Distance sampling

Distance sampling is a widely used group of closely related **methods for estimating the density and/or abundance** of biological populations. The main methods are *line-transect sampling* and *point-transect sampling* (also called *variable circular plot sampling*). These have been used successfully in a very diverse array of taxa, including trees, shrubs and herbs, insects, amphibians, reptiles, birds, fish, and marine and land mammals. In both cases, the basic idea is the same. One or more observers perform a standardized survey along a randomly located set of lines or points, searching for objects of interest (usually animals or clusters of animals). For each object detected, they record the distance from the line or point to the object. Not all the objects will be detected, but a fundamental assumption of the basic methods is that all objects that are actually on the line or point are detected. Intuitively, one would expect that objects become harder to detect with increasing distance from the line or point, resulting in fewer detections with increasing distance. The key to distance sampling analyses is to fit a *detection function* to the observed distances, and use this fitted function to estimate the proportion of objects missed during the survey. From here, we can readily obtain point and interval estimates for the density and abundance of objects in the survey area. The basic methods (sometimes called *standard* or *conventional distance sampling*) are described in detail in Ref. 1, which is an updated version of Ref. 2. Various extensions and more advanced methods are considered in Ref. 3. Free software, Distance [4], provides for the design and analysis of distance sampling surveys, implementing the methods described in Ref. 1 and many of those in Ref. 3.

Distance sampling is an extension of quadrat-based sampling methods. Two forms of quadrat sampling are *strip transects*, in which one or more observers move along a line, counting all objects within a predetermined distance of the line, and *point counts*, in which numbers of objects (usually birds or plants) in a circle about a point are counted. Population density is then estimated by dividing the total count by the total area surveyed. A fundamental assumption of these methods is that all objects within the strip or circle are counted. This assumption is difficult to meet for many populations, and cannot be tested using the survey data. Furthermore, especially for scarce species, the methods are wasteful because detections of objects beyond the strip or circle boundary are ignored. If the width of the strip or the radius of the circle is made sufficiently small to ensure that detection of any object within the surveyed area is almost certain, then a large proportion of detections are outside the surveyed area and so are ignored. Distance sampling extends quadrat-based methods by relaxing the assumption that all objects within the circle or strip are counted. By measuring distances to the objects that are observed, the probability of observing an object within the circle or strip can be estimated.

Another approach to estimating wildlife abundance involves **capture–recapture methods**. These are often more labor-intensive and more sensitive to failures of assumptions than distance sampling. However, they are applicable to some species that are not amenable to distance sampling methods, and can yield estimates of survival and recruitment rates, which distance sampling cannot do. Capture–recapture methods based on remote camera traps have been used successfully for cryptic or elusive species that show individual natural markings; DNA-based methods using hair-snares may also be possible. Capture–recapture can be efficient for populations that aggregate at some location each year, whereas distance sampling methods are more effective on dispersed populations. The two methods should therefore be seen as different tools for different purposes. We also note that various "hybrid" approaches take elements from both distance sampling and capture–recapture – these include **spatially explicit capture–recapture** and mark–recapture distance sampling (see the sections titled Related Methods and Standard Methods).

In **fish population estimation**, *catch per unit effort, catch-at-age*, and *catch-at-length* are all commonly used to estimate abundance [5], as they require that the commercial catch is sampled, which is more cost-effective than sampling the living fish. Acoustic

Based in part on the article "Distance sampling" by Len Thomas, Stephen T. Buckland, Kenneth P. Burnham, David R. Anderson, Jeffrey L. Laake, David L. Borches, and Samantha Strindberg, which appeared in the *Encyclopedia of Environmetrics*.

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.

This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vad033.pub2

surveys of fish schools can provide data amenable to distance sampling methods.

Alternative methods for estimating animal abundance are reviewed and compared in Refs 6–8.

First, the basic methods for line and point transects are considered, before briefly reviewing some related and extended methods.

Survey Design

As with any sampling exercise, obtaining reliable results from a distance sampling survey depends critically on good survey design. This relies on the fundamental sampling principles of replication and randomization. Systematic random placement of sampling units (lines or points) is preferable to a completely random design [1, 3], and stratification can be employed to reduce variance. The design process is facilitated by the software Distance, which has built-in (GIS) functionality and implements automated design algorithms [3, 9]. It can generate survey plans based on a range of point- and line-transect designs, as well as performing simulations to compare the efficiency of different designs and to investigate design properties such as probability of each location in the survey area being sampled (referred to as *coverage probability*).

Line-transect Sampling

In line-transect sampling, a series of (systematic) randomly located straight lines is traversed by one or more observers. This may be achieved in various ways, depending on the study species. In terrestrial studies, potential modes of movement include walking, horseback, all-terrain vehicle, airplane, and helicopter. Transect surveys in aquatic environments can be conducted by snorkeling or with SCUBA gear, from submarines, surface vessels, aircraft, or from sleds with mounted video units pulled underwater by a surface vessel. Passive acoustic surveys are gaining popularity for marine mammal species (and other taxa), in which case the "observer" is a towed hydrophone array. In the case of large observation platforms, there is typically a team of observers.

Estimation

Perpendicular distances x are determined from the line to each detected object of interest. In practice, (radial) detection distances r and detection angles θ are often recorded, from which perpendicular distances are calculated as $x = r \sin\theta$. Suppose k lines of lengths $l_1, \ldots, l_k$ (with $\sum l_j = L$) are positioned according to some randomized scheme, and that animals further than some distance w from the line (the truncation distance) are not recorded. Then, the surveyed area is $a = 2wL$, within which n animals are detected at perpendicular distances $x_1, \ldots, x_n$. Let P_a be the probability that a randomly chosen animal within the surveyed area is detected, and suppose an estimate $\hat{P}a$ is available. Animal density D is then estimated by

$$\hat{D} = \frac{n}{2wL\hat{P}_a} \quad (1)$$

To provide a framework for estimating P_a, we define the detection function $g(x)$ to be the probability that an object at distance x from the line is detected, $0 \le x \le w$, and assume that $g(0) = 1$ (i.e., that we are certain to detect an animal on the transect line). If we plot the recorded perpendicular distances in a histogram, then conceptually the problem is to specify a suitable model for $g(x)$ and to fit it to the perpendicular distance data. As shown in Figure 1, if we define $\mu = \int_0^w g(x)dx$, then $P_a = \mu/w$. The parameter μ is called the *effective strip (half-) width*;

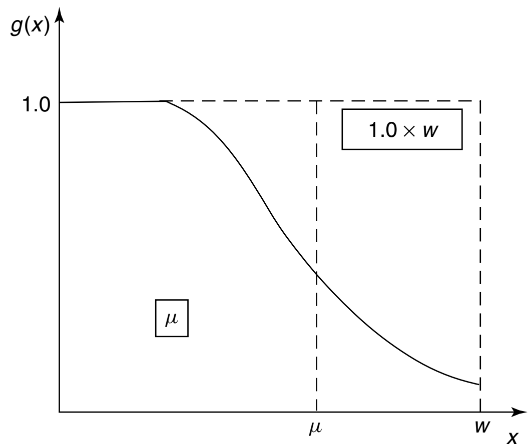

Figure 1 The area μ under the detection function $g(x)$, when expressed as a proportion of the area w of the rectangle, is the probability that an object within the surveyed area is detected; μ is also the effective strip width, and takes a value between 0 and w. (*Source*: Reproduced from Ref. 10. © John Wiley & Sons, Ltd, 1998.)

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.
This article is © 2013 John Wiley & Sons, Ltd.
This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vad033.pub2

it is the distance from the line for which as many objects are detected beyond μ as are missed within μ (Figure 1). Thus,

$$\hat{D} = \frac{n}{a\hat{P}_{\mathrm{a}}} = \frac{n}{2wL\hat{\mu}/w} = \frac{n}{2\hat{\mu}L} \qquad (2)$$

We now need to estimate μ. We can turn this into a more familiar estimation problem by noting that the probability density function (pdf) of perpendicular distances to detected objects, denoted by $f(x)$, is the detection function $g(x)$, rescaled so that it integrates to unity (*see* **Frequency curves**). That is, $f(x) = g(x)/\mu$. In particular, because we assume $g(0) = 1$, it follows that $f(0) = 1/\mu$ (Figure 2). Hence,

$$\hat{D} = \frac{n}{2\hat{\mu}L} = \frac{n\hat{f}(0)}{2L} \qquad (3)$$

The problem is reduced to modeling the pdf of perpendicular distances, and evaluating the fitted function at $x = 0$. The large literature for fitting density functions is now available to us. The Distance software uses the methods of Ref. 1, in which a parametric "key" function is selected and, if it fails to provide an adequate fit, polynomial or cosine series adjustments are added until the fit is judged to be satisfactory by one or more criteria.

Often, the perpendicular distances are recorded by distance category, so that each exact distance need not be measured, or data are grouped into distance categories before analysis. Standard likelihood methods for multinomial data are used to fit such "grouped"(or "interval") data [1].

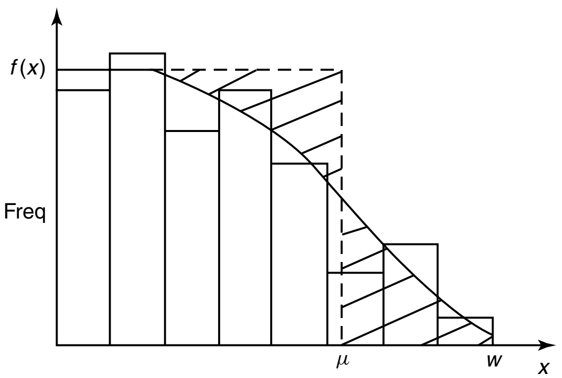

Figure 2 The pdf of perpendicular distances, $f(x)$, plotted on a histogram of perpendicular distance frequencies (scaled so that the total area of histogram bars is unity). The area below the curve is unity by definition. Because the two shaded areas are equal in size, the area of the rectangle, $\mu f(0)$, is also unity. Hence $\mu = 1/f(0)$. (*Source*: Reproduced from Ref. 10. © John Wiley & Sons, Ltd, 1998.)

Variance and Interval Estimation

The variance of $\hat{D}$ is well approximated [1] using the **delta method**:

$$\widehat{var}(\hat{D}) = \hat{D}^2 \left[\frac{\widehat{var}(n)}{n^2} + \frac{\widehat{var}[\hat{f}(0)]}{[\hat{f}(0)]^2} \right] \qquad (4)$$

The variance of n generally is estimated from the sample variance in encounter rates, n_j/l_j. Examination of a range of possible estimators of this variance [11] showed that one based on weighting by line lengths squared performed best when lines are placed at random within the study area. This is the default in the Distance software (versions 6 and higher). The authors also proposed and evaluated estimators for the case of systematic random parallel line placement (a recommended design strategy); an updated estimator with even better performance has since been developed [12].

Given that $f(0)$ is estimated by **maximum likelihood**, its variance can be estimated from the **information matrix**.

If we assume that $\hat{D}$ is **lognormally distributed**, approximately 95% confidence limits are given by $(\hat{D}/C, \hat{D}C)$ where

$$C = \exp\{1.96[\widehat{var}(\ln \hat{D})]^{0.5}\} \qquad (5)$$

with

$$\widehat{var}(\ln \hat{D}) = \ln \left[1 + \frac{\widehat{var}(\hat{D})}{\hat{D}^2} \right] \qquad (6)$$

Bootstrap resampling may also be used for variance and interval estimation. In this case, resamples are usually generated by sampling with replacement from the lines, so that independence between the lines is assumed, but independence between detections on the same line is not. If the model selection procedure for the detection function is applied independently to each resample, the bootstrap variance includes a component because of model selection uncertainty.

Cluster Size Estimation

Animals often occur in groups, which we term *clusters*. These may be flocks of birds, herds of ungulates,

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.
This article is © 2013 John Wiley & Sons, Ltd.
This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vad033.pub2

pods of whales, and so on. If one animal in a cluster is detected, then it is assumed that the whole cluster is detected, and the distance to the center of the cluster is recorded. Equation (3) then gives an estimate of the density of clusters. To obtain the estimated density of individuals, we must multiply by an estimate of mean cluster size in the population, $\mathrm{E}(s)$:

$$\hat{D} = \frac{n\hat{f}(0)\hat{\mathrm{E}}(s)}{2L} \tag{7}$$

Probability of detection is often a function of cluster size, so that the sample of detected cluster sizes exhibits size bias (larger clusters are easier to detect and so are overrepresented in the sample). In the absence of size bias, we can take $\hat{\mathrm{E}}(s) = \bar{s}$, the mean size of detected clusters. Several methods exist for estimating $\mathrm{E}(s)$ in the presence of size bias [1] (*see* **Size-biased sampling**). One that works well in practice is to regress $\log(s)$ on $\hat{\mathrm{g}}(x)$, the estimated probability of detection at distance x ignoring the effect of cluster size, and then predict $\log(s)$ when detection is certain, $\hat{\mathrm{g}}(x) = 1$, as there can be no size bias in that circumstance. The prediction is back-transformed using a bias adjustment. Another option to account for size bias is to include cluster size as an additional covariate in the detection function (see the section titled Multiple Covariate Distance Sampling).

Assumptions

The physical setting for line-transect sampling is idealized as follows:

1. N objects are distributed through an area of size A according to some stochastic process with average rate parameter $D = N/A$.
2. Lines, placed according to some randomized design, are surveyed and a sample of n objects is detected.

It is not necessary that the objects be randomly (i.e., Poisson) distributed. Rather, it is critical that the line or point be placed randomly with respect to the local distribution of objects. This ensures that objects in the surveyed strip are uniformly distributed with respect to distance from the line. Thus, if the strip has half-width w, object-to-line distances available for detection are uniformly distributed between zero and w. Random line placement also allows valid design-based extrapolation from the sample to the study area.

There are three key assumptions of the basic method. Many extensions have been developed (see section titled Extensions to Standard Methods) with the aim of relaxing these assumptions.

1. Objects directly on the line are always detected, i.e., $g(0) = 1$. Missing objects on the line causes a corresponding underestimation of D.
2. Objects do not move. Conceptually, distance sampling is a "snapshot" method: we would like to freeze animals in position while we conduct the survey. In practice, nonresponsive movement is not problematic provided it is slow relative to the speed of the observer. Responsive movement before detection is, however, problematic [13, 14].
3. Distances are measured accurately (for ungrouped distance data), or objects are correctly allocated to distance interval (for grouped data). Provided distance measurements are approximately unbiased, bias tends to be small in the presence of measurement errors [1, 15–17]. Biased measurements, if uncorrected, are problematic, and in cases when detection distances and angles are recorded, rounding of small angles to zero can also cause problems with estimation.

A fourth assumption is made in many derivations of estimators and variances: whether an object is detected is independent of whether any other object is detected. Point estimates of D are robust to violations of the assumption of independence, and robust variance estimates are obtained by taking the line to be the sampling unit, either by bootstrapping on lines, or by calculating a weighted sample variance of encounter rates by line.

It is also important that the detection function has a "shoulder"; that is, the probability of detection remains at or close to one initially as distance from the line increases from zero. This is not an assumption, but a property that allows more reliable estimation of object density.

Given the above, and assuming a suitably flexible method for estimating $g(x)$, the point and interval estimates of D are extremely robust to variation in $g(x)$ due to other factors such as observer, habitat, weather, and so on. This very useful property of standard distance sampling methods is known as *pooling robustness* [1, 3], and is not shared by capture–recapture methods, which are intrinsically

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.
This article is © 2013 John Wiley & Sons, Ltd.
This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vad033.pub2

nonrobust to unmodeled variation in detectability [18]. Large variations in density over the study area are also not a problem although if areas of differing density can be defined in advance and then stratification of survey effort should be used to increase precision.

Point-transect Sampling

In point-transect sampling, an observer visits a number of points, the locations of which are determined by some (systematic) random design. The method is usually (but not exclusively) used for songbird populations, in which typically many species are recorded and most detections are aural. By recording from points, the observer can concentrate on detecting the objects of interest, without having to navigate along a line, and without having to negotiate a randomly positioned line through possibly difficult terrain. The principal disadvantages are that detections made while traveling from one point to the next are not utilized, a problem especially for scarce species, and the method is unsuited to species that are generally detected by flushing them, or to species that typically change their location appreciably over the time period of the count (see below). An increasing application is in passive acoustic density estimation of cetaceans (and other taxa) from fixed locations, where recording devices may be left at a sample of sites for long time periods.

Estimation

Detection distances r are measured from the point to each detected object. Suppose the design comprises k points, and distances less than or equal to w are recorded. Then the surveyed area is $a = k\pi w^2$, within which n objects are detected. As for line-transect sampling, denote the probability that an object within the surveyed area a is detected by P_a with estimate $\hat{P}_\mathrm{a}$. Then, we estimate the object density D by

$$\hat{D} = \frac{n}{k\pi w^2 \hat{P}_\mathrm{a}} \tag{8}$$

We now define the detection function $g(r)$ to be the probability that an object at distance r from the point is detected, and we again assume that $g(0) = 1$. For line transects, the area of an incremental strip at distance x from the lines is $L\ \mathrm{d}x$, independent of x, which leads to the result that the pdf of distances differs from the detection function only in scale. By contrast, an incremental annulus at distance r from a point has area $2\pi r\ \mathrm{d}r$, proportional to r, so that the pdf of detection distances is $f(r) = 2\pi r g(r)/\nu$, where $\nu = 2\pi \int_0^w rg(r)\mathrm{d}r$. The respective shapes of the two functions $g(r)$ and $f(r)$ are illustrated in Figure 3. If we define an effective radius ρ, analogous to the effective strip width of line-transect sampling, then $\nu = \pi\rho^2$ is the effective area surveyed per point (Figure 4). Hence,

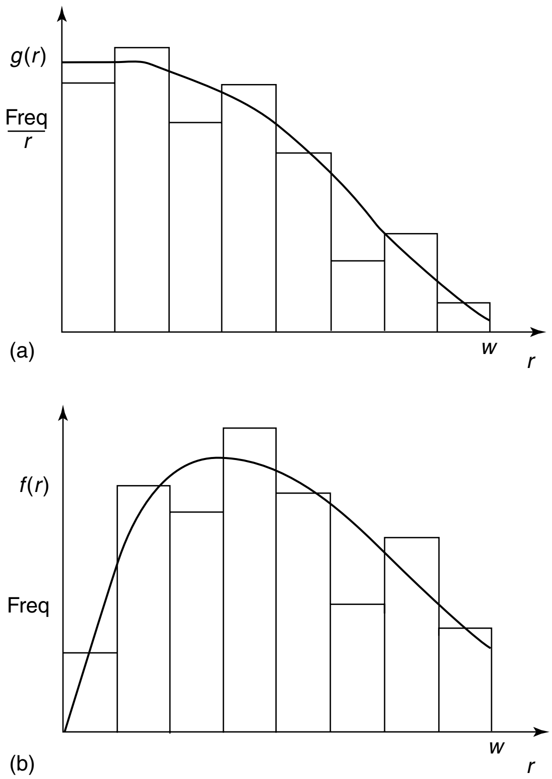

Figure 3 Histograms of detection distances from a point-transect survey. In (a), each histogram frequency has been scaled by dividing by the midpoint of the corresponding group interval. Also shown are the corresponding fits of the detection function [$g(r)$ in (a)] and the pdf of detection distances [$f(r)$ in (b)]. (*Source*: Reproduced from Ref. 10. © John Wiley & Sons, Ltd, 1998.)

$$\hat{D} = \frac{n}{a\hat{P}_\mathrm{a}} = \frac{n}{k\pi w^2 \pi\hat{\rho}^2/\pi w^2} = \frac{n}{k\hat{\nu}} \tag{9}$$

The area of the triangle in Figure 4 is $\rho^2 f'(0)/2$, where $f'(0)$ is the slope of $f(r)$ at $r = 0$. Since this is equal to the area under $f(r)$, which is unity, it

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.
This article is © 2013 John Wiley & Sons, Ltd.
This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vad033.pub2

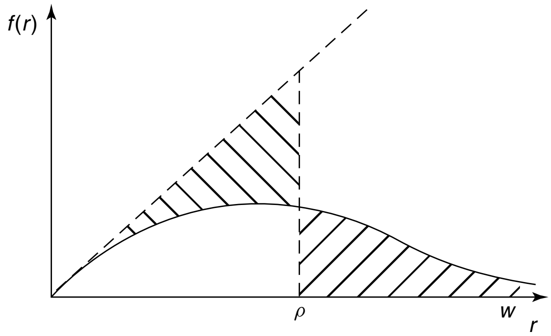

Figure 4 The pdf of detection distances, $f(r)$. The area under the curve is unity by definition. Because the two shaded areas are equal in size, the area of the triangle, $\rho^2 f'(0)/2$, is also unity. Hence $\nu = \pi\rho^2 = 2\pi/f'(0)$. (*Source*: Reproduced from Ref. 10. © John Wiley & Sons, Ltd, 1998.)

follows that $\nu = \pi\rho^2 = 2\pi/f'(0)$, and

$$\hat{D} = \frac{n\hat{f}'(0)}{2\pi k}$$

We therefore need to model the pdf of detection distances, and evaluate the slope of the fitted function at $r = 0$. The software Distance does this using the same set of models for the detection function as for line-transect sampling.

Variance and Interval Estimation

The methods for variance and interval estimation for line-transect sampling also apply to point transects with minor modifications. The variance of n is usually estimated from the sample variance in encounter rates between points. However, point-transect surveys are sometimes designed by defining a series of lines, as if a line-transect survey is to be carried out, then locating a series of points along each line. If the distance between neighboring points on the same line is smaller than the distance between neighboring points on different lines, then the data for all points on the same line should be pooled and the variance of n estimated from the sample variance in encounter rates between lines, weighted by the number of points on each line. Similarly, in this situation, bootstrap variance estimates should be calculated by resampling lines with replacement, rather than individual points.

Assumptions

Assumptions are virtually unchanged from those for line-transect sampling. As there, the standard analyses are very robust to failure of the assumption of independent detections. Point-transect sampling is, however, more subject to bias than line-transect sampling when objects move through the area around a point. In principle, we try to obtain a snapshot, locating each object at the position it occupied at one instant in time. However, the count is not instantaneous, because the observer needs time to detect all objects close to that point. If, during that time, movement brings new objects into the neighborhood of the point, then object density will be overestimated. To minimize bias, we recommend that the amount of time spent at the point before and after the snapshot instant be fixed in advance, and be as small as possible, given the requirement to detect all objects close to the point. An alternative approach is to use a cue counting method (see the section titled Related Methods), where the object counted is an instantaneous cue such as a bird song. Point-transect methods are used to estimate the density of cues (i.e., bird songs per unit area and time), and this is converted to animal density by dividing by an estimate of cue rate, obtained separately. The relative merits of these and other methods for songbirds are considered in Ref. 19.

Related Methods

Trapping webs [1] and trapping line transects [3] provide an alternative to traditional capture–recapture sampling for estimating animal density. In these, traps are placed along lines radiating from randomly chosen points (trapping webs) or with declining density either side of randomly chosen lines (trapping line transects), and the distances of trapped animals from the points or lines are used to estimate a distance sampling detection function. However, these methods suffer a number of potential biases [3], and have been largely superseded by **spatially explicit capture–recapture**.

Cue counting [20] was developed as an alternative to line-transect sampling for estimating whale abundance from sighting surveys. Observers on a ship or aircraft record all sighting cues within a sector ahead of the platform and their distance from the platform.

This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vad033.pub2

The cue used depends on species, but might be the blow of a whale at the surface. The sighting distances are converted into the estimated number of cues per unit time per unit area using a point-transect modeling framework. The cue rate is estimated from separate studies, in which individual animals or clusters are monitored over a period of time. Cue counting has also been used in passive acoustic point-transect surveys of songbirds [19] and whales [21].

Indirect methods are often used when the animals are rare, cryptic, or tend to move away before being detected. Instead of counting the animals, the objects counted are something produced by the animals, for example, animal dung (e.g., deer [22] or elephant [23]) or nests (e.g., great apes [23, 24]). To convert object density to animal density, one must then estimate two further parameters: object production rate and object disappearance rate, from separate studies.

Distance sampling can be viewed as a quadrat-based method where detection within the sampled quadrats is not assumed certain, and auxiliary information, the distances, is gathered and used to estimate detection probability. In some cases, it may be better (or more feasible) to estimate detection probability from other sources. An example is a *trapping point transect*, where an auxiliary survey is used to determine the detection function for traps set at known distances from radiotracked small mammals; this function is then used to estimate the effective area surveyed by a grid of traps in a main survey that covers the study area [25]. Similar methods have been used to estimate a passive acoustic detection function for cetaceans, using a sample of animals at known locations [26, 27]. A related method, used when animals can be lured in, for example, using a playback, is the *lure point transect* [28]. At the extreme, models based on theoretical considerations could be used to predict detectability, as with Ref. 29 for passive acoustic density estimation of sperm whales; however, this is clearly a method of last resort.

Related techniques sometimes used by botanists to estimate densities (and sometimes also termed distance sampling) are nearest-neighbor methods and point-to-nearest object methods [30]. These approaches do not involve modeling the detection function, and so are outside the definition of distance sampling used here.

Extensions to Standard Methods

The basic theory of distance sampling is now well established, as are the standard estimation and field methods [1]. More recent research (e.g., Ref. 3) has focused on methods for increasing precision and relaxing the assumptions of the standard methods, and on advanced design issues.

Multiple Covariate Distance Sampling

Generally, probability of detection is a function of many factors other than distance of the object from the line or point. We have considered briefly one other factor, cluster size, because if we do not allow for size bias in detection when objects occur in clusters, then our object density estimator may be biased. Other sources of heterogeneity contribute little to bias, provided $g(0) = 1$. Nevertheless, higher precision might be anticipated if additional covariates are recorded and their effects on $g(x)$ modeled. One approach, first used by Ref. 31, is to allow covariates to affect the scale of the detection function but not its shape. Building on this work, extensions to thedetection function estimation methods outlined in the section on line-transect sampling above allow the scale parameter of the key function to be a function of covariates [3, 32]. See also Ref. 33 for an accessible guide to these extensions.

Mark–recapture Distance Sampling

In some surveys, detection on the transect line is not certain ($g(0)<1$), either because observers miss potentially detectable animals because of environmental conditions, fatigue, and so on (*perception bias*), or because animals are unavailable for detection, for example, because they are underground or underwater (*availability bias*), or both.

In the case of perception bias, capture–recapture methods may be combined with distance sampling, through the use of two observation platforms [3, 34]. The platforms might be treated as mutually independent so that, provided that animals detected by both platforms (duplicate detections) can be identified, two-sample capture–recapture methods that incorporate covariates can be used. Bias in such methods is typically large enough to be of concern unless heterogeneity in detectability is well modeled. However, it is seldom possible to record covariates that

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.
This article is © 2013 John Wiley & Sons, Ltd.
This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vad033.pub2

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 separating the areas of search for the two platforms, and using one to set up trials for the other. Various analytic approaches are also potentially available [35]. The resulting binary data may then be modeled using logistic regression [36]. In some studies, the platform that sets up the trials could be provided, for example, by a radiotagging study, where locations of animals are known, or by an underwater acoustic array (so long as species could be identified accurately). In double-platform methods, Horvitz–Thompson-like estimators are used to estimate density, given the estimated probability of detection for each observation (*see* **Sampling, environmental**). All of the above work has taken place in the context of line transects, but similar methods for point transects are also available [37].

When animals are not available for detection by either platform, the above methods cannot correct for all the bias. Separate estimates of availability can be included in the analysis; alternatively the availability process can be modeled (e.g., Refs 38, 39).

Density Surface Modeling

The standard framework invokes properties of the sampling design to scale up from density in the surveyed transects to density in the study area. However, it may be advantageous to use **model-based methods** for this stage instead, particularly methods where spatially indexed covariates are used and hence a spatial density surface model is fitted. First, a model-based approach allows data collected from nonrandom surveys (platforms of opportunity) to be used; second, animal density may be related to habitat and environmental variables, potentially increasing precision and improving understanding of factors affecting abundance; third, abundance may be estimated for any subregion of interest, by integrating under the fitted spatial density surface; and last, for study areas that are small relative to the width of the transects, there may be information gain by jointly modeling detection probability and spatial variation in density.

One approach [40] is to conceptualize the distribution of animals as an inhomogeneous Poisson process, in which the detection function represents a thinning process. Joint modeling of detectability and density is possible (e.g., Refs 41–43), but in many cases the study area is large compared with the transect width, and there seems little loss in modeling detectability first using the observed distances, and then density variation over the study area conditional on the estimated detection function [3, 40]. Issues exist such as adequately modeling local small-scale density hotspots [44] and dealing with complex spatial topography [45].

Bayesian approaches to density surface modeling have been implemented [46, 47].

Other Extensions

Adaptive sampling [48] (*see* **Adaptive designs**) offers a means of increasing sample size, and hence increasing precision, by concentrating survey effort where most observations occur. Standard adaptive sampling methods can readily be extended to distance sampling surveys [48]. A major practical problem of adaptive sampling is that the required survey effort is not known in advance. Methods have been developed for line-transect sampling to allow a fixed total effort, using a variable adaptation rule [49]. An experimental trial on a survey of harbor porpoise in the Gulf of Maine yielded substantially more detections and somewhat better precision than did conventional line-transect sampling [49].

Distance sampling surveys are sometimes used as part of an experimental design, for example, to compare density between control and treatment plots. Methods have been developed for such situations, which allow for the dependence between sample units when parameters of the detection function are shared between units [50].

Nonrandom placement of samplers requires the use of density surface models for encounter rate, covered above. However, another potential consequence of not using random locations is that the density of objects with respect to distance from the line or point is not necessarily the standard form, and hence, the distribution of observed distances cannot be assumed to tell us directly about detectability. Other auxiliary sources of information about nonstandard density gradients can, however, be incorporated, as demonstrated by Refs. 15, 51 and 52.

Encyclopedia of Environmetrics, Online
This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vad033.pub2

Another assumption of the standard methods is that distances are measured without error. However, if there are errors, but information is available about their form, this can potentially be incorporated in estimation [15, 16, 53].

Although the standard methods are firmly established and perform well in many circumstances, there are still situations where neither they, nor the extensions above, are satisfactory. There is much research underway, and much left to be done, and hence the subject of distance sampling is still a lively one for statistics and ecology.

References

[1] Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L., & Thomas, L. (2001). *Introduction to Distance Sampling*, Oxford University Press, Oxford.

[2] Buckland, S.T., Anderson, D.R., Burnham, K.P., & Laake, J.L. (1993). *Distance Sampling: Estimating Abundance of Biological Populations*, Chapman & Hall, London, reprinted (1999) by Research Unit for Wildlife Population Assessment, St Andrews.

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

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

[5] Hilborn, R. & Walters, C.J. (1992). *Quantitative Fisheries Stock Assessment: Choice, Dynamics and Uncertainty*, Chapman & Hall, New York.

[6] Seber, G.A.F. (1982). *The Estimation of Animal Abundance and Related Parameters*, 2nd Edition, Macmillan, New York.

[7] Schwarz, C.J.& Seber, G.A.F. (1999). A review of estimating animal abundance II, *Statistical Science* **14**, 427–456.

[8] Borchers, D.L., Buckland, S.T., *&* Zucchini, W. (2002). *Estimating Animal Abundance: Closed Populations*, Springer-Verlag, London, p. 314.

[9] Strindberg, S. & Buckland, S.T. (2004). Zigzag survey designs in line transect sampling, *Journal of Agricultural, Biological and Environmental Statistics* **9**, 443–461.

[10] Buckland, S.T., Anderson, D.R., Burnham, K.P., & Laake, J.L. (1998). *Distance sampling, in Encyclopedia of Biostatistics*, P. Armitage & T. Colton, eds, John Wiley & Sons, Ltd, Chichester.

[11] Fewster, R.M., Buckland, S.T., Burnham, K.P., Borchers, D.L., Jupp, P.E., Laake, J.L., & Thomas, L. (2009). Estimating the encounter rate variance in distance sampling, *Biometrics* **65**, 225–236.

[12] Fewster, R.M. (2011). Variance estimation for systematic designs in spatial surveys, *Biometrics* **67**, 1518–1531.

[13] Buckland, S.T. (1992). Fitting density functions using polynomials, *Applied Statistics* **41**, 63–76.

[14] Fewster, R.M., Southwell, C., Borchers, D.L., Buckland, S.T., & 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.

[15] Marques, T.A. (2007). *Incorporating measurement error and density gradients in distance sampling surveys*, PhD thesis, University of St Andrews.

[16] Borchers, D.L., Marques, T.A., Gunnlaugsson, T., & Jupp, P.E. (2010). Estimating distance sampling detection functions when distances are measured with errors, *Journal of Agricultural, Biological, and Environmental Statistics* **15**, 346–361.

[17] Borchers, D.L., Pike, D., Gunlaugsson, T., & Vikingson, G. (2009). Minke whale abundance estimation from the NASS 1987 and 2001 cue counting surveys taking account of distance estimation errors, Vol. **7**, *North Atlantic Sightings Surveys (1987–2001)*, North Atlantic Marine Mammal Commission Publications, pp. 95–110.

[18] Link, W.A. (2003). Nonidentifiability of population size from capture–recapture data with heterogeneous detection probabilities, *Biometrics* **59**, 1123–1130.

[19] Buckland, S.T. (2006). Point transect surveys for songbirds: robust methodologies, *The Auk* **123**, 345–357.

[20] Hiby, A.R. (1985). An approach to estimating population densities of great whales from sighting surveys, *IMA Journal of Mathematics Applied in Medicine and Biology* **2**, 201–220.

[21] Marques, T.A., Thomas, L., Munger, L., Wiggins, S., & Hildebrand, J.A. (2011). Estimating North Pacific right whale (*Eubalaena japonica*) density using passive acoustic cue counting, *Endangered Species Research* **13**, 163–172.

[22] Marques, F.F.C., Buckland, S.T., Goffin, D., Dixon, C.E., Borchers, D.L., Mayle, B.A., & Peace, A.J. (2001). Estimating deer abundance from line transect surveys of dung: sika deer in southern Scotland, *Journal of Applied Ecology* **38**, 349–363.

[23] Stokes, E.J., Strindberg, S., Bakabana, P.C., Elkan, P.W., Iyenguet, F.C., Madzoké, B., Malanda, G.A., Mowawa, B.S., Moukoumbou, C., Ouakabadio, B.S., & Rainey, H.J. (2010). Monitoring great ape and elephant abundance at large spatial scales: measuring effectiveness of a conservation landscape, *PLoS ONE* **5**, e10294. DOI: 10.1371/journal.pone.0010294.

[24] Plumptre, A.J. (2000). Monitoring mammal populations with line transect techniques in African forests, *Journal of Applied Ecology* **37**, 356–368.

[25] Potts J.M. (2011). *Estimating abundance of rare, small mammals: a case study of the Key Largo woodrat (Neotoma floridana small)*, PhD thesis, University of St Andrews.

Encyclopedia of Environmetrics, Online © 2006 John Wiley & Sons, Ltd.

This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vad033.pub2

[26] Marques, T.A., Thomas, L., Ward, J., DiMarzio, N., & Tyack, P.L. (2009). Estimating cetacean population density using fixed passive acoustic sensors: an example with beaked whales, *Journal of the Acoustical Society of America* **125**, 1982–1994.

[27] Kyhn L.A., Tougaard, J., Thomas, L., Duve, L.R., Steinback, J., Amundin, M., Desportes, G., & Teilmann, J. (2012). From echolocation clicks to animal density – acoustic sampling of harbour porpoises with static dataloggers, *Journal of the Acoustical Society of America* **131**, 550–560.

[28] Buckland, S.T., Summers, R.W., Borchers, D.L., & Thomas, L. (2006). Point transect sampling with traps or lures, *Journal of Applied Ecology* **43**, 377–384.

[29] Küsel, E.T., Mellinger, D.K., Thomas, L., Marques, T.A., Moretti, D.J., & Ward, J. (2011). Cetacean population density from single fixed sensors using passive acoustics, *Journal of the Acoustical Society of America* **129**, 3610–3622.

[30] Diggle, P.J. (1983). *Statistical Analysis of Spatial Point Patterns*, Academic Press, London.

[31] Ramsey, F.L., Wildman, V., & Engbirng, J. (1987). Covariate adjustments to effective area in variable-area wildlife surveys, *Biometrics* **43**, 1–11.

[32] Marques, F.F.C. & Buckland, S.T. (2003). Incorporating covariates into standard line transect analysis, *Biometrics* **59**, 924–935.

[33] Marques, T.A., Thomas, L., Fancy, S.G., & Buckland, S.T. (2007). Improving estimates of bird density using multiple covariate distance sampling, *The Auk* **127**, 1229–1243.

[34] Borchers, D.L., Zucchini, W., & Fewster, R.M. (1998). Mark-recapture models for line transect surveys, *Biometrics* **54**, 1207–1220.

[35] Buckland, S.T., Laake, J.L., & Borchers, D.L. (2010). Double-observer line transect methods: levels of independence, *Biometrics* **66**, 169–177.

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

[37] Laake, J.L., Collier, B.A., Morrison, M.L., & Wilkins, R.N. (2011). Point-based mark–recapture distance sampling, *Journal of Agricultural, Biological and Environmental Statistics* **16**, 389–408.

[38] Schweder, T., Skaug, H.J., Langaas, M., & Dimakos, X. (1999). Simulated likelihood methods for complex double-platform line transect surveys, *Biometrics* **55**, 678–687.

[39] Kleppe, T., Skaug, H.J., & Okamura, H. (2010). Asymptotic bias of the hazard probability model under model mis-specification, *Journal of Cetacean Research and Management* **11**(3), 249–252.

[40] Hedley, S.L. & Buckland, S.T. (2004). Spatial models for line transect sampling. *Journal of Agricultural, Biological and Environmental Statistics* **9**, 181–199.

[41] Johnson, D.S., Laake, J.L., & Ver Hoef, J.M. (2010). A model–based approach for making ecological inference from distance sampling data, *Biometrics* **66**, 310–318.

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

[43] Royle, J.A., Dawson, D.K., & Bates, S. (2004). Modelling abundance effects in distance sampling, *Ecology* **85**, 1591–1597.

[44] Skaug, H.J. (2006). Markov modulated Poisson processes for clustered line transect data, *Environmental and Ecological Statistics* **13**, 199–211.

[45] Wood, S.N., Bravington, M.V., & Hedley, S.L. (2008). Soap film smoothing, *Journal of the Royal Statistical Society, Series B* **70**, 931–955.

[46] Niemi, A. & Fernandez, C. (2010). Bayesian spatial point process modelling of line transect survey data, *Journal of Agricultural, Biological and Environmental Statistics* **15**, 327–345.

[47] Royle, J.A. & Dorazio, R.M. (2008). *Hierarchical Inference and Modelling in Ecology*, Academic Press.

[48] Thompson, S.K. & Seber, G.A.F. (1996). *Adaptive Sampling*, John Wiley & Sons, Inc., New York.

[49] Pollard, J.H., Palka, D., & Buckland, S.T. (2002). Adaptive line transect sampling, *Biometrics* **58**, 862–870.

[50] Buckland, S.T., Russell, R.E., Dickson, B.G., Saab, V.A., Gorman, D.G., & Block, W.M. (2009). Analysing designed experiments in distance sampling, *Journal of Agricultural, Biological and Environmental Statistics* **14**, 432–442.

[51] Marques, T.A., Buckland, S.T., Borchers, D.L., Tosh, D., & McDonald, R.A. (2010). Point transect sampling along linear features, *Biometrics* **66**, 1247–1255.

[52] Cox, M.J., Borchers, D.L., Demer, D.A., Cutter, G.R., & Brierley, A.S. (2011). Estimating the density of Antarctic krill (*Euphausia superba*) from multi-beam echo-sounder observations using distance sampling methods, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **60**, 301–316.

[53] Marques, T.A. (2004). Predicting and correcting bias caused by measurement error in line transect sampling using multiplicative error models, *Biometrics* **60**, 757–763.

(*See also* **Abundance: population size and density estimation**; **Line-transect sampling, new approaches**; **Capture–recapture methodology**; **Capture–recapture models, spatially explicit**)

LEN THOMAS, STEPHEN T. BUCKLAND,
KENNETH P. BURNHAM, DAVID R. ANDERSON,
JEFFREY L. LAAKE, DAVID L. BORCHERS, &
SAMANTHA STRINDBERG

This article was published in *Encyclopedia of Environmetrics Second Edition* in 2012 by John Wiley & Sons, Ltd.
DOI: 10.1002/9780470057339.vad033.pub2