Shipboard Sightings Surveys: Methodological Developments to Meet Practical Needs

S.T. Buckland
University of St Andrews, School of Mathematics and Statistics
Mathematical Institute, North Haugh, St Andrews, UK
steve@mcs.st-and.ac.uk

1. Introduction

Shipboard sightings surveys are commonly used to estimate the abundance, and trends in abundance, of marine animals such as dolphins, whales, seals, turtles, seabirds and, occasionally, fish. The surveys are typically expensive, involving the hire of one or more ships, together with crew. Frequently, the survey area is large, and the target species sparsely distributed. There is therefore an onus on the statistician to devise efficient methods of estimating abundance.

2. Line transect sampling

By far the most widely used technique for such surveys is line transect sampling. The standard method proceeds as follows. The survey design comprises a series of k lines or transects of total length L, which are assumed to be randomly located in the study area, and the ship travels along each line in turn. A team of observers searches for the target animals. The distance y of each detected animal from the line is estimated, usually by calculating $y = r \sin q$, where r is the sighting distance of the animal from the ship and *è* the sighting angle of the animal from the line.

It is assumed that all animals on or very close to the line are detected, but probability of detection may decrease with distance from the line, out to some distance w (which might be infinite). A probability density function $f(y)$ is fitted to the distances of the n detected animals from the line, which allows the effective half-width $m = 1/f(0)$ of the surveyed strip to be estimated; we expect to detect as many animals beyond *ì* as we fail to detect within *ì* of the line. Animal density is then estimated as $\hat{D} = \dfrac{n}{2\hat{m}L}$.

When animals occur in well-defined groups, such as flocks or schools, then the distance of the group from the line is recorded, and the above formula yields an estimate of group density. We then use recorded group sizes to model the mean size $E(s)$ of groups in the population, and hence estimate animal density by

$$\hat{D} = \frac{n\hat{E}(s)}{2\hat{m}L} .$$

3. Covariate models for the detection process

Covariate models can be useful for improving precision of abundance estimates and for accommodating the effects of size-biased sampling in estimation. They also allow researchers to gain understanding in how different variables affect detectability. Heterogeneity in the sighting process may be modelled by using a probability density function $f(x)$ that includes a scale parameter (e.g. the half-normal model). We can then model the scale parameter as a function of covariates that contribute to the heterogeneity. The parameter $f(0)$ may be replaced by the mean value across detected animals, from which ì may be estimated. However, if animals occur in groups, group size is likely to be one of the covariates, in which case this approach is biased. This is due to size-bias: larger groups will tend to be over-represented in the sample, so that if estimation of $E(s)$ is not integrated with the covariate modelling, bias will result. We can avoid this problem by using a Horvitz-Thompson estimator in which the inclusion probabilities have been estimated. This yields $\hat{D} = \frac{\sum_{i=1}^{n} s_i \hat{f}(0 \mid z_i)}{2L}$, where s_i is the size of the i^{th} detected group and z_i is the vector of covariates for that group (which may include its size s_i) (Marques and Buckland, in prep.).

4. Double-platform methods

In many marine surveys, detection on the trackline is not certain because some animals are submerged when the ship passes. In this case, mark-recapture methods may be combined with distance sampling (Borchers, Zucchini and Fewster, 1998b). This is achieved by having two observation platforms. These might be treated as mutually independent, so that, if animals detected by both platforms (‘duplicate detections’) can be identified, two-sample mark-recapture methods that incorporate covariates can be used. Bias in such methods is typically large unless heterogeneity in detectability is well-modelled. However, it is seldom possible to record covariates that reflect this heterogeneity adequately; for example if a whale produces a blow that is particularly visible from one platform, due to light conditions or some other factor in the environment that is difficult to measure, then it will tend to be more visible from the other platform too, and abundance will be underestimated. These problems may be reduced by modelling individual detection cues (Schweder *et al.*, 1999; Skaug and Schweder, 1999) or by separating the areas of search for the two platforms, and using one to set up trials for the other (Borchers *et al.*, 1998a).

5. Spatial distance sampling models

Spatial distance sampling models are potentially useful for several reasons: animal density may be related to habitat and environmental variables, improving understanding of factors influencing abundance; animal abundance may be estimated for any sub-region of interest, simply by numerical integration of the estimated density surface over the sub-region; and they potentially allow analysis of data collected from transects that are non-random, such as in ‘platforms of opportunity’ surveys.

One way in which spatial models can be fitted to distance sampling data is to conceptualize the distribution of animals as an inhomogeneous Poisson process, in which the detection function represents a thinning process. If in the case of line transect sampling, the data are taken to be distances along the transect line between successive detections, this allows us to fit a spatial surface to these data. We can refine this further by conceptualizing the observations as ‘waiting areas’, that is the effective area surveyed between one detection and the next, where the effective width of the surveyed strip varies according to environmental conditions and observer effort (Hedley, Buckland and Borchers, 1999).

6. Automated design algorithms

Geographic Information Systems (GIS) are now widely available. To exploit them fully, designers of sightings surveys require automated design algorithms. This allows them to generate designs with known properties rapidly and simply. It also allows designs to be generated on the survey platform as they are required, which is an important advantage if the design must be able to accommodate rapidly changing circumstances, such as the moving ice edge in the Antarctic.

Shipboard surveys typically use continuous, zig-zag survey lines ('samplers'), so that costly ship time is not wasted in travelling from one line to the next. For convex survey regions or strata, a sampler with even coverage probability can be obtained by defining a principal axis for the design, and varying the angle of the sampler to that axis as the ship progresses through the area, to ensure that coverage probability is constant with respect to distance along the principal axis (Strindberg and Buckland, in prep.). By contrast, fixed-angle samplers do not give even coverage probability unless the survey region is rectangular. Fixed-waypoint samplers, which pass through equally spaced points on opposite sides of the survey region boundary, approximate an even coverage probability sampler more successfully than fixed-angle samplers.

If the survey region or stratum is not convex, but the region can be split into a small number of convex sections, an even-coverage sampler may be placed in each section, such that coverage probability is the same in each section. Another option is to place a convex hull around the region, and locate an even-coverage sampler within the hull. Sections of line outside the region would then not be covered.

7. Adaptive distance sampling

Animal populations are typically aggregated, and even large sightings surveys often yield small sample sizes of scarce species. These two features mean that abundance estimates are often imprecise. Adaptive sampling (Thompson and Seber, 1996) offers a means of increasing sample size, and hence improving both precision and bias, at no extra cost. Standard adaptive sampling methods can be extended to distance sampling. However, a major practical problem for shipboard adaptive surveys is that the required survey effort is not known in advance, whereas the ship is available for a predetermined number of days. Thus standard adaptive sampling will either finish early, so that ship time is wasted, or run out of effort, in which case the study region will be incompletely surveyed. Additionally for shipboard surveys, we would wish to avoid time off-effort, while the ship proceeds from one line to another; the loss of on-effort time may well more than offset any gain in efficiency from using adaptive sampling. Pollard and Buckland (1997) developed a method in which, when additional effort is triggered, the ship changes to a zig-zag (and hence continuous) course, centred on the nominal trackline. The angle of the zig-zag is a function of whether the ship is ahead of or behind schedule, allowing the total effort of the survey to be fixed in advance. This has the additional advantage that, if bad weather is encountered, the ship can make up for lost effort by reducing the amount of adaptive effort. In analysis, data from adaptive segments are downweighted to allow for the greater effort. Unlike standard adaptive sampling, the method is not design-unbiased, but simulations indicate that bias is small.

REFERENCES

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**, 1221-37.

Borchers, D.L., Zucchini, W. and Fewster, R.M. (1998b) Mark-recapture models for line transect surveys. *Biometrics* **54**, 1207-20.

Hedley, S.L., Buckland, S.T. and Borchers, D.L. (1999) Spatial modelling from line transect data. *Journal of Cetacean Research and Management* **1**, 255-64.

Marques, F.F.C. and Buckland, S.T. (in prep.) Incorporating covariates into standard line transect analyses.

Pollard, J.H. and Buckland, S.T. (1997) A strategy for adaptive sampling in shipboard line transect surveys. *Rep. int. Whal. Commn* **47**, 921-931.

Schweder, T., Skaug, H.J., Langaas, M. and Dimakos, X.K. (1999) Simulated likelihood methods for complex double-platform line transect surveys. *Biometrics* **55**, 678-87.

Skaug, H.J. and Schweder, T. (1999) Hazard models for line transect surveys with independent observers. *Biometrics* **55**, 29-36.

Strindberg, S. and Buckland, S.T. (in prep.) Optimised survey design: the zigzag sampler in line transect surveys.

Thompson, S.K. and Seber, G.A.F. (1996) *Adaptive Sampling*. Wiley, New York.

RESUME

Des enquêtes visuelles du bord de navire sont répandues pour évaluer l’abondance d’animaux marins tels que les dauphins, les baleines, les phoques, les tortues de mer, les oiseaux de mer et, de temps en temps, les poissons. La superficie des enquêtes est souvent vaste, et les enquêtes peuvent coûter cher, mais les tailles des échantillons sont en général petites. Le dessin de l’enquête et les méthodes d'analyse qui mènent à l’évaluation efficace sont très importantes, mais le statisticien faut aussi se rendre compte de l’hétérogénéité du comportement animal, des capacités de l’observateur, et de l’environnement. Nous passons en revue les développements récents et continus pour améliorer l'efficacité de telles enquêtes, y compris: les algorithmes conçus pour les échantillonneurs ‘zig-zag’ continus, pour réduire au minimum le temps ‘au-loin’ effort; les stratégies de prélèvement adaptives pour augmenter l’échantillon; méthodes covariate pour modeler l’hétérogénéité dans la détectabilité; les méthodes de double-plateforme pour les enquêtes dans lesquelles la détection des animaux sur le ligne de piste n'est pas sûre; et les modèles spatials qui relient la densité animale aux variables océanographiques, et qui analysent les données des ‘plateformes d’occasion’.