A field test of surveyors’ influence on estimates in line transect sampling

Anna Ringvall a,*, Ganapati P. Patil b, Charles Tuillie b

a Department of Forest Resource Management and Geomatics, Swedish University of Agricultural Sciences, 901 83 Umeå, Sweden
b Center for Statistical Ecology and Environmental Statistics, Department of Statistics, Pennsylvania State University, University Park, PA 16802, USA

Received 25 March 1999; received in revised form 1 November 1999; accepted 8 November 1999

Abstract

Line transect sampling is a well-known method for estimating densities of wildlife populations and can be a useful alternative to the traditional methods for timber cruising when assessing objects and species of interest in relation to biodiversity. This paper presents results from a field test of line transect sampling for inanimate objects in boreal forest, focusing on the surveyors’ influence on estimates under different surveying conditions. The method was tested by 11 surveyors in northern Sweden in two forest types and for two different object types. An underestimation of between −0.3 and −22.2% was found, depending on the surveying conditions and the model used for the detection function. The underestimation is partly due to violation of the assumption that all objects on or very close to the survey line are detected. It was found that systematic differences among surveyors were generally moderate. At most, the coefficient of variation for surveyor effect was 6.8%. Still, this size of surveyor effect can be problematic for long-term monitoring, where systematic differences among surveyors’ estimates may indicate literal changes or disguise real trends. When comparing line transect sampling with other sampling methods, the increased variation in estimates caused by surveyors’ systematic and random errors as well as the variation caused by model selection must be considered. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Multisource forest inventory; Long-term monitoring; Surveyor bias

1. Introduction

Biodiversity is becoming increasingly important in forest management planning. This has resulted in a demand for data on substrates of importance for biodiversity as well as on populations of certain species. Inventories in this area are often based on subjective methods that are heavily dependent on surveyor judgements. Long-term monitoring and modelling, however, require data of known quality, and such data can only be obtained by probabilistic sampling methods. Plot-based methods of probability sampling are commonly used when estimating parameters related to timber production. However, the objects and species of interest in relation to biodiversity and nature conservation are often sparse, making plot-based methods inefficient. There is thus a need to find other inventory methods to get sufficiently good estimates for sparse objects. Strip transects have been found to be a competitive alternative to circular plots.

* Corresponding author. Tel.: +46-90-786-68-38; fax: +46-90-77-81-16. E-mail address: anna.ringvall@engcon.su.se (A. Ringvall).
when sampling such things as coarse woody debris (Lämmis and Fries, 1995). However, some sampling elements can be difficult to detect and missed objects will lead to an underestimation of the true density. In order to be sure that all objects in a strip will be detected, the strip width must be small, leading to high variability in estimates. In this situation, line transect sampling can be a good alternative to strip surveys (Burnham et al., 1985).

Line transect sampling (LTS) is a well-known method for estimating densities of wildlife populations (Buckland et al., 1993). In LTS a transect is surveyed and the distance to all objects detected from the line is recorded. Even if some objects around the line are missed, unbiased estimates can be made with the help of a so-called detection function, \( g(x) \), that gives the probability of detecting an object given its distance \( x \) from the line. This probability is assumed to decrease with increasing distance from the surveyor. However, the true detection function is unknown, and a major problem in LTS is finding a good model for estimating \( g(x) \) from the observed distances. LTS has found its major applications in assessments of wildlife populations, but the method can also be used to estimate the density of inanimate objects.

The theory behind LTS and the modelling of the detection function is well described in the literature (see Burnham et al., 1980; Seber, 1982; Buckland, 1985; Buckland et al., 1993). The detection function can be modelled in many different ways. Parametric approaches can use the half-normal and the exponential function, while non-parametric methods can include Fourier series and kernel estimators. The software has been developed for estimating densities, variances, and confidence intervals (Laake et al., 1993).

Monitoring demands high data quality. It is desirable to use inventory methods that do not depend on surveyors' judgements and that are consistent for different surveyors, environmental conditions, and types of objects to be surveyed. In LTS, the number of objects detected from a transect depends on such factors as the sighting conditions and the surveyor's experience, fatigue, and motivation. Still, LTS theory allows for unbiased estimates as long as some basic requirements are fulfilled (Buckland et al., 1993, pp. 30–35). These are: (i) that all objects on or very close to the transect line must be detected, i.e., \( g(0) = 1 \); (ii) that objects should be detected at their initial position; and (iii) that distance measurements are accurate. For robust estimation, the detection function should also have a shoulder, that is, the detection probability should not drop off too quickly with increasing distance from the survey line.

The performance of LTS and different methods for estimating the detection function have been evaluated in several studies spanning a broad range of conditions. By surveying artificial objects from walked transects in a sagebrush area, Laake (1978) compared the sensitivity of different detection functions to animal movements. Also using artificial objects, Bergstedt and Anderson (1990) evaluated the performance of a Fourier series for the detection function in an aquatic environment using an underwater camera, and found no evidence of bias. Hone (1988) tested several models for the detection function for estimating densities of pig carcasses from aerial surveys. He concluded that the assumption of a probability of detection of one on the survey line might also hold in aerial surveys. In contrast, when estimating densities of male deer (Odocoileus hemionus) from the air, negative biases were found for most models used for the detection function (approximately −6 to −25%), probably due to missed animals close to the survey lines (White et al., 1989). A large underestimation, due to birds' attraction to the surveyor, was found in a test of a Fourier series for estimating densities of bobolinks (Dolichonyx oryzivorus) (Bolinger et al., 1988). Southwell (1994) evaluated walked transects for estimating the density of macropods (Macropus spp.) using different models for estimating the detection function. A high variation among estimates using different detection functions was observed, with the bias ranging from −15 to 17%.

In summary, the results from these field studies show that violation of the assumptions of perfect detectability on the survey line and no movement prior to detection can cause severe over- or underestimation.

Although the field performance of LTS has been investigated in several studies, these have often been of a limited size and no study has really focused on the surveyors' influence on estimates. However, Laake (1978) found that bias was surveyor-dependent, although it was not the major objective of the study.

In long-term monitoring, many different surveyors will probably be involved in a study over time. It is
therefore very important that there be no differences among surveyors' systematic errors. The aim of the present study was to investigate the consistency of LTS estimates when surveying inanimate objects from walked transects. In particular, the study aimed to assess the differences among different surveyors' systematic error components under different surveying conditions. The field study was made in the boreal zone of Sweden under controlled conditions with artificial objects.

2. Methods

2.1. Field study

Four forest stands situated north of Sivar (63°39′N, 20°33′E) were chosen for the study. The stands were selected to represent two different forest types with varying sighting conditions. Two were young stands of dense Scots pine (Pinus sylvestris) and lodgepole pine (Pinus contorta), ~30 years old with limited undergrowth, which made the sighting good at ground level, but worse at higher levels. The other two stands consisted of old, mixed Norway spruce (Picea abies) and Scots pine forest with blueberry (Vaccinium myrtillus) and wild rosemary (Rhododendron tomentosum) in the ground layer. In these two stands, sighting conditions at ground level were limited, although conditions were better at higher levels.

The goal was to test line transect sampling for an inanimate population under controlled conditions. Because we had no specific object in mind, artificial objects were used. These were gorse-shaped boards measuring 25 × 35 cm, painted in two different colours: (i) olive green to imitate objects that are difficult to detect; and (ii) steel blue to imitate objects that are easier to detect.

The objects were placed along 1000-m long transect lines within a zone of 25 m on each side of the survey lines. The lines were marked to enable every surveyor to walk exactly the same line. Along each line 115 objects were placed, resulting in a density of 23 objects ha⁻¹. The objects were placed randomly within the 1000 × 50 m area around each transect, with the only restriction being that every 5-m interval (distance from the line) should contain the same number of objects (Fig. 1). The objects were placed directly or almost directly on the ground (0–0.2 m above ground). Every object was marked with a number and the surveyor recorded this number when the object was detected. The exact location of each unit was known, and consequently the distances to objects found were not measured during the fieldwork. Hence, there are no measurement errors in this data.

The different types of object (green and blue) were placed on separate survey lines. For each type of object and forest type, four transects were randomly laid out, for a total of 16 transects. Buckland et al. (1993, p. 14) suggest that at least 60–80 observations are required for reliable estimates. In this study, the number of objects on each transect and the length of transects involved a trade-off between what was possible to perform in the field and a desire to have enough observations from each transect. The lines were therefore placed in such a way that it was possible to pool data from two lines with very similar conditions.

Eleven surveyors from the Swedish National Forest Inventory participated in the study. None of the surveyors had used LTS before, although they were trained for forest inventory in general. Each surveyor spent half a day on a training program in which they
were, among other things, instructed to ensure that all objects close to the transect line were detected and what could be an appropriate walking speed. They also walked specific transect lines. To avoid dependencies between individual detections, the surveyors were instructed to try not to look for other objects once they left the transect line to look for an object's number. On one day, each surveyor walked four transect lines, two in the morning and two in the afternoon. The only time limit given was that the surveyors had to finish two transects on a half-day. The transects were taken in a random order and the surveyors were strictly instructed to work individually. The surveyors were not tested for colour blindness, but none of them were known to be colour-blind.

2.2. Density estimation

The number of objects detected from one transect varied between 32 and 98, with an average of 63 detected objects per line. From the collected distance data, a separate density estimate was made for each transect as:

\[ \hat{D} = \frac{n \hat{f}(0)}{2L} \]  

(1)

where \( n \) is the number of objects detected from the survey line, \( \hat{f}(0) \) the estimated probability density function of the distance data, evaluated at distance zero, and \( L \) the length of the transect. The probability density function of observed distances, \( f(x) \), has the same shape as the detection function, \( g(x) \), but is scaled to integrate to one. For details, see Buckland et al. (1993, pp. 37–39).

It can be difficult to find a single function type that can be used to analyse all line transect data (Buckland, 1985). However, in a large-scale inventory it can be very tedious to find the best detection function for each separate transect or observer. Part of the aim of this study was, therefore, to see how well a single model type worked when applied to all surveyors. Four different methods for estimating \( f(0) \) were tested.

Three models for the detection function, two parametric and one non-parametric, were selected from the models available in the DISTANCE program. These were also recommended by Buckland et al. (1993, p. 46) as good detection functions. The models were selected after testing several combinations of key functions and adjustment terms in \textit{DISTANCE} and letting the program select the number of adjustment terms needed through a testing of all combinations of terms. For details of this procedure, see Buckland et al. (1993, pp. 41–51) and Laake et al. (1993). For each model, Aikake's information criterion (AIC) was calculated and compared with other models. The three models with the lowest AIC values in most cases were selected. A fourth approach to estimating \( f(0) \) used a kernel estimator (Silverman, 1986). A kernel estimator is a non-parametric smoothing technique that has been investigated for line transect data by Chen (1996) and Mack and Quang (1998).

A short overview of these four methods, two parametric forms for the detection function \( g(x) \) and two non-parametric estimators of \( f(0) \), follows:

- The half-normal function:

\[ g(x) = \exp \left( -\frac{x^2}{2\sigma^2} \right) \]

where \( x \) is the object's distance from the transect line, and \( \sigma \) a parameter to be estimated from the distance data.

- The hazard-rate function:

\[ g(x) = 1 - \exp \left( -\frac{x^a}{\mu} \right)^b \]

where \( a \) and \( b \) are parameters to be estimated from the distance data.

- The Fourier series estimator of \( f(0) \):

\[ \hat{f}(0) = \frac{1}{w} \sum_{j=1}^{m} \hat{a}_j \]

where \( w \) is the maximum distance recorded or the truncation point, and \( m \) the number of the \( a \) terms which are estimated from the collected data as:

\[ \hat{a}_j = \frac{2}{nw} \sum_{i=1}^{s} \cos \left( \frac{2\pi i j}{nm} \right) \]

Here only one \( a \) term was added, since this was the number favoured in most cases in the test.

- The kernel estimator of \( f(0) \):

\[ \hat{f}(0) = \frac{2}{nh} \sum_{i=1}^{s} K \left( \frac{x_i}{h} \right) \]
where \( h \) is the bandwidth that determines the degree of smoothing and \( K \) the kernel function. Here a normal kernel was used, i.e.:

\[
K \left( \frac{x}{h} \right) = \frac{1}{\sqrt{2\pi}} e^{-x^2/(2h^2)}.
\]

A separate estimate of the parameters and/or \( f(0) \) was made for each surveyor on each survey line. Then a separate density estimate was made for each transect and surveyor, according to formula (1), using the estimates of \( f(0) \) from each of the above methods. For the first three cases, the distance program (Laake et al., 1993) was used to estimate \( f(0) \). The selected models fitted the data well in most cases according to the \( \chi^2 \) goodness-of-fit test \((\rho > 0.1\) for all of the models for 157 out of 176 surveyor–transect combinations). In the fourth case, the IMSL routine for a Gaussian kernel was used to estimate \( f(0) \). The bandwidth was calculated according to suggestions given by Silverman (1986, p. 48) as \( h = 0.93m^{-1/5} \), where \( a \) is the minimum value of the sample standard deviation or the median/1.34. Examples of these four functions fitted to distance data are shown in Fig. 2.

To avoid poor results due to too few observations, \( f(0) \) and \( D \) were also estimated based on distance data pooled from the two most similar survey lines. Moreover, Buckland et al. (1993, p. 50) recommend a general truncation removing 5–10% of the largest observations. Density estimations were made separately for the untruncated data and for data with the largest 5% of observations truncated.

2.3. Statistical analyses

The difference between the estimated density and the true density was calculated for each observation. A variance analysis was then conducted for each forest and object type, using a random effect model (Montgomery, 1991). The variance of surveyors’ systematic errors is estimated by the variance component for surveyor effect in such a model. A separate analysis was made for each combination of the two forest types and the two object types since residual plots indicated that the variance of the random errors was not constant for the different cases. The model used to analyse the difference, \( \text{Diff}_{ij} \), was:

\[
\text{Diff}_{ij} = \mu + \alpha_i + \epsilon_{ij}
\]

where \( \mu \) is the overall bias, \( \alpha_i \) the effect of surveyor \( i \), and \( \epsilon_{ij} \) the random error for surveyor \( i \) on survey line \( j \). Thus, for each analysis, \( i \) takes 11 possible values and \( j \) takes four possible values. Surveyor effects were considered as random.

The analyses were made using the SAS procedure MIXED (SAS, 1992). Variance components were estimated using the restricted maximum likelihood (REML) method. An alternative analysis was done using the GLM procedure (SAS, 1989) to generate F-tests for the variance components, which in this case were estimated using the ANOVA method. These estimates of variance components are equal to the REML estimates if all ANOVA estimates are non-negative.

---

Fig. 2. Histograms of distance data together with the estimated probability density functions, \( f(r) \). These are estimated using (a) the half-normal function (—) and the hazard-rate function (—-—), and (b) the Fourier series (—) and the kernel estimator (—-—).
3. Results

Individual density estimates varied between 8.9 and 28.2 objects ha\(^{-1}\) while the true density was 23 objects ha\(^{-1}\). The mean values of individual density estimates for each detection function are shown in Table 1. For comparison, mean values are also shown for density estimates based on pooled distance data from the two most similar transects, and for density estimates based on truncated data. In all cases, there is an underestimation of the true density. Comparing the different cases showed that the effect of both truncation and pooling of data was small. Comparing the different detection functions revealed that, in this study, the Fourier series provided the estimates that were closest (on average) to the true density. The kernel estimator gave both a large negative bias and a large standard deviation.

Table 2 shows the results of the analysis of variance for each combination of forest and object type. Since

<table>
<thead>
<tr>
<th>Distance data</th>
<th>Model for detection function</th>
<th>Fourier series</th>
<th>Kernel estimator</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Half-normal</td>
<td>Hazard-rate</td>
<td></td>
</tr>
<tr>
<td>Untruncated</td>
<td>21.51 (2.72)(^\text{a})</td>
<td>20.14 (2.33)</td>
<td>21.99 (2.89)</td>
</tr>
<tr>
<td>Truncated 5%</td>
<td>21.00 (2.63)</td>
<td>20.43 (2.54)</td>
<td>21.75 (2.85)</td>
</tr>
<tr>
<td>Pooled data, untruncated</td>
<td>21.93 (3.31)</td>
<td>19.76 (2.13)</td>
<td>22.21 (2.60)</td>
</tr>
<tr>
<td>Pooled data, truncated 5%</td>
<td>21.25 (2.13)</td>
<td>19.98 (2.24)</td>
<td>21.97 (2.30)</td>
</tr>
</tbody>
</table>

\(^{a}\) The true density was 23 objects ha\(^{-1}\).

\(^{b}\) Standard deviations for mean densities, calculated as \(\sqrt{(N-1) \sum (D_i - \overline{D})^2}\), where \(D_i\) is the estimate for each observation, \(\overline{D}\) the mean density, \(N = 176\) for estimates based on single lines, and \(N = 88\) for estimates based on pooled data.

### Table 1
Mean values of individual density estimates using four different methods for estimating \(D\) 

### Table 2
Results from analysis of variance for each combination of forest and object type
the effect of truncation and pooling was small, only the results of analyses of density estimates based on untruncated data from single survey lines are shown. The negative bias varied with different conditions, but in both types of forest it was larger for the kind of object that was more difficult to detect. In general, the Fourier series provided the estimates that were closest to the true density and the hazard-rate function and kernel estimator had in this study higher biases than the other two functions. The performance of the different detection functions was not compared statistically.

Testing for surveyors' differences revealed that the Var(s) was small and not significantly different from zero in both forests for the more easily detected objects, and in the young pine forest also for the objects that were more difficult to detect. Surveyors' effects were slightly larger for the objects that were more difficult to detect in the old mixed forest, and were for two of the detection functions significant. Comparison of the functions showed that the size of the variance component for surveyors' effects differed less than the bias. The random error component was at the same level in all cases, except for objects that were easy to detect in the old mixed forest. In that case, all random error components but the kernel estimator's were considerably smaller.

4. Discussion

In general, densities were underestimated under all conditions considered in this study. Examination of the data shows that this was due to the fact that objects close to the survey lines were sometimes missed. Approximately 15% of the objects within 1 m of the centre line were not detected.

The major focus of this study was to determine whether or not surveyors influence the results in LTS. If different surveyors walk the same survey line, they will detect different objects, leading to slightly different estimates. This difference could be entirely random, or could include a systematic component. This study found that the systematic differences between surveyors were generally moderate. At most, the coefficient of variation for surveyor effect was 6.8%. However, in long-term monitoring, differences of this size can disguise real trends or indicate illusory changes. The surveyor effect, together with the random error component, will also have an effect on the precision of estimates. This must be taken into account when comparing the efficiency of LTS with other methods.

The differences among estimates for the kind of objects that are easy to detect and those that are difficult to detect can be explained by a slightly higher proportion of missed objects in the 0–2 m interval in the latter case. The underestimation was worst for the objects that were difficult to detect in the old mixed coniferous forest, where sighting conditions at ground level were poor, making the objects even more difficult to detect. Differences between different surveyors' systematic errors are partially, but not wholly, explicable in terms of the proportion of missed objects close to the survey line.

When the probability of detection close to the survey line is less than one, the actual probability of detection must be estimated in some way. Since this is a well-known problem in wildlife inventories, much effort has been spent on finding solutions. For example, Zahl (1989) estimates \( g(0) \) by varying sampling intensities. Recently, several alternatives that combine LTS theory with capture-recapture theory have been proposed (Alpijar-kräa and Pollock, 1996; Quang and Becker, 1997; Borchers et al., 1998; Skag and Schweder, 1999). In these approaches two observers survey the same transect independently and the information about which objects that were detected by surveyor 1, by surveyor 2 or by both is used for estimation. For aerial surveys where there is limited visibility beneath the aircraft, left-truncation has been proposed (Allredge and Gaten, 1985). A test of left truncation on this material resulted in extremely unreliable estimates.

Even if the data in an LTS are collected in a probabilistic manner, there will always be some subjectivity in the analysis phase. Decisions have to be made about data truncation, the grouping of data, and the choice of model for the detection function. Comparison of different analysts' estimates in a blind test yielded variations in their estimates of from 3.5 to 9%, expressed in terms of the coefficient of variation (Anderson and Southwell, 1995). In the present study, the differences among estimates based on different models for the detection function were rather large. The AIC value can be used to guide the choice of
model. In this study, a Fourier series was favoured for 93, and the half-normal-function for 62 of the 176 surveyor-transect combinations. The difference between the smallest and largest AIC value was however for most observations rather small, in general only around 1–2 units. Only for ca. 15 of the 176 cases were this difference larger than 2.5 units. Also, looking at individual observations shows that in many cases the AIC value did not favour the estimate closest to the true density. However, a discussion of the performance of different models might be less adequate if it were known that the probability of detection on the survey line is less than one in many cases.

A kernel estimator has the advantage of avoiding decisions about models and truncation points, but it did not work very well in this study. The kernel estimator is probably more sensitive to missed objects on the survey line than the other models since it is more locally affected by the data. The kernel estimator also seems to be more susceptible to random fluctuations in the spatial location of objects in relation to the survey line than are the other models. For example, in the old forest with objects easy to detect, the kernel estimator has a substantially larger random error component than the other models, arising from a large underestimation for all surveyors on one survey line. On this particular line, the random distribution of objects had resulted in a smaller proportion of objects being placed between 0 and 2 m from the line than between 2 and 5 m. Although the surveyors detected all the objects close to the line, the density in this interval was much smaller than the average density, which had a large influence on the kernel estimator and resulted in a severe underestimation for all surveyors on this line.

It is possible that better estimates could have been obtained if the truncation point, choice of model etc. were analysed separately for each observation. However, in a larger-scale inventory it might be impossible to analyse all the collected data individually, and therefore it is useful to know how a single model would perform if used for all the material.

For some surveyor-transect combinations, the number of objects detected was smaller than the recommended 60–80, but increasing the sample size by pooling observations from two survey lines did not improve the estimates very much. For some detection-functions, forest and object-types the biases became slightly smaller, but in other cases slightly larger. The same pattern was observed for the variance components for surveyor effect and random errors. However, one obvious improvement was for the kernel estimator in the old forest with objects easy to detect (the case just discussed), were the bias was only one-third, and the random error only one half for estimates based on pooled data.

5. Conclusions

Although the differences between surveyors’ systematic errors and the different levels of bias found in this study can be explained by varying degrees of fulfilment of the assumption that objects close to the transect are certain to be detected, these differences are still problematic when it comes to long-term monitoring. If surveyors violate the assumption to differing degrees, the differences in their estimates can give an illusion of trends that do not exist or disguise real trends. For the same reason, the systematic differences in estimates caused by the choice of method for estimating \( f(0) \) are problematic. For the difficult task of surveying wildlife populations, there may be no comparable alternatives to LTL, but for surveying inanimate populations, there are other reasonable alternatives to be considered. Any comparison of methods must take into account the increased variation in estimates caused by surveyors’ systematic and random errors as well as the variation caused by the detection model selected.

Acknowledgements

The authors wish to thank all the surveyors from the Swedish National Forest Inventory for their participation in this study. Special thanks go to Lars Davidson and Bo Hansson for their work in the preparation phase of the study. Thanks also to Yagve Kjellin and Kveta Nilsson for producing and painting the artificial objects, to Jörgen Wallerstein and Hampus Holmström for helping with preparations in the field, to Prof. Göran Stål for valuable comments that considerably improved the manuscript, and to two anonymous reviewers for their comments. Proper English Ltd. revised the language.
References


