THE UNIVERSITY *of* EDINBURGH

Efficient Model Fitting Approaches for Estimating Abundance and Demographic Rates for Marked and Unmarked Populations

Riki Herliansyah

Doctor of Philosophy
University of Edinburgh
April 2023

Declaration

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text.

(*Riki Herliansyah*)

To my mom Nor Halipah and my dad Juhriansyah, the most important figures of my life.

Acknowledgements

I want to start by thanking Ruth King and Stuart King, my supervisors, for their commitment to assisting me during this journey, especially during the challenging times since the pandemic started. The amount of time they dedicated was extraordinary, from quick responses to my emails to accommodating almost every meeting's requests. I knew that I often had problems with sophisticated mathematical equations but they were always able to illustrate it for me in a way that helped me easily understand the problems, and I am very grateful for that. In addition to offering me work-related comments and support, they often encouraged me to attend a variety of scientific conferences that were beneficial to me: (i) to disseminate my works for broader audience while teaching me how to create and simplify complicated works into presentable results; (ii) to know more people and their works in relevant subjects. Finally, I would like to thank them for their dedication in proofreading my works and providing incredibly detailed feedback.

I would like to thank my mom and dad and all family members for walking this path with me, supporting me mentally and being a beacon of light when I needed it. They have always believed that I will be doing well, and I have always been. I am grateful that Edinburgh has a wonderful Indonesian community that served as a support system for everyone. We assisted one another in times of need by sharing foods, helping one another move into and out of accommodations, and easing my homesickness. I want to thank my future partner for her assistance in every manner, including her patience in helping me with work-related challenges and in waiting for the completion of my PhD.

Last but not least, I want to express my gratitude to the Indonesia Endowment

Fund for Education (Lembaga Pengelola Dana Pendidikan—LPDP), Ministry of Finance Republic of Indonesia, for providing financial support for my PhD. The consistent and complete financial support I received from pre-departures through to the completion of my PhD journey means a lot to me. As a token of my gratitude for your support and for all the persons I named above, I am presenting my thesis.

Lay Summary

New statistical estimation procedures have been developed for analyzing two types of data commonly collected on animal populations to yield estimates of animal abundances and/or population demographic rate parameters, e.g., survival probabilities. Data collection to obtain information on such demographic/population aspects often involves repeatedly sampling the population over a series of sampling occasions. Observations at each sampling occasion often take the form of physical captures and/or visual sightings; and more recently motion-sensor camera images of animals. The associated data collected by such means are often referred to as capture-recapture data. In many studies individuals may be uniquely identifiable (either through natural markings or via applying some physical mark such as a tag/ring). However, this is not always the case, particularly with camera images, so that only counts relating to the number of encounters/sightings may be available when individuals are not uniquely identifiable.

Within such studies, due to imperfect detection, it is useful to separate out the different processes relating to (i) the true state of the ecosystem from (ii) the associated (imperfect) detection or observation of that process. The system process is of primary interest, as this describes the ecological processes of interest and hence informs ecological understanding. However, to infer the ecological system process, it is important to correctly account for the observation process(es) as this will impact the estimation of the components of the system of interest such as survival probabilities or level of abundance. In particular, incorporating individual heterogeneity to account for individual variability can be important within both the system and observation processes. Individual heterogeneity may reflect individual-level information, for example, weight, state of health and generic in-

formation.

Two specific challenges often arise within the estimation of the ecological quantities of interest, in relation to (i) missing information and (ii) the volume of available data volume. For example, in relation to the volume of data, the open-source eBird platform accumulated over 1 billion bird sightings in less than two decades (eBird, 2021). Algorithms that work well in lower dimensions (i.e., relatively small numbers of parameters or data size) may not necessarily work in such high dimensions (or volumes). This phenomenon is known as the curse of dimensionality. Therefore, current research areas of focus include improving the efficiency of standard model fitting approaches. On the other hand, missing information may arise in different forms e.g., individual-specific information such as weight and health status are typically unavailable when an individual is not observed. This missing information may create more complicated model structures thus leading to challenging computational algorithms.

In this thesis, we contribute to help in addressing these two challenges, with particular focus on the estimation of abundance and demographic parameters. We develop and present: (i) alternative model fitting algorithms that can handle such missing information in capture-recapture and camera trapping studies; and (ii) new efficient algorithms that are scalable to larger population and/or studies. The performance of these new approaches is compared to existing procedures using both simulated and real world data sets relating to a wide range of species including meadow voles, northern parula and barking deer for a range of capture-recapture type data. The new proposed algorithms resulted in considerable computational savings. For example, for the camera trap data relating to the barking deer data, the new algorithm took approximately 20 hours to run compared to 86 hours for the traditional, standard, algorithm implemented using bespoke code, while using black-box software led to substantially slower, and infeasible, implementation times.

Abstract

Capture-recapture studies and the use of motion-sensor camera traps are common and becoming increasing popular for collecting data on wildlife populations. In this thesis, we focus on data collected from capture-recapture and camera trapping studies, and develop model fitting algorithms that are efficient for estimating the animal abundance for each study.

In our first work, we restrict our study relating to individual heterogeneity in capture-recapture studies. We consider (i) continuous time-varying individual covariates and (ii) individual random effects. In general, the associated likelihood is not available in closed form but only expressible as an analytically intractable integral. The integration is specified over (i) the unknown individual covariate values (if an individual is not observed, its associated covariate value is also unknown) and (ii) the unobserved random effect terms. Previous approaches to dealing with these issues include numerical integration and Bayesian data augmentation techniques. However, as the number of individuals observed and/or capture occasions increases, these methods can become computationally expensive. Thus, we propose a new and efficient approach that approximates the analytically intractable integral in the likelihood via a Laplace approximation. We find that for the situations considered, the Laplace approximation performs as well as, or better, than alternative approaches, yet is substantially more efficient.

In the second work, we focus on spatially-related individual heterogeneity in camera trapping studies. However, animal identification is not always feasible in practice due to poor quality images and/or individuals not having uniquely identifiable physical characteristics. Spatially explicit models for unmarked individuals permit the estimation of animal density when individuals cannot be

uniquely identified. Due to the structure of these models, a Bayesian super-population (data augmentation) approach (SPA) is often used to fit the models to data, which involves specifying some reasonably "large" upper limit for the population. However, this approach presents computational challenges, particularly when dealing with larger populations, as demonstrated by the motivating dataset relating to barking deer (*Muntiacus muntjak*) collected in Ujung Kulon National Park (UKNP), Indonesia. In this second work, we develop new efficient algorithms in the Bayesian framework that do not require *a priori* specifying the upper population limit. We compare the performance of the different approaches using small datasets: relating to northern Parula and simulated data, and demonstrate that even with a relatively small dataset the new algorithms are consistently more efficient than the previous super-population approach. Finally, we apply the new algorithm to the large barking deer dataset, where the standard super-population approach is computationally very expensive. Our finding shows that the spatial density estimates of barking deer in the study area are between 12 and 13 animals per km^2 with 95% of credible interval ranging from 6 to 20 animals per km^2; the difference in the computational aspect between the two algorithms is particularly marked for the deer case study with the SPA algorithm taking substantially longer to implement compared to the new algorithm (by a factor of 4).

Contents

Chapter 1

Introduction

Estimating wildlife abundance is essential in ecological studies related to animals. In animal ecology, abundance estimation is one of the key mechanisms for population monitoring. Long-term monitoring of the wildlife population can be useful for studying how the population fluctuates over time, which becomes important in creating effective monitoring strategies. The simplest and most primitive method for estimating abundance is a direct count over a fixed defined area. For example, Jachmann (2002) used a block count to estimate wildlife abundance of African herbivores by counting the observed individuals within a fixed area; Pettorelli et al. (2007) and Tracey et al. (2008) conducted an aerial survey where an observer counted the wildlife from the aeroplane or took a snapshot of individuals within a certain area. However, such methods may produce data with high variations e.g., due to availability, or visibility of target individuals. On the other hand, conducting a complete census i.e., counting all individuals within the population at relatively large location or maintaining long-term monitoring activities are practically infeasible due to limited resources e.g., time and finance, and the characteristics of the population (and their habitat). As a result, in practice only a subset of the population is typically obtained via some survey method.

Traditionally, collecting data on wildlife population often required a substantial amount of resources. However, recent advances in technology have permitted a much greater amount of ecological data to be collected. Farley et al. (2018) reported that petabytes of ecological data coming from remote sensors has been successfully collected and been regarded as the largest generation of data in ecology, advancing at unimaginable speed. Another example where massive amounts of data are becoming increasingly available is in relation to citizen-science data, such as the eBird platform. The initiative was successful in gathering more than 684,300 (users) who reported over 1 billion bird sightings from more than 202 countries in 2021 (eBird, 2021). However, the modelling process is made more complex and difficult by these additional sorts of data. Since existing model fitting procedures may not scale up with enormous volumes of data, new analytical and computational features are therefore crucial. Thus, many current areas of research focuses in improving computational efficiency in model fitting algorithms.

In this thesis, we also focus on providing solutions in fitting moderate-high volume of ecological data for better efficiency while maintaining accuracy.

1.1 Underlying data structures

We focus on data collected from repeated surveys over time producing a series of sampling occasions; in particular, we focus on data collected from (i) capture-recapture studies; and (ii) camera-trapping studies. If the corresponding observed population can be uniquely identified, then we may construct individual histories of the associated species thus we refer them to as "marked population"; and otherwise we refer such population to "unmarked population". In this thesis, we consider marked capture-recapture; and unmarked camera-trapping data, and consider discrete-time models where individuals may be observed at regular time intervals e.g., daily, monthly or yearly.

1.1.1 Capture-recapture studies

The two-sample method is the simplest method used in capture-recapture experiments and used for estimating the population size. The two-sample method consists of two sampling occasions where the first occasion allows individual for marking or tagging before being released back to the population; and the second sample gives the recapture histories (Goudie and Goudie, 2007). This two-sample method was later developed into the T-sample occasions method allowing multiple recaptures. An early use of this theory was introduced by Schnabel (1938) for studying fish in a lake where observations were conducted at multiple points on the lake and taken at periodic intervals.

In capture-recapture studies, researchers go to the designed study area subsequently at different sampling times, also referred to as capture occasions, to identify and record all observed individuals. Observers "capture" individuals and mark them at the first encounter for future identification so that each individual is uniquely marked or tagged after first encounter. At the second sampling times, observed individuals are recorded and often labelled as "0" if the corresponding

individual is not observed during the sampling and "1" if it is observed. The survey is repeated until the final capture occasion, T. Individuals that have not been observed at any capture occasion are also denoted "0" and the number of such individuals is unknown therefore they are not reported. Thus, the data collected from capture-recapture studies are the set of capture histories i.e., from the first encounter to the final occasion for each individual observed, and these series are reported in 1 and 0 entries indicating their presences or absences as shown in Table 1.1.

In capture-recapture study, all observed individuals need to be identified uniquely. Some animals may have natural-physical markings that can be used for individual identification such as the spot patterns of cheetah or leopards (Balme et al., 2010); belly patterns of great crested newts (McCrea and Morgan, 2015) or unique stripe patterns of tigers (Karanth and Nichols, 1998). However, most animals do not have unique-natural markings. Therefore, marks such as a ring (e.g., for birds) or a tag (e.g., for sheep) are often applied to an individual at the first capture for identifying marks.

To illustrate a capture-recapture study, assume the known total population in the designed area is 6; and there are 4 distinct sampling times. The hypothetical structure of capture-recapture data can be seen in Table 1.1. From the table, we can see that the first observed individual (Unique id 1) was encountered at the first capture occasion but then unobserved at the second occasion before being seen on occasion 3 and unobserved again at the final occasion (1, 0, 1, 0). Individual 2 (Unique id 2) was captured and marked at the second occasion and only seen again at occasion 4 and so on. In this example, only 4 out of 6 individuals are observed during the study, the remaining 2 individuals are unobserved and in practice such numbers are unknown. We also note that additional information may be collected during the study. Such information can be at individual level (producing individual covariates) such as body mass or be global variables such as environmental covariates.

Table 1.1: Hypothetical capture-recapture data set showing 4 individuals captured within 4 capture occasions at given location.

Unique id	Occasion 1	Occasion 2	Occasion 3	Occasion 4
1	1	0	1	0
2	0	1	0	1
3	1	1	1	0
4	0	0	1	1

1.1.2 Camera-trapping studies

Obtaining capture histories of uniquely identified species is not always practically easy. The nature of the population, environmental conditions and limited resources makes it more challenging. A recent advancement in technology allows more possibilities in collecting wildlife population. One of these kinds that is quickly gaining popularity is motion-sensor cameras due to their non-invasive nature and cost-effectiveness. Additionally, the use of motion-sensor cameras for surveying allows for long-term wildlife monitoring which may be often of interests e.g., observing effects of environmental changes over time on the population. Cameras may record a wide range of information at the same time such as animals' presence and activities.

Early application of wildlife photography were introduced in the 1870s: (i) to photograph an endangered species, a captive quagga, at the London Zoo; (ii) to photograph rock-hopper penguins and breeding albatrosses. The latter are one of the earliest uses of wildlife photography for scientific purposes (Kucera and Barrett, 1993). George Shiras pioneered "camera trapping" as an early use in the 1890s where he developed a self-photographed camera using a trip wire and a flash system (Guggisberg, 1977). Camera systems have rapidly improved since then resulting in more modern camera trapping. Consequently, the use of camera trapping has been significantly increasing in animal ecology for population monitoring. For examples, camera trapping has been widely used for surveying different wild animals such as cat species e.g., tigers (Karanth and Nichols, 1998; Karanth et al., 2006; Jennelle et al., 2002), jaguars (Silver et al., 2004; Wallace et al., 2003), ocelots (Trolle and Kéry, 2003, 2005), mountain lions (Long et al., 2003), big cats and canids (Brassine and Parker, 2015; Pereira et al., 2022;

Di Bitetti et al., 2009); deer (Rahman, 2016; Rahman et al., 2017), pigs (Sweitzer et al., 2000) and grizzly bear (Mace et al., 1994).

In camera trapping, a collection of cameras is frequently placed along trails (at multiple locations) or at baited stations. Cameras capture individual animals on film and are set to record date and time of all photos (videos). Depending on the length of the study, battery-based cameras should be examined at least once a month, with the memory card and battery replaced (Karanth and Nichols, 1998). If natural-physical marks are available for individual identification, for example tigers (Karanth and Nichols, 1998; Karanth et al., 2006), ocelots (Trolle and Kéry, 2003, 2005) and leopards (Chapman and Balme, 2010), then individuals may be uniquely identified. Captured images of animals may be identified either manually by a human or more frequently recently, using specialised identification software e.g., HotsSpotter (Li and Stephens, 2003). The recorded data are then similar to capture-recapture data with individual capture histories for each individual observed within the study, but with the unique time recorded for each observation rather than sampling time, i.e. observations are recorded in continuous time. See Van dam Bates (2023) for further discussion of continuous time camera-trapping data and associated modelling approaches. Continuous-time hidden Markov models have also been developed and applied to capture-recapture studies with irregular sampling occasions (Mews et al., 2020). For further discussion of the use of hidden Markov models within such applications, see for example, Glennie et al. (2023).

However, the use of unique marks to uniquely identify individuals are often not available in practice and poor-quality images make it more challenging for absolute individual identification. In this case, the data of the camera-trapping studies corresponds to the time detections of the target species identified from the snapshots of each camera. Thus, the data are no longer capture histories of individual animals, but simply the times that individuals are observed by each camera trap. For illustration, four motion-sensor cameras are placed at a sample site and set to continuously record a (known) total of 10 animals in the area for three days. We discretise the sampling time into daily thus resulting in 3

Table 1.2: Hypothetical data structure for a camera-trapping study collected from four motion-sensor cameras during three sampling times.

Cameras	Day 1	Day 2	Day 3
1	4	5	6
2	3	2	4
3	4	6	3
4	1	3	5

sampling occasions (days). Photographed animals are collected and counted for each sampling time and each camera. Note that we assume multiple visits at each sampling time are permitted i.e., the same individual may appear more than once during a single capture occurrence. Table 1.2 shows a hypothetical data structure for a camera-trapping study after being reduced to the recorded data. Camera 1 captured 4 detections at the first sampling (Day 1) and 6 detections at the end of sampling (Day 3) while Camera 4 only detected 1 individual at Day 1 and so on. Similarly, trap-specific covariate information can also be collected during the study.

1.2 Basic Models

We consider two types of populations in this study: (i) closed populations, and (ii) open populations. We consider and describe the underlying models for each of these populations in turn below.

1.2.1 Closed populations

In closed populations, we assume that the population is constant throughout the study so that birth/death and immigration/emigration are not permitted within the study period. We note that if the sampling times are too far apart between successive samples, then the closure assumption is frequently violated in an ecological study. Typically, the main objective of designed models for closed populations is estimation of the fixed population size, N. In this section, we introduce the basic closed-population model for individually marked populations, denoted M_0. Let $\boldsymbol{x} = \{x_{it}; i = 1, \ldots, N; t = 1, \ldots, T\}$ denote a set of individual

capture histories for the study period with T capture occasions; and let p denote the capture probability. Assuming the capture probability is time-invariant and individuals are encountered independently, the likelihood of M_0 (Otis et al., 1978) is given by:

$$L(N, p; \boldsymbol{x}) \propto \frac{N!}{(N-n)!} \prod_{i=1}^{N} \prod_{t=1}^{T} p^{x_{it}} (1-p)^{1-x_{it}}, \tag{1.1}$$

where n is the total number of observed individuals. Note that in addition to the closure assumption, Otis et al. (1978) made two other assumptions related to identifying marks: (i) marks are not lost during the experiment; (ii) marks can be identified correctly at each sampling occasion. In addition to the estimation of the population size, it is important to consider the recapture probability in closed-population models as the misspecification in p may lead to biased estimates of N (McCrea and Morgan, 2015). Otis et al. (1978) described three other classes of models in addition to M_0 to account for different sources of variation in p: (i) M_t to account for temporal heterogeneity p_t; (ii) M_b to account for behavioural heterogeneity p_b (trap happy/shy); and (iii) M_h to account for individual heterogeneity p_h. The subscripts t, b and h correspond to the variation (heterogeneity) in p that we will discuss later in the next section. Combinations of different types of heterogeneity in p provide various alternative classes of model, and we discuss some of these in Chapter 2.

In addition to temporal, behavioural and individual heterogeneity, variation in space (i.e. spatial heterogeneity) can be incorporated into the capture probability, p, establishing spatial capture-recapture (SCR) models. The model structure of SCR resembles that of the closed population model, M_h, with individual spatial heterogeneity. Thus, the recapture probability, p, is now specified as an individual heterogeneity model. The idea of SCR model is to incorporate the spatial information into the capture probability i.e., individuals that are closer to traps are more likely to be detected (Borchers and Efford, 2008). Thus, the capture probability in SCR model is conditional on where individuals are during the sampling occasion. A quantity that describes the centre of an animal's movement activity (i.e., locations where animals traverse) is typically known as their "activity

centre", which is usually unknown. SCR models specify the capture probability as a function of the distance between an individual's activity centre and the associated traps. SCR not only provides an estimate of the total population size, but additionally is able to estimate the spatial density over the given sampling region.

The idea of incorporating spatial heterogeneity has been developed for unmarked populations, where individuals cannot be uniquely identified, allowing the spatial density estimation without the need of individual identification (Chandler and Royle, 2013). Such models are typically referred to as unmarked spatial capture-recapture or spatially explicit count model. We will discuss the unmarked case in more detail and the associated model, in Chapter 4.

1.2.2 Open populations

In open populations the closure restrictions are relaxed by allowing births, deaths and/or migration to occur during the study. For example, we may permit births or deaths within the study but assume that there is no temporary immigration or emigration. Estimation of the survival probabilities of individuals often becomes the parameter of interest in open populations, in contrast to the primary focus on abundance in closed populations. However, depending on the research question, the objective of open population models can also be the estimation to the population size. In this section, we introduce an open population model proposed by Cormack (1964), Jolly (1965), and Seber (1965) i.e., the Cormack-Jolly-Seber model (CJS) which focuses on the estimation of the survival probabilities. We recall that $\boldsymbol{x} = \{x_{it}; i = 1, \ldots, N; t = 1, \ldots, T\}$ denote a set of individual capture histories for the study period with T capture occasions. Now, let $\boldsymbol{\phi} = \{\phi_t; t = 1, \ldots, T-1\}$, where ϕ_t denotes the survival parameter i.e., the probability an individual is alive at time $t+1$ given that they are alive at time t. Similarly let $\boldsymbol{p} = \{p_t; t = 1, \ldots, T\}$, where p_t denotes the probability an individual is recaptured alive at time t, given that they are alive. Thus, the model assumes that all individuals present in the population has the same survival probability until the next capture occasion (given they are alive). Further, we denote

f_i and ℓ_i the first and the last time individual i is observed. Assuming a fully time-dependent model, the likelihood of the standard CJS model is given by:

$$L(N, \boldsymbol{\phi}, \boldsymbol{p}; \boldsymbol{x}) = \prod_{i=1}^{N} \left[\prod_{t=f_i}^{\ell_i - 1} \phi_t \prod_{t=f_i+1}^{\ell_i} p_t^{x_{it}} (1 - p_t)^{1 - x_{it}} \right] \chi_{i\ell_i}, \tag{1.2}$$

where $\chi_{i\ell_i}$ is the probability that individual i is not observed again after final occasion ℓ_i and is given by:

$$\chi_{i\ell_i} = 1 - \phi_{\ell_i} + \phi_{\ell_i}(1 - p_{\ell_i+1})\chi_{i\ell_i+1},$$

with $\chi_{iT} = 1$ for all $i = 1, \ldots, N$. Note that the CJS models can be easily extended by incorporating extra information e.g., recovery parameters such that capture histories x_{it} have an additional category indicating whether the corresponding individual i is dead at time t (Langrock and King, 2013); or recruitment parameters indicating the proportion of animals entering the population and are available for captures (Pledger et al., 2010; Worthington et al., 2019).

Similar to the closed population model, additional sources of heterogeneity (beyond temporal) can also be incorporated into $\boldsymbol{p}$ and $\boldsymbol{\phi}$ e.g., individual heterogeneity or combinations of both temporal and individual heterogeneity. A more detailed description of these models is discussed in Chapter 2 where we consider the CJS model with individual time-varying heterogeneous survival probabilities, i.e. where the survival probability is expressed as a function of an individual time-varying covariate.

1.3 Heterogeneity

In the previous section, we have briefly introduced the term heterogeneity. In this section, we describe and discuss different forms of heterogeneity in more detail.

Heterogeneity may arise in three different forms as described in King and McCrea (2019) and in Otis et al. (1978): temporal (p_t), behavioural (p_b) and individual (p_h). Temporal heterogeneity assumes that capture probabilities are time-variant over capture occasions which may be due to, for example, different

weather conditions that affects animals' behaviour, different search strategies or efforts related to capture occasions. Behavioural heterogeneity reflects changes in capture probabilities due to trap experiences such that independence assumption between capture occasions is no longer valid. For example, some animals may be more likely to be observed again in the future if they encounter trap-happy experience; or become less likely to be re-encountered at future occasions if animals experience trap-shy situation, unpleasant experience such as being trapped or having physical contacts. Note that in camera trapping studies we may not have such a trap shy response from animals since they are recorded without being physically contacted but a trap happy response is possible since sometimes baits are used to attract individuals.

Individual heterogeneity, with parameter dependence p_i with i denoting individual i, assumes the capture probability varies among individuals which may be due to individual characteristics of animals. Such variation among individuals may reflect, for example, relationship between the study locations and the home range size of individuals (activity centres), or characteristics of individuals (covariates) such as sex, weight, state of health and generic information (King and McCrea, 2019). King and McCrea (2019) further classified such covariates into: (i) deterministic covariate values, and (ii) stochastic covariate values. Behavioural heterogeneity is a special case of deterministic covariate where it permits the change in individual capture probabilities following its first capture. Examples of stochastic covariates includes, but not limited to, breeding and health status, hunger levels and location where such covariates may change in a stochastic manner e.g., hourly, daily, monthly, or even annually.

Having individual heterogeneity observed in capture-recapture studies and camera trapping is not always possible due to different reasons, especially for individual stochastic covariates. For example, in camera trapping collecting individual characteristics is nearly impossible since capture histories are collected from snapshot images and poor quality images makes it more challenging. However, we note that characteristics such as sex and patterns may be still possible to collect if natural marking between individuals is available for such identification.

Similarly, in capture-recapture studies, unless individuals are physically captured obtaining such individual-covariate information is also nearly impossible. However, again note that observing animal sex and/or breeding status may still be possible dependent on the species. Therefore, models with missing individual covariates would be often found in a wide range of examples. In this thesis, we also consider models with missing individual covariate that we will talk in more details how to fit such models in Chapters 2 and 4.

1.4 Model fitting approaches

In this section, we discuss two model fitting approaches in general. In particular, we focus on two distinct types of inference: (i) classical inference, and (ii) Bayesian inference.

1.4.1 Classical inference

Let $M(\boldsymbol{\theta})$ be any model with p-dimensional parameter vector $\boldsymbol{\theta}$ and $\boldsymbol{x}$ be the observed data. The model parameter estimates, $\hat{\boldsymbol{\theta}}$, are obtained by fitting the model $M(\boldsymbol{\theta})$ to data $\boldsymbol{x}$ by the method of maximum likelihood. The maximum likelihood estimation is used to find the parameters which the observed data have the highest joint probability, $f(\boldsymbol{x};\boldsymbol{\theta})$, by maximizing the likelihood function $L(\boldsymbol{\theta};\boldsymbol{x})$ as a function of $\boldsymbol{\theta}$. In practice, the natural logarithm of the likelihood function, $\ell(\boldsymbol{\theta};\boldsymbol{x})$, is often used since the maximum point of the log-likelihood occurs at the same point as the maximum of the likelihood function.

In theory, the maximum or (minimum) value is attained when the first derivative of the log-likelihood function is equal to zero assuming the log-likelihood function is differentiable such that:

$$\frac{\partial \ell(\boldsymbol{\theta};\boldsymbol{x})}{\partial \boldsymbol{\theta}} = 0. \tag{1.3}$$

The parameter estimates obtained from the maximum likelihood method are known as the maximum likelihood estimators (MLE) and often denoted as $\hat{\boldsymbol{\theta}}_{\text{mle}}$.

However, in practice solving Equation (1.3) is often not possible to do analytically. In particular, obtaining explicit maximum-likelihood estimates is often difficult when estimating capture-recapture models due to their complexity. We often require iterative methods for finding optimal parameters which maximizes the likelihood e.g., quasi Newton, Fisher scoring (Broyden, 1970; Osborne, 1992).

An important assumption for obtaining unique roots of Equation (1.3) such that the MLE is a (local) maximum is that the second derivative or the Hessian matrix, $H(\hat{\boldsymbol{\theta}}_{\text{mle}})$, of the log-likelihood is negative semi-definite. Additionally, one of the important properties of the MLE is that $\hat{\boldsymbol{\theta}}_{\text{mle}}$ converges in distribution to a normal distribution,

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\text{mle}} - \boldsymbol{\theta}) \rightarrow \text{N}_p(0, I^{-1}(\hat{\boldsymbol{\theta}}_{\text{mle}})),$$

where $I^{-1}(\boldsymbol{\theta})$ is the Fisher information matrix and n and p denote the sample size and the number of parameters respectively. Finally, we note that the observed Fisher information is the negative of the Hessian matrix i.e., $I(\hat{\boldsymbol{\theta}}_{\text{mle}}) = -H(\hat{\boldsymbol{\theta}}_{\text{mle}})$. In this thesis, we use the maximum likelihood method for estimating parameters of capture-recapture models in Chapter 2.

1.4.2 Bayesian inference

Unlike the maximum likelihood method where the parameters are assumed to have a fixed value which is estimated, the Bayesian approach assumes the parameter as a random variable and estimates its distribution. The idea is generally derived from the application of Bayes Theorem stating,

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{x})},$$

where $\pi(\boldsymbol{\theta})$ denotes the prior distribution for $\boldsymbol{\theta}$ that we assign, $\pi(\boldsymbol{\theta}|\boldsymbol{x})$ is the posterior distribution of $\boldsymbol{\theta}$ and $f(\boldsymbol{x}|\boldsymbol{\theta})$ denotes the likelihood function and is the same as $L(\boldsymbol{\theta}; \boldsymbol{x})$ defined in Section (1.4.1). The denominator term, $\pi(\boldsymbol{x})$, is the

marginal probability written mathematically as:

$$\pi(\boldsymbol{x}) = \int f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta},$$

for continuous data. For discrete data, the marginal probability is expressed as:

$$\pi(\boldsymbol{x}) = \sum f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}).$$

In practice, the denominator term can be dropped from the posterior distribution since it serves as a normalizing constant to the product of the likelihood and the prior. Thus, the posterior distribution is proportional to the joint distribution after dropping the marginal probability i.e.,

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) \propto f(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}). \tag{1.4}$$

In the Bayesian framework, inference is made based on information from the prior distribution and from the conditional probability of observed data (the likelihood); in classical inference, we solely rely on the likelihood of the observed data to estimate model parameters. Thus, in the Bayesian approach we may incorporate relevant knowledge about parameters, also refer to expert knowledge or prior beliefs, through the prior $\pi(\boldsymbol{\theta})$ without reference to the dataset. In general, obtaining the full posterior density is the main interest in Bayesian approach but it is very complicated. The posterior densities often do not have closed-form analytic forms except for a specific class of priors and models. However, exploring the full posteriors is possible numerically via Markov chain Monte Carlo (MCMC) simulations. The main idea of the MCMC is to draw random samples from the target distribution (the posterior) that cannot be drawn easily by constructing a Markov chain i.e., drawing dependent samples where the current sample is dependent upon the last sample. For inference, rather than consider the entire posterior, $\pi(\boldsymbol{\theta}|x)$, we often consider simpler summary statistics such as the marginal posterior distribution of individual parameters. We will use a Bayesian inference for estimating model parameters of camera-trapping studies

in Chapter 4.

1.5 Blackbox Software

In this thesis, some models are fitted using readily available packages from `R` software. We briefly describe three main packages for model fitting purposes: `rjags, nimble` and `TMB`.

1.5.1 `rjags`

The `rjags` package provides an interface from `R` to the JAGS (just another Gibbs sampler) library which allows for users to specify their own functions and distributions in performing Bayesian data analysis (Plummer et al., 2022). JAGS employs MCMC algorithm to draw a set of dependent samples from the target posterior distribution. Working with `rjags` means that we need to provide the model specification and priors along with some input values regarding observed data, number of iterations, number of chains and then it does the MCMC blackbox "magic". Then, generated samples can be extracted to summarise the posterior distribution.

1.5.2 `nimble`

The software package "Numerical inference for statistical models for Bayesian and likelihood estimation" (`nimble`) is an `R` package designed for carrying out both Bayesian and classical inference (Valpine et al., 2022). `nimble` is an extension of the BUGS language providing libraries for MCMC algorithms, particle filtering (sequential Monte Carlo) and Monte Carlo Expectation Maximization (MCEM). Similar to the software package `rjags`, `nimble` users provide the model specification and priors but with more flexibility e.g., customizing samplers and even writing algorithms when performing Bayesian data analysis. The Nimble compiler allows the model and algorithm written in `R` to be generated in C++ template during the compilation thus providing more efficient computation.

1.5.3 Template Model Builder (TMB)

Template Model Builder (TMB) is an R package designed for fitting complex statistical models especially models with random effects using a classical inference. It was built by Kristensen et al. (2016) where the principles are inspired by Automatic Differentiation Model Builder (ADMB) package. It uses some high-performance libraries including CppAD for automatic differentiation in C++, Matrix for sparse matrix calculation in R and Eigen for sparse matrix calculation in C++. Consequently, this package gives better performance in terms of speed for fitting models of high dimensional parameters and is much simpler when coding complex models compared to e.g., JAGS, ADMB (Kristensen et al., 2016). TMB users are required to have some basic understanding of the C++ programming language as the joint likelihood of the model must be written in C++. However, the other operations are performed in R e.g., reading data, compiling the model and optimization.

1.6 Thesis layout

The structure of this thesis is as follows: Chapter 2 discusses capture-recapture models for: (i) closed populations, and (ii) open populations considering different forms of heterogeneity including temporal, behavioural and/or individual (time-invariant and time-variant). Different approximations for integrating the unobserved individual heterogeneity are described in detail in this chapter. Laplace approximations are given in more details as a focus of the work.

Chapter 3 provides simulation studies fitted on closed-population and open-population models of capture-recapture data for comparing different approximations under several scenarios. A computational comparison is also reported. We give two real examples related to closed-population and open-population studies.

Chapter 4 describes spatial capture-recapture models for unmarked populations in camera trapping studies. Bayesian inferences such as Bayesian data augmentation and reversible jump MCMC are introduced as model fitting methods. The approach designed permits the direct specification of the prior information

on total population size; and also we develop efficient computational algorithm for model fitting.

Chapter 5 presents a simulation study to investigate the computational efficiency of the algorithms. We also demonstrate the approach initially on a small case study before applying to a case study of a large, unmarked data related to barking deer (*Muntiacus muntjac*) in Ujung Kulon National Park, Indonesia. A detail discussion on spatial density estimates of barking deer is given and reviewed.

The final chapter reviews possible and promising future works of the current research areas. Our works in Chapter 2 and 3 has been reviewed and published in The Journal of Agricultural, Biological and Environmental Statistics (JABES) while the second contribution in Chapter 4 and 5 are currently in submission to a statistics journal.

Chapter 2

Capture-Recapture Studies: Models and Methods

Capture-recapture surveys are often used when studying wildlife populations to understand the associated population dynamics necessary for management and conservation. These surveys involve repeatedly sampling the population over a series of capture occasions. Observations at each capture occasion may take the form of physical captures and/or visual sightings of animals. Individuals are uniquely identified, using, for example, a ring or tag applied at initial capture or via natural markings e.g., birds often ringed (BTO, 2018); butterflies may be marked using marking pens and ibex may be tagged on the ears or belly patterns of a Great crested newt (McCrea and Morgan, 2015). From the surveys, we obtain capture histories for each individual observed in the study. The observed data correspond to the set of observed histories, detailing whether or not they are observed at each capture occasion. We assume that there is no loss of marks; and no mis-classification across individuals during the study. See Section 1.1.1 for example of associated data and general notation used in this Chapter.

Capture-recapture studies may be assumed to be closed or open, dependent on whether the population is constant throughout the study; or may change due to births, deaths and/or emigration, or immigration, respectively. The corresponding parameters of interest typically differ between such studies with closed population models primarily focusing on abundance estimation; while open population models often focus on the estimation of survival probabilities, although these may also extend to abundance. For a review of these data and associated models see for example McCrea and Morgan (2015); Seber and Schofield (2019).

Incorporating heterogeneity i.e., temporal, behavioural and/or individual heterogeneity in capture-recapture models can be important to model the underlying system processes e.g., capture probabilities and/or survival probabilities. Omitting such sources of heterogeneity can lead to biased and/or misleading results (Rosenberg et al., 1995; Schwarz, 2001; Chao et al., 2001; White and Cooch, 2017). Otis et al. (1978) described three main sources of heterogeneity: temporal, behavioural and individual. For additional discussion see Section 1.3. In this work we focus on individual heterogeneity, which can often be incorporated via the use of observable characteristics, such as gender, breeding status or con-

dition. King and McCrea (2019) categorised observed individual covariates into 2×2 cross-classifications corresponding to continuous/discrete-valued and time-varying/invariant. We focus on the more challenging case of continuous-valued covariates. Missing data often arise for such continuous-valued and discrete-valued covariate information due to, for example, imperfect data collection or simply the structure of the experimental design. For example, for stochastic time-varying covariates, if an individual is not observed, the corresponding covariate is also unknown at that time. In general, for continuous-valued covariates, the observed data likelihood is only expressible as an analytically intractable integral leading to model-fitting challenges. Previous model-fitting approaches include using a Bayesian data augmentation (Bonner and Schwarz, 2006; King and Brooks, 2008; Bonner et al., 2010); an approximate discrete hidden Markov model (Langrock and King, 2013); and a two-step multiple imputation approach (Worthington et al., 2015). However, these approaches are computationally expensive and may lead to computational challenges for large datasets.

As an alternative to modeling individual heterogeneity with covariates, which may be missing or unobserved, the heterogeneity can be modelled via individual random effects. For example, the capture probabilities may be specified as a function of covariates or random effects as follows. For individual i, when there are observed covariate values, we may specify the associated capture probability, p_i, to be of the form:

$$\text{logit}(p_i) = \beta_0 + \beta_1 z_i,$$

where z_i denotes the covariate value for individual i. Alternatively, for individual random effects, we may specify:

$$\text{logit}(p_i) = \beta_0 + \epsilon_i,$$

where ϵ_i denotes the random effect term of the corresponding animal i respectively and ϵ_i has some associated distribution. The random effect terms may either be specified as a finite mixture model (Pledger, 2000; Pledger et al., 2003);

or an infinite mixture model (Coull and Agresti, 1999; Dorazio and Royle, 2003; King and Brooks, 2008). However, we note that identifiability issues may arise in terms of the distributional assumption of this heterogeneity, with different models leading to the same distribution for the observed data (Link, 2003, 2006), indicating that some sensitivity analyses are advisable in practice. For the finite case, a closed form expression for the likelihood is available, summing over the mixture components; for the infinite case the necessary integration is generally analytically intractable (though see Dorazio and Royle (2003) for a special case for closed models assuming a Beta distribution). Again, within this thesis we focus on the case of continuous-valued individual random effects. A variety of approaches have been applied to fit random effect individual heterogeneity models to data including conditional likelihood (Huggins and Hwang, 2011); Bayesian data augmentation (King and Brooks, 2008; Royle et al., 2007; Royle, 2008); numerical integration (Coull and Agresti, 1999; Gimenez and Choquet, 2010; White and Cooch, 2017) and combined numerical integration and data augmentation (Bonner and Schofield, 2014; King et al., 2016); King et al. (2022) recently integrated data augmentation and an importance sampling approach which itself uses an approximation of the likelihood using Monte Carlo integration or quadrature. Observed and unobserved heterogeneity can be jointly considered via a mixed model specification, including both covariate information and additional random effects with similar model-fitting tool applied e.g.,

$$\text{logit}(p_i) = \beta_0 + \beta_1 z_i + \epsilon_i$$

or

$$\text{logit}(p_{it}) = \beta_0 + \beta_1 z_{it} + \epsilon_i$$

if the capture probability is also time-varying. See for example King et al. (2006); Gimenez and Choquet (2010); Stoklosa et al. (2011) for additional discussion and examples.

Our contribution to the approaches taken to fit models with individual heterogeneity is the development of a computationally efficient Laplace approximation

for the analytically intractable integral in the likelihood function in the presence of individual heterogeneity for capture-recapture data, which we subsequently numerically maximise to obtain the MLEs of the parameters. We apply the approach to fit (i) a closed population model with individual random effects (using a higher order Laplace approximation for improved accuracy) and (ii) an open population model with time-varying continuous covariate information. For the open population model, we use the numerical automatic differentiation tool in the `R` package Template Model Builder (TMB; Kristensen et al., 2016) to approximate the likelihood. TMB builds on the approach of the Automatic Differentiation Model Builder (ADMB) package where the objective function is written in `C++`. See Section (1.5). The approach is scalable to both large dimensions and sample size i.e., it is designed to be fast for handling many random effects ($\approx 10^6$) and parameters ($\approx 10^3$) (Kristensen et al., 2016) since computing the most challenging computation (the second derivatives) is no longer expensive due to the automatic differentiation function within TMB.

2.1 Capture-Recapture Models

In this section, we first describe capture-recapture models, both a closed and open capture-recapture model, and consider a continuous-valued random effects model, and time-varying continuous-valued individual covariate model, respectively.

First, we provide a brief description of the general notation for capture-recapture studies, before describing the specific models for closed and open populations that we consider in detail with their associated observed data likelihoods in the following Section. We let $t = 1, \ldots, T$ denote the capture occasions within the study; and $i = 1, \ldots, n$ the observed individuals over all the capture occasions where n and T denote the total observed individual and capture occasions respectively. Then, for each observed individual $i = 1, \ldots, n$ and capture occa-

sion, $t = 1, \ldots, T$ we let,

$$x_{it}^{obs} = \begin{cases} 1 & \text{if individual } i \text{ is observed at time } t; \\ 0 & \text{if individual } i \text{ is unobserved at time } t. \end{cases}$$

The capture history associated with individual $i = 1, \ldots, n$ is denoted by $\boldsymbol{x}_i^{obs} = \{x_{it}^{obs} : t = 1, \ldots, T\}$; and the set of capture histories $\boldsymbol{x}^{obs} = \{x_{it}^{obs} : i = 1, \ldots, n; t = 1, \ldots, T\}$. Finally, we let f_i and ℓ_i denote the first and last time individual i is observed, respectively, for $i = 1, \ldots, n$.

In this thesis, we consider the individual heterogeneity model component denoted by $\boldsymbol{y} = \{\boldsymbol{y}^{obs}, \boldsymbol{y}^{unobs}\}$, where $\boldsymbol{y}^{obs}$ and $\boldsymbol{y}^{unobs}$ denote the observed and unobserved individual heterogeneity components, respectively. For example, $\boldsymbol{y}^{obs}$ may correspond to observed covariate values; while $\boldsymbol{y}^{unobs}$ may correspond to unobserved covariate values z_i or random effect terms ϵ_i within individual heterogeneity models. We assume that $\boldsymbol{y}^{unobs}$ is continuous-valued. We let $\boldsymbol{\psi}$ denote the associated individual heterogeneity model parameters; and $\boldsymbol{\theta}$ the model parameters to be estimated.

The corresponding data are given by $\{\boldsymbol{x}^{obs}, \boldsymbol{y}^{obs}\}$, with associated observed data likelihood,

$$f(\boldsymbol{x}^{obs}, \boldsymbol{y}^{obs}|\boldsymbol{\theta}, \boldsymbol{\psi}) = \int_{\boldsymbol{y}^{unobs}} f(\boldsymbol{x}^{obs}|\boldsymbol{\theta}, \boldsymbol{y}) f(\boldsymbol{y}|\boldsymbol{\psi}) d\boldsymbol{y}^{unobs}, \tag{2.1}$$

where $f(\boldsymbol{x}^{obs}|\boldsymbol{\theta}, \boldsymbol{y})$ denotes the complete data likelihood; and $f(\boldsymbol{y}|\boldsymbol{\psi})$ the random effect or covariate model component. This likelihood is, in general, analytically intractable. If there are discrete-valued elements of $\boldsymbol{y}^{unobs}$ the integral becomes a summation).

The associated model parameters, $\boldsymbol{\theta}$, depend on the specific capture-recapture or detection model being fitted. We describe the general set of parameters and indicate whether they apply to the closed or population models considered within this thesis:

- N = the total population size (*closed*);

- p_{it} = $\mathbb{P}$(individual i is observed at time t | available for capture at time t) (*open* and *closed*);

- $\phi_{it} = \mathbb{P}$(individual i is alive at time $t+1$ | alive at time t) (*open*).

We let $\boldsymbol{p} = \{p_{it} : i = 1, \ldots, N; t = 1, \ldots, T\}$, and similarly for $\boldsymbol{\phi}$.

2.1.1 Closed M_h-type models

We consider the closed models where the population of study is assumed to be fixed and closed to birth, death and/or migration during the study. In particular, we extend the model M_0 defined in Section 2.1 and restrict to the model M_{tbh}, proposed by Otis et al. (1978), where again the subscripts correspond to temporal, behavioural and individual heterogeneity, respectively. Additional heterogeneity can be modelled via observed covariates (Stoklosa et al., 2011), but we do not consider this case here. The total population size, N, is the parameter of primary interest. The capture probabilities, p_{it}, are expressed such that, $h(p_{it}) = \alpha_t + \lambda S_{it} + \epsilon_i$, where $S_{it} = 0$ if $t \leq f_i$; and $S_{it} = 1$ if $t > f_i$; and h denotes some function constraining the recapture probabilities to the interval $[0, 1]$. Within this work we assume a logistic relationship, so that $h \equiv$ logit. The α_t terms correspond to the temporal component; λ the behavioural component where $\lambda > 0$ signals the trap-happy response while $\lambda < 0$ signals the trap-shy response; and ϵ_i the individual random effect term which we assume to be of the form $\epsilon_i \sim \mathrm{N}(0, \sigma)$. The individual heterogeneity sub-models M_h, M_{th}, M_{bh} are obtained by setting restrictions on the parameters. For example, in the absence of a behavioural effect $\lambda = 0$; and when there are no temporal effects $\alpha_t = \alpha$ for all $t = 1, \ldots, T$. Let $\boldsymbol{x}^{unobs} = \{x_{it}^{unobs}; i = n+1, \ldots, N; t = 1, \ldots, T\}$ denote the unobserved capture histories with all entries equal to zero (i.e. $x_{it}^{unobs} = 0$ for all $t = 1, \ldots, T$ and all unobserved individuals $n+1, \ldots, N$) such that $\boldsymbol{x} = \{\boldsymbol{x}^{obs}, \boldsymbol{x}^{unobs}\}$ and $x_{it} = \{x_{it}^{obs}, x_{it}^{unobs}\}$. The conditional likelihood, given the capture probabilities, $\boldsymbol{p} = \{p_{it} : i = 1, \ldots, N, t = 1, \ldots, T\}$, and the total population size N is of

multinomial form (omitting the constant coefficient terms):

$$f(\boldsymbol{x}|N,\boldsymbol{p}) \propto \frac{N!}{(N-n)!}\prod_{i=1}^{N}\prod_{t=1}^{T} p_{it}^{x_{it}}(1-p_{it})^{(1-x_{it})}.$$

However, the capture probabilities are specified as random effect components. The model parameters are $\boldsymbol{\theta} = \{N, \boldsymbol{\alpha}, \lambda\}$ with individual heterogeneity model parameters $\boldsymbol{\psi} = \{\sigma\}$. Thus, assuming there is no additional observed individual covariate information within the study, we have $\boldsymbol{y}^{obs} = \emptyset$, and $\boldsymbol{y}^{unobs} = \{\epsilon_i : i = 1, \ldots, N\}$. Further, the associated random effect of an unobserved individual i.e., for $i = n+1, \ldots, N$ is denoted by ϵ_0, with associated capture probability at time t given by $h(p_{0t}) = \alpha_t + \epsilon_0$. The observed data likelihood of Equation (2.1) and unobserved data likelihood can be expressed as,

$$\begin{aligned} f(\boldsymbol{x}|\boldsymbol{\theta},\sigma) \quad \propto \quad & \frac{N!}{(N-n)!}\prod_{i=1}^{n}\left[\int f(\boldsymbol{x}_i^{obs}|\boldsymbol{\theta},\epsilon_i) f(\epsilon_i|\sigma)\, d\epsilon_i\right] \\ & \times \prod_{i=n+1}^{N}\left[\int f(\boldsymbol{x}_i^{unobs}|\boldsymbol{\theta},\epsilon_0) f(\epsilon_0|\sigma)\, d\epsilon_0\right], \end{aligned} \tag{2.2}$$

where $f(\epsilon_i|\sigma)$ denotes the density function for the unobserved heterogeneity process; and $f(\boldsymbol{x}_i|\boldsymbol{\theta},\epsilon_i)$ the probability of the associated capture history, such that

$$f(\boldsymbol{x}_i|\boldsymbol{\theta},\epsilon_i) = \prod_{t=1}^{T} p_{it}^{x_{it}}(1-p_{it})^{(1-x_{it})}.$$

We note that the likelihood can be written more efficiently by further considering only unique observed capture histories, but for notational simplicity we retain the product over all observed individuals within Equation (2.2).

Previous approaches for dealing with the intractable likelihood include the use of Gauss-Hermite quadrature to estimate the integral (Coull and Agresti, 1999). White and Cooch (2017) evaluated this approach further via simulation for different parameter values, and concluded that the results were generally unbiased except for relatively low recapture probabilities and/or few capture events. Further, they demonstrated that for larger individual heterogeneity variances a greater number of quadrature points are required to retain accuracy. Alternatively, a Bayesian data augmentation approach has been applied, treating the

individual heterogeneity terms as additional parameters (or auxiliary variables) and calculating the joint posterior distribution over both parameters and auxiliary variables. However, in this approach the number of auxiliary variables is also an unknown (it is equal to the unknown parameter, N), leading to the use of a reversible jump algorithm (King and Brooks, 2008) or super-population approach (Durban and Elston, 2005; Royle et al., 2007). To address these issues King et al. (2016) proposed a computationally efficient semi-complete data likelihood model fitting approach, combining a data augmentation approach for the individual heterogeneity terms of observed individuals, with a numerical integration scheme for the likelihood component corresponding to the unobserved individuals. A similar approach was proposed by Bonner and Schofield (2014), for the case of individual continuous covariates for closed populations (assumed constant within the study period), using a Monte Carlo approach to perform the numerical integration necessary for the likelihood component associated with the unobserved individuals.

2.1.2 Open capture-recapture models

We consider Cormack-Jolly-Seber-type (CJS) models (Cormack, 1964; Jolly, 1965; Seber, 1965) for open populations, which permit entries and exits from the population over the study period. We extend the model described in Section 2.1 and focus on the case where the survival probabilities are modelled as a function of individual time-varying continuous covariates, to explain the individual and temporal variability.

Recall that y_{it} denote the covariate value associated with individual $i = 1, \ldots, n$ at time $t = f_i, \ldots, T$; and set $\boldsymbol{y}_i = \{y_{it} : t = f_i, \ldots, T\}$. The survival probability is specified as a function of the covariate values such that $h(\phi(y_{it})) = \beta_0 + \beta_1 y_{it}$, for all $t = f_i, \ldots, T-1$ and $i = 1, \ldots, n$, where h denotes some function that maps to the interval $[0, 1]$; and β_0 and β_1 denote the corresponding regression parameters. Similarly we may also specify the recapture probabilities to be a function of the covariate such that $h(p(y_{it})) = \gamma_0 + \gamma_1 y_{it}$, for $t = f_i + 1, \ldots, T$ and $i = 1, \ldots, n$, assuming the same functional form for h

for simplicity; and where γ_0 and γ_1 are the associated regression parameters. For notational convenience, we let $\boldsymbol{\beta} = \{\beta_0, \beta_1\}$ and $\boldsymbol{\gamma} = \{\gamma_0, \gamma_1\}$. Further we define the stochastic model for the covariate values, assuming a first-order Markovian process, such that, for $t = f_i, \ldots, T-1$ and $i = 1, \ldots, n$,

$$y_{it+1}|y_{it} \sim N(y_{it} + \mu_t, \sigma_y). \tag{2.3}$$

Clearly, suitable covariate models will be dependent on the given covariate(s) and biological knowledge e.g., a Markov chain with transition Kernel (Bonner and Schwarz, 2006); an additive model (King et al., 2008); a random walk model (Bonner et al., 2010; Langrock and King, 2013). We note that in the case where the covariate value may not be observed at initial capture we also need to specify an initial state distribution for the initial covariate values. However, for simplicity, we assume the covariate values at initial capture are known for each individual, as is the case in our case study, but the approach is easily generalisable.

For capture-recapture studies we do not observe all the individual covariate values. Assuming that the covariate model is stochastic, if an individual is unobserved the associated covariate value is, by definition, also unknown; further if an individual is observed, the covariate value may still not be recorded unless individuals are physically captured so that physical-related information can be retained. Finally, we let $\boldsymbol{\zeta}_i^{obs}$ denote the set of occasions for which the covariate values are observed for individual i and $\boldsymbol{\zeta}_i^{unobs}$ the set of times following initial capture for which the covariate value is unknown for individual i.

To express the likelihood, we let $\boldsymbol{y}_i^{obs} = \{y_{it} : t \in \boldsymbol{\zeta}_i^{obs}\}$ and $\boldsymbol{y}_i^{unobs} = \{y_{it} : t \in \boldsymbol{\zeta}_i^{unobs}\}$ denote the observed and unobserved covariate values associated with individual $i = 1, \ldots, n$. The full set of observed and unobserved covariate values are $\boldsymbol{y}^{obs} = \{\boldsymbol{y}_i^{obs} : i = 1, \ldots, n\}$ and $\boldsymbol{y}^{unobs} = \{\boldsymbol{y}_i^{unobs} : i = 1, \ldots, n\}$. The model parameters are $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\gamma}\}$, with covariate parameters $\boldsymbol{\psi} = \{\mu_1, \ldots, \mu_{T-1}, \sigma_y\}$. The observed data likelihood in Equation (2.1) is given by,

$$f(\boldsymbol{x}^{obs}, \boldsymbol{y}^{obs}|\boldsymbol{\theta}, \boldsymbol{\psi}) = \prod_{i=1}^{n} \left[\int f(\boldsymbol{x}_i^{obs}|\boldsymbol{y}_i, \boldsymbol{\theta}) f(\boldsymbol{y}_i|\boldsymbol{\psi})\, d\boldsymbol{y}_i^{unobs} \right],$$

and is again analytically intractable. The term $f(\boldsymbol{x}_i^{obs}|\boldsymbol{y}_i, \boldsymbol{\theta})$ denotes the complete data likelihood component corresponding to the probability of the capture history $\boldsymbol{x}_i^{obs}$; and $f(\boldsymbol{y}_i|\boldsymbol{\psi})$ the joint probability density function of the covariate values associated with individual i. The probability of a given capture history, conditional on initial capture and all covariate values $\boldsymbol{y}_i$, is given by,

$$f(\boldsymbol{x}_i^{obs}|\boldsymbol{y}_i, \boldsymbol{\theta}) = \left[\prod_{t=f_i}^{\ell_i - 1} \phi(y_{it}) \right] \left[\prod_{t=f_i+1}^{\ell_i} p(y_{it})^{x_{it}^{obs}} (1 - p(y_{it}))^{1 - x_{it}^{obs}} \right] \chi_{i\ell_i},$$

where $\chi_{i\ell_i}$ denotes the probability that individual i is not recaptured again after time ℓ_i. We can express χ_{it} via the recursive function,

$$\chi_{it} = 1 - \phi(y_{it}) + \phi(y_{it})(1 - p(y_{it+1}))\chi_{it+1},$$

such that $\chi_{iT} = 1$, for all $i = 1, \ldots, n$. The covariate model component of the observed data likelihood, conditioning on the initial covariate value (which we assume to be known, but can easily be relaxed by the inclusion of an initial state distribution) is given by,

$$f(\boldsymbol{y}_i|\boldsymbol{\psi}) = \prod_{t=f_i}^{T-1} f(y_{it+1}|y_{it}, \boldsymbol{\psi}),$$

where $f(y_{it+1}|y_{it}, \boldsymbol{\psi})$ denotes the associated density for the given covariate model.

Previous attempts for dealing with missing covariate values include both classical and Bayesian model-fitting approaches. In particular, Catchpole et al. (2004) derived a conditional likelihood approach, conditioning on only the known observed covariate values. This approach is computationally fast but leads to a (potentially substantial) reduction in the precision of the parameter estimates

due to the amount of discarded information (Bonner et al., 2010); and cannot be applied when the observation process parameters (i.e. capture probabilities) are covariate dependent. To use all the available information, Worthington et al. (2015) consider a two-step multiple imputation approach, which involves initially fitting a model to only the observed covariate values and imputing the unobserved covariates before conditioning on these imputed values and using a complete-case likelihood approach for the associated capture histories. Alternatively Langrock and King (2013) numerically approximate the likelihood by finely discretising the integrals and estimate the integral via a hidden Markov model, providing improved parameter estimates. The integral can be made arbitrarily accurate by increasing the level of discretisation. However there is a trade-off between the accuracy of the estimate and the computational expense. Finally, within a Bayesian framework, a data augmentation approach has been applied, treating the missing covariate values as auxiliary variables and sampling from the joint posterior distribution of the parameters and auxiliary variables (Bonner and Schwarz, 2006; King et al., 2008).

2.2 Standard Approaches

In this thesis, we consider one standard model fitting algorithm for each model. In closed population model, Gaussian-Hermite quadrature (GHQ) is considered as a golden approach for a comparison with a proposed method while the Hidden Markov model (HMM) is chosen as a standard approach for a comparison in open population model.

2.2.1 Gaussian-Hermite quadrature

First, we begin to describe the standard approach to the marginal likelihood expressed in Equation (2.2) i.e., integrating out unobserved heterogeneity terms numerically using Gaussian-Hermite quadrature (Steen et al., 1969). The Gaussian-Hermite quadrature has been repeatedly used in the context of closed capture-recapture models in the presence of unobserved heterogeneity (Coull and Agresti,

1999; White and Cooch, 2017; McClintock et al., 2009). The implementation of Gauss-Hermite quadrature is rather straightforward, sampling the likelihood function at points in appropriate regions from certain distributions e.g. a Gaussian density. Let $g(y)$ be the objective (likelihood) function with y being the variable of interest such that the integral has a form of

$$I = \int g(y) \exp(-y^2)\, dy.$$

The above integral can be approximated by

$$\int g(y) \exp\{-y^2\}\, dy = \sum_{j=1}^{q} w_j g(v_j), \tag{2.4}$$

where v_j are the nodes that are symmetric around zero, and the roots of the Hermite polynomials with the associated weight w_j; and q denotes the number of quadrature points.

We may now use Equation (2.4) to approximate the integral of the M_h model defined in Equation (2.2). In this model, we assume the random effect terms to follow a normal distribution such that $\epsilon_i \sim \mathrm{N}(0, \sigma)$ for $i = 1, 2, \ldots, n$ and similarly $\epsilon_0 \sim \mathrm{N}(0, \sigma)$ for all unobserved individuals $i = n+1, \ldots, N$. In our case, the integral of a single individual has a form of

$$I = \int f(x|\boldsymbol{\theta}, \epsilon) \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{\epsilon^2}{2\sigma^2}\right\} d\epsilon.$$

Let $u = \frac{\epsilon}{\sqrt{2\sigma^2}}$ such that

$$I = \int f(x|\boldsymbol{\theta}, \sqrt{2\sigma^2}u) \frac{1}{\sqrt{\pi}} \exp\left\{-u^2\right\} du,$$

which has the same functional form shown in Equation (2.4). Thus, the approximation of Gaussian-Hermite quadrature to the integral of a single individual is given by:

$$\int f(x|\boldsymbol{\theta}, \sqrt{2\sigma^2}u) \frac{1}{\sqrt{\pi}} \exp\left\{-u^2\right\} du \approx \frac{1}{\sqrt{\pi}} \sum_{j=1}^{q} w_j f(x|\boldsymbol{\theta}, v_j\sqrt{2\sigma^2}). \tag{2.5}$$

Applying Equation (2.5) for each individual $i = 1, \ldots, N$, we have the closed approximation to Equation (2.2) given below:

$$L(\boldsymbol{\theta}, \sigma|\boldsymbol{x}) = \frac{N!}{(N-n)!} \prod_{i=1}^{n} \left[\sum_{j=1}^{q} \frac{w_j}{\sqrt{\pi}} f(\boldsymbol{x}_i^{obs}|\theta, v_j\sqrt{2\sigma^2}) \right] \times \prod_{i=n+1}^{N} \left[\sum_{j=1}^{q} \frac{w_j}{\sqrt{\pi}} f(\boldsymbol{x}_i^{unobs}|\theta, v_j\sqrt{2\sigma^2}) \right],$$

where

$$f(\boldsymbol{x}_i|\boldsymbol{\theta}, v_j\sqrt{2\sigma}) = \prod_{t=1}^{T} \left(\frac{1}{1+\exp(-\alpha_t - v_j\sqrt{2\sigma^2})} \right)^{x_{it}} \left(\frac{1}{1+\exp(\alpha_t + v_j\sqrt{2\sigma^2})} \right)^{1-x_{it}}.$$

Note that we use the R package `statmod` to compute the roots of the Hermite polynomial v_j and the associated weight w_j. Gauss-Hermite quadrature is the only approach considered for comparison with the proposed method in the closed M_h model. This method is very straightforward to implement with less computational burden when there is only one random effect included in the model.

2.2.2 The Hidden Markov Model

Now, we describe the second method, Hidden Markov model, for fitting the CJS model in open population. In CJS models with continuous covariates, the number of dimensions of integrals might vary from one individual to another, depending upon the availability of covariate information and increases with the increasing capture occasions, leading to computational complexity. For example, one individual may have more missing covariate than other individuals due to e.g., individuals may be observed more than others resulting in unique capture histories thus unique covariate information. Consequently, the number of integrals may also differ. Langrock and King (2013) approximates such models by discritizing covariates into fine range of intervals and re-expressing the likelihood into the Hidden Markov Model-matrix form. We first describe the definition and the general idea of the hidden Markov Model.

A hidden Markov model is defined by Zucchini et al. (2016) as one kind of

dependent mixture, also known as "Markov-dependent mixture" where a Markov process is used for modeling the underlying system. This hidden Markov model is structured based on two distinct processes: (i) state-dependent process which is observable, and (ii) unobserved (hidden) parameter process. Let X_t denote the observable state-dependent process and C_t denote the unobserved parameter process for $t = 1, 2, \ldots$. To learn about the hidden process C_t, we may only observe X_t. The current state C_t is dependent on the previous state C_{t-1} and

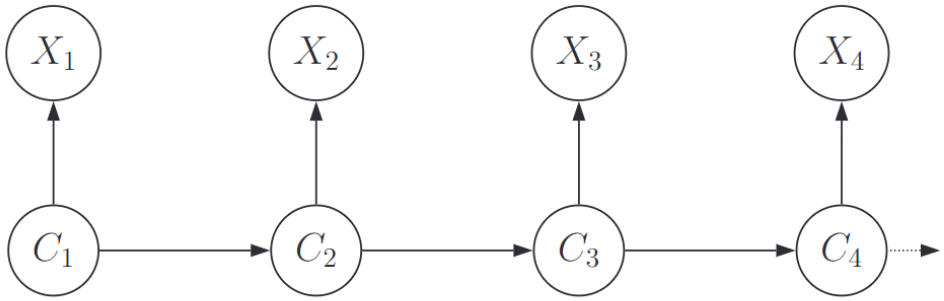


Figure 2.1: The directed graph of the standard Hidden Markov model (HMM)

the only component affecting the distribution X_t as illustrated in Figure (2.1) (Zucchini et al., 2016). Thus, the probability function X_t at time t assuming discrete-valued observations is given by:

$$Pr(X_t = x) = \sum_{i=1}^{m} Pr(C_t = i) Pr(X_t = x | C_t = i), \tag{2.6}$$

where $Pr(C_t = i)$ is the probability of the hidden Markov process being in state i, $Pr(X_t = x|C_t = i)$ denotes the probability function given the current hidden process at state i and time t and m denotes the number of states. Zucchini et al. (2016) re-write this expression in a matrix notation for convenience such that:

$$f(X_t = x) = \boldsymbol{\delta} \Gamma^{t-1} Q(x) \mathbf{1}', \tag{2.7}$$

where $\boldsymbol{\delta}$ is a vector with elements $f(C_t = 1)$, denoting the probability of the hidden process at initial state, $\mathbf{1}$ is a unit vector, and Γ^{t-1} denote the transition

probabilities between states at time $t-1$ i.e.,

$$\Gamma^{t-1} = \begin{pmatrix} \gamma_{11} & \cdots & \gamma_{1m} \\ \cdots & \cdots & \cdots \\ \gamma_{m1} & \cdots & \gamma_{mm} \end{pmatrix}$$

with

$$\gamma_{ij} = Pr(C_{s+1} = i | C_s = i).$$

Finally, let $Q(x)$ denote a diagonal matrix with ith element equal to $Pr(X_t = x|C_t = i)$. This expression can be easily generalised for continuous-valued observations by simply changing the summation in the Equation (2.6) into the integral.

Now, we follow the idea and modify the HMM form in general for mark-capture-recapture models used in Langrock and King (2013), omitting the recovery parameter. Suppose we define the state process for the survival process such that

$$S_t = \begin{cases} 1 & \text{if the individual is alive at time } t; \\ 0 & \text{if the individual is not available for capture at time } t, \end{cases}$$

and let Γ_t denote the transition of probability matrix between two survival states of a single individual such that

$$\Gamma^t = \begin{pmatrix} \phi_t & 1-\phi_t \\ 0 & 1 \end{pmatrix}.$$

Furthermore, we define the observation process x_t for a single individual i.e.,

$$x_t = \begin{cases} 1 & \text{if the individual is observed alive at time } t; \\ 0 & \text{if the individual is unobserved or observed dead at time } t, \end{cases}$$

and we let $Q(x_t)$ denote the corresponding transition probability matrix of the

recapture process at time t of a single individual such that

$$Q(x_t) = \begin{cases} \begin{pmatrix} 1-p_t & 0 \\ 0 & 1 \end{pmatrix} & \text{if } x_t = 0; \\ \begin{pmatrix} p_t & 0 \\ 0 & 0 \end{pmatrix} & \text{if } x_t = 1. \end{cases}$$

Therefore, the likelihood of the classical CJS model for observed individuals $i = 1, 2, \ldots, n$ can be written in the HMM-matrix form as expressed in Equation (2.7) as follows:

$$L(\boldsymbol{\Psi}|x_{it}, y_{it}) = \prod_{i=1}^{n} \boldsymbol{\delta} \left(\prod_{t=f_i+1}^{T} \Gamma_i^{t-1} Q(x_{it}) \right) \mathbf{1}',$$

where $\boldsymbol{\delta}$ represents the row vector with its elements 1 and 0 and $\mathbf{1}'$ is a column vector of 1 of length 2. Now, we extend the HMM expression for standard capture-recapture model by incorporating individual continuous time-varying covariates. Note that we only consider a single continuous covariate in this work for simplicity. Let B_0 and B_m denote the lower and the upper limit respectively containing all possible values of covariates with m being the number of discretization. Suppose $B_r = [B_{r-1}, B_r)$ denote associated intervals with b_r^* being the midpoint of intervals. Let $\Gamma_{it}^{(m)}$ be the system process matrix of size $(m+1) \times (m+1)$ of the i^{th} individual specified by

$$\Gamma_{it}^{(m)} = \begin{pmatrix} \phi_{it}(1)\omega_{it}(1,1) & \ldots & \phi_{it}(1)\omega_{it}(1,m) & 1-\phi_{it}(1) \\ \vdots & \ddots & \vdots & \vdots \\ \phi_{it}(m)\omega_{it}(m,1) & \ldots & \phi_{it}(m)\omega_{it}(m,m) & 1-\phi_{it}(1) \\ 0 & \ldots & 0 & 1 \end{pmatrix},$$

where

$$\omega_{it}(r,c) = \begin{cases} f(y_{i,t+1}|y_{it}) & \text{if } y_{i,t+1}, y_{it} \neq \emptyset, y_{it} \in B_r, y_{i,t+1} \in B_c; \\ f(y_{i,t+1}|b_r^*) & \text{if } y_{i,t+1} \neq \emptyset, y_{it} = \emptyset, y_{i,t+1} \in B_c; \\ f(y_{i,t+1} \in B_c|y_{it}) & \text{if } y_{i,t+1} = \emptyset, y_{it} \neq \emptyset, y_{it} \in B_r; \\ f(y_{i,t+1} \in B_c|b_r^*) & \text{if } y_{i,t+1}, y_{it} = \emptyset; \\ 0 & \text{otherwise.} \end{cases}$$

The function $f(y_{i,t+1}|.)$ denote the probability density function of the covariate process specified by some first-order Markov processes. If the covariate value of individuals at time $t+1$ is observed then the associated individual density function is simply a normal probability function. However, when such a covariate is not available at time $t+1$, the individual covariate process can be approximated by

$$f(y_{i,t+1} \in B_c|b_r^*) = \Phi\left(\frac{b_c - (b_r^* + \mu_t)}{\sigma}\right) - \Phi\left(\frac{b_{c-1} - (b_r^* + \mu_t)}{\sigma}\right),$$

where $\Phi(.)$ denotes the cumulative distribution function of the standard normal distribution. Further, we denote the survival process and recapture process respectively as follows:

$$\phi_{it}(r) = \begin{cases} h(\phi(y_{it})) = \beta_0 + \beta_1 y_{it} & \text{if } y_{it} \neq \emptyset, y_{it} \in B_r; \\ h(\phi(y_{it})) = \beta_0 + \beta_1 b_r^* & \text{if } y_{it} = \emptyset; \\ 0 & \text{otherwise,} \end{cases}$$

and,

$$p_{it}(r) = \begin{cases} h(p(y_{it})) = \gamma_0 + \gamma_1 y_{it} & \text{if } y_{it} \neq \emptyset, y_{it} \in B_r; \\ h(p(y_{it})) = \gamma_0 + \gamma_1 b_r^* & \text{if } y_{it} = \emptyset; \\ 0 & \text{otherwise.} \end{cases}$$

Now, we define the system matrix process for the recapture probability. Suppose

$Q^{(m)}(x_{it})$ be the diagonal matrix of size $(m+1)\times(m+1)$ defined below

$$Q^{(m)}(x_{it}^{obs}) = \begin{cases} \text{diag}(1-p_{it}(1),\ldots,1-p_{it}(m-1),1) & \text{if } x_{it}=0; \\ \text{diag}(p_{it}(1),\ldots,p_{it}(m-1),0) & \text{if } x_{it}=1. \end{cases}$$

The model can be easily extended by regressing the covariate on the recapture probability and such modification would not affect the HMM structure. We also assume that the covariate is always observed at initial states, therefore the row vector $\delta_i^{(m)}$ of length $m+1$ for individuals $i=1,2,\ldots,n$ can be written in the following form

$$\delta_i^{(m)} = \begin{cases} 1 & \text{if } y_{i,f_i} \in B_r; \\ 0 & \text{otherwise.} \end{cases}$$

Finally, putting all these together the likelihood of the continuous-covariate CJS model can be expressed again in the HMM-matrix type notation defined in Equation (2.7)

$$L(\boldsymbol{\Psi}|\boldsymbol{x}^{obs},\boldsymbol{y}) = \prod_{i=1}^{n} \boldsymbol{\delta}_i^{(m)} \left(\prod_{t=f_i}^{T} \Gamma_{i,t-1}^{(m)} Q^{(m)}(x_{it}^{obs}) \right) \mathbf{1}'_{m+1}, \tag{2.8}$$

where $\mathbf{1}'_{m+1}$ denote the column vector of 1 of size $m+1$. Similar to the GHQ, the HMM can be made arbitrarily accurate with large limits and large m. The smaller m means the faster the computation but at the risk of a poorer approximation, whereas the larger m means the more intensive the computational expense. The trade-off between the accuracy and the computational expense becomes the common challenge in discretization-based methods. Additionally, the HMM may not scale up well with high-dimensional spaces i.e., integrating two or more covariates may exponentially increase computational expense. This curse of dimensionality becomes another rising-potential issue when dealing with larger data.

2.3 Laplace Approximation

Now, we present a more efficient likelihood-based method for approximating the marginal likelihood for two capture-recapture models using Laplace approxima-

tions. The Laplace approximation is a numerical closed-form approximation of an integral and can be regarded as an alternative form of Gauss-Hermite quadrature (Liu, 1994). The underlying idea of the Laplace integration is to approximate the negative log-likelihood by a second (or higher) order Taylor expansion. When the negative-log likelihood is well approximated by a Gaussian curve the Laplace approximation can be shown to have high precision. In particular, in a study from Liu (1994) the standard error of the estimate was shown to be of order $O(m^{-1})$, where m denotes the number of observations. Adding in further leading terms can further improve the accuracy to the order of $O(m^{-2})$ (Wong and Li, 1992; Breslow and Lin, 1995; Raudenbush et al., 2000). Using an analytical expression for the Laplace approximation is generally only feasible when the dimension of the integral is small (for example, of dimension 2 or 3) due to the computational complexity in computing the higher order derivatives. However, numerical approximations of the required derivatives can be obtained using automatic differentiation. In particular we use the Template Model Builder (TMB) automatic differentiation tool developed by Kristensen et al. (2016) that enables the use of the Laplace approximation using a computationally efficient implementation. We apply the Laplace approximation to individual heterogeneity capture-recapture models applying the TMB tool for numerically calculating the derivatives, when these are analytically intractable. We note that Laplace approximations have been used previously for capture-recapture models but within a Bayesian context for approximating the marginal posterior densities of the parameter of interest (Smith, 1991; Chavez-Demoulin, 1999).

Now, we describe technical details of Laplace approximations. Suppose we have a smooth function $g(y)$ which can be expressed in the integration form as follows:

$$L = \int \exp\left(-\lambda g(y)\right) dy, \tag{2.9}$$

where $g(y)$ has a local minimum function at y^*. To approximate the function

$g(y)$, the Taylor expansion is used on the function of g around y^* such that

$$g(y) = g(y^*) + g'(y^*)(y - y^*) + \frac{g''(y^*)(y - y^*)^2}{2} + \ldots,$$

where y^* is a local minimum thus it follows that $g'(y^*) = 0$ and $g''(y^*) > 0$. The function of L can be rewritten as:

$$\begin{aligned} L \propto \int \exp\left(-\lambda g(y)\right) dy &= \int \exp\left(-\lambda \left[g(y^*) + \frac{g''(y^*)(y - y^*)^2}{2!} + \ldots\right]\right) dy \\ &= \exp\left(-\lambda g(y^*)\right) \int \exp\left(-\lambda \left[\frac{g''(y^*)(y - y^*)^2}{2!} + \frac{g^{(3)}(y^*)(y - y^*)^3}{3!} \ldots\right]\right) dy. \end{aligned}$$

Thus, if we only use the second order of the expansion, then we simply have the following expression

$$L = \exp\left(-\lambda g(y^*)\right) \int \exp\left(-\lambda \left[\frac{g''(y^*)(y - y^*)^2}{2}\right]\right) dy,$$

which is simply an unnormalized Gaussian density with $\sigma^2 = [\lambda g''(y^*)]^{-1}$. Therefore, the Laplace approximation of the second order on the integration can be expressed as

$$L = \exp\left(-\lambda g(y^*)\right) \sqrt{\frac{2\pi}{\lambda g''(y^*)}} \left\{1 + O\left(\frac{1}{\lambda}\right)\right\}. \tag{2.10}$$

Suppose now we want to use more leading orders of Taylor series for improved precision thus more accurate approximation such that

$$\begin{aligned} L &= \exp\left(-\lambda g(y^*)\right) \sqrt{\frac{2\pi}{\lambda g''(y^*)}} \\ &\quad \times E\left\{\exp\left(-\lambda \left[\frac{g^{(3)}(y^*)(y - y^*)^3}{3!} + \frac{g^{(4)}(y^*)(y - y^*)^4}{4!} + \ldots\right]\right)\right\} \end{aligned} \tag{2.11}$$

where E represents the expectation with respect to the normal density with mean zero and variance $\sigma^2 = [\lambda g''(y^*)]^{-1}$. The latter term on the right hand side can be expanded in a power series such that $\exp(X) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots$. Suppose we

define τ_3 and τ_4 respectively as follows:

$$\tau_3 = \frac{g^{(3)}(y^*)(y-y^*)^3}{3!}, \tau_4 = \frac{g^{(4)}(y^*)(y-y^*)^3}{4!}, \ldots,$$

such that Equation (2.11) can be rewritten as follows:

$$L = \exp\left(-\lambda g(y^*)\right)\sqrt{\frac{2\pi}{\lambda g''(y^*)}} \times E\left\{\exp\left(-\lambda\left[\tau_3 + \tau_4 + \tau_5 + \ldots\right]\right)\right\}.$$

Raudenbush et al. (2000) presented the general solution for the expectation which can be easily used up to the $m-th$ order given below:

$$\begin{aligned}
E(\tau_m) =& 0 \text{ for odd } m;\\
E(\tau_m) =& \frac{(m-1)(m-3)}{m!}(\sigma^2)^{m/2} g^{(m)}(y^*)\\
& \text{for even } m;\\
E(\tau_{m1}, \tau_{m2}) =& 0 \text{ for odd } (m_1+m_2);\\
E(\tau_{m1}, \tau_{m2}) =& \frac{(m_1+m_2-1)(m_1+m_2-3)}{m_1!m_2!}(\sigma^2)^{(m_1+m_2)/2} g^{(m_1)}(y^*)g^{(m_2)}(y^*)\\
& \text{for even } (m_1+m_2).
\end{aligned}$$

Let the order of expansion used in this case is up to the fourth order. Note that Raudenbush et al. (2000) suggested that $E\left[1-\lambda\tau_4+\frac{1}{2}\lambda^2\tau_3^2\right]$ to be highly accurate to approximate the integral; the expectation of odd orders disappear. Thus, we consider to use this approximation such that we have

$$\begin{aligned}
E(\tau_4) &= \frac{(3)(1)}{4!}(\sigma^2)^2 g^{(4)}(y^*) = \frac{3}{24}\frac{g^{(4)}(y^*)}{\lambda^2(g^{(2)}(y^*))^2};\\
E(\tau_3\tau_3) &= \frac{(5)(3)}{3!3!}(\sigma^2)^3 g^{(3)}(y^*)g^{(3)}(y^*) = \frac{5}{12}\frac{(g^{(3)}(y^*))^2}{\lambda^3(g^{(2)}(y^*))^3}.
\end{aligned}$$

In sum, the complete expression of the approximation to the integral up to the fourth order is

$$L = \exp\left(-\lambda g(y^*)\right)\sqrt{\frac{2\pi}{\lambda g''(y^*)}} \times \left[1 - \frac{3}{24}\frac{g^{(4)}(y^*)}{\lambda(g^{(2)}(y^*))^2} + \frac{5}{24}\frac{(g^{(3)}(y^*))^2}{\lambda(g^{(2)}(y^*))^3}\right]. \quad (2.12)$$

Note that we make a numerical comparison between the second and fourth order expansion for a particular application in Section 3.1.2.

This approximation can be straightforwardly extended for multivariate cases of a vector of random variable $\boldsymbol{y} = \{y_1, y_2, \ldots, y_p\}$. First, we rewrite the function of L given above in the form of a vector

$$\begin{aligned} L &= \int \exp\left(-\lambda g(\boldsymbol{y})\right) h(\boldsymbol{y})\, d\boldsymbol{y} \\ &= \int \exp\left(-\lambda \left[g(\boldsymbol{y}^*) + \frac{g''(\boldsymbol{y}^*)(\boldsymbol{y}-\boldsymbol{y}^*)^T(\boldsymbol{y}-\boldsymbol{y}^*)}{2} + \ldots \right]\right) d\boldsymbol{y} \\ &= \exp\left(-\lambda g(\boldsymbol{y}^*)\right) \int \exp\left(-\lambda \left[\frac{g''(\boldsymbol{y}^*)(\boldsymbol{y}-\boldsymbol{y}^*)^T(\boldsymbol{y}-\boldsymbol{y}^*)}{2} + \ldots \right]\right) d\boldsymbol{y}. \end{aligned}$$

The second order of the expansion for the multivariate case is given below

$$L = \exp\left(-\lambda g(\boldsymbol{y}^*)\right) \int \exp\left(-\lambda \left[\frac{g''(\boldsymbol{y}^*)(\boldsymbol{y}-\boldsymbol{y}^*)^T(\boldsymbol{y}-\boldsymbol{y}^*)}{2} \right]\right) d\boldsymbol{y},$$

which is also an unnormalized multivariate Gaussian density with $\Sigma = [\, g''(\boldsymbol{y}^*)]^{-1}$ and $g''(\boldsymbol{y}^*)$ is the hessian matrix of $g(\boldsymbol{y})$ evaluated at $\boldsymbol{y}^*$. Thus, the Laplace approximation of the second order on the vector $\boldsymbol{y}$ can be expressed as

$$\begin{aligned} L &= \exp\left(-\lambda g(\boldsymbol{y}^*)\right) \left(\frac{2\pi}{\lambda}\right)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}} \int \frac{1}{(\sqrt{2\pi})^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{(\boldsymbol{y}-\boldsymbol{y}^*)^T \Sigma^{-1} (\boldsymbol{y}-\boldsymbol{y}^*)}{2}\right) d\boldsymbol{y} \\ &= \exp\left(-\lambda g(\boldsymbol{y}^*)\right) \left(\frac{2\pi}{\lambda}\right)^{\frac{p}{2}} |g''(\boldsymbol{y}^*)|^{-\frac{1}{2}} \left\{ 1 + O\left(\frac{1}{\lambda}\right) \right\}. \end{aligned} \tag{2.13}$$

We also note that Laplace approximations can be thought as a special case of the saddlepoint approximation where the variable y expressed in Equation (2.9) is complex valued e.g., $y = a + ib$ (Strawderman, 2000; Barndorff-Nielsen and Cox, 1979). Similar to Laplace approximation, the saddlepoint method aims to solve the integral at $g'(y) = 0$ which are saddlepoints, maximizing the contribution to the integral L around these points.

2.3.1 Closed M_h-type models

We consider the general M_{tbh} model with corresponding likelihood specified in Equation (2.2). The integrand of the likelihood can be rewritten in exponential form, such that,

$$\begin{aligned} f(\boldsymbol{x}|\boldsymbol{\theta},\sigma) \;\propto\; & \frac{N!}{(N-n)!}\prod_{i=1}^{n}\left[\int \exp\{-g(\boldsymbol{x}_i^{obs},\epsilon_i|\boldsymbol{\theta},\sigma)\}\,d\epsilon_i\right] \\ & \times \prod_{i=n+1}^{N}\left[\int \exp\{-g(\boldsymbol{x}_i^{unobs},\epsilon_0|\boldsymbol{\theta},\sigma)\}\,d\epsilon_0\right], \end{aligned} \tag{2.14}$$

where,

$$g(\boldsymbol{x}_i,\epsilon_i|\boldsymbol{\theta},\sigma) = -\log f(\boldsymbol{x}_i|\boldsymbol{\theta},\epsilon_i) - \log f(\epsilon_i|\sigma),$$

for $i \in \{1,\ldots,N\}$. Dropping the subscript notation on i for notational brevity, let $\hat{\epsilon}$ denote the value of ϵ that minimises $g(\boldsymbol{x},\epsilon|\boldsymbol{\theta},\sigma)$ given model parameters $\boldsymbol{\theta}$ and heterogeneity parameter σ, so that $g'(\boldsymbol{x},\epsilon|\boldsymbol{\theta},\sigma) = 0$ and $g''(\boldsymbol{x},\epsilon|\boldsymbol{\theta}) > 0$. A second-order Taylor series expansion is given by,

$$g(\boldsymbol{x},\epsilon|\boldsymbol{\theta},\sigma) \approx g(\boldsymbol{x},\hat{\epsilon}|\boldsymbol{\theta},\sigma) + \frac{g''(\boldsymbol{x},\hat{\epsilon}|\boldsymbol{\theta},\sigma)(\epsilon-\hat{\epsilon})^2}{2}.$$

Laplace's method approximates the integrands in Equation (2.14) using the properties of normal density functions as derived in Equation (2.10) such that the contribution to the observed data likelihood takes the form,

$$\begin{aligned} \int \exp\{-g(\boldsymbol{x},\epsilon|\boldsymbol{\theta},\sigma)\}\,d\epsilon \;&\approx \exp\{-g(\boldsymbol{x},\hat{\epsilon}|\boldsymbol{\theta},\sigma)\}\int \exp\left\{\tfrac{(\epsilon-\hat{\epsilon})^2}{2g''(\boldsymbol{x},\hat{\epsilon}|\boldsymbol{\theta},\sigma)^{-1}}\right\}d\epsilon \\ &= \exp\{-g(\boldsymbol{x},\hat{\epsilon}|\boldsymbol{\theta},\sigma)\}\sqrt{\tfrac{2\pi}{g''(\boldsymbol{x},\hat{\epsilon}|\boldsymbol{\theta},\sigma)}}. \end{aligned} \tag{2.15}$$

To improve the accuracy of the approximation, we can also consider a higher-order Laplace approximation involving higher-order derivatives. We use the fourth-order Taylor expansion in Equation (2.12) to obtain the fourth-order Laplace approximation. Applying the one-dimensional Laplace approximation on the in-

tegral yields,

$$\int \exp\{-g(\boldsymbol{x},\epsilon|\boldsymbol{\theta},\sigma)\}\,d\epsilon \;\approx\; \exp\{-g(\boldsymbol{x},\hat{\epsilon}|\boldsymbol{\theta},\sigma)\}\sqrt{\frac{2\pi}{g''(\boldsymbol{x},\hat{\epsilon}|\boldsymbol{\theta},\sigma)}} \times\left[1+\frac{5(g^{(3)}(\boldsymbol{x},\hat{\epsilon}|\boldsymbol{\theta},\sigma))^2}{24(g''(\boldsymbol{x},\hat{\epsilon}|\boldsymbol{\theta},\sigma))^3}-\frac{3g^{(4)}(\boldsymbol{x},\hat{\epsilon}|\boldsymbol{\theta},\sigma)}{24(g''(\boldsymbol{x},\hat{\boldsymbol{\epsilon}}|\boldsymbol{\theta},\sigma))^2}\right], \tag{2.16}$$

where $g^{(3)}(.)$ and $g^{(4)}(.)$ denote the third and fourth derivatives with respect to the random effect term ϵ. A closed form expression for the fourth-order Laplace approximation is presented below where we assume the recapture probabilities are specified using the logistic function, so that $\text{logit}(p_{it}) = \alpha_t + \lambda S_{it} + \epsilon_i$.

We calculate the derivatives of the joint density of the complete data likelihood and associated random effects for M_h-type models with respect to the individual random effects to obtain the Laplace approximation of the (marginal) observed data likelihood. Recall that for $i = 1,\ldots,n$, we let $g(\boldsymbol{x}_i^{obs},\epsilon_i|\boldsymbol{\theta},\sigma)$ denote the objective function corresponding to the negative of the log of the joint density of the observed capture history for individual i, denoted $\boldsymbol{x}_i^{obs}$, and associated individual random effect density, ϵ_i, given the model model parameters, $\boldsymbol{\theta}$, and individual heterogeneity standard deviation, σ, i.e. $g(\boldsymbol{x}_i^{obs},\epsilon_i|\boldsymbol{\theta},\sigma) = -\log f(\boldsymbol{x}_i^{obs},\epsilon_i|\boldsymbol{\theta},\sigma)$. Similarly we let $g(\boldsymbol{x}_i^{unobs},\epsilon_0|\boldsymbol{\theta},\sigma)$ denote the analogous objective function for the null history, $\boldsymbol{x}_i^{unobs}$ and the associated ϵ_0 for all unobserved individuals $i = n+1,\ldots,N$. For $i = 1,\ldots,N$, the corresponding objective function is of the form:

$$\begin{aligned} g(\boldsymbol{x}_i,\epsilon_i|\boldsymbol{\theta},\sigma) &= -\log f(\boldsymbol{x}_i|\boldsymbol{\theta},\epsilon_i) - \log f(\epsilon_i|\sigma) \\ &= -\sum_{t=1}^{T}\left[x_{it}\log(p_{it}) + (1-x_{it})\log(1-p_{it})\right] + \frac{1}{2}\log(2\pi\sigma^2) + \frac{\epsilon_i^2}{2\sigma^2}, \end{aligned}$$

where again, $\boldsymbol{x}_i = \{\boldsymbol{x}_i^{obs},\boldsymbol{x}_i^{unobs}\}$ and $\epsilon_i = \epsilon_0$ for all unobserved individuals $i = n+1,\ldots,N$. For notational simplicity, we let $\eta_{it} = \alpha_t + \lambda S_{it} + \epsilon_i$. The capture

probabilities are assumed to be logistically regressed on the covariate, so that

$$p_{it} = \frac{\exp(\eta_{it})}{1+\exp(\eta_{it})} = \frac{1}{1+\exp(-\eta_{it})}; \text{ and } 1-p_{it} = \frac{1}{1+\exp(\eta_{it})}.$$

The first derivative of $g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma)$ with respect to ϵ_i is given by,

$$\begin{aligned}\frac{dg(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma)}{d\epsilon_i} &= \sum_{t=1}^{T}\left[\frac{-x_{it}\exp(-\eta_{it})}{1+\exp(-\eta_{it})} + \frac{(1-x_{it})\exp(\eta_{it})}{1+\exp(\eta_{it})}\right] + \frac{\epsilon_i}{\sigma^2} \\ &= \sum_{t=1}^{T}\left[\frac{-x_{it}}{1+\exp(\eta_{it})} + \frac{(1-x_{it})}{1+\exp(-\eta_{it})}\right] + \frac{\epsilon_i}{\sigma^2} \\ &= \sum_{t=1}^{T}(p_{it}-x_{it}) + \frac{\epsilon_i}{\sigma^2}.\end{aligned}$$

Similarly, the second derivative is given by:

$$\begin{aligned}g''(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma) = \frac{d^2 g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma)}{d\epsilon_i^2} &= \sum_{t=1}^{T}\left[\frac{\exp(-\eta_{it})}{\{1+\exp(-\eta_{it})\}^2}\right] + \frac{1}{\sigma^2} \\ &= \sum_{t=1}^{T}\left[p_{it}(1-p_{it})\right] + \frac{1}{\sigma^2}.\end{aligned}$$

Thus the negative log likelihood of the M_h-type model as given in Equation (2.2) can be estimated using the second order Laplace approximation,

$$\begin{aligned}-\log f(\boldsymbol{x}|\boldsymbol{\theta}, \sigma) = &-\log(N!) + \log(N-n)! \\ &+ \sum_{i=1}^{n}\left[g(\boldsymbol{x}_i^{obs}, \hat{\epsilon}_i|\boldsymbol{\theta}, \sigma) + \frac{1}{2}\log g''(\boldsymbol{x}_i^{obs}, \hat{\epsilon}_i|\boldsymbol{\theta}, \sigma) - \frac{1}{2}\log(2\pi)\right] \\ &+ \sum_{i=n+1}^{N}\left[g(\boldsymbol{x}_i^{unobs}, \hat{\epsilon}_0|\boldsymbol{\theta}, \sigma) + \frac{1}{2}\log g''(\boldsymbol{x}_i^{unobs}, \hat{\epsilon}_0|\boldsymbol{\theta}, \sigma) - \frac{1}{2}\log(2\pi)\right].\end{aligned}$$

To obtain the closed form of the fourth order Laplace approximation, we require higher order derivatives, in terms of the third and the fourth derivatives, with

respect to ϵ_i. For $i = 1, \ldots, N$, the third derivative of $g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma)$ is given by:

$$\begin{aligned} g^{(3)}(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma) &= \frac{d^3 g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma)}{d\epsilon_i^3} \\ &= \sum_{t=1}^{T} \left[\frac{\exp(\eta_{it})}{\{1+\exp(\eta_{it})\}^3} - \frac{\exp(-\eta_{it})}{\{1+\exp(-\eta_{it})\}^3} \right] \\ &= \sum_{t=1}^{T} \left[p_{it}\left(1-p_{it}\right)^2 - p_{it}^2\left(1-p_{it}\right) \right]. \end{aligned}$$

Similarly, the fourth derivative of $g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma)$ is given by:

$$\begin{aligned} g^{(4)}(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma) &= \frac{d^4 g(\boldsymbol{x}_i, \epsilon_i|\boldsymbol{\theta}, \sigma)}{d\epsilon_i^4} \\ &= \sum_{t=1}^{T} \left[\frac{\exp(\eta_{it})}{\{1+\exp(\eta_{it})\}^4} - \frac{4\exp(2\eta_{it})}{\{1+\exp(\eta_{it})\}^4} + \frac{\exp(-\eta_{it})}{\{1+\exp(-\eta_{it})\}^4} \right] \\ &= \sum_{t=1}^{T} \left[p_{it}(1-p_{it})^3 - 4p_{it}^2(1-p_{it})^2 + p_{it}^3(1-p_{it}) \right]. \end{aligned}$$

These analytic expressions can be substituted into the higher order Laplace approximation given in Equation (2.17). The negative likelihood function including the inner optimisation is coded in C++ utilizing the TMB library. The estimation of the model parameters is obtained by subsequently minimizing the negative log likelihood which is evaluated in TMB at given parameter values (for $\boldsymbol{\theta}$ and σ) using standard optimisation routines in `R`.

2.3.2 Open CJS model

Now, we apply the Laplace approximation for the open population CJS model with individual time-varying continuous covariates. Note that due to the complexity in approximating integrals in higher dimensions, we simply consider the second order approximation for this case. Recall that $\boldsymbol{y}_i = \{\boldsymbol{y}_i^{obs}, \boldsymbol{y}_i^{unobs}\}$. The observed data likelihood in Equation (2.1) can be expressed in the form,

$$f(\boldsymbol{x}^{obs}, \boldsymbol{y}_i^{obs}|\boldsymbol{\theta}, \boldsymbol{\psi}) = \prod_{i=1}^{n} \left[\int \exp\{-g(\boldsymbol{x}_i^{obs}, \boldsymbol{y}_i|\boldsymbol{\theta}, \boldsymbol{\psi})\}\, d\boldsymbol{y}_i^{unobs} \right],$$

such that for $i = 1, \ldots, n$,

$$g(\boldsymbol{x}_i^{obs}, \boldsymbol{y}_i|\boldsymbol{\theta}, \boldsymbol{\psi}) = -\sum_{t=f_i}^{\ell_i - 1} \log \phi(y_{it}) - \sum_{t=f_i+1}^{\ell_i} \{x_{it}^{obs} \log p(y_{it}) + (1 - x_{it}^{obs}) \log(1 - p(y_{it}))\}$$
$$- \log \chi_{i\ell_i} - \sum_{t=f_i}^{T-1} \log f(y_{it+1}|y_{it}, \boldsymbol{\psi}).$$

Applying the multivariate k-dimension second-order Laplace approximation given in Equation (2.13), we obtain,

$$\int \exp\{-g(\boldsymbol{x}_i^{obs}, \boldsymbol{y}_i|\boldsymbol{\theta}, \boldsymbol{\psi})\}\, d\boldsymbol{y}_i^{unobs} \approx \exp\{-g(\boldsymbol{x}_i^{obs}, \boldsymbol{y}_i^{obs}, \hat{\boldsymbol{y}}_i^{unobs}|\boldsymbol{\theta}, \boldsymbol{\psi})\}(2\pi)^{k/2}$$
$$\times |g''(\boldsymbol{x}^{obs}, \boldsymbol{y}_i^{obs}, \hat{\boldsymbol{y}}_i^{unobs}|\boldsymbol{\theta}, \boldsymbol{\psi})|^{-1/2},$$

where $\hat{\boldsymbol{y}}_i^{unobs}$ denotes the value of $\boldsymbol{y}_i^{unobs}$ that minimises $g(\boldsymbol{x}_i^{obs}, \boldsymbol{y}_i|\boldsymbol{\theta}, \boldsymbol{\psi})$. There is no closed form for the derivatives (due to the χ term) and so we use numerical automatic differentiation function provided in TMB, to evaluate the Laplace approximation for this model. Furthermore, obtaining higher order of Laplace approximations for multivariate cases in TMB is not feasible in the current version therefore we only consider the second order approximation.

Chapter 3

Capture-Recapture Studies: Simulation Study and Applications

In this Chapter, we present simulation studies for both closed and open population models under different scenarios e.g., different parameter settings. Then, we demonstrate the model fitting on real examples: (i) St. Andrews golf tees dataset from Borchers et al. (2002) for closed M_h model; and (ii) meadow voles from Nichols et al. (1992) for CJS model with a continuous covariate. We make some comparisons in terms of accuracy and computational expense among different model fitting algorithms and discuss the results. All methods use the R package `TMB` from Kristensen et al. (2016) for computational comparability for calculating the MLEs of the parameters and are coded in C++ and the bespoke R. All codes used in Chapter 3 can be found at `https://github.com/riki-herliansyah/capture-recapture.git`.

3.1 Simulation Study

First, we describe simulation studies considered for each model using 1000 datasets for each simulation. For the first simulation study, we consider closed population M_h-type models. In this setting, we apply both the second and fourth order Laplace approximations (LA2 and LA4, respectively) and compare with a Gauss-Hermite quadrature (GHQ) approximation. The second simulation is conducted on the open population CJS model with covariates where we compare the second order of Laplace approximation with an HMM approximation (Langrock and King, 2013). We discuss each of these cases in turn below.

3.1.1 Closed M_h-type models

We consider two simulation studies. The first investigates the performance of the Laplace approximation for the four individual heterogeneity models i.e., M_h, M_{th}, M_{bh} and M_{tbh}, given a set of parameter values; while the second focuses on the individual heterogeneity standard deviation, σ, given model M_h. For each dataset simulated we fit the generating model using both the second-order (LA2) and fourth-order (LA4) Laplace approximations and compare with a GHQ approach using 50 quadrature points as a "gold standard".

For each method we calculate the MLE and 95% confidence intervals using a non-parametric bootstrap approach. The main reasons for using a non-paramteric bootstrap simulation are to relax the asymptotic normality assumption on parameters of estimate and to avoid potentially unreliable confidence intervals derived directly from obtained standard errors as the estimate N may be often close to the boundary (King and McCrea, 2019). In particular, to compute the associated 95% confidence intervals, we simulate B bootstrap replicates of the data, such that each simulated dataset is of the same size as the original data (i.e. same number of observed individuals). Each bootstrap dataset is simulated by randomly drawing with replacement each observed capture history with equal probability. We fit the model to each bootstrap dataset and obtain the associated MLEs of the model parameters. We add the original MLEs of the parameters to this set of values, so that we have $B + 1$ parameter values corresponding to the MLEs of the parameters from the original and bootstrap datasets. The 95% confidence interval for each parameter is calculated as the associated lower and upper 2.5% quantile values of the $B + 1$ values. In practice we used $B = 999$. We also compute the associated coverage probabilities (CP) for the given parameters i.e., the proportion of the associated 95% confidence interval that contain the true value of the parameter, and the average relative bias (RB) for each parameter of interest given by:

$$\text{RB} = \frac{1}{1000} \sum_{sim=1}^{1000} \frac{(\hat{\theta}_{sim} - \theta)}{\theta},$$

where $\hat{\theta}_{sim}$ denotes the estimated parameter for the bootstrap dataset $sim = 1, \ldots, 999$ and original dataset (when $sim = 1000$).

For the first simulation study we set the total population size to be $N = 100$ individuals with $T = 6$ capture occasions and consider the four different individual heterogeneity models: M_h, M_{th}, M_{bh} and M_{tbh}. The parameters specified for the parameters were motivated by the snowshoe hare dataset (see for example, Cormack (1989); Baillargeon and Rivest (2007)), after fitting the given models to the data. For the time-invariant models (M_h and M_{bh}) we set $\alpha_t = \alpha = -1$ for $t = 1, \ldots, T$. For the time-varying models, for M_{th} we set $\boldsymbol{\alpha} = \{\alpha_t :$

$t = 1, \ldots, T\} = \{-1.75, -0.91, -1.44, -1.03, -1.22, -0.67\}$; and for M_{tbh}, $\boldsymbol{\alpha} = \{-1.49, -0.29, -0.44, 0.15, 0.10, 0.81)\}$, based on the parameter estimates from fitting the models to the data. For the models with a behavioural effect (M_{bh} and M_{tbh}), we specify a "trap happy" response, with $\lambda = 0.75$. Finally, for the individual heterogeneity component we set $\sigma = \sqrt{0.75}$. For models M_h and M_{bh} the probability of an individual not being observed within the study is 0.8; for models M_{th} this probability is 0.77; and for model M_{tbh}, 0.95. 1000 datasets were simulated for each model.

Table 3.1 provides the average relative bias (RB), 95% coverage probabilities (CP) and the average width of the 95% confidence intervals (width) for the 1000 simulated datasets for the two parameters of interest, population size, N, and individual heterogeneity standard deviation, σ, for each model and model-fitting approach. In general, across all approaches and models, the MLEs of the parameters, N and σ, appear to be consistently slightly negatively biased for all models. This bias is, however, consistently less for the LA4 and GHQ approaches, relative to LA2. The differences in RB become more distinct as shown in Figure 3.1. Given this larger bias, it is perhaps unsurprising that the associated coverage probabilities for LA2 are also consistently lower. This suggests that the second-order Taylor series expansion in the standard Laplace approximation is not sufficient to approximate the integral in the observed data likelihood of the M_h-type models. However adding in the higher-order Laplace approximation terms does improve the performance of the algorithm, with very similar performance between LA4 and GHQ in terms of both relative bias (though the bias appears to be very slightly less using the Laplace approximation) and coverage probabilities.

Table 3.1: Simulation results in terms of averaged relative bias (RB) and 95% coverage probabilities (CP) for 1000 simulated datasets for the different individual heterogeneity models fitted via Laplace approximations, second-order (LA2) and fourth-order (LA4), and Gauss-Hermite quadrature (GHQ).

Models	Methods	N			σ		
		RB	CP	Width	RB	CP	Width
M_h	LA2	-0.023	0.838	53.569	-0.205	0.833	1.093
	LA4	-0.004	0.884	36.689	-0.098	0.914	1.108
	GHQ	-0.004	0.882	47.833	-0.103	0.905	1.146
M_{th}	LA2	-0.003	0.854	75.585	-0.177	0.861	1.184
	LA4	0.002	0.893	45.736	-0.081	0.916	1.146
	GHQ	0.005	0.894	63.262	-0.084	0.910	1.221
M_{bh}	LA2	0.033	0.898	124.626	-0.142	0.875	1.349
	LA4	0.001	0.921	53.521	-0.075	0.930	1.086
	GHQ	0.005	0.920	66.280	-0.078	0.930	1.119
M_{tbh}	LA2	-0.010	0.859	35.314	-0.144	0.911	0.949
	LA4	-0.007	0.875	47.026	-0.039	0.957	0.994
	GHQ	-0.007	0.873	41.516	-0.044	0.950	0.986

For the second simulation study, we consider model M_h and investigate the performance of the Laplace approximations and GHQ for differing values of the individual heterogeneity standard deviation, σ. Increasing the individual heterogeneity increases the variability of the survival probabilities of individuals in the study and thus is likely to be increasingly challenging, see White and Cooch (2017) for further discussion. Motivated by the golf tees data (see Section (3.2.1)), we set $N = 250$, $T = 8$ and $\alpha = -1.5$ for the simulation study. We then consider a range of values for the individual heterogeneity standard deviation, such that $\sigma = 1, 1.5, 2$. For each σ value we simulate 1000 datasets.

Table 3.2 provides the average relative bias, the 95% coverage probabilities and mean 95% confidence intervals width for N and σ for the different model-fitting approaches. As for the previous simulation study, the LA2 approach has the poorest performance; while LA4 and GHQ perform better and have similar

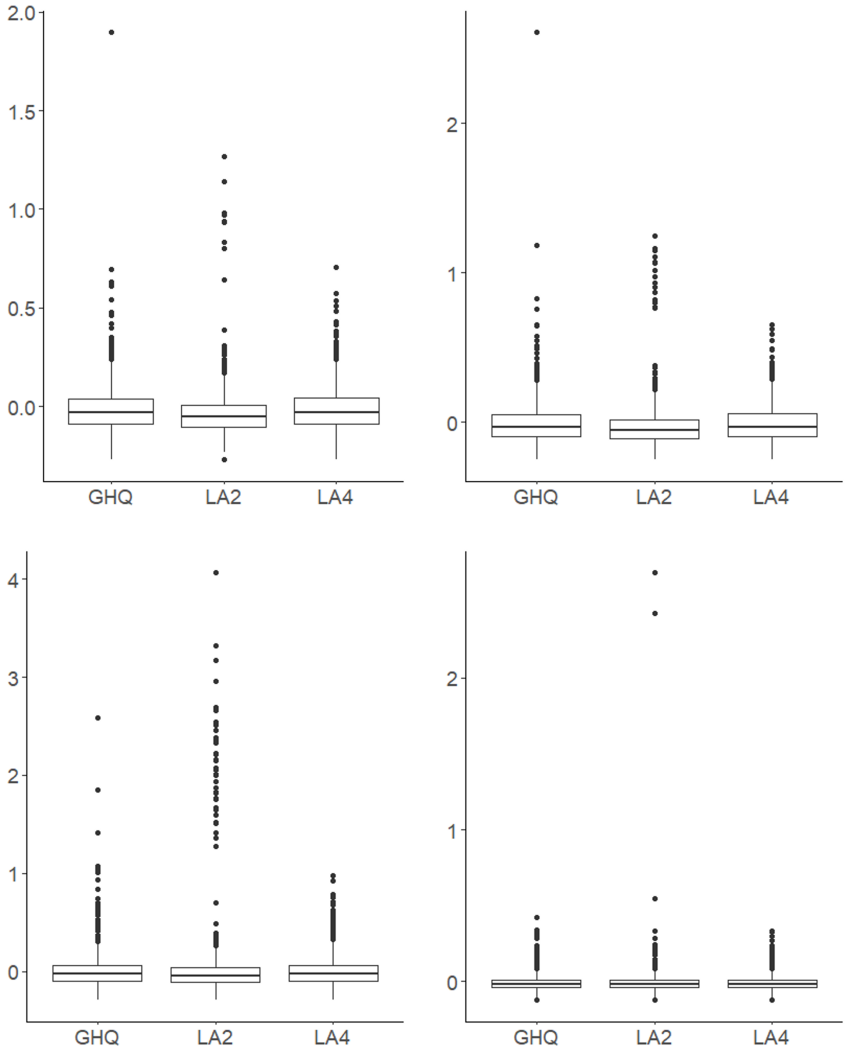


Figure 3.1: Distributions of relative bias (RB) of N fitted to 1000 simulated datasets for four different models: (top left) M_h, (top right) M_{th}, (bottom left) M_{bh}, and (bottom right) M_{tbh}.

relative biases and coverage probabilities. However, interestingly despite these similar coverage probabilities, for GHQ the width of the confidence interval increases as σ increases, due to the long tails for the upper bound; while this relationship is not present for LA4 with similar width confidence intervals across the different values of σ. These findings are consistent with the real data golf tees example in Section 3.2.1, where the 95% non-parametric confidence intervals of the quadrature approach are substantially wider for each of the four individual heterogeneity models. Finally we note that both the standard error of σ and relative bias of N both increase as σ increases for all the model-fitting approaches.

Table 3.2: Simulation results in terms of averaged relative bias (RB) and 95% coverage probabilities (CP) for 1000 simulated datasets from model M_h for values of $\sigma = 1, 1.5, 2$ and fitted via Laplace approximations, second-order (LA2) and fourth-order (LA4), and Gauss-Hermite quadrature (GHQ).

σ	Methods	N			σ		
		RB	CP	Width	RB	CP	Width
2.0	LA2	-0.053	0.883	84.145	-0.100	0.853	0.922
	LA4	0.010	0.929	75.001	0.008	0.952	1.007
	GHQ	0.017	0.929	158.29	-0.001	0.944	1.315
1.5	LA2	-0.034	0.879	93.322	-0.087	0.873	0.794
	LA4	-0.003	0.928	79.066	-0.013	0.935	0.768
	GHQ	-0.003	0.928	127.852	-0.019	0.933	0.942
1.0	LA2	-0.021	0.868	95.648	-0.084	0.869	0.682
	LA4	-0.002	0.903	76.665	-0.016	0.932	0.642
	GHQ	-0.002	0.903	85.726	-0.020	0.929	0.668

3.1.2 Comparison second and fourth order expansion on M_h model

To compare the likelihood between the second and fourth order expansions of the Laplace approximation, we fit the M_h model to snowshoe hare data setting α and σ to be fixed and obtain the corresponding likelihoods. In particular, we set $\alpha = -1$ and $\sigma = \{2, 4\}$, and evaluate the likelihood function for different values of N i.e., for $N \in [70, 150]$ as shown in Figure 3.2. From Figure 3.2 we note that when σ is "small" ($\sigma = 2$), the difference between two expansions is very small. However, for larger values of σ (e.g. $\sigma = 4$) the discrepancy between the second and fourth order expansion increases. Further, the mode of the likelihood function, as a function of N, shifts to the left, which would thus lead to a lower MLE estimate for the total population size (at least given fixed α and σ). This is the case for both the second and fourth order approximations but is substantially more marked for the larger value of σ. This finding is consistent with the results from Table 3.1 where the second order Laplace approximation underestimates N for model M_h. Thus, Breslow and Lin (1995) also suggested that the higher order expansion of Laplace approximation is required for bias-correction. Note that in this example we assume σ and α to be fixed to compare the difference between the second and fourth approximations. However, in these parameters will generally

be correlated with N in the hyperparameter space in practice, potentially leading to more complicated relationship.

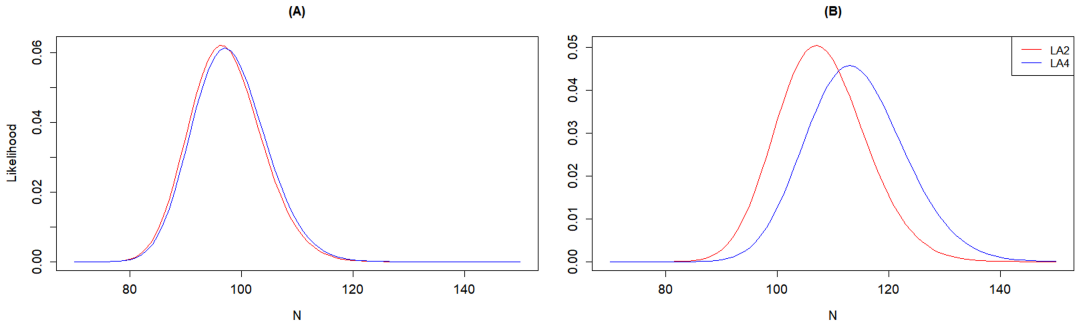


Figure 3.2: The likelihood comparison of second and fourth order expansions of Laplace approximation for M_h model fitted to snowshoe hare data, setting $\alpha = -1$; (A) $\sigma = 1$ and (B) $\sigma = 2$.

3.1.3 CJS model with missing continuous covariates

We consider the CJS model where we specify the survival probability as a function of a single individual covariate i.e., $\text{logit}(\phi_{it}(y_{it})) = \beta_0 + \beta_1 y_{it}$ for 2 different covariate models:

- Model 1: $y_{i,t+1}|y_{it} \sim N(y_{it}, \sigma_y)$ (a simple random walk);
- Model 2: $y_{i,t+1}|y_{it} \sim N(y_{it}+\mu_t, \sigma_y)$ (a random walk with additional temporal effects).

We consider a range of scenarios motivated by the real meadow vole data considered in Section 3.2.2: (i) we initially set $n = 200$ and consider a constant capture probability for two different regimes ($p = 0.5, 0.75$) for studies of length $T = 4, 6$; (ii) to investigate the sample size, we then set $T = 4$ and repeat the simulation study but increase n to 400 (with same constant capture probabilities as before). For all studies we set $\sigma_y = 1.2$. The initial covariate value for each individual at the time of initial capture is simulated from a Normal distribution with mean of 15 and standard deviation of 2, i.e. $y_{if_i} \sim \text{N}(15, 2)$ with f_i randomly sampled from $\{1, \ldots, T-1\}$. For covariate model 2 we simulate $\mu_t \sim \text{N}(0, 1)$ for $t = 1, \ldots, T-1$, independently for each simulated dataset. The survival probability was specified as a logistic regression on the individual covariate with regression parameters $\beta_0 = -3.0$ and $\beta_1 = 0.2$. For each model and

parameter combination we simulated 1000 datasets. For the HMM, we discretise the individual covariate value into $m = 20$ intervals, re-expressing the integral as a summation. The maximum likelihood estimates (MLE) can be easily obtained via standard optimisation algorithm. In this simulation, we consider several optimisation algorithms in R such as `optim` and `nlminb`. To compute the associated confidence interval (CI) of the MLE, we assume that the MLE is asymptotically normal such that

$$\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{SE(\hat{\boldsymbol{\beta}})} \sim \mathrm{N}(0, 1).$$

Thus, the Wald-type confidence interval of the MLE for 95% can be computed using the following equation:

$$\hat{\boldsymbol{\beta}} \pm 1.96 SE(\hat{\boldsymbol{\beta}}).$$

For each dataset, we fit the Laplace approximation and compared with an HMM-approximation (Langrock and King, 2013). Tables 3.3 and 3.4 provides the corresponding averaged relative biases (RB), 95% coverage probabilities (CP) and the mean 95% confidence interval widths of the regression coefficients, β_0 and β_1 across the generated datasets. Overall, both the Laplace and HMM approximations perform well with all coverage probabilities $> 94\%$ and small relative biases. Relative biases seem to be very identical across simulation studies considered between the second order Laplace approximation and the HMM as shown in Figure 3.3. Unsurprisingly, the relative biases are lower for smaller sampling occasions in scenario (i) with $n = 200$. The same finding is also expected in the scenario (ii) where the relative biases decrease as the sample sizes increase. The relative biases are consistently smaller for the Laplace approach, across the different scenarios considered, with the difference decreasing as p increases (as the amount of missing information decreases).

3.1.4 Computational comparisons

We compare the computational time of the Laplace approximations with other competing methods (GHQ for closed populations; and HMM approximations for

Table 3.3: Simulation results for CJS model with $n = 200$ considering varying values of $T = 4, 6$, $p = 0.5, 0.75$ for the two different covariate models. Presented are the relative bias (RB), 95% coverage probabilities (CP) and the mean 95% confidence interval widths (Width) of the regression parameters for the survival probabilities averaged over 1000 simulated datasets using the Laplace approximation and a hidden Markov model (HMM) approximation respectively. Model 1 corresponds to a simple random walk on the continuous covariate; model 2 to a random walk with additional temporal effects.

(a) Model 1

	p	Methods	β_0			β_1		
			RB	CP	Width	RB	CP	Width
$T = 4$	0.50	Laplace	0.060	0.955	5.585	0.064	0.956	0.378
		HMM	0.062	0.955	5.530	0.064	0.956	0.373
	0.75	Laplace	0.009	0.961	4.269	0.008	0.956	0.282
		HMM	0.010	0.961	4.266	0.009	0.956	0.282
$T = 6$	0.50	Laplace	0.036	0.962	4.485	0.031	0.962	0.293
		HMM	0.041	0.962	4.479	0.035	0.960	0.293
	0.75	Laplace	0.016	0.961	3.733	0.016	0.960	0.243
		HMM	0.017	0.961	3.731	0.016	0.960	0.244

(b) Model 2

	p	Methods	β_0			β_1		
			RB	CP	Width	RB	CP	Width
$T = 4$	0.50	Laplace	0.051	0.968	5.455	0.049	0.967	0.375
		HMM	0.084	0.966	5.545	0.098	0.969	0.384
	0.75	Laplace	0.020	0.959	4.237	0.018	0.959	0.284
		HMM	0.027	0.961	4.245	0.026	0.958	0.282
$T = 6$	0.50	Laplace	0.054	0.957	4.503	0.048	0.956	0.296
		HMM	0.078	0.959	4.513	0.074	0.956	0.298
	0.75	Laplace	0.036	0.943	3.732	0.036	0.945	0.245
		HMM	0.042	0.943	3.734	0.042	0.946	0.245

Table 3.4: Simulation results for CJS model with $T = 4$ considering varying values of $n = 200, 400$, $p = 0.5, 0.75$ for the two different covariate models. Presented are the relative bias (RB), 95% coverage probabilities (CP) and the mean 95% confidence interval widths (Width) of the regression parameters for the survival probabilities averaged over 1000 simulated datasets for the Laplace approximation and a hidden Markov model (HMM) approximation respectively. Model 1 corresponds to a simple random walk on the continuous covariate; model 2 to a random walk with additional temporal effects.

(a) Model 1

	p	Methods	β_0			β_1		
			RB	CP	Width	RB	CP	Width
$n = 200$	0.50	Laplace	0.060	0.955	5.585	0.064	0.956	0.378
		HMM	0.062	0.955	5.530	0.064	0.956	0.373
	0.75	Laplace	0.009	0.961	4.269	0.008	0.956	0.282
		HMM	0.010	0.961	4.266	0.009	0.956	0.282
$n = 400$	0.50	Laplace	0.028	0.958	3.737	0.030	0.957	0.250
		HMM	0.030	0.958	3.730	0.033	0.956	0.250
	0.75	Laplace	0.003	0.943	2.981	0.003	0.941	0.197
		HMM	0.004	0.943	2.980	0.004	0.941	0.197

(b) Model 2

	p	Methods	β_0			β_1		
			RB	CP	Width	RB	CP	Width
$n = 200$	0.50	Laplace	0.051	0.968	5.455	0.049	0.967	0.375
		HMM	0.084	0.966	5.545	0.098	0.969	0.384
	0.75	Laplace	0.020	0.959	4.237	0.018	0.959	0.284
		HMM	0.027	0.961	4.245	0.026	0.958	0.282
$n = 400$	0.50	Laplace	0.024	0.958	3.713	0.023	0.956	0.254
		HMM	0.050	0.960	3.746	0.053	0.956	0.258
	0.75	Laplace	0.016	0.949	2.971	0.016	0.949	0.199
		HMM	0.022	0.948	2.966	0.023	0.950	0.200

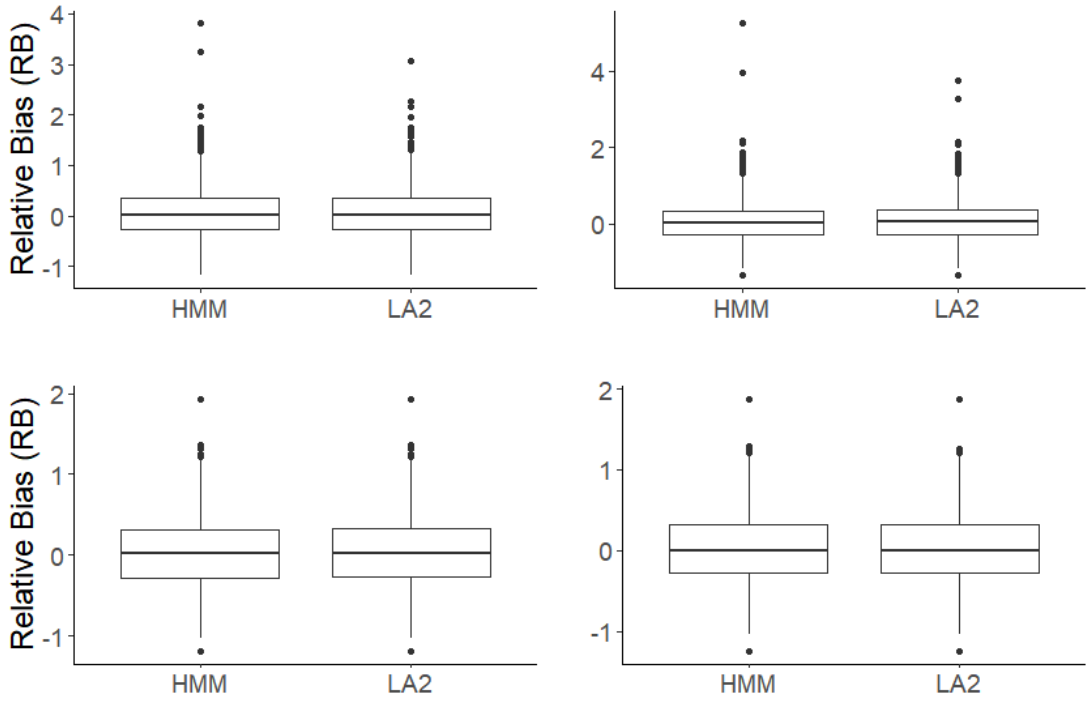


Figure 3.3: Distributions of relative bias (RB) of β_0 and β_1 for model 1 (top left and right) and of β_0 and β_1 for model 2 (bottom left and right) fitted to simulated dataset for $T = 4$ and $n = 200$.

open populations) for the simulation studies. Each algorithm was run on a 1.70 GHz Intel Core CPU i5-8350U computer with Windows 10. All methods were implemented in the TMB package for comparability.

For the closed population simulation study, we consider model M_h. The Laplace approximations (LA2 and LA4) are nearly twice as fast as the GHQ approach (using 50 quadrature points) in evaluating the log-likelihood function. For more complex models, the difference in computational speed is even greater. For example, for model M_{th}, the differential in computation speed increases to 8-10 times faster for the Laplace approximations compared to GHQ. Finally, we note that both approximations are "fast" in terms of absolute computational time for these models, with the maximisation of the likelihood to obtain the MLEs of the parameters in the order of seconds.

For the more computationally challenging CJS model with individual continuous covariates, we focus on the computational times associated with model 2. For example, to fit this model for a moderate sized dataset where $n = 500$ and

$T = 10$, the Laplace approximation took an average of 1.55 seconds to evaluate the observed data likelihood function; while the HMM approximation using 20 intervals took an average of 70.13 seconds (this is approximately 45 times slower than the Laplace approximation). The computational time of the HMM approach is dependent on the number of intervals used, and this relationship is non-linear. For example, if we increase the number of intervals from 20 to 30 the computational time of the HMM more than doubles, such that it is over 100 times slower than the Laplace approximation.

The Laplace approximation is consistently faster than the considered "gold-standard" approaches (GHQ for M_h-type models; and HMM approximation for the CJS model with individual time-varying continuous covariates). Further, the Laplace approximation requires no tuning parameters, while these alternatives do require some specification (e.g. number of quadrature points or intervals), with an associated trade-off between the computational time and accuracy.

3.2 Examples

We consider two case studies corresponding to the closed M_h-type models (St. Andrews golf tees data) and open CJS model with individual time-varying continuous covariates (meadow voles). We again compare the Laplace approximation with alternative approaches.

3.2.1 St. Andrews golf tees

We consider the St. Andrews golf tees data from Borchers et al. (2002). The dataset consists of $N = 250$ tees differing in size, colour and visibility resulting in individual capture heterogeneity. A total of $T = 8$ independent observers (i.e. capture occasions) were assigned to predefined transects and recorded all golf tees observed. A total of 546 observations were recorded and $n = 162$ unique tees observed in the study (additional size/colour data were not recorded).

Borchers et al. (2002) showed that omitting the presence of individual heterogeneity vastly underestimates the true population size, thus we consider the

set of closed population models with individual heterogeneity present. Table 3.5 provides the estimated population size and 95% non-parametric bootstrap confidence interval fitted to the four individual heterogeneity M_h-type models for the Laplace approximations and GHQ approach. In general, the higher order Laplace approximation (LA4) and the GHQ give relatively similar maximum likelihood estimates for N, varying from approximately 242 to 262; while the lower order Laplace approximation (LA2) gives somewhat varying estimates, as previously observed within the simulation study in Section 3.1.1. The higher order Laplace approximation tends to consistently produce slightly larger estimates of N than the GHQ approach for all individual heterogeneity models. For example, the estimates of N under M_h model are 251.3 and 242.4 for the LA4 and GHQ approaches, respectively.

The bootstrap confidence intervals for the GHQ approach are noticeably wider than the LA4 approximation, with a consistently substantially larger upper limit (we note that the lower limit is bounded by the number of observed individuals). In particular, the highest upper bound of the LA4 approximation is approximately 310, with the width ranging from 90 to 100; while the lowest upper bound in the GHQ approach is 358 and the widths generally double. King et al. (2008) report a similar uncertainty interval as the LA4 approximation, using a Bayesian approach with a model-averaged 95% credible interval of [194, 288], over the same set of models. Further, the estimate of σ for each of these models is relatively large (approximately 2 for all fitted models). We note that as observed in the second simulation study (see Section 3.1.1), as σ increases, the width of the 95% confidence intervals for N using the GHQ approach are increasingly larger than for the Laplace approximations, yet providing comparable coverage probabilities.

Finally, we compared the computational times. In general, the computational speed is fast in terms of absolute time, and on the order of milliseconds for each of the methods. However, comparing the Laplace approximations to GHQ, the Laplace approaches are approximately twice as fast for model M_h, increasing to fives times as fast for model M_{tbh}.

Table 3.5: The estimates of the total population of St. Andrews golf tees (the true population, $N = 250$) and associated non-parametric bootstrap 95% confidence interval (in brackets) for the four individual heterogeneity models using second-order (LA2) and fourth-order (LA4) Laplace approximations and Gauss-Hermite quadrature (GHQ) with 50 quadrature points.

Model	LA2	LA4	GHQ
M_h	224	251	242
	(192, 285)	(199, 284)	(198, 360)
M_{th}	224	254.1	242.6
	(191, 284)	(198, 282)	(197, 358)
M_{bh}	272	262	261
	(196, 353)	(203, 308)	(202, 406)
M_{tbh}	350	260	255
	(189, 439)	(198, 299)	(195, 450)

3.2.2 Meadow voles

We consider capture-recapture data collected on meadow voles (*Microtus pennsylvanicus*) at Patuxent Wildlife Research Center, Laurel, Maryland over $T = 4$ capture occasions from 1981-1982 (Nichols et al., 1992). A total of 512 voles were observed over the study period. When an individual was observed, its corresponding body mass was also recorded. We follow Bonner and Schwarz (2006) by classifying individuals as immature (body mass $\leq$ 22 g) and mature (body mass $>$ 22 g) and consider data only for the mature voles observed for the first time prior to the final capture occasion. This provides a total of $n = 199$ unique capture histories corresponding to a total of 450 observed sightings and associated body mass recordings. We note that approximately 40% of body mass recordings were unknown, following initial capture.

We follow Bonner and Schwarz (2006) and consider the model where the survival and recapture probabilities are dependent on body mass. In particular, we let y_{it} denote the body mass of individual $i = 1, \ldots, n$ at time $t = f_i, \ldots, T-1$ and set,

$$\text{logit}(\phi_{it}) = \beta_0 + \beta_1 y_{it}; \qquad \text{and} \qquad \text{logit}(p_{it+1}) = \gamma_0 + \gamma_1 y_{it+1}.$$

Similarly, we specify the model for body mass to be of the form,

$$y_{it+1}|y_{it} \sim \mathrm{N}(y_{it} + \mu_t, \sigma^2),$$

for $i = 1, \ldots, n$ and $t = f_i, \ldots, T-1$. The set of model parameters is $\boldsymbol{\theta} = \{\beta_0, \beta_1, \gamma_0, \gamma_1\}$, and the heterogeneity parameters $\boldsymbol{\psi} = \{\mu_1, \mu_2, \mu_3, \sigma^2\}$. All covariate values were recorded when an individual was observed, so we do not need to specify an initial distribution for body mass.

We compared the Laplace approximation, HMM-approximation using 20 intervals (Langrock and King, 2013), two-step multiple imputation approach (Worthington et al., 2015) and Bayesian data augmentation approach (as fitted by Bonner and Schwarz (2006)). Figure 3.4 provides the parameter estimates of the fitted CJS model for the different approaches, in terms of MLE/posterior mean and associated 95% confidence/credible interval. All approaches provide generally similar results, and in particular we note that the results for the Laplace approximation and HMM-approximation are very similar. There are however some differences in the results obtained via the two-step multiple imputation approach for the parameters associated with recapture probability in terms of the MLEs and with noticeably larger confidence intervals (the latter is also true for the estimation of σ). Some differences are not unexpected since the regression model parameters are estimated independently by fitting the given covariate model to the observed covariate values only, ignoring the capture observations.

Finally, we compared the associated computational times for each of the different approaches. To obtain the MLE of the parameters, the Laplace approximation takes < 1 second; the multiple imputation approach approximately 3 seconds; and the HMM-approximation approximately 9 seconds for these data. Thus, using 999 bootstrap replicates to obtain the 95% confidence intervals, the computation times are of the order of approximately 15 minutes; 45 minutes and 2.5 hours for the Laplace approximation, imputation approach and HMM-approximation, respectively. We note that the Bayesian data augmentation approach using a Markov chain Monte Carlo algorithm will depend on the updating algorithm used, number of iterations required for the Markov chain to converge to the sta-

tionary distribution and to obtain reliable posterior summary statistics with small Monte Carlo error following convergence and thus will be comparatively slower, particularly in relation to the fast Laplace approximation.

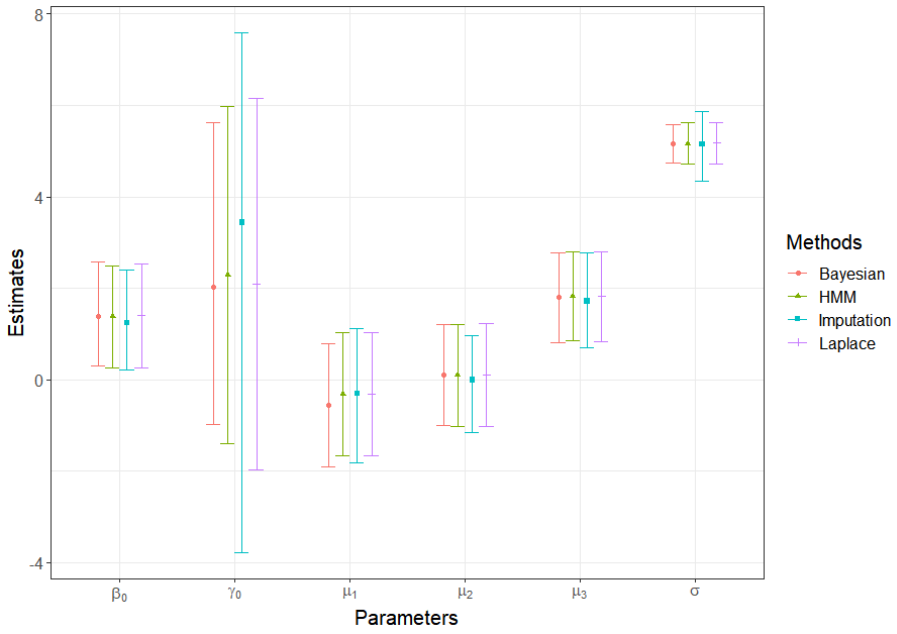
3.3 Discussion

In this first work, we describe a Laplace approximation to estimate the analytically intractable observed data likelihood for capture-recapture models in the presence of individual heterogeneity. The complexity of the likelihood function influences the order of the Laplace approximation (Taylor expansion) that can be analytically calculated. For closed population M_h-type models a fourth-order expression can be analytically derived; whereas for the CJS model with continuous time-varying covariates we use automatic differentiation (via TMB) to obtain the necessary numerical derivatives, as they are analytically intractable. The Laplace approximation for these models provide a reliable and efficient mechanism for evaluating the likelihood function, and hence obtaining the maximum likelihood estimates, and associated confidence intervals.

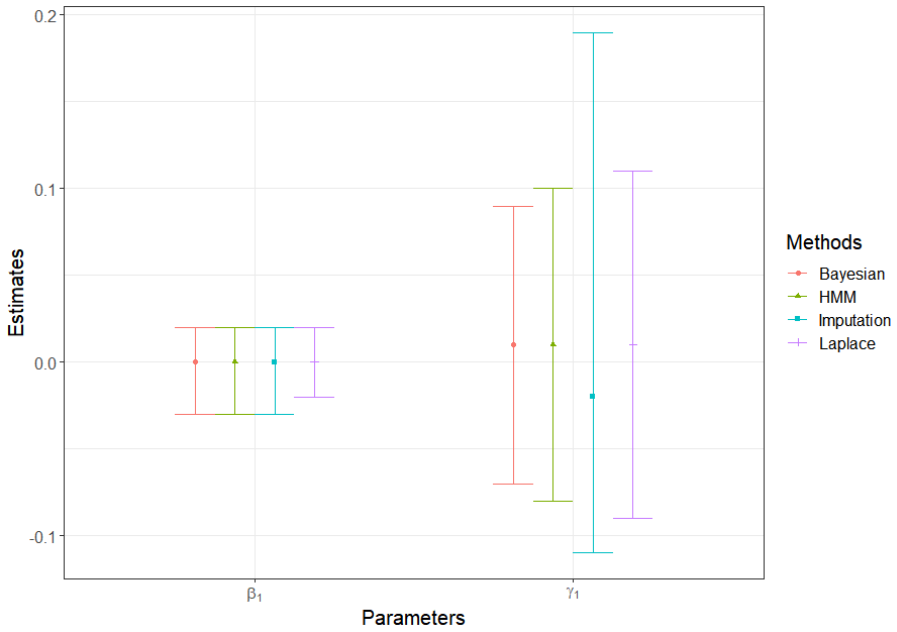
In particular, comparing this approach to the current "gold-standards", namely GHQ for M_h-type models, and an HMM-approximation for the CJS model with continuous covariates, the Laplace approximation consistently performs at least as well but at substantially lower computational cost. However, we note that for the M_h-type models, the fourth-order Laplace approximation is required for this performance. Further, for the scenarios considered, the coverage probabilities were essentially identical between the fourth-order Laplace approximation and the GHQ approach, yet the width of the associated 95% confidence intervals were significantly larger for the quadrature approach with increasing individual heterogeneity variability (i.e. σ). Wider confidence intervals for increasing variability is also discussed by White and Cooch (2017) when using the quadrature approach. Understanding this apparent difference is the focus of current research. It is possible that this phenomenon could be due to the sampled quadrature points within the GHQ approach remaining the same for different values of σ, i.e. the

approximation algorithm being used is independent of the value of σ. However this is not the case for the Laplace method, where the integral is approximated by evaluating the given function at the maximum point, which may depend on σ and other parameters in the model thus the precision of the approximation also changes as parameters change e.g., σ. In addition the Laplace approximation is scalable to higher dimensions and thus this approach is potentially a very attractive avenue to pursue for more complex models (for example, in the presence of multiple continuous individual covariates), particularly as the GHQ and the HMM approximation generally suffer from the curse of dimensionality (Langrock and King, 2013). Huber et al. (2004) explain that summation-based approaches such as GHQ and HMM may have larger bias when the dimension of integrals increases. The reason of this behaviour is due to GHQ and HMM-approximation being based on pre-specified and fixed quadrature points. These fixed points can easily become too coarse, particular in higher dimensions, such that the peak of the log-likelihood is missed.

Employing TMB for fitting closed and open population models improves the computational aspect and enables automatic Laplace approximations (second order expansion). TMB is designed for fitting complex statistical models, with or without random effects, using the classical likelihood method. Although the user needs to have some minimal knowledge of C++, since the objective functions i.e., negative log-likelihood functions are defined in C++, they do not need to derive the Hessian matrix as this is calculated numerically. Thus, in general, only minimal effort is required to translate use TMB templates. Further, Kristensen et al. (2016) provide various examples of how to use TMB for fitting different models, which are a useful source of learning guides for those new to TMB. The difficulties in writing C++ will depend on the complexity of the objective function of interest. However, post processing and plotting can still be implemented using `R`.



(a) Regression intercepts and covariate model parameter estimates



(b) Regression slope parameters for the survival and recapture probabilities

Figure 3.4: Comparison of the parameter estimates (MLE or posterior mean) and associated 95% uncertainty interval (non-parameteric confidence interval or symmetric credible interval) for the Laplace approximation, HMM-approximation, multiple imputation and Bayesian data augmentation approaches fitted to the meadow voles dataset.

Chapter 4

Camera-Trapping Studies: Models and Methods

Estimating the density of wildlife populations is essential in ecology for management and conservation. Camera trapping is increasingly becoming a preferred monitoring tool for sampling animal populations due to their non-invasive nature and efficiency (Srbek-Araujo and Chiarello, 2005; O'Connell et al., 2011). When individual animals are able to be uniquely identified, capture-recapture (CR) methods are commonly applied to obtain estimates of population size as we have shown in Chapters 2 and 3. CR models often incorporate different sources of heterogeneity such as individual, behavioural and/or temporal heterogeneity (Otis et al., 1978). The logical extension of abundance is density, where we enumerate the number of animals counted or surveyed per some areal unit. Let $\mathbb{S}$ be the state-space or observation window i.e., a region which contains possible values of locations of animals. The spatial density of animals, D, is then defined to be the total population size N divided by the total area of the state-space $|\mathbb{S}|$ i.e., $D = N/|\mathbb{S}|$. Reliable estimates of density, however, require information about the size of the area used by the target animal and the effective size of the area surveyed. For example, spatial capture-recapture (SCR) models, using an array of traps permits spatial density to be estimated (Efford, 2004; Borchers and Efford, 2008; Royle and Young, 2008; Borchers et al., 2014; Efford et al., 2016; Stevenson et al., 2021). However, standard CR-type approaches that require individuals to be uniquely identified are often infeasible in practice e.g., many species may be difficult to identify from camera trap images due to similar markings and/or poor quality images.

Various analytical approaches have been developed for population density estimation of unmarked individuals, including, for example, random encounter models (Rowcliffe et al., 2008), camera trap distance sampling (Howe et al., 2017), and random encounter and staying time (Nakashima et al., 2018). Such approaches often require the observations to be independent, which is often violated in practice (Palencia et al., 2021), and do not take into account spatial variability (Anderson, 2001; Pollock et al., 2002; Royle et al., 2014).

Spatially explicit models, also referred to as unmarked SCR, directly consider two underlying sources of heterogeneity: detectability and spatial heterogeneity

(Chandler and Royle, 2013). Conditional on the number of individuals in the study area, and their associated activity centres, the number of animals observed at each camera trap in a given time interval is assumed to be Poisson, with some specified mean. Fitting this model faces two challenges due to (i) the unobserved activity centers; and (ii) the total unknown number of individuals. A Bayesian data augmentation approach is often used in the presence of such challenges, which involves imputing the unknown activity centres and applying a super-population approach to deal with the unknown number of individuals (Ramsey et al., 2015; Evans and Rittenhouse, 2018; Connor et al., 2022). However, this approach does not scale to large populations, and can exhibit (very) slow and poor mixing within the Markov chain Monte Carlo (MCMC) algorithm.

We consider a large unmarked SCR dataset relating to barking deer (*Muntiacus muntjak*) from Ujung Kulon National Park (UKNP), Indonesia. A total of 1095 camera trap sightings over 77 cameras are recorded over a period of 4 months. The size of the population (in the thousands) is such that a super-population approach is computationally very demanding. Thus, motivated by these data, we develop a new efficient Bayesian model fitting approach, which removes the necessity of *a priori* setting an upper bound on the population size, and also permits the direct specification of a prior on the total population size. Different approaches have been developed to fit the model involving an unknown total population size. In particular, in the context of unmarked SCR dataset, Bayesian data augmentation has been primarily applied to obtain estimates of the posterior distribution of the model parameters (Chandler and Royle, 2013). More generally, for marked capture-recapture models, a trans-dimensional algorithm Reversible Jump MCMC, RJMCMC has been considered to update the total population size, which is an explicit parameters in the model (Fienberg et al., 1999; **?**; King and Brooks, 2008; McLaughlin, 2019). See, for example, Schofield and Barker (2014) for further discussion and a review of these, and additional, approaches.

The remainder of the Chapter is organised as follows. In Section 4.1 we describe the unmarked SCR model and associated assumptions, before describing

the different model fitting approaches in Section 4.2. In Section 4.2, we introduce two new reversible jump (RJ)MCMC algorithms and describe the algorithm implementation in R. Finally, we discuss prior distributions for each parameter in the model.

4.1 Spatially Explicit Models

First, we begin to describe the model and the associated assumptions, in particular we extend the model described in Section 2.1. Suppose that there are T sampling periods, and within each sampling period there are J camera detectors (sites), and we assume that their locations are the same over time, but this can be relaxed. The location of the camera traps is denoted by coordinates, $\boldsymbol{X} = \{\boldsymbol{X}_j\} \in \mathbb{R}^2$ for $j = 1, \ldots, J$. Individuals observed by the cameras are not uniquely identifiable, so that the data corresponds to the number of sightings observed by camera j at sampling period t, denoted by n_{jt}, for $j = 1, \ldots, J$ and $t = 1, \ldots, T$. Thus the full set of observed data are denoted by $\boldsymbol{n}^{obs} = \{n_{jt}^{obs} : j = 1 \ldots, J; t = 1, \ldots, T\}$. Camera traps were placed sufficiently close to each other such that individuals may be detected at multiple camera locations at each sampling period $t = 1, \ldots, T$. We define the (unobserved) encounter history for individual $i = 1, \ldots, N$, denoted by $\boldsymbol{x}_i^{unobs} = \{x_{ijt}^{unobs} : j = 1, \ldots, J; t = 1, \ldots, T\}$, such that, x_{ijt}^{unobs} denotes the number of times individual i is observed by detector $j = 1, \ldots, J$ for sampling occasion $t = 1, \ldots, T$.

Further, we define the latent variables $\boldsymbol{S}_i \in \mathbb{R}^2$, corresponding to the activity centre for individual $i = 1, \ldots, N$, representing the individual spatial heterogeneity. Royle et al. (2014) defined the activity centres of animals as the centroid of individuals' home ranges or the centroid of individuals' activities during the study period. Thus, the activity centre is also known as home range centres and is represented as point locations. Therefore, activity centres are generally unknown for any individual as they are not being tracked over time. However, this concept is not viewed the same as animals' territories as in the classical definition. Royle

et al. (2014) simply regard this as a conceptual device and a transient quantity i.e., locations where animals stay during the sampling time.

Following Chandler and Royle (2013), we also assume that for each individual $i = 1, \ldots, N$, given their activity centre $\boldsymbol{S}_i$, the number of times the individual is observed by trap $j = 1, \ldots, J$ at time $t = 1, \ldots, T$ is independent over detectors and individuals (and homogeneous over time) and has a Poisson distribution, such that,

$$x_{ijt}^{unobs} | \boldsymbol{S}_i, \boldsymbol{X}_j \sim \text{Poisson}(\lambda_{ij}(\boldsymbol{S}_i)).$$

The mean (or encounter rate), $\lambda_{ij}(\boldsymbol{S}_i)$ is specified to account for the correlation in counts from neighbouring detectors and spatial heterogeneity over individuals. In particular, we specify the encounter rate for individual i at trap j to be a function of the distance from the associated activity centre of the individual $\boldsymbol{S}_i$ and the trap location, $\boldsymbol{X}_j$. Specifying the encounter rate to be of half-normal form (Efford, 2004) we have that,

$$\lambda_{ij}(\boldsymbol{S}_i, \boldsymbol{X}_j) = \lambda_0 \exp\left(-\frac{||\boldsymbol{X}_j - \boldsymbol{S}_i||^2}{2\sigma^2}\right),$$

where λ_0 denotes an underlying baseline detection parameter and σ the scale parameter controlling the rate of decay in the detection rate.

In practice, we do not observe the individual encounter histories, $\boldsymbol{x}_i^{mis}$ due to the nature of unmarked populations (i.e., animals are not uniquely identifiable), but only the total trap-count data, $\boldsymbol{n}^{obs}$. The trap-count data can be regarded as (reduced) information summaries of the individual-level data via the relationship,

$$n_{jt}^{obs} = \sum_{i=1}^{N} x_{ijt}^{mis},$$

for $j = 1, \ldots, J$ and $t = 1, \ldots, T$. Using standard properties for independent Poisson random variables, the aggregate count-trap data can be modelled as:

$$n_{jt}^{obs} | N, \boldsymbol{S}, \boldsymbol{X} \sim \text{Poisson}(\Lambda_j(\boldsymbol{S})),$$

where $\Lambda_j(\boldsymbol{S}) = \sum_{i=1}^{N} \lambda_{ij}(\boldsymbol{S}_i)$. Note that Λ_j does not depend on the sampling periods t since we assume a time-invariant spatial explicit model. Thus we can simplify the model specification by considering the total trap counts over all sampling times, denoted by $n_{j.}^{obs} = \sum_{t=1}^{T} n_{jt}^{obs}$, for each trap $j = 1, \ldots, J$. Then,

$$n_{j.}^{obs} | N, \boldsymbol{S}, \boldsymbol{X} \sim \text{Poisson}\left(T\Lambda_j(\boldsymbol{S})\right),$$

independently for each $j = 1, \ldots, J$.

Note that we can also incorporate covariate information in the model parameters. For example, including trap or activity centre level covariates in the baseline detection rate, to account for changes in detection rate due to different environmental factors, at either the trap level (affecting detectability of individuals given the trap location) or activity centre level (to represent differences in detectability of individuals due to varying environment summarised by trap location) (Chandler and Royle, 2013; Evans and Rittenhouse, 2018; Connor et al., 2022). To complete the model specification, we assume the unobserved activity centres, $\boldsymbol{S}_i$ for $i = 1, \ldots, N$ are independent and uniformly distributed over the region, $\mathbb{S}$, (and do not change over the sampling period) so that

$$\boldsymbol{S}_i \sim \text{Uniform}(\mathbb{S}).$$

The corresponding complete data likelihood function of the data and (unobserved) activity centres is given by,

$$\begin{aligned} f(\boldsymbol{n}^{obs} | N, \lambda_0, \sigma, \boldsymbol{X}) &= f(\boldsymbol{n}^{obs} | N, \lambda_0, \sigma, \boldsymbol{S}) f(\boldsymbol{S}) \\ &= \left[\prod_{j=1}^{J} \left(\frac{\exp(-T\Lambda_j(\boldsymbol{S}))(T\Lambda_j(\boldsymbol{S}))^{n_{j.}^{obs}}}{n_{j.}^{obs}!} \right) \right] |\mathbb{S}|^{-N}, \quad (4.1) \end{aligned}$$

where $|\mathbb{S}|$ denotes the area over the region $\mathbb{S}$. The corresponding observed data likelihood is obtained by integrating out over the $\boldsymbol{S}$. Therefore, the observed data likelihood is obtained by integrating out activity centres in the Equation

(4.1) given as:

$$f(\boldsymbol{n}^{obs}|N,\lambda_0,\sigma,\boldsymbol{X}) = \int_{\boldsymbol{S}} \left[\prod_{j=1}^{J}\left(\frac{\exp(-T\Lambda_j(\boldsymbol{S}))(T\Lambda_j(\boldsymbol{S}))^{n_{j\cdot}^{obs}}}{n_{j.!}^{obs}}\right)\right]|\mathbb{S}|^{-N}d\boldsymbol{S}. \quad (4.2)$$

However, this integration can be very high dimensional, increasing as the number of traps increases (J) and number of animals increases (N). In the following section, we discuss Bayesian-based approaches for estimating the parameters when the likelihood is analytically intractable (via data augmentation).

4.2 Model fitting

The observed data likelihood expressed in Equation (4.2) is analytically intractable, so that we consider a Bayesian data augmentation (or complete data likelihood) approach (Tanner and Wong, 1987). Under this framework, we treat the latent variables, $\boldsymbol{S}$, corresponding to the activity centres over all individuals, N, as additional auxiliary variables, and form the joint posterior distribution over the model parameters and unknown activity centres,

$$\pi(N,\boldsymbol{S},\lambda_0,\sigma|\boldsymbol{n}^{obs}) \propto f(\boldsymbol{n}^{obs},\boldsymbol{S}|N,\lambda_0,\sigma)p(\lambda_0)p(\sigma)p(N), \quad (4.3)$$

where the joint likelihood of the observed data and activity centres is given in Equation (4.1) and $p(.)$ denotes the prior distribution of the corresponding parameters. We note that the posterior distribution is not of a fixed dimension, as it is now a function of the parameter N hence the number of activity centres $\boldsymbol{S}$ changes according to N for each iteration, so that traditional Markov chain Monte Carlo (MCMC) algorithms cannot be applied. Thus, this has led to the use of the super-population approach, which uses a further data augmentation step and defines an upper limit for the population size, once more returning to the fixed dimension case (Chandler and Royle, 2013). An alternative approach is to use a reversible jump (RJ)MCMC approach, which permits trans-dimensional moves. We initially describe the previous super-population approach and associated challenges, particularly for large datasets, before proposing an alternative

(and more computationally efficient) RJMCMC algorithm. We comment briefly on the conceptual differences of the approaches in terms of the prior specification on N, before comparing the performance of the different algorithms via a small data set relating to the Northern Parula.

4.2.1 Super-population approach

The super-population approach (SPA) has been applied to the unmarked SCR model fitted on various camera trapping data (Royle et al., 2009; Chandler and Royle, 2013; Ramsey et al., 2015; Evans and Rittenhouse, 2018; Connor et al., 2022). The idea is to use a fixed-dimensional parameter-expanded data augmentation approach by initially defining some upper limit for the population, M, typically referred to as the super-population. Then, each individual in the super-population has an associated activity centre and additional auxiliary variable to indicate whether or not it is a member of the population of interest.

Let $\boldsymbol{S} = \{\boldsymbol{S}_i : i = 1, \ldots, M\}$ denote associated activity centres of the augmented individuals and $\boldsymbol{z} = \{z_i : i = 1, \ldots, M\}$ denote the latent indicator variable such that $z_i = 1$ for $i = 1, \ldots, M$ if it is a member of the population N and $z_i = 0$ otherwise. Thus, the auxiliary variables z_i are specified to follow the Bernoulli distribution such that $z_i \sim \text{Bern}(\psi)$ for all augmented individuals $i = 1, \ldots, M$ with some probability ψ, and associate prior $p(\psi)$. Under parameter-expanded data augmentation, the realized total individuals becomes $N = \sum_{i=1}^{M} z_i$. Following this specification, the model can be rewritten as:

$$\boldsymbol{n}^{obs}|\boldsymbol{S}, \lambda_0, \sigma, \boldsymbol{z} \sim \text{Poisson}\left(\sum_{i=1}^{M} \lambda_{ij}(\boldsymbol{S}_i) z_i\right).$$

Thus, the conditional posterior density under super-population approach is:

$$\pi(\boldsymbol{S}, \lambda_0, \sigma, \boldsymbol{z}|\boldsymbol{n}^{obs}) \propto f(\boldsymbol{n}^{obs}, \boldsymbol{S}|\lambda_0, \sigma, \boldsymbol{z}) p(\boldsymbol{z}|\psi) p(\lambda_0) p(\sigma) p(\psi).$$

Although this approach leads to a fixed parameter dimension and so implementable using standard MCMC and software, there are also additional draw-

backs, relating to scalability and the prior specification on N. In particular, the super-population limit M needs to be specified *a priori* potentially leading to a large parameter space (see, for example, King et al. (2016) for further discussion); and the prior specification on N now an induced prior (via the prior specified on ϕ). In the following section, we describe an alternative RJMCMC model fitting approach that addresses both the scalability and prior specification issues.

4.2.2 Reversible Jump MCMC

We propose a RJMCMC algorithm (Green, 1995) for exploring the posterior distribution given in Equation (4.3). RJMCMC is a generalisation of the Metropolis-Hastings (MH) algorithm that permits model moves between different dimensions, which is required when updating the total population size N, due to the associated activity centre latent variables, $\boldsymbol{S}$. A similar updating algorithm was considered by King and Brooks (2008) in the presence of individual random effects and by McLaughlin (2019) for fitting marked SCR models. We describe the implementation of the RJMCMC algorithm, separated into two distinct move types: (i) single-update MH for model parameters $\{\sigma, \lambda_0, \boldsymbol{S}\}$; (ii) RJMCMC update for N. We briefly describe each of these moves in turn.

Single-update MH for model parameters $\{\sigma, \lambda_0, \boldsymbol{S}\}$

The Metropolis-Hasting algorithm is one of the most commonly used MCMC samplers for sampling from a posterior distribution. The objective of the algorithm is to generate random samples from some distribution when direct sampling is difficult. Generated samples can be used to approximate the (marginal) posterior distributions of the target parameters i.e., estimation via kernel density and/or via summary statistics of interest. The MH algorithm iterates from the initial state (initial parameters) to the final step by moving around the parameter space. However, different initial values may potentially bias results since the initial state of the MCMC algorithm is often far from its limiting stationary, invariant, distribution. Therefore, we implement a burn-in period to compensate for this. The idea is that the stationary distribution of the chains equal the distribution of

interest, so we run the chain until convergence then treat future realisations as a (dependent) sample from the distribution and use Monte Carlo integration to estimate the parameters of interest. We may also consider multiple chains for exploring the convergence of the MCMC. The algorithm involves two distinct moves within each iterative move: (i) propose new values for θ, and (ii) accept or reject the new values by following the given acceptance criteria. The first step requires the specification of a proposal distribution for the candidate value. One common choice is to consider a symmetric proposal distribution such that

$$\theta' = \theta + \epsilon,$$

where ϵ has the distribution $q(.)$ e.g., Uniform or Normal distribution. The corresponding acceptance probability is given by:

$$\alpha(\theta', \theta) = \min\left\{\frac{\pi(\theta'|\boldsymbol{x})q(\theta|\theta')}{\pi(\theta|\boldsymbol{x})q(\theta'|\theta)}, 1\right\},$$

and recall that in Section 1.4.2 $\pi(\theta|\boldsymbol{x})$ denotes the posterior distribution of the parameter θ given the observed data $\boldsymbol{x}$.

The above approach is used for updating parameters in the model $\{\sigma, \lambda_0, \boldsymbol{S}\}$ and given the proposal distributions. In particular, we use a single-update MH step for the parameters σ and λ_0 and consider a joint update for $\boldsymbol{S}_i$, for each $i = 1, \ldots, N$ (where N denotes the current value in the MCMC algorithm). For σ and λ_0 we use a random walk Normal proposal distribution, centred at the current location of the parameters where the associated variances are obtained via pilot-tuning i.e., $q(\sigma') \sim N(\sigma, \delta_1)$ and $q(\lambda_0') \sim N(\lambda_0, \delta_2)$. For $\boldsymbol{S}_i$, we use a bivariate Normal random walk, centred on the current activity centre location, with independent covariance matrix = κI, with κ determined via pilot-tuning. Note that we use *inside.owin()* function from `spatstat` R package (Baddeley et al., 2022) to check whether new activity centers are inside the region $\mathbb{S}$ when the state space is irregular (when updating $\boldsymbol{S}_i$ and adding in new activity centers in updating N). If the proposed activity centres $\boldsymbol{S}_i'$ lie outside the observation window $\mathbb{S}$, the move is immediately rejected.

RJMCMC update for N: fixed removal of activity centres

The next step is to update the total population size N. Updating the total population size N involves a change in dimension (recall that $\boldsymbol{S} = \{\boldsymbol{S}_i : i = 1, \ldots, N\}$). Thus, we consider a RJ step for updating N. We initially propose a new value for N, that we denote, N'. If $N' > N$, we further propose adding new activity centres $\boldsymbol{S}'$, using an independent proposal distribution (uniform over $\mathbb{S}$); else if $N' < N$ we remove (some) activity centres. We note that we consider two proposal distributions for updating the activity centres (in terms of those that are removed): a "fixed" proposal; and a "stochastic" proposal. We describe the fixed RJMCMC algorithm, and associated proposal distributions in further details as follows.

(a) Propose $N' = N + \epsilon$, where ϵ is an integer chosen from the interval $[-\delta_3, \delta_3]$ with an equal probability $p(\epsilon) = \frac{1}{2\delta_3}$ and note that this interval contains 0 which is excluded from the set of possible values, such that

$$\{N' \in [N - \delta_3, N + \delta_3] : N' \neq N\}.$$

The integer δ_3 controls the jump step and is chosen via pilot tuning.

(b) If $\epsilon > 0$, we update the elements of $\boldsymbol{S}'$ by setting $\boldsymbol{S}'_i = \boldsymbol{S}_i$ for $i = 1, \ldots, N$ and generating new values of $\boldsymbol{S}'$ for $i = N+1, \ldots, N+\epsilon$ i.e., using the prior distribution

$$\boldsymbol{S}' \sim \text{Uniform}(\mathbb{S}).$$

Then, we set $N = N'$ with probability $A = \min(1, \alpha(N', N))$ where:

$$\alpha(N', N) = \frac{\pi\left(\sigma, \lambda_0, N', \boldsymbol{S}' | \boldsymbol{n}^{obs}\right) q(N|N')}{\pi\left(\sigma, \lambda_0, N, \boldsymbol{S} | \boldsymbol{n}^{obs}\right) q(N'|N) q(\boldsymbol{S}'|\boldsymbol{S})},$$

otherwise we reject the move. If the proposal distribution of $\boldsymbol{S}$ is uniform over the region, the quantity $q(\boldsymbol{S}'|\boldsymbol{S})$ in the denominator will cancel with the prior in the numerator when the posterior distribution is expanded. The proposal distribution $q(N'|N)$ is simply equal to $p(\epsilon)$ and $q(N|N')$ is the

probability of reversing the move thus equal to $p(-\epsilon)$ therefore also cancel in the acceptance probability due to symmetric distribution on ϵ.

(c) Similarly, if $\epsilon < 0$, we set $\boldsymbol{S}' = \boldsymbol{S}$ for $i = 1, \ldots, N + \epsilon$ and set $N' = N$ with the probability value $A = \min(1, \alpha(N', N))$ where:

$$\alpha(N', N) = \frac{\pi\left(\sigma, \lambda_0, N', \boldsymbol{S}' | \boldsymbol{n}^{obs}\right) q(\boldsymbol{S}|\boldsymbol{S}') q(N|N')}{\pi\left(\sigma, \lambda_0, N, \boldsymbol{S}|\boldsymbol{n}^{obs}\right) q(N'|N)},$$

otherwise we reject the move. For example, let the current value of $N = 30$ (thus $\boldsymbol{S} = \{\boldsymbol{S}_1, \ldots, \boldsymbol{S}_{30}\}$) and $\epsilon = -5$ such that the propose value $N' = 25$, and the reduced set of activity centres is $\boldsymbol{S}_1, \ldots \boldsymbol{S}_{25}$.

We note that using multiple updating steps of N at each iteration was also considered (Gilks et al., 1995) to improve mixing. For convenience, we call this algorithm as a fixed RJMCMC. Next section, we introduce a modification on fixed RJMCMC algorithm.

RJMCMC update for N: Stochastic removal of activity centres

Here we consider the same reversible jump algorithm as described above but where the activity centres are stochastically removed when $\epsilon < 0$. For reversibility we describe the addition and removal of the activity centres as follows.

Step (b) is amended such that if $\epsilon > 0$ then we place the new activity centres randomly within the list of activity centres. The current state has N activity centres. Thus the first activity centre can be placed in $(N + 1)$ locations (i.e. before the first activity centre; between each successive activity centre; or after the final activity centre). The next activity centre (if $\epsilon \geq 2$) can be placed in $(N + 2)$ locations (as we include the previously added activity centre). And so on. Thus for general ϵ we have $(N + 1) \times ... \times (N + \epsilon)$ possible locations (each equally likely). However, the order that the activity centres can be added in is equal to $\epsilon!$ which will all give the same outcome which have to be accounted for. Thus, letting Tot denote the total number of possible ways (where the order is

unimportant as the final list is still the same) we can write,

$$Tot = \frac{(N+1) \times ... \times (N+\epsilon)}{\epsilon!} = \begin{pmatrix} N+\epsilon \\ \epsilon \end{pmatrix}.$$

Now as each outcome (in terms of location of the activity centres) are equally likely and there are a total number of Tot outcomes, the probability of any such outcome is simply $1/Tot$. Alternatively we can use the following rationale. In the proposed move we will have $N+\epsilon$ activity centres - and we need to work out the probability of the ϵ new centres have the given set of labels. This is again simply $1/Tot$ as there are Tot possibilities.

In the reverse move where we remove activity centres, suppose that we have $N+\epsilon$ possible activity centres and we wish to remove ϵ centres. Then, as each activity centre is chosen with equal probability, the probability of any given set of activity centre be removed is again simply $1/Tot$. Thus, in the acceptance probability the probabilities of moving between the different models (in terms of the set of activity centres) in the numerator and denominator simply cancel. We note that this is only true when activity centres are chosen with equal probability - so that we can use a counting arguments of equally likely events.

Finally, we comment that for this stochastic algorithm, an alternative way of thinking about the above is that the order of the activity centres are exchangeable - with the likelihood remaining constant to the ordering of the activity centres. Thus we could simply think about reordering the set of activity centres such that those added (when $\epsilon > 0$) are labelled $N+1,\ldots,N+\epsilon$; and conversely in the reverse move (when $\epsilon < 0$) we reorder the set of activity centres so that those that are removed are those listed at the end of the list. For convenience, we call this a stochastic RJMCMC to describe a stochastic removal on activity centres.

4.2.3 Bespoke R codes

For the implementation of super-population approach (SPA), there are several available MCMC black-box packages which can be used to fit the model such as `nimble` (Valpine et al., 2022) and `rjags` (Plummer et al., 2022) that we have

described in Section 1.5. However, in addition to the computational cost of specifying the upper super-population limit and additional indicator variables, $\boldsymbol{z}$, there is another hidden and highly considerable computational cost when fitting the unmarked SCR model, without having finer control over the MCMC algorithm. This is due to the structure of the likelihood in terms of a summation over all activity centre contributions for the Poisson means for each camera trap in the likelihood function. This summation is calculated when updating each individual activity centre, $\boldsymbol{S}_i$, and latent variable, z_i, for $i = 1, \ldots, M$. However, an efficient and new implementational trick can be applied within bespoke computational code for both the SPA and RJMCMC algorithms when updating the activity centres (and latent variables for SPA). In particular, the Poisson mean value can be stored and then only the *change* in the likelihood needs to be calculated when updating $\boldsymbol{S}_i$ (and z_i), simply considering only the contributions of the particular $\boldsymbol{S}_i$ (and z_i) values, as opposed to recalculating the mean by summing over all individuals. A similar computational tool can be applied when updating N within the RJMCMC step, considering only the difference to the Poisson mean contribution for the added/removed activity centres, when evaluating the likelihood function. These new implementational steps can be immediately incorporated within bespoke code (but typically not within more general black-box MCMC packages), providing significant improvements in computational efficiency. For example, running the SPA algorithm using bespoke `R` code, as opposed to `nimble`, decreased the computational time by a factor of 2 for the smaller northern Parula dataset ($M = 300$) and a factor of 20 ($M = 10000$) for the barking deer data. Thus, for meaningful computational time comparisons between the SPA and RJ approaches we consider their implementation using bespoke code, using the same MH updating steps for parameters in common.

4.2.4 Prior specification

Finally, we explain possible prior choices for model parameters $\{\sigma, \lambda_0, N\}$. First, we discuss the prior specification N for both SPA and RJMCMC. The prior specification on N in SPA is specified implicitly via the indicator variables $z_i \sim$

Bern(ϕ) for $i = 1, \ldots, M$, and associated prior specified on ϕ. For example, Chandler and Royle (2013) consider a general prior of the form, $\phi \sim \text{Beta}(a, b)$ which induces a Beta-Binomial(M, a, b) prior on N. A common choice is to set $a = 0.001$ and $b = 1$ which is a very close approximation to the scale (Jeffrey's) prior on N (Link, 2013).

However, within the RJMCMC algorithm, a prior is specified explicitly on the total population size, N, leading to greater flexibility and interpretability in the prior specification. For example, common choices may include Jeffrey's prior, Uniform prior or (hierarchical) Poisson prior for N. We consider the prior $N \sim \text{Poisson}(\mu)$, where $\mu \sim \Gamma(\alpha, \beta)$ (equivalent to a Negative-Binomial prior distribution) (King and Brooks, 2001; Royle, 2004). In general, the values of α and β can be specified to reflect prior knowledge; or be specified to induce a more uninformative prior.

Now, we move on the prior specification on σ. There are two typical priors specified on σ: (i) a Uniform (uninformative) prior i.e., Uniform$(0, \infty)$; and (ii) a Gamma (informative) prior i.e., Gamma(a, b). When no information related to σ available, we assign a Uniform prior. However, this may potentially lead to a convergence issue when running MCMC if the spatial correlation is relatively weak as noted in Ramsey et al. (2015). However, constructing informative prior on σ requires information about the home range size of target species and is often challenging in practice. The home range size can be defined as an area used by an individual at certain duration of time. Assuming a bivariate half-normal model on the encounter rate, we have that the term $||\boldsymbol{X}_j - \boldsymbol{S}_i||^2$ follows a chi-square distribution with 2 degrees of freedom (Royle et al., 2014)). Under a bivariate normal model we can compute all realized distances (the movement outcomes). In particular, $(1-\alpha)\%$ of the realized distances is within $\sigma\sqrt{q(\alpha, 2)}$ where $q(\alpha, 2)$ is the critical value of chi-square distribution with 2 degrees of freedom. Let $r_{1-\alpha}$ denote the quantity which contains $(1 - \alpha)\%$ of the movement outcomes such that

$$r_{1-\alpha} = \sigma\sqrt{q(\alpha, 2)},$$

Then we can compute the $(1-\alpha)\%$ area used around $\boldsymbol{S}$ by

$$A_{1-\alpha} = \pi r_{1-\alpha}^2.$$

Therefore, if information related to the home range size (area used by animals) is readily available, we can specify a and b on the Gamma prior to reflect the knowledge we have on individuals' space usages by the following equation:

$$\sigma = \frac{r_{1-\alpha}}{\sqrt{q(\alpha, 2)}}, \tag{4.4}$$

For example, we need 0.5% and 99.5% quantiles of chi-square distribution for $\alpha = 1\%$ to estimate a and b appropriately. Alternatively, we can introduce a proper Gamma$(0.001, 0.001)$ prior on σ as used in Royle et al. (2014) if information related to the space usage is unavailable.

For λ_0, the common choice of (uniformative) prior is the improper Uniform prior such that $\lambda_0 \sim \text{Uniform}(0, \infty)$ (Chandler and Royle, 2013). In this thesis, we also consider a normal prior on $\log(\lambda_0)$ i.e., $\log(\lambda_0) \sim \text{Normal}(0, \tau)$ for sufficiently large τ. Similarly, we can also introduce a proper Gamma$(0.001, 0.001)$ prior on λ_0 (Royle et al., 2014).

Chapter 5

Camera-Trapping Studies: Simulation Study and Applications

In this Chapter, first we perform a simulation study on the spatially explicit model for comparing the SPA and the RJMCMC algorithm in estimating the population size. Then, we demonstrate the model fitting on real datasets. For illustrating the real cases, we consider two case studies: (i) a small case study, Northern Parula, as used in Chandler and Royle (2013); and (ii) a large case study, barking deer collected in Ujung Kulon National park (UNKP), Indonesia. All codes used in Chapter 5 can be accessed in the author's GitHub `https://github.com/riki-herliansyah/unmarked`.

5.1 Simulation studies

We consider a simple simulation study to demonstrate the performance of RJMCMC algorithms and SPA fitted on spatially explicit models. The scenario settings of the simulation are given as follows: we set the total population size to be $N = 50$ individuals assuming $T = 10$ sampling occasions and $\lambda_0 = 0.6$. We generate $J = 100$ traps within the regular (square) state space with the size of $[60, 240] \times [60, 240]$ and a grid spacing of 20. Therefore, we assume and set the radius $r_{95} = 30$ to ensure sufficient spatial correlation between traps. The estimated σ is then computed using Equation (4.4):

$$\sigma = \frac{30}{\sqrt{5.99}} = 12.26.$$

We generate activity centres assuming a uniform distribution and a buffer of 5 within the state space such that $\boldsymbol{S}_i \sim \text{Uniform}(\mathbb{S})$ for $i = 1, \ldots, N$. Given all parameter values, 100 simulated datasets were generated and fitted on the unmarked SCR model.

As an illustration, we draw a spatial map given in Figure 5.1 displaying spatially correlated counts of individuals on a 20 unit grid from a single simulated dataset. A total of 100 traps (10×10) are placed in the area, and shown by black dots in the figure, with the size of the dot srepresenting the total number of detections at the given trap for the simulated dataset (ranging from 0 to 20 detections). The red crosses correspond to the (randomly generated) activity

centres for the given simulated data. The blue line border indicates the original spatial area while the red line border shows the additional buffer zone (5 units).

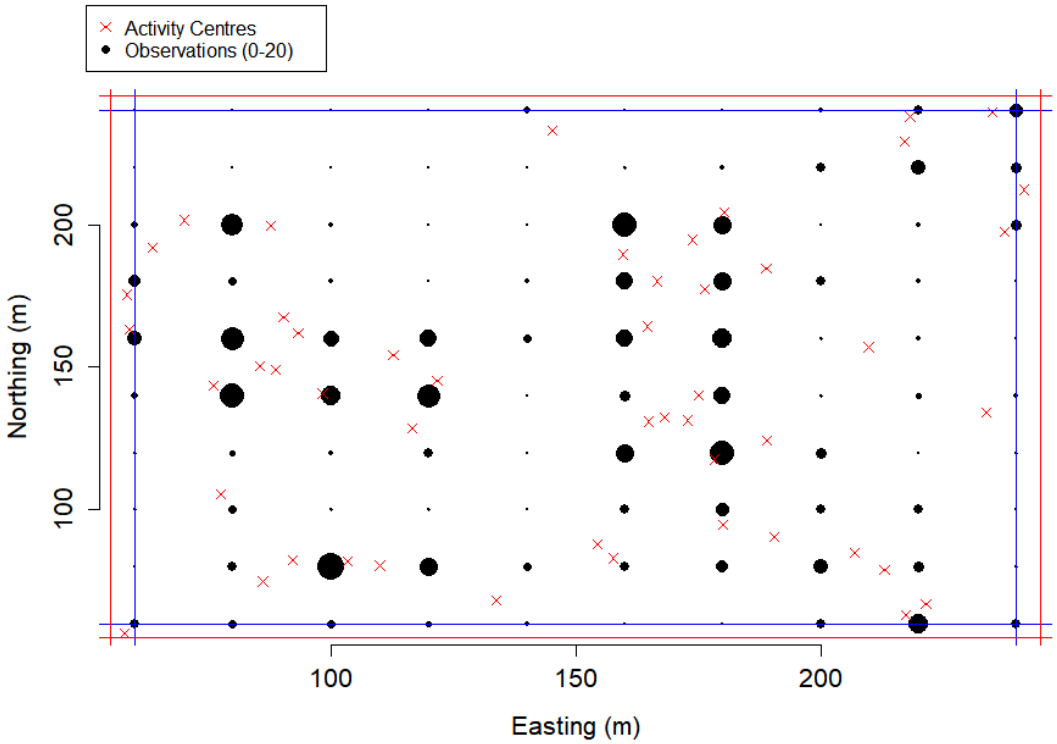


Figure 5.1: Spatially correlated counts of a single simulated data on a 20 unit grid at area of $[60, 240] \times [60, 240]$ (blue border) with a buffer of 5 unit (red border).

We assume a uniform prior on σ and λ_0 i.e., $\sigma, \lambda_0 \sim \text{Uniform}(0, 100)$. We also assume uniform prior on N for RJMCMC such that $N \sim \text{Uniform}(0, 150)$. We further consider multiple updates on N i.e., 5 times at each iteration to improve the mixing. For SPA, we used $M = 150$ as the upper limit and assume a $\text{Beta}(1, 1)$ prior for ψ such that $\psi \sim \text{Beta}(1, 1)$. We run the MCMC for 100,000 iterations with 10,000 additional burn-in period for three separate and independent chains.

Table 5.1 provides the statistical summaries i.e., relative bias (RB) for sample mean and mode, coverage probabilities (CP), effective sample size (ESS) and ESS per second (ESS/s) of corresponding 100 simulated datasets. We consider to focus on three main parameters of interest: σ, λ_0 and N for inspection. Our simulation results suggest that relative biases for posterior means are approximately the same for all parameters among different approaches. Additionally, we found that relative biases of N for posterior modes are much lower than the the relative biases

for posterior means by around 15% while it shows opposite behaviours in σ and λ_0. The reasonable explanation to this is that the posterior distribution of the total population N is often skewed thus the sample mode gives better estimation for the central tendency of the distribution than the mean while the posterior distribution of σ and λ_0 is approximately symmetric. Coverage probabilities shows that the

Table 5.1: Simulation results in terms of averaged relative bias (RB), 95% coverage probabilities (CP), effective sample size (ESS) and ESS per second (ESS/s) for 100 simulated datasets fitted via RJMCMC algorithms assuming stochastic and fixed removal proposal distribution, and super population approach (SPA).

Algorithms	Parameter	RB Mean	RB Mode	CP	ESS	ESS/s
Stochastic RJMCMC	σ	0.005	-0.073	92.00	1111	1.758
	N	0.176	0.015	90.00	416	0.616
	λ_0	-0.001	-0.289	94.00	1091	1.684
Fixed RJMCMC	σ	0.004	-0.086	93.00	984	1.607
	N	0.178	0.004	90.00	196	0.291
	λ_0	-0.004	-0.275	93.00	948	1.497
SPA	σ	0.006	-0.081	92.00	1853	0.806
	N	0.176	0.029	87.00	1487	0.647
	λ_0	0.003	-0.277	93.00	2144	0.933

accuracy is approximately 90% or higher indicating good fit. Our finding also suggests that the stochastic RJMCMC provides better mixing than the fixed RJMCMC; in particular we observe the effective sample size (ESS) and ESS per second for investigating the mixing. The stochastic RJMCMC produces ESS of approximately double for N while producing slightly better mixing for σ and λ_0 compared to the fixed RJMCMC. Compared to the SPA, the ESS per second produced by the RJMCMC algorithms is almost double for σ and λ_0. However, the ESS per second for N is almost the same as the stochastic RJMCMC which may be due to the choice of moderate M. In practice we may need to assume a reasonably large M to account for higher variability in the MCMC samples, and the larger M we choose the more expensive the computation.

5.2 Small case study: Northern Parula

To compare the performance of the different algorithms on a relatively small sample population, we now consider the well studied real data example relating to the Northern Parula (*Parula americana*). The Northern Parula is one of the wood warbler species, found in eastern Canada and US South to Florida. It has a unique feature of bright yellow colour with small body size as shown in Figure 5.2. This neotropical migratory wood warbler has weight between between 0.2 and 0.4 ounces and length between 4.3 and 4.7 inches, and can be found mainly in mature forest with hanging canopy.

Figure 5.2: Northern Parula. Image credit: Wikipedia.

The data consists of 226 detections detected within 105 trap stations over 3 survey periods. The distributions of traps and individual detection are shown in Figure 5.3. See Chandler and Royle (2013) for more details and associated SPA approach applied to analyse the data. We used $M = 300$ for the upper limit for the super population and $\delta = 10$ for the proposal distribution for the reversible

jump step on N. We considered two different priors on σ provided by Chandler and Royle (2013): (i) Uniform prior, $\sigma \sim \text{Uniform}(0, \infty)$ and (ii) the informative prior, $\sigma \sim \text{Gamma}(13, 10)$. For the baseline detection rate, λ_0, we specify the Uniform prior i.e., $\lambda_0 \sim \text{Uniform}(0, \infty)$. Finally, for the total population size, we specify $N \sim \text{Uniform}(0, 300)$ for the RJMCMC algorithms.

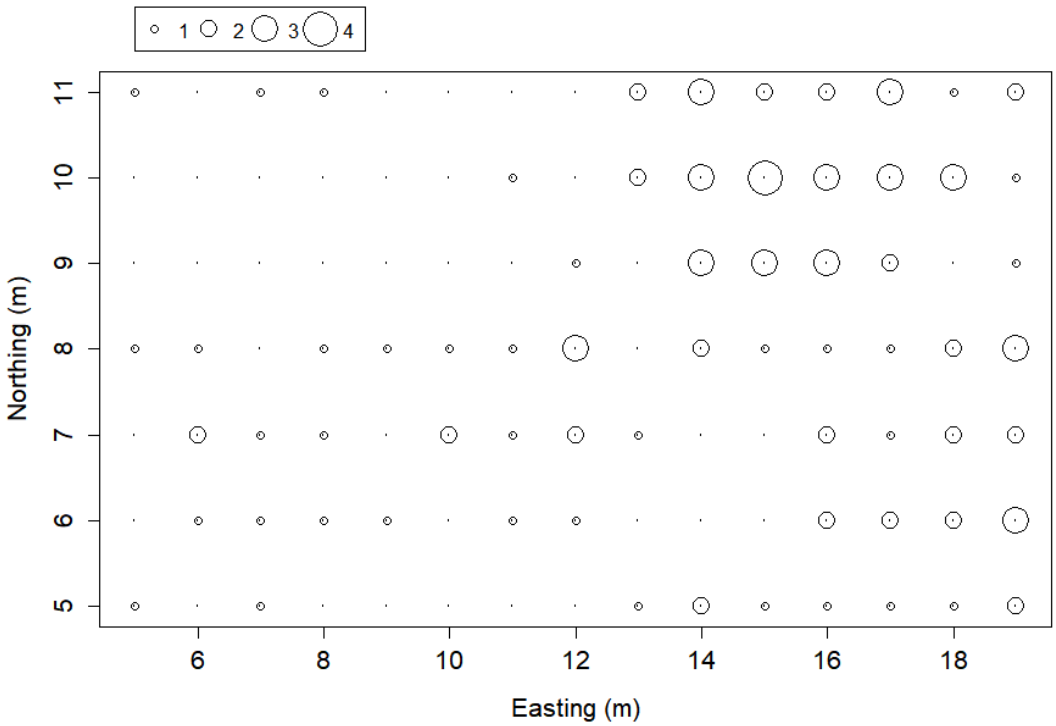


Figure 5.3: The distribution of counts of Northern Parula on a 1 m grid.

For fair comparisons, all algorithms were run under similar conditions and written in the bespoke `R` by applying the efficient likelihood calculation as explained in Section 4.2.3. Tuning parameters are chosen after some pilot tuning to obtain approximate mean acceptance rates between 0.2 and 0.4 for model parameters. In particular, we set tuning parameters equal to 0.1, 0.1 and 1 for σ, λ_0 and $\boldsymbol{S}$, respectively, for both SPA and RJMCMC algorithms with additional updates on N i.e., N was updated 4 times for each iteration to improve the mixing (Gilks et al., 1995). We run the MCMC algorithms for 300,000 iterations following an initial 10,000 iterations for burn-in for each algorithm using 3 separate and independent chains. For SPA the simulations took approximately 4 hours; while the RJMCMC algorithms took 0.6 hours each.

No issues were identified in terms of convergence or mixing of the different algorithms. Table 5.2 presents the corresponding posterior summary statistics of the fitted model for the Parula data for the different algorithms including the posterior mean (Mean), median (Med), standard deviation (SD) and the associated 95% credible intervals (95% CI). The posterior estimates are similar for each of the different algorithms, as would be expected. However, we note that the posterior estimates for N and σ appear to be influenced by the different prior specifications on σ (as previously noted by Chandler and Royle (2013)). In terms of computational efficiency, there are some noticeable differences when comparing the competing methods. In particular, we consider the effective sample size (ESS) and effective sample size per second (ESS/s). The RJMCMC algorithms consistently have a higher ESS/s compared to the SPA algorithm, typically of the magnitude of > 10. Further, the stochastic RJMCMC algorithm appears to have a marginally better performance than the fixed RJMCMC algorithm for all the model parameters as we consistently observe in simulation studies.

Table 5.2: Posterior summary statistics of model parameters fitted on the northern Parula data using super population approach (SPA) and RJMCMC algorithms (fixed proposal; stochastic proposal). Effective sample size (ESS) and effective sample size per second (ESS/s) are included for both methods for each prior.

Prior	Methods	Parameter	Mean	SD	Med	95% CI	ESS	ESS/s
Uniform	Fixed RJ	N	37.10	36.23	27.00	(2, 136)	1148	0.65
		σ	2.87	2.04	2.22	(0.96, 8.96)	328	0.18
		λ_0	0.27	0.14	0.24	(0.07, 0.59)	2798	1.58
	Stochastic RJ	N	38.22	36.65	28.00	(2, 138)	1523	0.84
		σ	2.73	1.89	2.14	(0.94, 8.61)	365	0.20
		λ_0	0.26	0.13	0.24	(0.07, 0.57)	3372	1.86
	SPA	N	37.59	38.90	27.00	(1, 142)	2196	0.12
		σ	3.33	2.87	2.15	(0.93, 10.76)	444	0.02
		λ_0	0.31	0.27	0.26	(0.07, 1.14)	1349	0.08
Gamma	Fixed RJ	N	64.57	39.36	54.00	(20, 170)	2048	0.74
		σ	1.33	0.27	1.30	(0.91, 1.95)	5011	1.82
		λ_0	0.29	0.13	0.27	(0.10, 0.59)	4215	1.53
	Stochastic RJ	N	64.17	37.60	54.00	(20, 164)	2974	1.09
		σ	1.33	0.26	1.29	(0.91, 1.94)	5773	2.12
		λ_0	0.29	0.13	0.27	(0.10, 0.59)	5031	1.85
	SPA	N	65.41	40.03	55.00	(20, 174)	5228	0.28
		σ	1.32	0.27	1.29	(0.90, 1.97)	7403	0.40
		λ_0	0.29	0.13	0.27	(0.10, 0.60)	6198	0.33

5.3 Large case study: Barking Deer

Now, we consider the large case study relating to the barking deer (*Muntiacus muntjak*). The barking deer, *Muntiacus muntjak*, is a primarily solitary animal (with no group larger than four animals observed), found mainly in Asia e.g., in India, Malaysia, Thailand, the Indonesian islands, Taiwan and Southern China.

They are commonly known to be territorial among males but may overlap with females, and during the mating season, the territory may overlap between males. The reasons for territory holding (in males) might include restriction of access to females for other males or (for both sexes) as a strategy to control access to food or water resources. Adult males as shown in Figure (5.4) may stand 80 cm at the shoulder while females smaller at 65 cm; with average adult deer between 98-120 cm (head and body) in size.

Figure 5.4: Adult male barking deer. Image credit: Shuttershock.

We begin by describing the study area and design including the characteristics e.g., seasons and ecosystems. The model specified in Chapter 4 is extended by incorporating the covariate information following a brief discussion of the observed data. Finally, we fit the model on the data using the RJMCMC and the SPA and discuss the findings.

5.3.1 Study area

Ujung Kulon National Park (UKNP) is the largest lowland rainforest in Java with the total area of approximately 120,551 ha of which about 44,337 ha is a marine zone. The UKNP is a triangular peninsula located at the southwest end of Java island, Indonesia, lying approximately at 6° 45′S by 105° 20′E. The study area is approximately 32,900 ha. Habitat structures of the study area can be divided into four vegetation types corresponding to primary forest, secondary forest, mangrove-swamp forest and beach forest. The primary and secondary forests account for 90% of the total area, with all of the camera traps located in these two habitat types as shown in Figure 5.5 with mangrove-swamp and beach forests being grouped as others.

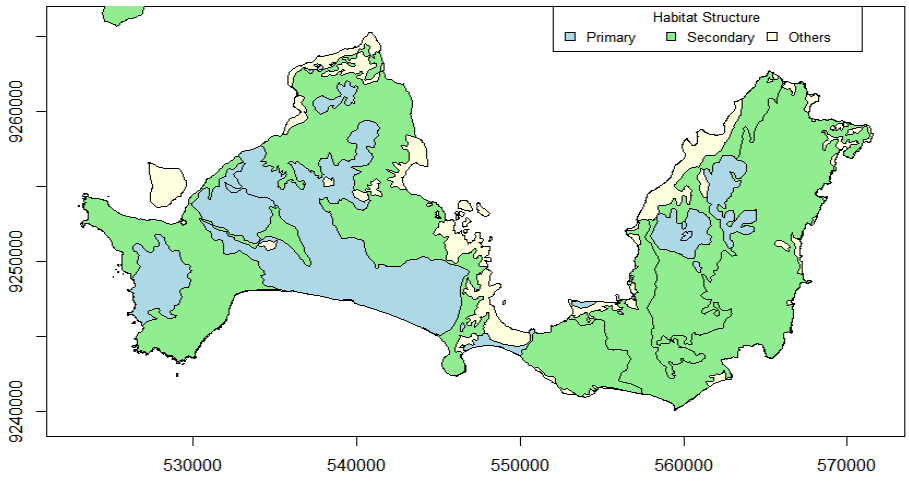


Figure 5.5: The area of Ujung Kulon National Park, Java, Indonesia with different habitat structures: (i) primary forest, (ii) secondary forest; and (iii) mangrove-swamp and beach forests (others).

There are two seasons corresponding to (i) the wet season occurring between October and April, with an average of approximately 400mm of rainfall per month and (ii) the dry season between the May-September, with approximately 100 mm rainfall per month (Rahman et al., 2017).

The study area was gridded into 1×1 km sites, providing a total of 329 sites. A total of 77 motion-sensor cameras (Bushnell Tropy Cam 119467 and 119405)

were distributed across the study region: 35 cameras in the primary forest; 42 cameras in the secondary forest. The spatial grid and camera trap locations are shown in Figure 5.6. The camera traps were placed 170 cm above ground and

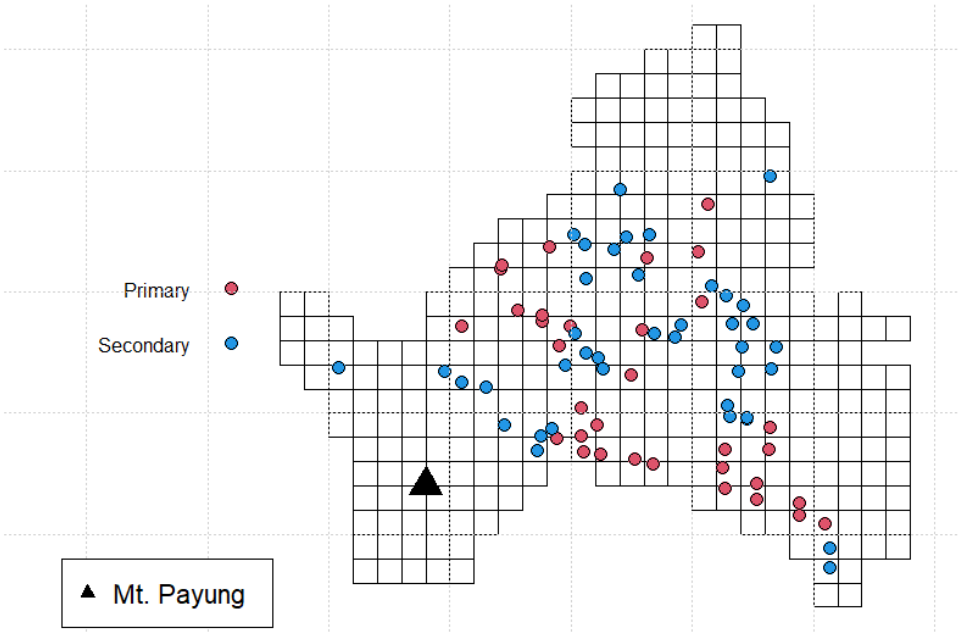


Figure 5.6: The study area in Ujung Kulon National Park, Java with 1×1 km^2 spacing grid. The points represent the camera trap locations distributed within the state space ensuring the sufficient spatial correlation between traps over different habitat. The black triangle represents Mt. Payung which is later excluded from the modeling.

fixed to a tree with a 10-20° angle. The survey was conducted from March to June, 2014 involving two season times: (i) wet season (March-April), and (ii) dry season (May-June). Cameras were checked once a month (approximately every 21-30 days) and the battery and/or memory card replaced if necessary. Poor quality photographs, where identification was uncertain were discarded. Further, repeat photographs of individuals within 1 hour were considered to be a single photographic event (Karanth and Nichols, 1998).

5.3.2 Data and model

A total of 1095 barking deer detections were recorded during the four months of sampling times; with 540 detections during the wet season and 555 detections during the dry season. Further, we note that for the wet season 344 detections were recorded in the primary forest and 196 in the secondary forest; corresponding to 64%/36% for the primary/secondary forest detections. For the dry season 231 detections were recorded in the primary forest and 324 detections in the secondary forest; corresponding to 42%/58% for the primary/secondary forest. Distributions of spatially correlated counts of barking deer at the UNKP for the two seasons can be seen in Figure 5.7.

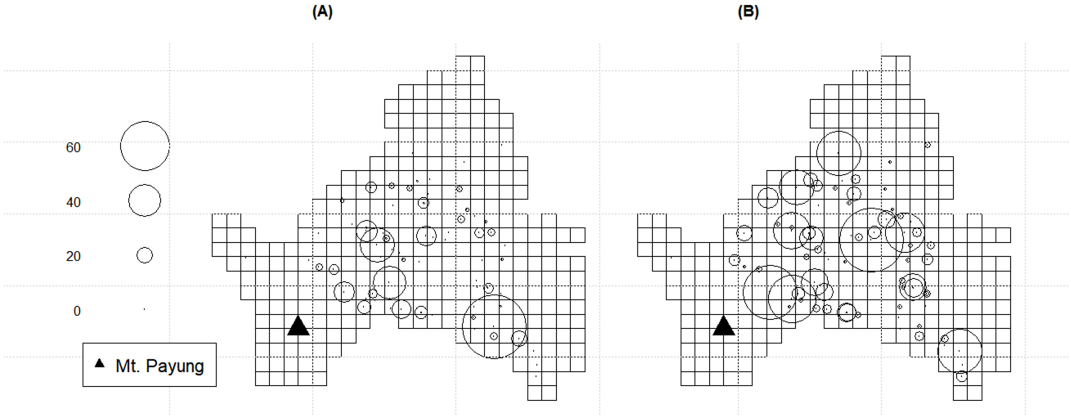


Figure 5.7: Spatially correlated counts of barking deer on a 1 km grid in at UNKP for two seasons: (A) wet and (B) dry. The black triangle represents Mt. Payung which is later excluded from the modeling.

These data are collected across two different seasons: wet and dry. We consider these separately due to the different weather conditions which may affect animal behaviour and/or detectability (Marcus Rowcliffe et al., 2011; Kays et al., 2020). In particular, we consider data relating to 2-month periods for each of season: March-April for the wet season and May-June for the dry season. We assume that the population is (approximately) closed for these periods (Silver et al., 2004; Soria-Díaz and Monroy-Vilchis, 2015; Rahman, 2016). Note that we consider a 7 day period for each sampling occasion thus over the 2 months this leads to 9 sampling occasions for the dry and wet season respectively.

We extend the initial model presented in Section 4.1 to incorporate the environmental covariate relating to habitat into the baseline detectability rate, such that,

$$\log(\lambda_{0j}) = \beta_0 + \beta_1 I(\text{habitat}_j = \text{primary}),$$

where β_1 corresponds to the parameter associated with the primary forest, relative to the secondary forest. For simplicity, let $\boldsymbol{\beta} = \{\beta_0, \beta_1\}$. The log specification on λ_0 implies that the original baseline rate for each forest can be obtained by taking exponential such that $\lambda_{sec} = \exp(\beta_0)$ and $\lambda_{pri} = \exp(\beta_0 + \beta_1)$ for corresponding baseline rates of primary and secondary forest, respectively. We let M_0 denote the standard model, as described in Section 4.1 (i.e. where $\beta_1 = 0$, so that $\lambda_{pri} = \lambda_{sec}$); and M_h the model where the baseline detection rate is a function of the habitat (at the given trap location).

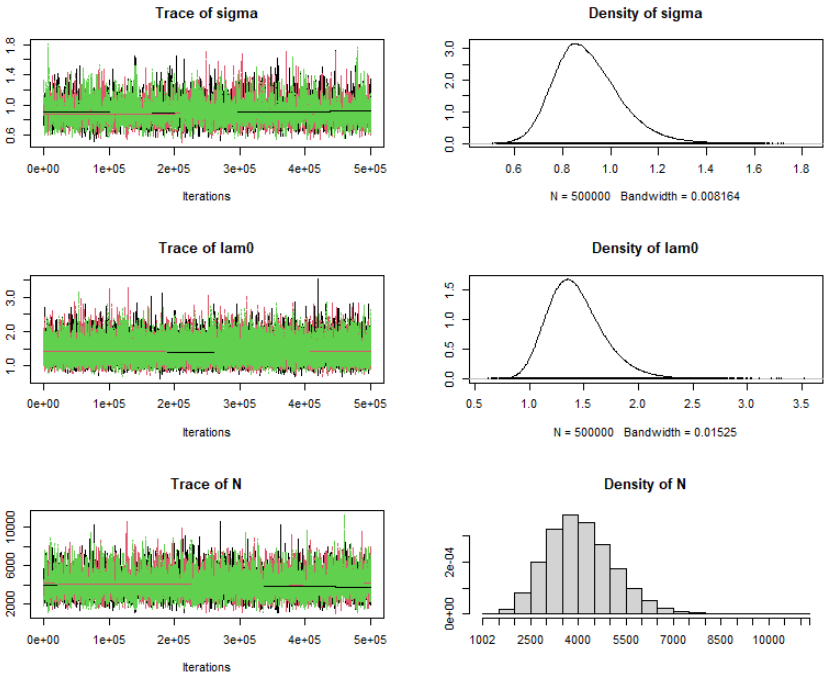
5.3.3 Stochastic RJMCMC algorithm

Due to the large number of individuals, we only consider one of the RJMCMC approaches, and given the performance for the previous Parula data, restrict attention to the stochastic RJMCMC algorithm.

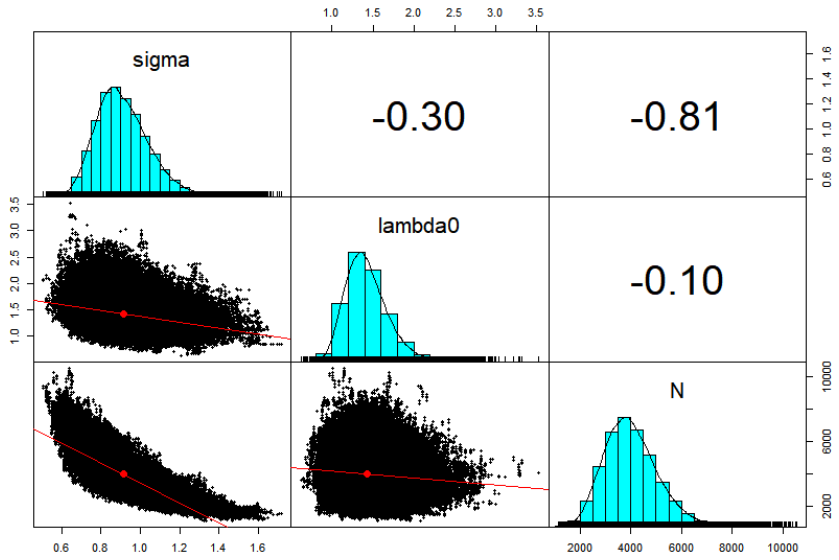
For model M_0, we specify a Normal prior for $\log(\lambda_0)$ i.e., $\log(\lambda_0) \sim \text{N}(0, 10)$; and for model M_h, we specify $\beta_k \sim \text{N}(0, 10)$ independently for $k = 0, 1$. We consider the same priors on the remaining parameters for both models. In particular we specify a Uniform prior on the scale parameter σ, so that $\sigma \sim \text{U}(0, \infty)$. For N we construct a weakly informative prior, using previous information relating to the barking deer in another national park, the Baluran National Park, Indonesia (Tyson, 2007) combined with information provided by park staff in UKNP. In particular the previous study for Baluran National Park suggested a density of approximately 25 per km^2, with 95% confidence interval $(15, 47)$ (from fairly limited data), with the density for the given barking deer in UKNP thought to be (potentially substantially) lower. Thus, assuming a prior density of half (=12.5) with a wide 95% uncertainty interval of (5, 17), led to the specification of $N \sim \text{Neg-Bin}(10, 0.0032)$, with prior mean of 3115, and 95% interval of (1491,

5325).

Each model was run for 500,000 iterations, following an initial burn-in of 10,000 iterations, using three independent chains each starting from different starting values. N was updated 15 times for each iteration to improve mixing (Gilks et al., 1995). The corresponding results are given in Table 5.3 relating to the posterior summary statistics of the parameters for each model and for each season: $M_0(\text{dry}), M_0(\text{wet}), M_h(\text{dry}), M_h(\text{wet})$. Convergence was checked using the Brooks-Gelman-Rubin statistic for each monitored parameter provided in the `R` coda package (Plummer et al., 2020). Figure 5.8 provides an example of posterior distributions of monitored parameters fitted to $M_0(\text{dry})$ and their correlation plots. There seems no issue in convergence observed from the trace plot as shown in Figure 5.8a. The R statistics of all parameters including covariate models also give a good estimate at approximately 1.01. However, we note that there seems to be a very strong negative correlation between σ and N as shown in Figure 5.8b (posterior correlation $= -0.81$), leading to poor mixing of the MCMC chains for σ and N and low effective sample size compared to λ_0. Finally, we record the computation times for fitting each model. Each simulation took approximately 20 hours to run for the dry-season models (M_0 and M_h) and slightly longer for the wet-season model i.e., 23 hours.



(a)



(b)

Figure 5.8: Posterior distribution checks on the dry model parameters (M_0) fitted to the barking deer data for 500,000 iterations. Figure (a) shows trace plots for convergence check, and (b) shows correlation plots.

First, we focus on model M_0 fitted to the data from each season. Given the

Table 5.3: Posterior summary statistics, corresponding to the mean, standard deviation, 2.5% (Q2.5), 50% (Q50) and 97.5% (Q97.5) quantiles and effective sample size (ESS) of the model parameters fitted on the barking deer data. The dry and wet seasons indicate the model fitting on the data for the corresponding season i.e., the data is separated into two seasons with and without habitat respectively.

Parameter	**Mean**	**SD**	**Q2.5**	**Q50**	**Q97.5**	**ESS**
M_0 (Dry Season)						
σ	0.90	0.14	0.68	0.89	1.22	2560
λ_0	1.43	0.26	1.00	1.40	2.02	10288
N	4073	1065	2251	3984	6396	3267
M_0 (Wet Season)						
σ	0.58	0.09	0.43	0.57	0.78	1112
λ_0	3.37	0.72	2.28	3.25	5.06	3114
N	4488	1161	2509	4391	7003	2619
M_h (Dry Season + Habitat)						
σ	0.89	0.14	0.66	0.88	1.20	2405
λ_{sec}	1.58	0.37	1.03	1.53	2.47	6626
λ_{pri}	1.35	0.31	0.87	1.31	2.08	15656
N	4108	1072	2296	4011	6484	3153
M_h (Wet Season + Habitat)						
σ	0.61	0.10	0.44	0.60	0.85	1118
λ_{sec}	1.67	0.41	1.07	1.61	2.67	3830
λ_{pri}	4.61	1.22	2.81	4.44	7.66	2703
N	4428	1197	2393	4319	7045	2445

estimates for the total population size, the associated estimates of the population density of barking deer i.e., $D = N/|\mathbf{S}|$ where $|\mathbf{S}|$ is the total area including the buffer zone (here $|\mathbf{S}| = 349$ km^2), in the study area are approximately 11.7 animals per km^2 in the dry season and 12.9 animals per km^2 in the wet season. The 95% credible intervals are highly overlapping between seasons with values ranging from approximately $(6.4, 18.3)$ and $(7.2, 20.1)$ individuals per km^2 for the dry and wet season respectively. In general the density of barking dear varies substantially, from 25 per km^2 in Baluran National Park, Indonesia (Tyson, 2007); to between 2.1-3.4 animals per km^2 in Nepal (Wegge and Storaas, 2009; Wegge and Mosand, 2015); 2.9 animals per km^2 in Sarawak (Dahaban et al., 1996); 3.1 animals per km^2 in Thailand (Srikosamatara, 1993). Despite similar population density estimates between the two seasons, the associated estimates for λ_0 and σ are noticeably different. In particular, for the wet season, the estimated scale parameter σ is considerably smaller than the dry season; while the baseline detection rate is substantially larger (a posterior mean > 2 times) than in the dry season. The interpretation of the scale parameter is related to the movement of barking deer i.e., the smaller σ in wet season may indicate a smaller movement range due to closer water sources and/or food availability, while the larger value of σ for the dry season may be a result of larger movement to search for water and/or food availability (Tyson, 2007). Similarly, the larger value of λ_0 for the wet season indicates an increase in catchability, compared to the dry season (Marcus Rowcliffe et al., 2011). This may again be potentially explained by animals having a smaller range due to plentiful resources during the wet season, resulting in smaller movements and higher frequency of cameras within their search/activity patterns.

We now consider the results associated with model M_h. We note that the inclusion of habitat type in the detection parameter in the model does not lead to any meaningful change to the total population size (and hence density estimates) and σ. However, there appears to be a substantial change in the estimates of the detection functions for the different habitats (primary and secondary forest) for the wet season; while similar estimates are obtained for the dry season for each

habitat, and which are comparable with the secondary forest in the wet season. The posterior mean for β_1 for the dry season (corresponding to the difference in detection between the primary and secondary forests) is equal to -0.24 with 95% credible interval $(-0.25, 0.48)$, suggesting no difference. However, for the wet season, the posterior mean for β_1 is 2.94 with 95% credible intervals of $(1.12, 5.86)$, indicating a much greater detection in traps in the primary forest habitat for barking deer in the wet season. This difference in detection between habitats in the wet season may again be related to the different usage of primary and secondary forests. For example, habitat preference is known to change seasonally which, again, may be related to food availability, resting or nesting sites and predator avoidance (Yokoyama et al., 2020).

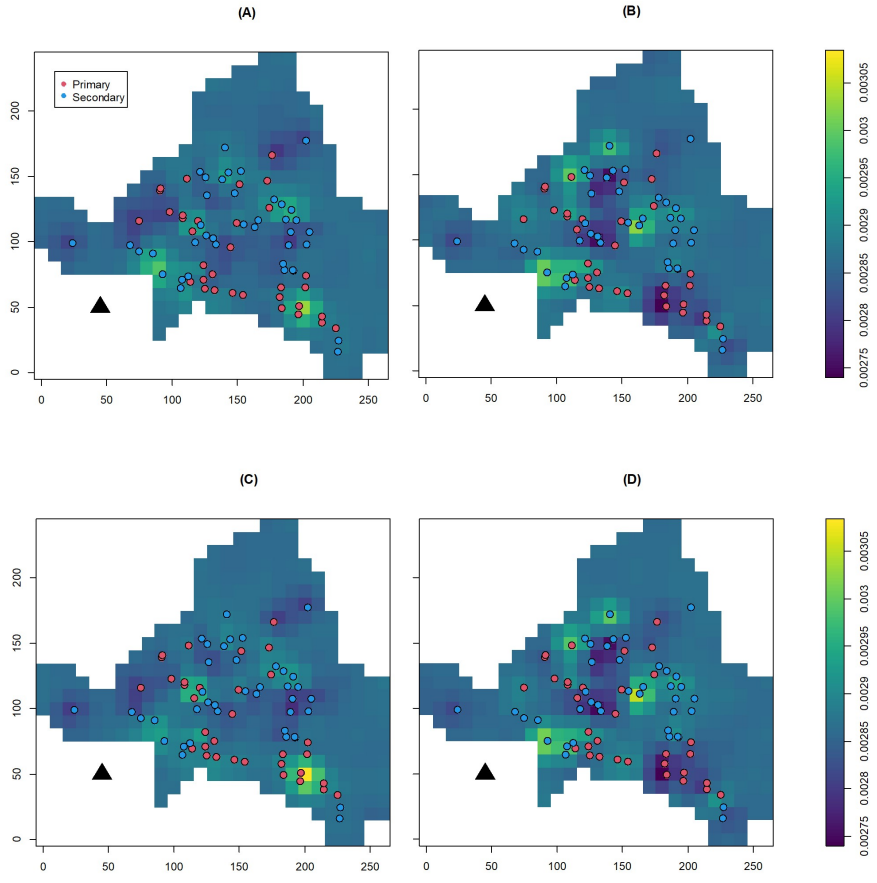


Figure 5.9: The estimated relative spatial densities of the barking deer for models M_0 (A and B) M_h, (C and D), corresponding to the wet (left) and the dry (right) seasons, respectively. The dots represent camera trap locations for primary forest (red) and secondary forest (blue). The black triangle represents Mt. Payung which is later excluded from the modeling.

To consider a more formal model selection approach in relation to the habitat covariate, a further RJMCMC step can be added to the algorithm, in order to obtain posterior model probabilities for M_0 and M_h. Implementing such an approach provides the associated posterior probabilities for model M_h of 0.54 for the dry season and 0.997 for the wet season. Figure 5.9 provides the corresponding spatial distribution of relative population densities for two fitted models, M_0 and M_h for each of the dry and wet seasons, respectively. We note that there appears to be a visible difference in density across the region between the two seasons (e.g. a higher density area in the south east during wet season compared to this being low density area during the dry season; and higher density patches in the centre and further north west in the dry season compared to the wet season). As expected, given the similarity in parameters for the dry season, there is little discernible contrast in the estimated densities between two models for the dry season. However, there are some minor differences observable for the wet season, most notably a small increase in the densities near traps in the primary forest areas.

5.3.4 Super population approach

For comparison purpose, we fit the model to the data using SPA but due to the computational expense, we restrict the analysis to M_0 for the dry season case. We specify similar priors to the RJMCMC algorithm for λ_0 and σ i.e., $\log(\lambda_0) \sim \text{N}(0, 10)$; $\sigma \sim \text{Uniform}(0, \infty)$. For the upper limit M, we assume a reasonably large upper limit with $M = 10000$. We introduce a similar weakly informative prior for N to reflect the knowledge we have on the population. In particular, we assume $\psi \sim \text{Beta}(6, 12)$ with a prior mean of 0.33, and 95% interval of (0.14, 0.56); this induces a prior on N with a mean of 3270 and 95% interval of (1421, 5596) to reflect the approximated spatial density of (5, 17). We note that a beta prior is a conjugate prior for $N \sim \text{Bin}(M, \psi)$. Finally, all parameters are tuned to produce approximately the same acceptance rate for updating the model parameters as the RJMCMC approach i.e., between 0.2 and 0.4.

We ran the MCMC simulations code for 500,000 iterations with an initial

Table 5.4: Posterior summary statistics, corresponding to the mean, standard deviation, median (Q50), 2.5% (Q2.5) and 97.5% (Q97.5) quantiles and effective sample size (ESS) of the model parameters fitted on M_0 (dry) using the SPA assuming $M = 10000$.

Parameter	**Mean**	**SD**	**Q2.5**	**Q50**	**Q97.5**	**ESS**
σ	0.95	0.16	0.71	0.93	1.31	1852
λ_0	1.49	0.29	1.04	1.45	2.16	9541
N	3746.90	1031.05	1910	3696	5871	1758
ψ	0.38	0.10	0.19	0.37	0.59	1777

10,000 iterations discarded for the burn-in period for three separate and independent chains. Table 5.4 shows the posterior summary statistics of the parameters for the fitted model. There seems no issue in convergence to the stationary distribution via an inspection of the trace plot and BGR statistic. The computational cost required for running a single chain in `R` using the more efficient bespoke likelihood calculation described in Section 4.2.3 was approximately 86 hours. Note that a single simulation originally took more than 50 hours for simply running 10,000 iterations on `nimble`, which does not incorporate the efficient computational calculation of the likelihood function.

Table 5.4 shows that the parameter estimates of σ and λ_0 are fairly similar to the stochastic RJMCMC algorithm, as we would expect allowing for Monte Carlo error and the minor difference in prior on N. In particular, the 95% credible intervals of the density (D) are very similar for both approaches i.e., between 6 and 17 individuals per km^2 obtained from the SPA and between 6 and 18 individuals per km^2 from the RJMCMC. On closer inspection, we note that the SPA, given a similar prior setting, produces slightly narrower credible intervals and smaller standard deviations (SD) for the posterior. Thus the estimates appear fairly similar but with some differences in the tails of the distribution. However, we would expect larger Monte Carlo errors with smaller effective sample sizes, with the ESS for SPA <2000 for and N; compared to >3000 for the RJMCMC approach (an increase of >1.5 times). Given the additional computational cost for each algorithm, we further observe that the ESS/minute are substantially higher for the RJMCMC compared to the SPA; >4 times higher for all parameters, and >8 times higher for N.

5.3.5 Sensitivity analysis

We consider a prior sensitivity analysis, varying the prior specified on the total population size, to investigate the influence of the weakly informative prior. In particular we consider two additional priors on N: (i) Neg-Bin(1, 0.001) with $\mu = 999$ and 95% interval (25, 3687); and (ii) Neg-Bin(5, 0.002) with $\mu = 2495$ and 95% interval (808, 5113).

Each assigned prior was fitted on the barking deer data for all models: M_0(dry), M_0(wet), M_h(dry), M_h(wet). We run 500,000 iterations plus 10,000 initial iterations as a burn-in period for a single chain using a stochastic RJMCMC algorithm. The posterior summary of statistics for all parameters on each assigned prior is given in Table 5.5 and Table 5.6 for models M_0 and M_h respectively.

In general, our study finds that the population density estimates i.e., posterior means and medians for the barking deer remained relatively robust, with significantly overlapping credible intervals across the different prior specifications for two given models. The posterior densities of the total population for model M_0 shown in Figure 5.10 also suggests highly overlapping posterior densities across the different priors. It is also noticed that there is an increased uncertainty in the population estimates (and associated density estimates) with the increasing vagueness of the priors.

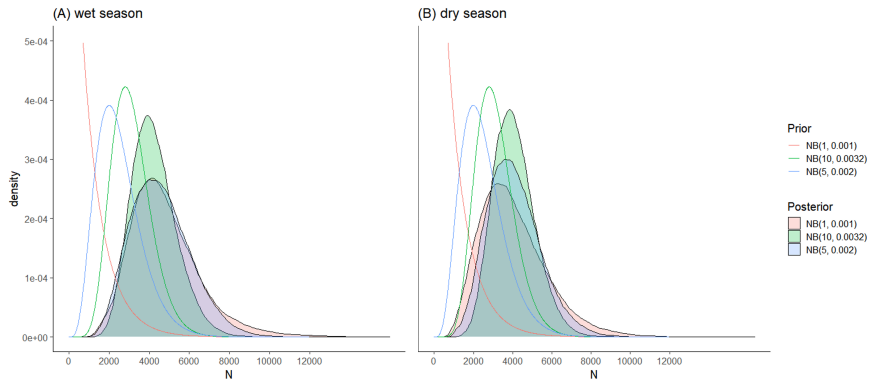


Figure 5.10: Posterior densities of the total population N for different choices of prior fitted on M_0 for (A) wet season (March-April) and (b) dry season (May-June).

Table 5.5: Posterior summaries of model parameters fitting on barking deer camera-trap data for the dry season and the wet season of the null model (M_0) for different choices of prior on N for assessing the sensitivity of the posterior distribution on the prior.

Prior on N	**Parameter**	**Mean**	**SD**	**Q2.5**	**Q50**	**Q97.5**
M_0 (Dry season)						
NB(10, 0.0032)	σ	0.90	0.14	0.67	0.88	1.22
	λ_0	1.43	0.26	1.00	1.41	2.03
	N	4129	1081	2269	4047	6440
	D	11.83	3.10	6.50	11.60	18.45
NB(5, 0.002)	σ	0.91	0.18	0.64	0.89	1.31
	λ_0	1.44	0.27	1.01	1.41	2.07
	N	4136	1404	1902	3967	7355
	D	11.85	4.02	5.45	11.37	21.07
NB(1, 0.001)	σ	0.95	0.26	0.60	0.91	1.60
	λ_0	1.44	0.27	1.01	1.41	2.05
	N	4099	1853	1245	3838	8528
	D	11.74	3.31	3.57	11.00	24.44
M_0 (Wet season)						
NB(10, 0.0032)	σ	0.57	0.09	0.42	0.56	0.78
	λ_0	3.39	0.73	2.28	3.27	5.08
	N	4605	1166	2556	4527	7112
	D	13.20	3.34	7.32	12.97	20.38
NB(5, 0.002)	σ	0.57	0.10	0.41	0.56	0.80
	λ_0	3.39	0.72	2.29	3.28	5.07
	N	4658	1370	2358	4551	7605
	D	13.35	3.92	6.76	13.04	21.79
NB(1, 0.001)	σ	0.57	0.13	0.37	0.54	0.89
	λ_0	3.38	0.73	2.27	3.26	5.07
	N	5164	2143	1846	4898	10120
	D	14.80	6.14	5.29	14.03	28.99

Table 5.6: Posterior summaries of model parameters fitting on barking deer camera-trap data for the dry season and the wet season of the habitat model (M_h) for different choices of prior on N for assessing the sensitivity of the posterior distribution on the prior.

Prior on N	**Parameter**	**Mean**	**SD**	**Q2.5**	**Q50**	**Q97.5**
M_h (Dry season)						
NB(10, 0.0032)	σ	0.88	0.13	0.66	0.87	1.17
	λ_{sec}	1.53	0.33	1.03	1.48	2.31
	λ_{pri}	1.37	0.32	0.87	1.32	2.12
	N	4180	1076	2362	4089	6521
	D	11.98	3.08	6.77	11.72	18.68
NB(5, 0.002)	σ	0.89	0.16	0.63	0.87	1.25
	λ_{sec}	1.53	0.34	1.03	1.48	2.36
	λ_{pri}	1.39	0.33	0.87	1.34	2.17
	N	4254	1368	2027	4090	7405
	D	12.19	3.92	5.81	11.72	21.22
NB(1, 0.001)	σ	0.88	0.16	0.56	0.85	1.36
	λ_{sec}	1.52	0.33	1.02	1.46	2.32
	λ_{pri}	1.38	0.33	0.87	1.33	2.14
	N	4664	2041	1711	4314	9693
	D	13.36	5.85	4.90	12.36	27.77
M_h (Wet season)						
NB(10, 0.0032)	σ	0.61	0.10	0.44	0.60	0.85
	λ_{sec}	1.70	0.44	1.07	1.63	2.79
	λ_{pri}	4.74	1.29	2.85	4.56	8.08
	N	4419	1182	2401	4318	7015
	D	12.66	3.39	6.88	12.37	20.10
NB(5, 0.002)	σ	0.61	0.12	0.42	0.59	0.90
	λ_{sec}	1.71	0.47	1.08	1.61	2.94
	λ_{pri}	4.66	1.24	2.83	4.50	7.60
	N	4596	1542	2098	4447	8087
	D	13.17	4.42	6.01	12.74	23.17
NB(1, 0.001)	σ	0.59	0.17	0.35	0.56	1.01
	λ_{sec}	1.70	0.42	1.08	1.62	2.76
	λ_{pri}	4.87	1.52	2.87	4.56	8.96
	N	5373	2490	1594	4996	11138
	D	15.4	7.13	4.57	14.32	31.91

5.4 Discussion

In this work, we provide a scalable Bayesian model-fitting algorithm for fitting spatial count models for unmarked individuals when the size of population and/or study presents computational challenges. The proposed trans-dimensional approach via an RJMCMC updating algorithm does not require an *a priori* upper limit on the population size and an associated data-augmentation expanded parameter space as for the traditional SPA commonly applied, permitting a greater scalability in applications. Further, the RJMCMC approach also permits a direct prior specification on the total population size, for which there may often be prior information, as for the barking deer case study. Bespoke code is required for implementing the RJMCMC algorithm, but this also provides the ability to include additional substantial computational savings within the updating of the activity centre parameters due to the particular structure of the likelihood. In particular, considering only differences in summations required (for the Poisson mean component) within the required likelihood calculations as opposed to a full recalculation of the likelihood term, provides a substantial computational saving, not generally possible in standard black-box MCMC packages. Demonstrable computational savings using these computational updates are also observed for the traditional SPA, so that (comparable) bespoke code is used for both the RJMCMC and SPA algorithms for consistency and informative comparability within this thesis. The improved comparative performance (in terms of ESS/sec) of the proposed RJMCMC algorithm compared to the alternative SPA approach is noticeable even on a relatively small dataset, as a result of the increase in computational time for the RJMCMC algorithm compared to the SPA. The improvement depends on the exact parameter and prior used by ranges from an improvement of approximately 3-20 for the cases we considered.

The differences between the traditional SPA and proposed RJMCMC algorithm are particularly marked for the motivating case study relating to barking deer. The SPA algorithm took substantially longer to implement compared to the RJMCMC algorithm (by a factor of 4) while having substantially lower effective sample size (ESS). The density estimate of the barking deer in UNKP is

substantially higher than many other regions, with ranges of approximately 2-3.5 animals per km^2, but substantially less than for the study at Baluran National Park, Indonesia with an estimated density of 25 per km^2. There are several possible factors which may influence the barking deer density (or equivalently population size) in UNKP. A previous study found that there is a relatively balanced sex ratio between adult males and females at 1.37:1 suggesting evidence of regular recruitment into the population (Rahman, 2016). In addition, the specific habitat within the national park may be a factor. In particular, the primary and secondary forest occupies $> 90\%$ of the park, with primary forest known to be dominated by emergent plants and tree species while palms and other fruit trees are mainly dominant in the secondary forest (Rahman et al., 2017). Although there is no record or data regarding food choices of barking deer at UKNP, a study at Baluran park, Indonesia found that trees, shrubs, grasses, forbs and climbers are frequently consumed by these species (Tyson, 2007), thus suggesting an abundant food supply for barking deer within the park. However, not all these foods can be found during the whole year, which may also be a factor in the spatial density of the deer within the park for different seasons (see Figure 5.9).

Chapter 6

Conclusions

This thesis introduced new efficient model fitting algorithms using classical and Bayesian approaches for fitting large, capture-recapture and camera-trap data where the observed data likelihood is analytically intractable. Computational efficiency continues to be an important consideration, as models become more complex and/or datasets/studies increase in size. Model fitting approaches that may work well in low-dimensional spaces may not necessarily work well in high-dimensional spaces. We answer such challenges by providing scalable likelihood and Bayesian methods.

In our first works in Chapters 2 and 3, we proposed Laplace approximations for fitting marked capture-recapture data with missing individual heterogeneity. In particular, we developed second-order and fourth-order Laplace methods as an approximation to the intractable observed data likelihood of the M_h and CJS models. Laplace approximations provide a reliable and efficient mechanism for evaluating such likelihood functions and are scalable to higher dimensions. With the help of readily-available `R` package such as `TMB`, the implementation of Laplace approximations becomes more straightforward and thus this approach is potentially a very attractive avenue to pursue for more complex models.

Our findings from the first works suggest that results are fairly consistent with what have been found in Warton et al. (2015). They described and compared several approaches in terms of speed and accuracy for fitting joint ecological mixed models, including Laplace approximations, adaptive quadrature, data imputation approaches and variational approximations. Within their applications considered, the Laplace approximation was a very computationally efficient approach, but less accurate for small samples; while adaptive quadrature appeared a reasonable compromise between accuracy and speed. The Expectation-Maximisation (EM) algorithm and Bayesian data augmentation were accurate but relatively slow. Variational approximations, appeared to be as computationally efficient as Laplace approximations whilst also providing moderately accurate estimates. Further, Niku et al. (2019) showed that variational approximations coded in `TMB` had smaller bias for negative binomial generalized linear latent variable models compared with the second order Laplace approximation. Exploring the use of

variational approximations for capture-recapture models is also an area of current research of this work, as well as investigating the robustness of Laplace approximation to models where the individual heterogeneity component is non-Gaussian. Altenatively, implementing Laplace approximations to marked SCR model is also an interesting direction for future research.

In Chapters 4 and 5, we developed more efficient Bayesian MCMC algorithms for fitting a large unmarked SCR dataset in camera-trapping studies which removes the necessity of *priori* setting an upper bound on the population size, and also permits the direct specification of a prior on the total population size. Our simulation studies and examples show that the new model fitting approaches, reversible jump MCMC, are consistently more efficient, yet attaining similarities in terms of mixing compared to the existing approach, super population algorithm. This alternate Bayesian method might provide a solution to the scalability challenge in the current Bayesian data augmentation. In addition, the new proposal distribution i.e., stochastic removal of activity centres on the reversible jump step for updating N which successfully improve the mixing, approximately twice better than the old proposal distribution, becomes more appealing. Providing a readily-used platform in `R` such as `nimble` for the RJMCMC when trans-dimensional algorithms are needed can be an interesting future research of this work.

However, a further and careful examination of model assumptions may be needed as violations to such assumptions may lead to bias estimates of the spatial density. As for traditional (marked) SCR, the spatial count model requires at least three assumptions: population closure, independence (over locations and individuals) and homogeneous activity centers (Chandler and Royle, 2013). It has been shown in the standard marked SCR that population density estimates are robust to low-moderate violations of these assumptions (Efford et al., 2016; Efford, 2019; Bischof et al., 2020; Theng et al., 2022). The closure assumption can be easily violated when longer periods of study, typically longer than 3 months for mammals, is conducted (Silver et al., 2004; Soria-Díaz and Monroy-Vilchis, 2015; Rahman, 2016). Depending on ecological questions, the model may be extended by allowing more relaxed assumptions on population where the population may

change over time (Dail and Madsen, 2011; Chandler et al., 2011).

The assumption of independent movement of individuals may be easily violated when animals move in (small or large) groups/herds. Recent research suggested that low to moderate levels of aggregation of individuals (group sizes < 8) introduce small biases in density estimation and the scale parameter σ in SCR (Bischof et al., 2020; Theng et al., 2022). These biases increase positively as the aggregation levels increase in particular when the group size is larger than 64, the bias is estimated to be over 20%. Additionally, there is a substantial decrease in coverage probabilities as the group size increases. Current research focuses on investigating unmarked spatial models for deviations to such modelling assumptions. Alternatively, the assumption of homogeneous activity centres implies that activity centres of animals do not change over time which may not be true in practice. This has lead to the proposal of a Markovian transient model for the activity centres in the traditional marked SCR (Royle et al., 2016; Theng et al., 2022). However, allowing activity centers to change over times (e.g. seasons) substantially increase computational efforts, and developing associated efficient computational tools, and/or alternative models, for unmarked SCR is also an area of future research of this work.

Bibliography

Anderson, D. R. (2001). The need to get the basics right in wildlife field studies. *Wildlife Society Bulletin*, 29:1294–1297.

Baddeley, A., Turner, R., and Rubak, E. (2022). spatstat: Spatial Point Pattern Analysis, Model-Fitting, Simulation, Tests.

Baillargeon, S. and Rivest, L.-P. (2007). Rcapture: Loglinear Models for Capture-Recapture in R. *Journal of Statistical Software*, 19:1–31.

Balme, G. A., Slotow, R., and Hunter, L. T. B. (2010). Edge effects and the impact of non-protected areas in carnivore conservation: leopards in the Phinda–Mkhuze Complex, South Africa. *Animal Conservation*, 13(3):315–323.

Barndorff-Nielsen, O. and Cox, D. R. (1979). Edgeworth and Saddle-Point Approximations with Statistical Applications. *Journal of the Royal Statistical Society. Series B (Methodological)*, 41(3):279–312.

Bischof, R., Dupont, P., Milleret, C., Chipperfield, J., and Royle, J. A. (2020). Consequences of ignoring group association in spatial capture–recapture analysis. *Wildlife Biology*, 2020(1).

Bonner, S. and Schofield, M. (2014). MC(MC)MC: exploring Monte Carlo integration within MCMC for mark–recapture models with individual covariates. *Methods in Ecology and Evolution*, 5(12):1305–1315. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/2041-210X.12095.

Bonner, S. J., Morgan, B. J. T., and King, R. (2010). Continuous Covariates in Mark-Recapture-Recovery Analysis: A Comparison of Methods. *Biometrics*, 66(4):1256–1265.

Bonner, S. J. and Schwarz, C. J. (2006). An Extension of the Cormack-Jolly-Seber Model for Continuous Covariates with Application to Microtus pennsylvanicus. *Biometrics*, 62(1):142–149.

Borchers, D., Buckland, S. T., and Zucchini, W. (2002). *Estimating Animal Abundance: Closed Populations*. Springer, London.

Borchers, D., Distiller, G., Foster, R., Harmsen, B., and Milazzo, L. (2014). Continuous-time spatially explicit capture–recapture models, with an application to a jaguar camera-trap survey. *Methods in Ecology and Evolution*, 5:656–665.

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

Brassine, E. and Parker, D. (2015). Trapping Elusive Cats: Using Intensive Camera Trapping to Estimate the Density of a Rare African Felid. *PLOS ONE*, 10(12):e0142508.

Breslow, N. E. and Lin, X. (1995). Bias Correction in Generalised Linear Mixed Models with a Single Component of Dispersion. *Biometrika*, 82:12.

Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms. *J. Inst. Maths Applics*, 6:222–231.

Catchpole, E. A., Morgan, B. J. T., and Coulson, T. (2004). Conditional methodology for individual case history data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 53(1):123–131.

Chandler, R. B. and Royle, J. A. (2013). Spatially explicit models for inference about density in unmarked or partially marked populations. *The Annals of Applied Statistics*, 7(2):936–954.

Chandler, R. B., Royle, J. A., and King, D. I. (2011). Inference about density and temporary emigration in unmarked populations. *Ecology*, 92(7):1429–1435.

Chao, A., Yip, P. S., Lee, S.-M., and Chu, W. (2001). Population size estimation based on estimating functions for closed capture–recapture models. *Journal of Statistical Planning and Inference*, 92(1-2):213–232.

Chapman, S. and Balme, G. (2010). An Estimate of Leopard Population Density in a Private Reserve in KwaZulu-Natal, South Africa, using Camera—Traps and Capture-Recapture Models. *South African Journal of Wildlife Research*, 40(2):114–120.

Chavez-Demoulin, V. (1999). Bayesian Inference for Small-Sample Capture-Recapture Data. *Biometrics*, 55(3):727–731. Publisher: [Wiley, International Biometric Society].

Connor, T., Division, W., Tripp, E., Bean, W. T., Saxon, B. J., Camarena, J., Donahue, A., Sarna-Wojcicki, D., Macaulay, L., Tripp, W., and Brashares, J. (2022). Estimating wildlife density as a function of environmental heterogeneity using unmarked data. *Remote Sensing*, 14(5):1087.

Cormack, R. M. (1964). Estimates of Survival from the Sighting of Marked Animals. *Biometrika*, 51:429–438.

Cormack, R. M. (1989). Log-Linear Models for Capture-Recapture. *Biometrics*, 45(2):395.

Coull, B. A. and Agresti, A. (1999). The Use of Mixed Logit Models to Reflect Heterogeneity in Capture-Recapture Studies. *Biometrics*, 55(1):294–301.

Dahaban, Z., Nordin, M., and Bennett, E. L. (1996). Immediate effects on wildlife of selective logging in a hill dipterocarp forest in Sarawak: mammals. In Edwards, D. S., Booth, W. E., and Choy, S. C., editors, *Tropical Rainforest Research — Current Issues: Proceedings of the Conference held in Bandar Seri Begawan, April 1993*, Monographiae Biologicae, pages 341–346. Springer Netherlands, Dordrecht.

Dail, D. and Madsen, L. (2011). Models for estimating abundance from repeated counts of an open metapopulation. *Biometrics*, 67:577–587.

Di Bitetti, M. S., Di Blanco, Y. E., Pereira, J. A., Paviolo, A., and Pírez, I. J. (2009). Time Partitioning Favors the Coexistence of Sympatric Crab-Eating Foxes (Cerdocyon thous) and Pampas Foxes (Lycalopex gymnocercus). *Journal of Mammalogy*, 90(2):479–490.

Dorazio, R. M. and Royle, J. A. (2003). Mixture Models for Estimating the Size of a Closed Population When Capture Rates Vary among Individuals. *Biometrics*, 59(2):351–364.

Durban, J. W. and Elston, D. A. (2005). Mark: Recapture with Occasion and Individual Effects: Abundance Estimation through Bayesian Model Selection in a Fixed Dimensional Parameter Space. *Journal of Agricultural, Biological, and Environmental Statistics*, 10(3):291–305.

eBird, T. (2021). eBird passes 1 billion bird observations - eBird.

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

Efford, M. G. (2019). Non-circular home ranges and the estimation of population density. *Ecology*, 100(2):e02580.

Efford, M. G., Dawson, D. K., Jhala, Y. V., and Qureshi, Q. (2016). Density-dependent home-range size revealed by spatially explicit capture-recapture. *Ecography*, 39(7):676–688.

Evans, M. J. and Rittenhouse, T. A. G. (2018). Evaluating spatially explicit density estimates of unmarked wildlife detected by remote cameras. *Journal of Applied Ecology*, 55(6):2565–2574.

Farley, S. S., Dawson, A., Goring, S. J., and Williams, J. W. (2018). Situating Ecology as a Big-Data Science: Current Advances, Challenges, and Solutions. *BioScience*, 68(8):563–576.

Fienberg, S. E., Johnson, M. S., and Junker, B. W. (1999). Classical Multilevel and Bayesian Approaches to Population Size Estimation Using Multiple Lists. *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 162(3):383–405. Publisher: [Wiley, Royal Statistical Society].

Gilks, W. R., Richardson, S., and Spiegelhalter, D. (1995). Introducing Markov Chain Monte Carlo. In *Markov Chain Monte Carlo in Practice*. CRC Press.

Gimenez, O. and Choquet, R. (2010). Individual heterogeneity in studies on marked animals using numerical integration: capture–recapture mixed models. *Ecology*, 91(4):951–957.

Glennie, R., Adam, T., Leos-Barajas, V., Michelot, T., Photopoulou, T., and McClintock, B. T. (2023). Hidden Markov models: Pitfalls and opportunities in ecology. *Methods in Ecology and Evolution*, 14(1):43–56.

Goudie, I. B. J. and Goudie, M. (2007). Who Captures the Marks for the Petersen Estimator? *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 170(3):825–839. Publisher: [Wiley, Royal Statistical Society].

Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82(4):711–732.

Guggisberg, C. A. (1977). *Early Wildlife Photographers*. Taplinger Publishing Company, Incorporated, New York.

Howe, E. J., Buckland, S. T., Després-Einspenner, M., and Kühl, H. S. (2017). Distance sampling with camera traps. *Methods in Ecology and Evolution*, 8(11):1558–1565.

Huber, P., Ronchetti, E., and Victoria-Feser, M.-P. (2004). Estimation of generalized linear latent variable models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(4):893–908.

Huggins, R. and Hwang, W.-H. (2011). A Review of the Use of Conditional Likelihood in Capture-Recapture Experiments. *International Statistical Review*, 79(3):385–400. Publisher: International Statistical Institute.

Jachmann, H. (2002). Comparison of aerial counts with ground counts for large African herbivores. *Journal of Applied Ecology*, 39(5):841–852. _eprint: https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1046/j.1365-2664.2002.00752.x.

Jennelle, C. S., Runge, M. C., and MacKenzie, D. I. (2002). The use of photographic rates to estimate densities of tigers and other cryptic mammals: a comment on misleading conclusions. *Animal Conservation*, 5(2):119–120. _eprint: https://zslpublications.onlinelibrary.wiley.com/doi/pdf/10.1017/S1367943002002160.

Jolly, G. M. (1965). Explicit Estimates from Capture-Recapture Data with Both Death and Immigration-Stochastic Model. *Biometrika*, 52:225–247.

Karanth, K. U. and Nichols, J. D. (1998). Estimation of Tiger Densities in India Using Photographic Captures and Recaptures. *Ecology*, 79(8):2852–2862.

Karanth, K. U., Nichols, J. D., Kumar, N. S., and Hines, J. E. (2006). Assessing Tiger Population Dynamics Using Photographic Capture–Recapture Sampling. *Ecology*, 87(11):2925–2937. _eprint: https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.1890/0012-9658%282006%2987%5B2925%3AATPDUP%5D2.0.CO%3B2.

Kays, R., Arbogast, B. S., Baker-Whatton, M., Beirne, C., Boone, H. M., Bowler, M., Burneo, S. F., Cove, M. V., Ding, P., Espinosa, S., Gonçalves, A. L. S., Hansen, C. P., Jansen, P. A., Kolowski, J. M., Knowles, T. W., Lima, M. G. M., Millspaugh, J., McShea, W. J., Pacifici, K., Parsons, A. W., Pease, B. S., Rovero, F., Santos, F., Schuttler, S. G., Sheil, D., Si, X., Snider, M., and Spironello, W. R. (2020). An empirical evaluation of camera trap study design: How many, how long and when? *Methods in Ecology and Evolution*, 11(6):700–713.

King, R. and Brooks, S. P. (2001). On the Bayesian analysis of population size. *Biomterika*, 88(2):317–336.

King, R. and Brooks, S. P. (2008). On the Bayesian Estimation of a Closed

Population Size in the Presence of Heterogeneity and Model Uncertainty. *Biometrics*, 64(3):816–824.

King, R., Brooks, S. P., and Coulson, T. (2008). Analyzing Complex Capture-Recapture Data in the Presence of Individual and Temporal Covariates and Model Uncertainty. *Biometrics*, 64(4):1187–1195.

King, R., Brooks, S. P., Morgan, B. J. T., and Coulson, T. (2006). Factors Influencing Soay Sheep Survival: A Bayesian Analysis. *Biometrics*, 62(1):211–220. Publisher: [Wiley, International Biometric Society].

King, R., McClintock, B. T., Kidney, D., and Borchers, D. (2016). Capture–recapture abundance estimation using a semi-complete data likelihood approach. *The Annals of Applied Statistics*, 10(1):264–285.

King, R. and McCrea, R. (2019). Capture–Recapture Methods and Models: Estimating Population Size. In *Handbook of Statistics*, volume 40, pages 33–83. Elsevier.

King, R., Sarzo, B., and Elvira, V. (2022). Large Data and (Not Even Very) Complex Ecological Models: When Worlds Collide.

Kristensen, K., Nielsen, A., Berg, C. W., Skaug, H., and Bell, B. M. (2016). **TMB** : Automatic Differentiation and Laplace Approximation. *Journal of Statistical Software*, 70(5).

Kucera, T. E. and Barrett, R. H. (1993). In My Experience: The Trailmaster® Camera System for Detecting Wildlife. *Wildlife Society Bulletin (1973-2006)*, 21(4):505–508.

Langrock, R. and King, R. (2013). Maximum likelihood estimation of mark–recapture–recovery models in the presence of continuous covariates. *The Annals of Applied Statistics*, 7(3):1709–1732.

Li, N. and Stephens, M. (2003). Modeling Linkage Disequilibrium and Identifying Recombination Hotspots Using Single-Nucleotide Polymorphism Data. *Genetics*, 165(4):2213–2233.

Link, W. A. (2003). Nonidentifiability of Population Size from Capture-Recapture Data with Heterogeneous Detection Probabilities. *Biometrics*, 59(4):1123–1130. Publisher: [Wiley, International Biometric Society].

Link, W. A. (2006). Rejoinder to "On Identifiability in Capture-Recapture Models". *Biometrics*, 62(3):936–939. Publisher: [Wiley, International Biometric Society].

Link, W. A. (2013). A cautionary note on the discrete uniform prior for the binomial *N*. *Ecology*, 94(10):2173–2179.

Liu, Q. (1994). A Note on Gauss-Hermite Quadrature. *Biometrika*, 81:7.

Long, E. S., Fecske, D. M., Sweitzer, R. A., Jenks, J. A., Pierce, B. M., and Bleich, V. C. (2003). Efficacy of Photographic Scent Stations to Detect Mountain Lions. *Western North American Naturalist*, 63(4):529–532. Publisher: Monte L. Bean Life Science Museum, Brigham Young University.

Mace, R. D., Minta, S. C., Manley, T. L., and Aune, K. E. (1994). Estimating Grizzly Bear Population Size Using Camera Sightings. *Wildlife Society Bulletin*, 22(1):74–83.

Marcus Rowcliffe, J., Carbone, C., Jansen, P. A., Kays, R., and Kranstauber, B. (2011). Quantifying the sensitivity of camera traps: an adapted distance sampling approach. *Methods in Ecology and Evolution*, 2(5):464–476.

McClintock, B. T., White, G. C., Burnham, K. P., and Pryde, M. A. (2009). A generalized mixed effects model of abundance for mark-resight data when sampling is without replacement. In Thomson, D. L., Cooch, E. G., and Conroy, M. J., editors, *Modeling demographic processes in marked populations, environmental and ecological statistics 3*, pages 271–289. Springer US, New York.

McCrea, R. S. and Morgan, B. J. T. (2015). *Analysis of Capture-Recapture Data*. CRC Press, Boca Raton.

McLaughlin, P. (2019). *On the topic of spatial capture-recapture modeling*. Doctoral Thesis, University of Connecticut.

Mews, S., Langrock, R., King, R., and Quick, N. (2020). Continuous-time multi-state capture-recapture models. arXiv:2002.10997 [stat].

Nakashima, Y., Fukasawa, K., and Samejima, H. (2018). Estimating animal density without individual recognition using information derivable exclusively from camera traps. *Journal of Applied Ecology*, 55(2):735–744.

Nichols, J. D., Sauer, J. R., Pollock, K. H., and Hestbeck, J. B. (1992). Estimating Transition Probabilities for Stage-Based Population Projection Matrices Using Capture-Recapture Data. *Ecology*, 73(1):306–312.

Niku, J., Brooks, W., Herliansyah, R., Hui, F. K. C., Taskinen, S., and Warton, D. I. (2019). Efficient estimation of generalized linear latent variable models. *PLOS ONE*, 14(5):e0216129.

Osborne, M. R. (1992). Fisher's Method of Scoring. *International Statistical Review / Revue Internationale de Statistique*, 60(1):99–117.

Otis, D. L., Burnham, K. P., White, G. C., and Anderson, D. R. (1978). Statistical Inference from Capture Data on Closed Animal Populations. *Wildlife Monographs*, (62):3–135. Publisher: [Wiley, Wildlife Society].

O'Connell, A. F., Nichols, J. D., and Karanth, K. U., editors (2011). *Camera Traps in Animal Ecology*. Springer Japan, Tokyo.

Palencia, P., Rowcliffe, J. M., Vicente, J., and Acevedo, P. (2021). Assessing the camera trap methodologies used to estimate density of unmarked populations. *Journal of Applied Ecology*, 58(8):1583–1592.

Pereira, K. S., Gibson, L., Biggs, D., Samarasinghe, D., and Braczkowski, A. R. (2022). Individual Identification of Large Felids in Field Studies: Common Methods, Challenges, and Implications for Conservation Science. *Frontiers in Ecology and Evolution*, 10.

Pettorelli, N., Côté, S. D., Gingras, A., Potvin, F., and Huot, J. (2007). Aerial Surveys Vs Hunting Statistics To Monitor Deer Density: The Example Of

Anticosti Island, Québec, Canada. *Wildlife Biology*, 13(3):321–327. Publisher: Nordic Board for Wildlife Research.

Pledger, S. (2000). Unified maximum likelihood estimates for closed capture-recapture models using mixtures. *Biometrics*, 56(2):434–442.

Pledger, S., Pollock, K. H., and Norris, J. L. (2003). Open capture-recapture models with heterogeneity: I. Cormack-Jolly-Seber model. *Biometrics*, 59(4):786–794.

Pledger, S., Pollock, K. H., and Norris, J. L. (2010). Open Capture–Recapture Models with Heterogeneity: II. Jolly–Seber Model. *Biometrics*, 66(3):883–890.

Plummer, M., Best, N., Cowles, K., Vines, K., Sarkar, D., Bates, D., Almond, R., and details, A. M. c. a. (2020). coda: Output Analysis and Diagnostics for MCMC.

Plummer, M., Stukalov, A., and Denwood, M. (2022). rjags: Bayesian Graphical Models using MCMC.

Pollock, K. H., Nichols, J. D., Simons, T. R., Farnsworth, G. L., Bailey, L. L., and Sauer, J. R. (2002). Large scale wildlife monitoring studies: statistical methods for design and analysis. *Environmetrics*, 13(2):105–119.

Rahman, D. A. (2016). *New insights into ecology and conservation status of Bawean deer (Axis kuhlii) and red muntjac (Muntiacus muntjak) in Indonesian tropical rainforest*. These de doctorat, Toulouse 3.

Rahman, D. A., Gonzalez, G., Haryono, M., Muhtarom, A., Firdaus, A. Y., and Aulagnier, S. (2017). Factors affecting seasonal habitat use, and predicted range of two tropical deer in Indonesian rainforest. *Acta Oecologica*, 82:41–51.

Ramsey, D. S. L., Caley, P. A., and Robley, A. (2015). Estimating population density from presence-absence data using a spatially explicit model: Estimating Density From Presence-Absence Data. *The Journal of Wildlife Management*, 79(3):491–499.

Raudenbush, S. W., Yang, M.-L., and Matheos, Y. (2000). Maximum Likelihood for Generalized Linear Models with Nested Random Effects via High-Order, Multivariate Laplace Approximation. *Journal of Computational and Graphical Statistics*, 9:131–157.

Rosenberg, D. K., Overton, W. S., and Anthony, R. G. (1995). Estimation of Animal Abundance When Capture Probabilities Are Low and Heterogeneous. *The Journal of Wildlife Management*, 59(2):252.

Rowcliffe, J. M., Field, J., Turvey, S. T., and Carbone, C. (2008). Estimating animal density using camera traps without the need for individual recognition. *Journal of Applied Ecology*, 45(4):1228–1236.

Royle, J. A. (2004). N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, 60(1):108–115.

Royle, J. A. (2008). Modeling individual effects in the Cormack-Jolly-Seber model: a state-space formulation. *Biometrics*, 64(2):364–370.

Royle, J. A., Chandler, R. B., Sollmann, R., and Gardner, B. (2014). *Spatial Capture-Recapture*. Elsevier.

Royle, J. A., Dorazio, R. M., and Link, W. A. (2007). Analysis of Multinomial Models With Unknown Index Using Data Augmentation. *Journal of Computational and Graphical Statistics*, 16(1):67–85.

Royle, J. A., Fuller, A. K., and Sutherland, C. (2016). Spatial capture-recapture models allowing Markovian transience or dispersal. *Population Ecology*, 58(1):53–62.

Royle, J. A., Karanth, K. U., Gopalaswamy, A. M., and Kumar, N. S. (2009). Bayesian inference in camera trapping studies for a class of spatial capture–recapture models. *Ecology*, 90(11):3233–3244.

Royle, J. A. and Young, K. V. (2008). A hierarchical model for spatial capture–recapture data. *Ecology*, 89(8):2281–2289.

Schnabel, Z. E. (1938). The Estimation of Total Fish Population of a Lake. *The American Mathematical Monthly*, 45(6):348–352. Publisher: Mathematical Association of America.

Schofield, M. R. and Barker, R. J. (2014). Hierarchical modeling of abundance in closed population capture–recapture models under heterogeneity. *Environmental and Ecological Statistics*, 21(3):435–451.

Schwarz, C. J. (2001). The Jolly-seber model: More than just abundance. *Journal of Agricultural, Biological, and Environmental Statistics*, 6(2):195–205.

Seber, G. A. F. (1965). A Note on the Multiple-Recapture Census. *Biometrika*, 52:249–259.

Seber, G. A. F. and Schofield, M. R. (2019). *Capture-Recapture: Parameter Estimation for Open Animal Populations*. Statistics for Biology and Health. Springer International Publishing, Cham.

Silver, S. C., Ostro, L. E. T., Marsh, L. K., Maffei, L., Noss, A. J., Kelly, M. J., Wallace, R. B., Gómez, H., and Ayala, G. (2004). The use of camera traps for estimating jaguar *Panthera onca* abundance and density using capture/recapture analysis. *Oryx*, 38(2):148–154.

Smith, P. J. (1991). Bayesian Analyses for a Multiple Capture-Recapture Model. *Biometrika*, 78(2):399–407. Publisher: [Oxford University Press, Biometrika Trust].

Soria-Díaz, L. and Monroy-Vilchis, O. (2015). Monitoring population density and activity pattern of white-tailed deer (Odocoileus virginianus) in Central Mexico, using camera trapping. *Mammalia*, 79(1):43–50.

Srbek-Araujo, A. C. and Chiarello, A. G. (2005). Is camera-trapping an efficient method for surveying mammals in Neotropical forests? A case study in south-eastern Brazil. *Journal of Tropical Ecology*, 21(1):121–125.

Srikosamatara, S. (1993). Density and biomass of large herbivores and other mammals in a dry tropical forest, western Thailand. *Journal of Tropical Ecology*, 9(1):33–43. Publisher: Cambridge University Press.

Steen, N. M., Byrne, G. D., and Gelbard, E. M. (1969). Gaussian quadratures for the integrals{ $\int_{0}^{} \(-)()$} and {$\int_{0}^{}\(-)()$}. *Mathematics of Computation*, 23(107):661–671.

Stevenson, B. C., Dam-Bates, P., Young, C. K. Y., 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.

Stoklosa, J., Hwang, W.-H., Wu, S.-H., and Huggins, R. (2011). Heterogeneous Capture—Recapture Models with Covariates: A Partial Likelihood Approach for Closed Populations. *Biometrics*, 67(4):1659–1665. Publisher: International Biometric Society.

Strawderman, R. L. (2000). Higher-Order Asymptotic Approximation: Laplace, Saddlepoint, and Related Methods. *Journal of the American Statistical Association*, 95(452):1358–1364.

Sweitzer, R. A., Van Vuren, D., Gardner, I. A., Boyce, W. M., and Waithman, J. D. (2000). Estimating Sizes of Wild Pig Populations in the North and Central Coast Regions of California. *The Journal of Wildlife Management*, 64(2):531–543.

Tanner, M. A. and Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, 82(398):528–540.

Theng, M., Milleret, C., Bracis, C., Cassey, P., and Delean, S. (2022). Confronting spatial capture–recapture models with realistic animal movement simulations. *Ecology*, 103(10):e3676. Publisher: John Wiley & Sons, Ltd.

Tracey, J. P., Fleming, P. J. S., Melville, G. J., Tracey, J. P., Fleming, P. J. S., and Melville, G. J. (2008). Accuracy of some aerial survey estimators: contrasts with known numbers. *Wildlife Research*, 35(4):377–384.

Trolle, M. and Kéry, M. (2003). Estimation of Ocelot Density in the Pantanal Using Capture-Recapture Analysis of Camera-Trapping Data. *Journal of Mammalogy*, 84(2):607–614.

Trolle, M. and Kéry, M. (2005). Camera-trap study of ocelot and other secretive mammals in the northern Pantanal. 69(3-4):409–416.

Tyson, M. J. (2007). *The ecology of Muntjak deer (Muntiacus muntjak) in Baluran National Park, Java and their interactions with other mammal*. PhD thesis, Manchester Metropolitan University.

Valpine, P. d., Paciorek, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F., Wehrhahn Cortes, C., Rodrìguez, A., Temple Lang, D., Paganin, S., Hug, J., Babu, J., Ponisio, L., and Sujan, P. (2022). nimble: MCMC, Particle Filtering, and Programmable Hierarchical Modeling.

Van dam Bates, P. (2023). *Counting processes for spatial capture-recapture*. PhD thesis, University of St Andrews.

Wallace, R. B., Gomez, H., Ayala, G., and Espinoza, F. (2003). Camera trapping for jaguar (panthera onca) in the tuichi valley, Bolivia. *J Neotropical Mammology*, 10:133–139.

Warton, D. I., Blanchet, F. G., O'Hara, R. B., Ovaskainen, O., Taskinen, S., Walker, S. C., and Hui, F. K. (2015). So Many Variables: Joint Modeling in Community Ecology. *Trends in Ecology & Evolution*, 30(12):766–779.

Wegge, P. and Mosand, H. (2015). Can the mating system of the size-monomorphic Indian muntjac (*Muntiacus muntjak*) be inferred from its social structure, spacing behaviour and habitat? A case study from lowland Nepal. *Ethology Ecology & Evolution*, 27(2):220–232.

Wegge, P. and Storaas, T. (2009). Sampling tiger ungulate prey by the distance method: lessons learned in Bardia National Park, Nepal. *Animal Conservation*, 12(1):78–84.

White, G. C. and Cooch, E. G. (2017). Population abundance estimation with heterogeneous encounter probabilities using numerical integration: Individual Heterogeneity and Abundance Estimation. *The Journal of Wildlife Management*, 81(2):322–336.

Wong, W. H. and Li, B. (1992). Laplace Expansion for Posterior Densities of Nonlinear Functions of Parameters. 7:7.

Worthington, H., King, R., and Buckland, S. T. (2015). Analysing Mark–Recapture–Recovery Data in the Presence of Missing Covariate Data Via Multiple Imputation. *Journal of Agricultural, Biological, and Environmental Statistics*, 20(1):28–46.

Worthington, H., McCrea, R., King, R., and Griffiths, R. (2019). Estimating abundance from multiple sampling capture-recapture data via a multi-state multi-period stopover model. *The Annals of Applied Statistics*, 13(4):2043–2064.

Yokoyama, Y., Nakashima, Y., Yajima, G., and Miyashita, T. (2020). Simultaneous estimation of seasonal population density, habitat preference and catchability of wild boars based on camera data and harvest records. *Royal Society Open Science*, 7(8):200579.

Zucchini, W., MacDonald, I. L., and Langrock, R. (2016). *Hidden Markov Models for Time Series: An Introduction Using R, Second Edition*. Chapman and Hall/CRC, New York, 2 edition.