nx-u,Ny-v]*Psiroot[Nx-u,Ny-v] ) *cosomegat[u,v]
      - ( s[u,v]*Psiroot[u,v]
        + s[Nx-u,Ny-v]*Psiroot[Nx-u,Ny-v] ) *sinomegat[u,v] ,
      ( r[u,v]*Psiroot[u,v]
        - r[Nx-u,Ny-v]*Psiroot[Nx-u,Ny-v] ) *sinomegat[u,v]
      + ( s[u,v]*Psiroot[u,v]
        - s[Nx-u,Ny-v]*Psiroot[Nx-u,Ny-v] ) *cosomegat[u,v] )
  END v LOOP
END u LOOP

```

Figure C.7: Pseudocode for looping over non-zero frequencies.

Figure C.8 shows the pseudocode for looping over all k_y values for the special case of $k_x = 0$ at frequency array index $u = 0$. Note that for $k_x = 0$, the $\pm k_y$ values are complex conjugates.

```

LOOP OVER v=1,Ny/2
  zhat(0,v) = COMPLEX(
    ( r(0,v)*Psiroot[0,v]
      + r[0,Ny-v]*Psiroot[0,Ny-v] ) *cosomegat[0,v]
    - ( s[0,v]*Psiroot[0,v]
      + s[0,Ny-v]*Psiroot[0,Ny-v] ) *sinomegat[0,v] ,
    ( r[0,v]*Psiroot[0,v]
      - r[0,Ny-v]*Psiroot[0,Ny-v] ) *sinomegat[0,v]
    + ( s[0,v]*Psiroot[0,v]
      - s[0,Ny-v]*Psiroot[0,Ny-v] ) *cosomegat[0,v] )
  zhat(0,Ny-v) = CONJ( zhat(0,v) )
END u LOOP

```

Figure C.8: Pseudocode for looping over all k_y values for the special case of $k_x = 0$ at frequency array index $u = 0$.

Figure C.9 shows the pseudocode for looping over all k_x values for the special cases of $k_y = 0$ at frequency array index $v = 0$. Note that for $k_y = 0$, the $\pm k_x$ values are complex conjugates.

```

LOOP OVER u=1,Nx/2
  zhat(u,0) = COMPLEX(
    ( r(u,0)*Psiroot[u,0]
      + r[Nx-u,0]*Psiroot[Nx-u,0] ) *cosomegat[u,0]
    - ( s[u,0]*Psiroot[u,0]
      + s[Nx-u,0]*Psiroot[Nx-u,0] ) *sinomegat[u,0] ,
    ( r[u,0]*Psiroot[u,0]
      - r[Nx-u,0]*Psiroot[Nx-u,0] ) *sinomegat[u,0]
    + ( s[u,0]*Psiroot[u,0]
      - s[Nx-u,0]*Psiroot[Nx-u,0] ) *cosomegat[u,0] )
  zhat(Nx-u,0) = CONJ( zhat(u,0) )
END v LOOP

```

Figure C.9: Pseudocode for looping over all k_x values for the special cases of $k_y = 0$ at frequency array index $v = 0$.

Finally, set the $(k_x, k_y) = (0, 0)$ value to 0, which sets the mean sea level to zero: $\text{zhat}[0,0] = \text{COMPLEX}(0.0, 0.0)$.

Array $\hat{z}[u,v] = \hat{z}(k_x(u), k_y(v))$ as just defined is Hermitian and has the frequencies in FFT order ready for input to an FFT routine.

Showing this level of detail may seem tedious, but it is absolutely critical that the $\hat{z}(\mathbf{k}, t)$ array be properly computed down to the last array element. Any error will show up in the generated sea surface as either a non-zero imaginary part of the complex array $Z[r, s] = \mathcal{D}\{\hat{z}[u, v]\}$ (easy to detect) or an incorrect sea surface elevation (often much harder to detect).

C.4.2 Example: A Two-Dimensional Sea Surface

As a specific example of the above algorithm, consider the following. Let us use a coarse grid sampling of $N_x \times N_y = 64 \times 64$ points, which makes it easier to see certain features in the associated plots. The physical region to be simulated is $L_x \times L_y = 100 \times 100$ m. The two-dimensional, one-sided variance spectrum of [Elfouhaily et al. \(1997\)](#) is used. Thus we have

$$\begin{aligned}\Psi(k_x, k_y) &= \frac{1}{k} \mathcal{S}(k) \Phi(k, \varphi) \triangleq \Psi(k, \varphi) \\ &= \Psi\left(k = \sqrt{k_x^2 + k_y^2}, \varphi = \tan^{-1}(k_y/k_x)\right).\end{aligned}$$

where $\mathcal{S}(k)$ is the omnidirectional spectrum and $\Phi(k, \varphi)$ is the nondimensional spreading function, as defined in Section B.3.

The upper left plot of Fig. C.10 shows this spectrum for a wind speed of 5 m s^{-1} and a fully developed sea state. The contours are of $\Psi_{1s}(k_x, k_y)$ evaluated at the discrete grid points; the $\frac{1}{2}\Delta k_x \Delta k_y$ factor seen in Eq. (C.13) has not yet been applied to create a discrete two-sided spectrum. The line along $k_y = 0$ corresponds to the spectrum plotted in Fig. B.4 in the examples of wave variance spectra. We see in both plots that for a 5 m s^{-1} wind, the spectrum peaks near $0.2 - 0.3 \text{ rad/m}$.

The plots of the real and imaginary parts of $\hat{z}(k_x(u), k_y(v))$ show that most of the variance is at low frequencies and that the amplitudes have the Hermitian symmetry illustrated in the two figures of the previous section. The lower right panel of the plot shows a contour plot of the sea surface generated from the inverse FFT of the amplitudes. The significant wave height for this surface realization is 0.60 m , in good agreement with the expected value given above for the Pierson-Moskowitz spectrum, which is similar to the Elfouhaily et al. spectrum in the gravity-wave region.

C.5 Spreading Function Effects

Figure C.10 shows a contour plot of a two-dimensional, one-sided variance spectrum $\Psi_{1s}(k_x, k_y)$ and a contour plot of a random surface generated from that variance spectrum. A particular spreading function is implicitly contained in that two-dimensional variance spectrum. The effect on the generated sea surface of the spreading function contained within $\Psi_{1s}(k_x, k_y)$ warrants discussion.

As we have seen (e.g. Eq. (B.27)), a 2-D variance spectrum is usually partitioned as

$$\Psi(k_x, k_y) = \frac{1}{k} \mathcal{S}(k) \Phi(k, \varphi) \triangleq \Psi(k, \varphi).$$

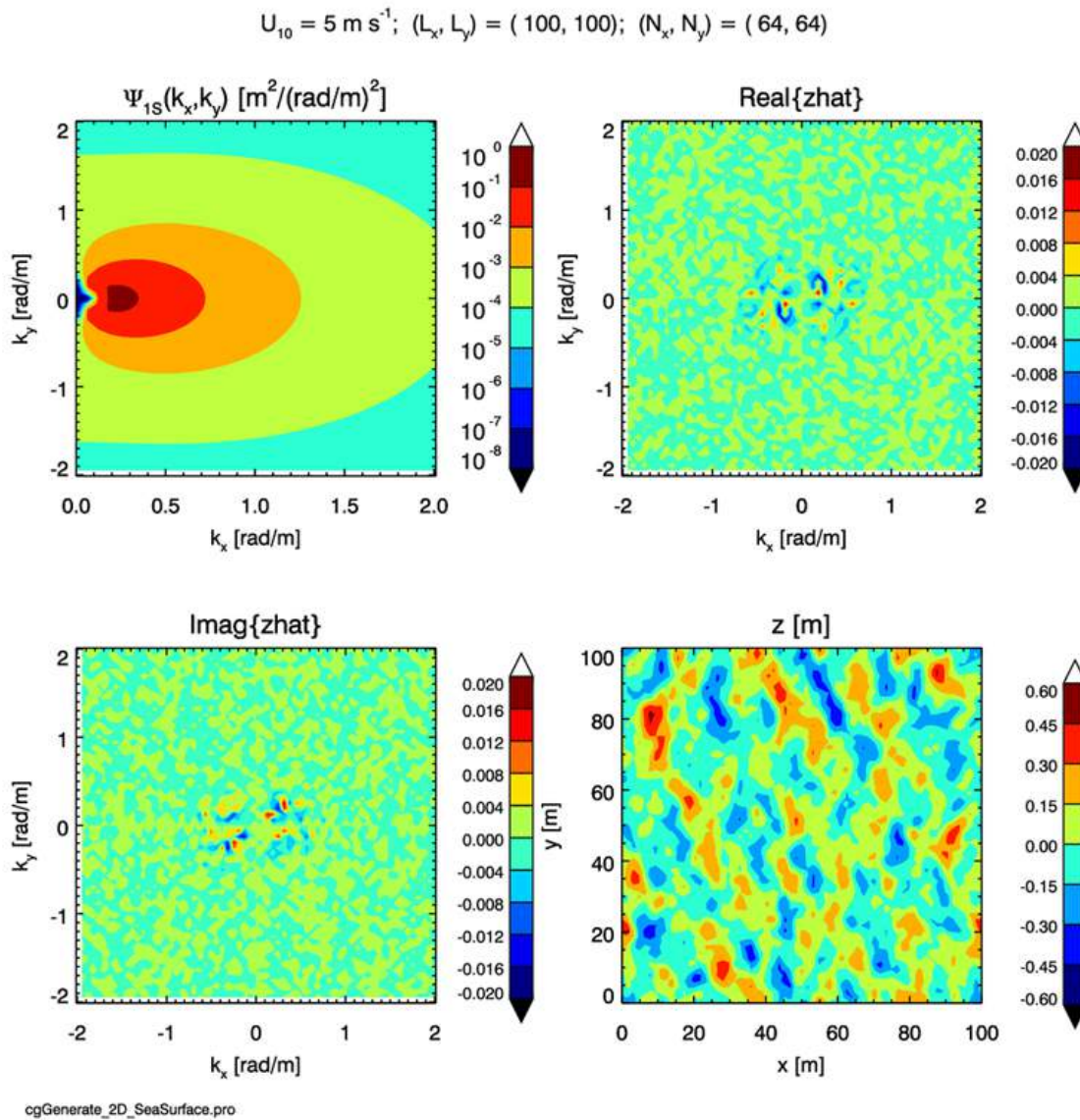


Figure C.10: Example two-dimensional sea surface generated from a 2-D, one-sided variance spectrum. The resolution is only 64×64 grid points, so as to make the features of the underlying spectrum and the Fourier amplitudes easier to see.

Here $\mathcal{S}(k)$ is the omnidirectional spectrum, and $\Phi(k, \varphi)$ is the nondimensional spreading function, which shows how waves of different frequencies propagate (or “spread out”) relative to the downwind direction at $\varphi = 0$.

One commonly used family of spreading functions is given by the “cosine-2S” functions of Longuet-Higgins et al. (1963), which have the form

$$\Phi(k, \varphi) = C_S \cos^{2S}(\varphi/2), \quad (\text{C.18})$$

where S is a spreading parameter that in general depends on k , wind speed, and wave age.

C_S is a normalizing coefficient that gives

$$\int_0^{2\pi} \Phi(k, \varphi) d\varphi = 1 \quad (\text{C.19})$$

for all k .

Figure C.11 shows the cosine-2S spreading functions for values of $S = 2$ and 20. These spreading functions are strongly asymmetric in φ , so that more variance (wave energy) is associated with downwind directions ($|\varphi| < 90^\circ$) than upwind ($|\varphi| > 90^\circ$). The larger the value of S , the more the waves propagate almost directly downwind ($\varphi = 0$), rather than at large angles relative to the downwind direction. However, the cosine-2S spreading functions always have a least a tiny bit of energy propagating in upwind directions, as can be seen for the $S = 2$ curves. This is crucial for the generation of time-dependent surfaces, as will be discussed on the next section.

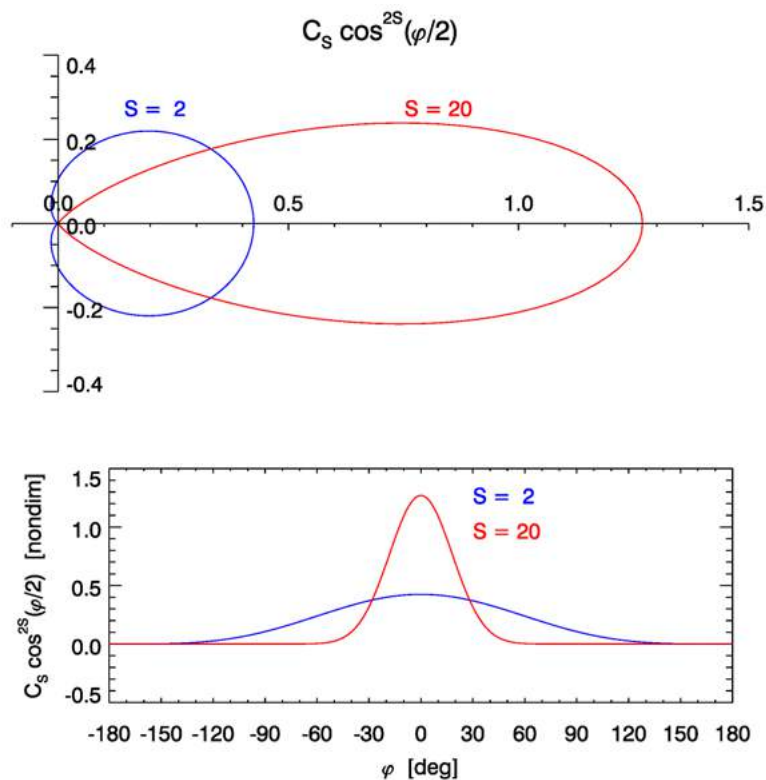


Figure C.11: The cosine-2S spreading functions for $S = 2$ and 20. Top panel: polar plot in φ ; bottom panel: linear in φ .

Figure C.12 shows a surface generated with the omnidirectional variance spectrum of Elfouhaily et al. (1997) (ECKV) as used in the previous section, combined with a cosine-2S spreading function (C.18) with $S = 2$ for all k values. The wind speed is 10 m s^{-1} . The simulated region is $100 \times 100 \text{ m}$ using 512×512 grid points. Note in this figure that the mean square slopes (mss) compare well with the corresponding Cox-Munk values shown in Table C.1. The mss values for the generated surface are computed from finite differences,

e.g.

$$mss_x(r, s) = \frac{z(r+1, s) - z(r, s)}{x(r+1) - x(r)}$$

averaged over all points of the 2-D surface grid. The $\langle \theta_x \rangle$ and $\langle \theta_y \rangle$ values shown in the figure are the average angles of the surface from the horizontal in the x and y directions. These are computed from

$$\theta_x(r, s) = \tan(mss_x(r, s)),$$

etc., averaged over all points of the surface.

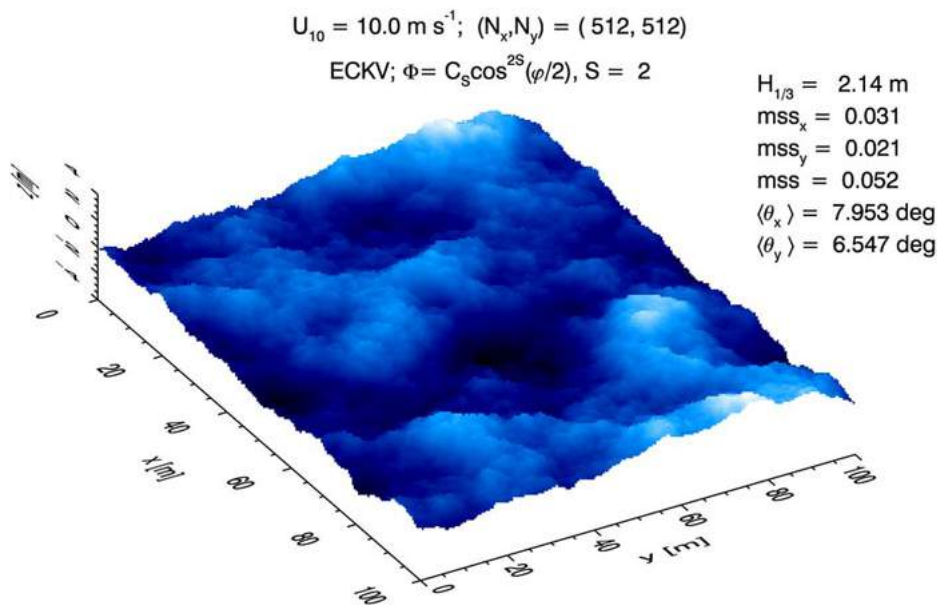


Figure C.12: A sea surface generated with the ECKV omnidirectional spectrum and a cosine-2S spreading function with $S = 2$. Compare with Fig. C.13. (Same as Fig. 13.13)

slope variable	DFT value	Cox-Munk formula	Cox-Munk value
mss_x	0.031	$0.0316U$	0.032
mss_y	0.021	$0.0192U$	0.019
mss	0.052	$0.001(3 + 5.12U)$	0.054

Table C.1: Comparison of Cox-Munk mean square slopes and values for the DFT-generated 2-D surface of Fig. C.12.

The spreading function used in Fig. C.12 was chosen (with a bit of trial and error) to give a distribution of along-wind and cross-wind slopes in close agreement with the Cox-Munk values (except for a small amount of Monte-Carlo noise). Figure C.13 shows a surface generated with a cosine-2S spreading function with $S = 20$; all other parameters were the same as for Fig. C.12. This S value gives wave propagation that is much more strongly in

the downwind direction $\varphi = 0$, as would be expected for long-wavelength gravity waves in a mature wave field. The surface waves thus have a visually more “linear” pattern, whereas the waves of Fig. C.12 appear more “lumpy” because waves are propagating in a wider range of angles φ from the downwind.

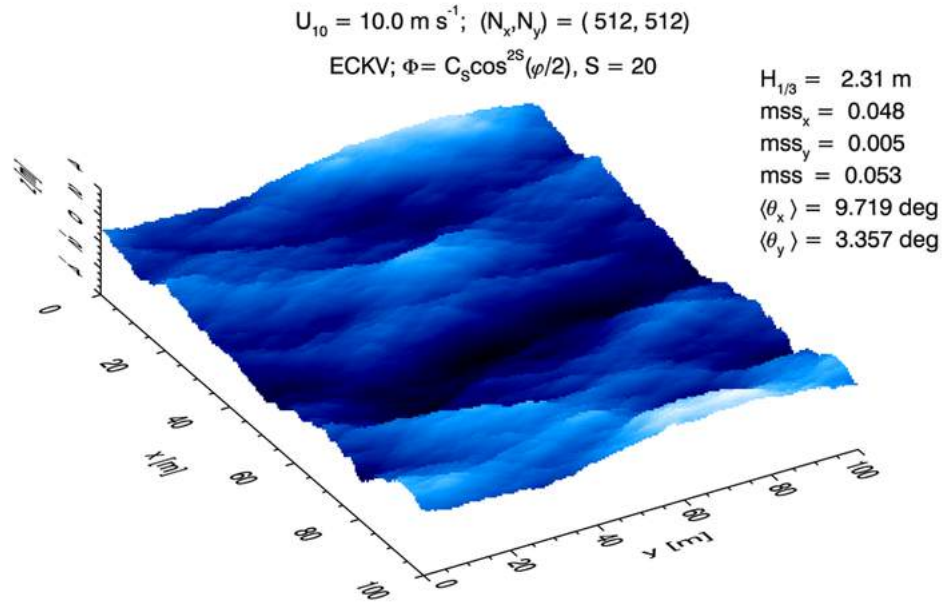


Figure C.13: A sea surface generated with the ECKV omnidirectional spectrum and a cosine-2S spreading function with $S = 20$.

As was shown in Section B.2, the total mean square slope depends only on the omnidirectional spectrum $\mathcal{S}(k)$. Thus the total mss is the same (except for Monte Carlo noise) in both figures C.12 and C.13, but most of the total slope is in the along-wind direction in Fig. C.13.

Real spreading functions are more complicated than the cosine-2S functions used here. In particular, some observations (Heron, 2006) of long-wave gravity waves tend to show a bimodal spreading about the downwind direction, which transitions to a more isotropic, unimodal spreading at shorter wavelengths. Although omnidirectional wave spectra are well grounded in observations, the choice of a spreading function is still something of a black art. You are free to choose any $\Phi(k, \varphi)$ so long as it satisfies the normalization condition (B.20).

C.6 Numerical Resolution

The question naturally arises: “How large a region L and how many points N must be used in generating sea surface realizations?” Not surprisingly, the general answer is, “It depends on your application.” It is computationally possible to create 2-D surfaces with sufficiently large N values in the x and y directions, but run times are prohibitive. Kay et al. (2011) created 2-D surfaces with $2^{16} \times 2^{16} = 65536 \times 65536$ points in each dimension, which sampled the variance spectrum for wavelengths from 200 m gravity waves to 3 mm

capillary waves. However, it took 6 hours to create just one surface on a 3 GHz computer. Note also that the storage of this surface requires $2^{32} = 4.3 \times 10^9$ numbers. Many ray tracing applications need to use tens of thousands to millions of surface realizations to get good statistical averages. Thus, the specific answer for brute-force scientific calculations is usually, “A bigger N than you can run on your computer.” This section shows one way to finesse certain calculations so that large- N calculations can be avoided.

For creating visually appealing sea surfaces for rendering into a movie scene, $N = 2^{10} = 1024$, $2^{11} = 2048$, or perhaps at most $2^{12} = 4096$ is usually adequate. This is because the visual impression of a sea surface is, to first order, determined by the height of the waves, which in turn is governed by the largest gravity waves for the given wind speed. If you are rendering an image, you don’t need to have a finer resolution for the sea surface than maps to an image pixel. For example, if you are going to simulate what a CCD with 1024×1024 pixels would record, and you are looking down onto the sea surface from above at an elevation where one CCD pixel sees a $0.1 \text{ m} \times 0.1 \text{ m}$ patch of the sea surface, then you need to resolve the sea surface with a grid spacing of at most 0.1 m. Modeling a $100 \text{ m} \times 100 \text{ m}$ patch of sea surface with a spatial resolution of 1024×1024 gives a grid spacing of 0.1 m, which would be adequate for the CCD image simulation.

Unfortunately, for accurate scientific calculations of sea surface reflectance, N may have to be much larger. This is because the surface reflectance depends to first order on the surface wave slope, and the slope is strongly affected by the smallest waves with spatial scales down to capillary size of a few millimeters.

The k value of the peak of the Pierson-Moskowitz k spectrum

$$\mathcal{S}_{\text{PM}}(k) = \frac{\alpha}{2k^3} \exp \left[-\beta \left(\frac{g}{k} \right)^2 \frac{1}{U_{19}^4} \right] \quad [\text{m}^2/(\text{rad}/\text{m})], \quad (\text{C.20})$$

is easily obtained from setting $d\mathcal{S}_{\text{PM}}(k)/dk = 0$ and solving for the value of k . This gives

$$k_{\text{p}} = \left(\frac{2\beta}{3} \right)^{0.5} \frac{g}{U_{19}^2},$$

which corresponds to a wavelength of $\Lambda_{\text{p}} = 2\pi/k_{\text{p}}$. Table C.2 shows the values of k_{p} and Λ_{p} for a few wind speeds.

$U_{19}[\text{m/s}]$	$k_{\text{p}} [\text{rad}/\text{m}]$	$\Lambda_{\text{p}} [\text{m}]$
5	0.276	23
10	0.069	91
15	0.031	205
20	0.017	364

Table C.2: Spatial frequencies and wavelengths corresponding to peak variance for the Pierson-Moskowitz spectrum $\mathcal{S}_{\text{PM}}(k)$.

If we use $L \approx 2\Lambda_{\text{p}}$ for the size of the spatial domain, we will resolve the large gravity waves, which contain most of the wave variance. (Keep in mind that density functions do not have a unique maximum: the location of the maximum of a density function depends on

which frequency variable is chosen. See the discussion of this in Section 2.5.) Differentiating the Pierson-Moskowitz ω spectrum

$$\mathcal{S}_{\text{PM}}(\omega) = \frac{\alpha g^2}{\omega^5} \exp \left[-\beta \left(\frac{g}{\omega U_{19}} \right)^4 \right] \quad [\text{m}^2/(\text{rad/s})]. \quad (\text{C.21})$$

and using the dispersion relation $\omega = \sqrt{gk}$ leads to $\Lambda_p = 2\pi(1.25/\beta)^{0.25}g/U_{19}^2$, which gives 73 m at $U_{19} = 10$. However, either version of Λ_p gives adequate guidance for our present purpose.)

Now consider how many points it takes to almost fully resolve, or account for, the variances of elevation and slope spectra when they are sampled for a DFT. For a specific example, take $U_{10} = 10$ m/s and $L = 200$ m and use the omnidirectional, or 1-D, part $\mathcal{S}(k)$ of the Elfouhaily et al. spectrum given by Eq. (B.28) and the associated equations of that section. This variance spectrum and the corresponding slope spectrum $k^2\mathcal{S}(k)$ are plotted as the blue curves in the next 3 figures.

As we have learned in Section B.2, the expected variance of the sea surface is given by

$$\langle z^2 \rangle = \int_0^\infty \mathcal{S}(k) dk, \quad (\text{C.22})$$

and the mean-square slope is given by

$$\langle \sigma^2 \rangle = mss = \int_0^\infty k^2 \mathcal{S}(k) dk. \quad (\text{C.23})$$

Such integrals usually must be numerically evaluated with the limits of $k = 0$ and ∞ being replaced by some small value k_0 and some large but finite value k_∞ . The present example uses $k_0 = 0.01$ and $k_\infty = 10^4$, with 10^6 evaluation points in between these limits. That gives an accurate evaluation of Eqs. (C.22) and (C.23) for the $U_{10} = 10$ m s⁻¹ spectra. The results are $\langle z^2 \rangle = 0.4296$ m² and $mss = 0.06011$ rad².

Now suppose we sample the spectrum using $N = 1024$ points, which leads eventually to a sea surface realization $z(x_r)$ with 1024 points. For $L = 200$ m the fundamental frequency is $k_f = 2\pi/L = 0.0314$ rad/m. This point is shown by the left-most red dot with the black center in Fig. C.14. For this L and N , the Nyquist frequency is $k_{\text{Ny}} = 2\pi/(2L/N) = 16.085$ rad/m, which is shown by the right-most black and red dot. These two end points and the $N - 2$ red vertical lines in between show the discrete points where the variance spectrum is sampled.

The elevation and slope variances that are accounted for in a sampling scheme with N points are given by

$$z^2(N) = \int_{k_f}^{k_{\text{Ny}}} \mathcal{S}(k) dk,$$

and

$$mss(N) = \int_{k_f}^{k_{\text{Ny}}} k^2 \mathcal{S}(k) dk.$$

In the present example, these integrals are $z^2(N = 1024) = 0.4219$ m² and $mss(N = 1024) = 0.02584$ rad². Thus the finite range of the sampled frequencies includes the fraction

$$f_E \triangleq \frac{z^2(N)}{\langle z^2 \rangle} = \frac{0.4219}{0.4296} = 0.982$$

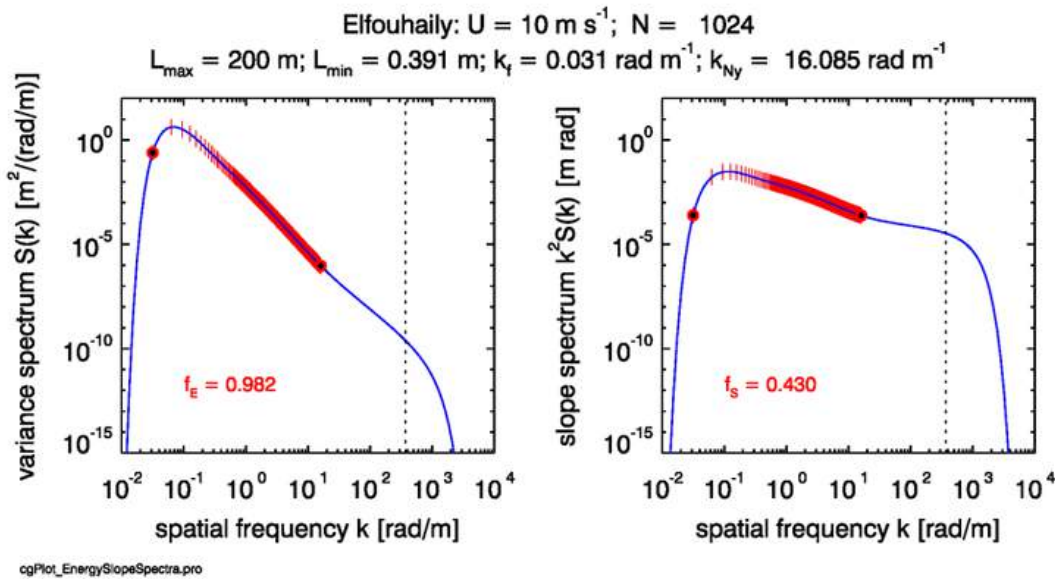


Figure C.14: Example sampling of elevation and slope spectra for $N = 1024$. The red dots and lines show the regions of the elevation and slope spectra sampled using 1024 points. The elevation spectrum is adequately sampled, but the slope spectrum is under sampled.

of the total variance of the sea surface elevation. However, the corresponding fraction of the sampled slope variance is just $f_S = 0.02584/0.06011 = 0.430$. Thus $N = 1024$ is sampling 98% of the elevation variance but only 43% of the slope variance. This sampling is acceptable if we are interested only in creating a sea surface that looks realistic to the eye. The large gravity waves will be correctly simulated, and that is what counts for visual appearance.

However, if our interest is ray tracing to simulate the optical reflectance and transmission properties of the sea surface, then we *must* also sample almost all of the slope variance. This is because, to first order, reflectance and transmission of light are governed by the slope of the sea surface, not by its height. If we under sample the slope variance, then the generated surface will be too smooth at the smallest spatial scales, and the computed optical properties will be incorrect. The brute-force approach to sampling more of the slope spectrum is to increase N .

Figure C.15 shows the sampling points when $N = 2^{16} = 65536$. Now the sampled points extend into the capillary-wave spatial scale: the smallest resolvable wave is 0.006 m. (Note that the fundamental frequency is fixed by the choice of L , which also fixes the spacing of the samples, $\Delta k = k_f$.) The elevation variance integral over the sampled ranges is unchanged (we are still missing a small bit of the variance from the very longest waves with frequencies below k_f). However, the slope integral is now $mss(N = 65536) = 0.05909 \text{ rad}^2$. Thus we are now sampling a fraction $f_S = 0.05909/0.06011 = 0.983$ of the slope variance. The resolution of both the elevation and slope spectra are now acceptable for almost any application. Although using a very large N is computationally feasible in one dimension, such a large N is computationally impracticable in two dimensions, when we would need a grid of size $N \times N$. In the present example, this would require $65536^2 = 4.3 \text{ Gbytes}$ of storage for just one real array, as well as a corresponding increase in run time for the

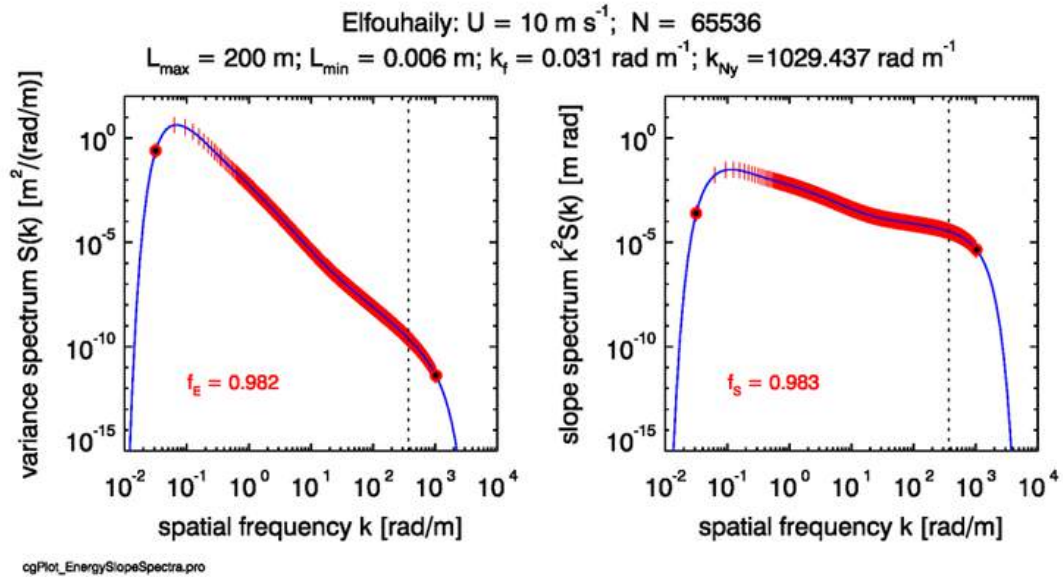


Figure C.15: Example sampling of elevation and slope spectra for $N = 65536$. The red dots and lines show the sampled regions of the elevation and slope spectra. Both the elevation and the slope spectra are adequately sampled.

FFT. Another approach to resolving the slope variance is therefore needed.

As we have just seen, the unsampled frequencies greater than the Nyquist frequency can account for a large part of the slope variance. These frequencies represent the smallest gravity and capillary waves. The amplitudes of these waves are small, so they contribute little to the total wave variance, but their slopes can be large. One way to account for these “missing” waves is to alias their variance into the waves of the highest frequencies. The highest frequency waves that are sampled will then contain too much variance, i.e., they will have amplitudes that are too large for their wavelengths, which will increase their slopes. One way to do this is as follows.

We seek an adjusted or re-scaled elevation spectrum

$$\tilde{\mathcal{S}}(k) \triangleq [1 + \delta(k)]\mathcal{S}(k) \quad (\text{C.24})$$

such that the integral of $k^2\tilde{\mathcal{S}}(k)$ over the sampled region equals the integral of the true $k^2\mathcal{S}(k)$ over the full spectral range. Then sampling the re-scaled spectrum $\tilde{\mathcal{S}}(k)$ will lead to the same slope variance as would be obtained from sampling the correct spectrum $\mathcal{S}(k)$ over the entire range of frequencies.

We can choose L so that the low frequencies are well sampled starting at the fundamental frequency $k_f = 2\pi/L$. There is thus no need to modify the low-frequency part of the variance spectrum, which if done, would adversely affect the total wave variance. We want to modify only the higher frequencies of the sampled region, which contribute the most to the wave slope but little to the wave elevation. A simple approach is to take $\delta(k)$ to be a linear function of k between the spectrum peak k_p and the highest sampled frequency k_{Ny} ,

and zero elsewhere:

$$\delta(k) \triangleq \begin{cases} 0 & \text{if } k \leq k_p \\ \delta_{N_y} \left(\frac{k-k_p}{k_{N_y}-k_p} \right) & \text{if } k > k_p \end{cases} \quad (\text{C.25})$$

δ_{N_y} is a parameter that depends on the spectrum (i.e., the wind speed), the size L of the spatial domain, and the number of samples N . We must determine the value of δ_{N_y} so that this $\delta(k)$ “adds back in” the unresolved slope variance. The added variance will be zero at the peak frequency k_p and largest at the Nyquist frequency. That is, the $\delta(k)$ function will make only a small change to the variance spectrum at the low frequencies, and the change will be largest near the highest sampled frequencies, which is consistent with the idea that the high frequency waves have the largest slopes. (There is nothing physical about the functional form of Eq. (C.25); it is simply a single-parameter *ad hoc* function that works reasonably well in practice. Nonlinear functions could be used, e.g. to concentrate more of the variance into the frequencies closest to the Nyquist frequency. However, those functions could have more than one free parameter to determine and, in any case, the end result is the same: a sea surface that reproduces the height and slope variances of the fully sampled spectrum.)

To determine δ_{N_y} we thus have

$$\begin{aligned} m_{ss} &= \int_0^\infty k^2 \mathcal{S}(k) dk \approx \int_{k_0}^{k_\infty} k^2 \mathcal{S}(k) dk = \int_{k_0}^{k_{N_y}} k^2 \mathcal{S}(k) dk + \int_{k_{N_y}}^{k_\infty} k^2 \mathcal{S}(k) dk \\ &\triangleq \int_{k_0}^{k_{N_y}} k^2 \tilde{\mathcal{S}}(k) dk = \int_{k_0}^{k_{N_y}} k^2 \mathcal{S}(k) dk + \delta_{N_y} \int_{k_p}^{k_{N_y}} k^2 \left(\frac{k-k_p}{k_{N_y}-k_p} \right) \mathcal{S}(k) dk. \end{aligned}$$

The right-most terms of these equations give (after recalling that $\delta(k) = 0$ for $k \leq k_p$)

$$\delta_{N_y} = \frac{\int_{k_{N_y}}^{k_\infty} k^2 \mathcal{S}(k) dk}{\int_{k_p}^{k_{N_y}} k^2 \left(\frac{k-k_p}{k_{N_y}-k_p} \right) \mathcal{S}(k) dk}. \quad (\text{C.26})$$

Figure C.16 shows example 1-D surface realizations created with $N = 1024$ samples and both the true and scaled variance spectra. The upper two panels reproduce Fig. C.14, except that the sampled points of the re-scaled spectra are shown in green. It is clear that the $\delta(k)$ function has added progressively more variance to the higher frequencies. The re-scaled variance spectrum does of course contain somewhat more elevation variance over the sampled region than does the true spectrum. As the green inset f_E number shows, this re-scaling has increased the fraction of sampled/true variance from 0.982 to 1.020. However, the f_S number in the upper-right panel shows that we are now sampling 99.5% of the slope variance, as opposed to 43% for the true spectrum. Having a bit too much total elevation variance is a good tradeoff for being able to model the optically important slope statistics.

The row 2 panel shows random realizations of the 1-D surfaces generated from these two spectra (with the same set of random numbers). The surface elevations $z(x_r)$ generated using the true spectrum (red line) and the re-scaled spectrum (green line) are almost indistinguishable at the scale of this figure. These significant wave heights for these two surface realizations compare well with the significant wave height of $H_{1/3} = 2.48$ m for a

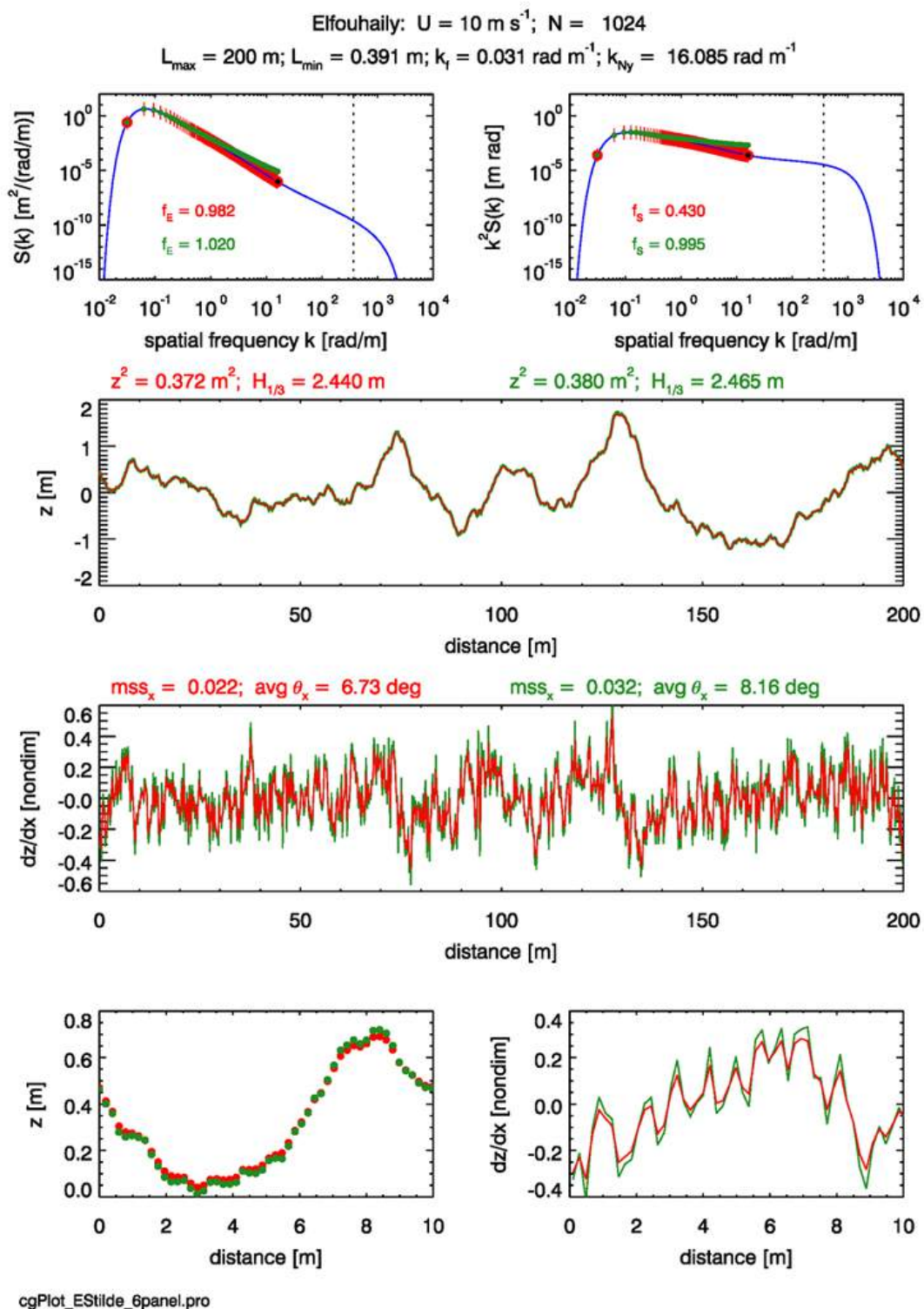


Figure C.16: Example 1-D surfaces and slopes created using true and adjusted variance spectra. See the preceding text for discussion.

Pierson-Moskowitz spectrum. The row 3 panel shows the surface slopes computed from finite differences of the discrete surface heights. The Cox-Munk along-wind mean square slope is given by $m_{ss_x} = 0.0316U_{12.5}$, which compares well with the value of 0.032 obtained with the re-scaled spectrum. However, the value obtained from the true spectrum is only 0.022, or 70% of the Cox-Munk value. The bottom two panels replot the 0-10 m sections of the z and dz/dx figures for better display of the details.

We thus see that when using the $\delta(k)$ correction to a variance spectrum, we can generate a surface that reproduces, to within a few percent of the theoretical values, both the surface elevation and slope statistics that would be obtained from the underlying true variance spectrum if it were sampled with enough points to fully resolve the elevation and slope variances. The variance adjustment technique described in this section was used in [Mobley \(2015\)](#). However, there are other approaches to the problem of unresolved spatial scales, e.g., [D'Alimonte and Kajiyama \(2016\)](#).

C.7 Time-Dependent Surfaces

One final step remains: the addition of time dependence to generate a sequence of the sea surface realizations. Many scientific applications do not need time dependence. This is the case when many independent random realizations of sea surfaces are used for Monte Carlo ray tracing to compute the average optical reflectance and transmittance properties of wind-blown sea surfaces. However, time dependence is crucial for applications such as computer-generated imagery for movies.

We already have the needed theory in hand from Section C.4 on Spectra to Surfaces: 2-D. The fundamental equations from that section are

$$\begin{aligned}\hat{z}_o(\mathbf{k}_{uv}) &\triangleq \frac{1}{\sqrt{2}}[\rho(\mathbf{k}_{uv}) + i\sigma(\mathbf{k}_{uv})] \sqrt{\frac{\Psi_{1s}(\mathbf{k} = \mathbf{k}_{uv})}{2} \Delta k_x \Delta k_y} \\ &= \frac{1}{\sqrt{2}}[\rho(\mathbf{k}_{uv}) + i\sigma(\mathbf{k}_{uv})] \sqrt{\Psi_{2s}(\mathbf{k}_{uv})},\end{aligned}\quad (\text{C.27})$$

and

$$\hat{z}(\mathbf{k}_{uv}, t) \triangleq \frac{1}{\sqrt{2}} [\hat{z}_o(\mathbf{k}_{uv})e^{-i\omega_{uv}t} + \hat{z}_o^*(-\mathbf{k}_{uv})e^{+i\omega_{uv}t}]. \quad (\text{C.28})$$

These quantities are evaluated as described in that section, but with the evaluations done at particular times t . Most importantly, these equations make no simplifying assumptions about the $\pm\mathbf{k}$ symmetry of the two-sided variance spectrum $\Psi_{2s}(\mathbf{k})$.

As we have learned, the Fourier transform of a snapshot of the sea surface gives a two-sided variance spectrum with identical values for $-\mathbf{k}$ and $+\mathbf{k}$. This represents equal amounts of energy propagating in the $-\mathbf{k}$ and $+\mathbf{k}$ directions, i.e., equal amounts of energy in waves propagating upwind and downwind. Such a situation in nature gives standing waves. Here also, if $\Psi_{2s}(-\mathbf{k}) = \Psi_{2s}(\mathbf{k})$, the time-dependent surface will be standing waves composed of waves of equal energy propagating the $+\mathbf{k}$ and $-\mathbf{k}$ directions. To obtain waves propagating downwind, as is the case in a real ocean, we must use an asymmetric two-sided spectrum with $\Psi_{2s}(-\mathbf{k}) \ll \Psi_{2s}(+\mathbf{k})$, so that almost all of the energy is propagating downwind. Note, however, that we cannot simply set $\Psi_{2s}(-\mathbf{k}) = 0$, which represents no energy at all propagating upwind. This is because $\Psi_{2s}(-\mathbf{k}) = 0$ destroys the Hermitian

property of Eq. (C.28). Thus we must conjure up an asymmetric variance spectrum that allows a nonzero (although perhaps negligible) amount of energy to propagate upwind.

An asymmetric two-sided variance spectrum can be constructed using an asymmetric spreading function $\Phi(k, \varphi)$ as described in Section C.5. Spreading functions of the form

$$\Phi(k, \varphi) = C_s \cos^{2S(k)}(\varphi/2) \quad (\text{C.29})$$

described there allow some energy to propagate in $-\mathbf{k}$ directions, i.e. when $|\varphi| > \pi/2$. With this choice of a spreading function, we can let the magnitude of $\Psi_{2s}(+\mathbf{k})$ equal the magnitude of the one-sided variance function $\Psi_{1s}(+\mathbf{k})$, which gives the total variance, because we assume that a negligible amount of the total energy propagates in $-\mathbf{k}$ directions. With this assumption, there is no division of $\Psi_{1s}(+\mathbf{k})$ by 2 in the first line of Eq. (C.13). That is, we are starting with a two-sided, asymmetric spectrum, not with a one-sided spectrum based on the assumption of upwind-downwind symmetry.

Once an asymmetric $\Psi_{2s}(\pm\mathbf{k})$ has been defined, we can evaluate Eq. (C.27) for a set of random numbers $\rho(\mathbf{k}_{uv})$ and $\sigma(\mathbf{k}_{uv})$. *This is done only once.* Then to generate a sequence of sea surface realizations for times $t = 0, \Delta t, 2\Delta t, \dots$, we multiply the time-independent amplitudes by the time-dependent cosines and sines as was shown previously in Eq. (C.17):

$$\begin{aligned} & 2 \hat{z}(u, v, t) \\ &= \left[\rho(u, v) \sqrt{\Psi_{2s}(u, v)} + \rho(N_x - u, N_y - v) \sqrt{\Psi_{2s}(N_x - u, N_y - v)} \right] \cos[\omega(k(u, v))t] \\ & - \left[\sigma(u, v) \sqrt{\Psi_{2s}(u, v)} + \sigma(N_x - u, N_y - v) \sqrt{\Psi_{2s}(N_x - u, N_y - v)} \right] \sin[\omega(k(u, v))t] \\ & + i \left\{ \left[\rho(u, v) \sqrt{\Psi_{2s}(u, v)} - \rho(N_x - u, N_y - v) \sqrt{\Psi_{2s}(N_x - u, N_y - v)} \right] \sin[\omega(k(u, v))t] \right. \\ & \left. + \left[\sigma(u, v) \sqrt{\Psi_{2s}(u, v)} - \sigma(N_x - u, N_y - v) \sqrt{\Psi_{2s}(N_x - u, N_y - v)} \right] \cos[\omega(k(u, v))t] \right\}. \end{aligned} \quad (\text{C.30})$$

We thus obtain the amplitudes $\hat{z}(k_{uv}, t)$ at the current time t . The inverse Fourier transform of this $\hat{z}(k_{uv}, t)$ gives the sea surface $z(\mathbf{x}_{rs})$ at time t .

The physics (or lack thereof) of this process is simple. We start with a realization of the sea surface at time zero. This surface contains waves of many discrete frequencies k_{uv} traveling in all directions φ_{uv} . Then to get the surface at any later time t , we simply propagate the sinusoidal waves of each frequency \mathbf{k}_{uv} in their original direction of travel through a phase change determined by the time step and the dispersion relation $\omega(k_{uv})$. For deep-water gravity waves, the dispersion relation is $\omega(k_{uv}) = \sqrt{gk_{uv}}$. It thus visually appears that the waves are propagating and the sea surface profile is changing with time. However, this Fourier transform technique is really just moving a collection of independent, non-interacting sinusoids through frequency-dependent phase angles to create a new surface realization from the sum of the phase-shifted sinusoids. In a real ocean, waves of difference frequencies can interact with each other (redistribute energy among frequencies) in highly complex and nonlinear ways, so that the sea surface statistics may not be time-independent. This is, in particular, how little waves grow to big waves when the wind begins to blow over a level surface. Modeling the nonlinear evolution of a sea surface is beyond the abilities of

Fourier transform techniques which are, at heart, just a mathematical artifice based on a time-independent directional variance density spectrum.

Figure C.17 shows a sequence of surface realizations for a 10 m s^{-1} wind speed and a spatial grid of size $L_x \times L_y = 100 \times 100 \text{ m}$, and a grid resolution of $N_x \times N_y = 256 \times 256$. A time series of images made with such a course grid could be useful for some non-scientific purposes.

We have seen that the spatial pattern of a sea surface generated by a Fourier transform is periodic. This is convenient for tiling a small computed region into a visually acceptable larger region. When time dependence is included and a finite-length time series of images is created, the sequence of images is not periodic in time because the various sinusoids comprising the surface do not have a common period. As pointed out by Tessendorf (2004), this can be remedied by “quantizing” the temporal frequency $\omega(k_{uv})$ as follows.

Let T_r be the length of time over which the time sequence of surface realizations should exactly repeat. The number of frames N_t in the video loop determines the time step between frames, $\Delta t = T_r/N_t$. For a smoothly moving sea surface, N_t must be large enough that Δt is less than about 0.1 s. Define $\omega_o \triangleq 2\pi/T_r$. For deep-water gravity waves, the true temporal frequency $\omega(k_{uv}) = \sqrt{gk_{uv}}$ is replaced by

$$\tilde{\omega}(k_{uv}) = \left\lfloor \frac{\sqrt{gk_{uv}}}{\omega_o} \right\rfloor \omega_o, \quad (\text{C.31})$$

where $\lfloor x \rfloor$ converts a real number x into its integer part (e.g., 15.23 is converted to 15). This operation slightly alters the temporal frequency of each wave component so that each component returns to exactly its initial shape after time T_r . A video loop can then be created from the sequence of surfaces, and the loop will match perfectly when the frame for time $T_r - \Delta t$ goes to the frame for time T_r , which is the same surface as time 0, after which the surfaces repeat. This adjustment to ω is greatest for the lowest frequencies, but the adjustment becomes smaller and smaller as the repeat time becomes larger and larger. It is thus easy to create a time-dependent small area of sea surface that can be tiled in both space and time to create an image of a larger sea surface over a longer time. This is good enough to fool movie-goers.

In order to employ a re-scaled variance spectrum as described in Section C.6, determine the value of the δ_{Ny} parameter using the omnidirectional variance spectrum, as seen in Eq. (C.24). Then apply the $\delta(k)$ correction to the directional spectrum $\Psi(k_x, k_y)$ with $k = k_{uv}$ for each $(k_x(u), k_y(v))$ point of the rectangular grid.

An example of a 20-second (simulated time) video loop created using all of these tricks can be seen on the corresponding Web Book page at [Time-dependent Surfaces](#). This video shows a time series of 2-D sea surface realizations generated using the Elfouhaily et al. (1997) variance spectrum with a frequency-dependent cosine-2S spreading function. This omnidirectional elevation variance spectrum was adjusted as described in Section C.6 at the higher spatial frequencies so that unresolved slope variance (from frequencies higher than the k_x and k_y Nyquist frequencies resolvable by a 512 by 512 DFT grid) is fully resolved. The true dispersion relation $\omega = \sqrt{gk}$ was quantized for each discrete spatial frequency so that the surface is exactly periodic after 20 seconds. Note in particular that the significant wave heights $H_{1/3}$ are slightly greater on average than the value of 2.25 m predicted by a Pierson-Moskowitz spectrum, which has somewhat less energy than the Elfouhaily et al. spectrum used here. Note also that the along-wind (mss_x) and cross-wind (mss_y) mean

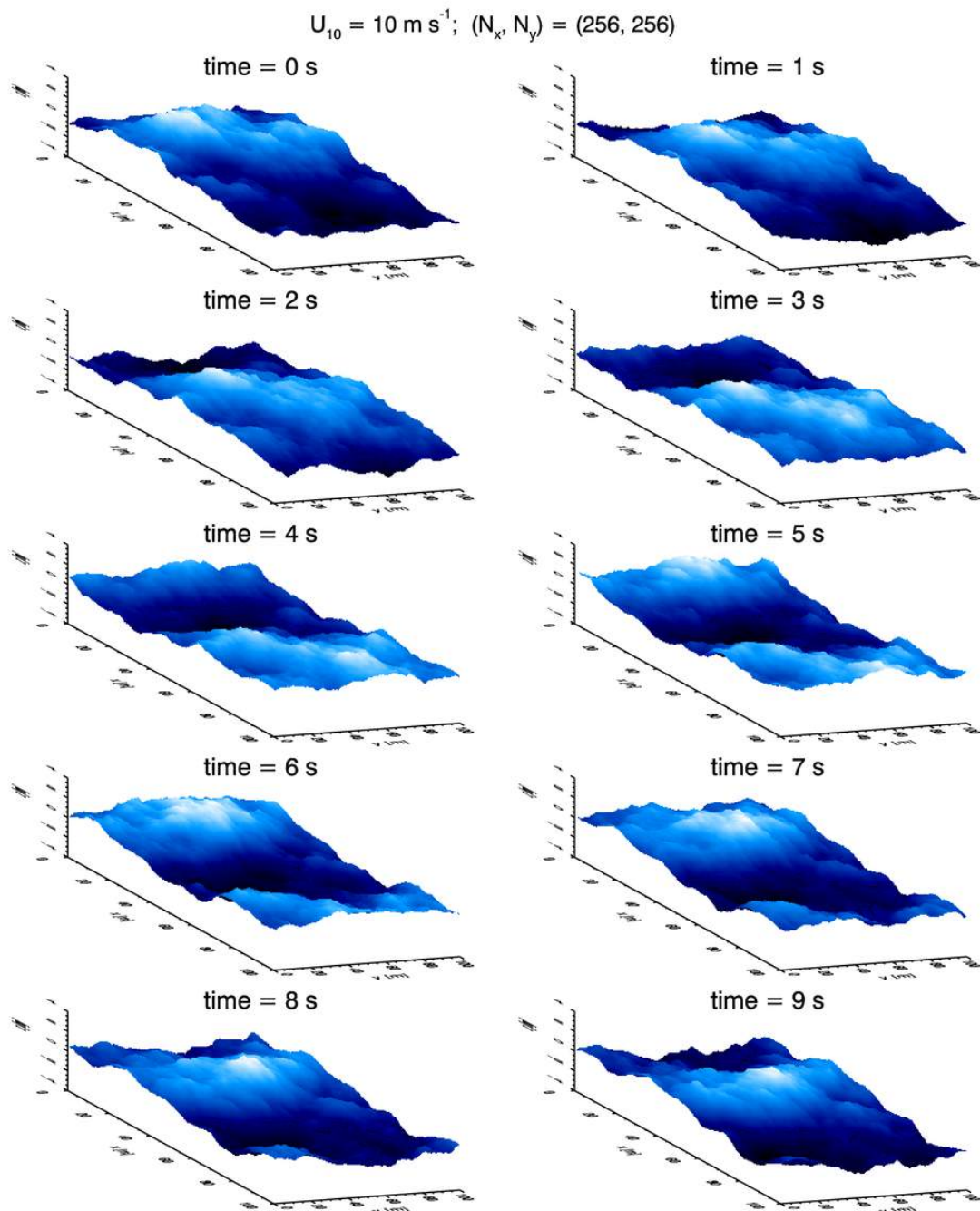


Figure C.17: A time-dependent sequence of 2-D sea surfaces for a 10 m s^{-1} wind speed. The physical domain is $100 \times 100 \text{ m}$; the sampling is 256×256 points. The vertical scale of the plots is -3 to $+3 \text{ m}$. The wind is blowing in the $+x$ direction, which is from the upper left to the lower right of each figure.

square slopes are very close to what is expected by Cox-Munk statistics: 0.031 and 0.019 respectively.

Modeling Sea Surfaces: Autocovariance Techniques

The preceding appendices described the statistical properties of random sea surfaces using variance spectra. This appendix begins the exploration of an alternate path to the specification of surface roughness properties. Clearly, wind-dependent variance spectra are applicable only to surface waves that are generated by wind. Consider, however, the surface of a flowing river. The river's surface can have ripples or waves generated by turbulence resulting from unstable shear flow induced by flow over a shallow bottom, or by eddies created as the water flows around rocks in the river. These water surfaces do not depend on the wind speed and can have different statistical properties, hence different optical properties, than wind-roughened water surfaces. Such surfaces can be described by their autocovariance functions.

Autocovariance functions can be converted to elevation variance spectra via the Wiener-Khinchin theorem as described below. Once a given autocovariance function has been converted to its equivalent elevation variance spectrum, the algorithms of Appendix C are immediately applicable, even though the variance spectrum is not wind-dependent. Indeed, this conversion enables the Fourier transform methods of the previous appendices to be used to generate random realizations of any surface, not just water surfaces.

As is often the case, there is a large gap between textbook theory—usually developed for continuous variables or an infinite sample size of discrete values—and its implementation in a computer program for a finite sample size of discrete variables. In particular, careful attention must be paid to the sampling of an autocovariance function in order to obtain the corresponding variance spectrum, or vice versa. I find it disappointing and frustrating (but not surprising) that numerical matters such as the effects of finite sample size, maximum lag size, and exactly how to sample spectra or autocovariances (in particular, the material in Section D.4) never seem to be mentioned in textbooks on digital signal processing or related subjects. It is left to the innocent student to spend a few weeks of unfunded time figuring out why various numerical results are not internally consistent or do not perfectly match the textbook theory.

This appendix begins with a discussion of autocovariances, and then the Wiener-Khinchin theorem is stated. The theorem is numerically illustrated on first using the

wind-dependent Pierson-Moskowitz elevation variance spectrum, for which certain values can be analytically calculated and used to check the numerical results. The modeling of water surfaces generated by shear-induced turbulence is then illustrated, again using analytical functions that allow for a rigorous check on the numerical results.

D.1 Autocovariance

The autocovariance $C_{zz}(\ell)$ of $z(r)$ is defined as

$$C_{zz}(\ell) \triangleq \mathcal{E}\{z(r)z(r+\ell)\}, \quad (\text{D.1})$$

where \mathcal{E} denotes statistical expectation and ℓ is the spatial lag. This definition in terms of the expectation holds for both continuous and discrete variables. For the present discussion with $z(r)$ being sea surface elevation, $C_{zz}(\ell)$ shows how strongly the sea surface elevation at one location is correlated to the elevation at a distance ℓ away. $C_{zz}(\ell)$ has units of m^2 , and $C_{zz}(\ell=0)$ is the variance of the surface elevation. The autocovariance is an even function of the lag: $C_{zz}(-\ell) = C_{zz}(\ell)$.

Consider an infinite sample of discrete zero-mean surface elevations spaced a distance Δx apart. The autocovariance is then computed by (e.g., Proakis and Manolakis, 1996, Eq. 2.6.3, with a minor change in their notation)

$$C_{zz}(\ell) = \sum_{r=-\infty}^{+\infty} z(r)z(r+\ell), \quad \text{for } \ell = 0, \pm 1, \pm 2, \dots$$

Here ℓ is indexing the lag distance in units of the sample spacing Δx . For a finite sample of N discrete values, the same authors define the sample autocovariance by (their Eq. 2.6.11)

$$C_{zz}(\ell) = \sum_{r=0}^{N-|\ell|-1} z(r)z(r+\ell). \quad (\text{D.2})$$

As usual, there are competing definitions. For a finite sample of N discrete values, perhaps with a non-zero mean m , the IDL autocorrelation function (A_CORRELATE) uses

$$C_{zz}(\ell_r) = \frac{1}{N} \sum_{r=0}^{N-|\ell|-1} [z(r) - m][z(r+\ell) - m] \quad \text{for } -(N-2) \leq \ell_r \leq N-2, \quad (\text{D.3})$$

where

$$m = \frac{1}{N} \sum_{r=0}^N z(r)$$

is the sample mean. MATLAB computes the autocovariance via

$$C_{zz}(\ell) = \frac{1}{N-1} \sum_{r=0}^{N-|\ell|-1} [z(r) - m][z(r+\ell) - m].$$

Note that the lag must be less than the length of the sample. (That is, the sample locations are at $x_r = r\Delta x$, $r = 0, \dots, N-1$ and the allowed lag distances are $\ell_r = r\Delta x$, $r = 0, \dots, N-$

2.) Note also the factor of $1/N$ in the IDL definition (D.3), which does not appear in Eq. (D.2), and which is a factor of $1/(N - 1)$ in the MATLAB version.

Nor is there even any consensus on the terms “autocovariance” and “autocorrelation.” Some authors (and this appendix) define the nondimensional autocorrelation $\rho_{zz}(\ell)$ as the autocovariance normalized by the variance, i.e.

$$\rho_{zz}(\ell) \triangleq \frac{C_{zz}(\ell)}{C_{zz}(0)}. \quad (\text{D.4})$$

However, Proakis and Manolakis (1996) call the autocovariance as used here the autocorrelation, and they call the autocorrelation of Eq. (D.4) the “normalized autocorrelation.” These sorts differences in the definitions and computations of autocovariances can cause much grief when comparing the numerical outputs of different computer codes, or numerical outputs with textbook examples.

D.2 The Wiener-Khinchin Theorem

Now that the autocovariance has been defined, the Wiener-Khinchin theorem can be stated: ***The Fourier transform of the autocovariance equals the variance spectral density function.*** Symbolically,

$$\mathcal{F}_\nu\{C_{zz}(\ell)\} = \mathcal{S}_{2s}(\nu). \quad (\text{D.5})$$

Indeed, some texts *define* the spectral density as the Fourier transform of the autocovariance. The inverse is of course

$$\mathcal{F}_\nu^{-1}\{\mathcal{S}_{2s}(\nu)\} = C_{zz}(\ell). \quad (\text{D.6})$$

Here \mathcal{S}_{2s} is a two-sided spectral density function as discussed in Section C.1.

It is important to note (as emphasized by the ν subscript on the Fourier transform operator \mathcal{F}) that the theorem is written with the ν version of the Fourier transform (Eq. (A.1) of the Fourier Transforms appendix), and the density function is written in terms of the spatial frequency ν , which has units of 1/meters. (In the time domain, the conjugate variables are time t in seconds and frequency f in 1/seconds = cycles/second = hertz.) The spectral density function for the angular spatial frequency $k = 2\pi\nu$ (or angular temporal frequency $\omega = 2\pi f$ in radians per second in the time domain) can be obtained by noting that corresponding intervals of the spectral densities contain the same amount of variance:

$$\mathcal{S}_{2s}(k)dk = \mathcal{S}_{2s}(\nu)d\nu,$$

which gives

$$\mathcal{S}_{2s}(k) = \frac{1}{2\pi}\mathcal{S}_{2s}(\nu = k/2\pi). \quad (\text{D.7})$$

Note that ℓ varies from $-\infty$ to $+\infty$ and, likewise, ν or k ranges over all negative and positive values. The variance spectrum obtained from the Fourier transform of an autocovariance function is therefore a two-sided spectrum.

Comment: In light of Eq. (D.7), the theorem stated in terms of angular spatial frequency k appears to be

$$\mathcal{F}_k\{C_{zz}(\ell)\} = 2\pi\mathcal{S}_{2s}(k), \quad (\text{D.8})$$

with the inverse

$$\mathcal{F}_k^{-1}\{\mathcal{S}_{2s}(k)\} = \frac{1}{2\pi}C_{zz}(\ell). \quad (\text{D.9})$$

I say “appears to be” because I’ve never actually seen the theorem written this way because the textbooks all seem to stick with x and ν (or t and f in the time domain). As [Press et al. \(1992\)](#) say in *Numerical Recipes* (p. 491), “There are fewer factors of 2π to remember if you use the (ν or) f convention, especially when we get to discretely sampled data...” In any case, Eqs. (D.8) and (D.9) are consistent with the k spectrum of Eq. (D.17) discussed below.

The theorem is usually proved in textbooks for continuous variables x and ν . However, in numerical application to a finite number of discrete samples, discrete variables x_r and ν_u or k_u are used, and proper attention must be paid to pesky factors of N , 2π , and bandwidth, and to the array ordering required by a particular FFT routine.

The continuous-variable Fourier transform in Eq. (D.5) gives a spectral density $\mathcal{S}_{2s}(\nu)$ with units of $\text{m}^2/(1/\text{m})$. However, if the theorem is written for a DFT of discrete data $C_{zz}(\ell_r) = C_{zz}(r)$,

$$\mathcal{D}_\nu\{C_{zz}(r)\} = \mathcal{S}_{2s}(\nu_u), \quad (\text{D.10})$$

then the resulting discrete spectrum $\mathcal{S}_{2s}(\nu_u) = \mathcal{S}_{2s}(u)$ has units of m^2 . Just as was discussed on the Fourier Transforms appendix, this discrete spectrum must be divided by the bandwidth $\Delta\nu$ in order to obtain the density at $\nu = \nu_u$. That is,

$$\mathcal{S}_{2s}(\nu = \nu_u) = \mathcal{S}_{2s}(u)/\Delta\nu. \quad (\text{D.11})$$

D.2.1 Example

[Horoshenkov et al. \(2013, Eq. 4\)](#) give an analytic formula for the autocorrelation function of surface waves generated by bottom-induced turbulence in shallow flowing water. In the notation of this section, this function is

$$\rho_{zz}(\ell) = \exp\left(-\frac{\ell^2}{2\sigma_w^2}\right) \cos\left(\frac{2\pi}{L_o}\ell\right). \quad (\text{D.12})$$

In their words, “ σ_w relates to the spatial radius of correlation (correlation length)” and “ L_o relates to the characteristic period in the surface wave pattern.” The average values for the physical conditions of the [Horoshenkov et al.](#) study are $\sigma_w = 0.22$ m and $L_o = 0.17$ m. [Note: Eq. (D.12) is [Horoshenkov’s](#) $W(\rho)$ as shown in their abstract and in their conclusions, where it has a factor of $1/2$ in the exponential. Their Eq (4) does not have the $1/2$. This is probably a typo since Gaussians usually have the form $\exp[-x^2/(2\sigma^2)]$.]

This autocorrelation function provides a nice example of how to use the Wiener-Khinchin theorem to obtain the corresponding variance spectrum. Equation (D.12), when converted to an autocovariance via a factor of the variance, $\langle z^2 \rangle = C_{zz}(0)$, (as seen in Eq. D.4) has the form

$$\begin{aligned} C_{zz}(\ell) &= C_{zz}(0) \exp\left(-\frac{\ell^2}{2\sigma_w^2}\right) \cos\left(\frac{2\pi}{L_o}\ell\right) \\ &= C_{zz}(0) \exp(-a^2\ell^2) \cos(q_o\ell), \end{aligned} \quad (\text{D.13})$$

where $a = 1/(\sqrt{2}\sigma_w)$ and $q_o = 2\pi/L_o$. This function has an easily computed analytical Fourier transform.

The continuous-variable Fourier transform of this $C_{zz}(\ell)$ is given by Eq. (A.1):

$$\mathcal{S}_{2s}(\nu) = \mathcal{F}_\nu\{C_{zz}(\ell)\} = \int_{-\infty}^{\infty} C_{zz}(\ell) e^{-i2\pi\nu\ell} d\ell. \quad (\text{D.14})$$

Here ℓ and ν are continuous variables; $\mathcal{S}_{2s}(\nu)$ has units of m^3 , which is interpreted as $\text{m}^2/(1/\text{m})$ as explained before. Note that this variance spectral density is two-sided, i.e., $-\infty < \nu < \infty$. Expanding the complex exponential via $e^{-i\theta} = \cos\theta - i\sin\theta$ gives

$$\begin{aligned} \mathcal{S}_{2s}(\nu) &= \int_{-\infty}^{\infty} C_{zz}(0) \exp(-a^2\ell^2) \cos(q_0\ell) \cos(2\pi\nu\ell) d\ell \\ &\quad - i \int_{-\infty}^{\infty} C_{zz}(0) \exp(-a^2\ell^2) \cos(q_0\ell) \sin(2\pi\nu\ell) d\ell. \end{aligned}$$

The imaginary term is zero because the integrand is an odd function of ℓ . Using the identity

$$\cos x \cos y = \frac{1}{2} [\cos(x+y) + \cos(x-y)]$$

gives

$$\mathcal{S}_{2s}(\nu) = 2C_{zz}(0) \int_0^{\infty} \exp(-a^2\ell^2) \frac{1}{2} \{\cos[(2\pi\nu + q_0)\ell] + \cos[(2\pi\nu - q_0)\ell]\} d\ell.$$

The integral

$$\int_0^{\infty} \exp(-a^2\ell^2) \cos(b\ell) d\ell = \frac{\sqrt{\pi} \exp[-b^2/(4a^2)]}{2a}$$

gives the Fourier transform of the $C_{zz}(\ell)$ of Eq. (D.13):

$$\mathcal{S}_{2s}(\nu) = \sqrt{\frac{\pi}{2}} \sigma_w C_{zz}(0) \left\{ \exp\left[-\frac{1}{2}(2\pi\sigma_w)^2(\nu + 1/L_o)^2\right] + \exp\left[-\frac{1}{2}(2\pi\sigma_w)^2(\nu - 1/L_o)^2\right] \right\}. \quad (\text{D.15})$$

This variance spectrum has maxima at $\nu = \pm 1/L_o$, where the value is very close to $\sqrt{\frac{\pi}{2}} \sigma_w C_{zz}(0)$. Figure D.1 plots this $C_{zz}(\ell)$ (Eq. D.13) and $\mathcal{S}_{2s}(\nu)$ (Eq. D.15) for typical values of $\sigma_w = 0.22$ m, $L_o = 0.17$ m, and $C_{zz}(0) = 2.5 \cdot 10^{-7}$ m^2 . Note that the subpeaks of the autocovariance lie at integer multiples of $\pm L_o$, and that the peaks of the spectrum are at $\pm 1/L_o$.

By definition, the integral over all frequencies of an elevation variance spectral density gives the total elevation variance $\langle z^2 \rangle$:

$$\langle z^2 \rangle = \int_{-\infty}^{\infty} \mathcal{S}_{2s}(\nu) d\nu.$$

This can be computed analytically for the spectrum of Eq. (D.15). The $\mathcal{S}_{2s}(\nu)$ spectrum of Eq. (D.15) is the sum of two identical Gaussians centered at different ν values. Consider the one centered at $\nu = 1/L_o$, which involves the exponential with the $\nu - 1/L_o$ term. The total variance is then twice the integral of this Gaussian:

$$\langle z^2 \rangle = 2\sqrt{\frac{\pi}{2}} \sigma_w C_{zz}(0) \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(2\pi\sigma_w)^2(\nu - 1/L_o)^2\right] d\nu.$$

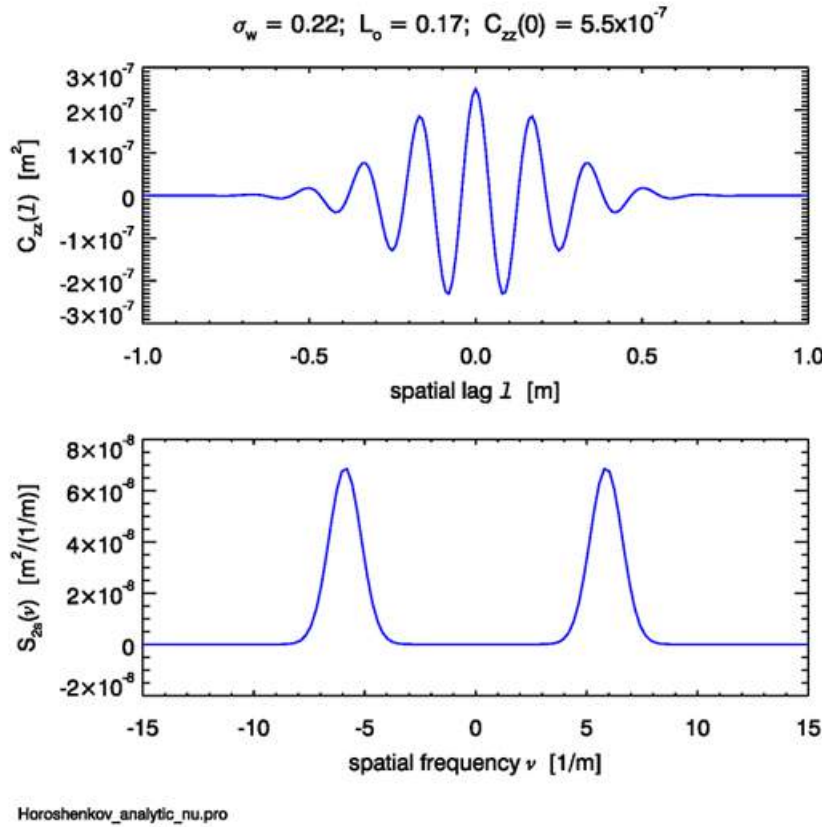


Figure D.1: The Horoshenkov autocovariance $C_{zz}(\ell)$ and elevation variance spectral density $S_{2s}(\nu)$ for typical parameter values taken from their Table 3.

Letting $x = \nu - 1/L_o$ gives

$$\langle z^2 \rangle = 2\sqrt{\frac{\pi}{2}}\sigma_w C_{zz}(0) \int_{-\infty}^{\infty} \exp[-c^2 x^2] dx,$$

where $c^2 = \frac{1}{2}(2\pi\sigma_w)^2$. The integral

$$\int_0^{\infty} \exp[-c^2 x^2] dx = \frac{\sqrt{\pi}}{2c}$$

then gives the final result:

$$\langle z^2 \rangle = 4\sqrt{\frac{\pi}{2}}\sigma_w C_{zz}(0) \frac{\sqrt{\pi}}{2\frac{1}{\sqrt{2}}2\pi\sigma_w} = C_{zz}(0). \quad (\text{D.16})$$

Thus starting with a variance of $C_{zz}(0)$ in the autocovariance of Eq. (D.13), obtaining the variance spectral density from the Fourier ν -transform of the autocovariance, and then integrating the variance spectral density over ν thus gives back the variance as originally specified in the autocovariance function.

However, if the above process is naively carried through starting (as in Eq. D.14) with the k -transform of Eq. (A.3), the end result (as in Eq. D.16) is $2\pi C_{zz}(0)$. This extra

factor of 2π is rectified by the $1/2\pi$ factor seen in Eq. (D.7). Thus the k version of the Horeshenkov spectral density is

$$\mathcal{S}_{2s}(k) = \frac{1}{2\pi} \sqrt{\frac{\pi}{2}} \sigma_w C_{zz}(0) \left\{ \exp \left[-\frac{1}{2} \sigma_w^2 (k + q_o)^2 \right] + \exp \left[-\frac{1}{2} \sigma_w^2 (k - q_o)^2 \right] \right\}. \quad (\text{D.17})$$

Integration of this $\mathcal{S}_{2s}(k)$ over all k then results in $\langle z^2 \rangle = C_{zz}(0)$, as required. The inverse k transform of the spectral density (D.17) gives $C_{zz}(\ell)/(2\pi)$, as expected from Eq. (D.9).

D.3 Numerical Example of the Wiener-Khinchin Theorem

This section gives a numerical example of the Wiener-Khinchin theorem, which leads into the details of how to sample autocovariance functions so that the resulting variance spectra meet the needs for surface generation.

The blue curve of Panel (a) in Fig. D.2 plots the one-sided Pierson-Moskowitz spectrum of Eq. (B.25) for wind speed of $U_{10} = 5 \text{ m s}^{-1}$. Using this spectrum, surfaces are generated at $N = 1024$ points over a region of length $L = 100 \text{ m}$. Note that N is a power of 2 as required for the use of the fast Fourier transform algorithm. The spacing of these points is at intervals of $\Delta x = L/N = 0.0944 \text{ m}$. The red dot at $\nu_1 = 1/L = 0.01 \text{ 1/m}$ is the fundamental frequency. The point at $\nu_{N/2} = 1/(2\Delta x) = 5.12 \text{ 1/m}$ is the Nyquist spatial frequency. The green vertical ticks show the locations of the remaining $N - 2$ points, which are evenly spaced at intervals of $\Delta\nu = 0.01 \text{ 1/m}$.

These discrete samples of the variance spectrum are then used as described in Section to create a random realization of the sea surface $z(r)$ at N points. One such surface, generated for a particular sequence of random variables, is shown in Fig. D.2(c). The periodogram of this surface, computed via Eqs. (C.2) and (C.3) of Section C.1 and Eq. (C.7), is shown in red in Fig. D.2(b). The blue curve in this panel is the one-sided spectrum $\mathcal{S}_{1s}(\nu)$ of Panel (a), replotted with linear axes for reference. The statistical noise in the periodogram is Gaussian distributed about the theoretical $\mathcal{S}_{1s}(\nu)$. These three panels of the figure are the essentially the same as Fig. C.4; the only difference is that the sequence of random numbers used to generate the surface is different and linear axes are used for the upper-right panel.

Equation (D.3) applied to the $z(r)$ of Panel (c) gives the autocovariance shown in red in Panel (d). This curve contains statistical noise. To obtain a theoretical curve for comparison, the Pierson-Moskowitz spectrum was sampled at 2048 points to insure coverage of most of the spectrum. The discrete Wiener-Khinchin theorem of Eq. (D.10) was then used to obtain the autocovariance from the discretely sampled spectrum:

$$C_{zz}(\ell_r) = \mathcal{D}_\nu^{-1} \{ \mathcal{S}_{2s}(u) \} = \mathcal{D}_\nu^{-1} \{ \mathcal{S}_{2s}(\nu = \nu_u) \Delta\nu \}. \quad (\text{D.18})$$

Here the DFT \mathcal{D}_ν^{-1} was computed via Eq. (A.8). Note that the discrete spectrum $\mathcal{S}_{2s}(u)$ was obtained by sampling the continuous spectral density at the desired ν values and then multiplying by the appropriate bandwidth. Because $\mathcal{S}_{2s}(u)$ is a real and even function of ν_u , its Fourier transform is also real and even. The result is shown as the green curve in this Panel (d). Equation (C.8) gives the total variance of $\langle z^2 \rangle = 0.0197 \text{ m}^2$ for $U_{10} = 5 \text{ m s}^{-1}$. The numerical result obtained by sampling $\mathcal{S}_{2s}(\nu)$ and taking the inverse Fourier transform as just described gives $C_{zz}(0) = \langle z^2 \rangle = 0.0178 \text{ m}^2$.

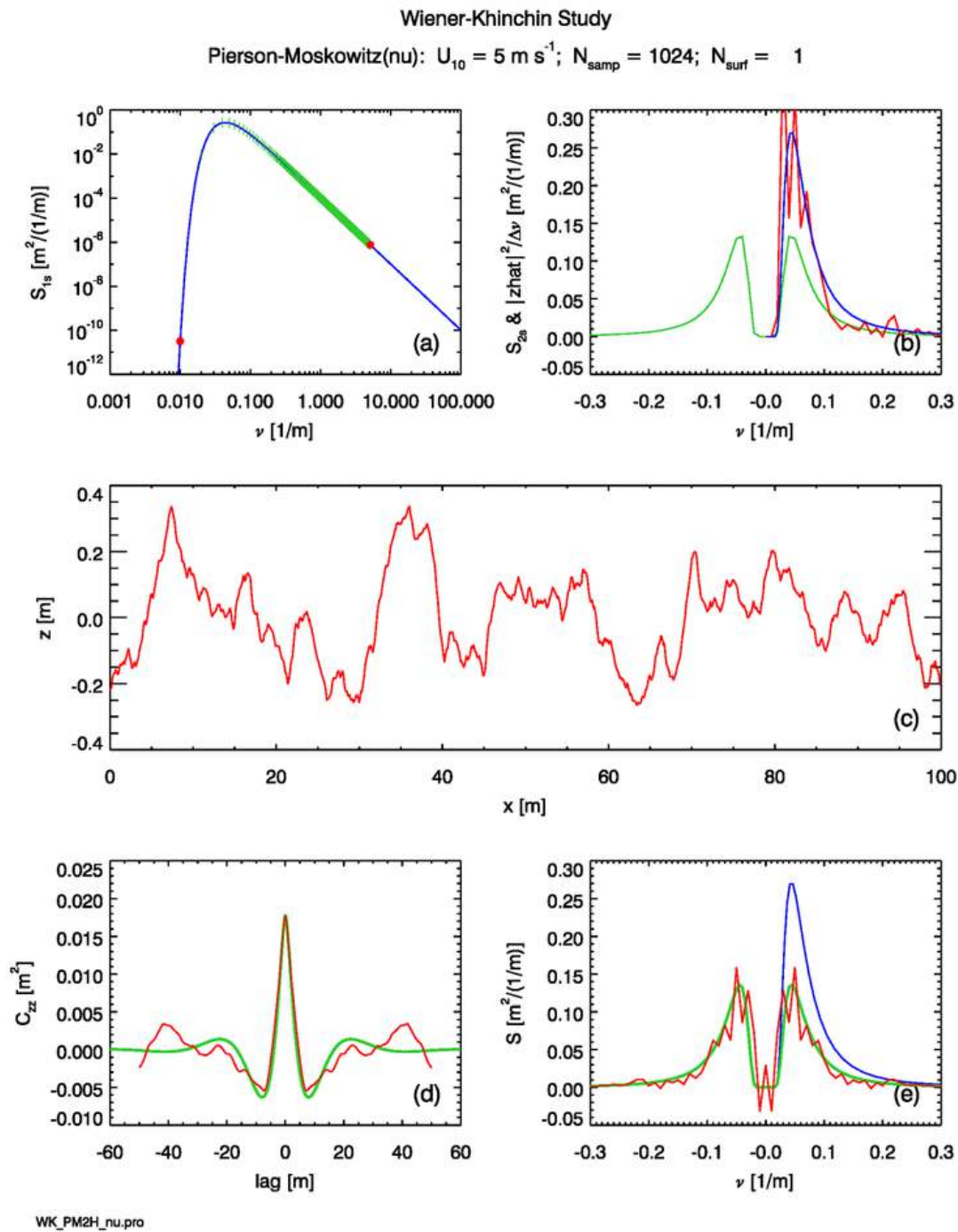


Figure D.2: Illustration of the Wiener-Khinchin theorem for a single realization of a random sea surface.

Latta and Bailie (1968) analytically computed the autocorrelation of the Pierson-Moskowitz spectrum in temporal form, but the result is a formula of horrible complexity consisting of the sum of five slowly converging infinite series, the terms of which are them-

selves are products of infinite series. That paper plots the numerically evaluated result in terms of an unspecified but normalized temporal lag, which makes comparison with the present results for spatial lag quantitatively impossible. However, [de Boer \(1969\)](#) obtained the spatial covariance of the Pierson-Moskowitz spectrum in the form of integrals of Bessel functions, which also require careful numerical evaluation. Figure D.3 shows their result for the autocovariance function of waves in the down-wind direction. Their plot is in terms of a nondimensional normalized lag distance $\xi_N = (2g/U^2)\ell$. The green curve of Fig. D.2(d) has a minimum of -0.0063 m^2 at $\ell = \pm 7.96 \text{ m}$. This translates to an autocorrelation of -0.31 at a normalized lag of $\xi_N = 6.3$. These values are in reasonable agreement with the minimum seen in Fig. D.3, keeping in mind that the curves in that figure were themselves generated on a 1960's era computer by difficult numerical integrations of unknown accuracy. The agreements for the variance and the location and magnitude of the minimum indicate that the numerically computed $C_{zz}^{PM}(\ell)$ is probably correct for all lags. (This numerical calculation will be verified again with greater accuracy in the discussion of the Horoshenkov spectra below, for which the exact autocovariance is known.)

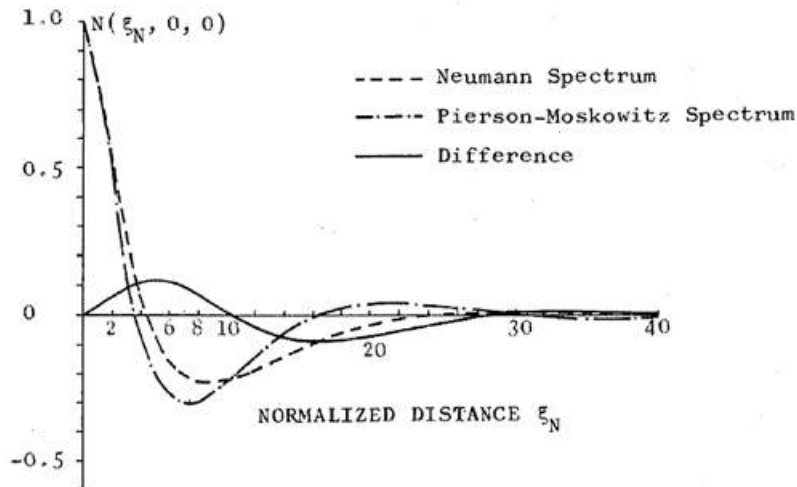


Figure D.3: Fig. 7 from [de Boer \(1969\)](#). The normalized lag distance is $\xi_N = (2g/U^2)\ell$.

Taking the DFT of the green curve in Fig. D.2(d) should give the two-sided spectrum $\mathcal{S}_{2s}(\nu)$ corresponding to the one-sided spectrum plotted in Panel (a). The green curve in Panel (e) of that figure shows the result (after dividing by the finite bandwidth, as mentioned previously), which is indeed one-half of the one-sided spectrum $\mathcal{S}_{1s}(\nu)$ (shown in blue). This provides a check on the correctness of a round-trip Fourier transform.

Taking the DFT of the red curve in Panel (d) gives a sample estimate of $\mathcal{S}_{2s}(\nu)$, which is shown in red in Panel (e). This curve has statistical noise, but it visually appears to be distributed about the theoretical value given by the green curve.

The statistical noise inherent in any single random realization of the sea surface and its autocovariance can be reduced by averaging the results from many surface realizations. Figure D.4 is the same as Fig. D.2, except that $N_{\text{surf}} = 100$ independent surfaces are generated. This reduces the statistical noise by a factor of $1/\sqrt{100}$. The red curve in Panel (b) shows the ensemble average periodogram for the 100 surfaces. It is clear that the average periodogram is in excellent agreement with the theoretical variance spectrum,

except for a small amount of remaining statistical noise.

The red curve of Panel (d) is the average autocovariance for the 100 surfaces. This curve is much closer to the theoretical (green) curve than the autocovariance for the single surface of Fig. D.2(d). The DFT of this average autocovariance is shown by the red curve in Panel (e). Again, this curve has much less noise and is closer to the (green) theoretical spectrum.

The statistical noise in the ensemble averages can be made arbitrarily small by averaging more and more surfaces. Figure D.5 shows that averages for 1,000 surfaces have noise levels in the periodogram, autocovariance, and spectrum derived from the autocovariance, that are almost unnoticeable at the scale of the figures.

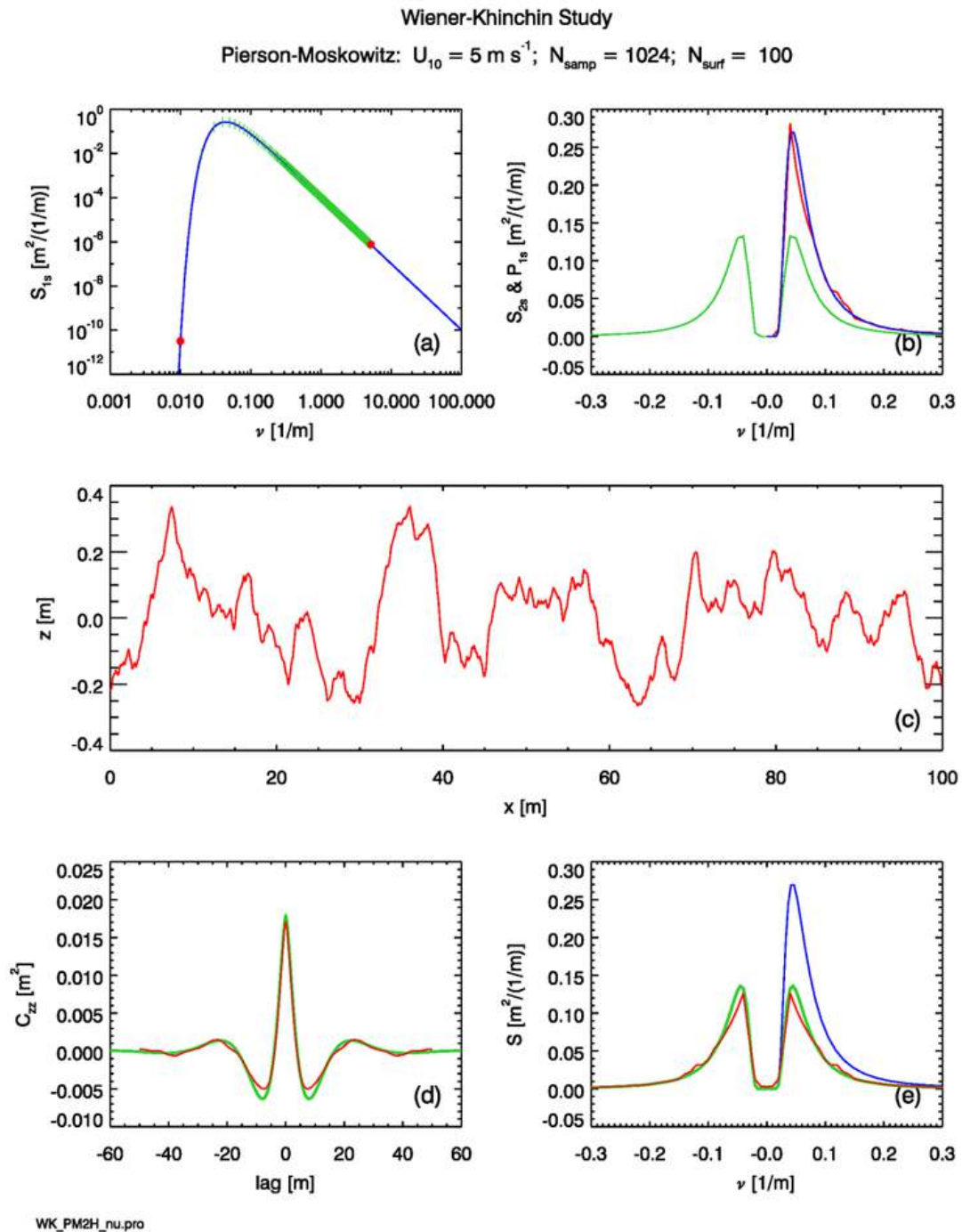


Figure D.4: Same as Fig. D.2, but for 100 surface realizations. The red curve in Panel (d) is the average of the autocovariances of the 100 surfaces. The red curve in Panel (e) is the Fourier transform of the 100-surface average of Panel (d). Only the first surface is plotted in Panel (c).

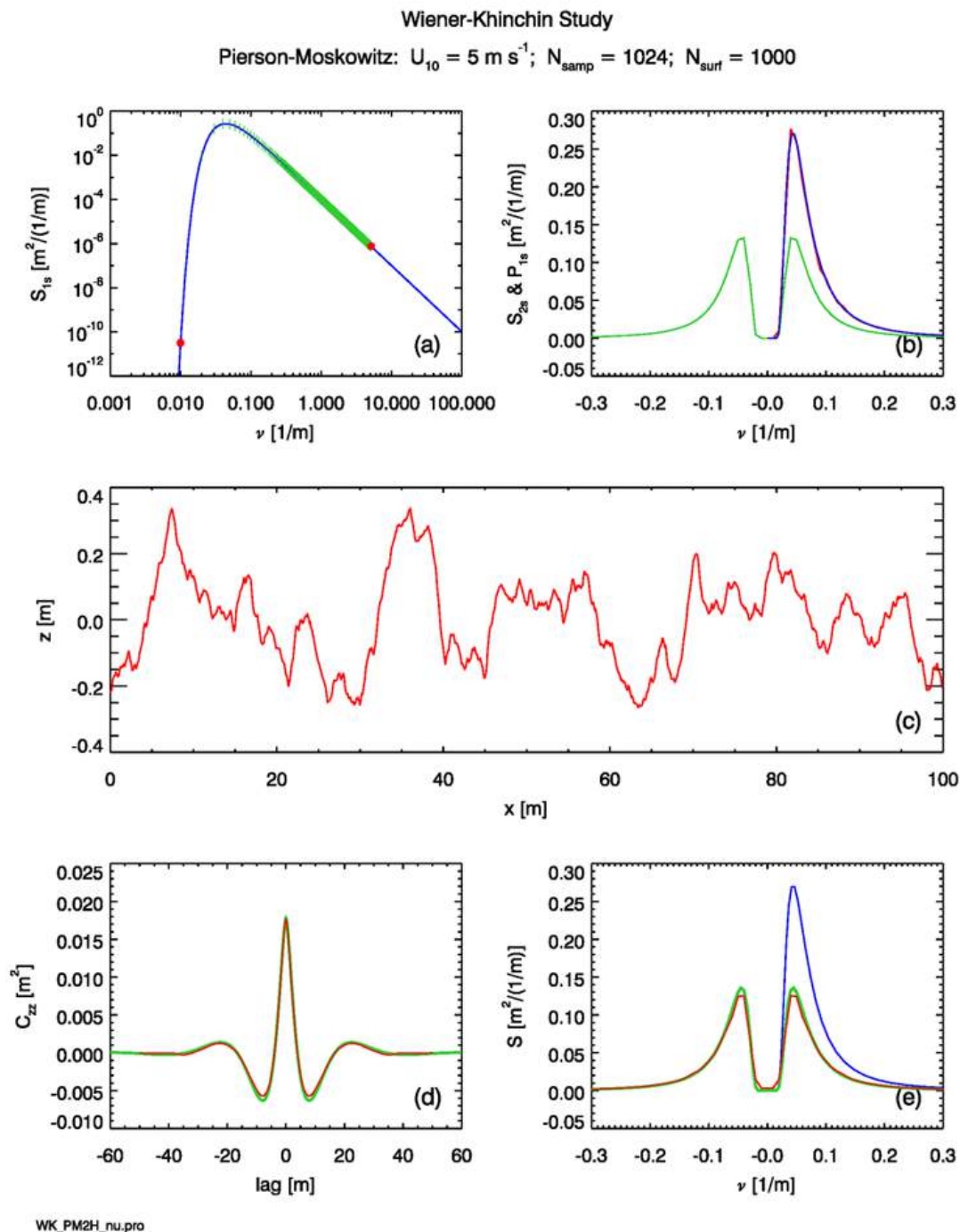


Figure D.5: Same as Fig. D.2, but for 1000 surface realizations. The red curve in Panel (d) is the average of the autocovariances of the 1000 surfaces. The red curve in Panel (e) is the Fourier transform of the 1000-surface average of Panel (d). Only the first surface is plotted in Panel (c).

D.4 Sampling Strategy and Computational Details

This section shows exactly how the calculations underlying Figs. D.2, D.4, and D.5 are performed. The devil is in the details, and these details are seldom if ever discussed in the literature. Consider the case of $N = 8$, which will allow individual points to be plotted. Of course, with so few sample points, the variance spectrum is not adequately sampled and the resulting sea surface is unphysical because it has far too little variance. However, the algorithms are the same for any value of N .

Consider first the generation of the random sea surface with N points. As discussed previously, the two-sided spectrum must be sampled at exactly N spatial frequencies, which are given by Eq. (A.15). The green dots in Fig. D.6(b) show these points for the case of $N = 8$. The frequency values, written in math order as in Eq. (A.14), are

$$\{\nu_u, u = 0, 1, \dots, N - 1\} = \left[-\frac{N}{2} + 1, \dots, -1, 0, 1, \dots, \frac{N}{2} \right] \Delta\nu, \quad (\text{D.19})$$

which for the choice of $L = 100$ m and $N = 8$ gives

$$\{\nu_u, u = 0, 1, \dots, 7\} = [-3, -2, -1, 0, 1, 2, 3, 4] \Delta\nu, \quad (\text{D.20})$$

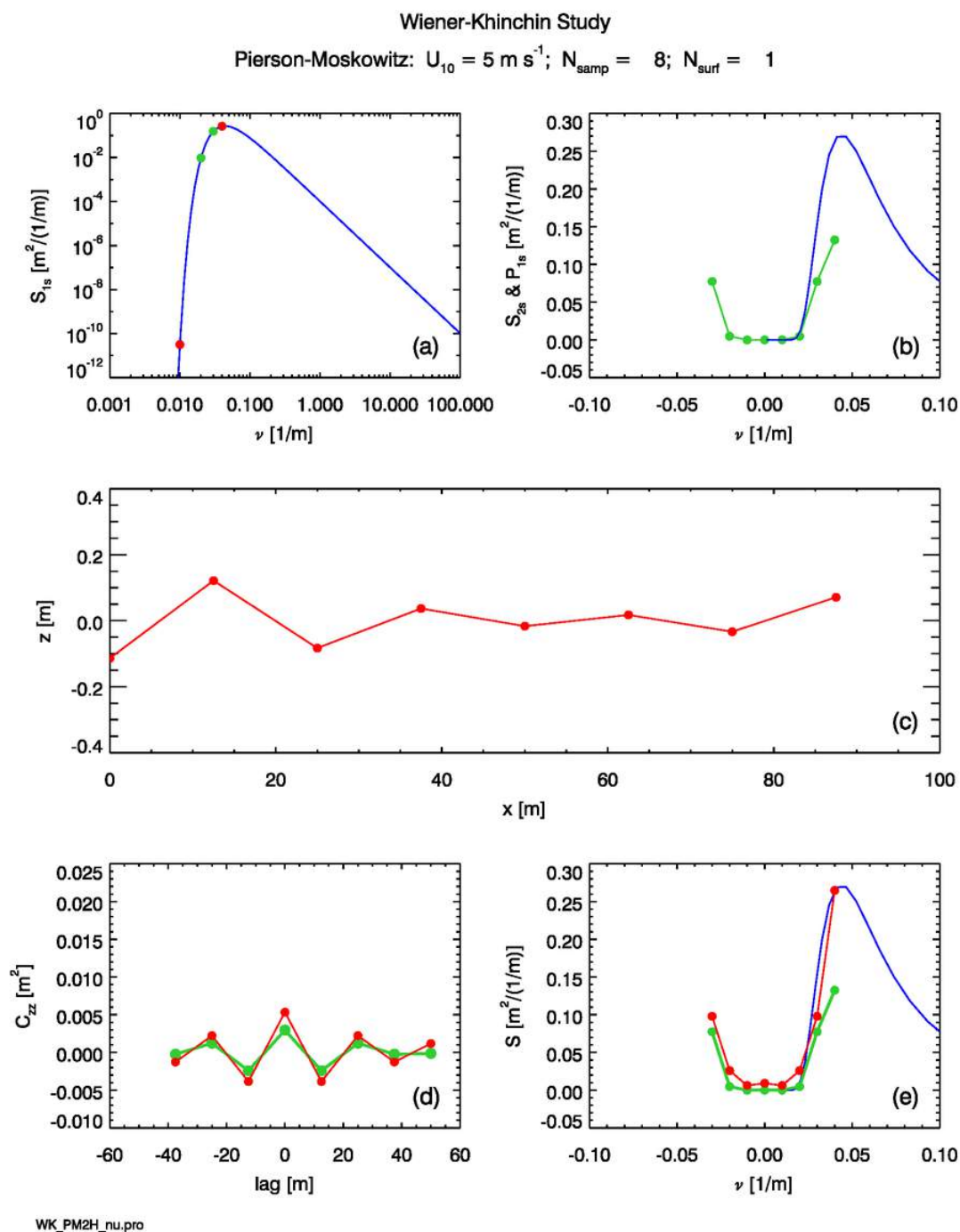
where $\Delta\nu = 1/L = 0.01 \text{ m}^{-1}$. Here braces $\{\dots\}$ denote a set of frequencies labeled by u values, and brackets $[\dots]$ denote an array of frequency values as shown. Note that the sampled frequencies are symmetric about $\nu = 0$, except for one “extra” point at index $u = N - 1$ or frequency $(N/2)\Delta\nu$. This value is the Nyquist frequency (which in IDL is stored as the last element of the frequency array in math order). Sampling the spectrum at *exactly* this pattern of frequencies guarantees that the spectral amplitudes generated from them are Hermitian, which in turn guarantees that the generated sea surface is real. The red dots in Panel (c) of the figure show the 8 surface elevations generated for a particular sequence of random numbers. The values are at $x_r = 0$ to $L - \Delta x$ for $r = 0$ to $N - 1$. Fourier-generated surfaces are inherently periodic, so that $z(L) = z(0)$.

Now take the inverse DFT as in Eq. (D.18) for the discrete spectrum given by the green dots in Panel (b). The result is the autocovariance values shown by the green dots in Panel (d). It is important to note that these $N = 8$ lag values follow the same pattern (in math order) as the frequencies:

$$\{\ell_r, r = 0, 1, \dots, 7\} = [-3, -2, -1, 0, 1, 2, 3, 4] \Delta x, \quad (\text{D.21})$$

where now $\Delta x = L/N = 12.5$ m. The lags are symmetric about $\ell = 0$, except for one “extra” point at $(N/2)\Delta x$. This is analogous the one extra value in the frequency spectrum at the Nyquist frequency. Taking the forward DFT of these 8 autocovariance values as in Eqs. (D.10) and (D.11) gives the green points plotted in Panel (e). These values are of course exactly the 8 points of the original spectrum, as shown by the green dots in Panel (b). This is just a check on the correct implementation of the “round trip” calculation of inverse and forward Fourier transforms.

Now suppose that we wish to compute the autocovariance of the surface elevations, and from that obtain an estimate of the variance spectrum via the Wiener-Khinchin theorem. This provides a more stringent test of the calculations because of the intermediate sea surface in between the variance spectrum and the autocovariance. The crucial observation is that when calling the IDL autocovariance routine `A_CORRELATE`, that routine must

Figure D.6: Illustration of sampling strategy for $N = 8$ sample points.

be given an array of the requested lag indices (lags in units of Δx) as seen in Eq. (D.21). Thus for an array of surfaces,

$$\text{zsurf} = [z(r = 0), z(r = 1), \dots, z(r = 7)], \quad (\text{D.22})$$

an array of lags

$$\text{lagindex} = [-3, -2, -1, 0, 1, 2, 3, 4], \quad (\text{D.23})$$

must be defined. The call to the IDL routine is then

$$\text{Czz} = \text{A_CORRLELATE}(\text{zsurf}, \text{lagindex}, /\text{COVARIANCE}). \quad (\text{D.24})$$

The IDL routine then returns an array of autocovariances at the lags shown in Eq. (D.21). These values are shown by the red dots in Fig. D.6(d). This Czz array returned by A_CORRLELATE has the same math order as the lagindex array. This array must next be shifted into the FFT order via the IDL shift function:

$$\text{CzzFFT} = \text{SHIFT}(\text{Czz}, -N/2 + 1). \quad (\text{D.25})$$

This array can now be given to the IDL FFT routine:

$$\text{S2s} = \text{FFT}(\text{CzzFFT}). \quad (\text{D.26})$$

The resulting S2s array is a complex 8-element array. The real part of S2s is $S_{2s}(u)$, with the frequencies in FFT order. The imaginary part is zero (to within a small roundoff error; values are typically less than 10^{-10}). This array is shifted back to math order and divided by $\Delta\nu$ to get the array plotted as the red dots in Panel (e) of the figure:

$$\text{S2splot} = \text{REAL_PART}(\text{SHIFT}(\text{S2s}, N/2 - 1))/\Delta\nu. \quad (\text{D.27})$$

It is always informative to take an “information count” of such operations. We started with a two-sided spectrum of 8 values. It is true that in the present time-independent simulations $S_{2s}(-\nu) = S_{2s}(+\nu)$ (except for the 0 and Nyquist frequencies, which are always special cases). However, this symmetry need not hold in general (and indeed is not the case when generating waves that propagate downwind, as explained previously). Thus these spectrum values represent 8 independent “pieces” of information in the form of 8 real numbers.

The 8 elevations of the sea surface are likewise 8 independent pieces of information.

Finally, the 8 covariances also comprise 8 pieces of information. Similarly to the variance spectrum, there is symmetry about the 0 lag, except for the value at the largest positive lag. However, again, the fact that $C_{zz}(-l_r) = C_{zz}(+l_r)$ represents two pieces of information: the value of $C_{zz}(+l_r)$ and the fact that $C_{zz}(-l_r)$ has the same value.

Thus the sampled variance spectrum $S_{2s}(u)$, the generated surface $z(r)$, and the surface autocovariance $C_{zz}(r)$ all contain the same amount of information, namely 8 real numbers. The various Fourier transforms and autocorrelation function show how to convert the information from one form to another.

D.4.1 Idle Speculations

It is certainly possible to sample in different ways. For example, surface correlations can be computed for all lags from $-L + \Delta x$ to $+L - \Delta x$, which gives $2N - 1$ total $C_{zz}(l_r)$ values. You can then take the FFT of that covariance and get a spectrum with $2N - 1$ values. However, I can guarantee you from two weeks of misery that the spectrum so obtained does not agree with the original $S_{2s}(\nu)$ spectrum. The $N - 1$ extra points added

by taking a greater range of correlations are in some way not independent of or consistent with the N independent pieces of information tallied above. That is to say, the sea surface contains only N pieces of information, and you cannot create more information simply by computing the autocovariance at more lag values. I vaguely remember reading somewhere that you should not compute autocovariances for lags greater than one-half of the data range. Note that the lag indices used above run from values of $(-N/2+1)\Delta x$ to $(N/2)\Delta x$, which indeed correspond to the $-L/2 + \Delta x$ to $+L/2$ data range. I suspect, but have never seen stated, that there is something going on here that is analogous to sampling at greater than the Nyquist frequency—You can do it, but it messes up the results in ways that are not immediately obvious.

Another possible way to compute the autocovariance for a given sea surface is to compute $C_{zz}(\ell_r)$ only for 0 and positive lags out to a maximum possible lag of $L - \Delta x$. This would again give N independent numbers. Autocovariances are real and even functions of the lag (symmetric about $\ell = 0$), which means that their Fourier transforms are also real and even. Since $e^{i\theta} = \cos\theta + i\sin\theta$, a Fourier transform can be written as the sum of a cosine transform plus i times a sine transform: $\mathcal{F}\{\cdot\} = \mathfrak{C}\{\cdot\} + i\mathfrak{S}\{\cdot\}$. Here the cosine transform \mathfrak{C} is defined as in Eq. (A.1) except that $e^{-i2\pi\nu x}$ is replaced by $\cos(2\pi\nu x)$; the sine transform \mathfrak{S} is defined in the same way but with $\sin(2\pi\nu x)$ replacing $e^{-i2\pi\nu x}$. For an even function, the sine components in the Fourier transform will all be zero. Thus it seems that Eq. (D.5) could be written as $\mathfrak{C}\{C_{zz}(\ell)\} = \mathcal{S}_{2s}(\nu)$. An example of this was seen above in the analytical computation of the Horoshenkov variance spectrum. However, there are four different algorithms for implementing the *discrete* cosine transform (DCT), which differ by how the discrete, finite- N sequence of points is assumed to be extended outside the domain for which $C_{zz}(\ell_r)$ is known. It seems that the present case of $C_{zz}(\ell_r)$, which is symmetric about $\ell_r = 0$, corresponds to the “Type I” DCT discussed at https://en.wikipedia.org/wiki/Discrete_cosine_transform or the “ y_1 ” extension seen in Fig. 2(a) of Makhoul (1980). The four different formulations of the DCT can be computed in four different ways by use of FFTs. Thus the use of a DCT in Eq. (D.5) opens a new can of worms. In any case, there is little or no penalty to be paid for sticking with a Fourier transform evaluated by an FFT routine in order to evaluate the DFTs as needed here. As a matter of practical necessity, the internal consistency of the spectra, surfaces, and autocovariances seen in the preceding figures (and to be seen below) indicate that the sampling scheme described above is correct, even if there may be equivalent ones.

D.4.2 Lessons Learned

The preceding simulations illustrated the Wiener-Khinchin theorem starting with a variance spectrum $\mathcal{S}_{1s}(\nu)$ (the Pierson-Moskowitz spectrum) and arriving at an autocovariance $C_{zz}(\ell)$ in two ways. The first way was to construct the corresponding two-sided spectrum $\mathcal{S}_{2s}(\nu)$ and then take the inverse Fourier transform to obtain the theoretical, noise-free $C_{zz}(\ell)$ via the Wiener-Khinchin theorem. The second way was to use $\mathcal{S}_{2s}(\nu)$ to generate a large number of random sea surfaces. The autocovariance of each random surface was computed by Eq. (D.3), and then the ensemble-average autocovariance was computed as the average of the individual autocovariances.

It is important to note that the size L of the spatial region and the number of sample points N must be chosen with care. As a rule, L must be large enough to cover several wavelengths of the longest wave that contains a significant amount of the total variance.

N must be large enough that the sampled points on the variance spectrum then reach far enough into the high-frequency end of the spectrum to cover the entire part of the spectrum that contributes a significant amount to the total variance. To see the effects of inadequate sampling, suppose we are concerned only with the short gravity and capillary waves, which are optically the most important because they have the highest slopes. If we are interested only in waves of wavelength ~ 1 m down to ~ 1 cm, it might then seem reasonable to let $L = 10$ m and $N = 1024$, which give $\Delta x \approx 1$ cm. The shortest resolvable wavelength is then $2\Delta x \approx 2$ cm. Figure D.7 shows an example surface and other quantities for this case.

However, now $\Delta\nu = 1/L = 0.1 \text{ m}^{-1}$ and the spectrum is sampled only at widely spaced points (the green dots in Fig. D.7(b)) that largely miss the peak of the variance spectrum. Consequently, the generated surface has too little variance compared to the real sea surface described by this spectrum. Also, the sample autocovariance function, shown by the red curve in Fig. D.7(d), computes the autocovariances only for lags up to $L/2 = 5$ m. This lag range does not capture the full autocovariance features of the real surface, for which the autocovariance is non-zero out to lags of ~ 40 m, as shown by the green curve in Panel (d). The spectrum estimated from the sample autocovariance (the red curve in Panel (e)) does reproduce the sampled spectrum (the green dots in Panel (b)), but this spectrum is not representative of the real sea surface.

Picking $L = 100$ m, as in the previous simulations, seems adequate for a wind speed of 5 m s^{-1} . This can be seen from the leftmost red point in Panel (a) of the previous plots, which is to the left of the spectrum maximum. However, $N = 1024$ then gives $\Delta\nu = 0.01 \text{ m}^{-1}$, and the last sampled point corresponds to a shortest resolvable wavelength of $2\Delta x \approx 20$ cm. If that is not adequate resolution for the problem at hand, there are two options. One option is to increase N , which costs more computer time to evaluate the FFTs. Increasing N by a factor of 8 to $N = 8192$ then gives $2\Delta x \approx 2.4$ cm, which might be adequate for the problem at hand. The time for an FFT is proportional to $N \log_2 N$, so that increase in N comes at a factor-of-ten increase in run time, which might be prohibitive if many surfaces must be generated. The other option is to account for the unsampled variance in some other way. One technique for doing that is to adjust the spectrum to account for the missing variance while keeping N relatively small. One technique for doing this is described in Section C.6 and in Mobley (2015).

These results can be summarized as follows:

- The size of the spatial domain, L , must be large enough to cover at least several wavelengths of the wave of maximum variance. The value of L sets the fundamental frequency $\nu_1 = 1/L$, which equals the frequency interval $\Delta\nu$.
- For the given fundamental frequency ν_1 , the number of spatial samples, N , must be large enough that the highest (Nyquist) frequency, $\nu_{N/2} = (N/2)\Delta\nu$ covers the domain of the variance spectrum for which the variance is non-negligible. This highest sampled frequency must also cover the highest frequency (shortest wavelength) needed for the problem at hand. The minimum resolvable wavelength is $2\Delta x = 2L/N$.

Of course, the need for large L and large N comes as the cost of increased computer time. Experimentation is necessary to determine what values are required for a particular physical situation.

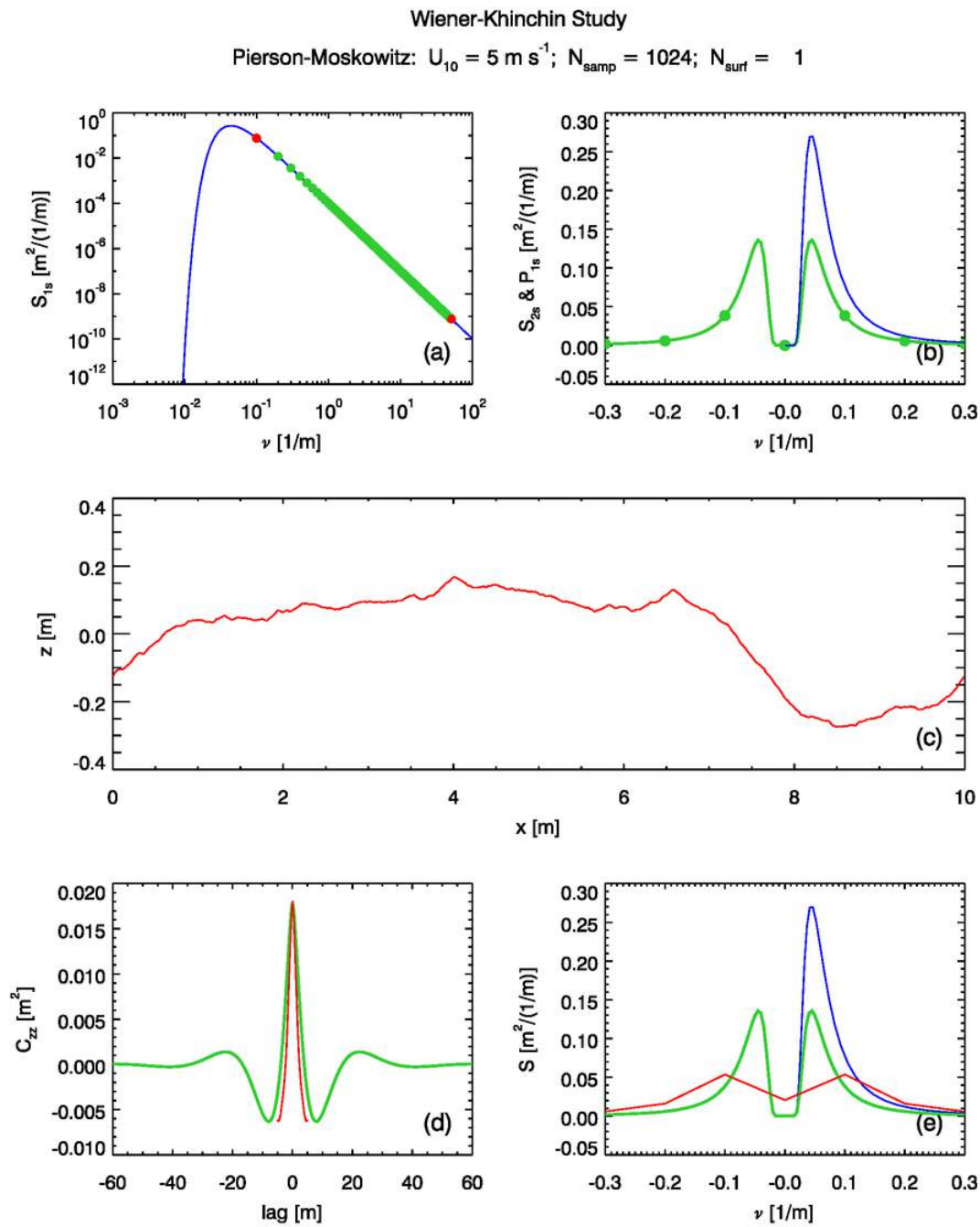


Figure D.7: Example simulation with inadequate sampling of the variance spectrum.

D.5 Turbulence-Generated Water Surfaces

This section illustrates the generation of random water surfaces beginning with the analytical autocovariance function of [Horoshenkov et al. \(2013\)](#),

$$C_{zz}(\ell) = C_{zz}(0) \exp\left(-\frac{\ell^2}{2\sigma_w^2}\right) \cos\left(\frac{2\pi}{L_o}\ell\right). \quad (\text{D.28})$$

This particular autocovariance has an analytical spectrum,

$$\mathcal{S}_{2s}(\nu) = \sqrt{\frac{\pi}{2}}\sigma_w C_{zz}(0) \left\{ \exp\left[-\frac{1}{2}(2\pi\sigma_w)^2(\nu + 1/L_o)^2\right] + \exp\left[-\frac{1}{2}(2\pi\sigma_w)^2(\nu - 1/L_o)^2\right] \right\}, \quad (\text{D.29})$$

which was computed in Section [D.2.1](#). Thus, for surface generation, the two-sided spectrum of Eq. [\(D.29\)](#) simply replaces the two-sided Pierson-Moskowitz spectrum used in the previous section, and the surface-generation calculations proceed as described previously. The autocovariance function is then not needed. However, if only the autocovariance is known or measured, then the needed variance spectrum must be obtained via the Wiener-Khinchin theorem. In the present study, knowing both the autocovariance and the spectral density as analytical functions provides a powerful check on the discrete numerical calculations of the same quantities.

The application of the above results is straightforward. If only the autocorrelation, rather than the autocovariance, is given, then a separate value of the surface elevation variance must be known. For the Horoshenkov study, a typical value of the surface variance is $C_{zz}(0) = 2.5 \cdot 10^{-7} \text{ m}^2$. (This is extremely small by oceanographic standards, but the surface waves in the Horoshenkov laboratory experiment had amplitudes of order 1 mm.) The previously cited parameter values of $\sigma_w = 0.22 \text{ m}$ and $L_o = 0.17 \text{ m}$ are used here. Since the characteristic spatial scales of σ_w and L_o are of order 0.2 m, a spatial region of length $L = 4 \text{ m}$ should be adequate to capture the spatial features of these surfaces. An N value of 1024 then gives the smallest resolvable wavelength as $2\Delta x \approx 0.8 \text{ cm}$, which is the scale of capillary waves. (Capillary waves have wavelengths in the range of a few millimeters to 2 cm.)

Figure [D.8](#) shows an example simulation based on the Horoshenkov variance spectrum [\(D.29\)](#). The layout is the same as for the Pierson-Moskowitz figures in Section [B.3.1](#). Panel (d) of the figure contains three autocovariance plots: The green curve is the inverse DFT of the sampled variance spectrum, which is shown in green in Panel (b). The red curve is the ensemble average autocovariance of 1000 water surface simulations. The purple curve is the theoretical autocovariance of Eq. [\(D.28\)](#). These three curves are indistinguishable at the scale of this plot. This nearly perfect agreement between autocovariance derived in three different ways indicates that the various numerical calculations are almost without doubt being done correctly.

The red curve in Panel (e) of the plot shows the variance spectrum derived via the Wiener-Khinchin theorem as the DFT of the ensemble-average autocovariance (the red curve in Panel (d)). Again, this curve is almost indistinguishable from the theoretical autocovariance, which is shown in green. Again, this agreement indicates that the DFTs are being computed correctly.

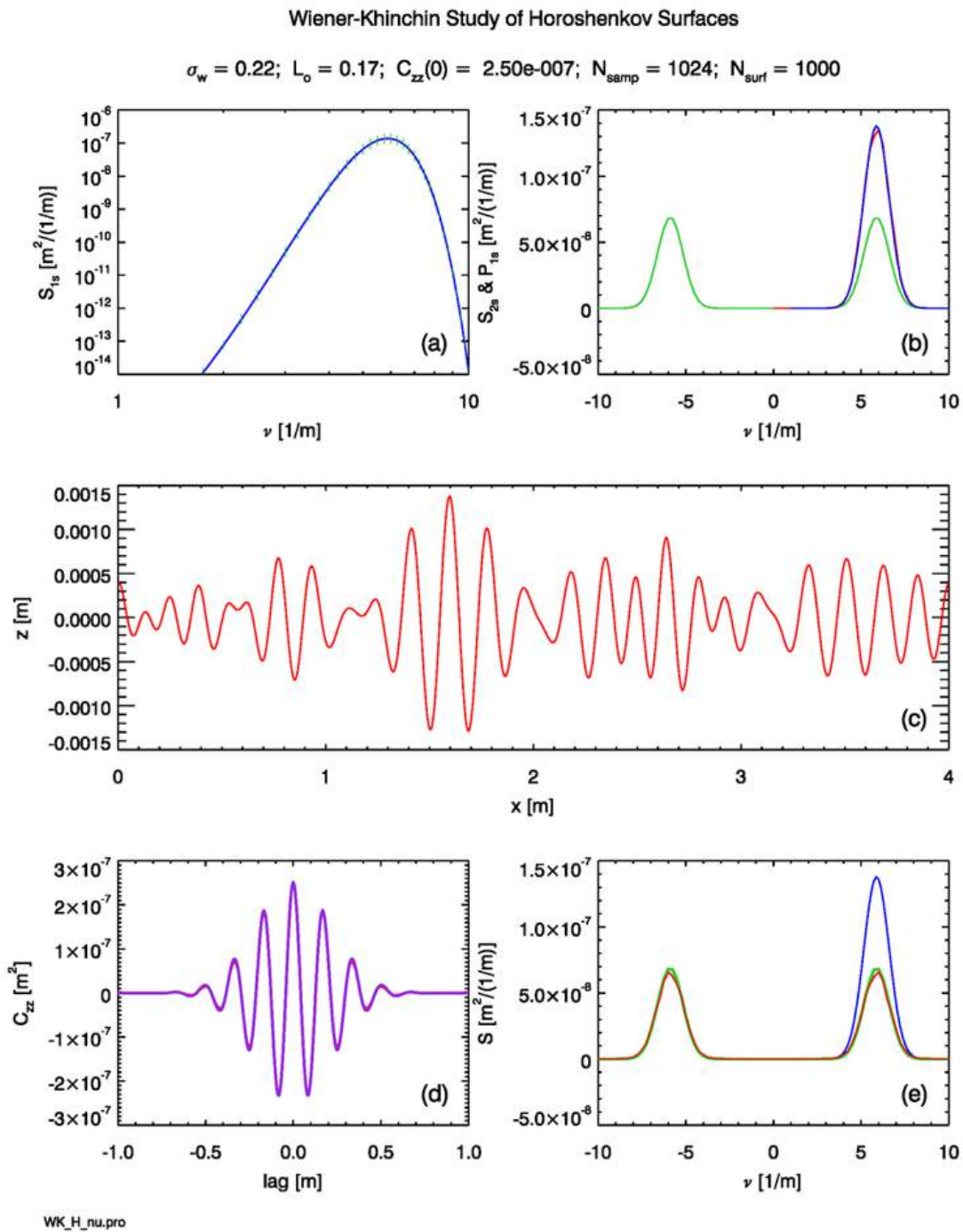


Figure D.8: Example of a turbulence-generated water surface based on the autocovariance function of Eq. (D.13). Compare the qualitative appearance of panel (c) with the sea surface shown in Fig. D.2.

D.5.1 Two-dimensional Water Surfaces

The IDL codes used in previous sections for generation of two-dimensional, time-independent water surfaces are formulated using a one-sided, two-dimensional elevation variance spectrum $\Psi(k_x, k_y)$ of the form (recall Eq. B.27)

$$\Psi(k_x, k_y) = \frac{1}{k} \mathcal{S}_{1s}(k) \Phi(k, \varphi). \quad (\text{D.30})$$

Here $\mathcal{S}_{1s}(k)$ is a one-sided omnidirectional spectrum and $\Phi(k, \varphi)$ is a nondimensional spreading function. To generate a 2-D, time-independent surface using the Horoshenkov model, the two-sided omnidirectional spectrum of Eq. (D.17) is multiplied by 2 to obtain a one-sided spectrum, which the IDL code evaluates only for the non-negative k_x values, i.e. for $-\pi/2 \leq \varphi \leq \pi/2$. The code then divides the result by 2 to get a two-sided spectrum and evaluates the $-k_x$ half plane of values by symmetry. Thus it is easy to replace an omnidirectional oceanographic $\mathcal{S}_{1s}(k)$ spectrum with that of Horoshenkov. There remains only the issue of what to use for a spreading function. There is no information about the spreading functions of turbulence-generated waves in the Horoshenkov et al. paper. There is no doubt some flow-induced difference in the waves in the “down-river” vs “cross-river” directions, just as there is in the “down-wind” vs “cross-wind” directions for wind-generated waves. However, pending further information on that difference, it is probably reasonable to use a frequency-independent, isotropic spreading function, $\Phi(\nu, \varphi) = \frac{1}{2\pi}$. With that assumption, two-dimensional surfaces can be generated.

Figure D.9 shows an example two-dimensional, turbulence-generated surface created with the $\sigma_w = 0.22$ m, $L_o = 0.17$ m and $C_{zz}(0) = 2.5 \cdot 10^{-7}$ m² values used for Fig. D.8. This particular 2-D surface realization has an elevation variance of $2.48 \cdot 10^{-7}$ m², which is close to the value of $C_{zz}(0)$ value used as input to the Horoshenkov spectrum. It is also noted that along any slice through the surface, there are about two dozen “bumps” in 4 m, just as seen in the 1-D surface realization of Fig. D.8. Figure D.10 shows the slice through the 2-D surface at $y = 2$. This surface is qualitatively like that of the middle panel of Fig. D.8. These results indicate that the 2-D calculations are correct.

The visual appearance of the Horoshenkov surface is strikingly different from the wind-generated sea surface seen in Fig. D.11, which is for a 5 m s^{-1} wind speed. In these plots, the surfaces have a factor-of-8 difference in the scaling of the surface elevation relative to the horizontal: 0.02 m vertical to 4 m horizontal = 0.005 for the Horoshenkov surface compared 4 m to 100 m = 0.04 for the wind-blown surface. This is purely for the visual appearance of the 3-D perspective plots. The Horoshenkov surface is actually quite smooth, with an average wave facet slope of only about 0.6 deg. The wind-blown surface has an average slope angle of about 3.7 deg in the along-wind direction and 2.9 deg in the cross-wind direction. (Keep in mind that for this simulation $\Delta x = 100/1024$, so the smallest resolvable wave has a wavelength of about 20 cm. Thus the smallest waves, which can have large slopes, are not resolved. An actual sea surface will therefore have larger average slopes.) Thus the Horoshenkov surface is smoother than the wind-blown surface, which suggests that turbulence-generated water surfaces may have significantly different optical reflectances than wind-generated surfaces. That hypothesis could be tested by ray tracing calculations based on surfaces like those of Figs. D.9 and D.11.

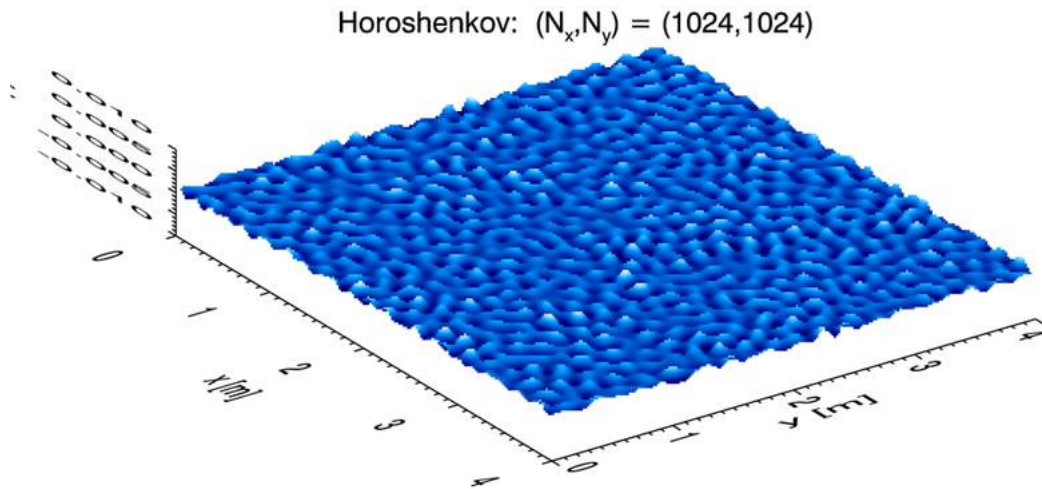


Figure D.9: A 2-D turbulence-generated surface. White is large positive surface elevations (wave crests) and dark blue is large negative values (wave troughs).

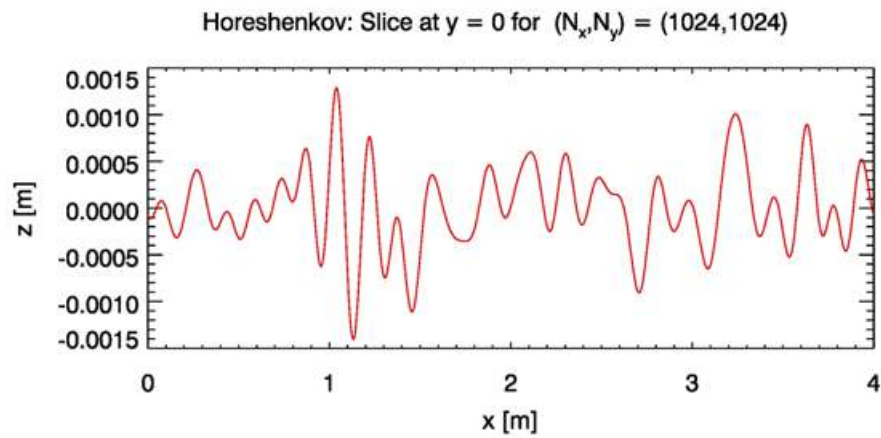


Figure D.10: A slice through the surface of Fig. D.9 at $y = 2$.

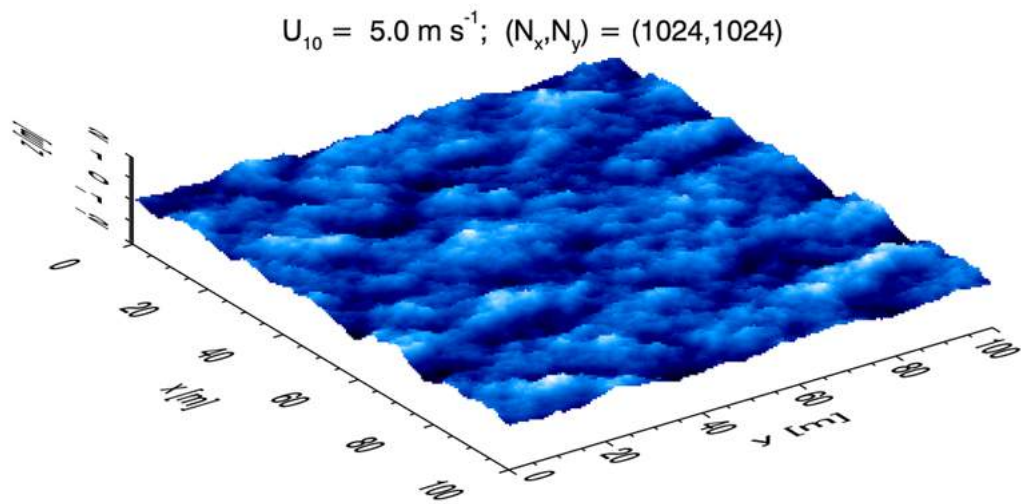


Figure D.11: A wind-generated surface for a wind speed of 5 m s^{-1} . Compare with Fig. D.9.

Monte Carlo Simulation

This book does not in general discuss solution methods for the RTE. However, the single-scattering approximation (SSA; Section 10.2) and the quasi-single-scattering approximation (QSSA; Section 10.3) are often referenced in the oceanographic and remote-sensing literature and were therefore included in the radiative transfer chapter. Monte Carlo solution techniques are frequently used in optical oceanography (and in many other fields) and therefore also warrant discussion. As an RTE solution technique, Monte Carlo techniques also could be included in the radiative transfer chapter. However, because of their importance and the amount of material to be discussed, Monte Carlo techniques deserve a chapter unto themselves.

As used to solve the RTE, Monte Carlo techniques refer to algorithms that use probability theory and random numbers to simulate the fates of numerous light rays propagating through a medium. Various averages over ensembles of large numbers of simulated ray trajectories give statistical estimates of radiances, irradiances, and other quantities of interest.

Monte Carlo techniques were developed in the 1940s for studies of neutron transport as needed for the design of nuclear weapons (Metropolis and Ulam, 1949; Eckhardt, 1987). The name Monte Carlo was at first a code name for this classified research. The name was well chosen because probability and statistics lie at the heart of both the simulation techniques and the gambling games played at the legendary Monte Carlo Casino in Monaco. Monte Carlo techniques are now highly developed and are used to solve many types of problems in physical and biological sciences; finance, economics, and business; engineering; computer graphics for movie production; and even in pure mathematics.

The essential feature of Monte Carlo simulation is that the known probability of occurrence of each separate event in a sequence of events is used to estimate the probability of the occurrence of the entire sequence. In the ray-tracing setting, the known probabilities that a light ray will travel a certain distance, be scattered through a certain angle, reflect off on a surface in a certain direction, etc., are used to estimate the probability that a ray emitted from a source at one location will travel through the medium and eventually be recorded by a detector at a different location.

The strengths of Monte Carlo techniques are that

- **They are conceptually simple.** The methods are based on a straightforward mimicry of nature, which in itself endows them with a certain elegance.
- **They are very general.** Monte Carlo simulations can be used to solve problems for any geometry (e.g., 3-D volumes with imbedded objects), incident lighting, scattering phase functions, etc. It is relatively easy to include polarization and time dependence.
- **They are instructive.** The solution algorithms highlight the fundamental processes of absorption and scattering, and they make clear the connections between the ray-level and the energy-level formulations of radiative transfer theory.
- **They are simple to program.** The resulting computer code can be relatively simple, and the tracing of rays is well suited to parallel processing.

The weaknesses of Monte Carlo techniques are that

- **They provide no insight into the underlying mathematical structure of radiative transfer theory.** The simulations simply accumulate the results of tracing large numbers of rays, each of which is independent of the others.
- **They can be computationally very inefficient.** Monte Carlo simulation is inherently a “brute force” technique. If care is not taken, much of the computational time can be expended tracing rays that never contribute to the solution because they never intercept a simulated detector.
- **They are not well suited for some types of problems.** For example, computations of radiance at large optical depths can require unacceptably large amounts of computer time because the number of solar rays penetrating the ocean decreases exponentially with the optical depth. Likewise, the simulation of a point source and a point (or very small) detector is difficult. Monte Carlo techniques are based on tracking individual rays in the geometric optics limit of physical optics and thus cannot address wave phenomena such as diffraction.

This appendix first discusses probability distributions and how they are sampled in Monte Carlo simulations. Subsequent sections discuss how ray tracks are simulated and how the statistical errors inherent in the computed results can be estimated. There are computational tricks for speeding up calculations and improving the accuracy of the statistical estimates. Two of the most important of these tricks—importance sampling and backward Monte Carlo ray tracing—are discussed in some detail.

The next appendix presents the mathematical details needed to carry out the ray tracing described in the present appendix.

E.1 Probability Functions

As a light ray travels through a medium, the distance it goes between interactions with the medium, whether it will be absorbed or scattered in a given interaction, the new direction it will travel after a scattering event, etc. are all random variables. In mathematics it is customary to let a capital letter, e.g. X , denote a random variable (such as the distance a ray travels) and to let a lower case letter, x in this case, denote a particular value of X ,

e.g. $x = 2.2$ m if X is distance. Let X be any such random variable that is defined over a range of values x_1 to x_2 . If X is distance traveled, $x_1 = 0$ and $x_2 = \infty$; if X is a polar scattering angle, $x_1 = 0$ and $x_2 = \pi$ or 180 deg, and so on.

The *probability density function* (PDF) for X , denoted $p_X(x)$, is a non-negative function such that $p_X(x)dx$ is the probability (a number between 0 and 1) that X has a value between x and $x + dx$. A PDF must satisfy the normalization

$$\int_{x_1}^{x_2} p_X(x)dx = 1. \quad (\text{E.1})$$

That is, the probability is 1 that X will have some value in its allowed domain. The *cumulative distribution function* (CDF) is a function $P_X(x)$ giving the probability that the random variable X will have a numerical value less than or equal to x . The CDF is obtained from the corresponding PDF via

$$CDF(x) = P_X(x) = \int_{x_1}^x p_X(x')dx'.$$

Note that a PDF has units of $1/\{\text{units of } x\}$ and can have a magnitude greater than 1 for some values of x , whereas a CDF is nondimensional and increases monotonically from 0 at $x = x_1$ to 1 at $x = x_2$.

The mean or average value of X is given by

$$\mu_X = \int_{x_1}^{x_2} x p_X(x)dx,$$

and the variance of X is given by

$$\sigma_X^2 = \int_{x_1}^{x_2} (x - \mu_X)^2 p_X(x)dx = \int_{x_1}^{x_2} x^2 p_X(x)dx - \mu_X^2.$$

Thus the variance is the mean of the square minus the square of the mean.

The PDF for a random variable that can have values only between 0 and 1 is fundamental to Monte Carlo simulation. Let \mathfrak{u} be the value of a random variable drawn from the unit interval between 0 and 1 such that \mathfrak{u} is equally likely to have any value $0 \leq \mathfrak{u} \leq 1$ on the interval from 0 to 1. The PDF for \mathfrak{u} is

$$p_{\mathfrak{u}}(\mathfrak{u}) = \begin{cases} 1 & \text{if } 0 \leq \mathfrak{u} \leq 1 \\ 0 & \text{if } \mathfrak{u} < 0 \text{ or } \mathfrak{u} > 1. \end{cases} \quad (\text{E.2})$$

The PDF $p_{\mathfrak{u}}$ is said to have a uniform 0 to 1 distribution, denoted by $p_{\mathfrak{u}} \sim \mathcal{U}[0, 1]$.

We wish to use a randomly drawn value of \mathfrak{u} to determine a value for a random variable X . This is done by regarding going from \mathfrak{u} to X as a change of variables. Then the probability that \mathfrak{u} lies between \mathfrak{u} and $\mathfrak{u} + d\mathfrak{u}$ is the same as the probability that X lies between x and $x + dx$. Thus

$$\int_0^{\mathfrak{u}} p_{\mathfrak{u}}(\mathfrak{u}')d\mathfrak{u}' = \int_{x_1}^x p_X(x')dx'$$

Because $p_{\mathcal{U}}(\mathbf{u})$ is known from Eq. (E.2), the left-hand integral can be evaluated to obtain

$$\mathbf{u} = \int_{x_1}^x p_X(x') dx' = P_X(x) = CDF(x) \quad (\text{E.3})$$

The fundamental principle of Monte Carlo simulation states that the equation $\mathbf{u} = P_X(x)$ uniquely determines x in such a manner that x lies in the interval x to $x + dx$ with probability $p_X(x)dx$. That is, drawing a value \mathbf{u} from a $\mathcal{U}[0, 1]$ distribution and then solving $\mathbf{u} = P_X(x)$ for x gives a randomly determined value of x that obeys the PDF for X .

The next sections illustrate how this principle is applied for specific examples of determining ray path lengths and scattering angles.

E.2 Determining Ray Path Lengths

Recall from the derivation of the RTE that radiance in a particular direction (θ, ϕ) decays due to absorption and scattering out of the beam according to

$$\frac{dL(r, \theta, \phi)}{dr} = -c(r)L(r, \theta, \phi),$$

which integrates to give

$$L(r, \theta, \phi) = L(0, \theta, \phi)e^{-\int_0^r c(r')dr'},$$

where $c(r)$ is the beam attenuation coefficient and r is the distance from some starting point. In terms of the optical path length $\tau = \int_0^r c(r')dr'$ this is

$$L(\tau, \theta, \phi) = L(0, \theta, \phi)e^{-\tau}.$$

This experimentally established exponential decay of radiance can be explained in terms of the fate of individual rays if the probability of any particular ray being absorbed or scattered out of the incident direction between τ and $\tau + d\tau$ is

$$p_T(\tau)d\tau = e^{-\tau}d\tau$$

Note that this $p_T(\tau)$ satisfies the normalization condition (E.1) with $x_1 = 0$ and $x_2 = \infty$. The corresponding CDF is $P_T(\tau) = 1 - e^{-\tau}$. Equation (E.3) now takes the form

$$\mathbf{u} = P_T(\tau) = 1 - e^{-\tau}.$$

Solving for τ gives

$$\tau = -\ln(1 - \mathbf{u}).$$

Because $1 - \mathbf{u}$ is also uniformly distributed on $[0, 1]$, we can just as well use

$$\tau = -\ln(\mathbf{u})$$

to determine τ . Note that since $\mathbf{u} \leq 1$, $\tau \geq 0$. Optical distances randomly chosen in this manner, when applied to many rays, are consistent with the exponential decay of radiance

with distance traveled. If the water is homogeneous, so that $c(r)$ does not depend on r , then $\tau = cr$ and the geometric distance a ray travels is given by

$$r = -\frac{1}{c} \ln(\mathbf{u}). \quad (\text{E.4})$$

Note that the average distance a ray travels is given by

$$\mu_T = \int_0^\infty \tau e^{-\tau} d\tau = 1$$

or, for homogeneous water,

$$\mu_R = \frac{1}{c}.$$

The average distance a ray travels between an absorption or scattering interaction with the water is called the *mean free path* between interactions. Likewise, the standard deviation σ_T , which is the square root of the variance σ_T^2 , of the optical distance traveled is also 1, or $\sigma_R = 1/c$. Thus rays travel on average one optical path length, or $1/c$ meters, but with a large spread of values about that distance.

E.3 Determining Scattering Angles

Scattering is an inherently three-dimensional process and must be specified by both polar (ψ) and azimuthal (α) scattering angles. The scattering phase function $\tilde{\beta}(\psi', \alpha' \rightarrow \psi, \alpha)$ can be interpreted as a PDF for scattering from an incident direction (ψ', α') to a final direction (ψ, α) , per unit of solid angle. If we pick a spherical coordinate system centered on the incident direction (ψ', α') and recall the expression for an element of solid angle in spherical coordinates, then we can write

$$\tilde{\beta}(\psi', \alpha' \rightarrow \psi, \alpha) d\Omega(\psi, \alpha) = \tilde{\beta}(\psi, \alpha) \sin \psi d\psi d\alpha.$$

Ocean water is usually well described as isotropic medium, which means that there are no optically preferred directions. (This is not the case, for example, in a cirrus cloud with non-randomly oriented ice crystals.) In that case, the polar and azimuthal scattering angles are independent, and we can write

$$\tilde{\beta}(\psi, \alpha) \sin \psi d\psi d\alpha = p_\Psi(\psi) d\psi p_A(\alpha) d\alpha$$

For an unpolarized beam, the azimuthal angle is equally likely to have any value from 0 to 360 deg, or 0 to 2π radians. Thus the PDF for azimuthal scattering is $p_A(\alpha) = 1/(2\pi)$, the CDF is $P_S(\alpha) = \alpha/(2\pi)$, and α is determined by

$$\alpha = 2\pi\mathbf{u}. \quad (\text{E.5})$$

Using this $p_X(\alpha)$ in the previous equation allows us to identify the PDF for the polar angle as

$$p_\Psi(\psi) = 2\pi\tilde{\beta}(\psi) \sin \psi$$

Recall from the discussion of the Volume Scattering Function in Section 3.1 that phase functions satisfy the normalization

$$2\pi \int_0^\pi \tilde{\beta}(\psi) \sin \psi d\psi = 1,$$

so this function $p_{\Psi}(\psi)$ is indeed a PDF. Therefore, to determine the polar scattering angle, we draw a $U[0, 1]$ random number u as always and solve

$$u = 2\pi \int_0^{\psi} \tilde{\beta}(\psi') \sin \psi' d\psi' \quad (\text{E.6})$$

for ψ .

In general, Eq. (E.6) must be solved numerically because of the complicated shape of most phase functions, or when the phase function is defined by tabulated data at a finite number of scattering angles and is fit with a spline (or other) function to generate a continuous function of scattering angle. However, a few commonly used phase functions allow this equation to be solved analytically. The simplest case, isotropic scattering, is instructive.

Isotropic Scattering. The phase function for isotropic scattering is $\tilde{\beta}(\psi) = \frac{1}{4\pi}$. Substituting this into Eq. (E.6) leads to

$$\psi = \cos^{-1}(1 - 2u) \quad (\text{E.7})$$

for the determination of the scattering angle. This result may look peculiar until it is remembered that *isotropic scattering means that scattering is equally likely into any element of solid angle, not equally likely at every scattering angle ψ* . Figure E.1 illustrates this important point. Scattering from a collimated beam is simulated two different ways. The scattering events occur at the center of a sphere, and the points plotted on the surface of the sphere show the scattering direction. The arrow at the “north pole” represents the direction of the unscattered beam and a scattering angle of 0. The blue lines are lines of constant scattering angle ψ , with the thick line at the “equator” being $\psi = 90$ deg. In the left panel, the polar scattering angle ψ was drawn from a $U[0, 180]$ distribution, i.e., any value of ψ between 0 and 180 deg was equally likely. The right panel drew ψ from the $\cos^{-1}(1 - 2u)$ distribution. In both simulations the azimuthal scattering angle α was drawn from a $U[0, 360]$ distribution. It is visually clear from the left panel that the uniform distribution of ψ values gives too many points near the north pole of the figure, i.e., too many points with scattering angles near 0. The right panel shows an even distribution of points per unit area of the surface of the sphere, i.e., per unit solid angle in any direction.

The widely used Fournier-Forand phase function, Eq. (6.15) described in Section 6.7, is

$$\begin{aligned} \tilde{\beta}^{\text{FF}}(\psi) = & \frac{1}{4\pi(1-\delta)^2\delta^\nu} \left[\nu(1-\delta) - (1-\delta^\nu) + [\delta(1-\delta^\nu) - \nu(1-\delta)] \sin^{-2}\left(\frac{\psi}{2}\right) \right] \\ & + \frac{1-\delta_{180}^\nu}{16\pi(\delta_{180}-1)\delta_{180}^\nu} (3\cos^2\psi - 1), \end{aligned}$$

where

$$\nu = \frac{3-\mu}{2} \quad \text{and} \quad \delta = \frac{4}{3(n-1)^2} \sin^2\left(\frac{\psi}{2}\right),$$

Here n is the real index of refraction of the particles, μ is the slope parameter of the hyperbolic distribution, and δ_{180} is δ evaluated at $\psi = 180$ deg. This phase function has

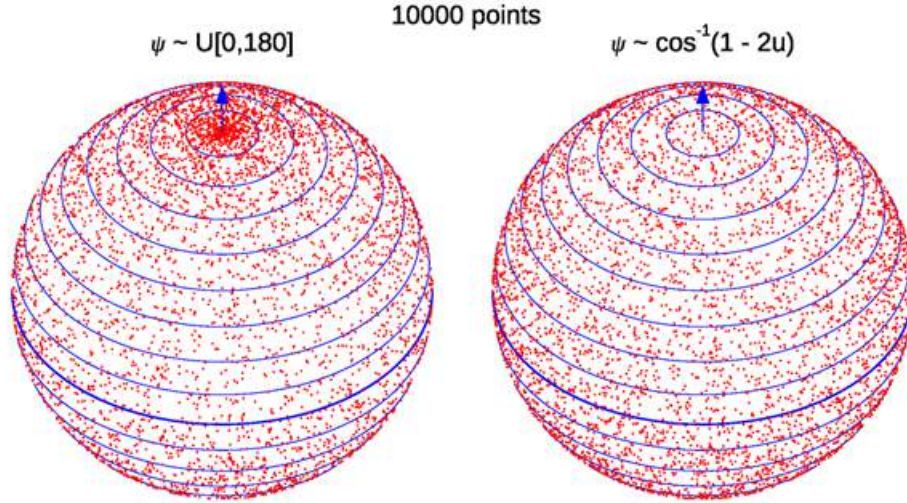


Figure E.1: Scattering directions for a collimated beam. The left panel is for a uniform distribution of polar scattering angles ψ . The right panel is for an isotropic distributions of scattering directions with ψ determined by Eq. (E.7).

an analytic CDF (Fournier and Jonasz, 1999)

$$P_{\Psi}^{\text{FF}}(\psi) = \frac{1}{(1-\delta)\delta^{\nu}} \left[(1-\delta^{\nu+1}) - (1-\delta^{\nu}) \sin^2(\psi/2) \right] + \frac{1}{8} \frac{1-\delta_{180}^{\nu}}{(\delta_{180}-1)\delta_{180}^{\nu}} \cos \psi \sin^2 \psi. \quad (\text{E.8})$$

However, solving $\mathbf{u} = P_{\Psi}^{\text{FF}}(\psi)$ for ψ (even if possible) would give a formula so complicated that it is numerically more efficient to use Eq. (E.6) for $P_{\Psi}^{\text{FF}}(\psi)$ to build up a table of ψ versus $P_{\Psi}^{\text{FF}}(\psi)$ values for closely spaced values of ψ and the given n and μ parameters, and then to interpolate within this table to obtain values of ψ as \mathbf{u} values are drawn in the course of a simulation. This is illustrated in Fig. E.2, which shows a Fourier-Forand CDF for values of n and μ that give a best fit to the Petzold average particle phase function phase function of Section 9.5.2. The blue arrows show how drawing a value of $\mathbf{u} = 0.7$ leads to a scattering angle of about 10 deg. When working with tabulated data for highly peaked phase functions, it is usually adequate to use linear interpolation between the tabulated values.

Table E.3 displays various phase functions and the formulas obtained from solving $\mathbf{u} = P_{\Psi}(\psi)$ used to determine the corresponding scattering angles.

Regarding the ψ formula for the cosine phase function, because $\sin^2 \psi$ can be written as $1 - \cos^2 \psi$, the equation for ψ can also be written as $\psi = \cos^{-1}(\sqrt{1-\mathbf{u}})$ or $\psi = \cos^{-1}(\sqrt{\mathbf{u}})$ after noting that $1-\mathbf{u}$ has the same statistical distribution as \mathbf{u} . Ray by ray the ψ values will be different for the same value of \mathbf{u} , but the statistical distribution of ψ values resulting from a large number of emitted rays will be the same, which is what matters in Monte Carlo simulations. In Section F.6, this same distribution of angles will be found for reflectance by a Lambertian surface.

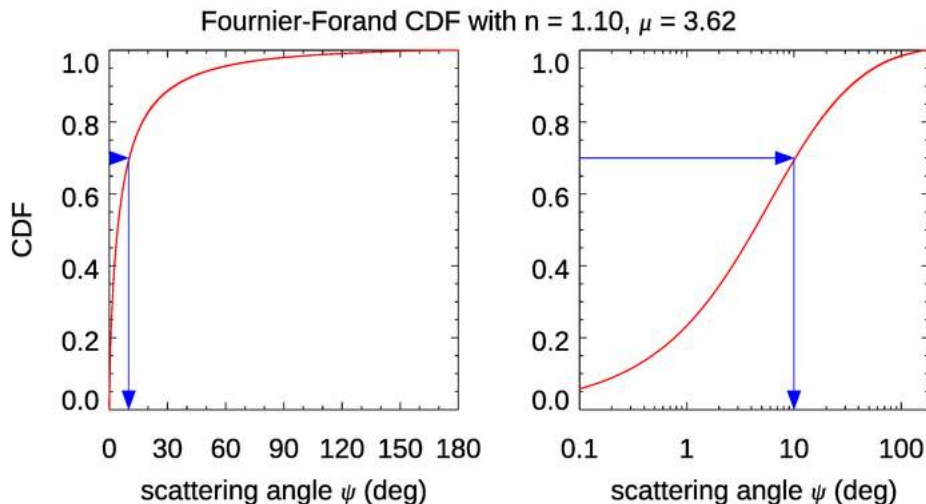


Figure E.2: A Fournier-Forand cumulative distribution function computed using Eq. (E.8). The horizontal blue arrow shows a value $u = 0.7$, which by Eq. (E.6) leads to a scattering angle of $\psi \approx 10$ deg, as shown by the downward blue arrow.

Name	Phase Function $\tilde{\beta}(\psi)$	ψ formula
isotropic	$\frac{1}{4\pi}$	$\psi = \cos^{-1}(1 - 2u)$
Henyey-Greenstein	$\frac{1}{4\pi} \frac{1-g^2}{(1+g^2-2g \cos \psi)^{3/2}}$	$\psi = \cos^{-1} \left[\frac{1+g^2}{2g} - \frac{1}{2g} \left(\frac{1-g^2}{1+g-2gu} \right)^2 \right]$ for $-1 \leq g \leq 1$, but $g \neq 0$
Rayleigh	$\frac{3}{16\pi} (1 + \cos^2 \psi)$	$\psi = \cos^{-1} \left[(A+B)^{1/3} + (A-B)^{1/3} \right]$ where $A = 2(1-2u)$ and $B = \sqrt{1+A^2}$
cosine	$\frac{1}{\pi} \cos \psi$ if $0 \leq \psi \leq \frac{\pi}{2}$ 0 if $\frac{\pi}{2} < \psi \leq \pi$	$\psi = \sin^{-1}(\sqrt{u})$ or $\psi = \cos^{-1}(\sqrt{u})$
arbitrary	any $\tilde{\beta}(\psi)$ that satisfies $2\pi \int_0^\pi \tilde{\beta}(\psi) \sin \psi = 1$	must solve $u = 2\pi \int_0^\psi \tilde{\beta}(\psi') \sin \psi' d\psi'$ numerically for ψ

Table E.1: Formulas for randomly choosing scattering angles ψ for commonly used phase functions $\tilde{\beta}(\psi)$. The PDF associated with $\tilde{\beta}(\psi)$ is always $2\pi\tilde{\beta}(\psi) \sin \psi$. u is a $\mathcal{U}[0, 1]$ random number.

E.4 Ray Tracing Options

The previous sections showed how to determine ray path lengths and scattering angles using the beam attenuation c , the scattering phase function $\hat{\beta}$, and a uniform $U[0, 1]$ random number generator. This section shows how to combine those two processes to create random ray paths through an absorbing and scattering medium.

There is more than one way to simulate ray paths, and each will give the same answer. However, some techniques can be numerically much more efficient than others. Indeed, a reasonable approach to developing a Monte Carlo algorithm for a particular problem is to

1. first figure out how to numerically simulate a process as it occurs in nature, and
2. then figure out how to simulate another, perhaps artificial, process that will give the same answer as the “natural” process, but with less computational time.

This section illustrates this two-step development process.

Consider first how rays propagate through a medium. Loosely speaking, a ray travels until it interacts with a particle, e.g. a molecule of water or chlorophyll. It is then either absorbed by the particle and disappears, or it is scattered into a new direction and continues on its way until it interacts with another particle.

Recall the albedo of single scattering, $\omega_o = b/c$. If there is no absorption, $b = c$ and $\omega_o = 1$. If there is no scattering, $b = 0$ and $\omega_o = 0$. $\omega_o = b/c$ thus can be interpreted as the *probability of ray survival in any particular interaction*. When a ray encounters a particle, we can randomly decide if the ray is to be absorbed or scattered as follows:

1. Draw a random number u from a $U[0, 1]$ distribution.
2. Compare u with ω_o .
 - If $u \leq \omega_o$, then the ray is scattered.
 - If $u > \omega_o$, then the ray is absorbed.

If the ray is absorbed, tracing stops and a new ray is emitted from the source and tracing begins anew. If the ray is scattered, two new $U[0, 1]$ random numbers are drawn and used to determine new polar and azimuthal scattering directions ψ and α as shown in the previous section. Another random number is used along with c to determine the distance traveled before another interaction.

Figure E.3 illustrates this process for two rays, which also introduces the geometry to be used in numerical simulations below. A source emits a collimated beam of rays, which are then recorded by an annular, ring, or “bullseye” detector some distance away. The red ray is emitted by the source, undergoes one scattering, and is then absorbed by a particle. The green ray is emitted, undergoes two scatterings, and is recorded by a detector.

This process mimics what happens in nature. Call this “Type 1” ray tracing (there are no standard names for ways of tracing rays). ***Note that all of the computations used to trace the absorbed ray are wasted because the ray never reached the detector.*** Nature can afford to trace innumerable rays and waste some by absorption, but that is not advisable for most numerical simulations. We therefore seek other ways to trace rays.

The previous section showed that the mean free path or average distance between interactions with the medium is $1/c$. These interactions can result in either absorption or

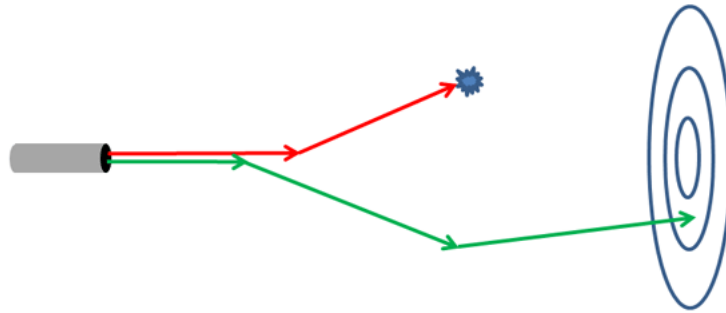


Figure E.3: Illustration of Type 1 ray tracing. The red ray is absorbed and the green one reaches the target.

scattering of the ray, as just described. Rather than tracing one ray at a time as nature does, consider a source emitting “bundles” of many rays (often called “photon packets” in the literature). Then view each interaction as having a fraction $1 - \omega_o$ of the rays in the bundle being absorbed, and the remaining fraction ω_o being scattered, all in the same direction. Let the bundle be emitted with an initial weight of $w = 1$, which can represent one unit of energy, power, or some number of rays. At each interaction, the current weight w is multiplied by ω_o to account for the loss of energy or number of rays by absorption (that is, a fraction ω_o continues onward). The scattered bundle then carries a reduced weight. If the ray bundle reaches the target, the current weight w is tallied. Another bundle is then emitted from the source and traced. This tracing process, which we’ll call “Type 2,” is illustrated by the green ray track in Fig. E.4. After two scatterings, as illustrated, the detected ray bundle has weight $w = \omega_o^2$.

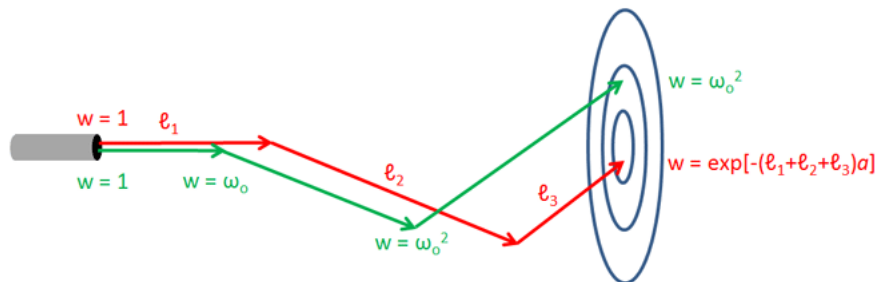


Figure E.4: Illustration of Type 2 ray tracing in green and Type 3 in red.

A third ray-tracing process can be envisioned. The mean distance traveled between scattering events is $1/b$. We can thus use $1/b$ and a random number to determine the distance between scattering interactions, and the initial weight of $w = 1$ is not changed at each interaction because all rays in the bundle are viewed as being scattered. Then, if the ray bundle reaches the target, absorption is treated as a continuous process occurring along the entire ray path. Assuming homogeneous water, the final weight tallied is then $e^{(-la)}$, where ℓ is the total path length in meters and a is the absorption coefficient. The red track in Fig. E.4 illustrates this “Type 3” ray tracing. The red track shows a total

path length of $\ell_1 + \ell_2 + \ell_3$, so the final weight is $e^{-[(\ell_1 + \ell_2 + \ell_3)a]}$.

The two tracks in Fig. E.4 are drawn as though each track were generated by exactly the same sequence of random numbers. Because $1/b > 1/c$, the individual Type 3 ray paths will be greater than the Type 2 paths. The scattering angles are the same. Thus these two tracing types clearly lead to different results, ray bundle by ray bundle. However, *numerical simulation of many ray bundles shows all three of these ray tracing types yields the same distribution of energy at the detector.*

To summarize, the three types of ray tracing considered here are

Type 1: Individual rays are tracked, and rays can be absorbed.

Type 2: Ray bundles are tracked, with a bundle weight being multiplied by ω_o at each interaction.

Type 3: Ray bundles are tracked, with track lengths determined by the mean free path for scattering, no weighting at scattering events, and absorption treated as a continuous process based on total path length.

E.4.1 Numerical Comparison of Tracking Types

To illustrate the results obtained for different ways of tracking rays, a Monte Carlo code was written to simulate the energy received by an annular target as illustrated in Figs. E.3 and E.4. For the simulations shown here, the IOPs were defined by $a = 0.2 \text{ m}^{-1}$, $b = 0.8 \text{ m}^{-1}$, and a Fournier-Forand phase function with parameter values $(n, \mu) = (\text{index of refraction, slope of Junge distribution})$ chosen to give a good fit to the Petzold average particle phase function. Thus $\omega_o = 0.8$, and optical distance τ is numerically the same as geometrical distance ℓ in meters. A run was made with 10^6 rays being sent from the source and using Type 1 ray tracing. rays were traced until they were absorbed. Figure E.5 shows some of the resulting statistics.

The red histogram shows the percent of rays that traveled an optical distance $\tau_1 \leq \tau \leq \tau_2$ between interactions, for a bin size of $\tau_2 - \tau_1 = 1$. The theoretically expected fraction of rays traveling an optical distance between τ_1 and τ_2 between interactions is

$$\int_{\tau_1}^{\tau_2} e^{-\tau} d\tau = e^{-\tau_1} - e^{-\tau_2} \quad (\text{E.9})$$

The red dots in the figure are the expected values given by this formula. The shortest ray path length between interactions was $\tau = 1.192 \cdot 10^{-7}$ and the longest was 16.69. The blue histogram shows the distribution of total distances traveled until the rays were absorbed. Thus the value for the first bar shows that about 18% of the rays were absorbed after going a total optical distance between 0 and 1. Note that this distance can represent more than one interaction, i.e., a ray being scattered one or more times before being absorbed. Both the red and the blue histograms sum to 100%. As shown on the previous section, the mean distance traveled between interactions is $\tau = 1$, or $1/c$ in meters. For this particular simulation the actual average was $\tau = 0.9974$ (or 0.9974 m for these IOPs). The small difference is statistical noise resulting from the finite sample size of the numerical simulation. Likewise, the average distance traveled until the rays are absorbed is $1/a$. For the present case of $a = 0.2 \text{ m}^{-1}$ this gives 5 m. The average for this simulation was 4.9949 m. Since $c = 1 \text{ m}^{-1}$ for this simulation, another way to view this is that the rays were on average scattered four times before being absorbed on the fifth interaction.

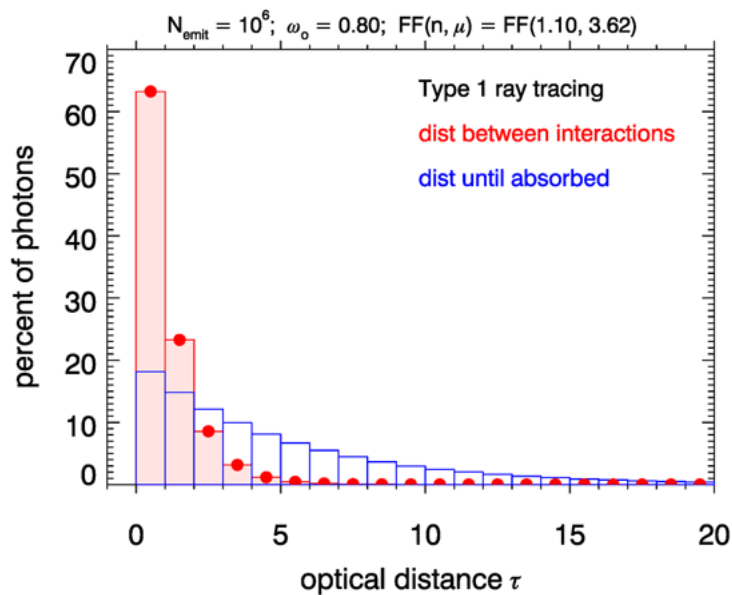


Figure E.5: Example distributions of the optical distances between interactions and of the total distances traveled before absorption for Type 1 ray tracing.

We next compare results for the three different ways of tracking rays. Because oceanic phase functions scatter much more light at small scattering angles than at large angles, most rays that are scattered just a few times will hit the detector near its center. To even out the numbers of rays (or power) detected by each ring, an annular target was defined with a logarithmic spacing for the radii of the detector rings. A logarithmic spacing is often used in instruments (e.g., the LISST particle sizer) so that each detector ring receives roughly equal amounts of power, which reduces the dynamic range needed for the instrument design. The detector simulated here had $N_{\text{rings}} = 10$ rings with the smallest ring radius being $r_{\text{min}} = 0.1$ and the largest being $r_{\text{max}} = 10$. This detector is placed in a target plane some distance z_T from the source and centered on the optical axis of the source, as shown in Figs. E.3 and E.4. The rays crossing the detector plane at some distance r from the detector center are tallied in bins as follows:

Bin 0: Unscattered rays that hit the detector at $r = 0$.

Bin 1: Scattered rays that hit the target plane inside the first detector ring, i.e. at $0 < r < r_{\text{min}}$.

Bins 2, ..., $N_{\text{rings}} + 1$: rays that hit detector rings 1 to N_{rings} .

Bin $N_{\text{rings}} + 2$: rays that hit the detector plane outside the outer ring, i.e. at $r > r_{\text{max}}$.

Simulations were made with $N_{\text{emit}} = 10^6$ rays (for Type 1) or ray bundles (Types 2 and 3) emitted from the collimated source. Figure E.6 shows the distribution of rays or bundles received anywhere in the detector plane as a function of the number of scatterings, for the three types of tracing and for target plane distances of $z_T = 5$ and 15. The left panel shows that for $z_T = 5$ one or two percent of rays (depending on the tracing type)

reach the target plane without being scattered. Most rays are scattered 3 or 4 times, and very few rays are scattered more than 10 times. The right panel shows that for $z_T = 15$, almost no rays reach the target plane unscattered, and most undergo 5 to 25 scatterings, with a peak around 10 or 15, depending on the way the rays are traced. For Type 1 ray tracing, almost no rays are scattered more than 30 times. Note that the probability of surviving 30 scatterings is $\omega_0^{30} = 0.0012$. For tracing Types 2 and 3, which never have ray bundles absorbed, there are broad tails in the number of scatterings, although for Type 2 a bundle being scattered 40 times has an almost negligible weight of $w = \omega_0^{40} = 0.00013$.

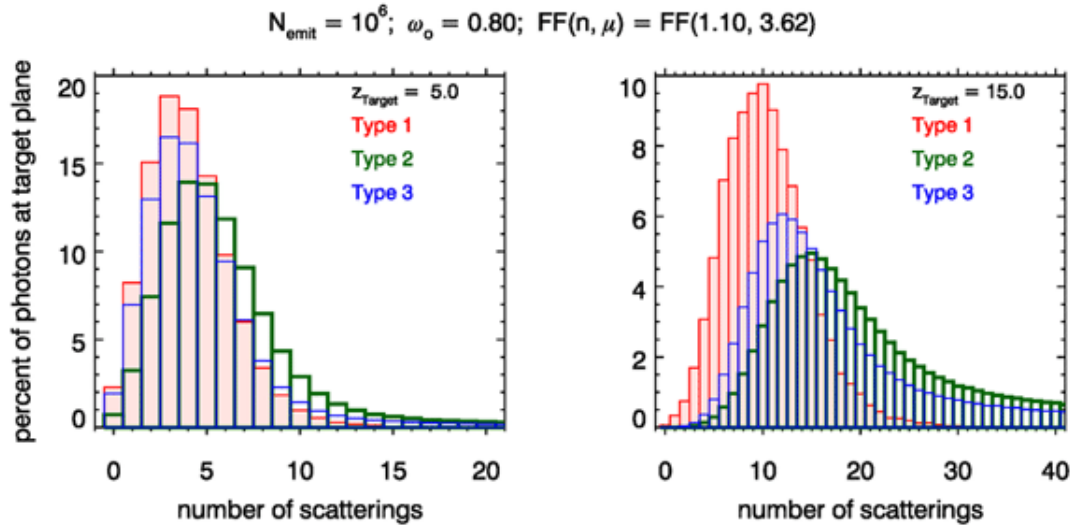


Figure E.6: Distribution of rays reaching the target plane as a function of the number of scatterings, for the three scattering types and two target plane distances.

The left panel of Fig. E.7 shows the intersection points of the rays reaching a 20 by 20 area of the target plane for the first 10^4 emitted rays, tracing Type 1, and the target plane at $z_T = 5$. Note that of the 10^4 emitted rays, only 2973 reached the target plane (of which 2966 are in the area plotted). This is less than 30% of the rays making a contribution to the answer of how much power is detected; i.e., 70% of the calculations were wasted. The ray-intersection dots are color-coded to show the number of times each plotted point was scattered. The two black circles show off-axis angles of 30° and 60° . Very few rays were scattered more than about 30° off of the optical axis. The right-hand panel shows the distributions of the rays reaching the target plane by number of scatterings and total distance traveled. Every ray must of course travel a distance of at least $\tau = 5$ before reaching the target plane, and most rays were scattered several times.

Figure E.8 shows the same distributions for Type 2 scattering. Note that now 93% of the emitted ray bundles eventually intersect the target plane. Only 7% of the emitted rays were wasted. Those rays ended up being scattered into directions away from the target plane (either by backscattering or by multiple large-angle forward scatterings). Figure E.9 shows the results for Type 3 tracing. The distributions are similar to those for Type 2, but with over 94% of the rays reaching the target plane.

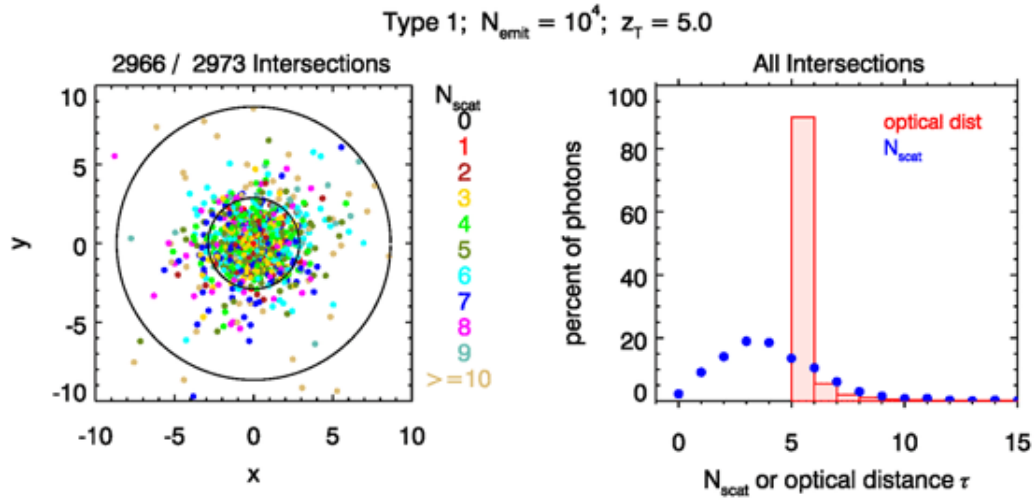


Figure E.7: Distribution of rays reaching the target plane at $z_T = 5$ for 10^4 emitted rays and Type 1 tracing. The left panel shows the spatial distribution of points where the rays intersected the target plane. The two black circles are drawn at 30° and 60° angles off of the optical axis. The right panel shows the distributions of number of scatterings and total distance traveled.

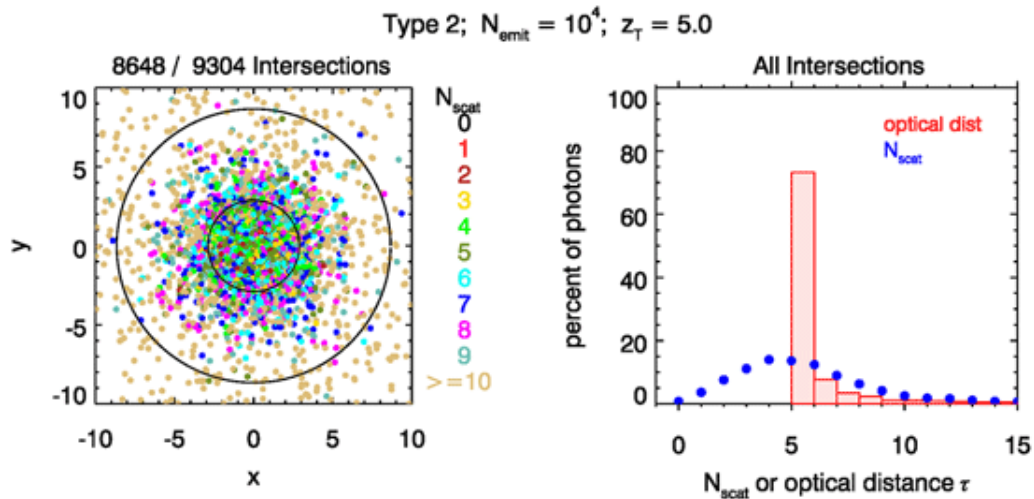


Figure E.8: Same as Fig. E.7 but for Type 2 tracing.

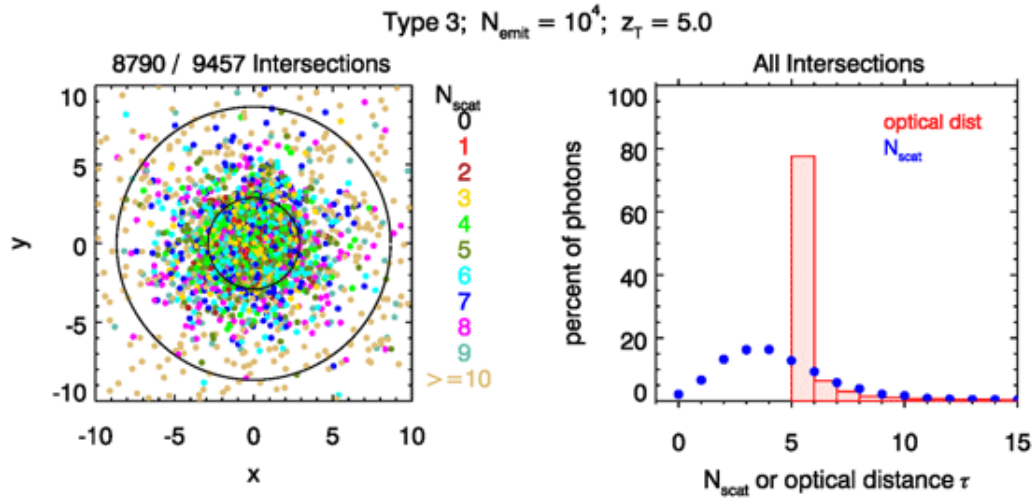


Figure E.9: Same as Fig. E.7 but for Type 3 tracing.

Figures E.10-E.12 show the corresponding results when the target plane is at $z_T = 15$. Now, for Type 1 tracing, only 179 of 10^4 emitted rays ever reached the target plane. Over 98% of the ray-tracing calculations were wasted! Note also that there is obvious statistical noise in the distributions of the right panel, due to the small number of rays used to compute the statistics. For Types 2 and 3 about 71% and 82%, respectively, of the emitted rays eventually reach the target plane. The statistical noise is now much smaller (but still noticeable) because of the larger number of ray bundles.

The total optical distance distributions for Types 2 and 3 show broad tails. In both cases, fewer than one fourth of the rays made it to the target plane after traveling a total distance of $\tau = 15$ –16. About 30% of the rays underwent 30 or more scatterings and traveled a distance of $\tau \geq 30$. These broad tails illustrate the phenomenon of pulse stretching in time-dependent problems. If we think of all N_{emit} rays being emitted simultaneously, then the longer distances traveled correspond directly to later arrival times at the target. Pulse stretching is an important limiting factor in time-dependent applications such as lidar bathymetry or communications with high-frequency light pulses.

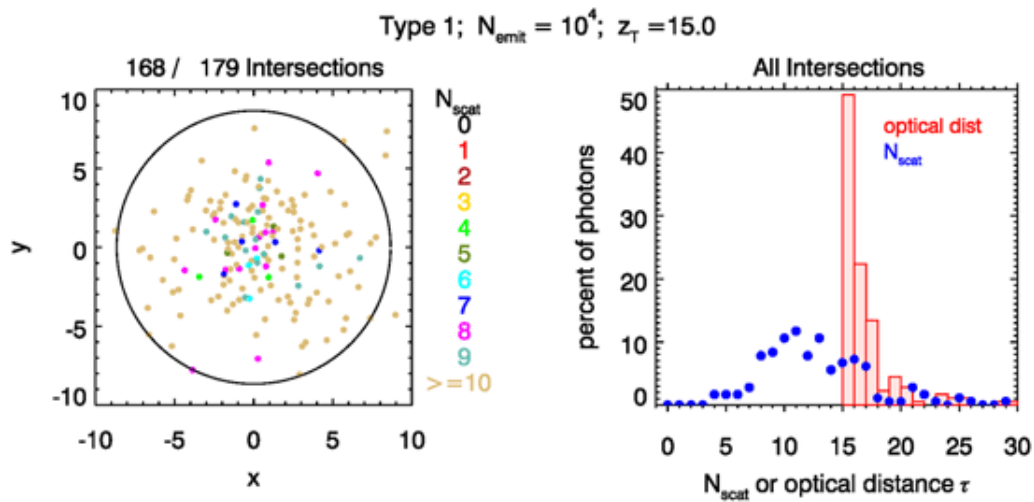


Figure E.10: Same as Fig. E.7 but for a target plane at $z_T = 15$. The black ring in the left panel is 30° off of the optical axis.

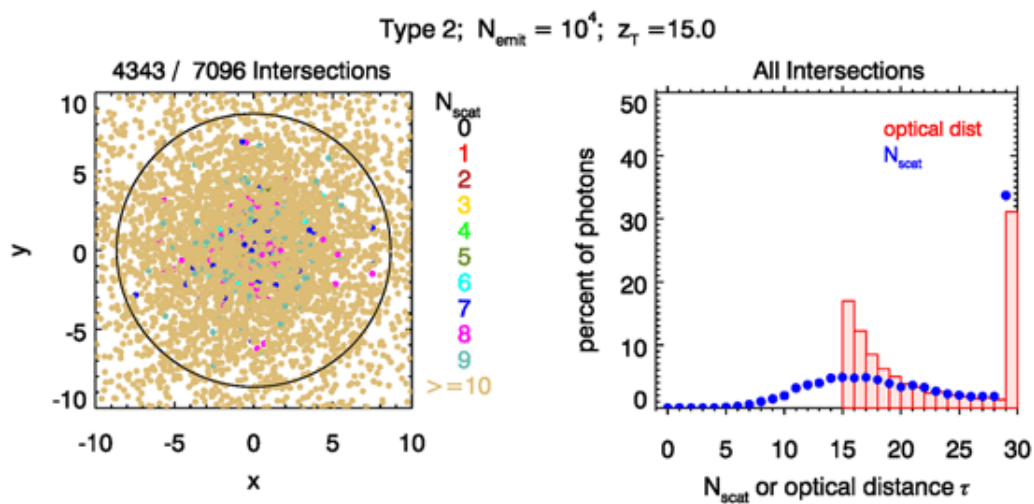


Figure E.11: Same as Fig. E.10 but for Type 2 tracing.

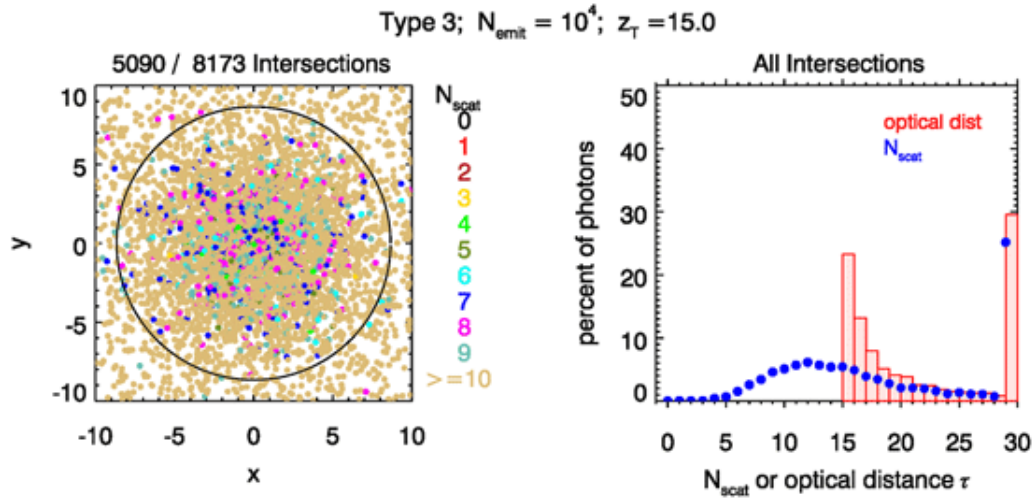


Figure E.12: Same as Fig. E.10 but for type 3 tracing.

Figure E.13 shows the distributions of numbers of rays at the target plane and the corresponding power (or energy) for the three tracing types and the detector at $z_T = 5$. The left panel shows the distributions as a function of the detector ring radii, and the right panel is the same information as a function of the bin number defined above. Recall that the first abscissa point is unscattered rays, the second is scattered rays inside the first detector ring, and the last plotted point is for rays outside the last detector ring. The solid-line histogram represents the 10 detector rings for $r_{\text{min}} = 0.1 < r < r_{\text{max}} = 10$.

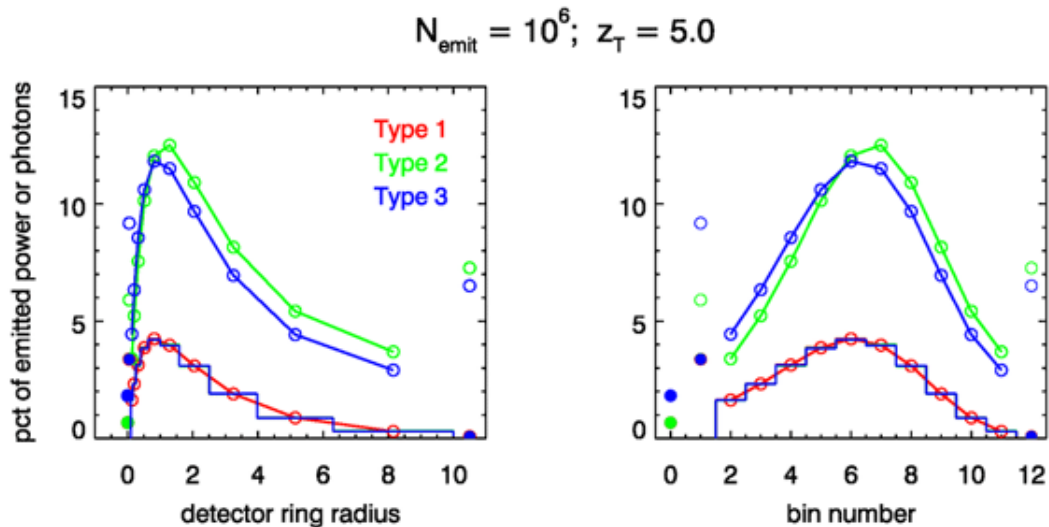


Figure E.13: Distributions of number of rays (open circles) and power (histograms) for the target plane at $z_T = 5$, for the three types of tracing. The left panel shows the distributions as a function of radius from the optical axis. The right panel displays the same information as a function of the bin number.

There are several things to notice in Fig. E.13. First, the distributions of the numbers of rays (open circles) are different for the three tracing types. For Type 1, the distribution of the number of rays is the same as the power distribution because the rays all retain their initial weight of $w = 1$. Thus power detected is simply the number of rays detected. For Types 2 and 3, more rays are detected, but each is weighted less to account for absorption along the way. Finally—and most importantly—the power distributions for these three tracing types are identical to within a small amount of statistical noise, which is not visible in these plots.

Figure E.14 shows the power distributions for the detector at $z_T = 15$ but only $N_{\text{emit}} = 10^4$ rays emitted. There are obvious differences in the distributions for the three tracing types. However, if $N_{\text{emit}} = 10^6$ rays are traced, as in Fig. E.15 these differences almost disappear. This indicates that *the three ways to trace rays all give identical predictions of the detected power, to within some amount of statistical noise, which can be reduced by tracing more rays.*

However, *the computation time required by the three tracing types can vary greatly.* Recall from Fig. E.7 that about 30% of the rays reached the target plane at $z_T = 5$ for Type 1 tracing, but that over 90% reached the target plane for Types 2 and 3 (Figs. E.8 and E.9). Thus, if we require a certain number of detected rays to achieve some desired level of statistical noise, we would have to emit and trace over three times as many rays for Type 1 tracing (hence three times the computer time) as for Type 2 or 3. For the detector at $z_T = 15$ and Type 1 tracing, fewer than 2% of the emitted rays reached the target plane (Fig. E.10), whereas about 80% of the rays reached the target plane for Types 2 and 3 (Figs. E.11 and E.12). Getting the same number of rays on target would thus require emitting over 40 times as many rays (hence 40 times the computation time) for Type 1 as for Types 2 or 3.

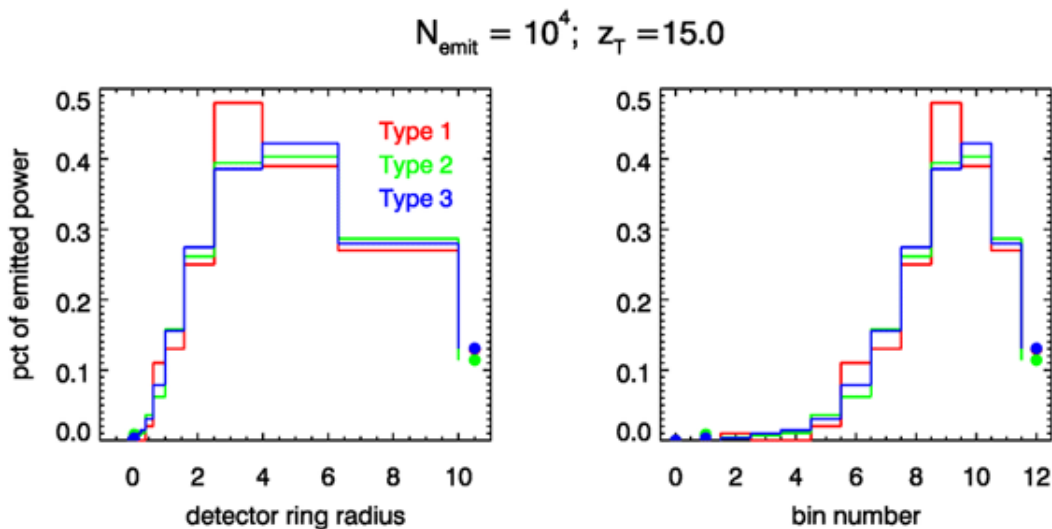


Figure E.14: Energy distributions for $z_T = 15$ and $N_{\text{emit}} = 10^4$, for the three tracing types. There are obvious differences in the distributions.

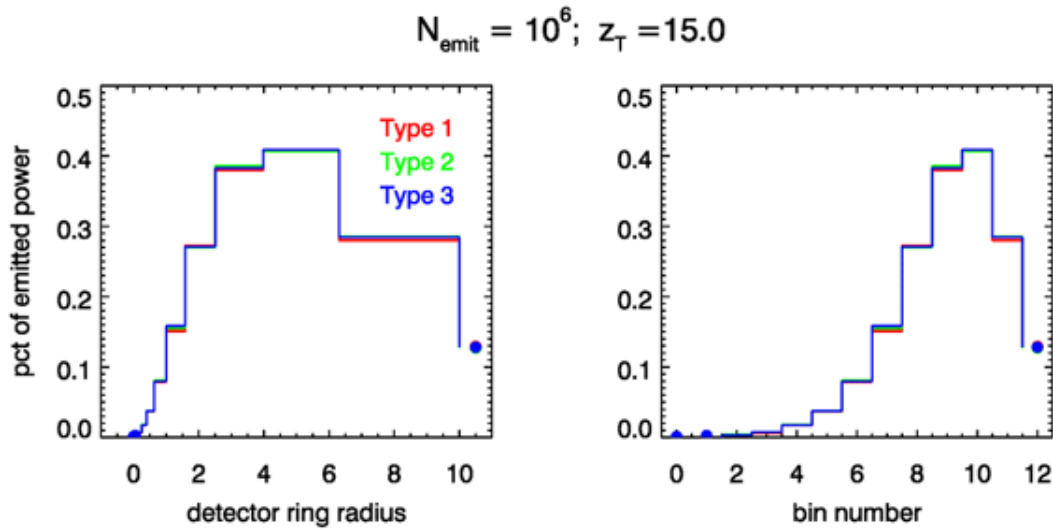


Figure E.15: Energy distributions for $z_T = 15$ and $N_{\text{emit}} = 10^6$, for the three tracing types. The differences seen in Fig. E.14 have almost disappeared.

We have now shown that several ways of tracing rays can be devised and that each gives the same distribution of power or energy at a detector some distance away from the source. However, an intelligent choice of the ray tracing algorithm can greatly reduce the needed computations. Moreover, the computation differences depend on the particular problem, e.g., on detector distance from the source as shown here (or on the IOPs, not shown here). However, we can do still more to reduce computation times, which leads us to the topic of variance reduction techniques, which begins in Section E.6. First, however, we consider how to estimate the errors in Monte Carlo simulations.

E.5 Error Estimation

This section shows how to estimate the errors in Monte Carlo simulations. General results from probability theory are illustrated with numerical examples.

To be specific, suppose that we need to estimate the fraction of power emitted by a light source that will be received by a detector. The answer of course depends on the water IOPs; the size, orientation, and location of the detector relative to the light source; the angular distribution of the emitted light; and perhaps on other things like boundary surfaces that can reflect or absorb light scattered onto the boundary. The numerical results to be shown below use the source and detector geometry shown in Fig. E.16. The source emits a collimated bundle of rays toward a detector that is $\tau = 5$ optical path lengths away and is 0.1 optical path lengths in diameter. The water IOPs have $a = 0.2 \text{ m}^{-1}$ and $b = 0.8 \text{ m}^{-1}$ so that $c = 1 \text{ m}^{-1}$ and one meter is one optical path length. The albedo of single scattering is $\omega_0 = 0.8$. The scattering phase function is the One-Term Henyey-Greenstein phase function of Eq. (6.12) with a scattering-angle mean cosine of $g = 0.8$.

In the simulations, N rays will be emitted from the source, each with an initial weight of $w = 1$. Most of those rays will miss the detector, as illustrated by the blue arrows in Fig.

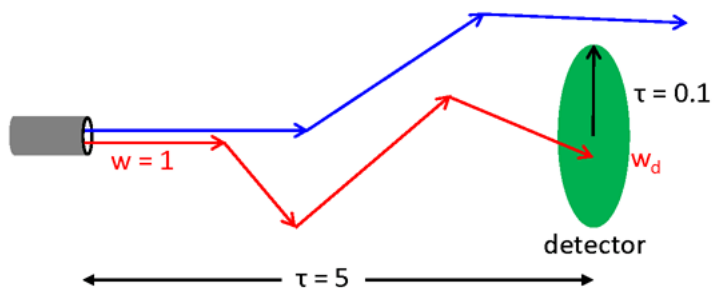


Figure E.16: Source and detector geometry used for numerical simulations.

E.16. However, some rays will hit the detector, as illustrated by the red arrows, at which time their current weight will be tallied to the accumulating total weight w_d received by the detector. After all rays have been traced, the Monte Carlo estimate of the fraction of emitted power received by the detector is simply $f_d = w_d/N$.

If we do only one simulation tracing, say, $N = 10^4$ ray bundles, then the resulting estimate of f_d is all we have. In particular, we have no estimate of the statistical error in the estimated f_d .

E.5.1 Probability Theory

To develop a quantitative error estimate for the result of a Monte Carlo simulation, we begin with a review of some results from basic probability theory. Recalling the notation introduced on in Section E.1, let $p_W(w)$ be the probability density function (PDF) for random variable W . For concreteness, let W be the ray bundle weight received by the detector, and lower case w represents a specific value of the weight W , e.g., $w = 0.72$. In the present example, $p_W(w)$ is the PDF that a ray strikes the detector with a weight w , $0 \leq w \leq 1$. A value of $w = 1$ means that the original ray of weight 1 has lost nothing to absorption between the source and the detector; a value of $w = 0.1$ means that the ray has lost 90% of its initial weight to absorption. Rays that miss the detector do not contribute to accumulating weight or power received by the detector and do not enter into the calculations below; they are simply wasted computer time. Note that we have no idea what mathematical form $p_W(w)$ has: it results from a complicated sequence of randomly determined ray path lengths and scattering angles.

For any continuous PDF $p_W(w)$ the *expected* or *mean value* of W is defined as

$$\text{mean}(W) \equiv \mu \equiv \mathcal{E}\{W\} = \int w p_W(w) dw, \quad (\text{E.10})$$

where \mathcal{E} denotes expected value and the integral is over all values for which W is defined. If the random variable is discrete, the integral is replaced by a sum over all allowed values of W . Similarly, the *variance* of W is defined as

$$\text{var}(W) \equiv \sigma^2 \equiv \mathcal{E}\{(W - \mu)^2\} = \int (w - \mu)^2 p_W(w) dw = \mathcal{E}\{W^2\} - [\mathcal{E}\{W\}]^2.$$

Note that if c is a constant, then

$$\mathcal{E}\{cW\} = c\mathcal{E}\{W\} \quad \text{and} \quad \text{var}(cW) = c^2 \text{var}(W).$$

Greek letters μ and σ^2 are used to denote the *true* or *population* mean and variance of a PDF.

Suppose that N_d ray bundles actually reach the detector. The total weight received by the detector is then given by the sum of these randomly determined weights:

$$S_{N_d} = \sum_{i=1}^{N_d} w_d(i),$$

where $w_d(i)$ is the weight of ray bundle i when it reached the detector.

Each of the N ray bundles emitted by the source and traced to completion is independent of the others. In particular, a different sequence of random numbers is used to determine the path lengths and scattering angles for each emitted bundle. Moreover, the underlying PDFs for ray path length and scattering angles (i.e., the IOPs) are the same for each bundle. The random variables are then said to be *independent and identically distributed* (iid), and S_{N_d} is said to be a random sample of size N_d of random variable W .

The linearity of the expectation (i.e., the integral of a sum is the sum of the integrals) means that for iid random variables such as W ,

$$\mathcal{E}\{S_{N_d}\} = N_d \mathcal{E}\{W\} = N_d \mu \quad \text{and} \quad \text{var}(S_{N_d}) = N_d \text{var}(W) = N_d \sigma^2. \quad (\text{E.11})$$

In the Monte Carlo simulation, the *sample mean*, i.e. the estimate of the average detected weight obtained by from the N_d detected ray bundles is

$$m_{N_d} \equiv \frac{1}{N_d} S_{N_d} = \frac{1}{N_d} \sum_{i=1}^{N_d} w_d(i).$$

Equation (E.11) now gives two extremely important results. First,

$$\mathcal{E}\{m_{N_d}\} = \frac{1}{N_d} \mathcal{E}\{S_{N_d}\} = \mu. \quad (\text{E.12})$$

That is, the expectation of the sample mean m_{N_d} is equal to the true mean μ . The sample mean is then said to be an *unbiased estimator* of the true mean of the PDF. Second,

$$\text{var}(m_{N_d}) = \text{var}\left(\frac{1}{N_d} S_{N_d}\right) = \frac{1}{N_d^2} \text{var}(S_{N_d}) = \frac{\sigma^2}{N_d}. \quad (\text{E.13})$$

Thus, the variance of the sample mean goes to zero as $N_d \rightarrow \infty$, that is, as more and more ray bundles are *detected*. In other words, ***the Monte Carlo estimate of the average power received by the detector is guaranteed to give a result that can be made arbitrarily close to the correct result if enough rays are detected.*** This result is known as *the law of large numbers*. Again, you can emit and trace all the rays you want, but if they don't hit the detector, they don't count.

It is often convenient to think in terms of the standard deviation, e.g., when plotting data and showing the spread of values. The standard deviation of the error in m_{N_d} is

$$s_{N_d} = \sqrt{\text{var}(m_{N_d})} = \frac{\sigma}{\sqrt{N_d}}. \quad (\text{E.14})$$

The dependence of the standard deviation of the estimate on $1/\sqrt{N_d}$ is a very general and important result. However, this “approach to the correct value” is very slow. If we want to reduce the standard deviation of the error in the estimated average power received by the detector by a factor of 10, we must detect 100 times as many rays. That can be computationally very expensive.

It is to be emphasized that result (E.13) that the variance of a sample mean equals the true variance divided by the sample size holds for any situation for which the individual samples are independent and identically distributed random variables.

Note, of course, that if we knew the PDF for the received power, $p_W(w)$, then we could simply evaluate Eq. (E.10) to get the desired true mean μ , and no Monte Carlo simulation would be required.

Finally, it must be remembered that the discussion here assumes that “all else is the same” when considering the number of detected rays. For example, we emit and trace more rays without changing the physics of the simulation. The section on importance sampling presents ways to increase the number of detected rays and thereby reduce the variance, but with a change in the physics that sometimes may invalidate the simple $1/\sqrt{N_d}$ dependence.

E.5.2 Numerical Examples

Numerical simulations were performed for the geometry and conditions described for Fig. E.16. For these simulations, tracing type 1 of the previous section was used. That is, ray bundles were traced until they either hit the target (still with weight $w = 1$) or were absorbed. For a given number N of emitted ray bundles, various numbers N_{run} of independent runs were done. That is, N rays were traced and the number N_d of detected ray bundles and their weights were tallied for each run. The fraction of emitted power received by the detector was then computed by the total detected weight for the run divided by N . Then another run was made with everything the same except that a different sequence of random numbers was used (i.e., each run was started with a different seed for the $\mathcal{U}[0, 1]$ random number generator used to determine path lengths and scattering angles).

The first set of simulations used $N = 10,000$ emitted ray bundles for each run, with $N_{\text{run}} = 10, 100, 1000$, and $10,000$ runs being made. Figure E.17 shows the distributions of the sample means m_{N_d} and other information for these four sets of run numbers. The upper left panel of the figure is for only $N_{\text{run}} = 10$ runs, or trials, or simulations. In this panel, the histogram shows that one run, or 0.1 of the total number of runs, gave an estimated fraction between 0.0088 and 0.0090 of the emitted power; two runs, or 0.2 of the total, gave a fraction between 0.0104 and 0.0106, and so on. As the number of runs increases, the estimates of the fraction of power received range from slightly less than 0.008 to slightly more than 0.014, with most estimates centering somewhere near 0.0108.

As the number of runs increases, something very remarkable happens: *the distribution of the fraction of the emitted power appears to be approaching a Gaussian or normal shape, even though the underlying PDF $p_W(w)$ is certainly not Gaussian.* This distribution can be thought of as the distribution of errors in the estimated mean of the distribution, or the distribution of $\mathcal{E}\{m_{N_d} - \mu\}$, where μ is the unknown true mean of the distribution $p_W(w)$. This approach to a Gaussian distribution is a consequence of the *central limit theorem*. *The central limit theorem states that the **sum** of a large number of independently distributed random variables with finite means and variances is approximately normally distributed regardless of what the distribution of the random variable itself may be.* This is

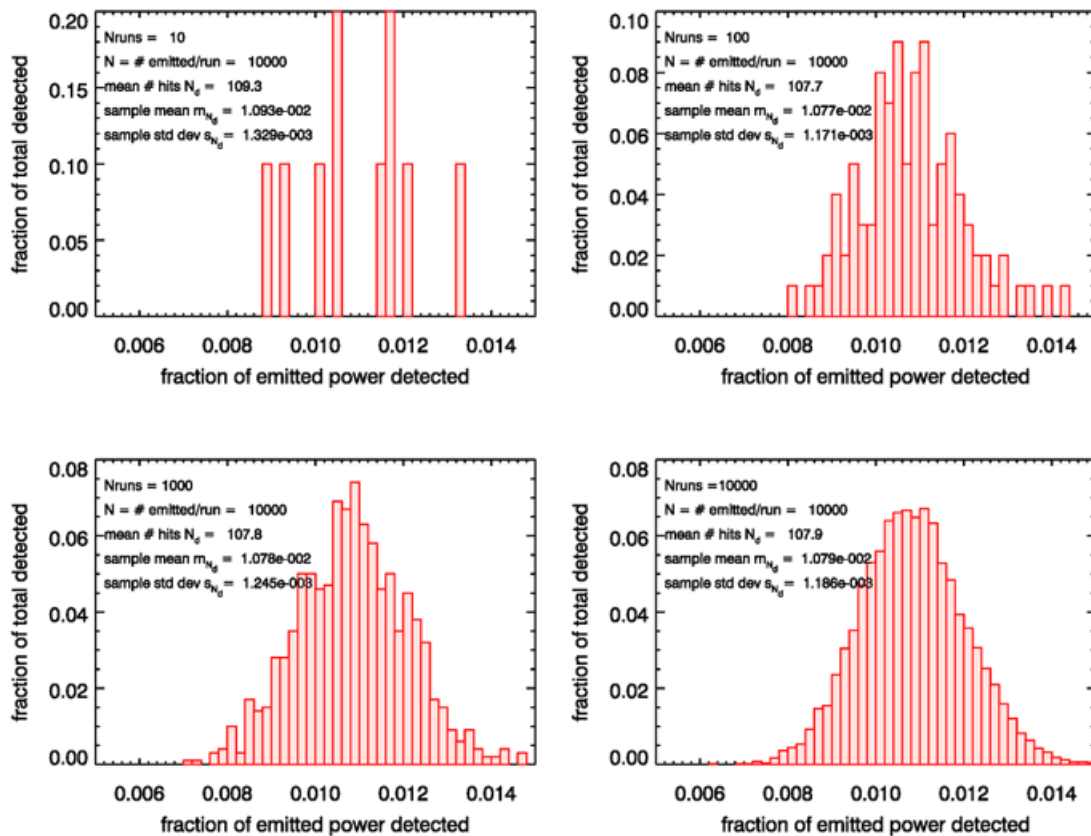


Figure E.17: Estimates of the fraction of detected power for four sets of runs with $N = 10^4$ emitted rays in each run.

one of the most profound results in probability theory. Indeed, it explains why so many natural phenomena tend to have a Gaussian shape. Phenomena as disparate as average student exam scores, noise in electrical circuits, daily water usage in a city, or the fraction of people who develop cancer can all result from sums of many individual contributions. Such sums then tend toward a Gaussian distribution as the number of individual contributions increases. The theorem was first proved for a specific PDF in 1733, but it was not proven to hold for all PDFs (having finite means and variances) until the early 1900's¹.

Figure E.18 shows the corresponding results for series of $N_{\text{run}} = 10, 100, 1000$, and 10,000 runs being made, but with each run now having $N = 100,000$ emitted rays. Now the spread in the estimated values is much less, from slightly less than 0.010 to about 0.012, again centering somewhere around 0.0108. Again, we see the approach to a Gaussian shape as more and more runs are made, but the Gaussian has a narrower width, i.e. less variance about the mean. The standard deviation of the sample estimates of the mean is usually called the “standard error of the mean.”

¹By the way, in spite of what you see in the tabloid press, there is nothing in probability theory called “the law of averages.” The central limit theorem, or perhaps the law of large numbers mentioned previously, is maybe the closest thing to the often invoked but mythical “law of averages.”

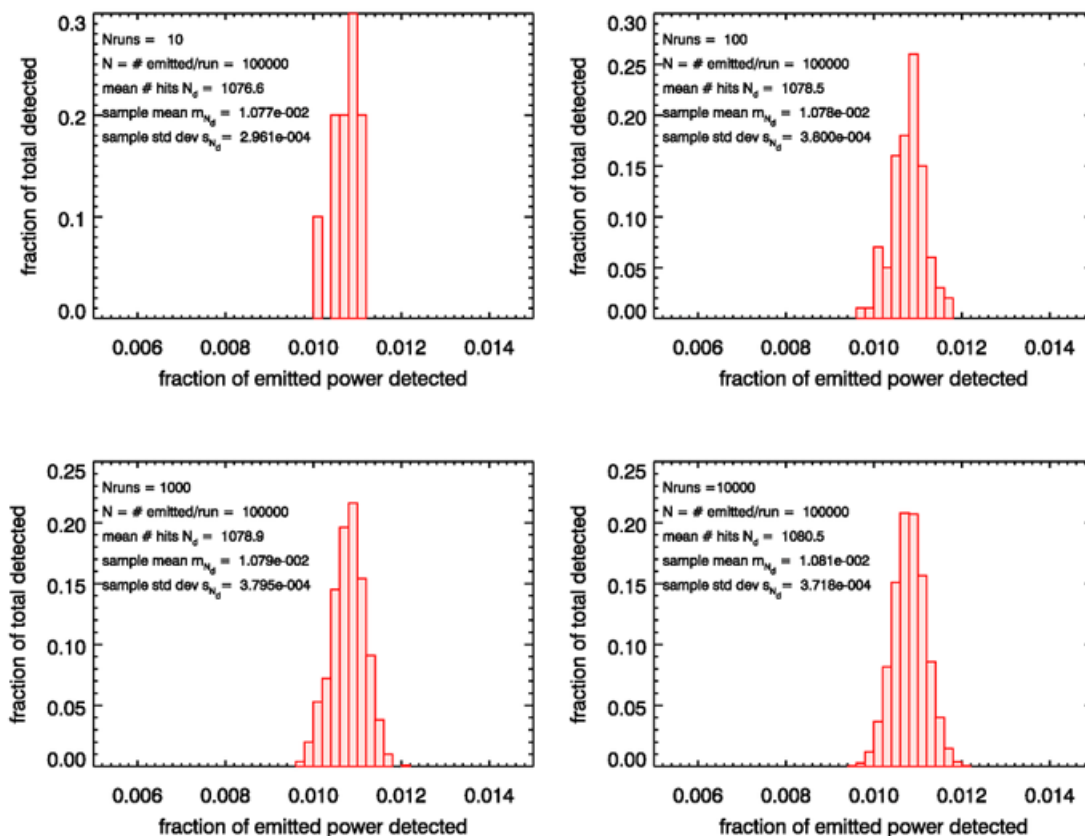


Figure E.18: Estimates of the fraction of detected power for four sets of runs with $N = 10^5$ emitted rays in each run.

The lower right panel of Fig. E.17 shows that the sample standard deviation for the case of 10,000 runs each with 10,000 emitted rays is $s_{N_d} = 1.186 \cdot 10^{-3}$ and the average number of detected rays is $N_d = 107.9$. The corresponding panel of Fig. E.18 shows $s_{N_d} = 3.718 \cdot 10^{-4}$ and $N_d = 1080.5$. The ratio of these sample standard deviations is

$$\frac{s_{N_d}(N_d = 107.9)}{s_{N_d}(N_d = 1080.5)} = \frac{1.186 \cdot 10^{-3}}{3.718 \cdot 10^{-4}} = 3.19.$$

The corresponding ratio of $\frac{\sigma}{\sqrt{N_d}}$ values is

$$\frac{\frac{\sigma}{\sqrt{N_d=107.9}}}{\frac{\sigma}{\sqrt{N_d=1080.5}}} = \sqrt{\frac{1080.5}{107.5}} = 3.16.$$

This nicely illustrates the dependence of the sample standard deviation, or the standard error of the mean, on the square root of the number of ray bundles detected, just as predicted by Eq. (E.14).

E.5.3 Error Estimation

In the present example of the fraction of emitted power received by a detector, the central limit theorem guarantees that the errors in the fraction of received power computed by many Monte Carlo runs approaches a Gaussian. We can thus use all of the results for Gaussian, or normal, probability distributions to estimate the errors in the Monte Carlo results.

It is often desirable to know the probability that the computed sample mean m_{N_d} is within some specified amount, say 1 standard deviation, of the (unknown) true mean μ . Conversely, we may want to compute the error range so that the sample mean is within that error range of the true mean with some specified probability. Such questions can be answered starting with the statement

$$\text{Prob}\{\mu - \beta s_{N_d} \leq m_{N_d} \leq \mu + \beta s_{N_d}\} = 1 - \alpha.$$

This equation states that the probability is $1 - \alpha$ that the sample mean is within a range βs_{N_d} of the true mean μ ; β is a fraction of the sample standard deviation s_{N_d} . In other words, the probability is $1 - \alpha$ that $\mu = m_{N_d} \pm \beta s_{N_d}$. The central limit theorem guarantees that, if the sample size is large enough, the deviation of the sample mean from the true mean, $m_{N_d} - \mu$, is approximately normally distributed:

$$\text{PDF}(m_{N_d} - \mu) \approx \frac{1}{\sqrt{2\pi}s_{N_d}} \exp\left\{-\frac{(m_{N_d} - \mu)^2}{2s_{N_d}^2}\right\}. \quad (\text{E.15})$$

Assuming that we have enough samples to get a good approximation to the normal distribution, the probability that m_{N_d} is greater than μ by an amount βs_{N_d} is then

$$\text{Prob}\{m_{N_d} - \mu \geq \beta s_{N_d}\} = \frac{1}{\sqrt{2\pi}s_{N_d}} \int_{\beta s_{N_d}}^{\infty} \exp\left\{-\frac{t^2}{2s_{N_d}^2}\right\} dt.$$

Letting $y = t/s_{N_d}$ gives

$$\text{Prob}\{m_{N_d} - \mu \geq \beta s_{N_d}\} = \frac{1}{\sqrt{2\pi}} \int_{\beta}^{\infty} \exp\left\{-\frac{y^2}{2}\right\} dy \equiv Q(\beta). \quad (\text{E.16})$$

The $Q(\beta)$ integral in the last equation cannot be performed analytically, but it is tabulated in probability texts, and software packages such as MATLAB and IDL have routines to compute it. It is also common to find tables and subroutines for

$$\Phi(\beta) \equiv \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\beta} \exp\left\{-\frac{y^2}{2}\right\} dy.$$

Note that $Q(\beta) + \Phi(\beta) = 1$.

Let us now apply Eq. (E.16) to various examples. We first compute the probability that the sample mean is within one standard deviation of the true mean. Letting $\beta = 1$, we compute the probability $\alpha/2$ that $m_{N_d} - \mu$ lies in the “right-hand tail” of the normal distribution beyond $\beta = 1$. This is the area shaded in red in Fig. E.19.

This probability is

$$\text{Prob}\{m_{N_d} - \mu \geq s_{N_d}\} = Q(1.0) = \frac{\alpha}{2}.$$

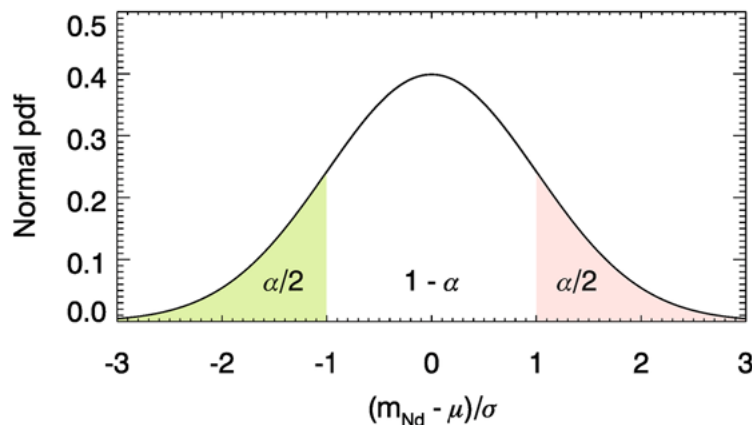


Figure E.19: The Gaussian or normal distribution of Eq. E.15

$Q(1) \approx 0.1587$, so $\alpha = 2Q(\beta) = 0.3174$. The Gaussian distribution is symmetric about the mean, so $Prob\{m_{N_d} - \mu \leq -s_{N_d}\}$ that $m_{N_d} - \mu$ lies in the left-hand (green-shaded) tail of the distribution also equals 0.1587. Thus the probability that $m_{N_d} - \mu$ does *not* lie in either tail of the distribution, i.e. that $\mu \leq m_{N_d} \pm s_{N_d}$ is $1 - \alpha/2 - \alpha/2 = 0.6826$.

For the simulations of Fig. E.17, the lower right panel shows values of $m_{N_d} = 0.0179$ and $s_{N_d} = 1.186 \cdot 10^{-3}$. Thus there is a roughly 68% chance that the true fraction of received power is within the range $0.01079 \pm 1.186 \cdot 10^{-3}$. For the corresponding run of Fig. E.18, which traced ten times as many ray bundles, the corresponding numbers are $0.01081 \pm 3.718 \cdot 10^{-4}$. Thus, in the last set of simulations, we are 68% certain that the sample mean is within about $\pm 3.4\%$ of the true mean.

Suppose we need the probability that we are within, say, 5% of the correct value. For the lower right simulation of Fig. E.18, 0.05 of 0.01081 is about 0.00054. From $0.00054 = \beta 3.178 \cdot 10^{-4}$ we get $\beta = 1.454$. $Q(1.454) = 0.0729 = \alpha/2$, so the probability of being within 5% is $1 - \alpha = 0.854$. If being 85% confident that the Monte Carlo estimate of the mean is within 5% of the true value is adequate for your application, then you are done. If you need to be 95% confident that you are within 5% of correct, then you need to continue tracing rays until you get enough rays on the target to reduce the sample variance to a value small enough to achieve the desired 95% confidence.

As a final example, we might ask how big is the error so that we can say that we are within that range with 90% certainly. We now set $1 - \alpha = 0.9$, and solve

$$Q(\beta) = \frac{\alpha}{2} = 0.05$$

Again, the inverse of $Q(\beta)$ is also tabulated. This equation gives $\beta = 1.645$. From the last panel of Fig. E.18 we then get $\beta s_{N_d} = 1.645 \cdot 3.718 \cdot 10^{-4} = 6.12 \cdot 10^{-4}$, so that $\mu = 0.01081 \pm 6.12 \cdot 10^{-4}$ with 90% confidence.

As a final caveat to this section, keep in mind that the central limit theorem says that the error becomes Gaussian as the number of samples, N_d in the present examples, becomes very large. How large is large enough depends on the particular problem and the user's accuracy requirement. In the present examples, 10,000 runs each with 10,000 or

more emitted rays, resulting in 100 or more detected rays for each run, gives distributions that visually appear close to Gaussian (the lower right panels of the preceding two figures). There are various ways to quantify how close a data distribution is to a Gaussian, but that is a topic for somewhere else. Just do a search on “normality tests.”

E.6 Importance Sampling

As we saw on the Error Estimation Section E.5, the standard error of the mean in a Monte Carlo estimate depends on $1/\sqrt{N}$, where N is the number of samples. In the present discussion, N is the number of rays reaching a detector. In addition, we have seen that rays that are traced but do not intersect the detector—i.e., do not generate a sample of the underlying PDF—do not contribute to the answer and are wasted calculations.

The general topic of *importance sampling* considers ways to generate and trace rays so that more rays reach the detector, thus increasing N and thereby reducing the statistical error in the estimated quantity, while simultaneously reducing the number of wasted rays.

E.6.1 Theory of Importance Sampling

Let $PDF(x)$ be the probability distribution function for variable x . The basic idea of importance sampling is to “over sample” the parts of the PDF that are most “important,” i.e., those parts of the PDF that send rays in the direction of the detector. Those parts of the PDF that send rays in directions away from the detector are under sampled. Thus more rays are sent towards the detector, which increases the number detected and therefore reduces the variance in the estimated quantity, and fewer are sent in directions that never reach the detector. However, this process samples the PDF in a biased or incorrect manner compared to the physical process being studied. To account for the biased sampling of the PDF, the weight of each ray is adjusted to keep the final answer correct.

Mathematically, importance sampling is described as follows. The mean of any function $f(x)$ of random variable x is by definition

$$\langle f \rangle = \int f(x) PDF(x) dx,$$

where the integration is over the full range of x (e.g., over 0 to π if x is the polar scattering angle). This can be rewritten as

$$\begin{aligned} \langle f \rangle &= \int f(x) \frac{PDF(x)}{PDF_b(x)} PDF_b(x) dx \\ &\equiv \int f(x) w(x) PDF_b(x) dx = \langle fw \rangle_b. \end{aligned}$$

$PDF_b(x)$ is the biased PDF used to generate random values of x , and $w(x)$ is the weight given to each biased sample of x . The weight $w(x)$ is just the ratio of the unbiased to the biased PDFs. *Note that the estimate of the mean of f when sampled by the correct, unbiased PDF is the same as the estimate of the weighted function, fw , when sampled with the biased PDF.* The estimate $\langle f \rangle = \langle fw \rangle_b$ is thus unbiased, even though PDF_b is biased. Because the sampling is done using a biased PDF, importance sampling is sometimes called biased

sampling. However, the estimate of $\langle f \rangle$ is still unbiased, even though the sampling uses a biased PDF. The term importance sampling is therefore preferred in recent literature.

If the biased PDF is chosen well, it will increase the number of detected rays and thereby reduce the error in the estimate. However, each detected ray will have a smaller weight, so that the product of more detected rays times a smaller weight for each remains the same as for a smaller number of higher weighted rays that would be generated by the original PDF. As we will see in the following numerical examples, this basic idea often works well. However, in practice, the biasing can sometimes be pushed to extremes and actually increase the error in the estimate.

E.6.2 Example for an Embedded Point Source

The first numerical example is based on [Gordon \(1987\)](#). This paper provides a good starting point for examining in detail the advantages and pitfalls of importance sampling.

Gordon was interested in computing the irradiance pattern of upwelling light at the sea surface generated by an isotropic point source at some depth in the water. Figure [E.20](#) shows the geometry of his problem. The left panel of the figure shows an isotropic point source at depth z emitting rays equally in all directions. For an isotropic source, half of all rays are emitted in downward directions. Few of those rays will be scattered in directions that eventually reach the sea surface. The same holds true for many rays emitted in upward, but nearly horizontal directions. Such rays are illustrated by the red arrows. Only rays emitted in nearly straight upward directions are likely to reach the surface and contribute to the estimate of upwelling irradiance $E_u(r)$ as a function of radial distance r away from the source location. The green arrow illustrates such a ray.

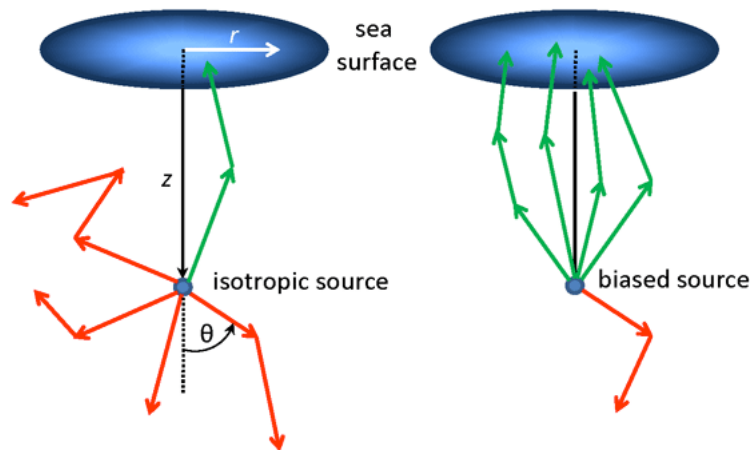


Figure E.20: Geometry of Gordon's simulation of irradiance at the sea surface due to an imbedded isotropic point light source in the water. The left panel illustrates the isotropic source, with only a few rays (green arrow) reaching the sea surface. The right panel illustrates the biased source, with most rays being emitted into upward directions.

To increase the number of rays emitted almost straight upward, Gordon chose a biased PDF for the probability of emission of a ray at polar angle θ , measured from $\theta = 0$ in the nadir direction (z is measured positive downward from the sea surface, and θ is measured

from the $+z$ direction). This isotropic emission is mathematically the same as isotropic scattering. Recall from Section E.3 that the PDF for polar scattering angle ψ is $2\pi\tilde{\beta}\sin\psi$. Taking ψ for the polar angle of emission θ and recalling that the phase function for isotropic scatter is $\tilde{\beta} = 1/4\pi$ shows that the PDF for the polar emission angle is $PDF(\theta) = \frac{1}{2}\sin\theta$. Gordon wished to have a biased PDF $PDF_b(\theta)$ that would emit more rays into upward directions $90\text{ deg} < \theta \leq 180\text{ deg}$. There are many functions that could give such a behavior, but the one Gordon chose is

$$PDF_b(\theta) = \frac{\sqrt{1-\epsilon^2}}{\pi(1+\epsilon\cos\theta)}, \quad (\text{E.17})$$

where $0 \leq \epsilon < 1$ is a parameter to be determined as described below. The weighting function is then

$$w(\theta) = \frac{PDF(\theta)}{PDF_b(\theta)} = \frac{\pi}{2\sqrt{1-\epsilon^2}}(1+\epsilon\cos\theta)\sin\theta.$$

(Gordon did not say why he chose this particular function, nor did he give the value of ϵ used in his calculations.)

The right panel of Fig. E.20 illustrates biased emission, with many more rays being emitted into upward directions and reaching the sea surface, and relatively few rays being emitted into downward directions.

The red curves in Fig. E.21 show the unbiased emission function $PDF(\theta) = \frac{1}{2}\sin\theta$, for which each emitted ray is given an initial weight of $w = 1$. The other curves show the biased functions $PDF_b(\theta)$ and the angle-dependent initial weights for several values of ϵ . Note that as ϵ approaches 1, most rays are emitted at angles near $\theta = 180\text{ deg}$, i.e. in near-zenith directions. However, rays emitted at angles near 180 deg are given very small weights, whereas the relatively few rays emitted in roughly horizontal and downward directions are given large weights (except for rays emitted near $\theta = 0$).

Monte Carlo simulations were performed for an unbiased isotropic source and for biased sources with various ϵ values. Figure E.22 shows the results for a point source at $\tau = 5$ optical depths below the surface and for water IOPs defined by a Fournier-Forand phase function with a backscatter fraction of 0.02 and an albedo of single scatter of $\omega_o = 0.8$. Independent runs were done for isotropic emission and for values of $\epsilon = 0.8$ and 0.99 in Eq. (E.17). Each run had 10^6 rays emitted by the source. The rays reaching the sea surface were tallied for a ring (or “bullseye”) detector centered above the source and having equally spaced differences in ring diameters of $\Delta r = 1$ in optical distance.

The left panel of the figure shows the number of rays received by each detector ring. For the isotropic source, at most about 40,000 rays were received by any ring (near radius = 5), and fewer than 10,000 rays were received by the innermost ring. As ϵ increases, the rays are emitted in a more and more vertical direction, and the numbers of rays received by the inner rings increase. The innermost ring received about 94,000 rays for $\epsilon = 0.8$ and 270,000 rays for $\epsilon = 0.99$. We thus naively expect that the variance of the estimated irradiance in the innermost ring is improved by a factor of $\sqrt{270,000/10,000} \approx 5$ with $\epsilon = 0.99$, compared to unbiased isotropic emitting. Or, conversely, we can get the same number of detected rays for one-fifth of the emitted rays, i.e., for one-fifth of the computer time.

However, there is no free lunch. The outer rings, greater than about 5 or 6 optical distances in these simulations, received fewer rays with the biased emission than for isotropic

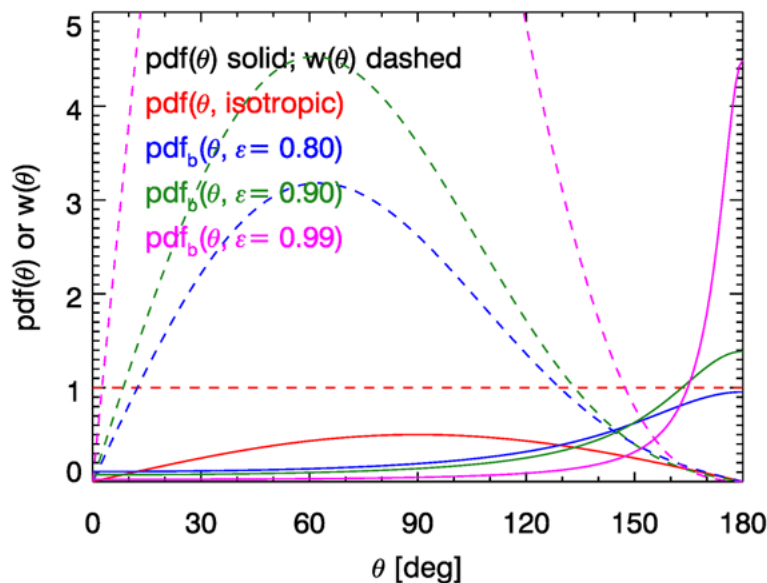


Figure E.21: Gordon's probability distribution functions for polar angle of ray emission (solid lines) and the corresponding ray packet weights for selected ϵ values (dashed lines).

emission. Thus the variance in the estimates of irradiance for the outer rings actually increases when using a biased source emission. This is simply because the biased near-zenith emission and subsequent near-forward scattering sends most of the rays towards the inner rings when ϵ nears 1.

The middle panel of the figure shows the fraction of the emitted power that is detected for each ring. Although the *numbers* of rays received by the rings varies according to the value of ϵ , the power received does not (except for statistical noise, which is almost unnoticeable in this figure). The right panel shows the irradiance $E_u(r)$ received in each ring normalized to the total power emitted by the source, Φ_{emit} . This is just the values of the middle panel divided by the area of each ring. This panel corresponds to Fig. 6 in Gordon's paper, except that the IOPs are different.

There is more to be learned from this simple example. Suppose we are interested only in the magnitude of the irradiance received at the ocean surface directly above the source. We might then consider only the power received by the innermost ring of the detector seen in the previous figure, i.e. detector radius ≤ 1 in those plots.

Consider the variance in the estimated fraction of the emitted power detected (i.e., of the surface irradiance, after dividing by the constant collection area) by the innermost ring of the previous figure, as a function of ϵ . Simulations were performed for isotropic emission and for values of ϵ from 0.0 to 0.999 (note that $\epsilon = 0$ in Eq. (E.17) is not the same as isotropic emission). Each simulation comprised 100 runs with 10^4 photo packets emitted for each run. The red dots in Fig. E.23 show the standard deviation in the estimated value of the fraction of emitted power received by the innermost ring as computed from the values in each of the 100 runs. The open red circles show the corresponding results from an independent set of 100 runs, which started with a different seed for random number generation. Those two sets of points give a qualitative idea of the amount of statistical

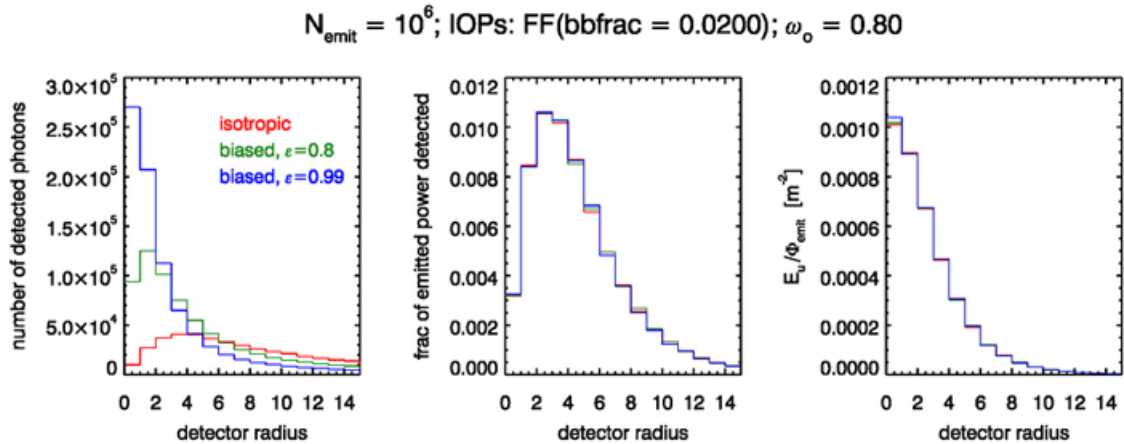


Figure E.22: Comparison of ray counts (left), fraction of emitted power detected (middle), and normalized irradiance (right) for a point source at 5 optical depths. Quantities are binned into concentric rings of differing by 1 optical depth in radius.

noise in these results. For isotropic emission an average of only 99 rays reached the detector out of the 10,000 emitted in each run (for the first set of runs with the first seed; for the second set the average was 101). The mean of the 100 runs was $3.257 \cdot 10^{-3}$ and the standard deviation was $3.493 \cdot 10^{-4}$. When biased emission is used, the number of detected rays increases with ϵ as shown by the green dots and right-hand axis in the figure. The standard deviation decreases with increasing ϵ up to values around $\epsilon = 0.8$ or 0.9 . However, beyond $\epsilon = 0.9$ the standard deviation begins to increase, and even becomes greater than for isotropic emission when $\epsilon = 0.999$, even though almost 4,000, or 40% of the emitted rays reach the detector. This seems to contradict the idea that increasing the number of detected rays reduces the variance.

It was already mentioned on the previous section on error estimation that the dependence of the standard deviation of an estimate on $1/\sqrt{N}$ assumes that “all else is the same” in the simulation. In the present case, we are changing the physics of the simulation (namely the emission pattern of the source) by using a biased source, in which case all else is not the same. As was shown in the theory section, this change in physics does not affect the estimated mean value, but it does affect the variance of that estimate over many runs. The use of a biased emission pattern does increase the number of detected rays and thereby reduce the standard deviation of the estimated quantity, but only up to a point. When ϵ is very close to 1, only a very few rays are emitted in downward directions. However, those rays can have very large weights: for $\epsilon = 0.999$ the maximum weight is 45.6. rays emitted within 10 deg of the zenith ($\theta > 170$ deg in Fig. E.21) have weights less than 0.07, and those emitted within 5 deg of the zenith have weights less than 0.008. Thus the downward-emitted rays can have weights hundreds or even thousands of times larger than the vast majority upward-emitted rays. The large numbers of low-weight detected rays do reduce the error in the estimate. However, when ϵ is very close to 1, one of the downward-emitted but very-large-weight rays may occasionally be backscattered in just the right direction to hit the detector and make a large contribution to the fraction of the emitted power detected. Such rays are few in number, so their contribution fluctuates from simulation to

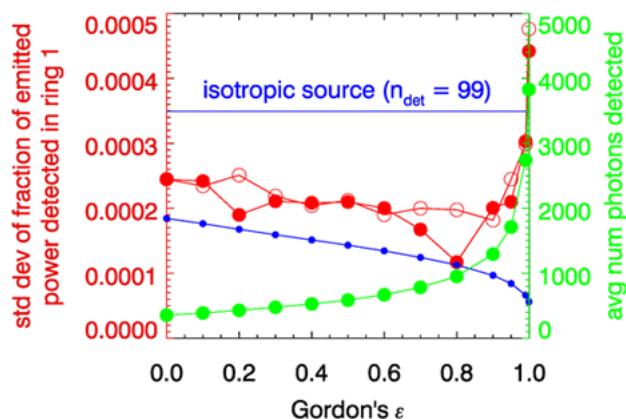


Figure E.23: Variance in the estimated fraction of emitted power detected by the innermost ring. The red dots are one set of 100 runs; the open circles are an independent set of runs. The blue dots are the theoretical reduction in the standard deviation based on a $1/\sqrt{N}$ dependence. The green dots are the numbers of detected rays from the first set of runs; the numbers from the second set are almost identical.

simulation, and the difference of even a few rays may significantly changed the estimate. These rare rays then increase the standard deviation of the computed quantity, and the benefits of the biased source emission are lost.

The blue dots in Fig. E.23 show the reduction in the standard deviation as expected due only to the increased numbers of detected rays, i.e. a dependence on $1/\sqrt{N_\epsilon/N_{\text{iso}}}$. The actual improvement is less than the value that would be obtained if the runs were just emitting and detecting more rays for an isotropically emitting source (i.e., for the “all else is the same” situation). Nevertheless, the reduction in the standard deviation is roughly a factor of two, which is significant for many applications.

Figure E.24 shows the numbers of detected rays binned by the weights of the detected rays for three of the runs, and for $\epsilon = 0.999$ and 0.8. For $\epsilon = 0.999$, the left panel shows that most of the detected rays have weights in the range of 10^{-5} to 0.1. However, a very few (2, 4 and 5 ray packets in these particular runs) have weights between 1 and 10. One ray with a weight of 1 contributes as much to the estimated irradiance as 1000 rays with a weight of 10^{-3} . The numbers of medium-weight rays are stable for the different runs, but the numbers of the highest-weight rays fluctuate greatly. The fluctuation in the number of high-weight rays begins to increase the variance for ϵ values very near 1. The numbers of very-low-weight rays also varies greatly from one run to the next, but those fluctuations have a negligible effect on the variance because the weights are so small. The right panel of this figure shows that for $\epsilon = 0.8$, most of the detected rays have weights in the range of 0.001 to 0.1, the numbers of rays in each bin is almost the same for each run, and there are no rays with weights greater than 1. Thus $\epsilon = 0.8$ gives a stable number of medium-weight rays and a reduction in the standard deviation of the estimated detected weight, as desired. Note that in each case the estimate of the fraction of power received is $(3.27 \pm 0.01) \cdot 10^{-3}$, so the estimated mean is almost independent of ϵ .

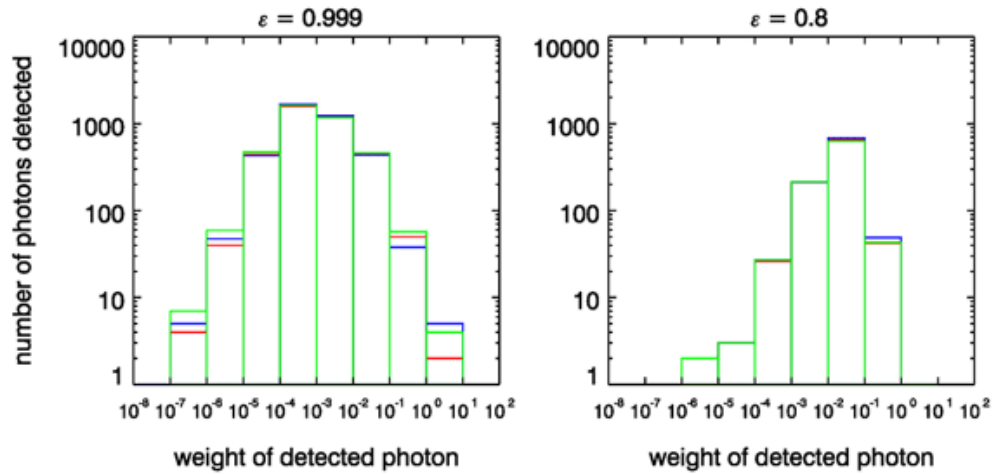


Figure E.24: Numbers of rays detected in various weight bins for three runs, which are shown by different colors.

Figure E.25 gives yet another view of this situation. The left panel shows the histogram of the 100 estimates of the fraction of emitted power detected by the innermost ring for isotropic emission. The middle panel shows a tighter histogram for $\epsilon = 0.8$, i.e., a reduced error in the estimate. The right panel shows the histogram for $\epsilon = 0.999$; the spread is now wider and the error in the estimate has increased.

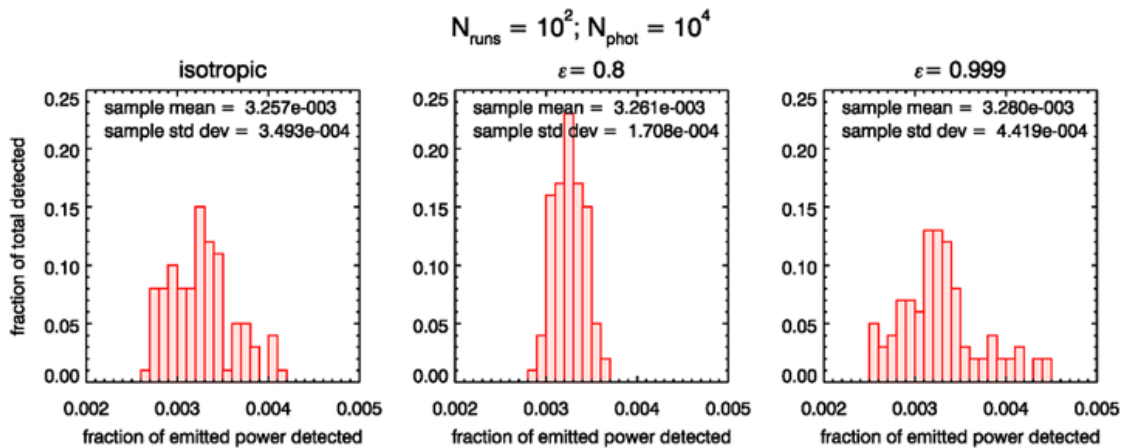


Figure E.25: Histograms of the 100 estimates of the fraction of power detected by the innermost ring, for isotropic emission and for $\epsilon = 0.8$ and 0.999 .

The behavior of the estimated error when doing importance sampling can be succinctly summed up as *the tails of the distribution matter*. That is to say, importance sampling can in many cases give more detected rays and a corresponding decrease in the error of the estimated quantity. However, the biasing can be pushed too far. In the present case, if ϵ is too close to 1, the rare but high-weight rays—those in the high-weight tail of the distribution of ray weights in Fig. E.24—can overpower the reduction of variance obtained

by the numerous but low-weight rays in the main part of the distribution of ray weights.

There is no general rule for determining how much biasing can be used or what value of a parameter like ϵ gives the minimum error in the estimated quantity. However, the broad near-minimum seen in Fig. E.23 shows that any value of $\epsilon \leq 0.9$ will give a better estimate than naive isotropic emission, and fortunately the exact value used is not critical. A suitable value of a biasing parameter such as ϵ must be determined by numerical experimentation for a particular class of problems. However, this effort is worthwhile if many simulations must be made and computer time or accuracy are critical.

E.6.3 Example for Backscattering

Now consider an example of using importance sampling in the simulation of backscattering, as might be needed for the design and evaluation of an instrument to measure the backscatter coefficient b_b . This example is inspired by the Monte Carlo simulations used in Gainusa Bogdan and Boss (2011), who did not use importance sampling.

The generic geometry for simulating a backscattering sensor is shown in Fig. E.26. This figure illustrates a source emitting ray packets in a collimated beam. The distance a ray packet travels between interactions with the medium is randomly determined using the ray path length algorithm described in Section E.2. Those ray packets are then scattered according to the chosen phase function $\tilde{\beta}(\psi, \alpha)$, where ψ is the polar scattering angle and α is the azimuthal scattering angle. Ray tracing is done by the “Type 2” technique of Section E.4. That is, at each scattering, the initial ray weight $w = 1$ is multiplied by the albedo of single scattering $\omega_o = b/c$, where b is the scattering coefficient and c is the beam attenuation coefficient. This accounts for energy loss due to absorption.

Oceanic phase functions are highly peaked in the forward scattering direction. Thus most ray packets undergo multiple forward scatterings and continue to travel away from the source and detectors. The green arrows in Fig. E.26 illustrate such a ray trajectory. Only rarely will a ray be backscattered in just such a manner that it is eventually detected, as illustrated by the red arrows in the figure. For typical oceanic conditions, only 0.5% to 3% of rays are backscattered in any given interaction. For a spatially small detector, very few of the backscattered rays will actually intersect the detector. The end result is that almost every ray emitted from the source is wasted computation because the ray never reaches the detector.

These simulations can be improved by use of importance sampling, as follows. The ray packets are emitted by the source according to whatever distribution is chosen (e.g., an idealized collimated point source or a beam profile and distribution of emitted directions that describes a particular light source such as an LED). Those rays travel an initial distance determined by the beam attenuation coefficient and the random number drawn. Then, *on the first scattering only*, the scattered direction is determined using a biased phase function $\tilde{\beta}_b$ that gives an increased number of backscattered rays. This “reverses” many of the initial rays, all of which are traveling away from the detector. Subsequent scatterings then use the normal phase function for the water body. This gives an increased number of ray packets traveling in the general direction of the detector, and thus an increased number of detected rays. This is illustrated in Fig. E.27. The green arrow is the initial ray emitted by the source. The blue arrow is the first-scattered ray, whose direction is determined using the biased phase function. The red arrows are subsequent scatterings of the ray.

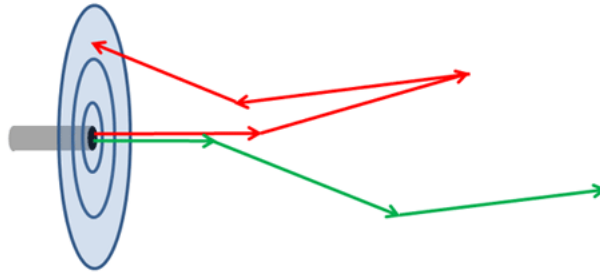


Figure E.26: Illustration of a source emitting a collimated beam at the center of a circular detector array. The green arrows show a ray undergoing multiple forward scatterings and traveling away from the detector. The red arrows show a ray being backscattered into the detector.

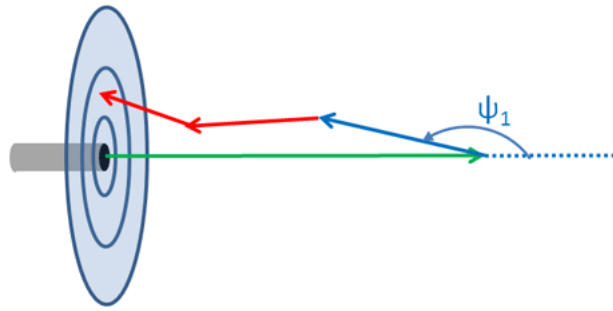


Figure E.27: Illustration of an emitted ray (green arrow) being backscattered by a biased phase function at the first scattering event (blue arrow). ψ_1 is the scattering angle drawn from the biased phase function. After the first scattering, the normal phase function is used (red arrows).

In oceanic simulations, the unbiased PDF is highly forward scattering. To reverse some of the initial rays, we use a biased PDF that enhances backscatter. The One-Term Henyey-Greenstein (OTHG, Eq. 6.12) phase function with asymmetry parameter $g = \langle \cos \psi \rangle$,

$$\tilde{\beta}_{\text{OTHG}}(\psi) = \frac{1 - g^2}{4\pi(1 + g^2 - 2g \cos \psi)^{3/2}},$$

gives a convenient analytical phase function to use for $\tilde{\beta}_b$. The parameter g plays the same role as ϵ in the previous example. A value of $g = 0$ gives isotropic scattering (50% backscatter); a negative g gives more backscatter than forward scatter. Numerical investigations show that the results are not sensitive to the exact form of the biased phase function, so long as the biasing is not pushed to extremes (such as using $g = -0.9$ in the OTHG, which gives 98% backscatter). Just as in the previous example, if the biasing is pushed too far, the small-angle forward-scattered rays have very large weights. The rare one of these rays that is subsequently backscattered (by normal scattering) into the detector gives a large fluctuation in the computed mean, which offsets the reduction in variance obtained by detecting many more low-weight rays.

The green curve of Fig. E.28 shows a Fournier-Forand phase function with a backscatter fraction of $\tilde{b}_b = 0.0183$; this phase function is typical of ocean water and gives a good fit to the Petzold average-particle phase function. The red curve shows a OTHG phase function with $g = -0.3$, which gives a backscatter fraction of $\tilde{b}_b = 0.714$. When used as the biased phase function $\tilde{\beta}_b$, this OTHG gives 40 times more backscattered rays at the first scattering event as does the Fournier-Forand phase function. The blue curve shows the corresponding weighting function $w(\psi) = \tilde{\beta}_{\text{FF}}(\psi)/\tilde{\beta}_{\text{OTHG}}(\psi)$. Note that the biased backscattered ray packets are given weights less than 0.1, whereas the very small angle forward scattered rays can have weights as large as 1000.

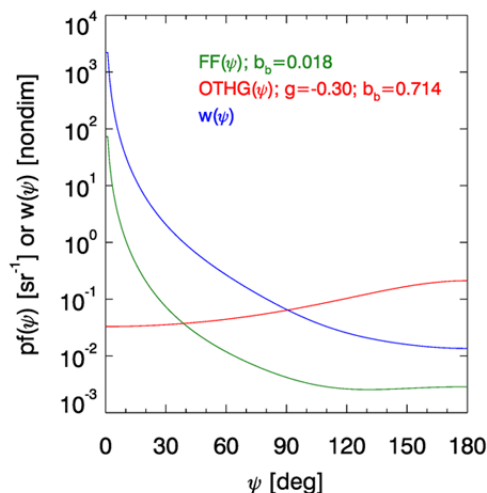


Figure E.28: Example phase functions used for normal scattering and biased backscattering. FF is a Fournier-Forand phase function as used for normal scattering; OTHG is a One-Term Henyey-Greenstein phase function used for biased backscatter; and w is the weighting function used to weight the rays at the first scattering.

Figure E.29 and Table E.2 show example simulations with and without biased first scatterings. The target was an annular (“bullseye”) detector like the one in Fig. E.26, but with five rings each of 1 cm radial width. This is similar to the proposed sensor design in Gainusa Bogdan and Boss (2011). The water properties were defined by a Fournier-Forand phase function with a backscatter fraction of 0.0183, which is shown in Fig. E.28. The absorption coefficient was $a = 0.2 \text{ m}^{-1}$ and the scattering coefficient was $b = 0.8 \text{ m}^{-1}$. The albedo of single scattering is then $\omega_o = 0.8$, which is typical of ocean waters at blue or green wavelengths. Three simulations were done: one without biased scattering and two with first scatterings biased with a OTHG phase functions with either $g = +0.3$ or $g = -0.3$. The OTHG with $g = +0.3$ has a backscatter fraction of $\tilde{b}_b = 0.286$, and $g = -0.3$ has $\tilde{b}_b = 0.714$. The left panel of the figure shows the numbers of rays received by each detector ring; the right panel is the percent of emitted power received by each ring. Note that 12.9 times more rays reach the detector when biased first scattering is used with $g = +0.3$ (green curve), and 52.6 times more for $g = -0.3$ (blue curve) than for unbiased scattering (red curve). However, the fraction of emitted power received by the detector (over all 5 rings) remains the same, 0.0645%, to within less than 0.5% percent of Monte Carlo noise.

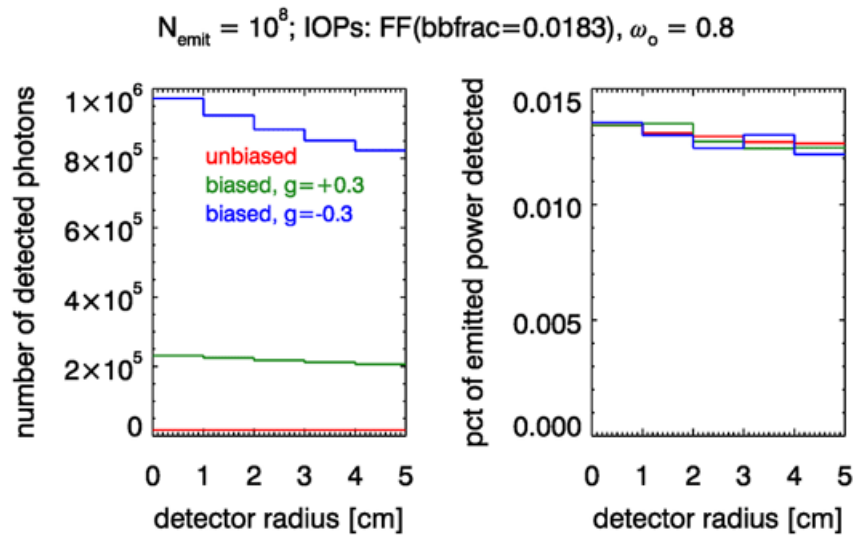


Figure E.29: Comparison of the number of detected rays (left panel) and the percent of detected power (right panel) for biased and unbiased first scatterings. Each run had 10^8 ray packets emitted by the source as a collimated beam.

First Scattering	N_{det} (all rings)	Increase Factor	% Emitted Power Detected (all rings)
unbiased	84,645	—	0.0648
biased, $g = +0.3$	1,093,114	12.9	0.0645
biased, $g = -0.3$	4,451,875	52.6	0.0642

Table E.2: Comparison of detected power for biased and unbiased first scatterings. Each run had 10^8 ray packets emitted by the source as a collimated beam. N_{det} is the number of ray packets that reached the detector (total for all 5 detector rings).

Conversely, for a given number of rays detected, biased first scattering allows that number to be detected with fewer rays being emitted and traced to completion, i.e., with less computer time. For unbiased scattering, running the Monte Carlo code with unbiased scattering until 100,000 rays reached the detector (total over all 5 rings) required the emission of $1.18 \cdot 10^8$ ray packets and a total run time of 6632 sec (on a 2.4 GHz PC). However, use of biased scattering with $g = -0.3$ required emission of only $2.23 \cdot 10^6$ rays, and a total run time of 105 seconds. The percent of emitted power that was detected was the same to within 3% in both cases, but the run time savings was a factor of 63.

E.7 Backward Monte Carlo Ray Tracing

Backward Monte Carlo (BMC) ray tracing is one of the conceptually simplest, yet most powerful, techniques for improving the accuracy and decreasing the computer times in Monte Carlo simulations. For simulation of a perfectly collimated source (e.g., an idealized laser beam) it is the only simulation technique available. The BMC technique is outlined in this section and one numerical example is shown. The mathematical details are presented in Appendix F.

E.7.1 The Essence of Backward Monte Carlo Simulation

The essential idea of backward Monte Carlo (BMC) simulation is seen in Fig. E.30. A detector is shown at some location in the water, oriented in some direction, and having a field of view and response as a function of the off-axis angle. This could be a radiance sensor with a narrow field of view (as illustrated), or it could be a plane irradiance sensor with a hemispherical field of view and a cosine response, and so on. The red arrows in the figure represent a light ray coming from the Sun or sky, passing through the sea surface, scattering within the water, and reflecting off of the bottom. Note that none of these red rays enters the detector. The computation of the red rays is therefore wasted. Undetected rays are not a problem in Nature because an enormous number of rays (photons) is available for detection in the near-surface region of a Sun-lit ocean, so that undetected rays are of no concern. (Recall from Section 1.3.3 the estimate of 10^{21} photons $\text{m}^{-2} \text{s}^{-1}$ near the sea surface on a sunny day.) Numerically, however, we can generate only a relatively small number of rays because of limited computer power. Thus in computations, every ray must, if possible, be made to contribute to the estimate of the radiometric variable of interest.

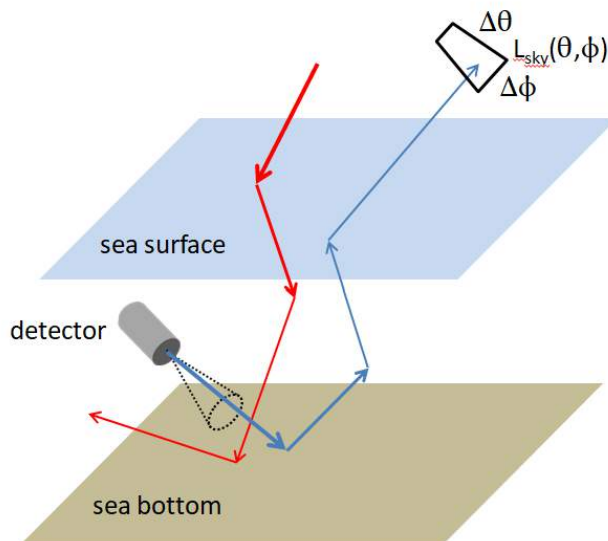


Figure E.30: The essence of backward Monte Carlo simulation.

The [Helmholtz Reciprocity Principle](#) (which is based on the time-invariance of Maxwell's Equations) states that if a ray generated at some point A reaches some other point B (after any number or kind of interactions), then a ray leaving B in the opposite direction will

retrace the path of the original ray back to A. Furthermore, if a fraction f of the energy leaving A reaches B, then the same fraction of the energy of the reversed ray from B will reach A.

Therefore, rather than trace rays from the sky to the detector, as Nature does, we can trace rays *from the detector to the sky*. Rays are emitted from the detector with an angular distribution determined by the angular response of the detector. For example, to simulate a plane irradiance detector with a cosine response, we emit rays from the detector with a corresponding cosine angular emission pattern. If in Nature a ray coming from some particular sky direction still has 2% of its original energy left when it reaches the detector, then the reversed ray emitted from the detector and reaching the sky in the same direction will have 2% of its emitted energy left when it reaches the sky.

In BMC simulation we therefore

1. Emit a large number of rays from the sensor in an angular pattern determined by the radiometric variable of interest.
2. Each emitted ray has an initial weight of $w = 1$.
3. Trace each ray to completion using standard Monte Carlo ray tracing techniques as described in the previous sections and in Appendix F.
4. Tally the fraction of detector-emitted energy accumulated in each angular bin of a set of (θ, ϕ) sky bins.
5. After computing the fraction of energy reaching each sky bin, multiply that weight by the sky radiance of the bin and apply scale factors for the variable being computed (Table F.2). This gives the fraction of the sky radiance from that bin that reaches the detector.

The blue arrows in Fig. E.30 represent a ray emitted by the detector and eventually reaching the sky within some directional bin of size $\Delta\theta, \Delta\phi$, centered on the nominal direction (θ, ϕ) , where the radiance is $L_{\text{sky}}(\theta, \phi)$.

This procedure does not guarantee that every detector-emitted ray will eventually reach the sky. Some rays may go off in directions that never cross the sea surface, and those computations are wasted. However, by emitting rays from the sensor, any ray that does reach the sky contributes to the estimation of the radiometric variable. There are no wasted rays coming from the sky and missing the sensor, as is the case for the red rays in Fig. E.30.

The mathematical details of how these calculations are actually carried out can be messy; they are therefore presented in Appendix F.

E.7.2 Example BMC Simulations

BMC simulations have been used extensively to compute underwater light fields in three-dimensional geometries. A few examples are

- **Gordon (1985)** used BMC to evaluate the effect of ship hulls on underwater irradiances. (This is the paper that brought the BMC technique to the attention of the oceanography community.)

- Mobley and Sundman (2003) used BMC in a study of the effects of 3-D bottoms (both sloping bottoms and level bottoms with a non-uniform bottom reflectance) on reflected, upwelling radiances.
- Light et al. (2003) used BMC to simulate light fields in 3-D sea ice cylindrical geometries.
- Gordon and Ding (1992) and Leathers et al. (2004) used BMC to quantify the effects of instrument self-shading on in-water measurements of upwelling radiances.
- Lesser et al. (2018, 2021) used BMC to compute various irradiances incident onto 3-D coral reefs.

For a specific numerical example, consider the coral reef geometry seen in Fig. E.31. $(\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}})$ is an ocean coordinate system used to define reef the geometry and for ray tracing. The green half-planes represent the horizontal top and vertical wall of an idealized coral wall reef. The blue plane is the sea surface. The sensor (light detector) can be placed at any location (x_s, y_s, z_s) on the surface of the reef or within the water column. The object labeled S_d illustrates a sensor looking upward so as to collect downward traveling rays (e.g., to measured E_d) at some distance x_s in “front” of the reef wall (e.g. at location $(x_s, y_s = 0, z_s)$). The object labeled S_h represents a sensor on the face of the reef wall and looking horizontally away from the wall (e.g. at location $(x_s = 0, y_s = 0, z_s)$). These sensors have a normal, or viewing direction, $\hat{\mathbf{n}}$ that is defined by a polar angle θ_s and azimuthal angle ϕ_s in the ocean system. The Sun can be “in front of” the reef, so that the wall receives the Sun’s direct light. The Sun can be “behind” the reef, as shown in the figure, so that the reef wall is in its own shadow. The Sun-shadow boundary is illustrated by the gray half plane. In simulations, the detectors can simulate plane, hemispherical scalar, or scalar irradiances, or radiance in any direction, depending on the angular distribution of the emitted rays.

Those who model coral reef primary production often use PAR values derived from open-ocean downwelling plane irradiances, denoted here as PAR_d , with a scale factor applied to convert the value of PAR_d to a PAR value, PAR_h , incident onto a vertical reef wall. The scale factor is often taken to be 0.25 based on a very limited set of measurements (Brakel, 1979). One purpose of the Lesser et al. (2021) study was to evaluate how much a value of PAR_d measured in the 1-D geometry far away from the reef wall is affected by the wall itself, and what factors should be used to convert an open ocean PAR_d to PAR_h at the face of the reef wall.

A series of BMC runs was made for this geometry using a BMC code developed for simulation of light fields near 3-D coral reefs (Mobley, 2018). The BMC code has a general 3-D reef geometry “built in” (Fig. E.31 is a simplified version of the general geometry). As described in Lesser et al. (2021), the BMC code was run with spectrally dependent IOPs measured by Russell et al. (2019) in wall reef areas of the Pacific Ocean. The full set of runs included Sun zenith and azimuthal angles representing sunrise to sunset. Sensors of different types were placed at various depths and orientations. The runs covered 400 to 700 nm by 10 nm. The sensor types included plane, hemispherical scalar, and spherical scalar irradiance detectors. The spectral irradiances from those runs (in $\text{W m}^{-2} \text{nm}^{-1}$) were used to compute the corresponding PAR values in (in $\mu\text{mol quanta m}^{-2} \text{s}^{-1}$).

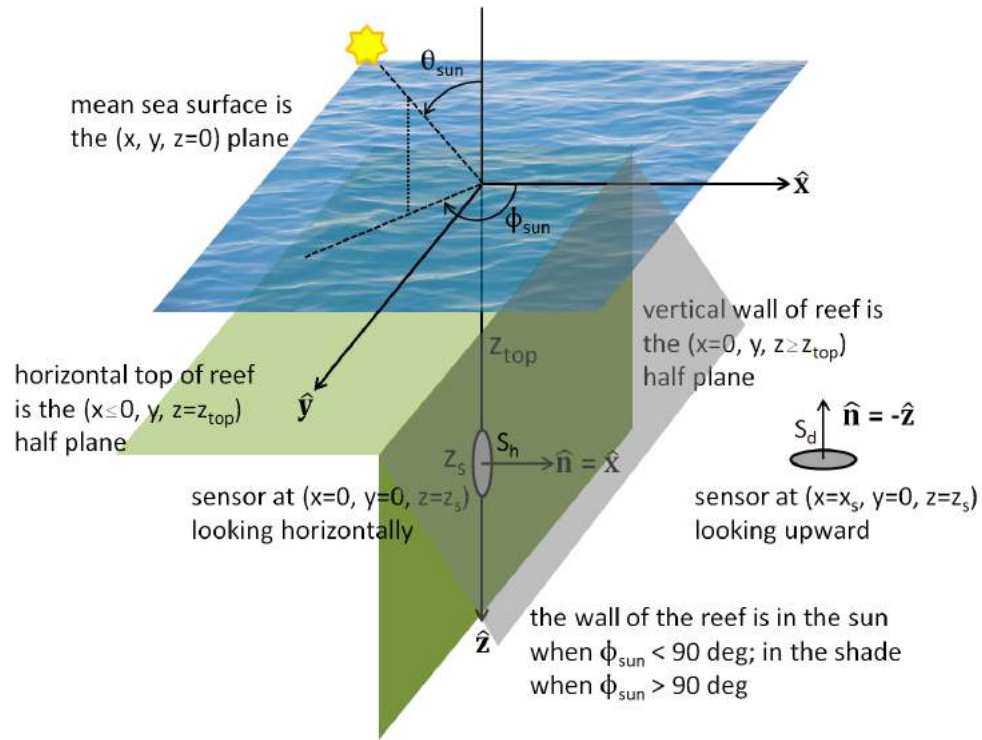


Figure E.31: Coral wall reef 3-D geometry used in simulations.

Figure E.32 shows the results from one series of runs. The runs simulated downwelling plane irradiance sensors (measuring $E_d(\lambda)$, leading to PAR_d), plane irradiance sensors looking horizontally away from the reef wall (leading to PAR_h) and hemispherical scalar irradiance sensors looking horizontally away from the reef wall (leading to PAR_{oh}). These sensors were placed from just next to the reef wall, out to a distance $x_s = 100$ m away from the wall. For the output in this figure, the Sun was at a 30 deg zenith angle and either in front of the reef wall ($\phi = 0$) or behind the wall ($\phi = 180$), in which case the reef wall is in its own shadow. At $z_s = 50$ m depth, the water in front of the wall is in shadow out to about $x = 20$ m, which is indicated by the gray shaded area in the figure.

The solid lines in the figure show how PAR_d at $z_s = 50$ m decreases as the sensor gets close to the reef wall so that the radiance distribution becomes 3-D. The symbols show the distances x_s away from the reef wall where the BMC code computed the irradiances. At $x_s = 100$ m the reef is optically very far away and there is no noticeable effect on $PAR_d \approx 23 \mu\text{mol quanta m}^{-2} \text{s}^{-1}$, which is independent of the Sun's azimuthal angle. However, as the sensor is moved closer to the reef wall PAR_d decreases to about $14 \mu\text{mol quanta m}^{-2} \text{s}^{-1}$ when the Sun is in front of the reef, and to about $5 \mu\text{mol quanta m}^{-2} \text{s}^{-1}$ when the reef wall is in its own shadow. The corresponding PAR_h values are about 8 and 3.5, respectively. Thus the scale factor that converts an open ocean PAR_d to a reef-wall PAR_h is $8/23 = 0.35$ for the wall in the Sun and $3.5/23 = 0.15$ for the wall in the shade. Although these particular values do on average equal the commonly used value of 0.25, it is emphasized that these conversion factors depend on the water IOPs and on the sensor depth and

type. The figure also shows the values for a hemispherical scalar irradiance sensor, which might be a more suitable measure of the light available to the coral. The BMC values at $x_s = 100$ m agree to within 2% with values computed by the 1-D HydroLight code; the small discrepancy is due to statistical noise. Such comparisons provide a check on the correctness of the BMC 3-D code.

This example is sufficient to show the power of backward Monte Carlo ray tracing for 3-D geometries. Many more runs are required in order to draw any general conclusion about irradiances near a coral reef.

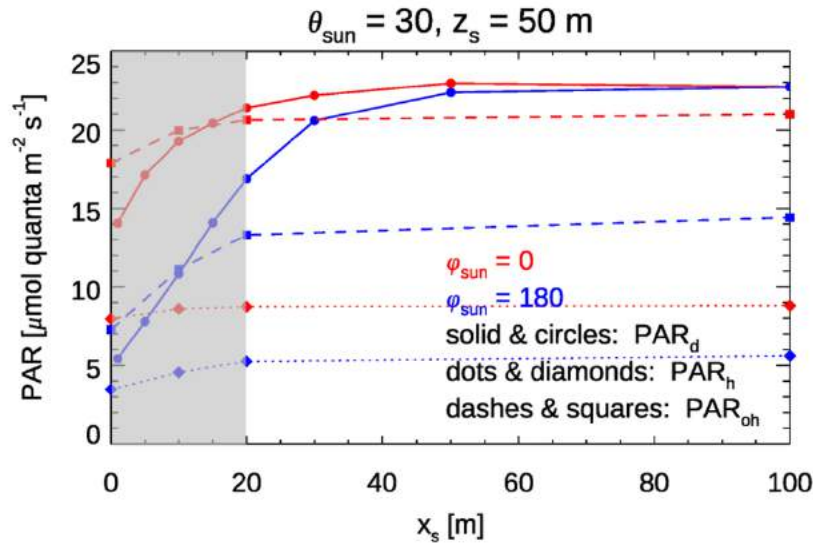


Figure E.32: Effect of the 3-D coral structure on various measures of PAR.

In closing it is noted that the BMC technique is computationally most efficient when the source (the sky in the above example) has a large spatial or angular extent and the sensor is spatially small (or even a point source) or a radiometer with a small field of view. If both the source and the sensor are small, or well collimated, then neither forward nor backward Monte Carlo simulation is numerically efficient.

Monte Carlo Ray Tracing

The mathematical details of Monte Carlo ray tracing are not something most oceanographers need to know. However, if you ever need to write a computer program for either forward or backward Monte Carlo simulation, you may think this appendix is the most useful chapter in the book.

The preceding Monte Carlo appendix explains how to determine ray path lengths (Section E.2) and scattering angles (Section E.3), and it explains three options for weighting and tracking rays within the water (Section E.4). However, those ray tracing calculations must be performed with reference to a coordinate system that is appropriate for the particular problem at hand.

In most Monte Carlo simulations it is necessary repeatedly to determine whether a ray intersects a surface (e.g., the sea bottom, the sea surface, or the surface of an object in the water). If a ray intersects a solid surface, then a reflected ray must be created in accordance with the BRDF of the surface. If a ray intersects the sea surface, a reflected ray is always generated, and usually a transmitted ray is also generated). It is these calculations that are described in this appendix.

The first step is to pick an overall “ocean” coordinate system convenient for the geometry of the problem being studied. Global coordinate systems were discussed in general in Section 1.4.1.1. An example of a global system tailored to a specific problem is seen in Fig. E.31, where a Cartesian (\hat{x} , \hat{y} , \hat{z}) system was used to define the 3-D geometry of a coral reef. In that particular system, the origin of the coordinate system is at the mean sea surface, \hat{x} points away from the reef wall, and \hat{z} is positive downward. “Local” coordinate systems as discussed in Section 1.4.1.2 will also be needed for the calculation of scattering within the water body and for reflection by surfaces.

There are innumerable possibilities for light ray paths and interactions with the bottom and sea surface and within the water body. Ray tracing in a geometry like that of Fig. E.31 boils down to repeatedly finding the intersections of lines (the rays) and planes (the mean sea surface, the reef top or side, etc.). The needed geometric entities are now defined in detail.

F.1 General Ray-Plane Intersections

A *line* \mathbf{L} refers to the set of co-linear points defined by two points $\mathbf{r}_1 = [x_1, y_1, z_1]^T$ and $\mathbf{r}_2 = [x_2, y_2, z_2]^T$ in 3-D space. (A superscript T denotes transpose, which in the present case gives column vectors.) A *ray* \mathbf{R} is the directed line segment from point \mathbf{r}_1 to point \mathbf{r}_2 ; the initial point of the ray is \mathbf{r}_1 and the final point is \mathbf{r}_2 . A ray represents a beam of light, which travels in a straight line between scattering or reflection events and which is attenuated by absorption as it passes through the water. A ray is created at \mathbf{r}_1 (by initialization, reflection by a surface, or scattering within the water) and travels to \mathbf{r}_2 , where it can be scattered again, unless it intersects a surface along the way. Note that a line is infinitely long, whereas a ray is a line segment with finite length with a direction.

Suppose that the problem has a mean sea surface at depth $z = 0$ and a level sea bottom at depth z_B ($\hat{\mathbf{z}}$ is positive downward). Given an initial ray point \mathbf{r}_1 , a new ray is generated (in any of several ways, as described below), whose endpoint is \mathbf{r}_2 . If $z_2 \geq z_B$, the ray has intersected the bottom. When that occurs, a reflected ray is generated as described in Section F.2. If $z_2 \leq 0$, the ray has intersected the mean sea surface. Transmitted and reflected rays are then computed as described in Section F.3. If $0 < z_2 < z_B$, then the ray is within the water column. A new ray is then generated by scattering as described in Section F.4. The new ray generated by reflection or scattering is then traced in the same manner as the previous ray.

A line \mathbf{L} can be written in parametric form as

$$\mathbf{L} = \mathbf{r}_1 + (\mathbf{r}_2 - \mathbf{r}_1)t, \quad -\infty < t < \infty. \quad (\text{F.1})$$

The corresponding ray \mathbf{R} is then

$$\mathbf{R} = \mathbf{r}_1 + (\mathbf{r}_2 - \mathbf{r}_1)t, \quad 0 < t \leq 1. \quad (\text{F.2})$$

A plane is determined by three non-colinear points $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$; $\mathbf{p}_i = [p_{ix}, p_{iy}, p_{iz}]^T, i = 1, 2, 3$. A plane \mathbf{P} can be written in parametric form as

$$\mathbf{P} = \mathbf{p}_1 + (\mathbf{p}_2 - \mathbf{p}_1)u + (\mathbf{p}_3 - \mathbf{p}_1)v, \quad -\infty < u, v < +\infty. \quad (\text{F.3})$$

The intersection of a line and a plane can be found by setting the parametric forms for a line and a plane equal:

$$\mathbf{r}_1 + (\mathbf{r}_2 - \mathbf{r}_1)t = \mathbf{p}_1 + (\mathbf{p}_2 - \mathbf{p}_1)u + (\mathbf{p}_3 - \mathbf{p}_1)v.$$

This can be written in matrix form as

$$\begin{bmatrix} x_1 - p_{1x} \\ y_1 - p_{1y} \\ z_1 - p_{1z} \end{bmatrix} = \begin{bmatrix} x_1 - x_2 & p_{2x} - p_{1x} & p_{3x} - p_{1x} \\ y_1 - y_2 & p_{2y} - p_{1y} & p_{3y} - p_{1y} \\ z_1 - z_2 & p_{2z} - p_{1z} & p_{3z} - p_{1z} \end{bmatrix} \begin{bmatrix} t \\ u \\ v \end{bmatrix}. \quad (\text{F.4})$$

After solving this equation for t, u, v , the value of t gives the distance from \mathbf{r}_1 at which the *line* intersects the plane. A negative value of t means that the plane lies “behind” the starting point of the ray, i.e., the plane is in the opposite direction of the ray. For example, if the ray is heading upward, the the sea bottom is “behind” the ray. If $t > 1$, the plane is

“in front of” the ray, but the ray does not reach the plane. If $0 \leq t \leq 1$, the associated *ray* intersects the plane. Thus if $t > 1$ or $t < 0$, the line intersects the plane, but the ray does not.

The point of intersection of the line with the plane is

$$\mathbf{L}_P = \mathbf{r}_1 + (\mathbf{r}_2 - \mathbf{r}_1)t. \quad (\text{F.5})$$

where the value of t is given by the solution of Eq. (F.4). The solution values of u, v , when substituted into Eq. (F.3), give the point in the plane where the line intersects it.

We are free to choose points $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ as desired, so long as they lie in the plane and are not colinear. Therefore Eq. (F.4) can be tailored to the mean sea surface and sea bottom planes so as to simplify the matrix inversions required to obtain t, u , and v .

F.1.1 Ray-Bottom Intersections

The plane of the sea bottom can be defined by the three points

$$\begin{aligned} \mathbf{p}_1 &= [0, 0, z_B]^T, \\ \mathbf{p}_2 &= [0, 1, z_B]^T, \\ \mathbf{p}_3 &= [1, 0, z_B]^T. \end{aligned}$$

Equation (F.4) then simplifies to

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 - z_B \end{bmatrix} = \begin{bmatrix} x_1 - x_2 & 0 & 1 \\ y_1 - y_2 & 1 & 0 \\ z_1 - z_2 & 0 & 0 \end{bmatrix} \begin{bmatrix} t_B \\ u_B \\ v_B \end{bmatrix}, \quad (\text{F.6})$$

which has the solution

$$\begin{bmatrix} t_B \\ u_B \\ v_B \end{bmatrix} = \frac{1}{z_1 - z_2} \begin{bmatrix} 0 & 0 & 1 \\ 0 & (z_1 - z_2) & -(y_1 - y_2) \\ (z_1 - z_2) & 0 & -(x_1 - x_2) \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 - z_B \end{bmatrix}.$$

The value of t for the ray-bottom intersection is

$$t_B = \frac{z_1 - z_B}{z_1 - z_2}. \quad (\text{F.7})$$

Recalling Eq. (F.5), the point $\mathbf{x}_B = [x_B, y_B, z_B]^T$ where the line intersects the bottom is

$$x_B = x_1 + (x_2 - x_1) \left(\frac{z_1 - z_B}{z_1 - z_2} \right), \quad (\text{F.8a})$$

$$y_B = y_1 + (y_2 - y_1) \left(\frac{z_1 - z_B}{z_1 - z_2} \right), \quad (\text{F.8b})$$

$$z_B = z_B. \quad (\text{F.8c})$$

If the ray has $z_1 = z_2$ (to within the numerical roundoff error of the calculations), the ray is parallel to the bottom, and there is no intersection for ray tracing. This can also be

seen from the determinant of the matrix in Eq. (F.6), $D = -(z_1 - z_2)$, which is then zero, indicating that the equation has no solution. Although such events are rare, computer codes must always check for such rays and treat them as special cases.

If the ray intersects the bottom, then a reflected ray is generated with polar and azimuthal angles of reflection using the BRDF of the bottom as described in Section F.8. Material surfaces are often assumed to be Lambertian reflectors (Section 13.48), in which case the angles of reflection are determined as described in Section F.2.

F.1.2 Ray-Sea Surface Intersections

The plane containing the mean sea surface can be defined by the three points

$$\begin{aligned}\mathbf{p}_1 &= [0, 0, 0]^T, \\ \mathbf{p}_2 &= [0, 1, 0]^T, \\ \mathbf{p}_3 &= [1, 0, 0]^T.\end{aligned}$$

Equation (F.4) then simplifies to

$$\begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} x_1 - x_2 & 0 & 1 \\ y_1 - y_2 & 1 & 0 \\ z_1 - z_2 & 0 & 0 \end{bmatrix} \begin{bmatrix} t_S \\ u_S \\ v_S \end{bmatrix},$$

which has the solution

$$\begin{bmatrix} t_S \\ u_S \\ v_S \end{bmatrix} = \frac{1}{z_1 - z_2} \begin{bmatrix} 0 & 0 & 1 \\ 0 & (z_1 - z_2) & -(y_1 - y_2) \\ (z_1 - z_2) & 0 & -(x_1 - x_2) \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix}.$$

The value of t for the ray intersection with the air-water surface is

$$t_S = \frac{z_1}{z_1 - z_2}. \quad (\text{F.9})$$

Recalling Eq. (F.5), the point $\mathbf{L}_S = [x_S, y_S, z_S]^T$ where the line intersects the mean sea surface is

$$x_S = x_1 + (x_2 - x_1) \left(\frac{z_1}{z_1 - z_2} \right), \quad (\text{F.10a})$$

$$y_S = y_1 + (y_2 - y_1) \left(\frac{z_1}{z_1 - z_2} \right), \quad (\text{F.10b})$$

$$z_S = 0. \quad (\text{F.10c})$$

If the ray intersects the mean sea surface, then reflected and (perhaps) transmitted rays are created as described in Section F.3.

F.2 Reflection by Lambertian Surfaces

If a ray is found to intersect the bottom, it is then reflected at the point of intersection. Assume that the bottom is a Lambertian surface (Section 13.7) with irradiance reflectance R_B . In general this reflectance depends on wavelength and is determined by the type of bottom material being modeled. A Lambertian surface reflects an incident ray into randomly determined polar and azimuthal angles according to Eqs. (F.52) and (F.51):

$$\theta_r = \cos^{-1}(\sqrt{u_1}), \quad (\text{F.11})$$

$$\phi_r = 2\pi u_2. \quad (\text{F.12})$$

Here u_1 and u_2 are independently drawn uniform $[0,1]$ random numbers. Equation (F.11) is certainly not intuitive, but this distribution of reflected polar angles can be understood in the context of the discussion of Section 13.7. The polar angle θ_r is measured relative to the outward normal to the Lambertian surface at $\theta_r = 0$. That is, θ_r is measured from the $-\hat{\mathbf{z}}$ direction. A convenient local “bottom” coordinate system then has $(\hat{\mathbf{x}}_B, \hat{\mathbf{y}}_B, \hat{\mathbf{z}}_B) = (\hat{\mathbf{x}}, -\hat{\mathbf{y}}, -\hat{\mathbf{z}})$ in terms of the ocean coordinate system. After computation of (θ_r, ϕ_r) from the equations above, the direction of the reflected ray in the ocean system is given by

$$\theta = \pi - \theta_r \quad (\text{F.13})$$

$$\phi = \pi + \phi_r \text{ modulo } \pi. \quad (\text{F.14})$$

The modulo π on the last equation keeps the ϕ values in the $-\pi$ to π range, which is equivalent to the 0 to 2π range for ϕ_r .

The distance traveled by the reflected ray is given by (E.4):

$$\rho_r = -\frac{1}{c} \ln u_3, \quad (\text{F.15})$$

where c is the beam attenuation coefficient of the water. This ρ_r is the distance the newly created reflected ray will travel through the water body, unless it intersects another surface.

The weight w_r of the reflected ray is

$$w_r = R_B w_i, \quad (\text{F.16})$$

where w_i is the weight of the incident ray, and R_B is the irradiance reflectance of the Lambertian bottom.

Equations (F.13) and (F.14) give the direction of the reflected ray in the ocean coordinate system. The length of the ray is given by Eq. (F.15). This length is independent of coordinate system, so we can write $\rho = \rho_r$. These (ρ, θ, ϕ) values give the location of the end point of the reflected ray in spherical coordinates, *relative to the point of reflection* $[x_B, x_B, x_B]^T$, which is computed by Eq. (F.8). The location of the reflected ray endpoint in the ocean system is then given by

$$\mathbf{r}_2 = \begin{bmatrix} x_B \\ y_B \\ z_B \end{bmatrix} + \rho \begin{bmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{bmatrix}. \quad (\text{F.17})$$

F.3 Reflection and Transmission by the Sea Surface

If a ray crosses the mean sea surface at $z = 0$, it is processed using the slope statistics for a wind-blown sea surface, Snell's law of refraction, and Fresnel's law of reflection for unpolarized light. The Cox-Munk wind speed-wave slope statistics of Section 13.4 are typically used to model the sea surface and will be used here. In all cases there is a ray reflected back into the water. The weight of the reflected ray is determined from the angle of the incident ray onto the (usually) tilted water surface and Fresnel's law. There is usually a ray transmitted into the air, whose direction in the air is determined by Snell's law and whose weight is obtained from one minus the fraction of the reflected-ray weight. The weight of the transmitted ray is added to an accumulating weight sum for the sky $(\Delta\theta, \Delta\phi)$ bin in the direction of the transmitted ray. In some cases there may be total internal reflection, in which case the reflected ray has the same weight as the incident ray and there is no transmitted ray. These calculations proceed as follows.

First, a unit vector normal to the sea surface is randomly generated in accordance with the Cox-Munk wind speed-wave slope law for the wind speed U of the simulation. Let $\eta(x, y)$ be the elevation of the wind-blown sea surface, and let the wind be blowing in the $+\hat{\mathbf{x}}$ direction. Then the along-wind and cross-wind slopes of the sea surface are

$$\eta_a = \frac{\partial\eta}{\partial x} \quad \text{and} \quad \eta_c = \frac{\partial\eta}{\partial y}.$$

The slopes η_a and η_c have a Gaussian distribution about a zero mean with variances according to Eq. (13.41):

$$\sigma_a^2 = 3.16 \cdot 10^{-3}U \quad \text{and} \quad \sigma_c^2 = 1.92 \cdot 10^{-3}U, \quad (\text{F.18})$$

where U is in meters per second at 12.5 m above mean sea level. Let \mathbf{r} be a random number drawn from a Gaussian distribution with zero mean and unit variance, denoted by $\mathbf{r} \sim \mathcal{N}(0, 1)$. Then if \mathbf{r}_1 and \mathbf{r}_2 are independently drawn random numbers from $\mathcal{N}(0, 1)$,

$$\eta_a = \mathbf{r}_1 \sigma_a \quad \text{and} \quad \eta_c = \mathbf{r}_2 \sigma_c \quad (\text{F.19})$$

have the required variances for sea surface slopes. The associated outward (facing the sky) normal to the sea surface is then

$$\hat{\mathbf{n}} = \frac{\eta_a \hat{\mathbf{x}} + \eta_c \hat{\mathbf{y}} - \hat{\mathbf{z}}}{[\eta_a^2 + \eta_c^2 + 1]^{1/2}}. \quad (\text{F.20})$$

Note that $\hat{\mathbf{n}} = -\hat{\mathbf{z}}$ if the sea surface is level. When a ray crosses the mean sea surface at $z = 0$, two Gaussian random numbers are drawn and used in Eq. (F.19) to generate random values of η_a and η_c . These values are then used in Eq. (F.20) to generate the outward normal of a random sea surface wave facet $\hat{\mathbf{n}}$, which is then used in Eqs. (F.21) and (F.24)-(F.26) below to determine the angles of ray transmission and reflection.

Let $\hat{\boldsymbol{\xi}}'_i$, $\hat{\boldsymbol{\xi}}'_r$ and $\hat{\boldsymbol{\xi}}'_t$ be unit vectors in the directions of the incident, reflected, and transmitted rays, respectively. (As before, a primed variable indicates the incident ray, and unprimed variables indicate scattered (reflected or transmitted) rays.) The incident ray that reaches the sea surface is traveling in direction (θ, ϕ) in the ocean coordinate system. This direction is described by a unit vector

$$\begin{aligned}\hat{\xi}' &= \xi'_x \hat{x} + \xi'_y \hat{y} + \xi'_z \hat{z} \\ &= \sin \theta \cos \phi \hat{x} + \sin \theta \sin \phi \hat{y} + \cos \theta \hat{z}.\end{aligned}$$

The dot product between \hat{n} and $\hat{\xi}'$ gives the angle of incidence θ' of the ray onto the sloping sea surface:

$$\theta' = \cos^{-1}(\hat{\xi}' \cdot \hat{n}). \tag{F.21}$$

This is the angle relative to the normal to the (generally sloping) surface at the point where the incident ray is reflected back into the water by the underside of the sea surface. There is also usually a transmitted ray whose direction relative to the surface normal is given by Snell's law

$$\theta_t = \sin^{-1}(n_w \sin \theta'), \tag{F.22}$$

where n_w is the real index of refraction of the water. If the argument of the inverse sine is greater than 1, then there is total internal reflection and no ray is transmitted into the air.

The incident and transmitted angles θ' and θ_t are then used in Fresnel's formula (13.11) for the reflectance of the sea surface,

$$R_F(\theta') = \frac{1}{2} \left(\left[\frac{\sin(\theta' - \theta_t)}{\sin(\theta' + \theta_t)} \right]^2 + \left[\frac{\tan(\theta' - \theta_t)}{\tan(\theta' + \theta_t)} \right]^2 \right), \tag{F.23}$$

which holds for $\theta' \neq 0$. If $\theta' = 0$, the Fresnel reflectance is given by (13.12)

$$R_F(\theta' = 0) = \left(\frac{n_w - 1}{n_w + 1} \right)^2.$$

The weight w of the incident ray is multiplied by R_F to get the weight of the reflected ray. The weight of the transmitted ray is then $(1 - R_F)w$. The weights of the reflected plus transmitted rays thus equal the weight of the incident ray; in other words, energy is conserved at the sea surface.

For a ray incident onto the surface from the water side, the directions of $\hat{\xi}_r$ and $\hat{\xi}_t$ relative to the tilted wave facet are given by Eqs. (13.7)–(13.9):

$$\hat{\xi}_r = \hat{\xi}' - 2(\hat{\xi}' \cdot \hat{n}) \hat{n} \tag{F.24}$$

$$\hat{\xi}_t = n_w \hat{\xi}' - c \hat{n}, \tag{F.25}$$

where

$$c = n_w \hat{\xi}' \cdot \hat{n} - \left[(n_w \hat{\xi}' \cdot \hat{n})^2 - n_w^2 + 1 \right]^{1/2}. \tag{F.26}$$

Note that both of these vectors lie in the plane determined by $\hat{\xi}'$ and \hat{n} .

Both $\hat{\xi}'$ and \hat{n} are specified in the ocean coordinate system. Therefore the equations for $\hat{\xi}_r$ and $\hat{\xi}_t$ give values that are also in the ocean coordinate system.

If the surface is not level (i.e., the wind speed U is not zero), it can happen that the surface is tilted so much that a near-grazing incident ray is reflected into a slightly upward direction, or that a transmitted ray is traveling in a slightly downward direction. In a real ocean, these rays would encounter the sea surface again at a nearby point and be reflected and transmitted again. In computer codes, such rays can be “wrapped around” for further

intersections with the sea surface, or that can be flagged as anomalous and are dropped. Simulation results show that there are only a few anomalous rays out of every 1000 rays incident onto the sea surface, so dropping these rays has a negligible effect on the computed radiometric values.

F.4 In-water Ray Scattering

If the ray end point \mathbf{r}_2 is within the water, then the ray is scattered at its endpoint. The generation of new rays by in-water scattering is central to the ray-tracing process, and these calculations are somewhat complicated because of the coordinate system transformations. It is therefore worthwhile to review in detail how ray scattering is effected in Monte Carlo codes. Although the coordinate system transformations developed in this section are standard (e.g. Bower, 2012; Cheston, 1964), they can be hard to find unless you know what you are looking for. The needed equations are derived here for completeness.

Recalling Eq. (F.2), let \mathbf{R}' denote an initial ray with starting point \mathbf{r}_1 and ending point \mathbf{r}_2 . This ray is traveling in direction (θ, ϕ) in the chosen global (ocean) coordinate system, as seen in Fig. F.1. An in-water ray can be anywhere in the water and traveling in any direction. It is therefore necessary to define a local (at the point of scattering) coordinate system for scattering calculations, as was described in Section 1.4.1.2 and Fig. 1.7. The scattering angles ψ and α will then be applied in this system to determine the direction of the scattered ray \mathbf{R} .

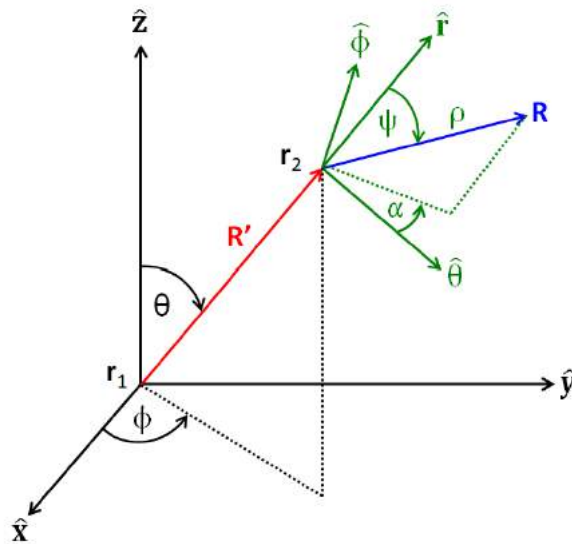


Figure F.1: Coordinate systems used to describe the scattering of an in-water ray. The initial or unscattered ray \mathbf{R}' is shown in red; the scattered ray \mathbf{R} is blue. The ocean coordinate system and angles measured in this system are in black, and the local coordinate system and angles measured in the local system are green.

In the *ocean* system, ray \mathbf{R}' has components (R'_x, R'_y, R'_z) :

$$\begin{aligned}\mathbf{R}' &= R'_x \hat{\mathbf{x}} + R'_y \hat{\mathbf{y}} + R'_z \hat{\mathbf{z}} \\ &= R' \sin \theta \cos \phi \hat{\mathbf{x}} + R' \sin \theta \sin \phi \hat{\mathbf{y}} + R' \cos \theta \hat{\mathbf{z}},\end{aligned}\quad (\text{F.27})$$

where the Cartesian components R'_x etc., in terms of spherical coordinates come from inspection of Fig. F.1. R' is the length of the vector \mathbf{R}' .

A convenient *local* coordinate system $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}, \hat{\mathbf{r}})$ for scattered rays is constructed as described in Section 1.4.1.2; the equations developed there are repeated here for ease of reference. The radial unit vector

$$\hat{\mathbf{r}} = \frac{\mathbf{R}'}{R'} = \frac{\mathbf{r}_2 - \mathbf{r}_1}{|\mathbf{r}_2 - \mathbf{r}_1|}$$

is in the same direction as the initial ray. The azimuthal unit vector $\hat{\boldsymbol{\phi}}$ is defined by the cross product of the ocean coordinate system $\hat{\mathbf{z}}$ and the incident vector's direction:

$$\hat{\boldsymbol{\phi}} = \frac{\hat{\mathbf{z}} \times \hat{\mathbf{r}}}{|\hat{\mathbf{z}} \times \hat{\mathbf{r}}|}.$$

This vector points in the direction of increasing ϕ values. (If the unscattered vector is in the same direction as $\hat{\mathbf{z}}$, the direction of $\hat{\boldsymbol{\phi}}$ can be chosen at random.) The polar unit vector is then given by

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\phi}} \times \hat{\mathbf{r}}.$$

This vector points in the direction of increasing θ values. The $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}, \hat{\mathbf{r}})$ system is then an orthogonal system of coordinates in which the scattering angles can be applied to define the direction of the scattered ray. However, these directions are not fixed in the ocean system; they depend on the direction of the unscattered ray.

To express the local coordinate system directions in the ocean system, note that the radial unit vector $\hat{\mathbf{r}}$ can be thought of as the normalized rate of change \mathbf{R}' with respect to length R' :

$$\hat{\mathbf{r}} = \frac{1}{\left| \frac{\partial \mathbf{R}'}{\partial R'} \right|} \frac{\partial \mathbf{R}'}{\partial R'}.$$

Equation (F.27) then gives

$$\hat{\mathbf{r}} = \sin \theta \cos \phi \hat{\mathbf{x}} + \sin \theta \sin \phi \hat{\mathbf{y}} + \cos \theta \hat{\mathbf{z}}, \quad (\text{F.28})$$

after noting from Eq. (F.27) that $|\partial \mathbf{R}' / \partial R'| = 1$. In the same fashion, $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\phi}}$ can be written as

$$\hat{\boldsymbol{\theta}} = \frac{1}{\left| \frac{\partial \mathbf{R}'}{\partial \theta} \right|} \frac{\partial \mathbf{R}'}{\partial \theta} \quad \text{and} \quad \hat{\boldsymbol{\phi}} = \frac{1}{\left| \frac{\partial \mathbf{R}'}{\partial \phi} \right|} \frac{\partial \mathbf{R}'}{\partial \phi}.$$

After noting that $|\partial \mathbf{R}' / \partial \theta| = R'$ and $|\partial \mathbf{R}' / \partial \phi| = R' \sin \theta$, the same process gives

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \cos \theta \cos \phi \hat{\mathbf{x}} + \cos \theta \sin \phi \hat{\mathbf{y}} - \sin \theta \hat{\mathbf{z}} \\ \hat{\boldsymbol{\phi}} &= -\sin \theta \hat{\mathbf{x}} + \cos \theta \hat{\mathbf{y}}.\end{aligned}$$

Now consider an arbitrary vector point function $\mathbf{G}(\mathbf{R}')$. In the present case¹, $\mathbf{G}(\mathbf{R}')$ will be the end point of the scattered vector \mathbf{R} at location \mathbf{R}' . In the Cartesian ocean system, $\mathbf{G}(\mathbf{R}')$ is written in component form as

$$\mathbf{G}(\mathbf{R}') = G'_x \hat{\mathbf{x}} + G'_y \hat{\mathbf{y}} + G'_z \hat{\mathbf{z}}.$$

In the local system, $\mathbf{G}(\mathbf{R}')$ is written as

$$\mathbf{G}(\mathbf{R}') = G'_r \hat{\mathbf{r}} + G'_\theta \hat{\boldsymbol{\theta}} + G'_\phi \hat{\boldsymbol{\phi}}.$$

The relations between the components of $\mathbf{G}(\mathbf{R}')$ in these two coordinate systems are obtained as follows. The radial component of $\mathbf{G}(\mathbf{R}')$ is

$$G'_r = \mathbf{G}(\mathbf{R}') \cdot \hat{\mathbf{r}} = G'_x(\hat{\mathbf{x}} \cdot \hat{\mathbf{r}}) + G'_y(\hat{\mathbf{y}} \cdot \hat{\mathbf{r}}) + G'_z(\hat{\mathbf{z}} \cdot \hat{\mathbf{r}}).$$

However, from Eq. (F.28)

$$\begin{bmatrix} \hat{\mathbf{x}} \cdot \hat{\mathbf{r}} \\ \hat{\mathbf{y}} \cdot \hat{\mathbf{r}} \\ \hat{\mathbf{z}} \cdot \hat{\mathbf{r}} \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \phi \\ \sin \theta \sin \phi \\ \cos \theta \end{bmatrix},$$

so that

$$G'_r = G'_x \sin \theta \cos \phi + G'_y \sin \theta \sin \phi + G'_z \cos \theta.$$

In the same fashion we find

$$\begin{aligned} G'_\theta &= G'_x \cos \theta \cos \phi + G'_y \cos \theta \sin \phi - G'_z \sin \theta \\ G'_\phi &= -G'_x \sin \phi + G'_y \cos \phi. \end{aligned}$$

Writing these results in matrix form gives

$$\begin{bmatrix} G'_r \\ G'_\theta \\ G'_\phi \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \phi & \sin \theta \sin \phi & \cos \theta \\ \cos \theta \cos \phi & \cos \theta \sin \phi & -\sin \theta \\ -\sin \phi & \cos \phi & 0 \end{bmatrix} \begin{bmatrix} G'_x \\ G'_y \\ G'_z \end{bmatrix}.$$

The inverse gives (G'_x, G'_y, G'_z) in terms of $(G'_r, G'_\theta, G'_\phi)$:

$$\begin{bmatrix} G'_x \\ G'_y \\ G'_z \end{bmatrix} = \begin{bmatrix} \sin \theta \cos \phi & \cos \theta \cos \phi & -\sin \phi \\ \sin \theta \sin \phi & \cos \theta \sin \phi & \cos \phi \\ \cos \theta & -\sin \theta & 0 \end{bmatrix} \begin{bmatrix} G'_r \\ G'_\theta \\ G'_\phi \end{bmatrix}. \quad (\text{F.29})$$

Now let the vector point function $\mathbf{G}(\mathbf{R}')$ be the end point of the scattered ray \mathbf{R} at location \mathbf{R}' . The end point has magnitude (length of the scattered ray) ρ and direction (ψ, α) in the local system. Inspection of Fig. F.1 shows that

$$\mathbf{G}(\mathbf{R}') = \begin{bmatrix} G'_r \\ G'_\theta \\ G'_\phi \end{bmatrix} = \mathbf{R} = \begin{bmatrix} \rho \cos \psi \\ \rho \sin \psi \cos \alpha \\ \rho \sin \psi \sin \alpha \end{bmatrix}.$$

¹In other problems, $\mathbf{G}(\mathbf{R}')$ might be, for example, the value of an electric field at spatial point \mathbf{R}' . Then \mathbf{R} would be the strength of the electric field and (ψ, α) would give the direction of the electric field at \mathbf{R}' .

Inserting these values into Eq. (F.29) and remembering that the origin of the $(\hat{\mathbf{r}}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}})$ system is at point \mathbf{r}_2 gives

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} + \begin{bmatrix} \sin \theta \cos \phi & \cos \theta \cos \phi & -\sin \phi \\ \sin \theta \sin \phi & \cos \theta \sin \phi & \cos \phi \\ \cos \theta & -\sin \theta & 0 \end{bmatrix} \begin{bmatrix} \rho \cos \psi \\ \rho \sin \psi \cos \alpha \\ \rho \sin \psi \sin \alpha \end{bmatrix}. \quad (\text{F.30})$$

This equation gives the location of the end point of the scattered ray \mathbf{R} in the ocean coordinate system. To continue tracing the scattered ray, the unscattered point \mathbf{r}_2 becomes the new \mathbf{r}_1 , and the scattered ray endpoint at $\mathbf{R} = [x, y, z]^T$ becomes the new \mathbf{r}_2 .

Some readers may not be familiar with the transformation of Eq. (F.30). It is therefore emphasized that the familiar transformation from spherical (ρ, ψ, α) to Cartesian (θ, ϕ, r) coordinates (you can think of the $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\phi}}, \hat{\mathbf{r}})$ system of Fig. F.1 as a local Cartesian system),

$$\begin{bmatrix} \theta \\ \phi \\ r \end{bmatrix} = \rho \begin{bmatrix} \sin \psi \cos \alpha \\ \sin \psi \sin \alpha \\ \cos \psi \end{bmatrix},$$

gives the coordinates of the endpoint of the scattered ray in the *local* coordinate (θ, ϕ, r) system used for the scattering calculations, whereas Eq. (F.30) gives the coordinates of the endpoint of the scattered ray in the *ocean* (x, y, z) coordinate system used for ray tracing.

The ray length ρ is determined by Eq. (E.4), ψ is determined according to the scattering phase function used in the simulation, and α is randomly chosen to be azimuthally isotropic. Equation (F.30) is then used to obtain the end point of the scattered ray in the ocean system.

As described in Section E.3, the polar scattering angle ψ is determined for an arbitrary scattering phase function $\tilde{\beta}(\psi)$ by Eq. (E.6), and the azimuthal scattering angle α is determined by Eq. (E.5). The determination of ψ usually must be determined numerically. In practice, this often can be done most efficiently by first building a look-up table of values of the integral of Eq. (E.6) for closely spaced values of ψ from 0 to π . This look-up table is then the cumulative distribution function for the phase function, which is viewed as a probability distribution function for the scattering angle ψ . The random number u can then be used in this look-up table to obtain the value of ψ .

For Type 2 ray tracing as described in Section E.4, at each scattering, the current ray weight is multiplied by the albedo of single scattering, ω_o . This multiplication of the weight by ω_o accounts for the loss of ray energy to absorption as the ray passes through the water.

F.5 Use of the Preceding Equations

It may be worthwhile to summarize the use of the preceding equations as employed in a Monte Carlo code. The algorithm is as follows.

Initial Ray: An initial ray starting at point $\mathbf{r}_1 = [x_1, y_1, z_1]^T$ is predicted to reach final point $\mathbf{r}_2 = [x_2, y_2, z_2]^T$, with both points expressed in the ocean coordinate system.

Testing: This ray is examined to see if it has intersected the bottom, the mean sea surface, or is within the water. The weight is checked to see if the ray is less than the minimum weight w_{\min} , in which case it is dropped and new ray is initialized.

Ray-Bottom Intersection: Suppose it is found that this initial ray intersects the bottom. The point of intersection of the ray with the bottom, $\mathbf{p}_B = [x_B, y_B, z_B]^T$, is given by Eqs. (F.8). These coordinates are in the ocean system.

Bottom Reflection: The ray is reflected into polar and azimuthal directions (θ_r, ϕ_r) according to Eqs. (F.11) and (F.12). These directions are relative to the bottom coordinate system at the point of intersection. The reflected ray will travel a distance ρ_r given by Eq. (F.15); the weight of the reflected ray is given by Eq. (F.16). The directions of the reflected ray are converted to directions in the ocean system via Eqs. (F.13) and (F.14).

New Ray Initial Point: The intersection point \mathbf{p}_B becomes the initial point $\mathbf{r}_1 = [x_1, y_1, z_1]^T$ for the reflected ray.

New Ray Final Point: The end point of the new ray in the ocean system is then given by Eq. (F.17).

Recursion: The new ray from \mathbf{r}_1 to \mathbf{r}_2 is now traced just as was the initial ray.

This process is the same for rays that intersect the sea surface or that are within the water. In either case, the appropriate equations of Sections F.3 or F.4 are used to compute new rays, either by reflection and transmission at the sea surface, or by scattering within the water. Each new ray is then tracked in the same manner as the previous one.

F.6 Ray Initialization in Backward Monte Carlo Simulations

When a ray is initialized from a detector, values of polar and azimuthal angles (γ, α) are randomly determined according to the detector's angular response. These values of (γ, α) are in a local coordinate system for the detector. If the detector is within the water, this will be a system defined by the detector's orientation (θ_D, ϕ_D) in the ocean system. In any case, the initial ray can be treated just as if it were a scattered ray created by a scattering event at the location of the detector. The initial ray's ρ value is determined using Eq. (F.15), and the initial weight is $w = 1$. The initial ray endpoint is then converted into ocean coordinates using transformation (F.30), just as for scattered rays.

A plane irradiance sensor (cosine detector) has a detecting *surface material* for which each *point* is equally sensitive to radiance from any direction; the cosine response comes from the change in the apparent detector surface area for off-axis viewing directions. The detector material is assumed to be azimuthally isotropic. Let the polar (off-axis) angular

response of the detector material be $\sigma(\gamma)$. Then the material of a cosine detector has $\sigma(\gamma) = \sigma_o$, independent of γ , which gives a cosine response of the detector as a whole, including the geometric effect of the apparent detector surface area for different viewing directions. In backward Monte Carlo ray tracing, rays are emitted from the detector location in an angular pattern that mimics the response of the detector to incoming rays.

In general for an azimuthally isotropic sensor whose polar (off-axis) angular response is $\sigma(\gamma')$, the random value of γ is determined by solving (Modest, 1993, Eq. 19.32)

$$\mathbf{u} = \frac{\int_0^\gamma \sigma(\gamma') \cos \gamma' \sin \gamma' d\gamma'}{\int_0^{\pi/2} \sigma(\gamma') \cos \gamma' \sin \gamma' d\gamma'} , \quad (\text{F.31})$$

where \mathbf{u} is a uniform [0,1] random number. (See Section 13.7 for further discussion of how a cosine-reflecting material leads to a surface whose reflected *radiance* is independent of viewing direction.)

In this case of a cosine collector or emitter, the collector material response function is

$$\sigma(\gamma) = \begin{cases} \sigma_o & \text{if } 0 \leq \gamma \leq \pi/2 \\ 0 & \text{if } \pi/2 < \gamma \leq \pi . \end{cases}$$

Using this in Eq. (F.31) yields

$$\mathbf{u} = \sin^2 \gamma ,$$

or

$$\gamma = \sin^{-1}(\sqrt{\mathbf{u}}) . \quad (\text{F.32})$$

This peculiar looking formula for γ is precisely what is needed to make the angular distribution of rays emitted from a detector in a backward Monte Carlo simulation mimic the cosine response of a plane irradiance sensor.

Consider a “top-hat” detector with a field of view of half angle γ_D and whose collecting surface is equally sensitive to radiance from any direction, like that of a cosine collector. (In other words, we can place a Gershun tube over a cosine detector to limit the field of view and create a radiance detector.) That is,

$$\sigma(\gamma) = \begin{cases} \sigma_o & \text{if } 0 \leq \gamma \leq \gamma_D \\ 0 & \text{if } \gamma_D < \gamma \leq \pi . \end{cases}$$

Equation (F.31) then yields

$$\gamma = \sin^{-1}(\sin \gamma_D \sqrt{\mathbf{u}}) . \quad (\text{F.33})$$

This distribution of emission angles is used to simulate a radiance detector with a viewing direction half-angle of γ_D . Note that Eq. (F.33) reduces to Eq. (F.32) if the detector field of view opens up to $\gamma_D = 90$ deg.

Another way to view emission from the detector is to think of the emitted ray as a ray that has been created by scattering at the location of the point detector. For example, one can imagine a ray incident onto the detector surface from “behind” or “within” the surface

material, and then being scattered (emitted) outward from the surface into the water. For a phase function that scatters (emits) in a cosine pattern, the scattering phase function is

$$\tilde{\beta}(\psi) = \begin{cases} \frac{1}{\pi} \cos \psi & \text{if } 0 \leq \psi \leq \pi/2 \\ 0 & \text{if } \pi/2 < \psi \leq \pi. \end{cases}$$

Inserting this phase function into Eq. (E.6) leads again to $\psi = \sin^{-1}(\sqrt{u})$, with the scattering angle ψ now playing the role of the emission angle γ . Recall Table E.3 for a cosine scattering angle.

The viewpoint of emission as a scattering process readily yields the formulas needed for emission by scalar irradiance detectors. For a hemispherical scalar irradiance detector, corresponding to isotropic emission into $0 \leq \gamma \leq \pi/2$, the equivalent scattering phase function is

$$\tilde{\beta}(\gamma) = \begin{cases} \frac{1}{2\pi} & \text{if } 0 \leq \gamma \leq \pi/2 \\ 0 & \text{if } \pi/2 < \gamma \leq \pi. \end{cases}$$

Inserting this phase function into Eq. (E.6) gives

$$\gamma = \cos^{-1}(1 - u) \quad \text{or} \quad \gamma = \cos^{-1}(u).$$

Note that this γ can range from 0 to 90 deg. For a scalar irradiance detector, the equivalent phase function is $\tilde{\beta}(\gamma) = 1/(4\pi), 0 \leq \gamma \leq \pi$, and the resulting equation for the emission angle is

$$\gamma = \cos^{-1}(1 - 2u),$$

which gives γ values between 0 and 180 deg.

In all cases, the azimuthal angle is determined from

$$\alpha = 2\pi u.$$

For convenience of reference, Table F.1 collects these results for the functions used to generate emission angles for various types of sensors.

F.7 Processing Ray Weights in Backward Monte Carlo Calculations

In forward Monte Carlo calculations the final ray weights are accumulated in angular bins, and those weights then give the fractions of total emitted power reaching each bin. That power is then converted to the radiometric variable of interest.

The situation is more complicated in backward Monte Carlo simulations. This section considers the common case of an in-water sensor and illumination by the sky. It has been qualitatively stated that ray weights are accumulated in each sky angular bin, and that these weights are used to determine how much of the sky radiance reaches the in-water sensor. Exactly how this is done is rather subtle and warrants further comment.

The BMC technique rests on a reciprocity relation developed in Case (1957), which was exploited in the ocean setting in Gordon (1985). The discussion here again follows Gordon

Sensor Response	Radiometric Variable	Formula
cosine 0 to $\pi/2$	E_d or E_u	$\gamma = \sin^{-1}(\sqrt{u})$
cosine 0 to γ_D	radiance L	$\gamma = \sin^{-1}(\sin \gamma_D \sqrt{u})$
isotropic 0 to $\pi/2$	E_{od} or E_{ou}	$\gamma = \cos^{-1}(u)$
isotropic 0 to π	E_o	$\gamma = \cos^{-1}(1 - 2u)$
Lambertian reflector	bottom reflection	$\theta_r = \sin^{-1}(\sqrt{u})$

Table F.1: Formulas for randomly choosing sensor emission angles γ for commonly used radiometric variables, and for reflection angles θ_r for Lambertian surfaces. u is a $\mathcal{U}[0, 1]$ random number.

and will start with the basic reciprocity principle and will outline the development that leads to the simple result actually used in the BMC calculations. The presentation here has more discussion and a somewhat different formulation of the final result than that of Gordon.

Assume that the ocean coordinate system has depth positive downward. Let $\hat{\mathbf{n}}$ be the unit outward normal to the mean sea surface, i.e., $\hat{\mathbf{n}} = -\hat{\mathbf{z}}$. Let $\hat{\boldsymbol{\xi}}$ be a unit direction vector. For light rays traveling downward, $\hat{\boldsymbol{\xi}} \cdot \hat{\mathbf{n}} < 0$, and $\hat{\boldsymbol{\xi}} \cdot \hat{\mathbf{n}} > 0$ for rays traveling upward. Let \vec{x}_B represent the spatial location of the mean sea surface (the boundary of the ocean volume), and let \vec{x}_D represent the in-water location of the detector.

Figure F.2 shows the original (forward) and adjoint (backward) ray tracing problems used to develop the BMC technique. This figure shows the various quantities seen in Case's reciprocity relation below.

Case's general reciprocity relation is

$$\begin{aligned}
 & \int_{\hat{\boldsymbol{\xi}} \cdot \hat{\mathbf{n}} < 0} d\Omega(\hat{\boldsymbol{\xi}}) \int_B dB |\hat{\boldsymbol{\xi}} \cdot \hat{\mathbf{n}}| \left[L_1(\vec{x}_B, \hat{\boldsymbol{\xi}}) L_2(\vec{x}_B, -\hat{\boldsymbol{\xi}}) - L_2(\vec{x}_B, \hat{\boldsymbol{\xi}}) L_1(\vec{x}_B, -\hat{\boldsymbol{\xi}}) \right] \\
 = & \int_{\Xi} d\Omega(\hat{\boldsymbol{\xi}}) \int_V dV \left[\frac{1}{n_w^2} L_1(\vec{x}, -\hat{\boldsymbol{\xi}}) S_2(\vec{x}, \hat{\boldsymbol{\xi}}) - \frac{1}{n_w^2} L_2(\vec{x}, \hat{\boldsymbol{\xi}}) S_1(\vec{x}, -\hat{\boldsymbol{\xi}}) \right]. \quad (\text{F.34})
 \end{aligned}$$

In this equation, subscript 1 refers to the original problem, and subscript 2 refers to the adjoint problem. The area integration over “B” represents an integration over the bounding surface (the sea surface) of the ocean, and the volume integration over “V” represents an integration over the volume of the water column. On the left hand side of the equation, the integration over direction with $\hat{\boldsymbol{\xi}} \cdot \hat{\mathbf{n}} < 0$ is an integration over all downward directions, i.e., directions incident onto the sea surface. n_w is the water index of refraction. The radiances and sources in this equation are as follows:

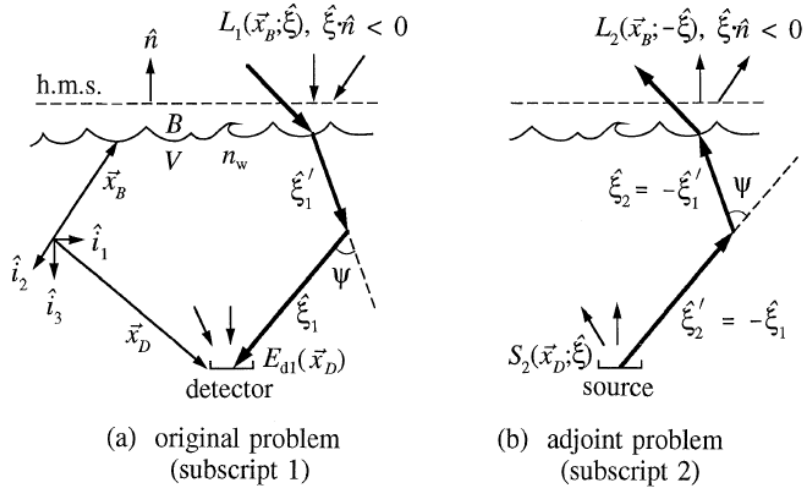


Figure F.2: Illustration of the original (forward) and adjoint (time-reversed) problems used to develop the BMC method. This figure shows the in-water irradiance E_d being computed. Reproduced from Fig. 6.2 of *Light and Water*.

- $L_1(\vec{x}_B, \hat{\xi})$ is the incident sky radiance falling onto the sea surface (since the integration is over downward directions, $\hat{\xi} \cdot \hat{\mathbf{n}} < 0$) in the original problem; thus $L_1(\vec{x}_B, \hat{\xi}) = L_{\text{sky}}(\hat{\xi}) \neq 0$.
- $L_1(\vec{x}_B, -\hat{\xi})$ is the water-leaving radiance in the original problem ($-\hat{\xi}$ is upward for the integration over $\hat{\xi} \cdot \hat{\mathbf{n}} < 0$).
- $L_2(\vec{x}_B, \hat{\xi})$ is the radiance incident onto the sea surface in the adjoint problem. This is zero because the light source in the adjoint problem is the emitting detector, not the sky.
- $L_2(\vec{x}_B, -\hat{\xi})$ is the water-leaving radiance in the adjoint problem. ($-\hat{\xi}$ is upward since the integration is over downward directions.) This is the water-leaving radiance generated by the BMC ray tracing. Call $L_2(\vec{x}_B, \hat{\xi}) = L_{2w}(\vec{x}_B, \hat{\xi})$ as a reminder that this is water-leaving (subscript “w”) radiance.
- $L_1(\vec{x}, -\hat{\xi})$ is the in-water radiance in the original problem.
- $L_2(\vec{x}, \hat{\xi})$ is the in-water radiance in the adjoint problem.
- $S_1(\vec{x}, -\hat{\xi})$ is the internal source in the original problem; this is 0.
- $S_2(\vec{x}, \hat{\xi})$ is the internal source in the adjoint problem. This generates the radiance emitted by the detector.

Consider the left hand side (LHS) of Eq. (F.34). $L_2(\vec{x}_B, \hat{\xi})$ is the radiance incident onto the sea surface in the adjoint problem. However, in the adjoint problem, there is water-leaving radiance from the internal source, but no incident sky radiance, so this term is zero. Thus the LHS reduces to

$$LHS = \int_{\hat{\xi} \cdot \hat{\mathbf{n}} < 0} d\Omega(\hat{\xi}) \int_B dB |\hat{\xi} \cdot \hat{\mathbf{n}}| \left[L_1(\vec{x}_B, \hat{\xi}) L_2(\vec{x}_B, -\hat{\xi}) \right].$$

This can be rewritten as

$$LHS = \int_B dB \left[\int_{\hat{\boldsymbol{\xi}} \cdot \hat{\mathbf{n}} < 0} d\Omega(\hat{\boldsymbol{\xi}}) |\hat{\boldsymbol{\xi}} \cdot \hat{\mathbf{n}}| L_2(\vec{x}_B, -\hat{\boldsymbol{\xi}}) \right] L_{\text{sky}}(\hat{\boldsymbol{\xi}}).$$

The quantity in brackets is the irradiance leaving the sea surface in the $-\hat{\boldsymbol{\xi}}$ (upward) direction from an area element dB of the sea surface at surface location \vec{x}_B . This irradiance is weighted by the sky radiance $L_{\text{sky}}(\hat{\boldsymbol{\xi}})$ in the corresponding (ray-reversed) downward direction. Suppose the BMC code has the sky directions are discretized into finite θ, ϕ bins of size 10 deg in θ except for the polar cap and the near-horizon band, and 15 deg in ϕ . Call these bins $Q(u, v)$, with u labeling the discrete θ bins and v labeling the ϕ bins. Collecting all rays from any point on the sea surface corresponds to the integration over the boundary B in the equation.

Note that $dB d\Omega |\hat{\boldsymbol{\xi}} \cdot \hat{\mathbf{n}}| L_w^*(\vec{x}_B, -\hat{\boldsymbol{\xi}})$ has units of power. The integrals over the sea surface (B) and direction (Ω) give the total power leaving the sea surface. The power leaving the sea surface in the direction of each $Q(u, v)$ bin is the weight $w(u, v)$ accumulated during the ray tracing. The total power leaving the sea surface is then the sum of the weights $w(u, v)$ over all u, v directions. Let $L_{\text{sky}}(u, v)$ be the directionally averaged sky radiance over bin $Q(u, v)$. Then the integral over direction in the last LHS equation can be discretized as a sum over the sky bins:

$$LHS = \sum_u \sum_v w(u, v) L_{\text{sky}}(u, v). \quad (\text{F.35})$$

In this equation, the weights are computed by ray tracing, and the sky radiances are obtained from a model of sky radiance as a function of solar zenith angle and atmospheric conditions. Note that the evaluation of the LHS of the reciprocity relation does not depend on the type of in-water radiometric variable being computed.

Now consider the right hand side (RHS) of Eq. (F.34). The internal source in the original problem, $S_1(\vec{x}, -\hat{\boldsymbol{\xi}})$, is zero. This reduces the RHS to

$$RHS = \int_{\Xi} d\Omega(\hat{\boldsymbol{\xi}}) \int_V dV \frac{1}{n_w^2} L_1(\vec{x}, -\hat{\boldsymbol{\xi}}) S_2(\vec{x}, \hat{\boldsymbol{\xi}}). \quad (\text{F.36})$$

To make further progress in evaluating the RHS, we must pick a particular source function for the adjoint problem. *This is where the type of radiometric variable being estimated enters the problem.*

Suppose we wish to compute the downwelling plane irradiance at some point $\vec{x} = \vec{x}_D$ in the water body; call this $E_d(\vec{x}_D)$. We have seen in Section F.6 that simulation of E_d , which is measured by a cosine-response detector, requires that the emission function also have a cosine emission pattern. Let this emission or source function be

$$S_2(\vec{x}, \hat{\boldsymbol{\xi}}) = \begin{cases} -J_o \hat{\boldsymbol{\xi}} \cdot \hat{\mathbf{z}} \delta(\vec{x} - \vec{x}_D) & \text{if } \hat{\boldsymbol{\xi}} \cdot \hat{\mathbf{z}} < 0 \\ 0 & \text{if } \hat{\boldsymbol{\xi}} \cdot \hat{\mathbf{z}} > 0. \end{cases} \quad (\text{F.37})$$

Here J_o is a scale factor (which will be determined below by the total energy emitted by the source). For downwelling E_d , the emitted rays are upward, so $\hat{\boldsymbol{\xi}} \cdot \hat{\mathbf{z}} < 0$ and the source term is positive. The Dirac delta function $\delta(\vec{x} - \vec{x}_D)$ places the source at the detector location \vec{x}_D . Inserting this source function into Eq. (F.36), noting that the delta function when

integrated over the volume of the ocean just picks out the detector location, and noting that the source function is 0 for $\hat{\xi} \cdot \hat{z} > 0$, the RHS becomes

$$\begin{aligned}
RHS &= \int_{\hat{\xi} \cdot \hat{z} < 0} d\Omega \int_V dV \frac{1}{n_w^2} L_1(\vec{x}, -\hat{\xi}) \left[-J_o \hat{\xi} \cdot \hat{z} \delta(\vec{x} - \vec{x}_D) \right] \\
&= -\frac{J_o}{n_w^2} \int_{\hat{\xi} \cdot \hat{z} < 0} d\Omega L_1(\vec{x}_D, -\hat{\xi}) \hat{\xi} \cdot \hat{z} \\
&= \frac{J_o}{n_w^2} \int_{\hat{\xi} \cdot \hat{z} > 0} d\Omega L_1(\vec{x}_D, \hat{\xi}) \hat{\xi} \cdot \hat{z} \\
&= \frac{J_o}{n_w^2} E_d(\vec{x}_D). \tag{F.38}
\end{aligned}$$

The last equation makes use of the definition of E_d as a cosine-weighted integral of the radiance over the downwelling directions.

The total power emitted by the source function in the adjoint problem, P_2 , for the source of Eq. (F.37) is

$$\begin{aligned}
P_2 &= \int_{\hat{\xi} \cdot \hat{z} < 0} d\Omega \int_V dV S_2(\vec{x}, \hat{\xi}) \tag{F.39} \\
&= -J_o \int_{\hat{\xi} \cdot \hat{z} < 0} d\Omega \hat{\xi} \cdot \hat{z} \\
&= J_o \int_0^{\pi/2} \int_0^{2\pi} \cos \theta \sin \theta d\theta d\phi \\
&= \pi J_o. \tag{F.40}
\end{aligned}$$

Equating the last RHS form from Eq. (F.38) and the LHS form from Eq. (F.35), and using $J_o = P_2/\pi$ from the last equation, gives the desired estimate of E_d :

$$E_d = \pi n_w^2 \sum_u \sum_v \frac{w(u, v)}{P_2} L_{\text{sky}}(u, v).$$

The weight array $w(u, v)$ gives the total weights accumulated in each sky bin; tracing more rays gives larger $w(u, v)$ values. If each emitted ray has an initial weight of 1, then the total emitted power is $P_2 = N_{\text{rays}}$, where N_{rays} is the total number of rays emitted by the detector. The quantity

$$f(u, v) = \frac{w(u, v)}{N_{\text{rays}}} \tag{F.41}$$

is then the *fraction* of the emitted power that is received by bin $Q(u, v)$.

The procedure is similar for other radiometric variables. For estimation of E_u the source function is chosen to be

$$S_2(\vec{x}, \hat{\xi}) = \begin{cases} 0 & \text{if } \hat{\xi} \cdot \hat{z} < 0 \\ J_o \hat{\xi} \cdot \hat{z} \delta(\vec{x} - \vec{x}_D) & \text{if } \hat{\xi} \cdot \hat{z} > 0. \end{cases} \tag{F.42}$$

This again leads to $P_2 = \pi J_o$ as in Eq. (F.40). For estimation of scalar irradiances E_{od} and E_{ou} , the source functions are those of Eq. (F.37) and (F.42), but without the $\hat{\xi} \cdot \hat{z}$

factors. These sources when used in Eq. (F.39) lead to $P_2 = 2\pi J_o$. The source function for computation of total scalar irradiance E_o is

$$S_2(\vec{x}, \hat{\xi}) = J_o \delta(\vec{x} - \vec{x}_D) \text{ for all } \hat{\xi},$$

which leads to $P_2 = 4\pi J_o$. Finally, the source function for computation of radiance by a sensor at location \vec{x}_D pointed in direction $\hat{\xi}_D$ is

$$S_2(\vec{x}, \hat{\xi}) = J_o \delta(\hat{\xi} - \hat{\xi}_D) \delta(\vec{x} - \vec{x}_D) \text{ for all } \hat{\xi}.$$

This source, when used in Eq. (F.39), leads to $P_2 = J_o$. In this case, J_o is the total power emitted.

Thus all of these radiometric variables are estimated by an equation of the form

$$\text{Radiometric variable} = f_\pi n_w^2 \sum_u \sum_v f(u, v) L_{\text{sky}}(u, v), \quad (\text{F.43})$$

where the scale factor $f_\pi = 1, \pi, 2\pi$, or 4π , depending on the variable being estimated.

Equation (F.43) is the fundamental equation used to process the accumulated weights in a BMC simulation. The normalized weights show what fraction of the sky radiance from each direction contributes to the in-water radiometric variable of interest. The n_w^2 factor accounts for the decrease in solid angle when crossing the sea surface from air to water (the n^2 law for radiance; see Section 13.1.2). The f_π factor accounts for the effects of the detector angular response.

Note that no mention has been made of the inherent optical properties or their spatial distribution. Therefore Eq. (F.43) is completely general and applicable to any water body.

For ease of reference, Table F.2 summarizes these results.

Radiometric Quantity	Emission Pattern	scale factor f_π
E_d	cosine upward	π
E_u	cosine downward	π
E_{od}	isotropic upward	2π
E_{ou}	isotropic downward	2π
E_o	isotropic	4π
L	collimated	1

Table F.2: Emission patterns and scale factors for various radiometric quantities in backward Monte Carlo calculations.

F.8 The BRDF as a PDF

Equation (13.44) of the BRDF Section 13.6,

$$L_r(\theta_r, \phi_r) = \int_{2\pi_i} L_i(\theta_i, \phi_i) BRDF(\theta_i, \phi_i, \theta_r, \phi_r) \cos \theta_i d\Omega_i,$$

shows how the BRDF is used in the radiative transfer equation (e.g., in HydroLight), which is always working with *radiances*. In Monte Carlo simulations, you are tracking many individual *rays* as they interact with the medium and its boundary surfaces. In this case, the BRDF must be used as a *probability distribution function* (PDF) to determine the direction and weight of the reflected ray whenever a ray hits the boundary surface. This is a tricky business, and the step-by-step process is as follows².

F.8.1 Computing the Reflected Ray Weight and Direction from a BRDF

The problem statement is simple:

Given: A ray with weight w_i is incident onto the surface in direction (θ_i, ϕ_i) . The BRDF of the surface is known.

Needed: The weight w_r and direction (θ_r, ϕ_r) of the reflected ray.

Since the input direction (θ_i, ϕ_i) is known, $BRDF(\theta_i, \phi_i, \theta_r, \phi_r)$ can be viewed as an (unnormalized) bivariate PDF for the reflected angles θ_r and ϕ_r . Note that, in general, these angles are correlated. Proceed as follows:

1. Compute the directional-hemispherical reflectance for the given (θ_i, ϕ_i) :

$$\begin{aligned} \rho^{\text{dh}}(\theta_i, \phi_i) &= \iint_{2\pi_i} BRDF(\theta_i, \phi_i, \theta_r, \phi_r) \cos \theta_r d\Omega_r \\ &= \int_0^{2\pi} \int_0^{\pi/2} BRDF(\theta_i, \phi_i, \theta_r, \phi_r) \cos \theta_r \sin \theta_r d\theta_r d\phi_r. \end{aligned} \quad (\text{F.44})$$

2. The reflected ray weight is then

$$w_r = \rho^{\text{dh}}(\theta_i, \phi_i) w_i. \quad (\text{F.45})$$

3. Compute the cumulative distribution function (CDF) for ϕ_r by

$$CDF_\phi(\phi_r) = \frac{1}{\rho^{\text{dh}}(\theta_i, \phi_i)} \int_0^{\phi_r} \int_0^{\pi/2} BRDF(\theta_i, \phi_i, \theta, \phi) \cos \theta \sin \theta d\theta d\phi. \quad (\text{F.46})$$

Note that the directional-hemispherical reflectance is being used to convert the BRDF into a normalized bivariate PDF for θ_r and ϕ_r . We are then “integrating out” the θ_r dependence to leave a PDF for ϕ_r , which is then being used to construct the CDF for ϕ_r .

²This development is based on Modest (1993).

4. Draw a random number u from a uniform $[0,1]$ distribution. Solve the equation

$$u = CDF_{\phi}(\phi_r) \quad (\text{F.47})$$

for ϕ_r . This is the randomly determined azimuthal angle of the reflected ray.

5. Compute the CDF for angle θ_r from

$$CDF_{\theta}(\theta_r) = \frac{\int_0^{\theta_r} BRDF(\theta_i, \phi_i, \theta, \phi_r) \cos \theta \sin \theta d\theta}{\int_0^{\pi/2} BRDF(\theta_i, \phi_i, \theta, \phi_r) \cos \theta \sin \theta d\theta} . \quad (\text{F.48})$$

Note that the angle ϕ_r determined in step 4 is used in the BRDF in Eq. (F.48) when evaluating the θ integrals. This accounts for the correlation between θ_r and ϕ_r in the determination of the reflection angles.

6. Draw a new random number u from a uniform $[0,1]$ distribution and solve the equation

$$u = CDF_{\theta}(\theta_r) \quad (\text{F.49})$$

for θ_r . This is the randomly determined polar angle of the reflected ray. You can now send the new ray on its way.

For all but the simplest BRDFs, Eqs. (F.44) to (F.49) all must be evaluated numerically for each ray, which can be an enormous computer cost when billions of rays are being traced.

F.8.2 A Simple BRDF Example

The Minnaert BRDF³ is

$$BRDF_{\text{Minnaert}}(\theta_i, \phi_i, \theta_r, \phi_r) = \frac{\rho}{\pi} (\cos \theta_i \cos \theta_r)^k . \quad (\text{F.50})$$

Note that for $k = 0$ this reduces to the Lambertian BRDF of Section 13.7.

Equations (F.44) to (F.49) can be evaluated analytically for the Minnaert BRDF. Equation (F.44) evaluates to

$$\rho^{\text{dh}} = \frac{2\rho}{k+2} \cos^k \theta_i ,$$

which reduces to $\rho^{\text{dh}} = \rho$ for a Lambertian surface. Equation (F.46) gives just

$$CDF_{\phi}(\phi_r) = \frac{\phi_r}{2\pi} .$$

Inserting this into Eq. (F.47) and solving for ϕ_r gives

$$\phi_r = 2\pi u . \quad (\text{F.51})$$

Thus the azimuthal angle is uniformly distributed over 2π radians. The CDF for θ_r as given by (F.48) is

$$CDF_{\theta}(\theta_r) = 1 - \cos^{k+2} \theta_r .$$

³This BRDF was invented to explain the curious fact that the full moon appears almost uniformly bright from the center to the edge of the lunar disk. If the lunar dust were a Lambertian reflector, the full moon would appear bright at the center and darker at the edge. However, the Minnaert BRDF agrees with observation over only a limited range of angles.

Equation (F.49) then gives

$$\theta_r = \cos^{-1} \left(\sqrt[k+2]{\mathbf{u}} \right),$$

after noting that $1 - \mathbf{u}$ has the same uniform distribution as \mathbf{u} .

For a Lambertian surface ($k = 0$ in the previous equation), the randomly generated θ_r angles are determined via

$$\theta_r = \cos^{-1} \left(\sqrt{\mathbf{u}} \right), \tag{F.52}$$

which has been discussed in Section E.3. Although not intuitive, this distribution is precisely what is necessary to make the number of reflected rays *per unit solid angle* (which leads to the reflected radiance) proportional to $\cos \theta_r$, as required for a Lambertian surface.

Image Prediction

Chapter 16 outlined the basics of photometry and the associated “classical” visibility theory. Visibility theory as developed in Section 16.6.2 assumes a human observer, and it implicitly assumes that there is adequate ambient lighting to see the target. Whether or not the target is visible then depends only on the apparent visual contrast between the target and the background as given by some measure of contrast such as the Weber contrast of Eq. (16.15) or the Michelson contrast of Eq. (16.20).

Two questions now arise:

1. How must the theory be modified if the observer is not a human eye but is, say, a digital camera?
2. How must the theory be modified if the ambient light is very faint so that there may not be enough light for an eye or camera to see the target regardless of its contrast? This could be the case at night, or deep in the ocean where almost no sunlight penetrates.

The first of these questions in principle has a simple answer. Many of the previous results such as Eq. 16.1 can be used if the photopic luminosity function $K_{cd} \bar{y}(\lambda)$ is replaced by the calibrated spectral response of the camera. An example of a normalized camera response curve is shown in the left panel of Fig. 16.6. However, the hard part is to get the camera scale factor corresponding to the luminous efficacy K_{cd} . Good luck getting a camera manufacturer to give you that. Those folks do not speak the language seen here, they may not have ever determined the camera equivalent of K_{cd} because no customer ever asks for it, and it would probably be a company trade secret anyway.

The best we can do here is take a closer look at the problem of predicting how a digital image of an object will change in brightness, sharpness, and color due to the processes of absorption and scattering as the viewing distance increases. I call this problem *image prediction*, to distinguish it from *image processing* and *image analysis*. Image processing refers to the manipulation of a given digital image, usually in order to improve some measure of its quality. Image analysis is the extraction of information from an image, often using image processing techniques. Image processing and image analysis are applied to a given, usually degraded, image. Image degradation includes darkening of the original image due to absorption of light by water, blurring of image detail due to scattering by the water,

and changes in color due to the wavelength dependence of the water optical properties. The job of image prediction is to compute the degraded image for a given environment, starting with an undegraded image as would be seen at zero distance or through a vacuum.

These topics of this appendix thus include the role of the point spread function (PSF) in image degradation, the boundary conditions commonly used in image prediction calculations, the use of Fourier transfer techniques, and various computational issues. The final sections introduce the modulation transfer function (MTF). These sections are illustrated with predictions of image appearance for various situations. The numerical examples are based on simulations of images under water, but the mathematics are applicable to any absorbing and scattering medium and to any level of ambient light.

As is too often the case, textbooks often present the theory only for continuous functions and say little if anything about the important differences when the same ideas are applied to discrete functions of finite spatial extents, i.e., to digital images. As seen in Section A.5, there are important differences in the Fourier convolution theorem for continuous and discrete cases. No publication I have found says a word about how to convert a point spread function (PSF), which has units of inverse meters squared, into a corresponding non-dimensional function needed for use in the Fourier convolution theorem. The literature also says very little about boundary effects on image prediction. The following sections give those details as I have worked them out so that the two techniques of convolution and Fourier transforms give the same prediction—or close to the same in some cases, but with clearly understood causes for the differences.

G.1 The Importance of the PSF in Image Prediction

The PSF plays a fundamental role in the prediction of how an object appears when seen through an absorbing and scattering medium such as water. Suppose you are looking in a particular direction at point (x_o, y_o) in Fig. G.1 from a distance z_o away from the shark. Most of the light detected in this viewing direction probably comes from point (x_o, y_o) in the image. However, every other point in the scene contributes at least a small amount of light to the detected signal in the viewing direction (x_o, y_o) because of scattering from those other directions into the viewing direction. How much light those other points contribute to a given direction is determined by the value of their PSFs at the angular distances away from the viewing direction.

Let $\mathcal{J}(x, y, 0)$ represent the “bright-dark” pattern of the shark image when seen at distance $z = 0$, or as seen through a vacuum. Thus for a gray-scale digital image and 8 bit resolution, $\mathcal{J} = 0$ for a black pixel and $\mathcal{J} = 2^8 - 1 = 255$ for a white pixel. When viewing a particular point (x_o, y_o) in the image through a vacuum at some distance z_o from the shark, the detected radiance comes only from point (x_o, y_o) and has the value $\mathcal{J}(x_o, y_o, 0)$. However, when viewing point (x_o, y_o) in the through a scattering medium, every point in the scene contributes to the radiance seen at (x_o, y_o) . The concentric white circles in the figure represent contours of $PSF(R, \psi)$. Point (x_1, y_1) thus contributes a radiance of $\mathcal{J}(x_1, y_1, 0) PSF(R_1, \psi_1)$, where angle $\psi_1 = \tan^{-1}(\rho_1/z)$ and $R_1 = z_o/\cos(\psi_1)$. Likewise point (x_2, y_2) contributes radiance $\mathcal{J}(x_2, y_2, 0) PSF(R_2, \psi_2)$, where angle $\psi_2 = \tan^{-1}(\rho_2/z)$ and $R_2 = z_o/\cos(\psi_2)$. The total radiance seen looking toward point (x_o, y_o) from a distance

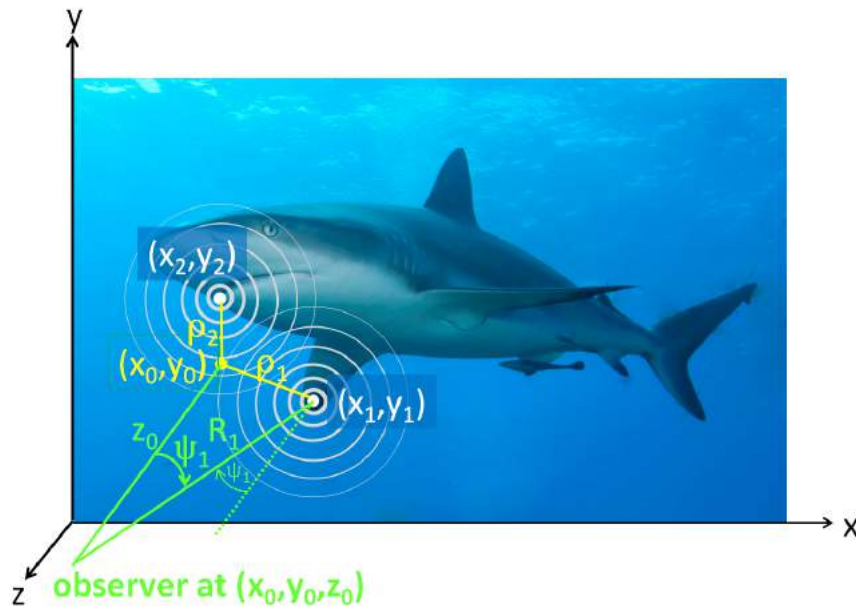


Figure G.1: Illustration of how the PSF describes the contributions of all points (x, y) in an image to the signal at a particular point (x_o, y_o) . The white circles surrounding points (x_1, y_1) and (x_2, y_2) represent the PSFs of those points.

z_o away is given by summing up the contributions from all points in the image:

$$\mathfrak{I}(x_o, y_o, z_o) = \int_{\text{all } x} \int_{\text{all } y} \mathfrak{I}(x, y, 0) PSF(x_o - x, y_o - y, R(x_o - x, y_o - y)) dx dy. \quad (\text{G.1})$$

The connection between points in the image at distance z_o away and the distance and angular variables used for PSFs is

$$\rho = \sqrt{(x_o - x)^2 + (y_o - y)^2}, \quad \psi = \tan^{-1} \left(\frac{\rho}{z_o} \right), \quad R = \frac{z_o}{\cos(\psi)}$$

so that

$$PSF(x_o - x, y_o - y, R) = PSF(R, \psi).$$

Thus, *given the inherent appearance of an object (the appearance at distance $z = 0$ or through a vacuum) and the PSF of the medium, the appearance of the object as seen through the medium can be computed.* In other words, **the point spread function completely characterizes the effect of the environment on an image.**

It is usually implicitly assumed that the field of view (FOV) of the camera (or eye) taking the image is small enough that the range R from the viewing point (x_o, y_o, z_o) to point $(x, y, 0)$ in the image plane is approximately the distance z_o from the observer to the image plane. A 50 mm camera lens with a “full-frame” CCD (corresponding to 35 mm film) has a FOV of 47 deg from one corner of the image to the opposite corner. If the coordinate system is located at the center of the image, the maximum off-axis angle is then 23.5 deg, in which case $R = z_o / \cos(23.5) = 1.09z_o$, and there is less than a 10% error for any pixel in using $R = z_o$. For an 80 mm lens, the error is less than 4%. The assumption

that $R \approx z_o$ is always hiding in the background but never seems to be mentioned. Most publications do not even show the distance in the argument of the PSF; it is just assumed to be the distance from the observer to the image plane. This approximation is adequate for most image prediction. The exception would be the use of a very wide angle lens, in which case the range from the edges of the image to a viewing direction near the center could be much greater than z_o , and the contributions from the edges of the image would be reduced because of the longer ranges through the water. In the equations below I will write R as the distance variable in the PSF to keep the equations general.

Note that in a vacuum and with a perfect optical system, the PSF reduces to a 2-D Dirac delta function: $PSF(x_o - x, y_o - y) = \delta(x_o - x) \delta(y_o - y)$, in which case

$$\mathfrak{I}(x_o, y_o, z_o) = \int_{\text{all } x} \int_{\text{all } y} \mathfrak{I}(x, y, 0) \delta(x_o - x) \delta(y_o - y) dx dy = \mathfrak{I}(x_o, y_o, 0).$$

Thus the image at distance z_o is the same as at distance 0.

G.2 Image Boundary Conditions

We now begin to dig into the ugly details that must be understood in order to write a computer program to predict the appearance of a given digital image. It seems that these details are never discussed in the standard texts, but they are absolutely crucial to getting correct results.

Figure G.2 illustrates the shark of Fig. G.1, now including the surrounding ocean, as might be seen by a diver or a camera. Suppose we are interested just in the shark, so we take a digital photograph that includes the shark and some of the surrounding area. The white dashed line around the shark indicates the area captured in the photograph. This image is of size N_x by N_y pixels, captures a physical area of L_x by L_y meters at the distance z_o of the shark from the camera. I'll refer to the plane a distance z_o from the camera as the object plane, i.e., the (x, y) plane containing the object being photographed, with distance z_o being measured along the z axis, which is pointing towards the camera. Of course, the camera is really collecting light from the full 3-D volume of the ocean, not just from a plane, but we can treat the light as though it is coming from a plane, and the term object plane is well established in optics.

The inherent optical properties (IOPs) of the water are assumed known. These IOPs are the absorption coefficient and the volume scattering function, or some equivalent such as the beam attenuation coefficient, the albedo of single scattering, and the scattering phase function. Knowing the complete set of IOPs, we can compute the point spread function $PSF(R, \psi)$ of the water as a function of off-axis viewing direction ψ and viewing range R .

Suppose that we are looking at a particular pixel in the image of the shark, say the tip of its nose. As shown in the previous section the light received by the camera pixel that images the tip of the shark's nose comes not just from light directly transmitted from the shark's nose to the camera, but also from *all* points in the object plane because of scattering. That is, light from the scuba diver, the submarine, the sea surface, and the bottom can be scattered into the viewing direction of the shark's nose.

Comment on notation: It is common to use the letters i, j, k for dummy summation indices. However, I use i for $\sqrt{-1}$. I will therefore use r and s for indices on spatial variables, e.g., image pixel (x_r, y_s) or (r, s) , and u and v for indices on frequency variables,



Figure G.2: Illustration of a shark as seen in the ocean. The white dotted lines around the shark represent the part of the entire x, y object plane captured by a digital camera.

e.g., μ_u or ν_v . Moreover, since we are dealing with digital (discretized) images, I'll start with the discrete forms of the equations, and show the corresponding results for continuous variables only as needed to point out certain differences between discrete and continuous formulations.

G.3 Convolution

The term *convolution* refers to a mathematical operation that shows how two functions “overlap” to produce a third function. However, as with Fourier transforms, there are important differences for functions of continuous versus discrete variables, and for functions defined on infinite versus finite domains. In the present discussion, these issues are related to the choice of boundary conditions for extending an image beyond its boundary.

G.3.1 Serial Convolution

As before, let $\mathcal{J}(x_r, y_s, R)$ or $\mathcal{J}(r, s, R)$ represent the image at pixel (r, s) , at physical location (x_r, y_s) of the object plane, when viewed from a distance R . $\mathcal{J}(r, s, 0)$ is then the pattern of bright and dark values (typically values from 0 to 255 at a given wavelength) at pixel (r, s) of the image, when seen from zero distance. This is the “actual” or “true” image of the object, as would be seen in the absence of the water. $\mathcal{J}(r, s, 0)$ is assumed given; the problem is to predict what the object will look like when viewed from distance or range R through water with given IOPs and given external lighting conditions (e.g., solar zenith and azimuthal angle and sky conditions). When looking at pixel (r, s) of the scene from a distance R away, the magnitude of the image at pixel (r, s) is given by (recall the continuous variable form in Eq. (G.1))

$$\mathcal{J}(r, s, R) = \sum_{r'=-\infty}^{+\infty} \sum_{s'=-\infty}^{+\infty} \mathcal{J}(r', s', 0) PSF(r - r', s - s', R) \Delta x \Delta y. \quad (\text{G.2})$$

Equation (G.2) is the fundamental equation for predicting what an object will look like when viewed from a distance through an absorbing and scattering medium like water. This is the equation Nature uses. Note that the summations extend over the entire object plane. This equation is usually, although imprecisely, called a *convolution*¹. However, there are two variants of Eq. (G.2), so to be specific I will call Eq. (G.2) the *infinite, discrete, serial convolution* or *infinite, serial convolution summation*.

Equation (G.2) gives exactly the correct prediction of $\mathfrak{J}(r, s, R)$, given $\mathfrak{J}(r, s, 0)$ and the point spread function, but now we encounter the first problem. The camera has not photographed the entire object plane, but just the region containing the shark as outlined in Fig. G.2. We assume that we know what the shark looks like from distance 0, but only for a finite region of the object plane. If we pick the origin of the (x, y) coordinate system to be at the lower left of the image, then we know $\mathfrak{J}(r, s, 0)$ for the pixel range of $r = 0, 1, \dots, N_x - 1$ and $s = 0, 1, \dots, N_y - 1$. (We could just as well pick the origin of the (x, y) system at the middle of the image, for example, in which case $r = -N_x/2, \dots, -1, 0, 1, \dots, N_x/2 - 1$ and $s = -N_y/2, \dots, -1, 0, 1, \dots, N_y/2 - 1$. The choice of the origin does not matter.) Given the image over just a finite area of the object plane, we are forced to truncate the summations of Eq. (G.2) to just the range where the image is available. Equation (G.2) then becomes

$$\mathfrak{J}(r, s, R) = \sum_{r'=0}^{N_x-1} \sum_{s'=0}^{N_y-1} \mathfrak{J}(r', s', 0) PSF(r - r', s - s', R) \Delta x \Delta y, \quad (\text{G.3})$$

for $(r, s) = (0, \dots, N_x - 1, 0, \dots, N_y - 1)$. I will call this equation a *truncated, serial convolution summation* or a *truncated, discrete, serial convolution*. The sums of Eq. (G.3) can be evaluated with the understanding that whenever $r' > r$ or $s' > s$, which would lead to negative indices for the PSF, the term is simply set to zero. The omission of the contributions from pixels outside the area of the image means that there is no contribution to $\mathfrak{J}(r, s, R)$ from the these pixels. This is equivalent to assuming that the ocean is black outside the image, as seen in Fig. G.3. The consequences of omitting contributions from outside of the imaged area will be seen in Section G.6.2.

The standard notation for such a summation of two *non-dimensional* functions $f(r, s)$ and $g(r, s)$ is

$$f * g = f(r, s) * g(r, s) \triangleq \sum_{r'=0}^{N_x-1} \sum_{s'=0}^{N_y-1} f(r', s') g(r - r', s - s'). \quad (\text{G.4})$$

In the present case, $f(r, s)$ is the image and $g(r, s)$ is derived from the PSF in a way that will be seen in Eq. (G.8).

G.3.2 Cyclic Convolution

Rather than just omit the contributions of the object plane lying outside of the imaged area, another possibility is to extend the imaged area in all directions. One way to do this is to tile the entire plane with repeated copies of the imaged area, as illustrated in Fig. G.4.

¹Equation (G.2) is the discrete version of the convolution integral for continuous functions:

$$\mathfrak{J}(x, y, R) = \int_{x'=-\infty}^{+\infty} \int_{y'=-\infty}^{+\infty} \mathfrak{J}(x', y', 0) PSF(x - x', y - y', R) dx' dy'.$$



Figure G.3: The object plane as modeled by a truncated discrete serial convolution. The black area (extending to ∞ in all directions) is not included in the serial convolution summations of Eq.(G.3).



Figure G.4: The object plane as modeled by cyclic convolution and Fourier transforms. The multiple sharks (extending to ∞ in all directions) illustrate the periodicity inherent in the use of cyclic convolution and Fourier transforms.

The convolution equivalent of tiling the plane as in Fig. G.4 is called *cyclic convolution*. For this choice of boundary conditions, Eq. G.3 is replaced by

$$\mathfrak{J}(r, s, R) = \sum_{r'=0}^{N_x-1} \sum_{s'=0}^{N_y-1} \mathfrak{J}(r', s', 0) PSF(r - r' + N_x H(r' - r), s - s' + N_y H(s' - s), R) \Delta x \Delta y, \quad (\text{G.5})$$

and Eq. (G.4) is replaced by

$$f(r, s) \otimes g(r, s) \triangleq \sum_{r'=0}^{N_x-1} \sum_{s'=0}^{N_y-1} f(r', s', 0) g(r-r'+N_x H(r'-r), s-s'+N_y H(s'-s), R), \quad (\text{G.6})$$

where $H(k)$ is the Heaviside step function defined in Section A.5 by

$$H(\ell) \triangleq \begin{cases} 0 & \text{if } \ell \leq 0 \\ 1 & \text{if } \ell > 0 \end{cases}$$

We are now dealing with 2-D functions (images) that are periodic in space. For 2-D functions, the x and y directions wrap around independently. The Heaviside step functions are used to wrap the indices of g whenever $r - r'$ is outside the range of 0 to $N_x - 1$ or $s - s'$ is outside the range of 0 to $N_y - 1$. For example, if $r = 2$ and $r' = 4$ (and $s \geq s'$) then $r - r' = -2$, but $g(-2, s - s')$ is not defined. The value of $2 - 4 + N_x H(4 - 2) = -2 + N_x$, so the value of $g(N_x - 2, s - s')$, which is defined, will be used.

In the present case, both f and g have the same dimensions² N_x by N_y . Equation G.6 is called the 2-D *cyclic convolution* of two discrete functions $f(r, s)$ and $g(r, s)$. The symbol \otimes is used to distinguish cyclic convolution from serial, which uses symbol $*$. Note that two discrete functions f and g of finite size have both serial and cyclic convolutions, but $f * g \neq f \otimes g$. This is illustrated in Fig. G.5.

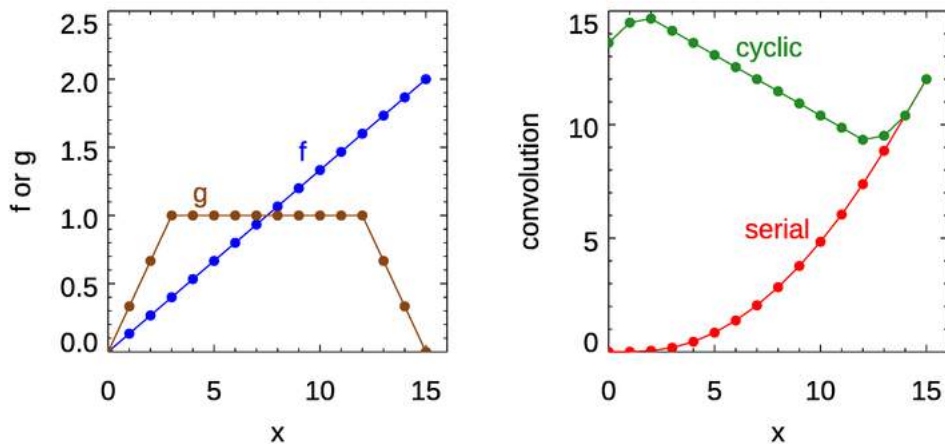


Figure G.5: Illustration of the difference in discrete serial and cyclic convolutions for 1-D functions f and g with $N = 16$ points in each.

As Bracewell (1986, page 363) notes in discussion of the 1-D case, the $NH(r' - r)$ term is explicitly required in computations, although it is commonly omitted when writing

²In general, discrete convolutions do not need to have the same number of points for f and g ; see Bracewell (1986, Chapter 3) for a full discussion. The IDL routine CONVOL uses flags to select the type of convolution and how to extend the data, e.g., EDGE.ZERO = 1 for serial with zero values outside the image as in Fig. G.3, EDGE.WRAP = 1 for cyclic as in Fig. G.4, plus several other options for how to extend the data, which are not relevant here.

equations because the wrapping of indices is understood when doing cyclic convolution. “Understood” maybe, but very confusing when it comes time to write a computer program and you forget that you must account for negative indices when indexing arrays. I will also usually omit the Heaviside functions for brevity of notation but they must be included in computer programs that carry out these calculations..

The spatial periodicity implicit in the use of cyclic convolution is the same as is implicit in the use of Fourier transforms. A Fourier transform represents an image $\mathcal{J}(r, s)$ as sums of sines and cosines whose fundamental wavelengths are L_x and L_y in the x and y directions. The use of Fourier transforms, i.e. sines and cosines, implies that the image is periodic in space, as illustrated in Fig. G.4. In this case, there is a contribution to the image $\mathcal{J}(r, s, R)$ from all points of the object plane, but the contribution comes from an infinite array of sharks, not from the scuba diver or submarine seen in Fig. (G.2).

The computer time required to evaluate either Eq.(G.4) or (G.6) is proportional to $N_x^2 \times N_y^2$. That is, there are $N_x \times N_y$ points to be predicted, each of which requires summations over $N_x \times N_y$ points. This is computationally prohibitive for N_x and N_y larger than a few hundred, and a typical digital image may have several thousand pixels in each direction. The next section develops the mathematical tools needed to evaluate the cyclic convolution summations of Eq. (G.6) using the extremely efficient fast Fourier transform (FFT).

In summary, it must be kept in mind there are three physical problems involved in image prediction:

1. the image as seen in the real world (Fig. G.2),
2. the image as modeled by serial convolution (Fig. G.3),
3. the image as modeled by cyclic convolution or Fourier transforms (Fig. G.4).

These are three physically different radiative transfer problems, and we cannot expect that the serial convolution and cyclic convolution or Fourier techniques will agree exactly with each other, and neither will agree exactly with the image in nature. However, cyclic convolution and Fourier transforms will give identical results because both describe an infinite tiling of the of image. In other words, there are different boundary effects for these three problems due to the three different visual environments outside the area of interest. The best we can hope for is that the contributions from areas of the object plane outside of the given image will be small compared to the contributions from areas within the image. The importance of these boundary effects will be studied in Section G.6.2.

G.4 Image Prediction Using the Convolution Theorem and FFTs

Depending on your choice of boundary conditions, either Eq. (G.3) or Eq. (G.5) is the basis for computing the appearance of an image seen at some distance from the object plane. However, those equations require evaluation for each of the $N_x \times N_y$ image pixels, and each pixel receives a contribution from every other pixel. The computation time is therefore proportional to $N_x^2 \times N_y^2$. The computer time required to evaluate these equations is prohibitive for large images (N_x and N_y greater than 1000 or so).

Let us choose the boundary condition corresponding to the tiling seen in Fig. G.4, which corresponds to cyclic convolution as seen in Eq. (G.6). Combining these equations, we can write

$$\begin{aligned}\mathfrak{J}(r, s, R) &= \mathfrak{J}(r, s, 0) \otimes \{PSF(r, s, R) \Delta x \Delta y\} \\ &\triangleq \mathfrak{J}(r, s, 0) \otimes PSF_{\text{nd}}(r, s, R).\end{aligned}\quad (\text{G.7})$$

The last equation defines a nondimensional PSF as the dimensional PSF times the grid resolution $\Delta x \Delta y$:

$$PSF_{\text{nd}}(r, s, R) \triangleq PSF(r, s, R) \Delta x \Delta y. \quad (\text{G.8})$$

Equation (G.7) is just compact notation for the summations seen in Eq. (G.5), so nothing has yet been gained numerically. However, the extremely important Fourier convolution theorem of Eq. (A.22) states that the Discrete Fourier Transform (DFT) of a cyclic convolution is $N_x N_y$ times the product of the DFTs of the two functions being convolved. When applied to Eq. (G.7), this gives

$$\mathcal{D}\{\mathfrak{J}(r, s, R)\} = N_x N_y \mathcal{D}\{\mathfrak{J}(r, s, 0)\} \odot \mathcal{D}\{PSF_{\text{nd}}(r, s, R)\} \quad (\text{G.9})$$

This can be written as

$$\hat{\mathfrak{J}}(u, v, R) = N_x N_y \hat{\mathfrak{J}}(u, v, 0) \odot \widehat{PSF}_{\text{nd}}(u, v, R), \quad (\text{G.10})$$

where the hat on the symbols denotes the Fourier transform, and the (u, v) arguments label the spatial frequencies in the (x, y) directions. $\hat{\mathfrak{J}}(u, v, 0)$ and $\widehat{PSF}_{\text{nd}}(u, v, R)$ are both 2-D arrays of size $N_x \times N_y$, i.e. matrices. However, the product seen in Eq. (G.10) is an element-by-element (spatial frequency by spatial frequency) multiplication, not matrix multiplication. This is denoted by the \odot symbol, which is sometimes called the Hadamard product of the arrays.

The importance of Eq. (G.9) is that the 2-D DFTs can be evaluated using the Fast Fourier Transform (FFT) algorithm. A 2-D FFT requires computer time proportional to $N_x \log_2 N_x \times N_y \log_2 N_y$. This is much faster for large N_x and N_y . For example, if processing a $N_x = 1024 \times N_y = 1024$ image, the brute force summations of Eq. (G.5) require the order of 10^{12} calculations, whereas the FFTs require the order of 10^8 calculations, which is a factor of 10^4 less computer time.

To predict the image as seen at a distance, the procedure can be summarized as

1. Compute the DFT of the original image $\mathfrak{J}(r, s, 0)$ to get $\hat{\mathfrak{J}}(u, v, 0)$
2. Compute the DFT of the non-dimensional PSF, $PSF_{\text{nd}}(r, s, R)$ to get $\widehat{PSF}_{\text{nd}}(u, v, R)$
3. Multiply the two DFS element by element, including the $N_x N_y$ factors, to get $\hat{\mathfrak{J}}(u, v, R)$
4. Take the inverse DFT of $\hat{\mathfrak{J}}(u, v, R)$ (using the inverse FFT) to get the final image $\mathfrak{J}(r, s, R)$.

All of this can be summarized in one equation:

$$\mathfrak{J}(r, s, R) = \mathcal{D}^{-1}\{\hat{\mathfrak{J}}(u, v, R)\} = \mathcal{D}^{-1}\{N_x N_y \hat{\mathfrak{J}}(u, v, 0) \odot \widehat{PSF}_{\text{nd}}(u, v, R)\} \quad (\text{G.11})$$

A more detailed procedure will be given in the next section.

G.5 Computational Issues

As is usually the case when going from elegant theory to a computer program, confusing computational details arise in the application of Eq. (G.11). These are addressed in this section.

G.5.1 Converting $PSF(R, \psi)$ to $PSF(r, s, R)$

The first issue to resolve is how to obtain the PSF in Cartesian coordinates, $PSF(x_r, y_s, R) = PSF(r, s, R)$, from a PSF given in polar coordinates, $PSF(R, \psi)$. Point spread functions are usually assumed to be azimuthally symmetric, in which case they are given as functions of the off-axis angle ψ and range R . However, the FFT requires PSFs as functions of spatial Cartesian coordinates (r, s) and range. The conversion of $PSF(R, \psi)$ to $PSF(r, s, R)$ leads to ambiguities, and I can find not a single word in any publication or web site discussing the “correct” way to obtain $PSF(r, s, R)$ from $PSF(R, \psi)$.

Figure G.6 shows three (x, y, z) coordinate systems that could be used to specify pixel locations in an image. The dotted lines represent pixels in an image with $N_x = 8$ and $N_y = 4$ pixels. The small circles in some of the lower left pixels are the pixel centers. In the previous equations, I have placed the origin of the (x, y) system at the lower left of the image, as shown by the black axes in the figure. This choice gives the (x, y) values of the pixel centers as $[(r + \frac{1}{2})\Delta x, (s + \frac{1}{2})\Delta y]$ for $r = 0, 1, \dots, N_x - 1$ and $s = 0, 1, \dots, N_y - 1$. The choice of axes origin affects only the labeling of the pixel locations, it does not affect the DFT of $\mathcal{J}(r, s)$ because the DFT simply takes an N_x by N_y array of numbers and carries out the evaluation of Eq. (A.10). If the origin of the axes is placed the center of the image as shown by the blue axes in the figure, the pixel centers would still be at $[(r + \frac{1}{2})\Delta x, (s + \frac{1}{2})\Delta y]$, but with $r = -N_x/2, \dots, 0, \dots, N_x/2 - 1$ and $s = -N_y/2, \dots, 0, \dots, N_y/2 - 1$.

When evaluating the convolution summations (G.5), each point (r, s) is getting its own PSF via the $PSF(r' - r, s' - s, R)$ factor, so it is only the differences in pixel locations that are needed, not their absolute values in any particular coordinate system. In this case, the distance between the pixels in the object plane is

$$\rho(r, s, r', s') = \sqrt{[(r - r')\Delta x]^2 + [(s - s')\Delta y]^2}.$$

Then for a viewing range of R , the angle ψ of pixel (r', s') as seen looking at pixel (r, s) is

$$\psi(r, s, r', s') = \tan^{-1} \frac{\rho(r, s, r', s')}{R}. \quad (\text{G.12})$$

This value of $\psi(r, s, r', s')$ is the value of ψ to be used in $PSF(R, \psi)$ to obtain the PSF for that pair of pixels. No explicit conversion of $PSF(R, \psi)$ to $PSF(r, s, R)$ is needed.

However, in application of the DFT formulas, we must first obtain an N_x by N_y array of PSF values $PSF(r, s, R)$, and that requires specification of a particular coordinate system. Consider a specific example for the $(r, s) = (3, 1)$ pixel shown in Fig. G.6, which is where the origin of the red system is located in the figure. If the black coordinate system seen in the figure is used, then the value of ρ to be used in Eq. (G.12) to obtain the needed value of ψ is

$$\rho(3, 1) = \sqrt{[(3.5\Delta x)^2 + (1.5\Delta y)^2]},$$

This distance is shown by the longer green arrow. If the blue system centered at $(x, y) = (L_x/2, L_y/2)$ is used, then that same pixel has a ρ value of $\rho(3, 1) = \sqrt{[(-\frac{1}{2}\Delta x)^2 + (-\frac{1}{2}\Delta y)^2]}$,

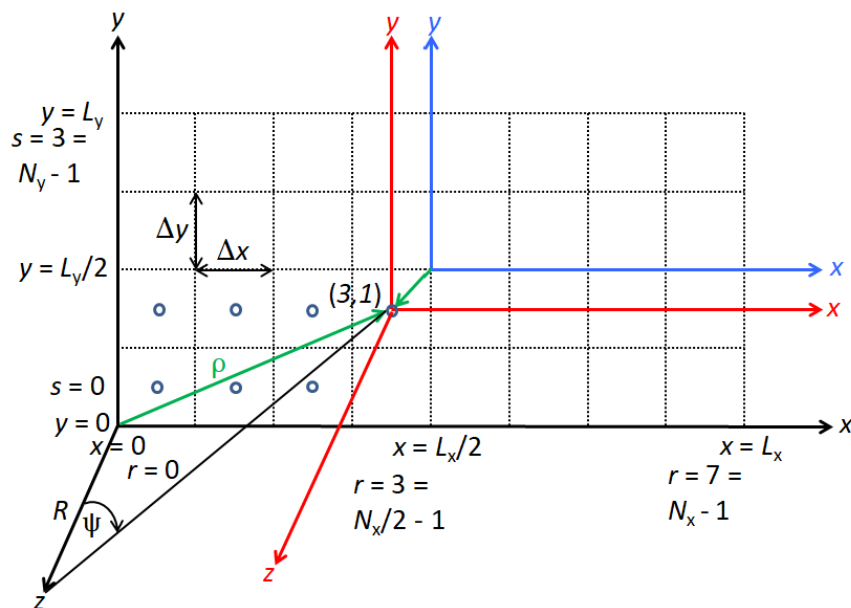


Figure G.6: Possible coordinate systems for labeling pixels and defining $PSF(x, y)$.

which is shown by the shorter green arrow. If the red system is used, then this particular pixel is at the origin, so $\rho = 0$ and $\psi = 0$ is to be used in $PSF(\psi, R)$. Thus, depending on the choice of origin for the coordinate system, you get a different array of numbers for $PSF(r, s, R)$, hence a different prediction for the degraded image $\mathfrak{J}(r, s, R)$. This is philosophically unsatisfactory because *the image in nature cannot depend on the choice of a coordinate system*.

Numerical experimentation shows the following. If the origin is chosen at one corner of the image, say the lower left as for the black system in Fig. G.6, then the predicted image has a gradient of brightness from that corner to the opposite corner, which is clearly incorrect. Moreover, this choice of origin does not reproduce the azimuthal symmetry of the PSF; it is really capturing just one quadrant of the PSF. If the origin is chosen at the exact center of the image, as in the blue system, then the agreement between the convolution prediction of Eq. (G.5) and the FFT prediction of Eq. (G.11) is generally close, but there is a difference of a few percent (at least for my test images) that seems due to an offset of half a pixel between the two predictions. This seems to be due to the fact that $\psi = 0$ occurs in the convolution sums whenever $r = r', s = s'$. However, ψ is never exactly zero for the blue system because the minimum distance from the origin to a pixel center is $\sqrt{(\frac{1}{2}\Delta x)^2 + (\frac{1}{2}\Delta y)^2}$, as illustrated by the short green arrow in the figure. If the origin is shifted to the center of one of the pixels nearest the center of the image, such as the red system, then the agreement between convolution and FFT predictions is almost perfect (discounting the effects of boundaries, to be discussed in the next section).

There is another indication that picking the origin at the center of a pixel as near as possible to the center of the image (e.g., the red system) is the correct choice. For an

azimuthally symmetric function, the 2-D Fourier transform is real³. With this choice, the imaginary part of the transform of $PSF(x, y)$, which depends only on $r = \sqrt{x^2 + y^2}$ and not on $\theta = \tan^{-1}(y/x)$, is identically zero. With any other choice for the origin (e.g., the blue system), the imaginary part of the transform is non-zero.

Figures in the literature (e.g., [Zisserman, 2014](#), page 5) always show $PSF(x, y)$ contoured as a “bull’s eye” in the middle of the (x, y) domain, without any further comment. The implication is that the coordinate system for defining $PSF(r, s)$ is chosen at the middle of the (x, y) region of interest. That choice does seem to work, so numerically things seem to be OK. But philosophically, there is a dependence of $PSF(r, s, R)$ on the choice of coordinate system, which makes it difficult for me to sleep at night. I have not resolved this issue in a way that is satisfactory to my understanding, so I would be most appreciative if some dear reader of this section could clarify this matter for me.

Philosophy aside, $PSF(r, s, R)$ can be determined as follows. Pick an (x, y) coordinate system centered at or near the center of the image. Then for each image pixel, obtain the distance ρ of the pixel center from the origin via

$$\rho = \sqrt{x_r^2 + y_s^2}. \quad (\text{G.13})$$

using whatever (r, s) indexing scheme is desired. Then obtain the off-axis angle ψ via

$$\psi = \tan^{-1} \frac{\rho}{R}. \quad (\text{G.14})$$

Use this value of ψ in the known $PSF(R, \psi)$ to get the values of $PSI(r, s, R)$ for use in Eqs. (G.8) and (G.9). Note that all pixels having the same value of ρ have the same value of $PSF(r, s, R)$, independent of the azimuthal angle of the pixel relative to the x axis of the coordinate system. (There will generally be 4 such pixels for each ρ value.) The generated $PSF(r, s, R)$ thus has the required azimuthal symmetry.

G.5.2 Image Shifting

There is a final issue to discuss before we get down to the business of image prediction. The upper left panel of Fig. G.7 shows an $N_x \times N_y = 32 \times 32$ test pattern used for code debugging. The upper right panel shows this image as seen through water with particular optical properties and rescaled so that the maximum and minimum pixel values are 255 and 0 to show the greatest detail. The lower left panel shows the final image after processing either by the cyclic convolution of Eq. (G.6) or by the FFT Eq. (G.11), with the original image as $f(r, s)$ and the non-dimensional PSF of Eq. (G.8) as $g(r, s)$. Note that the final image has undergone a left circular shift by $N_x/2 + 1$ pixels and a downward circular shift by $N_y/2 + 1$ pixels. This strange shifting of pixel locations does not occur when the image is computed by serial convolution using Eq. (G.3).

The cause of this shifting is explained as follows using Fig. G.8. The 1-D equivalent of Eqs. (G.6) and (G.7) is (omitting the distance arguments)

$$\mathfrak{J}(r) \circledast PSF(r) \triangleq \sum_{r'=0}^{N-1} \mathfrak{J}(r') PSF(r - r' + NH(r' - r)), \quad (\text{G.15})$$

³To show this, convert the 2-D transform in Cartesian coordinates (x, y) of a function $f(x, y)$ to polar coordinates (r, θ) and a function $f(r, \theta) = f(r)$. This leads to what is called a Hankel transform of zero order.

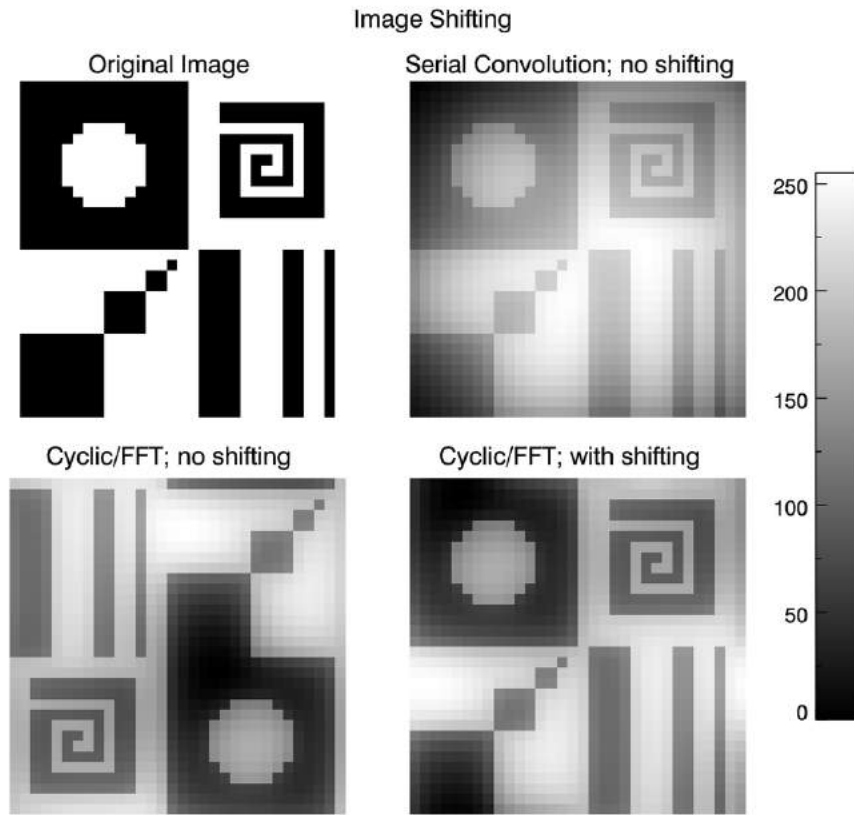


Figure G.7: Effect of convolution type on image shifting. Upper left: a 32x32 test image. Upper right: the degraded image as computed by serial the convolution of Eq. (G.3) as seen through water with particular optical properties. Lower left: the degraded image as computed by either cyclic convolution or FFTs, without any image shifting performed. Lower right: the degraded image as computed by either cyclic convolution or FFTs, after shifting of the original image.

In the upper-left panel of Fig. G.8, the green dots represent an $N = 32$ pixel image $\mathcal{J}(r)$ increasing in brightness with increasing index r . The blue dots represent a nondimensional PSF. The red dots show the 1-D circular convolution as computed by Eq. (G.15). With cyclic convolution, whenever the PSF $r - r'$ index in Eq. (G.15) is negative, the index is circularly shifted, or wraps around, to a positive value in the right half of the PSF before multiplication by the image $\mathcal{J}(r)$. This causes the monotonically increasing $\mathcal{J}(r)$ to be circularly shifted by $N/2$ pixels so that the larger “right half” of the convolved image is now in the left half of the r domain, and the smaller “left half” of the convolved image is in the right half of the r domain. The same thing happens in two dimensions. Whenever the PSF $r - r'$ index in Eq. (G.6) is negative, the index is circularly shifted, or wraps around, to a positive value in the right half of the PSF. Ditto for $s - s'$ being less than 0, in which case the $s - s'$ value wraps around to the top half, respectively, of the PSF. This wrapping of indices between the left/right and top/bottom halves of the PSF causes a wrapping of

the 4 quadrants of the final image. Note that this is a shifting of the 4 quadrants of the original image; it is not a rotation of the original image. The use of FFTs via Eq. (G.11) is exactly equivalent to the use of cyclic convolution, so a degraded image computed by either cyclic convolution (Eq. G.6) or FFTs (Eq. G.11) shows the shift.

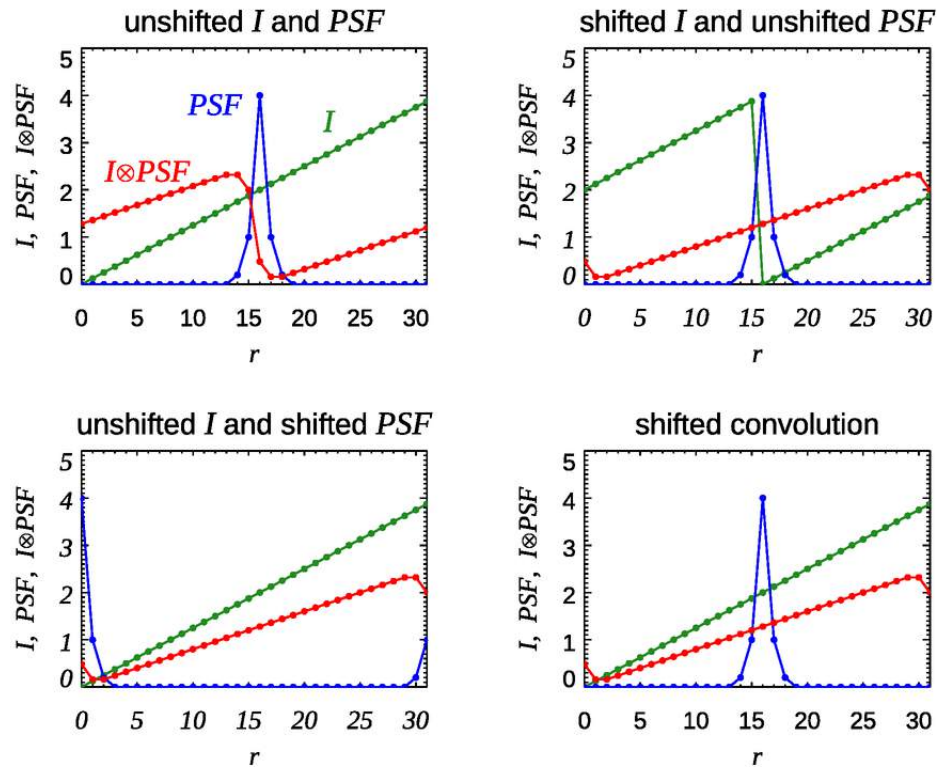


Figure G.8: Illustration of image shifting cause by 1-D circular convolution. The green dots in the upper-left panel represent an $N = 32$ pixel image $\mathcal{J}(r)$ increasing in brightness with increasing index r . The blue dots represent a PSF. The red dots are the circular convolution as computed by Eq. (G.15). The upper-right panel shows the green image shifted before convolution; the lower-left panel shows the PSF shifted before convolution; the lower-right panel shows the final convolution being shifted.

Fortunately the fix for this situation is simple: before processing the original image, do a right circular shift of the image by $N_x/2$ pixels and a top circular shift by $N_y/2$ pixels⁴. Because convolution is commutative, i.e. $f \circledast g = g \circledast f$, it does not matter whether you shift the original image $\mathcal{J}(r, s, 0)$ or the $PSF(r, s, R)$. The upper-right panel of Fig. G.8 shows the original image shifted before convolution; the convolution is now as intuitive expected without any shifting. The lower-left panel of the figure shows the original image unshifted and the PSF shifted before convolution. You can also wait until the end of

⁴In IDL this shift is performed by the command $\mathcal{J}_{\text{shifted}}(r, s) = \text{SHIFT}(\mathcal{J}(r, s), N_x/2, N_y/2)$. In MATLAB, the command is CIRCSHIFT. However, in Fortran 95, the CSHIFT function with positive arguments moves the elements left and down, rather than right and up. So in Fortran the command is $\mathcal{J}_{\text{shifted}}(r, s) = \text{CSHIFT}(\mathcal{J}(r, s), -N_x/2, -N_y/2)$.

processing and shift the final image $\mathcal{J}(r, s, R)$; the result is shown in the lower right panel. You get the same unshifted final result by each of these three shifting options.

However you choose to do it, shift you must or the final image will have a quadrant-shifted layout like that in the lower left of Fig. G.7. In the following simulations using FFTs or cyclic convolution, I shift the original image at the start of processing. Note that no shifting is done if brute force serial convolution is being used (Eq. G.3). Only one shift is needed; shifting both the original image and the PSF leads back to a shifted final image.

G.5.3 Summary of the FFT Algorithm

The above developments give three ways to compute the degraded image $\mathcal{J}(r, s, R)$ given the image at zero viewing distance, $\mathcal{J}(r, s, 0)$, and the PSF at distance R . The first is via brute-force serial convolution via Eq. (G.3). The second is via brute-force cyclic convolution via Eq. (G.6). The third is via the use of FFTs via the following steps:

1. For each pixel (r, s) of the image, compute $PSF(r, s, R)$ via Eqs. (G.13) and (G.14), using a coordinate system centered as near as possible to the center of the image.
2. Compute the nondimensional function $PSF_{nd}(r, s, R)$ using Eq. (G.8).
3. Shift the original image $\mathcal{J}(r, s, 0)$ to the right by $N_x/2 + 1$ pixels and upward by $N_y/2 + 1$ pixels.
4. Compute the DFT of the shifted original image $\mathcal{J}(r, s, 0)$.
5. Compute the DFT of $PSF_{nd}(r, s, R)$.
6. Multiply these two DFTs together element by element, including the $N_x N_y$ factors seen in Eq. (G.10).
7. Take the inverse DFT of this product as seen in Eq. (G.11) to obtain the degraded image $\mathcal{J}(r, s, R)$.

I can find no text or publication that presents all of the results developed above, in particular

1. the $N_x N_y$ factor in the convolution theorem (G.10) (other than one line in Bracewell (1986) for the 1-D version),
2. the $\Delta x \Delta y$ factors on the PSF seen in Eq. (G.8), and
3. the spatial shifting of pixels in the final image when FFTs are used.

If you do not include these $N_x N_y$ and $\Delta x \Delta y$ factors, the FFT predictions will be off by a factor of $N_x N_y \Delta x \Delta y = L_x L_y$ from the convolution summation predictions of Eq. (G.5). It was this disagreement that led me to re-examine everything from scratch and to write this appendix.

In all fairness, much image prediction is done with a PSF that is normalized to 1 at $\psi = 0$. In that case, and ignoring the factors just mentioned, the predicted image $\mathcal{J}(r, s, R)$ does have the correct *relative* pattern of brightness. Using a normalized PSF and ignoring the other factors accounts for the effects of scattering in blurring an image, which is determined by the shape of the PSF. However, a PSF normalized to 1 at $\psi = 0$ cannot account for the quantitative effect of absorption in darkening the image, which depends on the magnitude of the PSF. If you are interested only in the ratio of the brightest pixel to the

darkest, then you do get the correct answer. However, it costs no more in computer time to work with a PSF in units of $1/\text{meter}^2$ and to include the $N_x N_y$ and $\Delta x \Delta y$ factors, in which case the image $\mathfrak{J}(r, s, R)$ has the correct *absolute* brightness. That is, the prediction accounts for the effects of both absorption and scattering on the final image. If all you care about is a brightness ratio, you can obtain that from the properly computed absolute brightness values at the last step of your analysis. Finally, if you do not account for the circular shifting inherent in cyclic convolution and FFTs, then your final image will be spatially shifted as seen above, although relative pixel brightnesses between particular pixels will be correct.

G.6 Understanding Image Predictions

We now have the theory in hand. The remaining sections of this chapter compare images $\mathfrak{J}(r, s, R)$ computed both by the brute-force truncated serial convolution sum of Eq. (G.3) and by the cyclic convolution/FFT algorithm just described, for different PSFs, for different lighting environments, and so on. These simulations illustrate the effects of boundary conditions, water IOPs, and ambient lighting on image degradation. The processing of PNG and JPEG files will also be discussed.

G.6.1 Example Point Spread Functions

The following comparisons use PSFs corresponding to various water IOPs and viewing distances. Figure G.9 shows PSFs computed by Monte Carlo ray tracing for a homogeneous water body with a Petzold average particle phase function and a single-scattering albedo of $\omega_o = 0.8$. The left panel shows the PSF for the first 15 deg of ψ and for nondimensional optical distances between the source and detector of $\tau = cR = 0.1, 1, 5, 10,$ and 20 (c is the beam attenuation coefficient). The colored dots are the centers of the angular bins used to tally the simulated light rays in the Monte Carlo simulations. (The bin collecting rays from $\psi = 0$ to 0.1 deg is tallied as $\psi = 0$, which cannot be plotted in the log scale plot.) These curves show that the magnitude of the PSF decreases as τ increases because of absorption. The flattening of the shape of the curves with increasing τ comes from multiple scattering. The right panel of the figure shows the PSF values normalized to 1 at $\psi = 0$. These curves show that the shape of the PSF starts out very highly peaked near $\psi = 0$ for small τ and eventually becomes relatively flat in ψ as the optical distance increases.

The nearly linear shapes of these PSFs when plotted on log-log axes suggest that a model of the form

$$PSF(\psi, \tau) = \exp[A(\tau) + B(\tau) \ln(\psi)] \quad (\text{G.16})$$

should provide a good fit to the Monte Carlo values. The black lines in the left panel of Fig. G.9 show the model curves when $A(\tau)$ and $B(\tau)$ are determined by fitting the PSF values at 0.15 and 12.5 deg. The model of Eq. (G.16) blows up at $\psi = 0$, so for $\psi < 0.001$ rad, the Monte Carlo value at $\psi = 0$ is used. Equation (G.16) is used to model PSFs in the simulations to follow purely as an expedient to have a simple formula for $PSF(\psi, \tau)$, so that interpolation does not need to be done between the Monte Carlo values. This is more than adequate for the generic studies of this section. For comparison with actual imagery taken in the field, PSFs can be generated using measured (or modeled) IOPs for the particular water body, with interpolation done between the computed values.

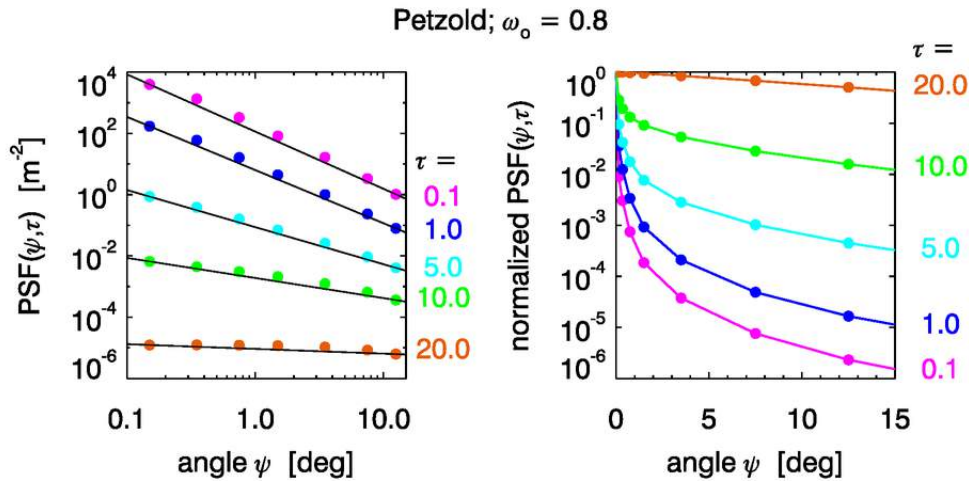


Figure G.9: Example point spread functions. Left panel: The colored dots show PSFs for selected optical distances $\tau = cR$ for $\psi = 0$ to 15 deg, as determined by Monte Carlo simulation. The black lines are the model fits of Eq.(G.16). Right panel: the same curves normalized to 1 at $\psi = 0$.

G.6.2 Boundary Effects

As was illustrated by the figures in Section G.2, the serial convolution and cyclic convolution or FFT calculations are actually solving different radiative transfer problems: the first for a black background outside the image, and the second for a periodic background. This section examines the impact of those differences on the region of interest.

A 64 by 64 pixel uniform gray-scale image with a brightness of 220 (in the range of 0 for black to 255 for white) was created and used for $\mathcal{J}(r, s, 0)$. The left panel of Fig. G.10 shows the gray image $\mathcal{J}(r, s, 0)$. The right panel shows $\mathcal{J}(r, s, R)$ as computed by the serial convolution sum (G.3) as seen at a distance of $R = 5$ m for water with a beam attenuation coefficient of $c = 1 \text{ m}^{-1}$, so that the optical distance is $\tau = 5$. The maximum value (pixels near the center of the image) for $\mathcal{J}(r, s, R)$ is 10.385 for $\tau = 5$, down from 220 in the original image. The minimum value (at the corners of the image) is 5.929. If these values are rounded to the nearest gray scale levels of 10 and 6 and displayed, the final image visually appears almost black, as can be seen in the gray-scale bar at the right of the figure. Therefore, to show the relative appearance of the final image, I rescaled the absolute values of the pixels so that the maximum value at the center has the original value of 220. The right panel of the figure shows the resulting rescaled $\mathcal{J}(r, s, R)$. There is clearly a darkening of the image around the border, which is caused by the surrounding black area of the object plane. However, there is no edge darkening at all for $\mathcal{J}(r, s, R)$ as computed by the FFT technique (image not shown). This is because the periodicity of the FFT has filled the object plane with copies of the original uniformly gray image, so that the entire object plane is uniformly gray and there is no difference between the image $\mathcal{J}(r, s, 0)$ and the entire object plane. All pixels of the FFT image have the value of 10.385 for $\tau = 5$, the same as the center pixels of the convolution image.

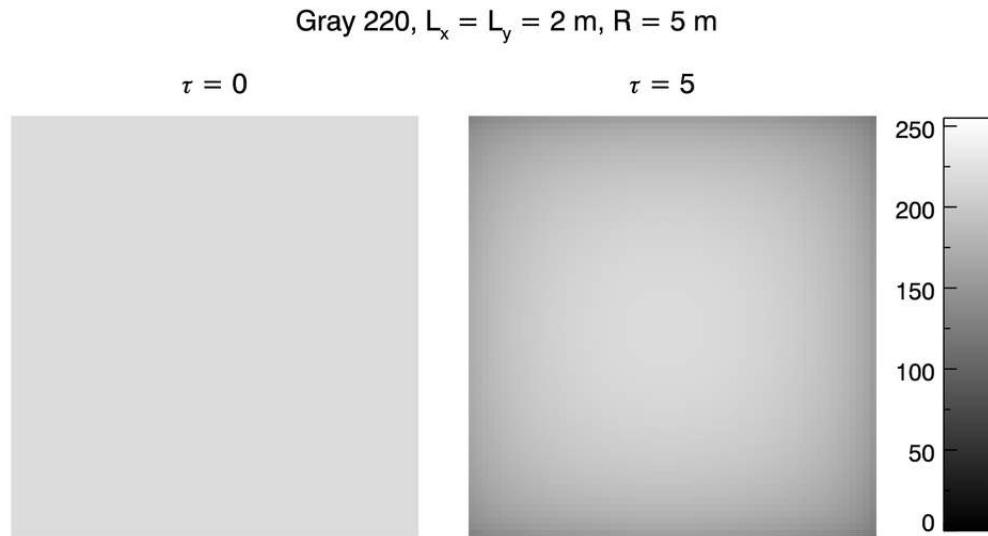


Figure G.10: The left image shows a uniformly gray image $\mathcal{J}(r, s, 0)$ with a gray scale level of 220. The right panel shows the degraded $\mathcal{J}(r, s, \tau = 5)$ computed by serial convolution, with the center pixel normalized to 220.

Figure G.11 shows cross sections through the center of the image at y pixel $s = N_y/2$, for various values of τ . That is, we are holding the viewing distance R constant at 5 m, and changing the water properties via the value of c . The left panel shows the absolute values of the FFT and convolution cross sections for $\tau = 5$. The convolution pixels at $r = 0$ (and at $r = 63 = N_x - 1$) have the value 7.587, so there is a reduction of 27% brightness from the center to the edge of this transect. The right panel of the figure shows cross sections of $\mathcal{J}(r, s, \tau)$ for FFT and for serial convolution for $\tau = 0.1, 1, 5$. The pixel values for each τ have been normalized so that the center pixel is 220. When $\tau = 0.1$ there is a darkening of only about 2% from the center to the edge. The PSF for $\tau = 0.1$ is extremely peaked near $\psi = 0$, so there is very little scattering into a given direction from more than a few pixels away; the boundaries then do not affect the central part of the image. For $\tau = 1$ the reduction is 9%, and for $\tau = 5$ it is 27%. For these larger τ values, the PSF is less peaked and there is more scattering into the central part of the figure from pixels near the boundary. For τ values larger than 5, the percent difference in the edge versus center pixels becomes less and less: 21% at $\tau = 10$ and 6% at $\tau = 20$. This is because the overall image is becoming very dark and all pixels are approaching zero absolute values because of absorption. It is clear from these simulations that the boundary effects can be tens of percent, depending on the image and the water IOPs.

It must be emphasized that there will also be edge effects for the FFT image in the general case where the image is not uniformly gray. However, the FFT edge effects will be different than the convolution edge effects. Because of these differences, *the serial convolution and FFT algorithms will never give identical final images*. However, they will agree closely near the image centers if the IOPs, image size, and viewing geometry are such that the edge effects do not penetrate all the way to the center of the image (e.g., the case

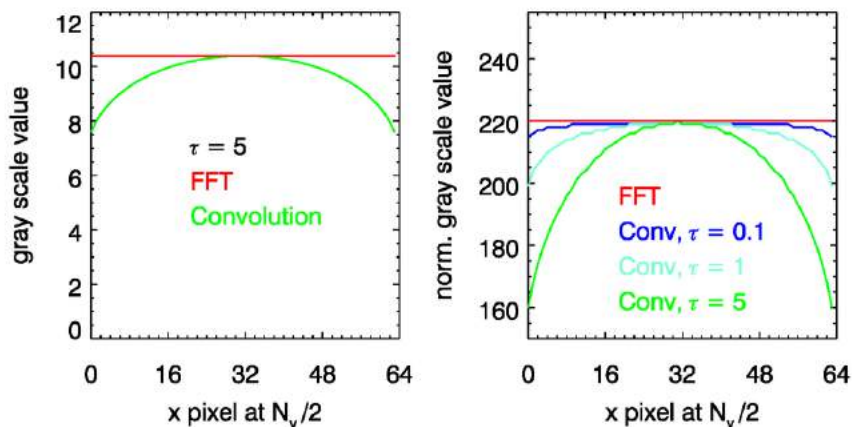


Figure G.11: The left panel shows a cross section of the absolute pixel values of the degraded image $\mathfrak{J}(r, s, \tau = 5)$ as computed by the FFT technique (red) and by serial convolution (green). The right panel shows cross sections of degraded images for FFT and for serial convolution with $\tau = 0.1, 1, 5$. The pixel values for each τ have been normalized so that the center pixel is 220.

for $\tau = 0.1$ in the right panel of Fig. G.11. Unfortunately, it is easy to have situations where the boundary effects influence the entire image (e.g., the $\tau = 5$ in Fig. G.11). Only by comparing convolution and FFT images can some idea be gained of the importance of the boundaries on the central part of the image in any given simulation.

The image processing texts I have seen (which admittedly do not include the computer graphics literature) say very little to nothing about boundary effects. The only comment I’ve found on this is in Szeliski (2021, pages 114-115). He mentions several ways to extend an image beyond its original size. Although not explicitly stated, his idea seems to be that you add a “picture frame” around the original image to create a larger image with the chosen boundary conditions, process the larger image, and then keep only the area of the original image. His “padding” options, as he calls them, are

- *Zero*: Add zeros around the original image. This is implicit in the serial convolution sum, as was seen in Fig. G.3, and Szeliski points out that this option leads to a darkening of the image borders, as was seen in Fig. G.10.
- *Cyclic*: Wrap (repeat or tile) the original image, as was seen in Fig. G.4. This is implicit with the use of Fourier transforms applied to the original image.
- *Constant*: Set all pixels outside the original image to a specified border value (other than zero).
- *Clamp*: Repeat the edge pixels indefinitely. That is, what ever value the image pixel has at $(0, s)$ will be used for all values of $(r < 0, s)$, and so on.
- *Mirror*: Reflect pixels across the image edge.
- *Extend*: Extend the image by subtracting the mirrored version from the edge pixel value.

Szeliski mentions that in computer graphics, these options are called “wrapping modes.”

G.6.3 Point Spread Function Effects

The point spread functions seen in Fig. G.9 show the highly peaked nature of oceanic PSFs. A very sharp peak (e.g., for very small τ values) at means that only neighboring pixels contribute significantly to the light received by a given pixel. That is, only small-angle forward scattering is significant in the image degradation. A broad peak (large τ values) means that pixels far from the viewing direction can contribute significant amounts of light to the viewing direction; i.e., multiple scattering is important.

In some fields of optics, it is common to use Gaussian PSFs, e.g. to model blurring by camera or microscope lenses. A Gaussian PSF has the form

$$PSF(\psi) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}\psi^2\right), \quad (\text{G.17})$$

where the standard deviation σ controls the angular spread of the Gaussian. Figure G.12 compares the oceanic PSF of Fig. G.9 for $\tau = 5$ with the Gaussian of Eq. (G.17) for $\sigma = 0.015$. The Gaussian PSF has much more scattering at small angles (ψ less than a few degrees), and much less at larger angles, than does the oceanic PSF for $\tau = 5$.

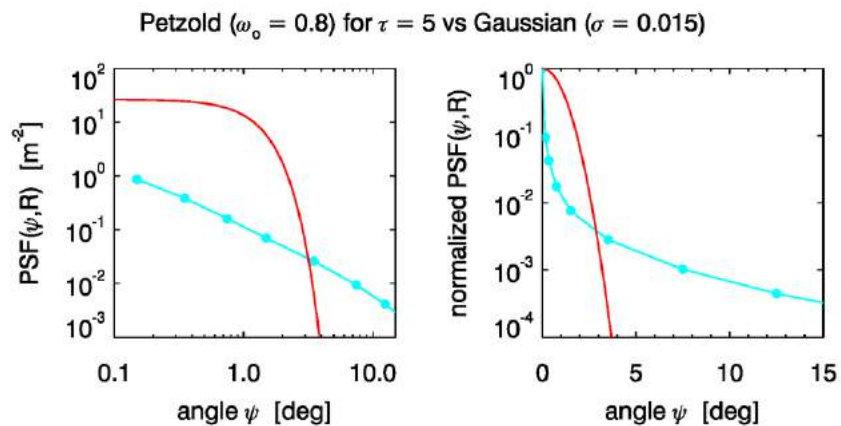


Figure G.12: The oceanic PSF for $\tau = 5$ in Fig. G.9 (aqua dots) and a Gaussian PSF for $\sigma = 0.015$ (red line). Left panel, absolute values; right panel, normalized values.

To study the visual effects of these two PSFs, a standard United States Air Force 1951 resolution chart⁵ was cropped to 256 by 256 pixels for convenient processing with FFTs. The upper left panel of Fig. G.13 shows this image, which is used as $\mathcal{J}(r, s, 0)$. The gray line at $s = 156$ shows the location used for the transects seen in Fig. G.14. This transect was chosen because it cuts through both high and low spatial frequency patterns. The gray line itself was not in the image as processed. The upper right panel of the figure shows the degraded image $\mathcal{J}(r, s, \tau = 5)$ as obtained by brute-force serial convolution (Eq. G.3). The lower left panel shows the corresponding result obtained by FFTs. These two images are visually much the same, but the darkening around the edge of the final image is noticeable in the convolution image. The lower right panel shows the result when the

⁵Obtained as a PNG image at <https://commons.wikimedia.org/wiki/File:USAF-1951.svg>

Gaussian PSF of the left panel of Fig. G.12 is used in the FFT processing. The images for the ocean PSF are much sharper (retain the higher spatial frequencies) than the image for the Gaussian PSF. The large PSF at small angles for the Gaussian blurs out the high frequencies. Indeed, a Gaussian is sometimes used as a low-pass filter on an image for that reason. The important details of how this image was processed and the way in which the absolute values were converted to gray scale values between 0 and 255 is discussed in Section G.6.5.2.

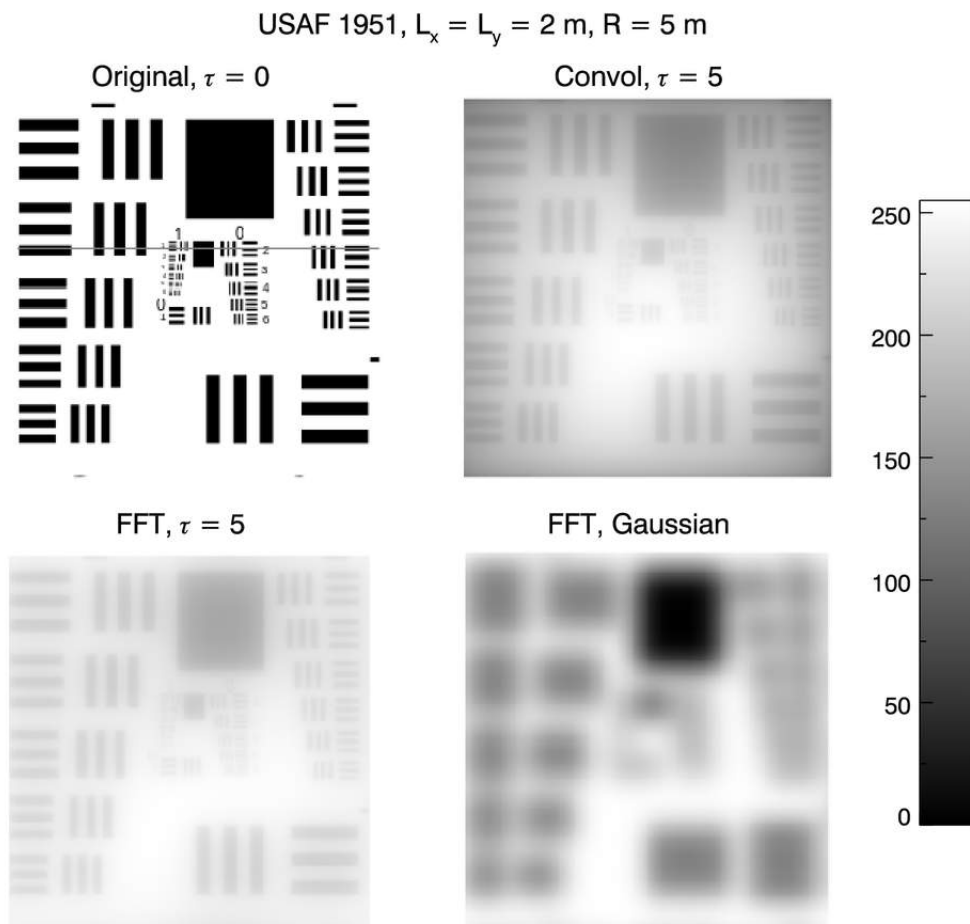


Figure G.13: USAF 1951 resolution chart degraded in various ways. Upper left: the original image $\mathcal{J}(r, s, 0)$. The gray line at $s = 156$ shows the line of the transects plotted in Fig. G.14. Upper right: $\mathcal{J}(r, s, \tau = 5)$ as computed by the serial convolution sum of Eq. (G.3). Lower left: $\mathcal{J}(r, s, \tau = 5)$ as computed by FFTs. Lower right: $\mathcal{J}(r, s)$ as computed by FFTs with the Gaussian PSF of Fig. G.12.

Figure G.14 shows the pixel values along the transect at $s = 156$, which is the gray line in the upper left panel of Fig. G.13. Working from left to right in the left panel and comparing with the upper left panel of Fig. G.13, we see that the transect first passes

through a horizontal black bar (lower pixel values), then 3 narrow black bars, then through an area of small features—i.e. high spatial frequencies—a white area (large values), and finally 3 more vertical and one horizontal bar. The peaks and valleys of the transects show that both the FFT (red) and convolution (green) degraded images preserve all but the very highest spatial frequencies. That is, there is some blurring of the finest spatial features in the middle of the resolution chart, but overall the bars are identifiable except at the smallest scales. The black line at the bottom of the left panel shows the ratio of the convolution prediction to that of the FFT. The ratio is greater than 0.97 over the center third of the image and greater than 0.90 over the center two thirds of the image. However, at the very edges of the image, the convolution value is only 64% of the FFT value because of the darkening of the convolution image caused by the surrounding black boundary. The values for the Gaussian PSF are much larger than those of the ocean PSF and are not shown.

The right panel of the figure shows the transects for all three images, where for each image the absolute values were rescaled so that the brightest pixel in each image is 255. (The brightest pixels are in the whitest area below and left of the image centers, not along the transect line.) For the Gaussian PSF, the normalized curve captures only the largest-scale dark and bright regions, corresponding to the visual impression seen in the lower right panel of Fig. G.13. The details of even the largest of the sets of 3 bars is lost. This is because the Gaussian PSF has much more scattering at small angles ψ , as was seen in Fig. G.12. The Gaussian PSF is not a realistic PSF for most ocean waters.

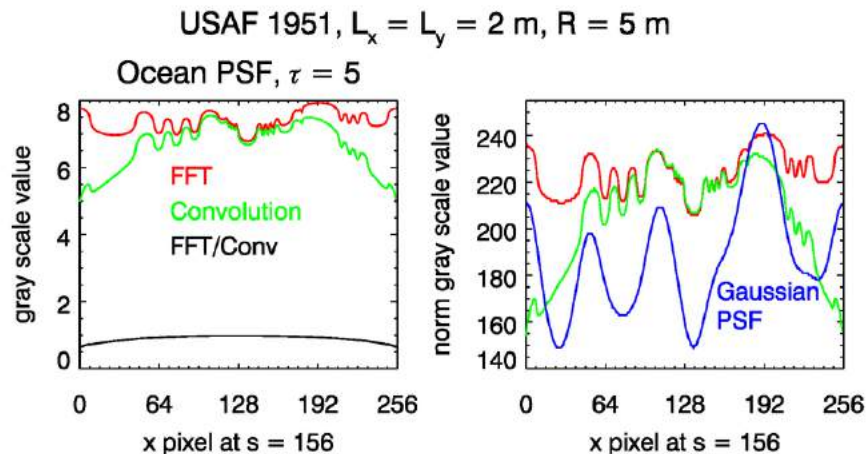


Figure G.14: Cross sections through the resolution chart images of Fig. G.13 at $s = 156$, which is the gray line seen in the upper left panel of Fig. G.13. Left panel: Convolution (green) versus FFT (red) for the ocean PSF for $\tau = 5$. The black line is the ratio of convolution to FFT. The absolute values for the Gaussian PSF are larger and are not shown at this ordinate scale. Right panel: Normalized values computed by scaling the absolute values so that the brightest pixel anywhere in the image is 255 for each of the three images.

G.6.4 Timing

As has been stated, the use of FFTs to evaluate the factors in Eq. (G.11) is *much* faster than the use of either serial or cyclic convolution summations. Table G.1 gives a few examples of run times. Both convolution and FFT times exclude the time needed to evaluate the PSF. If the PSF is evaluated within the convolution summations, the run time roughly doubles. The FFT times include computing the two direct transforms $\hat{\mathcal{J}}(u, v, 0)$ and $\widehat{PSF}_{nd}(u, v, R)$, their multiplication, and the inverse transform of $\hat{\mathcal{J}}(u, v, R)$ to obtain the final $\mathcal{I}(r, s, R)$. This table should make clear why the development of the FFT by Cooley and Tukey (1965) was voted the most important development in applied mathematics in the 20th century.

Given the enormous speed advantage of the FFT algorithm, there is no need to consider convolution summations as anything more than a check on the FFT computer code. The simulations in Sec. G.7 therefore use only FFTs.

Example	image size $N_x \times N_y$	convolution time	FFT time
Fig. G.7	32×32	1.44 sec	< 0.00001 sec
Fig. G.10	64×64	25.6 sec	< 0.00001 sec
Fig. G.13	256×256	6481 sec	0.017 sec
Fig. G.17	512×512	est. 28.8 hours	0.094 sec

Table G.1: Example run times for some of the figures in this tutorial. Run times refer to a single panel in multipanel figures.

G.6.5 Image Formats

Image formats can be confusing. I'll discuss only the two most relevant to underwater imaging because these formats are what are likely to come out of a camera or be found on the web, and which are used in this note.

G.6.5.1 JPEG files

Color JPEG files have 3 *channels*, which contain the amounts of red, green, and blue (RGB) in the image, with the R, G, and B values ranging from 0 (none of that color; the pixel is “turned off”) to 255 (maximum brightness) for each color component. A JPEG image is fundamentally just an array of numbers dimensioned $(3, N_x, N_y)$, where N_x and N_y are the image size in pixels, and the 3 channels contains the brightness values for the R , G , and B channels, respectively.

To process a color JPEG image, i.e. to predict how the image will appear through water from a distance R , extract each of the individual color channels and process them separately as N_x by N_y gray scale images using a PSF appropriate for the water at each of the central R, G, B wavelengths. (I'll use an italic R for the range, and a Latin R for red.) Note that different colors will attenuate and blur differently because of the wavelength dependent IOPs. Thus red will usually be absorbed fastest (due to absorption by water) but scatter less, and green or blue may scatter more but with less absorption. After predicting how much of each color is transmitted to range R , recombine the three components into the 3

channels of the JPEG format, and then display that JPEG file to obtain a color rendition of the image at range R .

It is often the case for large viewing ranges that the final pixel values are very small numbers, in which case the final image will appear very dark with almost no detail visible. To make an image that “looks better,” i.e., is brighter and shows more detail, the final pixel values for each channel can be multiplied by some factor. If that is done, each channel must be multiplied by the same factor so that the relative amounts of R, G, and B remain the same as in the actual image. Otherwise the color of the final image will be distorted.

JPEG files also can contain metadata such as the time and location where the photo was taken, camera settings, or an embedded color table to be used to display the image. The image may be “true color,” which means that the R, G, B values are the amount of each color, 0 to 255 for each channel. This is called the “decomposed” color model and gives $256^3 = 16,777,216$ possible colors. However, an image may have “indexed color,” which means that the number in the channel is the index of an RGB color triple in a color table, which typically has just 256 possible colors. Moreover, image display software (such as Windows Photo Viewer) sometimes seems to have a mind of its own and may decide to transform an image in some unexpected way beyond the user’s control. If you expect to get a final image that looks like a blue or green ocean but instead get a sickly yellow or lurid purple, then the problem probably lies with the color table, not with the mathematics.

G.6.5.2 PNG files

PNG files are a bit trickier. PNG files have 4 channels. The first three are for the RGB colors just as for JPEG files, and channel 4 is called an *alpha channel*. The alpha channel is often called the “transparency” of the image, but the numbers are really “opaqueness.” These numbers can be used to fade from one image to another in a time sequence, or to overlay two images so that you can see both like a double exposure on film. An opaqueness of 255 means maximum opaqueness, so that “nothing gets through;” such a pixel displays as black. An opaqueness of 0 means that the image is transparent; such a pixel displays as white. Image display software of course understands this, so that a PNG file displays correctly when the alpha channel is used.

The USAF 1951 resolution chart used in Section G.6.3 is a PNG file which, when displayed, looks black and white, as seen in the left panel of Fig. G.15 (same as in Fig. G.13). Since this is a black and white image, the 3 color channels are not used (the values are set to 0 for each pixel); the alpha channel is used to create the black and white image. This channel has a pattern of 0’s and 255’s that generate the white (value 0) and black (value 255) pixels when displayed (or values between 0 and 255 for shades of gray). If channel 4 is extracted and used as a gray scale image, the pixel values are “reversed” from the customary gray scale convention of 0 being black and 255 being white. The recommended way to process such an image is as follows:

Option 1 is to reverse the opaqueness channel to get a gray scale image, process, rescale (if needed), and reverse again to get back to opaqueness for display as a PNG image. This is what was done in creating the images of Fig. G.13. The alpha channel of the resolution chart PNG file was reversed via gray scale value = 255 minus alpha channel value. After the convolution processing, the degraded gray scale image $\mathcal{J}(r, s, R)$ had the maximum and minimum gray scale values 8.228 and 3.478, respectively. If these values are simply converted to the nearest integer values of 8 and 3 for use in an image, the resulting image

looks visually almost black and no detail can be seen. It is therefore customary to rescale the actual pixel values to give a broader range of values, which then visually show more of the image pattern. In the simulations of Fig. G.13, this was done by arbitrarily rescaling the pixel values so that the lightest pixel of the degraded image $\mathcal{J}(r, s, R)$ has a value of 255. This was done via⁶

$$\mathcal{J}_{\text{rescaled}}(r, s, R) = \text{INT}\{255 \mathcal{J}(r, s, R) / \text{MAX}\{\mathcal{J}\} + 0.5\} \quad (\text{G.18})$$

If this is done, the rescaled image has a maximum value of 255 and a minimum of 108. These values were then reversed again to create opaqueness, and then inserted into a PNG format file as the alpha channel. That PNG file was then displayed to create the images seen in Fig. G.13 for the various processing options of convolution vs FFT and ocean vs Gaussian PSF.

Option 2 is to process the opaqueness values as they are, rescale to the maximum value, and display. If the opaqueness channel is extracted and processed without alteration, the final maximum pixel value is 8.389 and the final minimum value is 5.401. As before, if these values are simply converted to the nearest integer values of 5 and 2, the resulting image has almost no visible detail. If the maximum image value of 8.389 is again converted to 255 via Eq. (G.18), the minimum is 167. This is a narrower range of values than was obtained when processing the gray scale image as just described. Keep in mind that this analysis is based on 255 displaying as black in the PNG, so the scaling has really set the darkest pixel to completely black, rather than setting the lightest pixel to white as in the normal gray scale viewpoint of the first option.

The degraded images for these two processing and rescaling options are shown in Fig. G.15. These two options do not yield the same degraded images because they are solving different radiative transfer problems. For example, a white Secchi disk viewed against a dark background is not the same as a black Secchi disk viewed against a light background. Even if the maximum and minimum pixel values are in the same locations (e.g., at the center of the Secchi disk and at the corner of the image), the brightness values at those pixels will be different, and they won't even have the same relative values, as was just seen. It can be argued in the present case that the physically realistic way to process and display the PNG data is to reverse the opaqueness values to get a gray scale that corresponds to what the eye would see, process those values, rescale in some way if needed, and then reverse again before display as an alpha channel in a PNG file.

However, there is no one unique way to process the data in an image, and there is no one unique or best way to rescale the degraded image values in order to create an image that “looks good.” Other scalings will be used in Sections G.7.1 and G.7.2. The business of scaling data to get nice looking images requires care. IDL for example has a function BYTSCL (byte scale) that will take your data and rescale it so that the smallest data value maps to 0 (black) and the largest maps to 255 (white). That shows the greatest amount of detail in the displayed image, and that's what I used to create the images in Fig. G.7, where I was interested only in the spatial pattern. However, use of such a function throws out the magnitude information in the data. That is, the difference in the largest and smallest data values might be 10%, or it might be a factor 10, but you can't know from looking at

⁶INT represents a generic function that converts the floating point values of $\mathcal{J}(r, s, R)$ into integers, as required for an image file. In IDL and MATLAB this is implemented via the FIX function. The 0.5 is needed so that the value is converted to the nearest integer, because FIX rounds downward.

the figure. My personal preference is to manually rescale my data in a way that preserves relative magnitudes in some way. Then I know exactly what the image is showing. There is always a trade off between getting an image that looks good and one that maintains fidelity to the physics.

The message here is that you must understand exactly what the image data describe, and then process in a way that makes physical sense for the problem at hand.

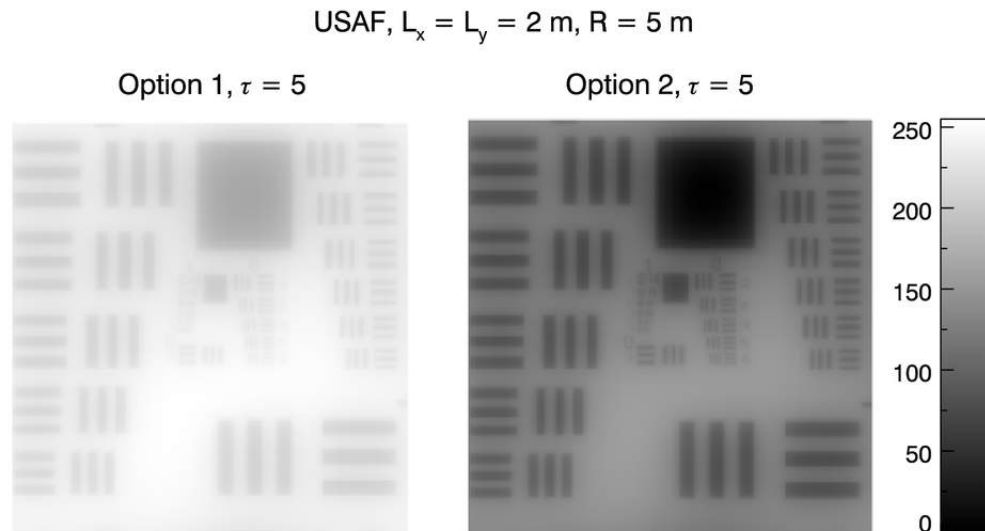


Figure G.15: The USAF resolution chart processed in two different ways. Left: Option 1, using a gray scale obtained from a reversing the PNG alpha channel. The brightest pixel of the degraded image was rescaled to 255. Right panel: the alpha channel processed as is. The rescaling sets the darkest pixel to 0.

G.7 Example Image Simulations

We now consider two realistic simulations for objects in typical ocean waters. The first is a simulation of a Secchi disk viewed as a gray scale image; the second simulates a color image of a sponge.

G.7.1 Secchi Disk Gray-Scale Image Simulations

Consider the simulation of a Secchi disk as seen by a broadband instrument such as the Teledyne Explorer Pro dim-light camera, which outputs a gray scale image. When lowering a Secchi disk from above the surface, there are several complicating factors: (1) the disk is viewed through the sea surface, (2) the IOPs may be depth dependent, and (3) the background radiance depends on depth even if the IOPs are homogeneous. We can eliminate these factors by viewing the disk horizontally at a given depth. The sea surface roughness and IOP stratification are then irrelevant. Assuming that the IOPs and external lighting are horizontally homogeneous, the background radiance is independent of the

viewing range.

Figure G.16 shows the geometry. The camera is looking horizontally in the $+x$ or $\phi = 0$ azimuthal direction at a distance R from the Secchi disk. The Sun can be either in front of the disk, as in the figure, so that the side of the disk seen by the observer is illuminated; or the Sun can be behind the disk so that the observer sees the shaded side. The Sun generates an ambient or background radiance $L_b(\theta = 90, \phi = 180)$ as shown by the thick green arrow in the figure (remember that θ, ϕ give the direction of light propagation, not the viewing direction). Radiance $L(\theta, \phi < 90)$ is incident onto the side of the disk being observed, as shown by the dull green arrows. This radiance generates a plane irradiance $E(\theta = 90, \phi = 0)$ onto the disk. If the disk is a Lambertian reflector, the target radiance reflected toward the observer is

$$L_t(\theta = 90, \phi = 180) = \frac{R_{SD}}{\pi} E(\theta = 90, \phi = 0),$$

where R_{SD} is the reflectance of the Secchi disk. The background and target radiances are used to generate the image of the Secchi disk in the object plane, at range $R = 0$.

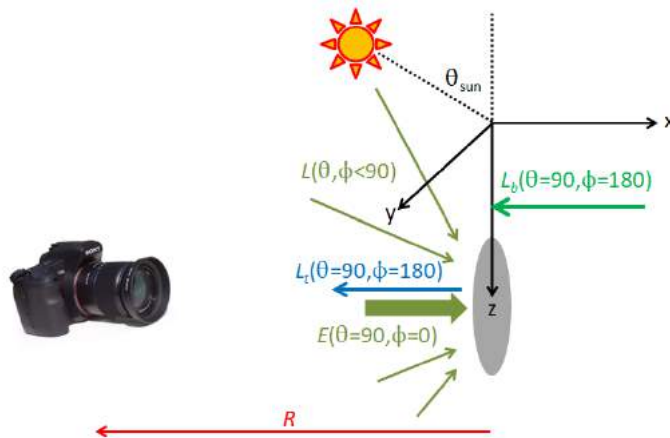


Figure G.16: Geometry for simulation of a Secchi disk when viewed horizontally within the water.

To generate some numbers for numerical simulation, HydroLight was run for the following conditions. Case 1 water with a chlorophyll concentration of 2 mg m^{-3} was used to define the IOPs at a wavelength of 550 nm . The Sun was at a zenith angle of 50 deg in a clear sky. Two runs were done, one with the Sun in front of the disk and one with the Sun in behind. The disk and viewing path were at 15 m depth. The disk was assumed to have a typical Secchi disk reflectance of $R_{SD} = 0.80$. The radiance and irradiance values for these two runs are shown in Table 2.

When the Sun is in front of the disk, the viewing direction is away from the Sun and the background radiance is relatively dark. The front side of the disk is illuminated and is brighter than the background by a factor of 4.514 in the present case. Thus the disk will appear bright against a darker background. Suppose the camera is adjusted so that when the disk is viewed from zero distance, the disk itself has a gray scale value of $GS_t = 220$. The background brightness must be a factor of 4.514 less, so $GS_b = 220/4.514 = 49$. When

the Sun is behind the disk, the camera is looking toward the Sun and the disk is in its own shadow. Then the background is brighter than the disk, and the disk will appear as dark against a brighter background. If the camera is now adjusted so that the bright background has a gray scale value of 220, the disk brightness will be $220 \times 0.528 = 116$. Finally, assume that the disk has a radius of 0.15 m and that the camera field of view sees a 2 m by 2 m region of the object plane, and that the camera image is 512 by 512 pixels. These values were used to generate simulated gray scale images of a Secchi disk.

variable	units	Sun in front	Sun behind
L_b	$\text{W m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$	0.00404	0.01152
E	$\text{W m}^{-2} \text{nm}^{-1}$	0.07162	0.02638
L_t	$\text{W m}^{-2} \text{sr}^{-1} \text{nm}^{-1}$	0.01824	0.00608
L_t/L_b		4.514	0.528
GS_t		220	116
GS_b		49	220

Table G.2: Radiometric quantities used for the Secchi disk simulations. GS_t and GS_b are the gray scale values of the target and background, respectively, for the disk at the object plane at $R = 0$.

Figure G.17 shows the Secchi disk for the Sun in front of the disk and for viewing ranges corresponding to optical distances of $\tau = 0, 2, 5$ and 10. These images were computed using the PSFs seen in Fig. G.9 and the FFT algorithm. The contrast shown in the figure panels was computed using the gray scale values at the center of the disk and at a background point two radii (0.3 m) from the center:

$$C(\tau) = \frac{GS_t(\tau) - GS_b(\tau)}{GS_b(\tau)}.$$

Thus the inherent contrast at $\tau = 0$ is $C(0) = (220 - 49)/49 = 3.49$. As τ increases the image becomes darker and darker. For display, the gray scale values were rescaled so that the background brightness at the background reference point was held constant at 49, while the disk brightness decreases. Visually, the disk is easily detectable through $\tau = 5$, but there is only a faint hint of the disk at $\tau = 10$.

Figure G.18 shows the equivalent sequence of images for the case of the Sun behind the disk. Again, the disk is visible at $\tau = 5$, but it is undetectable at $\tau = 10$. Remember that the contrast is negative if the background is brighter than the target. It is the absolute value that is usually used to determine visibility.

Figure G.19 shows cross sections of the normalized gray scale values for these two sets of simulations.

In these simulations the range was held constant while the water IOPs were changed. If the camera lens is zoomed so that the size of the object plane remains constant, this is equivalent to holding the IOPs constant and increasing the range. In practice, the water IOPs would be constant and the range would increase. If the camera field of view is held

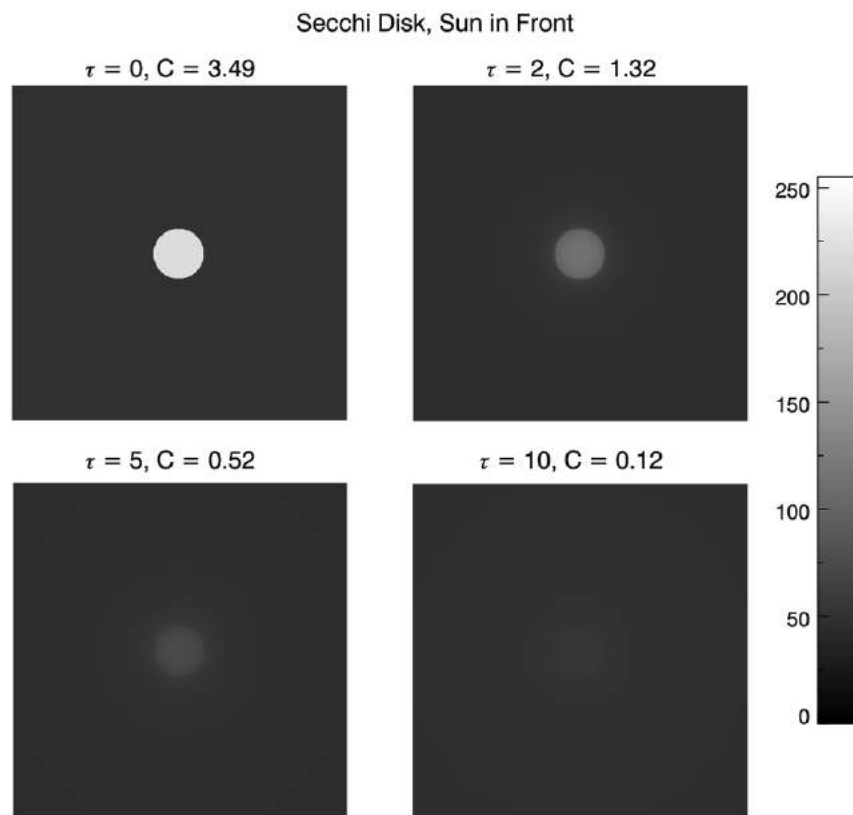


Figure G.17: Images of the Secchi disk when the Sun is in front and the disk appears lighter than the background.

fixed, the apparent size (angular extent) of the disk would then decrease with increasing range. That brings into play another factor, which is the size of the disk. A smaller disk corresponds to a higher spatial frequency and will thus disappear sooner as scattering blurs out the high spatial frequencies. The effect of disk size on Secchi disk visibility has been studied by [Hou et al. \(2007\)](#) using modulation transfer function (MTF; see Section G.8) techniques equivalent to the PSF approach developed here.

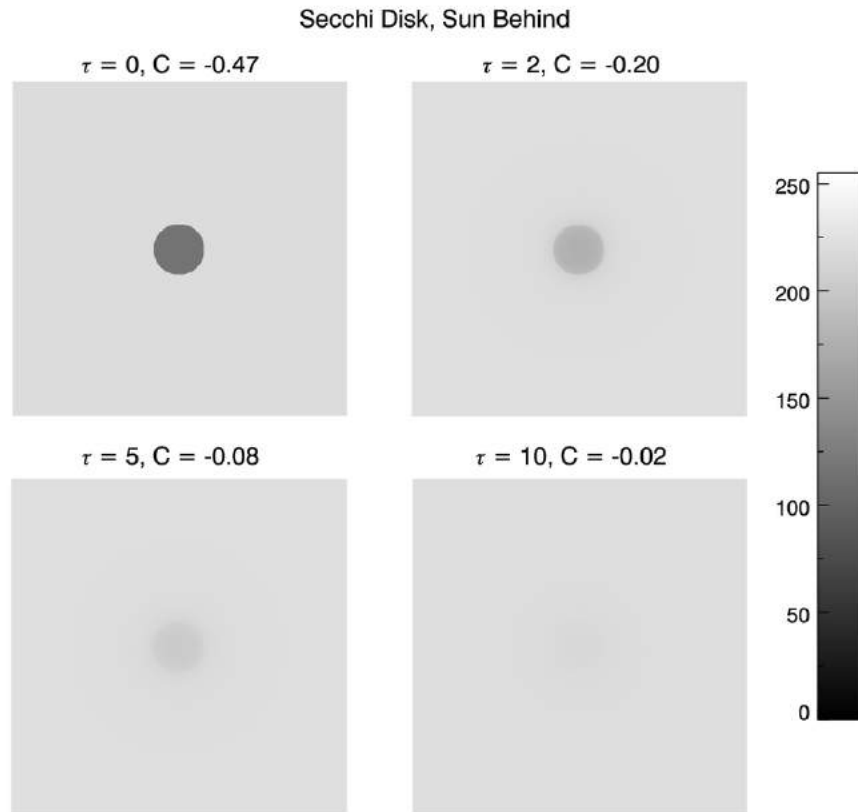


Figure G.18: Images of the Secchi disk when the Sun is behind and the disk appears darker than the background.

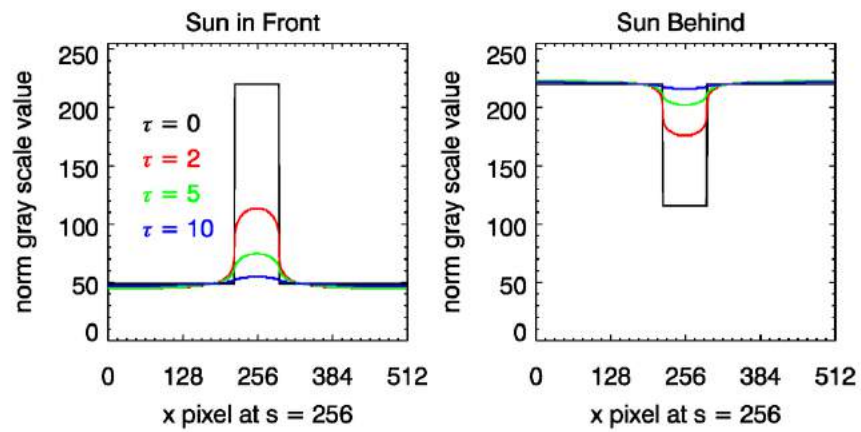


Figure G.19: Normalized cross sections through the Secchi disk for the case of the Sun in front of the disk and behind the disk.

G.7.2 Color Image Simulation

Now consider the degradation with range of a color image. In addition to the decrease in brightness with increasing range, there will also be a shift in the color because different wavelengths absorb and scatter differently, and thus have different PSFs. Table 3 shows the IOPs for 450, 550, and 650 nm computed for water with a chlorophyll concentration of 0.5 mg m^{-3} , according to the “new Case 1” IOP model in HydroLight. The scattering phase function is a Fournier-Forand phase function determined by the wavelength-dependent backscatter fraction b_b/b .

IOP	450 nm	550 nm	650 nm
a [1/m]	0.0469	0.0650	0.3490
b [1/m]	0.2638	0.2463	0.2335
c [1/m]	0.3107	0.3113	0.5825
b_b/b	0.00968	0.0113	0.0153

Table G.3: IOPs used for generation of wavelength-dependent PSFs. Values are for Case 1 water with a chlorophyll concentration of 0.5 mg m^{-3} .

Figure G.20 shows the PSFs for a viewing range of $R = 5 \text{ m}$ computed using the IOPs of Table G.3. The blue and green PSFs are similar in magnitude, but the red PSF is much less than the blue and green PSFs. Table G.3 shows that the three wavelengths have about the same total scattering coefficient, but the red wavelength has over 5 times as much absorption as the green, and over 7 times as much absorption as the blue. Thus the reduced magnitude of the red PSF is due to the increased absorption by pure water at 650 nm.

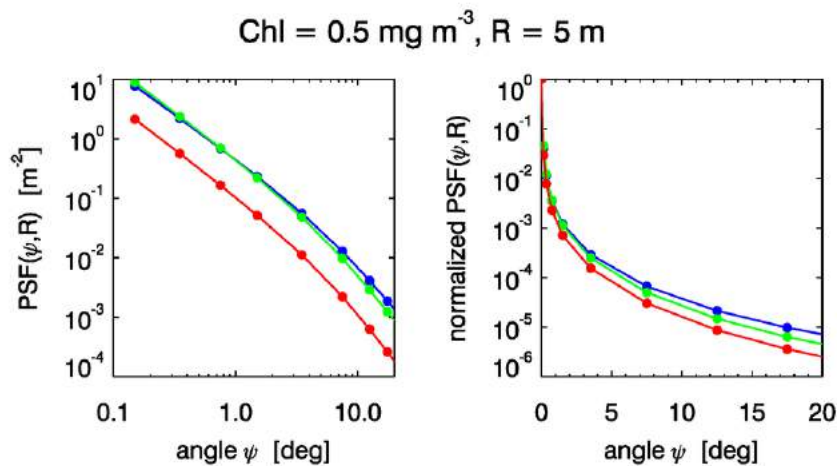


Figure G.20: Wavelength dependent PSFs for the IOPs of Table 3. Left panel: absolute values; right panel: normalized values. The blue curves are for 450 nm, the green for 550 nm, and the red for 650 nm.

These PSFs were used to process the blue, green, and red bands of the color image of red sponge⁷ seen in the left panel of Fig. G.21. The right panel of the figure shows the predicted image for a viewing range of 5 m. Note that the water has absorbed essentially all of the red light, so the red sponge appears almost black. The fine detail of the background seen in the original image has been blurred out by scattering. Figure G.22 shows a transect across the middle of these images. The plots at $R = 5$ m make clear that the high spatial frequencies have been damped out, and that almost no red light remains. The absolute magnitudes for the RGB bands were less than 10 at 5 m, which gives a very dark image, so each band was rescaled by a factor of 10 in order to create the brighter image seen on the right of Fig. G.21. This rescaling preserves the relative amounts of R, G, and B, and is the equivalent of changing the camera sensitivity.

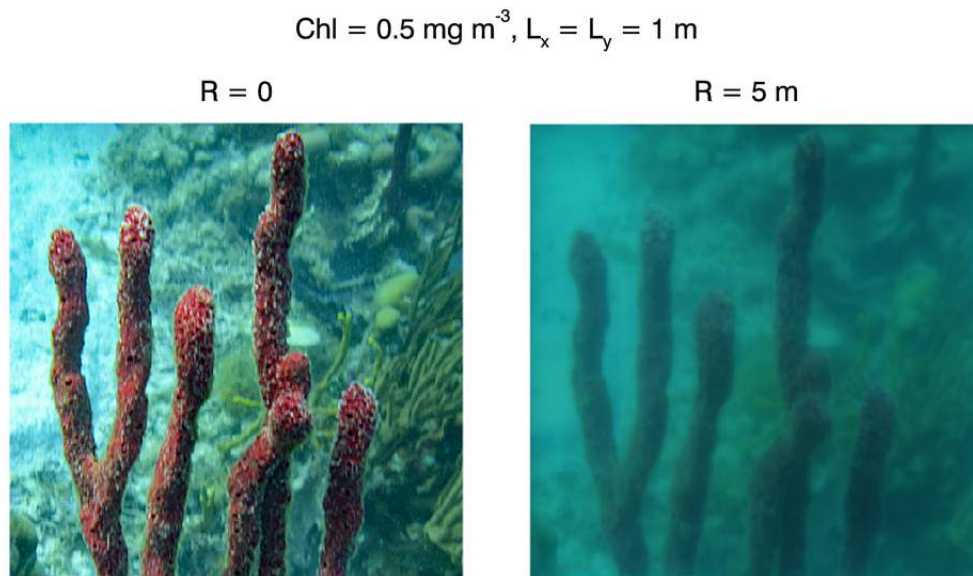


Figure G.21: Left panel: the original image of a red sponge, photographed from a distance of less than 1 m. Right panel: the predicted image at a range of $R = 5$ m.

⁷Photographed, I might brag, by the author while on a sea kayak expedition along the Kuna Yala (Caribbean) coast of Panama. Thanks to Mike Lesser for the species identification as *Amphimedon compressa*.

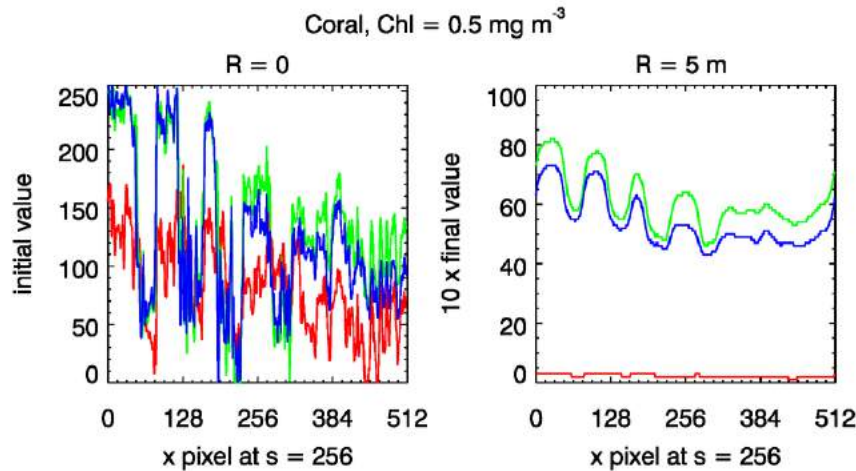


Figure G.22: The left panel shows a cross section at $s = 256$ through the initial coral image ($R = 0$). The right panel shows the pixel values for a range of $R = 5$ m. These values are rescaled by a factor of 10 from the actual values.

G.8 The Modulation Transfer Function (MTF)

The primary purpose of the preceding sections is to show how to predict the appearance of an image as a function of water optical properties and distance. Deciding whether or not an object is visible in the image is a separate but related problem. As we have seen, the point spread function is central to image prediction, whether used in convolution summations or Fourier transforms. The PSF is also the basis for computing the modulation transfer function (MTF), which quantifies how well object contrast is transmitted as a function of spatial frequency. The MTF is first explained by a description of how it can be measured. Its relation to the PSF is then presented.

G.8.1 Measuring the MFT

Figure G.23 shows three sets of parallel lines or bars, which are shaded from black to white in a sinusoidal brightness pattern⁸. The upper left panel of the plot shows 4 periods of the bright to dark to bright pattern in a length L . The upper middle and right panels show 8 and 16 periods in length L . The lower three panels show the signals detected by an instrument as it scans across the sets of lines if the lines are observed at range 0 (or through a vacuum). The signal might be digital counts or volts or irradiance, depending on the type of measuring instrument. The distance L might be millimeters or meters. If you are viewing the lines from a distance R , you can convert distance L to angle $\psi = \tan^{-1}(L/R)$. Here I'll assume that L is in meters as might be appropriate for imaging centimeter-sized to meter-sized objects underwater. In that case, the spatial frequencies of the three sets of lines are 4 lines or cycles per meter, 8 cycles per meter, and 16 cycles per meter. The spatial frequency ν would then be written as $\nu = 4$ cycles/m, or sometimes just $\nu = 4 \text{ m}^{-1}$, and so on, since the number of lines or cycles is a nondimensional number. In discussions

⁸To be honest, I created these bars in Powerpoint by making a line fuzzy, so the pattern shown here is just an approximation to a sinusoidal brightness pattern.

of camera lenses or microscopes, you will often see units such as lp/mm, meaning line pairs (i.e. cycles) per millimeter, or lp/mrad, meaning line pairs per milliradian. You are free to choose whatever measure of spatial frequency is convenient for your problem. For the examples here, I'll use cycles per meter.

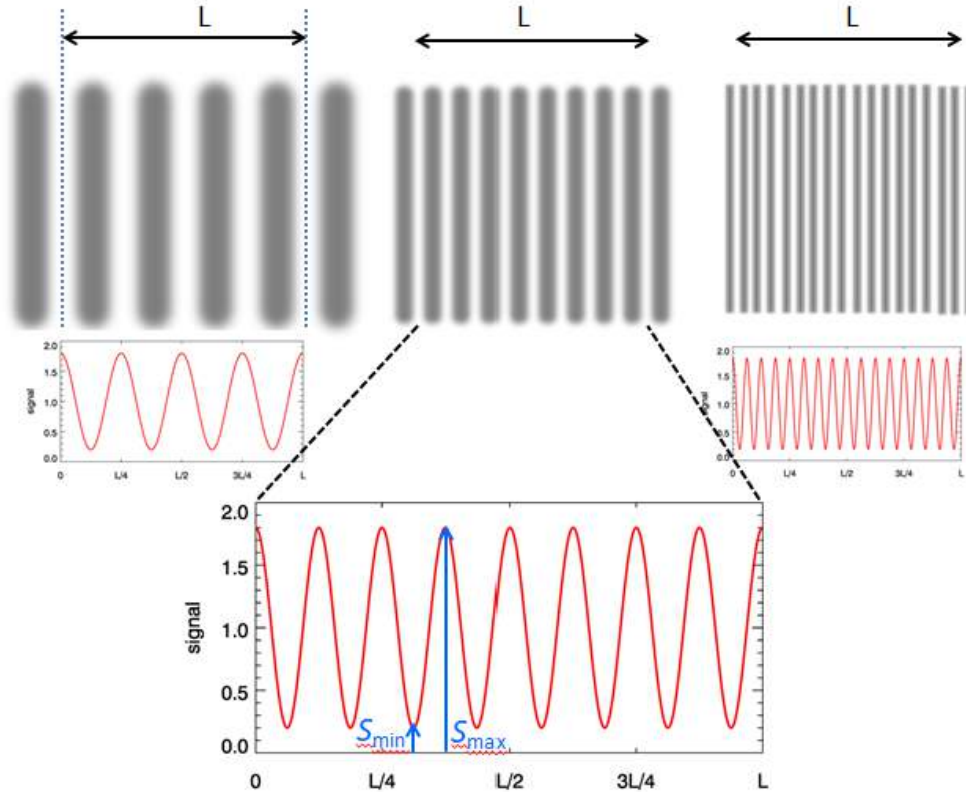


Figure G.23: Resolution chart line pattern at range 0. Upper left: 4 cycles per length L ; upper middle: 8 cycles per length L ; upper right: 16 cycles per length L . Lower three panels: The signals recorded as an instrument scans across the line patterns. If L is one meter, the three sets of lines corresponds to spatial frequencies of 4, 8, and 16 cycles/m.

In Fig. G.23 the instrument measures a full-amplitude signal at each spatial frequency. As drawn, the signal has an average or background value (the “DC signal”) of 1.0 and the deviation from the background ranges from 0.2 to 1.8 (these values of course depend on the sensor calibration). It is the deviation from the background that determines the “visibility” of the bars, i.e., how well the bright and dark bars can be distinguished. Let $S_{\max}(0, \nu)$ and $S_{\min}(0, \nu)$ be the maximum and minimum signal strengths measured at range 0 for spatial frequency ν . The inherent (i.e., as seen from 0 distance or through a vacuum) contrast is defined to be

$$M(0, \nu) \triangleq \frac{S_{\max}(0, \nu) - S_{\min}(0, \nu)}{S_{\max}(0, \nu) + S_{\min}(0, \nu)}. \quad (\text{G.19})$$

This measure of contrast is called the Michelson contrast or the modulation depth. In general, $M(0, \nu) < 1$ because the recorded signal is not zero at the minimum. For the bar

patterns of Fig. G.23, the inherent contrast at each frequency is

$$M(0, \nu) = \frac{1.8 - 0.2}{1.8 + 0.2} = 0.8.$$

Figure G.24 shows how the resolution lines of Fig. G.23 might appear from a distance R when seen through an absorbing and scattering medium. The lines of the resolution chart of Fig. G.23 are less bright due to absorption and blurred together because of scattering.

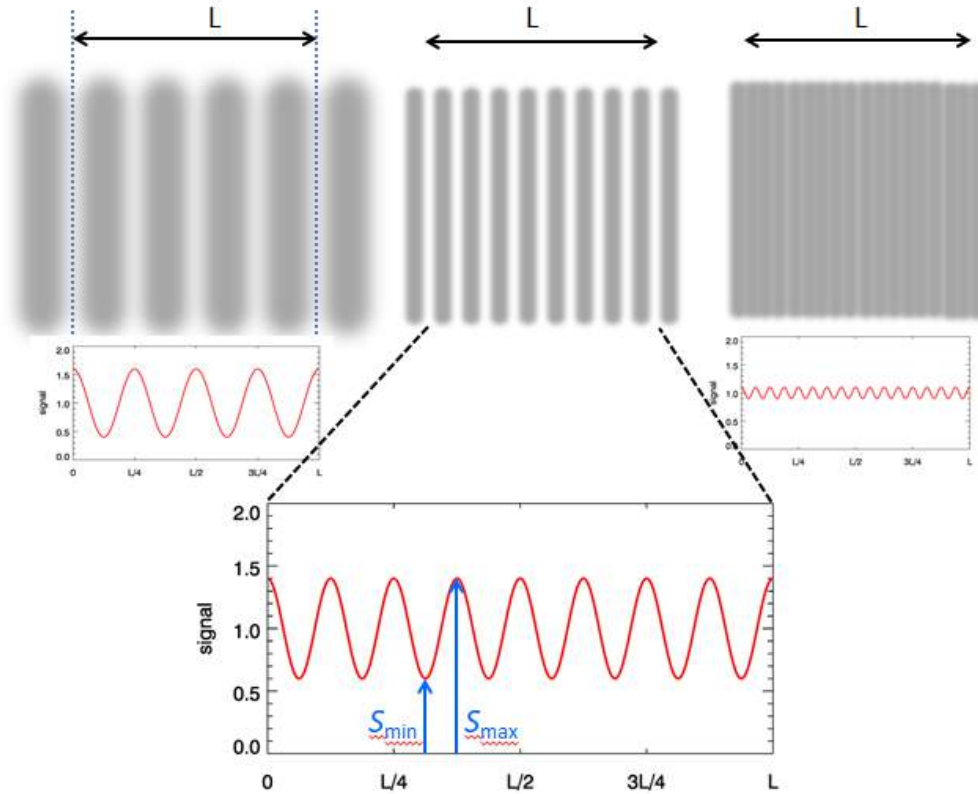


Figure G.24: Resolution chart line pattern at range R . Now the lines of the resolution chart of Fig. G.23 are less bright due to absorption and blurred together because of scattering.

Now the contrast is

$$M(R, \nu) = \frac{S_{\max}(R, \nu) - S_{\min}(R, \nu)}{S_{\max}(R, \nu) + S_{\min}(R, \nu)}. \quad (\text{G.20})$$

For the examples of Fig. G.24, the contrast for a frequency of 4 cycles/m is

$$M(R, 4 \text{ cycles/m}) \triangleq \frac{1.6 - 0.4}{1.6 + 0.4} = 0.6.$$

For the other two frequencies we get $M(R, 8 \text{ cycles/m}) = (1.4 - 0.6)/(1.4 + 0.6) = 0.4$ and $M(R, 16] : \text{cycles/m}) = (1.1 - 0.9)/(1.1 + 0.9) = 0.1$.

The *modulation transfer function* (MTF) is defined as the contrast at range R normalized by the contrast at range 0:

$$MTF(R, \nu) \triangleq \frac{M(R, \nu)}{M(0, \nu)}. \quad (\text{G.21})$$

The MTF shows how well the system (water, camera lens, human eye) transmits inherent contrast as a function of spatial frequency. The MTF is often described as giving the response of the system to sinusoids of different spatial frequencies. Yet another way of saying the same thing is that the MTF is the weighting function applied by the system to different frequencies relative to the weighing function for zero frequency. For the present example, using the numbers from above, we have $MTF(R, 4 \text{ cycles/m}) = 0.6/0.8 = 0.75$. For the examples at 8 and 16 cycles/m the MTF is 0.5 and 0.125, respectively.

In the three panels of Fig. G.23 the lines of the three panels are each clearly distinguished. In the left and middle panels of Fig. G.24, the lines are still distinguishable. However, for the pattern of lines at $\nu = 16$ cycles/m the lines are so blurred together than we may decide that they are not really distinguishable. Thus we conclude (for this example) that a spatial frequency of 8 cycles/m is “resolved” or “visible”, but a spatial frequency of 16 cycles/m is not resolved. So the cutoff for resolved vs not resolved, or visible vs not visible, is somewhere between an MTF value of 0.5 and 0.125.

G.8.2 Relation Between the MTF and the PSF

Remember that the Fourier transform takes in a function of spatial position and returns a function of spatial frequency (or a function of angle and returns a function of angular frequency). The MTF specifies how well contrast is transmitted as a function of spatial frequency. The $PSF(x_r, y_s)$ is a function of position and its DFT $\widehat{PSF}(\nu_u, \mu_v)$ is a function of spatial frequency. We saw in Fig. G.12 that different PSFs blur the same image differently. These observations hint that there may be a relation between the Fourier transform of the PSF and the MTF, and indeed there is. I’ll simply give you the answer.

The *Optical Transfer Function* (OTF) is by definition the Fourier transform of the point spread function:

$$OTF(\nu, \mu) \triangleq \mathcal{F}\{PSF(x, y)\},$$

or for a discrete PSF, as we are using here:

$$OTF(u, v) \triangleq \mathcal{D}\{PSF(r, s)\}.$$

The OTF describes how a system responds to inputs of various spatial frequencies. Here the “system” can be a camera lens, your eye, or a slab of ocean water of thickness R . In general the OTF is a complex function, i.e., it has both real and imaginary parts, and its values can be positive or negative. A complex number $c = a + ib$ can always be written as $|c|e^{i\phi}$, where the absolute value of c is $|c| = \sqrt{c^*c} = \sqrt{a^2 + b^2}$; $c^* = a - ib$ is the complex conjugate of c , and $\phi = \tan^{-1}(b/a)$ is the phase angle. In the same fashion, the OTF can be written as

$$OTF(\nu, \mu) = |OTF(\nu, \mu)|e^{iPTF(\nu, \mu)}.$$

Here PTF is called the *phase transfer function* and—guess what!—the absolute value of the OTF is the MTF. The MTF tells how the system responds in *magnitude* to inputs of

different frequencies, and the PTF contains phase information. The PTF does not concern us for the ocean so long as the PSF is azimuthally symmetric, as it almost always is for the ocean⁹. This is because the 2-D Fourier transform of an azimuthally symmetric function (here the PSF) is real valued, so the *PTF* is zero. Thus we have the very important result that

$$\begin{aligned} MTF(u, v) &= |\mathcal{D}\{PSF(r, s)\}| = |\widehat{PSF}(u, v)| \\ &= \left[\widehat{PSF}^*(u, v) \odot \widehat{PSF}(u, v) \right]^{\frac{1}{2}}. \end{aligned} \quad (\text{G.22})$$

As before, the \odot indicates frequency-by-frequency multiplication. When computing the MFT via Eq. (G.22) it is fine to use a normalized PSF because the MFT is normalized to 1 at zero frequency.

G.9 Numerical Examples of MTFs

Section A.3.1 discusses the order in which FFT subroutines return their values as a function of frequency. Because FFTs are usually the based of computing MTFs, the same frequency ordering must be considered before we can get from the 2-D $MTF(u, v)$ to a 1-D function of spatial frequency that corresponds to what is measured by Eqs. (G.20) and (G.21). FFT routines return the Fourier amplitudes for the spatial frequency magnitudes $\nu_0 = 0$ (the DC signal), the fundamental frequency $\nu_1 = 1/L_x$, and the harmonics of the fundamental frequency through to the Nyquist frequency $\nu_{N_x/2} = 1/(2\Delta x) = 1/(2L_x/N_x)$, and there are both positive and negative frequency pairs for a total of N_x frequencies. This “FFT frequency order” is shown in Eq. (A.14). There are corresponding formulas for the y direction. For plotting, it is convenient to use the “math frequency order”, which is shown in Eq. (A.14).

The orders of the frequencies in the $MTF(u, v)$ array, which come from FFTs of the PSF, are in the FFT frequency order. Thus the $(\nu_u, \nu_v) = (u, v) = (0, 0)$ value is at the “lower left” of the matrix. This can be a confusing mess the first time you encounter it, so a simple numerical example is warranted. Figure G.25 shows an example generated for the case of $L_x = L_y = 2$ m and $N_x = N_y = 8$ (and for the PSF based in the Petzold phase function with $\omega_o = 0.8$ and for $\tau = 5$). For these values, the fundamental frequency in the x direction is $\nu_1 = 1/L_x = 0.5$ cycles/m, with the same value for μ_1 . The Nyquist frequency is $\nu_{N_x/2} = 1/(2\Delta x) = 2.0$ cycles/m, with the same value for the y direction. Thus to extract the MTF values corresponding to the magnitude of the frequency $\nu = \sqrt{\nu_u^2 + \mu_v^2}$, we can use the values in the red box in the bottom row of the figure to get the MFT values from the DC signal to the Nyquist frequency. We could also read off the same values along the first column.

Figure G.26 shows three example PSFs and their corresponding MTFs, computed as above. These functions all have the same Petzold phase function and $\omega_o = 0.8$, but are for optical distances of $\tau = 2, 5$, and 10. For these runs, $L_x = L_y = 2$ m and $N_x = N_y = 128$. The most highly peaked PSF ($\tau = 2$ in blue) has the least scattering at large angles. Thus the MTF has relatively high values (greater than 0.33) over the entire range of frequencies,

⁹In a camera lens, aberrations may give a PSF that is not azimuthally symmetric. I can tell that my aging eyes have PSFs that are not azimuthally symmetric simply by looking at a point light source such as a star; I see a blurred and not azimuthally symmetric pattern of light around the star.

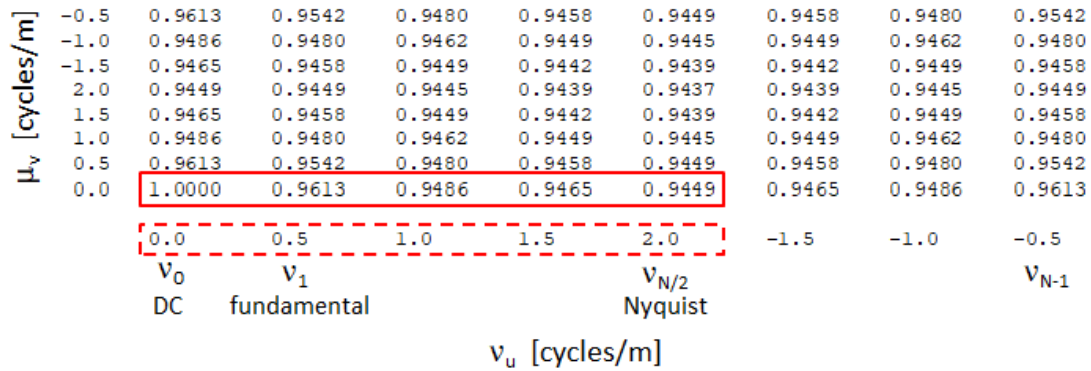


Figure G.25: Example 2-D MFT showing the FFT frequency order. For this example, $N_x = N_y = 8$ and $L_x = L_y = 2$ m.

which means that the high frequencies are transmitted well. The $\tau = 8$ PSF is much less peaked, indicating that multiple scattering is important. The corresponding MTF is less than 0.1 for almost all frequencies, and less than 0.01 at the higher frequencies. Consequently, only the very lowest frequencies are well transmitted; the high frequencies are blurred out by scattering.

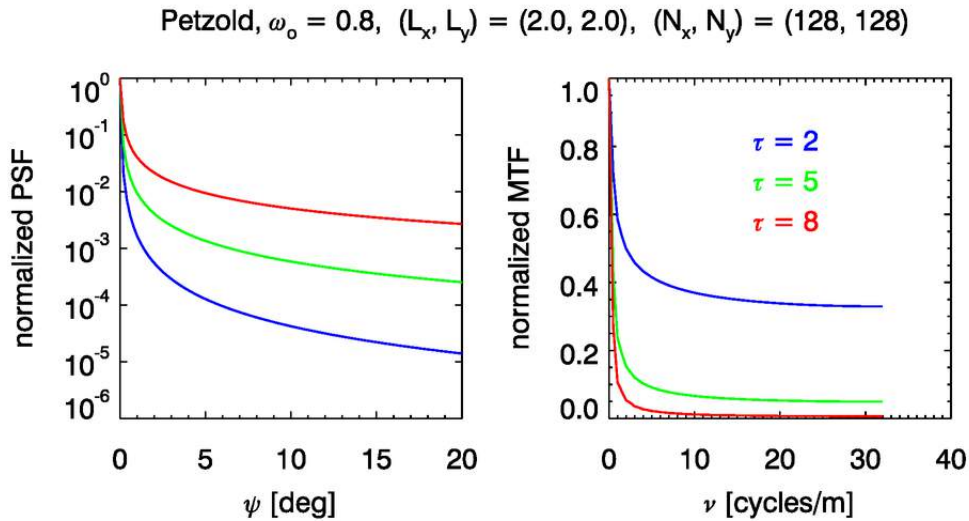


Figure G.26: Normalized PSF and MTF for optical distances of $\tau = 2, 5,$ and 8 . The spatial frequencies are in the dashed red box; the corresponding MTF values are in the solid red box.

Figure G.27 shows the 256 by 256 pixel USAF resolution chart as processed by these PSFs. The loss of the high spatial frequencies as τ goes from 2 to 5 to 8 is obvious and gives some intuitive feel for the implications of the corresponding MTFs. Each of these predicted images (other than the original image) was normalized so that the brightest pixel in the image was set to 220 on the gray scale.

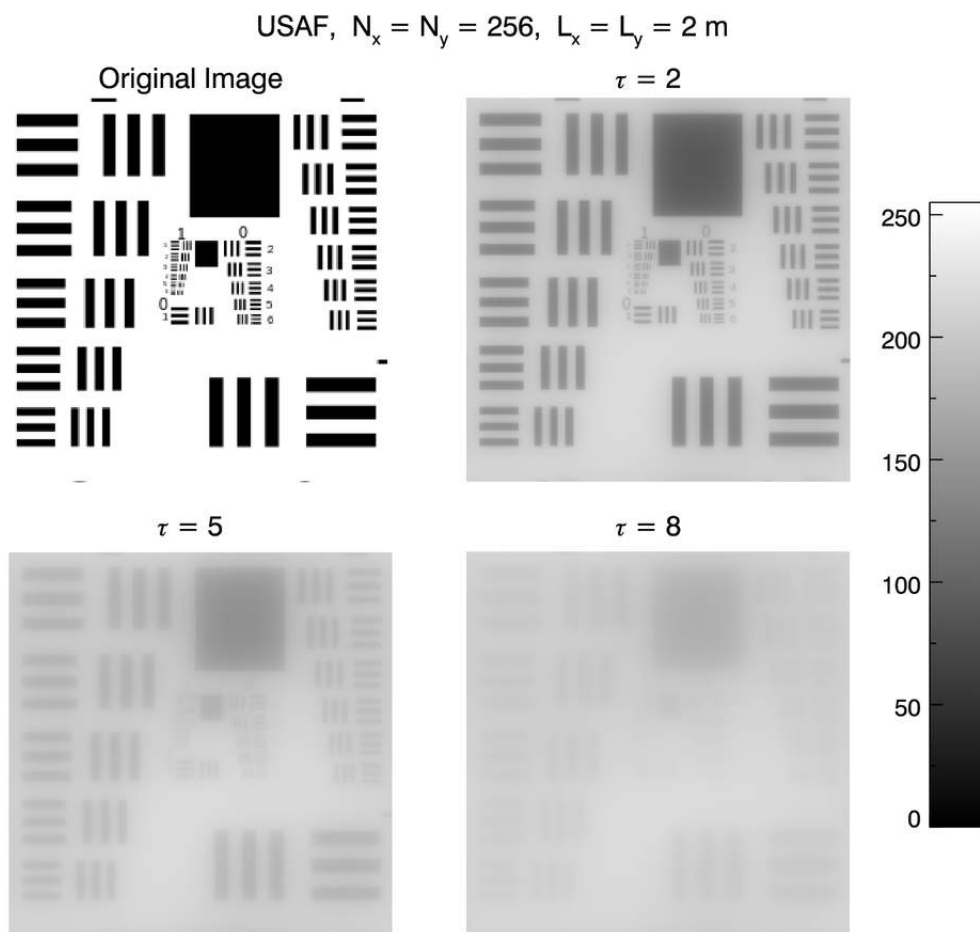


Figure G.27: Image degradation for different optical distances of $\tau = 2, 5$, and 8 . The corresponding PSFs and MTFs are seen in Fig. G.26.

MTFs have a very useful property. Suppose you are looking at a Secchi disk from above the sea surface as into descends into the water. The MTF for the “optical system” of the eye + sea surface + water column gives you information on the depth where the disk will disappear from view (e.g., Hou et al., 2007). If the various optical components—here the eye, the water surface, and the water column—are independent (as they are here), then the MTF for the total system is the product of the MTFs for the individual components:

$$MTF_{\text{total}} = MTF_{\text{eye}} \odot MTF_{\text{sea surface}} \odot MTF_{\text{water column}} .$$

Here, as before, the \odot indicates element-by-element multiplication.

Much more can be said about OTFs and MTFs. Indeed, entire books have been devoted to measuring and modeling MTFs (e.g., Boreman, 2001). Hou (2013) gives a good overview of models of MTFs appropriate for ocean waters.

I’ll close this section with a story. The MTF is the standard measure of “quality” in optical components such as camera lenses. I once needed to evaluate which 1000 mm

telephoto lens (costing several thousand dollars) would be best for an experiment that was to measure the PSF over multi-kilometer path lengths just above the surf zone, where breaking waves inject lots of small droplets into the atmosphere. Getting high resolution for very small angles, i.e., for high spatial frequencies, was critical. The best lens for that job would be the one with the highest MTF at high spatial frequencies. So I contacted Nikon, Cannon, Pentax, et al. and asked them to send me the MTFs for particular lenses. Not a single manufacturer would let me see the MTFs for their lenses. Bottom line: comparison shopping based on something other than advertising hype is not allowed. In the final analysis, it didn't matter. The proposal wasn't funded.

Bibliography

- E. Aas. Influence of shape and structure on light scattering by marine particles. Technical Report Report 53, Univ. Copenhagen Inst. Geophys., 1984.
- E. Aas. Refractive index of phytoplankton derived from its metabolite composition. *J. Plankton Research*, 18(12):2223–2249, 1996.
- P. K. Acharya, A. Berk, L. S. Bernstein, M. W. Matthew, S. M. Adler-Golden, D. C. Robertson, G. P. Anderson, J. H. Chetwynd, F. X. Kneizys, E. P. Shettle, L. W. Abreu, W. O. Gallery, J. E. A. Selby, and S. A. Clough. MODTRAN User’s Manual, Versions 3.7 and 4.0. Technical report, Air Force Res. Lab., Hanscom Air Force Base, MA, 1998.
- J. Acker. *The Color of the Atmosphere with the Ocean Below: A History of NASA’s Ocean Color Missions*. Independently published, 2015. ISBN 1507699220.
- S. G. Ackleson and R. W. Spinrad. Size and refractive index of individual marine particulates: a flow cytometric approach. *Applied Optics*, 27(7):1270–1277, 1988.
- G. Adler, A. A. Riziq, C. Erlick, and Y. Rudich. Effect of intrinsic organic carbon on the optical properties of fresh diesel soot. *Proc. Nat. Acad. Sci.*, 107(15):6699–6704, 2010.
- Y. C. Agrawal and H. C. Pottsmith. Instruments for particle size and settling velocity observations in sediment transport. *Marine Geology*, 168:89–114, 2000.
- A. Ahmad, B. A. Franz, C. R. McClain, E. J. Kwaitkowska, J. Werdell, E. P. Shettle, and B. N. Holben. New aerosol models for the retrieval of aerosol optical thickness and normalized water-leaving radiances from SeaWiFS and MODIS sensors over coastal regions and open oceans. *Applied Optics*, 49(29):5545–5560, 2010a. Errata to Table 3 are in Ahmad et al. (2010b).
- A. Ahmad, B. A. Franz, C. R. McClain, E. J. Kwaitkowska, J. Werdell, E. P. Shettle, and B. N. Holben. New aerosol models for the retrieval of aerosol optical thickness and normalized water-leaving radiances from SeaWiFS and MODIS sensors over coastal regions and open oceans: publishers note. *Applied Optics*, 50(5):626, 2010b.
- Z. Ahmad and R. S. Fraser. An iterative radiative transfer code for ocean-atmosphere systems. *J. Atmos. Sci.*, 39:656–665, 1982.
- Z. Ahmad, C. R. McClain, J. R. Herman, B. A. Franz, E. J. Kwaitkowska, W. D. Robinson, E. J. Bucsela, and M. Tzortziou. Atmospheric correction for NO₂ absorption in retrieving

- water-leaving reflectances from the SeaWiFS and MODIS measurements. *Applied Optics*, 39(26):6504–6512, 2007.
- Y.-H. Ahn. *Proprietes optiques des particules biologiques et minerales presentes dans l’ocean. Application: inversion de la reflectance*. PhD thesis, Univ. Pierre and Marie Curie, Paris, France, 1999.
- Y.-H. Ahn, A. Bricaud, and A. Morel. Light backscattering efficiency and related properties of some phytoplankters. *Deep-Sea Research*, 39(11/12):1835–1855, 1992.
- J. H. G. M. Alves and M. L. Banner. Revisiting the Pierson-Moskowitz asymptotic limits for fully developed wind waves. *J. Phys. Oceanogr.*, 33:1301–1323, 2003.
- D. Antoine, M. Chami, H. Claustre, F. d’Ortenzio, A. Morel, G. Bcu, B. Gentili, F. Louis, J. Ras, E. Roussier, A. J. Scott, D. Tailliez, S. B. Hooker, P. Guevel, J.-F. Dest, C. Dempsey, and D. Adams. BOUSSOLE: A joint CNRS-INSU, ESA, CNES, and NASA ocean color calibration and validation activity. Technical Report TM-2006-214147, NASA, December 2006.
- T. Aoki, T. Kitamura, S. Matsuno, K. Mitsu, Y. Ohashi, A. Okada, D. R. Cady, J. G. Learned, D. O’Connor, S. Dye, P. W. Gorham, M. McMurdo, R. Mitiguy, M. Webster, C. Wilson, and P. Grieder. Background light measurements in the deep ocean. *Il Nuovo Cimento C*, 9:642–652, 1986. doi: 10.1007/BF02514882.
- T. Aoyama, M. Hayakawa, T. Kinoshita, and M. Nio. Tenth-order electron anomalous magnetic moment: Contribution of diagrams without closed lepton loops. *Phys. Rev. D*, 91, 2015. doi: 10.1103/PhysRevD.91.033006.
- J. R. Apel. *Principles of Ocean Physics*, volume 38 of *International Geophysics Series*. Academic Press, 1987.
- R. E. Apfel. The tensile strength of liquids. *Scientific American*, 227:58–62, 1972.
- R. A. Arnone, A. M. Wood, and R. W. Gould Jr. Water mass classification. *Oceanography*, 17(2):14–15, 2004.
- W. Arnott and P. L. Marston. Optical glory of small freely rising gas bubbles in water: observed and computed cross-polarized backscattering patterns. *J. Opt. Soc. Am.*, 5: 496–506, 1998.
- W. P. Arnott and P. L. Marston. Backscattering of laser light from freely rising spherical and spheroidal air bubbles in water. In *Ocean Optics IX*, volume 925, pages 296–307. SPIE Proc., 1988.
- W. P. Arnott and P. L. Marston. Unfolded optical glory of spheroids: backscattering of laser light from freely rising spheroidal air bubbles in water. *Applied Optics*, 30:3429–3442, 1991.
- A. B. Arons and M. B. Peppard. Concerning an heuristic point of view toward the emission and transformation of light. *Am. J. Phys.*, 33:367, 1969. English translation of Einstein, A. (1905), *Ann. Phys.* 17, 132.

- C. P. Artlett and H. M. Pask. New approach to remote sensing of temperature and salinity in natural water samples. *Optics Express*, 25(3):2840–2851, 2017. doi: 10.1364/OE.25.002840.
- A. P. Arya. *Fundamentals of Nuclear Physics*. Allyn and Bacon, 1966.
- S. Asano and M. Sato. Light scattering by randomly oriented spheroidal particles. *Applied Optics*, 19(6):962–974, 1980.
- S. Asano and G. Yamamoto. Light scattering by a spheroidal particle. *Applied Optics*, 14(1):29–49, 1975.
- R. W. Austin and G. Halikas. The index of refraction of seawater. Technical Report Ref. 76-1, Scripps Inst. Oceanogr., 1976.
- R. W. Austin and T. J. Petzold. The determination of the diffuse attenuation coefficient of sea water using the Coastal Zone Color Scanner. In J. F. R. Goser, editor, *Oceanography from Space COSPAR/SCOR/IUCRM Symposium Plenum*, volume 13, pages 239–256. Springer, 1981. doi: 10.1007/978-1-4613-3315-9_29.
- R. W. Austin and T. J. Petzold. Spectral dependence of the diffuse attenuation coefficient of light in ocean water. *Optical Engineering*, 25(3):471–479, 1986.
- D. M. Avouris and J. D. Ortiz. Validation of 2015 Lake Erie MODIS image spectral decomposition using visible derivative spectroscopy and field campaign data. *J. Great Lakes Res.*, 45:466–479, 2019. doi: 10.1016/j.jglr.2019.02.005.
- M. Babin. Phytoplankton fluorescence: Theory, current literature and in situ measurement. In M. Babin, C. S. Roesler, and J. J. Cullen, editors, *Real-time coastal observing systems for ecosystem dynamics and harmful algal blooms*. UNESCO, 2008.
- M. Babin and D. Stramski. Variations in the mass-specific absorption coefficient of mineral particles suspended in water. *Limnol. Oceanogr.*, 49:756–767, 2004.
- M. Babin, A. Morel, V. Fournier-Sicre, F. Fell, and D. Stramski. Light scattering properties of marine particles in coastal and open ocean waters as related to the particle mass concentration. *Limnol. Oceanogr.*, 48:843–859, 2003a.
- M. Babin, D. Stramski, G. M. Ferrari, H. Claustre, A. Bricaud, G. Obolensky, and N. Hoepffner. Variations in the light absorption coefficients of phytoplankton, nonalgal particles, and dissolved organic matter in coastal waters around Europe. *J. Geophys. Res.*, 108:3211, 2003b. doi: 10.1029/2001JC000882.
- H. P. Bachelder and E. Swift. Estimated near-surface mesoplanktonic bioluminescence in the western North Atlantic during July 1986. *Limnol. Oceanogr.*, 34:113–128, 1989.
- H. Bader. The hyperbolic distribution of particle sizes. *J. Geophys. Res.*, 75(15):2822, 1970. doi: 10.1029/JC075i015p02822.
- S. W. Bailey, B. A. Franz, and P. J. Werdell. Estimation of near-infrared water-leaving reflectance for satellite ocean color data processing. *Optics Express*, 18(7):7521–7527, 2010.

- W. M. Balch, K. A. Kilpatrick, P. Holligan, D. Harbour, and E. Fernandez. The 1991 coccolithophore bloom in the central North Atlantic. 1. Relating optics to coccolith concentration. *Limnol. Oceanogr*, 41(8):1684–1696, 1996.
- S. Baldy. Bubbles in the close vicinity of breaking waves: Statistical characteristics of the generation and dispersion mechanism. *J. Geophys. Res.*, 93:8239–8248, 1988.
- S. Baldy and M. Bourguel. Measurement of bubbles in a stationary field of breaking waves by a laser-based single-particle scattering technique. *J. Geophys. Res.*, 90:1037–1047, 1985.
- M. L. Banner. Equilibrium spectra of wind waves. *J. Phys. Oceanogr.*, 20:966–984, 1990.
- Yu. N. Barabanenkov, Yu. A. Kravtsov, V. D. Orzin, and A. I. Saichev. Enhanced backscattering in optics. In E. Wolf, editor, *Progress in Optics XXIX*, pages 65–197. Elsevier, 1991. doi: 10.1016/S0079-6638(08)70006-4.
- J. S. Bartlett, K. J. Voss, S. Sathyendranath, and A. Vodacek. Raman scattering by pure water and seawater. *Applied Optics*, 37(15):3324–3332, 1998.
- M. Bass [Ed.]. *Handbook of Optics, Volume 1: Fundamentals, Techniques, and Design*. McGraw-Hill, 1995.
- W. Beebe. A half mile down. *The National Geographic Magazine*, LXVI(6):661–704, 1934.
- A. Beer. Bestimmung der absorption des rothen lichts in farbigen flüssigkeiten. *Ann. Physik Chemie*, 162(5):77–88, 1852. [Determination of the absorption of red light in colored liquids] (in German).
- M. Bellacicco, M. Cornec, E. Organelli, R. J. W. Brewin, G. Neukermans, G. Volpe, M. Barbieux, A. Poteau, C. Schmechtig, F. D’Ortenzion, S. Marullo, H. Claustre, and J. Pitarch. Global variability of optical backscattering by nonalgal particles from a BiogeochemicalArgo data set. *Geophys. Res. Letters*, 46:9767–9776, 2019. doi: 10.1029/2019GL084078.
- M. J. Berg, C. M. Sorensen, and A. Chakrabarti. A new explanation of the extinction paradox. *J. Quant. Spectros. Rad. Transfer*, 112:1170–1181, 2011.
- M. V. Berry and A. K. Geim. Of flying frogs and levitrons. *European J. Phys.*, 18(4):307, 1997. doi: 10.1088/0143-0807/18/4/012/meta.
- R. Bertel. Description d’un spectrographe sous-marin pour les recherches qualitatives de la lumire a différentes profondeurs de la mer. *Annales de L’Institut Océanographique*, III: 294–303, 1912.
- R. R. Bidigare, M. E. Ondrusek, J. H. Morrow, and D. Kiefer. In vivo absorption properties of algal pigments. In *Ocean Optics X*, volume 1302, pages 290–302. SPIE Proc., 1990.
- C. E. Binding, D. G. Bowers, and E. G. Mitchelson-Jacob. Estimating suspended sediment concentrations from ocean colour measurements in moderately turbid waters: The impact of variable particle scattering properties. *Rem. Sens. Environ.*, 94:373–383, 2005.

- W. P. Bissett, K. L. Carder, J. J. Walsh, and D. A. Dieterle. Carbon cycling in the upper waters of the Sargasso Sea: II. Numerical simulation of the apparent and inherent optical properties. *Deep-Sea Res. I*, 46:271–317, 1999.
- K. M. Bisson, E. Boss, P. J. Werdell, A. Ibrahim, and J. J. Behrenfeld. Particulate backscattering in the global ocean: A comparison of independent assessments. *Geophys. Res. Letters*, 48, 2021. doi: 10.1029/2020GL090909.
- J. L. Blanchard, R. F. Heneghan, J. D. Everett, R. Trebilco, and A. J. Richardson. From bacteria to whales: Using functional size spectra to model marine ecosystems. *Trends in Ecology & Evolution*, 32(3):174–186, 2017. doi: S0169534716302361.
- N. V. Blough and S. A. Green. Spectroscopic characterization and remote sensing of nonliving organic matter. In R. G. Zepp and C. Sonntag, editors, *Role of Nonliving Organic Matter in the Earth's Carbon Cycle*, pages 23–45. Wiley, 1995.
- N. V. Blough and R. Del Vecchio. Chromophoric DOM in the coastal environment. In D. A. Hansell and C. A. Carlson, editors, *Biogeochemistry of Marine Dissolved Organic Matter*, pages 509–546. Academic Press, 2002.
- B. A. Bodhaine, N. B. Wood, E. G. Dutton, and J. R. Slusser. On Rayleigh optical depth calculations. *J. Atmos. Oceanic Technol.*, 16, 1999.
- C. F. Bohren and E. Clothiaux. *Fundamentals of Atmospheric Radiation*. Wylie-VCH, 2006.
- C. F. Bohren and A. B. Fraser. The color of the sky. *The Physics Teacher*, 1985.
- C. F. Bohren and D. R. Huffman. *Absorption and Scattering of Light by Small Particles*. John Wiley and Sons, 1983.
- C. F. Bohren and S. B. Singham. Backscattering by nonspherical particles: a review of method and suggested new approaches. *J. Geophys. Res.*, 96(D3):5269–5277, 1991.
- G. D. Boreman. *Modulation Transfer Function in Optical and Electro-Optical Systems*, volume TT52. SPIE Press, 2001.
- E. Boss and W. S. Pegau. Relationship of light scattering at an angle in the backward direction to the backscattering coefficient. *Applied Optics*, 40:5503–5507, 2001.
- E. Boss, M. S. Twardowski, and S. Herring. Shape of the particulate beam attenuation spectrum and its inversion to obtain the shape of the particulate size distribution. *Applied Optics*, 40:4885–4893, 2001. URL http://misclab.umeoce.maine.edu/boss/scientific_articles.php#2001.
- E. Boss, W. S. Pegau, M. Lee, M. Twardowski, E. Shybanov, G. Korotaev, and F. Baratange. Particulate backscattering ratio at LEO 15 and its use to study particle composition and distribution. *J. Geophys. Res.*, 109(C1), 2004a. doi: 10.1029/2002JC001415. URL http://misclab.umeoce.maine.edu/boss/scientific_articles.php#2004.

- E. Boss, D. Stramski, T. Bergmann, W. S. Pegau, and M. Lewis. Why should we measure the optical backscattering coefficient? *Oceanography*, 17(2):44–49, 2004b.
- E. Boss, W. S. Pegau, and P. Hill. Effect of particulate aggregation in aquatic environments on the beam attenuation and its utility as a proxy for particulate mass. *Optics Express*, 17(11):9408–9420, 2009a. URL http://misclab.umeoce.maine.edu/boss/scientific_articles.php#2009.
- E. Boss, W. H. Slade, M. Behrenfeld, and G. Dall’Olmo. Acceptance angle effects on the beam attenuation in the ocean. *Optics Express*, 17(3):1535–1550, 2009b. URL http://misclab.umeoce.maine.edu/boss/scientific_articles.php#2009.
- E. Boss, L. Taylor, S. Gilbert, K. Gundersen, N. Hawley, C. Janzen, T. Johenger, H. Purcell, C. Robertson, D. W. H. Schar, G. J. Smith, and M. N. Tamburri. Comparison of inherent optical properties as a surrogate for particulate matter concentration in coastal waters. *Limnol. Oceanogr. Methods*, 2009c.
- P. Bouguer. *Essai d’optique sur la gradation de la lumiere*. Claude Jombert, Paris, France, 1729. [Optics essay on the attenuation of light] (in French).
- A. F. Bower. *Applied Mechanics of Solids*. 2012. URL <http://www.solidmechanics.org/>. on line.
- D. G. Bowers and C. E. Binding. The optical properties of mineral suspended particles: a review and synthesis. *Estuarine, Coastal and Shelf Sci.*, 67:219–230, 2006.
- D. G. Bowers and H. L. Brett. The relationship between CDOM and salinity in estuaries: an analytical and graphical solution. *J. Marine Systems*, 73:1–7, 2008.
- D. G. Bowers, K. M. Braithwaite, W. A. M. Nimmo-Smith, and G. W. Graham. The optical efficiency of flocs in shelf seas and estuaries. *Estuarine, Coastal and Shelf Sci.*, 91(3):341–350, 2010.
- R. N. Bracewell. *The Fourier Transform and Its Applications, Second Edition, Revised*. McGraw Hill, Inc., 1986.
- H. Bradner, M. Bartlett, G. Blackington, J. Clem, D. Karl, J. Learned, A. Lewitus, S. Matsuno, D. O’Conner, W. Peatman, M. Reichle, C. Roos, J. Waters, M. Webster, and M. Yarbrough. Bioluminescence profile in the deep Pacific Ocean. *Deep-sea Res.*, 34(11):1831–1840, 1987.
- W. H. Brakel. Small-scale spatial variation in light available to coral reef benthos: quantum irradiance measurements from a Jamaican reef. *Bull. Mar. Sci.*, 29:406–413, 1979.
- N. Breitz and H. Medwin. Instrumentation for the acoustical measurement of bubble densities under breaking waves. *J. Acoust. Soc. Am.*, 86:739–743, 1989.
- R. J. W. Brewin, G. Dall’Olmo, S. Sathyendranath, and J. J. Hardman-Mountford. Particle backscattering as a function of chlorophyll and phytoplankton size structure in the open-ocean. *Optics Express*, 20(16):17632–17652, 2012.

- A. Bricaud and A. Morel. Light attenuation and scattering by phytoplanktonic cells: a theoretical modeling. *Applied Optics*, 25(4):571–580, 1986.
- A. Bricaud and D. Stramski. Spectral absorption coefficients of living phytoplankton and nonalgal biogenous matter: A comparison between the Peru upwelling area and the Sargasso Sea. *Limnol. Oceanogr.*, 35(3):562–582, 1990.
- A. Bricaud, A. Morel, and L. Prieur. Absorption by dissolved organic matter of the sea (yellow substance) in the UV and visible domains. *Limnol. Oceanogr.*, 26:43–53, 1981.
- A. Bricaud, A. Morel, and L. Prieur. Optical-efficiency factors of some phytoplankters. *Limnol. Oceanogr.*, 28:816–832, 1983.
- A. Bricaud, J. R. V. Zaneveld, and J. C. Kitchen. Backscattering efficiency of coccolithophorids: use of a three-layered model. In *Ocean Optics XI*, volume 1750, pages 27–33. SPIE, 1992.
- A. Bricaud, M. Babin, A. Morel, and H. Claustre. Variability in the chlorophyll-specific absorption coefficients of natural phytoplankton: Analysis and parameterization. *J. Geophys. Res.* 100(C7), 100(C7):13321–13332, 1995.
- A. Bricaud, A. Morel, M. Babin, K. Allali, and H. Claustre. Variation of light absorption by suspended particles with chlorophyll a concentration in oceanic (case 1) waters: Analysis and implications for bio-optical models. *J. Geophys. Res.*, 103(C13):31033–31044, 1998.
- A. Bricaud, H. Claustre, J. Ras, and K. Oubelkheir. Natural variability of phytoplanktonic absorption in oceanic waters: Influence of the size structure of algal populations. *J. Geophys. Res. Oceans*, 109(C11), 2004.
- J. W. Brown and R. V. Churchill. *Complex Variables and Applications, Eighth Edition*. McGraw-Hill, 2009. ISBN 978-0-07-305194-9. URL https://www.academia.edu/36209325/Brown_Churchill_Complex_Variables_and_Application_8th_edition_pdf. Available as a pdf.
- O. B. Brown and H. R. Gordon. Size-refractive index distribution of clear coastal water particulates from light scattering. *Applied Optics*, 13(12):2874–2881, 1974.
- W. Budde. *Optical Radiation Measurements, Vol. 4. Physical Detectors of Optical Radiation*. Academic Press, 1983.
- H. Buiteveld, J. H. M. Hakvoort, and M. Donze. The optical properties of pure water. In *Ocean Optics XII*, volume 2258, pages 174–183. SPIE Proc., 1994.
- J. W. Campbell. The lognormal distribution as a model for bio-optical variability in the sea. *J. Geophys. Res.*, 100(C7):13237–13254, 1995.
- K. L. Carder, R. G. Steward, G. R. Harvey, and P. B. Ortner. Marine humic and fulvic acids: their effects on remote sensing of ocean chlorophyll. *Limnol. Oceanogr.*, 34:68–81, 1989.
- J. Cartmill and M.-Y. Su. Bubble size distribution under saltwater and freshwater breaking waves. *Dyn. Atmos. Oceans*, 20:25–31, 1993.

- K. M. Case. Transfer problems and the reciprocity principle. *Rev. Modern Phys.*, 29(4): 651–658, 1957.
- P. Cermeno and F. G. Figueiras. Species richness and cell-size distribution: size structure of phytoplankton communities. *Marine Ecology Progress Series*, 357:79–85, 2008. doi: 10.3354/meps07293.
- I. Cetinić, M. J. Perry, N. T. Briggs, E. Kallin, E. A. D’Asaro, and C. M. Lee. Particulate organic carbon and inherent optical properties during 2008 North Atlantic Bloom Experiment. *J. Geophys. Res.*, 117(C06028), 2012. doi: 10.1029/2011JC007771.
- M. Chami, A. Thirouard, and T. Harmel. POLVSM (Polarized Volume Scattering Meter) instrument: an innovative device to measure the directional and polarized scattering properties of hydrosols. *Optics Express*, 22(21):26403–26428, 2014.
- S. Chandrasekhar. *Radiative Transfer*. Dover, 1960.
- G. C. Chang, T. D. Dickey, E. Boss, C. D. Mobley, and W. S. Pegau. Toward closure of upwelling radiance in coastal waters. *Applied Optics*, 42(9):1574–1582, 2003.
- P. A. Cherenkov. Visible light from clear liquids under the action of gamma radiation. *Dokl. Akad. Nauk SSSR*, 2(8):451–454, 1934. doi: <https://doi.org/10.1103/PhysRev.52.378>. Republished in *Phys. Rev.* 52, 378 in 1937.
- W. B. Cheston. *Elementary Theory of Electric and Magnetic Fields*. Wiley, 1964.
- S. W. Chisholm, R. J. Olson, E. R. Zettler, R. Goericke, J. B. Waterbury, and N. A. Welschmeyer. A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature*, 334:340–343, 1988.
- A. M. Ciotti, M. R. Lewis, and J. J. Cullen. Assessment of the relationships between dominant cell size in natural phytoplankton communities and the spectral shape of the absorption coefficient. *Limnol. Oceanogr.*, 47(2):404–417, 2002.
- W. R. Clavano, E. Boss, and L. Karp-Boss. Inherent optical properties of non-spherical marine-like particles — from theory to observations. *Oceanogr. Mar. Biol. Annual Rev.*, 45:1–38, 2007.
- P. G. Coble. Marine optical biogeochemistry: the chemistry of ocean color. *Chem. Rev.*, 107:402–418, 2007.
- J. W. Cooley and J. W. Tukey. An algorithm for the machine calculation of complex Fourier series. *Math. Comput.*, 19:297–301, 1965. doi: 10.2307/2003354.
- C. Cox and W. Munk. Statistics of the sea surface derived from Sun glitter. *Sears Found. J. Marine Research*, 13(2):198–227, 1954a.
- C. Cox and W. Munk. Measurement of the roughness of the sea surface from photographs of the Sun’s glitter. *J. Opt. Soc. Amer. A*, 44(11):838–850, 1954b.
- C. Cox and W. Munk. Some problems in optical oceanography. *Sears Found. J. Marine Research*, 14(1):63–78, 1955.

- G. B. Crawford and D. M. Farmer. On the spatial distribution of ocean bubbles. *J. Geophys. Res.*, 92:8231–8243, 1987.
- T. W. Cronin and N. Shashar. The linearly polarized light field in clear, tropical marine waters: spatial and temporal variation of light intensity, degree of polarization and e-vector angle. *J. Exp. Biol.*, 204:2461–2467, 2001.
- J. J. Cullen. Observation and prediction of harmful algal blooms. In M. Babin, C. S. Roesler, and J. J. Cullen, editors, *Real-time coastal observing systems for ecosystem dynamics and harmful algal blooms*. UNESCO, 2008.
- D. B. Curtis, B. Meland, M. Aycibin, N. P. Arnold, V. H. Grassian, M. A. Yound, and P. D. Kleiber. A laboratory investigation of light scattering from representative components of mineral dust aerosol at a wavelength of 550 nm. *J. Geophys. Res.*, 113 D08210, 2008. doi: 10.1029/2007JD009387.
- D. D’Alimonte and T. Kajiyama. Effects of light polarization and waves slope statistics on the reflectance factor of the sea surface. *Optics Express*, 24(8):7922–7942, 2016.
- G. Dall’Olmo, T. K. Westberry, M. J. Behrenfeld, E. Boss, and W. H. Slade. Significant contribution of large particles to optical backscattering in the open ocean. *Biogeosciences*, 6:947–967, 2009.
- G. E. Davis. Scattering of light by an air bubble in water. *J. Opt. Soc. Am.*, 45:572–581, 1955.
- B. Davison and J. B. Sykes. *Neutron Transport Theory*. Oxford, 1957.
- J. G. de Boer. On the correlation function in time and space of wind-generated ocean waves. Technical Report 160, SACLANT ASW Research Centre, 1969. URL <http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=AD0865249>.
- D. B. Deane. Sound generation and air entrainment by breaking waves in the surf zone. *J. Acoust. Soc. Am.*, 192:2671–2689, 1997.
- C. A. Deering and J. M. Stoker. Let’s agree on the casing of lidar. *LiDAR News Magazine*, 4(6):1–4, 2014.
- A. G. Dekker, S. R. Phinn, J. Anstee, W. P. Bissett, V. E. Brando, B. Casey, P. Fearn, J. Hedley, W. Klonowski, Z. P. Lee, M. Lynch, M. Lyons, C. D. Mobley, and C. Roelfsema. Intercomparison of shallow water bathymetry, hydro-optics, and benthos mapping techniques in Australian and Caribbean coastal environments. *Limnol. Oceanogr. Methods*, 9:396–425, 2011.
- C. E. Del Castillo and R. L. Miller. On the use of ocean colour remote sensing to measure the transport of dissolved organic carbon by the Mississippi River plume. *Rem. Sens. Environ.*, 112:836–844, 2008.
- E. J. Denton and F. J. Warren. The photosensitive pigments in the retinae of deep-sea fish. *J. Mar. Biol. Assoc. U.K.*, 36:651–662, 1957.
- E. L. Dereniak and D. G. Crowe. *Optical Radiation Detectors*. John Wiley and Sons, 1984.

- R. A. Desiderio. Application of the Raman scattering coefficient of water to calculations in marine optics. *Applied Optics*, 39:1893–1894, 2000.
- H. M. Dierssen, R. M. Kudela, J. P. Ryan, and R. C. Zimmerman. Red and black tides: Quantitative analysis of water-leaving radiance and perceived color for phytoplankton, colored dissolved organic matter, and suspended sediment. *Limnol. Oceanogr.*, 51(6): 2646–2659, 2006. doi: 10.4319/lo.2006.51.6.2646.
- L. S. Dolin. Theory of lidar method for measurement of the modulation transfer function of water types. *Applied Optics*, 52(2):199–207, 2013.
- D. Doxaran, J.-M. Froidefond, P. Castaing, and M. Babin. Dynamics of the turbidity maximum zone in a macrotidal estuary (the Gironde, France): observations from field and MODIS satellite data. *Estuarine, Coastal and Shelf Sci.*, 81:321–332, 2009.
- S. Q. Duntley. Measurements of the distribution of water wave slopes. *J. Opt. Soc. Amer.*, 44(7):574–575, 1954.
- S. Q. Duntley, R. J. Uhl, R. W. Austin, A. R. Boileau, and J. E. Tyler. An underwater photometer. page 904. Optical Soc. Amer., 1955. Abstract only of presentation at the Fortieth Annual Meeting of the O.S.A.
- L. M. N. Duysens. The flattening effect of the absorption spectra of suspensions as compared to that of solutions. *Biochim. Biophys. Acta*, 19:1–12, 1956.
- Eastman Kodak Co. *Wratten Light Filters*. Eastman Kodak Co., Rochester, New York, fourth, revised edition, 1920.
- R. Eckhardt. Stan Ulam, John von Neumann, and the Monte Carlo Method. *Los Alamos Science, Special Issue 1987*, pages 131–143, 1987. URL http://library.sciencemadness.org/lanl1_a/lib-www/pubs/00326867.pdf.
- S. Eibenberger, S. Gerlich, M. Arndt, M. Mayor, and J. Txen. Matter–wave interference of particles selected from a molecular library with masses exceeding 10 000 amu. *Physical Chemistry Chemical Physics*, 15(35):14696–14700, 2013. doi: 10.1039/C3CP51500A.
- A. Einstein. Theorie der Opaleszenz von homogenen Flüssigkeiten und Flüssigkeitsgemischen in der Nähe des kritischen Zustandes. *Ann. Physik*, 33: 1275–1298, 1910.
- R. Eisberg and R. Resnick. *Quantum Physics of Atoms, Molecules, Solids, Nuclei, and Particles; Second Edition*. John Wiley and Sons, 1985.
- M. A. Eleveld, A. B. Ruescas, A. Hommersom, T. S. Moore, S. W. M. Peters, and C. Brockmann. An optical classification tool for global lake waters. *Remote Sensing*, 9:420, 2017. doi: 10.3390/rs9050420.
- T. Elfouhaily, B. Chapron, K. Katsaros, and D. Vandemark. A unified directional spectrum for long and short wind-driven waves. *J. Geophys. Res.*, 102(C7):15781–15796, 1997.
- H. A. Erikson. Light intensity at different depths in lake water. *J. Optical Soc. Amer.*, 23: 170–177, 1933.

- M. L. Estapa, E. Boss, L. M. Meyer, and C. S. Roesler. Role of iron and organic carbon in mass-specific light absorption by particulate matter from Louisiana coastal waters. *Limnol. Oceanogr.*, 5757(1):97–112, 2012. doi: 10.4319/lo.2012.57.1.0097.
- P. G. Falkowski and D. A. Kiefer. Chlorophyll-a fluorescence in phytoplankton – relationship to photosynthesis and biomass. *J. Plankton Res.*, 7(5):715–731, 1985.
- P. G. Falkowski and T. G. Owens. Light–shade adaptation — two strategies in marine-phytoplankton. *Plant Physiol.*, 66(4):592–595, 1980.
- P. G. Falkowski and J. A. Raven. *Aquatic Photosynthesis: 2nd Edition*. Princeton University Press, 2007.
- P. G. Falkowski, H. Lin, and M. Y. Gorbunov. What limits photosynthetic energy conversion efficiency in nature? Lessons from the oceans. *Philosoph. Trans. Royal Soc. B*, page 372, 2016. doi: 10.1098/rstb.2016.0376.
- Y. Fan, W. Li, K. J. Voss, C. K. Gatebe, and K. Stamnes. Neural network method to correct bidirectional effects in water-leaving radiance. *Applied Optics*, 55(1):10–21, 2016.
- R. P. Feynman. *QED: The Strange Story of Light and Matter*. Princeton Univ. Press, 1985. Latest edition is 2014.
- R. P. Feynman, R. B. Leighton, and M. Sands. *The Feynman Lectures on Physics, vol 2*. Addison-Wesley, 1964.
- D. J. Finney. Dimensions of statistics. *J. Royal Stat. Soc. C (Applied Statistics)*, 26(3): 285–289, 1977. doi: 10.2307/2346969.
- P. D. Fitzgerald. Numerical approximation of Kramers-Kronig relations to transform discretized absorption data. *Archive for Physics*, 2020. URL <https://arxiv.org/abs/2012.02369>.
- P. J. Flatau, J. Piskozub, and J. R. V. Zaneveld. Asymptotic light field in the presence of a bubble-layer. *Optics Express*, 5:120–124, 1999.
- P. J. Flatau, M. Flatau, J. R. V. Zaneveld, and C. D. Mobley. Remote sensing of bubble clouds in sea water. *Quart. J. Royal. Meteorol. Soc.*, 126:2511–2523, 2000.
- D. Fleisch. *A Student’s Guide to Maxwell’s Equations*. Cambridge Univ. Press, 2008. URL <https://www.princeton.edu/ssp/joseph-henry-project/eddy-currents/A-students-guide-to-maxwells-equations-D.-FleischLEISC.pdf>.
- E. N. Flory, P. S. Hill, T. G. Milligan, and J. Grant. The relation between floc area and backscattering during a spring bloom. *Deep-Sea Res.*, 51:213–223, 2004.
- G. Fournier. Derivation of an explicit expression for the Fournier-Forand phase function in terms of the mean cosine. Technical report, Defence R&D Canada - Valcartier, Quebec, Canada, 2011.
- G. Fournier and J. L. Forand. Analytic phase function for ocean water. In J. S. Jaffee, editor, *Ocean Optics XII*, volume 2258, pages 194–201. SPIE Proc., 1994.

- G. Fournier and M. Jonasz. Computer based underwater imaging analysis. In G. Gilbert, editor, *Airborne and In-water Underwater Imaging*, volume 3761, pages 62–77. SPIE Proc., 1999.
- F. E. Fox and K. Herzfeld. Gas bubbles with organic skin as cavitation nuclei. *J. Acoust. Soc. Am.*, 26:984–989, 1954.
- I. M. Frank and I. E. Tamm. Coherent radiation of fast electrons in a medium. *Dokl. Akad. Nauk SSSR*, 14:109, 1937.
- T. M. Frank and E. A. Widder. UV light in the deepsea: In situ measurements of downwelling irradiance in relation to the visual threshold sensitivity of UVsensitive crustaceans. *Marine and Freshwater Behaviour and Physiology*, 27(2-3):189–197, 1996. doi: 10.1080/10236249609378964.
- B. Franz. Atmospheric Correction Discussion: Where we are today and where we might be in 2.5 years. Presentation at the PACE Science Team Meeting; Jan 2014, January 2014. URL https://pace.oceansciences.org/docs/sci2015_franz.pdf.
- B. Franz, S. W. Bailey, P. J. Werdell, and C. R. McClain. Sensor-independent approach to the vicarious calibration of satellite ocean color radiometry. *Applied Optics*, 46(22): 5068–5082, 2007.
- R. M. Friedman. *The Politics of Excellence: Behind the Nobel Prize in Science*. W H Freeman and Company, 2001. ISBN 978-0716731030.
- R. Frouin, M. Schwindling, and P. Y. Dechamps. Spectral reflectance of sea foam in the visible and near infrared: In situ measurements and remote sensing implications. *J. Geophys. Res.*, 101(C6):14361–14371, 1996.
- M. Fujii, E. Boss, and F. Chai. The value of adding optics to ecosystem models: a case study. *Biogeosciences Discuss.*, 4:1585–1631, 2007. URL <http://www.biogeosciences-discuss.net/4/1585/2007/bgd-4-1585-2007.pdf>.
- A. Gainusa Bogdan and E. S. Boss. Evaluation of a compact sensor for backscattering and absorption. *Applied Optics*, 50(21):3758–3772, 2011.
- G. Gamow. *Thirty Years that Shook Physics*. Dover, 1985. ISBN 978-0486248950.
- B.-C. Gao and A. F. H. Goetz. Column atmospheric water vapor and vegetation liquid water retrievals from airborne imaging spectrometer data. *J. Geophys Res.*, 95(D4): 3549–3564, 1990.
- B.-C. Gao, M. J. Montes, Z. Ahmad, and C. O. Davis. Atmospheric correction algorithm for hyperspectral remote sensing of ocean color from space. *Applied Optics*, 39(6):887–896, 2000.
- R. D. M. Garcia. Fresnel boundary and interface conditions for polarized radiative transfer in a multilayer medium. *J. Quant. Spectros. Rad. Trans.*, 113:306–307, 2012.

- Y. Ge, H. R. Gordon, and K. J. Voss. Simulation of inelastic-scattering contributions to the irradiance field in the ocean: variation in Fraunhofer line depths. *Applied Optics*, 32:4028–4036, 1993.
- A. A. Gershun. Fundamental ideas of the theory of a light field (vector methods of photometric calculations). *Izvestiya Akad. Nauk SSSR*, pages 417–430, 1936. In Russian.
- A. A. Gershun. The light field. *J. Math. and Phys.*, XVIII(2):51–151, 1939. Translation of Gershun (1936) by P. Moon and G. Timoshenko.
- J. A. Goodman, M. J. Montes, and S. L. Ustin. Applying TAFKAA for atmospheric correction of AVIRIS over coral ecosystems in the Hawaiian Islands. AVIRIS conference presentation; pdf online, 2003. URL <https://ntrs.nasa.gov/api/citations/20050192446/downloads/20050192446.pdf>.
- J. W. Goodman. *Introduction to Fourier Optics, Second Edition*. McGraw Hill, 1996. ISBN 0-07-024254-2.
- H. Gordon. Ship perturbation of irradiance measurements at sea. 1: Monte Carlo simulations. *Applied Optics*, 24(23):4172–4182, 1985.
- H. R. Gordon. Simple calculation of the diffuse reflectance of the ocean. *Applied Optics*, 12(12):2803–2804, 1973.
- H. R. Gordon. Diffuse reflectance of the ocean: The theory of its augmentation by chlorophyll a fluorescence at 685 nm. *Applied Optics*, 18(8):1161–1166, 1979.
- H. R. Gordon. Bio-optical model describing the distribution of irradiance at the sea surface resulting from a point source imbedded in the ocean. *Applied Optics*, 26(19):4133–4148, 1987.
- H. R. Gordon. Can the Lambert-Beer law be applied to the diffuse attenuation coefficient of ocean water? *Limnol. Oceanogr.*, 34(8):1389–1409, 1989.
- H. R. Gordon. Modeling and simulating radiative transfer in the ocean. In R.W. Spinrad, K.L. Carder, and M.J. Perry, editors, *Ocean Optics*, pages 1–39. Oxford, 1994a.
- H. R. Gordon. Equivalence of the point and beam spread function of scattering media: a formal demonstration. *Applied Optics*, 33(6):1120–1122, 1994b.
- H. R. Gordon. Remote sensing of ocean color: a methodology for dealing with broad spectral bands and significant out-of-band response. *Applied Optics*, 34(36):8363–8374, 1995.
- H. R. Gordon. Atmospheric correction of ocean color imagery in the Earth observing system era. *J. Geophys. Res.*, 102D:17081–17106, 1997.
- H. R. Gordon. Normalized water-leaving radiance: revisiting the influence of surface roughness. *Applied Optics*, 44(2):241–248, 2005.
- H. R. Gordon. Backscattering of light from disklike particles: is fine-scale structure or gross morphology more important? *Applied Optics*, 45:7166–7173, 2006.

- H. R. Gordon. Rayleigh-Gans scattering approximation: surprisingly useful for understanding backscattering from disk-like particles. *Optics Express*, 15(9):5572–5588, 2007.
- H. R. Gordon. *Physical Principles of Ocean Color Remote Sensing*. IOCCG, 2019. URL https://ioccg.org/wp-content/uploads/2019/11/gordon_book_nov_20_2019.pdf.
- H. R. Gordon and D. K. Clark. Clear water radiances for atmospheric correction of Coastal Zone Color Scanner imagery. *Applied Optics*, 20:4175–4780, 1981.
- H. R. Gordon and K. Ding. Self-shading of in-water optical instruments. *Limnol. Oceanogr.*, 1992.
- H. R. Gordon and T. Du. Light scattering by nonspherical particles: Application to coccoliths detached from *Emiliania huxleyi*. *Limnol. Oceanogr.*, 46:1438–1454, 2001.
- H. R. Gordon and B. Franz. Remote sensing of ocean color: Assessment of the water-leaving radiance bidirectional effects on the atmospheric diffuse transmittance for SeaWiFS and MODIS intercomparisons. *Rem. Sens. Environ.*, 112:2677–2685, 2008.
- H. R. Gordon and W. R. McCluney. Estimation of the depth of sunlight penetration in the sea for remote sensing. *Applied Optics*, 14(2):413–416, 1975.
- H. R. Gordon and A. Morel. *Remote Assessment of Ocean Color for Interpretation of Satellite Visible Imagery: A Review*, volume 4. Springer-Verlag, 1983.
- H. R. Gordon. and M. Wang. Surface-roughness considerations for atmospheric correction of ocean color sensors. I: the Rayleigh-scattering component. *Applied Optics*, 31(21):4247–4260, 1992.
- H. R. Gordon and M. Wang. Retrieval of water-leaving radiance and aerosol optical thickness over the oceans with SeaWiFS: a preliminary algorithm. *Applied Optics*, 33:443–452, 1994a.
- H. R. Gordon and M. Wang. Influence of oceanic whitecaps on atmospheric correction of ocean-color sensor. *Applied Optics*, 33(33):7754–7763, 1994b.
- H. R. Gordon, O. B. Brown, R. H. Evans, J. W. Brown, R. C. Smith, K. S. Baker, and D. K. Clark. A semianalytical radiance model of ocean color. *J. Geophys. Res.*, 93(D9):10909–10924, 1988.
- H. R. Gordon, T. Du, and T. Zhang. Atmospheric correction of ocean color sensors: analysis of the effects of residual instrument polarization sensitivity. *Applied Optics*, 36(27):6938–6948, 1997a.
- H. R. Gordon, T. Du, and T. Zhang. Remote sensing of ocean color and aerosol properties: resolving the issue of aerosol absorption. *Applied Optics*, 36(33):8670–8684, 1997b.
- H. R. Gordon, M. R. Lewis, S. D. McLean, M. S. Twardowski, S. A. Freeman, K. J. Voss, and G. C. Boynton. Spectra of particulate backscattering in natural waters. *Optics Express*, 17(18):16192–16208, 2009.

- R. Greenler. *Rainbows, Halos and Glories*, volume PM321. SPIE Press, 2020. Republication of the original 1980 edition.
- W. Gregg and K. L. Carder. A simple solar irradiance model for cloudless maritime atmospheres. *Limnol. Oceanogr.*, 38(5):1657–1675, 1990.
- D. J. Griffiths. *Introduction to Electrodynamics*. Prentice-Hall, 1981.
- D. J. Griffiths. *Introduction to Elementary Particles, Second Edition*. Wiley, 2008.
- L. Guidi, L. Stemann, G. A. Jackson, F. Ibanez, H. Claustre, L. Legendre, M. Picheral, and G. Gorski. Effects of phytoplankton community on production, size, and export of large aggregates: a world-ocean analysis. *Limnol. Oceanogr.*, 54:1951–1963, 2009.
- S. H. D. Haddock, M. A. Moline, and J. F. Case. Bioluminescence in the sea. *Annual Review of Marine Science*, 2:443–493, 2010. doi: 10.1146/annurev-marine-120308-081028.
- M. A. Haines and B. D. Johnson. Injected bubble populations in seawater and fresh water measured by a photographic method. *J. Geophys. Res.*, 100:7057–7068, 1995.
- D. Halliday and R. Resnick. *Fundamentals of Physics, Third Edition*. John Wiley and Sons, 1988.
- D. Hanneke, S. Fogwell Hoogerheide, and G. Gabrielse. Cavity control of a single-electron quantum cyclotron: Measuring the electron magnetic moment. *Phys. Rev. A*, 83, 2011. doi: 10.1103/PhysRevA.83.052122.
- J. E. Hansen. Multiple scattering of polarized light in planetary atmospheres. Part II. Sunlight reflected by terrestrial water clouds. *J. Atmos. Sci.*, 28:1400–1426, 1971.
- J. E. Hansen and L. D. Travis. Light scattering in planetary atmospheres. *Space Sci. Rev.*, 16:527–610, 1974.
- B. Hapke. *Theory of Reflectance and Emittance Spectroscopy*. Cambridge Univ. Press, 1993.
- T. Harmel, M. Hieronymi, W. Slade, R. Röttgers, F. Roullier, and M. Chami. Laboratory experiments for inter-comparison of three volume scattering meters to measure angular scattering properties of hydrosols. *Optics Express*, 24(2):A234–A256, 2016. doi: 10.1364/OE.24.00A234.
- T. Harmel, J. Agagliate, M. Hieronymi, and P. Gernez. Two-term Reynolds-McCormick phase function parameterization better describes light scattering by microalgae and mineral hydrosols. *Optics Letters*, 46(8):1860–1863, 2021.
- A. H. Harvey, J. S. Gallagher, and J. M. H. Levelt Sengers. Revised formulation of the refractive index of water and steam as a function of wavelength, temperature and density. *J. Phys. Chem. Ref. Data*, 27:761–774, 1998. doi: 10.1063/1.556029.
- A. Hatcher, P. Hill, and J. Grant. Optical backscatter of marine flocs. *J. Sea Res.*, 46: 1–12, 2001.

- D. Haubrich, J. Musser, and E. S. Fry. Instrumentation to measure the backscattering coefficient b_b for arbitrary phase functions. *Applied Optics*, 50(21):4134–4147, 2011.
- S. K. Hawes. Quantum Fluorescence Efficiencies of Marine Fulvic and Humic Acids. Master's thesis, Univ. South Florida, 1992.
- S. K. Hawes, K. L. Carder, and G. R. Harvey. Quantum fluorescence efficiencies of fulvic and humic acids: effects on ocean color and fluorometric detection. In *Ocean Optics XI*, volume 1750, pages 212–223. SPIE Proc., 1992.
- E. Hecht. *Optics, Second Edition*. Addison-Wesley, 1989. 1987 Edition, reprinted with corrections.
- J. D. Hedley and C. D. Mobley. *HydroLight 6.0 - EcoLight 6.0 Technical Documentation*. Numerical Optics, Ltd., Belmont House, 19 West Street, Tiverton, EX16 8AA, UK, December 2019a. URL <https://www.numopt.com/>.
- J. D. Hedley and C. D. Mobley. *HydroLight 6.0 - EcoLight 6.0 Users' Guide*. Numerical Optics, Ltd., Belmont House, 19 West Street, Tiverton, EX16 8AA, UK, December 2019b. URL <https://www.numopt.com/>.
- L. C. Henyey and J. L. Greenstein. Diffuse radiation in the galaxy. *Astrophys. J.*, 93:70–83, 1941.
- M. L. Heron. Short-wave ocean wave slope models for use in remote sensing data analysis. *IEEE Transactions Geosci. Rem. Sens.*, 44(7):1962–1973, 2006.
- S. G. Herring. A systematic survey of the modeled optical properties of nonspherical marine-like particles. Master's thesis, Oregon State Univ, 2002.
- N. Hoepffner and Sathyendranath. Effect of pigment composition on absorption properties of phytoplankton. *Mar. Ecol. Prog. Ser.*, 73(1):11–23, 1991.
- F. E. Hoge, C. W. Wright, W. B. Krabill, R. R. Buntzen, G. D. Gilbert, R. N. Swift, J. K. Yungel, and R. E. Berry. Airborne lidar detection of subsurface oceanic scattering layers. *Applied Optics*, 27:3969–3977, 1988.
- N. Højerslev. A spectral light absorption meter for measurements in the sea. *Limnol. Oceanogr.*, 20(6):1027–1034, 1975.
- N. Højerslev and J. R. V. Zaneveld. A theoretical proof of the existence of the submarine asymptotic daylight field. Technical Report 34, Københavns Univ. Inst. Fysisk Oceanog., Copenhagen, 1977.
- J. W. Holl. Nuclei and Cavitation. *J. Basic Eng.*, 92:681–688, 1970.
- L. H. Holthuijsen. *Waves in Oceanic and Coastal Waters*. Cambridge Univ. Press, 2007. doi: <https://doi.org/10.1017/CBO9780511618536>.
- K. V. Horoshenkov, A. Nichols, S. J. Tait, and G. A. Maximov. The pattern of surface waves in a shallow free surface flow. *J. Geophys. Res. Earth Surf.*, 118:1864–1876, 2013.

- W. Hou. *Ocean Sensing and Monitoring: Optics and Other Methods*, volume TT98. SPIE Press, 2013.
- W. Hou, K. L. Carder, and D. K. Costello. Scattering phase function of very large particles in the ocean. In S. G. Ackleson and R. Frouin, editors, *Ocean Optics XIII*, volume 2263, pages 579–584. SPIE, 1996.
- W. Hou, Z.-P. Lee, and A. W. Weidemann. Why does the Secchi disk disappear? An imaging perspective. *Optics Express*, 15(6):2791–2802, 2007.
- W. Hou, D. J. Gray, A. D. Weidemann, and R. A. Arnone. Comparison and validation of point spread models for imaging in natural waters. *Optics Express*, 16(13):9958–9965, 2008.
- L Hu, X. Zhang, Y. Xiong, D. J. Gray, and M.-X. He. Variability of relationship between the volume scattering function at 180° and the backscattering coefficient for aquatic particles. *Applied Optics*, 59(10):C31–C41, 2020. doi: 10.1364/AO.383229.
- Y.-X. Hu, D. Winker, P. Yang, B. Baum, L. Poole, and L. Vann. Identification of cloud phase from PICASSO-CENA lidar depolarization: a multiple scattering sensitivity study. *J. Quant. Spectros. Rad. Trans.*, 70:569–579, 2001.
- P. A. Huang, D. W. Wang, E. J. Walsh, W. B. Krabill, and R. N. Swift. Airborne measurements of the wavenumber spectra of ocean surface waves. Part II: Directional distribution. *J. Phys. Oceanogr.*, 30:2768–2787, 2000.
- T. B. Huffman and D. L. Zveare. Sound Speed Dispersion, Attenuation and Inferred Microbubbles in the Upper Ocean. Master’s thesis, Naval Postgraduate School, 1974.
- P. D. T. Huibers. Models for the wavelength dependence of the index of refraction of water. *Applied Optics*, 36(16):3785–3787, 1997. doi: 10.1364/AO.36.003785.
- Y. Huot, C. A. Brown, and J. J. Cullen. New algorithms for MODIS sun-induced chlorophyll fluorescence and a comparison with present data products. *Limnol. Oceanogr. Methods*, 3:108–130, 2005.
- Y. Huot, A. Morel, M. S. Twardowski, D. Stramski, and R. A. Reynolds. Particle optical backscattering along a chlorophyll gradient in the upper layer of the eastern south pacific ocean. *Biogeosciences*, 5:495–507, 2008.
- R. Iturriaga and D. A. Siegel. Microphotometric characterization of phytoplankton and detrital absorption properties in the Sargasso Sea. *Limnol. Oceanogr.*, 34(8):1706–1726, 1989.
- G. A. Jackson, R. Maffione, D. K. Costello, A. L. Alldredge, B. E. Logan, and H. G. Dam. Particle size spectra between $1\ \mu\text{m}$ and $1\ \text{cm}$ at Monterey Bay determined using multiple instruments. *Deep Sea Res. Part I Oceanogr. Res. Papers*, 44(11):1739–1767, 1997.
- J. D. Jackson. *Classical Electrodynamics*. John Wiley and Sons, 1962.
- J. D. Jackson. *Classical Electrodynamics, Second Edition*. John Wiley and Sons, 1975.

- S. W. Jeffrey and M. Vesk. Introduction to marine phytoplankton and their pigment signatures. In *Phytoplankton Pigments in Oceanography: Guidelines to Modern Methods*, Monographs on Oceanographic Methodology, pages 37–84. UNESCO, 1997.
- S. W. Jeffrey, S. W. Wright, and M. Zapata. *Phytoplankton Pigments*, chapter Microalgal classes and their signature pigments, pages 3–77. Cambridge Univ. Press, 2012. doi: 10.1017/CBO9780511732263.
- N. G. Jerlov. *Optical Oceanography*. 5. Elsevier Oceanography, 1968.
- N. G. Jerlov. *Marine Optics*. 14. Elsevier Oceanography, 1976.
- N. G. Jerlov and K. Nygard. A quanta and energy meter for photosynthetic studies. Technical Report 10, Copenhagen Univ. Inst. Fysisk Oceanog., 1969.
- G. Johnsen, N. B. Nelson, R. V. M. Jovine, and B. B. Prezelin. Chromoprotein-dependent and pigment-dependent modeling of spectral light-absorption in 2 dinoflagellates, *Prorocentrum minimum* and *Heterocapsa pygmaea*. *Mar. Ecol. Prog. Ser.*, 114(3):245–258, 1994.
- B. D. Johnson. Bubble populations: Background and breaking waves. In E. C. Monahan and G. MacNiocail, editors, *Oceanic Whitecaps and Their Role in Air-sea Exchange Processes*, pages 69–73. D. Reidel, 1986.
- B. D. Johnson and R. C. Cooke. Bubble populations and spectra in coastal waters: A photographic approach. *J. Geophys. Res.*, 84:3761–3766, 1979.
- B. D. Johnson and R. C. Cooke. Generation of stabilized microbubbles in seawater. *Science*, 213:209–211, 1981.
- N. G. Johnson and G. Liljequist. On the angular distribution of submarine daylight and on the total submarine illumination. *Svenska Hydrografisk Biol. Komm. Skrifter*, 14:3–15, 1938.
- M. Jonasz. Particle-size distributions in the Baltic. *Tellus*, 35B:346–358, 1983.
- M. Jonasz. Nonsphericity of suspended marine particles and its influence on light scattering. *Limnol. Oceanogr.*, 32:1059–1065, 1987.
- M. Jonasz and G. R. Fournier. *Light Scattering by Particles in Water: Theoretical and Experimental Foundations*. Academic Press, 2007.
- C. Junge. Die Rolle der Aerosole und der gasförmigen Beimengungen der Luft im Spurenstoffhaushalt der Troposphäre. *Tellus*, 5(1):1–26, 1953. doi: 10.3402/tellusa.v5i1.8567.
- C. Junge. The size distribution and aging of natural aerosols as determined from electrical and optical data on the atmosphere. *J. Meteorology*, 12:13–25, 1955.
- G. W. Kattawar. A three-parameter analytic phase function for multiple scattering calculations. *J. Quant. Spectrosc. Radiat. Trans.*, 15:839–849, 1975.
- G. W. Kattawar. Polarization of Light in the Ocean. In R. W. Spinrad, K. L. Carder, and M. J. Perry, editors, *Ocean Optics*, pages 202–225. Oxford Univ. Press, 1994.

- G. W. Kattawar and C. N. Adams. Stokes vector calculations of the submarine light field in an atmosphere-ocean with scattering according to a Rayleigh phase matrix: Effect of interface refractive index on radiance and polarization. *Limnol. Oceanogr.*, 34:1453–1472, 1989.
- G. W. Kattawar and G. N. Plass. Asymptotic radiance and polarization in optically thick media: ocean and clouds. *Applied Optics*, 15(12):3166–3178, 1976.
- G. W. Kattawar and X. Xu. Filling in of Fraunhofer lines in the ocean by Raman scattering. *Applied Optics*, 31(30):6491–6500, 1992.
- S. Kay, J. Hedley, S. Lavender, and A. Nimmo-Smith. Light transfer at the ocean surface modeled using high resolution sea surface realizations. *Optics Express*, 19:6493–6505, 2011.
- K. L. Kelly. Color designations for lights. *J. Opt. Soc. Amer.*, 33(11):627–632, 1943.
- M. G. Kelly and P. Tett. *Bioluminescence in Action*, chapter Bioluminescence in the Ocean, pages 399–417. Academic Press, 1978.
- W. H. Kelly, G. B. Beard, and R. A. Peters. The beta decay of K-40. *Nuclear Physics*, 11: 492–498, 1959.
- M. Kerker. *The Scattering of Light and Other Electromagnetic Radiation*. Academic Press, 1969. 666 pages.
- A. Khelifa and P. S. Hill. Models for effective density and settling velocity of flocs. *J. Hydraul. Res.*, 44:390–401, 2006.
- R. Kidd, J. Ardini, and A. Anton. Evolution of the modern photon. *Amer. J. Phys.*, 17 (1):27–35, 1989.
- D. A. Kiefer and R. A. Reynolds. Advances in understanding phytoplankton fluorescence and photosynthesis. In P. G. Falkowski and A. D. Woodhead, editors, *Primary Productivity and Biogeochemical Cycles in the Sea*, pages 155–174. Plenum Press, 1992.
- D. L. Kingsbury and P. L. Marston. Mie scattering near the critical angle of bubbles in water. *J. Opt. Soc. Amer.*, 71:358–361, 1981.
- B. Kinsman. *Wind Waves: Their Generation and Propagation on the Ocean Surface*. Prentice-Hall, Inc., 1965.
- J. T. O. Kirk. A theoretical analysis of the contribution of algal cells to the attenuation of light within natural waters. III. Cylindrical and spheroidal cells. *New Phytologist*, 77: 341–358, 1976.
- J. T. O. Kirk. Spectral adsorption properties of natural waters: contribution of the soluble and particulate fractions to light absorption in some inland waters of south-eastern Australia. *Australian Journal of Marine and Freshwater Research*, 31(3):287–296, 1980. doi: 10.1071/MF9800287.

- J. T. O. Kirk. *Light and Photosynthesis in Aquatic Ecosystems, Second Edition*. Cambridge Univ. Press, 1994.
- G. J. Kirkpatrick, D. F. Millie, M. A. Moline, and O. Schofield. Optical discrimination of a phytoplankton species in natural mixed population. *Limnol. Oceanogr.*, 45(2):467–471, 2000.
- M. Kishino, M. Takahashi, N. Okami, and S. Ichimura. Estimation of the spectral absorption coefficients of phytoplankton in the sea. *Bulletin of Marine Sci.*, 37(2):634–642, 1985.
- J. C. Kitchen and J. R. V. Zaneveld. A three-layered sphere model of the optical properties of phytoplankton. *Limnol. Oceanogr.*, 37(8):1680–1690, 1992.
- P. Koepke. Effective reflectance of oceanic whitecaps. *Applied Optics*, 23(11):1816–1824, 1984.
- I. Koike, H. Shigemitsu, T. Kazuki, and K. Kazuhiro. Role of sub-micrometre particles in the ocean. *Nature*, 345:242–244, 1990.
- A. A. Kokhanovsky and E. P. Zege. Optical properties of aerosol particles: a review of approximate analytical solutions. *J. Aerosol. Sci.*, 28(1):1–21, 1997.
- D. A. Kolovayev. Investigation of the concentration and statistical size distribution of wind-produced bubbles in the near-surface ocean. *Oceanology*, 15:659–661, 1976. English translation.
- I. Kostakis, M. Twardowski, C. Roesler, R. Röttgers, D. Stramski, D. McKee, A. Tonizzo, and S. Drapeau. Hyperspectral optical absorption closure experiment in complex coastal waters. *Limnol. Oceanogr. Methods*, 19(3):589–625, 2021. doi: 10.1002/lom3.10447.
- T. Kutser, D. C. Pierson, K. Y. Kallio, A. Reinart, and S. Sobek. Mapping lake CDOM by satellite remote sensing. *Remote Sensing Environ.*, 94:535–540, 2005.
- E. Lamarre and W. K. Melville. Air entrainment and dissipation in breaking waves. *Nature*, 351:469–472, 1991.
- W. Lamb. Anti-photon. *Applied Physics B*, 60:77–84, 1995.
- J. H. Lambert. *Photometria sive de mensura et gradibus luminis, colorum et umbrae*. Eberhardt Klett, Augsburg, Germany, 1760. [Photometry, or, On the measure and gradations of light intensity, colors, and shade] (in Latin).
- D. Lapota, J. R. Losee, and M. R. Geiger. Bioluminescence displays induced by pulsed light. *Limnol. Oceanogr.*, 31(4):887–889, 1986.
- D. Lapota, C. Galt, J. R. Losee, H. D. Huddel, J. K. Orzech, and K. H. Neilson. Observations and measurements of planktonic bioluminescence in and around a milky sea. *J. Exp. Mar. Biol. Ecol.*, 119:55–81, 1988.
- J. C. Larsen. An introduction to electromagnetic induction in the ocean. *Phys. Earth Planetary Interiors*, 7:389–398, 1973.

- J. C. Larsen. Transport and Heat Flux of the Florida Current at 27 degrees N Derived from Cross-Stream Voltages and Profiling Data: Theory and Observations. *Phil. Trans. Royal Soc. London A*, 339(1650):169–236, 1992. doi: 10.1098/rsta.1992.0007.
- P. Latimer. Experimental tests of a theoretical method for predicting light scattering by aggregates. *Applied Optics*, 24(19):3231–3239, 1985.
- G. E. Latta and J. A. Bailie. On the autocorrelation functions of wind generated ocean waves. *Zeit. Angew. Math.*, 19:575–586, 1968.
- J. Learned. Attenuation of Cherenkov photons in the ocean. Internal Report DIR-10-81, Hawaii DUMAND Center, Univ. of Hawaii, Honolulu, 11 pages., 1981.
- R. A. Leathers, T. V. Downes, and C. D. Mobley. Self-shading correction for oceanographic upwelling radiometers. *Optics Express*, 12(20):4709–4718, 2004. doi: 10.1364/OPEX.12.004709.
- M. E. Lee and M. R. Lewis. A new method for the measurement of the optical volume scattering function in the upper ocean. *Atmos. Ocean. Technol.*, 20:563–571, 2003.
- Z. Lee, C. Hu, S. Shang, K. Du, M. Lewis, R. Arnone, and R. Brewin. Penetration of UV-visible solar radiation in the global oceans: insights from ocean color remote sensing. *J. Geophys. Res. Oceans*, 118:4241–4255, 2013. doi: 10.1002/jgrc.20308.
- Z.-P. Lee, R. A. Arnone, C.-M. Hu, P. J. Werdell, and B. Lubac. Uncertainties of optical parameters and their propagations in an analytical ocean color inversion algorithm. *Applied Optics*, 49(3):369–381, 2010.
- Z.-P. Lee, K. Du, K. J. Voss, G. Zibordi, B. Lubac, R. Arnone, and A. Weidemann. An inherent-optical-property-centered approach to correct the angular effects in water-leaving radiance. *Applied Optics*, 50(19):3155–3167, 2011.
- Z.-P. Lee, S. Shang, C. Hu, K. Du, A. Weidemann, W. Hou, J. Lin, and G. Lin. Secchi disk depth: A new theory and mechanistic model for underwater visibility. *Rem. Sens. Environ.*, 169:139–149, 2015.
- G. D. Leeuw and L. H. Cohen. Bubble size distribution in coastal seas. In B. Jahne and E. C. Monahan, editors, *Air-Water Gas Transfer*. AEON Verlag and Studio, 1995.
- R. B. Leighton. *Principles of Modern Physics*. McGraw-Hill, 1959.
- M. Lesser, C. D. Mobley, J. D. Hedley, and M. Slattery. Incident light on mesophotic corals is constrained by reef topography and colony morphology. *Mar. Ecol. Prog. Series*, pages 13445–13454, 2021. doi: 10.1002/ece3.8066.
- M. P. Lesser, M. Slattery, and C. D. Mobley. Biodiversity and functional ecology of mesophotic coral reefs. *Annual Review of Ecology, Evolution, and Systematics*, 49:49–71, 2018.
- C. Li, W. Cao, J. Yu, T. Ke, G. Lu, Y. Yang, and C. Guo. An instrument for in situ measuring of the volume scattering function of water: design, calibration and primary experiments. *Sensors*, 12:4514–4533, 2012. doi: 10.3390/s120404514.

- L. Li, D. Stramski, and M. Darecki. Characterization of the light field and apparent optical properties in the ocean euphotic layer based on hyperspectral measurements of irradiance quartet. *Applied Sciences*, 8:2677–2710, 2018. doi: 10.3390/app8122677.
- X. Li and B. E. Logan. Size distributions and fractal properties of particles during a simulated phytoplankton bloom in a mesocosm. *Deep-Sea Res. II*, 42(1):125–138, 1995. doi: 10.1016/0967-0645(95)00008-E.
- R. L. Liboff. *Introductory Quantum Mechanics*. Addison-Wesley, 1980.
- B. Light, G. A. Maykut, and T. C. Grenfell. A two-dimensional Monte Carlo model of radiative transfer in sea ice. *J. Geophys. Res. Oceans*, 108(C7):3219–3236, 2003. doi: 10.1029/2002JC001513.
- J. B. Liley. Fitting size distributions to optical particle counter data. *Aerosol Sci. Tech.*, 17(2):84–92, 1992. doi: 1080/02786829208959562.
- P. Lilienfeld. Gustav Mie: The Person. *Applied Optics*, 30(33):4696–4698, 1991.
- S. C. Ling and H. P. Pao. Study of micro-bubbles in the North Sea. In B. R. Kerman, editor, *Sea Surface Sound*, pages 197–210. Kluwer Academic Publishers, 1988.
- H. Loisel and A. Morel. Light scattering and chlorophyll concentration in case 1 waters: A reexamination. *Limnol. Oceanogr.*, 43(5):847–858, 1998.
- F. Lombard, E. Boss, A. M. Waite, M. Vogt, J. Uitz, L. Stemann, H. M. Sosik, J. Schulz, J-B. Romagnan, M. Picheral, J. Pearlman, M. D. Ohman, B. Niehoff, K. O. Möller, P. Miloslavich, A. Lara-Lopez, R. Kudela, R. M. Lopes, R. Kiko, L. Karp-Boss, J. S. Jaffe, M. H. Iversen, J-O. Irisson, K. Fennel, H. Hauss, L. Guidi, G. Gorsky, L. C. Giering, P. Gaube, S. Gallagher, G. Dubelaar, R. K. Cowen, F. Carlotti, C. Briseo-Avena, L. Berline, K. Benoit-Bird, N. Bax, S. Batten, S. D. Ayata, L. F. Artigas, and W. Appeltans. Globally consistent quantitative observations of planktonic ecosystem. *Front. Mar. Sci.*, 6, 2019. doi: 10.3389/fmars.2019.00196F.
- M. S. Longuet-Higgins, D. E. Cartwright, and N. D. Smith. Observations of the directional spectrum of sea waves using the motions of a flotation buoy. In *Ocean Wave Spectra*, pages 111–136. Prentice-Hall, 1963.
- J. K. Lotsberg, E. Marken, J. J. Stamnes, S. R. Erga, K. Aursland, and C. Olseng. Laboratory measurements of light scattering from marine particles. *Limnol. Oceanogr. Methods*, 5:34–40, 2007.
- B. Lubac and H. Loisel. Variability and classification of remote sensing reflectance spectra in the eastern English Channel and southern North Sea. *Rem. Sens. Environ.*, 110(1): 45–58, 2007. doi: 10.1016/j.rse.2007.02.012.
- R. V. Lynch III. The occurrence and distribution of surface bioluminescence in the oceans during 1966 through 1972. Technical Report 8210, Naval Research Lab., 1978. URL <https://apps.dtic.mil/sti/pdfs/ADA054638.pdf>.

- I. MacCallum, A. Cunningham, and D. McKee. The measurement and modelling of light scattering by phytoplankton cells at narrow forward angles. *J. Optics A: Pure Appl. Opt.*, 6:698–702, 2004.
- R. A. Maffione. *Optical Backscattering and Submerged Source Techniques for Characterizing the Optical Properties of the Ocean*. phdthesis, Oregon State Univ., August 1996.
- R. A. Maffione and D. R. Dana. In-situ characterization of optical backscattering and attenuation for lidar applications. In V. I Feigl and Yu. Kopilevich, editors, *Laser Remote Sensing of Natural Waters*, volume 2964. Proc. SPIE, November 1996.
- R. A. Maffione and D. R. Dana. Instruments and methods for measuring the backward-scattering coefficient of ocean waters,. *Applied Optics*, 36:6057–6067, 1997.
- R. A. Maffione and R. C. Honey. Instrument for measuring the volume scattering function in the backward direction. In G. D. Gilbert, editor, *Ocean Optics XI*, volume 1750, pages 15–26. SPIE Proc., 1992. doi: 10.1117/12.140650.
- R. A. Maffione, K. J. Voss, and R. C. Honey. Measurement of the spectral absorption coefficient in the ocean with an isotropic light source. *Applied Optics*, 32(18):3273–3279, 1993.
- F. Maggi. Variable fractal dimension: A major control for floc structure and flocculation kinematics of suspended cohesive sediment. *J. Geophys. Res.*, 112(C7):C07012, 2007.
- J. Makhoul. A fast cosine transform in one and two dimensions. *IEEE Trans. Acoustics, Speech, Signal Process.*, 28(1):27–34, 1980.
- A. Mannino, R. Vandermeulen, A. Neeley, and M. Lomas. Optical and pigment phytoplankton hyperspectral library. Talk at the NASA Ocean Color Research Team Meeting, Busan, Korea, April 2019. URL <https://oceancolor.gsfc.nasa.gov/meetings/scienceteam/ocrt/apr2019>.
- S. Maritorena, A. Morel, and B. Gentili. Determination of the fluorescence quantum yield by oceanic phytoplankton in their natural habitat. *Applied Optics*, 39(36):6725–6737, 2000.
- B. R. Marshall and R. C. Smith. Raman scattering and in-water ocean optical properties. *Applied Optics*, 29:71–84, 1990.
- P. L. Marston. Critical angle scattering by a bubble: physical-optics approximation and observations. *J. Opt. Soc. Amer.*, 69:1205–1211, 1979.
- P. L. Marston and D. L. Kingsbury. Scattering by a bubble in water near the critical angle: Interference effects. *J. Opt. Soc. Amer.*, 71:192–196, 1981.
- P. L. Marston, D. S. Langley, and D. L. Kingsbury. Light scattering by bubbles in liquids: Mie theory, physical-optics approximations and experiments. *J. Appl. Sci. Res.*, 38: 373–383, 1982.

- P. L. Marston, B. Billette, and C. Dean. Scattering of light by a coated bubble in water near the critical and Brewster scattering angles. In *Ocean Optics IX*, volume 925, pages 308–316. SPIE Proc., 1988.
- S. R. Massel. On the geometry of ocean surface waves. *Oceanologia*, 53(2):521–548, 2011.
- B. R. Masters. Lord Rayleigh: A Scientific Life. *Optics and Photonics*, 2009.
- C. F. Matta, L. Massa, A. V. Gubskaya, and E. Knoll. Can one take the logarithm or the sine of a dimensioned quantity or a unit? Dimensional analysis involving transcendental functions. *J. Chem. Educ.*, 88(1):6770, 2011. doi: 10.1021/ed1000476.
- J. C. Maxwell. A dynamical theory of the electromagnetic field. *Proc. Royal Soc. London*, pages 459–512, 1864. URL https://en.wikisource.org/wiki/A_Dynamical_Theory_of_the_Electromagnetic_Field.
- J. W. McLean and K. J. Voss. Point spread function in ocean water: comparison between theory and experiment. *Applied Optics*, 30(15):2027–2030, 1991.
- J. W. McLean, J. D. Freeman, and R. E. Walker. Beam spread function with time dispersion. *Applied Optics*, 37(21):4701–4711, 1998.
- R. M. Measures. *Laser Remote Sensing: Fundamentals and Applications*. Krieger, 1992.
- H. Medwin. In situ acoustic measurements of bubble populations in coastal ocean waters. *J. Geophys. Res.*, 75:599–611, 1970. doi: 10.1029/JC075i003p00599.
- H. Medwin. In situ acoustic measurements of microbubbles at sea. *J. Geophys. Res.*, 82: 971–976, 1977.
- G. Meister, E. J. Kwiatkowska, B. A. Franz, F. S. Patt, G. C. Feldman, and C. R. McClain. Moderate-resolution imaging spectroradiometer ocean color polarization correction. *Applied Optics*, 44(26):5524–5535, 2005.
- F. Melin and V. Vantrepotte. How optically diverse is the coastal ocean? *Rem. Sens. Environ.*, 160:235–251, 2015. doi: 10.1016/j.rse.2015.01.023.
- L. E. Mertens. *In-water Photography: Theory and Practice*. Wiley-Interscience, 1970.
- L. E. Mertens and R. S. Replogle. Use of point spread and beam spread functions for analysis of imaging systems in water. *J. Opt. Soc. Amer.*, 67(8):1105–1117, 1977.
- N. Metropolis and S. Ulam. The Monte Carlo method. *J. Amer. Statistical Assoc.*, 44 (247):335–341, 1949.
- J. R. Meyer-Arendt. Radiometer and photometry: units and conversion factors. *Applied Optics*, 7:2081–2084, 1968.
- G. Mie. Beiträge zur optik trüber medien, speziell kolloidaler metallösungen. *Ann. Physik, Vierte Folge*, 25(3):377–445, 1908.

- S. D. Miller, S. H. D. Haddock, C. D. Elvidge, and T. F. Lee. Detection of a bioluminescent milky sea from space. *Proc. Nat. Acad. Sci. U.S.A.*, pages 14181–14184, 2005. doi: 10.1073/pnas.0507253102.
- S. D. Miller, S. H. D. Haddock, W. C. Straka III, C. J. Seaman, C. L. Combs, M. Wang, W. Shi, and S.-H. Nam. Honing in on bioluminescent milkyseas from space. *Scientific Reports*, 11:15443–15452, 2021. doi: 10.1038/s41598-021-94823-z.
- M. I. Mishchenko. Multiple scattering by particles embedded in an absorbing medium. 2. Radiative transfer equation. *J. Quant. Spectros. Rad. Trans.*, 108:2386–2390, 2008a.
- M. I. Mishchenko. Multiple scattering, radiative transfer, and weak localization in discrete random media: unified microphysical approach. *Rev. Geophys.*, 46:1–33, 2008b.
- M. I. Mishchenko. 125 years of radiative transfer: Enduring triumphs and persisting misconceptions. In *Radiation Processes in the Atmosphere and Ocean.*, volume Conf. Proc. 1531, pages 11–18. Amer. Inst. Physics, 2013. doi: 10.1063/1.4804696.
- M. I. Mishchenko. Directional radiometry and radiative transfer: The convoluted path from centuries-old phenomenology to physical optics. *J. Quant. Spectros. Rad. Trans.*, 146:4–33, 2014.
- M. I. Mishchenko and L. D. Travis. Gustav Mie and the evolving discipline of electromagnecit scattering by particles. *Bull. Amer. Meteorological Soc.*, pages 1853–1861, December 2008.
- M. I. Mishchenko, L. D. Travis, R. A. Kahn, and R. A. West. Modeling phase functions for dustlike tropospheric aerosols using a shape mixture of randomly oriented polydisperse spheroids. *J. Geophys. Res.*, 102(D14):16831–16847, 1997.
- M. I. Mishchenko, L. D. Travis, and A. A. Lacis. *Scattering, Absorption, and Emission of Light by Small Particles*. Cambridge Univ. Press, 2002. URL <https://pubs.giss.nasa.gov/abs/mi06300n.html>.
- M. I. Mishchenko, J. M. Dlugach, M. A. Yurkin, L. Bi, B. Cairns, L. Liu, R. L. Panetta, L. D. Travis, P. Yang, and N. T. Zakharova. First principles modeling of electromagnetic scattering by discrete and discretely heterogeneous random media. *Physics Reports*, 632: 1–75, 2016. doi: 10.1016/j.physrep.2016.04.002.
- E. A. Moberg and H. M. Sosik. Distance maps to estimate cell volume from two-dimensional plankton images. *Limnol. Oceanogr. Methods*, 10:278–288, 2012. doi: 10.4319/lom.2012.10.278.
- C. D. Mobley. A numerical model for the computation of radiance distributions in natural waters with wind-roughened surfaces. *Limnol. Oceanogr.*, 34(8):1473–1483, 1989.
- C. D. Mobley. *Light and Water: Radiative Transfer in Natural Waters*. Academic Press, 1994. Download as a pdf from the Ocean Optics Web Book.
- C. D. Mobley. Monte Carlo simulation of a point light source in an infinite medium. Technical report, SRI International, May 1996.

- C. D. Mobley. Estimation of the remote-sensing reflectance from above-surface measurements. *Applied Optics*, 38(36):7442–7455, 1999.
- C. D. Mobley. Fast light calculations for ocean ecosystem and inverse models. *Optics Express*, 19(20):18927–18944, 2011.
- C. D. Mobley. HydroPol Mathematical Documentation: Invariant Imbedding Theory for the Vector Radiative Transfer Equation. Technical report, Sequoia Scientific, Inc., Bellevue, WA, 2014.
- C. D. Mobley. Polarized reflectance and transmittance properties of wind-blown sea surfaces. *Applied Optics*, 54(15):4828–4849, 2015.
- C. D. Mobley. Modeling Sea Surfaces: A Tutorial on Fourier Transform Methods, Version 2. Technical report, Sequoia Scientific, Inc., Bellevue, WA, 2016.
- C. D. Mobley. Monte Carlo Ray Tracing for Coral Reef Light Simulations. Technical report, Sequoia Scientific, Inc., Bellevue, WA, 2018.
- C. D. Mobley and L. K. Sundman. Effects of optically shallow bottoms on upwelling radiances: Inhomogeneous and sloping bottoms. *Limnol. Oceanogr.*, 41(1, Part 2):329–336, 2003.
- C. D. Mobley, B. Gentili, H. R. Gordon, Z. Jin, G. W. Kattawar, A. Morel, P. Reinersman, K. Stamnes, and R. Stavn. Comparison of numerical models for the computation of underwater light fields. *Applied Optics*, 32(36):7484–7504, 1993.
- C. D. Mobley, L. K. Sundman, and E. Boss. Phase function effects on oceanic light fields. *Applied Optics*, 41(6):1035–1050, 2002.
- C. D. Mobley, H. Zhang, and K. J. Voss. Effects of optically shallow bottoms on upwelling radiances: Bidirectional reflectance distribution function effects. *Limnol. Oceanogr.*, 48(1, part 2):337–345, 2003.
- C. D. Mobley, D. Stramski, W. P. Bisset, and E. Boss. Optical modeling of ocean waters: Is the Case 1–Case-2 classification still useful? *Oceanography*, 17(2):60–67, 2004.
- C. D. Mobley, L. K. Sundman, C. O. Davis, J. H. Bowles, T. V. Downes, R. A. Leathers, M. J. Montes, W. P. Bissett, D. D. R. Kohler, R. P. Reid, E. M. Louchard, and A. Gleason. Interpretation of hyperspectral remote-sensing imagery by spectrum matching and look-up-tables. *Applied Optics*, 44(17):3576–3592, 2005.
- C. D. Mobley, F. Chai, P. Xiu, and L. K. Sundman. Impact of improved light calculations on predicted phytoplankton growth and heating in an idealized upwelling-downwelling channel geometry. *J. Geophys. Res. Oceans*, 120, 2015. doi: 10.1002/2014JC010588.
- C. D. Mobley, P. J. Werdell, B. Franz, Z. Ahmad, and S. Bailey. Atmospheric correction for satellite ocean color radiometry. Technical Report NASA/TM-2016-217551, NASA, June 2016.
- M. F. Modest. *Radiative Heat Transfer*. McGraw-Hill, 1993.

- M. J. Montes and B.-C. Gao. NRL atmospheric correction algorithms for oceans: TAFKAA users' guide. Technical report, U. S. Naval Res. Lab., 2004.
- M. J. Montes, B.-C. Gao, C. O. Davis, and M. Moline. Analysis of AVIRIS data from LEO-15 using TAFKAA atmospheric correction. AVIRIS Conference presentation, pdf online, 2003. URL <https://ntrs.nasa.gov/api/citations/20050192437/downloads/20050192437.pdf>.
- C. A. Moore, R. C. Honey, D. M. Hancock, S. Damron, R. Hilbers, and S. P. Tucker. Instrumentation for measuring in-situ sea truth for laser radar applications. In M. A. Blizard, editor, *Ocean Optics VII*, volume 489, pages 343–354. Proceedings of SPIE, 1984.
- T. S. Moore, J. W. Campbell, and H. Feng. A fuzzy logic classification scheme for selecting and blending satellite ocean color algorithms. *IEEE Trans. Geosci. Rem. Sens.*, 39(8):1764–1776, 2001.
- T. S. Moore, J. W. Campbell, and M. D. Dowell. A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. *Rem. Sens. Environ.*, 113(11):2424–2430, 2009. doi: 10.1016/j.rse.2009.07.016.
- T. S. Moore, M. D. Dowell, S. Bradt, and A. R. Verdu. An optical water type framework for selecting and blending retrievals from bio-optical algorithms in lakes and coastal waters. *Rem. Sens. Environ.*, 143:97–111, 2014. doi: 10.1016/j.rse.2013.11.021.
- A. Morel. Note au sujet des constantes de diffusion de la lumiere pour l'eau et l'eau de mer optiquement pures. *Cahiers Océanographiques*, XX(2):157–162, 1968.
- A. Morel. Optical modeling of the upper ocean in relation to its biogenous matter content (case 1 waters). *J. Geophys. Res.*, 93(C9):10749–10768, 1988.
- A. Morel. *Introduction to optical properties in the sea: theoretical aspects*, chapter Chapter 5, pages 109–151. UNESCO Publishing, 2008.
- A. Morel and A. Bricaud. Theoretical results concerning light-absorption in a discrete medium, and application to specific absorption of phytoplankton. *Deep-Sea Res., Part A*, 28(11):1375–1393, 1981a.
- A. Morel and A. Bricaud. Theoretical results concerning the optics of phytoplankton, with special reference to remote sensing applications. In *Oceanography from Space COSPAR/SCOR/IUCRM Symposium Plenum*, pages 383–327, 1981b.
- A. Morel and A. Bricaud. Inherent optical properties of algal cells including picoplankton theoretical and experimental results. *Canadian Bull. Fish. Aquat. Sci.*, 214:521–560, 1986.
- A. Morel and B. Gentili. Diffuse reflectance of oceanic waters: its dependance on Sun angle as influenced by molecular scattering contribution. *Applied Optics*, 30(30):4427–4438, 1991.
- A. Morel and B. Gentili. Diffuse reflectance of oceanic waters: II: Bidirectional aspects. *Applied Optics*, 32(33):6864–6879, 1993.

- A. Morel and B. Gentili. Diffuse reflectance of oceanic waters: III: Implication of bidirectionality for the remote-sensing problem. *Applied Optics*, 35(24):4850–4862, 1996.
- A. Morel and S. Maritorena. Bio-optical properties of oceanic waters: A reappraisal. *J. Geophys. Res.*, 106(C4):7163–7180, 2001.
- A. Morel and L. Prieur. Analysis of variations in ocean color. *Limnol. Oceanogr.*, 22(4):709–722, 1977.
- A. Morel and R. C. Smith. Relation between total quanta and total energy for aquatic photosynthesis. *Limnol. Oceanogr.*, 19(4):591, 1974.
- A. Morel and R. C. Smith. Terminology and units in optical oceanography. *Marine Geodesy*, 5(4):335–349, 1982.
- A. Morel, D. Antoine, and B. Gentili. Bidirectional reflectance of oceanic waters: accounting for Raman emission and varying particle scattering phase function. *Applied Optics*, 41(30):6289–6306, 2002.
- A. Morel, H. Claustre, D. Antoine, and B. Gentili. Natural variability of bio-optical properties in Case 1 waters: attenuation and reflectance within the visible and near-UV spectral domains, as observed in South Pacific and Mediterranean waters. *Biogeosciences*, 4:913–925, 2007.
- J. R. Morrison. In situ determination of the quantum yield of phytoplankton chlorophyll a fluorescence: A simple algorithm, observations, and a model. *Limnol. Oceanogr.*, 48(2):618–633, 2003.
- J. R. Morrison and N. B. Nelson. Seasonal cycle of phytoplankton UV absorption and the Bermuda Atlantic Time-series Study (BATS) site. *Limnol. Oceanogr.*, 49(1):215–224, 2004.
- P. J. Mulhearn. Distribution of microbubbles in coastal waters. *J. Geophys. Res.*, 86:6429–6434, 1981.
- Z. Musilova, F. Cortesi, M. Matschiner, W. I. L. Davies, J. S. Patel, S. M. Stieb, F. de Busserolles, M. Malmstrøm, O. K. Tørresen, C. J. Brown, J. K. Mountford, R. Hanel, D. L. Stenkamp, K. S. Jakobsen, K. L. Carleton, S. Jentoft, J. Marshall, and W. Salzburger. Vision using multiple distinct rod opsins in deep-sea fishes. *Science*, 364:588–592, 2019.
- National Research Council. *Assessing the Requirements for Sustained Ocean Color Research and Operations*. National Academies Press, 2011.
- B. Nechad, K. G. Ruddick, and Y. Park. Calibration and validation of a generic multi-sensory algorithm for mapping of total suspended matter in turbid waters. *Rem. Sens. Environ.*, 114:854–866, 2010.
- A. R. Neeley and A. Maninno, editors. *Volume 1: Inherent Optical Property Measurements and Protocols: Absorption Coefficient*, volume 1.0 of *IOCCG Protocol Series*. IOCCG, 2018. doi: 10.25607/OBP-119.

- N. B. Nelson and P. G. Coble. Optical analysis of chromophoric dissolved organic matter. In O. Wurl, editor, *Practical Guidelines for the Analysis of Seawater*, volume Chapter 5. CRC Press, 2009.
- N. B. Nelson and D. A. Siegel. Chromophoric DOM in the open ocean. In C. A. Carlson, editor, *Biogeochemistry of Marine Dissolved Organic Matter*, pages 547–578. Academic Press, 2002.
- G. Neumann and W. J. Pierson, Jr. *Principles of Physical Oceanography*. Prentice-Hall, 1966.
- R. A. Neville and J. F. R. Gower. Passive remote-sensing of phytoplankton via chlorophyll α fluorescence. *J. Geophys. Res. C: Oceans Atmos.*, 82(24):3487–3493, 1977.
- F. E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsburg, and T. Limperis. Geometrical considerations and nomenclature for reflectance. Technical Report Monograph 160, U. S. National Bureau of Standards, 1977.
- T. J. O’Hern. *Cavitation Inception Scale Effects I. Nuclei Distribution in Natural Waters II. Cavitation Inception in a Turbulent Shear Flow*. PhD thesis, California Institute of Technology, Pasadena, California, 1987.
- T. J. O’Hern., L. d’Agostino, and A. J. Acosta. Comparison of holographic and Coulter counter measurement of cavitation nuclei in the ocean. *J. Fluids Eng.*, 110:200–207, 1988.
- T. Oishi. Significant relationship between the backward scattering coefficient of sea water and the scatterance at 120 deg. *Applied Optics*, 29(31):4658–4665, 1990.
- R. J. Olson and H. M. Sosik. A submersible imaging-in-flow instrument to analyze nano and microplankton: Imaging FlowCytobot. *Limnol. Oceanogr. Methods*, pages 195–203, 2007.
- J. E. O’Reilly, S. Maritorena, B. G. Mitchell, D. A. Siegel, K. L. Carder, S. A. Garver, M. Kahru, and C. McClain. Ocean chlorophyll algorithms for SeaWiFS. *J. Geophys. Res.*, 103(C11):24937–24953, 1998.
- E. Organelli, G. Dall’Olmo, R. J. W. Brewin, G. A. Tarran, E. Boss, and A. Bricaud. The open-ocean missing backscattering is in the structural complexity of particles. *Nature Communications*, 2018. doi: 10.1038/s41467-018-07814-6.
- M. Ostrovska. Model of the dependence of the sun-induced chlorophyll a fluorescence quantum yield on the environmental factors in the sea. *Optics Express*, 20(21):23301–23317, 2012.
- J. M. Palmer. Getting intense on intensity. *Metrologia*, 30:371, 1993.
- J. M. Palmer. Radiometry and photometry: Units and conversions. In M. Bass, editor, *Handbook of Optics, Second Edition, Volume III*, chapter 7. McGraw-Hill, 2000. ISBN 0-07-135408-5.

- F. S. Patt, R. A. Barnes, Jr. R. E. Eplee, B. A. Franz, G. C. Feldman, S. W. Bailey, J. Gales, J. Werdell, M. Wang, R. Frouin, R. P. Stumpf, R. A. Arnone, Jr. R. W. Gould, P. M. Martinolich, V. Ransibrahmanakul, J. E. O'Reilly, and J. A. Yoder. Algorithm updates for the fourth SeaWiFS data reprocessing. Technical report, NASA, 2003. SeaWiFS Postlaunch Technical Report Series, S. B. Hooker and E. R. Firestone [Eds].
- T. J. Petzold. Volume scattering functions for selected ocean waters. Technical Report SIO Rept. 72-78., Scripts Inst. Oceanogr., 1972.
- T. J. Petzold. Volume scattering functions for selected ocean waters. In J. E. Tyler, editor, *Light in the Sea*, volume Chapter 12. Dowden, Hutchinson and Ross, 1977.
- A. D. Phelps and T. G. Leighton. Oceanic bubble population measurements using a buoy-deployed combination frequency technique. *IEEE J. Oceanic Eng.*, 23:400–410, 1998.
- A. D. Phelps, D. G. Ramble, and T. G. Leighton. The use of a combination frequency technique to measure the surf zone bubble population. *J. Acoust. Soc. Am*, 101:1981–1989, 1997.
- W. J. Pierson and L. Moskowitz. A proposed spectral form for fully developed wind seas based on the similarity theory of S. A. Kitaigorodskii. *J. Geophys. Res.*, 69:5181–5190, 1964.
- J. Piskozub, D. Stramski, E. Terrill, and W. K. Melville. Small-scale effects of underwater bubble clouds on ocean reflectance: 3-D modeling results. *Optics Express*, 17:11747–11752, 2009.
- J. Pitarch. A review of Secchi's contribution to marine optics and the foundation of Secchi disk science. *Oceanography*, 33(3), 2020. doi: 10.5670/oceanog.2020.301.
- T. Platt, N. Hoepfner, V. Stuart, and C. Brown. Why Ocean Colour? The Societal Benefits of Ocean-Colour Technology. Technical Report Report 7, IOCCG, 2008.
- R. M. Pope and E. S. Fry. Absorption spectrum (380-700 nm) of pure water. II. Integrating cavity measurement. *Applied Optics*, 36(33):8710–8723, 1997.
- D. K. Prasad and K. Agarwal. Classification of hyperspectral or trichromatic measurements of ocean color data into spectral classes. *Sensors*, 16(3):413, 2016. doi: 10.3390/s16030413.
- R. W. Preisendorfer. Theoretical proof of the existence of characteristic diffuse light in natural waters. *Sears Foundation J. Marine Res.*, 18(8):1–9, 1959.
- R. W. Preisendorfer. *Radiative Transfer Theory on Discrete Spaces*. Pergamon, 1965.
- R. W. Preisendorfer. *Hydrologic Optics*. NOAA Pacific Mar. Environ. Lab., Seattle, WA., 1976. In 6 volumes: Vol. 1: Introduction, 218 pp; Vol. 2: Foundations, 400 pp; Vol. 3: Solutions, 246 pp; Vol. 4: Imbeddings, 207 pp; Vol. 5: Properties, 296 pp; Vol. 6: Surfaces. Can be downloaded from the Ocean Optics Web Book.
- R. W. Preisendorfer. Secchi disk science: Visual optics of natural waters. *Limnol. Oceanogr.*, 31(5):909–926, 1986.

- R. W. Preisendorfer. *Principal Component Analysis in Meteorology and Oceanography*. Elsevier, 1988.
- R. W. Preisendorfer and C. D. Mobley. Unpolarized Irradiance Reflectances and Glitter Patterns of Random Capillary Waves on Lakes and Seas, by Monte Carlo Simulation. Technical Report Tech. Memo. ERL-PMEL-63; published as Preisendorfer and Mobley (1986), NOAA Pacific Mar. Environ. Lab., Seattle, WA, 1985. NTIS PB86-123577.
- R. W. Preisendorfer and C. D. Mobley. Albedos and glitter patterns of a wind-roughened sea surface. *J. Phys. Oceanogr.*, 16(7):1293–1316, 1986. Reply in *J. Phys. Oceanogr.* 17(4), 551.
- W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in Fortran: The Art of Scientific Computing, Second Edition*. Cambridge Univ. Press, 1992.
- B. B. Prezelin and R. S. Alberte. Photosynthetic characteristics and organization of chlorophyll in marine dinoflagellates. *Proc. Natl. Acad. Sci. U. S. A.*, 75(4):1801–1804, 1978.
- L. Prieur and A. Morel. Etude theorique du regime asymptotique: relations entre caracteristiques optiques et coefficient d'extinction relatif a la penetration de la lumiere du jour. *Cah. Oceanogr.*, 23:35–47, 1971.
- L. Prieur and S. Sathyendranath. An optical classification of coastal and oceanic waters based on the specific spectral absorption of phytoplankton pigments, dissolved organic matter, and other particulate materials. *Limnol. Oceanogr.*, 26(4):671–689, 1981.
- J. G. Proakis and D. G. Manolakis. *Digital Signal Processing: Principles, Algorithms, and Applications, Third Edition*. Prentice Hall, 1996.
- E. M. Purcell and C. R. Pennypacker. Scattering and absorption of light by nonspherical dielectric grains. *Astrophys. Journal*, 186:705–714, 1973.
- X. Quan and E. S. Fry. Empirical equation for the index of refraction of seawater. *Applied Optics*, 34:3477–3480, 1995.
- A. Quirantes and S. Bernard. Light scattering by marine algae: Two-layer spherical and nonspherical models. *J. Quant. Spectros. Rad. Trans.*, 89:311–321, 2004.
- A. Quirantes and S. Bernard. Light-scattering methods for modelling algal particles as a collection of coated and/or nonspherical particles. *J. Quant. Spectros. Rad. Trans.*, 100: 315–324, 2006.
- E. D. Rainville. *Elementary Differential Equations, Third Edition*. Macmillan, 1964.
- Lord Rayleigh. On the transmission of light through an atmosphere containing small particles in suspension, and on the origin of the blue of the sky. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 47(287):375–384, 1899. doi: 10.1080/14786449908621276.
- L. O. Reynolds and N. J. McCormick. Approximate two-parameter phase function for light scattering. *J. Optical Soc. Amer.*, 70(10):1206–1212, 1980.

- R. A. Reynolds, D. Stramski, V. M. Wright, and S. B. Wozniak. Measurements and characterization of particle size distributions in coastal waters. *J. Geophys. Res.*, 115 (C08024), 2010. doi: 10.1029/2009JC005930.
- R. A. Reynolds, D. Stramski, and G. Neukermans. Optical backscattering by particles in arctic seawater and relationships to particle mass concentration, size distribution, and bulk composition. *Limnol. Oceanogr.*, 61:1869–1890, 2016. doi: 10.1002/lno.10341.
- J. A. Richards and X. Jia. *Remote Sensing Digital Image Analysis: An Introduction. 4th Edition*. Springer, 1996.
- C. S. Roesler and M. J. Perry. In situ phytoplankton absorption, fluorescence emission, and particulate backscattering spectra determined from reflectance. *J. Geophys. Res.*, 100(C7):13279–13294, 1995.
- C. S. Roesler, M. J. Perry, and K. L. Carder. Modeling in situ phytoplankton absorption from total absorption spectra in productive inland marine waters. *Limnol. Oceanogr.*, 34(8):1510–1523, 1989.
- R. Roettgers, D. McKee, and C. Utschig. Temperature and salinity correction coefficients for light absorption by water in the visible to infrared spectral region. *Optics Express*, 22(21):25093–25108, 2014.
- H. Runyan, R. A. Reynolds, and D. Stramski. Evaluation of particle size distribution metrics to estimate the relative contributions of different size fractions based on measurements in Arctic waters. *J. Geophys. Res. Oceans*, 125, 2020. doi: 10.1029/2020JC016218.
- B. J. Russell, H. M. Dierssen, and E. J. Hochberg. Water column optical properties of Pacific coral reefs across geomorphic zones and in comparison to offshore waters. *Remote Sensing*, 11:1757–1782, 2019. doi: 10.3390/rs11151757.
- O. Sachs. *The Island of the Colorblind*. Knopf, 1996.
- J. J. Sakurai. *Advanced Quantum Mechanics*. Addison-Wesley, 1967.
- B. T. San and M. L. Suzen. Evaluation of different atmospheric correction algorithms for EO-1 Hyperion imagery. *Internat. Archives Photogrammetry, Rem. Sens. Spatial Info. Sci.*, XXXVIII (Part 8, 2010. Kyoto, Japan.
- R. Sanchez and N. McCormick. Analytic beam spread function for ocean optics applications. *Applied Optics*, 41(30):6276–6288, 2002.
- S. Santabarbara, W. Remelli, A. A. Petrova, and A. P. Casazza. Influence of the wavelength of excitation and fluorescence emission detection on the estimation of fluorescence-based physiological parameters in different classes of photosynthetic organisms. In N. Grigoryeva, editor, *Fluorescence Methods for Investigation of Living Cells and Microorganisms*. IntechOpen, 2020. doi: 10.5772/intechopen.93230.
- S. Sathyendranath and T. Platt. The spectral irradiance field at the surface and in the interior of the ocean: a model for applications in oceanography and remote sensing. *J. Geophys. Res.*, 93(C8):9270–9280, 1988.

- S. Sathyendranath and T. Platt. Angular distribution of the submarine light field: modification by multiple scattering. *Proc. Royal Soc. London A*, 433:287–297, 1991.
- P. M. Saunders. Practical conversion of pressure to depth. *J. Phys. Oceanog.*, 11(4): 573–574, 1981. doi: 10.1175/1520-0485(1981)011<0573:PCOPTD>2.0.CO;2.
- R. Sauzède, H. Claustre, J. Uitz, C. Jamet, G. Dall’Olmo, F. D’Ortenzio, B. Gentili, A. Poteau, and C. Schmechtig. A neural network-based method for merging ocean color and Argo data to extend surface bio-optical properties to depth: Retrieval of the particulate backscattering coefficient. *J. Geophys. Res. Oceans*, 121:2552–2571, 2016. doi: 10.1002/2015JC011408.
- A. Schuster. On lunar and solar periodicities of earthquakes. *Proc. Royal Soc. London*, 61.:369–377, 455–465, 1897.
- A. Schuster. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *J. Geophys. Res.*, 3(1):13–41, 1898.
- J. N. Schwarz, P. Kowalczuk, S. Kaczmarek, G. F. Cota, B. G. Mitchell, M. Kahru, F. P. Chavez, A. Cunningham, D. McKee, P. Gege, M. Kishino, D. A. Phinney, and R. Raine. Two models for absorption by coloured dissolved organic matter (CDOM). *Oceanologia*, 44(2):209–241, 2002.
- D. J. Segelstein. The Complex Refractive Index of Water. Master’s thesis, University of Missouri, Kansas City, 1981.
- P. Service. The Wright-Guild experiments and the development of the CIE 1931 RGB and XYZ color spaces. Online pdf, 2016. URL https://philservice.typepad.com/Wright-Guild_and_CIE_RGB_and_XYZ.pages.pdf.
- R. W. Sheldon, A. Prakash, and W. H. Sutcliff. Size distribution of particles in ocean. *Limnol. Oceanogr.*, 17:327–340, 1972.
- Y. T. Shen, S. Gowing, and B. Eckstein. Cavitation Susceptibility Measurements of Ocean, Lake and Laboratory Waters. Technical report, David Taylor Naval Ship Research and Development Center, 1986.
- E. P. Shettle and R. W. Fenn. Models for the aerosols of the lower atmosphere and the effects of humidity variation on their optical properties. Technical report, Air Force Geophys. Res. Lab., Hanscom Air Force Base, MA, 1979. Environ. Res. Paper 676, AFGL-TR-79-0214.
- K. S. Shifrin. *Physical Optics of Ocean Water*. American Institute of Physics, 1988.
- W. H. Slade, Y. C. Agrawal, and O. A. Mikkelsen. Comparison of measured and theoretical scattering and polarization properties of narrow size range irregular sediment particles. pages 1–6. IEEE Oceans–San Diego, IEEE, 2013.
- R. C. Smith and K. Baker. The bio-optical state of ocean waters and remote sensing. *Limnol. Oceanogr.*, 23(2), 1978.

- B. H. Soffer and D. K. Lynch. Some paradoxes, errors, and resolutions concerning the spectral optimization of human vision. *Am. J. Phys.*, 67(11):946–953, 1999.
- M. G. Solonenko and C. D. Mobley. Inherent optical properties of Jerlov water types. *Applied Optics*, 54(17):5392–5401, 2015.
- A. Sommerfeld. *Electrodynamics: Lectures on Theoretical Physics, Vol III*. Academic Press, 1952. Translated by E. G. Ramberg.
- H. M. Sosik. Characterizing seawater constituents from optical properties. In M. Babin, C. S. Roesler, and J. J. Cullen, editors, *Real-time coastal observing systems for ecosystem dynamics and harmful algal blooms*. UNESCO, 2008.
- R. H. Stavin and A. D. Weidemann. Optical modeling of clear ocean light fields: Raman scattering effects. *Applied Optics*, 27:4002–4010, 1988.
- C. A. Stedmon, C. L. Osburn, and T. Kragh. Tracing water mass mixing in the Baltic – North Sea transition zone using the optical properties of coloured dissolved organic matter. *Estuarine, Coastal and Shelf Sci.*, 87:156–162, 2010.
- M. Stokes, M. Anderson, S. Chandrasekar, and R. Motta. A standard default color space for the internet - sRGB. online, 1996. URL <https://www.w3.org/Graphics/Color/sRGB>.
- B. Stout and N. Bonod. Gustav Mie: the man, the theory. *Photoniques*, 101:22–26, 2020. doi: 10.1051/photon/202010122.
- M. Stramska and T. Petelski. Observations of oceanic whitecaps in the north polar waters of the Atlantic. *J. Geophys. Res.*, 108(C3), 2003. doi: 10.1029/2002JC001321.
- D. Stramski and J. Tegowski. Effects of intermittent entrainment of air bubbles by breaking wind waves on ocean reflectance and underwater light field. *J. Geophys. Res.*, 106(C12): 31345–31360, 2001.
- D. Stramski and S. B. Wozniak. On the role of colloidal particles in light scattering in the ocean. *Limnol. Oceanogr.* 50, 50:1581–1591, 2005.
- D. Stramski, A. Morel, and A. Bricaud. Modeling the light attenuation and scattering by spherical phytoplanktonic cells: a retrieval of the bulk refractive index. *Applied Optics*, 27(19):3954–3958, 1986.
- D. Stramski, R. A. Reynolds, M. Kahru, and B. G. Mitchell. Estimation of particulate organic carbon in the ocean from satellite remote sensing. *Science*, 285(5425):239–242, 1999. doi: 10.1126/science.285.5425.239.
- D. Stramski, A. Bricaud, and A. Morel. Modeling the inherent optical properties of the ocean based on the detailed composition of the planktonic community. *Applied Optics*, 40(18):2929–2945, 2001.
- D. Stramski, E. Boss, D. Bogucki, and K. J. Voss. The role of seawater constituents in light backscattering in the ocean. *Prog. Oceanogr.*, 61:27–56, 2004a.

- D. Stramski, S. Wozinak, and P. J. Flatau. Optical properties of Asian mineral dust suspended in seawater. *Limnol. Oceanogr.*, 49(3):749–755, 2004b.
- D. Stramski, M. Babin, and S. Wozinak. Variations in the optical properties of terrigenous mineral-rich particulate matter suspended in seawater. *Limnol. Oceanogr.*, 52(6):2418–2433, 2007.
- D. Stramski, R. A. Reynolds, S. Kaczmarek, J. Uitz, and G. Zheng. Correction of path-length amplification in the filter-pad technique for measurements of particulate absorption coefficient in the visible spectral region. *Applied Optics*, 54:6763–6782, 2015. doi: 10.1364/AO.54.006763.
- J. W. Strutt (Lord Rayleigh). On the light from the sky, its polarization and colour. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 41(271):107–120, 1871. doi: 10.1080/14786447108640452.
- R. Stumpf and J. R. Pennock. Calibration of a general optical equation for remote sensing of suspended sediments in a moderately turbid estuary. *J. Geophys. Res.*, 94:14363–14371, 1989.
- M. Y. Su and J. Cartmil. Low-frequency underwater sound speed variations due to oceanic bubbles. In M. J. Buckingham and J. R. Potter, editors, *Sea Surface Sound*, pages 351–365. World Scientific, 1994.
- M. Y. Su, S. C. Ling, and J. Cartmill. Optical microbubble measurements in the North Sea. In B. R. Kerman, editor, *Sea Surface Sound*, pages 211–223. Kluwer Academic Publishers, 1988.
- A. Subramaniam, C. W. Brown, R. R. Hood, E. J. Carpenter, and D. G. Capone. Detecting Trichodesmium blooms in SeaWiFS imagery. *Deep-Sea Res., Part II*, 49(1-3):107–121, 2002.
- S. Sugihara, M. Kishino, and N. Okami. Contribution of Raman scattering to upward irradiance in the sea. *J. Oceanogr. Soc. Japan*, 40:397–404, 1984.
- J. M. Sullivan and M. S. Twardowski. Angular shape of the oceanic particulate volume scattering function in the backward direction. *Applied Optics*, 48(35):6811–6819, 2009.
- J. M. Sullivan, M. S. Twardowski, J. R. V. Zeneveld, and C. C. Moore. *Measuring Optical Backscattering in Water*, chapter Light Scattering Reviews 7: Radiative Transfer and Optical Properties of Atmosphere and Underlying Surface, pages 189–224. Springer Praxis Books, 2013. doi: 10.1007/978-3-642-21907-8_6.
- B. Sun, G. W. Kattawar, and P. Yang. Radiance and polarization in the diffusion region with an arbitrary scattering phase matrix. *J. Quant. Spectros. Rad. Trans.*, 183:154–161, 2016. doi: 10.1016/j.jqsrt.2016.06.021.
- D. Sun, Y. Li, Q. Wang, H. Lyu, C. Le, C. Huang, and S. Gong. Partitioning particulate scattering and absorption into contributions of phytoplankton and non-algal particles in winter in Lake Taihu (China). *Hydrobiologia*, 644:337–349, 2010. doi: 10.1007/s10750-010-0198-7.

- R. Szeliski. Computer Vision: Algorithms and Applications, Second Edition (draft). online, 2021. URL <http://szeliski.org/Book/>.
- Y. Takano and K.-N. Liou. Solar radiative transfer in cirrus clouds. Part I: Single-scattering and optical properties of hexagonal ice crystals. *J. Atmos. Sci.*, 46(1):3–19, 1989.
- C. Tamburini et al. Deep-sea bioluminescence blooms after dense water formation at the ocean surface. *PLoS ONE*, 8(7), 2013. doi: 10.1371/journal.pone.0067523. The paper lists 166 authors.
- H. Tan, R. Doerffer, T. Oishi, and A. Tanaka. A new approach to measure the volume scattering function. *Optics Express*, 21(16):18697–18711, 2013. doi: 10.1364/OE.21.018697.
- S. Tan, J. Zhu, L. Tsang, and S. V. Nghiem. Numerical simulation of Maxwell’s equation in 3D (NMM3D) applied to active and passive remote sensing of terrestrial snow and snow on sea ice. 2016. doi: 10.1109/PIERS.2016.7735538.
- E. J. Terrill and W. K. Melville. A broadband acoustic technique for measuring bubble size distributions: laboratory and shallow water measurements. *J. Atmos. Oceanic Technol.*, 17:220–239, 2000.
- E. J. Terrill, W. K. Melville, and D. Stramski. Bubble entrainment by breaking waves and their influence on optical scattering in the upper ocean. *J. Geophys. Res.*, 106(16): 815–816, 823, 2001.
- J. Tessendorf. Simulating Ocean Water. Technical report, 2004. URL <http://jtessen.people.clemson.edu/reports.html>. SIGGRAPH Course Notes.
- S. A. Thorpe. On the clouds of bubbles formed by breaking wind waves in deep water, and their role in air-sea transfer. *Philos. Trans. R. Soc. London Series A*, 304:155–210, 1982.
- S. A. Thorpe. Measurements with an automatically recording inverted echo sounder; ARIES and the bubble clouds. *J. Phys. Oceanogr.* 16, 16:1462–1478, 1986.
- S. A. Thorpe and A. J. Hall. The characteristics of breaking waves, bubble clouds, and near-surface currents observed using side-scan sonar. *Cont. Shelf Res.*, 1:353–384, 1983.
- S. A. Thorpe and P. N. Humphries. Bubbles and breaking waves. *Nature*, 283:463–465, 1980.
- U. A. Timofeeva and F. I. Gorobetz. The relationship between the coefficient of extinction of collimated and diffuse fluxes of light. *Izv. Akad. Nauk, USSR, Series Geofizik*, 3: 291–296, 1967.
- M. C. Tomlinson, T. T. Wynne, and R. P. Stumpf. An evaluation of remote sensing techniques for enhanced detection of the toxic dinoflagellate, *Karenia brevis*. *Rem. Sens. Environ.*, 113(3):598–609, 2009.

- A. Tonizzo, M. Twardowski, S. McLean, K. Voss, M. Lewis, and C. Trees. Closure and uncertainty assessment for ocean color reflectance using measured volume scattering functions and reflective tube absorption coefficients with novel correction for scattering. *Applied Optics*, 56(1):130–146, 2017. doi: 10.1364/AO.56.000130.
- A. Tonomura, T. Endo, and T. Kawasaki. Demonstration of single-electron buildup of an interference pattern. *Amer. J. Phys.*, 47(2):117–120, 1989.
- D. A. Toole, D. A. Siegel, D. W. Menzies, J. J. Neumann, and R. C. Smith. Remote-sensing reflectance determinations in the coastal ocean environment: impact of instrumental characteristics and environmental variability. *Applied Optics*, 39(3):456–469, 2000.
- D. H. Towne. *Wave Phenomena*. Addison-Wesley, 1967.
- M. S. Twardowski, E. Boss, J. B. Macdonald, W. S. Pegau, A. H. Barnard, and J. R. V. Zaneveld. A model for estimating bulk refractive index from the optical backscattering ratio and the implications for understanding particle composition in case I and case II waters. *J. Geophys. Res., Oceans*, 106(C7):14129–14142, 2001.
- M. S. Twardowski, E. Boss, J. M. Sullivan, and P. L. Donaghay. Modeling the spectral shape of absorption by chromophoric dissolved organic matter (CDOM). *Mar. Chem.*, 89:69–88, 2004.
- M. S. Twardowski, X. Zhang, S. Vagle, J. Sullivan, S. Freeman, H. Czerski, Y. You, L. Bi, and G. Kattawar. The optical volume scattering function in a surf zone inverted to derive sediment and bubble particle subpopulations. *J. Geophys. Res.*, 117(C00H17), 2012. doi: 10.1029/2011JC007347.
- J. E. Tyler. Radiance distribution as a function of depth in an underwater environment. *Bull. Scripps Inst. Ocean.* 7(5), 7(5):363–412, 1960. Download pdf from the Web Book.
- J. E. Tyler and R. C. Smith. *Measurements of Spectral Irradiance Underwater*. Gordon and Breach, New York, 1970.
- M. Tzortziou, J. R. Herman, C. L. Gallegos, P. J. Neale, A. Subramaniam, L. W. Harding Jr., and Z. Ahmad. Bio-optics of the Chesapeake Bay from measurements and radiative transfer closure. *Estuarine, Coastal and Shelf Sci.*, 68:348–362, 2006.
- J. Uitz, H. Claustre, A. Moreal, and S. B. Hooker. Vertical distribution of phytoplankton communities in open ocean: An assessment based on surface chlorophyll. *J. Geophys. Res.*, 111(C08005), 2006. doi: 10.1029/2005JC003207.
- O. Ulloa, S. Sathyendranath, and T. Platt. Effect of the particle-size distribution on the backscattering ratio in seawater. *Applied Optics*, 33(30):7070–7077, 1994.
- M. v. Smoluchowski. Molekular-kinetische Theorie der Opaleszenz von Gasen im kritischen Zustande, sowie einiger verwandter Erscheinungen. *Ann. Physik*, 25:205–226, 1908.
- S. Vagle and D. M. Farmer. The measurement of bubble-size distributions by acoustical backscatter. *J. Atmos. Oceanic Technol.*, 9:630–644, 1992.

- S. Vagle and D. M. Farmer. A comparison of four methods for bubble size and void fraction measurements. *IEEE J. Oceanic Eng.*, 23:211–222, 1998.
- R. D. Vaillancourt, C. W. Brown, R. R. L. Guillard, and W. M. Balch. Light backscattering properties of marine phytoplankton: relationships to cell size, chemical composition and taxonomy. *J. Plankton Res.*, 26(2):191–212, 2004.
- H. C. van de Hulst. *Light Scattering by Small Particles*. John Wiley and Sons, New York, 1957.
- H. C. van de Hulst. *Multiple Light Scattering: Tables, Formulas, and Applications*. Academic Press, 1980. Vol. 1, pages 1-299; Vol. 2, 300-739.
- H. C. van de Hulst. *Light Scattering by Small Particles*. Dover, 1981. Reprint of the 1957 edition.
- R. A. Vandermeulen, A. Mannino, S. E. Craig, and P. J. Werdell. 150 shades of green: Using the full spectrum of remote sensing reflectance to elucidate color shifts in the ocean. *Rem. Sens. Environ.*, 247., 2020. doi: 10.1016/j.rse.2020.111900.
- A. P. Vasilkov, J. R. Herman, Z. Ahmad, M. Kahru, and B. G. Mitchell. Assessment of the ultraviolet radiation field in ocean waters from space-based measurements and full radiative-transfer calculations. *Applied Optics*, 44(14):2863–2869, 2005.
- A. Vodacek, M. D. DeGrandpre, E. T. Peltzer, R. K. Nelson, and N. V. Blough. Seasonal variation of CDOM and DOC in the Middle Atlantic Bight: Terrestrial inputs and photooxidation. *Limnol. Oceanogr.*, 42:674–686, 1997.
- M. Vollmer. Physics of the microwave oven. *Physics Education*, 39(1):74–81, 2004.
- H. Volten, J. F. de Haan, J. W. Hoovenier, R. Schreurs, and W. Vassen. Laboratory measurements of angular distributions of light scattered by phytoplankton and silt. *Limnol. Oceanogr.* 43, 43:1180–1197, 1998.
- K. J. Voss. Use of the radiance distribution to measure the optical absorption coefficient in the ocean. *Limnol. Oceanogr.*, 34(8):1614–1622, 1989.
- K. J. Voss. Simple empirical model for the oceanic point spread function. *Applied Optics*, 30(18):2647–2651, 1991.
- K. J. Voss. A spectral model of the beam attenuation coefficient in the ocean and coastal waters. *Limnol. Oceanogr.*, 37(3):501–509, 1992.
- K. J. Voss and A. L. Chapin. Measurement of the point spread function in the ocean. *Applied Optics*, 29(25):3638–3642, 1990.
- K. J. Voss and A. L. Chapin. Upwelling radiance distribution camera system, NURADS. *Optics Express*, 13(11):4250–4262, 2005. doi: 10.1364/OPEX.13.004250.
- K. J. Voss and E. S. Fry. Measurement of the Mueller matrix for ocean water. *Applied Optics*, 23(23):4427–4439, 1984.

- K. J. Voss, A. Morel, and D. Antoine. Detailed validation of the bidirectional effect in various Case 1 waters for application to ocean color imagery. *Biogeosciences*, 4:781–789, 2007.
- J. R. Waaland and D. Branton. Gas vacuole development in a blue-green alga. *Science*, 163:1339–1341, 1969.
- R. E. Walker. *Marine Light Field Statistics*. Wiley-Interscience, 1994.
- G. E. Walrafen. Raman spectral studies of the effects of temperature on water structure. *J. Chem. Phys.*, 47(1):118–121, 1967.
- A. L. Walsh and P. J. Mulhearn. Photographic measurements of bubble populations from breaking wind waves at sea. *J. Geophys. Res.* 92, 92:14553–14565, 1987.
- G. Wang and J. Moisan. Remote sensing of phytoplankton pigments. *Intechopen*, 2021. doi: 10.5772/intechopen.95381.
- M. Wang. The Rayleigh lookup tables for the SeaWiFS data processing: accounting for the effects of ocean surface roughness. *Internat. J. Remote Sensing*, 23(13):2693–2702, 2002.
- M. Wang. A refinement for the Rayleigh radiance computation with variation of the atmospheric pressure. *Internat. J. Remote Sensing*, 26(24):5651–5653, 2005.
- M. Wang and S. W. Bailey. Correction of sun glint contamination of the SeaWiFS ocean and atmosphere products. *Applied Optics*, 40(27):4790–4798, 2001.
- J. Wei, Z.-P. Lee, and S. Shang. A system to measure the data quality of spectral remote-sensing reflectance of aquatic environments. *J. Geophys. Res. Oceans*, 121(11):8189–8207, 2016.
- M. R. Wernand. *Poisedon's Paintbox: Historical Archives of Ocean Color in Global-Change Perspective*. PhD thesis, Univ. of Utrecht, 2011.
- G. C. Whipple. *The Microscopy of Drinking Water*. Wylie, 1899.
- A. L. Whitmire, E. Boss, T. J. Cowles, and W. S. Pegau. Spectral variability of the particulate backscattering ratio. *Optics Express*, 15(11):7019–7031, 2007.
- E. A. Widder. Bioluminescence and the pelagic visual environment. *Marine and Freshwater Behaviour and Physiology*, 35(1):1–26, 2002. doi: 10.1080/10236240290025581.
- E. A. Widder. A look back at quantifying oceanic bioluminescence: Seeing the light, flashes of insight and other bad puns. *Marine Tech. Soc. Journal*, 40(2):136–137, 2006. doi: 10.4031/002533206787353321.
- E. A. Widder. Bioluminescence in the ocean: Origins of biological, chemical, and ecological diversity. *Science*, 328(5979):704–708, 2010. doi: 10.1126/science.1174269.
- E. A. Widder, M. I. Latz, and J. F. Case. Marine bioluminescence spectra measured with an optical multichannel detection system. *Biological Bulletin*, 165:791–810, 1983.

- K. Witkowski, T. Krol, A. Zielinski, and E. Kuten. A light-scattering matrix for unicellular marine phytoplankton. *Limnol. Oceanogr.*, 43:859–869, 1998.
- L.-A. Wu, G. L. Long, Q. Gong, and G.-C. Guo. Optics in ancient China. *AAPPS Bulletin*, 25(4), October 2015. doi: 10.22661/AAPPSBL.2015.25.4.06. URL <http://www.epsnews.eu/2015/10/optics-in-ancient-china/>. News of the Association of Asia Pacific Physical Societies.
- P. J. Wyatt and C. Jackson. Discrimination of phytoplankton via light-scattering properties. *Limnol. Oceanogr.*, 34:96–112, 1989.
- Z. Xu and D. K. P. Yue. Analytical solution of beam spread function for ocean light radiative transfer. *Optics Express*, 23(14):17966–17978, 2015.
- Y. Z. Yacobi, J. J. Alberts, M. Takacs, and M. McElvaine. Absorption spectroscopy of colored dissolved organic carbon in Georgia (USA) rivers: the impact of molecular size distribution. *J. Limnol.*, 62:41–46, 2003.
- B. Yan, B. Chen, and K. Stamnes. Role of oceanic air bubbles in atmospheric correction of ocean color imagery. *Applied Optics*, 41(12):2202–2212, 2002. doi: 10.1364/AO.41.002202.
- H. Yang and H. R. Gordon. Remote sensing of ocean color: Assessment of the water-leaving radiance bidirectional effects on the atmospheric diuse transmittance. *Applied Optics*, 36(30):7887–7897, 1997.
- H. Ye, J. Li, T. Li, Q. Shen, J. Zhu, X. Wang, F. Zhang, J. Zhang, and B. Zhang. Spectral classification of the Yellow Sea and implications for coastal ocean color remote sensing. *Remote Sensing*, 8(4):321, 2016. doi: 10.3390/rs8040321.
- C. S. Yentsch. Measurement of visible light absorption by particulate matter in the ocean. *Limnol. Oceanogr.*, 7(62):207–216, 1962. doi: 10.4319/lo.1962.7.2.0207.
- C. S. Yentsch. Why is the measurement of fluorescence important to the study of biological oceanography? In R. W. Spinrad, K. L. Carder, and M. J. Perry, editors, *Ocean Optics*, pages 148–164. Oxford Univ. Press, 1994.
- Y. You, A. Tonizzo, A. A. Gilerson, M. E. Cummings, P. Brady, J. M. Sullivan, M. S. Twardowski, H. M. Diersen, S. A. Ahmad, and G. W. Kattawar. Measurements and simulations of polarization states of underwater light in clear oceanic water. *Applied Optics*, 50(24):4873–4893, 2011.
- D. E. Yount. Skins of varying permeability: A stabilization mechanism for gas cavitation nuclei. *J. Acoust. Soc. Am.*, 65:1429–1439, 1979.
- A. Zajonc. Light reconsidered. *Optics and Photonics News*, pages S2–S5, 2003.
- J. R. V. Zaneveld and W. S. Pegau. Robust underwater visibility parameter. *Optics Express*, 11(23), 2003.

- J. R. V. Zaneveld, J. C. Kitchen, and C. C. Moore. The scattering error correction of reflecting-tube absorption meters. In J. S. Jaffe, editor, *Ocean Optics XII*, volume 2258, pages 44–55. SPIE Proc., 1994.
- P.-W. Zhai, G. W. Kattawar, and Y. Hu. Comment on the transmission matrix for a dielectric interface. *J. Quant. Spectros. Rad. Trans.*, 113:1981–1984, 2012.
- H. Zhang, K. J. Voss, R. P. Reid, and E. M. Louchard. Bidirectional reflectance measurements of sediments in the vicinity of Lee Stocking Island, Bahamas. *Limnol. Oceanogr.*, 41(1, (part 2)):380–389, 2003.
- X. Zhang and L. Hu. Estimating scattering of pure water from density fluctuation of the refractive index. *Optics Express*, 17:1671–1678, 2009.
- X. Zhang, M. R. Lewis, and B. D. Johnson. Influence of bubbles on scattering of light in the ocean. *Applied Optics*, 37:6525–6536, 1998.
- X. Zhang, M. R. Lewis, M. Lee, B. D. Johnson, and G. Korotaev. Volume scattering function of natural bubble populations. *Limnol. Oceanogr.*, 47:1273–1282, 2002.
- X. Zhang, M. Lewis, W. P. Bissett, B. Johnson, and D. Kohler. Optical influence of ship wakes. *Applied Optics*, 43:3122–3132, 2004.
- X. Zhang, L. Hu, and M.-X. He. Scattering by pure seawater: Effect of salinity. *Optics Express*, 17(7):5698–5710, 2009.
- X. Zhang, G. R. Fournier, and D. J. Gray. Interpretation of scattering by oceanic particles around 120 degrees and its implication in ocean color studies. *Optics Express*, 25(4):A191–A199, 2017. doi: 10.1364/OE.25.00A191.
- X. Zhang, D. Stramski, R. A. Reynolds, and E. R. Blocker. Light scattering by pure water and seawater: the depolarization ratio and its variation with salinity. *Applied Optics*, 58(4):991–1004, 2019.
- X. Zhang, L. Hu, Y. Xiong, Y. Huot, and D. Gray. Experimental estimates of optical backscattering associated with submicron particles in clear oceanic waters. *Geophys. Res. Letters*, 47, 2020. doi: 10.1029/2020GL087100.
- X. Zhang, L. Hu, D. Gray, and Y. Xiong. Shape of particle backscattering in the North Pacific Ocean: the χ factor. *Applied Optics*, 60(5):1260–1266, 2021. doi: 10.1364/AO.414695.
- A. Zisserman. Lecture 3: Image restoration. On line, 2014. URL <http://www.robots.ox.ac.uk/~az/lectures/ia/lect3.pdf>.
- V. M. Zoloratev and A. V. Demin. Optical constants of water over a broad range of wavelengths, 0.1 \AA – 1 m. *Spectroscopy (USSR)*, 43(2):157–161, 1977.

Index

- Absorbance, [110](#), [164](#), [166](#)
- Absorptance, [98](#)
- Absorption
 - by CDOM, [152](#), [266](#), [288](#)
 - by minerals, [271](#), [300](#)
 - by non-algal particles, [151](#), [269](#)
 - by phytoplankton, [151](#), [262](#)
 - by water, [256](#)
 - importance of, [149](#)
 - measurement of, [161](#)
 - physics of, [153](#)
 - pigment packaging effect on, [260](#)
 - quantum mechanics terminology, [154](#)
 - relation to particle size and density, [303](#)
- Absorption Coefficient
 - defined, [99](#)
 - derived from Maxwell's equations, [419](#)
 - measurement, [109](#)
- Aggregates, [280](#)
- Anomalous Dispersion, [428](#)
- Apparent Optical Properties
 - average cosines, [138](#)
 - defined, [117](#)
 - diffuse attenuation functions, [120](#)
 - reflectances, [132](#)
- Asymptotic Radiance Distribution, [63](#), [370](#)
- Atmospheric Correction, [529](#)
 - aerosols, [572](#)
 - atmospheric transmittances, [552](#)
 - BRDF effect, [545](#)
 - empirical line fit, [590](#)
 - non-absorbing gases, [562](#)
 - normalized reflectance, [541](#)
 - out-of-band response, [581](#)
 - overview of the problem, [530](#)
 - polarization effects, [587](#)
 - pressure effects, [563](#)
 - problem formulation, [538](#)
 - radiative transfer techniques, [594](#)
 - Sun glint, [569](#)
 - surface reflectance effects, [563](#)
 - vicarious calibration, [555](#)
 - whitecaps, [571](#)
- Autocovariance, [703](#)
 - modeling turbulence-generated water surfaces, [721](#)
 - Wiener-Khinchin theorem, [705](#)
- Backscattering, [182](#)
 - backscattering coefficient defined, [102](#), [182](#)
 - backscattering coefficient, measurement, [183](#)
 - backscattering fraction, defined, [102](#)
 - models for, [185](#)
- Beam Attenuation Coefficient
 - compared to diffuse attenuation, [125](#)
 - defined, [99](#)
 - errors in measurement, [109](#)
 - measurement, [106](#)
- Beam Spread Function, [344](#)
 - equivalence to the PSF, [347](#)
 - use in modeling lidar, [348](#)
- Beer's Law, [356](#)
- Bidirectional Reflectance Distribution Function (BRDF), [494](#)
 - BRDFs as PDFs, [788](#)
 - Lambertian, [498](#)
- Bioluminescence, [85](#)
- Blackbody Radiation, [71](#)
- Bouguer's Law, [356](#)
- Bubbles, [274](#)

- Case1-Case2 water classification, 143
- Cherenkov Radiation, 89
- Classification
 - marine particles, 247
 - water bodies, 140
- Closure, 395
- Colored Dissolved Organic Matter (CDOM), 266
 - absorption by CDOM, 266, 283, 288
 - elastic scattering by CDOM, 269
 - in coastal waters, 296
- Convolution, 637
 - cyclic, 796
 - Fourier 1-D convolution theorem, 637
 - Fourier convolution theorem, 799
 - serial, 795
- Coordinate Systems, 19
 - for image analysis, 801
 - for polarization, 47
 - for scattering of polarized light, 49
 - global, 19
 - local, 22
- Density Functions, 82
- Diffuse Attenuation Functions
 - compared to beam attenuation, 125
 - defined, 120
 - dependence on IOPs and environmental conditions, 120
 - Gordon's normalization of K_d , 127
 - inelastic scattering effects on, 242
 - models for K_d , 128
- Dirac Delta Function, 25, 502, 794
- Dispersion, 420
 - anomalous, 428
 - dispersion relations, 422
 - general development, 426
 - Kramers-Kronig relations, 432
 - phase and group speeds, 420
- Electric and Magnetic Fields
 - complex representation, 416
 - fundamentals, 402
 - terminology, 413
- Energy Conservation, 389
 - across the air-water surface, 390
 - expressed as plane irradiance, 391
 - expressed as radiance, 394
 - expressed as scalar irradiance, 392
 - within the water, 390
- Extinction Paradox, 454, 467
- Fluorescence, 221
 - by phytoplankton, 265
 - CDOM fluorescence, 232
 - CDOM fluorescence effects, 241
 - CDOM fluorescence excitation-emission function, 236
 - CDOM fluorescence scattering coefficient, 232
 - CDOM wavelength redistribution function, 233
 - chlorophyll fluorescence, 227
 - chlorophyll fluorescence effects, 238
 - chlorophyll fluorescence emission function, 231
 - chlorophyll fluorescence quantum efficiency, 228
 - chlorophyll fluorescence scattering coefficient, 227
 - chlorophyll fluorescence wavelength redistribution function, 227
 - Jablonski diagram, 222
 - physics of, 222
 - quenching, 230
- Forel-Ule water color classification, 141
- Fourier Transforms, 629
 - continuous transforms, 631
 - convolution, 637
 - discrete transforms, 632
 - frequency ordering, 635
 - negative frequencies, 636
 - notation, 629
 - Parseval's relations, 637
- Fresnel Reflectance
 - as a BRDF, 502
 - for polarized light, 477
 - for unpolarized light, 476
- Geometric Optics Approximation, 466
- Geometry
 - coordinate systems, 19
 - for scattering of polarized light, 46
 - solid angle, 24

- Gershun's Law, 387
 use in computing heating rates, 389
 used in computing absorption, 397
- Heaviside Step Function, 637
- HydroLight
 description of, 379
 example output: K functions, 120, 242
 example output: fluorescence effects, 238, 241
 example output: irradiances, 237, 238, 390
 example output: normalized reflectance $[\rho_w]_N^{\text{ex}}$, 549
 example output: photometric variables, 617
 example output: radiances, 60, 64, 377, 394
 example output: Raman scatter, 237
 example output: reflectances, 135, 237
 example output: Secchi depth, 626, 818
- Image Prediction, 791
 boundary conditions, 794
 color images, 822
 cyclic convolution, 796
 cyclic shifting, 803
 examples of image degradation, 807
 Fourier convolution theorem, 799
 image formats, 814
 modulation transfer function, 824
 PSF role in, 792
 serial convolution, 795
- Index of Refraction
 of water, 252
- Inelastic Scattering
 effects on K_d , 242
 general formalism, 211
 Raman scattering, 213
- Inherent Optical Properties
 additivity, 113
 biogeochemical models, 292
 classic IOP model for Case 1 water, 281
 conceptual definitions, 98
 historical notation, 100
 new IOP model for Case 1 water, 284
 of bubbles, 274
 of CDOM, 266
 of coastal waters, 295
 of minerals, 271
 of non-algal particles, 269
 of phytoplankton, 259
 of water, 252
 operational definitions, 106
- Instruments
 ac-s, 110
 ac9, 397
 acs, 397
 Forel-Ule color scale, 141
 General Angle Scattering Meter (GASM), 112
 ICAM, 397
 LISST, 208
 LISST-VSF, 113
 PSICAM, 110, 397
 spectrophotometer, 110, 163
 Volume Scattering Meter, 113, 195, 201
- Intensity
 definition, 34
 in definition of the VSF, 101
- Irradiance
 cosine law, 30
 example data, 57
 plane irradiance, definition, 30
 r-squared law, 441
 scalar irradiance, definition, 32
 vector irradiance, definition, 34
- Irradiance Reflectance R
 dependence on IOPs and environmental conditions, 136
- Irradiance Reflectance R
 defined, 132
 dependence on IOPs and environmental conditions, 135
- Jablonski Diagram, 155, 222
- Jerlov water type classification, 142
- Kramers-Kronig Relations, 432
- Lambert's Law, 356
- Lidar Equation, 348
- Light
 electromagnetic formulation, 332, 407

- historical understanding, 1
- Lunar Spectrum, 80
- Maxwell's Equations
 - in a vacuum, 402
 - in dielectric media, 408
 - Mie's solution, 437
 - overview, 332
 - plane wave solutions, 414
- Mie Theory
 - examples, 448
 - extinction paradox, 454
 - geometric optics approximation, 466
 - geometry, 439
 - overview, 175
 - radar cross section, 447
 - Rayleigh's approximation, 459
 - Rayleigh-Gans approximation, 463
 - solution, 441
 - statement of the problem, 438
 - used with particle size distributions, 447
- Minerals
 - absorption by, 271, 300
 - remote sensing of, 304
 - scattering by, 272, 301
- Modulation Transfer Function, 824
 - Computing the MTF from the PSF, 827
 - measuring the MTF, 824
- Monte Carlo Simulation, 357, 727
 - backward ray tracing, 764
 - BRDFs as PDFs, 788
 - determining path lengths, 730
 - determining scattering angles, 731
 - error estimation, 745
 - importance sampling, 753
 - in-water scattering, 776
 - Lambertian surfaces, 773
 - probability functions, 728
 - ray tracing, 769
 - ray tracing options, 735
 - ray weights in BMC calculations, 782
 - ray-plane intersections, 770
 - sea-surface interactions, 774
- n-squared law for radiance, 474
- Non-algal Particles (NAP), 269
- Nyquist Frequency, 646
- Optical Density, 110
- Optical Depth, 22, 329
- Orthogonality of Sines and Cosines, 640
- Parseval's Relations, 637
- Particle Size Distributions, 305
 - area and volume, 308
 - creating from data, 317
 - cumulative and number, 307
 - fitting a power law, 320
 - models for, 309
 - non-uniqueness of particle size, 305
 - parametric description, 318
 - Sauter mean diameter, 314
 - used in Mie theory, 447
- Petzold's VSF measurements, 192
- Phase and Group Speeds, 421
 - for light in water, 427
- Phase Function, 102
 - Fournier-Forand, 199
 - Henye-Greenstein, 196
 - Petzold average-particle, 195
 - Reynolds-McCormick, 198
- Photometry, 597
 - chromaticity, 606
 - CIE chromaticity coordinates, 607
 - contrast, 622
 - converting chromaticity to RGB values, 612
 - human color vision, 598
 - law of contrast transmittance, 622
 - luminance difference law, 620
 - luminance transfer equation, 616
 - luminosity functions, 602
 - metameric spectra, 610
 - Secchi disk, 623
- Photons
 - as elementary particles, 10
 - physical properties, 18
 - single-photon interference, 8
- Photosynthetically Available Radiation (PAR)
 - definition, 35
- Phytoplankton, 259
 - absorption, 260
 - fluorescence, 260
 - pigment packaging, 260
 - pigments, 260

- scattering, 260
- Point Functions, 82
- Point Spread Function, 344
 - equivalence to the BSF, 347
 - polar to Cartesian coordinates, 801
- Polarization
 - defined, 40
 - in radiative transfer equations, 333
 - scattering geometry, 46
 - Stokes vector, 39
- Poynting Vector, 419
- Quantum Electrodynamics (QED), 7, 331
- Quasi-Single-Scattering Approximation, 366
- r-squared law for irradiance, 441
- Radiance, 27
 - asymptotic distribution, 370
 - conceptual definition, 30
 - example data, 60
 - limitations of, 54
 - n-squared law, 474
 - operational definition, 29
 - visualizing, 64
- Radiative Transfer Equations
 - Gershun's law solution, 387
 - QSSA solution technique, 366
 - scalar RTE, heuristic derivation, 324
 - scalar, rigorous derivation, 340
 - scalar, standard forms, 328
 - solution techniques, 355
 - SSA solution technique, 359
 - vector, for mirror-symmetric media, 335
 - vector, most general, 330
- Radiative Transfer Theory, 323
 - closure, 395
 - radiative processes, 324
 - scalar RTE, heuristic derivation, 324
 - scalar RTE, rigorous derivation, 340
 - solving the RTE, 355
- Radiometer
 - Gershun tube, 27
 - instrument types, 12
 - modern design, 28
- Radiometry
 - frequency variables, 37
 - geometrical radiometry, 27
 - geometry, 19
 - intensity, definition, 34
 - plane irradiance, 30
 - scalar irradiance, 32
 - Stokes vector, 39
 - terminology and notation, 36
 - vector irradiance, 34
- Raman Scattering, 213
 - effect on in-water irradiance, 237
 - effect on remote-sensing reflectance, 237
 - interpretation of in-water emission, 218
 - phase function, 216
 - scattering coefficient, 213
 - temperature and salinity dependence, 220
 - wavelength redistribution function, 214
- Rayleigh Scattering Approximation, 459
 - applicability of, 461
 - obtained from Mie theory, 460
- Rayleigh-Gans Approximation, 463
- Reflectance
 - BRDF, 494
 - exact normalized, 541
 - Fresnel's equations for polarized light, 477
 - Fresnel's equations for unpolarized light, 476
 - irradiance reflectance R , defined, 132
 - Lambertian, 498
 - measures of, 132, 500
 - remote-sensing reflectance R_{rs} defined, 134
- Refraction
 - Snell's law, 173, 471
- Remote Sensing, 505
 - as an inverse problems, 512
 - data processing levels, 509
 - data resolution, 508
 - terminology, 507
 - thematic mapping, 520
 - validation, 510
- Remote-sensing Reflectance R_{rs}
 - defined, 134
 - dependence on IOPs and environmental conditions, 135, 136
 - exact normalized, 541
- Scatterance, 98

- Scattering
 - backscattering coefficient defined, 102
 - by bubbles, 274
 - by homogeneous spheres, 175
 - by irregular particles, 177
 - by minerals, 272, 301
 - by non-algal particles, 269
 - by phytoplankton, 263
 - by pure water, 178
 - by turbulence, 180
 - by water, 257
 - Compton scattering, 5
 - Einstein-Smoluchowski theory, 178
 - Fourier-Forand phase function, 199
 - Henyey-Greenstein phase function, 196
 - importance of, 169
 - index of refraction, 172
 - Mie theory, 437
 - models for, 173
 - Petzold's measurements of the VSF, 192
 - phase function, 102
 - physical basis of, 172
 - relation to particle size and density, 302
 - Reynolds-McCormick phase function, 198
 - scattering coefficient, defined, 102
 - Snell's law, 173
- Scattering Angle
 - computed from directions, 21
- Scattering Coefficient
 - defined, 99
 - measurement, 111
- Secchi Disk, 623
 - black and white disks, 627
 - classical theory of Secchi depth, 624
 - examples of visibility, 817
 - Lee et al. theory of Secchi depth, 626
- Single-scattering Albedo
 - defined, 99
- Single-scattering Approximation, 359
- Snell's law, 173, 472
- Solar Energy
 - creation, 76
 - spectrum, 77, 82
- Solid Angle
 - defined, 24
 - differential element of, 25
 - of a spherical cap, 25
- Source Function
 - for bioluminescence, 87
 - for CDOM fluorescence, 232
 - for chlorophyll fluorescence, 227
 - for Raman scattering, 213
 - general formalism, 211
 - in the SRTE, 328, 342
 - in the VRTE, 334
- Spectrophotometry, 163
- Successive-Order-of-Scattering Method, 359
- Surfaces, 471
 - BRDF, 494
 - Cox-Munk slope statistics, 486
 - Lambertian, 498
 - level sea surface, 471
 - modeling resolution, 691
 - modeling time-dependent surfaces, 698
 - modeling wind-blown sea surfaces, 488
 - modeling with autocovariance functions, 703, 721
 - modeling with wave variance spectra, 665
 - spreading functions, 687
 - turbulence-generated, 721
- Ternary diagrams for water classification, 145
- Thematic Mapping, 520
- Transmittance, 99
- Turbulence
 - effects on water surfaces, 721
 - scattering by, 180
- Visibility, 597, 791
 - contrast, 622
 - Secchi disk, 623
 - Secchi disk examples, 817
- Volume Scattering Function
 - asymmetry parameter, 103
 - conceptual definition, 101
 - example plots, 103
 - measurement, 111
 - phase function, 102
- Water, 252
 - absorption, 256
 - elastic scattering, 257
 - imaginary index of refraction, 255

- real index of refraction, [253](#)
- Water Classification Schemes, [140](#)
 - Case 1 and Case 2, [143](#)
 - Forel-Ule color classification, [141](#)
 - Jerlov water type, [142](#)
 - Ternary diagrams, [145](#)
- Wave Variance Spectra, [643](#)
 - elevation variance spectra, [648](#)
 - Elfouhaily et al. spectrum, [659](#)
 - Pierson-Moskowitz spectrum, [656](#)
 - slope variance spectra, [653](#)
 - spreading functions, [662](#)
 - theory, [648](#)
- Wiener-Khinchin Theorem, [705](#)
 - analytical example, [706](#)
 - numerical example, [709](#)