Master's Dissertation

Bridging Theory and Practice: A Portfolio Dissertation on Acoustic Spatial Capture-Recapture with an Interactive Tutorial

Author

Ané Cloete

Supervisor

Dr David Borchers

I hereby certify that this dissertation, which is 4108 words in length, has been composed by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a degree. This project was conducted by me at the University of St Andrews from June 2023 to August 2023 towards fulfilment of the requirements of the University of St Andrews for the degree of M.Sc. Statistical Ecology under the supervision of Dr David Borchers.

August 28, 2023

Acknowledgements

Distance sampling (DS) infers detection probability from data on distances between the detected animals and detectors, and typically relies on visual observation of animals (Marques et al., 2013; Buckland et al., 2015). Mark-recapture (MR) methods generate data in the form of capture-histories, which give information on the probability of detection and is an analytical framework for a diverse array of survey techniques, such as physical capture, camera traps, and non-invasive genetic sampling (Royle et al., 2013).Traditionally, MR methods did not produce reliable density estimates due to their non-spatial nature (see Royle et al. (2013) for a comprehensive explanation). The development of spatially explicit capture-recapture (SECR or SCR) enabled studies using mark-recapture techniques to estimate density in a sound statistical manner (Efford, 2004). Consequently, for the remainder of this document, mark-recapture and capture-recapture will refer to spatially explicit capture-recapture and will be used interchangeably.

Contents

1 Report

1.1 Introduction

Estimating animal abundance and density is vital in understanding and managing wildlife populations, making it a crucial aspect of ecological research and conservation efforts (Buckland et al., 2015). It is almost always impossible to conduct a complete count of an animal population, so population size estimation is required through survey sampling (Borchers et al., 2015; Buckland et al., 2015). A multitude of methods have been developed over the years, including plot sampling, removal methods, distance sampling and capture-recapture methods (Schwarz and Seber, 1999; Borchers et al., 2002). The latter two methods are arguably the most commonly used techniques and rely on estimating the probability of detecting an animal to estimate density (Marques et al., 2013; Borchers et al., 2015; Buckland et al., 2015; Stevenson et al., 2015).

Distance sampling (DS) infers detection probability from data on distances between the detected animals and detectors, and typically relies on visual observation of animals (Marques et al., 2013; Buckland et al., 2015). Capture-recapture (CR) methods generate data in the form of capture histories containing captures and recaptures of animals, which give information on the probability of detection. CR is an analytical framework for a diverse array of survey techniques, such as physical capture, camera traps, and non-invasive genetic sampling (Lukacs and Burnham, 2005; Green et al., 2020; Ruprecht et al., 2021). Traditionally, CR methods did not produce reliable density estimates due to their non-spatial nature (see Royle et al. (2013) for a comprehensive explanation). The development of spatially explicit capture-recapture (SECR or SCR) enabled studies using capture-recapture techniques to estimate density in a sound statistical manner (Efford, 2004). Consequently, for the remainder of this document, capture-recapture will refer to spatially explicit capture-recapture and will be used interchangeably.

Passive acoustic surveying involves recording animal calls and is particularly useful in situations where species are difficult to sample by sight or other means of capture but produce regular vocalisations, such as the rare and elusive vaquita species (Thomas et al., 2017). Visual observation may be challenging due to, for example, camouflaging or a nocturnal lifestyle (Leaper et al., 1992; Mellinger et al., 2007; Rice et al., 2017) and the capture of individuals may be infeasible for several reasons, including morphology, behaviour or welfare concerns (Putman, 1995; Hammond, 2009; Toy et al., 2017). Marques et al. (2013) further suggest three reasons why acoustic sampling may be preferable even when visual observation or conventional capture methods are suitable: (1) greater range of detectability for animals which produce loud or regular vocalisations, (2) light conditions do not dictate surveying, and (3) amenability to automated data collection and processing facilitates the analysis of vast quantities of data. Passive acoustics is also inherently impartial and non-intrusive, leading to negligible observer bias and minimal disturbance to the population (Sebastián-González et al., 2018; Pérez-Granados and Traba, 2021; Clink et al., 2023).

Although acoustic surveying has many benefits, it is not immune to the common phenomenon in ecological studies of imperfect detection. No matter how rigorous and thorough a survey, some individuals will undoubtedly remain un-captured (here, captured is used interchangeably with detected), and unless imperfect detection is accounted for, estimates of density will be biased (Kellner and Swihart, 2014; Tourani, 2022). Similarly, not all calls will be recorded, and for accurate estimates of density, it is essential to incorporate undetected calls (Stevenson et al., 2021). Distance sampling and capture-recapture methods account for this phenomenon by explicitly modelling the observation process through a detection function. In acoustic surveys, locations of recorded calls are not typically observed and, although supplementary information recorded can be utilized to estimate call locations, estimates of distances between detectors and detected locations are not uniformly precise (e.g. see Sebastián-González et al., 2018). Consequently, distance sampling does not present an ideal framework for analysing acoustic surveys, as the assumption of exact distances is likely to be violated (Stevenson et al., 2015; Stevenson et al., 2021). Efford, Dawson and Borchers (2009) were the first to suggest using SCR methods for analysing data obtained via an acoustic survey with an array of microphones.

Since their seminal work, acoustic SCR (aSCR) has undergone rapid development in terms of model formulation alongside technological advances in recording devices (e.g. Darras et al., 2019) and algorithmic procedures for noise reduction, call recognition and identification (Priyadarshani et al., 2018; Stowell et al., 2018; Clink et al., 2023). However, the adoption of aSCR (and new ecological survey methods in general) in the field has been relatively slow. According to a review by Tourani (2022), out of 364 publications on SCR applications, a mere 2% were applied to acoustic surveys. A contributing factor to its slow integration may be the convenience and familiarity of more established methods. Practitioners often gravitate towards conventional techniques they are already comfortable with and have been using for an extended period. The difference in design of acoustic surveys and the biological mechanisms responsible for generating the observed data is a potential barrier. The inertia associated with transitioning to a new approach, such as aSCR, can be challenging to overcome without proper guidance and support. Thus, I believe one prominent factor is the lack of easily accessible educational resources specifically focused on acoustic spatial capture-recapture. While there have been notable advancements in the field, the absence of comprehensive learning materials has hindered the widespread implementation of aSCR methods.

In light of these challenges, this report has two main goals. The first objective is to offer a high-level overview of acoustic spatial capture-recapture. The reader is assumed to already possess an in-depth understanding of CR concepts and SCR. The second goal is to identify the key elements necessary for effective learning and application, with the intent of creating an interactive tutorial. It is important to note that throughout the report, capture and detection are used interchangeably.

1.2 Acoustic spatial capture-recapture

In the realm of spatial capture-recapture, acoustic SCR stands out as a distinctive subset, where the detections are animal calls, sometimes referred to as vocalisations or cues (Marques et al., 2013; Royle et al., 2013; Petersma et al., 2023). Standard SCR surveys consist of multiple survey occasions, where several detectors are operated during each occasion. The resulting data collected is whether an individual was detected at a detector and on which occasion, such that an individual's capture (or detection) history in space and over time is recorded. The recaptures (or re-detections) across detectors enable SCR models to estimate detection probability, and the spatial features inherent in the survey design and resulting detections allow a sensible estimate of density. Whether re-detections occur across space or time (or both) depends on the type of detector used (Efford, Borchers and Byrom, 2009; Efford, Dawson and Borchers, 2009).

Autonomous recording units (ARUs) used in acoustic surveying are considered proximity detectors as a call can be detected at multiple devices, allowing for simultaneous detections and re-detections across space and within a single survey occasion (Dawson and Efford, 2009; Efford, Dawson and Borchers, 2009). As a result, acoustic capture histories are typically not indexed by time and describe whether a call was detected at each ARU, i.e. call capture histories (Stevenson et al., 2015). However, vocalisations need not be the detection unit (Figure 1). By using acoustics, surveys can focus on detecting individual animals or groups of animals, provided that calls can be classified as belonging to unique individuals or groups. The observed data is then indexed by call, individual/group and detector, translating to, for example, an individual's call capture history (Stevenson et al., 2021). Consequently, the analysis and formulation of the aSCR model differs when the detection unit is not calls (Figure 1). It is important to note here that in all cases, the ultimate density estimate is of calling animals and not the overall population. Conversion to population density requires knowing or estimating the fraction of calling individuals (Stevenson et al., 2015).

In standard SCR, detections are conditional on the animal's activity or home range centre, which is unobserved and thus treated as a latent variable. In aSCR, the latent locations are either animal/group or call locations, depending on the detection unit (Figure 1). When call locations are modelled, the model is termed an aSCR-CD model as calling density is estimated (CD). Two such models have been developed by Stevenson et al. (2015) and Efford, Dawson and Borchers (2009). The latter is not considered in this report, as it discards informative re-detections of calls to avoid the dependence between call locations discussed below.

Unlike the number of animals detected in SCR studies, the number of calls (n_c) detected in aSCR studies is highly unlikely to follow a Poisson point process (PPP). The distribution is unknown because the number of unique individuals detected is unknown, so Stevenson et al. (2015) construct a likelihood conditional on the number of calls allowing for parameter estimation without assuming a distribution for n_c. However, the conditional likelihood is intractable due to the dependence induced by multiple calls from one animal coming from the same location and since individual identities are unknown (Stevenson et al., 2015). A simplified likelihood assuming independence between call locations is formulated instead, leading to an aSCR-CD model assuming a PPP governs call locations. The rationale for this is that assuming independence for non-independent data typically does not significantly influence estimator bias, and the authors show that despite the misspecification, the resulting density estimates are unbiased. However, appropriate variance estimates require a parametric bootstrap procedure.

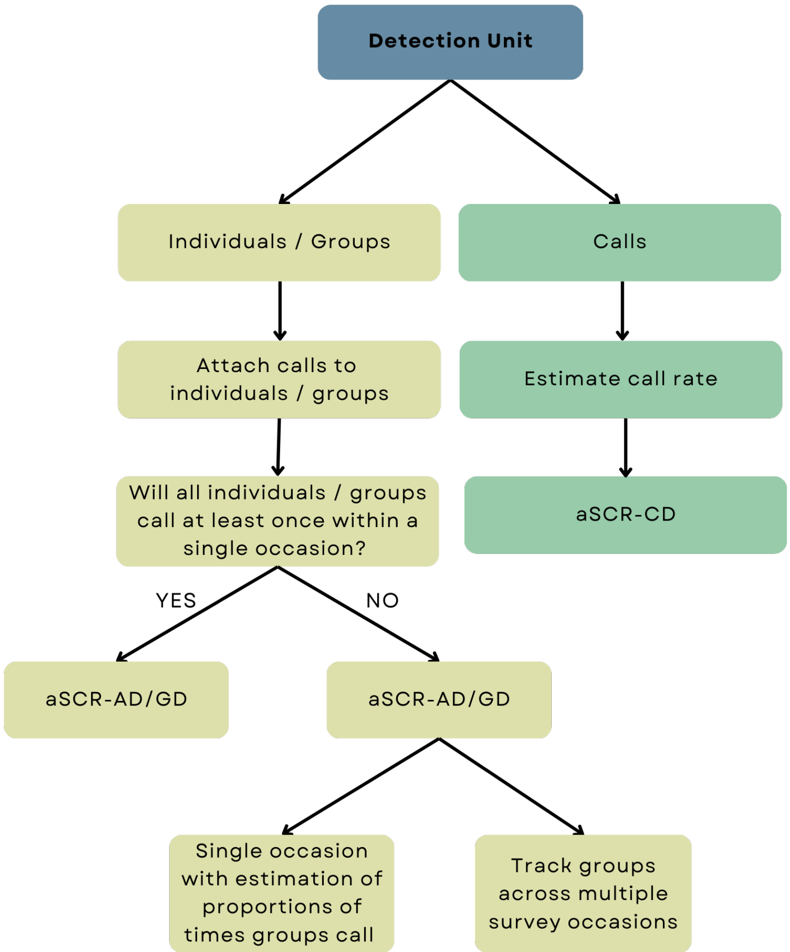

Figure 1: Diagram depicting the acoustic spatial capture-recapture (aSCR) workflow and model depending on the detection unit of an acoustic survey. There are two types of detection units, animals (either as individuals or groups) and calls. The resulting estimate of density is thus either animal or group density (AD/GD) or calling density (CD). When the detection unit is animals, the analysis also depends on whether one can assume that all individuals or groups will call at least once within a single survey occasion. If not, additional data collection is required, as shown in the diagram and described in Section 1.2.3.

Stevenson et al. (2021) developed the aSCR model formulation for animal density (AD) to avoid the issue of dependence between call locations by directly modelling animal locations instead of call locations, allowing for the number of animals detected (n) to be modelled appropriately by a PPP. This also leads to direct estimation of animal density without collecting data on call rate. Note, I rename the model aSCR-AD/GD to include group density (GD) as groups of animals can be viewed theoretically as individuals, thus leading to the same analysis but a different interpretation of the resulting density estimate. They further motivate the development of the aSCR-AD/GD model by arguing that estimating an inhomogeneous density surface is sensible only when modelling animal locations and not call locations. It is possible to fit an inhomogeneous Poisson point process to call locations in a manner analogous to standard SCR models. However, this could lead to spatial confounding, wherein calling clusters from the same individual would be indistinguishable from clusters formed by many calling individuals (Stevenson et al., 2021). Directly modelling animal locations allows the aSCR-AD/GD model to accommodate inhomogeneous processes and circumvent the dependence between call locations but requires that calls can be linked to individuals or groups (Figure 1). It is not always possible to match calls to individuals or groups; even if possible, there may be considerable uncertainty in individual identification. In that case, aSCR-CD is a reasonable and efficient alternative to aSCR-AD/GD (Stevenson et al., 2021).

Typically, recording devices can collect additional data associated with detected calls, and the inclusion of such auxiliary information has been shown to improve estimates of the latent location variables (Efford, Dawson and Borchers, 2009; Stevenson et al., 2015; Borchers et al., 2015; Stevenson et al., 2021). Signal strength (SS) and time of arrival (TOA) have been used in previous analyses as they are informative of animal and call locations. On average, the louder the received signal at a detector, the closer the calling individual. Similarly, earlier detection of a call suggests increased proximity between the calling animal and the detector. In aSCR-CD, signal strength, in addition to improving the efficiency of estimates, may be further included in the detection function when it is used to determine a detection threshold above which a received acoustic signal is considered to be a detection (Stevenson et al., 2015).

The following section contains notation to describe and formulate aSCR models; for a comprehensive explanation of the model formulations, see Appendix A and B.

We consider a survey of duration T with m acoustic recording devices which are situated at known locations, $\mathbf{X} = (x_1, \ldots, x_k)$, within the survey region $S \subset \mathbb{R}^2$. The survey region, in line with SCR, is defined as the set of all locations where calling individuals (or groups) can be detected by the detector array (Dawson and Efford, 2009; Efford, Dawson and Borchers, 2009; Stevenson et al., 2015; Stevenson et al., 2021).

1.2.1 Calling Density (aSCR-CD)

In the aSCR-CD model developed by Stevenson et al. (2015), signal strength and time of arrival are recorded together with detections of acoustic signals. Vocalisations of the species are only considered detections if their received signal strength exceeds a certain threshold, c, ensuring that calls can be easily recognised. The resulting data is the number of unique vocalisations (n_c), the detected calls' capture histories denoted by $\mathbf{Y}$ and the supplementary information recorded by the devices.

The capture histories are in the form, y_{ik}, where $y_{ik} = 1$ if call $i \in \{1, \ldots, n_c\}$ call was detected at device $k \in \{1, \ldots, m\}$ and zero otherwise. The i^{th} call detected at the k^{th} microphone also has an associated signal strength, s_{ik}, and time of arrival, z_{ik}, where the set of all SS and TOA observations are contained in $\mathbf{S}$ and $\mathbf{Z}$. Note, time of arrival measurements are with respect to the start of the survey. The observed data for the i^{th} call is therefore contained in $\mathbf{y_i}, \mathbf{s_i}$ and $\mathbf{z_i}$, where each vector consists of m elements, one for each detector. The detection locations, i.e. call locations, are taken as latent variables and are denoted by $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_{n_c})$. Each $\mathbf{u}_i$ comprises Cartesian coordinates within the survey region S.

Following the approach of Borchers and Efford (2008), the authors construct a likelihood conditional on n_c with model parameters, γ and ϕ, relating to the SS and TOA processes. However, the conditional likelihood is intractable due to the reasons described previously, and to address the complexity of the likelihood function, Stevenson et al. (2015) construct a simplified likelihood by assuming independence between call locations:

$$L_s(\boldsymbol{\gamma}, \boldsymbol{\phi}) = \prod_{i=1}^{n_e} \int_S f_s(\boldsymbol{s}_i \mid \boldsymbol{y}_i, \boldsymbol{u}_i) f_z(z_i \mid \boldsymbol{y}_i, \boldsymbol{u}_i) f_y(\boldsymbol{y}_i \mid \boldsymbol{u}_i) f_u(\boldsymbol{u}_i) d\boldsymbol{u}_i. \tag{1}$$

The first component of the simplified likelihood, hereafter referred to as simply the likelihood, describes the probability density of signal strengths based on the capture histories and latent call locations (f_s). The observation is disregarded if the signal strength does not exceed the threshold and the probability density function (PDF) is set to one. Otherwise, the PDF for observed signal strength is defined as a truncated Gaussian distribution with standard deviation represented by σ_s:

$$f_s\left(s_{ik} \mid y_{ik} = 1, \boldsymbol{u}_i\right) = \frac{1}{\sigma_s} f_n\left(\frac{s_{ik} - E\left(s_{ik} \mid \boldsymbol{u}_i\right)}{\sigma_s}\right)\left(1 - \Phi\left(\frac{c - E\left(s_{ik} \mid \boldsymbol{u}_i\right)}{\sigma_s}\right)\right)^{-1}. \tag{2}$$

In the above, the probability and cumulative density functions of the standard normal distribution are denoted by f_n and Φ, respectively.

The expected signal strength given the call location can be modelled via a link function with distance between microphone and call locations (d) as a covariate, i.e. $E(s_{ik} \mid \boldsymbol{u}_i) = h^{-1}(\beta_0 - \beta_s d)$, but any other reasonable specifications are allowed (Efford, Borchers and Byrom, 2009; Stevenson et al., 2015).

The second component (f_z) describes the TOA process and is defined as:

$$f(z_i \mid m_i, \boldsymbol{u}_i) = \begin{cases} \frac{\left(2\pi\sigma_z^2\right)^{(1-m_i)/2}}{2T\sqrt{m_i}} \exp\left(\sum_{\{k:y_{ik}=1\}} \frac{\left(\delta_{ik}(\boldsymbol{u}_i)-\bar{\delta}_i\right)^2}{-2\sigma_t^2}\right) & m_i > 1 \\ 1 & m_i = 1 \end{cases} \tag{3}$$

where the number of microphones that detected the i^{th} call is denoted by m_i, the expected call time is denoted by, δ_{ik}, and $\bar{\delta}_i$ is the corresponding mean. The above PDF is only parametrically defined when m_i is greater than one because it is *differences* between arrival times, i.e. when the call is detected at multiple microphones, that provide useful information about the location.

When using the SS threshold to define a detection, the probability of detection is directly related to the probability of the signal strength surpassing the threshold. The detection function, $g_c(d;\gamma)$, which describes the probability of detection given the distance between call location and microphone, thus incorporates the model specification of the expected received signal strength:

$$g_c(d;\gamma) = 1 - \Phi\left(\frac{c - h^{-1}\left(\beta_{0s} - \beta_{1s} d\right)}{\sigma_s}\right) \tag{4}$$

It is important to note if signal strength data are not collected, the detection function can be specified as, for example, the hazard half-normal function, or any appropriate parametric form.

The conditional distribution of a call capture history, y_{ik}, is then defined in terms of the specified detection function:

$$f(y_{ik} \mid \boldsymbol{u}_i) = \begin{cases} g\left(d_k\left(\boldsymbol{u}_i\right)\gamma\right) & y_{ik} = 1 \\ 1 - g\left(d_k\left(\boldsymbol{u}_i\right)\gamma\right) & y_{ik} = 0 \end{cases} \tag{5}$$

and the resulting joint mass function assuming independence between detected calls is:

$$f(\boldsymbol{y}_i \mid \boldsymbol{u}_i) = \frac{\prod_{k=1}^{K} f(y_{ik} \mid \boldsymbol{u}_i)}{p_c(\boldsymbol{u}_i;\boldsymbol{\gamma})}. \tag{6}$$

Given that only capture histories where at least one detection is made are observed, the above PMF is zero-truncated by including in the denominator the probability of detecting a call originating from $\mathbf{u}_i$ at all:

$$p_c(\boldsymbol{u}) = 1 - \prod_{k=1}^{m} [1 - g_c(d_k(\mathbf{u}))] \tag{7}$$

Finally, call locations are assumed to be realisations of a homogenous Poisson point process (HPPP) thinned by the detection probability surface:

$$f(\boldsymbol{u}_i) \propto p_c(\boldsymbol{u}_i; \boldsymbol{\gamma}) = \frac{p_c(\boldsymbol{u}_i; \boldsymbol{\gamma})}{a(\boldsymbol{\gamma})}, \tag{8}$$

where the effective sampling area (ESA), denoted by $a(\boldsymbol{\gamma})$, is defined as, $\int_A p_c(\boldsymbol{u}_i; \boldsymbol{\gamma})\, du$ and thus acts as the normalising constant. Each component of the likelihood in Equation 20 has now been described.

To obtain parameter estimates of γ and ϕ, the log-likelihood is maximised:

$$(\widehat{\boldsymbol{\gamma}}, \widehat{\boldsymbol{\phi}}) = \arg\max_{\gamma,\phi} \log\left(L_s(\boldsymbol{\gamma}, \boldsymbol{\phi})\right), \tag{9}$$

and with a Horvitz-Thompson-like estimator, calling density (D_c), is calculated using the estimate of γ:

$$\widehat{D}_c = \frac{n_c}{a(\widehat{\gamma})T} \tag{10}$$

The interpretation of D_c is that of the number of calls per unit area per unit time. Density of calling individuals, D_a, is then estimated simply be dividing with the species call rate, c_r. The call rate can be known beforehand or estimated from independently collected data. Standard errors and confidence intervals are estimated via a parametric bootstrap procedure using the parameters estimates and call rate data (see Appendix A for more details).

1.2.2 Animal Density (aSCR-AD/GD)

In this section, the terms individual(s) or animal(s) will be used, but the same applies to group(s). When estimating calling density, it is required that the number of unique calls is identifiable (n_c). When individuals are the detection unit, two additional requirements are that calls can be matched to individuals and that individuals do not move during the survey. In other words, calls from one individual originate from the same location. The capture histories are thus represented by y_{ijk}, such that if the i^{th} call, $i \in \{1, \ldots, c_i\}$, of the j^{th} individuals, $j \in \{1, \ldots, n\}$, was detected at the k^{th} microphone, $k \in \{1, \ldots, m\}$, $y_{ijk} = 1$ and zero otherwise (Stevenson et al., 2021). The same modification applies to any supplementary information also recorded by the detectors, here TOA is included, and z_{ik} from the previous section gains an additional index for the j^{th} individual, z_{ijk}.

The observed data comprise the number of detected animals (n), the number of calls produced by each individual, $\mathbf{c} = (c_1, \ldots, c_n)$, the individuals' capture histories, $\mathbf{Y} = (Y_1, \ldots, Y_n)$, and the TOA measurements, ($\mathbf{Z} = (Z_1, \ldots, Z_n)$). The latent locations, $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$, are now the physical locations of the animals during the survey and not the individual call locations.

The model likelihood, with parameters contained in $\boldsymbol{\theta}$, takes the following form:

$$L(\boldsymbol{\theta}) = f_n(\mathbf{n}) \prod_{j=1}^{n} \int_S f_z(\mathbf{Z}_i \mid \mathbf{Y}_i, \mathbf{c}_i, \mathbf{u}_i) f_y(\mathbf{Y}_i \mid \mathbf{c}_i, \mathbf{u}_i) f_c(\mathbf{c}_i \mid \mathbf{u}_i) f_u(\mathbf{u}_i)\, d\mathbf{u}_i. \tag{11}$$

The above assumes that individuals' locations are independent, but there is dependence between call locations within and between individuals. See Appendix B for a detailed model formulation.

The call detection probability, $g_c(d_k(\mathbf{u}))$, is unchanged except for the interpretation of $\mathbf{u}$ and can be modelled using any suitable detection function as before. Detections of the same calls are still assumed to be independent across microphones, and so, the probability that a call produced at $\mathbf{u}$ is detected at all remains as specified before in Equation 7.

The probability mass function for the number of detected calls, c_i, of the i^{th} individual conditional on the latent animal location is calculated as follows:

$$f_c(c_i \mid \mathbf{u}_i) = \sum_{C_i=1}^{\infty} f(c_i \mid C_i, \mathbf{u}_i) f(C_i), \tag{12}$$

where C_i is the unobserved total number of calls produced by individual i. It is assumed that the PMF for c_i conditional on C_i above is a zero truncated binomial with probability of success equal to $p_c(\mathbf{u})$ and the call production process, $f(C_i)$, follows a Poisson distribution with parameter $c_r T$. Consequently, the following holds true:

$$(c_i \mid \mathbf{u}_i) \sim \text{ ZTPoisson}\{c_r T p_c(\mathbf{u}_i)\} \tag{13}$$

The PMF for the capture history matrices, conditional on the latent locations, is equivalent to that of call capture histories in aSCR-CD (Equation 5), as the individual capture histories are also realisations of a Bernoulli trial with probability of success governed by the detection function:

$$\mathbf{y}_{ijk}|\mathbf{u} \sim \text{Bernoulli}[g_c(d(\mathbf{u}_i))]. \tag{14}$$

Additionally, as with aSCR-CD, the observed capture histories are zero-truncated by incorporating the probability of detecting the i^{th} call produced at $\mathbf{u}_i$:

$$f\left(\mathbf{Y}_i \mid c_i, \mathbf{u}_i\right) = \frac{\prod_{j=1}^{c_i} f\left(\mathbf{y}_{ij} \mid \mathbf{u}_i\right)}{p_c(\mathbf{u}_i)}. \tag{15}$$

It is assumed that animal locations are realisations of an inhomogeneous PPP with intensity parameter, $D(\mathbf{u})$, describing the density of individuals at location $\mathbf{u} \in \mathbf{S}$. Locations of detected animals are realisations of the same process, but one that is thinned by the probability of detecting an animal at $\mathbf{u}$, i.e. $p(\mathbf{u})$. Accordingly, the number of detected animals is modelled by a Poisson PMF with parameter equal to, $D(\mathbf{u})p(\mathbf{u})$, the intensity function of the thinned PPP:

$$f\left(n\right) = \frac{\left\{\int_S D(\mathbf{u})p(\mathbf{u})d\mathbf{u}\right\}^n \exp\left\{-\int_S D(\mathbf{u})p(\mathbf{u})d\mathbf{u}\right\}}{n!}, \tag{16}$$

and the PDF of the latent locations is equal to the normalised intensity function:

$$f\left(\mathbf{s}_i\right) = \frac{D\left(\mathbf{u}_i\right)p\left(\mathbf{u}_i\right)}{\int_S D(\mathbf{u})p(\mathbf{u})d\mathbf{u}}. \tag{17}$$

Spatial heterogeneity in animal density can be modelled via a log link function with spatial covariates measured at locations $\mathbf{u}$.

As before, $f\left(\mathbf{Z}_i \mid \mathbf{Y}_i, \mathbf{c}_i, \mathbf{u}_i\right)$, describes the conditional distribution of the TOA process and its formulation is exactly as in aSCR-CD. The only difference being that the PDF is now defined for the i^{th} call of the j^{th} individual:

$$f\left(\boldsymbol{Z}_i \mid \mathbf{Y}_i, c_i, \boldsymbol{u}_i\right) = \prod_{j=1}^{c_i} f\left(\boldsymbol{z}_{ij} \mid \boldsymbol{y}_{ij}, \boldsymbol{u}_i\right). \tag{18}$$

Finally, the model parameters contained in $\boldsymbol{\theta}$ thus consist of the regression coefficients, $\boldsymbol{\beta}$, the parameters of the chosen detection function (here λ_0 and σ), the call rate c_r and standard deviation of the TOA measurements, σ_z. Estimates are obtained via maximisation of the log likelihood:

$$(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda_0}}, \widehat{\boldsymbol{\sigma}}, \widehat{\boldsymbol{c_r}}, \widehat{\boldsymbol{\sigma_z}}) = \arg\max_{\boldsymbol{\theta}} \log\left(L_s(\boldsymbol{\theta})\right), \tag{19}$$

and variance estimates and confidence intervals are obtained as in standard maximum likelihood analysis.

1.2.3 aSCR-AD/GD with additional data collection

The above model formulation assumes that all individuals or groups call at least once within a single survey. However, this assumption does not apply to all species, and increasing survey length to the extent that the assumption holds is typically not feasible. For instance, gibbons (Family Hylobatidae) occur in groups that engage in duetted singing and not all groups will sing every day (Brockelman and Srikosamatara, 1993). Seasonal and weather-dependent changes in singing frequency also occur between and within gibbon groups (Raemaekers et al., 1984; Cheyne, 2008). For these types of wildlife populations, a single survey occasion only samples the groups (or individuals) that called during that particular occasion, leading to biased density estimates.

To address this issue, one can estimate the probability of calling during a survey occasion and correct for the silent groups accordingly. One method of estimating calling probability requires that surveyors conduct focal follows of detected groups from the aSCR survey over multiple additional occasions and determine the average proportion of occasions the groups called. However, it is essential to note that this method may lead to biased results as if only a few groups have been detected, the sample may not be representative of the entire gibbon population (Vu et al., 2018). Alternatively, multiple survey occasions can be conducted to ensure that all units in the survey region are likely to have been detected, which thus requires that individuals or groups can be linked across occasions (Jiang et al., 2006; Kidney et al., 2016).

1.3 Discussion

This section identifies key elements for inclusion in the tutorial and provides an overview of its intended structure, highlighting its broader scope beyond that of the dissertation. Only chapters one to five will be developed to meet the thesis requirements within the given time frame. The ultimate goal is to develop the entire tutorial.

The tutorial will be designed to progress naturally from chapter to chapter; therefore, learners are expected to work through them sequentially. The intended audience of the tutorial is assumed to have some knowledge of CR concepts but may be largely unfamiliar with SCR, thus necessitating an introduction in the context of acoustic sampling. The first chapter will set the scene by introducing the concept of estimating animal abundance and acoustic surveying.

Chapter 1: Setting the Scene

- Introduction to aSCR

At the core of all SCR models lies the fundamental relationship between detection probability and distance. This concept will be elucidated in Chapter 2 by first exploring a basic CR model and demonstrating how abundance estimation is possible using equal detection probability from capture history data. It will become apparent that assuming an equal probability of capture across a detector array is inappropriate, and the relationship between distance and detections will be discussed.

Chapter 2: Probability of detection

- Probability of detection & basic capture-recapture analysis
- Distance dependent detection

Chapter 2 will serve as a natural introduction to SCR; thus, the SCR framework will be presented in the following chapter. Instead of providing an exhaustive explanation of SCR, the tutorial will offer a brief intuitive introduction, mainly in the context of acoustic surveying, to foster an understanding of SCR and facilitate the formulation of the aSCR methodology.

Chapter 3: Introduction to SCR

- Survey structure
- Model structure

The chapter will discuss how the SCR framework considers the variability in detection probability resulting from distances via the detection function and incorporates a defined sampling area and a density model. The primary goal is to explain the SCR framework to the learner while minimizing the use of complex statistical terminology. When discussing the density model, the distinction between call and animal locations will be mentioned, leading to the idea of the detection unit. Following this, Chapter 4 will expand upon the detection unit and explain the different analyses and interpretations of density estimates resulting from each detection unit. Additionally, the recording and use of supplementary data in acoustic surveying will be mentioned.

Chapter 4: Overview of aSCR

- Detection units
- Calling density
- Animal density
- Group density

The subsequent chapters (5, 6, 7, and 8) will be dedicated to each analysis method outlined in this report, and models will be implemented using the newly developed package, `acre`, in R (R Core Team, 2022). Chapter 4 will centre on aSCR-CD, commencing with a step-by-step introduction of the package structure using a simple aSCR-CD model. The following chapter will incorporate supplementary information into the aSCR-CD model.

Chapter 5: aSCR Calling Density

- Introduction to the acre package using aSCR-CD models

Chapter 6: aSCR-CD with supplementary data

- Fitting aSCR-CD models with signal strength and time of arrival data

Chapter 7 will present aSCR-AD/GD in a format consistent with the previous chapter while highlighting the key differences compared to aSCR-CD. Chapter 8 will delve into the additional requirements necessary when one cannot assume that all individuals or groups will call at least once during a single survey occasion.

Chapter 7: aSCR-AD

- Difference in data requirements
- Simple model
- Advanced model with supplementary data

Chapter 8: aSCR-AD/GD with additional data collection

- Estimating calling probability
- Tracking groups

Most of the statistical and mathematical details presented in this report are unnecessary for inclusion in the tutorial. The structure and elements discussed should suffice to facilitate the adoption of aSCR in practice. The tutorial intends not to delve into the intricacies and differences of the aSCR models. Instead, it will focus on bridging the gap between theory and practice, empowering practitioners to confidently integrate acoustic spatial capture-recapture into their ecological studies and conservation efforts.

4003 words

2 Free choice element

The online tutorial is a combination of the free choice elements web page and teaching materials. To access the tutorial, please click **here**.

3 Reflective Essay

Throughout my honour's thesis, which focused on Bayesian dynamic occupancy models, I discovered that they remained underutilised in ecological literature despite their considerable power and flexibility. Several factors, including these models' computational demands and time-intensive nature, contribute to this. However, even reaching the stage where these challenges become apparent necessitates a solid understanding of the methodology and its application. For individuals not well-versed in statistics or coding, these models can appear quite daunting, and unfortunately, there is a scarcity of easily accessible online resources addressing them. While a commendable textbook is dedicated to these models, it remains out of reach for those without university affiliations or the financial means to acquire it.

This observation extended beyond dynamic occupancy models, revealing a broader pattern concerning novel statistical ecology methods. Typically, gaining information about these techniques requires deciphering statistical publications, attending in-person workshops or, at the most extreme, enrolling in a university programme. Beyond these, practitioners may still gravitate toward more established methods. Contributing factors may be the familiarity of these methods, the lack of comprehensive comparative studies, limited availability of open access code and data or user-friendly software interfaces and packages. One significant element may be the lack of accessible educational resources guiding practitioners. This scarcity of free educational resources might be due to funding limitations or experts' time constraints, compounded by the fact that statistical ecology is a relatively small field without significant commercial profit motives.

Motivated by these insights, I opted for a portfolio dissertation to create something with tangible utility for practitioners. My goal was to move beyond the traditional demonstration of proficiency in a specific statistical method and outside my comfort zone to produce something accessible and comprehensible for practitioners in the field. Thus, I built a tutorial introducing an innovative statistical ecology method: acoustic spatial capture-recapture.

The education landscape evolves rapidly, spurred by technological advancements and unforeseen transformative events. Particularly for subjects intertwined with coding, the learning trajectory is shifting towards online platforms, gradually overshadowing traditional in-person approaches. In response to this trend, I've developed an online tutorial that emphasises real-time interaction with code using the *learnr* package in R, eliminating any software installation hassles. This tutorial encourages active learning through hands-on interactive components, exercises and quizzes. It's designed to accommodate varying learning speeds, ensuring everyone can progress at their own pace. Moreover, being hosted online provides free and unrestricted access for all learners.

However, while my tutorial serves as an accessible introduction to the method, it's essential to note its limitations beyond the intended structure. The tutorial would benefit from a comprehensive chapter on survey design and data pre-processing, two critical components in applying these methods effectively. Although the tutorial is aimed at practitioners who may not understand more complex statistical concepts, understanding the method's full utility requires a deep dive into its mathematical foundation and underlying assumptions. To include such explanations in the tutorial, the tutorial would need to have a chapter introducing statistical modelling and probability distributions, significantly increasing the its scope. I acknowledge that this tool isn't a one-size-fits-all solution, but it is my endeavour to bolster the adoption of aSCR.

Over the past few months, I've not only gained confidence in my abilities but also acquired new skills. Teaching something I've only recently learned was challenging. I needed to understand the topic well enough to effectively explain it to an audience who might not have an extensive statistical background. I had to draw upon the experience and knowledge I gained throughout my academic career and prove to myself that I understand these concepts and can demonstrate them to others in a straightforward manner. This experience has taught me to trust in myself and emphasised the importance of planning while remaining open to change and adaptation in the face of unexpected challenges. Interspersed throughout the tutorial are interactive web applications built with the R package, *Shiny*. Prior to working on my dissertation, I had limited experience with web application development. Now, I can confidently incorporate them into my projects. Additionally, I gained experience with other programming languages such as HTML and CSS.

I've also confronted some of my weaknesses. When I come across something that seems straightforward, but I don't grasp it right away, I often hesitate to ask for help, fearing that I've somehow fallen short. Sometimes, it's possible to dodge and avoid the issue, but I needed to be thorough here. To ensure I covered all the bases, I had to overcome my fear of failure and the discomfort of appearing inadequate and simply reach out for assistance. In the end, I discovered that many people have felt the same. It turns out that even the individuals I hold in high esteem face similar struggles.

In conclusion, the underutilisation of novel statistical models in ecological studies, such as acoustic SCR, underscores a broader gap in accessible educational resources. While there are inherent challenges in comprehending and applying these models, especially for those unfamiliar with intricate statistical concepts, their benefits are undeniable. My journey in crafting an online tutorial has not only been about bridging this knowledge gap but also a personal exploration of my strengths, weaknesses, and the ever-evolving landscape of education in the digital age. Through my endeavours, I've reinforced the significance of making complex methods more approachable and the profound impact of continuous learning and adaptation.

888 words

References

Borchers, D. L., Buckland, S. T., Zucchini, W. and Borchers, D. (2002). *Estimating animal abundance: closed populations*, Vol. 13, Springer.

Borchers, D. L. and Efford, M. (2008). Spatially explicit maximum likelihood methods for capture–recapture studies, *Biometrics* **64**(2): 377–385.

Borchers, D. L., Stevenson, B., Kidney, D., Thomas, L. and Marques, T. A. (2015). A unifying model for capture–recapture and distance sampling surveys of wildlife populations, *Journal of the American Statistical Association* **110**(509): 195–204.

Brockelman, W. Y. and Srikosamatara, S. (1993). Estimation of density of gibbon groups by use of loud songs, *American Journal of Primatology* **29**(2): 93–108.

Buckland, S. T., Rexstad, E. A., Marques, T. A., Oedekoven, C. S. et al. (2015). *Distance sampling: methods and applications*, Vol. 431, Springer.

Cheyne, S. M. (2008). Effects of meteorology, astronomical variables, location and human disturbance on the singing apes: Hylobates albibarbis, *American Journal of Primatology: Official Journal of the American Society of Primatologists* **70**(4): 386–392.

Clink, D. J., Kier, I., Ahmad, A. H. and Klinck, H. (2023). A workflow for the automated detection and classification of female gibbon calls from long-term acoustic recordings, *Frontiers in Ecology and Evolution* **11**: 1071640.

Darras, K., Batáry, P., Furnas, B. J., Grass, I., Mulyani, Y. A. and Tscharntke, T. (2019). Autonomous sound recording outperforms human observation for sampling birds: a systematic map and user guide, *Ecological Applications* **29**(6): e01954.

Dawson, D. K. and Efford, M. G. (2009). Bird population density estimated from acoustic signals, *Journal of Applied Ecology* **46**(6): 1201–1209.

Efford, M. (2004). Density estimation in live-trapping studies, *Oikos* **106**(3): 598–610.

Efford, M. G., Borchers, D. L. and Byrom, A. E. (2009). Density estimation by spatially explicit capture–recapture: likelihood-based methods, *Modeling demographic processes in marked populations* pp. 255–269.

Efford, M. G., Dawson, D. K. and Borchers, D. L. (2009). Population density estimated from locations of individuals on a passive detector array, *Ecology* **90**(10): 2676–2682.

Green, A. M., Chynoweth, M. W. and Şekercioğlu, Ç. H. (2020). Spatially explicit capture-recapture through camera trapping: a review of benchmark analyses for wildlife density estimation, *Frontiers in Ecology and Evolution* p. 473.

Hammond, P. S. (2009). Mark–recapture, *Encyclopedia of marine mammals*, Elsevier, pp. 705–709.

Jiang, X., Luo, Z., Zhao, S., Li, R. and Liu, C. (2006). Status and distribution pattern of black crested gibbon (nomascus concolor jingdongensis) in wuliang mountains, yunnan, china: implication for conservation, *Primates* **47**: 264–271.

Juodakis, J., Castro, I. and Marsland, S. (2021). Precision as a metric for acoustic survey design using occupancy or spatial capture-recapture, *Environmental and Ecological Statistics* **28**: 587–608.

Kellner, K. F. and Swihart, R. K. (2014). Accounting for imperfect detection in ecology: a quantitative review, *PloS one* **9**(10): e111436.

Kidney, D., Rawson, B. M., Borchers, D. L., Stevenson, B. C., Marques, T. A. and Thomas, L. (2016). An efficient acoustic density estimation method with human detectors applied to gibbons in cambodia, *PloS one* **11**(5): e0155066.

Leaper, R., Chappell, J. and Gordon, J. (1992). *The development of practical techniques for surveying sperm whale populations acoustically*, University of Oxford, Department of Zoology, Wildlife Conservation Research Unit.

Lukacs, P. M. and Burnham, K. P. (2005). Review of capture–recapture methods applicable to noninvasive genetic sampling, *Molecular ecology* **14**(13): 3909–3919.

Marques, T. A., Thomas, L., Martin, S. W., Mellinger, D. K., Ward, J. A., Moretti, D. J., Harris, D. and Tyack, P. L. (2013). Estimating animal population density using passive acoustics, *Biological reviews* **88**(2): 287–309.

Mellinger, D. K., Stafford, K. M., Moore, S. E., Dziak, R. P. and Matsumoto, H. (2007). An overview of fixed passive acoustic observation methods for cetaceans, *Oceanography* **20**(4): 36–45.

Pérez-Granados, C. and Traba, J. (2021). Estimating bird density using passive acoustic monitoring: a review of methods and suggestions for further research, *Ibis* **163**(3): 765–783.

Petersma, F. T., Thomas, L., Thode, A. M., Harris, D., Marques, T. A., Cheoo, G. V. and Kim, K. H. (2023). Accommodating false positives within acoustic spatial capture–recapture, with variable source levels, noisy bearings and an inhomogeneous spatial density, *Journal of Agricultural, Biological and Environmental Statistics* pp. 1–20.

Priyadarshani, N., Marsland, S. and Castro, I. (2018). Automated birdsong recognition in complex acoustic environments: a review, *Journal of Avian Biology* **49**(5): jav–01447.

Putman, R. (1995). Ethical considerations and animal welfare in ecological field studies, *Biodiversity & Conservation* **4**: 903–915.

R Core Team (2022). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
URL: *https://www.R-project.org/*

Raemaekers, J. J., Raemaekers, P. M. and Haimoff, E. H. (1984). Loud calls of the gibbon (hylobates lar): repertoire, organisation and context, *Behaviour* **91**(1-3): 146–189.

Rice, A. N., Soldevilla, M. S. and Quinlan, J. A. (2017). Nocturnal patterns in fish chorusing off the coasts of georgia and eastern florida, *Bulletin of Marine Science* **93**(2): 455–474.

Royle, J. A., Chandler, R. B., Sollmann, R. and Gardner, B. (2013). *Spatial capture-recapture*, Academic Press.

Ruprecht, J. S., Eriksson, C. E., Forrester, T. D., Clark, D. A., Wisdom, M. J., Rowland, M. M., Johnson, B. K. and Levi, T. (2021). Evaluating and integrating spatial capture–recapture models with data of variable individual identifiability, *Ecological Applications* **31**(7): e02405.

Schwarz, C. J. and Seber, G. A. (1999). A review of estimating animal abundance iii, *Statistical Science* **14**(4): 427–456.

Sebastián-González, E., Camp, R. J., Tanimoto, A. M., de Oliveira, P. M., Lima, B. B., Marques, T. A. and Hart, P. J. (2018). Density estimation of sound-producing terrestrial animals using single automatic acoustic recorders and distance sampling, *Avian Conservation and Ecology* .

Stevenson, B. C., Borchers, D. L., Altwegg, R., Swift, R. J., Gillespie, D. M. and Measey, G. J. (2015). A general framework for animal density estimation from acoustic detections across a fixed microphone array, *Methods in Ecology and Evolution* **6**(1): 38–48.

Stevenson, B. C., van Dam-Bates, P., Young, C. K. and Measey, J. (2021). A spatial capture–recapture model to estimate call rate and population density from passive acoustic surveys, *Methods in Ecology and Evolution* **12**(3): 432–442.

Stowell, D., Wood, M., Pamuła, H., Stylianou, Y., Glotin, H. and Orme, D. (2018). Automatic acoustic detection of birds through deep learning: the first bird audio detection challenge. methods ecol evol 10 (3): 368–380.

Thomas, L., Jaramillo-Legorreta, A., Cardenas-Hinojosa, G., Nieto-Garcia, E., Rojas-Bracho, L., Ver Hoef, J. M., Moore, J., Taylor, B., Barlow, J. and Tregenza, N. (2017). Last call: Passive acoustic monitoring shows continued rapid decline of critically endangered vaquita, *The Journal of the Acoustical Society of America* **142**(5): EL512–EL517.

Tourani, M. (2022). A review of spatial capture–recapture: Ecological insights, limitations, and prospects, *Ecology and Evolution* **12**(1): e8468.

Toy, D. L., Roche, E. A. and Dovichin, C. M. (2017). Small high-definition video cameras as a tool to resight uniquely marked interior least terns (sternula antillarum athalassos), *Waterbirds* **40**(2): 180–186.

Vu, T. T., Tran, L. M., Nguyen, M. D., Van Tran, D. and Ta, N. T. (2018). Improving the estimation of calling probability and correction factors in gibbon monitoring using the auditory point count method, *International Journal of Primatology* **39**: 222–236.

A ASCR-CD Model formulation

We consider a survey of duration T with K acoustic recording devices which are situated at known locations, $\mathbf{X} = (x_1, \ldots, x_k)$, within the survey region $S \subset \mathbb{R}^2$. The survey region, in line with SCR, is defined as the set of all locations where calling individuals (or groups) can be detected by the detector array (Dawson and Efford, 2009; Efford, Dawson and Borchers, 2009; Stevenson et al., 2015; Stevenson et al., 2021).

In the ASCR-CD model developed by Stevenson et al. (2015), signal strength and time of arrival are recorded together with detections of acoustic signals. Vocalisations of the species are only considered detections if their received signal strength exceeds a certain threshold, c, which ensures that calls can be recognized with ease. The resulting data is the number of unique vocalisations (n_c), the detected calls' capture histories denoted by $\mathbf{Y}$ and the supplementary information recorded by the devices.

The capture histories are in the form, y_{ik}, where $y_{ik} = 1$ if call $i \in \{1, \ldots, n_c\}$ call was detected at device $k \in \{1, \ldots, m\}$ and 0 otherwise. Each detected call at a microphone also has an associated signal strength, s_{ik}, and time of arrival, z_{ik}, where the set of all SS and TOA observations are contained in $\mathbf{S}$ and $\mathbf{Z}$. Note, time of arrival measurements are with respect to the start of the survey. The observed data for the i^{th} call is therefore contained in $\mathbf{y_i}, \mathbf{s_i}$ and $\mathbf{z_i}$, where each vector consists of m elements, one for each detector. The detection locations, i.e. call locations, are taken as latent variables and are denoted by $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_{n_c})$. Each $\mathbf{u}_i$ comprise Cartesian coordinates within the survey region S.

Given that call locations of the same individual are not independent and the number of unique individuals detected is not known, evaluating a distribution for the number of unique calls observed is infeasible. Following the approach of Borchers and Efford (2008), the authors construct a likelihood that is conditional on n_c with model parameters, γ and ϕ, relating to the SS and TOA processes. The likelihood is expressed in terms of the marginal distribution of n_c and the conditional distribution of the observed data given the model parameters:

$$L_n(\boldsymbol{\gamma}, \boldsymbol{\phi}) = f(\mathbf{Y}, \mathbf{S}, \mathbf{Z} \mid n_c),$$

where the joint density of the observed data is calculated by integrating over all possible values of the unobserved latent call locations, $\mathbf{U}$:

$$
\begin{aligned}
L_n(\boldsymbol{\gamma}, \boldsymbol{\phi}) &= \int_{S^{n_c}} f(\mathbf{Y}, \mathbf{U}, \mathbf{S}, \mathbf{Z} \mid n_c) d\boldsymbol{U} \\
&= \int_{S^{x_c}} f(\mathbf{Y}, \mathbf{S}, \mathbf{Z} \mid n_c) f\left(\boldsymbol{U} \mid n_c\right) d\boldsymbol{U} \\
&= \int_{S^{n_c}} f\left(\boldsymbol{S}, \boldsymbol{Z} \mid \boldsymbol{Y}, \boldsymbol{U}, n_c\right) f\left(\boldsymbol{Y} \mid \boldsymbol{U}, n_c\right) f\left(\boldsymbol{U} \mid n_c\right) d\boldsymbol{U}.
\end{aligned}
$$

Marginalization over $\mathbf{U}$ is performed given that the capture histories, signal strength and time of arrival observations all depend on the call locations which are considered latent variables. Furthermore, the SS and TOA processes are assumed as independent, thus leading to the final form of the conditional likelihood in Equation.

$$
L_n(\gamma, \boldsymbol{\phi}) = \int_{S^{n_c}} f\left(\boldsymbol{S} \mid \mathbf{Y}, \boldsymbol{U}, n_c\right) f\left(\boldsymbol{Z} \mid \mathbf{Y}, \boldsymbol{U}, n_c\right) f\left(\boldsymbol{Y} \mid \boldsymbol{X}, n_c\right) f\left(\boldsymbol{U} \mid n_c\right) d\boldsymbol{U}.
$$

The above likelihood is intractable due to its high dimensionality and as the dependence between calls is not known, the joint density of call locations cannot be specified (Stevenson et al., 2015; Juodakis et al., 2021). To address the complexity of the likelihood function, Stevenson et al. (2015) construct a simplified likelihood by assuming independence between call locations. The joint probability density of call locations is thus simplified to a product of the individual probability densities: $f\left(\boldsymbol{U} \mid n_c\right) = \prod_{i=1}^{n_c} f\left(\boldsymbol{u}_i\right)$. Consequently, the $2n_c$-dimensional integral can be separated into a product of n_c integrals of 2 dimensions:

$$
L_s(\boldsymbol{\gamma}, \boldsymbol{\phi}) = \prod_{i=1}^{n_e} \int_S f_s(\boldsymbol{s}_i \mid \boldsymbol{y}_i, \boldsymbol{u}_i) f_z(z_i \mid \boldsymbol{y}_i, \boldsymbol{u}_i) f_y(\boldsymbol{y}_i \mid \boldsymbol{u}_i) f_u(\boldsymbol{u}_i) d\boldsymbol{u}_i. \tag{20}
$$

A.1 Signal strength

The first component of the simplified likelihood, hereafter referred to as the likelihood, describes the joint probability of the observed signal strengths given the capture histories and the latent call locations:

$$
f_s(\boldsymbol{s}_i \mid \boldsymbol{y}_i, \boldsymbol{u}_i) = \prod_{k=1}^{m} f_s(s_{ik} \mid y_{ik}, \boldsymbol{u}_i),
$$

Only if the signal strength is above the threshold, c, and thus $y_{ik} = 1$ is s_{ik} observed. Consequently, when $s_{ij} < c$ (and $y_{ik} = 0$), the observation is dropped, and the term $f(s_{ik}|y_{ik}, \mathbf{u}_i)$ is set to one. The probability density of an observed signal strength is then defined to be a truncated Gaussian distribution:

$$f_s\left(s_{ik} \mid y_{ik}=1, \boldsymbol{u}_i\right) = \frac{1}{\sigma_s} f_n\left(\frac{s_{ik} - E\left(s_{ik} \mid \boldsymbol{u}_i\right)}{\sigma_s}\right)\left(1 - \Phi\left(\frac{c - E\left(s_{ik} \mid \boldsymbol{u}_i\right)}{\sigma_s}\right)\right)^{-1},$$

where the probability and cumulative density functions of the standard normal distribution are denoted by f_n and Φ, respectively. In order to account for measurement error, $s_{ik} \mid \boldsymbol{u}_i$ is assumed to follow a normal distribution with standard deviation, σ_s:

$$s_{ij} \mid \boldsymbol{x}_i \sim N\left(E\left(s_{ik} \mid \boldsymbol{x}_i\right), \sigma_s\right).$$

Intuitively, the expectation of s_{ik} given the location $\mathbf{u}_i$ can be any consistently decreasing function of distance. The further away a call location, the weaker the received signal strength. Stevenson et al. (2015) specify the expectation as a generalized linear model with a distance covariate, $d_k(\mathbf{u}_i)$, describing the distance between the microphone k and call location i,

$$E\left(s_{ik} \mid \boldsymbol{u}_i\right) = h^{-1}\left(\beta_{0s} - \beta_{1s} d_j\left(\boldsymbol{u}_i\right)\right),$$

where β_{0s}, β_{1s} are the associated regression coefficients and h^{-1} is the inverse link function.

A.2 Time of arrival

To formulate the PDF describing the data generating process of the time-of-arrival observations, it is important to note that the relative position of a call can only be deduced from the differences in precise arrival times as opposed to a single detection time (Stevenson et al., 2015). Only locations of calls that were recorded at more than two microphones can be informed by time-of-arrival data. Given that $b_i = \sum_{k=1}^{m} y_{ik}$ (the number of detectors that captured the i^{th} call) and the TOA data is only dependent on the capture histories through d_i, the following holds true:

$$f_z\left(\boldsymbol{z}_i \mid y_i, \boldsymbol{u}_i\right) \equiv f_z\left(\boldsymbol{z}_i \mid d_i, \boldsymbol{u}_i\right),$$

The joint probability density of the TOA observations is thus set to one if the number of microphones that detected call i is less than two (i.e. $d_i = 1$). When $d_i > 1$ uncertainty in the observed TOA data is assumed to follow a Gaussian distribution with parameter, σ_z, leading to the following PDF:

$$f_z\left(z_i \mid j_i > 1, \boldsymbol{u}_i\right) = \frac{\left(2\pi\sigma_z^2\right)^{(1-j_i)/2}}{2T\sqrt{j_i}} \exp\left(\sum_{\{j:y_{ij}=1\}} \frac{\left(\delta_{ij}\left(\boldsymbol{u}_i\right) - \bar{\delta}_i\right)^2}{-2\sigma_t^2}\right),$$

where the expected call time is denoted by, δ_{ij} defined as: $z_{ij} - d_j\left(\boldsymbol{x}_i\right)/v$ where v is the speed of sound, and $\bar{\delta}_i$ is the corresponding mean.

A.3 Detection function

The probability of detection is directly related to the probability of signal strength surpassing the threshold, such that the detection function, $g(d;\gamma)$, describing the probability of detection given the distance between call location and microphone, d, is defined as:

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

Consequently, the probability of a single observation given the unobserved call location can be described as follows:

$$f_y\left(y_{ij} \mid \boldsymbol{u}_i\right) = \begin{cases} g\left(d_j\left(\boldsymbol{u}_i\right)\gamma\right) & y_{ij} = 1 \\ 1 - g\left(d_j\left(\boldsymbol{u}_i\right)\gamma\right) & y_{ij} = 0 \end{cases}$$

and the resulting joint mass assuming independence between detected calls,

$$f_y\left(\boldsymbol{y}_i \mid \boldsymbol{u}_i\right) = \frac{\prod_{j=1}^{J} f_y\left(y_{ij} \mid \boldsymbol{u}_i\right)}{p.\left(\boldsymbol{u}_i;\boldsymbol{\gamma}\right)},$$

incorporates in the denominator the probability of detecting a call produced at $\mathbf{u}_i$ at all, since capture histories are observed conditional on detection.

A.4 Call locations

Animal locations are assumed to be realisations of a homogenous Poisson point process with constant intensity across space. However, the intensity of detected calls, and thus the point process for call locations of detected animals, is not uniform throughout the survey region. Instead, intensity of detected calls varies across the survey regions, peaking at locations nearest to the detectors. The point process is thus filtered by the probability of detection such that the PDF for a location, u_i, resulting from this point process, takes on the following from:

$$f_u\left(\boldsymbol{u}_i\right) \propto p\left(\boldsymbol{u}_i ; \boldsymbol{\gamma}\right)=\frac{p .\left(\boldsymbol{u}_i ; \boldsymbol{\gamma}\right)}{a(\boldsymbol{\gamma})},$$

where the effective sampling area (ESA), denoted by $a(\boldsymbol{\gamma})$, is defined as, $\int_A p.\left(\boldsymbol{u}_i ; \boldsymbol{\gamma}\right) du$ and thus acts to normalise the intensity surface.

A.5 Estimation

The log of the simplified likelihood is then maximised in order to obtain parameters estimates of γ and ϕ:

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

and by using a Horvitz-Thompson-like estimator, calling density (D_c), is obtained using the estimate of γ:

$$\widehat{D}_c=\frac{n_c}{a(\widehat{\gamma}) T}$$

Density of calling individuals, D_a, is then estimated simply be dividing with the species call rate, c_r. The call rate can be known beforehand or estimated from independently collected data.

A.6 Bootstrap procedure

Standard errors and confidence intervals are estimated via a parametric bootstrap procedure using the parameters estimates and call rate data. The procedure uses the model estimates and call rate data to simulate the underlying data generating process and generate new data. From this data, the same model is refitted, and new parameter estimates are obtained. The procedure is implemented N times, resulting in N estimates of each parameter. Appropriate variance estimates are thus obtained by taking the standard error (SE) of the sampling distributions of each parameter and using the SE to calculate confidence intervals based on a normal approximation. See the exact steps taken in Stevenson et al., 2015.

B ASCR-AD Model Formulation

Consider a survey of duration T with m acoustic recording devices which are situated at known Cartesian coordinates, $\mathbf{X} = (x_1, \ldots, x_k)$, within the survey region $S \subset \mathbb{R}^2$. The survey region, in line with SCR, is defined as the set of all locations where calling individuals (or groups) can be detected by the detector array (Dawson and Efford, 2009; Efford, Dawson and Borchers, 2009; Stevenson et al., 2015; Stevenson et al., 2021).

When estimating calling density, it is necessary to identify the number of unique calls (i.e. n_c). When individuals (or groups) serve as the detection unit, two additional requirements come into play. Note, the term individual(s) will be used, but the same applies to group(s). First, calls should be able to be matched to specific individuals. Second, the assumption is made that the individuals remain stationary throughout the survey. The capture histories are thus indexed by the calling individual, such that $y_{ijk} = 1$ if the i^{th} call, $i \in \{1, \ldots, c_i\}$, of the j^{th} individual, $j \in \{1, \ldots, n\}$, was detected at the k^{th} microphone, $k \in \{1, \ldots, m\}$ and zero otherwise. The same modification applies to any supplementary information also recorded by the detectors, here TOA is included, and z_i from the previous section gains an additional index for the j^{th} individual, z_{ijk}.

The observed data comprise the number of detected animals (n), the number of calls produced by each individual, $\mathbf{c} = (c_1, \ldots, c_n)$, the individuals' capture histories $\mathbf{Y} = (Y_1, \ldots, Y_n)$ and the TOA measurements ($\mathbf{Z} = (Z_1, \ldots, Z_n)$). Each $\mathbf{Y}_1$ is a capture history matrix consisting of row vectors $\mathbf{y_{ij}}$ of length m whose elements are the y_ijks. The same applies to each, Z_i where the elements of the row vectors of each block row $\mathbf{Z_i}$ matrix are the z_{ij}s. The physical location of each animal whose calls were detected are now the latent variables and will also be referred to with, $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$ where each u_i' are Cartesian coordinates.

The model likelihood, with parameters contained in $\boldsymbol{\theta}$, takes the following form:

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= f\left(\mathbf{n}, \mathbf{c}, \mathbf{Y}, \mathbf{Z}\right) && (21)\\
&= f\left(\mathbf{n}\right) \prod_{j=1}^{n} \int_S f\left(\mathbf{c}_i, \mathbf{Y}_i, \mathbf{Z}_i, \mathbf{u}_i\right) d\mathbf{u}_i, && (22)\\
&= f\left(\mathbf{n}\right) \prod_{j=1}^{n} \int_S f\left(\mathbf{Z}_i \mid \mathbf{Y}_i, \mathbf{c}_i, \mathbf{u}_i\right) f\left(\mathbf{Y}_i \mid \mathbf{c}_i, \mathbf{u}_i\right) f\left(\mathbf{c}_i \mid \mathbf{u}_i\right) f\left(\mathbf{u}_i\right) d\mathbf{u}_i. && (23)
\end{aligned}
$$

In the above, it assumed that individuals' locations are independent of one another, but there is dependence between call locations both within and between individuals. The latent variables are marginalised out of the likelihood by integrating over all possible u_i's within the survey region.

B.1 The Detection function

The detection process is modelled using a hazard half-normal function:

$$g_c(d(\mathbf{x}_k, \mathbf{s})) = 1 - \exp\left\{-\lambda_0 \exp\left(\frac{-d(\mathbf{x}_k, \mathbf{u})^2}{2\sigma^2}\right)\right\}$$

where $d(\mathbf{x}_k, \mathbf{u})$ is the Euclidean distance between the k^{th} detector and animal location $\mathbf{u}$. The above detection function is the call detection function and describes the probability a call originating from $\mathbf{u}$ is detected by a microphone at distance d. This, together with assuming detections of the same calls are independent (as in ASCR-CD), leads to the formulation of the probability that a call produced at $\mathbf{u}$ is detected at all:

$$p_c(\boldsymbol{u}) = 1 - \prod_{k=1}^{m} \left[1 - g_c\left\{d\left(\boldsymbol{x}_k, \boldsymbol{u}\right)\right\}\right]$$

B.2 Call and animal detection probability

The probability mass function for the number of detected calls, c_i, of the i^{th} individual conditional on the latent animal location is calculated as follows:

$$f\left(c_i \mid \mathbf{u}_i\right) = \sum_{C_i=1}^{\infty} f\left(c_i \mid C_i, \mathbf{u}_i\right) f\left(C_i\right),$$

where C_i is the unobserved total number of calls produced by individual i. The PMF for c_i conditional on C_i and $\mathbf{u}_i$ is described by a binomial distribution truncated at zero with parameter, $p_c(\mathbf{u})$. It is assumed that the call production process can be described by Poisson distribution with parameter $c_r T$, i.e. $C_i \sim \text{Poisson}(c_r T)$, although alternative distributions can be assumed for the call production process. The call rate, c_r, is taken as the average number of vocalisations made per individual and unit of time. Logically, the call detection process is a filtered version of the latter and thus, the above sum can be expressed as a zero truncated Poisson process:

$$\left(c_i \mid \mathbf{u}_i\right) \sim \text{ ZTPoisson}\left\{c_r T p_c\left(\mathbf{u}_i\right)\right\}$$

Following from this, the probability of detecting an animal at position $\mathbf{u}$, i.e. capturing at least one of its calls using any of the detectors, is equivalent to the inverse probability of none of its vocalisations being detected:

$$p(\mathbf{u}) = 1 - f(c = 0|\mathbf{u})$$

and the effective sampling area is defined as $\int_S p(\mathbf{u})d\mathbf{u}$.

B.3 Capture histories

The PMF for the capture history matrices, conditional on the latent locations, is equivalent to that of ASCR-CD as the call capture histories are also realisations of a Bernoulli trial with probability of success governed by the detection function:

$$\mathbf{y}_{ij}|\mathbf{u} \sim \text{Bernoulli}[g_c(d(\mathbf{x}_k, \mathbf{u}_i))].$$

Additionally, as with ASCR-CD, the observed capture histories are zero-truncated by incorporating the probability of detecting the i^{th} call produced at $\mathbf{u}_i$ in the denominator:

$$f\left(\mathbf{Y}_i \mid c_i, \mathbf{u}_i\right) = \frac{\prod_{j=1}^{c_i} f\left(\mathbf{y}_{ij} \mid \mathbf{u}_i\right)}{p_c(\mathbf{u}_i)}.$$

B.4 Animal locations and number of animals detected

It is assumed that animal locations are realisations of an inhomogeneous Poisson point process (PPP) with intensity parameter, $D(u)$, describing the density of individuals at location $\mathbf{u} \in \mathbf{S}$. Locations of detected animals are realisations of the same process, but one that is thinned by the probability of detecting an animal at $\mathbf{u}$, i.e. $p(\mathbf{u})$. Accordingly, the thinned PPP has an intensity function equal to $D(\mathbf{u})p(\mathbf{u})$ and the expect number of points generated by this process, i.e. the number of animals detected, is equal to:

$$f\left(n\right) = \frac{\left\{\int_S D(\mathbf{u})p(\mathbf{u})d\mathbf{u}\right\}^n \exp\left\{-\int_S D(\mathbf{u})p(\mathbf{u})d\mathbf{u}\right\}}{n!},$$

and the PDF of the latent location variables is equal to the normalised intensity function of the filtered PPP:

$$f\left(\mathbf{s}_i\right)=\frac{D\left(\mathbf{u}_i\right) p\left(\mathbf{u}_i\right)}{\int_S D(\mathbf{u}) p(\mathbf{u}) d \mathbf{u}}.$$

B.5 Animal density

Spatial heterogeneity in animal density can be modelled through a log link function:

$$\log\{D(u)\}=\beta_0+\sum_{q=1}^{Q} \beta_q x_q(u)$$

where $x_q(\mathbf{u})$ is any spatial covariate measured at location $\mathbf{u}$. Associated regression coefficients are represented by β_0 and β_q.

B.6 Time of arrival

As before, $f\left(\mathbf{Z}_i \mid \mathbf{Y}_i, \mathbf{c}_i, \mathbf{u}_i\right)$, describes the conditional distribution of the TOA process and its formulation is exactly as in ASCR-CD. The only difference being that the PDF is now defined for the i^{th} call of the j^{th} individual:

$$f\left(\boldsymbol{Z}_i \mid \mathbf{Y}_i, c_i, \boldsymbol{u}_i\right)=\prod_{j=1}^{c_i} f\left(\boldsymbol{z}_{ij} \mid \boldsymbol{y}_{ij}, \boldsymbol{u}_i\right).$$

B.7 Estimation

The model parameters contained in $\boldsymbol{\theta}$ consist of the regression coefficients, $\boldsymbol{\beta}$, the parameters of the HH detection function, λ_0 and σ, the call rate c_r and standard deviation of the TOA measurements, σ_z. Estimates are obtained via maximisation of the log likelihood:

$$(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\lambda_0}}, \widehat{\boldsymbol{\sigma}}, \widehat{\boldsymbol{c_r}}, \widehat{\boldsymbol{\sigma_z}})=\arg \max_{\boldsymbol{\theta}} \log \left(L_s(\boldsymbol{\theta})\right),$$

and variance estimates and confidence intervals are obtained as in standard maximum likelihood analysis.