Biometrics 66, 1247–1255
December 2010
DOI: 10.1111/j.1541-0420.2009.01381.x

Point Transect Sampling Along Linear Features

T. A. Marques,[1,2,*] S. T. Buckland,[1] D. L. Borchers,[1] D. Tosh,[3] and R. A. McDonald[4]

[1]Centre for Research into Ecological and Environmental Modelling, The Observatory, University of St Andrews, St Andrews KY16 9LZ, Scotland
[2]Centro de Estatística e Aplicações da Universidade de Lisboa, Bloco C6, Piso 4, Campo Grande, 1749-016 Lisboa, Portugal
[3]Quercus, School of Biological Sciences, Queen's University, Belfast, 97 Lisburn Road, Belfast BT9 7BL, United Kingdom
[4]The Food and Environment Research Agency, Sand Hutton, York YO41 1LZ, United Kingdom
*email: tiago@mcs.st-and.ac.uk

Summary. Distance sampling is a widely used methodology for assessing animal abundance. A key requirement of distance sampling is that samplers (lines or points) are placed according to a randomized design, which ensures that samplers are positioned independently of animals. Often samplers are placed along linear features such as roads, so that bias is expected if animals are not uniformly distributed with respect to distance from the linear feature. We present an approach for analyzing distance data from a survey when the samplers are points placed along a linear feature. Based on results from a simulation study and from a survey of Irish hares in Northern Ireland conducted from roads, we conclude that large bias may result if the position of samplers is not randomized, and analysis methods fail to account for nonuniformity.

Key words: Distance sampling; Hare survey; Linear features; Nonrandom samplers; Nonuniform density; Point transects; Uniformity.

1. Introduction

Irish hares (*Lepus timidus hibernicus*) are a subspecies of the mountain hare that occurs only in Ireland. Surveys of Irish hares in Northern Ireland are conducted from points along roads for practical reasons. Because of disturbance along roads, and because hares tend to avoid field edges, we cannot assume that hare densities are representative along roads. The methods of this article were motivated by the need to estimate hare densities, allowing for the biased form of sampling.

Given increased human impacts on natural systems, there is a growing need for efficient methods to monitor populations of wild animals. Most often, we cannot count all N animals in the area of interest A, and so we must draw inferences based on a sample from a covered region $a < A$. If detection in the covered region is certain, simple plot sampling leads to a density ($D = N/A$) estimator given by the number of animals detected, n, divided by the area covered, $\hat{D} = n/a$. The uncertainty in this design-based density estimator for the area of interest A arises solely from the sampling design. If animals in the sampling units can remain undetected, then given n detected animals, a density estimator is given by the above estimator divided by $\hat{P}$, the estimated probability of detecting an animal, given it is in the covered area,

$$\hat{D} = \frac{n}{a\hat{P}}. \qquad (1)$$

This is a Horvitz–Thompson-like (HTL) estimator (Borchers, Buckland, and Zucchini, 2002, p. 143–144): the probabilities are estimated instead of known by design as in conventional Horvitz–Thompson estimators. Distance sampling (DS) is one of the most widely used methods to estimate these probabilities and hence density (Buckland et al., 2001). An abundance estimator is readily obtained by multiplying the corresponding density estimator by the size of the area of interest.

Conventional distance sampling (CDS) methods use the recorded distances to detected animals from either lines or points randomly or systematically distributed in the study area. These distances are used to model a detection function, $g(z)$, which represents the probability of detecting an animal, given it is at a distance z from the line or point (z represents either a radial distance from a point or a perpendicular distance from a line). It can be shown that the probability of detecting an animal, given that it is in the covered area, is the mean value of the detection function, with respect to the distribution of distances available for detection, $\pi(z)$, i.e.,

$$P = \mathbb{E}_z[g(z)] = \int_0^w g(z)\pi(z)\,dz, \quad 0 \le z \le w \qquad (2)$$

where w represents a truncation distance (distances larger than w are ignored in the analysis; note that for line transects, we use absolute distance from the line, and so do not differentiate whether a distance was to the left of the line or to the right). Several assumptions need to be satisfied for asymptotically unbiased estimation, namely $g(0) = 1$, animals are detected at their initial location before any response to the observer, and distances are measured without error. Buckland et al. (2001) discuss these assumptions in detail.

A fundamental condition for CDS methods to work, often ignored in practice, is the random location of samplers, independently of the population of interest, which makes it reasonable to treat $\pi(z)$ as known. This has been referred to as the uniformity assumption, because in the case of line transect sampling, it results in a uniform $(0,w)$ distribution of animals with respect to distance from the line. For additional details see also Fewster et al. (2008). Thus for line transect sampling, $\pi(z) = 1/w$. For point transects, the distribution is triangular, $\pi(z) = 2z/w^2$.

Responses of animals to roads have been widely reported (e.g., Yost and Wright, 2001; Ortega and Capen, 2002; Marsh and Beckman, 2004). Nonetheless, DS methods have frequently been used along trails or roads (e.g., Borralho, Rego, and Pinto, 1996; Heydon, Reynolds, and Short, 2000; Ruette, Stahl, and Albaret, 2003; Ward, White, and Critchley, 2004), and although sometimes the problem of violating the assumption of uniformity is recognized (e.g., Heydon et al., 2000; Ward et al., 2004), the practical and logistical advantages are usually considered to outweigh the disadvantages (e.g., Heydon et al., 2000). If animal density is atypical near roads or other linear features, an analysis under the uniformity assumption can lead to considerable bias. However, practitioners rarely consider the consequences for their population estimates of such biases. Because $g(z)$ and $\pi(z)$ appear only as a product in equation (2), they cannot both be estimated from conventional distance data.

Considering a survey design comprising points randomly spaced along a linear feature, or systematically spaced with a random start, it is possible to estimate both $g(z)$ and $\pi(z)$, provided the sighting angle with respect to the linear feature is recorded in addition to the radial distance. We concentrate on point transects as the line transect case requires additional assumptions about the search/detection process.

Surveys along roads or paths provide a common example of when $\pi(z)$ is unknown. Others include cetacean or bird surveys from shore, and points along riparian habitats. Other examples include surveys of a three-dimensional environment in which animals are not uniformly distributed with respect to the third dimension (e.g., radar surveys of flying birds or sonar surveys of fish). In this article, the main concern is the bias in DS density estimates on a strip of width $2w$ around the linear feature, due to failure of the uniformity assumption. The more general problem of nonrepresentative samples, leading to biases when making inferences for areas larger than the area covered, is not the key focus of this article. However, under special circumstances (Section 2.5), our methods might be useful in reducing this bias.

The structure of the article is as follows. We develop a likelihood that incorporates a nonuniform distribution model along with the detectability model in Section 2. We then present a simulation study in Section 3 and an analysis of the 2005 survey of Irish hares in Section 4, in both cases comparing our methods with CDS estimators. We finish with a discussion of some issues related to the use of these methods, future research problems, and the pitfalls of using nonrandom samplers, such as road-based surveys, for estimating animal abundance.

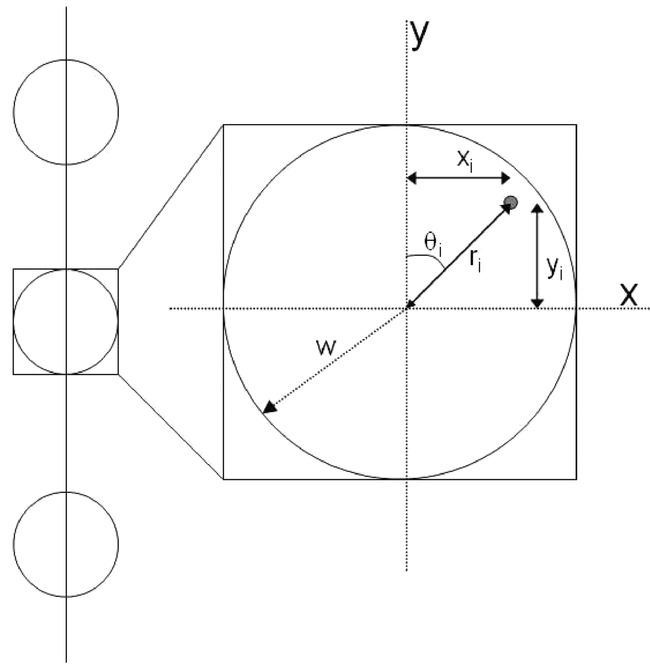

Figure 1. A linear feature along which point transects are placed. A point is blown up and the distances required to define the methods shown (x, perpendicular distance with respect to the linear feature; y, distance along the linear feature; r, radial distance to the animal; θ, sighting angle with respect to the linear feature).

2. Methods

2.1 *Preliminaries*

Consider k point transects laid along the linear feature (Figure 1), with CDS assumptions holding, except that animal density is assumed to be a function of distance x from the linear feature; we represent this function, after folding along the linear feature, by $D(x),\ 0 \leq x \leq w$. Random placement of points with respect to distance y along the linear feature allows us to assume that this function is independent of y. Assume that the detection function is, as usual for point transects, a function of the radial distance alone, $g(r \mid \underset{\sim}{\phi}_1)$, with associated vector of unknown parameters $\underset{\sim}{\phi}_1$. For clarity, we will usually omit the parameters $\underset{\sim}{\phi}_1$, and denote the detection function by $g(r)$.

2.2 *Bivariate Likelihood*

Conventional point transect methods involve the maximization of a likelihood (with respect to $\underset{\sim}{\phi}_1$ alone) based on the *probability density function* (*pdf*) of detected distances in the circle. A model is assumed for the detection function $g(r)$ and $\pi_r(r) = \frac{2r}{w^2}$ is assumed known by design (see Borchers and Burnham, 2004, for further details about these functions; for details about models typically used for $g(r)$, see Buckland et al., 2001, p. 47). This can be generalized to a bivariate likelihood, which assumes that r and θ are independent and the detection probability does not depend on θ, leading to

$$\prod_{i=1}^{n} \pi_\theta(\theta_i) \frac{g(r_i)\pi_r(r_i)}{\displaystyle\int_R g(r)\pi_r(r)\,dr} = \mathcal{L}_\theta \mathcal{L}_r. \tag{3}$$

Because $g(r)$ and $\pi_r(r)$ appear as a product, they cannot be estimated separately from $\mathcal{L}_r$, the CDS likelihood. $\mathcal{L}_\theta$, not usually considered because there is no direct interest in $\pi_\theta(\theta)$, allows us to test whether $\pi_\theta(\theta)$ is uniform, which one would expect under CDS. Failure of such uniformity could indicate the need for our methods.

Consider a square with sides of length $2w$ located so that the linear feature runs along its center. Once the square has been placed along the linear feature, define a Cartesian coordinates system with origin at the square's center (see Figure 1). It is convenient to define x and y to be the absolute values of the Cartesian coordinates.

We represent the *pdf* of x for objects in the square (whether detected or not) by $d(x \mid \underline{\phi}_2)$ on the interval $(0, w)$, indexed by a parameter $\underline{\phi}_2$, which we often drop for readability. Random placement of the square along the linear feature results in a uniform *pdf* of object locations in the y-dimension, given x: $\pi^*(y \mid x) = 1/w$ and hence a joint *pdf* $\pi^*(x, y) = d(x)/w$.

Because observations beyond radial distance w are either ignored in analysis or not recorded, we require the joint *pdf* of x and y conditional on the event $r \le w$, which we call "c." The *pdf* of interest is then $\pi_{x|c}(x \mid c)\pi_{y|x,c}(y \mid x, c)$. For brevity we refer to this as $\pi_{x,y}(x, y)$ (rather than $\pi_{x,y|c}(x, y \mid c)$). From Bayes' theorem,

$$\pi_{x|c}(x \mid c) = \frac{\dfrac{\sqrt{w^2 - x^2}}{w} d(x)}{P_c} \tag{4}$$

where $P_c = \int_0^w \sqrt{w^2 - x^2} d(x)/w\,dx$. Also $\pi_{y|x,c}(y \mid x, c) = \frac{1}{\sqrt{w^2 - x^2}}$, so for objects in a circle of radius r about an observer located at the origin, $\pi_{x,y}(x, y) = \frac{1}{w} d(x)/P_c$.

Because observers usually record (r, θ) and because detection probability depends on r, it is convenient to work in terms of (r, θ) rather than (x, y). We transform variables to obtain $\pi_{r,\theta}(r, \theta) = rd(r \sin\theta)/wP_c$. Using Bayes' theorem (and canceling wP_c), we then obtain the joint *pdf* of *observed* (r, θ) as

$$f_{r,\theta}(r, \theta \mid \underline{\phi}_1, \underline{\phi}_2) = \frac{rd(r\sin\theta)g(r)}{\displaystyle\int_0^w \int_0^{\pi/2} rd(r\sin\theta)g(r)\,d\theta\,dr}. \tag{5}$$

The above formulation explicitly allows the inclusion of a nonuniform model for $d(x)$ in the likelihood that follows. Being cast in a likelihood framework, we can use tools for model selection such as Akaike's information criterion (AIC).

2.3 *Estimating Density in the Covered Circles*

Given the n observed (r, θ) pairs, assuming a parametric form for the $d(x)$ and $g(r)$, we can use the joint distribution of θ and r to build a likelihood that can be maximized to estimate the unknown parameters ($\underline{\phi}_1$ and $\underline{\phi}_2$) as

$$\mathcal{L}(\underline{\phi}_1, \underline{\phi}_2 \mid \underline{r}, \underline{\theta}) = \prod_{i=1}^{n} \frac{\pi(r_i, \theta_i)g(r_i)}{\displaystyle\int_R \int_\theta \pi(r, \theta)g(r)\,d\theta\,dr}. \tag{6}$$

Analogous to the conventional case,

$$P = \int_0^w g(r \mid \underline{\phi}_1)\pi_r(r \mid \underline{\phi}_2)\,dr \tag{7}$$

where now $\pi_r(r) = r\int_0^{\pi/2} d(r\sin\theta)d\theta/wP_c$.

Given the maximum likelihood estimators (MLE's) for the parameters of interest, we can substitute these in equation (7), leading to the following estimator for P:

$$\hat{P} = \int_0^w g(r \mid \hat{\underline{\phi}}_1)\pi_r(r \mid \hat{\underline{\phi}}_2)\,dr. \tag{8}$$

A density estimator is then obtained using an HTL estimator as in equation (1). Note, however, that this estimator applies only to the covered circles and yet our interest is in density in the strip of width $2w$ centered on the linear feature. We address this issue next.

2.4 *Estimating Density in the Vicinity of the Linear Feature*

To estimate density in the vicinity of the linear feature, we must estimate P_c, the probability that an animal is in the circle, given that it is in the square of side $2w$ that contains it. The estimated density in the strip of half-width w about the linear feature is then

$$\hat{D} = \frac{n}{a\hat{P}\hat{P}_c} \tag{9}$$

where $\hat{P}_c$ is obtained using the estimate of $d(x)$ and a is the area of the k squares containing the covered circles ($k4w^2$). (Note that for constant $d(x)$, as in CDS, $P_c = \pi/4$.)

This density estimator is valid only in the strip of width $2w$ centered on the linear feature. Because the linear feature is not randomly located, the estimator will not in general be representative of density in the wider survey region. If an estimate of abundance for this wider region is required, the approach of the next section might be adopted.

2.5 *Extrapolating Results for a Wider Area*

When samplers are placed along unrepresentative linear features, extrapolation of density estimates to the wider area is problematic. If the influence of the linear feature on animal density does not extend beyond distance w, our methods allow more defensible estimation for the wider area than methods that ignore the nonuniform animal distribution.

The density D in a strip of length L and half-width w about the linear feature is given by

$$D = \frac{\displaystyle\int_0^L 2\int_0^w D(x)\,dx\,dy}{2Lw} = \frac{\displaystyle\int_0^w D(x)\,dx}{w} \tag{10}$$

which leads to the following estimator for $D(x)$

$$\hat{D}(x) = \hat{D}w\hat{d}(x) \tag{11}$$

as $d(x) = D(x)/\int_0^w D(x)\,dx$.

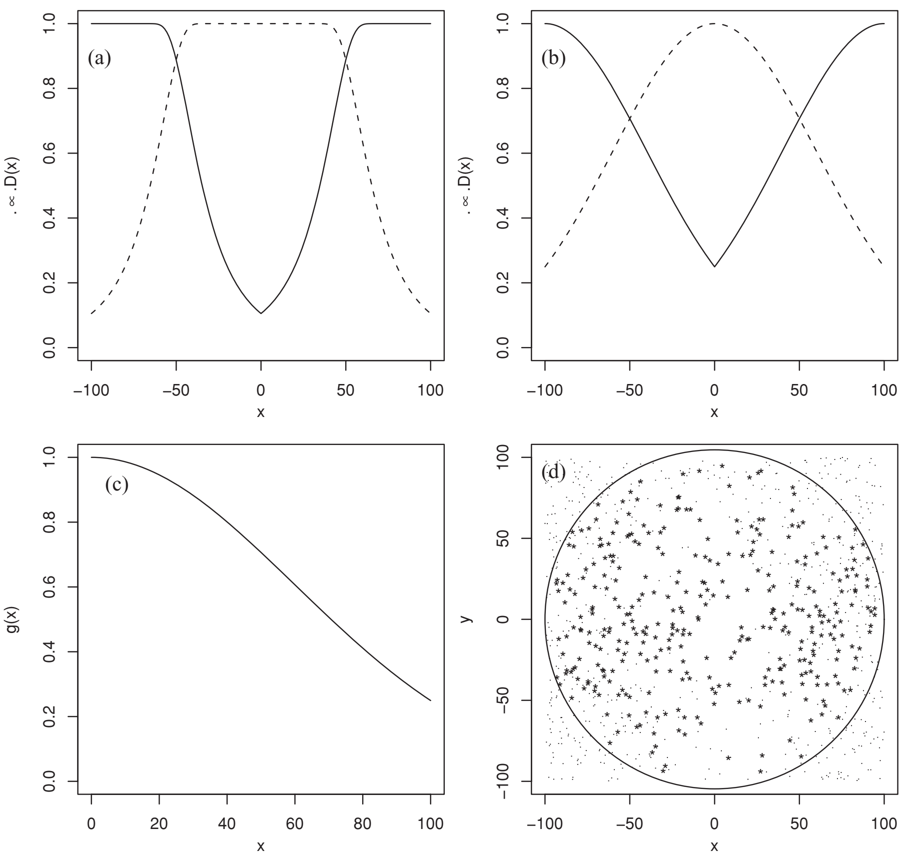

Figure 2. Details of simulation settings. (a) Example of HR density model, used in scenarios 3 and 4; (b) Example of HN density model, used in scenarios 1, 2, 7, and 8. Solid and dashed lines represent, respectively, avoidance and attraction towards the linear feature; (c) Example of HN detection function, used in scenarios 1–6 and 11–12; (d) Realization of a population under scenario 2, with the detected animals shown as *. The linear feature is along $x = 0$.

In contrast with the design-based approach used in conventional methods, this leads to a model-based density estimate for the wider survey region. If we can assume that the influence of the linear feature has disappeared at w, then $\hat{D}(w)$ is a better estimate of density in the wider region than $\hat{D}$. Inspection of the estimated $d(x)$ may allow assessment of whether or not that is a reasonable assumption.

2.6 *Variance Estimation*

A straightforward approach for estimating the variance of the proposed estimators is a nonparametric bootstrap, resampling at the point level. This presents several advantages over methods based on MLE properties: (1) it assumes that the points are independent sampling units, which yields more robust inference than assuming individual detections are independent, required by MLE; (2) it can readily incorporate complex survey spatio-temporal correlation structures (e.g., same points repeated over time); (3) it avoids the need to calculate the variance of complicated functions of the estimated parameters from the estimated information matrix; (4) it avoids the approximation arising from using the delta method to combine the components of variance corresponding to n and P, as in CDS; and (5) it is particularly suited for model averaging. A bootstrap procedure readily extendible to this case is given by Buckland et al. (2001, p. 82–84).

3. A Simulation Experiment

3.1 *Simulation Settings*

Consider a square study area of side 200 m, the center of which is, without loss of generality, at (0,0). There is a linear feature along the y-axis ($x = 0$). Animals have some $d(x)$ and are uniformly distributed in the y direction. To test our methods, a design with one single point transect located at the center of the study area is considered. For fitting the detection and density models, this is equivalent to many points along a linear feature, which for analysis purposes are essentially stacked on top of each other (i.e., density at the single point is the sum of densities at all the points).

A constant population size of $N = 1000$ animals was simulated. The nonuniform density was assumed to be one of four types, determined by two factors (see Figure 2a, b): (1) animals either avoid or prefer the linear feature and (2) have either a hazard rate (HR) or half-normal (HN) form for $d(x)$. For each animal, the x coordinate was generated from these distributions, while the y coordinate was generated from a uniform on (−100,100). The detection function was assumed to be HN (Figure 2c). Given the detected radial distances, r, a truncation distance of $w = \min(100, \max(r))$ was used.

The parameters for the 12 main simulation scenarios are shown in Table 1. For each scenario, the following procedure

Table 1

Simulation scenarios (Sc) considered, as a function of density model $d(x)$ and detection function $g(x)$ parameters. HN stands for half-normal and HR for hazard rate. Parameter values for model HN are indicated by subscript; for HR, parameter values are $\sigma = 60$ and $b = 4.3$. The $d(x)$ model superscript T indicates animals are attracted to linear feature, and F indicates animals avoid it. P is the probability of detecting an animal, given that it is in the circle. P_c is the probability that an animal is in the circle, given that it is in the square that contains it. $\mathbb{E}(n)$ is the average number of detected animals by simulation. True abundance is $N = 1000$ for all scenarios.

Sc	Model	P	P_c	$\mathbb{E}(n)$
1	$HN^T_{60}HN_{60}$	0.58	0.86	496
2	$HN^F_{60}HN_{60}$	0.50	0.72	355
3	$HR^T\, HN_{60}$	0.59	0.89	522
4	$HR^F\, HN_{60}$	0.48	0.70	335
5	$HN^T_{30}HN_{60}$	0.63	0.95	599
6	$HN^F_{30}HN_{60}$	0.41	0.58	238
7	$HN^T_{60}HN_{30}$	0.22	0.86	185
8	$HN^F_{60}HN_{30}$	0.13	0.72	94
9	$HN^T_{30}HN_{30}$	0.28	0.95	266
10	$HN^F_{30}HN_{30}$	0.06	0.58	32
11	$HN^T_{900}HN_{60}$	0.54	0.79	425
12	$HN^F_{900}HN_{60}$	0.54	0.78	424

was repeated 100 times: the animals' positions were generated, the detection process simulated, and the relevant distances to detected animals used to estimate abundance (N), either assuming a uniform distribution as in CDS or using our methods. To evaluate the impact of model misspecification on our methods, estimation was carried out using the following options for modeling $d(x)$: (1) the true model, (2) the wrong model (i.e., HR if true was HN, and vice versa), and (3) using Akaike's information criterion ($AIC = -2\log_e \mathcal{L} + 2p$, where $\mathcal{L}$ is the likelihood and p the number of parameters estimated) to select from these two. Figure 2d presents a single realization of a population under scenario 2, with locations of both detected and undetected animals shown.

3.2 *Simulation Results*

The estimated population size and corresponding mean percentage bias for conventional and proposed methods can be compared in Table 2. The improvement over conventional methods is substantial for all scenarios tested, even if the wrong density model is assumed. AIC proved a good tool to select between candidate models for the density model, except for scenarios leading to very small sample sizes. However, even if the true HN model was considered, the estimator was still appreciably biased in scenario 10. This results from a combination of the nature of the estimator and small sample size. As pointed out by others (e.g., Borchers, 1996; Marques and Buckland, 2003), HTL estimators are positively biased, and this bias can be considerable if the probabilities in the denominator are small, as in scenario 10 (cf. Table 1). Also, given that all the scenarios were identical with respect to true abundance, sample size was proportional to the true inclusion probabilities, and scenario 10 has an expected sample size (32) that is smaller than the minimum number of 60 recommended by Buckland et al. (2001, p. 14).

To investigate this issue further, we considered two additional simulation exercises (note $d(x)$ was assumed known for these): (1) 24 additional scenarios, in which P took values intermediate between those of scenarios shown in Table 1 and (2) 30 additional scenarios, with the same parameters as scenario 10, but with true abundance in the range 1200 to 7000 (with a corresponding range for $\mathbb{E}(n)$ of 38 to 224), compared with 1000 ($\mathbb{E}(n) = 32$) in the original simulation. In Figure 3a, the bias in estimated abundance is shown as a function of true P, for all 36 scenarios considered. It can be seen that values of P of around 15–20% or greater are required for reliable estimation of D. Note that the bias of our methods is nonetheless considerably smaller than that for conventional methods, for all scenarios run. The effect of sample size on estimation bias is clear in Figure 3b, where for sample sizes larger than around 150, the bias becomes small.

For scenarios 11 and 12, $d(x)$ is, for practical purposes, virtually uniform. It is reassuring to see that our methods still work relatively well under the more conventional setting, even if there is not much information in the data to estimate the density parameter ($\underline{\phi}_2$), leading to large variance in estimates of $\underline{\phi}_2$. However, their use would not be recommended in such situations because extra parameters need to be estimated leading to worse precision than for the conventional methods (cf. Table 1). Note the slight bias for these scenarios: when true density falls off very slowly with distance and we use a density model that allows only declining density, sampling variance can result in an estimated density that falls off much more quickly than truth but not in one that falls off much more slowly than truth, because the density model is constrained to be nonincreasing with distance. This results in an estimated density that falls off too fast on average. Similarly, when using a density model that assumes increasing density and the true density gradient is very small, the result is an estimated density that increases too quickly on average.

4. Applying the Methods to Hare Survey Data

In 2005, a survey of Irish hares was carried out in Northern Ireland. In total, 5421 point transects were surveyed along roads, and 210 clusters (314 individual hares) were detected. Here we focus only on estimating cluster density.

These animals tend to avoid field boundaries, as the distance data suggest (Figure 4a). We used our methods to estimate density within the covered region and, using equation (11), to estimate density at $w = 150$ m from the road. Estimates were compared with those from CDS.

We considered four models, the first (M1) being a CDS model, using an HN detection function with one cosine adjustment term added to improve fit, implemented in distance (Thomas et al., 2009). Three models with nonuniform density were considered. All have an HN detection function. Two of these have $d(x)$ corresponding to road avoidance, the first (M2) having $d(x)$ with HN form and the second (M3)

Table 2

Results for the 12 *simulation scenarios* (*Sc*). *N: 2.5th percentile of the estimated values, the mean and the 97.5th percentile, using the Louis and Zeger* (2009) *style; B: percentage bias. First two columns assume $d(x)$ to be HN, next two columns assume HR, next two columns use AIC to choose between HN and HR, and last two columns assume uniformity for $\pi(x)$* (*i.e., CDS*).

	HN assumed		HR assumed		AIC		CDS	
Sc	N	B	N	B	N	B	N	B
1	$_{983}1004_{1024}$	0.4	$_{993}1012_{1031}$	1.2	$_{982}1002_{1021}$	0.2	$_{1421}1440_{1458}$	44.0
2	$_{971}1001_{1031}$	0.1	$_{945}977_{1009}$	−2.3	$_{965}996_{1028}$	−0.4	$_{582}594_{605}$	−40.6
3	$_{991}1008_{1026}$	0.8	$_{968}983_{999}$	−1.7	$_{970}986_{1001}$	−1.4	$_{1559}1578_{1597}$	57.8
4	$_{1206}1245_{1285}$	24.5	$_{969}998_{1026}$	−0.2	$_{988}1023_{1058}$	2.3	$_{452}459_{466}$	−54.1
5	$_{979}991_{1003}$	−0.9	$_{999}1016_{1032}$	1.6	$_{980}992_{1004}$	−0.8	$_{2305}2331_{2356}$	133.1
6	$_{962}1016_{1070}$	1.6	$_{783}831_{879}$	−16.9	$_{958}1013_{1068}$	1.3	$_{300}303_{307}$	−69.7
7	$_{953}993_{1034}$	−0.7	$_{924}961_{999}$	−3.9	$_{936}977_{1017}$	−2.3	$_{1435}1464_{1493}$	46.4
8	$_{962}1032_{1102}$	3.2	$_{1181}1362_{1544}$	36.2	$_{1064}1211_{1358}$	21.1	$_{483}499_{515}$	−50.1
9	$_{966}991_{1016}$	−0.9	$_{1013}1044_{1075}$	4.4	$_{969}995_{1021}$	−0.5	$_{2436}2472_{2508}$	147.2
10	$_{839}1394_{1948}$	39.4	$_{998}2054_{3110}$	105.4	$_{782}1832_{2882}$	83.2	$_{69}73_{77}$	−92.7
11	$_{927}944_{961}$	−5.6	$_{844}1007_{1171}$	0.7	$_{922}940_{958}$	−6.0	$_{972}988_{1004}$	−1.2
12	$_{1020}1039_{1059}$	3.9	$_{1014}1033_{1052}$	3.3	$_{1019}1038_{1057}$	3.8	$_{985}1002_{1018}$	0.2

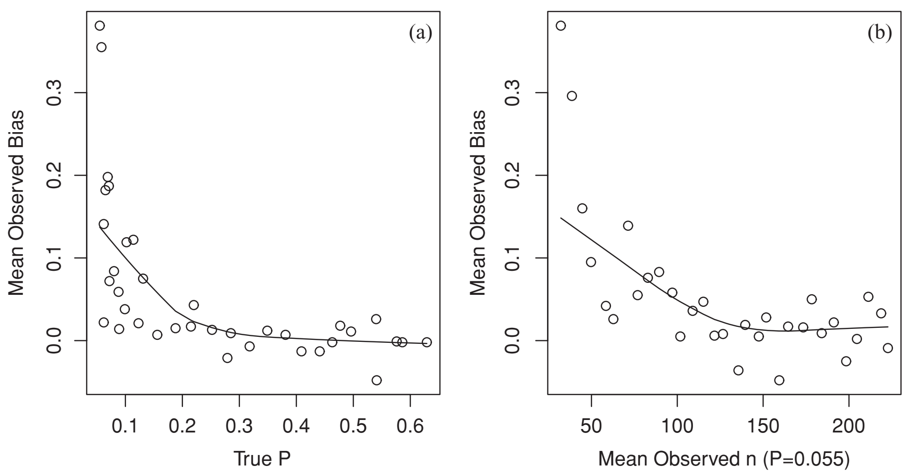

Figure 3. Mean observed percentage bias as a function of (a) True P; (b) Sample size. In (a), the data correspond to the 12 scenarios in Table 1, plus 24 extra scenarios spanning intermediate values of P. In (b), the data correspond to scenario 10 with N varying from 1000 to 7000. The line in both panels is a standard lowess smooth.

having $d(x)$ with HR form. Our final model (M4) has

$$d(x) = \frac{1 - \beta \times f(x,\sigma)}{\displaystyle\int_0^w 1 - \beta \times f(x,\sigma)\,dx} \qquad (12)$$

where $f(x,\sigma) = \exp[-(x/\sigma)^2]/\sqrt{2\pi\sigma^2}$ is the normal density with mean 0 and standard deviation σ. Like M2, M4 is based on the normal distribution, but unlike M2 and M3 it allows easier parameter interpretation. This model accommodates both attraction to ($\beta < 0$), and avoidance of ($\beta > 0$), the linear feature (note $\beta = 0$ gives a uniform).

The results obtained for these models, after numerical maximization of the appropriate likelihoods, are shown in Table 3. Variances were obtained using the standard empirical variance estimator for the CDS model M1 and a bootstrap resampling procedure for the remaining models, considering points as the resampling units. The percentile method was used to obtain the 95% confidence intervals.

These results show that the conventional method underestimates abundance considerably. The nonuniform density models lead to density estimates more than 100% larger than those for M1, for the strip of width $2w = 300$ m centered on the road. Furthermore, estimates from models M2 to M4 indicate that density estimates obtained using CDS methods will result in severe underestimation of density for the wider region. The considerable difference in point estimates obtained from models M2 to M4 is somewhat disappointing. The fit of model M4 is preferred by AIC, and $\Delta AIC = 3.9$ for model M2, the second-best model, indicating a strong preference for M4. For comparison, $\Delta AIC > 80$ for CDS models. Model M3 is not supported by the data (cf. Table 3). M4 is arguably the most interesting of the three models considered, suggesting that it is likely that the effect of the road is almost absent for distances approaching $w = 150$ m. Unfortunately, the

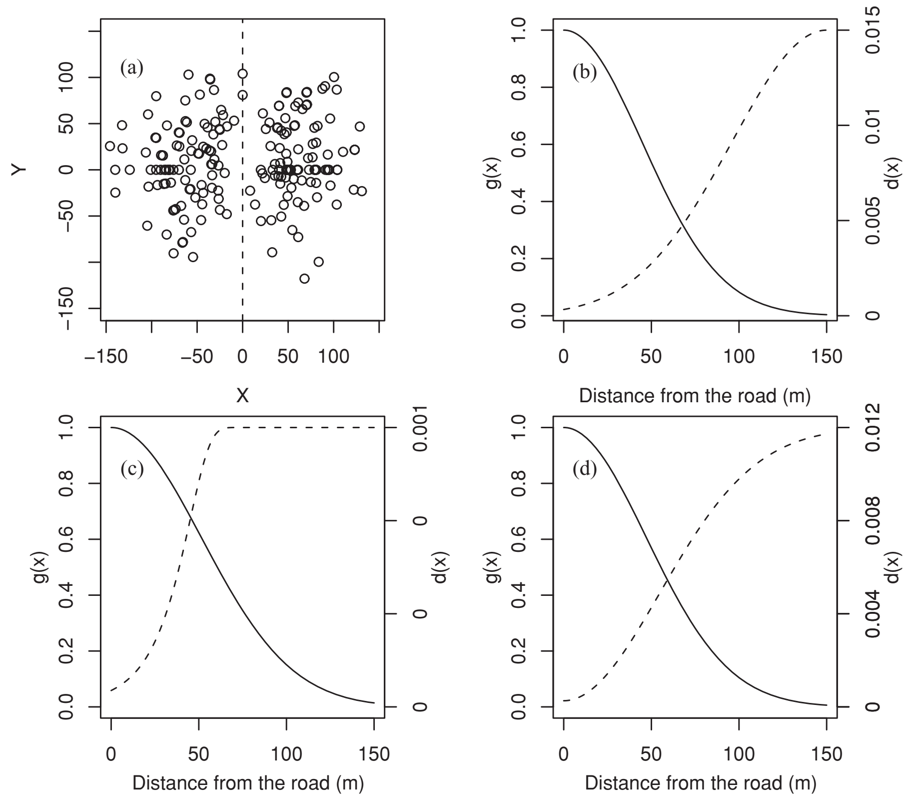

Figure 4. Data and estimated models in the hares example. (a) The positions of the detected hares with respect to the point, considered to be centered at (0,0). The dashed line represents the road position with respect to the point. Data have been truncated so that $r \leq 150$ m; (b) to (d) The estimated density model (dashed line) and detection function (solid line), considering: (b) an HN; (c) an HR; and (d) a normal-based model for $d(x)$.

Table 3

Results of the analysis of the hare data for four models (M): CDS assuming uniformity (M1); proposed method with animals avoiding the road according to an HN gradient (M2), HR gradient (M3), and a normal-based gradient (M4). AIC: Akaike information criterion (AIC weights shown in parentheses); $d(x)$: estimated parameters for the density model; $g(x)$: estimated parameters for the detection function; $\hat{P}$: estimated probability of detecting a hare cluster, given that it is in the circle of radius $w = 150$ m; $\hat{P}_c$: estimated probability that a cluster is in the circle given that it is in the square containing it; $\hat{D}$: estimated cluster density in the strip of width $2w = 300$ m centered on the road; $\hat{D}_{150}$: estimated cluster density at 150 m from the road. Corresponding 95% confidence intervals for density estimates are shown using the Louis and Zeger (2009) style.

M	ΔAIC	$d(x)$	$g(x)$	$\hat{P}$	$\hat{P}_c$	$\hat{D}$	$\hat{D}_{150}$
M1	-	-	66.45,−0.11	0.42	-	$_{0.71}1.24_{2.15}$	$_{0.71}1.24_{2.15}$
M2	3.83 (0.126)	54.2	44.8	0.075	0.626	$_{4.99}8.673_{14.41}$	$_{10.01}19.249_{35.02}$
M3	7.87 (0.017)	106.3,8.1	51.4	0.166	0.723	$_{2.19}3.382_{7.46}$	$_{2.68}4.437_{12.60}$
M4	0 (0.857)	134.2,54.7	47.1	0.106	0.671	$_{2.95}5.727_{13.54}$	$_{3.98}10.074_{40.38}$

95% confidence interval for density is wide, reflecting poor precision.

5. Discussion

The results presented show that if DS data are analyzed using conventional methods when points are located along a linear feature, substantial bias may result due to nonuniform distribution of the animals with respect to the linear feature. Our methods are useful, for example, when point transects are placed along roads or rivers that affect animal distribution. This commonly occurs in published studies (Ruette et al., 2003), but the possible impacts of nonuniform distribution on population estimates are seldom assessed. Our methods assume that the detection function is independent of θ. If that is not the case, $d(x)$ and the detection function are not separately estimable using the proposed methods. Note that this creates problems when sectors of the circle around each point that are closer to the road are more or less visible than sectors away from the road. Marques (2007) presents such an example and a way to deal with it.

The simulations show that the methods work better in some cases than others. They should work best when the true underlying P is around 0.15–0.20 or higher, due to inherent positive bias of HTL estimators for small P. (More specifically, the HTL estimator bias is a function of the coefficient of variation in the P estimate; for small values of P, the relative precision is usually poor.) Hence, the new method should work better for animals that tend to have higher density near the linear feature than away from it. If sample size is large, as in the hare survey example, the bias in P, and therefore in D, will be smaller.

For the hare survey example, despite previous knowledge about the species and the data suggesting that CDS methods should not be used, the goodness-of-fit statistics associated with the CDS analysis were nonsignificant (at the usual 5% significance level). However, a Kolmogorov–Smirnov test of the uniformity of the sighting angles distribution (expected if the uniform density holds) identifies a nonuniform distribution of animals with respect to the road (in this case $p < 10^{-4}$), reinforcing the suggestion that sighting angles have relevant information in point transect surveys (e.g., Buckland et al., 2001, p. 275). (Note that these might also allow identification of responsive movement.) Hence we recommend that in point transect surveys, sighting angles should be recorded and analyzed.

Because detection probabilities seem close to 0 at 150 m, an analysis was implemented for different truncation values, from 99 to 147 m, in 3 m intervals (analysis not shown). As more severe truncation is used, estimates from the different models become more similar, making estimation less dependent on the particular model used. Reducing the truncation distance introduces other problems, as there is a loss of precision and it becomes less likely that the nonuniformity caused by the presence of the linear feature has disappeared by the truncation distance. Hence the choice of an appropriate truncation distance for our methods might be a more influential decision than for conventional methods.

There are several possible options when there is a density gradient with respect to a given linear feature. Notwithstanding logistical constraints in the field, the ideal approach remains to avoid the problem by design; place line transects either perpendicular to the linear feature, or randomly with respect to the feature so that animals are distributed uniformly with respect to distance from the transect. Another strategy is to use multiple independent observers to estimate directly the probability of detection (as a function of distance, and of other covariates), and use an HTL estimator to estimate abundance (Borchers et al., 1998). This approach does not require a model for estimation of $d(x)$, and can in fact be used to estimate $d(x)$ (Laake and Borchers, 2004, p. 184), but for many surveys, ensuring independence of observers or controlling for potential increased levels of response to multiple surveyors might not be feasible.

We focused here on the estimation of abundance in the vicinity of the linear feature, but usually one is interested in making inferences over much wider areas. Estimation can be seen as a two-stage process: (1) estimation of the probability of detecting an animal given that it is in the covered area (model based) and (2) estimation of abundance in the wider survey region (usually design based), given estimates from (1). Our methods can model $D(x)$, via $d(x)$, as a function of distance from the linear feature. This can be used in the second stage for model-based inference beyond $x = w$, although estimation for uncovered areas along the road, within w of the road, is still design based.

Investigators conducting DS studies in which location of samplers is not random with respect to the animals' locations should routinely assess whether animals are distributed independently of the line or point. If nonrandom samplers are used, this should be clearly stated when reporting results, and the possible implications for the results, given the characteristics of the study species, discussed.

The work developed here suggests further investigation. Although the simulation scenarios considered are illustrations of the potential bias, the bias will depend on the true $d(x)$. Investigation into other realistic models to describe the density gradients in real studies, as well as into nonparametric approaches to estimate $d(x)$, would be useful. The hare data demonstrate the potential difficulties in applying our methods to real data: alternative models for the nonuniform density lead to considerable variation in the results. An approach that incorporates model uncertainty in the estimates might be used to account for this. This can be done using a bootstrap procedure, in which the choice among the competing models is done for each bootstrap resample (Buckland, Burnham, and Augustin, 1997). Additionally, independent information on $d(x)$ will help to differentiate between candidate models. Extending the likelihood presented to include data from transects perpendicular to the road (and therefore sampling directly from $d(x)$) is straightforward, and should lead to more reliable estimates, helping to separate the shapes of $g(r)$ and $d(x)$. Related methods for line transect sampling are described in Marques (2007) and Buckland et al. (2007). Other models rather than the HN were not investigated for $g(r)$. Further investigation is needed to assess the impact of a different true detection function, and how easy it is to choose among different candidate models, when applying the methods to real data sets. As stressed by a reviewer, another area that needs additional work is the evaluation of the proposed bootstrap variance estimators, but this is too computer intensive to be included as part of the present study. We dealt here with point transects. If density is markedly different in the vicinity of the linear feature, then substantial bias in abundance estimates is anticipated if line transects are placed along it. Further research is required to deal with this case, because the sighting angles in line transect sampling convey information both on $d(x)$ and on the search process, and such information cannot easily be disentangled, except by imposing additional assumptions on the search pattern.

Acknowledgements

TAM thanks Dinis Pestana for guidance, and was sponsored by FCT and FSE, III Quadro Comunitário de Apoio (SFRH/BD/10812/2002), and Centro de Estatística e Aplicações da Universidade de Lisboa. Luisa Chaves' comments contributed to the readability of the manuscript. We thank S. Brown, J. Preston, A. Portig, and W. I. Montgomery for assistance and advice regarding the hare survey example. The hare survey work was supported by the Quercus partnership between the Northern Ireland Environment Agency and Queen's University Belfast.

References

Borchers, D. L. (1996). Line transect estimation with uncertain detection on the trackline. Ph.D. Thesis, University of Cape Town, Cape Town.

Borchers, D. L. and Burnham, K. P. (2004). General formulation for distance sampling. In *Advanced Distance Sampling*, S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas (eds), 6–30. Oxford: Oxford University Press.

Borchers, D. L., Buckland, S. T., Goedhart, P. W., Clarke, E. D., and Hedley, S. L. (1998). Horvitz-Thompson estimators for double-platform line transect surveys. *Biometrics* **54,** 1221–1237.

Borchers, D. L., Buckland, S. T., and Zucchini, W. (2002). *Estimating Animal Abundance*. London: Springer.

Borralho, R., Rego, F., and Pinto, P. V. (1996). Is driven transect sampling suitable for estimating red-legged partridge *Alectoris rufa* densities? *Wildlife Biology* **2,** 259–268.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: An integral part of inference. *Biometrics* **53,** 603–618.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. (2001). *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford: Oxford University Press.

Buckland, S. T., Borchers, D. L., Johnston, A., Henrys, P. A., and Marques, T. A. (2007). Line transect methods for plant surveys. *Biometrics* **63,** 989–998.

Fewster, R. M., Southwell, C., Borchers, D. L., Buckland, S. T., and Pople, A. R. (2008). The influence of animal mobility on the assumption of uniform distances in aerial line transect surveys. *Wildlife Research* **35,** 275–288.

Heydon, M. J., Reynolds, J. C., and Short, M. J. (2000). Variation in abundance of foxes (*Vulpes vulpes*) between three regions of rural Britain, in relation to landscape and other variables. *Journal of Zoology* **251,** 253–264.

Laake, J. L. and Borchers, D. L. (2004). Methods for incomplete detection at distance zero. In *Advanced Distance Sampling*, S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. Thomas (eds), 108–189. Oxford: Oxford University Press.

Louis, T. A. and Zeger, S. L. (2009). Effective communication of standard errors and confidence intervals. *Biostatistics* **10,** 1–2.

Marques, T. A. (2007). Incorporating measurement error and density gradients in distance sampling surveys. Ph.D. Thesis, University of St. Andrews, St. Andrews, U.K.

Marques, F. F. C. and Buckland, S. T. (2003). Incorporating covariates into standard line transect analyses. *Biometrics* **59,** 924–935.

Marsh, D. M. and Beckman, N. G. (2004). Effects of forest roads on the abundance and activity of terrestrial salamanders. *Ecological Applications* **14,** 1882–1891.

Ortega, Y. K. and Capen, D. E. (2002). Roads as edges: Effects on birds in forested landscapes. *Forest Science* **48,** 381–390.

Ruette, S., Stahl, P., and Albaret, M. (2003). Applying distance-sampling methods to spotlight counts of red foxes. *Journal of Applied Ecology* **40,** 32–43.

Thomas, L., Buckland, S. T., Rexstad, E. A., Laake, J. L., Strindberg, S., Hedley, S. L., Bishop, J. R. B., Marques, T. A., and Burnham, K. P. (2009). Distance software: Design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology*, doi: 10.1111/j.1365-2664.2009.01737.x.

Ward, A. I., White, P. C. L., and Critchley, C. H. (2004). Roe deer *Capreolus capreolus* behaviour affects density estimates from distance sampling surveys. *Mammal Review* **34,** 315–319.

Yost, A. C. and Wright, R. G. (2001). Moose, caribou, and grizzly bear distribution in relation to road traffic in Denali National Park, Alaska. *Arctic* **54,** 41–48.

Received January 2009. *Revised November* 2009.
Accepted November 2009.