Spatially explicit capture-recapture methods to estimate minke whale density from data collected at bottom-mounted hydrophones

Tiago A. Marques, Len Thomas, Stephen W. Martin, David K. Mellinger, Susan Jarvis, Ronald P. Morrissey, Carroll-Anne Ciminello, Nancy Dimarzio

Manuscript Number: JORN-D-10-00021R2

Title: Spatially explicit capture recapture methods to estimate minke whale abundance from data collected at bottom mounted hydrophones

Article Type: EURING Proceedings

Keywords: Bayesian statistics; density; minke whale; passive acoustic monitoring; spatially explicit capture recapture; WinBUGS

Corresponding Author: Dr. Tiago André Marques,

Corresponding Author's Institution: University of St Andrews

First Author: Tiago André Marques

Order of Authors: Tiago André Marques; Len Thomas; Steve W Martin; David K Mellinger; Susan Jarvis; Ron P Morrissey; Carol-Anne Ciminello; Nancy DiMarzio

Abstract: Estimation of cetacean abundance or density using visual methods can be cost ineffective under many scenarios. Methods based on acoustic data have recently been proposed as an alternative, and could potentially be more effective for visually elusive species that produce loud sounds. Motivated by a data set of minke whale (\textit{Balaenoptera acutorostrata}) boing sounds detected at multiple hydrophones at the U.S. Navy's Pacific Missile Range Facility (PMRF), we present an approach to estimate density or abundance based on spatially explicit capture recapture (SECR) methods. We implement the proposed methods in both a likelihood and a Bayesian framework. The point estimates for abundance and detection parameters from both implementation methods are very similar and agree well with current knowledge about the species. The two implementation approaches are compared in a small simulation study. While the Bayesian approach might be easier to generalize, the likelihood approach is faster to implement (at least in simple cases like the one presented here) and more readily amenable to model selection. SECR methods seem to be a strong candidate for estimating density from acoustic data where recaptures of sound at multiple hydrophones/microphones are available, and we anticipate further development of related methodologies in the future.

Journal of Ornithology manuscript No.
(will be inserted by the editor)

Spatially explicit capture recapture methods to estimate minke whale density from data collected at bottom mounted hydrophones

Tiago A Marques · Len Thomas · Stephen W. Martin · David. K. Mellinger · Susan Jarvis · Ronald P. Morrissey · Carroll-Anne Ciminello · Nancy DiMarzio

Received: date / Accepted: date

Tiago A Marques · Len Thomas (✉)
Centre for Research into Ecological and Environmental Modelling, The Observatory, Buchanan Gardens, University of St Andrews, Fife, KY16 9LZ, Scotland
Tel.: 0044 1334 461801
Fax: 0044 1334 461800
E-mail: len@mcs.st-and.ac.uk

Stephen W. Martin
Space and Naval Warfare Systems Center Pacific, 53560 Hull Street, San Diego, California 92152, U.S.A.

David. K. Mellinger
Cooperative Institute for Marine Resources Studies, Oregon State University, 2030 SE Marine Science Drive, Newport, Oregon 97365, U.S.A.

Susan Jarvis · Ronald P. Morrissey · Carol-Anne Ciminello · Nancy DiMarzio
Naval Undersea Warfare Center Division, 1176 Howell Street, Newport, Rhode Island 02841, U.S.A.

Abstract Estimation of cetacean abundance or density using visual methods can be cost ineffective under many scenarios. Methods based on acoustic data have recently been proposed as an alternative, and could potentially be more effective for visually elusive species that produce loud sounds. Motivated by a data set of minke whale (*Balaenoptera acutorostrata*) boing sounds detected at multiple hydrophones at the U.S. Navy's Pacific Missile Range Facility (PMRF), we present an approach to estimate density or abundance based on spatially explicit capture recapture (SECR) methods. We implement the proposed methods in both a likelihood and a Bayesian framework. The point estimates for abundance and detection parameters from both implementation methods are very similar and agree well with current knowledge about the species. The two implementation approaches are compared in a small simulation study. While the Bayesian approach might be easier to generalize, the likelihood approach is faster to implement (at least in simple cases like the one presented here) and more readily amenable to model selection. SECR methods seem to be a strong candidate for estimating density from acoustic data where recaptures of sound at multiple acoustic sensors are available, and we anticipate further development of related methodologies.

Keywords minke whale · passive acoustic monitoring · proximity detector · secr · spatially explicit capture recapture · OpenBUGS

1 Introduction

The estimation of animal density and abundance is a fundamental requirement for effective management and conservation decisions. However, this is particularly challenging for many cetacean species, which typically occur over very large areas, at low densities, and spend a large proportion of their time submersed. All this makes them especially challenging to survey using standard visual methods. These include distance sampling methods, namely shipboard and aerial surveys in which line transects or cue counting approaches are used (see Buckland et al (2001) for details), as well as capture-recapture methods (e.g. Evans and Hammond, 2004) based on photo-ID or DNA. While working well under certain circumstances, all of these methods have several shortcomings. Low encounter rates create problems in analysis and low precision in the estimates, and surveys are restricted to good weather and daylight conditions. This makes them cost ineffective for many scenarios.

In recent years, acoustic data has been proposed as having information about density (Mellinger et al, 2007). Common sense alone suggests that the amount of animal-produced sound (however it is measured) might act as an index of animal abundance. The challenge is to find ways to convert that amount of sound to animal density. Using sound to detect and localize animals from towed hydrophone arrays has been successfully implemented for sperm whales (e.g. Barlow and Taylor, 2005). However, this approach does not really differ from an analysis perspective from conventional line transect distance sampling. On the other hand, Marques et al (2009) presented the first example in which data from fixed hydrophones was used to estimate cetacean density, using an approach akin to cue counting. Unlike for conventional cue counting, the detection function was estimated using a regression based approach using a sample of data for which the animal locations and vocalizations were known from acoustic dive tags.

If one has an array of fixed hydrophones, sounds detected at multiple hydrophones can be seen as capture-recapture data. Each sound can be assigned a capture history, say a 1 for each hydrophone where it was detected and a 0 on hydrophones where it was missed. This assumes we can tell when the same sound is received at multiple hydrophones, say from timing and/or frequency information, just as we assume we can tell when the same individual is sighted in a photo ID study. Standard capture-recapture analyses could be undertaken, but there are several reasons to focus on the use of spatially explicit capture-recapture (SECR, Borchers and Efford (2008); Borchers (this volume)) methods for estimating density. Firstly, SECR methods explicitly model the dependence of capture probability on distance, thereby reducing the un-modelled heterogeneity that usually hinders capture-recapture analysis. For acoustic data we may expect distance of the sound source to be a major component of capture probability. Secondly, SECR methods estimate density and abundance over an explicitly-defined area, as opposed to traditional methods where the area sampled is not clearly defined and hence converting abundance to density is problematic. SECR has received considerable attention recently, both from classical likelihood (e.g. Efford et al, 2008; Borchers and Efford, 2008; Efford et al, 2009) and Bayesian (e.g. Royle and Young, 2008; Royle, 2009; Royle et al, 2009a) perspectives. In particular Dawson and Efford (2009) have applied likelihood-based methods to acoustic data, estimating bird density from a set of 4 microphones.

In this paper, we compare Bayesian and likelihood based approaches to SECR, using as motivation the estimation of density of sounds produced by common minke whales (*Balaenoptera acutorostrata*) off the coast of Kauai, Hawaii. Minke whales are one of the most abundant baleen whale species worldwide, but they are also one of the smallest and can be very difficult to detect using standard visual survey methods. Although commonly sighted in high latitude waters, they are rarely seen in tropical and sub-tropical areas, despite being heard there during winter and spring. This is particularly true around the Hawaiian Islands, where extensive aerial and shipboard surveys (e.g. Mobley Jr. et al, 1999; Rankin et al, 2007) have produced only a handful of sightings, but the characteristic “boing” sound attributed to minke whales (Barlow and Taylor, 2005) can be detected readily in the right season (e.g. Rankin et al, 2007). Therefore, methods based on acoustic rather than visual detections might prove more effective at estimating their abundance.

The data used here come from a set of 16 bottom-mounted hydrophones that are part of the U.S. Navy’s Pacific Missile Range Facility (PMRF), an instrumented testing range located along the western shore of Kauai. The hydrophones are part of the Barking Sands Underwater Range Expansion (BSURE), which extends northwest of the island, covering approximately 2,300km^2 and having water depths to 4,500m throughout most of its area. While the hydrophones were designed for tracking underwater objects such as submarines and torpedos, they are capable of detecting minke whale boing vocalizations, and are therefore well suited to study this cryptic cetacean.

The paper is structured as follows. The next section describes the Bayesian and likelihood based approaches to SECR. These are then applied to the case study data in section 3. Section 4 presents a simulation study evaluating performance of the two approaches in a simple scenario that mimics the case study. Lastly, a discussion section gives the main conclusions and suggests potential avenues for future investigation.

It is not our intention in this paper to provide a definitive estimate of minke whale (sound) density in the study area; instead we focus on establishing the utility of the SECR methodology and comparing approaches to analysis. This is (to our knowledge) the first time that: (1) both a Bayesian and likelihood SECR implementation are directly compared, (2) a Bayesian SECR model has been applied to acoustic data, here using proximity detectors (see below for details) and (3) SECR is proposed to estimate density of cetaceans.

2 SECR models and inference

In this section, we outline the models and estimation approaches. A non-technical introduction to SECR models is given by Borchers (this volume); more details of the likelihood-based methods are in Borchers and Efford (2008) and Efford et al (2009); details of a similar Bayesian method (with different types of detectors) is in Royle and Young (2008) and Royle et al (2009a).

As initially conceived, SECR models consider that animals' home range centres are at unobserved locations $\mathbf{X} = (X_1, X_2, ..., X_N)$, where X_i represents the position of animal i (i.e., its cartesian coordinates in 2-dimensional space). Inference is focused on estimating N, the number of home range centres in a given area A (e.g. Borchers and Efford, 2008; Royle and Young, 2008), as well as density, $D = N/A$. In the current scenario, the equivalent to the home range centres are the actual sound source locations. Hence the focus of our estimate using SECR is the number and density of sound sources within a given area A and time period T. The estimation of the actual animal abundance requires dividing the estimated sound source abundance $\hat{N}$ by T and sound production rate, and we do not deal with this here.

The Bayesian and likelihood approaches have several differences (details below), so we deal with them separately in the sections below. However, they both use the same model for sound detection, as follows. Consider an array of K hydrophones, each with known location. A sound produced at location X is detected at hydrophone k ($k = 1, \ldots, K$) with probability $p_k(X; \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is a vector of detection parameters. Hydrophones operate independently, so that the probability a sound is detected on at least one hydrophone is $p_{\cdot}(X; \boldsymbol{\theta}) = 1 - \prod_{k=1}^{K} (1 - p_k(X; \boldsymbol{\theta}))$. Detection probability is assumed to be a non-increasing function of horizontal distance $d_{X,k}$ between sound source at X and hydrophone k: $p_k(X;\boldsymbol{\theta}) = g(d_{X,k};\boldsymbol{\theta})$. There are many candidate models for the distance detection function g; here we use three, all of which have a long history in the classical distance sampling literature:

1. half-normal $g(d;\boldsymbol{\theta}) = g_0 \exp(-d^2/(2\sigma^2))$ with $\boldsymbol{\theta} = (g_0, \sigma)$
2. hazard rate $g(d;\boldsymbol{\theta}) = g_0(1 - \exp(-(d/\sigma)^{-z}))$ with $\boldsymbol{\theta} = (g_0, \sigma, z)$
3. negative exponential $g(d;\boldsymbol{\theta}) = g_0 \exp(-d/\sigma))$ with $\boldsymbol{\theta} = (g_0, \sigma)$

Detectors such as this, where an object (in this case a sound) can be "captured" on more than one detector and where the detector can capture many objects in one "trapping session", are termed "proximity detectors" by Efford et al (2008). Other detectors (not relevant to passive acoustics) are described in that paper. For simplicity in what follows, we consider only one "trapping session", although generalization to multiple sessions (with potentially varying animal densities and/or detection parameters) is simple.

Assume n sounds in the period of interest were detected on one or more hydrophones. Let ω_{ik} represent the detection of the ith sound at the kth hydrophone, such that $\omega_{ik} = 1$ if the sound was detected, 0 otherwise. $\boldsymbol{\omega}_i$ is the capture history of the ith sound, and $\boldsymbol{\omega}$ is all the recorded capture histories.

2.1 Likelihood-based methods

For simplicity, we assume that sound source locations are distributed in space according to a homogeneous Poisson process with intensity D. Extension to an inhomogeneous process is conceptually straightforward, and is given by Borchers and Efford (2008). The joint likelihood for D and the detection parameters $\boldsymbol{\theta}$ given the capture histories

$\boldsymbol{\omega}$ can be written

$$L(D, \boldsymbol{\theta}|n, \boldsymbol{\omega}) = Pr(n|D, \boldsymbol{\theta})Pr(\boldsymbol{\omega}|n, \boldsymbol{\theta}) \tag{1}$$

where $Pr(n|D, \boldsymbol{\theta})$ is the marginal distribution of the number of sound sources detected, n, and $Pr(\boldsymbol{\omega}|n, \boldsymbol{\theta})$ is the conditional distribution of the capture histories given n. (In the inhomogeneous Poisson case the latter distribution will also depend on the spatial intensity parameters.)

Given the assumption that sound source locations follow a Poisson process, then n is the outcome of a thinned Poisson process, which has a Poisson distribution with parameter $Da(\boldsymbol{\theta})$, where $a(\boldsymbol{\theta}) = \int_{X \in A} p.(X; \boldsymbol{\theta})dX$ has an intuitive interpretation as the "effective sample area" (see Borchers, this volume, for details). The area A needs to be large enough that no detections can occur from outside of it; in practice Efford (2009) suggests it is sufficient to define A as a rectangle with limits formed by buffering the hydrophones at a distance w such that $g(w) < 0.01$. The first term in (1) is thus

$$Pr(n|D, \theta) = n!^{-1} Da(\theta)^n \exp\left(-Da(\theta)\right). \tag{2}$$

Assuming independence between detections, the second term in (1) can be written

$$Pr(\boldsymbol{\omega}|n, \boldsymbol{\theta}) = \binom{n}{n_1, \ldots, n_C} a(\boldsymbol{\theta})^{-n} \int_{X \in A} \prod_{i=1}^{n} Pr\left(\boldsymbol{\omega}_i | X, \boldsymbol{\theta}\right) dX \tag{3}$$

where the first part is the multinomial coefficient (with $n_1, \ldots, n_C$ representing the frequency of each of the C unique capture histories), the second part ($a(\boldsymbol{\theta})^{-n}$) is there because we condition on the number of observed capture histories, and the remainder is the probability of obtaining capture history $\boldsymbol{\omega}_i$ given sound source location X, integrated over all possible locations. Since hydrophones operate independently,

$$Pr\left(\boldsymbol{\omega}_i | X, \boldsymbol{\theta}\right) = \prod_{k=1}^{K} p_k(X; \boldsymbol{\theta})^{\omega_{ik}} \left(1 - p_k(X; \boldsymbol{\theta})\right)^{(1-\omega_{ik})}. \tag{4}$$

Note that, because the sound source locations are not known, they are integrated out of the likelihood. In practice, to reduce computational effort during maximization, the likelihood is evaluated over a discrete grid of points, and the integrations become sums (Efford et al, 2009). The choice of the grid size is a compromise between computational efficiency and no influence on the results.

One approach to estimation of D, which we call the "full likelihood" approach, is joint maximization of the parameters D and $\boldsymbol{\theta}$ in (1). Variances on parameters can be estimated from the inverse of the information matrix and profile likelihoods can be used to obtain confidence intervals.

An alternative (e.g. Borchers and Efford, 2008) when D is homogeneous is to maximize the conditional likelihood

$$L\left(\boldsymbol{\theta}|n,\boldsymbol{\omega}\right) \propto a(\boldsymbol{\theta})^{-n} \int_{X \in A} \prod_{i=1}^{n} Pr\left(\boldsymbol{\omega}_i|X,\boldsymbol{\theta}\right) dX \tag{5}$$

to obtain estimates of $\boldsymbol{\theta}$ and hence $\hat{a} = a(\hat{\boldsymbol{\theta}})$. From this, D can be estimated using the Horvitz-Thompson-like estimator $\hat{D} = n/\hat{a}$. While the estimates derived from both full and conditional likelihoods are equivalent, the Horvitz-Thompson-like formulation permits different variance estimators for $\hat{D}$ than the full likelihood. Borchers and Efford (2008) suggest two in their supplementary materials, one assuming fixed N in the study area, and the other random N. Both have design-based and model-based components (see Discussion), and can be expected to be more robust than the full likelihood estimator to departures from a Poisson animal distribution. (We note in passing that these two estimators are sometimes referred to as "binomial" and "Poisson", e.g., in Efford (2009), but we prefer to use the terms fixed-N and random-N as neither assumes animals follow binomial or Poisson distributions.) Confidence intervals can be obtained by assuming D follows a normal or log-normal distribution.

All of the above inference can be carried out using the `secr` package (Efford, 2009) in R (R Development Core Team, 2009).

2.2 Bayesian methods

There are several differences between the model described in the previous section and that used here. Firstly, the Bayesian model is parameterised in terms of abundance N within an area A, rather than density D, and this N is assumed to be a fixed quantity. This leads to a binomial likelihood for n given N, rather than the Poisson likelihood for n given D in the previous section.. Secondly, data augmentation is used to deal with the un-observed capture histories and sound source locations. Thirdly, being Bayesian, prior distributions are used on all unknown parameters, although since uniform priors with widely spaced limits are used, the posterior and likelihood surfaces will have the same shape within the truncation bounds.

Let $\widetilde{\boldsymbol{\omega}}$ be the capture histories of the N sound sources in the study area A. n of these are observed, and therefore have one or more non-zero elements; the remaining $(N - n)$ contain only zeros. The equivalent of (3), conditioning on the sound source locations, is then

$$Pr(\widetilde{\boldsymbol{\omega}}|N, \mathbf{X}, \boldsymbol{\theta}) = \binom{N}{n0, \ldots, n_C} \prod_{i=1}^{N} Pr\left(\widetilde{\boldsymbol{\omega}}_i | X_i, \boldsymbol{\theta}\right). \tag{6}$$

Inference is based on the joint posterior distribution

$$\begin{aligned} Pr(N, \mathbf{X}, \boldsymbol{\theta}|\boldsymbol{\omega}) &\propto Pr(N, \mathbf{X}, \boldsymbol{\theta}) Pr(\widetilde{\boldsymbol{\omega}}|\boldsymbol{\omega}, N) L(N, \mathbf{X}, \boldsymbol{\theta}|\widetilde{\boldsymbol{\omega}}) \\ &= Pr(N) Pr(\mathbf{X}|N) Pr(\boldsymbol{\theta}) L(N, \mathbf{X}, \boldsymbol{\theta}|\widetilde{\boldsymbol{\omega}}) \end{aligned} \tag{7}$$

where a discrete uniform prior distribution is used for $Pr(N)$, with lower bound 0 and an (arbitrarily high) upper bound M, and uniform prior distributions are used for $Pr(\mathbf{X}|N)$ and $Pr(\boldsymbol{\theta})$. Note that $Pr(\widetilde{\boldsymbol{\omega}}|\boldsymbol{\omega}, N) = 1$, which is why it disappears from the second line, and $L(N, \mathbf{X}, \boldsymbol{\theta}|\widetilde{\boldsymbol{\omega}})$ has the same form as (6) except that $\widetilde{\boldsymbol{\omega}}$ is the fixed variable.

In practice, the fact that the dimension of both $\widetilde{\boldsymbol{\omega}}$ and $\mathbf{X}$ depend on N raises computational issues. These are sidestepped by a further data augmentation, where (M-N) additional all-zero capture histories and sound source locations are added (see Royle et al (2007) for the general framework, and Royle et al (2009b) and Royle and Young (2008) for applications). Let M be the fixed size of a superpopulation of sound sources, with capture histories $\widetilde{\boldsymbol{\omega}}^*$ and locations $\mathbf{X}^*$. n of the capture histories are observed; the remaining $(M - n)$ contain only zeros. Let $\mathbf{z} = (z_1, \ldots, z_M)$ be a vector of indicator variables, such that $z_i = 1$ if sound source i is part of the population N, 0 otherwise. This means N is now a derived parameter in the model: $N = \sum_{i=1}^{M} z_i$. Let the z_i for each sound source follow a Bernoulli distribution with parameter ψ. Inference is then based on the joint posterior

$$\begin{aligned} Pr(\mathbf{z}, \mathbf{X}^*, \boldsymbol{\theta}, \psi|\boldsymbol{\omega}, M) &\propto Pr(\mathbf{z}, \mathbf{X}^*, \boldsymbol{\theta}, \psi)Pr(\widetilde{\boldsymbol{\omega}}^*|\boldsymbol{\omega}, M)L(\mathbf{z}, \mathbf{X}^*, \boldsymbol{\theta}, \psi|\widetilde{\boldsymbol{\omega}}^*) \\ &= Pr(\mathbf{X}^*)Pr(\boldsymbol{\theta})Pr(\psi)Pr(\mathbf{z}|\psi)L(\mathbf{z}, \mathbf{X}^*, \boldsymbol{\theta}|\widetilde{\boldsymbol{\omega}}^*) \qquad (8) \end{aligned}$$

where uniform prior distributions are used for $Pr(\mathbf{X}^*)$ and $Pr(\boldsymbol{\theta})$, a uniform (0,1) prior is used for $Pr(\psi)$, $Pr(\widetilde{\boldsymbol{\omega}}^*|\boldsymbol{\omega}, M) = 1$, and

$$L(\mathbf{z}, \mathbf{X}^*, \boldsymbol{\theta}|\widetilde{\boldsymbol{\omega}}^*) = \binom{M}{n0, \ldots, n_C} \prod_{i=1}^{M} z_i Pr\left(\widetilde{\boldsymbol{\omega}}_i^*|X_i^*, \boldsymbol{\theta}\right). \qquad (9)$$

The marginal posterior distributions of N, θ and $\mathbf{X}$ from (8) are then same as (7), but implementation is greatly simplified.

The most convenient route to fitting the Bayesian models to data is via short programs written in OpenBUGS (Thomas et al, 2006). An example program is provided as an appendix.

Table 1 Summary of case study data.

Date	Start time (GMT)	# detections	# unique boings (n)	Capture frequency 1	2	3	4	5-9	10-14
5 Mar 06	22:15:55	63	14	2	2	4	2	2	2
13 Mar 06	23:05:28	75	12	1	3	0	0	6	2
19 Apr 06	02:59:40	6	5	4	1	0	0	0	0
19 Apr 06	04:29:40	12	10	8	2	0	0	0	0
16 Apr 07	03:52:20	41	9	2	2	0	1	3	1
21 Apr 07	02:49:43	36	7	1	0	1	1	3	1
Total		233	57	18	10	5	4	14	6

3 Case study

3.1 Case study methods

The data come from six 10-minutes sample periods, taken from March and April 2006 and 2007 (Table 1). For this simple analysis, we collapsed data over sampling occasions and treated them as a single 1 hour period. Sounds were recorded at 16 bottom-mounted hydrophones in BSURE, spaced from 8km to 18km apart and arranged in two lines (Figure 1). A custom-developed detector and classifier (Mellinger et al., unpublished) was utilized to detect minke whale boing vocalizations on the multiple hydrophones. The boing detector outputs included detailed timing and frequency content information. This information was utilized to make initial manual associations (i.e., determine whether detections at different hydrophones were of the same sound). The association outputs served as inputs to the SECR analysis.

Likelihood-based models were fit using the `secr` package (version 1.2.10) in R (version 2.9.2). As mentioned in section 2.1, it is necessary to define a buffer distance w

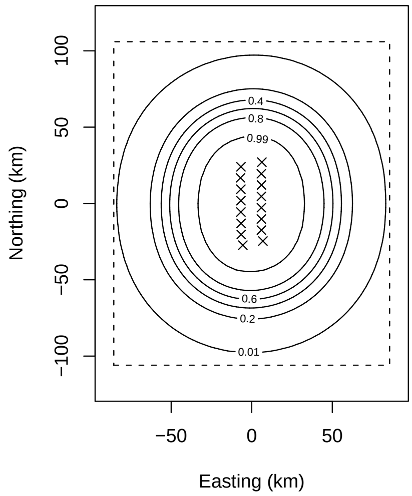

Fig. 1 Layout of BSURE case study hydrophones (crosses), solid contour lines showing probability of detecting a sound from that location with one or more hydrophones (denoted $p_{\cdot}(X;\boldsymbol{\theta})$ in the text) estimated from a likelihood-based analysis with the half-normal detection function model, and the dashed rectangle showing the 80km buffer used in that analyis.

for integration, and Efford (2009) advises setting it such that $g(w) < 0.01$. We took an iterative approach: fitting detection function models with increasing values for w and determining when values of log-likelihood, σ and D stabilized. We found that, for the half-normal, hazard rate, and negative exponential models, $g(w) = 0.01$ corresponded to $w \approx 80, 110, \text{and} 150\text{km}$ respectively, and at these distances the log-likelihood and parameter estimates were stable to three significant figures. In our case this was good enough for model selection, because there were large differences among models; however had the models been closer, more accuracy and hence large buffer distances would have been required. Given the above buffer distances, we fit the models by maximizing the conditional likelihood (5) and selected the best model on the basis of minimum Akaike information criterion for small sample sizes (AICc). Minke whale boings are readily detectable on the bottom-mounted hydrophones, and it is highly plausible that all boings produced at zero distance are detected with certainty. Initial analyses, where g_0 was estimated, gave values of $\hat{g}_0 > 0.99$. We therefore fixed this parameter at 1.0. After obtaining maximum likelihood estimates of the remaining detection function parameters, density was estimated using the Horvitz-Thompson-like formulation described in section 2.1, with conditional (fixed-N) variance estimate.

Bayesian models were fit in OpenBUGS version 3.0.3 using the Appendix code. Since model selection is not straightforward in OpenBUGS (Deviance information criterion is not available for discrete nodes) only the half-normal detection function model was fit (the best fitting model using likelihood based methods), using a buffer of $w = 80$km and fixing $g_0 = 1$. The superpopulation size, M was chosen in a similar manner to the buffer: increasing values were tried until further changes had no affect on inference. A useful shortcut diagnostic was to check that the posterior upper 97.5th quantile for ψ was well away from its upper bound of 1. We found that a value of $M \approx 2\hat{N}$ worked well; this amounted to adding 350 artificial all-zero capture histories to the dataset (to give M=407). We set a uniform prior on σ with lower bound 0 and upper bound again large enough that it did not affect posterior estimates – for these data a value of 50 km was used (this value was obtained by trial and error, making sure the posterior results were not constrained by this choice). Starting values for the MCMC chain were set by choosing values for all quantities at random from their priors. Convergence was checked informally, by starting multiple chains from random start points, examining trace plots, and checking that resulting parameter estimates from different chains were indistinguishable except for Monte-Carlo error. Initial investigations showed that 3000 samples was a sufficient burn-in to ensure convergence to the target posterior distribution and that keeping 100,000 samples after that was sufficient for 3 significant figure accuracy in parameter estimates.

3.2 Case study results

There were 233 detections of 57 individual boing sounds in the test dataset (Table 1).

For the likelihood-based implementation, the half-normal detection function model was strongly favoured, with the next best model, the hazard rate, having a ΔAICc of > 13, and the negative exponential model trailing a distant third (Table 2). Estimated probability of detecting a sound for locations within A from the half-normal model is shown in Figure 1. The estimated detection functions (illustrated in Figure 2), and hence densities, were quite different among the three models, with the hazard rate giving a density estimate about 40% lower than the half-normal, with the negative exponential being about 50% lower again. Estimated density from the half normal model was 47.88 sounds per hour per 10,000km^2 (SE 10.60). This corresponds to 179 sounds per hour within the 80km buffer area used for that model ($A = 37,283$ km^2).

The Bayesian implementation of the half-normal model gave very similar results to the likelihood-based implementation. The posterior mean estimate of ψ was 0.46 with 95% central posterior interval of 0.28-0.69. The upper value was well away from 1.0, providing reassurance that a large enough number of artificial zero capture histories had been used.

Table 2 Results from analysis of case study data. Values in brackets after estimates are standard errors. Density ($\hat{D}$) units are sounds per hour per 10,000km^2. σ represents the scale parameter of the 3 models considered, z represents the shape parameter of the Hazard-Rate model.

Model	ΔAICc	Param. estimates	$\hat{D}$	CI($\hat{D}$)
Likelihood-based method				
Half-normal	0	$\sigma = 21.37(2.27)$	47.88 (10.60)	40.23-72.10
Hazard rate	13.11	$\sigma = 27.36(3.84)$, $z = 3.60(0.41)$	28.45 (7.48)	13.80-43.10
Negative exponential	46.71	$\sigma = 32.56(11.87)$	13.62 (8.13)	3.25-62.98
Bayesian method				
Half-normal	-	$\sigma = 21.72(2.50)$	48.46 (10.26)	30.04-70.54

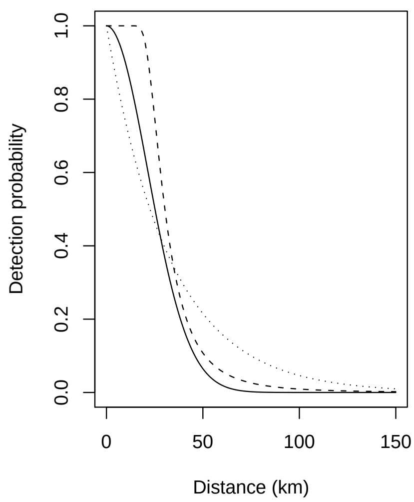

Fig. 2 Estimated half-normal (solid line), hazard rate (dashed line) and negative exponential (dotted line) detection functions fit by maximum likelihood to the case study data.

4 Simulation study

4.1 Simulation study methods

To check the performance of the different approaches under known conditions, we undertook a small simulation study. We simulated 100 replicate populations of $N = 175$ objects located at random within a rectangular study area defined by an 80km buffer around the hydrophones. This corresponds to a true density of $D = 46.94$ per 10,000km^2. Hydrophones were located at the same positions as previously, and we simulated detection of each object at each hydrophone according to a half-normal detection function with $\sigma = 20$. The simulated data were analyzed in the same way as the real data, except for the following. For the likelihood-based analysis, we fit only a half-normal detection function. We were interested in comparing variance estimates from both the unconditional likelihood, and the conditional likelihood with assumed random and fixed N, so we recorded all three. For the Bayesian analysis, we used only 10,000 samples after burn-in, to save computer time. No thinning was used. Informal checks for convergence consisted in, for some of the simulated data sets, starting the Monte Carlo chains at different points and checking they converged on similar values.

4.2 Simulation study results

The simulated data comprised a mean of 187.23 detections (SD 30.57) of 49.78 detected objects (SD 6.23), slightly lower than the numbers in the case study (because a slightly lower D and σ was used).

Both likelihood and Bayesian methods gave very similar estimates on average, with negligible bias in estimates of both σ and D (Table 3). The standard deviation

Table 3 Summary of results from simulation study. For the mean estimated SD of the density estimate ($\hat{D}$) and corresponding confidence interval (CI) coverage, the three values represent respectively the full likelihood, binomial and Poisson based estimates. σ represents the scale parameter of the Half-Normal model.

Statistic	Likelihood-based method	Bayesian method
Mean $\hat{\sigma}$ (true value 20)	20.11	20.34
SD $\hat{\sigma}$	1.84	1.89
Mean estimated SD $\hat{\sigma}$	1.95	2.00
95% CI coverage $\hat{\sigma}$	0.96	0.96
Mean $\hat{D}$ (true value 46.94)	47.28	48.00
SD $\hat{D}$	8.87	9.13
Mean estimated SD $\hat{D}$	10.00; 9.27; 9.29	9.29
95% CI coverage $\hat{D}$	0.99; 0.95; 0.94	0.95

of the estimates (i.e., the actual standard deviation of the 100 replicate estimates of σ and D for each method) was also very similar between methods. The Bayesian method also did a good job of estimating the standard deviation: the mean of the estimated standard deviation on $\hat{D}$ was 9.29, while the actual standard deviation was 9.13. There was a suggestion that the full likelihood method slightly over-estimated the standard deviation: the mean estimate was 10.00 while the true value was 8.87. This led to slightly high 95% confidence interval coverage for the full likelihood method (0.99); by contrast the coverage of the 95% posterior credibility interval was exactly at the expected 0.95. The standard deviations of the derived estimates of D from the conditional likelihood formulation were very similar, and closer to truth (9.27 for the binomial and 9.29 for the Poisson), and the corresponding confidence interval coverage was better.

5 Discussion

Although there is no ground truth with which to compare the case study estimates, the results appear reasonable given what is known about minke whale acoustic behaviour. We expect certain detection at zero horizonal distance, and that was backed up by preliminary SECR analyses. The fitted half-normal detection function parameter estimate of 21.4 (likelihood-based) or 21.7 (Bayesian) is reasonable, corresponding to a detection probability of about 0.95 at 10km, 0.5 at 25km and 0.1 at 45km (Figure 2). Calls produced at constant source level and homogeneous propagation and background noise conditions will all be detectable to a certain distance, beyond which the received level falls below the threshold set for the detector and they are no longer detectable. The resulting "step" detection function would be best fit by the hazard rate model, with large (> 5) values of the z parameter. However, variation in source levels, propagation and noise all contribute to a "rounding off" of the detection shoulder, leading to an average detection function that is closer to the half-normal form (see, e.g. Burnham et al, 2004, figure 11.2). It is possible that more flexible models, such as finite mixtures or the semiparametric families used in conventional distance sampling, may provide a better fit. Future work on the case study will focus on applying more complex models of the detection process (such as time-varying detection) to a larger sample of data; parallel field work is also in progress that, if successful, will provide an estimate of animal call rate and potentially allow estimation of animal, rather than call, density.

We assume in this work that the manual association of sounds was done without error, i.e. that no detected sounds were incorrectly associated as having the same sound source, and that all detected sounds from a given source were identified and associated. This seems a reasonable assumption given the amount of human effort put into this task. We envisage that applications of these methods in the future might be based on an automated association procedure. If one can characterize the association phase and an eventual association error process, it should be possible to include this directly in the estimation procedure. Therefore the precision in the estimates would include a component due to mis-association.

We also considered sound and hydrophone locations to exist in two-dimensional (horizontal) space, which is clearly a simplification for hydrophones located at around 4.5km depth and whales diving in the top few hundred meters of water. If the main determinant of detection probability is direct distance, rather than horizontal distance, then variation in whale and hydrophone depth represent un-modelled sources of heterogeneity. However, in our case, compared with other sources of variation (such as in source level and propagation conditions) this seems quite minor. In other cases (with deeper diving whales and shallower hydrophones, or smaller σ) it may be more important, in which case the methods could be extended to three dimensions, with additional assumptions about the depth distribution of sound sources being required.

Another simplification was that our analysis assumed a homogeneous density, even failing to account for the islands of Kauai and Niihau, both of which occur within the study area. This is acceptable given the preliminary nature of the test case analysis; however, in future work, we hope to explore the relationship between biologically relevant covariates such as depth and density. We will also need to account for the masking effect of islands on sound propagation.

Our simple simulation study showed that both likelihood and Bayesian methods yield unbiased estimates and standard deviations when their assumptions are met (or nearly met in the case of the likelihood method, which assumes random N when it was fixed in the simulations). Our estimates were based on likelihood modes in the former method and posterior means in the latter, but as with many analyses, this did not seem to generate a significant difference in estimates. Under small sample sizes, where likelihoods might be severely skewed, the posterior mode might be a better candidate than posterior mean in the Bayesian method.

Mean estimated standard deviation of D in the full likelihood method was high compared with the actual value (the standard deviation of the estimated D's), but this is understandable given that the method assumes population size in the study area is a Poisson random variable when it was actually fixed in the simulation. The two conditional likelihood variance estimators produced estimates of standard deviation that were smaller on average than the full likelihood estimates, and closer both to the actual standard deviation of the likelihood estimator and to the estimate from the Bayesian method. To understand these differences requires some discussion of the form of the conditional likelihood variance estimators. The fixed-N estimator has two components: the first is design-derived and reflects the uncertainty in D arising from sampling only a proportion a of the study area A, assuming each animal is sampled independently; the second is model-derived and reflects the additional uncertainty due to estimating the parameters of a. The random-N estimator also has two components, with the first arising from an assumption that the variance of n is equal to its mean (i.e., that animal locations are independent of one another), and the second component being the same as the second component of the fixed-N estimator. It turns out that the first component of the fixed-N estimator is just $(1 - \hat{a}/A)$ times the first component of the random-N estimator – this term can be thought of as a finite population correction factor that will cause the first component of the fixed-N estimator to go to zero when all of A is sampled. Hence, it is inevitable that the fixed-N estimator produces smaller estimates of variance than the random-N estimator, as we found. Because our simulations used a fixed N, all assumptions of the fixed-N estimator were met, and hence it was not surprising that it produced estimates of standard deviation in D that were close, on average, to the actual value. It was more surprising that the random-N estimator also performed well and this estimator deserves further investigation.

Given this initial investigation, there appears to be little difference between likelihood and Bayesian approaches. One major drawback of the Bayesian implementation in OpenBUGS is that there is no ready method of selecting among different candidate detection function models (or alternative models for spatial density distribution if a non-homogeneous distribution is assumed). In addition, the model formulation used here, with augmentation of the observed capture histories with a large number of artificial all-zero histories, results in rather long computation times (although still short when compared with the time required to collect and process the acoustic data). It is important to check that enough augmentation is used, that wide enough priors are set on detection parameters, and that the burn-in time and number of samples are sufficient to yield reliable estimates. By contrast, the likelihood-based methods are rather easier to implement, thanks to the `secr` R package. Model selection via AICc (or other criteria) is straightforward, convergence appeared reliable in the examples we used, and fitting was much faster than in OpenBUGS. For both methods one must check that an appropriately large buffer is used around the sample locations, and for the likelihood-based method one can also vary the number of grid points used in numerical integration. Our suspicion is that the Bayesian approach will make it easier to handle complex scenarios such as random effects in the detection process or mis-association of sounds – both are examples where data augmentation can potentially provide an elegant solution, enabling inference to proceed by integrating out the complicating factors with relative ease. However, for simpler applications it appears at present that the likelihood-based approach is more convenient.

We checked the performance of the methods under ideal conditions, where the model assumptions were met and sample sizes were reasonably large. It would be useful to determine how well they perform under more challenging situations, such as alternative detection models, mis-association of calls, inhomogeneity in spatial density and small sample sizes. Previous simulation studies of other varieties of SECR methods have shown them to be reasonably robust to various challenges (e.g. Efford et al, 2008).

For these methods to work, the optimal spacing between the acoustic sensors is a function of the scale of the detection process because the information about the detection process lies essentially in the "recaptures". Provided the acoustic data from multiple hydrophones can be seen as capture histories, the SECR approach becomes a natural one to estimate density. We predict in the future that statistical methods, sound processing, survey design and even hydrophone hardware might be developed and optimized with this goal in mind. The sound processing algorithms used here should be easily adapted to other scenarios. The hardware technology required to implement similar approaches still needs some development, and therefore the application of these methods outside a setting like a navy range is still not straightforward. The development of cheap and easily deployable sensors is desirable.

When a sound is detected at three or more hydrophones with appropriate geometry, then given precise information about arrival time and assumptions about sound propagation, it becomes possible to estimate the sound source location (in 2 dimensions; more detections are required for 3-d localization). Detection of echoes at a single hydrophone can also potentially be used to provide additional information about location. In the current study, we make no use of this information, but the SECR methods could be extended to utilize such information when available, potentially yielding more precise inferences. Dawson and Efford (2009) have shown how information about sound source distance that is contained in the relative received amplitude can be used to improve inference. Information on bearing from vector-sensing hydrophones could also potentially be used.

Passive acoustic methods have enormous potential to provide estimates of density and abundance in situations not readily amenable to surveys by other modalities. In many cases, however, we are limited by our knowledge of the vocal behaviour of the animals, e.g., call rates. Nevertheless, there is a great deal of research interest in the area and many ongoing studies aimed at increasing our knowledge. We anticipate that passive acoustic density estimation will be increasingly applied in future years.

Acknowledgements This research was undertaken as part of the DECAF project (Density Estimation for Cetaceans from passive Acoustic Fixed sensors), funded under the National Oceanographic Partnership Program jointly by the Joint Industry Programme and US National Marine Fisheries Service. We thank the other DECAF project members for their many contributions to this work. Murray Efford and Andy Royle provided prompt response to any issues that arose relating to implementation of the analysis. We are particularly grateful to David Borchers, who kindly stepped in to give the underlying EURING talk at short notice when TAM had to cancel his attendance at the last moment. Two anonymous reviewers and the session chairs have provided many helpful comments which improved the quality of this work.

References

Barlow J, Taylor B (2005) Estimates of sperm whale abundance in the northeastern temperate Pacific from a combined acoustic and visual survey. Mar Mamm Sci

21:429–445

Borchers D, Efford M (2008) Spatially explicit maximum likelihood methods for capture-recapture studies. Biometrics 64:377–385

Borchers DL (this volume) A non-technical overview of spatially explicit capture-recapture models. J Ornithol

Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (2001) Introduction to distance sampling - Estimating abundance of biological populations. Oxford University Press, Oxford

Burnham KP, Buckland ST, Laake JL, Borchers DL, Marques TA, Bishop JRB, Thomas L (2004) Further topics in distance sampling. In: Buckland ST, Anderson DR, Burnham KP, Laake JL, Borchers DL, Thomas L (eds) Advanced Distance Sampling, Oxford University Press, Oxford, pp 307–392

Dawson DK, Efford MG (2009) Bird population density estimated from acoustic signals. J App Ecol 46:1201–1209

Efford MG (2009) secr - Spatially Explicit Capture-Recapture in R, version 1.2.10. Department of Zoology, University of Otago, Dunedin, New Zealand

Efford MG, Borchers DL, Byrom AE (2008) Modeling Demographic Processes in Marked Populations, Springer, New York, chap Density estimation by spatially explicit capture-recapture: likelihood-based methods, pp 255–269. Environmental and Ecological Statistics

Efford MG, Dawson DK, Borchers DL (2009) Population density estimated from locations of individuals on a passive detector array. Ecology 90:2676–2682

Evans PGH, Hammond PS (2004) Monitoring cetaceans in european waters. Mamm Rev 34:131–156

Marques TA, Thomas L, Ward J, DiMarzio N, Tyack PL (2009) Estimating cetacean population density using fixed passive acoustic sensors: an example with Blainville's beaked whales. J Ac Soc Am 125:1982–1994

Mellinger DK, Stafford KM, Moore SE, Dziak RP, Matsumoto H (2007) An overview of fixed passive acoustic observation methods for cetaceans. Oceanography 20:36–45

Mobley Jr JR, Grotefendt RA, Forestell, H P, Frankel AS (1999) Results of aerial surveys of marine mammals in the major Hawaiian Islands (1993-98). Tech. rep., Final Report to the Acoustic Thermometry of Ocean Climate Program (ATOC MMRP)

R Development Core Team (2009) R: A language and environment for statistical computing URL http://www.R-project.org, ISBN 3-900051-07-0

Rankin S, Norris TF, Smultea MA, Oedekoven C, Zoidis AM, Silva E, Rivers J (2007) A visual sighting and acoustic detections of minke whales, *Balaenoptera acutorostrata* (Cetacea: *Balaenopteridae*), in nearshore Hawaiian waters. Pac Sci 61:395–398

Royle JA (2009) Analysis of capture-recapture models with individual covariates using data augmentation. Biometrics 65:267–274

Royle JA, Young KV (2008) A hierarchical model for spatial capture-recapture data. Ecology 89:2281–2289

Royle JA, Dorazio RM, Link WA (2007) Analysis of multinomial models with unknown index using data augmentation. J Comp Graph Stat 16:67–85

Royle JA, Karanth KU, Gopalaswamy AM, Kumar NS (2009a) Bayesian inference in camera trapping studies for a class of spatial capture-recapture models. Ecology 90:3233–3244

Royle JA, Nichols JD, Karanth KU, Gopalaswamy AM (2009b) A hierarchical model for estimating density in camera-trap studies. J App Ecol 46:118–127

Thomas A, OHara B, Ligges U, Sturtz S (2006) Making BUGS open. R News 6:12–17

Appendix: Example OpenBUGS code

This is the code used to run the application example in the Bayesian framework, a passive acoustic SECR with half-normal (HN) detection function. The user must input as data the following objects (object names in the code given inside brackets): (1) the number of detected animals (n), (2) the boundaries of the region over which integration takes place (Xl, Xu and Yl, Yu), (3) the upper bound on the prior for sigma ($maxSigma$), (4) the number of added all 0's capture histories required for data augmentation ($nzeroes$), (5) the traps locations ($traps$, the trap x and y coordinates need to be respectively in columns 1 and 2), (6) the area over which abundance is estimated ($Area$) and (7) the capture histories (Y, a matrix in which position i, k is 1 if animal i was detected on trap k, and 0 otherwise). The random variables involved for which priors are required are (1) the inclusion probability (psi), (2) the HN detection function parameter ($sigma$), (3) a vector of latent indicator variables associated with each of M (=n+$nzeroes$) animals (z) and (4) the M animals location (respectively x and y coordinates ($x1$ and $x2$).

The model specification is:

```
model {
    #prior for inclusion parameter
    psi~dunif(0,1)
    #prior for HN detection function;
    sigma~dunif(1,maxSigma)
    #for each sound in the augmented capture history
    for(i in 1:(n+nzeroes)){
        z[i]~dbern(psi) #inclusion indicator
```

```
        #draw a sound location
        x1[i]~dunif(Xl,Xu)
        x2[i]~dunif(Yl,Yu)
        #for each trap
        for(k in 1:K){
            #calculate distance from trap to sound location
            dist2[i,k]<-(pow(x1[i]-traps[k,1],2)+pow(x2[i]-traps[k,2],2))
            #work out prob of detection given location
            pdet[i,k]<- exp(-dist2[i,k]/(2*sigma*sigma))*z[i]
            #connect to observed detections
            Y[i,k]~dbern(pdet[i,k])}}
    #estimate sound abundance and density
    N<-sum(z[1:(n+nzeroes)])
    D<-N/Area}
```

Here is the revised version of our manuscript. We addressed the two minor points raised by the session chairs.