Masters Dissertation

Acoustic Spatial Capture-Recapture

Author

Ané Cloete

Supervisor

Dr David Borchers

I hereby certify that this dissertation, which is approximately 4000 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.

July 1, 2023

Contents

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 mark-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). Mark-recapture (MR) methods generate data in the form of capture-histories, which give information on the probability of detection and is an analysis framework for a diverse array of survey techniques, such as physical capture, camera traps, and DNA 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.

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 that can be picked up by passive acoustic recording devices, 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 (source) 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 capturing is suitable: (1) greater range of detectability for animals which produce loud or regular vocalisations, (2) light conditions do not dictate surveying, and (3) automated data collection and processing in passive acoustics facilitates the analysis of vast quantities of data.

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 mark-recapture methods account for this phenomenon by explicitly modelling the observation process through a detection function. For acoustic surveys, locations of recorded calls are not typically observed and, although supplementary information recorded can be used to estimate the call location, estimates of the distances do not all have high precision (Stevenson et al., 2015 ; Stevenson et al., 2021). Consequently, distance sampling does not present an ideal framework for analysing acoustic surveys, as the assumption of accurate distances is likely to be violated. Efford et al. (2009) were the first to suggest using SECR methods for analysing data obtained via an acoustic survey with an array of microphones.

Since their seminal work, acoustic SECR (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). However, the adoption of aSCR (and new ecological survey methods in general) in the field has been relatively slow, and this can be attributed to several reasons. 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. Another contributing factor 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 inertia associated with transitioning to a new approach, such as acoustic spatial capture-recapture, can be challenging to overcome without proper guidance and support.

In light of these challenges, the primary objective of this report is to provide a high-level overview of acoustic spatial capture-recapture and to identify key elements essential for effective learning and application. The ultimate goal is to develop a set of interactive tutorials that bridge the gap in educational resources and empower practitioners to confidently incorporate acoustic spatial capture-recapture into their ecological studies and conservation efforts.

2 Acoustic spatial Capture-Recapture

In the realm of spatial capture-recapture (SCR), 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). 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 trap and on which occasion, such that an individual's capture (or encounter) history in space and over time is recorded. This repeated temporal sampling is what enables SCR methods to estimate detection probability, and the spatial features inherent in the survey design and the detections allow a sensible estimate of density.

In acoustic surveying, only a single survey occasion is required as calls can be detected at multiple recording devices (detectors), resulting in simultaneous captures and recaptures (Dawson and Efford, 2009; Efford et al., 2009). As a result, acoustic capture histories are typically not indexed by time and describe whether a call was detected at each microphone, i.e. call capture histories (Stevenson et al., 2015). However, vocalisations need not be the detection unit. 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 microphone, translating to, for example, an individual's call capture history (Stevenson et al., 2021). The analysis that follows and the resulting estimate of density is dependent on the target, so it is essential to determine the detection unit beforehand (Figure 1).

When the detection unit is individual animals, the density estimate is that of animal density, hence the name ASCR-AD, where AD stands for animal density (Figure 1). A similar analysis is performed when the detection unit is groups of animals, which results in group density (ASCR-GD). For the latter, the model formulation is the same as ASCR-AD, as groups can be viewed theoretically as individuals, thus leading to the same analysis but a different interpretation of the resulting density estimate. However, the analysis may be slightly more complicated, as it may not be realistic to assume that all individuals or groups will call within a single survey occasion. Additional parameters are then required to either convert to absolute group density or estimate it directly, the latter of which requires tracking the groups across multiple occasions. Alternatively, choosing groups within a single survey occasion requires estimating the proportion of time groups call within the survey occasion, so the resulting density of vocalising groups within the survey period is converted to group density. Lastly, calling density (ASCR-CD) is estimated when cues are considered the detection unit and density estimation requires independently estimating the calling rate of the target species (Figure 1). It is important to note here that in all cases, the ultimate estimate of density is of calling animals and not the overall population. Ad hoc conversion to population density requires knowing or estimating the fraction of calling individuals (Stevenson et al., 2015).

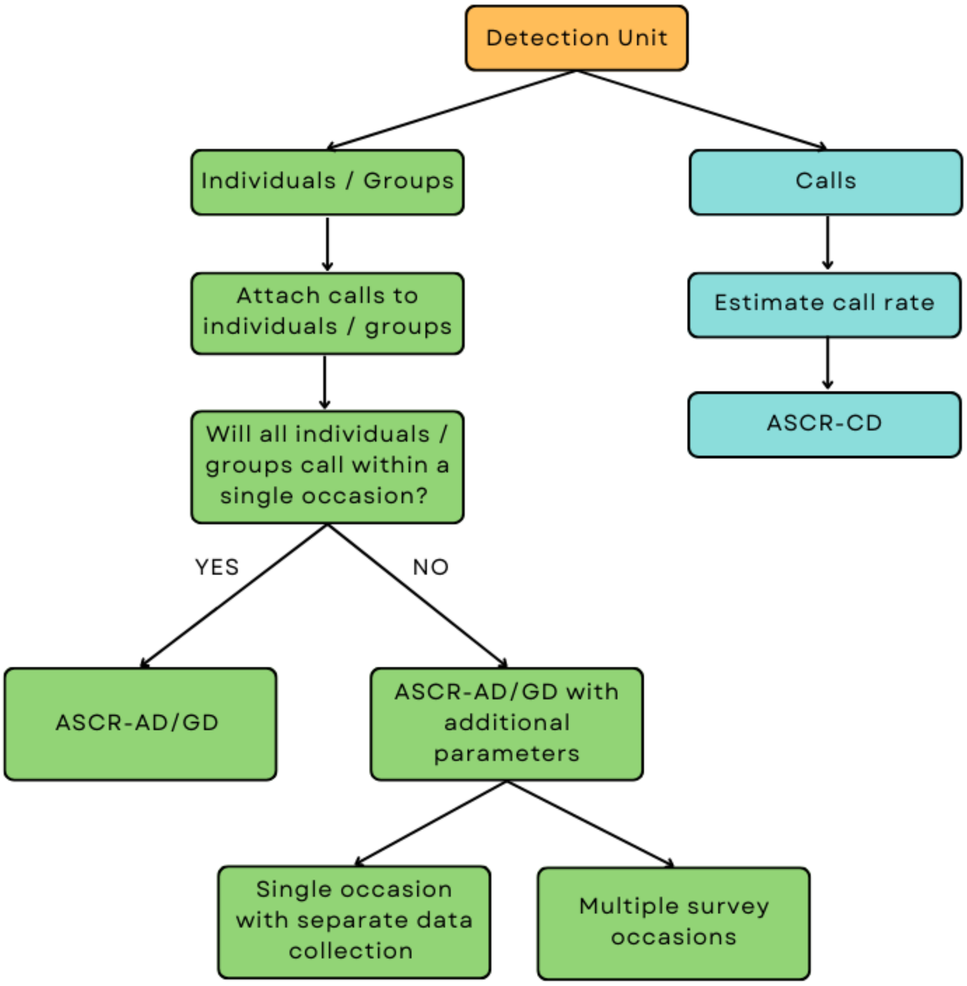

Figure 1: Illustration of the nomenclature and generation of Quarter Degree Grid Cells (QDGC) from

In standard SCR, detections are conditional on the animal's activity or home range centre, which are unobserved and thus treated as latent variables. In acoustic surveying, the latent variables become the animal locations (ASCR-AD) or the call locations (ASCR-CD). Stevenson et al. (2021) developed the model formulation for ASCR-AD for two reasons. Firstly, the additional data collection of ASCR-CD may be laborious and expensive. Secondly, the likelihood on which the ASCR-CD model is based is built upon the assumption that call locations are realisations of a homogeneous Poisson point process which means that call locations are considered to be independent of each other (Stevenson et al., 2015; Stevenson et al., 2021). In reality, call locations are almost certainly not independent; calls from the same individuals obviously originate from related locations. Directly modelling animal locations allows the ASCR-AD model to accommodate inhomogeneous animal density surfaces and direct estimation of animal density, but this requires that calls can be linked to individuals or groups (Figure 1).

When using calls as the detection unit and, consequently, call locations as the latent variables, assuming dependence results in an intractable likelihood, and so Stevenson et al. (2015) construct a simplified likelihood for ASCR-CD based on the assumption of independence. The resulting density estimates are shown to be unbiased despite the misspecification, but variance estimates require a bootstrap procedure. Additionally, since it is not an actual likelihood, other likelihood-based tools (such as AIC) cannot be used in subsequent analysis. Regardless, 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 (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 et al., 2009; Stevenson 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. Logically, 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 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. Considering that this project aims to unify educational resources and enhance the accessibility of ASCR literature, model and survey design notation will be presented (as much as possible) following the widely used textbook on spatial capture-recapture by Royle et al. (2013). The textbook also includes a section on acoustic SCR (see Chapter 9). At the time, the only published model for ASCR was that described in Efford et al. (2009) and consequently, the models presented in this report developed by Stevenson et al. (2015) and Stevenson et al. (2021) are not included.

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 et al., 2009; Stevenson et al., 2015; Stevenson et al., 2021).

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 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.

Due to the lack of independence among call locations and the unknown number of unique individuals detected, specifying a distribution for the number of unique calls observed is not possible. 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, respectively. However, the conditional likelihood is intractable due to its high dimensionality and the unknown dependence between call locations (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:

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

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

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

and 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, $E\left(s_{ik} \mid \boldsymbol{u}_i\right)$, is modelled using a link function with the distance between microphone and call location as a covariate.

The second component of the likelihood describing the TOA process, is defined as follows:

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

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 PDF for call location is defined piecewise. This is because only the differences between arrival times, such as when the call is detected at multiple microphones, 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, $g(d;\gamma)$, which describes the probability of detection given the distance between call location and microphone, d, is thus defined as:

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

Consequently, the conditional distribution of a call capture history y_{ik} is:

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

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

$$f\left(\boldsymbol{y}_i \mid \boldsymbol{u}_i\right)=\frac{\prod_{k=1}^{K} f\left(y_{ik} \mid \boldsymbol{u}_i\right)}{p\left(\boldsymbol{u}_i;\boldsymbol{\gamma}\right)},$$

Given that only capture histories where at least one detection is made is 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 $\left(p\left(\boldsymbol{u}_i;\boldsymbol{\gamma}\right)\right)$.

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

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

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 as the normalising constant. Each component of the likelihood in Equation 5 has now been described. It is important to note if signal strength data are not collected, the detection function can be specified as the typical hazard half-normal function, resulting in a more familiar SCR analysis with model parameters of g_0 and σ.

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:

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

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).

2.2 Animal Density (ASCR-AD)

When estimating calling density, it is required that the number of unique calls is identifiable (i.e. n_c). When individuals (or groups) are the detection unit, two additional requirements are that calls can be matched to individuals and individuals do not move during the survey. Hereafter, the term individual(s) or animal(s) will be used, but the same applies to group(s). 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 physical location of each animal whose calls were detected are now the latent variables and will also be referred to with $\mathbf{U} = (\mathbf{u}_1, \ldots, \mathbf{u}_n)$. Furthermore, the locations of the m microphones are denoted by $\mathbf{x}_k$ and are also Cartesian coordinates.

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

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= f\left(\mathbf{n}, \mathbf{c}, \mathbf{Y}, \mathbf{Z}\right) & (2)\\
&= f\left(\mathbf{n}\right) \prod_{j=1}^{n} \int_S f\left(\mathbf{c}_i, \mathbf{Y}_i, \mathbf{Z}_i, \mathbf{u}_i\right) d\mathbf{u}_i, & (3)\\
&= f\left(\mathbf{n}\right) \prod_{j=1}^{n} \int_S f\left(\mathbf{Z}_i \mid \mathbf{Y}_i, \mathbf{c}_i, \mathbf{u}_i\right) f\left(\mathbf{Y}_i \mid \mathbf{c}_i, \mathbf{u}_i\right) f\left(\mathbf{c}_i \mid \mathbf{u}_i\right) f\left(\mathbf{u}_i\right) d\mathbf{u}_i. & (4)
\end{aligned}
$$

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. See Appendix B for a detailed model formulation.

The detection process is modelled using a hazard half-normal (HH) 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 $\mathbf{u}$. As detections of the same calls are independent across microphones (as in ASCR-CD), the probability that a call produced at $\mathbf{u}$ is detected at all:

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

The probability mass function for the number of detected calls, c_i, of the $i^t h$ 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. 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:

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

The PMF for the capture history matrices, conditional on the latent locations, is equivalent to that of call capture histories in ASCR-CD as the individual 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 $\mathbf{u}_i$:

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

It is assumed that animal locations are realisations of an inhomogeneous 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:

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

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.

As before, $f(\mathbf{Z}_i \mid \mathbf{Y}_i, \mathbf{c}_i, \mathbf{u}_i)$, 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:

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

Finally, the model parameters contained in $\boldsymbol{\theta}$ thus 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.

2.3 Group Density (ASCR-GD)

3 Discussion

This report assumes that the reader possesses an in-depth understanding of mark-capture concepts and SCR. However, the tutorial's intended audience is assumed to have some knowledge of MR concepts but may be largely unfamiliar with SCR, thus necessitating their introduction in the context of acoustic sampling. To begin with, the learner must grasp the relationship between detection probability and distance, as it is fundamental in all SCR models. The tutorial will elucidate this concept by introducing MR concepts and demonstrating how one can crudely estimate abundance using average detection probability from the resulting data. It will become apparent that assuming an equal probability of capture across the detector array is inappropriate, and the tutorial will showcase how a histogram of distances can provide valuable insights into the trend of detection probability. Finally, the idea of the effective sampling area will be introduced.

Chapter 1: Setting the scene

1. Introduction to MR concepts using acoustic sampling
2. Capture histories and calculation of detection probability
3. Heterogeneity in call detection probability
4. Distance dependent detection
5. Effective sampling area

Chapter One serves as a natural introduction to SCR; thus, the SCR framework will be presented in the following chapter. A thorough explanation of non-acoustic SCR will not be included; I believe only a brief introduction in the context of acoustic surveying is necessary to foster an understanding of SCR and aid in formulating the ASCR methodology. Additionally, plenty of online literature and educational resources are available on the topic of SCR (e.g. Royle et al., 2013).

Chapter 2: A Brief Introduction to SCR

1. Detection function
2. Effective sampling area and buffer zone / mask
3. Density model

The learner will understand that 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. Following this, Chapter Three will explain how, depending on the survey's detection unit, the resulting density estimate is different and expand upon the flow diagram presented in Figure 1. Additionally, the recording and use of supplementary data in acoustic surveying will be mentioned.

> **Chapter 3: Overview of ASCR**
>
> 1. Detection units and their different analysis
> 2. Supplementary data

The remaining chapters (4, 5, and 6) will be dedicated to each analysis method outlined in this report. Chapter 4 will centre on ASCR-CD, commencing with an explanation of the specific detection function used and an exploration of how density is modelled. To facilitate comprehension, a straightforward analysis in R will be introduced, allowing the learner to navigate the analysis process step by step. Subsequent analyses will increase in complexity through a two-step process. First, the supplementary data will be incorporated. Then, the signal strength detection function introduced in Section 2.1 will be applied.

> **Chapter 4: ASCR-CD**
>
> 1. Introduction
> 2. Simple analysis
> 3. Analysis with additional data
> 4. Advanced analysis (signal strength detection function?)

Chapter five will present ASCR-AD in a format consistent with the previous chapter while highlighting the key differences compared to ASCR-CD. Chapter six will then explain how the ASCR-AD analysis and interpretation of results slightly differ when the detection unit is groups of animals and what must be done to end with a sensible estimate of density ultimately.

> **Chapter 5: ASCR-AD**
>
> 1. Introduction
> 2. Simple analysis
> 3. Analysis with additional data

Chapter 6: ASCR-GD

1. Introduction
2. Example analysis

This section provided an overview of the intended structure of the tutorial, highlighting its broader scope beyond that of the dissertation. Only chapters one and four will be developed to meet the thesis requirements within the given time frame. The tutorial is designed to progress naturally from chapter to chapter; therefore, learners are expected to work through them sequentially. In other words, chapter four should not be viewed as a standalone chapter. The ultimate goal is to develop the entire tutorial. The R package implemented throughout the tutorial, `acre`, can implement other methods not presented in this tutorial and include other data sources, such as estimated bearings and distances in analysis.

References

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

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

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

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

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

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

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

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

P. S. Hammond. Mark–recapture. In *Encyclopedia of marine mammals*, pages 705–709. Elsevier, 2009.

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

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

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

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

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

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

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

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

B. C. Stevenson, P. van Dam-Bates, C. K. Young, and J. Measey. 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, 2021.

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

L. Thomas, A. Jaramillo-Legorreta, G. Cardenas-Hinojosa, E. Nieto-Garcia, L. Rojas-Bracho, J. M. Ver Hoef, J. Moore, B. Taylor, J. Barlow, and N. Tregenza. 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, 2017.

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

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

A ASCR-CD Model formulation

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 et al., 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 are not independent and the number of unique individuals detected is not known, it is not possible to specify a distribution for the number of unique calls detected. 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:

$$
L_s(\boldsymbol{\gamma}, \boldsymbol{\phi}) = \prod_{i=1}^{n_e} \int_S f(\boldsymbol{s}_i \mid \boldsymbol{y}_i, \boldsymbol{u}_i) f(z_i \mid \boldsymbol{y}_i, \boldsymbol{u}_i) f(\boldsymbol{y}_i \mid \boldsymbol{u}_i) f(\boldsymbol{u}_i) d\boldsymbol{u}_i. \tag{5}
$$

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:

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

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:

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

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 of link function (package).

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:

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

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:

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

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:

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

and the resulting joint mass assuming independence between detected calls,

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

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:

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

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.

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 et al., 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 $y_i jk$s. 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:

$$
\begin{aligned}
L(\boldsymbol{\theta}) &= f\left(\mathbf{n}, \mathbf{c}, \mathbf{Y}, \mathbf{Z}\right) && (6)\\
&= f\left(\mathbf{n}\right) \prod_{j=1}^{n} \int_S f\left(\mathbf{c}_i, \mathbf{Y}_i, \mathbf{Z}_i, \mathbf{u}_i\right) d\mathbf{u}_i, && (7)\\
&= f\left(\mathbf{n}\right) \prod_{j=1}^{n} \int_S f\left(\mathbf{Z}_i \mid \mathbf{Y}_i, \mathbf{c}_i, \mathbf{u}_i\right) f\left(\mathbf{Y}_i \mid \mathbf{c}_i, \mathbf{u}_i\right) f\left(\mathbf{c}_i \mid \mathbf{u}_i\right) f\left(\mathbf{u}_i\right) d\mathbf{u}_i. && (8)
\end{aligned}
$$

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:

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

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

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

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, $f(\mathbf{Z}_i \mid \mathbf{Y}_i, \mathbf{c}_i, \mathbf{u}_i)$, 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:

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

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:

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

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