

Double-Observer Line Transect Surveys with Markov-Modulated Poisson Process Models for Animal Availability

D. L. Borchers* and R. Langrock

Centre for Research into Ecological and Environmental Modelling, The Observatory, Buchanan Gardens,
University of St Andrews, Fife, KY16 9LZ, Scotland

**email*: dlb@st-andrews.ac.uk

SUMMARY. We develop maximum likelihood methods for line transect surveys in which animals go undetected at distance zero, either because they are stochastically unavailable while within view or because they are missed when they are available. These incorporate a Markov-modulated Poisson process model for animal availability, allowing more clustered availability events than is possible with Poisson availability models. They include a mark-recapture component arising from the independent-observer survey, leading to more accurate estimation of detection probability given availability. We develop models for situations in which (a) multiple detections of the same individual are possible and (b) some or all of the availability process parameters are estimated from the line transect survey itself, rather than from independent data. We investigate estimator performance by simulation, and compare the multiple-detection estimators with estimators that use only initial detections of individuals, and with a single-observer estimator. Simultaneous estimation of detection function parameters and availability model parameters is shown to be feasible from the line transect survey alone with multiple detections and double-observer data but not with single-observer data. Recording multiple detections of individuals improves estimator precision substantially when estimating the availability model parameters from survey data, and we recommend that these data be gathered. We apply the methods to estimate detection probability from a double-observer survey of North Atlantic minke whales, and find that double-observer data greatly improve estimator precision here too.

KEY WORDS: Abundance estimation; Availability bias; Cox point process; Mark-recapture; Maximum likelihood.

1. Introduction

Line transect (LT) surveys are one of the most widely used methods of estimating animal abundance (see Buckland et al., 2001, 2014, for an overview of LT and related methods). A persistent problem with LT surveys of marine mammals is failure of one of the key assumptions of standard LT theory, namely that animals at perpendicular distance zero are always detected (Buckland et al., 2001). If this probability is g_0 and true density is D , and if all other LT assumptions are met, then the expected value of a density estimate obtained using standard LT methods is $g_0 D$ (Buckland et al., 2014).

Mark-recapture distance-sampling (MRDS) methods are widely used for dealing with the problem of $g_0 < 1$: see Burt et al. (2015) and references therein for an overview. They involve independent observers acting as “capture occasions” and rely on “recaptures” (or “duplicates”) to estimate g_0 . They are susceptible to bias when there is correlation in detections by the two observers, due to latent variable(s) that are not in the model (“unmodelled heterogeneity”). Correlation can be induced by individual-level or environmental variables that affect the detection probabilities of both observers, and by animals being stochastically available for detection when animal availability to one observer is correlated with its availability to the other. We focus on the issue of availability in this article but include a brief discussion of unmodelled heterogeneity due to other latent variables in Section 6.

Stochastic availability can be dealt with by incorporating a model for the availability process. We refer to such models

as “availability process line transect” (APLT) models. APLT models were initially developed assuming a Poisson availability process (Skaug and Schweder, 1999; Okamura, 2003; Okamura et al., 2003), but more recent work has shown that an assumption of Poisson availability can result in substantially biased estimates of detection probability if it is violated (Okamura et al., 2012; Borchers et al., 2013; Langrock et al., 2013). To accommodate more clustered availability series than is possible with a Poisson model, Hiby and Lovell (1998) developed a continuous-time Markov process model, Borchers et al. (2013) developed a discrete-time hidden Markov model (HMM) and Langrock et al. (2013) developed a Markov-modulated Poisson process (MMPP) model for availability.

APLT models consider detection probability as a function of radial distance, and Langrock et al. (2013) show that when using a single observer it is sometimes, but not always, possible to estimate probability of detection at radial distance zero. A more reliable way of estimating this probability is to incorporate a mark-recapture (MR) component in APLT models. This is done by using two independent observers, as with MRDS methods. Some such methods do exist, but all either assume Poisson availability (Schweder et al., 1996; Skaug and Schweder, 1999) or incorporate the availability process non-parametrically by resampling from availability time series data (Schweder et al., 1999; Okamura et al., 2012).

We extend the MMPP model of Langrock et al. (2013) to accommodate double-observer surveys by introducing an MR component. We illustrate using the minke whale survey

considered by Skaug and Schweder (1999), re-analyzing their data using a model that allows availability events to be more clustered than under a Poisson process. The latter is a nested special case of our model. Our model also accommodates multiple detections of the same animal at different times; these are informative about the availability process. We investigate estimator performance by simulation when the availability process is assumed known, when the rate parameters of the MMPP availability model are estimated and parameters of the Markov component are known, and when all the parameters of the MMPP are estimated.

2. Model Formulation

2.1. Animal Availability Model

We detail the model for the general case in which animals can be in any of N states, although in the application below only the two-state case is considered. Animals switch between N “behavioral” states as determined by an N -state Markov chain in continuous time, with infinitesimal generator matrix $\mathbf{Q} = (q_{ij})_{i,j=1,\dots,N}$. The duration of a stay in state k ($k = 1, \dots, N$) is exponentially distributed with parameter $-q_{kk}$. The initial state (which below will be the state the animal is in when entering the detectable range) is determined by the initial state distribution vector $\boldsymbol{\pi}$, whose k th element is the probability of initially being in state k . Given that the survey start time is random with respect to animal state, it is reasonable to assume $\boldsymbol{\pi}$ to be the stationary distribution of the process. This distribution gives the probabilities of the process being in each state at a randomly chosen time. In the case $N = 2$, the stationary probability of the process being in state 1 is $\pi_1 = q_{21}/(q_{12} + q_{21})$.

The availability model of Hiby and Lovell (1998) is the above process with two states. To complete the MMPP, a Poisson process is superimposed on this model, such that availability events occur according to a Poisson process with state-dependent rate λ_k when the animal is in state k . For example, in a two-state model, with state 1 corresponding to the animal resting/breathing at the surface, and state 2 corresponding to foraging dives, λ_1 will typically be much larger than λ_2 ; in fact, the latter will often be zero (which would correspond to animals not surfacing at all while in foraging state).

2.2. Embedding the MMPP in a Distance-Sampling Survey

As is the case with most LT models, we assume that animals do not change their perpendicular distance from the line while within detectable range. Our model also assumes that if they move in the forward (along-transect) direction, they all move at a constant speed (which may be zero). To establish our notation, consider a whale at perpendicular distance x that enters the observers’ field of view at forward distance $y_{\max}(x)$ and is detected at forward distance y , by which time it has been in view for a distance $l = y_{\max}(x) - y$ (see Figure 1).

Conditional on an animal at x being available at a distance l from $y_{\max}(x)$, we assume that it is detected with probability $h(l, x)$. Although it is a probability, we refer to h as the detection hazard function in order to distinguish it from conventional detection probability functions on line transect

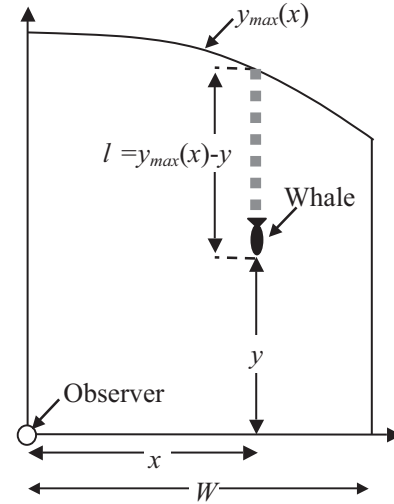


Figure 1. Schematic representation of the survey process. Observers move in the y -direction, animals remain at fixed perpendicular distances (x), and observers can see no farther ahead than the curve $y_{\max}(x)$. When an animal at x is detected at forward distance y , it has been in view for a distance $l = y_{\max}(x) - y$.

surveys, which depend only on perpendicular distance and not on forward distance. We thus assume that the observed detections are realizations of MMPPs that operate on the forward distance scale after being thinned according to the detection hazard function $h(l, x)$. The MMPP described above is assumed to generate the actual availability events, and for an event at (l, x) a Bernoulli distribution with success probability $h(l, x)$ determines whether that event is detected. We assume that conditional on availability at (l, x) , detections are independent between observers so that in the case of two observers, with probabilities of detecting a surfacing (availability event) given by $h_1(l, x)$ and $h_2(l, x)$ respectively for the two observers, the probability that any observer detects the surfacing is $h(l, x) = 1 - \{1 - h_1(l, x)\}\{1 - h_2(l, x)\}$. (We discuss how violation of this independence assumption might be dealt with in Section 6.) It follows that the distances (y or l) at detections are the events of a Markov-modulated *nonhomogeneous* Poisson process with state- and distance-dependent rate parameters $\lambda_k h(l, x)$ ($k = 1, \dots, N$), and with parameters of the underlying Markov chain being $\{q_{ij}\}$, $i, j = 1, \dots, N$ (cf. Langrock et al., 2013, for more details).

3. Estimation

3.1. Density and Cumulative Distribution of Sightings

As will become clear below, in order to estimate the model parameters, we need, among other things, the density function for the forward distances (or distances from $y_{\max}(x)$) of detections and the cumulative distribution for the distance at first detection. Although there are closed-form expressions for these in the case of homogeneous MMPPs, there are in general no closed-form expressions available for these in the case of non-homogeneous MMPPs. However, Langrock et al. (2013) showed that these can be approximated arbitrarily

accurately using an $h(l, x)$ that with respect to l is piecewise constant on the intervals $[i_0, i_1), [i_1, i_2), \dots, [i_{p-1}, i_p]$ (with $i_0 = 0$ and $i_p = y_{\max}(x)$). Let $h^{(r)}(x)$ denote the constant value of $h(l, x)$ on the r th interval, $[i_{r-1}, i_r)$. Let $\mathbf{\Lambda}(x)^{(r)} = \text{diag}(\lambda_1 \cdot h^{(r)}(x), \dots, \lambda_N \cdot h^{(r)}(x))$. Langrock et al. (2013) showed that

- (i) the joint density of detections of an animal (or of a group of animals) occurring at distances $\mathbf{l} = (l_1 < l_2 < \dots < l_d)$, with $i_{R_j} \leq l_j \leq i_{R_{j+1}}$ for $j = 1, \dots, d$, is given by

$$f_{\mathbf{l}}(\mathbf{l}|x) = \boldsymbol{\pi} \prod_{j=1}^d \{ \mathcal{P}(l_{j-1}, l_j|x) \mathbf{\Lambda}(x)^{(R_{j+1})} \} \mathcal{P}(l_d, y_{\max}|x) \mathbf{1}^t, \tag{1}$$

where $l_0 = 0$, $\mathbf{1} \in \mathbb{R}^N$ is a row vector of ones and

$$\mathcal{P}(l_{j-1}, l_j|x) = \begin{cases} \exp[\{\mathbf{Q} - \mathbf{\Lambda}(x)^{(R_{j+1})}\}(l_j - l_{j-1})] & \text{if } l_{j-1} \text{ and } l_j \text{ are in the same interval} \\ \exp[\{\mathbf{Q} - \mathbf{\Lambda}(x)^{(R_{j-1+1})}\}(i_{R_{j-1+1}} - l_{j-1})] \\ \cdot \prod_{r=R_{j-1}+1}^{R_j-1} \exp[\{\mathbf{Q} - \mathbf{\Lambda}(x)^{(r+1)}\}(i_{r+1} - i_r)] \\ \cdot \exp[\{\mathbf{Q} - \mathbf{\Lambda}(x)^{(R_j+1)}\}(l_j - i_{R_j})] & \text{otherwise} \end{cases}$$

- (ii) the cumulative distribution of distance l to first detection is $F_l(l|x) = 1 - \boldsymbol{\pi} \mathcal{P}(0, l|x) \mathbf{1}^t$

We can therefore approximate $f_{\mathbf{l}}(\mathbf{l}|x)$ and $F_l(l|x)$ arbitrarily accurately using a function $h(l, x)$ that is a step function with respect to l by using enough steps.

In addition to generating a vector of observed locations \mathbf{l} for a detected animal at perpendicular distance x , a double-observer survey also generates a ‘‘capture history’’ for each detection, i.e., a realization of a random variable ω_j for detection j , such that $\omega_j = 1$ if seen by observer 1 only, $\omega_j = 2$ if seen by observer 2 only, and $\omega_j = 3$ if seen by both. Conditional on detection at l_j , the probability mass function for ω_j is trinomial, with index 1 and dependence on x as follows:

$$p(\omega_j|l_j, x) = \begin{cases} \frac{h_1(l_j, x)\{1 - h_2(l_j, x)\}}{h(l_j, x)} & \text{if } \omega_j = 1; \\ \frac{\{1 - h_1(l_j, x)\}h_2(l_j, x)}{h(l_j, x)} & \text{if } \omega_j = 2; \\ \frac{h_1(l_j, x)h_2(l_j, x)}{h(l_j, x)} & \text{if } \omega_j = 3. \end{cases}$$

Then, if we let $\boldsymbol{\omega} = (\omega_1, \omega_2, \dots, \omega_d)$ denote the capture histories associated with the detections of an animal (or of a group of animals) made at distances $\mathbf{l} = (l_1 < l_2 < \dots < l_d)$, the joint density of \mathbf{l} and $\boldsymbol{\omega}$, given x , is

$$f_{\mathbf{l}, \boldsymbol{\omega}}(\mathbf{l}, \boldsymbol{\omega}|x) = \boldsymbol{\pi} \prod_{j=1}^d \{ \mathcal{P}(l_{j-1}, l_j|x) \mathbf{\Lambda}(x)^{(R_{j+1})} p(\omega_j|l_j, x) \} \times \mathcal{P}(l_d, y_{\max}|x) \mathbf{1}^t \tag{2}$$

where both $h_1(l, x)$ and $h_2(l, x)$ are piecewise constant with respect to l on the intervals $[i_0, i_1), [i_1, i_2), \dots, [i_{p-1}, i_p]$ (with $i_0 = 0$ and $i_p = y_{\max}(x)$).

The proof is straightforward, since conditional on \mathbf{l} (and x), and assuming independent detections both between observers and of the same animal on different occasions, the probability for $\boldsymbol{\omega}$ is $f_{\boldsymbol{\omega}|\mathbf{l}}(\boldsymbol{\omega}|\mathbf{l}, x) = \prod_{j=1}^d p(\omega_j|l_j, x)$ and Equation (2) is just the product of this conditional probability mass function and the density function $f_{\mathbf{l}}(\mathbf{l}|x)$, given in Equation (1). (We deal with dependence of detections of the same animal on different occasions by allowing the hazard to change after first detection—see Sections 4 and 6 below.)

Since with this model, each animal generates binary trials (with ‘‘success’’ being detection of a surfacing event) for each observer, irrespective of whether the other observer has detected the animal, duplicates need not involve both observers detecting the same surfacing.

3.2. Likelihood

To obtain the full likelihood, for observed capture histories and forward and perpendicular distances at which animals were detected, we condition on the fact that the detection was made within maximum perpendicular distance and within $[0, y_{\max}(x)]$ in the forward distance dimension. Let $\pi(x)$ denote the pdf of perpendicular distances x of all animals within some maximum distance W from the transect line. For detected animals i , $i = 1, \dots, n$, let $\mathbf{l}_i = (l_{i1}, \dots, l_{id_i})$ and $\boldsymbol{\omega}_i = (\omega_{i1}, \dots, \omega_{id_i})$ denote detection distances and capture histories, respectively, and let x_i be the perpendicular detection distance. Furthermore, let $\boldsymbol{\phi}$ be the vector comprising the parameters of the MMPP for availability, let $\boldsymbol{\theta}$ be the vector of parameters of the detection hazards $h_1(l, x)$ and $h_2(l, x)$, and let $\boldsymbol{\beta} = (\boldsymbol{\phi}, \boldsymbol{\theta})$. Then, making the dependence of $f_{\mathbf{l}, \boldsymbol{\omega}}(\cdot)$ and $F_l(\cdot)$ on $\boldsymbol{\beta}$ explicit, the likelihood for $\boldsymbol{\beta}$, conditional on detection within perpendicular distance W and forward distance $y_{\max}(x_i)$, is given by

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{i=1}^n \frac{f_{\mathbf{l}, \boldsymbol{\omega}}(\mathbf{l}_i, \boldsymbol{\omega}_i|x_i; \boldsymbol{\beta}) \pi(x_i)}{\int_0^W F_l(y_{\max}(z)|z; \boldsymbol{\beta}) \pi(z) dz} \tag{3}$$

This follows from the definition of conditional probability (cf. Langrock et al., 2013). The denominator can be computed via numerical integration, and it simplifies if transect lines are placed randomly, in which case $\pi(x) = 1/W$ and $\pi(x)$ cancels.

3.3. Parameter Estimation

Data that allow *in situ* estimation of availability process parameters are very rarely gathered on line transect surveys. Auxiliary data containing information on the surfacing pattern are most commonly used to model the availability process (cf. Skaug and Schweder, 1999). A typical source for such data is GPS tagging. (See Section 3.4 for details

of how multiple sets of tag data can be incorporated in our model.) We recommend that data that allows estimation of the availability process be gathered on the survey itself if this is possible. This gives the analyst more options: to use these data alone for availability process estimation, to use them with some availability parameters estimated from auxiliary data, or to use only availability parameters estimated from auxiliary data (when these are available). (See Sections 4 and 5 for more details, and Section 6 for further discussion.)

If we take some or all of the availability process parameters $\phi = (\{q_{jk}\}, \{\lambda_j\})$, $j, k = 1, \dots, N$ states, to be known, then Equation (3) is a likelihood for the detection hazard parameter vector θ and whichever (if any) parameters of the availability process are not known. We can then obtain maximum likelihood estimates of θ and any unknown availability process parameters by numerically maximizing Equation (3). Confidence intervals for the parameters can be obtained either based on the Hessian of the log-likelihood or by bootstrapping. If the hazard function parameters are obtained conditionally on ϕ (obtained from auxiliary data), then a two-stage bootstrap should be used rather than the Hessian, with the first stage involving obtaining bootstrap resamples of $\hat{\phi}$ and the second stage involving bootstrapping the sightings survey data and re-estimation of θ (nonparametrically, with transects as sampling units, for example), conditional on randomly selected resamples of $\hat{\phi}$.

Two detection hazard forms suggested by Skaug and Schweder (1999) (given in terms of forward distance $y = y_{\max}(x) - l$) are the exponential power and inverse power models:

$$\text{exponential power: } h(y, x) = \theta_1 \exp\left(-\frac{x^{\theta_3} + y^{\theta_3}}{\theta_2^{\theta_3}}\right) \quad 0 \leq \theta_1 \leq 1, \theta_2, \theta_3 > 0 \quad (4)$$

$$\text{inverse power: } h(y, x) = \theta_1 \left(\frac{\theta_2}{\sqrt{\theta_2^2 + x^2 + y^2}}\right)^{\theta_3} \quad 0 \leq \theta_1 \leq 1, \theta_2 > 0, \theta_3 > 1. \quad (5)$$

In the case of automated detectors (acoustic detectors, for example), the assumption of repeated independent detections of the same animal by the same observer is plausible, but this within-observer, within-animal independence is unlikely to hold for human observers. Initial detection is likely to increase the probability of subsequent detections by humans. This can be dealt with by assuming that for each observer the hazard function changes after the initial detection, with a set of parameters θ_{ini} determining the hazard function until the initial sighting, and another set of parameters θ_{sub} determining the hazard function subsequent to the initial sighting. The formulae given above change accordingly.

In order to obtain an abundance estimate from a fitted nonhomogeneous MMPP using a conditional likelihood approach (see Borchers and Burnham, 2004, for a discussion of conditional *vs* full likelihood), we require the effective strip half-width, $\mu = \int_0^W F_l(y_{\max}(z), z; \theta) dz$. The probability of detecting an animal that is within distance W of the transect line is μ/W (Buckland et al., 2001). For given model parameters, the value of μ can easily be computed using numerical integration. The uncertainty in μ can be quantified through bootstrapping.

3.4. Marginalizing Over Individual-Level Availability

It is common to have separate estimates of availability process parameters for a number of different animals. In this case, individual-level heterogeneity in animal availability patterns can readily be accommodated by evaluating the likelihood Equation (3) separately for each set of availability process parameters and summing the likelihoods. This amounts to marginalizing the likelihood over the empirical distribution of individual-level availability model parameters. See Borchers et al. (2013) and Langrock et al. (2013) for details.

If some availability model parameters (ϕ_a , say) are estimated directly from availability time series and others (ϕ_b , say) are estimated by maximizing the likelihood given in Equation (3) conditional on these, then the above approach to dealing with individual-level variation in availability parameters would need to be modified. One could still marginalize non-parametrically over ϕ_b (as above) but one would need to model the conditional distribution of ϕ_a given ϕ_b parametrically and marginalize the likelihood over ϕ_a for each individual ϕ_b .

4. Simulation Experiments

4.1. Simulation Design

We conduct two sets of simulations, one with a setup that results in the density of distances in view, l , at initial detection of animals on the trackline, i.e., $f_l(l_1|x=0) = \frac{d}{dl} F_l(l|x=0)$,

having mode roughly in the middle of the interval $[0, y_{\max}(x)]$ (Scenario I) and the other for which $f_l(l_1|x=0)$ is monotonically increasing as l increases, up to a mode that is close to $y=0$ (Scenario II). Whether or not $f_l(l_1|x=0)$ has a mode close to $y=0$ depends on the detection process and availability process and is not under the control of the surveyor or analyst. Examples of surveys with modes at $y=0$ and $y>0$ can be found in Figure 3 of Borchers et al. (2013) and Figure 2 of Rekdal et al. (2014).

The densities for Scenarios I and II are shown in Figure 2. Note that in Figure 2, we have not conditioned on animals being detected within detectable range, hence these are not pdfs (because they do not integrate to unity—their integral is the probability of being detected by forward distance zero). Estimation of parameters turns out to be more difficult in Scenario II, and we included both these scenarios in this study to emphasize some practical aspects of surveys that need to be considered when using the suggested approach (see discussion at the end of this section).

For each of the two scenarios, we consider single-observer and double-observer estimators, and within each of these we

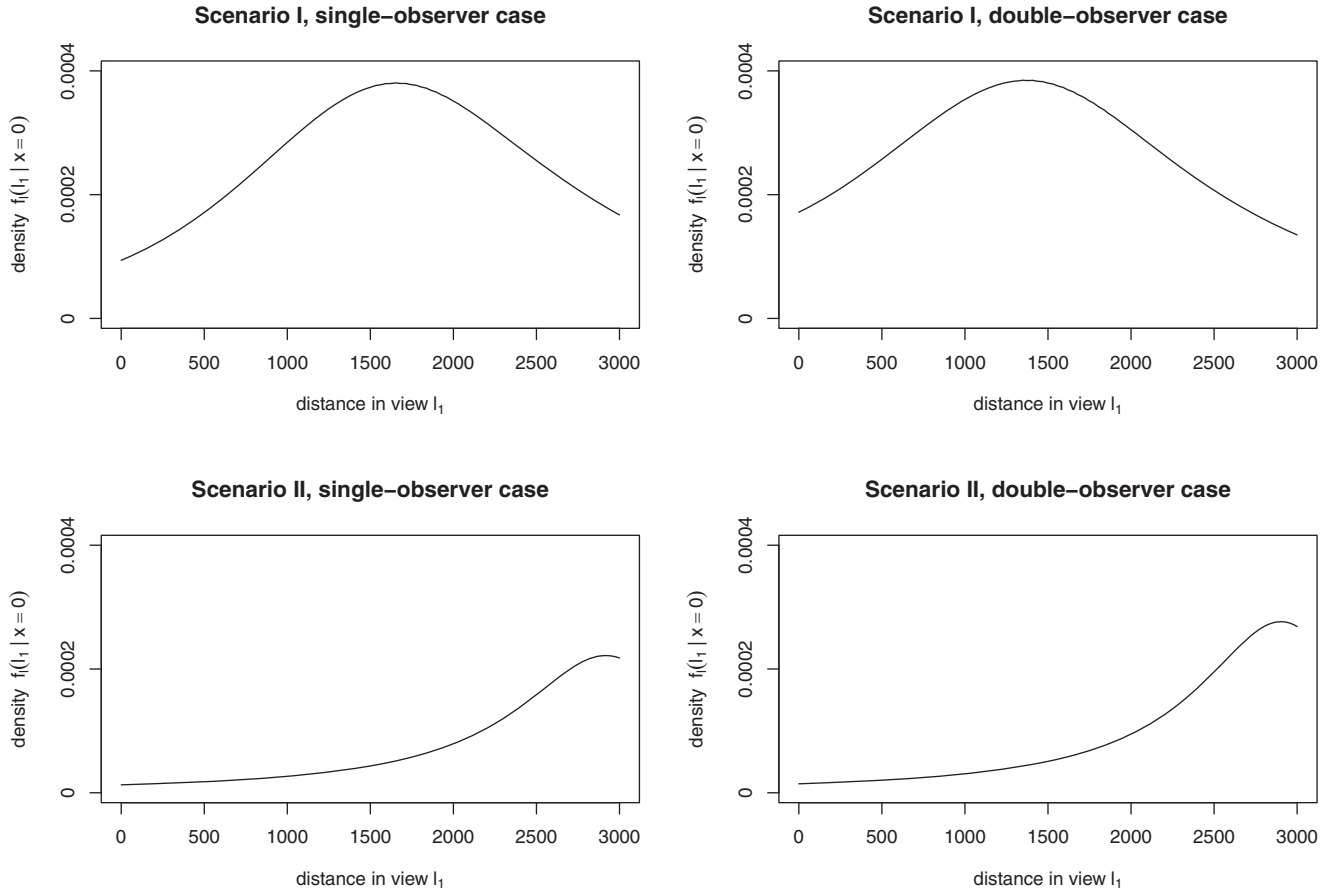


Figure 2. Densities of distances in view, l , at initial detection of animals on the trackline, i.e., $f_l(l_1|x=0) = \frac{d}{dl}F_l(l|x=0)$, in Scenarios I and II and for both single- and double-observer scenarios.

consider two kinds of data, namely data from the initial sighting of an animal only or data from multiple sightings (“IS” vs. “MS”), and when multiple sightings data are used, we either have observers detecting all availability events after their initial sightings with certainty (“ $h(\cdot) = 1$ ”) and specify this in the estimator, or we have $h(l, x)$ increasing after initial detection (but to values less than 1), and estimate two sets of hazard function parameters for each observer; one before first detection, the other after (“ $h(\cdot) < 1$ ”). Finally, for each of the above scenarios and cases, three kinds of estimators were considered, corresponding to different amounts of information on the availability process being available: one in which the MMPP availability process is fully known to the estimator (“fully known availability”: “FKA”), one in which the Poisson rates of the MMPP availability model are estimated and parameters of the state process are known (“partially known availability”: “PKA”), and one in which all the parameters of the MMPP are estimated (“not known availability”: “NKA”). The two single-observer estimators with FKA and MS (those for $h(\cdot) = 1$ and $h(\cdot) < 1$) are not investigated as in these cases the detection of further availability events subsequent to the initial detection does not add any information about the effective strip half-width, compared to the IS case: the performance of the estimator is identical to that of the single-observer estimator in the FKA/IS case.

We therefore have sixteen cases for each of the two scenarios (cf. Tables 1 and 2). The simulations were constructed for situations similar to those from the minke whale survey considered by Skaug and Schweder (1999). For each case, 100 simulations were conducted, where in each simulation run animals passed within perpendicular distance W of observers until 300 animals were detected. The inverse power form (5) of the detection hazard $h(y, x)$ was used, and the shapes of the functions used in the two different scenarios are shown in Figure 3. (Only the functions associated with the detection probability at the initial sighting are shown; in scenarios with multiple sightings and $h(\cdot) < 1$, the hazard functions have similar shapes but are slightly higher for surfacings subsequent to the initial sighting.) We fixed $W = 3000$ and $y_{\max}(x) = 3000$ for all x . The MMPP parameters were specified as $q_{12} = 0.002, q_{21} = 0.001, \lambda_1 = 0.01, \lambda_2 = 0$ in Scenario I and as $q_{12} = 0.002, q_{21} = 0.001, \lambda_1 = 0.002, \lambda_2 = 0$ in Scenario II. In both scenarios, there are hence no surfacings in state 2, the “diving” state. The expected dive cycle duration in both scenarios is 1500 units. Furthermore, in both scenarios 2/3 of the time is spent in state 2, the state in which animals cannot be detected because they do not come to the surface. In state 1, surfacings occur on average every 100 units in Scenario I and every 500 units in Scenario II. Animals on the trackline ($x = 0$) are detected with probability approximately 0.79 and

Table 1

Simulation scenario I: percentage biases and coefficients of variation (in parentheses) of ESHW estimates for observer 1 (true ESHW: 1618) obtained in different scenarios (100 simulation runs in each scenario); FKA, PKA, and NKA stand for full/partial/no knowledge of the availability process, and IS and MS stand for “initial sighting(s) only” and “multiple sightings”, respectively. In case of multiple sightings, $h_s(\cdot)$ indicates whether or not detections of an individual’s surfacings following its initial detection were considered to be certain ($h_s(\cdot) = 1$) or probabilistic but with changed parameter values of the detection hazard function ($h_s(\cdot) < 1$).

Single-observer estimator			
	IS	MS $h_s(\cdot) < 1$	MS $h_s(\cdot) = 1$
FKA	-3.32 (0.13)	/	/
PKA	2.63 (0.16)	-4.29 (0.09)	-3.69 (0.09)
NKA	-39.59 (0.84)	-4.48 (0.10)	-3.98 (0.09)
Double-observer estimator			
	IS	MS $h_s(\cdot) < 1$	MS $h_s(\cdot) = 1$
FKA	0.61 (0.03)	0.18 (0.03)	0.07 (0.03)
PKA	1.11 (0.05)	0.59 (0.04)	0.44 (0.03)
NKA	0.71 (0.08)	-0.36 (0.05)	0.15 (0.04)

Table 2

Simulation scenario II: percentage biases and coefficients of variation (in parentheses) of ESHW estimates for observer 1 (true ESHW: 251) obtained in different scenarios (100 simulation runs in each scenario); FKA, PKA, and NKA stand for full/partial/no knowledge of the availability process, and IS and MS stand for “initial sighting(s) only” and “multiple sightings”, respectively. In case of multiple sightings, $h_s(\cdot)$ indicates whether or not detections of an individual’s surfacings following its initial detection were considered to be certain ($h_s(\cdot) = 1$) or probabilistic but with changed parameter values of the detection hazard function ($h_s(\cdot) < 1$).

Single-observer estimator			
	IS	MS $h_s(\cdot) < 1$	MS $h_s(\cdot) = 1$
FKA	-29.93 (1.06)	/	/
PKA	36.25 (1.17)	-39.71 (1.12)	-20.58 (0.91)
NKA	14.33 (1.37)	-51.69 (1.35)	-19.35 (0.92)
Double-observer estimator			
	IS	MS $h_s(\cdot) < 1$	MS $h_s(\cdot) = 1$
FKA	-1.18 (0.12)	1.28 (0.12)	1.29 (0.13)
PKA	1.20 (0.27)	2.93 (0.20)	1.08 (0.15)
NKA	-32.77 (0.83)	-4.93 (0.37)	1.99 (0.18)

0.83 in Scenario I with one and with two observers, respectively, and with probability approximately 0.22 and 0.27 in Scenario II with one and with two observers, respectively.

In some scenarios in which not all parameters are identifiable, it is still possible to estimate effective strip half-width. For example, in the PKA/IS single-observer case, it is possible to estimate $\lambda_1\theta_1$, the product of the surfacing rate and the intercept of the detection hazard function, even though λ_1 and θ_1 are not individually identifiable. In these cases, we reparameterized the model in terms of these estimatable compound parameters.

4.2. Simulation Results

Simulation results are shown in 1 and 2. In all cases, we anticipate better estimator performance in Scenario I than Scenario II, because the mode of $f_i(l_1|x=0)$ is well ahead of the observer in Scenario I but not Scenario II. Estimation of the probability of detection at $x = 0$ amounts to estimation of $F_l(y_{\max}(0)|x=0)$, i.e., estimation of the proportion of the area under $f_i(l|x=0)$ from 0 to ∞ that occurs before $l = y_{\max}(0)$. To estimate this proportion, one has to effectively estimate the area under $f_i(l|x=0)$ in that part of the support of $f_i(l|x=0)$ from which one cannot get data (i.e., for $l > y_{\max}$). If there is a clear decline in $f_i(l|x=0)$ before $l = y_{\max}(0)$ (the region from which one does get data), then given that $f_i(l|x=0)$ is unimodal and smooth, the data are somewhat informative about the area under $f_i(l|x=0)$ beyond $l = y_{\max}$, and so one is better able to estimate the proportion than if there is no decline. By contrast, if $f_i(l|x=0)$ increases monotonically with l for $l < y_{\max}(0)$, there is very little information in the data about where the mode of $f_i(l|x=0)$ occurs, how high it is, and how long it takes for $f_i(l|x=0)$ to decrease to zero. This makes Scenario II more difficult than Scenario I.

We start by considering the results for the IS (initial sighting only) simulations. The double-observer estimator is less biased (much less in the case of Scenario II) and in all cases much more precise than the single-observer estimator. Indeed, the performance of the single-observer estimator is too poor in Scenario II to be practically useful (with high bias and coefficients of variation greater than 100% in all cases). The double-observer estimator is very much better than the single-observer estimator in the IS case. However, the high bias (-32%) and coefficient of variation (83%) of the double-observer estimator with no knowledge of the availability process (NKA) in Scenario II means it may be of questionable utility with initial sightings only in scenarios like this when the availability process is unknown.

The picture is similar in the case of MS ($h_s(\cdot) < 1$). The performance of the single-observer estimator improves to the extent that it is now practically useful in Scenario I (although not Scenario II), but it remains substantially worse than that of the double-observer estimator. In addition, for all of the practically useful estimators, performance is better in the case of MS ($h_s(\cdot) < 1$) than IS, and when the estimator has no knowledge of the MMPP (rows NKA in Table 1), it is quite substantially better. Moreover, the double-observer estimator performs well in the NKA case for both Scenarios. Improved performance is not surprising: when one is ignorant of the MMPP parameters, getting repeat detections of individuals’

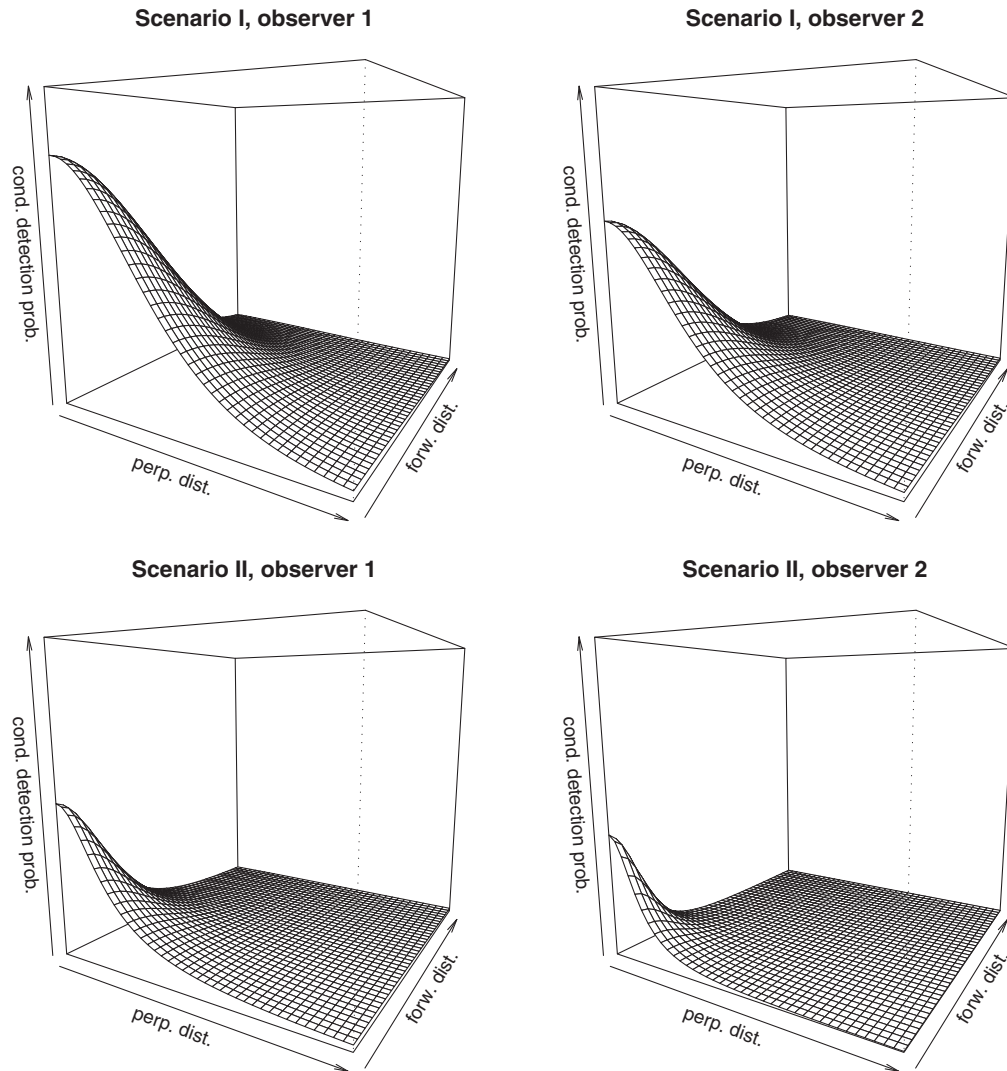


Figure 3. Detection hazard function, i.e., the probability of a surfacing at forward distance l and perpendicular distance x leading to a sighting, in Scenarios I and II and in each case for the two observers at the initial sighting.

availability events improves one's ability to draw inferences about these parameters.

The performance of the estimators in the MS ($h_s(\cdot) = 1$) simulations is broadly similar to that with MS ($h_s(\cdot) < 1$), but estimators are for the most part slightly less biased and slightly less variable. With biases of around 20% and coefficients of variation of about 0.9, the single-observer estimator remains practically unusable in Scenario II.

5. An Analysis of a Double-Observer Survey of Minke Whales

To estimate minke whale abundance in the Northeastern Atlantic, line transect surveys with two independent observer platforms were conducted from 1996 to 2001 (see Skaug et al., 2004). Transects were located using a systematic design with random start point, and the randomization makes it reasonable to assume that animals are uniformly distributed with respect to perpendicular distance from the lines (i.e., $p(x) = 1/W$; here $W = 2000$). After its initial sighting, each

animal was continually observed, with the encounter history giving all the animal's sighting events until it had either been seen by observers on both platforms or it had left the detectable range (case MS in the simulations above, with the slight difference that here observation of an animal stopped when it was seen from both platforms). In this survey, it can reasonably be assumed (Skaug, *pers. commun.*) that all availability events of an animal after its initial detection will be detected by the platform making the initial detection ($h_s(\cdot) = 1$ in the simulations above). A total of 870 minke whales were detected in the survey, with between 1 and 8 sighting events recorded per animal (1.7 on average).

Separate data on minke whale surfacing times were gathered by VHF tagging of eight different minke whales (Schweder et al., 1999). Using these auxiliary data, we first implemented the FKA scenario, fitting a two-state MMPP to the auxiliary data (with one state being diving, in which no surfacing events occur), and then embedding this MMPP in the double-observer shipboard survey analysis. Beaufort

Table 3

Minke whale data: estimated effective strip half-width for both platforms combined (μ), estimated mean state dwell-times (in seconds) in the surfacing and diving states (d_{surf} and d_{dive} , respectively), and estimated mean inter-surfacing times in the surfacing state (λ_1); all estimates are given with 95% confidence intervals (obtained through bootstrapping).

	FKA		PKA		NKA	
	estimate	95% C.I.	estimate	95% C.I.	estimate	95% C.I.
μ for $B \in \{0, 1\}$	681	(600, 762)	282	(228, 375)	267	(143, 382)
μ for $B = 2$	504	(444, 572)	199	(160, 258)	188	(100, 268)
μ for $B \in \{3, 4\}$	351	(315, 395)	131	(109, 168)	124	(68, 173)
d_{surf}	100	(44, 231)	136	(106, 184)	141	(102, 205)
d_{dive}	53	(31, 83)	255	(205, 287)	285	(154, 715)
λ_1	36	(29, 43)	68	(60, 77)	68	(60, 77)

sea state, B , was included in detection hazard functions as a factor with three levels: $B \in \{0, 1\}$ (calm or light air), $B = 2$ (light breeze), and $B \in \{3, 4\}$ (gentle or moderate breeze); the survey effort only involved Beaufort states 0–4. Following Skaug et al. (2004), the sea state was assumed to influence the scale parameter of the hazard function (Equations (4) or (5)): $\log(\theta_{i,2}) = \beta_{i,0} + \beta_{i,1} I_{\{B \in \{0,1\}\}} + \beta_{i,2} I_{\{B \in \{3,4\}\}}$, where the subscript i , $i = 1, 2$, indicates the i -th observer, and where I denotes the indicator function and $\beta_{i,0}$ is the value of the predictor for $B = 2$.

The likelihood was maximized numerically using `nlm` in R. The exponential power hazard model (Equation (4)) yielded $\log(\mathcal{L}(\boldsymbol{\beta})) = -17446.45$, while the inverse power hazard model (Equation (5)) gave $\log(\mathcal{L}(\boldsymbol{\beta})) = -17432.99$; in the following, we focus on the models that use the inverse power model. The likelihood ratio test for the hypothesis that the sea state has no influence on the scale parameter of the hazard function, i.e., that $\beta_{1,1} = \beta_{1,2} = \beta_{2,1} = \beta_{2,2} = 0$, yields the test statistic 125.5 and hence a p-value of 0, and the hypothesis is rejected.

We also implemented the scenarios PKA and NKA. With PKA, the model was fitted to the survey data conditional on the mean dive cycle length of 391.5 seconds. This value was chosen based on Christiansen et al. (2011), who fitted HMMs to inter-surfacing times and reported mean surfacing intervals of 43 seconds in the surfacing state and 155 seconds in the deep dive state, with a dive cycle comprising “five to six regular dives [in the surfacing state] followed by one deep dive.” Table 3 gives the estimates and associated confidence intervals of the MMPP availability process parameters and of the effective strip half-widths, for the three different scenarios considered (FKA, PKA, and NKA). To obtain interval estimates, we conducted, in each of the three scenarios, a bootstrap with 1000 iterations, in each iteration drawing a non-parametric bootstrap sample from the distances associated with sightings (i.e., sampling with replacement from the observed triplets of forward distances, perpendicular distances, and associated covariate value), and, in case of the FKA scenario, an additional parametric bootstrap sample from the models for surfacings (i.e., generating observations from the fitted MMPP availability model and then refitting the model). In case of the PKA scenario, the uncertainty in the expected dive cycle length was accounted for using the uncertainty quantification given in Christiansen et al. (2011).

The differences between the FKA estimates on the one hand and the PKA and NKA estimates on the other are striking. ESHW estimates for PKA and NKA are about 40% of those from FKA and would result in density estimates being about 2.5 times larger for PKA and NKA than FKA. The difference between FKA and PKA estimates is a consequence of the fact that mean dive cycle length obtained by fitting to the minke VHF tag data is 153 seconds, while that of Christiansen et al. (2011) is 2.55 times longer. Estimates of mean dive cycle length and other parameters for NKA are broadly consistent with those from PKA. The differences in dive cycle length and related parameters may be a consequence of the diving behavior at the time and location of the survey being different from those of VHF-tagged animals (which were obtained at different times and locations)—in which case the PKA or NKA estimates are more appropriate. They might also be a consequence of failure of the assumption that all surfacings after initial detection are seen—in which case something closer to the FKA estimates may be appropriate.

6. Discussion

We have extended the approach developed by Langrock et al. (2013) for single-observer LT surveys with stochastic animal availability, to allow for MR data generated by double observers. Viewed from an MR perspective, the model is a continuous-time variant of the closed-population conditional MR likelihoods of Huggins (1989) and Alho (1990), with detection probability that changes with time, with an MMPP temporary immigration/emigration structure (rather than the completely random or Markovian immigration/emigration structures of Kendall et al., 1997, for example), and with an observed random effect (namely x) in detection probability. As such, it can be classed as a continuous-time MR model of type M_{th} (Otis et al., 1978) with temporary immigration and emigration. If we allow the parameters of the hazards to change after first detection, then it is of type M_{tth} .

The new approach was shown to improve estimation precision and allows more reliable estimation of ESHW and abundance. It also allows estimation using limited or no independent information on the availability process in some circumstances, particularly when data include multiple sightings of animals by the same observer. Including multiple sightings roughly halved estimator coefficient of variation

in our (simulation) Scenario I and reduced it by a factor of between 2 and 4 in our Scenario II.

The estimates of ESHW from the minke survey were sensitive to assumptions about the availability process. As it is often the case that little is known about how availability patterns vary in time and space, using availability process estimates external to the survey may involve strong assumptions, which if violated, can result in biased ESHW and abundance estimates. There is therefore merit in being able to estimate availability process parameters simultaneously with detection hazard parameters from the LT survey data. The cost of incorporating resightings to do this is the need to model how detection probability changes for subsequent sightings, except in cases where detection probability does not change after initial detection. We considered only the case in which all hazard parameters change after first sighting and then stay unchanged. It is possible that only some parameters change, and/or that change depends on how many times an animal has been detected by that point and/or how much time has passed. In principle one can construct hazards accordingly and choose between models using model selection criteria. Sample size may of course limit the ability to distinguish between models. Nevertheless, using resightings data seems to us something worth doing as these data contain information on availability and one may avoid substantial bias by estimating availability *in situ*.

In the light of our analyses, we recommend: (i) that double-observer survey methods be used when this is feasible and (ii) that consideration be given to instructing observers to record all detections of animals, not just the initial detection. If it is infeasible to record multiple detections in areas of high animal density, we recommend that this be attempted only in lower density areas. Detection probability estimates from such data will rely on the assumption that availability processes are similar in high- and low-density areas, and surveyors should consider whether this is a plausible assumption in the context of their surveys.

A final issue, to which we alluded in Section 1, is bias due to unmodelled heterogeneity. We can investigate how much of the bias from unmodelled stochastic availability has been accounted for by modeling the availability process, by comparing detection probability from our model ($p = \mu/W$) with detection probability from the so-called “full independence” (FI) model (see Borchers et al., 2006; Burt et al., 2015), which assumes independence conditional on x : $p_{FI} = \int_0^W p_{FI}(x)\pi(x)dx$, where $p_{FI}(x) = 1 - \{1 - p_1(x)\}\{1 - p_2(x)\}$, $p_i(x) = F_{i(i)}(y_{\max}(z), z; \theta)$, and $F_{i(i)}()$ is the same as $F_i()$, but evaluated using $h_i()$ in place of $h()$ ($i = 1, 2$). For Scenarios I and II, p is 5 and 10% smaller than p_{FI} , respectively, suggesting that p_{FI} is biased by 5% to 10% due to neglecting stochastic availability. (Bias can be much larger; we constructed other scenarios with bias of more than 30%.)

Neglect of other variables may also cause bias. We incorporated sea state in our analysis of minke sightings but there may be unrecorded variables that affect detectability of animals. MRDS estimators deal with this using “point independence” (PI) models (see Borchers et al., 2006; Burt et al., 2015, for details). These model the shape of the detection function using a model, $p_{\text{shape}}(x)$, with intercept 1, for the

distribution of detections by any observer in the x dimension (without using recapture data or assuming independence between observers’ detections), and a separate FI model, $p(x)$, that uses the recapture data and assumes independent detections, given x . They are combined thus to estimate detection probability: $p(0)p_{\text{shape}}(x)$. One could do the same at the level of the detection hazard, conditional on (l, x) . For this one needs a model $h_{\text{shape}}(l, x)$, with intercept 1, for the distribution of detections by any observer in the (l, x) plane. A PI hazard function might then be $h(y_{\max}(0), 0)h_{\text{shape}}(l, x)$, where $h(y_{\max}(0), 0)$ is $h(l, x) = 1 - \{1 - h_1(l, x)\}\{1 - h_2(l, x)\}$ evaluated at $(l = y_{\max}(0), x = 0)$.

There are, however, difficulties that arise when using PI at the hazard level rather than the perpendicular distance detection function level. For example, having a single $h_{\text{shape}}(l, x)$ is in general inadequate because animals that are undetected by l should have a different $h_{\text{shape}}(l, x)$ from animals seen by observer 1 only by l , and these may have a different $h_{\text{shape}}(l, x)$ from animals seen by 2 only by l . In general, one would require three related $h_{\text{shape}}(l, x)$ functions (and need to estimate up to three sets of parameters), and it is not obvious how they should be related, without making the kind of independence assumption that one is trying to avoid by constructing a PI model. This is a topic for future research.

7. Supplementary Materials

The R code used to fit the models and generate the results of this article is available at the *Biometrics* website on Wiley Online Library.

ACKNOWLEDGEMENTS

This work was part-funded by EPSRC Grant EP/I000917/1. We are grateful to Nils Oien for giving us access to the minke whale data, and to Hans J. Skaug for providing a detailed description of the data and for useful discussions of this work.

REFERENCES

- Alho, J. M. (1990). Logistic regression in capture–recapture models. *Biometrics* **46**, 623–635.
- Borchers, D. L. and Burnham, K. P. (2004). General formulation for distance sampling. In *Advanced Distance Sampling*, S. T. Buckland, D. R. Anderson, K. P. Burnham, J. L. Laake, D. L. Borchers, and L. J. Thomas (eds), Oxford, UK: Oxford University Press.
- Borchers, D. L., Laake, J. L., Southwell, C., and Paxton, C. G. M. (2006). Accommodating unmodelled heterogeneity in double-observer distance sampling surveys. *Biometrics* **62**, 372–378.
- Borchers, D. L., Zucchini, W., Heide-Jorgensen, M. P., Canadas, A., and Langrock, R. (2013). Using hidden markov models to deal with availability bias on line transect surveys. *Biometrics* **69**, 703–713.
- Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. J. (2001). *Introduction to Distance Sampling*. London, Oxford: Oxford University Press.

- Buckland, S. T., Anderson, D. R., Burnham, K. P., Laake, J. L., Borchers, D. L., and Thomas, L. J. (2014). *Advanced Distance Sampling*. London, Oxford: Oxford University Press.
- Burt, M., Borchers, D., Jenkins, K., and Marques, T. (2015). Using mark-recapture distance sampling methods on line transect surveys. *Methods in Ecology and Evolution* **5**, 1180–1191.
- Christiansen, F., Rasmussen, M., and Lusseau, D. (2011). Inferring surface time of minke whales from inter-surfacing interval data using a hidden markov model. Technical Report, IWC paper SC/63/RMP6.
- Hiby, L. and Lovell, P. (1998). Using aircraft in tandem formation to estimate abundance of harbour porpoise. *Biometrics* **54**, 1280–1289.
- Huggins, R. M. (1989). On the statistical analysis of capture experiments. *Biometrika* **76**, 133–140.
- Kendall, W. L., Nichols, J. D., and Hines, J. E. (1997). Estimating temporary emigration using capture–recapture data with Pollock’s robust design. *Ecology* **78**, 563–578.
- Langrock, R., Borchers, D. L., and Skaug, H. J. (2013). Markov-modulated nonhomogeneous poisson processes for modeling detections in surveys of marine mammal abundance. *Journal of the American Statistical Association* **108**, 840–851.
- Okamura, H. (2003). A line transect method to estimate abundance of long-diving animals. *Fisheries Science* **69**, 1176–1181.
- Okamura, H., Kitakado, T., Hiramatsu, K., and Mori, M. (2003). Abundance estimation of diving animals by the double-platform line transect method. *Biometrics* **59**, 512–520.
- Okamura, H., Minamikawa, S., Skaug, H. J., and Kishiro, T. (2012). Abundance estimation of long-diving animals using line transect methods. *Biometrics* **68**, 504–513.
- Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). Statistical inference from capture data on closed animal populations. *Wildlife Monographs* **62**, 1–135.
- Rekdal, S., Hansen, R., Borchers, D., Bachmann, L., Laidre, K., Wiig, O., Nielsen, N., Fossette, S., Tervo, O., and Heide-Jorgensen, M. (2014). Trends in bowhead whales in west greenland: Aerial surveys vs. genetic capture–recapture analyses. *Marine Mammal Science* **31**, 133–154.
- Schweder, T., Hagen, G., Helgeland, J., and Koppervik, I. (1996). Abundance estimation of northeastern Atlantic minke whales. *Report of the International Whaling Commission* **46**, 391–405.
- Schweder, T., Skaug, H. J., Langaas, M., and Dimakos, X. K. (1999). Simulated likelihood methods for complex double-platform line transect surveys. *Biometrics* **55**, 678–687.
- Skaug, H. and Schweder, T. (1999). Hazard models for line transect surveys with independent observers. *Biometrics* **55**, 29–36.
- Skaug, H. J., Oien, N., Schweder, T., and Bothun, F. G. (2004). Current abundance of minke whales (*balaenoptera acutorostrata*) in the northeast atlantic; variability in time and space. *Canadian Journal of Fisheries and Aquatic Sciences* **61**, 870–886.

Received March 2014. Revised April 2015. Accepted April 2015.