Biometrics 60, 757–763
September 2004

Predicting and Correcting Bias Caused by Measurement Error in Line Transect Sampling Using Multiplicative Error Models

Tiago A. Marques

Centro de Estatística e Aplicações da Universidade de Lisboa, Bloco C2, Campo Grande, 1749-016 Lisboa, Portugal
Centre for Research into Ecological and Environmental Modelling, The Observatory,
University of St. Andrews, St. Andrews KY16 9LZ, Scotland
email: tiago@mcs.st-and.ac.uk

Summary. Line transect sampling is one of the most widely used methods for animal abundance assessment. Standard estimation methods assume certain detection on the transect, no animal movement, and no measurement errors. Failure of the assumptions can cause substantial bias. In this work, the effect of error measurement on line transect estimators is investigated. Based on considerations of the process generating the errors, a multiplicative error model is presented and a simple way of correcting estimates based on knowledge of the error distribution is proposed. Using beta models for the error distribution, the effect of errors and of the proposed correction is assessed by simulation. Adequate confidence intervals for the corrected estimates are obtained using a bootstrap variance estimate for the correction and the delta method. As noted by Chen (1998, *Biometrics* **54,** 899–908), even unbiased estimators of the distances might lead to biased density estimators, depending on the actual error distribution. In contrast with the findings of Chen, who used an additive model, unbiased estimation of distances, given a multiplicative model, lead to overestimation of density. Some error distributions result in observed distance distributions that make efficient estimation impossible, by removing the shoulder present in the original detection function. This indicates the need to improve field methods to reduce measurement error. An application of the new methods to a real data set is presented.

Key words: Beta distribution; Distance sampling; Line transect sampling; Measurement error; Multiplicative models.

1. Introduction

Biologists and managers face a rising need to assess animal abundance efficiently. One of the most widespread methodologies used for this purpose is line transect sampling. In this methodology, transects of total length L are traversed by one or more observers who record the distances from the line to every animal sighted. Using these distances, an estimator of animal density is given by

$$\hat{D} = \frac{n\hat{f}(0)}{2L}, \tag{1}$$

where n is the number of animals sighted and $\hat{f}(0)$ is an estimator of the probability density function (p.d.f.) of detected distances, evaluated at distance 0. Consider the detection function $g(x)$, defined as the probability of detecting an animal at a distance x from the line. Assuming animals are uniformly distributed with respect to the lines, we can think of $f(x)$ as a rescaling of the detection function. The estimation procedures are described in detail by Buckland et al. (2001).

This estimator is consistent if several key assumptions are met. Provided transect lines are randomly allocated, independently of animal distribution, these assumptions are: (1) animals on the transect line are detected with probability 1; (2) there is no animal movement in response to the observers prior to detection; and (3) measurements are exact. Strictly speaking, sightings should also be independent, but this assumption can be relaxed if robust procedures of variance estimation are adopted (Buckland et al., 2001). The effects of violating the first two assumptions, and ways to deal with them, have been the subject of recent work (e.g., Lake, 1978; Buckland and Turnock, 1992; Quang and Becker, 1997; Borchers et al., 1998a,b). The third assumption has been addressed with less emphasis. The work of Chen (1998) was the first to deal exclusively with it, and he concluded that even a random additive error, with zero mean, leads to an underestimation of density, and that this effect cannot be reduced by increasing the sample size. Based on some knowledge of the error distribution, and using the method of moments, he derived a corrected estimator for the density. The model proposed by Chen can be seen as a particular case of the model proposed by Alpizar-Jara (1997). This author, using an approach based on the SIMEX algorithm (Cook and Stefanski, 1994), also presented a way of deriving a corrected estimate of density. Recently, in the context of line transects where the objects of interest are groups, Chen and Cowling (2001) dealt with the simultaneous effects of errors in distances and group sizes. Schweder (1996, 1997) dealt with measurement errors in radial distances, in relation to cue-based methods for estimating minke whale abundance. Hiby, Ward, and Lovell (1989) developed a method to account for error measurement in cue-counting methods using grouped data, by incorporating a measurement error function, with unbiased multiplicative Gaussian errors, in the estimation of the detection function.

In this article, I discuss the ways in which errors are generated in the context of line transects and, under certain conditions, propose a multiplicative error model for them. The effect of such errors on the density estimator is evaluated and a corrected estimator of density is derived (Section 2.1). I describe two special cases of beta models, and using them as examples, look at the effects on the estimation procedure (Section 2.2). Results are illustrated with simulations of animal populations of known density (Section 3). An application of the methods is given in Section 4 and results are discussed in Section 5.

2. Modeling Measurement Errors

2.1 *Predicting the Effect of Multiplicative Errors*

Let an error be defined as $X - Y$, where X is the true distance, and Y is the estimated distance. In a distance sampling context, four types of error can be identified: (1) recording/data handling errors, (2) rounding errors, (3) biased random errors, and (4) unbiased random errors. The first kind we hope to identify and remove before using the data for estimation. The second is a special case of errors common in distance sampling, where some convenient values, such as 0, 10, 50, 100, etc., are consistently preferred. If rounding is substantial, especially at zero, reliable estimation becomes impossible. This work does not address these two cases. The other two types, random errors, arise for several different reasons and can potentially lead to bias. While the effect of biased distance estimation is more pronounced and leads directly to biased abundance estimation, bias in distance estimation can often be easily removed by estimating it from experiments. Unbiased errors can nonetheless cause substantial bias that is difficult to remove. This work presents an approach that can simultaneously correct for biased and unbiased errors.

Random errors arise from the inability to record precise distances. If the resulting error associated with a given distance is independent of the original distance, an additive model, $Y = X + R$, where the error distribution, R, is independent of X, might be adequate to deal with it. This is the case presented by Chen (1998), where errors arise from uncertainty in GPS recordings. No measurement is exact, so there is always some kind of additive error in any distance measurement, but given proper field methods it is plausible to assume that this additive error is negligible, compared with other possible sources of bias. On the other hand, if the evaluation of a distance becomes more uncertain as that distance increases, a multiplicative error model might be adequate, accounting for the fact that the resulting error will tend to be proportional to the original true distance. This is likely to be the case if distances are visually estimated, a common procedure in distance sampling studies, although not encouraged (Buckland et al., 2001, p. 265). For such situations, I propose a multiplicative model for the error

$$Y = XR,$$

where X and R are assumed independent. This plausible form for the effect of the error allows a simple correction for the bias in density estimation that might be introduced by the errors.

As can be seen from (1), we are interested in estimating the value of the p.d.f. of detected distances at 0. If we use the observed distances (Y_i, $i = 1, 2, \ldots, n$), we estimate $f_Y(0)$, but what we really want is the p.d.f. of true distances, $f_X(0)$. With the proposed error model, $f_X(0)$ is proportional to $f_Y(0)$. As R and X are independent and $R \geq 0$, the distribution of the observed distances with errors is

$$f_Y(y) = \int_0^{+\infty} f_X\left(\frac{y}{r}\right) \frac{f_R(r)}{r}\, \mathrm{d}r,$$

a standard random variable transformation result, where $f_R(r)$ is the p.d.f. of errors. Since we are interested in the value of $f_Y(0)$, we get

$$f_Y(0) = f_X(0) \int_0^{+\infty} \frac{f_R(r)}{r}\, \mathrm{d}r = f_X(0)K. \qquad (2)$$

This expression requires that $E(|R|^{-1})$ exists, reflecting restrictions to possible models for R, as for values that K might take for each assumed distribution of R. Given the distribution of R, K can be evaluated. If measurement errors are ignored, equation (1) yields a biased estimator of density ($\hat{D}_{\mathrm{e}}$),

$$\hat{D}_{\mathrm{e}} = \frac{n\hat{f}_Y(0)}{2L}.$$

From expression (2), a corrected estimator of density is

$$\hat{D}_{\mathrm{c}} = \frac{n\hat{f}_Y(0)}{2LK} = \frac{\hat{D}_{\mathrm{e}}}{K}. \qquad (3)$$

Following a suggestion from the associate editor, a suitable estimator for K^{-1} is the harmonic mean of a sample of R's,

$$\hat{K}^{-1} = \frac{1}{S}\sum_{s=1}^{S} \frac{1}{r_s}, \qquad (4)$$

where S is the number of observations with both true and error distances. This estimator has the advantage that it is unnecessary to assume a specific distribution for R, and its variance can be easily obtained by bootstrap, and then incorporated in the estimation procedure using the delta method, using standard software for analyzing line transect data such as `Distance 4` (Thomas et al., 2002).

The final corrected density estimator is therefore

$$\hat{D}_{\mathrm{c}} = \frac{\hat{D}_{\mathrm{e}}}{\dfrac{1}{S}\displaystyle\sum_{s=1}^{S} \frac{x_s}{y_s}} = D_{\mathrm{e}} M_{\mathrm{H}}(r_s), \qquad (5)$$

where $M_{\mathrm{H}}(r_s)$ represents the harmonic mean of a sample of r's.

2.2 *Models for the Error*

Although K can be estimated without the need to assume a specific distribution for the errors, it is necessary to do so in order to evaluate the effect of different error structures in the density estimator. As a first approach, we assume two models for the distribution of R

$$\text{model I: } R = (0.5 + U), \quad \text{model II: } R = 2U,$$

where $U \sim Be\ (\theta_1, \theta_2)$, $\theta_1, \theta_2 \in (0, +\infty)$. Models with both parameters less than 1 are not useful, since situations in which these error distributions would arise are unlikely. We consider only the case where θ_1 and $\theta_2 \geq 1$. Under these models, a symmetric beta results in unbiased estimation of distances. The estimated distance takes the maximum value of 1.5 (model I) or 2 times (model II) the original value. The resulting errors are therefore dependent on the values of the original observations, although X and R are independent. The beta family contains a wide choice of shapes, and depending on parameter values, $E(X)$ can be larger, equal, or lower than $E(Y)$, therefore allowing simulation of a range of different situations.

Even if the observations are unbiased estimators of the distances, the density estimator may be biased. This is due to the fact that observations are unbiased if $\theta_1 = \theta_2$, and the estimator of population density is asymptotically unbiased if $K = 1$, but neither condition implies the other.

Expression (2) can be developed for both models. For model I

$$f_Y(0) = f_X(0) \int_0^{+\infty} \frac{f_R(r)}{|r|}\, \mathrm{d}r = f_X(0) \int_0^{+\infty} \frac{f_U(r - 0.5)}{r}\, \mathrm{d}r$$

$$= f_X(0) \int_{0.5}^{1.5} \frac{1}{B(\theta_1, \theta_2)} \frac{(r - 0.5)^{\theta_1 - 1}(1.5 - r)^{\theta_2 - 1}}{r}\, \mathrm{d}r \quad (6)$$

and for model II

$$f_Y(0) = f_X(0) \int_0^{+\infty} \frac{f_R(r)}{|r|}\, \mathrm{d}r = f_X(0) \int_0^{+\infty} \frac{\frac{1}{2} f_U(r/2)}{r}\, \mathrm{d}r$$

$$= f_X(0) \int_0^{2} \frac{1}{B(\theta_1, \theta_2)} \frac{1}{2} \frac{(r/2)^{\theta_1 - 1}(1 - r/2)^{\theta_2 - 1}}{r}\, \mathrm{d}r.$$

For model II, substituting $t = r/2$ and simplifying leads to

$$= f_X(0) \frac{1}{2} \frac{(\theta_1 + \theta_2 - 1)}{(\theta_1 - 1)}. \quad (7)$$

With available information on R, we can evaluate the bias introduced by the errors and derive corrected estimators. In practice, the distribution of R will not be known, but it can be estimated using experiments with objects placed at known distances. Alternatively, it may be possible to record some distances both using an accurate method and the less precise method used on the whole survey.

K can be calculated from $f_U(u)$. Surfaces for models I and II are presented in Figure 1. We can see that the effect of the error is more pronounced for model II for the same values of the parameters. In both cases, density estimates are positively biased if the error process is unbiased ($\theta_1 = \theta_2$). In some cases, $E(Y) > E(X)$ results in positively biased density estimates. This reflects the impact of the specific form of error distribution on the estimation process. It is interesting to note that for some values of the parameters, $K = 1$, even though $E(Y) > E(X)$.

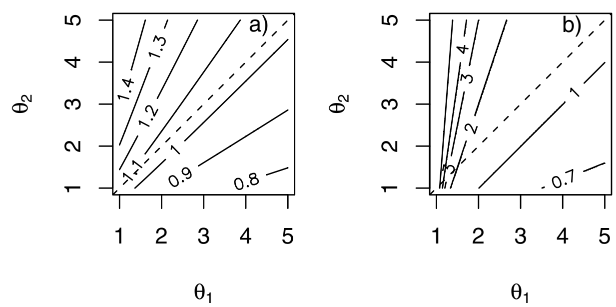

Figure 1. Values of K, under model I (a) and model II (b), for small values of the parameters of U. The dashed line indicates unbiased estimation of distances ($\theta_1 = \theta_2$). Areas above and below the line correspond respectively to underestimation and overestimation of distances. $K < 1$ corresponds to uncorrected density estimates being underestimated, and $K > 1$ corresponds to uncorrected density estimates being overestimated.

An alternative for models I and II would be to consider $R \sim Ga(\alpha, \beta)$

$$f_R(r) = \frac{x^{\alpha - 1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)}, \quad \alpha, \beta > 0, \quad r > 0.$$

We can show that $E(|R|^{-1}) = 1/\beta(\alpha - 1)$, and therefore a simple estimator for K is $1/\hat{\beta}(\hat{\alpha} - 1)$, and using (3) a corrected estimator for density is easily obtained. As a referee pointed out, the gamma model is more elegant and transparent than models I and II above. The beta models are, however, easier to use in simulations as there is a simple relation between parameter values and under- or overestimation of distances.

3. Simulation Study

A simple simulation study was performed to evaluate the methods with populations of known density.

A population of 10,000 animals was randomly generated on a square with side 1000 m, giving a density of 100 animals per hectare. The study area was divided into 25 nonoverlapping squares, and in each of these squares a transect of 200 m was randomly selected. In each row of squares the transect was randomly generated for the first square and in the subsequent squares it was systematically placed with respect to the first one. In analysis, a truncation distance of 10 m was used. To avoid edge effects, only transects placed at more than 10 m from the edge of the square were considered. The detection probability of each animal (p_i) was calculated, using the perpendicular distance to the nearest transect, based on a half-normal detection function ($\sigma = 5$). For each animal, a random number between 0 and 1 (u_i) was generated and the animal was detected if $p_i > u_i$. This process was repeated 100 times, resulting in 100 independent data sets. The average number of animals detected in each "survey" was 593, standard deviation 20.1. The data generated were considered to be error free.

I then generated errors with the following distributions: beta(1,1), beta(3,2), and beta(5,5). The choice for these particular models was arbitrary, but had a rationale. The distribution beta(1,1) was used as an extreme case, beta(5,5) as estimation of distances is unbiased, but density estimation is biased, and beta(3,2) as estimation of distances is biased, but nonetheless estimation of density should be unbiased. For each data set without error, one error set was independently generated, and introduced as postulated above (for models I and II), giving five contaminated sets. (The case of beta(1,1) for model II was not implemented, since K would be infinite.)

To preclude analyst influence, the data sets were analyzed using a standard analysis in `Distance 4` (Thomas et al., 2002), as described below. The models considered for the detection function were half-normal+cosine, uniform+cosine, and hazard rate+simple polynomial, and the one with lowest AIC selected. The variance for encounter rate was calculated analytically based on replicate lines. In the analysis of contaminated data, the largest 5% of distances was truncated, as otherwise some models required several adjustment terms to provide an adequate fit of the data.

The analysis of the error-free data led to an average estimated density of 98.6 animals per hectare, with a standard error of 0.65. The actual coverage for the 95% CI was 93%.

For the contaminated data sets, only 23 transects were used to estimate density, and the remaining two were used as a separate experiment, where true and contaminated distances were evaluated. This resulted on average on 516.2 (standard deviation 19.4) observations to estimate density and 49.6 (standard deviation 7.1) observations to estimate K. K was estimated using the harmonic mean estimator on the sample of R's resulting from the two transects. The variance of K estimates was obtained by bootstrap (999 resamples), and the appropriate variance for corrected density, incorporating this extra variance component, was obtained using the delta method.

The true, mean estimated (based on the nonparametric estimator and the appropriate beta model—by maximizing a beta likelihood of the R's and then evaluating equations [6] or [7], depending on the model used to generate the errors), and mean observed K for each combination of model and error is presented in Table 1. Also shown is the coverage of the 95% confidence interval, based both on corrected and uncorrected analysis. The nonparametric estimator for K and the parametric beta-based estimator showed no differences, justifying the nonparametric estimator when the true model is unknown. An increased coverage with the use of the proposed correction is present in all cases. Figure 2 shows the error-based density estimates and the corrected density estimates using the harmonic mean estimator, showing that the correction reduced the bias in most cases. It can be seen that the results were very close to the expected ones, validating the effects of errors predicted and the proposed correction. However, in some cases, true K and observed K were slightly different. Especially in the case of beta(3,2) for model II, a K of 1 was expected but an average K of 0.952 was obtained. These unexpected results will be considered in Section 5. Notice that even in this case coverage was increased, due to an increased variance related to the estimation of K.

4. Practical Application of the Correction

I considered a study area of 1677.12 m^2 with 250 golf tee groups randomly distributed, resulting in a density of 0.149 tees/m^2. Two strata, with different abundances ($N = 130$ and $N = 120$, in areas of, respectively, 1057.12 and 620 m^2), were surveyed for golf tees by eight independent observers, which were considered as a single-pooled observer, resulting in 125 sightings. The original tee data set includes group size, color, and visibility, but these were ignored for the purpose of this study. I am in fact estimating tee-group density, but that is irrelevant for this application, and in the following I will refer only to tees and tee density. There were, respectively, six and five transects in each stratum, and the width of the transects was 4 m. Further details of this data set can be found in Borchers, Buckland, and Zucchini (2002). Initial analysis of the data reveals a serious $g(0)$ problem. As $g(0)$ problems are a side issue in this work, I simply estimated it (and its variance) from the data and used it as a multiplier (Buckland et al., 2001, p. 57) in `Distance 4` (Thomas et al., 2002). In order to mimic real life applications, I assumed that one transect in each stratum (transects 2 and 10, with a total of 22 detections) was part of an experiment to collect data on measurement error, in order to estimate K. Thus, estimated distances (mean for those observers who saw each tee) and real distances were available for 22 observations. The remaining nine transects (with 103 detections) were used in the usual way to derive a density estimate. Using `Distance 4`, estimates of density were obtained, both considering true distances and distances with errors (Table 2). The errors led to an overestimation of density of 16.8%. Using the harmonic mean of R, K was estimated as 1.123, and therefore a corrected point estimate of density is, using expression (3), 0.155. A bootstrap

Table 1

True K (TK), mean estimated K using the harmonic mean (EHK) or the true beta model (EBK), and mean observed K (OK), under the combinations of errors and models (I—model I, II—model II) considered. Density (D), density coefficient of variation (DCV), and Coverage of 95% CI (C 95%), for corrected and uncorrected analysis. Mean estimated density based on true distances is 98.6 animals/ha. True density is 100 animals/ha.

	TK	EHK	EBK	OK	D[a]	DCV[a]	C 95%[a]
Beta(1,1), I	1.099	1.105	1.104	1.078	96.5/106.4	7.46/5.83	89/81
Beta(5,5), I	1.024	1.021	1.021	1.030	99.4/101.5	6.76/6.76	95/94
Beta(5,5), II	1.125	1.126	1.125	1.096	96.4/108.0	8.69/6.17	89/78
Beta(3,2), I	0.944	0.943	0.943	0.941	98.5/92.8	6.83/6.13	94/75
Beta(3,2), II	1.000	1.008	1.007	0.952	94.2/93.9	11.04/6.68	90/75

[a]Corrected analysis/uncorrected analysis.

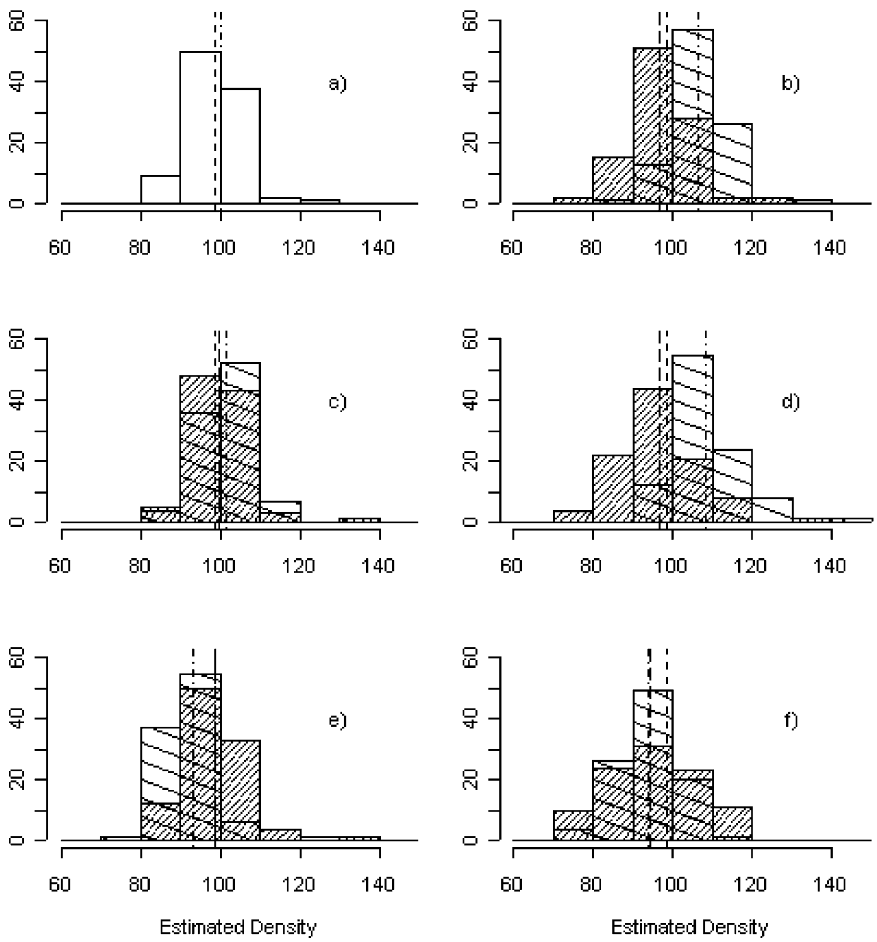

Figure 2. Error-based density estimates (lighter histograms) and the corrected estimates (darker histograms) using the harmonic mean estimator. (a) True distances. (b) Error beta(1,1), model I. (c) Error beta(5,5), model I. (d) Error beta(5,5), model II. (e) Error beta(3,2), model I. (f) Error beta(3,2), model II. Dash line—mean value of true distance estimates. Dash-dot line—mean value of error distance estimates. Long dash line—mean value of corrected error distance estimates. True $D =$ 100 animals/ha.

variance for K (999 bootstrap samples) was calculated, and assuming K and D_e independent, we used $\hat{K}$ as a multiplier in the analysis. As can be seen from Table 2, the corrected results now lie closer to the true values; 72% of the error bias in density was removed by using the correction, and the impact on precision was negligible.

Although I have used the harmonic mean to estimate K, assuming either a beta or a gamma model leads to very similar estimates for K, respectively, 1.115 and 1.117.

5. Discussion

In practical applications of distance sampling, underestimation or overestimation of density are frequently reported to be related to assumption violations, such as undetected animals on the transect (e.g., Anderson et al., 2001) or animal movement (e.g., Southwell, 1994; Langbein et al., 1999). However, I am not aware of observed bias being attributed to measurement errors, although several authors recognize its influence in the estimation procedures (e.g., Alpizar-Jara, 1997; Chen, 1998; Buckland et al., 2001). This fact alone is an indication of how the effect of errors is underestimated or neglected when considering distance sampling, and demonstrates the need to study the problem.

Chen (1998), using an additive model, concludes that when estimation of distances is unbiased, errors lead to underestimation of density. This work shows that with a multiplicative error unbiased estimation of distances leads to overestimation of density, a fact that shows that the model relating contaminated and error-free data is responsible for a different effect on the final estimates. As with Chen's error model, the effect

Table 2

Results of the analysis performed in Distance 4 considering the true distances, distances with errors (uncorrected estimator), and distances with errors with the proposed correction. True $D = 0.149$.

Analysis	$\hat{D}$	DCV	D 95% CI	RMSE
True distances	0.149	0.195	(0.100,0.222)	0.029
Error distances—D_e	0.174	0.195	(0.117,0.260)	0.042
Error distances—D_c	0.155	0.217	(0.100,0.238)	0.039

of the error cannot be reduced by increasing sample size. As unbiased errors lead to bias, when testing to evaluate whether errors influence the estimation process, it is not sufficient to see whether there are significant differences between mean actual and estimated distances (as in Borralho, Rego, and Pinto, 1996); rather we should test whether density estimates calculated from these two data sets are significantly different (as in Heydon, Reynolds, and Short, 2000).

Another consequence of the errors is the change produced in the detection function. The model selected by the software when considering the true versus the contaminated distances was often different, a fact also detected by Alpizar-Jara (personal communication), generally needing more complex models in the latter case. It was interesting to notice that the resulting detection function after the introduction of errors (the one you would actually see based on field data) can lack the shoulder that it originally had, as illustrated in Figure 3. This is an unexpected consequence of the errors. Depending on the effect of errors, the resulting detection function can have a shape that leads to unstable estimation even when the detection process itself is smooth and has a shoulder.

The use of a beta for the error model should not be interpreted as supporting the fact that this model is likely to best represent reality. Other models, such as gamma, or even Gaussian or log-normal, might be appropriate under certain conditions. The use of the beta is justified as a flexible model to generate error distributions with intuitive characteristics. Field trials, where error-free and contaminated distances are recorded, are needed to pursue this matter further. Nonetheless, the fact that an estimator that assumes no distribution for R is available is reassuring.

The problem identified in the simulation study, not only but especially regarding the beta(3,2), model II, where the 95% CI for mean K did not include the predicted K, might be related to two different aspects of the analysis. The first one is the fact that when we consider the error introduced under models I and II, the latter error can severely change the shape of the observed distance distribution. This fact is immediately understood by a graphical analysis of the resulting distances under the two distinct models. If we consider, as an example, the distribution of exact distances (Figure 3a) and contaminated distances (Figure 3b and 3c), we can see that the shape of observed distances under model II (Figure 3c) closely resembles a negative exponential. Since the exponential was not selected as a candidate to model the distances in the simulation study, we could never obtain the changes in density that would be expected theoretically. Adding to this fact, we have to consider also that the true K is asymptotically derived, in the sense that it would be completely expressed if the sample size was infinite.

Using a $\hat{K}$ based on separate field trials, where conditions might differ from the actual survey, might not be entirely appropriate, as it is usually argued that actual survey conditions should lead to greater variability in the estimation of distances (e.g., Schweder, 1996, 1997). Therefore, when performed, these trials should, as much as possible, mimic true survey conditions. If in a real life application, the data on the errors are collected during the actual survey, and common data are used, then a bootstrap procedure that takes this into account could be used. Problems might arise here as the sample size to estimate K might be too small if care is not taken. If one specific transect is used to estimate K, and then it is not selected in a resample, there are no data available to estimate K. Therefore, if a joint experiment is conducted, we should make an effort to have a subset of observations with truth and error distances in most transects. The minimum number of such distances in order to achieve good results was not investigated, but an educated guess calls for a sample of at least 30 observations, randomly chosen during the survey.

In the golf tee example, we were able to reduce the error-induced bias by 72%. The natural price to pay is the increase in variance (see Table 2) due to the fact that an extra parameter (K) is being estimated. Nonetheless, as can be in this case, in applications of the methods it is unlikely that this leads to a large decrease in precision, because the main source of variation in $\hat{D}_c$ is still the one associated with the variance of encounter rate.

The effect of errors in line transect sampling is not as important as in point transect sampling or cue counting (Buckland et al., 2001). However, this work shows that it can be the source of important bias even in line transect sampling, therefore establishing that further work is needed to evaluate and correct the effects of errors on estimates of density in point transect sampling and cue counting. This and similar work might help to establish procedures that eliminate or reduce the effect of errors in distance sampling estimates. Nonetheless, emphasis should be made that the appropriate procedure would be to take every measure possible to minimize errors at the sampling stage, not to rely on analysis that corrects for these errors. Therefore, training and calibrating observers, a

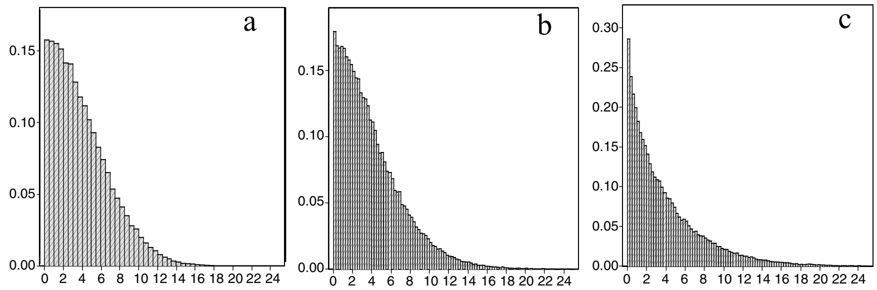

Figure 3. Histograms with 100,000 observations, considered as adequate representations of the underlying distribution. (a) No error (*Half-normal*, $\sigma = 5$). (b) Error beta(1.5,1.5), model I. (c) Error beta(1.5, 1.5), model II.

precise definition of the transect line, and better technology to measure distances should always be considered.

Acknowledgements

I thank Steve Buckland, David Borchers, Silvio Velosa, Dinis Pestana, and Luisa Chaves for valuable comments and suggestions that improved the quality of this work, and Miguel Bernal for the use of the golf tees data set, collected as part of his Masters project. Two referees and an associate editor provided comments that greatly improved this manuscript. This research was sponsored by FCT and FSE, III Quadro Comunitário de Apoio (SFRH/BM/2068/2000 and SFRH/BD/10812/2002), and Centro de Estatística e Aplicações da Universidade de Lisboa.

Résumé

L'échantillonnage sur transect linéaire est une des méthodes les plus utilisées pour estimer l'effectif d'une population animale. Les méthodes d'estimation standard supposent une détection certaine le long du transect, aucun mouvement animal, et aucune erreur de mesure. Le non-respect de ces hypothèses peut générer des biais importants. Nous étudions ici l'effet d'erreurs de mesure sur les estimateurs d'abondance. A partir du processus générant des erreurs, nous présentons un modèle d'erreur multiplicatif, et la connaissance de la distribution de l'erreur nous permet de corriger simplement les estimateurs. Les effets des erreurs et de la correction proposée sont ensuite étudiés par simulations, en générant des erreurs selon une loi beta. Nous utilisons un estimateur par bootstrap de la variance, et la méthode delta, afin de générer des intervalles de confiance corrects pour les estimateurs corrigés. Comme le remarque Chen (1998, *Biometrics* **54,** 899–908), même des estimateurs sans biais des distances peuvent conduire à des estimateurs biaisés de la densité, selon la distribution vraie de l'erreur. Contrairement aux conclusions de Chen, qui utilisait un modèle additif, des estimations sans biais des distances, avec un modèle multiplicatif, résultent en une surestimation de la densité. Certaines distributions de l'erreur résultent dans des distributions de distances observées qui empêchent une estimation efficace, en gommant le plateau qui existe dans la fonction de détection originale. Il est donc nécessaire d'améliorer les méthodes de terrain, afin de réduire les erreurs de mesure. Nous appliquons les méthodes proposées à un jeu de données réel.

References

Alpizar-Jara, R. (1997). *Assessing Assumption Violation in Line Transect Sampling*. Ph.D. Thesis, North Carolina State University, Raleigh.

Anderson, D. R., Burnham, K. P., Lubow, B. C., Thomas, L., Corn, P. S., Medica, P. A., and Marlow, R. W. (2001). Field trials of line transect methods applied to estimation of desert tortoise abundance. *Journal of Wildlife Management* **65,** 583–597.

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

Borchers, D. L., Zucchini, W., and Fewster, R. (1998b). Mark-recapture models for line transect surveys. *Biometrics* **54,** 1207–1220.

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. and Turnock, B. J. (1992). A robust line transect method. *Biometrics* **48,** 901–910.

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

Chen, S. X. (1998). Measurement errors in line transect sampling. *Biometrics* **54,** 899–908.

Chen, S. X. and Cowling, A. (2001). Measurement errors in line transect surveys where detectability varies with distance and size. *Biometrics* **57,** 732–742.

Cook, J. R. and Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* **89,** 1314–1328.

Heydon, M. J., Reynolds, J., 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.

Hiby, L., Ward, A., and Lovell, P. (1989). Analysis of the North Atlantic sightings survey 1987: Aerial survey results. *Report of the International Whaling Commission* **39,** 447–455.

Laake, J. L. (1978). Line transect sampling estimators robust to animal movement. Master's Thesis, Utah State University, Logan.

Langbein, J., Hutchings, M. R., Harris, S., Stoate, C., Tapper, S. C., and Wray, S. (1999). Techniques for assessing the abundance of brown hares *Lepus europaeus*. *Mammal Review* **29,** 93–116.

Quang, P. X. and Becker, E. (1997). Combining line transect and double count sampling techniques for aerial surveys. *Journal of Agricultural, Biological and Environmental Statistics* **2,** 1–20.

Schweder, T. (1996). A note on a buoy-sighting experiment in the North Sea in 1990. *Report of the International Whaling Commission* **46,** 383–385.

Schweder, T. (1997). Measurement error models for the Norwegian minke whale survey in 1995. *Report of the International Whaling Commission* **47,** 485–488.

Southwell, C. (1994). Evaluation of walked line transect counts for estimating macropod density. *Journal of Wildlife Management* **58,** 348–356.

Thomas, L., Laake, J. L., Strindberg, S., Marques, F. F. C., Buckland, S. T., Borchers, D. L., Burnham, K. P., Hedley, S. L., and Pollard, J. H. (2002). `Distance 4.0`, Release 2. University of St. Andrews, U.K. Research Unit for Wildlife Population Assessment. Available at `http://www.ruwpa.st-and.ac.uk/distance/`.

Received May 2003. *Revised March* 2004.
Accepted March 2004.