Conventional Distance Sampling[1]

Tiago A. Marques[a,b], Stephen T. Buckland[a], David L. Borchers[a],
Eric Rexstad[a] and Len Thomas[a]

[a]Research Unit for Wildlife Population Assessment, Centre for Research into
Ecological and Environmental Modelling, The Observatory, University of St
Andrews, St Andrews, KY16 9LZ, Scotland
[b] Centro de Estatística e Aplicações da Universidade de Lisboa, Faculdade de
Ciências da Universidade de Lisboa, Bloco C6 - Piso 4, Campo Grande, 1749-016
Lisboa, Portugal

Distance sampling is a widely used methodology for estimating animal density or abundance. Its name derives from the fact that the information used for inference are the recorded distances to objects of interest, usually animals, obtained by surveying lines or points. The methods are also particularly suited to plants or immotile objects, as the assumptions involved (see below for details) are more easily met. In the case of lines the perpendicular distances to detected animals are recorded, while in the case of points the radial distances from the point to detected animals are recorded. A key underlying concept is the detection function, usually denoted $g(y)$ (here y represents either a radial or perpendicular distance from the line or point). This represents the probability of detecting an animal of interest, given that it is at a distance y from the transect. This function is closely related to the probability density function (pdf) of the detected distances, $f(y)$, as

$$f(y) = \frac{g(y)\pi(y)}{\int_0^w g(y)\pi(y)dy},\tag{1}$$

where $\pi(y)$ is the distribution of distances available for detection and w is a truncation distance, beyond which distances are not considered in the analysis. The above pdf provides the basis of a likelihood from which the parameters of the detection function can be estimated. An important and often overlooked consideration is that $\pi(y)$ is assumed known. This is enforced by design, as the random placement of transects, independently of the animal population, leads to a distribution which is uniform in the case of line transects and triangular in the case of point transects (see Buckland *et al.* (2001) for further details).

Given the n distances to detected animals, density can be estimated by

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

in the case of line transects with total transect length L, where $\hat{f}(0)$ is the estimated pdf evaluated at 0 distance, and by

---

$$\hat{D} = \frac{n\hat{h}(0)}{2k\pi} \tag{3}$$

in the case of k point transects, where $\hat{h}(0)$ is the slope of the estimated pdf evaluated at 0 distance (Buckland *et al.*, 2001). This is a useful result because we can then use all the statistical tools that are available to estimate a pdf in order to obtain density estimates. So one can consider plausible candidate models for the detection function and then use standard maximum likelihood to obtain estimates for the corresponding parameters and therefore density estimates.

The most common software to analyze distance sampling data, Distance Thomas *et al.* (2010), uses the semi-parametric key+series adjustment formulation from Buckland (1992), in which a number of parametric models are considered as a first approximation and then some expansion series terms are added to improve the fit to the data. Standard model selection tools and goodness-of-fit tests are available for assisting in model selection.

Variance estimates can be obtained using a delta method approximation to combine the individual variances of the random components in the formulas above (i.e. n and either $\hat{f}(0)$ or $\hat{h}(0)$; for details on obtaining each component variance, see Buckland *et al.* (2001)). In some of the more complex scenarios, one must use resampling methods based on the non-parametric bootstrap, which are also available in the software.

Given a sufficiently large number of transects randomly allocated independently of the population of interest, estimators are asymptotically unbiased if (1) all animals on the transect are detected, i.e., $g(0) = 1$, (2) sampling is an instantaneous process (typically it is enough if animal movement is slow relative to the observer movement), and (3) distances are measured without error (see Buckland *et al.* (2001) for further details about assumptions). Other assumptions, like the fact that all detections are independent events, are strictly required as the methods are based on maximum likelihood, but the methods are extraordinarily robust to their failure (Buckland, 2006). Failure of the $g(0) = 1$ assumption leads to underestimation of density. Violation of the movement and measurement error assumption have similar consequences. Underestimation of distances and undetected responsive movement towards the observers lead to overestimation of density, and overestimation of distances and undetected movement away from the observer lead to underestimation of density. Random movement and random measurement error usually leads to overestimation of density. Naturally the bias depends on the extent to which the assumptions are violated. Most of the current research in the field is aimed at relaxing or avoiding the need for such assumptions. As there are no free lunches in statistics, these come at the expense of more elaborate methods, additional data demands and additional assumptions.

Further details about conventional distance sampling, including dealing with clustered populations, cue counting methods and field methods aspects, can be found in Buckland *et al.* (2001), while advanced methods, including the use of multiple covariates in the detection function, double platform meth-

ods to deal with situations in which $g(0) < 1$, spatial models, automated survey design, and many other specialized topics, are covered in Buckland *et al.* (2004). An extended list of distance sampling related references can be found at http://www.creem.st-and.ac.uk/tiago/webpages/distancesamplingreferences.html.

References

Buckland, S. T. (1992). Fitting density functions with polynomials. *Applied Statistics*, **41**, 63–76.

Buckland, S. T. (2006). Point transect surveys for songbirds: robust methodologies. *The Auk*, **123**(2), 345–345.

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 University Press, Oxford.

Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D., and Thomas, L. (2004). *Advanced Distance Sampling*. Oxford University Press, Oxford.

Thomas, L., Buckland, S. T., Rexstad, E. A., Laake, J. L., Strindberg, S., Hedley, S. L., Bishop, J. R., Marques, T. A., and Burnham, K. P. (2010). Distance software: design and analysis of distance sampling surveys for estimating population size. *Journal of Applied Ecology*, **47**, 5–14.