Distance Sampling with a Random Scale Detection Function

Cornelia S. Oedekoven Jeffrey L. Laake Hans J. Skaug

Received: date / Accepted: date

Abstract

Distance sampling was developed to estimate wildlife abundance from observational surveys with uncertain detection in the search area. We present novel analysis methods for estimating detection probabilities that make use of random effects models to allow for unmodeled heterogeneity in detection. The scale parameter of the half-normal detection function is modeled by means of an intercept plus an error term varying with detections, normally distributed with zero mean and unknown variance. In contrast to conventional distance sampling methods, our approach can deal with long-tailed detection functions without truncation. Compared to a fixed effect covariate approach, we think of the random effect as a covariate with unknown values and integrate over the random effect. We expand the random scale to a mixed scale model by adding fixed effect covariates.

We analyzed simulated data with large sample sizes to demonstrate that the code performs correctly for random and mixed effect models. We also generated replicate simulations with more practical sample sizes (~100) and compared the random scale half-normal with the hazard rate detection function. As expected each estimation model was best for different simulation models. We illustrate the mixed effect modeling approach using harbor porpoise vessel survey data where the mixed effect model provided an improved model fit in comparison to a fixed effect model with the same covariates. We propose that a random or mixed effect model of the detection function scale be adopted as one of the standard approaches for fitting detection functions in distance sampling.

Keywords: Abundance estimation AD Model Builder Half-normal Harbor porpoise detections Heterogeneity in detection probabilities Mixed effects

1 Introduction

Distance sampling was developed to estimate wildlife abundance from observational surveys with visibility bias (Buckland et al., 2001). This visibility bias may occur in the case that the observer misses objects within the search area owing to imperfect detection. In this paper we present novel analysis methods for estimating detection probabilities that make use of random effects models.

The two most common distance sampling methods are line and point transect sampling. For line transects, the observer travels down the line and records all perpendicular distances from the line to the detections of the species of interest. For point transects, the observer remains at the point for a fixed amount of time and records all radial distances from the point to the detections of the species of interest. For brevity, we will speak of objects below where each object may consist of single animals (or plants) or clusters of these. Here, we assume that all objects on the line or point are detected with certainty.

Using conventional distance sampling (CDS) methods, the first step of analyzing distance sampling data generally consists of fitting a probability density function (pdf) $f(x)$ to the sample of observed distances to infer the detection probability (Buckland et al., 2001). This function is determined by $g(x)$ and $h(x)$, where $g(x)$ is the probability of seeing an object at distance x given the object is at that distance and $h(x)$ is the probability that the object is at distance x. The pdf $f(x)$ is given by:

$$f(x) = \frac{g(x)h(x)}{\int g(u)h(u)du},$$

which is the probability density for seeing an object at x conditional on the fact that it was seen somewhere. Random placement of a sufficiently large number of lines or points within the study area allows us to assume a uniform distribution of objects locally at the line or points. For lines, this means that $h(x) = 1/w$ where w is the strip half-width and for points $h(x) = 2\pi x/(\pi w^2)$ where w is the radius of the circle. Misspecification of $h(x)$ can be caused e.g., by presuming randomly placed transects while surveying along linear features such as roadsides where animals are not evenly distributed with increasing distance from the line. This can lead to bias in estimating detection probability and, hence, to bias in estimating abundance (Marques et al., 2010). However, from here on, we will refer to line transect sampling although the methods we describe are the same for points with the adjustment for a different $h(x)$. With $h(x) = 1/w$, $f(x)$ simplifies to

$$f(x) = \frac{g(x)}{\int_0^w g(u)du}. \tag{1}$$

With the additional assumption that detection at x=0 is perfect (i.e. $g(0)$=1), $f(x)$ evaluated at distance zero is given by:

$$f(0) = \frac{1}{\int_0^w g(u)du}. \tag{2}$$

For n observations from strips of total length L and width $2w$, the estimator of object density within the total search area is:

$$D = \frac{n}{2wLp} = \frac{n}{2wL\int_0^w g(u)\frac{1}{w}du} = \frac{nf(0)}{2L}, \tag{3}$$

where $p = \int_0^w g(u)\frac{1}{w}du$ is the average detection probability. Note that n refers to the number of detected objects. In the case that objects consist of clusters of size larger than one, eq (3) needs to be multiplied with the expected cluster size to estimate density of individuals. Using the design-based approach from Buckland et al. (2001), object abundance in the study area may be obtained by multiplying D from eq (3) with the size of the study area. The quantity $\mu = \int_0^w g(u)du = wp$ is called the effective strip width (ESW), but is actually a half-strip width for each side of the line.

However, when not considering cluster size, p is the only quantity from eq (3) that requires estimation, while n, w and L are known. Hence, it is important to fit a flexible detection function that allows reliable estimation of p. Using CDS methods, this was generally accomplished by comparing the fits of multiple key-adjustment term combinations (see section 2 for details). However, two main methods have been developed that allow modeling heterogeneity in detection probabilities by including observable covariates in the detection function model (Marques and Buckland, 2003) or by using mixture models (Miller and Thomas, submitted).

In the following we begin by summarizing and comparing these existing methods for fitting detection functions (section 2). This sets the stage for section 3.1 where we propose a new method, i.e. the random scale detection function. We discuss the likelihood for this function (section 3.2) and expand the random scale to a mixed scale model (section 3.3). Furthermore, we demonstrate our proposed methods in a simulation study (section 4) and apply the mixed effect approach to harbor porpoise (*Phocoena phocoena*) detections in comparison to the equivalent fixed effect approach (section 5).

2 Existing methods for fitting flexible detection functions

Currently there are three primary ways to fit detection functions for distance sampling data. The most common is the key function and adjustment series described in Buckland et al. (2001). The general formula is:

$$g(x) = \frac{k(x)(1 + \sum_{j=1}^{m} a_j p_j(x))}{k(0)(1 + \sum_{j=1}^{m} a_j p_j(0))}$$

where $k(x)$ is a key function, $p_j(x)$ is a series of adjustment functions with coefficients a_j and m the total number of adjustment terms fitted. The denominator scales the function such that $g(0)$=1 although this denominator is not necessary for fitting because it cancels in eq (1). An example is a half-normal key function and a cosine adjustment series

$$g(x) = \frac{\exp(-(x/\gamma)^2/2)(1 + \sum_{j=1}^{m} a_j \cos(j\pi x/w))}{(1 + \sum_{j=1}^{m} a_j)}$$

where γ is the scale parameter of the half-normal key function. This key-adjustment approach allows for flexible fitting to the observed distances. It does, however, require defining a truncation width (w), imposing non-linear constraints to maintain monotonicity (i.e. $g(x_1) \geq g(x_2)$ for all $w \geq x_2 > x_1$) and ensuring that $1 \geq g(x) > 0$. In addition, it has been shown that fitting of detection functions with long tails is problematic with this approach.

A second approach is to include a vector of explanatory covariates $\boldsymbol{z}$ in the scale parameter of the half-normal or hazard-rate detection function (Marques and Buckland, 2003). An example using a half-normal detection function is:

$$g(x|\boldsymbol{z}) = \exp(-[x/\exp(\boldsymbol{z'\beta})]^2/2) \tag{4}$$

where $\boldsymbol{z}'$ denotes the vector transpose and $\boldsymbol{\beta}$ is a parameter vector of the same length as $\boldsymbol{z}$. In comparison to the previous approach, no adjustment series need be used and the single parameter scale of the half-normal function (or the hazard-rate) is replaced with $\exp(\boldsymbol{z'\beta})$. Hence, the scale of the detection function is adjusted for each detected object depending on the observed covariate values during the detection.

The model in eq (4) is conditional on $\boldsymbol{z}$; hence, it is essential that $\boldsymbol{z}$ is independent of x (i.e., $h(x|\boldsymbol{z}) = h(x)$) (Borchers and Burnham, 2004). An obvious example where this fails is animal behavior that might differ with x (e.g. responsive movement of the animals to the observer). This approach provides monotone detection functions without constraints, does not require truncation and is suitable for fitting long tails. It has the added advantage of providing better small-area estimates of density when the covariates vary spatially (Hedley and Buckland, 2004). On the other hand, the covariate approach does depend on being able to identify and measure covariates that affect detection probability (Marques and Buckland, 2003; Marques et al., 2007).

If there is any remaining lack of fit, the first and second approaches can be combined using covariates in the key function and a series adjustment (Marques et al., 2007, e.g.). However, it is then subject to the same problems as the key-adjustment approach where the constraints may become even more problematic as they depend on the explanatory covariate values. Even if the function is constrained correctly for all observed values of $\boldsymbol{z}$, predictions for unobserved values of $\boldsymbol{z}$ may yield invalid probabilities due to the addition of adjustment functions.

The third approach is rather recent and involves fitting a mixture of m detection functions (Miller and Thomas, submitted) along the lines of Pledger (2000) for capture-recapture models. Here, the detection function can be represented as:

$$g(x) = \sum_{j=1}^{m} \pi_j g'_j(x)$$

where $\sum_{j=1}^{m} \pi_j = 1$ and $g'_j(x)$ is a properly specified detection function. As long as each component detection function is monotone, $g(x)$ will be monotone.

3 Random and mixed scale models

3.1 Random scale detection function

An additional approach we present here is to use random effects in the detection function scale to allow for unmodeled heterogeneity in detection. Consider a half-normal detection function where the scale parameter is modeled by means of an intercept β plus an error term ϵ, varying with detections, normally distributed with zero mean and unknown variance ($\epsilon \sim N(0, \sigma_\epsilon)$):

$$g(x|\epsilon) = \exp(-x^2/(2\gamma(\epsilon)^2)). \tag{5}$$

The scale is now modeled as:

$$\gamma(\epsilon) = \exp(\beta + \epsilon).$$

We assume a normal distribution for ϵ and use $N(\epsilon, 0, \sigma_\epsilon)$ as shorthand for the normal density function evaluated at ϵ with mean zero and standard deviation σ_ϵ. Considering that long-tails may result from some objects with high detection probabilities out to great distances or some conditions under which objects are detectable at great distances, we argue that this random scale will be able to cope with long-tailed detection functions (i.e. with large values for ϵ).

3.2 Likelihood formulation for the random scale model

Using the random scale detection function, the marginalized likelihood for the sample of n observed distances can be derived directly from equations 2.39 and 2.40 in Borchers and Burnham (2004). In comparison with the covariate approach using fixed effects from above, we think of the random effect as a covariate with unknown values and integrate over the random effect. This is accomplished by including an integral over the unknown random effect in both the numerator and denominator:

$$L_g(\beta, \sigma_\epsilon) = \prod_{i=1}^{n} \frac{\int_{-\infty}^{\infty} g(x_i|\epsilon)\, N(\epsilon, 0, \sigma_\epsilon) d\epsilon}{\int_{-\infty}^{\infty} \int_0^w g(u|\epsilon) du\, N(\epsilon, 0, \sigma_\epsilon) d\epsilon}, \tag{6}$$

where the x_i refer to the distances to the detected objects with $i = 1, 2, ..., n$. We denote L_g with subscript g indicating that here we use a properly defined detection function $g(x|\epsilon)$ with $g(0) = 1$ (for comparison see eq (13) Appendix 1, Supporting Information where we present an alternative formulation, L_f where the scale mixture is applied to the probability density from eq (1) rather than to the detection function). In this formulation (eq (6)) we denote the scale intercept with β_g. The numerator of eq (6) is the marginal detection function evaluated at x_i:

$$\int\limits_{-\infty}^{\infty} g(x_i|\epsilon) N(\epsilon, 0, \sigma_\epsilon) d\epsilon, \tag{7}$$

while the denominator of eq (6), divided by w, is the marginal probability that the object was seen within truncation width w:

$$\int\limits_{-\infty}^{\infty} \int_0^w g(u|\epsilon) du\, N(\epsilon, 0, \sigma_\epsilon) \frac{1}{w} d\epsilon. \tag{8}$$

We note that in contrast to point transects, the availability function for line transects $h(x) = 1/w$ from eqs (7) and (8) cancel in eq (6).

3.3 Mixed scale detection function

A mixed effects model in which observed covariates ($\boldsymbol{z}$) are included in the detection function can be accomplished by combining the covariate model from above (eq (4)) with the random scale model (eq (5)) using:

$$\gamma(\epsilon, \boldsymbol{z}) = \exp(\boldsymbol{z'\beta} + \epsilon). \tag{9}$$

where $\boldsymbol{z}$, $\boldsymbol{\beta}$ and ϵ are as before. Note that here the intercept β from eq (5) is replaced with $\boldsymbol{z'\beta}$. The half-normal detection function with a mixed scale can now be written as:

$$g(x|\boldsymbol{z}, \epsilon) = \exp(-[x/\exp(\boldsymbol{z'\beta} + \epsilon)]^2/2). \tag{10}$$

In this case, the likelihood is conditional on the observed covariate values. Building upon the likelihood formulation from eq (6), the likelihood for the sample of n observations is now given by:

$$L_g(\boldsymbol{\beta}, \sigma_\epsilon|\boldsymbol{z}) = \prod_{i=1}^{n} \frac{\int_{-\infty}^{\infty} g(x_i|\boldsymbol{z}, \epsilon)\, N(\epsilon, 0, \sigma_\epsilon)d\epsilon}{\int_{-\infty}^{\infty}\int_0^w g(u|\boldsymbol{z}, \epsilon)du\, N(\epsilon, 0, \sigma_\epsilon)d\epsilon} \tag{11}$$

3.4 Density estimators using a random or mixed scale

Using a random scale detection function, an estimate of object density D can be obtained using eq (8) in place of p in eq (3) giving:

$$D = \frac{n}{2wL\int_{-\infty}^{\infty}\int_0^w g(u|\epsilon)du\, N(\epsilon, 0, \sigma_\epsilon)\frac{1}{w}d\epsilon}.$$

When explanatory covariates are included for the mixed scale approach, the Horvitz-Thompson-like estimator (eq 2.44 in Borchers and Burnham, 2004) can be used to estimate object density:

$$D = \sum_{i=1}^{n} \frac{1}{2wLp_i} = \sum_{i=1}^{n} \frac{1}{2wL\int_{-\infty}^{\infty}\int_0^w g(u|\epsilon, z_i)du\, N(\epsilon, 0, \sigma_\epsilon)\frac{1}{w}d\epsilon}, \tag{12}$$

where for each of $i = 1, 2, ..., n$ objects, 1 is divided by the probability that it is detected p_i, which are then summed up over all n. For the mixed scale approach, the numerator of eq (12) needs to be replaced with s_i, the size of the ith object, in the case that cluster sizes are larger than 1 and density of individuals is estimated.

4 Simulation study

The R package *RandomScale* (`https://github.com/jlaake/RandomScale`) contains code for fitting models using maximum likelihood, for plotting the fitted model and for estimating abundance in the covered area using eq (12) multiplied by $2wL$. Some of the functions of this package are based on the L_g formulation from eq (6), while other functions use L_f, where the scale mixture is applied to the probability density from eq (1). In Appendix S1 (Supporting Information) we define L_f in eq (13) and provide a proof and simulations that show that L_f yields the same MLE as L_g in the case of the half-normal detection function in combination with normal random effects; however, L_f was more stable numerically than L_g in our simulations. There is no guarantee that L_f will approximate L_g for non-Gaussian detection functions, and the method should be regarded as approximate and used with caution in this case.

The underlying programs used to maximize L_g and L_f were developed with the software package ADMB

(Fournier et al., 2012). L_g can also be fitted solely with R code in the package. ADMB allows flexible specification with random effects (Fournier et al., 2012). By default ADMB integrates the likelihood using the Laplace approximation, but for L_g and L_f it was necessary to use the more accurate Gauss-Hermite adaptive quadrature which is also part of ADMB. Some additional C++ code to enable the use of multinomial weights with Gauss-Hermite integration for the random effects is contained in the package. With simulation we compare the results from the R and ADMB code obtained with the two different formulations (Appendix 1, Supporting Information). We used examples with simulated data for random and mixed effects with large sample sizes so the results and comparisons were only slightly affected by sampling variability (Section 4.1). We also provide replicate simulations from various detection functions and compare the results from the half-normal random scale detection function with the hazard rate detection function (Section 4.2). All of the code used in this manuscript is provided in the package (use help(RandomScale)).

4.1 Fitting random and mixed scale detection functions

The following is an example of a mixed effects model that can only be fitted with the ADMB code and L_f (see eq (13) Appendix 1, Supporting Information) in the *RandomScale* package. We simulated distances for 536 detected objects from a half-normal detection function with random scale ($log(\sigma_\epsilon) = -0.5$) truncated at $w = 50$ where the distances of the first 438 detected objects were from a population with $N = 2000$ with a larger scale intercept $\beta_g = 2$ compared to the last 98 objects from a population of $N = 1000$ with $\beta_g = 1$. The subsets of the data are distinguished by including a two-level factor covariate with values 0 and 1 for the first and second subset, respectively. All objects have the same random effect distribution. We fit models to the data with the covariate (mixed model) and without the covariate (random model), both using L_f.

The fit of the detection functions averaged over all data look similar for both models (Fig. 1) but the model with the covariate is clearly better with a $\triangle$AIC of 33.34. The estimate of abundance from the model with the covariate is 3212 (se=261.6) and without the covariate is 3267 (se=276). For the mixed effect model the estimated standard deviation (0.57) is smaller than the same quantity for the random effect model (0.67) which absorbs the heterogeneity due to the missing covariate into the random effect.

The total abundance estimates are similar, but when abundance is estimated for each type of object (with covariate: 2176 (se = 182.9) and 1036.1 (se = 155.1); without covariate: 2669 (se = 231) and 597.2 (se = 70.9) the importance of including the covariate becomes obvious. When using the model with the covariate, the model fits tighter to the observed data (Fig. 2) in particular for the subset of the data with the smaller sample size, i.e. the subset of the data with covariate value = 1. On the other hand, for the model without the covariate predicted detection probabilities are too low for distances near zero and too high for larger distances which results in an underestimate of abundance of those with covariate value 1. Likewise, the estimated abundance for objects with covariate value 0 is too high.

4.2 Simulation comparison with hazard rate

The random scale half-normal detection function has two parameters and is thus more flexible than a half-normal with a single parameter. The hazard rate which is often used to represent detection functions also has two parameters, so a simulation comparison of the alternative two-parameter models is worthwhile. We simulated data from a t-distribution with 3, 5, and 10 degrees of freedom, also from a random scale half-normal (β_g = -0.5; σ_ϵ = 0.5) and from a hazard rate ($g(x) = 1 - exp(-(x/\sigma)^{-p})$; $\sigma = 0.7$; $p = 2.5$). We simulated 500 replicates for each detection function with expected sample sizes of 60-90 and 130-180 by varying the true abundance (N) for the scenario. The distances were generated using rejection sampling with w=40 and the parameters were chosen so the largest observed distance would not exceed 20. The number detected (n) and the largest observed distance (w) would vary so they are summarized as means in the results (Table 1). For each data set we fit the random scale half-normal with the ADMB code from the *RandomScale* package using L_g eq (6) and L_f (see eq (13) from Appendix 1, Supporting Information) and the hazard rate detection function using the *mrds* package (Laake et al., 2013) using a transect width (w) equal to the largest observed distance and twice the largest observed distance to approximate an infinite width. We measured the proportion of replicates in which Akaike's Information Criterion (AIC) was smaller for L_g versus L_f and vice versa. Even though they should produce the same likelihood value we have found that our ADMB implementation of L_f has better convergence than L_g. We also compared the proportion of replicates in which AIC was smaller for L_f versus the hazard rate model. For the random-scale half-normal model we computed the percent relative bias (PRB=$100(N - \hat{N})/N$) and its simulation standard error and root mean squared error ($\sqrt{(var(\hat{N}) - (\bar{\hat{N}} - N)^2)}$ expressed as a percentage of N. We also computed the same quantities using the estimate from the model with the smallest AIC for each replicate. In comparing abundance estimates to the true value we used N/w which scales with the width of the transect that was used.

As expected, the random scale half-normal and hazard rate did best when the data were generated from the fitted model. In general, when generating data under a different distribution, the hazard rate tended to underestimate and the random scale half-normal model tended to over-estimate abundance. However, when AIC was used to select the best model, the average bias was typically less than 5% and often within simulation error. The bias of the average was largest when data were generated from the hazard rate, because the random scale half-normal tended to over-estimate the intercept and abundance because the hazard rate detection function has a long tail and then flattens near x=0. For these same scenarios, the ADMB code for L_g had substantial problems with convergence in comparison to L_f. In fewer than 0.2% of the 10000 simulations did the L_g code produce a smaller negative log-likelihood than L_f. When w was set to twice the largest observed distance, the random scale half-normal performed better with less bias and the hazard rate performed worse with more negative bias except when the hazard rate was used to generate the simulated distances. In real data applications we never know the true detection function, so it is useful to have a set of models to examine and use a model selection criterion like AIC.

5 Application to harbor porpoise data

In 2002, a small boat survey for harbor porpoise (*Phocoena phocoena*) was conducted in waters of the Strait of Juan de Fuca and around the San Juan Islands in Washington state, USA. Three observers surveyed along a set of systematically placed lines with an observer standing on the bow and at the starboard and port sides. When harbor porpoise were detected, the angle from the line to the harbor porpoise was measured with an angle board and the radial distance to the detection was estimated visually. Observers were trained and tested in visual distance estimation but for this example, we ignore the error in distance estimation. The angle and radial distance was converted to perpendicular distance. In addition to distance, the number of harbor porpoise (size) was recorded for each detection.

A total of 477 harbor porpoise groups were detected with group size varying from 1 to 6. We fitted a model with a half-normal detection function and used group size as a covariate. We fitted a fixed effect detection function with the *mrds* package (Laake et al. 2013) and a mixed effects detection function with the *RandomScale* package. The *mrds* package requires a finite width, so to make the AIC values equivalent we set w=443.2 the largest distance for each analysis. The fit of the detection functions (Table 2) look similar (Fig. 3) but the model that includes the random effect is slightly better with a $\triangle$AIC of 2.6. The estimate of harbor porpoise group abundance within the 886.4 meter strip is 1243 (se = 59) for the fixed effect model and 1360 (se = 93) for the mixed effect model. The higher abundance estimate resulted from the slightly steeper estimated detection function (Fig. 3).

6 Discussion

Incorporating a random effect in the scale of the detection function extends the covariate approach of Marques and Buckland (2003) to enable modeling of additional unspecified and typically unknown sources of heterogeneity in detection probability. This removes the need to select an arbitrary truncation width which is typically needed for the CDS key-adjustment function fitting (Buckland et al., 2001). The random and mixed effects modeling can be used with other detection functions such as the hazard function (Buckland et al., 2001) as long as the parametrization includes a scale parameter (x/σ); although it could also be applied to the shape parameter in the hazard function. The models can be easily extended to allow covariates to be included for the random effects standard deviation σ_ϵ. For example, heterogeneity in detection probability may be enhanced or reduced as a function of weather, habitat or other covariates. We propose that a random or mixed effect model of the detection function scale be adopted as one of the standard approaches for fitting detection functions in distance sampling.

7 Ackwlowledgements

We thank Steve Buckland for reviewing a nearly final draft of the paper. Cornelia Oedekoven was supported by a studentship jointly funded by the University of St Andrews and EP-SRC, through the National Centre for Statistical Ecology (EPSRC grant EP/C522702/1). Hans Skaug thanks the Center for Stock Assessment Research for facilitating his visit to University of California, Santa Cruz.

References

Borchers, D. and Burnham, K. (2004). *Advanced Distance Sampling.*, chapter General formulation for distance sampling. Oxford University Press, Oxford.

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

Fournier, D. A., S. H. J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M. N., Nielsen, A., and Sibert, J. (2012). Ad model builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software* **27,** 233–249.

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

Laake, J., Borchers, D., Thomas, L., Miller, D., and Bishop, J. (2013). *mrds: Mark-Recapture Distance Sampling.* R package version 2.1.4.

Marques, F. F. C. and Buckland, S. T. (2003). Incorporating covariates into standard line transect analyses. *Biometrics* **53,** 924–935.

Marques, T. A., Buckland, S. T., Borchers, D. L., Tosh, D., and McDonald, R. A. (2010). Point transect sampling along linear features. *Biometrics* page no.

Marques, T. A., Thomas, L., Fancy, S. G., and Buckland, S. T. (2007). Improving estimates of bird densities using multiple covariate distance sampling. *The Auk* **124,** 1229–1243.

Miller, D. L. and Thomas, L. Mixture models for distance sampling detection functions. Unpublished manuscript.

Pledger (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics* **56,** 434–442.

Table 1: Percent relative bias (PRB) and root mean square error (RMSE) as proportion of true abundance for random scale half-normal and hazard rate detection function models for distances generated from t-distribution, random scale half-normal and hazard rate detection functions. Each value is the summary for 500 replicate simulations; $\bar{w}$ and $\bar{n}$ are the mean truncation distance and mean sample size. The subscripts F, G and HR refer to L_f, L_g and the hazard rate. AVG subscript represents the values in which the estimate was generated from the model that had the lowest AIC for each replicate. Data were generated from a t-distribution with listed degrees of freedom (t(df)), a random scale half-normal (hn) with β_g = -0.5 and $\sigma_\epsilon = 0.5$, and a hazard rate (hr; $g(x) = 1 - exp(-(x/\sigma)^{-p}$ with $\sigma = 0.7$ and $p = 2.5$).

Function	$\bar{w}$	$\bar{n}$	PRB_F	PRB_{HR}	PRB_{AVG}	$se(PRB_{AVG})$	$AIC_F < AIC_{HR}$	$AIC_F < AIC_G$	$AIC_G < AIC_F$	$RMSE_F$	$RMSE_{HR}$	$RMSE_{AVG}$
t(df=3)	8.80	136.46	8.67	-9.80	-2.25	0.84	0.39	0.14	0.00	16.67	18.61	18.87
	17.23	136.46	6.23	-13.01	-0.97	0.80	0.63	0.16	0.00	17.97	17.02	17.81
t(df=5)	5.23	132.30	7.46	-7.92	-0.86	0.77	0.46	0.06	0.00	15.73	16.89	17.13
	10.47	132.30	4.04	-12.50	-0.21	0.68	0.76	0.09	0.00	17.34	14.62	15.29
t(df=10)	3.70	128.09	4.88	-7.98	-2.11	0.68	0.46	0.08	0.00	14.97	14.47	15.26
	7.41	128.09	1.54	-13.58	-1.19	0.60	0.84	0.07	0.00	17.82	12.74	13.57
t(df=3)	6.86	68.18	15.08	-3.78	2.11	1.34	0.29	0.10	0.00	23.95	33.18	30.01
	13.63	68.18	8.84	-10.11	0.43	1.13	0.57	0.10	0.00	21.95	26.57	25.24
t(df=5)	4.32	66.16	11.24	-3.64	1.74	1.14	0.34	0.06	0.01	22.19	27.56	25.66
	8.64	66.16	4.51	-11.48	-0.59	1.01	0.70	0.11	0.01	21.69	22.20	22.59
t(df=10)	3.24	63.65	10.92	-1.82	2.37	2.55	0.34	0.03	0.00	23.77	27.04	25.58
	6.48	63.65	3.62	-12.01	0.00	2.12	0.77	0.08	0.03	21.33	20.14	21.23
hn	4.52	172.60	3.33	-12.28	-2.51	0.69	0.62	0.06	0.00	17.21	14.27	15.68
	9.03	172.60	0.02	-16.62	-2.03	0.63	0.89	0.12	0.00	20.11	13.43	14.32
hn	3.79	86.50	8.28	-7.32	-0.80	1.02	0.39	0.04	0.00	20.00	23.58	22.85
	7.59	86.50	2.09	-13.69	-1.41	0.91	0.79	0.07	0.00	21.57	19.87	20.47
hz	15.67	183.22	20.22	2.87	3.11	1.09	0.01	0.80	0.00	10.96	28.48	11.34
	28.48	183.22	19.44	1.94	4.28	1.30	0.09	0.82	0.00	10.68	27.64	13.72
hz	12.57	91.44	24.97	4.01	5.32	0.87	0.05	0.56	0.00	17.98	35.94	20.17
	22.68	91.44	23.32	2.40	5.76	0.90	0.14	0.55	0.00	17.12	34.27	20.93

Table 2: Parameter estimates, standard errors for fixed (AIC=5375.7) and mixed effect (AIC=5373.1) models fitted to harbor porpoise vessel survey data.

	Fixed-effect		Mixed-effect	
	Estimate	Std error	Estimate	Std error
Intercept	4.772	0.069	4.722	0.096
Size	0.084	0.037	0.088	0.052
$\log(\sigma_\epsilon)$			-1.250	0.304

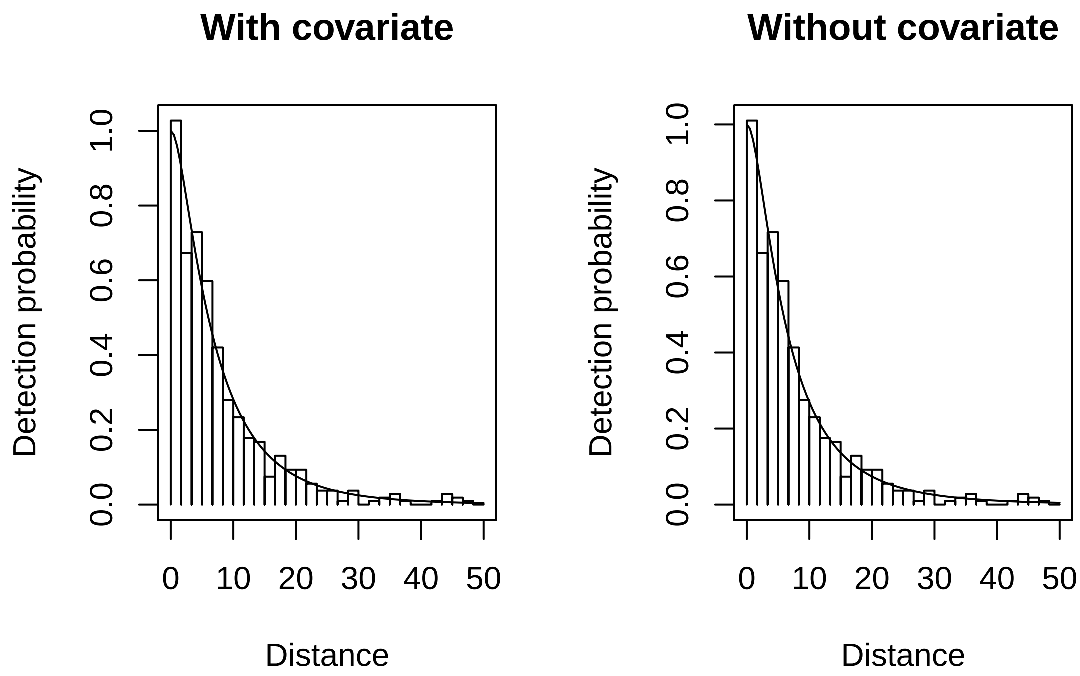

Figure 1: Average detection functions fitted to simulated data with ADMB code using L_f (Appendix 1, Supporting Information) with and without the covariate.

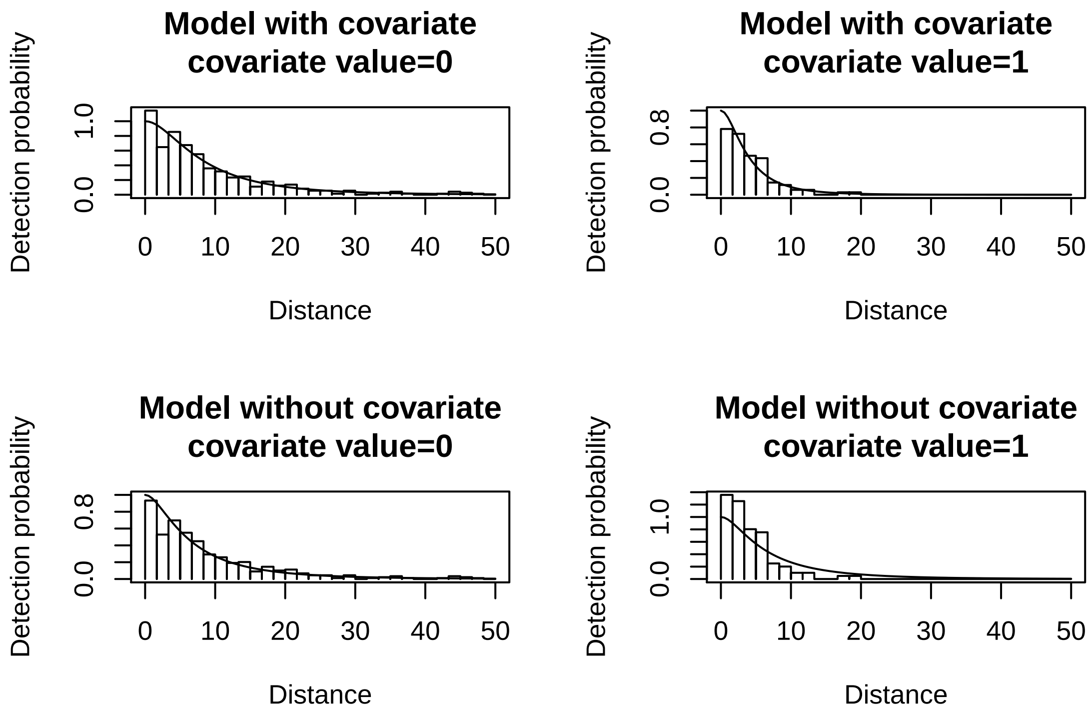

Figure 2: Detection functions fitted to simulated data with ADMB code using L_f (Appendix 1, Supporting Information) with (top row of plots) and without the covariate (bottom row of plots) shown for covariate values 0 and 1.

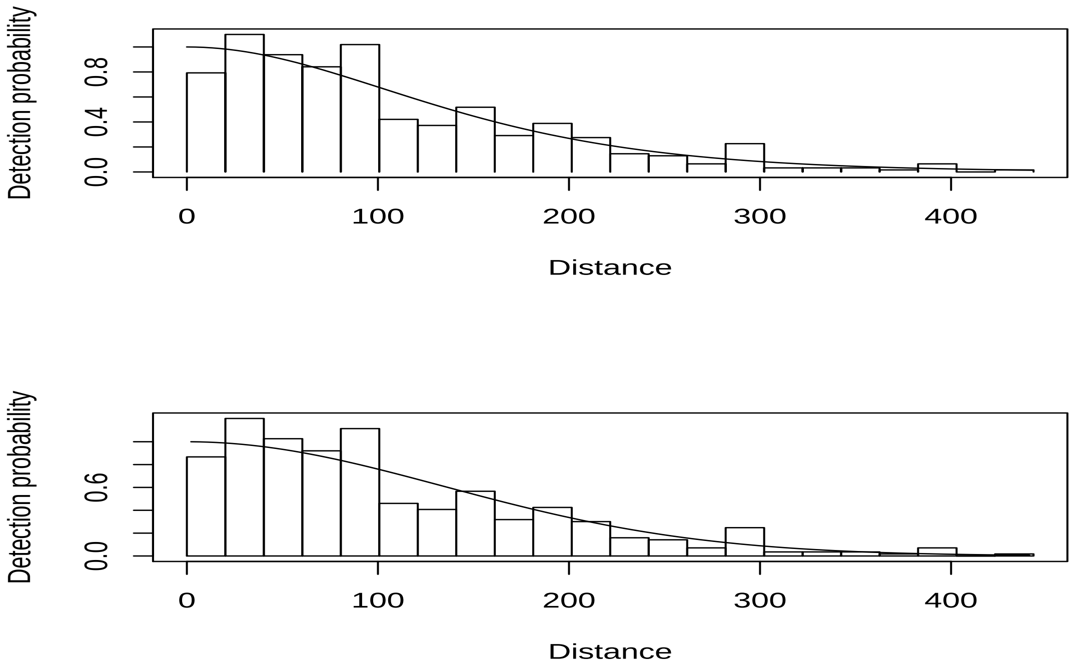

Figure 3: Detection functions fitted to harbor porpoise vessel survey data. The upper panel is the mixed effects model and lower panel is the fixed effects model. Both include group size as a covariate.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix S1: Comparison between two likelihood formulations for the random scale detection function