Running title: Animal density estimation from a fixed microphone array
Number of words: 7063
Number of tables: 3
Number of figures: 3
Number of references: 38

A general framework for animal density estimation from acoustic detections across a fixed microphone array

Ben C. Stevenson[1*] David L. Borchers[1]

Res Altwegg[2] René J. Swift[3] Douglas M. Gillespie[3]

G. John Measey[4]

1. *Centre for Research into Ecological and Environmental Modelling, University of St Andrews, The Observatory, Buchanan Gardens, St Andrews, Fife KY16 9LZ, United Kingdom*
2. *Statistics in Ecology, Environment and Conservation, Department of Statistical Sciences, and African Climate and Development Initiative, University of Cape Town, Rondebosch 7701, South Africa*
3. *Sea Mammal Research Unit, Scottish Oceans Institute, University of St Andrews, St Andrews, Fife KY16 8LB, United Kingdom*
4. *Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Stellenbosch, South Africa*

*Corresponding author.
Ben C. Stevenson
bcs5@st-andrews.ac.uk
+44 1334 461844

Abstract

1. Acoustic monitoring can be an efficient, cheap, non-invasive alternative to physical trapping of individuals. Spatially explicity capture-recapture (SECR) methods have been proposed to estimate calling animal abundance and density from data collected by a fixed array of microphones. However, these methods make some assumptions that are unlikely to hold in many situations, and the consequences of violating these are yet to be investigated.

2. We generalize existing acoustic SECR methodology, enabling these methods to be used in a much wider variety of situations. We incorporate time of arrival (TOA) data collected by the microphone array, increasing the precision of calling animal estimates. We use our method to estimate calling male density of the Cape peninsula moss frog *Arthroleptella lightfooti*.

3. Our method gives rise to an estimator of calling animal density that has negligible bias, and 95% confidence intervals with appropriate coverage. We show that using TOA information can substantially improve estimate precision.

4. Our analysis of the *A. lightfooti* data provides the first statistically rigorous estimate of calling male density for an anuran population using a microphone array. This method fills a methodological gap in the monitoring of frog populations, and is applicable to acoustic monitoring of other species that call or vocalize.

Key-words: Anura, Bootstrap, frog advertisement call, maximum likelihood, Pyxicephalidae, spatially explicit capture-recapture, time of arrival

1 Introduction

Population size is one of the most important variables in ecology and a critical factor for conservation decision making. Distance sampling and capture-recapture are both well-established methods used for the estimation of animal abundance and density. Both approaches calculate estimates of detection probabilities, and these provide information about how many animals in the survey area were undetected. Estimates of abundance and density are then straightforward to calculate. One particular point of difference is that distance sampling uses locations of detected individuals in *space*, while typically capture-recapture records the initial capture, and subsequent recaptures, of individuals at various points in *time*. The relatively recent introduction of spatially explicit capture-recapture (SECR) methods (Efford, 2004; Borchers & Efford, 2008; Royle & Young, 2008; Royle *et al.*, 2013, see Borchers, 2012, for a non-technical overview) has married the spatial component of distance sampling and the temporal nature of capture-recapture approaches. Indeed, Borchers *et al.* (in press) linked the two under a unifying model to show that they exist at opposite ends of a spectrum of methods, which vary with the amount of spatial information employed.

Data collected from SECR surveys are records (known as the capture histories) of where and when each individual was detected. Detection may occur in a variety of ways, e.g., by physical capture, or from visual recognition of a particular individual. SECR methods treat animal activity centres as unobserved latent variables, and the positions of detectors that did (and did not) detect a particular individual are informative about its location; an individual's activity centre is likely to be close to the detectors at which it was detected.

Efford *et al.* (2009) first proposed the application of SECR methods to detection data collected without physically capturing the animals themselves, but from an acoustic survey using an array of microphones (see Section 9.4, Royle *et al.*, 2013, for a summary of acoustic SECR methods). This is appealing when the species of interest is visually cryptic and difficult to trap physically, but is acoustically detectable. Moreover, it is less disruptive and invasive than physical capture. When individuals can be detected (virtually) simultaneously on multiple detectors (e.g., by virtue of the same call being recorded at multiple microphones), then "recaptures" (or, more accurately, redetections) occur at different points in space rather than across time, thus removing the need for multiple survey occasions. This has the advantage of substantially reducing the cost of fieldwork. In this case, the capture histories simply indicate which microphones detected each call, and no longer have a temporal component. The latent locations are no longer considered activity centres, but simply the physical location of the individual when the call was made. The use of SECR for these data is advantageous over competing approaches (e.g., distance sampling) as these often assume that the locations can be determined without error, and this does not hold in many cases.

The method of Efford *et al.* (2009) used signal strengths (i.e., the loudness of a recieved call at a microphone) to improve estimates of indviduals' locations: Microphones that received a stronger signal of a particular call are likely to be closer to the latent source locations than those that recevied a weaker signal. Such additional information is capable of improving the precision of parameter estimates (Borchers *et al.*, in press).

Naturally, acoustic detection methods are unable to estimate the density of non-calling individuals. Any density estimates obtained from acoustic surveys therefore correspond to the density of calling individuals, or density of calls themselves (i.e., calls per unit area per unit time), rather than overall population density. If the proportion of individuals in the population that call is known (or can be estimated) then it is straightforward to convert estimated calling animal density to population density. Otherwise, the utility of measures related to abundance or density (e.g., relative abundance indices) has been shown for a variety of taxa, of which only subsets of the populations are acoustically detectable.

For example, females do not call for almost all anuran species. It is therefore only possible to obtain an estimate of calling male density from an acoustic survey. Nevertheless, qualitative estimates of call density (i.e., density recorded on a categorical scale) for frog populations have been found to correlate well with capture-recapture estimates (Grafe & Meuche, 2005), and male chorus participation is the best known determinant of mating success in many frog species (Halliday & Tejedo, 1995). As a result, call density is often used as a proxy for frog density (e.g., Corn *et al.*, 2000; Crouch & Paton, 2002; Pellet *et al.*, 2007).

Further examples of taxa for which measures related to abundance and density have been estimated using acoustic methods include birds (e.g., Buckland, 2006; Celis-Murillo *et al.*, 2009; Dawson & Efford, 2009), cetaceans (e.g., Harris *et al.*, 2013; Martin *et al.*, 2013), insects (e.g., Fischer *et al.*, 1997), and primates (e.g., Phoonjampa *et al.*, 2011). See Marques *et al.* (2013) for an overview on the use of passive acoustics for the estimation of population density.

While the method of Efford *et al.* (2009) shows promise in estimating calling animal abundance and density using fixed arrays of acoustic detectors, a major practical issue was not addressed in this work: The method as described is only appropriate if each individual makes exactly one call. The likelihood presented assumes independent detections between calls, thus independence between call locations. This is unlikely to hold when individuals emit more than a single call, as locations of calls made by the same individual are almost certainly related. This issue was not explicitly acknowledged, and as a result the subsequent analyses presented by Marques *et al.* (2012), Martin *et al.* (2013), and Dawson & Efford (2009), which all apply the method of Efford *et al.* (2009), are problematic. We outline these studies below.

Marques *et al.* (2012) and Martin *et al.* (2013) applied acoustic SECR methods to data collected by underwater hyrdophones, which detected vocalizations from minke whales *Balaenoptera acutorostrata* Lacépède. As the location of a whale's call is likely to be close to the location of its previous call, this analysis suffers the assumption violation mentioned above. The consequences of this violation are not clear.

Furthermore, calls were treated as the unit of detection meaning that each call (rather than each individual) was given its own capture history. The resulting density estimate was therefore of call density rather than calling whale density. Distance sampling analyses have previously used independently estimated call rates to convert from call density to calling animal density (e.g., Buckland, 2006), and Efford *et al.* (2009) suggest using the same approach. The efficacy of this approach in an SECR setting is yet to be investigated, and a way of estimating variance of animal density estimates generated in this way has not yet been proposed.

Dawson & Efford (2009) estimated density of singing ovenbirds *Seiurus aurocapilla* (Linnaeus) using small arrays of microphones. Only the first detection from each individual was retained for analysis. The authors claim that this allows for the direct estimation of calling animal density, as calling individuals are now the unit of detection. There are two problems with this practice: First, it can only be carried out in situations where individuals are recognizable from their calls, and on many surveys this is not the case. Second, detection probabilities calculated using this method correspond to calls, but when calling animals are the unit of detection it is necessary to calculate the detection probabilities of *individuals* instead. The approach of Dawson & Efford (2009) ignores a temporal component of individual-level detection – the longer the survey, the more likely it is that *at least one call* from a particular individual will be detected. Individuals are detectable multiple times (i.e., every time they call) while calls are not, so call detection probabilities are necessarily smaller than individual detection probabilities. This results in the overestimation of the density of undetected individuals, causing (potentially substantial) positive bias in calling animal density estimates.

Putting the method of Efford *et al.* (2009) into practice is therefore problematic. It is necessary to investigate the consequences of violating assumptions of call location independence, and propose suitable estimators based on acoustic detection data from a microphone array. In this manuscript we present a general method that gives rise to estimators of calling animal density. We also develop methodology that can be used to estimate variance of the proposed estimators. We show by simulation that both perform well under reasonable assumptions.

An additional improvement is possible, which we also incorporate into our estimator. While Efford *et al.* (2009) suggest the use of received signal strengths to further inform call locations (in addition to detection locations), Borchers *et al.* (in press) demonstrate the utility of time of arrival information in this regard. Multichannel arrays are capable of recording the precise times at which a signal is detected by each individual microphone, and subtle differences between these times are informative about the location of the sound source. For example, a call's source location is likely to be closest to the microphone with the earliest detection time. The use of such auxiliary data informative on call locations in acoustic SECR is further motivated by Fewster & Jupp (2013), who show that incorporating response data from additional sources leads to estimators that are asymptotically more efficient. Indeed, we show via simulation that our estimator has less bias and is more precise when it incorporates time of arrival data.

We use our method to estimate calling male density of the Cape peninsula moss frog *Arthroleptella lightfooti* (Boulenger) from an acoustic survey. The genus *Arthroleptella* (moss frogs; family Pyxicephalidae) are tiny (adults are typically 7–8 mm total length), visually cryptic, and inhabit seepages on mountain tops in South Africa's Western Cape Province (Channing, 2004). Due to the region's topography, many species are severely range restricted, endemic to individual mountains, such that most of the genus are on the IUCN red list (1 Critically Endangered, 1 Vulnerable, 3 Near Threatened, and 2 Least Concern; Measey, 2011).

Individuals are extremely hard to find (approximately 3–4 person-hours per individual), and therefore prohibitively expensive to monitor via direct observation. However, males can be heard calling throughout the austral winter from within montane seepages, making an acoustic survey ideal. Movement of individuals is minimal over the course of such surveys; during physical searches frogs appear to call from the same precise locations (Measey, pers. obs.). Currently, these populations are monitored with a subjective estimate of calling male abundance (Measey *et al.*, 2011). Such subjective methods are typically employed in anuran monitoring methodologies (Dorcas *et al.*, 2009). These estimates have no corresponding measure of estimate uncertainty. Additionally, there is no formal way of accurately determining the survey area within which individuals are detected, and so estimates of calling male density are not available. Indeed, Dorcas *et al.* (2009) conclude that current auditory monitoring approaches to surveying anuran populations are restricted in their ability to estimate abundance or density. At present, no method exists that is capable of generating both point and interval estimates of either call or calling male density in a statistically rigorous manner. For the genus *Arthroleptella* (amongst others), this problem is further compounded by the lack of any method capable of identifying individuals from their calls, so it is not known how many different individuals have been detected. The method we present overcomes these problems.

2 Materials and methods

2.1 Overview

Our method has three main components:

1. An acoustic SECR survey from which call density is estimated (Section 2.3).

2. Estimation of the average call rate (Section 2.4), allowing for conversion of the call density estimate into a calling animal density estimate.

3. A parametric bootstrap procedure (Section 2.5) for variance estimation.

Once call density is estimated in Step 1, establishing an estimate for the mean call rate in Step 2 allows for the estimation of calling animal density. Measures of parameter uncertainty (such as standard errors and confidence intervals) are calculated using a parametric bootstrap approach. Parameter estimates from both Step 1 and Step 2 are required in order to carry out this procedure.

The SECR model we present for Step 1 assumes that individual calls are identifiable, i.e., it is known whether or not two detections at different microphones are of the same call. Some acoustic pre-processing is required in order to ascertain how many unique calls were detected across the array, and which of these were detected by each of the microphones. The details of this process will vary from study to study depending factors such as the acoustic properties of the focal species' calls. We describe a simple method in Section 2.6 which is suitable for our survey of *A. lightfooti*.

We do not assume that individuals are identifiable, i.e., our method does not require knowledge of whether or not two detected calls were made by the same animal. This is more difficult than identifying calls; there is less information available from which to determine individual identification, and one must contend with between-call variation in whatever acoustic properties of the calls are measured.

2.2 Notation and terminology

We consider a survey of duration T with k microphones placed at known locations within the survey region $A \subset \mathbb{R}^2$. Vocalizations from members of the focal species are detected by these microphones, and measurements of the received signal strength and time of arrival are collected for each detection. A detection is defined to be a received acoustic signal of a call that has a strength above a particular threshold, c, so that is easily identifiable above any background noise. Detections with strengths below this threshold are discarded.

The observed data comprises the number of unique calls detected, n_c, capture histories of the detected calls, $\boldsymbol{\Omega}$, recorded signal strengths, $\boldsymbol{Y}$, and times of arrival measured from some reference point (typically the beginning of the survey), $\boldsymbol{Z}$. These are defined as follows.

Let ω_{ij} be 1 if call $i \in \{1, \ldots, n_c\}$ was detected at microphone $j \in \{1, \ldots, k\}$, and 0 otherwise. We denote $\boldsymbol{\omega}_i = (\omega_{i1}, \ldots, \omega_{ik})$ as the capture history for the i^{th} call on the k detectors, and $\boldsymbol{\Omega}$ contains the capture histories for all n_c calls. If the i^{th} call was detected by the j^{th} microphone then we also observe y_{ij} and z_{ij}, the measured signal strength and the recorded time of arrival from the start of the survey, respectively. The sets of all these observations are given by $\boldsymbol{Y}$ and $\boldsymbol{Z}$, and $\boldsymbol{y}_i$ and $\boldsymbol{z}_i$ contain the signal strength and time of arrival information associated with the i^{th} call.

The detected calls have unobserved locations $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_c})$, where $\boldsymbol{x}_i \in A$ provides the Cartesian coordinates of the location at which the i^{th} call was made. We also use $\boldsymbol{x}$ generically to denote a particular location within the survey region. Note that locations of calls emitted by the same individual cannot be considered independent. As it is not known which calls were made by the same individual, call locations in general are not independent.

The parameter vector $\boldsymbol{\theta} = (D_c, \boldsymbol{\gamma}, \boldsymbol{\phi})$ is estimated from the acoustic survey data. The scalar D_c is call density (calls per unit area per unit time), which is assumed to be constant across the survey area covered by the array (although see Section 4.5 for some discussion on modelling spatial variation in calling animal and call density), while the vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$ contain parameters associated with the signal strength and time of arrival processes respectively.

The detection function and the effective sampling area (ESA) play important roles in both SECR and distance sampling, and so they are worth briefly introducing here. The detection function $g(d; \boldsymbol{\gamma})$ gives the probability that a call is detected by a microphone, given that their respective locations are separated by distance d. This is usually a monotonic decreasing function as calls further from a microphone are usually less detectable. Here we use the signal strength detection function (Efford *et al.*, 2009, further detail provided in Section 2.3.1), and this depends on the signal strength parameters $\boldsymbol{\gamma}$. Assuming independence across microphones, the probability that a call made at $\boldsymbol{x}$ is detected at all is therefore $p.(\boldsymbol{x};\boldsymbol{\gamma}) = 1 - \prod_{j=1}^{k} 1 - g(d_j(\boldsymbol{x});\boldsymbol{\gamma})$, where $d_j(\boldsymbol{x})$ is the distance between the location $\boldsymbol{x}$ and the j^{th} microphone. The ESA depends on the detection function, and is given by $a(\boldsymbol{\gamma}) = \int_A p.(\boldsymbol{x};\boldsymbol{\gamma})d\boldsymbol{x}$ (Borchers & Efford, 2008; Borchers, 2012).

The average call rate of calling members of the population at the time of the survey, μ_r, is estimated from a separate, independent sample of n_r call rates, $\boldsymbol{r} = (r_1, \cdots, r_{n_r})$. If $\boldsymbol{r}$ is used to estimate a parametric distribution for population call rates, then the vector $\boldsymbol{\psi}$ holds the associated parameters. The final parameter of interest is calling animal density, D_a.

Throughout this manuscript we do not explicitly differentiate between a random variable and its observed value, instead this should be clear from its context. Likewise, we use the function $f(\cdot)$ to generically denote any probability density function (PDF) or probability mass function (PMF) without explicit differentiation. The random variable(s) that $f(\cdot)$ is associated with should be clear from its argument(s). From Equation (2) onwards we omit the indexing of parameters in PDFs and PMFs for clarity.

2.3 Call density estimator

The estimator we propose for $\boldsymbol{\theta}$ is based on an SECR model, which we describe in this section.

The full likelihood is the joint density of the data collected from the acoustic survey, as a function of the model parameters:

$$\begin{aligned} L(\boldsymbol{\theta}) &= f(n_c, \boldsymbol{\Omega}, \boldsymbol{Y}, \boldsymbol{Z}; \boldsymbol{\theta}) \\ &= f(n_c; D_c, \boldsymbol{\gamma})\, f(\boldsymbol{\Omega}, \boldsymbol{Y}, \boldsymbol{Z}|n_c; \boldsymbol{\gamma}, \boldsymbol{\phi}). \end{aligned} \tag{1}$$

Note that D_c does not appear in the second term of Equation (1). This is a consequence of assuming that call density is constant over the survey area (Borchers & Efford, 2008).

SECR approaches often assume that the number of animals detected is a Poisson random variable, as animal locations are considered a realization of a Poisson point process. Because we do not know how many unique individuals have been detected, the distribution of the random variable n_c is not known (indeed, it is certainly not a Poisson random variable if individuals call more than once, see Appendix C). This issue is linked to the dependence of within-animal call locations; independence in call locations implies that said locations are a realization of a Poisson point process, but any dependence violates this.

We use the so-called *conditional likelihood* approach of Borchers & Efford (2008), which we extend here to include signal strength and time of arrival information. This allows for estimation of $\boldsymbol{\theta}$ without any distributional assumption on n_c, by conditioning on n_c itself. Parameters $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$ can be estimated directly using this likelihood, which is the second term in Equation (1):

$$L_n(\boldsymbol{\gamma}, \boldsymbol{\phi}) = f(\boldsymbol{\Omega}, \boldsymbol{Y}, \boldsymbol{Z} | n_c). \tag{2}$$

Once the estimate $\widehat{\boldsymbol{\gamma}}$ has been obtained, an estimate of D_c can then be calculated using a Horvitz-Thompson-like estimator. This is accomplished by dividing the number of detected calls by the estimated ESA and the survey length, i.e.,

$$\widehat{D}_c = \frac{n_c}{a(\widehat{\boldsymbol{\gamma}})T}. \tag{3}$$

Estimates for SECR model parameters that are obtained via maximization of the full likelihood are in fact equal to those obtained via maximization of the conditional likelihood and use of a Horvitz-Thomson-like estimator (Borchers & Efford, 2008), so there is no practical difference in the two approaches if we are only interested in point estimates (though note that his only holds when density is assumed constant across the survey area). Indeed, specifying the distribution for the number of detections (here denoted as n_c) only serves to allow calculation of estimate uncertainty; here $\widehat{D}_c$ depends on n_c, and so uncertainty in $\widehat{D}_c$ is subject to the variance of n_c.

Let us now describe the conditional likelihood, Equation (2), in further detail. The capture histories, $\boldsymbol{\Omega}$, received signal strengths, $\boldsymbol{Y}$, and times of arrival, $\boldsymbol{Z}$, all depend on the call locations $\boldsymbol{X}$: The closer a call is made to a microphone, the higher the probability of detection, the louder expected received signal strength, and the earlier the expected measured time of arrival. We therefore obtain the joint density of $\boldsymbol{\Omega}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$, conditional on n_c, by marginalizing over $\boldsymbol{X}$:

$$
\begin{aligned}
L_n(\boldsymbol{\gamma}, \boldsymbol{\phi}) &= \int_{A^{n_c}} f(\boldsymbol{\Omega}, \boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z} | n_c) \, d\boldsymbol{X} \\
&= \int_{A^{n_c}} f(\boldsymbol{\Omega}, \boldsymbol{Y}, \boldsymbol{Z} | \boldsymbol{X}, n_c) \, f(\boldsymbol{X} | n_c) \, d\boldsymbol{X} \\
&= \int_{A^{n_c}} f(\boldsymbol{Y}, \boldsymbol{Z} | \boldsymbol{\Omega}, \boldsymbol{X}, n_c) \, f(\boldsymbol{\Omega} | \boldsymbol{X}, n_c) \, f(\boldsymbol{X} | n_c) \, d\boldsymbol{X}.
\end{aligned}
$$

By assuming independence between the detected calls' recorded signal strengths and times of arrival, conditional on $\boldsymbol{X}$ (i.e., the time of a call's detection does not depend on its strength) we obtain

$$
L_n(\boldsymbol{\gamma}, \boldsymbol{\phi}) = \int_{A^{n_c}} f(\boldsymbol{Y} | \boldsymbol{\Omega}, \boldsymbol{X}, n_c) \, f(\boldsymbol{Z} | \boldsymbol{\Omega}, \boldsymbol{X}, n_c) \, f(\boldsymbol{\Omega} | \boldsymbol{X}, n_c) \, f(\boldsymbol{X} | n_c) \; d\boldsymbol{X}.
$$

The conditional likelihood presented above is intractable for two reasons: i) In general, the joint density of the call locations, $f(\boldsymbol{X} | n_c)$, is unknown due to the uncertain identification problem; we are unable to allocate calls to individuals (see Section 2.2), and ii) The integral is of dimension $2n_c$, usually rendering any method of its approximation too computationally expensive to be feasible.

Instead, we compute the *simplified likelihood* which overcomes these two problems by treating call locations as if they are independent. Justification for this is that treating non-independent data as if they are independent often has minimal effect on the bias of an estimator (though variance estimates may be affected substantially). This gives $f(\boldsymbol{X}|n_c) = \prod_{i=1}^{n_c} f(\boldsymbol{x}_i)$, and results in a separable integral, allowing for the evaluation of a product of n_c 2-dimensional integrals instead of a single $2n_c$-dimensional integral:

$$L_s(\boldsymbol{\gamma}, \boldsymbol{\phi}) = \prod_{i=1}^{n_c} \int_A f(\boldsymbol{y}_i|\boldsymbol{\omega}_i, \boldsymbol{x}_i)\, f(\boldsymbol{z}_i|\boldsymbol{\omega}_i, \boldsymbol{x}_i)\, f(\boldsymbol{\omega}_i|\boldsymbol{x}_i)\, f(\boldsymbol{x}_i)\, d\boldsymbol{x}_i. \tag{4}$$

Estimates for $\boldsymbol{\gamma}$ and $\boldsymbol{\phi}$ are found by maximising the log of the simplified likelihood function, i.e.,

$$(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\phi}}) = \arg\max_{\boldsymbol{\gamma}, \boldsymbol{\phi}} \log\left(L_s(\boldsymbol{\gamma}, \boldsymbol{\phi})\right), \tag{5}$$

and our estimator for D_c remains as shown in Equation (3).

In situations where call locations *can* be considered independent, the conditional and simplified likelihoods are equivalent. Otherwise, the simplified likelihood is not a true likelihood per se, and should not be treated as such. That is, any further likelihood-based inference (such as the calculation of standard errors based on the curvature of the log-likelihood surface at the maximum likelihood estimate, or likelihood-based information criteria) should not be directly used.

The following sections focus on providing further details about each term that appears in the integrand of Equation (4).

2.3.1 Signal strength

The use of signal strength to improve estimator precision in SECR models was first proposed by Efford *et al.* (2009).

Assuming independence between received signal strengths (see Section 4.4 for discussion on this point), the first component of the integrand in Equation (4) is

$$f(\boldsymbol{y}_i|\boldsymbol{\omega}_i, \boldsymbol{x}_i) = \prod_{j=1}^{k} f(y_{ij}|\omega_{ij}, \boldsymbol{x}_i).$$

The expected received signal strength of the i^{th} call at the j^{th} microphone can be any sensible monotonic decreasing function of $d_j(\boldsymbol{x_i})$, the distance between the j^{th} microphone and the location of the i^{th} call. Here we simply use

$$E(y_{ij}|\boldsymbol{x}_i) = h^{-1}(\beta_{0s} - \beta_{1s} d_j(\boldsymbol{x}_i)),$$

where $h^{-1}(\cdot)$ is the inverse of a link function (typically chosen to be either the identity or log function). We account for Gaussian measurement error in the received signal strengths, i.e.,

$$y_{ij}|\boldsymbol{x}_i \sim N(E(y_{ij}|\boldsymbol{x}_i), \sigma_s).$$

The parameter vector $\boldsymbol{\gamma}$ therefore comprises β_{0s}, β_{1s}, and σ_s which have direct signal strength interpretations: β_{0s} is the source signal strength of calls (on the link function's scale), β_{1s} is the loss of strength per metre travelled due to signal propagation (on the link function's scale), and σ_s is the standard deviation of the normal distribution used to account for signal measurement error.

However, recall that y_{ij} is only observed if the received signal strength exceeds the microphone threshold of detection, i.e., if and only if $y_{ij} > c$ (or, equivalently, $\omega_{ij} = 1$). Otherwise, y_{ij} is discarded and ω_{ij} is set to 0. Therefore, we set $f(y_{ij}|\omega_{ij} = 0, \boldsymbol{x}_i)$ to 1, and $(y_{ij}|\omega_{ij} = 1, \mathbf{x}_i)$ is a random variable from a truncated normal distribution, giving

$$f(y_{ij}|\omega_{ij} = 1, \boldsymbol{x}_i) = \frac{1}{\sigma_s} f_n \left(\frac{y_{ij} - E(y_{ij}|\boldsymbol{x}_i)}{\sigma_s} \right) \left(1 - \Phi \left(\frac{c - E(y_{ij}|\boldsymbol{x}_i)}{\sigma_s} \right) \right)^{-1}, \tag{6}$$

where $f_n(\cdot)$ and $\Phi(\cdot)$ are the PDF and the cumulative density function of the standard normal distribution, respectively.

2.3.2 Probability of detection

Based on the previous section, Efford *et al.* (2009) proposed the *signal strength detection function*, to be used when signal strength information has been collected by the detectors during an SECR survey. This takes the form

$$g(d; \boldsymbol{\gamma}) = 1 - \Phi \left(\frac{c - h^{-1}(\beta_{0s} - \beta_{1s} d)}{\sigma_s} \right),$$

thus giving the probability of a call's received signal strength exceeding c (and, therefore, the probability of detection).

The i^{th} capture history, $\boldsymbol{\omega}_i$, is only observed if the i^{th} call is detected, i.e., if $\sum_{j=1}^{k} \omega_{ij} > 0$. Thus, we observe $\boldsymbol{\omega}_i$ conditional on detection, and so $f(\boldsymbol{\omega}_i|\boldsymbol{x}_i)$ must incorporate the probability of detection in the denominator. Assuming independent detections of each call across all microphones, the third component of the integrand in Equation 4 is therefore

$$f(\boldsymbol{\omega_i}|\boldsymbol{x}_i) = \frac{\prod_{j=1}^{k} f(\omega_{ij}|\boldsymbol{x}_i)}{p.(\boldsymbol{x}_i; \boldsymbol{\gamma})}.$$

As ω_{ij} is 1 if the i^{th} call is detected by the j^{th} microphone, and 0 otherwise, we have

$$f(\omega_{ij}|\boldsymbol{x}_i) = \begin{cases} g(d_j(\boldsymbol{x}_i);\boldsymbol{\gamma}) & \omega_{ij} = 1, \\ 1 - g(d_j(\boldsymbol{x}_i);\boldsymbol{\gamma}) & \omega_{ij} = 0. \end{cases} \tag{7}$$

2.3.3 Time of arrival

A single detection time on its own is not informative on call location. It is only *differences* between precise arrival times that provide information about the relative position of a call in relation to the locations of the microphones at which it was detected. Time of arrival data are therefore only informative for calls detected at two or more microphones; the arrival times, $\boldsymbol{z}_i$, depend on $\boldsymbol{\omega}_i$ through m_i, the number of microphones that detected the i^{th} call, i.e., $m_i = \sum_{j=1}^{k} \omega_{ij}$, $m_i \in \{1, \cdots, k\}$. Therefore $f(\boldsymbol{z}_i|\boldsymbol{\omega}_i, \boldsymbol{x}_i) \equiv f(\boldsymbol{z}_i|m_i, \boldsymbol{x}_i)$, and we set $f(\boldsymbol{z}_i|m_i = 1, \boldsymbol{x}_i)$ to 1.

Information about call locations improves the precision of parameter estimates, though here we do not assume that times of arrival allow perfect triangulation of call locations. Instead, we account for uncertainty in recorded times of arrival due to Gaussian measurement error, controlled by the parameter σ_t. For calls detected at two or more microphones, inference can be made by marginalizing over the time the call was made, a latent variable, and this integral is available in closed form (see the online supplementary material of Borchers *et al.*, in press),

$$f(\boldsymbol{z}_i|m_i > 1, \boldsymbol{x}_i) = \frac{(2\pi\sigma_t^2)^{(1-m_i)/2}}{2T\sqrt{m_i}} \exp\left(\sum_{\{j:\omega_{ij}=1\}} \frac{(\delta_{ij}(\boldsymbol{x}_i) - \bar{\delta}_i)^2}{-2\sigma_t^2} \right). \tag{8}$$

The term $\delta_{ij}(\boldsymbol{x}_i)$ is the expected call production time, given call location $\boldsymbol{x}_i$, and the time of arrival collected by detector j, i.e., $\delta_{ij}(\boldsymbol{x}_i) = z_{ij} - d_j(\boldsymbol{x}_i)/v$, where v is the speed of sound. The average across all detectors on which a detection was made is $\bar{\delta}_i$.

2.3.4 Call locations

We assume individuals' locations are a realization of a homogeneous Poisson point process across the survey area, A. As the dependence between call locations is not clear, it is not possible to specify their joint density, $f(\boldsymbol{X})$, from data collected by the acoustic survey alone. Under the simplified likelihood (Equation (4)) this is now tractable: $\boldsymbol{X}$ itself is a realization of a filtered homogeneous Poisson point process – the intensity of *emitted* calls is constant across the survey area, but the intensity of *detected* calls is highest closest to the microphones. The filtering is therefore through the detection probability surface (see Section 2.2). We now have $f(\boldsymbol{X}) = \prod_{i=1}^{n_c} f(\boldsymbol{x}_i)$, and $f(\boldsymbol{x}_i)$ is proportional to the intensity of the point process, i.e., $f(\boldsymbol{x}_i) \propto p.(\boldsymbol{x}_i; \boldsymbol{\gamma})$. As $a(\boldsymbol{\gamma}) = \int_A p.(\boldsymbol{x}; \boldsymbol{\gamma}) d\boldsymbol{x}$, the ESA is the normalizing constant, and we obtain

$$f(\boldsymbol{x}_i) = \frac{p.(\boldsymbol{x}_i; \boldsymbol{\gamma})}{a(\boldsymbol{\gamma})}.$$

We have now provided details for all terms in the integrand of the simplified likelihood, Equation (4).

2.4 Calling animal density estimator

Although call density, D_c may be informative in situations where a species' call rate is of primary interest, it is usually the density of calling individuals per unit area, D_a that is required.

First used in distance sampling by Hiby (1985), a common method used to obtain an estimate for calling animal density from call density involves dividing call density by the average call rate across the calling population, i.e., $\widehat{D}_a = \widehat{D}_c / \widehat{\mu}_r$ (see Buckland *et al.*, 2001, pp. 191–197). See Appendix B for justification for this estimator from its asymptotic properties.

If μ_r is not known *a priori* then it must be estimated separately from call rate data, $\boldsymbol{r}$, collected independently of the acoustic survey. In the simplest case, the sample mean $\bar{r} = n_r^{-1} \sum_{i=1}^{n_r} r_i$ is an estimator for μ_r. If the average call rate is known to vary (e.g., perhaps due to covariates such as rainfall, season, or temperature) then it is important to observe $\boldsymbol{r}$ at the same time as the acoustic survey. Alternatively, given call rate data collected across a range of such covariates, a model could be fitted to estimate the average call rate for specific conditions of a future survey, thereby reducing future field effort.

In any case, for calculation of variance estimates (Section 2.5) one has to simulate call rate data from whatever model is used to estimate μ_r. In the case of taking a simple random sample of n_r call rates, this can be done using the empirical distribution function (EDF). Otherwise, if a parametric model has been fitted to $\boldsymbol{r}$ (potentially using covariates, as described above), then such data can be generated from $f(\boldsymbol{r}; \widehat{\boldsymbol{\psi}})$.

2.5 The bootstrap procedure

We calculate estimate uncertainty (i.e., standard errors and confidence intervals for the model parameters) using a parametric bootstrap. By combining parameter estimates from Sections 2.3 and 2.4, we can simulate data in a way that mimics the real data generation process, including dependencies in call locations.

Here we use the superscript * to denote simulated data, or parameters estimated from simulated data. We propose the following algorithm:

1. Simulate animal locations as a realization of a homogeneous Poisson point process with intensity $\widehat{D}_a$.

2. Determine the number of calls made by each individual by simulating call rates from either the EDF of $\boldsymbol{r}$ or $f(\boldsymbol{r}; \widehat{\boldsymbol{\psi}})$.

3. Generate $\boldsymbol{X}^*$ by repeating each location from Step 1 the appropriate number of times, given by Step 2.

4. Obtain $\boldsymbol{\Omega^*}$ by simulating from $f(\omega_{ij}|\boldsymbol{x}_i^*;\widehat{\boldsymbol{\gamma}})$ (Equation (7)). Omit all rows from $\boldsymbol{\Omega}^*$ and $\boldsymbol{X}^*$ that are associated with undetected calls.

5. Obtain $\boldsymbol{Y}^*$ by simulating from $f(y_{ij}|\omega_{ij}^* = 1, \boldsymbol{x}_i^*;\widehat{\boldsymbol{\gamma}})$ (Equation (6)), and $\boldsymbol{Z}^*$ by simulating from $f(\boldsymbol{z}_i|\boldsymbol{\omega}_i^*, \boldsymbol{x}_i^*;\widehat{\boldsymbol{\phi}})$ (Equation (8)) for all detections.

6. Calculate $\widehat{\boldsymbol{\theta}}^*$ from $\boldsymbol{\Omega}^*$, $\boldsymbol{Y}^*$, and $\boldsymbol{Z}^*$ using Equations (3) and (5).

7. Obtain $\boldsymbol{r}^*$ by simulating from either its EDF or $f(\boldsymbol{r};\widehat{\boldsymbol{\psi}})$, calculate $\widehat{\boldsymbol{\psi}}^*$ and therefore $\widehat{\mu}_r^*$.

8. Calculate $\widehat{D}_a^* = \widehat{D}_c^*/\widehat{\mu}_r^*$.

9. Repeat the above steps R times and save the parameter estimates from each iteration.

Here we treat D_a as the sole parameter of interest, but in practice the following holds for any other estimated parameter. Let the saved density estimates from the simulated data be $\widehat{\boldsymbol{D}}_a^* = (\widehat{D}_{a1}^*, \widehat{D}_{a2}^*, ..., \widehat{D}_{aR}^*)$. Bias can be estimated by subtracting the parameter estimate from the mean of the estimates from the bootstrap samples (Davison & Hinkley, 1997), i.e., $\bar{D}_a^* - \widehat{D}_a$, where $\bar{D}_a^* = R^{-1}\sum_{i=1}^{R} D_i^*$.

Confidence intervals can be calculated using any suitable bootstrap confidence interval method, many of which are outlined by Davison & Hinkley (1997). The simplest approach is to calculate confidence intervals based on a normal approximation, using $\mathrm{SD}(\widehat{\boldsymbol{D}}_a^*)$ as the standard error. Note that the normal approximation may be more suitable for a transformation of $\widehat{D}_a$ (e.g., $\log(\widehat{D}_a)$), and so a back-transformation of a confidence interval based on this transformed parameter may have better coverage properties. Other possible approaches include the so-called *basic* and *percentile* methods, although note that the latter requires R to be larger in comparison to the normal approximation and basic methods.

Note that Step 5 above makes the assumption that individuals do not move over the course of the survey. See Section 4.2 for some discussion on accounting for animal movement.

2.6 Application to *Arthroleptella lightfooti*

We use the method presented above to generate estimates of call and calling male density of *A. lightfooti*, and estimate associated variances.

2.6.1 Equipment and survey design

The data we use were generated from a 25 s subset of a recording carried out on 16 May, 2012.

The recording was made using an array of six Audio-Technica AT8004 handheld omni-directional dynamic microphones, connected to a DR-680 8-Track portable field audio recorder via Hosa Technology STX-350F Professional 1/4 inch TRS male to XLR female cables. Each of the six microphones were placed in microphone holders which were fastened atop 1 m tall wooden dowels. The immediate vicinity was vacated during the recording. The configuration of our array is shown in Figure 1.

2.6.2 Acoustic pre-processing

The open-source software package PAMGUARD (Gillespie *et al.*, 2009, see www.pamguard.org) was used in order to identify calls of *A. lightfooti* males, which have a signature frequency of 3.8 KHz. The first 600 s of the recording were ignored in case any disturbance to the frogs during set-up affected calling behavior. Furthermore, a detection was only recorded if the strength of the received signal was above a threshold of 130 units. Along with signal strengths, precise times of signal arrival (accurate to 2.083×10^{-5} s) were also recorded for each detection.

In order to construct the observed $\boldsymbol{\Omega}$, $\boldsymbol{Y}$, and $\boldsymbol{Z}$, it was necessary to determine which detected sounds on different microphones were of the same call from the same frog. As individuals are not recognizable from their calls, this was done as follows: If two calls were detected within $d/330$ seconds of one another by two microphones that were d meters apart, then they are assumed to have the same source (using 330 ms^{-1} as the speed of sound in air).

Note that this approach to call identification will never result in detections of the same call being attributed to different frogs, however there is potential for calls from different frogs to be falsely identified as the same individual. This is unlikely, however, as calls from males are temporally negatively correlated; they tend to call in turn in an attempt to increase their likelihood of being heard by a female (Altwegg & Measey, pers. obs.).

2.6.3 Bootstrap details

No individual call rate data were collected concurrently with the acoustic survey. Instead, we use call rate data collected at another time and location so that we are able to demonstrate the methods described in Sections 2.4 and 2.5. Call rate data were obtained by finding locations of 8 individual calling males and placing a microphone in close proximity; this ensured that all calls they emitted were detected, and were easily identifiable from calls of other males.

We ran the bootstrap procedure (Section 2.5) for 10 000 iterations in order to reduce the relative Monte Carlo error associated with the standard error (calculated using Equation (9) in Koehler *et al.*, 2009) to below 1%.

2.7 Simulation study

We test our method using a simulation study. A total of 1 000 datasets were independently simulated using Steps 1–5 and Step 7 in Section 2.5. Values used for the simulation parameters were set at the corresponding estimates obtained from the real data analysis. For each simulated dataset, we use the method we outline above to obtain both point estimates and confidence intervals for D_a and D_c. We used 500 bootstrap repetitions for each iteration in order to prevent the simulation from being prohibitively time-consuming. For comparison, we also calculate confidence intervals based on the approach of Efford *et al.* (2009), which ignores the dependence between call locations.

We also conduct a simulation study to investigate the impact of using time of arrival information in addition to the signal strength data. A total of 10 000 datasets were independently simulated, the same way as above, and two estimates of both D_a and D_c were obtained from each: One from a model that used time of arrival information, and another from a model that did not.

2.8 Software implementation

Implementation (in Section 3, below) of the methods we present was accomplished using the R package admbsecr (Stevenson & Borchers, 2014, see https://github.com/b-steve/admbsecr). This software can be used to obtain parameter estimates via numerical maximization of the log of the simplified likelihood. Optimization is carried out by a call to an executable generated by AD Model Builder (Fournier *et al.*, 2012). Numerical integration is used to approximate marginalization over call locations.

The code used to carry out analysis of the *A. lightfooti* data can be found in Appendix A.

3 Results

3.1 Real data analysis

A total of 225 unique calls were detected by the six microphones over the course of the 25 s survey.

Density parameter estimates, their associated standard errors, and estimated biases (obtained from the bootstrap procedure of Section 2.5) are provided in Table 1. We use $\widehat{\gamma}$ to plot the detection function, shown in Figure 2. To illustrate the utility of the time of arrival information, we plot uncertainty surrounding the estimation of a location of one of the detected calls in Figure 1.

Normal QQ plots for $\widehat{\boldsymbol{D}}_a^*$ and $\widehat{\boldsymbol{D}}_c^*$ both indicated approximate normality, and so confidence intervals based on a normal approximation using the standard errors shown in Table 1 were deemed to be appropriate. Setting the nominal coverage at 95%, this approach gave an interval of (239.42, 492.75) for D_a and an interval of (65.06, 133.23) for D_c; D_a is calling males per hectare and D_c is calls per hectare per second.

3.2 Simulation study

We show the performance of a number of confidence interval calculation methods in Table 2. Coverage is only significantly different (at the 5% level) to the nominal 95% coverage rate for both intervals calculated using the basic bootstrap method, and for naïve confidence intervals that rely on call locations being independent (as per the method of Efford *et al.*, 2009).

Estimates of bias, variance, and mean square error of the estimators investigated in the second simulation study are shown in Table 3. The estimator that utilizes time of arrival data is more precise and less biased. Estimated sampling distributions of the estimates obtained both with and without the time of arrival information are shown in Figure 3.

4 Discussion

4.1 Summary

The method we have proposed to estimate calling animal density from a fixed microphone array relies on maximizing a simplified likelihood (Equation (4)). We then use a parametric bootstrap (Section 2.5) to account for dependence between call locations.

In our simulation studies, parameter estimates were shown to have negligible bias (in all cases, bias was estimated at substantially less than 1% of the estimate sizes; see Tables 1 and 3). Note that this is despite the simplified likelihood treating call locations as independent. Our findings suggest that density estimates obtained via acoustic SECR methods are robust to this violation. The bootstrap confidence interval methods generated intervals with coverage close to their nominal level (Table 2). Indeed, these easily outperformed the method that does not account for dependence among call locations in the construction of confidence intervals.

Using time of arrival information led to decreased bias and substantially increased precision in density estimates (Figure 3, Table 3) in comparison to the approach of Efford *et al.* (2009). In applications like ours, time of arrival data is far more informative on animal location than trap location and signal strength information (Figure 1). With more information on where calls are located the detection function parameters can be estimated more precisely. In turn, this results in higher precision estimates of the ESA, call density, and calling animal density.

4.2 Animal movement

The approach we present here assumes that calls made by the same individual are associated with the same location, which is a reasonable assumption for our case study of *A. lightfooti*. A natural extension is to account for animal movement. We outline two ways of doing this here.

The first is to adjust our bootstrap method. This requires the fitting of a movement model (e.g., Jonsen *et al.*, 2005; McClintock *et al.*, 2012, see King, 2014, for an overview) to independently collected data, explaining between-call animal movement patterns. Rather than the bootstrap procedure allocating all calls to the same location, movement can be introduced using parameter estimates from the movement model, resulting in appropriate variance estimates. However, we recognize that this may represent an infeasible amount of field effort in addition to the acoustic survey.

If individuals can be identified from their calls, then the analysis of Ergon & Gardner (in press) suggests an alternative way forward. A new SECR approach was used to analyze live-trapping data of field voles *Microtus agrestis* (Linnaeus), where individuals' home range centres moved (due to a dispersion model) from one survey session to the next. Similar approaches could possibly be used to account for animal movement in acoustic SECR surveys. There are complications, however, associated with detections in continuous time rather than allowing movement across discrete sessions: One must integrate over all possible paths an individual could have taken between detection occasions, considerably increasing computational complexity.

4.3 Inference via the conditional likelihood

It would be beneficial to propose estimators based on the maximization of the conditional likelihood (Equation (2)) rather than the simplified likelihood (Equation (4)). Such an approach would deal directly with call location dependence, removing the need to collect data or make restrictive assumptions about call rates and animal movement. Under a classical framework, this would also result in maximization of a true likelihood, allowing for use of further likelihood-based inference.

It is not clear how this could be achieved when animal identification is not possible; a solution to this so-called *unknown identification problem* would present a significant breakthrough. One possible approach is to use a reversible jump Markov chain Monte Carlo procedure under a Bayesian framework. The number of unique detected individuals, as well as the allocations of calls to individuals, would vary from iteration to iteration. Alternatively, inference could potentially be made using methods that deal with the estimation of parameters from intractable likelihoods (e.g., the synthetic likelihood approach of Wood, 2010).

Otherwise, if animal identities *can* be determined, possible methods of incorporating animal movement and call rate into the conditional likelihood are a little clearer. The dependence between latent locations of calls from the same individual is obvious under the assumption of no animal movement, and potentially estimable via a movement model otherwise.

Direct estimation of the average call rate, μ_r (and therefore calling animal density) is also likely to be possible from the acoustic survey. In order to obtain this, one must specify a distribution with mean μ_r for the number of calls made by individuals to account for the call production process. This is then filtered by the detection process, giving rise to the observed data and call identities.

4.4 Further generalizations

Our method is more general than that of Efford *et al.* (2009), as we do not rely on assumptions regarding independence of call locations for variance estimation. Further generalizations are possible, and we outline two of them here. First, our method assumes that individuals all emit calls with the same strength, β_{0s}, which may not hold. Second, there is the issue of directional calling: The orientation of an individual may result in the loss of strength per metre, β_{1s}, due to signal propagation at a lower rate in some directions. Our method assumes signal propagation occurs uniformly across all directions.

It is likely that further latent variables will be required to fit models appropriate for either case, i.e., latent call source strengths or latent individual orientations, respectively. With additional latent variables comes further computational complexity: Under a classical framework these must be integrated out of the likelihood. A Bayesian approach presents a viable alternative; latent variables can be sampled from rather than marginalized over, which is potentially simpler.

4.5 Spatiotemporal changes in density

In some situations it is not necessarily animal density that is of particular ecological interest, but rather temporal or spatial variation in density. Our method can be used to make inference in either case. Independent microphone arrays set out at various points in time and space will generate separate density estimates, from which temporal and spatial shifts of animal abundance can be determined.

There is also potential for an alternative: In general, SECR methods are capable of directly estimating a density intensity *surface*, rather than a constant intensity over the survey area. We have skirted this issue for brevity; assuming a constant density is reasonable in many cases over small survey areas.

4.6 Analysis of *Arthroleptella lightfooti* data

Regarding the survey of *A. lightfooti*, our method obtained an estimate of 366.08 calling males per hectare. Alternative methods used to monitor abundance of threatened species in the genus *Arthroleptella* make use of auditory estimates (Measey *et al.*, 2011). Trained practitioners stand at a set locale and listen to an assemblage, placing call abundance into a category (Dorcas *et al.*, 2009); the assemblage calling in this study was assessed using this method, falling into the highest category, > 100 individuals. It is difficult to compare the two estimates as this abundance cannot be converted into a density.

Our estimates of call density and calling male density are associated with coefficients of variation of approximately 17.5% from just 25 s worth of recording using only six microphones (Table 1). The relatively high precision of $\widehat{D}_a$ is in part due to the fact that variance in the recorded call rates, $\boldsymbol{r}$, was very low as individual *A. lightfooti* call at very regular intervals. This allowed for a precise estimate of μ_r which was used in the calculation of $\widehat{D}_a$. Uncertainties associated with our density estimators decrease as survey length and n_r increase (see Appendix B, Figure 1).

4.7 Concluding remarks

Our method advances acoustic SECR methodology by improving estimator precision via time of arrival information, and by proposing an unbiased estimator for calling animal density. Our confidence intervals account for dependence in call locations, which had previously been ignored. Our analysis here is the first to provide reliable point and interval estimates of both the call and calling male density of a frog species from an acoustic survey. This approach is general, and can be applied to estimate calling animal density for a wide variety of species.

Acknowledgements

We thank both Tanya Scott and Paula Strauss for their assistance in the field.

Data collection was made possible due to access facilitated by CapeNature (Permit No. AAA007-00084-0035) and South African National Parks (Table Mountain National Park).

Funding for the frog survey was received from the National Geographic Society/Waitt Grants Program (No. W184-11). The EPSRC and NERC helped to fund this research through a PhD grant (No. EP/1000917/1).

B.C.S. wishes to acknowledge the Auckland Grammar School board of trustees for further financial assistance through the F.W.W. Rhodes Scholarship.

G.J.M. would like to thank the South African National Biodiversity Institute and the DST-NRF Centre of Excellence for Invasion Biology.

We are grateful to two anonymous reviewers, whose comments lead to substantial improvements to the content and structure of this manuscript.

References

Borchers, D.L. (2012) A non-technical overview of spatially explicit capture-recapture models. *Journal of Ornithology*, **152**, 435–444.

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

Borchers, D.L., Stevenson, B.C., Kidney, D., Thomas, L. & Marques, T.A. (in press) A unifying model for capture-recapture and distance sampling surveys of wildlife populations. *Journal of the American Statistical Association*.

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

Buckland, S.T., Anderson, D.R., Burnham, K.P., Laake, J.L., Borchers, D.L. & Thomas, L. (2001) *Introduction to Distance Sampling: Estimating Abundance of Biological Populations*. Oxford University Press, Oxford.

Celis-Murillo, A., Deppe, J.L. & Allen, M.F. (2009) Using soundscape recordings to estimate bird species abundance, richness, and composition. *Journal of Field Ornithology*, **80**, 64–78.

Channing, A. (2004) Genus *Arthroleptella*. L.R. Minter, M. Burger, J.A. Harrison, H.H. Braack, P.J. Bishop & D. Kloepfer, eds., *Atlas and Red Data Book of the Frogs of South Africa, Lesotho and Swaziland*, pp. 206–219. Smithsonian Institute, Washington D.C.

Corn, P.S., Muths, E. & Iko, W.M. (2000) A comparison in Colorado of three methods to monitor breeding amphibians. *Northwestern Naturalist*, **81**, 22–30.

Crouch, W.B. & Paton, P.W.C. (2002) Assessing the use of call surveys to monitor breeding anurans in Rhode Island. *Journal of Herpetology*, **36**, 185–192.

Davison, A.C. & Hinkley, D.V. (1997) *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge.

Dawson, D.K. & Efford, M.G. (2009) Bird population density estimated from acoustic signals. *Journal of Applied Ecology*, **46**, 1201–1209.

Dorcas, M.E., Price, S.J., Walls, S.C. & Barichivich, W.J. (2009) Auditory monitoring of anuran populations. C.K. Dodd, ed., *Amphibian Ecology and Conservation: A Handbook of Techniques*, pp. 281–298. Oxford University Press, New York.

Efford, M.G. (2004) Density estimation in live-trapping studies. *Oikos*, **106**, 598–610.

Efford, M.G., Dawson, D.K. & Borchers, D.L. (2009) Population density estimated from locations of individuals on a passive detector array. *Ecology*, **90**, 2676–2682.

Ergon, T. & Gardner, B. (in press) Separating mortality and emigration: Modelling space use, dispersal and survival with robust-design spatial capture-recapture data. *Methods in Ecology and Evolution*.

Fewster, R.M. & Jupp, P.E. (2013) Information on parameters of interest decreases under transformations. *Journal of Multivariate Analysis*, **120**, 34–39.

Fischer, F.P., Schulz, U., Schubert, H., Knapp, P. & Schmoger, M. (1997) Quantitative assessment of grassland quality: Acoustic determination of population sizes of orthopteran indicator species. *Ecological Applications*, **7**, 909–920.

Fournier, D.A., Skaug, H.J., Ancheta, J., Ianelli, J., Magnusson, A., Maunder, M.N., Nielsen, A. & Sibert, J. (2012) AD Model Builder: Using automatic differentiation for statistical inference of highly parameterized complex nonlinear models. *Optimization Methods and Software*, **27**, 233–249.

Gillespie, D., Mellinger, D.K., Gordon, J., McLaren, D., Redmond, P., McHugh, R., Trinder, P., Deng, X. & Thode, A. (2009) PAMGUARD: Semiautomated, open source software for real-time acoustic detection and localization of cetaceans. *Journal of the Acoustical Society of America*, **125**, 2547–2547.

Grafe, T.U. & Meuche, I. (2005) Chorus tenure and estimates of population size of male European tree frogs *Hyla arborea*: Implications for conservation. *Amphibia-Reptilia*, **26**, 437–444.

Halliday, T.R. & Tejedo, M. (1995) Intrasexual selection and alternative mating behaviour. H. Heatwole & B.K. Sullivan, eds., *Amphibian Biology. Vol. II: Social Behaviour*, pp. 419–468. Surrey Beatty, Chipping Norton.

Harris, D., Matias, L., Marques, T., Thomas, L., Harwood, J. & Geissler, W.H. (2013) Applying distance sampling to fin whale calls recorded by single seismic instruments in the northeast Atlantic. *Journal of the Acoustical Society of America*, **134**, 3522.

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

Jonsen, I.D., Flemming, J.M. & Myers, R.A. (2005) Robust state-space modeling of animal movement data. *Ecology*, **86**, 2874–2880.

King, R. (2014) Statistical ecology. *Annual Review of Statistics and its Application*, **1**, 401–426.

Koehler, E., Brown, E. & Haneuse, S.J.P.A. (2009) On the assessment of monte carlo error in simulation-based statistical analyses. *The American Statistician*, **63**, 155–162.

Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Jarvis, S., Morrissey, R.P., Ciminello, C. & DiMarzio, N. (2012) Spatially explicit capture-recapture methods to estimate minke whale density from data collected at bottom-mounted hydrophones. *Journal of Ornithology*, **152**, 445–455.

Marques, T.A., Thomas, L., Martin, S.W., Mellinger, D.K., Ward, J.A., Morettie, D.J., Harris, D. & Tyack, P.L. (2013) Estimating animal population density using passive acoustics. *Biological Reviews*, **88**, 287–309.

Martin, S.W., Marques, T.A., Thomas, L., Morrissey, R.P., Jarvis, S., DiMarzio, N., Moretti, D. & Mellinger, D.K. (2013) Estimating minke whale (*Balaenoptera acutorostrata*) boing sound density using passive acoustic sensors. *Marine Mammal Science*, **29**, 142–158.

McClintock, B.T., King, R., Thomas, L., Matthiopoulos, J., McConnell, B.J. & Morales, J.M. (2012) A general discrete-time modeling framework for animal movement using multistate random walks. *Ecological Monographs*, **82**, 335–349.

Measey, G.J. (2011) Appendix 1. G.J. Measey, ed., *Ensuring a Future for South Africa's frogs: A Strategy for Conservation Research.* SANBI Biodiversity Series 19, pp. 42–83. South African National Biodiversity Institute, Pretoria.

Measey, G.J., Weldon, C., Morgan, D., Channing, A., Harvey, J. & Turner, A. (2011) Conservation and ecological studies. G.J. Measey, ed., *Ensuring a Future for South Africa's frogs: A Strategy for Conservation Research.* SANBI Biodiversity Series 19, pp. 18–28. South African National Biodiversity Institute, Pretoria.

Pellet, J., Helfer, V. & Yannic, G. (2007) Estimating population size in the European tree frog (*Hyla arborea*) using individual recognition and chorus counts. *Amphibia-Reptilia*, **28**, 287–294.

Phoonjampa, R., Koenig, A., Brockelman, W.Y., Borries, C., Gale, G.A., Carroll, J.P. & Savini, T. (2011) Pileated gibbon density in relation to habitat characteristics in post-logging forest recovery. *Biotropica*, **43**, 619–627.

Royle, J.A., Chandler, R.B., Sollmann, R. & Gardner, B. (2013) *Spatial Capture-Recapture.* Academic Press, Amsterdam.

Royle, J.A. & Young, K.V. (2008) A hierarchical model for spatial capture-recapture data. *Ecology*, **89**, 2281–2289.

Stevenson, B.C. & Borchers, D.L. (2014) *SECR models with supplementary location information.* R package version 1.0.3.

Wood, S.N. (2010) Statistical inference for noisy nonlinear ecological dynamic systems. *Nature*, **466**, 1102–1104.

Table 1 Parameter estimates, standard errors, and estimated biases from analysis of the *A. lightfooti* data. Parameters above the horizontal line are those that were estimated directly from the acoustic survey. D_c is in calls per hectare per second, D_a is in calling males per hectare, σ_t is in milliseconds, and μ_r is in calls per individual per 25 s.

Parameter	Estimate	Standard error	Bias (%)
D_c	99.15	17.39	0.59
β_{0s}	156.57	1.81	-0.14
β_{1s}	2.67	0.18	-0.22
σ_s	11.50	0.44	-0.07
σ_t	1.96	0.12	0.60
D_a	366.08	64.63	0.62
μ_r	6.77	0.12	0.01

Table 2 Coverage of various confidence interval methods for the parameters D_a and D_c. Nominal coverage was set at 95%. The methods above the horizontal line are calculated from the bootstrap approach from Section 2.5; the naïve method assumes independence between call locations.

CI method	D_a	D_c
Basic	0.924	0.927
Normal	0.942	0.941
Percentile	0.942	0.946
Naive	–	0.729

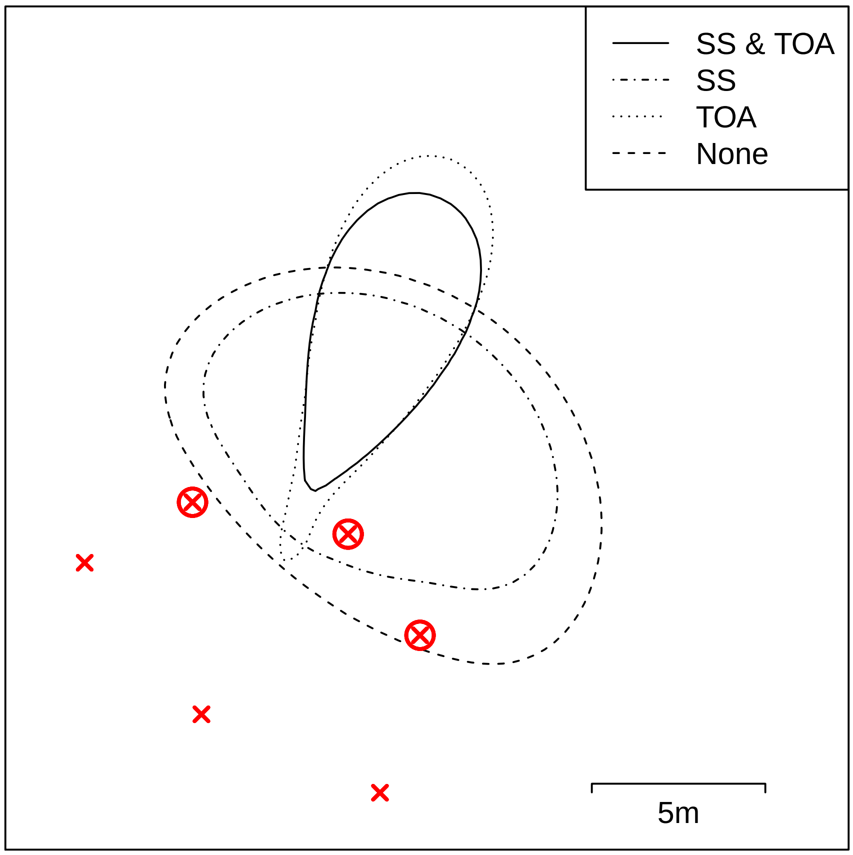

Fig. 1 Estimated locations of a detected call from SECR models with various levels of supplementary information. Crosses show the microphone locations, while circled crosses indicate the microphones at which this particular call was detected. Each contour shows the area within which the call was estimated to have originated with a probability of 0.9. As more additional data is used, the area inside the contour decreases, representing a more precise location estimate.

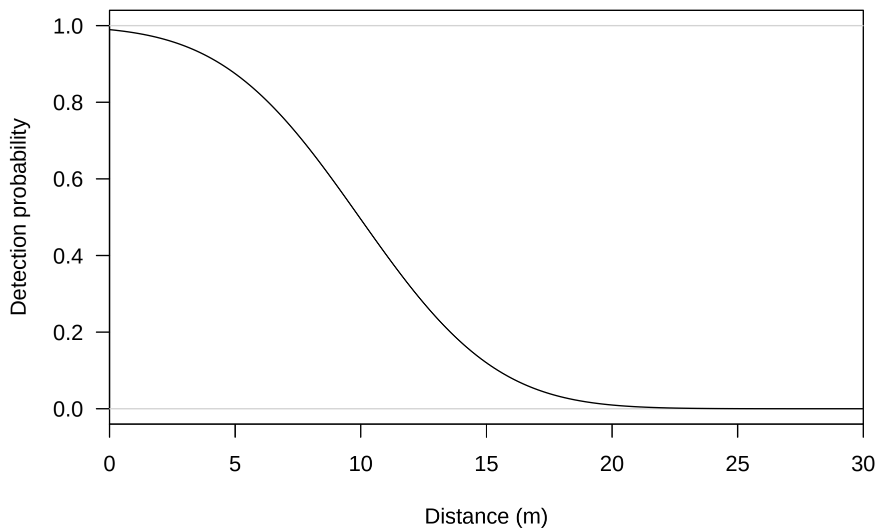

Fig. 2 Estimated detection function, $g(d;\widehat{\gamma})$, from the *A. lightfooti* data.

Table 3 Performance of D_a estimators with and without the use of time of arrival data. Calculated bias is $\widehat{E}(\widehat{D}_a - D_a)$ as a percentage of D_a. CV gives the coefficient of variation as a percentage. MSE gives the mean square error. The simulated data were generated with D_a set at 366.08.

Estimator	Bias (%)	CV (%)	MSE
With TOA	0.62	17.65	4181.73
Without TOA	2.93	23.08	7256.95

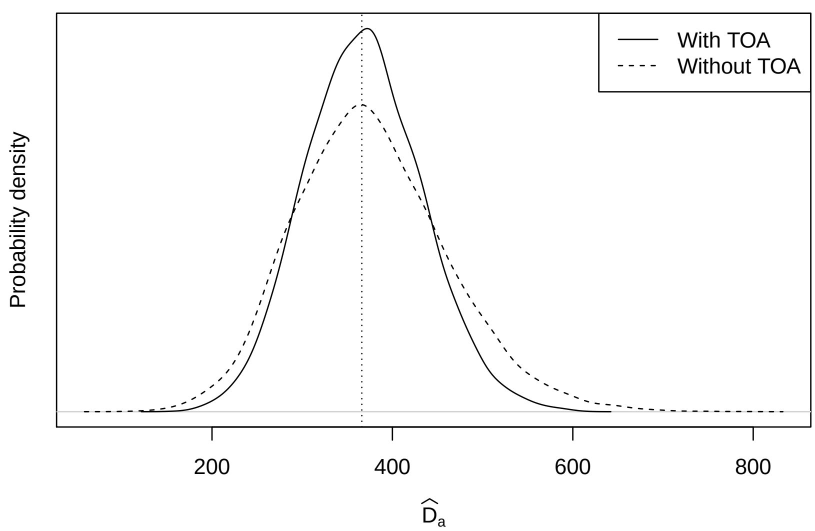

Fig. 3 Estimated sampling distributions of $\widehat{D}_a$ for models with and without time of arrival information incorporated. The dotted vertical line shows the value of D_a used to generate the simulated data.