

Intermittence et dynamique multi-échelles des données haute fréquence à l'aide de la décomposition modale empirique

François G. Schmitt, Yongxiang Huang

Résumé

Les données haute fréquence ont des propriétés complexes, possédant de fortes fluctuations multi-échelles. Elles possèdent également de nombreuses valeurs manquantes. Nous présentons ici une méthode permettant de caractériser l'intermittence des fluctuations, à multi-échelles, tenant compte des valeurs manquantes et des forçages périodiques. Cette approche, appelée décomposition modale empirique, associée à l'analyse spectrale de Hilbert, est présentée ici et son utilisation est illustrée par un exemple provenant des données MAREL Carnot : données de température marine et atmosphérique, et données de fluorescence et d'oxygène dissous.

Introduction

Les géosciences sont caractérisées par de fortes fluctuations, à de nombreuses échelles, associées à une forte complexité. Des méthodes adéquates sont nécessaires pour appréhender ces fortes fluctuations. Ceci est d'autant plus vrai de l'océanographie côtière, sujette à différents forçages, déterministes comme les cycles annuels, diurnes, tidaux, et stochastiques via par exemple la turbulence. Les séries temporelles en point fixe s'apparentent à des mesures eulériennes, donnant accès via leurs mesure à long terme, à de nombreuses échelles de temps et indirectement d'espace. Ces séries possèdent des valeurs manquantes, en raison de pannes régulières, et de maintenances. Ces valeurs manquantes sont de longueur variable¹, et pour cette raison ici nous ne choisissons pas de combler les valeurs manquantes (par une méthode ou une autre), mais d'utiliser une méthodologie pour l'analyse de données, compatible avec l'irrégularité de la fréquence de mesure. Cette méthode permet également de caractériser l'intensité des fluctuations (leur intermittence), et n'est pas trop sensible aux forçages périodiques. Ceci signifie que l'intermittence de la partie stochastique

peut être considérée, même en présence d'un forçage important (forçage diurne ou tidal par exemple)².

Cette approche a déjà été appliquée à des données dans la zone de déferlement côtière³, des données hydrologiques⁴, dans le cadre de la turbulence⁵, dans le cadre de l'énergie éolienne⁶, et pour des données haute fréquence en milieu côtier⁷. Ci-dessous, nous présentons cette méthodologie, puis nous l'appliquons à titre d'illustration à des données MAREL Carnot.

1. DUR G., SCHMITT F. G., SOUSSI S., « Analysis of high frequency temperature time series in the Seine estuary from the Marel autonomous monitoring buoy », *Hydrobiologia*, 588, 2007, p. 59-68 ; SCHMITT F. G., DUR G., SOUSSI S., ZONGO S.B., « Statistical properties of turbidity, oxygen and pH fluctuations in the Seine river estuary (France) », *Physica A*, 387, 2008, p. 6613-6623.

2. HUANG Y., SCHMITT F. G., LU Z., LIU Y., « An amplitude-frequency study of turbulent scaling intermittency using Hilbert spectral analysis », *EPL*, 84, 2008, 40010; SCHMITT F. G., HUANG Y., *Stochastic analysis of scaling time series, from turbulence theory to applications*, Cambridge University Press, 2016.

3. SCHMITT F. G., HUANG Y., LU Z., LIU Y., FERNANDEZ N., « Analysis of turbulent fluctuations and their intermittency properties in the surf zone using empirical mode decomposition », *Journal of Marine Systems*, 77, 2009, p. 473-481.

4. HUANG Y.X., SCHMITT F. G., LU Z. M., LIU Y.L., « Analysis of daily river flow fluctuations using empirical mode decomposition and arbitrary order Hilbert spectral analysis », *Journal of Hydrology*, 373, 2009, p. 103-111.

5. HUANG *et al.*, 2008, *op. cit.* ; HUANG Y., SCHMITT F. G., HERMAND J.-P., GAGNE Y., LU Z. M., LIU Y. L., « Arbitrary order Hilbert spectral analysis for time series possessing scaling statistics: a comparison study with detrended fluctuation analysis and wavelet leaders », *Physical Review E*, 84, 2011, 016208.

6. CALIF R., SCHMITT F. G., HUANG Y., « Characterization of wind energy fluctuations using arbitrary-order Hilbert spectral analysis », *Physica A*, 392, 2013, p. 4106-4120.

7. HUANG Y., SCHMITT F. G., « Application of an empirical mode decomposition based time dependent intrinsic correlation to marine data », *Journal of Marine Systems* 130, 2014, p. 90-100.

Quatre séries ont été choisies pour cela : des données de température océanique, atmosphérique, de fluorescence et d'oxygène dissous.

Présentation de la méthode

Décomposition modale empirique

La décomposition modale empirique (*empirical mode decomposition*, EMD) est une méthode introduite en 1998 par des chercheurs de la NASA, Norden Huang et ses collaborateurs, pour analyser des séries temporelles non-linéaires et non-stationnaires⁸. L'objectif de cette méthode est de décomposer une série temporelle en une somme de sous-séries temporelles, appelées modes (*intrinsic mode functions*, IMF). Ces modes ne sont pas des fonctions sinus comme pour la décomposition de Fourier, ni des fonctions prédéfinies comme pour les ondelettes, mais sont des séries approximativement monochromatiques, ce qui signifie qu'elles sont localisées en fréquence. Cette méthode a rencontré un grand succès dans les sciences de la nature et de la société, avec plus de 6 500 citations de l'article original, selon le Web of Science. Elle ne fonctionne pas pour toutes les séries, mais pour les séries environnementales, elle a montré son efficacité.

Le cadre général de cette méthode d'analyse est de considérer un signal en tant que superposition de différentes oscillations quasi monochromatiques (les modes). Ces modes sont extraits des données d'origine par l'application d'un algorithme, que nous détaillons pas ici⁹. L'algorithme fonctionne de façon itérative, des hautes fréquences vers les basses fréquences. À la fin de la décomposition, la série temporelle d'origine s'exprime comme la somme d'un nombre fini de modes $C_i(t)$ et d'un résidu $r(t)$:

$$X(t) = \sum_{i=1}^p C_i(t) + r(t) \quad (1)$$

Le nombre p de modes nécessaires pour une série temporelle de longueur N est de l'ordre de $\log_2 N$. Cette méthode peut aussi être utilisée pour filtrer des données (en excluant les premiers modes à haute fréquence) ou pour considérer les fluctuations à

certaines échelles (en étudiant les modes correspondant à ces échelles)¹⁰.

Caractérisation de l'intermittence par des analyses temps-fréquence-amplitude

Cette décomposition se complète par une seconde étape, visant à extraire localement (pour chaque temps t) une amplitude et une fréquence instantanées. Pour cela, pour chaque mode, une transformation de Hilbert est appliquée :

$$\hat{X}(t) = \frac{1}{\pi} \int \frac{X(u) du}{t - u} \quad (2)$$

Cette transformée de Hilbert est utilisée pour construire un signal analytique de la façon suivante :

$$Z(t) = X(t) + i\hat{X}(t) = A(t)e^{i\theta(t)} \quad (3)$$

où $A(t)$ et $\theta(t)$ sont respectivement l'amplitude et la phase, estimées localement, pour chaque temps t . La phase peut être dérivée pour donner la fréquence instantanée $\omega(t) = \theta'(t)/2\pi$. Nous avons ainsi une méthode d'analyse temps-fréquence-amplitude, fournissant, pour chaque temps t , en fait plusieurs valeurs instantanées de l'amplitude et de la fréquence, puisqu'il y en a une pour chaque mode. Ceci fournit donc, pour chaque temps t , p valeurs instantanées de l'amplitude et de la fréquence.

Ce type de décomposition permet de fournir une densité de probabilité $p(\omega, A)$ en 2D dans un espace amplitude-fréquence ; cette densité de probabilité permet d'extraire l'énergie associée à une fréquence donnée $p(\omega, A)$:

$$h(\omega) = \int_0^\infty p(\omega, A) A^2 dA \quad (4)$$

Ceci fournit une approche alternative à l'utilisation de la transformée de Fourier pour estimer la densité spectrale d'énergie $E(f)$. La comparaison entre l'approche via la décomposition modale empirique et la transformation de Hilbert (EMD-HSA) et l'approche en Fourier, a été réalisée pour des données simulées et des données de turbulence¹¹. L'approche EMD-HSA est attrayante par le fait qu'elle peut fonctionner lorsque certaines données sont manquantes ; elle ne requiert par un échantillonnage régulier, comme la méthode FFT (Fast Fourier transform), souvent implémentée pour réaliser les transformations de Fourier.

Nous avons implémenté, dans le cadre de la thèse de Yongxiang Huang soutenue en 2009, une

8. HUANG N.E., SHEN Z., LONG S.R., WU M.C., SHIH H.H., ZHENG Q., YEN N.C., TUNG C.C., LIU H.H., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proceedings of the Royal Society A, 454(1971), 1998, p. 903-995.

9. Voir l'article d'origine, HUANG *et al.*, 1998, *op. cit.* ; voir aussi SCHMITT, HUANG, 2016, *op. cit.*

10. FLANDRIN P., RILLING G., GONCALVÈS P., Empirical mode decomposition as a filter bank, IEEE Signal Processing Lett., 11(2), 2004, p. 112-114.

11. HUANG *et al.*, 2008, *op. cit.* ; SCHMITT, HUANG, 2016, *op. cit.*

généralisation de cette approche permettant de considérer l'intermittence des fluctuations dans le cadre de la méthode EMD-HSA. Il s'agit de considérer les moments d'ordre $q > 0$, plutôt que le moment d'ordre 2, et donc l'expression suivante :

$$L_q(\omega) = \int_0^\infty p(\omega, A) A^q dA \quad (5)$$

En cas d'invariance d'échelle, ces moments d'ordre q de l'amplitude caractérisent l'intermittence via la relation en loi de puissance suivante¹² :

$$L_q(\omega) \approx \omega^{-1-\zeta(q)} \quad (6)$$

où $\zeta(q)$ est l'exposant d'intermittence, associé aux fluctuations de la série analysée, classiquement estimé via des fonctions de structure¹³. L'estimation de cette fonction, pour une gamme de fréquence donnée sur laquelle l'invariance d'échelle est constatée, permet de caractériser l'intensité des fluctuations.

Cette fonction peut aussi être comparée à différents modèles de façon à déterminer quel modèle s'applique le mieux. Par exemple, pour un mouvement Brownien, $\zeta(q) = q/2$; pour un mouvement Brownien fractionnaire, $\zeta(q) = qH$, où $\zeta(1) = H$ ($0 < H < 1$) est appelé exposant de Hurst, et correspond à l'invariance d'échelle des fluctuations moyennes. Ces modèles sont caractérisés par une fonction linéaire, et sont appelés self-affines. D'autres modèles existent, pour lesquels la fonction est bilinéaire, et pour des modèles multifractals, la fonction est non-linéaire et concave ou convexe. Plus le moment est élevé, plus les fluctuations considérées sont importantes : les moments faibles caractérisent les petites fluctuations et les fortes fluctuations les extrêmes. Le fait que la fonction $\zeta(q)$ soit concave est une caractéristique de ce qui est appelé multifractalité et représente le fait que les forts moments sont associés à des fluctuations plus importantes que ce qui serait attendu si la fonction était linéaire. La non-linéarité représente donc un indicateur de l'intermittence des fluctuations de la série analysée.

Il est maintenant classique et bien reconnu que la turbulence Eulérienne se caractérise par un exposant $\zeta(q)$ non linéaire et convexe, avec H proche de $1/3$ ¹⁴. Notons enfin un autre avantage de la méthode EMD-HSA pour l'étude des fluctuations stochastiques en présence d'un forçage périodique (comme le forçage tidal ou diurne) : il a été montré que cette méthode est assez peu perturbée par un forçage

périodique localisé en fréquence, et qu'elle permettait d'extraire les exposants caractérisant les fluctuations sous-jacentes¹⁵. Ci-dessous, cette méthode est appliquée, à titre d'illustration, à plusieurs séries issues des données MAREL Carnot.

Application aux séries MAREL Carnot

Nous considérons ici les séries suivantes issues de la base MAREL Carnot proposée dans le cadre du colloque : la température de l'air, la température marine, la fluorescence, et l'oxygène dissous (pourcentage de saturation). Les données considérées correspondent à une portion de 3 années, recueillies du 1^{er} janvier 2007 au 31 décembre 2009. Le total de données est théoriquement de 78911, mais en raison des valeurs manquantes, il n'y a que 81,6 % des données pour l'oxygène dissous, 88,3 pour la fluorescence, 88,1 pour la température de l'air et 90 % pour la température marine. La figure II.5.1 représente les données analysées. Les différentes séries montrent un cycle annuel marqué, sauf les données d'oxygène. On visualise également les blooms printaniers associés à de fortes valeurs de la fluorescence vers les mois d'avril-juin chaque année.

Régimes invariants d'échelle

La figure II.5.2 montre le moment d'ordre 1 estimé via la méthode EMD-HSA, pour les données de température. On constate des valeurs de $H = 0,2$ pour la température atmosphérique, et $H = 0,4$ pour la température marine, pour des périodes allant de 6 heures à 1 mois pour la température atmosphérique, et de 1 jour à 3 mois pour la température océanique. Pour des échelles de l'ordre de 3 mois, les deux spectres se rejoignent. On constate également que le cycle journalier est visible pour les deux séries, tandis que le cycle tidal est identifiable par un léger pic pour les données marines. Pour les échelles inférieures à la journée, le spectre des données atmosphériques est plus pentu que pour les données marines. La série des différences entre température marine et atmosphérique est également considérée, avec un spectre très proche de celui de la température atmosphérique, sauf pour les échelles supérieures à 3 mois. La figure II.5.1b représente la densité de probabilité de la différence de température : celle-ci a des valeurs assez importantes, allant de -12 à +11 °C, avec une moyenne négative

12. HUANG *et al.*, 2008, *op. cit.* ; SCHMITT, HUANG, 2016, *op. cit.*

13. MONIN A.S., YAGLOM A.M., *Statistical Fluid Mechanics* vol. II. MIT Press, 1971.

14. FRISCH U., *Turbulence : the legacy of A.N. Kolmogorov*, Cambridge University Press, 1995 ; SCHMITT, HUANG, *op. cit.*, 2016.

15. HUANG *et al.*, 2008, *op. cit.* ; 2011 *op. cit.* ; SCHMITT, HUANG 2016, *op. cit.*

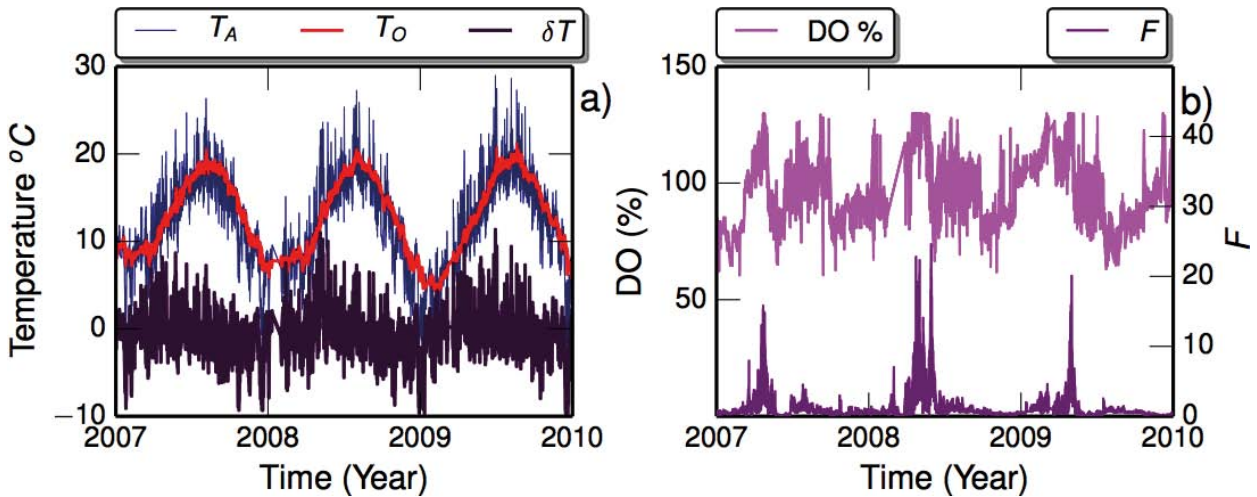


Figure II.5.1. (a) Données de température atmosphérique et océanique (°C), et leur différence. (b) Données d'oxygène dissous (pourcentage de saturation) et données de fluorescence (FFu).

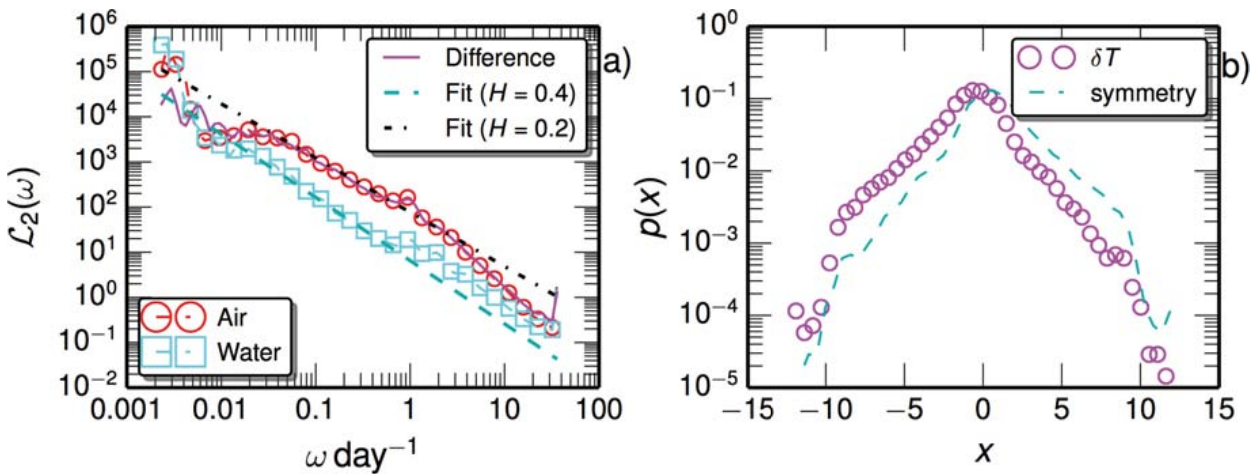


Figure II.5.2. (a) Analyse spectrale de Hilbert, pour le moment d'ordre 1, de la température atmosphérique, océanique, et de la différence ; des ajustements en loi de puissance sont représentés, correspondant à $H=0.2$ et 0.4 respectivement pour la température atmosphérique et océanique. (b) la densité de probabilité de la différence entre température atmosphérique et océanique ; la densité est dissymétrique, et suit approximativement une exponentielle.

(-0,76 °C). La densité symétrique est représentée, montrant que la densité de probabilité n'est pas symétrique, avec des valeurs négatives plus fréquentes que les valeurs positives, indiquant que les eaux marines sont plus froides que l'atmosphère. La densité de probabilité, pour les valeurs positives aussi bien que négatives, est approximativement une exponentielle.

En ce qui concerne la dynamique de l'oxygène dissous, comme discuté ailleurs¹⁶, cette quantité est à la fois advectée par la turbulence, et influencée par de nombreux facteurs, comme la température de l'eau, la salinité, la photosynthèse, la respiration, la turbidité, la pénétration de la lumière, et les échanges entre atmosphère et eaux marines. La dynamique est donc

complexe et aucune théorie n'existe pour décrire sa dynamique à petites échelles. Rappelons que des valeurs de la saturation supérieures à 100 % correspondent à des épisodes de supersaturation associées à une forte production primaire et des fortes oxygénations, tandis que des valeurs inférieures à 50 % indiquent un danger pour la vie aquatiques associées à de l'hypoxie.

La figure II.5.3 représente les analyses spectrales de Hilbert des données d'oxygène dissous et de fluorescence, pour le premier moment, mettant en évidence la valeur de H : on constate des régimes invariants d'échelle avec $H=0,42$ et $0,5$ pour l'oxygène dissous et la fluorescence respectivement, pour des régimes portant sur deux décades, pour les échelles supérieures au jour. On voit l'influence du cycle journalier pour

16. SCHMITT *et al.*, 2008, *op. cit.*

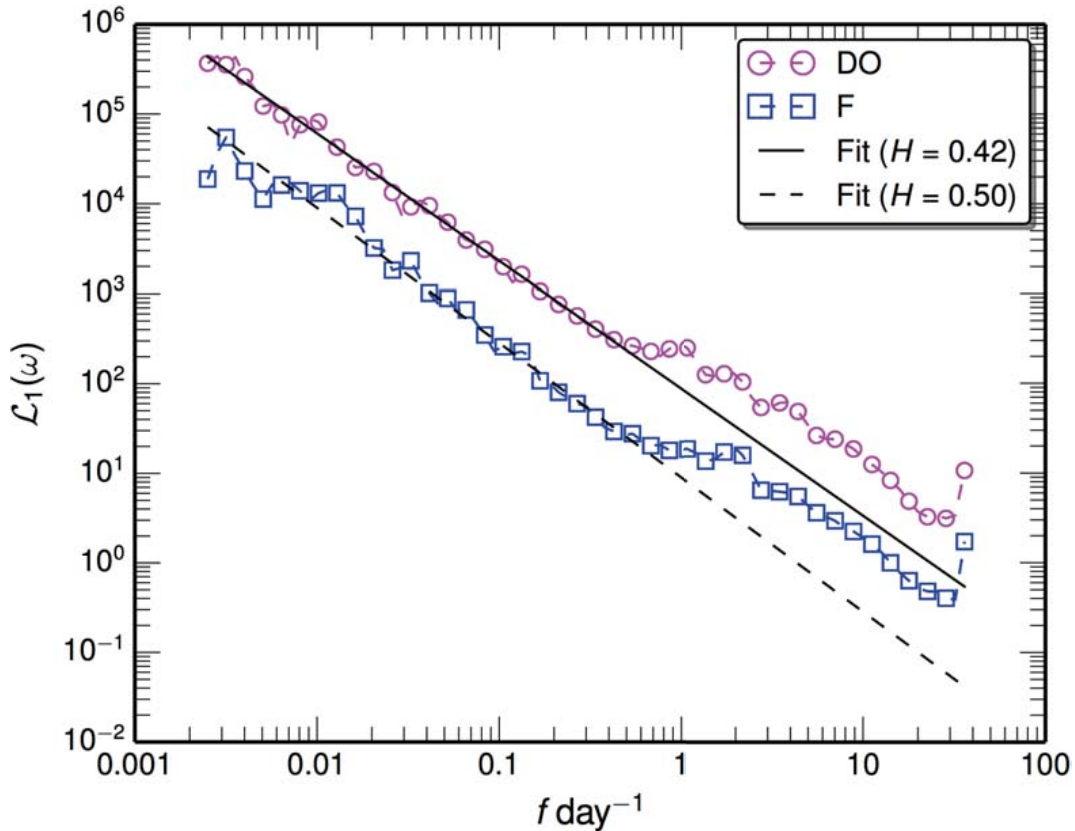


Figure II.5.3. Analyse spectrale de Hilbert pour le moment d'ordre 1, de l'oxygène dissous et la fluorescence. Les ajustements en loi de puissance indiquent $H=0.42$ et 0.5 pour l'oxygène dissous et la fluorescence respectivement.

l'oxygène dissous, mais pas pour la fluorescence, plus influencée par le cycle tidal.

Intermittence

Nous considérons ici l'intermittence des séries à l'aide de la méthode EMD-HSA. Pour les données de température, les exposants d'intermittence n'ont pas été représentés ici, en raison d'un régime invariant d'échelle trop réduit pour les moments d'ordre élevés. Pour l'oxygène dissous et la fluorescence, l'analyse EMD-HSA a été réalisée pour des moments allant de 0 à 4 (figure II.5.4). Le régime invariant d'échelle choisi va de 1,8 à 315 jours pour l'oxygène dissous, et de 1,2 à 167 jours pour la fluorescence. Ces échelles sont comprises entre les échelles de forçage journalier et annuel.

Les exposants extraits sont représentés en figure II.5.5, montrant une forte non-linéarité indiquant une multifractalité caractérisant l'intermittence de ces fluctuations sur les échelles considérées. Une droite de pente 0,42 est représentée, indiquant une possible influence de la turbulence (pour laquelle une pente de $1/3$ est théoriquement attendue).

Conclusions

Il est de plus en plus reconnu que les données haute fréquence dans l'environnement marin côtier sont nécessaires pour mieux appréhender la dynamique des différents paramètres importants pour l'écosystème. Les données récoltées possèdent deux caractéristiques : de fortes fluctuations, à toutes les échelles, et de nombreuses valeurs manquantes en raison de pannes des systèmes et des maintenances nécessaires. Pour analyser ces séries complexes, des méthodes adéquates sont nécessaires. Ici nous avons présenté une méthode issue du domaine du traitement du signal, et aussi de la turbulence. Cette méthode, appelée EMD-HSA, peut être appliquée à des données possédant des valeurs manquantes, fonctionne lorsque des forçages déterministes sont importants, et permet de caractériser l'intermittence et les fortes fluctuations des champs considérés. Nous avons ici illustré cette méthode par l'analyse de séries provenant de la base MAREL Carnot : température atmosphérique, température marine, pourcentage de saturation en oxygène et fluorescence. Les codes Matlab utilisés sont libres d'utilisation¹⁷.

17. <https://github.com/lanlankai>

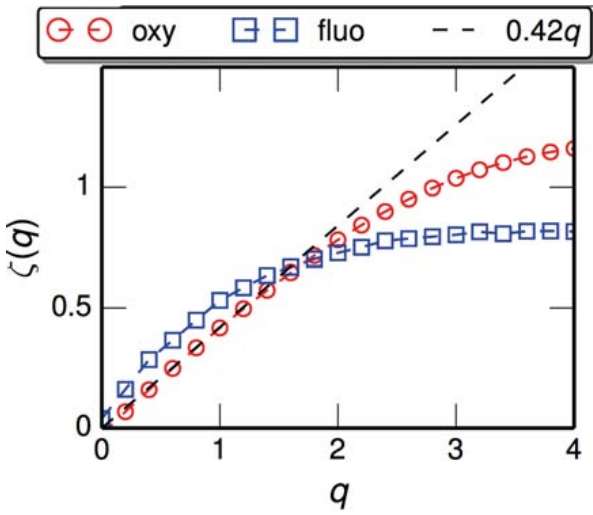
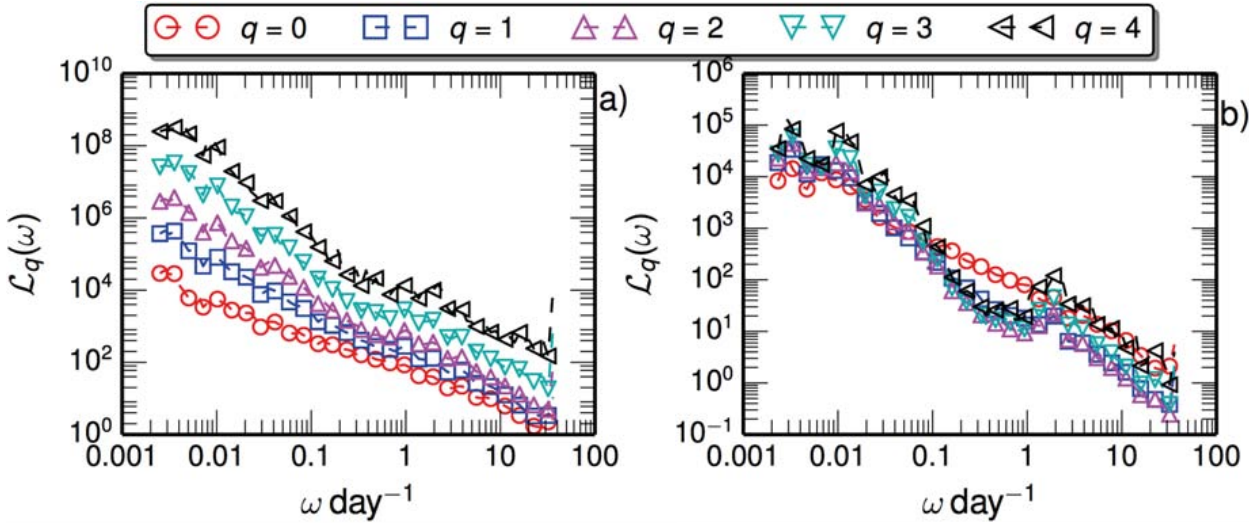


Figure II.5.4. Analyse spectrale de Hilbert représentée pour les données d'oxygène dissous (a) et de fluorescence (b) pour les moments allant de 0 à 4 : les régimes invariants d'échelle considérés vont de l'échelle annuelle à journalière.

Figure II.5.5. Exposants invariants d'échelle estimés pour les données d'oxygène dissous et de fluorescence, à partir de la méthode EMD-HSA. Une droite de pente 0.42 est représentée pour comparaison.

Références bibliographiques

- CALIF R., SCHMITT F. G., HUANG Y., « Characterization of wind energy fluctuations using arbitrary-order Hilbert spectral analysis », *Physica A*, 392, 2013, p. 4106-4120.
- DUR G., SCHMITT F. G., SOUSSI S., « Analysis of high frequency temperature time series in the Seine estuary from the Marel autonomous monitoring buoy », *Hydrobiologia*, 588, 2007, p. 59-68.
- FLANDRIN P., RILLING G., GONCALVÈS P., Empirical mode decomposition as a filter bank, *IEEE Signal Processing Lett.*, 11(2), 2004, p. 112-114
- FRISCH U., *Turbulence : the legacy of A.N. Kolmogorov*, Cambridge University Press, 1995.
- HUANG N.E., SHEN Z., LONG S.R., WU M.C., SHIH H.H., ZHENG Q., YEN N.C., TUNG C.C., LIU H.H., « The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis », *Proceedings of the Royal Society A*, 454(1971), 1998, p. 903-995.
- HUANG Y., SCHMITT F. G., « Application of an empirical mode decomposition based time dependent intrinsic correlation to marine data », *Journal of Marine Systems* 130, 2014, p. 90-100.
- HUANG Y., SCHMITT F. G., LU Z., LIU Y., « An amplitude-frequency study of turbulent scaling intermittency using Hilbert spectral analysis », *Europhysics Letters (EPL)*, 84, 2008, 40010.
- HUANG Y.X., SCHMITT F. G., LU Z. M., LIU Y. L., « Analysis of daily river flow fluctuations using empirical mode decomposition and arbitrary order Hilbert spectral analysis », *Journal of Hydrology*, 373, 2009, p. 103-111.
- HUANG Y., SCHMITT F. G., HERMAND J.-P., GAGNE Y., LU Z. M., LIU Y. L., « Arbitrary order Hilbert spectral analysis for time series possessing scaling statistics: a comparison study with detrended fluctuation analysis and wavelet leaders », *Physical Review E*, 84, 2011, 016208.
- MONIN A.S., YAGLOM A.M., *Statistical Fluid Mechanics vol. II*. MIT Press, 1971.
- SCHMITT F. G., HUANG Y., *Stochastic analysis of scaling time series, from turbulence theory to applications*, Cambridge University Press, 2016.
- SCHMITT F. G., DUR G., SOUSSI S., ZONGO S.B., « Statistical properties of turbidity, oxygen and pH fluctuations in the Seine river estuary (France) », *Physica A*, 387, 2008, p. 6613-6623.
- SCHMITT F. G., HUANG Y., LU Z., LIU Y., FERNANDEZ N., « Analysis of turbulent fluctuations and their intermittency properties in the surf zone using empirical mode decomposition », *Journal of Marine Systems*, 77, 2009, p. 473-481.

François G Schmitt

CNRS, UMR LOG 8187, Université de Lille 1,
Université du Littoral Côte d'Opale, 28 avenue Foch,
62930 Wimereux, France.

Yongxiang Huang

Xiamen University, State Key Laboratory of Marine
Environmental Sciences, Xiamen University,
Xiamen 361102, China.