# Incorporating animal movement into distance sampling

R. Glennie[1], S. T. Buckland[1], R. Langrock[2], T. Gerrodette[3], L. T. Ballance[3], S. J. Chivers[3], M. D. Scott[3], and W. F. Perrin[3]

[1]University of St Andrews, Centre for Research into Ecological and Environmental Modelling
[2]Bielefeld University
[3]Southwest Fisheries Science Center, NOAA Fisheries

## Abstract

Distance sampling is a popular statistical method to estimate the density of wild animal populations. Conventional distance sampling represents animals as fixed points in space that are detected with an unknown probability that depends on the distance between the observer and the animal. Animal movement, responsive or non-responsive to the observer, can cause substantial bias in density estimation. Methods to correct for responsive animal movement exist, but none account for non-responsive movement independent of the observer. Here, an explicit animal movement model is incorporated into distance sampling, combining distance sampling survey data with independently obtained animal telemetry data. A detection probability that depends on the entire unobserved path the animal travels is derived in continuous space-time. The intractable integration over all possible animal paths is approximated by a hidden Markov model. A simulation study shows the method to be negligibly biased (less than 5%) in scenarios where conventional distance sampling overestimates abundance by up to 100%. The method is applied to a line transect survey of spotted dolphins (*Stenella attenuata attenuata*) in the eastern tropical Pacific.

## 1 Introduction

Distance sampling is a statistical method used to estimate the population density of wild animals (Buckland et al., 2015). It is applied to a wide variety of taxa, e.g., seabirds, cetaceans, primates, and ungulates. Many conservation and management studies depend on the accuracy of distance sampling inference; yet, the statistical method relies on a key assumption that is significantly violated in many applications.

Distance sampling is a snapshot method: the survey is assumed to occur instantaneously. Animals are idealised as static points that are detected with unknown probability by an observer, who stands at a point or traverses a line within the study region. Surveyed transects, lines or points, are placed according to a randomised design such that animals are distributed independently of the observer. For line transects, animals are distributed uniformly around the line, in point transects they follow a triangular distribution. Given this, the decline in the number of detections as distance from the observer increases is solely due to a change in the probability of detection; thus, the recorded locations are used to estimate this probability and, ultimately, animal density. In short, distance sampling is a thinned point process model with unknown thinning probability (Hedley and Buckland, 2004; Yuan et al., 2016). The assumption that the survey is a temporal snapshot of the animal population is central to distance sampling theory; the method, however, is applied to surveys of mobile animal populations where transects are surveyed over a time interval within which animals may have moved a significant distance.

Animal movement can be in response to an observer's presence (Turnock and Quinn, 1991), attraction or avoidance, or can be of the animal's own accord, independent of the observer. Responsive movement is a well-known problem, and specific survey techniques, searching further along line transects to see animals before they respond or remaining at point transects long enough for animals to resume normal behaviour, are recommended to mitigate bias in density estimates (Buckland et al., 2005). Furthermore, double-observer methods exist that can account for responsive movement (Palka and Hammond, 2001). In comparison, movement independent of the observer has received little attention. For point transects, a snapshot method is recommended to reduce bias in estimates (Buckland, 2006); however, many surveys do not employ this method, and it does not suit technological advances where observation technology may survey a point for a considerable time. Splitting continuous surveying periods at points into discrete snapshots involves subjective judgments that can affect the inferences obtained (Howe et al., 2017). Alternatively, cue-counting is used as it is not biased by non-responsive animal movement, but this relies on the animal population having a clearly defined cue. For line transects, a rule-of-thumb, based on a limited simulation study, deems surveys on animals that move at less than half the observer's speed to be free of substantial bias (Hiby, 1982); yet, observer speed is often constrained by the transport chosen and the terrain covered. Thus, density can be unavoidably overestimated due to animal movement. This overestimation is not caused by counting the same animal more than once, but by more animals entering the transect from outside and recorded locations leading to a biased estimated detection function. Surveys of mobile animals record greater numbers of unique individuals, compared to a hypothetically immobile population, and animals are recorded closer to the observer; both

effects led to positive bias in density estimation (Glennie et al., 2015). This calls into question inference drawn from surveys where animal movement is undeniable, and precludes the use of distance sampling on populations of fast-moving animals and on studies where transects are surveyed over a long time period.

Previous work has considered only how movement affects the number of animals seen, not where they are seen (Yapp, 1956). Random encounter models (Lucas et al., 2015), where animals are assumed to move in randomly-orientated straight lines at constant speed (Hutchinson and Waser, 2007), can provide estimates of density, corrected for movement, given the count of animals seen and an independent estimate of animal speed. A detection probability can also be included, but must be ascertained independently. An advantage of distance sampling is that the detection probability can be estimated from the data. Yet, this probability, when movement is admitted, depends on the entire path the animal has travelled whilst the transect is surveyed; this path, other than the single location observed when the animal is detected, is unobserved. Thus, any estimation of detection probability that accounts for animal movement must include a continuous space-time hidden process to describe the animal's trajectory. Furthermore, to calculate the proportion of animals never seen, the method must average over all possible animal trajectories. This can be achieved by specifying an explicit model for the encounter process (Gurarie and Ovaskainen, 2012).

Hidden Markov models (HMMs) (Zucchini and MacDonald, 2009) are used for time series data that arise from an unobserved (or partially observed) stochastic spatial process. In particular, HMMs are used to analyse animal telemetry data (Langrock et al., 2012), animal locations recorded over time, where the paths taken by animals between recorded locations are averaged over according to the movement model specified. Distance sampling observations consist of a single such recorded location and the animal's path until detection is unobserved; thus, if independent information on animal movement is collected, distance sampling can be viewed as a HMM, where animal paths are a hidden process and detection is the observed process. Tagging and tracking of animals is becoming more common as the technology reduces in price and size. HMMs can allow this auxiliary information to improve distance sampling estimation. Pedersen et al. (2011) developed a spatial HMM with an unobserved, diffusive movement process which is described by a stochastic partial differential equation. The intractable continuous-spacetime likelihood is approximated by discretising space into a large number of spatial cells (Eydeland, 1994); the computations involved are costly, constraining the level of discretisation attainable.

Here, a spatial HMM that incorporates animal movement into distance sampling using independently obtained animal movement data is presented. The continuous-spacetime likelihood and its discrete approximation are described. A simulation study compares this method to conventional distance sampling, and the method is applied

to a line transect survey of spotted dolphins (*Stenella attenuata attenuata*) in the eastern tropical Pacific (ETP) (Gerrodette and Forcada, 2005; Gerrodette et al., 2008).

# 2 Methods

Suppose $n$ animals are detected in total over the survey. The goal is to estimate the total abundance in the survey region, $N$. In distance sampling surveys, observers search transects and record the location of any animal they encounter. In line transect sampling, the perpendicular distance from the line to each encounter is recorded; in point transects, the radial distance is recorded. Use of two-dimensional location data, forward distance on line transects and angle of detection on points, is seldom used (Borchers and Cox, 2017). Furthermore, the time of a detection, though routinely recorded, is not used. Here, a model is developed that uses both pieces of information: for the $i^{\text{th}}$ detected animal, let $\vec{x}_i$ be the two-dimensional location of the animal, when detected, relative to the observer and $t_i$ be the time between the observer beginning to survey the transect that animal $i$ was seen on and the time animal $i$ was detected.

Recording the single location an animal is encountered provides no information about how the animal moves: independent data is required on animal movement. Suppose $m$ animals are tracked or tagged and their movement paths recorded over time. It is assumed the movement of these tagged animals is representative of the movement of any animal in the study area. Note it is not required that the tagged animals be members of the surveyed population.

## 2.1 Model

In conventional distance sampling (CDS), the probability density function (*pdf*) of the recorded animal locations is estimated. A detection function, $g(x)$, is defined as the conditional probability an animal is detected given it resides at location $x$. For line transects, location is defined as the perpendicular distance the animal is from the line; for point transects, it is defined as the radial distance from the point. The probability density of the observed distances is then given by

$$f(x) = \frac{g(x)\lambda(x)}{\int_{\mathcal{P}} g(x)\lambda(x) \, \mathrm{d}x}$$

where $\mathcal{P}$ is the set of all animal locations and $\lambda$ is the probability density function of the animal's location. CDS makes the design-based assumption that transects are placed according to a randomised scheme. This implies for line transects that $\lambda$ is the density of a uniform distribution over $\mathcal{P}$ and for point transects a triangular distribution over $\mathcal{P}$.

A convenient functional form is chosen for the detection function and its parameters estimated by maximum likelihood. This form is chosen *ad hoc* to be half-normal or exponential with some polynomial adjustments. Alternatively, an explicit model for the detection process can be

specified by a two-dimensional hazard-rate function $h(x, t)$ (Borchers and Cox, 2017; Skaug and Schweder, 1999), which describes the detection rate of an animal residing at point $x$ at time $t$. This is equivalent to a survival process where death is interpreted as detection and the detection intensity varies over time and space. In practice, CDS analyses do not use the time of detections, a detection function can be derived from a given hazard by integrating over the time interval that an animal is at risk of detection. Nonetheless, here, the term CDS is used to refer to 2D hazard models also. The hazard commonly depends on the radial distance between the observer and the animal, $r(\vec{x}, t)$, such that the hazard is infinite at zero radius and decreases with increasing radius.

Here, the recorded detection times and the two-dimensional recorded location are used to estimate the search process. Detection times are required since animals recorded at later times on the transect have had longer to move and so may have originated at a further distance. Rather than condition on the animal residing at a single fixed point, the detection probability is derived conditional on the animal travelling a fixed path over space. Given an animal travels a path $\vec{x}$ and is seen at location $\vec{x}_\tau$ at time $\tau$, the conditional *pdf* is

$$g(\vec{x}, \tau) = S_\tau(\vec{x}) h(\vec{x}_\tau, \tau)$$

where

$$S_t(\vec{x}) = \exp\left( - \int_0^t h(\vec{x}_s, s) \, \mathrm{d}s \right)$$

is the probability of the animal eluding detection until time $t$. Notice, detection probability now depends on time and the entire trajectory of the animal. For brevity, we term the model presented here MDS: movement in distance sampling.

Similar to CDS, if in place of assuming the distribution of animals, we assume the movement process is known, and in place of conditioning on the location of an animal, we condition on the path an animal has taken, the *pdf* of the observed encounter on a transect of duration $T$ is given by:

$$f(\vec{x}, \tau) = \frac{g(\vec{x}, \tau) \Lambda(\vec{x})}{\int_\chi \int_0^T g(\vec{y}, t) \, \mathrm{d}t \, \mathrm{d}\Lambda(\vec{y})} \tag{1}$$

where $\Lambda$ is the probability measure over all (measurable) paths, $\chi$, that an animal could have taken, that is, $\Lambda(\vec{x})$ is the probability of path $\vec{x}$. The denominator is the probability an animal is seen at some time on the transect and is required as we do not observe those animals that were never encountered.

Yet, the paths of animals are unobserved and distance sampling surveys provide no information on how animals move. Thus, independent animal movement data is required to determine $\Lambda$. Here, animal movement is described by diffusion with average speed parameter $\nu$. This makes $\Lambda$, the probability of a given path occurring, mathematically tractable: it is a Gaussian probability distribution. For telemetry data, one can condition on the initial location of the animal and compute the likelihood, $\mathcal{L}_\Lambda$

easily (Okubo and Levin, 2013); for distance sampling surveys, the initial locations are assumed to be independently distributed with respect to the transect.

From the distance sampling survey, given animal $i$ was recorded in location $\vec{x}_i$ at time $\tau_i$, the likelihood for the detection parameters, $\boldsymbol{\theta}$, is obtained by averaging over all possible animal paths:

$$\mathcal{L}_{\boldsymbol{\theta}} = \prod_{i=1}^n \frac{\int_{\chi_i} g(\vec{x}, \tau_i) \, \mathrm{d}\Lambda(\vec{x})}{p_i}$$

where $\chi_i$ is the space of all measurable paths that pass through location $\vec{x}_i$ at time $\tau_i$, and $p_i = 1 - \int_\chi S_{T_i}(\vec{x}) \, \mathrm{d}\Lambda(\vec{x})$ is the probability the encounter with animal $i$ occurs at some time when the transect is surveyed for total time $T_i$.

Assuming the distance sampling survey and the animal telemetry data are independent, the combined likelihood $\mathcal{L}_{\boldsymbol{\theta}, \boldsymbol{\lambda}} = \mathcal{L}_{\boldsymbol{\theta}} \mathcal{L}_{\boldsymbol{\lambda}}$ can be maximised to obtain maximum likelihood estimates, $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\lambda}})$. Methods in section 2.2 describe the approximations used to compute the likelihood. Maximising the combined likelihood means that uncertainty in the movement parameters is propagated to the density estimation. Abundance can be estimated using either approach already available in distance sampling: a model-based estimator or a Horvitz-Thompson-like estimator.

The Horvitz-Thompson-like estimator of abundance is

$$\hat{N} = \sum_{i=1}^n \frac{1}{\hat{p}_i}$$

A sandwich estimator for the variance of $\hat{N}$ can be derived analogously to the estimator used in CDS (Fewster et al., 2008).

Alternatively, an explicit model for abundance can be integrated (Buckland et al., 2016). For example, the Poisson process:

$$\mathcal{L}_N = \frac{(N p_{enc})^n \exp(-N p_{enc})}{n!}$$

where $p_{enc}$ is the average probability of detection during the survey. This gives a model-based estimate of abundance by maximising the likelihood:

$$\mathcal{L}_{\boldsymbol{\theta}, \boldsymbol{\lambda}, N} = \mathcal{L}_{\boldsymbol{\lambda}} \frac{N^n \exp(-N p_{enc})}{n!} \prod_{i=1}^n \int_{\chi_i} g(\vec{x}, \tau_i) \, \mathrm{d}\mathbb{P}(\vec{x})$$

## 2.2 Computation

The continuous space-time likelihood is analytically intractable. The integral over all possible animal paths can be approximated by quadrature. A buffer region around each transect is discretised into $K$ cells of length and width

$\Delta x$ and time is discretised into steps of duration $\Delta t$. Animal movement occurs at these time-steps, thus animals are stationary within time-steps.

Notice that all integrals to be approximated are of the form

$$\int_\chi S_t(\vec{\mathbf{x}}) \, d\Lambda(\vec{\mathbf{x}})$$

for some time $t$. Approximating animal movement by jumps at each time-step, the function $S$ can be separated into components:

$$S_t(\vec{\mathbf{x}}) = s_0(\vec{x}_0) s_{\Delta t}(\vec{x}_{\Delta t}) s_{2\Delta t}(\vec{x}_{2\Delta t}) \dots s_t(\vec{x}_t)$$

where $s_l(\vec{x}) = \exp\left(-\int_l^{l+\Delta t} h(\vec{x}, u) \, du\right)$ is the probability of eluding detection during a single time-step given the animal is at location $\vec{x}$. Notice, observer movement can be integrated analytically within each time-step, thus only animal movement is discretised. Let $\mathbf{P}_t$ be a $K \times K$ diagonal matrix with $k^{\text{th}}$ diagonal $s_t(\vec{y}_k)$ where $\vec{y}_k$ is the centre location of spatial cell $k$.

The integration over all paths is replaced with a sum over all paths on the discrete grid. When discretised, diffusive animal movement can be approximated by a continuous-time Markov chain (Pedersen et al., 2011). The transition rate matrix of this chain, $\mathbf{G}$, is block-Toeplitz in structure, and computing the matrix exponential, to obtain the transition probability matrix $\mathbf{\Gamma} = \exp(\mathbf{G}\Delta t)$, is computationally demanding.

The integral is approximated by a spatial HMM likelihood:

$$\int_\chi S_t(\mathbf{X}) \, d\Lambda = \mathbf{p}_0 \mathbf{P}_0 \mathbf{\Gamma} \mathbf{P}_1 \mathbf{\Gamma} \dots \mathbf{P}_t \mathbf{\Gamma} \mathbf{1} \qquad (2)$$

where $\mathbf{p}_0$ is $1 \times K$ row vector representing the initial distribution of animals on the grid with respect to the transect and $\mathbf{1}$ is a $K \times 1$ column vector of ones.

In distance sampling surveys, the range of detection distances can be small compared to the distances an animal can travel during a survey period. The former makes it necessary that $\Delta x$ be adequately small while the latter requires the buffer around the transect, and so the grid, to be large. Ultimately, $K$ is large. This makes the matrix calculations in equation (2) computationally demanding. The block-Toeplitz structure of $\mathbf{G}$ can be exploited to accelerate computations using a 2D discrete Fourier transform (Lee, 1986) and the Krylov subspace approximation to the matrix exponential (see Appendix).

For spatial HMMs, there is no analytical bound on the error of this approximation. Practical advice is to reduce the discretisation until the inference obtained no longer significantly changes.

## 2.3 Simulation study

A simulation study is conducted to demonstrate the performance of MDS compared to CDS for two particular distance sampling surveys. The magnitude of the bias in CDS depends on the relationship between relative animal speed, transect width, and the shape of the detection function (Glennie et al., 2015). This simulation study considers the effect of animal speed when all other factors are fixed.

A study population of 100 animals in 100 square kilometres is simulated. Animals move according to a diffusion process with average speed varying from 0.5 metres per second to 4.0 metres per second.

Two distance sampling surveys were simulated on this population: a line transect study and a point transect study. The line transect survey consists of 50 transects of length 1 kilometre; the observer traverses each line at speed 1 metre per second. Hence, simulation scenarios cover relative animal speeds of 50% to 400% the speed of the observer. For the point transect survey, 100 points were surveyed, each for 5 minutes.

In both surveys, the hazard of detecting an animal at a radial distance $r$ is given by $\alpha r^{-\beta}$ for parameters $\alpha, \beta > 0$. Detection parameters were chosen such that for a hypothetically immobile animal population, the effective area searched was approximately 0.015 square kilometres. This corresponds to a line transect with half-width 30 metres and point transect with radius 100 metres.

The independent animal telemetry data required was simulated from the diffusion process, recording the location of ten tagged animals every minute for one hour.

One hundred simulations were performed. A distance sampling model with no animal movement was fit to each simulated data set. For fair comparison, a two-dimensional hazard was used in this model also, unlike CDS models where a one-dimensional hazard is commonly used. A MDS model, as presented in this paper, was fit to each data set with the auxiliary movement data.

The relative bias, mean square error and confidence interval coverage was estimated for each scenario, with and without movement incorporated.

## 2.4 Application: spotted dolphins

The presented method is applied to a 2006 shipboard line transect survey conducted in the eastern tropical Pacific on spotted dolphins, estimating the abundance within the core area, as defined by Gerrodette and Forcada (2005).

Dolphin schools are treated as the individual unit of detection and group size estimated separately. This is a standard approach in CDS; however, incorporating movement makes the approach more questionable. In particular, we assume that schools do not fuse or break-up during the time the observer surveys each transect. Furthermore, we assume the movement model, informed by tags on single individuals, describes the movement of a school as a whole.

The radial distance and angle to each detected spotted dolphin school was recorded. Only sightings in Beaufort state 2 or less were retained. Once an encounter occurs, observers cease to survey and approach the detected school in 'closing mode'. The times when surveying ceased and resumed are recorded. This affects the detection probability. Short breaks in effort cannot be ignored, nor can

the transect be subdivided. The former leads to underestimation of detection probability and the latter leads to overestimation. The following hazard accounts for this interrupted effort:

$$h(\vec{x}, t) = \alpha r(\vec{x}, t)^{-\beta} e(t)$$

where $r(\vec{x}, t)$ is the radial distance at time $t$ between the observer and location $\vec{x}$, and $e(t) = 1$ when the observer is on effort at time $t$ and zero otherwise. Thus, animal movement during off-effort time is accounted for.

The location of the ship was recorded every ten minutes. It is assumed the ship travels in a straight line at a constant speed between these records. The movement model does not account for movement caused by ocean current; it is assumed that the animals and ship are drifting in the same direction and at the same rate, thus this movement has no effect on their relative positions.

Independent tag data is collected on nineteen spotted dolphins (Scott and Chivers, 2009) providing fixed locations at approximately 15 minute intervals over 1–2 days.

School abundance is estimated using a Horvitz-Thompson-like estimator:

$$\hat{N} = \sum_{i=1}^{n} \frac{1}{\hat{p}_i}$$

where $n$ is the number of detected schools and $p_i$ is the probability school $i$ is detected. The total number of spotted dolphins is estimated as $\hat{N}\hat{s}$ where $\hat{s}$ is the estimated mean school size.

Goodness of fit is evaluated by a chi-squared test. The perpendicular distance from the transect line is subdivided into discrete bins; the observed and expected number of sightings in each bin is compared.

# 3 Results

## 3.1 Simulation study

**Line transect simulation**

CDS estimators of detection probability and abundance had bias > 10% when animal speed exceeded 1.5 metres per second and CDS overestimated abundance by > 100% for speed > 3.5 metres per second. In contrast, MDS led to < 5% bias for all scenarios (Figure 1). Mean square error for CDS estimators was dominated by their bias; MDS showed constant MSE across all animal speeds (Figure 2).

Confidence interval coverage across all parameters for CDS was less than 40% for speeds over 1.0 metre per second and fell to 0% for speeds over 2.0 metres per second. MDS coverage was nominal within 1% for all parameters and across all simulation scenarios.
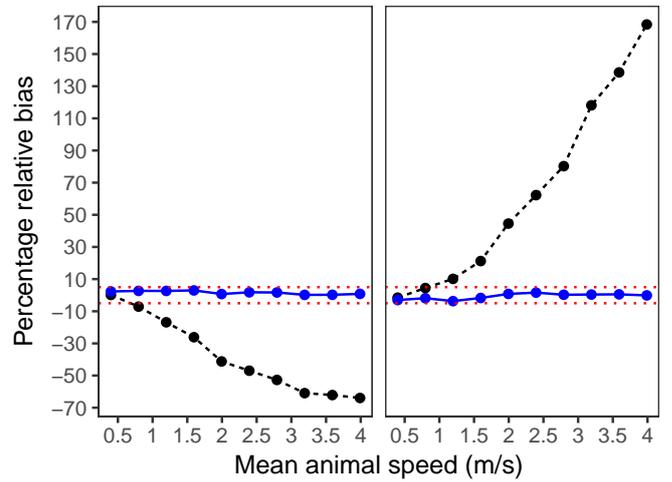


**Figure 1.** *Percentage relative bias in estimated mean detection probability (left pane) and abundance (right pane) for conventional distance sampling (dashed line) and distance sampling with movement incorporated (solid line) against animal speed (metres per second) estimated from 100 simulations of a line transect survey of 50 transects with truncation width 30 metres and observer speed 1 metre per second. Dotted lines mark 5% relative bias.*
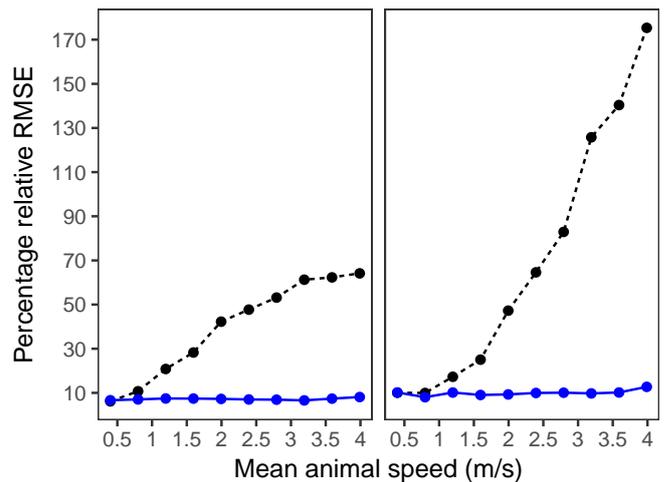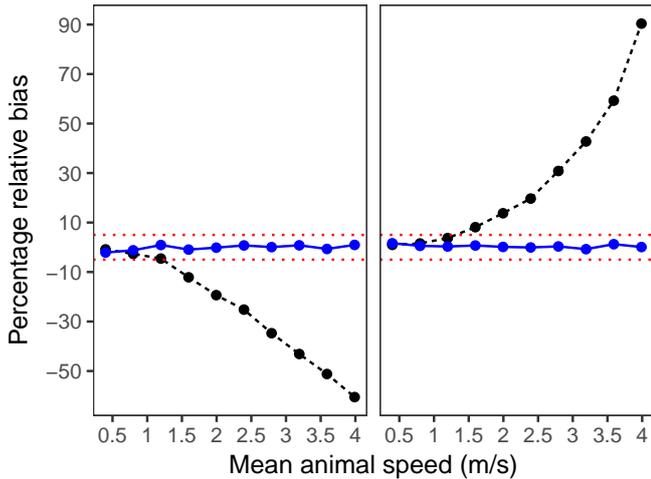


**Figure 2.** *Percentage relative root mean square error (RMSE) in estimated mean detection probability (left pane) and abundance (right pane) for conventional distance sampling (dashed line) and distance sampling with movement incorporated (solid line) against animal speed (metres per second) estimated from 100 simulations of a line transect survey of 50 transects with truncation width 30 metres and observer speed 1 metre per second.*

**Point transect simulation**

CDS point transect sampling behaved similarly with bias > 10% for animal speed > 2 metres per second and bias reaching 90% for speed around 4 metres per second. Incorporating movement reduced bias to < 5% across all scenarios and MSE varied negligibly. CDS 95% confidence interval coverage was poor (< 45%) for all parameters when animal speed exceeded 2 metres per second, while

coverage was nominal for all parameters when movement was incorporated.



**Figure 3.** *Percentage relative bias in estimated detection probability (left pane) and abundance (right pane) for conventional distance sampling (dashed line) and distance sampling with movement incorporated (solid line) against animal speed (metres per second) estimated from 100 simulations of a point transect survey with 100 transects of radius 100 metres, surveyed each for 5 minutes. Dotted lines mark 5% relative bias.*
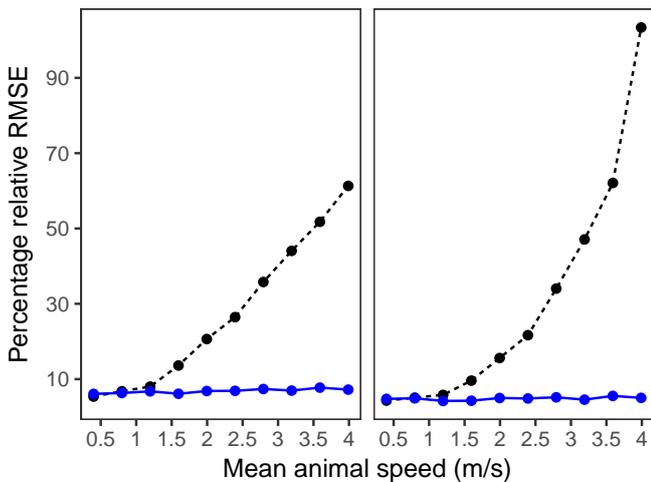


**Figure 4.** *Percentage relative root mean square error (RMSE) in estimated mean detection probability (left pane) and abundance (right pane) for conventional distance sampling (dashed line) and distance sampling with movement incorporated (solid line) against mean animal speed (metres per second) estimated from 100 simulations of a point transect survey with 100 transects of radius 100 metres, surveyed each for 5 minutes.*

## 3.2 Application: spotted dolphins

Estimated abundance of spotted dolphin schools in the core area, as defined by Gerrodette and Forcada (2005), differed between models (Table 1). Mean group size was 164 with large variability due to the rare detection of very large schools (> 2000 individuals). Incorporating movement reduced the abundance estimate by 22% (33292 animals). The coefficient of variation (CV) for the abundance estimator was reduced (by 7.4%) when animal movement was accounted for. The average speed of the ship was 17km/h; the estimated average speed of each spotted dolphin was 7.4km/h. The large reduction in the abundance estimate indicates that even though the dolphins move relatively slowly compared to the ship, bias can be substantial, because, whilst being surveyed, they can move a large distance compared to the width of the transect. This highlights the danger of assessing whether movement is a problem based solely on relative animal speed; MDS can account for the interdependent effects of animal speed, transect width, and detection function shape.

For comparison, the expected number of sightings within each 0.5 kilometre from the transect line was calculated (Figure 5). MDS had a similar goodness of fit as CDS to the data (chi-squared test gives p-value of 0.31 for CDS and 0.35 for MDS); however, the estimated detection function differs considerably between the two methods (Figure 6). The CDS estimated detection function has a narrower shoulder and smaller detection scale indicating that animal movement has caused negative bias in the estimation of detection probability. If the survey had indeed taken place in a snapshot of time, CDS estimates the probability of an animal being detected, given it is inside the transect, to be 0.47; MDS estimates this to be 0.62. Note, this deficiency does not result in a marked difference in goodness-of-fit to the observed data, but has an important effect on the final abundance estimate.
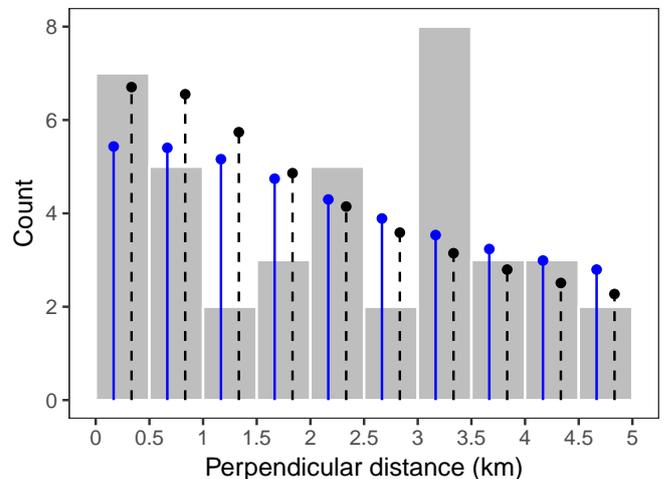


**Figure 5.** *Observed number of spotted dolphin schools sighted in each 0.5 km perpendicular distance from the transect line (shaded bars) with expected number of sightings from conventional distance sampling model (dashed lines) and distance sampling with movement (solid lines)*

## 4 Discussion

Distance sampling surveys on mobile animal populations should not ignore animal movement. The simulation study
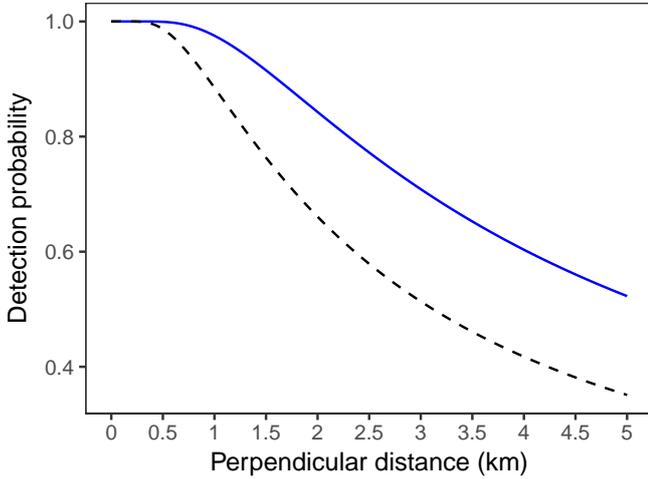
***Figure 6.*** *Estimated detection function for a hypothetically immobile spotted dolphin population for conventional distance sampling (dashed line) and distance sampling with movement (solid line)*

|  | Estimate | CV(%) | LCL | UCL |
|---|---|---|---|---|
| CDS density | 910 | 19.9 | 430 | 983 |
| MDS density | 707 | 12.5 | 533 | 880 |

***Table 1.*** *Maximum likelihood estimates of spotted dolphin school density (per $10^6 km^2$) with coefficient of variation (CV) and lower and upper 95% confidence interval bounds for conventional distance sampling (CDS) and distance sampling with movement (MDS)*

demonstrates the remarkable bias that non-responsive animal movement can cause. Incorporating this movement into distance sampling can mitigate this bias and remove the subjective judgment of when movement bias may have occurred and to what extent.

The cost of this improvement in estimation is the need for additional information on animal movement. The expense and practicality of collecting such data depends on the species to be surveyed. Tag data on cetaceans, ungulates, and seabirds is becoming more common. When no such data is available, a 'plug-in' estimator of the movement parameters could be used, similar to the multipliers used in CDS. Accounting for movement on imperfect knowledge is better than ignoring it completely. The movement information need not come from animals in the same population as that surveyed by distance sampling, nor be collected in the same time period. Nevertheless, it is assumed tagged animals behave as representative members of the study population; thus, it is recommended that movement information be collected around the same time the distance sampling survey is conducted and on animals that are members of the study population.

Estimating a detection process that depends on time necessarily requires detection times be recorded. In practice, this information is often recorded in the field, but not used in CDS models. Here, this information is essential and any application of the model would require this data to be collected. Furthermore, as for the ETP survey, periods of on and off effort must be recorded. When the observer pauses

their efforts, animals continue to move; failing to record this information leads to underestimation of the detection probability. Records of the observer's location over time must also be kept. Furthermore, the relative location of detected animals should be recorded relative to the observer; this is contrary to practice in CDS where measurements are made relative to the point or line. Also, animal location must be recorded in two-dimensional space.

### Assumptions

The assumptions made in the theory presented are synonymous with those made in CDS. Violations of these assumptions will cause bias in the inference obtained.

1. The path an animal travels is independent of the observer: animals do not respond to the observer and their movement is independent of the transect placement, that is, surveying does not preferentially take place in areas animals would avoid or be attracted to.

2. Animals at zero radius are detected: this assumption can be violated for animals that are not always available for detection, for example, a diving cetacean can be missed by an observer in a ship directly above it (Borchers et al., 2013; Barlow, 2015).

3. Location measurements are exact: this assumption applies to observed locations of animals on the distance sampling survey and the recorded locations of tracked animals. Observation error in animal telemetry data is common and can be accounted for (Johnson et al., 2008). Models for measurement error in distance sampling can also be incorporated (Marques, 2004).

4. Animal movement is diffusive: the simple model that animal movement is a spatially-invariant, isotropic diffusion process is violated by many animal populations; more realistic movement models can be considered. No matter what movement model is incorporated, one assumes that all animals in the survey move according to the specified model. Departures from the movement model could cause detection probability to be biased.

5. Sampling is representative and independent: for the distance sampling survey, this assumption requires transects be placed according to a randomised design, that transects be independent, and that animals be independent. For animals that travel in groups, treating groups as the independent unit to be sampled may be a better choice. For the movement model, it is assumed that tagged animals move independently and that the sample of tagged animals be representative of the surveyed population. One can use telemetry from tagged animals who are not members of the surveyed population, but only with the assumption these animals exhibit movement patterns similar to those animals surveyed by distance sampling.

**Model extensions**

The model formulation is flexible and can include existing extensions of conventional distance sampling. Here, only a hazard that depends on radial distance was considered. A hazard that depends on angle and radius could be used. For point transects, any detection function can be used to define a hazard (Borchers et al., 2015). Yet, for line transects, the method relies on an analytical form for $S_t$: observer movement makes integration of the hazard more complicated. This could be avoided by discretising observer movement similarly to animal movement, that is, the observer only moves at time-steps and not within time-steps; however, this can introduce significant bias when discretisation is rough and often requires much lower levels of discretisation than otherwise necessary.

Another popular extension is multiple-covariate distance sampling (Marques and Buckland, 2003). Including covariates in the detection process could be handled similarly with this model where parameters in the hazard can depend on environmental conditions over space and time. Given this, it is important to highlight that any covariates included would need to be known for each time and each location in space. This may require one to assume that covariates are constant in the time between their recording. Additional information on the movement of detected individuals can also be incorporated. Observed locations or observed directions of travel can be used to improve estimation of the unknown path each animal took. Furthermore, responsive movement could be accounted for if the response, how an animal's location changes over time with respect to the observer, is recorded for one or more focal individuals.

Including covariate information in the movement model is more computationally demanding. The HMM computational algorithm used depends on the block-Toeplitz structure of the transition rate matrix, $\mathbf{G}$. A block-Toeplitz structure is equivalent to assuming that the transition rates are spatially invariant. Relaxing this assumption increases computation time and limits the level of discretisation that can be practically obtained.

Additionally, this assumption limits the range of possible movement models that can be considered. Bias in density estimation from animal movement is at its worst when animal movement is persistent in a single direction. A movement process with persistent movement would require the discretisation of a 4-dimensional space: location and velocity. Given current computational resources, it is likely any practical discretisation of this space would be too coarse to provide good estimators of detection parameters. Nevertheless, for many taxa, animal movement is more complex than can be described by a diffusion process, and so extending MDS to include advection-diffusion or Ornstein-Uhlenbeck animal movement models could improve the inference obtained.

Finally, the methods presented can be extended to double-observer distance sampling, where two observers survey the transect simultaneously. If multiple sightings of an individual by different observers can be matched together, then only animal paths that pass through these multiple observed locations need be considered, providing information on animal movement directly from the distance sampling data and improving the estimation of each animal's detection probability.

**Conclusion**

Animal movement can be incorporated with distance sampling. The presented theoretical framework provides a basis for further development and the computational approach discussed makes the method applicable. Accounting for animal movement can mitigate the bias it causes in the surveys where distance sampling is applied, and it can widen the application of distance sampling to animals whose movement has so far prohibited its use.

# References

Barlow, J. (2015). Inferring trackline detection probabilities, g (0), for cetaceans from apparent densities in different survey conditions. *Marine Mammal Science*, 31(3):923–943.

Borchers, D. L. and Cox, M. (2017). Distance sampling detection functions: 2D or not 2D? *Biometrics*, Early View.

Borchers, D. L., Stevenson, B. C., Kidney, D., Thomas, L., and Marques, T. A. (2015). A unifying model for capture–recapture and distance sampling surveys of wildlife populations. *Journal of the American Statistical Association*, 110(509):195204.

Borchers, D. L., Zucchini, W., Heide-Jørgensen, M. P., Cañadas, A., and Langrock, R. (2013). Using hidden Markov models to deal with availability bias on line transect surveys. *Biometrics*, 69(3):703713.

Buckland, S. T. (2006). Point-transect surveys for songbirds: robust methodologies. *The Auk*, 123(2):345–357.

Buckland, S. T., Anderson, D. R., Burnham, K. P., and Laake, J. L. (2005). *Distance sampling*. Wiley Online Library.

Buckland, S. T., Oedekoven, C. S., and Borchers, D. L. (2016). Model-based distance sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, 21(1):58–75.

Buckland, S. T., Rexstad, E. A., Marques, T. A., and Oedekoven, C. S. (2015). *Distance sampling: methods and applications*. Springer.

Eydeland, A. (1994). A fast algorithm for computing integrals in function spaces: financial applications. *Computational Economics*, 7(4):277–285.

Fewster, R. M., Buckland, S. T., Burnham, K. P., Borchers, D. L., Jupp, P. E., Laake, J. L., and Thomas, L. (2008). Estimating the encounter rate variance in distance sampling. *Biometrics*, 65(1):225236.

Gerrodette, T. and Forcada, J. (2005). Non-recovery of two spotted and spinner dolphin populations in the eastern tropical pacific ocean. *Marine Ecology Progress Series*, 291:1–21.

Gerrodette, T., Watters, G., Perryman, W., and Ballance, L. (2008). Estimates of 2006 dolphin abundance in the eastern tropical pacific, with revised estimates from 1986-2003. Technical report, NOAA, NMFS Technical Memorandum NOAA-TM-NMFS-SWFSC-422.

Glennie, R., Buckland, S. T., and Thomas, L. (2015). The effect of animal movement on line transect estimates of abundance. *PloS one*, 10(3):e0121333.

Gurarie, E. and Ovaskainen, O. (2012). Towards a general formalization of encounter rates in ecology. *Theoretical Ecology*, 6(2):189202.

Hedley, S. L. and Buckland, S. T. (2004). Spatial models for line transect sampling. *Journal of Agricultural, Biological, and Environmental Statistics*, 9(2):181–199.

Hiby, A. (1982). The effect of random whale movement on density estimates obtained from whale sighting surveys. *REP. INT. WHAL. COMMN.*, 32:791–794.

Howe, E., S.T., B., Desprs-Einspenner, M.-L., and Khl, H. (2017). Distance sampling with camera traps (in press). *Methods in Ecology and Evolution*.

Hutchinson, J. M. C. and Waser, P. M. (2007). Use, misuse and extensions of "ideal gas" models of animal encounter. *Biological Reviews*, 82(3):335359.

Johnson, D. S., London, J. M., Lea, M.-A., and Durban, J. W. (2008). Continuous-time correlated random walk model for animal telemetry data. *Ecology*, 89(5):1208–1215.

Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., and Morales, J. M. (2012). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*, 93(11):2336–2342.

Lee, D. (1986). Fast multiplication of a recursive block Toeplitz matrix by a vector and its application. *Journal of Complexity*, 2(4):295–305.

Lucas, T. C., Moorcroft, E. A., Freeman, R., Rowcliffe, J. M., and Jones, K. E. (2015). A generalised random encounter model for estimating animal density with remote sensor data. *Methods in Ecology and Evolution*, 6(5):500–509.

Marques, F. F. and Buckland, S. T. (2003). Incorporating covariates into standard line transect analyses. *Biometrics*, 59(4):924–935.

Marques, T. A. (2004). Predicting and correcting bias caused by measurement error in line transect sampling using multiplicative error models. *Biometrics*, 60(3):757–763.

Okubo, A. and Levin, S. A. (2013). *Diffusion and ecological problems: modern perspectives*, volume 14. Springer Science & Business Media.

Palka, D. and Hammond, P. (2001). Accounting for responsive movement in line transect estimates of abundance. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(4):777–787.

Pedersen, M. W., Patterson, T. A., Thygesen, U. H., and Madsen, H. (2011). Estimating animal behavior and residency from movement data. *Oikos*, 120(9):12811290.

Scott, M. D. and Chivers, S. J. (2009). Movements and diving behavior of pelagic spotted dolphins. *Marine Mammal Science*, 25(1):137–160.

Skaug, H. J. and Schweder, T. (1999). Hazard models for line transect surveys with independent observers. *Biometrics*, 55(1):29–36.

Turnock, B. J. and Quinn, T. J. (1991). The effect of responsive movement on abundance estimation using line transect sampling. *Biometrics*, pages 701–715.

Yapp, W. B. (1956). The theory of line transects. *Bird Study*, 3(2):93104.

Yuan, Y., Bachl, F., Lindgren, F., Brochers, D., Illian, J., Buckland, S., Rue, H., and Gerrodette, T. (2016). Point process models for spatio-temporal distance sampling data. *arXiv preprint arXiv:1604.06013*.

Zucchini, W. and MacDonald, I. L. (2009). *Hidden Markov models for time series: an introduction using R*, volume 22. CRC press Boca Raton.