# Bayesian hierarchical models for ecological data: estimating population size, spatial and temporal patterns

University of Kent

**Fabian Ricardo Ketwaroo**

School of Mathematics, Statistics and Actuarial Science

University of Kent

This dissertation is submitted for the degree of

*Doctor of Philosophy*

April 2023

This dissertation is dedicated to my late grandmother, Deowlene Harrilall, who has been a constant source of inspiration and support throughout my life. Her unwavering love and kindness have been instrumental in shaping who I am today and have been invaluable in helping me navigate the challenges of my academic journey. This dissertation is a testament to her influence and a small tribute to the impact she has had on my life.

# Acknowledgements

On a personal note, to my parents, Pradeep and Salome, I would like to express my deepest gratitude and appreciation for your unwavering support, encouragement, and sacrifices throughout my academic journey. Without your love, guidance, and belief in my abilities, I would not have been able to achieve this significant milestone in my life. You have supported me financially, emotionally and always believed in me even when I doubted myself. To my siblings and other family members, thank you for all your unwavering love and support.

I am equally grateful to my wife Afeena, for her unwavering support and love throughout my academic journey. Her patience, understanding, and constant encouragement have been the pillars of my success. She has been there for me in the good times and the bad, offering a listening ear, a shoulder to cry on, and a cheerleader when I needed it the most. Her sacrifices and dedication have made it possible for me to pursue this dream and I will always be grateful for her unwavering love and support. Without you, this achievement would not have been possible.

# Abstract

The work in this thesis presents three manuscripts, described in Chapters 2 to 4.

Chapter 2 presents an evaluation of the popular N-mixture model in a Bayesian framework to corroborate and extend issues concerning N-mixture models previously discussed in a classical framework. Specifically, the chapter focuses on prior specification, when no prior information is available, as well as on model selection. For prior specification, a novel objective prior that is proper is implemented and tested, and its performance is compared to approximations of the Jeffreys prior. Model selection of an extensive class of N-mixture models is performed using the Watanable-Akaike information criterion (WAIC) in a wide range of scenarios.

Chapter 3 presents a Bayesian hierarchical modelling framework for count data on species that exhibit temporary emigration (TE) at a site with temporally replicated sampling. This modelling framework accounts for observation error and models TE parametrically and non-parametrically to provide estimates of temporal population size. Temporal models and Dirichlet process mixture models are introduced to model TE parametrically and non-parametrically, respectively. Both of these approaches give rise to interesting ecological interpretations of TE. Additionally, using an efficient Bayesian variable selection algorithm, this modelling framework is further extended to identify important predictors of observation error.

Chapter 4 presents a Bayesian spatial model that simultaneously models disease dynamics and population dynamics using spatial capture-recapture data and imperfect diagnostic tests. Accounting for observation error in both detection and diagnostic tests, this framework enables a better understanding of how disease dynamics relate to population demographics in spatiotemporal contexts at an individual level. Specifically, disease transmission is modelled as a function of population density.

The supplementary material for each paper is presented in the appendix.

# Table of contents

# List of figures

# List of tables

# Chapter 1

# Introduction

With a growing number of wildlife species in decline (Almond et al., 2020; Thomas, 2013), coupled with global warming, there is an increasing need to better understand changes in populations, how environmental changes are affecting populations, identify species in need of protection, and develop or evaluate management practices, policies and guidelines. This is essential to ensure the long-term survival of many species and the maintenance of healthy ecosystems. Consequently, this thesis is a collection of three manuscripts on novel Bayesian methods and their applications to statistical ecology for different ecological data, providing novel and improved tools for better monitoring of wildlife populations.

We consider one of the most popular/important metrics used to understand wildlife populations: population size. Population size refers to the number of individuals of a species living in a given area or habitat at a given time. Populations can be considered open or closed. In a closed population, population size does not change during the study period, while in an open population, size can change due to the *BIDE* concept (Carroll and Conroy, 2011): births ($B$), deaths ($D$), immigration ($I$), and emigration ($E$). Consequently, in a closed population, individuals remain in the study area and there are no births or deaths during the study period, whilst in an open population individuals may be born, leave, and potentially return to, the study area, reproduce, and die during the study period. With closed populations, the primary aim is to estimate population size while in open populations the aim is to understand the dynamics of population size. Additionally, individual characteristics such as disease state, sex, age, etc. can play a significant role in population dynamics.

Changes in population size can occur in both time and space leading to temporal and spatial patterns. Spatial patterns refer to the distribution of individuals, populations, and communities in space whilst temporal patterns refer to the changes or variations in populations over time. Understanding spatial patterns can help explain complex processes such as individual space use and understanding temporal patterns can provide information on changes across time such as seasonality and species behavior.

We consider population density to investigate spatial patterns. Population density refers to the number of individuals in a population per unit of area or volume. It is a measure of how crowded or dispersed a population is within its environment and can play an important role in many ecological processes such as competition for resources, predation, disease transmission, and population growth. For temporal patterns, we consider temporary emigration (TE). Emigration is when an individual leaves the study area. This can be permanent or temporary. Thus, TE is when individuals can return to the study area after emigration. A better understanding of TE can inform ecological processes such as seasonal migration, temporary foraging, or breeding. If more than one site is sampled, the data can be used to estimate the spatial patterns of individuals and if sites are sampled across time, the data can be used to estimate temporal patterns.

The characteristics of the population, as well as time and financial constraints, determine the choice of the sampling method. Capture-recapture (CR) is one popular sampling method that we considered in this thesis. CR data involves sampling a location multiple times where a sample of individuals is captured and then releasing them back into the population with newly captured individuals marked in some way (e.g., tagging or banding). CR data provides an individual encounter history that displays the occasions where an individual was captured and not captured. This allows populations to be studied at an individual level and enables precise estimation of population size, capture probability, and survival. Additionally, CR data with the geographic locations where sampling was conducted is called spatial capture-recapture (SCR) data. Another commonly used sampling method, which we consider in two of the chapters, is count data. Count data entails taking repeated counts of unmarked individuals within a specific area. It is relatively less costly and time-consuming than CR and in some cases more practical. Generally, estimates obtained from count data are less precise than CR data but still allow accurate inference.

Generally, sampling methods, especially those considered in this thesis produce observation errors as the probability of observing, either capturing or counting, an individual that is present is typically much lower than 1. Failure to account for observation error can lead to biased or inaccurate estimates, with important implications for conservation and management efforts. Hence, statistical models developed in the literature (Nichols (1992); Royle (2004b)) and in this thesis account for observation errors. One way to account for observation error is by estimating the probability of capture or detection, depending on the model, for each individual and each sampling occasion. The detection probability can be a function of individual characteristics, such as sex, disease status, etc., and environmental covariates, such as temperature, precipitation, etc.

To perform inference, we use Bayesian hierarchical models (BHM). BHM are statistical models that are well suited for statistical ecology as they can be used to analyze complex data structures that have multiple levels of variation and dependence, such as longitudinal or spatial data, to provide robust statistical inference on the underlying mechanisms driving observed patterns in the data. BHM adopts the Bayesian approach to statistical inference which involves updating prior beliefs about model parameters based on observed data, using Bayes' theorem. The parameters of a BHM are modelled as random variables with prior distributions and corresponding hyperparameters that can themselves be modelled as random variables, resulting in a hierarchical structure. We employ Markov chain Monte Carlo (MCMC) methods to fit the models to the data. MCMC is a type of Monte Carlo method that generates random samples to estimate a quantity of interest. As a result, BHM allows for the modeling of relationships between multiple levels of processes and data, while also incorporating uncertainty and variability in both data and model parameters. In particular, they can account for observation errors by including measurement error terms in the model and estimating uncertainty in the parameters.

One of the main advantages of BHMs is that they allow prior knowledge to be incorporated at each level of the hierarchy, which can improve the estimation of the parameters and reduce uncertainty. However, care needs to be taken about how prior information is incorporated into the prior distribution(s). When prior knowledge is not available, as is in most cases in ecology, vague priors can be used to express the lack of knowledge. In more recent years, due to the increase in popularity and accessibility, objective priors can also be used to express a lack of prior knowledge. In

comparison to vague priors, objective priors are often chosen based on mathematical properties, such as invariance under reparameterization. As a result, objective priors may be more appropriate in this scenario.

Another key advantage of BHM is that model and variable selection can be performed within a single framework. Model selection refers to the process of choosing the best model from a set of candidate models, based on their ability to fit the data and make accurate predictions. BHM enables model selection by using Bayesian model comparison techniques such as the Bayes factor (Kass and Raftery, 1995) and Deviance Information Criterion (Spiegelhalter et al., 2002) that selects the most parsimonious model that best explains the data by balancing model complexity with goodness of fit. Similarly, BHM can perform variable selection which is the process of choosing the most important variables from a larger set of potential predictor variables by assigning prior probabilities to each possible set of variables and then calculating the posterior probabilities of each set of variables given the data (O'hara and Sillanpää). Overall, BHM provides a rigorous and probabilistic framework for performing model and variable selection. This helps to avoid overfitting and selection bias, allowing for more accurate and reliable statistical inference of complex datasets.

Chapters 2 and 3 focus on BHM for count data. In the second chapter, we focus on spatially replicated counts over time for closed populations whilst, in the third chapter, we focus on temporally replicated counts without spatial replication for open populations. The fourth chapter focuses on BHM for SCR data.

In the second chapter, we are concerned with fitting N-mixture models in a Bayesian framework. Accounting for observation error, N-mixture models are a popular class of hierarchical models as they provide an attractive framework to obtain estimates of population size using only count data. However, due to the reduced information in count data, a number of modelling issues have been found with N-mixture models in a classical inference setting. These issues include identifiability (Dennis et al., 2015) and model selection (MKe). This chapter aims to tackle the issue of model selection of N-mixture models in a Bayesian framework as well as the important choice of prior distributions when prior information is not available. It is important to evaluate statistical models to ensure that conclusions are accurate, reliable, suitable, and appropriately communicated. Chapter 2 has been written as a paper to be submitted to the journal Methods in Ecology and Evolution.

In the third chapter, we are interested in modelling TE from roost count data to estimate population size. Roost count data is a type of count data where individuals of a species are counted whilst exhibiting TE at their roost. Notably, not accounting for TE can result in positively biased estimates of population size (Chandler et al., 2011). Chandler et al. (2011) developed an N-mixture model that accounts for TE using count data. However, this model often assumes TE is constant throughout the study or is modelled independently at each sampling occasion. Consequently, in this chapter, we aim to extend this model by considering two classes of models for TE, a parametric and a non-parametric model that assume TE heterogeneity, enabling an intuitive ecological interpretation of TE whilst being parsimonious. Additionally, when accounting for observation error via a set of variables, the identification of important variables can be a valuable tool to understand species behaviour as well as to avoid overfitting. Overfitting can lead to high uncertainty around model parameters. Thus, to accomplish this, we consider Bayesian variable selection on observation error. Chapter 3 has been written as a paper to be submitted to the journal Annals of Applied Statistics.

In the fourth chapter, using SCR data and imperfect diagnostic tests, we are interested in modelling disease dynamics with a particular focus on the effect of population density on disease transmission. Disease dynamics is the study of how infectious diseases spread and evolve over time within a population. It is important to understand the influence of population density on disease transmission as this allows for the development of effective strategies for controlling and preventing the spread of infectious diseases. However, this is challenging as modelling population density can be difficult as the number of individuals and their spatial location are unobservable, whilst modelling disease dynamics can be difficult as the disease status of individuals is typically imperfectly observed. Thus, this chapter aims to tackle these challenges by proposing a novel spatially explicit capture-recapture (SCR) model that allows for the simultaneous modelling of population demographics and disease dynamics within a spatiotemporal context. SCR models are hierarchical models. They consider the collection of individuals in a population as a latent point process of where an individual moves, distributed within a region of interest. Accounting for the spatial nature of sampling together with the spatial distribution of individuals, SCR models are a powerful tool for estimating population size and population density. Chapter 4 has been written as a paper to be submitted to the journal Environmetrics.

**Chapter 2**

# Specifying and selecting N-mixture models in a Bayesian framework

# Abstract

Using only spatially replicated counts from unmarked individuals, N-mixture models provide an attractive framework to obtain estimates of population size by accounting for imperfect detection. The robustness of N-mixture models has been examined in detail in a classical inference framework. However, to our knowledge, only a small number of such studies have been carried out on N-mixture models in a Bayesian setting. In this paper, we consider fitting N-mixture models within a Bayesian framework. To aid implementation, we apply a new proper objective prior distribution to N-mixture models. Using simulated data, we compare this new proper objective prior to approximations of the popular objective prior, Jeffreys prior, and find that these prior distributions perform similarly in terms of model inference. Importantly, we find that when the detection probability is small, using priors that are concentrated at zero, even with large variance, expected population size can be considerably underestimated. Large estimates of expected population size were also found, evident by the bimodal density of posterior medians obtained for simulated data. Additionally, we consider an extensive class of N-mixture models and investigate model selection using the Watanable-Akaike Information Criterion (WAIC) in a wide range of scenarios to examine the sensitivity of WAIC to likelihood specification. We find that WAIC computed from the conditional likelihood produces misleading results favoring more complicated models than the true model. Contrary, WAIC computed using the marginal likelihood correctly selects the true model with a high probability. Hence, model selection of N-mixture models should be obtained from WAIC using the marginal likelihood, not the conditional likelihood. We demonstrate the usefulness/importance of employing these methods in two real datasets. Hence, this work can be considered a template for how to specify and select N-mixture models in a Bayesian context.

## 2.1   Introduction

A fundamental objective of many wildlife population monitoring programs and ecological studies is to estimate the size of a population. This is essential for the development and communication of management practices and guidelines. However, monitoring wildlife populations is challenging and costly, as the probability of detecting individuals in the monitored population is typically less than one. Survey sampling, which involves counting unmarked individuals in a given area over a specified period of time is relatively lower in cost and effort in comparison with other sampling methods, such as capture-recapture sampling and removal sampling.

Using count data from survey sampling, N-mixture models (Royle, 2004a) are a class of hierarchical models that accounts for imperfect detection, allowing estimation of population size in a cost-effective way. N-mixture models have been used for a number of purposes, including evaluation of conservation actions (Romano et al., 2017), understanding population size and population dynamics (Studds et al., 2017), population prediction to conservation scenarios (Ladin et al., 2016) and to forecast shifts in species distributions (Hunter et al., 2017).

The performance of N-mixture models in a classical setting has been investigated in detail. Dennis et al. (2015) showed that infinite estimates of population size can arise when the probability of detection and the number of times the population is sampled are small. Barker et al. (2018) demonstrated the inability of count data to discriminate between different hierarchical models, even when these models yield substantially different estimates of population size. Knape et al. (2018) highlighted that estimated population size can be severely i) underestimated if the fitted model does not account for over-dispersion in the population process, when that is present or ii) overestimated if the fitted model does not account for over-dispersion in the detection process, when that is present. However, to our knowledge, only a small number of studies have investigated N-mixture models in a Bayesian framework (see for example Link et al., 2018; Toribio et al., 2012, who studied the robustness of the N-mixture model in a Bayesian setting). Thus, we consider fitting an extensive class of N-mixture models in a Bayesian framework, specifically focusing on prior specification and model selection, which are key aspects of Bayesian modelling.

An important question in Bayesian model building is how does one choose a prior distribution $p(\theta)$ for parameter $\theta$? One can either be subjective: choosing priors that reflect some subjective opinion about $\theta$ (before data are collected) or objective: finding prior distributions that formally express ignorance about $\theta$. Subjective priors have the appeal of using prior information to increase estimation precision without compromising accuracy (Morris et al., 2015), resulting in larger effective sample sizes and saved resources. However, care needs to be taken about how prior information is incorporated into the prior distribution, especially where there is limited prior information as, in the case of sparse data, which is often true in ecological applications, the prior can have a strong effect on the posterior distribution. Additionally, it can be difficult to quantify prior effects in practice.

Contrarily, objective and vague priors are two classes of priors that allow Bayesian inference when information about $\theta$ is not available. These priors aim to avoid bias in parameter estimation by placing less emphasis on prior beliefs and more emphasis on the data. Based on mathematical properties, objective priors are designed to reflect minimal information, and have certain mathematical properties, discussed in this chapter. On the other hand, vague priors are deliberately chosen to convey no prior knowledge about the parameter being estimated, such as a flat prior or one with a very long tail, but without necessarily exhibiting the same mathematical properties as an objective prior. As a result, an objective prior may be more appropriate to express prior ignorance. Uniform distributions or normal distributions with large variances are common examples of vague priors. The Jeffreys prior (Jeffreys, 1946) is a popular objective prior designed to be invariant under reparameterization.

Notably, the majority of objective priors are improper (Leisen et al., 2018). A proper prior is a well-defined probability distribution as it integrates to 1 over the support of the parameter, whereas an improper prior has an infinite integral over the support of the parameter. In general, improper priors are not a problem as long as the resulting posterior is a proper probability distribution, from which one can derive moments such as the posterior mean. However, as of present, general results that allow one to assess if a given improper prior results in a proper posterior are yet to be developed (Leisen et al., 2018). Hence, caution is needed when using improper objective priors as spurious inference can be obtained. In addition, improper prior distributions cannot be applied in mixture models and model selection via Bayes factors (Leisen et al., 2018). Thus, this limits the use of many objective priors.

Banner et al. (2020) highlighted the use of priors in ecology and found vague priors are more often used in ecology than subjective priors. Both vague and objective priors have been used in N-mixture models: MKe demonstrated N-mixture models in a Bayesian framework by using priors that are approximations to Jeffreys prior, Link et al. (2018) used improper objective priors to study the robustness of N-mixture models, Toribio et al. (2012) used vague priors on the log and logit scale, which in turn resembled the Jeffreys prior on the original scale, to study the robustness of a Bayesian approach to fitting N-mixture models for pseudo-replicated count data. McCaffery et al. (2016) also used vague priors on the log and logit scale to analyze Lek count data.

In this paper, using a recently developed proper objective prior (Walker and Villa, 2021) and vague priors that are approximations to the Jeffreys prior, we test these priors and investigate the effect of prior choice in N-mixture models via an extensive simulation study.

N-mixture models can be relatively easily built in a Bayesian setting, but different models can result in substantially different estimates of population size (Ketwaroo, 2019). Therefore, it is imperative to have measures that allow one to compare models. Predictive accuracy measures can be used to compare models. Predictive accuracy measures simply compute how well a model estimated from available data generalises to out-of-sample data. However, the availability of out-of-sample data is often limited. One common way to overcome this deficiency is to use the sample data twice; once to fit the statistical model and again to test its predictive power. The issue here is that this can lead to over-fitting. Hence, predictive accuracy measures that use the data twice need to account for over-fitting. One such predictive accuracy measure is the Watanable-Akaike information criterion (WAIC, Watanabe and Opper, 2010). WAIC is often used in popular software such as NIMBLE (de Valpine et al., 2017) and Stan (Carpenter et al., 2017). Importantly, Ariyo et al. (2020) recently showed via an extensive simulation study that the marginal likelihood (averaging over latent variables) is superior to the conditional likelihood (given latent variables) when using WAIC to select the true longitudinal model. In addition, Millar (2018) showed using over-dispersed count data that WAIC computed using the conditional likelihood is an unreliable tool for model selection and recommended using WAIC computed using the marginal likelihood. Thus, in this paper, we investigate whether WAIC can be used to select among the different N-mixture models considered and whether WAIC for N-mixture models is sensitive to the likelihood specification in a wide range of scenarios.

We fit N-mixture models considered using Markov Chain Monte Carlo (MCMC) methods provided by the R package NIMBLE (de Valpine et al., 2017) version 0.10.0.

Finally, we consider two real data sets, yellow-bellied toads (Ketwaroo, 2019) and Swiss great tits (MKe), and we investigate the usefulness/importance of employing these methods in each case.

The paper is organised as follows: Section 2.2 provides a detailed description of the different N-mixture models considered, prior specification, and model selection. Simulation results are presented in Section 2.3 and the results for the two case studies are presented in Section 2.4. Section 2.5 concludes the paper and provides ideas for potential future directions.

## 2.2   Materials and Methods

### 2.2.1   N-mixture Models

Assuming population closure, N-mixture models estimate population size and account for imperfect detection using only replicated counts at multiple sites. N-mixture models are composed of two key processes: a population size process describing the spatial variation in the number of individuals among sites and a detection process describing the detection of individuals at each site (Royle, 2004b). Count data (hereafter $C_{ij}$) are obtained at $i = 1, \ldots, M$ sites with $j = 1, \ldots, J$ sampling occasions at each site.

For the population size process, it is assumed that the local population size at site $i$ (hereafter $N_i$) is an independent random variable with a chosen discrete probability function $g$. That is,

$$N_i \sim g(N; \lambda_i, \gamma)$$

where $\lambda_i$ represents the expected population size at site $i$ and $\gamma$ represents an optional parameter for over-dispersion in the population size process. In order to avoid over-parametrization, $\lambda_i$ may be

---

The work in this chapter is a continuation of the same author's MSc project, but an extension of it in a number of ways. Specifically, the N-mixture models considered were introduced in the MSc thesis, and the analysis of the yellow-bellied toads data set was first presented in the MSc thesis. However, the introduction of the new proper objective prior within the context of N-mixture models, the comparison between this prior and approximations to Jefferys prior using simulation and real data, as well as the model selection discussion and results using conditional and marginal WAIC correspond to new work presented in this chapter.

common to all sites, or it may be expressed as a function of site-specific covariates. In this paper, we consider the options introduced and considered in Ketwaroo (2019) for both the population size process and the detection process. Specifically, the Poisson and Negative binomial distributions for $g$ as well as the lesser-known Discrete Weibull distribution (Nakagawa and Osaki, 1975).

The Discrete Weibull (DW) distribution developed by Nakagawa and Osaki (1975) is the discrete form of the continuous Weibull distribution that is popular in survival analysis and failure time studies (Peluso et al., 2019). In this paper, we focus on the type 1 DW distribution, the most commonly used type in the literature. Let $Y$ be a random variable that follows a (type 1) DW distribution, the cumulative distribution is defined as:

$$F(y;q,b) = \begin{cases} 1 - q^{(y+1)^b} & \text{for } y = 0,1,2,3,\ldots, \\ 0 & \text{otherwise} \end{cases}$$

and the probability mass function is defined as:

$$f(y;q,b) = \begin{cases} q^{y^b} - q^{(y+1)^b} & \text{for } y = 0,1,2,3,\ldots, \\ 0 & \text{otherwise} \end{cases}$$

where $0 < q < 1$ and $b > 0$.

Importantly, Kalktawi (2017) highlighted the flexibility of the DW distribution to model count data; relative to the Poisson distribution, the DW distribution can be used to model over-, under-, and equi-dispersed data. Regarding the parameters $(q,b)$ of the DW distribution, Peluso et al. (2019) show that if:

1. $0 < b \leq 1$ there is over-dispersion, regardless of the value of $q$,

2. $b \geq 3$ there is under-dispersion, regardless of the value of $q$ and

3. $1 < b < 3$, depending on the value of $q$ there is under-dispersion or over-dispersion.

For the detection process, it is assumed

$$C_{ij} \sim h(C; N_i, p_{ij}, \rho) \tag{2.2.1}$$

where $h$ is a discrete probability distribution, $p_{ij}$ represents the probability of detecting an individual at site $i$ and sampling occasion $j$ and $\rho$ represents an optional parameter for over-dispersion in the detection process. We consider the Binomial and the Beta-Binomial (BB) distributions for $h$. The Binomial distribution is most commonly used to describe the detection process, assuming independence of detection. Using a Binomial detection process, $p_{ij}$ can be assumed to be constant across all sites and sampling occasions, or in a logistic regression framework, it can be expressed as a function of site and sampling occasion specific covariates.

Martin et al. (2011) showed that the BB distribution can serve as a detection process for modelling the correlating behaviour of individuals, thus relaxing the assumption of independent detection of individuals by the Binomial distribution. The BB detection process accomplishes this by modelling

$$p_{ij} \sim \text{Beta}(\alpha, \beta)$$

for $\alpha, \beta > 0$. Therefore, the BB detection process can also be used to model heterogeneity in detection probabilities (Ketwaroo, 2019; Martin et al., 2011). In addition, $\rho$ represents the degree to which individual behaviours or site attributes correlate with each other, which could affect detection (Martin et al., 2011), and is defined as

$$\rho = \frac{1}{\alpha + \beta + 1}$$

Notably, the BB distribution does not allow the distinction between correlations in individual behaviour and attributes of the site that could affect detection.

Assuming $N_i$ are independent random variables with discrete probability function $g(N_i; \lambda_i, \gamma)$, and $C_{ij}$ are conditionally dependent on $N_i$ with discrete probability function $h(N_i, p_{ij}, \rho)$, the marginal likelihood can be written as:

$$L(p_{ij}, \lambda_i, \rho, \gamma; C_{ij}) = \prod_{i=1}^{M} \left\{ \sum_{N_i=\max_j C_{ij}}^{\infty} \left( \prod_{j=1}^{J} h(C_{ij}; N_i, p_{ij}, \rho) \right) g(N_i; \lambda_i, \gamma) \right\}. \qquad (2.2.2)$$

This marginal likelihood takes into account all values for population size at each site and in reality, an upper bound can be chosen when fitting N-mixture models using the marginal likelihood. In a

Bayesian setting, $N_i$ can be treated as a latent variable that can be sampled via MCMC methods, hence avoiding the need for the infinite sum or truncation. The full conditional likelihood for N-mixture models can then be written as:

$$L(p_{ij}, \rho; N_i, C_{ij}) = \prod_{i=1}^{M} \left( \prod_{j=1}^{J} h(C_{ij}; N_i, p_{ij}, \rho) \right). \tag{2.2.3}$$

where there is no longer the need to marginalise over $N_i$, as in equation (2.2.2), and now $g$ from equation (2.2.2) serves as the prior distribution for $N_i$ in this conditional model. Bayesian inference using the marginal likelihood has the appeal of being similar to the maximum likelihood approach and in some cases, faster than sampling latent variables using the conditional likelihood (Ponisio et al., 2020). However, for N-mixture models, Ponisio et al. (2020) showed that in a Bayesian setting, marginalization is generally less computationally efficient than sampling $N_i$. This is possibly due to the computational cost of summing over the range of possible values of $N_i$ when the chosen upper bound is large.

Table 2.1 N-mixture models developed/implemented in Ketwaroo (2019) considered in this paper.

| N-mixture model | Model for population size process | Model for detection process |
| --- | --- | --- |
| P-B | Poisson ($\lambda$) | Binomial($N_i$, $p$) |
| DW-B | Discrete Weibull ($q$, $b$) | Binomial ($N_i$, $p$) |
| NB-B | Negative Binomial ($r$, $s$) | Binomial ($N_i$, $p$) |
| P-BB | Poisson ($\lambda$) | Beta - Binomial ($N_i$, $p_{ij}$, $\rho$) |
| DW-BB | Discrete Weibull ($q$, $b$) | Beta - Binomial ($N_i$, $p_{ij}$, $\rho$) |

Table 2.1 displays the list of N-mixture models investigated in this paper. We assume $\lambda_i$ to be constant for all sites for all models and for models with a Binomal detection process, we assume $p_{ij}$ to be constant across sites and sampling occasions. The P-B model is one of the most popular N-mixture models and it assumes equi-dispersion in the population size and detection processes. The NB-B model is also popular as it accounts for over-dispersion in the population size process relative to the Poisson distribution. The DW-B model offers more flexibility by accounting for over-, under-, and equi-dispersion in the population size process relative to the Poisson distribution. The P-BB model

accounts for over-dispersion in the detection process, and the DW-BB model has the advantage of accounting for over-dispersion in the detection process as well as under -, equi-, or under-dispersion in the population size process relative to the Poisson distribution.

### 2.2.2 Objective Prior Distributions

Jeffreys Prior (Jeffreys, 1946) - An obvious candidate for an objective prior is to use a flat prior $p(\theta) \propto c, c > 0$ such that $\int p(\theta)d\theta = \infty$. This flat prior is an improper prior and not transformation invariant. Instead, Jeffreys (1946) derived prior distributions that are transformation invariant. The Jeffreys prior is the most popular objective prior and can be defined as:

$$p(\theta) \propto \sqrt{|I(\theta)|}$$

where $I(\theta) = -E\left[\frac{\partial^2 log p(x|\theta)}{\partial\theta\partial\theta^T}|\theta\right]$ is the Fisher information where $p(x|\theta)$ denotes the likelihood. For a Poisson distribution with mean $\lambda$, the Fisher information $I(\lambda) = \frac{1}{\lambda}$, and so the Jeffreys prior is the *improper prior*, $p(\lambda) \propto \frac{1}{\lambda^2}$. This prior can be approximated by a Gamma$(\varepsilon_1, \varepsilon_2)$ where $\varepsilon_1, \varepsilon_2 \approx 0$ such as Gamma$(0.5, 0.00001)$ (Spiegelhalter et al., 2003). The Jeffreys prior yields sensible posterior distributions in scenarios where there is only one parameter of interest. However, it produces posteriors with poor performance when the parameter space has two or more dimensions (Leisen et al., 2018).

Walker and Villa (2021) recently developed a novel proper objective (OB) prior for continuous parameters by considering the connection between information, divergence and scoring rules. Let $\Theta = (0, \infty)$ be the parameter space of interest such that $\theta \in \Theta$. For some constant $a > 0$, the OB prior can be defined as

$$p(\theta) = \frac{a}{(a+\theta)^2}.$$

Setting $a = 1$ results in a heavy-tailed distribution as shown in Fig. 2.1. This distribution shape allows it to behave similarly to standard improper objective priors such as Jeffreys priors and reference priors (Berger et al., 2009), where a reference prior is an objective prior designed to maximize some measure of distance between the posterior and prior to allow the data to have maximum effect on the posterior. Measures such as the Kullback-Leibler divergence (Kullback and Leibler, 1951) or the Hellinger

distance (Beran, 1977) can be used to construct reference priors. Reference priors and Jeffreys priors are only equivalent for one-dimensional parameters.

Walker and Villa (2021) showed that this novel objective prior performed almost equivalently to the Jeffreys prior on simulated data. Unlike improper objective prior distributions, this novel objective prior distribution is proper, guaranteeing a proper posterior distribution.
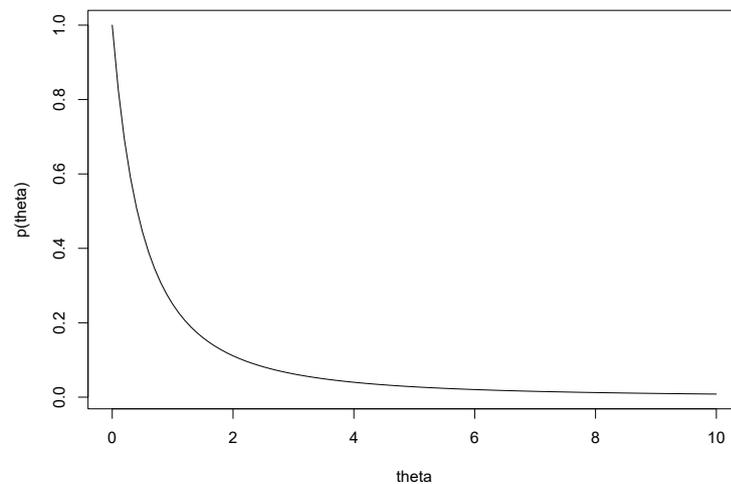


Fig. 2.1 The OB prior $p(\theta) = 1/(1+\theta)^2$ for a parameter defined in $(0, \infty)$.

### 2.2.3    Model Selection via WAIC

WAIC, also called the "widely available information" criterion, is a fully Bayesian predictive accuracy measure estimator based on the log posterior predictive distribution (Watanabe and Opper, 2010). To mathematically define WAIC, let $\theta$ represent all model parameters, $y_1, \ldots, y_n$ denote the sample data, $f$ be the true model, $\tilde{y}$ be the future data that could be observed, and $p_{post}(\tilde{y}) = \int p(\tilde{y}_i|\theta)p(\theta|y)d\theta$ be the posterior predictive distribution where $\tilde{y}_i$ denotes future data point $i$. Since the future $\tilde{y}_i$ is unknown, the expected log predicted density(elpd) can be used as a measure of predictive accuracy (Gelman et al., 2014):

$$\text{elpd} = E_f(\log p_{post}(\tilde{y}_i)) = \int \log p_{post}(\tilde{y}_i)f(\tilde{y}_i)d\tilde{y}_i$$

For the $n$ new data points, elpd is computed for each data point to establish the predictive accuracy measure of that data set:

$$\text{Expected log pointwise predicted density (elppd)} = \sum_{i=1}^{n} E_f(\log p_{post}(\tilde{y}_i))$$

However, the log posterior predictive density is unknown as the likelihood $p(\tilde{y}_i|\theta)$ cannot be computed. For this reason, the prediction accuracy of a fitted model can be summarised using the log pointwise predictive density(lppd):

$$\text{lppd} = \log \prod_{i=1}^{n} p_{post}(y_i) = \sum_{i=1}^{n} \log \int p(y_i|\theta)p(\theta|y)d\theta$$

In practice, draws from the posterior distribution can be used to evaluate lppd. Let $\theta^s$, for $s = 1, \ldots, S$ be the draws from the posterior distribution, then the computed lppd ($\widehat{\text{lppd}}$) can be defined as:

$$\widehat{\text{lppd}} = \sum_{i=1}^{n} \log \left( \frac{1}{S} \sum_{s=1}^{S} p(y_i|\theta^s) \right)$$

Accordingly, WAIC estimates the expected log pointwise predictive density elppd as the log pointwise predictive distribution lppd with a bias adjustment $\widehat{\text{elppd}}_{\text{WAIC}} = \widehat{\text{lppd}} - p_{\text{WAIC}}$. Two estimates of the bias adjustment have been proposed in the literature (Gelman et al., 2014). In this paper, we use the following bias adjustment:

$$p_{\text{WAIC}} = \sum_{i=1}^{n} \text{var}_{post}(\log p(y_i|\theta)), \tag{2.2.4}$$

which can be computed by:

$$\text{computed } p_{\text{WAIC}} = \sum_{i=1}^{n} V_{s=1}^{S}(\log p(y_i|\theta^s))$$

where $V_{s=1}^{S}$ represents the posterior sample variance. Thus, $p_{\text{WAIC}}$ can be easily computed by summing the posterior variance of the log predictive density over all data points $y_i$. See Gelman et al. (2014) for more information on the other bias adjustment.

Hence, WAIC can be generally expressed as

$$\text{WAIC} = -2(\text{lppd} - p_{\text{WAIC}}).$$ (2.2.5)

Specifically, conditional WAIC (cWAIC) and marginal WAIC (mWAIC) can be expressed as

$$\text{cWAIC} = -2(\text{lppd}_c - p_{\text{cWAIC}})$$ (2.2.6)

$$\text{mWAIC} = -2(\text{lppd}_m - p_{\text{mWAIC}})$$ (2.2.7)

where $\text{lppd}_c, p_{\text{cWAIC}}$ are computed using the conditional likelihood (equation (2.2.3)) and $\text{lppd}_m, p_{\text{mWAIC}}$ are computed using the marginal likelihood (equation (2.2.2)). Both cWAIC and mWAIC can be computed by using MCMC samples from the fitted conditional model, and this is the approach employed in this work.

Notably, WAIC (equation (2.2.5)) is on the deviance scale, making it comparable with other measures of deviance such as the Akaike information criterion (AIC), and the Deviance information criterion (DIC). The model with the lowest WAIC is considered the best model considering all models. In addition, as opposed to conditioning on a single point as is done in AIC and DIC, WAIC has the advantage of averaging over the entire posterior distribution, making it more appropriate for Bayesian models and particularly useful for complex models with many parameters. The notable weakness of WAIC is that its calculation depends on the independence assumption of data given the parameters, making it unclear how to compute for structured data settings such as time series, spatial, and network data.

As WAIC is an information criterion, we assess the strength of evidence for each model using delta WAIC and Akaike weights. Assuming there are $M$ candidate models, delta WAIC for the $m^{th}$ candidate model ($\Delta_m$) can be computed as $\Delta_m = \text{WAIC}_m - \text{WAIC}^*$ where $\text{WAIC}^*$ is the minimum WAIC among the $M$ candidate models.

Akaike weights, denoted by $\omega_m$, can be computed as:

$$\omega_m = \frac{\exp(-0.5\Delta_m)}{\sum_{i=1}^{M} \exp(-0.5\Delta_i)}.$$

That is, $\omega_m$, is the ratio of a candidate model's delta WAIC relative to the sum of the delta WAICs for all candidate models.

## 2.3    Simulation Study

We consider two extensive simulation cases: 1) to investigate model performance when using the OB prior and priors that are approximations to the Jeffreys prior for $\lambda$ in the P-B N-mixture model, and 2) to investigate whether WAIC is a reliable tool for model selection of N-mixture models and whether its performance depends on which likelihood calculation, conditional or marginal, is employed.

In both cases, we fit models using MCMC methods provided by R package NIMBLE (de Valpine et al., 2017) version 0.10.0 and use the full conditional N-mixture model (Equation (2.2.2)) as it was found to be more computationally efficient than the marginalized N-mixture model (Equation (2.2.3))(Ponisio et al., 2020). cWAIC and mWAIC are computed using MCMC samples from the fitted conditional model. To evaluate inference quality, we use the posterior median for each parameter since the conditional posterior distributions for $\lambda$ and $p$ were found to be skewed, and use $\hat{\lambda}$ and $\hat{p}$ to denote the median of the posterior medians over the simulation set. We also calculate 95% posterior credible interval coverage ($Cov_\theta$), residual mean square error $\left( RSME_\theta = \frac{\sqrt{\sum_{i=1}^{\text{nsim}}(\hat{\theta}_i - \theta)^2/\text{nsim}}}{\theta} \right)$, and median relative bias $\left( B_\theta = \text{median}\left( \frac{\hat{\theta} - \theta}{\theta} \right) \right)$, where $\theta$ is the true parameter value, $\hat{\theta}$ is the posterior median and nsim is the number of simulation runs.

### 2.3.1    Case 1 - Comparison Of Prior Distributions

For $\lambda$, we use the OB prior, and the following approximations to the Jeffreys priors: Gamma$(0.001, 0.001)$ and Gamma$(0.5, 0.00001)$, and for $p$ we use a Uniform$(0, 1)$ prior. We set $M = 20$, $J = 5$, and perform 100 simulation runs for each scenario: $\lambda = (5, 100, 500)$ and $p = (0.1, 0.25, 0.6)$. For $\lambda = (5, 100), p = (0.1, 0.25)$, we run 515000 MCMC iterations with burn-in of 15000 and thinning of 10 for 1 chain. For $\lambda = (5, 100), p = 0.6$, we run 115000 MCMC iterations with burn-in of 15000 and thinning of 10 for 1 chain. We run 815000 MCMC iterations with burn-in of 105000 and thinning of 20 for 1 chain for $\lambda = 500, p = (0.1, 0.25, 0.6)$. Different MCMC settings were chosen so that the effective sample size was similar between the different simulation scenarios.

Table 2.2 Simulation results using the OB prior.

| $\lambda$ | $p$ | $\hat{\lambda}$ | $Cov_\lambda$ | $RMSE_\lambda$ | $B_\lambda$ | $\hat{p}$ | $Cov_p$ | $RMSE_p$ | $B_p$ |
|---|---|---|---|---|---|---|---|---|---|
| 5 | 0.1 | 3.215 | 92 | 0.399 | −0.357 | 0.149 | 92 | 0.798 | 0.494 |
| 5 | 0.25 | 4.503 | 94 | 0.273 | −0.099 | 0.284 | 97 | 0.284 | 0.135 |
| 5 | 0.6 | 4.934 | 96 | 0.154 | −0.013 | 0.594 | 96 | 0.087 | −0.009 |
| 100 | 0.1 | 57.000 | 89 | 0.441 | −0.423 | 0.175 | 89 | 1.004 | 0.747 |
| 100 | 0.25 | 86.739 | 92 | 0.284 | −0.133 | 0.291 | 91 | 0.369 | 0.167 |
| 100 | 0.6 | 100.326 | 96 | 0.115 | 0.003 | 0.605 | 97 | 0.096 | 0.008 |
| 500 | 0.1 | 315.925 | 90 | 0.415 | −0.368 | 0.160 | 91 | 0.859 | 0.601 |
| 500 | 0.25 | 466.172 | 94 | 0.326 | −0.067 | 0.268 | 93 | 0.322 | 0.073 |
| 500 | 0.6 | 498.572 | 97 | 0.108 | −0.002 | 0.601 | 96 | 0.098 | 0.002 |

Table 2.3 Simulation results using the Gamma$(0.001, 0.001)$ prior.

| $\lambda$ | $p$ | $\hat{\lambda}$ | $Cov_\lambda$ | $RMSE_\lambda$ | $B_\lambda$ | $\hat{p}$ | $Cov_p$ | $RMSE_p$ | $B_p$ |
|---|---|---|---|---|---|---|---|---|---|
| 5 | 0.1 | 3.862 | 95 | 0.468 | −0.227 | 0.124 | 96 | 0.706 | 0.237 |
| 5 | 0.25 | 4.857 | 95 | 0.407 | −0.028 | 0.268 | 97 | 0.297 | 0.072 |
| 5 | 0.6 | 5.001 | 96 | 0.169 | 0.000 | 0.591 | 96 | 0.092 | −0.016 |
| 100 | 0.1 | 69.914 | 99 | 0.407 | −0.300 | 0.138 | 98 | 0.778 | 0.380 |
| 100 | 0.25 | 92.502 | 92 | 0.358 | −0.075 | 0.276 | 91 | 0.349 | 0.103 |
| 100 | 0.6 | 101.664 | 96 | 0.122 | 0.016 | 0.598 | 96 | 0.101 | −0.003 |
| 500 | 0.1 | 348.831 | 96 | 0.367 | −0.302 | 0.143 | 95 | 0.727 | 0.426 |
| 500 | 0.25 | 472.884 | 94 | 0.293 | −0.054 | 0.263 | 94 | 0.297 | 0.052 |
| 500 | 0.6 | 501.903 | 96 | 0.109 | 0.004 | 0.598 | 95 | 0.099 | −0.003 |

Table 2.4 Simulation results using the Gamma$(0.5, 0.00001)$ prior.

| $\lambda$ | $p$ | $\hat{\lambda}$ | $Cov_\lambda$ | $RMSE_\lambda$ | $B_\lambda$ | $\hat{p}$ | $Cov_p$ | $RMSE_p$ | $B_p$ |
|---|---|---|---|---|---|---|---|---|---|
| 5 | 0.1 | 5.187 | 95 | 1.055 | 0.037 | 0.092 | 98 | 0.669 | $-0.084$ |
| 5 | 0.25 | 5.175 | 96 | 0.689 | 0.035 | 0.249 | 95 | 0.334 | $-0.003$ |
| 5 | 0.6 | 5.045 | 96 | 0.181 | 0.009 | 0.589 | 96 | 0.095 | $-0.017$ |
| 100 | 0.1 | 89.843 | 99 | 0.702 | $-0.101$ | 0.115 | 99 | 0.639 | 0.145 |
| 100 | 0.25 | 102.855 | 94 | 0.549 | 0.028 | 0.249 | 95 | 0.365 | $-0.002$ |
| 100 | 0.6 | 102.726 | 96 | 0.152 | 0.027 | 0.592 | 96 | 0.111 | $-0.013$ |
| 500 | 0.1 | 503.416 | 97 | 1.151 | 0.007 | 0.099 | 97 | 0.558 | $-0.007$ |
| 500 | 0.25 | 567.462 | 92 | 0.972 | 0.135 | 0.221 | 90 | 0.366 | $-0.113$ |
| 500 | 0.6 | 507.927 | 95 | 0.124 | 0.016 | 0.591 | 95 | 0.107 | $-0.015$ |

From Tables 2.2, 2.3, and 2.4 it can be seen that the OB prior and Gamma priors perform similarly in terms of inference at high and low levels of detection probability. Notably, when $p$ is small and priors for $\lambda$ are concentrated at zero, as is the case for all prior distributions considered here, $\lambda$ can be severely underestimated, as can also be seen in Fig. 2.2, which displays the density plots of the posterior medians of $\lambda$ from the 100 runs for the OB prior. In addition, looking at Fig. 2.2, we can see that large estimates of $\lambda$ are also obtained when $p$ is low, evident in the tails/ bi-modal density of the distribution of posterior medians. This corroborates the results found by Dennis et al. (2015) in a classical setting, who found that the maximum likelihood estimates of population size can tend to infinity when detection probability is small. Additionally, looking at Fig.2.2, it can also be seen that there are cases when $\lambda$ is estimated well. Hence, the results demonstrate that the distribution of posterior medians obtained for $\lambda$ has two or maybe even three modes, for the first time demonstrating the substantial risk of underestimating $\lambda$ when detection probability is small.

(a) $p = 0.1$          (b) $p = 0.25$          (c) $p = 0.6$
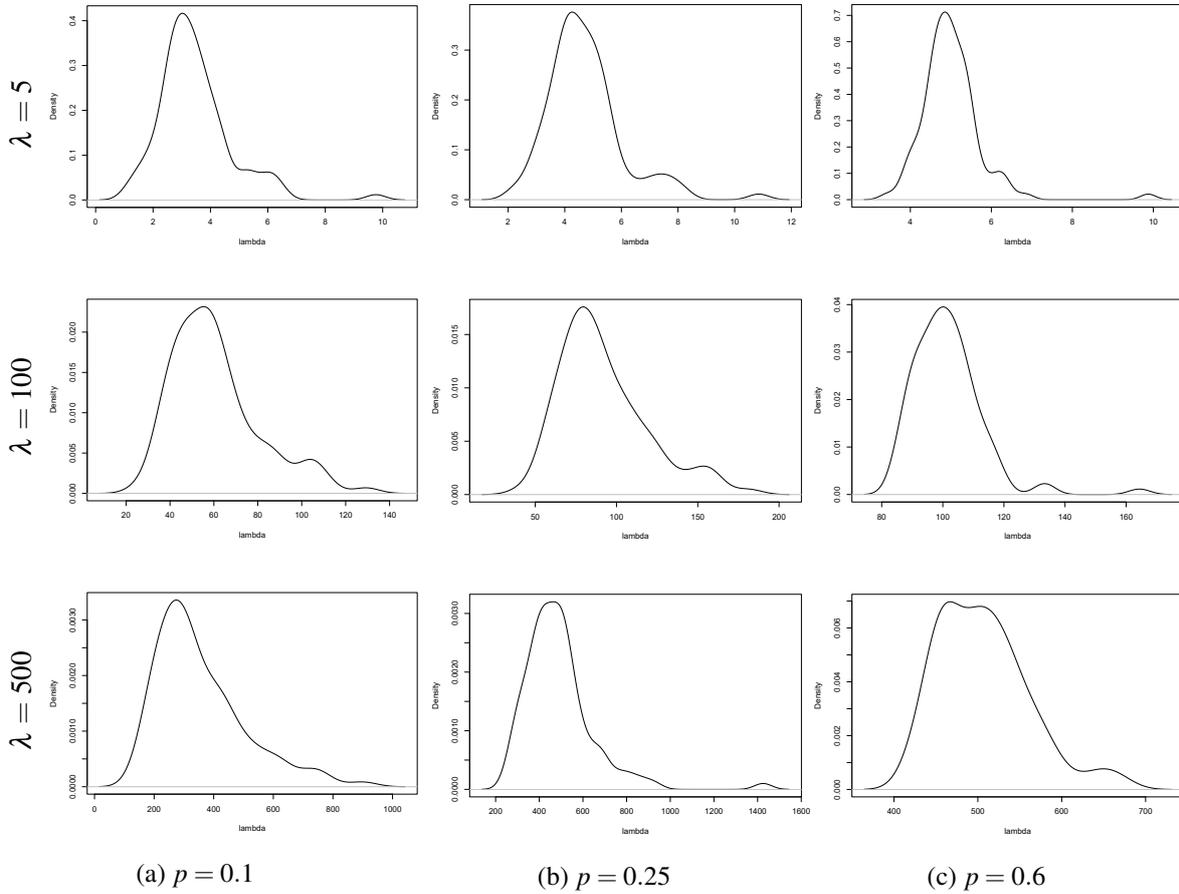
Fig. 2.2 Density plots of the posterior medians of $\lambda$ obtained using the OB prior.

### 2.3.2   Case 2 - Model Selection via WAIC

We consider four simulation scenarios:

- Scenario 1: Over-dispersion in the population size process; the true model is the Discrete Weibull Binomial (DW-B) N-mixture model.

- Scenario 2: Over-dispersion in the detection process; the true model is the Poisson Beta-Binomial (P-BB) N-mixture model.

- Scenario 3: Equi-dispersion in the population size process; the true model is the Poisson Binomial (P-B) N-mixture model.

- Scenario 4: Under-dispersion in the population size process; the true model is the DW-B
  N-mixture model.

In each scenario, 100 data sets were simulated from the true model, and the class of N-mixture models considered in this paper (Table 2.1) are fitted to each data set. Setting $M = 50$, $J = 5$, data generating model parameters for each case are shown in Table 2.5. For cases $1, 3$ and $4$ data were simulated using $p = (0.25, 0.6)$ to investigate model selection when $p$ is high and low. Similarly, for case 2, the Beta-Binomial parameters $\alpha = (3, 1)$ and $\beta = (2, 3)$ were chosen such that the mean detection probability is $0.25, 0.6$ respectively. For scenarios 1 and 4, data were generated with an expected population size of 4.325 and 9.564 respectively. Parameters in the parameter space $(0, \infty)$ were assigned the OB prior, and parameters in the parameter space $(0, 1)$ were assigned a Uniform$(0, 1)$ prior. MCMC settings for each scenario are given in A.1 of the Supplementary material. In each scenario, we compute the cWAIC and mWAIC for each N-mixture model, report the proportion of times each model was selected %WAIC, median $\Delta$ WAIC and median WAIC weights ($\omega_{\text{WAIC}}$) for both cWAIC and mWAIC. We use expected population size ($\lambda$) and $p$ to compare inference quality between models. For simplicity, we let $p$ represent the mean detection probability for BB models. We define 'Best by mWAIC' and 'Best by cWAIC' to be inferences of models selected by mWAIC and cWAIC respectively.

Table 2.5 Data generating model parameters for each model.

| Scenario | Model | Parameters |
|---------:|:-----:|:----------:|
| 1 | DW-B | $q = 0.75, b = 0.95$ |
| 2 | P-BB | $\lambda = 5$ |
| 3 | P-B | $\lambda = 5$ |
| 4 | DW-B | $q = 0.9999, b = 4$ |

**Scenario 1- Over-dispersion in the population size process**

As can be seen from Tables 2.6 and 2.7, when there was over-dispersion in the population size process, cWAIC strongly favoured the more complicated model, the P-BB model, which gave poor inference, instead of the true model. On the other hand, mWAIC selected the correct model with

higher probability and better inference. The ability of mWAIC to select the true model was reduced with low $p$, but it selected a similar model that accommodates overdispersion in the population size process and produced good inference. Model inference results (Table 2.7) also agree with the findings of Knape et al. (2018), that is, models that do not accommodate overdispersion in the population size process, when overdispersion is present, underestimate expected population size.

Table 2.6 Scenario 1 model selection results when the true model is the DW-B N-mixture model.

| $p$ | Model | %cWAIC | %mWAIC | ΔcWAIC | Δ mWAIC | $\omega_{cWAIC}$ | $\omega_{mWAIC}$ |
|---|---|---|---|---|---|---|---|
| 0.6 | P-B | 0 | 0 | 25.133 | 203.843 | 0 | 0 |
| | DW-B | 0 | 87 | 31.056 | 0 | 0 | 0.549 |
| | NB-B | 0 | 8 | 30.497 | 0.419 | 0 | 0.419 |
| | P-BB | 98 | 0 | 0 | 203.388 | 1 | 0 |
| | DW-BB | 2 | 5 | 162.001 | 135.517 | 0 | 0 |
| 0.25 | P-B | 0 | 0 | 35.974 | 26.222 | 0 | 0 |
| | DW-B | 0 | 29 | 41.345 | 0.205 | 0 | 0.421 |
| | NB-B | 0 | 64 | 40.645 | 0 | 0 | 0.453 |
| | P-BB | 100 | 1 | 0 | 16.553 | 0.999 | 0 |
| | DW-BB | 0 | 6 | 16.938 | 2.909 | 0 | 0.102 |

Table 2.7 Scenario 1 model inference results when the true model is the DW-B N-mixture model.

| $p$ | Model | $\hat{\lambda}$ | $Cov_\lambda$ | $RMSE_\lambda$ | $B_\lambda$ | $\hat{p}$ | $Cov_p$ | $RMSE_p$ | $B_p$ |
|-----|-------|------|--------|---------|---------|------|--------|--------|--------|
| 0.6 | P-B | 3.005 | 7 | 0.334 | $-0.305$ | 0.668 | 36 | 0.126 | 0.114 |
| | DW-B | 4.450 | 93 | 0.144 | 0.029 | 0.600 | 95 | 0.071 | 0.0003 |
| | NB-B | 3.377 | 68 | 0.276 | $-0.220$ | 0.603 | 95 | 0.070 | 0.006 |
| | P-BB | 2.907 | 5 | 0.355 | $-0.328$ | 0.683 | 18 | 0.145 | 0.138 |
| | DW-BB | 4.963 | 62 | 0.217 | 0.147 | 0.633 | 84 | 0.075 | 0.055 |
| | Best by mWAIC | 4.259 | 93 | 0.146 | $-0.015$ | 0.603 | 95 | 0.070 | 0.005 |
| | Best by cWAIC | 2.908 | 5 | 0.355 | $-0.327$ | 0.681 | 19 | 0.144 | 0.136 |
| 0.25 | P-B | 2.005 | 1 | 0.537 | $-0.537$ | 0.416 | 8 | 0.679 | 0.663 |
| | DW-B | 4.356 | 90 | 0.312 | 0.007 | 0.254 | 96 | 0.278 | 0.016 |
| | NB-B | 3.159 | 81 | 0.353 | $-0.269$ | 0.267 | 95 | 0.275 | 0.067 |
| | P-BB | 1.828 | 0 | 0.579 | $-0.577$ | 0.440 | 2 | 0.777 | 0.762 |
| | DW-BB | 3.335 | 76 | 0.256 | $-0.228$ | 0.358 | 66 | 0.490 | 0.433 |
| | Best by mWAIC | 3.467 | 78 | 0.424 | $-0.198$ | 0.254 | 96 | 0.289 | 0.055 |
| | Best by cWAIC | 1.828 | 0 | 0.579 | $-0.577$ | 0.440 | 2 | 0.777 | 0.762 |

**Scenario 2 - Over-dispersion in the detection process**

Looking at Tables 2.8 and 2.9, it can be seen that the cWAIC again strongly favoured the more complicated model, ie the DW-BB model, whilst mWAIC selected the correct model at least 3 times more and produced better inference than models selected by cWAIC. In addition, models that did not accommodate over-dispersion in the detection process over-estimated expected population size, agreeing with Knape et al. (2018).

Table 2.8 Scenario 2 model selection results when the true model is the P-BB N-mixture model.

| $p$ | Model | %cWAIC | %mWAIC | $\Delta$cWAIC | $\Delta$ mWAIC | $\omega_{\text{cWAIC}}$ | $\omega_{\text{mWAIC}}$ |
|------|-------|--------|--------|---------|---------|---------|---------|
| 0.6 | P-B | 0 | 1 | 134.983 | 18.899 | 0 | 0 |
|  | DW-B | 0 | 0 | 132.142 | 12.376 | 0 | 0.002 |
|  | NB-B | 0 | 3 | 135.862 | 11.597 | 0 | 0.002 |
|  | P-BB | 26 | 83 | 2.742 | 0 | 0.202 | 0.643 |
|  | DW-BB | 74 | 13 | 0 | 1.600 | 0.798 | 0.291 |
| 0.25 | P-B | 0 | 0 | 154.726 | 38.776 | 0 | 0 |
|  | DW-B | 0 | 0 | 141.092 | 25.334 | 0 | 0 |
|  | NB-B | 0 | 0 | 139.177 | 20.955 | 0 | 0 |
|  | P-BB | 23 | 61 | 8.565 | 0 | 0.014 | 0.644 |
|  | DW-BB | 77 | 39 | 0 | 1.185 | 0.986 | 0.357 |

Table 2.9 Scenario 2 model inference results when the true model is the P-BB N-mixture model.

| $p$ | Model | $\hat{\lambda}$ | $\text{Cov}_\lambda$ | $\text{RMSE}_\lambda$ | $B_\lambda$ | $\hat{p}$ | $\text{Cov}_p$ | $\text{RMSE}_p$ | $B_p$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.6 | P-B | 7.644 | 6 | 0.655 | 0.528 | 0.388 | 1 | 0.367 | $-0.353$ |
| | DW-B | 9.603 | 0 | 1.044 | 0.921 | 0.337 | 0 | 0.436 | $-0.437$ |
| | NB-B | 8.687 | 2 | 0.930 | 0.737 | 0.342 | 0 | 0.440 | $-0.429$ |
| | P-BB | 4.953 | 96 | 0.086 | $-0.009$ | 0.591 | 96 | 0.053 | -0.015 |
| | DW-BB | 5.852 | 25 | 0.208 | 0.170 | 0.593 | 96 | 0.051 | $-0.012$ |
| | Best by mWAIC | 5.182 | 84 | 0.129 | 0.036 | 0.591 | 92 | 0.077 | $-0.015$ |
| | Best by cWAIC | 5.736 | 41 | 0.188 | 0.147 | 0.593 | 96 | 0.050 | $-0.0108$ |
| 0.25 | P-B | 11.280 | 1 | 1.798 | 1.608 | 0.111 | 7 | 0.568 | $-0.556$ |
| | DW-B | 26.937 | 0 | 5.226 | 4.387 | 0.046 | 0 | 0.812 | $-0.814$ |
| | NB-B | 38.048 | 0 | 7.619 | 6.609 | 0.031 | 0 | 0.864 | $-0.874$ |
| | P-BB | 4.774 | 96 | 0.283 | $-0.045$ | 0.263 | 96 | 0.228 | 0.051 |
| | DW-BB | 5.563 | 86 | 0.220 | 0.113 | 0.273 | 96 | 0.206 | 0.093 |
| | Best by mWAIC | 4.251 | 96 | 0.232 | $-0.149$ | 0.267 | 97 | 0.198 | 0.071 |
| | Best by cWAIC | 4.219 | 94 | 0.244 | $-0.156$ | 0.277 | 96 | 0.213 | 0.109 |

**Scenario 3: Equi-dispersion in the population size process**

Tables 2.10 and 2.11 show cWAIC strongly favouring the more complicated model, the DW-BB model which, compared to the true model, gave poorer inference while mWAIC selected the true model more often, in favour of models that fit the data best. In addition, models selected by mWAIC produced better inference than models selected by cWAIC.

Table 2.10 Scenario 3 model selection results when the true model is the P-B N-mixture model.

| $p$ | Model | %cWAIC | %mWAIC | $\Delta$cWAIC | $\Delta$ mWAIC | $\omega_{\text{cWAIC}}$ | $\omega_{\text{mWAIC}}$ |
|------|-------|--------|--------|---------|---------|---------|---------|
| 0.6 | P-B | 0 | 63 | 28.854 | 0 | 0 | 0.360 |
| | DW-B | 0 | 14 | 25.960 | 1.514 | 0 | 0.179 |
| | NB-B | 0 | 17 | 29.521 | 0.360 | 0 | 0.320 |
| | P-BB | 12 | 4 | 2.715 | 3.722 | 0.205 | 0.060 |
| | DW-BB | 88 | 2 | 0 | 4.559 | 0.795 | 0.041 |
| 0.25 | P-B | 0 | 70 | 35.211 | 0 | 0 | 0.463 |
| | DW-B | 0 | 9 | 31.197 | 2.201 | 0 | 0.171 |
| | NB-B | 0 | 6 | 39.293 | 2.513 | 0 | 0.144 |
| | P-BB | 2 | 5 | 13.148 | 4.392 | 0.001 | 0.059 |
| | DW-BB | 98 | 10 | 0 | 4.034 | 0.998 | 0.006 |

Table 2.11 Scenario 3 model inference results when the true model is the P-B N-mixture model.

| $p$ | Model | $\hat{\lambda}$ | $\text{Cov}_\lambda$ | $\text{RMSE}_\lambda$ | $B_\lambda$ | $\hat{p}$ | $\text{Cov}_p$ | $\text{RMSE}_p$ | $B_p$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.6 | P-B | 5.036 | 94 | 0.095 | 0.007 | 0.593 | 93 | 0.062 | $-0.012$ |
| | DW-B | 5.944 | 33 | 0.212 | 0.188 | 0.599 | 93 | 0.059 | 0.000 |
| | NB-B | 5.050 | 94 | 0.097 | 0.010 | 0.590 | 92 | 0.064 | $-0.016$ |
| | P-BB | 4.635 | 90 | 0.098 | $-0.073$ | 0.640 | 82 | 0.082 | 0.067 |
| | DW-BB | 5.525 | 61 | 0.133 | 0.104 | 0.647 | 72 | 0.090 | 0.080 |
| | Best by mWAIC | 5.168 | 84 | 0.119 | 0.034 | 0.598 | 91 | 0.065 | $-0.002$ |
| | Best by cWAIC | 5.422 | 65 | 0.126 | 0.084 | 0.645 | 70 | 0.090 | 0.080 |
| 0.25 | P-B | 4.786 | 96 | 0.197 | $-0.043$ | 0.259 | 94 | 0.214 | 0.036 |
| | DW-B | 5.317 | 94 | 0.195 | 0.063 | 0.282 | 92 | 0.260 | 0.128 |
| | NB-B | 6.056 | 96 | 0.476 | 0.211 | 0.209 | 94 | 0.295 | $-0.165$ |
| | P-BB | 4.774 | 96 | 0.283 | $-0.045$ | 0.262 | 96 | 0.228 | 0.051 |
| | DW-BB | 4.145 | 74 | 0.189 | $-0.171$ | 0.273 | 96 | 0.206 | 0.093 |
| | Best by mWAIC | 4.845 | 92 | 0.177 | $-0.031$ | 0.260 | 93 | 0.212 | 0.039 |
| | Best by cWAIC | 4.185 | 75 | 0.193 | $-0.165$ | 0.263 | 95 | 0.202 | 0.090 |

**Scenario 4: Under-dispersion in the population size process**

As shown in Tables 2.12 and 2.13, cWAIC once again favoured the more complicated model instead of the true whereas mWAIC had a stronger preference for the true model and a better preference for models with good inference than models selected by cWAIC.

Table 2.12 Scenario 4 model selection results when the true model is the DW-B N-mixture model.

| $p$ | Model | %cWAIC | %mWAIC | $\Delta$cWAIC | $\Delta$ mWAIC | $\omega_{\text{cWAIC}}$ | $\omega_{\text{mWAIC}}$ |
|-----|-------|--------|--------|-------|--------|-------|-------|
| 0.6 | P-B | 0 | 79 | 36.733 | 0 | 0 | 0.581 |
|     | DW-B | 0 | 16 | 33.225 | 2.103 | 0 | 0.216 |
|     | NB-B | 0 | 0 | 38.395 | 4.846 | 0 | 0.054 |
|     | P-BB | 6 | 2 | 8.202 | 4.560 | 0.016 | 0.062 |
|     | DW-BB | 94 | 3 | 0 | 5.111 | 0.984 | 0.045 |
| 0.25 | P-B | 0 | 84 | 42.759 | 0 | 0 | 0.618 |
|     | DW-B | 0 | 5 | 41.055 | 3.900 | 0 | 0.083 |
|     | NB-B | 0 | 5 | 45.617 | 2.958 | 0 | 0.155 |
|     | P-BB | 0 | 1 | 25.645 | 8.039 | 0 | 0.011 |
|     | DW-BB | 100 | 5 | 0 | 5.174 | 1 | 0.047 |

Table 2.13 Scenario 4 model inference results when the true model is the DW-B N-mixture model.

| $p$ | Model | $\hat{\lambda}$ | $\text{Cov}_\lambda$ | $\text{RMSE}_\lambda$ | $B_\lambda$ | $\hat{p}$ | $\text{Cov}_p$ | $\text{RMSE}_p$ | $B_p$ |
|---|---|---|---|---|---|---|---|---|---|
| 0.6 | P-B | 8.807 | 77 | 0.112 | −0.079 | 0.584 | 96 | 0.064 | −0.025 |
| | DW-B | 9.409 | 94 | 0.068 | −0.016 | 0.600 | 96 | 0.058 | 0.001 |
| | NB-B | 8.979 | 86 | 0.102 | −0.061 | 0.578 | 92 | 0.073 | −0.034 |
| | P-BB | 7.795 | 31 | 0.185 | −0.185 | 0.657 | 77 | 0.103 | 0.095 |
| | DW-BB | 8.450 | 51 | 0.118 | −0.116 | 0.673 | 57 | 0.125 | 0.122 |
| | Best by mWAIC | 8.818 | 78 | 0.111 | −0.078 | 0.594 | 92 | 0.074 | −0.009 |
| | Best by cWAIC | 8.369 | 47 | 0.124 | −0.125 | 0.710 | 6 | 0.182 | 0.184 |
| 0.25 | P-B | 9.084 | 94 | 0.271 | −0.050 | 0.238 | 93 | 0.238 | −0.047 |
| | DW-B | 5.562 | 54 | 0.415 | −0.418 | 0.305 | 86 | 0.311 | 0.219 |
| | NB-B | 12.116 | 91 | 0.635 | 0.266 | 0.179 | 85 | 0.356 | −0.285 |
| | P-BB | 6.149 | 62 | 0.363 | −0.357 | 0.349 | 75 | 0.477 | 0.397 |
| | DW-BB | 5.747 | 7 | 0.401 | −0.399 | 0.445 | 7 | 0.814 | 0.781 |
| | Best by mWAIC | 8.630 | 86 | 0.324 | −0.097 | 0.239 | 84 | 0.373 | −0.040 |
| | Best by cWAIC | 5.747 | 7 | 0.401 | −0.399 | 0.445 | 7 | 0.814 | 0.781 |

From this extensive simulation study, it can be seen that mWAIC selected the correct model with a high probabiliy while cWAIC favoured the more complicated model that often gave poor inference. Hence, model selection via WAIC for N-mixture models should be performed using the marginal likelihood as cWAIC can favour unnecessarily complicated models. Importantly, these scenarios demonstrate that one can select between different N-mixture models with different model inferences using mWAIC.

## 2.4 Case Studies

We consider two case studies: yellow-bellied toads and Swiss great tits. We apply all N-mixture models defined in Table 2.1 to both data, assuming the expected population size to be constant across sites and detection probability for Binomial models to be constant across sites and sampling occasions.

We fit models using the conditional likelihood (Equation (2.2.3)) and using MCMC samples from the fitted model, we perform model selection using both cWAIC and mWAIC. We choose the OB prior for continuous parameters with parameter space $(0, \infty)$, whereas parameters with parameter space $(0, 1)$ are assigned a Uniform$(0, 1)$ prior. Additionally, for the yellow-bellied toad, we investigate the prior sensitivity of parameters with parameter space $(0, \infty)$ by using an approximation to the Jeffreys prior, Gamma$(0.001, 0.001)$. To assess model fit, we use posterior predictive goodness of fit: we define $\tau_i = \sum_{j=1}^{J} C_{i,j}$ and using MCMC samples, we simulate counts, and hence $\tau_i$, from our model and compare these to the observed data. A model fits the data well if it produces similar $\tau_i$ values to the observed data. MCMC settings used for both case studies are presented in A.2 of the Supplementary material. We assess convergence using Gelman and Rubin's convergence diagnostic (Gelman and Rubin, 1992).

### 2.4.1  Yellow-bellied Toads

In 2018, survey sampling of five populations of yellow-bellied toads (*Bombina variegata*) was conducted at 27 sites from the end of May to the beginning of July. Each site was sampled 4 times during the period of study. Sites were represented by ponds or tanks located in a variety of habitats, mainly vineyards, and meadows, in the Italian Alps.

With convergence achieved for all model parameters, it can be seen from Table 2.14 that all models considered produce different estimates of expected population size and detection probability, highlighting the need to select the correct model to avoid erroneous inference. Notably, both cWAIC and mWAIC were in agreement strongly favoring the DW-BB model with cWAIC and mWAIC weights of 1.0 and the least support to the P-B model with cWAIC and mWAIC weights of 0. Additionally, the OB prior and the Jeffreys prior approximation (Gamma$(0.001, 0.001)$) give similar model inference and WAIC values. Posterior predictive goodness of fit indicated all models except the P-BB fitted the data well. Fig. 2.3 displays the GOF plot for the DW-BB model and it can be seen that the true value is captured between the 5th and 95th quantile for all sites for the DW-BB model. The P-BB model lack of fit is evident in the large estimates of expected population size.

Table 2.14 Model results from analysing yellow-bellied toads data. Values within the brackets represent the 95% posterior credible interval. For BB models, detection probability represents mean detection probability.

| Prior | Model | cWAIC | mWAIC | Detection Probability | Expected population size | $p$ correlation. |
|-------|-------|-------|-------|----------------------|--------------------------|------------------|
| OB | | | | | | |
| | P-B | 542.328 | 905.735 | $0.637(0.585, 0.686)$ | $10.759(9.375, 12.218)$ | - |
| | DW-B | 461.211 | 528.078 | $0.186(0.051, 0.360)$ | $38.825(16.730, 145.290)$ | - |
| | NB-B | 461.535 | 529.078 | $0.201(0.054, 0.371)$ | $34.574(15.847, 130.823)$ | - |
| | P-BB | 491.621 | 654.837 | $0.038(0.010, 0.085)$ | $238.333(83.459, 653.727)$ | $0.049(0.013, 0.109)$ |
| | DW-BB | 338.121 | 502.857 | $0.662(0.555, 0.743)$ | $11.111(7.360, 18.410)$ | $0.182(0.085, 0.289)$ |
| | | | | | | |
| Gamma(0.001,0.001) | | | | | | |
| | P-B | 544.649 | 907.358 | $0.635(0.579, 0.688)$ | $10.818(9.401, 12.341)$ | - |
| | DW-B | 461.215 | 528.126 | $0.193(0.054, 0.365)$ | $37.553(16.39, 137.21)$ | - |
| | NB-B | 461.377 | 529.158 | $0.197(0.053, 0.366)$ | $45.425(15.977, 132.666)$ | - |
| | P-BB | 494.398 | 654.739 | $0.018(0.003, 0.057)$ | $672.845(122.529, 1951.403)$ | $0.023(0.004, 0.073)$ |
| | DW-BB | 344.243 | 507.252 | $0.649(0.511, 0.738)$ | $11.328(7.460, 19.100)$ | $0.163(0.061, 0.280)$ |
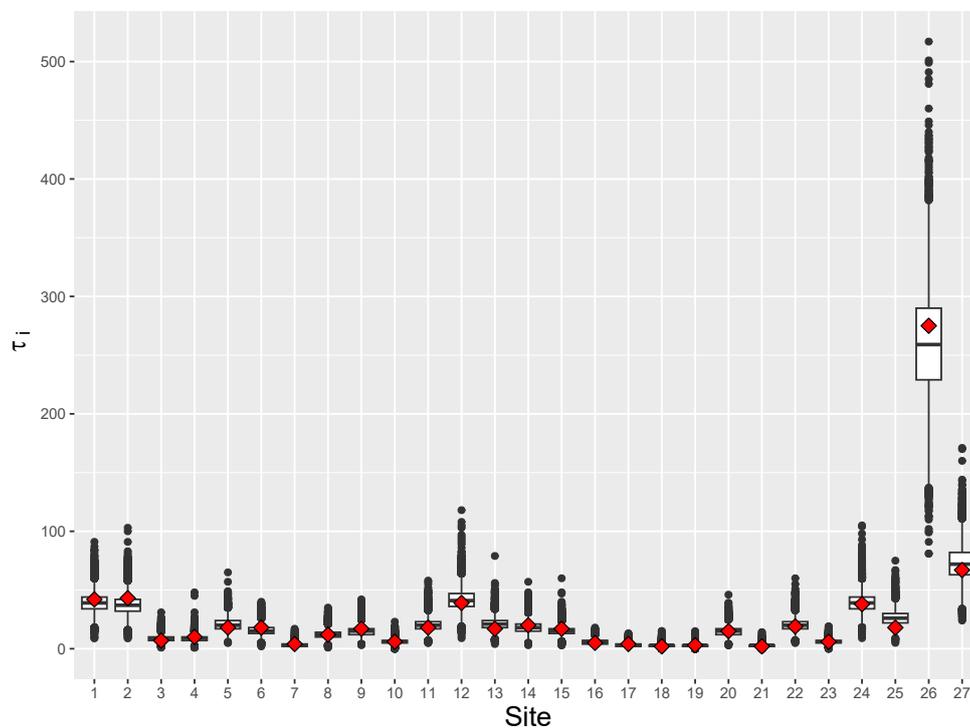


Fig. 2.3 GOF plot for the DWBB model. Red diamonds represent the observed values and boxplots represent the simulated values.

### 2.4.2   Swiss Great Tits

The Swiss great tits data were collected in the Swiss breeding bird survey MHB from 2013. The Swiss common bird breeding survey MHB is based on a sample of 267 1-km$^2$ areas. Volunteers survey a quadrant-specific route, composed of 263 sites, three times during the breeding season.

The Swiss great tits data was analysed by Kéry and Royle (2017) where they highlighted the good-fit-bad-prediction dilemma. Using covariates on both expected population size and detection probability, they analysed this data set using three models: P-B, ZIP-B and NB-B where they found that the best-fitted model (NB-B) via AIC produced unrealistic estimates of population size. To come to this conclusion, they performed residual diagnostic checks. They found that the residual diagnostic checks for the P-B and ZIP-B models looked much better than those of the NB-B model, despite the much better fit (GOF test) and predictive ability (measured by AIC) of the NB-B model. Thus, we investigate this good-fit-bad-prediction dilemma in a Bayesian framework using methods considered in this paper.

Convergence was achieved for all model parameters. From Table 2.15, we also find that the NB-B model was favored by both cWAIC and mWAIC over the P-B model, and the NB-B also produced large values of expected population size. To evaluate model fit, we simulate data from the model and compute the 95% coverage of $\tau_{1:M}$. For the P-B model, 33.34% of the sites captured the observed values while for the NB-B model, 72.09% of the sites captured the observed values. Contrary to MKe, we consider BB models, and as can be seen from Table 2.15, the DW-BB model was strongly supported as the best model amongst all models by both cWAIC and mWAIC with weights of 1.0. The DW-BB model produced inference similar to the P-B model but fitted the data well with 72.09% of the sites capturing the true value. This motivates the use of the DW-BB model as the good-fit-bad prediction dilemma observed by MKe may be due to the violation of the independence detection assumption in the Binomial detection process. Hence, our findings in both case studies suggest that model selection and model fit are in agreement with model selection favouring the model with the better fit.

Table 2.15 Model results from analysing Swiss great tits data. Values within the brackets represent the 95% credible interval. For BB models, detection probability represents mean detection probability.

| Model | cWAIC | mWAIC | Detection Probability | Expected population size | $p$ correlation |
|-------|-------|-------|-----------------------|--------------------------|-----------------|
| P-B | 3689.985 | 6020.880 | $0.641(0.621, 0.661)$ | $10.142(9.679, 10.616)$ | - |
| DW-B | 2968.517 | 3600.062 | $0.063(0.016, 0.132)$ | $123.412(53.72, 460.90)$ | - |
| NB-B | 2954.897 | 3579.046 | $0.045(0.011, 0.103)$ | $202.140(61.714, 624.950)$ | - |
| P-BB | 2663.500 | 4311.279 | $0.298(0.262, 0.335)$ | $20.131(18.184, 22.558)$ | $0.476(0.425, 0.526)$ |
| DW-BB | 2606.225 | 3564.125 | $0.416(0.249, 0.560)$ | $17.799(12.550, 30.460)$ | $0.054(0.015, 0.116)$ |

## 2.5 Discussion

As N-mixture models provide an attractive framework to gain inference on population size by using only replicated counts from unmarked individuals. A large number of studies have been carried out on N-mixture models in a classical setting, resulting in the identification of issues such as computational aspects of model fitting, model selection, sensitivity to overdispersion, etc. However, to our knowledge, few studies have been conducted in a Bayesian setting to investigate N-mixture models. N-mixture models have also become easier to fit in a Bayesian framework with the advent of software such as NIMBLE (de Valpine et al., 2017) and Stan (Carpenter et al., 2017). Hence, in this paper, we considered fitting an extensive class of N-mixture models in a Bayesian framework to corroborate and extend issues concerning N-mixture models obtained in a classical framework.

Moreover, we have performed extensive simulation studies to investigate the choice of prior distributions and model selection in N-mixture models. We implemented a novel proper objective prior, the OB prior, and compared its performance to approximations of the popular Jeffreys priors. We found these priors performed similarly in terms of inference. Importantly, when $p$ is small, we found that $\lambda$ can be considerably underestimated in addition to well-known cases of $\lambda$ being overestimated, a finding we believe to be previously unknown. We further investigated model selection via WAIC, considering both the conditional and marginal WAIC criteria, cWAIC and mWAIC, respectively. We found that cWAIC can lead to misleading results that favour the more complicated model while mWAIC selected the true model with a high probability. Hence, mWAIC should be used instead of cWAIC to select between competing N-mixture models.

Finally, we considered these methods in two case studies. We found the OB prior and a Jeffreys prior approximation produced similar inference results and model selection results as cWAIC and mWAIC were in agreement for the case study considered. In addition, contrary to the good-fit-bad-prediction highlighted by Kéry and Royle (2017), we find model selection via WAIC to be in agreement with the model goodness of fit when the DW-BB model is considered in the model list. Future work can be focused on developing Bayesian goodness-of-fit measures to check model assumptions of N-mixture models.

Notably, Vehtari et al. (2017) highlighted checks that can be done to investigate the stability of WAIC. Ariyo et al. (2022b) highlighted WAIC sensitivity to the choice of prior in Bayesian linear mixed models for longitudinal data. Thus, a possible avenue for future work can be to investigate the stability of cWAIC and mWAIC and WAIC sensitivity to the choice of prior in N-mixture models.

Another important avenue for future work is the identifiability of N-mixture models in a Bayesian framework as identifiability issues have been found in a classical setting (Barker et al., 2018; Dennis et al., 2015). Thus, future work can be focused on investigating identifiability of N-mixture models in a Bayesian setting using methods such as data cloning (Lele et al., 2007).

Overall, N-mixture models are a powerful tool for estimating population size. However, like any tool, care must be taken. This work highlights that in a Bayesian framework, care needs to be taken with the choice of prior distributions and advocates the use of mWAIC to select between models.

# Chapter 3

# A new modelling framework for roost count data

# Abstract

Roost counts, where individuals of a species are counted whilst arriving or departing from their roost site, are an important monitoring tool for several species around the world. However, the raw count data are an underestimate of the size of the monitored population at any one time because of individuals temporarily not using the roost (temporary emigration, TE) and because the probability of detection of individuals, even when using the roost, is typically much lower than one (observation error). In this paper, we develop a novel modelling framework for estimating population size, from roost count data, while accounting for both TE and observation error. Our framework builds on the popular class of N-mixture models but extends them in a number of ways. Specifically, we introduce two model classes for TE, a parametric, which relies on temporal models, and a non-parametric, which relies on Dirichlet process mixture models. Both model classes give rise to interesting ecological interpretations of the TE pattern while being parsimonious in terms of the number of parameters required to model the pattern. When accounting for observation error, we use mixed-effects models and implement an efficient Bayesian variable selection algorithm for identifying important predictors for the probability of detection. We demonstrate our new modelling framework using an extensive simulation study, which highlights the importance of using mixed-effects models for the probability of detection and illustrates the performance of the model when estimating population size and underlying TE patterns. We also assess the ability of the corresponding variable selection algorithm to identify important predictors under different scenarios for observation error and its corresponding model. When fitted to two motivating data sets of parrots, our results provide new insights into how each species uses the roost throughout the year, on changes in population size between and within years, and on important predictors for observation error.

## 3.1   Introduction

The loss of Earth's biological diversity negatively impacts ecosystem services that are vital for human health and prosperity (Cardinale et al., 2012). This global issue is recognised by International agreements and policy frameworks including the Convention on Biological Diversity (CBD) and the United Nations Sustainable Development Goals (SDGs), which call upon all United Nations Member States to take urgent action to restore and protect habitats and to halt further biodiversity loss (sdgs.un.org).

With an increasing number of species suffering population declines (Almond et al., 2020; Thomas, 2013), it is paramount to develop innovative monitoring methods in order to characterise population dynamics, understand how environmental changes affect populations, identify species that require protection, and develop or appraise management practices, policies and guidelines (Jetz et al., 2019). However, some highly mobile species such as parrots (Dénes et al., 2018) and bats (Kunz, 1982) can be challenging to monitor because they are not individually identifiable and they often feed and nest in low densities among inaccessible habitats such as forest canopies (Dénes et al., 2018). Consequently, one of the only opportunities to survey such species at scale is at communal roosts where large numbers of individuals may regularly come together and interact socially for reasons including predator avoidance, cooperative breeding, information exchange, informing foraging strategy and meeting thermoregulatory demands (Beauchamp, 1999; Kunz, 1982; Salinas-Melgoza et al., 2013; Seixas and Mourao, 2018). During these surveys, individuals are observed and counted as they arrive or depart from their roost, which is a more cost-effective method than others, such as capture-mark-recapture, which can be impractical for such species (Kunz, 1982).

In practice, the specific methodology includes multiple simultaneous counts obtained by one or more observers positioned at one or more vantage points. In addition, due to the challenging nature of performing roost counts and the costs and challenges of identifying and accessing a roost, only a single roost is often monitored for a given species, with counts typically collected under different environmental conditions (Berg and Angel, 2006; Cougill and Marsden, 2004; Matuzak and Brightsmith, 2007). These roost survey counts cannot serve as an index of population size due to individuals exhibiting TE, and hence becoming temporarily unavailable for detection, and due to

observation error, with the probability of detecting individuals that are available for detection typically being much lower than one. Therefore, statistical modelling needs to be employed for inferring population size and TE patterns from roost count data. This is the aim of this paper, as we describe below.

Count data for closed populations that do not exhibit TE are often analyzed using standard N-mixture models (Royle, 2004b), which can estimate population size using spatially-replicated counts over time by accounting for observation error. The time-for-space substitution N-mixture model (Kéry and Royle, 2015) uses temporally replicated counts without spatial replication, giving temporal estimates of population size and enabling estimation of a single population trend, but also does not account for TE. However, Chandler et al. (2011) showed that failure to account for TE can result in positively biased estimates of population size.

Roost survey sampling usually takes place under Pollock's robust design (Pollock, 1982), with several short secondary periods, eg days, across various primary periods, eg months. The population size is then assumed constant across secondary periods within the same primary period (closed population) but can change between primary periods (open population) due to births, deaths, immigration, or permanent emigration. In this case, Chandler et al. (2011) extended the standard N-mixture models to account for TE. This model has two processes: an ecological process for the latent number of individuals present and available for detection, and an observation process, for the available individuals detected. The proportions of individuals in the population in any given primary period that are available for detection on each secondary period are either assumed constant for the duration of the study period (Chandler et al., 2011) or are estimated separately of each other, requiring one parameter to be estimated for each primary period (Kéry and Royle, 2020). However, the first option may be too restrictive and the latter is parameter-greedy, and does not allow for an intuitive ecological interpretation of the results. Finally, existing models do not provide information on TE cyclical patterns, where certain primary periods of each year correspond to certain levels of TE. Identifying and inferring these cyclical patterns can give new insights into the behaviors of the species, such as breeding patterns and seasonal availability of foods.

Naturally, detection probability, and hence observation error (with the two terms used interchangeably in this paper), is expected to vary between sampling occasions as a response to changes in

environmental and weather conditions or effort. This variation can be captured within a logistic regression model accounting for the effect of covariates, such as time of sampling and weather conditions at the time of surveying (see for example Kéry and Royle, 2020; Neubauer et al., 2022). All of the existing modelling approaches can account for the effect of covariates (referred to as variables or predictors in the literature and in this paper) on detection probability through fixed effects models for a given variable set. However, it is unlikely that these fixed effects will capture all of the variation in detection, as other, unobserved or unobservable effects, such as the behaviour of the surveyed species, can have a substantial impact on observation error. The impact of using fixed-effects is unclear when the model for observation error is misspecified, that is when important variables for observation error are omitted, which is likely to be the case in reality. Additionally, the potential set of variables to be considered as predictors for observation error can be large, and hence corresponding tools are required to identify the subset of important variables in the model.

Motivated by two roost count survey data sets of parrots, in this paper we develop a novel modeling framework that can be used to estimate time-varying population size at a site, while accounting for TE and observation error. We extend the TE N-mixture model developed by Chandler et al. (2011) by proposing two model classes: a parametric approach, which employs different temporal models that account for temporal auto-correlation of different order, and a non-parametric approach based on the Dirichlet process (DP) prior (Ferguson, 1973) that allows us to cluster the primary periods according to roost use by the surveyed individuals, and leads to interesting ecological insights about the behavior of the population.

To account for variation in observation error, in addition to that captured by a fixed-effects model, we introduce a mixed-effects logistic regression model on the detection probability. Additionally, we implement a recent efficient Bayesian variable selection (BVS) algorithm, the Bayesian Group Lasso Spike and Slab (BGLSS) (Liquet et al., 2017; Xu and Ghosh, 2015), to perform variable selection for the probability of detection in this mixed-effects model framework.

We implement our novel modelling framework in a Bayesian setting using Markov Chain Monte Carlo (MCMC) methods via R package NIMBLE (de Valpine et al., 2017) version 0.13.0.

We present an extensive simulation study that assesses the performance of the proposed models in estimating population size and TE patterns under different scenarios, such as when the model for

observation error is misspecified. For the first time in N-mixture models and related literature, we highlight the risks of using misspecified fixed-effects models for observation error and demonstrate how the risks are mitigated by instead using mixed-effects models, as we propose in this paper. We also demonstrate the performance of our proposed variable selection approach in identifying important predictors for observation error in our novel mixed-effects modelling framework under these scenarios.

Finally, we apply our new modelling framework to two case studies, considering roost count data on Ecuadorian Amazon parrots *Amazona lilacina* and on Orange-winged Amazon parrots *Amazona amazonica*. We use cross-validation to select the most appropriate model for the TE pattern in each case and obtain interesting ecological results on temporal population sizes, TE trends, and cyclical patterns and identify important predictors affecting observation error.

The paper is organized as follows. In Section 3.2 we define our new modelling framework, including background on the methods on which it builds. Simulation results are presented in Section 3.3 and the results for the two case studies are presented in Section 3.4. Section 3.5 concludes the paper and provides ideas for potential future directions.

## 3.2 Models

Sampling follows Pollock's robust design (Pollock, 1982) with $T$ open primary periods (e.g. months) and $J$ closed secondary periods (e.g. days within a month). Often, studies can have $Y$ additional top-level primary periods, e.g. $Y$ years, with $T$ primary periods, e.g. months, and $J$ secondary periods, e.g. days within them. The data are summarised in counts $C_{j,t,y}$ of individuals detected on secondary occasion $j$, primary period $t$, within top-level primary period $y$.

We assume there is an overall super-population of $M$ individuals that can visit the roost at least once during the survey period. These $M$ individuals can contribute to the $Y$ super-population sizes ($\kappa_y$, $y = 1, \ldots, Y$), indicating the number of individuals that can visit the roost at least once in each top-level primary period and denote the probability that an individual from the super-population has used the roost at least once in top-level primary period $y$ by $\delta_y$. Conditional on $\kappa_y$, we denote the number of individuals using the roost in primary period $t$ within top-level primary period $y$ by $N_{t,y} \sim \text{Bin}(\kappa_y, \theta_{t,y})$

(temporal population size), with $\theta_{t,y}$ referred to as the availability parameters (meaning that these individuals are available for detection in that primary period). Finally, individuals that use the roost in primary period $t$ within top-level primary period $y$ are detected on secondary occasion $j$ with probability $p_{j,t,y}$. The hierarchical representation of the model is given in equation (3.2.1), while a graphical representation of the model is given in Fig 3.1.

$$M \sim \text{Poisson}(\lambda)$$

$$\kappa_y \sim \text{Binomial}(M, \delta_y)$$

$$N_{t,y} \sim \text{Binomial}(\kappa_y, \theta_{t,y})$$

$$C_{j,t,y} \sim \text{Binomial}(N_{t,y}, p_{j,t,y}) \tag{3.2.1}$$

The $\kappa_y$ variables allow us to study the availability pattern within each top-level primary period, conditional on the corresponding population size, and hence identify changes in availability patterns across top-level primary periods, without these changes being confounded to changes in population size. When there are no top-level primary periods, this model can be simplified by dropping the $\kappa_y$ level, i.e. setting $\kappa_y = M \; \forall y$, and the $y$ subscript in all subsequent levels.

The main novelty of our proposed framework lies in the way in which we model detection probability, as described in Section 3.2.1, and the availability parameters, as described in Section 3.2.2.
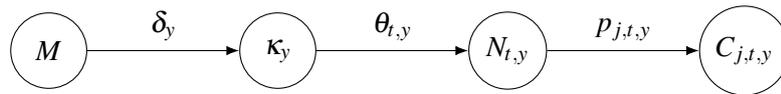


Fig. 3.1 Graphical model representation

## 3.2.1 Detection Probability

The model of equation (3.2.1) is a function of the detection probability on secondary occasion $j$, primary period $t$, top-level primary period $y$, $p_{j,t,y}$. This probability cannot be freely varying, as that introduces more parameters than we can estimate into the model. Instead, it can be assumed as constant for all $j, t, y$ or, more realistically, as a function of variables (covariates), which can vary between secondary and/or primary periods, within a logistic regression framework, as for example in

Kéry and Royle (2020). However, it is likely that, in practice, such models are misspecified, and that the variables considered are only a subset of the variables that affect detection probability in the field. In such cases, as we demonstrate in our simulation study in Section 3.3, the estimation of population size can be substantially biased, and for that reason we propose the use of a mixed effects model:

$$\text{logit}(p_{j,t,y}) = \eta_{j,t,y} = \mu + \sum_{g=1}^{G} X_{j,t,y,g} \beta_g + \varepsilon_{j,t,y} \tag{3.2.2}$$

where $g = 1, \ldots, G$ are continuous/categorical variables, such that variable $g$ requires $C_g$ coefficients to model its effect, so that if $g$ is a continuous variable, $C_g = 1$, and if $g$ is a categorical variable, $C_g$ is its number of levels (excluding baseline). Finally, $\beta_g$ is the $(C_g \times 1)$ vector corresponding to the logistic regression coefficients for variable $g$, $X_{j,t,y,g}$ is the vector of length $C_g$ containing variable $g$ on occasion $t, y, j$, and $\varepsilon_{j,t,y} \sim \text{Normal}(0, \sigma_\varepsilon^2)$ are corresponding independent random effects.

The inclusion of the random effect terms allows for any variability in detection probability that is not captured by the variables considered by the fixed effects to be absorbed by the random effect variance, which, as we demonstrate using simulation, leads to reliable inference on population size, even when the detection probability model is misspecified. However, an overparameterised fixed effects model can lead to increased uncertainty around variable effects and population size, and therefore, we suggest the use of a Bayesian variable selection algorithm, and specifically of the Bayesian Group Lasso Spike-and-Slab (BGLSS) algorithm (Xu and Ghosh, 2015), for identifying important predictor variables for $p$. The BGLSS places a prior on each group of coefficients, where a group can consist of coefficients introduced to model the effect of a categorical variable and can number a single coefficient in the case of continuous variables. This prior is given in equation (3.2.3) below, and more details are provided in B.2 of the Supplementary material.

$$\beta_g | \tau_g^2 \sim (1 - \gamma_g) \delta_0(\beta_g) + \gamma_g N(0, \tau_g^2 I_{C_g})$$
$$\tau_g^2 \sim \text{Gamma}\left(\frac{C_g + 1}{2}, \frac{\psi^2}{2}\right)$$
$$\gamma_g \sim \text{Bernoulli}(\phi_g)$$
$$\psi \sim \text{Gamma}(a, b) \tag{3.2.3}$$

where $\gamma_g$ is a binary variable that indicates whether variable $g$ is included (1) in the model or not (0), $\delta_0(\beta_g)$ denotes a point mass at $0 \in \mathbb{R}^{C_g}$, $I_{C_g}$ is the identity matrix ($C_g \times C_g$), $\psi$ is the shrinkage parameter, and $\phi_g$ is the prior inclusion probability, which can be fixed to 0.5 or can be assigned a uniform or Beta prior distribution.

The BGLSS accommodates group-level variable selection by using a spike and slab prior (Mitchell and Beauchamp, 1988), with coefficients exactly zero for excluded variables, and the Bayesian group lasso (BGL) (Casella et al., 2010) for included variables, enforcing the $L_1$ penalization (Tibshirani, 1996), giving more parsimonious models. This Bayesian formulation can reduce the computational cost by proposing a prior on $\psi$ rather than testing several values and choosing the best value by cross-validation. In addition, the BGLS produces reliable standard errors of coefficients without any extra cost in comparison to the frequentist group lasso (Yuan and Lin, 2006).

### 3.2.2 Availability Parameters

We propose two model classes for modelling the availability parameters, a nonparametric approach and a parametric approach, both of which are described below. We define $\theta_\ell = \theta_{t,y}$, with $\ell = t + T(y-1)$ for $\ell = 1, \ldots T \cdot Y$ to model correlation in the availability parameters for the whole time series, across primary periods. When there are no top-level primary periods, $Y = 1$ and $\theta_\ell = \theta_t$ for $\ell = 1, \ldots, T$. Table 3.1 provides the terminology used hereafter for each model considered for the availability parameters.

Table 3.1 Models proposed for availability parameters.

| Notation | Model |
|----------|-------|
| DP | Dirichlet process (DP) mixture model |
| RW1 | Random walk of order 1 |
| RW2 | Random walk of order 2 |
| Cor | Across level correlation model |
| AR1 | Auto-regressive model of order 1 |

**Nonparametric approach**

We model availability non-parametrically via a Beta-Dirichlet process (DP) mixture model (Kottas, 2006). This formulation expresses the distribution of availability parameters as a mixture model, and provides a flexible and robust specification of the corresponding density, by describing it as a mixture model with an unknown number of components, with primary periods clustered according to their corresponding availability parameters, eg low, medium, and high. This is ecologically relevant as it enables the study of TE trends and hence roost use patterns throughout the season(s).

The Beta DP mixture model can be represented using the Chinese restaurant process (CRP) algorithm, which relies on the inferred cluster allocation variables, $z_\ell$, $\ell = 1, \ldots, T \cdot Y$, indicating the cluster to which primary period $\ell$ has been allocated. The CRP is used to represent the sequential way in which cases, i.e. periods in our case, are allocated to clusters, with the number of clusters being infinite a priori, but finite in practice and inferred as part of the process. The corresponding model for the availability parameters is given in equation (3.2.4).

$$\theta_\ell | \tilde{\gamma}, \tilde{\psi}, z_\ell \sim \text{Beta}(\tilde{\gamma}_{z_\ell}, \tilde{\psi}_{z_\ell}), \quad \ell = 1, \ldots, (T \cdot Y)$$

$$z_\ell \sim \text{CRP}(\alpha), \quad \alpha \sim \text{Gamma}(\zeta, \tau)$$

$$\tilde{\gamma}_k \sim \text{Gamma}(\mu, \nu), \quad \tilde{\psi}_k \sim \text{Gamma}(\vartheta, \omega), \quad k = 1, \ldots, K.$$

$$(3.2.4)$$

where $\zeta, \tau, \mu, \nu, \vartheta, \omega \in \mathbb{R}$ and $K \leq (T \cdot Y)$. More details are provided in B.1 of the Supplementary material.

**Parametric approach**

Alternatively, availability can be modelled parametrically using temporal models, specifically random walk models and auto-regressive models. These temporal models share information across primary periods by accounting for temporal auto-correlation, which is meaningful ecologically as, as also mentioned above, the availability pattern is expected to be smooth and allows for borrowing strength in cases where the data are sparse.

1. Random walk models, which enable estimation of non-linear temporal trends retaining the smoothing-varying feature that is present in observed time series data. As highlighted in Fahrmeir and Lang (2001), random walk models can be rewritten in an undirected symmetric form, as a one-dimensional version of the spatial intrinsic conditional autoregressive (ICAR) model (Besag, 1974). Generally, random walk models can be defined as a set of conditional probability distributions under the ICAR models as

$$\theta_\ell | \theta_{-\ell}, \sigma^2, W^{\text{RW}} \sim N\left[\frac{\sum_{n=1}^{T \cdot Y} w_{\ell n} \theta_n}{w_{\ell+}}, \frac{\sigma^2}{w_{\ell+}}\right], \ell = 1, \ldots, T \cdot Y. \qquad (3.2.5)$$

where $W^{\text{RW}}$ represents the temporal weights matrix with entry $\omega_{\ell n}$ in the $\ell$th row and the $n$th column, $w_{\ell+}$ is the sum of the elements in the $\ell$th row, $\sigma^2$ is the ICAR variance and $\sigma^2/\omega_{\ell+}$ is the conditional variance.

Consequently, random walk models possess the same set of properties as the ICAR model. That is, positive auto-correlation is assumed via a chosen $W$ that imposes a neighbourhood structure on time points in the study period and determines the amount of information borrowed from other time points. This shared information across temporal neighbours results in temporally smooth time trends, with estimation of $\theta_\ell$ borrowing information from past time points eg. $(\ell-1, \ell-2)$ but also from future time points eg. $(\ell+1, \ell+2)$, provided that these time points are within the study period. In addition, as the conditional variance increases, $\theta_\ell$ can deviate more from its neighbours, producing a temporal pattern that is less smooth but more flexible. This model representation allows us to infer the variance of the ICAR model ($\sigma^2$) and $\theta_\ell \, \forall \ell$.

- Random walk of order 1 (RW1) can be defined as an ICAR model with binary weights, $W^{\text{RW1}}$, such that the entry $\omega_{\ell,n} = 1$ if points $\ell, n$ are neighbours and 0 otherwise. In the RW1 model, each $\ell$ has 2 neighbours $\ell-1, \ell+1$, except the first and the last, which only have one neighbour, adjacent to the right and left respectively. The binary temporal weights matrix, $W^{\text{RW1}}$, assumes that equal strength of information is borrowed from

adjacent neighbours.

- Random walk of order 2 (RW2). Similarly, the RW2 model can be defined as an ICAR

    model but with a general weights matrix ($W^{RW2}$). The elements in $W^{RW2}$ are derived from

    the conditional distributions of each $\theta_\ell$ conditioned on all other parameters in $\theta$ and the

    variance $\sigma^2$ (conditional distributions listed in B.4.2 of the Supplementary material). The

    elements are the coefficients in the numerator of the conditional mean for $\theta_\ell$. As can be

    seen in equation (3.2.5), the conditional variance depends on the number of neighbours,

    hence, the RW2 model generally produces smoother temporal trends than the RW1 model

    as it borrows information from more time points. In addition, using a general weights ma-

    trix instead of a binary weights matrix specifies the strength of the information borrowed,

    with more information borrowed from close neighbours.

- Across level correlation (Cor) model. We extend the RW1 model to allow a time point

    to borrow information from other specific time points, in addition to $\ell - 1, \ell + 1$ time

    points given time points are within the study period. For instance, this allows a specific

    month in a year to be correlated to months directly before and after that month, but also

    the same month across years. This model is defined similarly to the RW1 model with a

    binary weights matrix ($W^{Cor}$) such that the entry $\omega_{\ell,n} = 1$ if points $\ell, n$ are neighbours

    and 0 otherwise, where neighbours in this case are the adjacent time points, but also time

    points that are $c$ time periods apart, where $c = 12$ in the case of monthly patterns across

    years. Therefore, the first time point are neighbour with $(\ell + 1, \ell + qc)$ time points, the

    last time point with $(\ell - 1, \ell - qc)$ neighbours and others with $(\ell - 1, \ell + 1, \ell \pm qc)$ for

    $q = 1, \ldots, ((T \cdot Y)/c) - 1$, provided time points are within the study period.

2. Auto-regressive models. An auto-regressive model of order 1 (AR1) on the set of time-specific parameters can be defined as

$$\theta_\ell = \rho\theta_{\ell-1} + \varepsilon_\ell, \ \ell = 2, \ldots, T \cdot Y, \quad (3.2.6)$$

$$\theta_1 \sim \mathrm{N}(0, \sigma_1^2(1 - \rho^2))$$

where $\rho$ is the temporal correlation coefficient ($|\rho| < 1$) and $\varepsilon_\ell \sim N(0, \sigma^2)$ are iid noise effect terms. The RW1 model is a subset of the AR1 model when $\rho = 1$. As such, the AR1 is a more flexible model as it accommodates both positive and negative temporal auto-correlation. However, if positive auto-correlation is present, the RW1 model is preferable as one fewer parameter needs to be estimated.

### 3.2.3 Inference

We fit models in a Bayesian framework using MCMC methods via R package NIMBLE (de Valpine et al., 2017) version 0.13.0. Specifically, for variables assigned an ICAR model, we follow NIMBLE's recommendation and update these variables without the zero constraints and then centering (Paciorek, 2009). We implement the Beta mixture DP model by using the collapsed sampler (Neal, 2000) provided in NIMBLE. We use methods developed by Wade and Ghahramani (2018) to summarise DP cluster results. We employ median thresholding in variable selection (Barbieri and Berger, 2004), that is, $\mathrm{Pr}(\gamma_g = 1|y) > 0.5$, $g = 1\ldots,G$ to identify significant variables.

## 3.3 Simulation Study

In this section, we present an extensive simulation study to explore a number of different cases, listed in Table 3.2. For each case, we perform 50 simulation runs and we set $T = 36, J = 8$, assuming no top-level primary periods with $\lambda = 100$ and consider high and low detection levels, $p \approx (0.6, 0.3)$, with $p$ as a function of covariates (variables). The coefficients for fixed effects are set as: $\beta = (\beta_1 = 1.25, \beta_2 = 0.2, \beta_3 = 2, \beta_4 = 0, \beta_5 = -0.6, \beta_6 = 0.5, \beta_7 = -1, \beta_8 = 0)$ with the first five corresponding to continuous variables, $x_1, \ldots, x_5$, and last three to categorical variables, $x_6$ and $x_7$, with two and three

levels, respectively. Continuous variables were generated from a standard normal distribution and categorical variables from a multinomial distribution with equal probabilities. To obtain the desired level of average detection, as stated above, the intercepts, $\beta_0$, were set to $(0.75, -1.5)$, for high and low detection probability, respectively. To introduce misspecification in the model for detection, variables $x_1$ and $x_7$ were not included in the model in each of the two cases described in Table 3.2. When the DP model was used to generate the data, we specified two clusters of equal size (18) from $\text{Beta}(10, 10)$ and $\text{Beta}(10, 1)$ respectively. When the RW1 model was used to generate data, we set $\sigma = 1$.

The following prior distributions were used in all cases: $\lambda \sim \text{Gamma}(0.01, 0.01)$, $\psi \sim \text{Gamma}(0.001, 0.001)$, $\phi_g = 0.5, \beta_1 \sim \text{Normal}(0, 2), \sigma \sim \text{Uniform}(0, 15), \alpha \sim \text{Gamma}(1, 1)$, $\tilde{\gamma}_k \sim \text{Gamma}(2, 0.1), \tilde{\psi}_k \sim \text{Gamma}(2, 0.1)$. The MCMC settings in terms of the number of iterations, burn-in, and thinning in each case are reported in B.5 of the Supplementary material.

Table 3.2 Simulation settings.

| Case | Description |
| --- | --- |
| 1 | Comparing estimation of population size under different models for the availability parameters when the correct model for these parameters is fitted to the data and we do not perform variable selection and<br>a) the model for detection probability is correctly specified.<br>b) the model for detection probability is misspecified (fixed vs mixed effects models). |
| 2 | Assessing the performance of BGLSS in variable selection under the RW1 model for the availability parameters when<br>a) the model for detection probability is correctly specified.<br>b) the model for detection probability is misspecified (mixed effects model). |

We use median relative bias and median 95% credible interval (CI) coverage to summarise the estimation of population size and of covariate effects. We also use median misclassification rate for summarising the DP mixture clustering and the BGLSS performance. The detailed results of the simulation study for each case are presented in B.5 of the Supplementary material and the key findings are summarised in Table 3.3 and discussed below.

Table 3.3 Median relative bias and median 95% CI coverage of population size and covariate coefficients for each simulation scenario and setting for detection probability, as described in Table 3.2. CS: correctly specified; MS: misspecified; FE: fixed effects; ME: mixed effects.

| Case | Model for $\theta$ | Model for $p$ | Average $p$ | Parameters | RB | Coverage |
|---|---|---|---|---|---|---|
| 1. a) | | | | | | |
| | | | 0.6 | Coefficients | 0.008 | 94 |
| | | | | Population size | −0.001 | 100 |
| | DP | CS - FE | | | | |
| | | | 0.3 | Coefficients | 0.002 | 98 |
| | | | | Population size | −0.002 | 98 |
| | | | 0.6 | Coefficients | 0.003 | 96 |
| | | | | Population size | −0.001 | 98 |
| | RW1 | CS - FE | | | | |
| | | | 0.3 | Coefficients | −0.003 | 96 |
| | | | | Population size | −0.005 | 98 |
| 1. b) | | | | | | |
| | | | 0.6 | Coefficients | −0.787 | 2 |
| | | | | Population size | 8.928 | 0 |
| | DP | MS - FE | | | | |
| | | | 0.3 | Coefficients | −0.590 | 4 |
| | | | | Population size | 4.417 | 0 |
| | | | 0.6 | Coefficients | −0.745 | 4 |
| | | | | Population size | 6.066 | 2 |
| | RW1 | MS - FE | | | | |
| | | | 0.3 | Coefficients | −0.502 | 12 |
| | | | | Population size | 3.347 | 4 |
| | | | 0.6 | Coefficients | 0.036 | 94 |
| | | | | Population size | −0.005 | 98 |
| | DP | MS - ME | | | | |
| | | | 0.3 | Coefficients | 0.011 | 90 |
| | | | | Population size | −0.019 | 90 |
| | | | 0.6 | Coefficients | 0.046 | 98 |
| | | | | Population size | −0.004 | 100 |
| | RW1 | MS - ME | | | | |
| | | | 0.3 | Coefficients | 0.043 | 96 |
| | | | | Population size | 0.013 | 92 |
| 2. a) | | | | | | |
| | | | 0.6 | Coefficients | 0.001 | 96 |
| | | | | Population size | −0.001 | 98 |
| | RW1 | CS - FE | | | | |
| | | | 0.3 | Fixed effects | −0.006 | 96 |
| | | | | Population size | −0.001 | 98 |
| 2. b) | | | | | | |
| | | | 0.6 | Coefficients | | |
| | | | | Population size | −0.005 | 100 |
| | RW1 | MS - ME | | | | |
| | | | 0.3 | Coefficients | | |
| | | | | Population size | 0.031 | 90 |

**Case 1**

When the model for detection probability is correctly specified (a), both the DP and the RW1 models perform well in terms of inference, with low median relative bias and high coverage for covariate coefficients and population size. The DP mixture model has a low misclassification rate, on average equal to 0.055 for both levels of detection. In addition, the standard deviation of the RW1 model ($\sigma$) is also estimated well with low relative bias $(0.011, -0.035)$ and high coverage $(0.98, 1)$ at high and low levels of detection respectively. Consequently, this scenario shows that both models for the availability parameters perform well in terms of inference when the model for detection probability is correctly specified.

However, when the model for detection probability is misspecified (b) and a fixed effects detection model is used, estimation of population size is considerably positively biased with very poor coverage in all cases. Similarly, covariate coefficients are estimated with high bias and low coverage and the DP mixture model performs poorly, with a misclassification rate on average equal to $(0.111, 0.444)$ for high and low detection probability, respectively. However, using a mixed effects model for detection probability corrects for the misspecification and produces population size and covariate coefficient estimates with negligible bias and high coverage. The DP mixture model also performs better, with a misclassification rate on average equal to $(0.055, 0.111)$ for high and low detection probability, respectively.

**Case 2**

Similarly, when the model for detection probability is correctly specified (a), BGLSS performs well in identifying both significant (strong and weak) and non-significant effects with mean misclassification rates of 0 across both levels of detection. As such, population size and covariate coefficients are estimated well in all cases.

When the model for detection probability is misspecified (b) and a mixed effects detection model is employed, BGLSS has, as expected, lower power to identify weak effects ($\beta_2 = 0.2$) with average misclassification rate $(0.38, 0.4)$ at high and low detection probability, respectively, but still high power to identify strong effects with average misclassification rate 0 at both levels of detection. In

addition, the power to identify non-significant variables also declines, with a mean misclassification rate $(0.1, 0.06)$ at high and low detection levels respectively. However, importantly, inference on population size is unaffected in all cases when mixed effects models for detection probability are employed.

## 3.4 Case Studies

### 3.4.1 Ecuadorian Amazon Parrots

We consider roost count data collected as part of an ongoing conservation project for the Ecuadorian Amazon parrot (*Amazona lilacina*) in Ecuador (Biddle et al., 2020, 2021a,b). Counts were obtained from a single site close to the El Salado Mangrove Reserve, where parrots roost overnight, for 36 consecutive months between 2016 and 2019. Each year, surveys took place between November and October, with surveys taking place on three to five days within each month, and two surveys being performed each day, AM and PM. We assume that the population is closed within each month, but open between months.

We model the data using the model defined in equation (3.2.1), fitting all models listed in Table 3.1 and use k-fold cross-validation to select the most appropriate model for the availability parameters. In each case, we consider a mixed effects model for detection probability, and perform variable selection via BGLSS, considering the following variables: median temperature, average relative humidity, visibility, average wind speed, rain/drizzle, storm/thunder (taken from the Simon Bolivar weather station approximately 14km from the roost site (https://www.tutiempo.net/clima/01-1999/ws-842030.html), time of sampling (AM/PM), and weather recorded by the observer at the roost site (clear, cloud, rain, sunshine). The prior distributions were set as described in the simulation study.

k-fold cross-validation was performed by splitting the data into monthly subsets ($k = 36$) and using root mean square error (RMSE) to evaluate the predictive accuracy of the models considered when leaving one month out at a time. RW1 was selected as the model with the lowest RMSE, as seen in Table 3.4. RW1, RW2, and Cor are the top three models, having similar RMSE values. Notably, all these models considered produced similar estimates of population size, BVS results, and model fit. Consequently, we display the results obtained from the RW1 model in the paper, while the

results obtained from the other models are presented in B.6.1 of the Supplementary material, with the exception of the DP model clustering results, which are shown in Table 3.5 and discussed as they provide us with new insights about the use of the roost throughout and across years.

Table 3.4 Ecuadorian Amazon parrots case study. Cross-validation results.

| Model | DP | RW1 | RW2 | Cor | AR1 |
|-------|--------|--------|--------|--------|--------|
| RMSE | 66.850 | 59.925 | 61.157 | 62.599 | 66.436 |

Fig 3.2a shows posterior summaries of the month-specific population sizes, $N_1, \ldots, N_{36}$, obtained from the RW1 model. The pattern suggests two peaks in the year, January/February/March and then June/July/August. The first peak, which is more consistent across years, could represent chicks fledging and returning to the roost with the adults, while the second peak, which varies more between years, could represent social gathering before the breeding season, giving opportunities for time to create breeding pairs and highlighting the importance of these communal roosts for the formation of new breeding pairs.

We assessed the fit of models using posterior predictive goodness of fit. For that, we define *monthly rate* to be the sum of the counts obtained in a month divided by the number of surveys in that particular month. Using MCMC samples, we simulated counts, and hence rates, from our models and compared these to the observed rates. Fig 3.2b displays that the RW1 model fits the data well as it produces similar monthly rates to the observed rates, with the true values falling within the 95% posterior credible interval of simulated values and with no consistent pattern of bias observed.

(a) Population size

(b) GOF
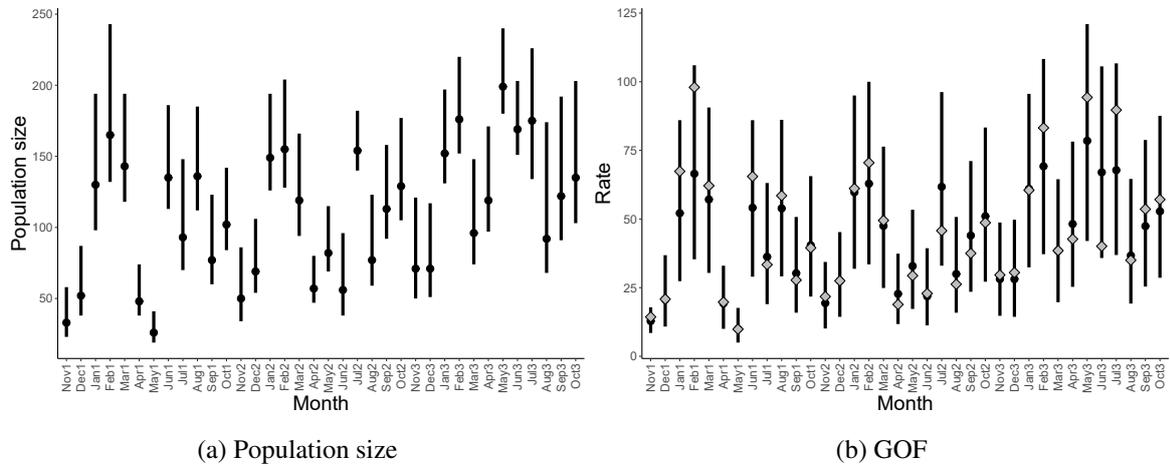
Fig. 3.2 Ecuadorian Amazon parrots case study. (a) The black dots represent the posterior mean population size for each month and the thick bands represent the corresponding 95% posterior credible interval. (b) The diamonds are the observed monthly rates and the thick bands represent the 95% intervals of simulated monthly rates. In both cases, the x-axis represents the months in each year with months ending in 1, 2, and 3 denoting months in the 1st, 2nd, and 3rd year, respectively.

The results of the RW1 model are consistent with the clustering output of the DP model (Table 3.5), where two clusters of equal size (18) have been identified for each year. These correspond to months with low (L) and months with high (H) availability probabilities, with the clustering pattern fairly consistent across years and agreeing with the general trend identified by the RW1 model. Locating and observing individual nests for this species can be difficult, and hence this clustering pattern of the overall roosting population provides supportive evidence to reports of seasonal breeding behaviour. The first peak corresponds with months when chicks fledge from nests (January / February / March) and so is likely to represent population recruitment, whilst the second peak in October occurs just before breeding pairs start to nest together in the dry forest and could represent an increase in attendance at the social roost to form or strengthen pair bonds. Due to the fluctuating nature of this particular roost site, accounting for detection probability allows us to identify robust patterns for ecological interpretation that would not be visible clearly in the raw data, helping conservation managers to determine breeding phenology more broadly so that efforts can be more focused on finding nest cavities and documenting breeding success at the right time of year. In other Amazon parrot species roost attendance is also linked with food availability (i.e. in times of food scarcity,

roost attendance is greater to allow information sharing) so it is also possible that fluctuating food availability in this seasonal climate may drive high/low distinction.

Table 3.5 Ecuadorian Amazon parrots case study. Cluster allocations from the DP model.

| | Months | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | Nov | Dec | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct |
| 1 | L | L | H | H | H | L | L | H | L | H | L | H |
| 2 | L | L | H | H | H | L | L | L | H | L | H | H |
| 3 | L | L | H | H | L | L | H | H | H | L | L | H |

Baseline detection probability is fairly low (posterior mean = 0.365 with $(0.261, 0.460)$ 95% posterior credible interval). Rain, storm, and time of sampling are identified as important predictors for observation error with posterior inclusion probabilities (PIP): $0.562, 0.613$, and $0.691$ respectively, but all with 95% posterior credible intervals covering 0 (Table 3.6). Rain, storm, and surveying in PM instead of AM have an estimated positive effect on the probability of detection. The presence of rain and storm can force parrots to fly lower down in the sky and land close to the observation point to gain shelter, increasing the probability of detection. Higher detection probability in PM than in AM is possibly due to the character of final destination: in the PM parrots are flying to one communal roost while in the AM parrots fly in multiple directions based on food dispersal and nest location, making it more difficult to detect them.

Table 3.6 Ecuadorian Amazon parrots case study. Posterior summaries of coefficients for the detection probability model.

| Coefficient | Mean | SD | 95% PCI |
|---|---|---|---|
| Intercept | $-0.553$ | 0.229 | $(-1.040, -0.159)$ |
| Median Temperature | 0.006 | 0.028 | $(-0.039, 0.091)$ |
| Humidity | 0.003 | 0.027 | $(-0.049, 0.076)$ |
| Visibility | $-0.001$ | 0.026 | $(-0.067, 0.056)$ |
| Wind Speed | $-0.022$ | 0.056 | $(-0.211, 0.029)$ |
| Rain | 0.049 | 0.105 | $(-0.034, 0.355)$ |
| Storm | 0.164 | 0.308 | $(-0.034, 1.050)$ |
| Time-PM | 0.106 | 0.139 | $(-0.011, 0.438)$ |
| Weather-Cloud | $-0.011$ | 0.045 | $(-0.144, 0.028)$ |
| Weather-Rain | 0.000 | 0.045 | $(-0.078, 0.085)$ |
| Weather-Sunshine | $-0.001$ | 0.043 | $(-0.087, 0.078)$ |

### 3.4.2 Orange-Winged Amazon Parrots

We next consider roost count data from Orange-winged Amazon parrots (*Amazona amazonica*) in Brazil. Counts were collected from a single site at an island near Belém, Pará between September 2004 and September 2005, with 96 surveys conducted (54 in the afternoon and 42 in the morning) across 50 weeks. More details can be found in De Moura et al. (2010). We assume that the population is closed within each week, but open between weeks. Therefore, in this case, the primary periods correspond to weeks, and there are no top-level primary periods. Detection probability is modelled as a function of the following categorical covariates: Cloud (cloudy, partially cloudy, no cloud), wind (strong wind, medium wind, low wind), rain (yes, no) and time of sampling (AM or PM).

k-fold cross-validation, performed by leaving one week out at the time ($k = 50$), again selected RW1 as the best model as seen in Table 3.7. We note that the Cor model is not an option in this case as the data are collected in a single year, so we cannot model correlation between weeks across different years. All models considered produced similar estimates of temporal population size, with a similar

model fit. We display the results produced from the RW1 model in the main body of the paper, with the results obtained from the other models in B.6.2 of the Supplementary material.

Table 3.7 Orange-winged Amazon parrots case study. Cross-validation results.

| Model | DP | RW1 | RW2 | AR1 |
|-------|------|------|------|------|
| RMSE | 1283.779 | 1267.940 | 1345.571 | 1277.701 |



(a) Population size                                   (b) GOF
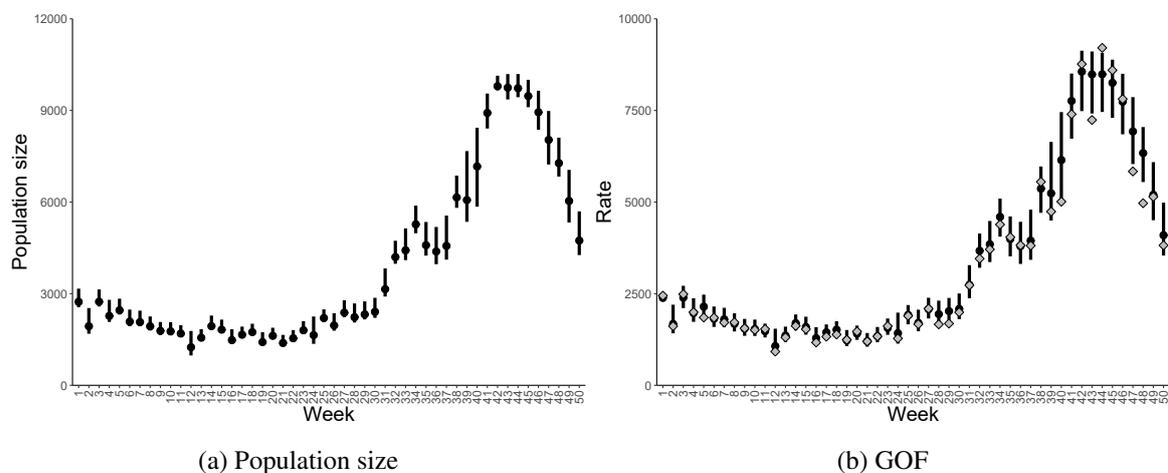
Fig. 3.3 Orange-winged Amazon parrots case study. (a) The black dots represent the posterior mean population size each week and the thick bands represent the corresponding 95% posterior credible interval. (b) The diamonds are the observed weekly rates and the thick bands represent the 95% intervals of simulated weekly rates.

Fig 3.3a shows the posterior summaries of the temporal population size estimates obtained for each week using the RW1 model. The primary factor influencing the fluctuation in population size at the roosting site is the breeding season (De Moura et al., 2010). Consequently, the period of low population size (weeks 1-31) is possibly when paired individuals leave the roost in search of a nest, where they breed, nest, and rear young until the nestlings can fly. This long period of low population size may be due to the asynchronous reproduction of Orange-winged Amazons. The period of high population size (weeks 41-48) corresponds to the return of pairs with young, while the period of medium population size (weeks 32-40 and 49-50) corresponds to the time when individuals start returning with young (weeks 32-40) and when individuals start to disperse (weeks 49-50). Finally, like the Ecuadorian Amazon parrots, we use posterior goodness of fit to assess model fit, defining

weekly rate to be the sum of counts obtained in a week divided by the number of surveys in that particular week. Fig 3.3b suggests that the RW1 model fits the data well as it produced similar weekly rates to the observed rates for the majority of the weeks.

Baseline detection probability was estimated as high (posterior mean = 0.862 with $(0.789, 0.913)$ 95% posterior credible interval), possibly because in this case parrots were counted from a boat by a minimum of three teams of two observers, each team oriented in a different direction. Predictors cloud, rain and time were the only ones with PIP>0.5, but only marginally so $(0.535, 0.511, 0.511,$ respectively), and their coefficients are estimated close to 0. In this case, rain and surveying PM instead of AM decreased the probability of detection (Table 3.8), and we discuss this result and compare it to that obtained for the Ecuadorian Amazon parrots in Section 3.5.

Table 3.8 Orange-winged Amazon parrots case study. Posterior summaries of coefficients for detection probability.

| Coefficient | Mean | SD | 95% PCI |
|---|---|---|---|
| Intercept | 1.830 | 0.279 | (1.320, 2.350) |
| Partially cloudy | 0.007 | 0.049 | $(-0.061, 0.149)$ |
| Cloudy | $-0.019$ | 0.071 | $(-0.268, 0.040)$ |
| Low wind | 0.009 | 0.048 | $(-0.034, 0.176)$ |
| Strong wind | 0.001 | 0.030 | $(-0.059, 0.074)$ |
| Rain-Yes | $-0.016$ | 0.078 | $(-0.288, 0.046)$ |
| Time-PM | $-0.002$ | 0.039 | $(-0.116, 0.096)$ |

## 3.5 Discussion

Roost count surveys are widely used and, for certain populations, are the only viable monitoring tool, as individuals may nest in elevated cavities in trees or cliffs that are difficult to find, reach, and capture (Dénes et al., 2018). In this paper, we have developed a new modelling framework for roost count survey data that accounts for observation error and TE, non-parametrically and parametrically to provide key estimates of population size, information on TE trends, and predictors of detection

via variable selection. All of these estimates can serve as fundamental tools in adaptive wildlife monitoring, conservation, and management.

Moreover, we have performed an extensive simulation study to assess the performance of our novel modelling framework under different scenarios. When the model for detection probability is correctly specified, reliable estimates of population size and patterns of TE are obtained using both the nonparametric and parametric approaches introduced in the paper, even when the probability of detection is low. However, when the model for detection probability is misspecified, which is likely to be the case in practice, our results demonstrate the importance of using a mixed effect model for the probability of detection, so that the random effects part can absorb the lack of fit introduced by omitting important predictors for observation error. Failure to employ a mixed-effects model, in this case, gives rise to highly biased estimates of population size.

We applied our modelling framework to two case studies on parrots. We found substantially different sizes of detection probabilities and variable effects on detection. The observation methods and roost site characteristics for each parrot species can explain in part these differences. Detection probability was much higher for the Orange-winged amazons, which were counted by a team of six people from a boat directly under the flight path between the mainland and an island roost, vastly reducing the chance of missing individuals. Detection however was lower for the Ecuadorian Amazon parrots, which were counted by two people from an observation tower on the mainland, where birds fly over and amongst buildings and human development to patches of scattered mangroves interspersed with aquaculture. Detection probability was higher for the Ecuadorian Amazon parrots when surveyed in the afternoon, whereas for the Orange-winged amazons, they were marginally more detectable during morning surveys. This again corresponds to the observation methods and direction of travel associated with the AM and PM surveys - with both cases showing higher detection probability when observers are at closest proximity to the roost i.e. on the observation tower when birds are departing from the mainland dry forest (Ecuadorian amazon parrots), on the boat when birds are departing from the mangrove roost (Orange-winged amazons). The effect of rain also differed, increasing the detection probability for the Ecuadorian Amazon parrots but decreasing it for the Orange-winged Amazon parrots. This can be attributed to the differing flight path birds have to make, with a 1km flight over a water body not possible in the rain (orange-winged Amazon parrots), thus lowering

detection probability, whereas a shorter flight that can be taken lower down and in shorter stages by landing on trees on the edge of the town close to the observation tower (Ecuadorian Amazon parrots) making birds more detectable.

Similarly, we identified differences in phenology between the two species, with the roost use pattern of Ecuadorian Amazon parrots being described by a two-mixture model, whereas that of Orange-winged Amazon parrots by a three-mixture model, when the DP approach is used to describe TE. This can be due to different levels of population and habitat fragmentation. There was a large difference in the population size between the two species, with the Ecuadorian Amazon parrots being just a few hundred birds, whilst the Orange-winged Amazon parrots population consists of over ten thousand birds. The Ecuadorian Amazon parrots have faced a 60 percent population decline at this roost site in the past two decades, in part attributed to habitat fragmentation, with the feeding, nesting, and roosting areas now occurring amongst a highly transformed landscape on the edges of a large city, vastly different to the relatively undisturbed roosting habitat of the Orange-winged Amazon parrots.

We have demonstrated our new modelling framework on parrot data, but bats and other species are also routinely monitored in the same way. The model can be readily fitted to such data and can be extended to account for data from multiple sites, when these are available, and to account for spatial correlation between sites. Spatial models such as the ICAR and the Besag, York and Mollié (BYM) model (Besag et al., 1991) can be considered to account for spatial correlation.

Variable selection on detection probability via BGLSS performed well when the model is correctly specified or when misspecified and a mixed effect model is used for detection. BGLSS had lower power to identify weaker effects when using a mixed effect model for observation error. Additionally, BGLSS can only identify significant categorical covariates not significant levels of categorical variables. We also considered Bayesian Sparse Group selection (BSGS). BSGS developed by Chen et al. (2016) enables variable selection of both continuous and categorical variables. It has the advantage of identifying both significant categorical covariates and their relative levels. However, results shown in B.3 of the Supplementary material suggest that BGLSS generally outperforms BSGS. Performance of other BVS methods such as the variable selection method of Griffin et al. (2020) can also be investigated in this scenario. Thus, future work can be focused on investigating/improving BVS methods when using a mixed-effect model.

The Beta DP mixture model in this framework enables our model to perform clustering of primary periods independently for top-level primary periods, and hence treats the observations as being from one long time series, with clusters, as a result, independent across top-level primary periods. An alternative would be to implement a hierarchical Dirichlet process (HDP) model (Teh et al., 2004), which allows clusters with the same locations but potentially different weights to be identified across top-level primary periods, providing a natural way to capture dependence between top-level primary periods.

Another direction of future work is model selection. The proposed options for modelling the availability patterns define different, competing models (Table 3.1), for the TE pattern, each with its own advantages. We use the well-established approach of cross-validation to select between competing models. However, cross-validation can be computationally intensive as it requires fitting the model multiple times. Other model selection methods such as the Watanabe-Akaike information criterion (WAIC) (Watanabe and Opper, 2010) only require fitting the model once and can be easily computed using popular software, such as NIMBLE and Stan (Carpenter et al., 2017). Notably, WAIC computation relies on the independence assumption of data given the parameters. This assumption is often violated in temporal models where dependence among the data is a key modelling feature. Hence, future work can be focused on investigating/developing efficient model selection methods for temporally correlated data.

**Chapter 4**

# Modelling disease dynamics from spatially explicit capture-recapture data

# Abstract

One of the main aims of wildlife disease ecology is to identify how disease dynamics vary over space and time and as a function of population density. However, monitoring spatio-temporal and density-dependent disease dynamics in the wild is challenging because the number of individuals, their disease status and their spatial locations are unobservable, or only imperfectly observed. In this paper, we develop a novel spatially explicit capture-recapture (SCR) model motivated by an SCR data set on European badgers (*Meles Meles*), naturally infected with bovine tuberculosis (bTB). Our model accounts for the observation process of individuals as a function of their latent activity centres, and for their imperfectly observed disease state and its effect on demographic rates and behaviour. This framework has the advantage of simultaneously modelling population demographics and disease dynamics within a spatial context. It can therefore give estimates of critical parameters such as population size; local and global density by disease status and hence spatially-explicit disease prevalence; disease transmission probabilities as functions of density; and demographic rates as functions of disease status. Our findings for the badger population suggest that infected individuals have a lower survival probability but a wider home range than uninfected individuals, whereas we find no clear evidence of density-dependence in disease transmission. We also present an extensive simulation study, considering different scenarios of disease transmission within the population, and our findings highlight the importance of accounting for spatial variation in disease transmission and individual disease or general disease states when these affect demographic rates. Therefore, our new model enables a better understanding of how wildlife disease dynamics are linked to population demographics within a spatio-temporal context.

## 4.1  Introduction

Linking host population and disease dynamics is a key aim of wildlife disease ecology. Infectious disease can directly influence host population density through impacts on demographic vital rates (Manlove et al., 2016; McCallum et al., 2007; Vredenburg et al., 2010), but in turn disease dynamics can also vary over space and time and with population density. Quantifying how pathogen transmission varies with population density is important because of its implication for the conservation and management of wildlife populations (McCallum, 2016; Silk et al., 2019). For example, pathogen transmission can vary across a continuum from being greater at higher population densities (density-dependent) to independent of population density (frequency-dependent) (Hopkins et al., 2020). Pathogens with density-dependent transmission are unlikely to drive their hosts to extinction, while this is not true when transmission is instead frequency-dependent (De Castro and Bolker, 2005; McCallum, 2012)).

Despite the value of information on the relationship between infection dynamics, demographic rates, and population density, we still lack detailed knowledge of wildlife populations due to two main challenges in this context. Challenge I is the difficulty of estimating population size, and hence population density, and demographic rates of wild populations where not every individual is captured or known about at different points in time. Challenge II is associated with the study of incidence and prevalence of infectious disease in a host population, often in situations where diagnostic tests vary in their sensitivity and specificity (Choquet et al., 2013; Drewe et al., 2010; Enøe et al., 2000).

Capture-recapture (CR) models have been one of the main tools developed to deal with Challenge I. Traditional capture-recapture (CR) models essentially represent "fish bowl" sampling, that is, a system that is unconnected to the spatial structure of the population. These models do not account for the spatial nature of sampling nor the spatial distribution of individuals (Royle et al., 2018). Consequently, they do not allow for study of many vital spatial processes of the population, such as density, movement and dispersal of individuals. This weakness of CR models has been overcome by the development of spatially-explicit capture-recapture (SCR) models (Borchers and Efford, 2008; Efford, 2004). SCR models are hierarchical models that consider the collection of individuals in a

population as a latent point process of where an individual moves (home range), which is centred on their individual activity centre (AC), distributed within some region of interest.

In SCR data, individual ACs are unknown, and thus are considered latent variables in corresponding models. SCR models can be fitted in a classical framework, where the ACs are marginalised from the likelihood by integration (Borchers and Efford, 2008), or in a Bayesian framework, where the ACs are explicitly estimated along with other unknown parameters and random variables using Markov chain Monte Carlo (MCMC) methods (Royle and Young, 2008). Once inferred, the ACs can be used to estimate spatial population processes such as density, which is the number of ACs per unit area of the region of interest. Additionally, conditional on the latent ACs, the probability of observing or encountering an individual is modelled as a function of the distance between the individual's AC and the location of each trap. Consequently, SCR models take into consideration the spatial nature of sampling as well as spatial distribution of individual ACs to allow for the study of spatial population processes, which is arguably equally important to the study of demographic population rates, with the formal link between state model and observational model allowing for better inference on the former and more robust accommodation of the latter (Sutherland et al., 2019).

However, existing SCR models do not currently deal with Challenge II, and hence do not account for additional data targeting disease status that are collected on each captured individual when aiming to study disease dynamics. These data can be in the form of individual measurements or of test results, and are typically also prone to error, and hence are only an imperfect observation of an individual's status. Additionally, as is always the case in CR data, these individual-level measurements are only available when individuals are caught on a particular occasion, and are missing for all others.

In this paper, we develop a novel SCR model that accounts for the observation process of individuals, tackling Challenge I, as well as their imperfectly observed disease status, hence also tackling Challenge II. Our new modelling framework allows the simultaneous modelling of population demographics and disease dynamics within a spatiotemporal context. This makes it possible to simultaneously test hypotheses related to spatial and density-related variation in disease transmission alongside examining variation in survival and individual detection probabilities as a function of individual (latent) disease state.

We perform an extensive simulation study to assess model performance for a number of scenarios. Our results demonstrate the quality of inference in our proposed model, but also reveal the requirements in terms of data size and effect size to have sufficient power to identify density-dependence in disease transmission. We highlight that, when demographic rates are dependent upon individual disease status, existing SCR models, which do not account for that dependence, yield substantially biased estimates of population density.

When fitted to a motivating case study of European badgers (*Meles meles*), naturally infected with bovine tuberculosis (bTB) at Woodchester Park in Gloucestershire, UK (Delahay et al., 2013; McDonald et al., 2018), our findings highlight that infected individuals have a lower survival probability, but a larger home range. We also infer that population size has been steadily decreasing in recent years but that disease prevalence has remained constant, finding no clear evidence of density-dependence in disease trasmission in the population during this recent period.

The paper is structured as follows: in Section 4.2 we describe the case study that motivated the work in this paper, in Section 4.3 we introduce the new model and discuss our inference approach, while Sections 4.4 and 4.5 present simulation and case study results, respectively. Section 4.6 discusses the results from our simulations and case study in the context of wildlife disease ecology, and suggests directions for future work.

## 4.2   Data Collection and Processing

The Woodchester Park study area is located on the Cotswold limestone escarpment in Gloucestershire, South-west England. Approximately 7 km$^2$ of its area has been used to monitor badgers in a consistent manner since 1981 (Rogers et al., 1998). The majority of this study area comprises of mixed woodland, grassland and arable farmland (Delahay et al., 2006). Badger population density at Woodchester Park is particularly high (McDonald et al., 2018) and main badger setts are relatively regularly spaced throughout the study area.

The badger population is monitored by CR sampling, which enables demographic and epidemiological data collection on the resident badger population. To enable consistent trapping, the study area has been divided into three zones of approximately equal size, and each zone is trapped four times

each year from May to January. Trapping is suspended each year in the spring, February to April inclusive, to avoid catching dependent cubs that cannot be left underground for protracted periods (Woodroffe et al., 2006). To determine which setts are active and how many traps to deploy, a sett activity survey is conducted in each zone before each trapping event. More traps than that which are likely to be needed (i.e. saturation trapping) are used as a general guideline for the number of traps deployed per sett.

Box traps constructed of steel mesh with spring-loaded doors are used as traps. They are dug into the substrate close to each active sett and baited with peanuts for four to eight days to habituate badgers to their presence (Rogers et al., 1997). On the last day of baiting, the traps are set for two consecutive nights and each trap is checked on the following morning. Once captured, newly caught badgers are permanently marked with a unique ID tattoo on the abdomen (Cheeseman et al., 1982).

Once captured, three tests are currently used to test for bovine tuberculosis (bTB): interferon-gamma immunoassay (Ifn, Dalley et al., 2008), used since 2006 to detect a cell-mediated immune response, Dual Path Platform test (DPP®, Chembio.inc), used since 2015 to test for antibodies and M. bovis culture (Cul, Gallagher and Horwill, 1977) used since 1976. Ifn and DPP® use blood samples whilst Cul use samples of sputum, feces, urine, and swabs of abscesses and wounds. Each test is imperfect, resulting in false positive and false negative errors, making it difficult to infer an individual's disease state from the tests alone (Ashford et al., 2020; Drewe et al., 2010).

Following examination and disease diagnostic testing, badgers caught during the first night of trapping are held overnight and released the following morning. This prevents badgers from being re-captured on the second trapping night. Badgers caught during the second night are released the following day. Badgers are released where captured, following a period of recovery and following approval by the Named Animal Care and Welfare Officer (NACWO), Named Veterinary Surgeon (NVS), or another experienced person with delegated authority.

## 4.3 Model

SCR models assume that a population of $i = 1, \ldots, N_t$ individuals are monitored at $t = 1, \ldots, T$ sampling occasions and $j = 1, \ldots, J$ sampling locations, and each individual has an associated spatial

location within a spatial domain ($S$), representing its AC $s_{i,t} = [s_{i,t,x}, s_{i,t,y}]$. The collection of ACs can be thought of as a statistical spatial point pattern that describes how individuals are distributed within $S$. This statistical point process is often referred to as the state model. Here, we define our model in a Bayesian framework using data augmentation (DA, Royle and Dorazio, 2012) and let $i = 1, \ldots, M$ be "pseudo-individuals" that potentially could belong to $N_t$. In what follows, $i = 1, \ldots, M$ indexes individuals, $t = 1, \ldots, T$ sampling occasions and $j = 1 \ldots, J$ sampling locations.

Our model has two key latent states: presence, $z_{i,t}$, and disease status, $d_{i,t}$, defined as

$$z_{i,t} = \begin{cases} 1 & \text{alive} \\ 0 & \text{unrecruited/dead} \end{cases}$$

$$d_{i,t} | z_{i,t=1} = \begin{cases} 1 & \text{infected and alive} \\ 0 & \text{uninfected and alive.} \end{cases}$$

We assume that individual ACs do not change over time by modelling

$$s_{it} = s_i \sim \text{Uniform}(S) \; \forall i, t$$

Naturally, observation $y_{i,j,t}$ (equal to 1 if individual $i$ was caught on sampling occasion $t$ and sampling location $j$ and 0 otherwise) depends on the corresponding presence state and AC

$$y_{i,j,t} | z_{i,t} \sim \text{Bernoulli}(p(x_j, s_i) z_{i,t})$$

where here $p(x_j, s_i)$, which corresponds to the probability that individual $i$ is caught on sampling occasion $t$ and sampling location $j$, conditional on being present, is modelled by the half-normal function (Efford, 2004)

$$p(x_j, s_i) = p_{0_{d_{i,t}}} \exp\left( -\frac{1}{2\sigma_{d_{i,t}}^2} \| x_j - s_i \|^2 \right)$$

where $p_{0_{d_{i,t}}}$ is the baseline encounter probability and $\sigma_{d_{i,t}}$ represents the rate at which detection probability declines as Euclidean distance from the AC increases. We model both of these parameters dependent on the individual disease status at occasion $t$, allowing disease status to potentially affect behaviour in terms of space use.

Finally, we model the result of test $Q$, $\omega_{i,t}^Q$, for individual $i$ on occasion $t$, conditional on their disease status, as a Bernoulli($\omega_{i,t}^Q$) random variable, with

$$\omega_{i,t}^Q | d_{i,t} = \begin{cases} 1 - q_{00}^Q, & d_{i,t} = 0 \\ q_{11}^Q, & d_{i,t} = 1 \end{cases}$$

where we refer to the probability of a true positive result by test $Q$ as $q_{11}^Q$ (sensitivity of test $Q$) and to the corresponding probability of a true negative result as $q_{00}^Q$ (specificity of test $Q$), and $Q \in \{\text{DPP}^\circledR, \text{Ifn}, \text{Cul}\}$. Following Buzdugan et al. (2017), we assume independence between tests and hence define the joint distribution of the three test results as the product of the marginal Bernoulli distributions. The sensitivity and specificity of each test are inferred parameters, thus, enabling the diagnostic accuracy of each test to be evaluated. This formulation accounts for imperfect tests and enables a higher diagnostic accuracy than single test use (Drewe et al., 2010).

We model the transition between latent states accordingly, so that at $t = 1$

$$z_{i,1} \sim \text{Bernoulli}(\gamma_1)$$

$$d_{i,1} | z_{i,1} \sim \text{Bernoulli}(z_{i,1} \delta_I)$$

where $\gamma_1$ is the recruitment probability that a "pseudo-individual" is in the population at the start of the study and $\delta_I$ is the probability of being infected at the start of the study. For $t \geq 2$,

$$z_{i,t} \sim \text{Bernoulli}(\phi_{d_{i,t-1}} z_{i,t-1} + \gamma_t \alpha_{i,t})$$

$$d_{i,t} \sim \text{Bernoulli}(z_{i,t} [d_{i,t-1} + \{(1 - d_{i,t-1}) \psi_{i,t-1}\}])$$

where $\phi_{d_{i,t-1}}$ is the probability of survival from occasion $t - 1$ to $t$ conditional on disease status on occasion $t - 1$ for individual $i$, $\psi_{i,t}$ is the disease transmission probability, that is the probability that

an individual that is uninfected on occasion $t-1$ becomes infected by occasion $t$, $\gamma_t$ is the recruitment probability that a "pseudo-individual" is first recruited, and hence is first available for capture, on occasion $t$ and $\alpha_{i,t}$ is a latent indicator variable of whether an individual is available to be recruited or not on occasion $t$. We define $\alpha_{i,t} = \left(1 - \mathrm{I}\left(\sum_{t=1}^{t-1}(z_{i,t}) > 0\right)\right)$ such that $\alpha_{i,t} = 1$ if individual $i$ is available to be recruited on occasion $t$, $\alpha_{i,t} = 0$ otherwise to ensure an individual can only be recruited once.

We note that, clearly, only individuals that are alive and uninfected can become infected and, as is the case in our motivating data, once infected, individuals cannot become uninfected. To investigate the relationship between density and disease transmission, we model $\psi_{i,t}$ as a function of population density on each occasion. We discretize the study space using a grid, and create $R$ non-overlapping habitat cells. We denote the cell in which individual AC $i$ falls by $c_i$, with $c_i \in \{1,\ldots,R\}$. Local density of grid cell $r$, $r = 1,\ldots,R$, on occasion $t$ is defined as $\ell_{r,t} = \sum_{i=1}^M I(z_{i,t} = 1, c_i = r)$, where $I(z_{i,t} = 1, c_i = r)$ is an indicator variable equal to 1 if individual $i$ is alive and its AC falls within cell $r$, and 0 otherwise. Thus, we build a logistic regression model for the probability of disease transmission

$$\mathrm{logit}(\psi_{i,t}) = \beta_0 + \beta_1 \ell_{c_i,t} \tag{4.3.1}$$

such that an uninfected individual can become infected due to its local density $(\ell_{c_i,t})$ at the individual AC location in its habitat cell. The coefficient, $\beta_1$ determines the direction and size of the effect of local density on disease transition probability.

Finally, population size on occasion $t$, $N_t$, can be estimated as $N_t = \sum_i z_{i,t}$ and population density on occasion $t$, $D_t$, as $D_t = N_t/\mathrm{area}(S)$, while the corresponding sizes of the infected population $(N_t^{\mathrm{i}})$ and of the uninfected population $(N_t^{\mathrm{u}})$ can be estimated as $N_t^{\mathrm{i}} = \sum_i z_{i,t} d_{i,t}$ and $N_t^{\mathrm{u}} = \sum_i z_{i,t}(1 - d_{i,t})$, respectively. Hence, disease prevalence $(D_t^{\mathrm{i}})$ can be estimated as $D_t^{\mathrm{i}} = N_t^{\mathrm{i}}/N_t$. Density estimates for infected/uninfected individuals can be easily computed as well as realized disease density maps.

We fit models in a Bayesian framework using MCMC methods via R package NIMBLE (de Valpine et al., 2017) version 0.13.0. Additionally, to increase the computational efficiency of using a Bayesian implementation via DA, we skip unnecessary calculations, vectorized computation and performed block sampling on correlation parameters (Turek et al., 2021) when appropriate. We employ user-

defined NIMBLE functions to reduce the total number of nodes in the model and improve MCMC efficiency. We use the R package nimbleSCR (Bischof et al., 2020) version 0.2.1 to create habitat grids and for the computation of local density. To improve convergence and mixing, we use a coarse habitat grid to provide the model with a large number of latent density points to serve as a covariate on disease transmission probability. We also center latent density to improve computation by reducing the correlation between the intercept and fixed effect. Random walk block samplers are assigned to $(q_{11}^{\text{DPP} \circledR}, q_{11}^{\text{Ifn}}, q_{11}^{\text{Cul}})$, $(q_{00}^{\text{DPP} \circledR}, q_{00}^{\text{Ifn}}, q_{00}^{\text{Cul}})$ and $(p_{0d_i}, \sigma_{d_i})$ to improve MCMC efficiency.

## 4.4   Simulation Study

We performed a simulation study to assess the performance of the proposed modelling framework in estimating population density and all other model parameters, as well as the impact on estimation when the effect of (local) population density on transmission probability is ignored and, more importantly, when density dependent disease transmission, disease status and its effect on other model parameters are ignored altogether.

We refer to our proposed model as $M(\psi_\ell)$, to the model that does not account for density-dependence in $\psi$ as model $M(\psi_0)$ and to the standard open SCR model that does not account for density dependence and disease status as model $M(\text{SRC0})$.

We investigate model performance at both high and low population levels, $M = (1000, 500)$, given high and low density effects on disease transmission, $\beta_1 = (0.25, 0.1)$. We set $T = 8$ and the rest of the parameter values are set as: $\phi_{d_i=0} = 0.9$, $\phi_{d_i=1} = 0.8$, $\gamma_1 = 0.4$, $\gamma_2 = \gamma_{2:4} = 0.1$, $\gamma_3 = \gamma_5 = 0.2$, $\gamma_4 = \gamma_{6:T} = 0.15$, $p_{0d_i=0} = 0.5$, $p_{0d_i=1} = 0.2$, $\sigma_{d_i=0} = 0.5$, $\sigma_{d_i=1} = 1$, $\delta_I = 0.15$, $\beta_0 = -2.5$, $q_{11}^{\text{DPP}} = 0.492$, $q_{11}^{\text{Ifn}} = 0.809$, $q_{11}^{\text{Cul}} = 0.1$, $q_{00}^{\text{DPP}} = 0.931$, $q_{00}^{\text{Ifn}} = 0.936$ and $q_{00}^{\text{Cul}} = 0.999$. This setting results in mean$(N_{1:T}) \approx (400, 200)$ for high and low density, respectively.

We use an $11 \times 11$ habitat grid in which we center a $7 \times 7$ trapping grid resulting in a buffer width of 2 distance units. For each case, we perform 10 simulation runs and use relative bias $\left(\text{RB} = \frac{\bar{\theta} - \theta}{\theta}\right)$ to measure relative error and coefficient of variation $\left(\text{CV} = \frac{\text{SD}(\hat{\theta})}{|\bar{\theta}|}\right)$ to measure relative precision, where $\theta$ is the true parameter value, $\bar{\theta}$ is the mean and $\text{SD}(\hat{\theta})$ is the standard deviation of the posterior distribution obtained, across the 10 runs.

From Tables C.1 and C.2 in the Supplementary material and Fig. 4.1, we can see that our proposed modelling approach performs well in estimating all demographic and disease-dynamics related parameters. Estimation of the density effect ($\beta_1$) is possible but is more challenging than of other parameters, as also reported by Milleret et al. (2023). As expected, the quality of inference is best in larger populations and when density effects are high. When the population size is low and/or the true value of $\beta_1$ is low, estimation of $\beta_1$ is computationally challenging, requiring more MCMC samples to reach convergence and to obtain sufficient effective sample sizes. Consequently, this simulation study demonstrates our model's ability to simultaneously model population demographics and disease dynamics within a spatio-temporal context.



(a) RB                    (b) CV
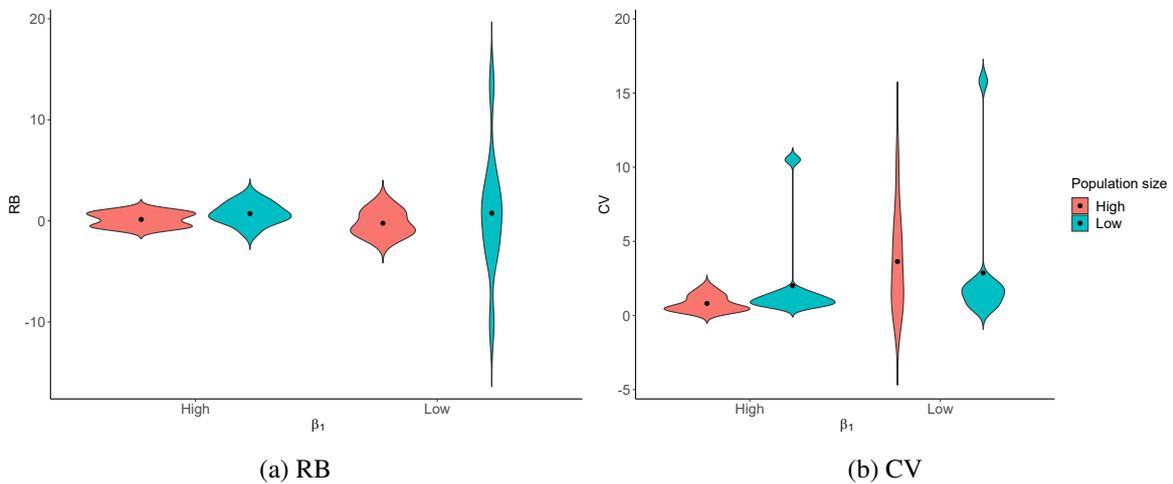
Fig. 4.1 Violin plots of RB, (a), and CV, (b), for $\beta_1$ when using our proposed model, M($\psi_\ell$). Dots represent the median in each case.

From Figs. 4.2 and 4.3, it can be seen that the M(SCR0) model underestimates population size for the first half of the study period, but tends to overestimate it at the end in all cases. This result highlights the need to account for the spatial variation in disease transmission and the disease status of individuals when that is linked to demographic parameters and/or space use. All three models had similar CV for $N$. The M($\psi_0$) model shows a small RB for $N$ in some cases, but overall has a similar performance to model M($\psi_\ell$). As shown in the Supplementary material Fig. C.1, we simulated the density values, which serve as the latent covariate in model M($\psi_\ell$), using a realistic scenario of small and gradual changes over time, with activity centres simulated from a homogeneous Poisson process, as is the standard assumption of SCR models, including the one in this paper. As a result, especially

when the true value for $\beta_1$ is low, density does not vary dramatically between sampling occasions or between grid cells, and hence, ignoring its effect in the $M(\psi_0)$ model does not lead to substantial bias in the estimation of $N$.



(a) High Density RB

(b) High Density CV

(c) Low Density RB
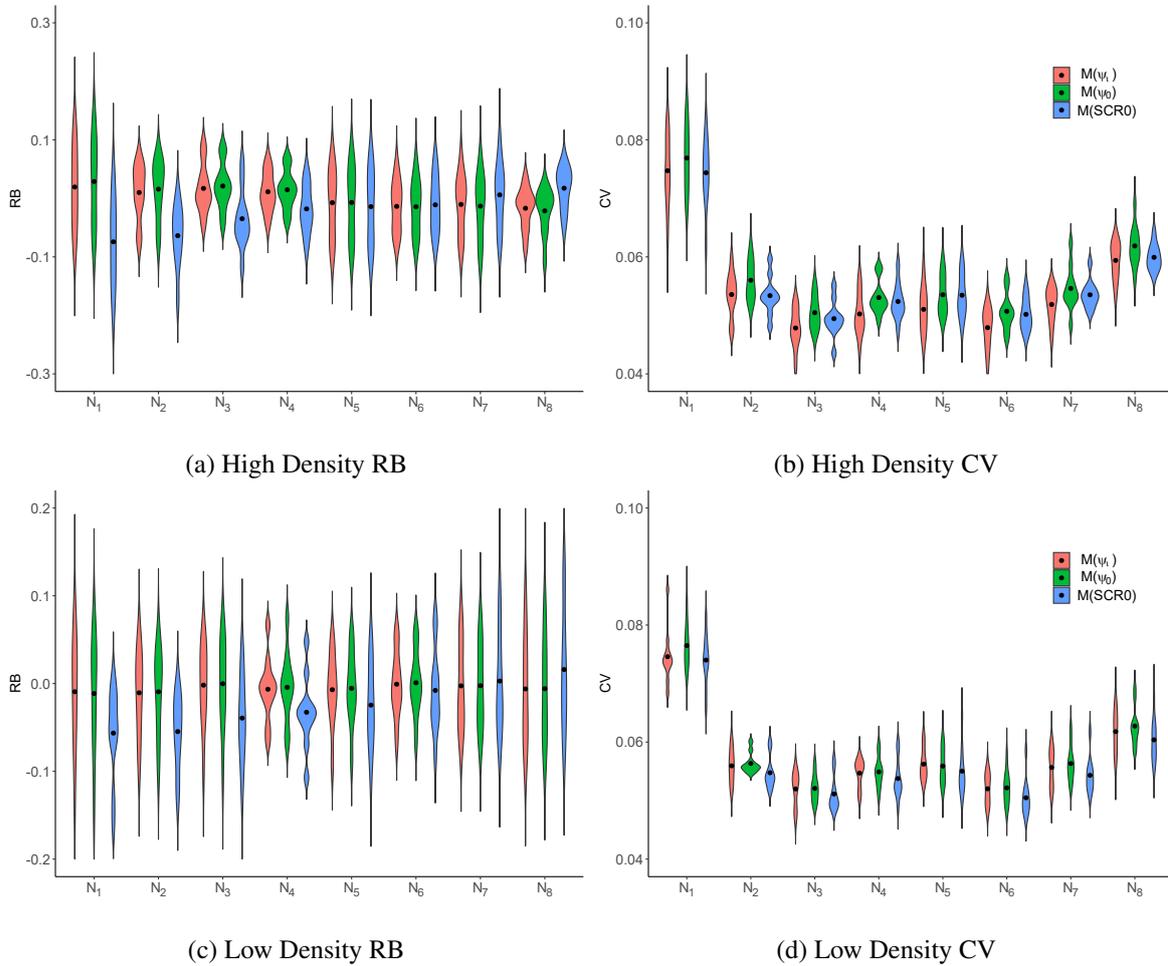
(d) Low Density CV

Fig. 4.2 RB and CV of population size, $N$, at high population size data ($N \approx 400$), high and low density effect levels ($\beta_1 = 0.25$, and $\beta_1 = 0.1$, respectively) for three models: our proposed model, $M(\psi_\ell)$, the model that does not account for density-dependence in disease transmission, $M(\psi_0)$ and the model that does not account for disease status, M(SCR0).

(a) High Density RB

(b) High Density CV



(c) Low Density RB
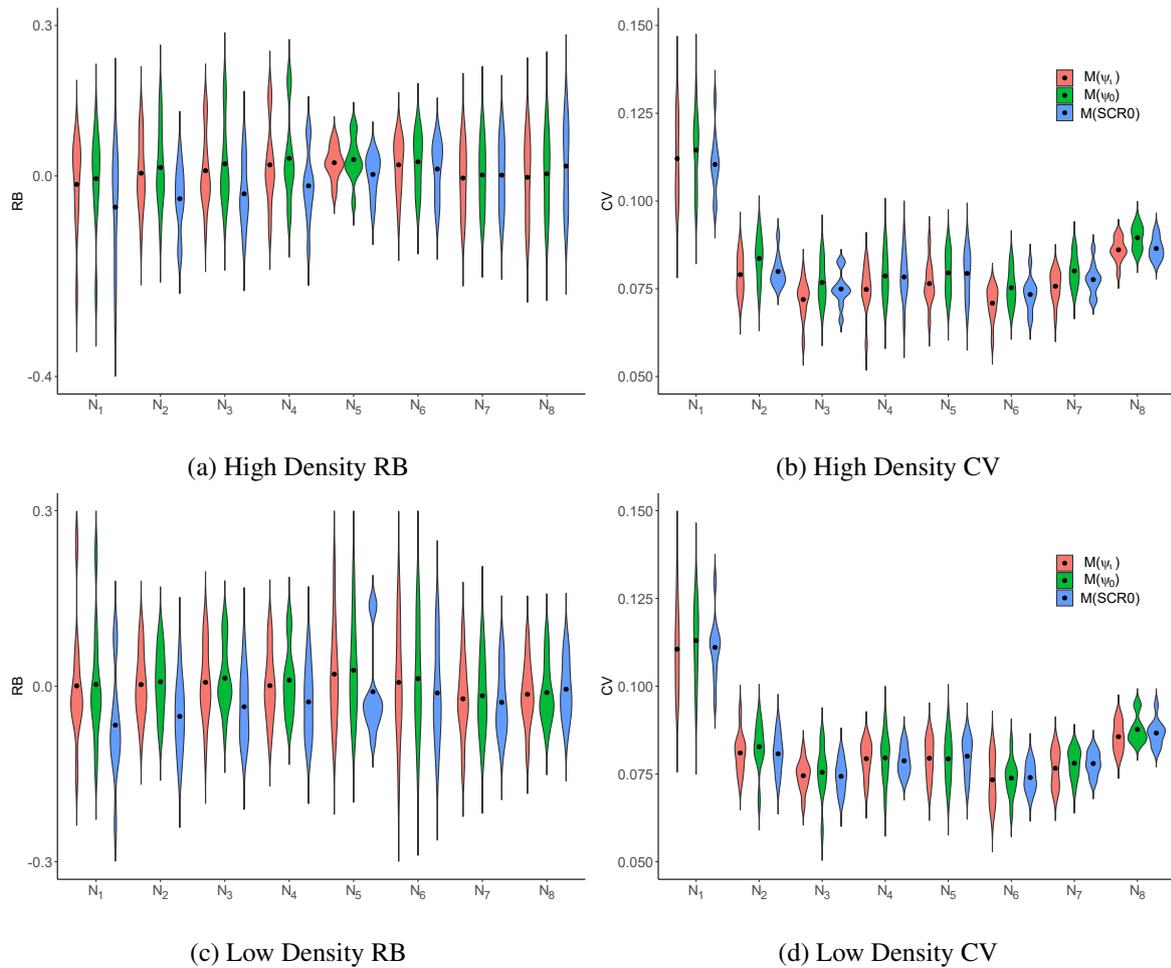
(d) Low Density CV

Fig. 4.3 RB and CV of population size, $N$, at low population size data ($N \approx 200$), high and low density effect levels ($\beta_1 = 0.25$, and $\beta_1 = 0.1$, respectively) for three models: our proposed model, $M(\psi_\ell)$, the model that does not account for density-dependence in disease transmission, $M(\psi_0)$ and the model that does not account for disease status, M(SCR0).

## 4.5   Case Study

We analyse SCR data from the badger case study from 2014-2018 using the modelling framework defined in Section 4.3. These years were selected as they correspond to a period when the population was undisturbed by management interventions and the types of tests for TB employed did not change, as opposed to the pre-2014 years where different tests were employed and the post-2018 years where management interventions started to take place in the surrounding area. We use an $11 \times 8$ habitat grid in which we place a $7 \times 7$ trapping grid. Prior settings are provided in C.2.2 of the Supplementary material where informative priors for the sensitivity and specificity of each test were used by taking information from Drewe et al. (2010) and Ashford et al. (2020).

Table 4.1 displays some of the posterior estimates obtained. Caterpillar plots of these parameter posterior summaries are also shown in C.2.3 of the Supplementary material. All parameters converged according to Gelman and Rubin's convergence diagnostic (Gelman and Rubin, 1992), displayed good mixing, and had effective sample size (ESS) $\geq 500$, with the exception of the coefficients in the model for disease transmission (ESS $\approx 100$).

Our results indicate that during the study period (2014-2018) the Woodchester badger population was in decline (Fig. 4.4a), with both the number of uninfected (Fig. 4.4b) and infected individuals (Fig. 4.4c) decreasing over the course of the study. Disease prevalence during this period remained relatively stable, albeit with some weak evidence for a decline from approximately 20% to closer to 15% (Fig. 4.4d).

Model results also confirmed known differences between infected and uninfected badgers in their behaviour and survival. There was a significant difference between baseline detection and scale parameter for uninfected and infected individuals as shown in Table 4.1. That is, infected individuals are less likely to be caught at their activity center and have a larger home range area than uninfected individuals. Specifically, assuming a circular home range area, the effective home range area for an uninfected badger was found to be $0.334 \, \text{km}^2$ ( 95% PCI : $(0.297 \, \text{km}^2, 0.377 \, \text{km}^2)$) while the effective home range area for an infected badger was found to be $1.586 \, \text{km}^2$ ( 95% PCI : $(1.234 \, \text{km}^2, 2.08 \, \text{km}^2)$). Infected individuals also had a lower survival probability ($\phi$) than uninfected individuals (Table 4.1), with the lack of overlap of the 95% PCIs indicating that these differences are statistically significant.

Table 4.1 Case study. Posterior summaries of model parameters.

| Parameters | Mean | St.Dev | 95% PCI |
|---|---|---|---|
| $\gamma_1$ | 0.329 | 0.028 | $(0.274, 0.384)$ |
| $\gamma_2$ | 0.011 | 0.008 | $(0.001, 0.031)$ |
| $\gamma_3$ | 0.115 | 0.029 | $(0.062, 0.173)$ |
| $\gamma_4$ | 0.006 | 0.005 | $(0.002, 0.020)$ |
| $\gamma_5$ | 0.066 | 0.027 | $(0.020, 0.123)$ |
| $\gamma_6$ | 0.030 | 0.013 | $(0.009, 0.057)$ |
| $\gamma_7$ | 0.134 | 0.037 | $(0.069, 0.217)$ |
| $\gamma_8$ | 0.010 | 0.009 | $(0.002, 0.035)$ |
| | | | |
| $\phi_{d_i=0}$ | 0.901 | 0.012 | $(0.877, 0.925)$ |
| $\phi_{d_i=1}$ | 0.811 | 0.029 | $(0.754, 0.867)$ |
| | | | |
| $p0_{d_i=0}$ | 0.782 | 0.048 | $(0.693, 0.878)$ |
| $p0_{d_i=1}$ | 0.185 | 0.029 | $(0.134, 0.248)$ |
| $\sigma_{d_i=0}$ | 0.267 | 0.008 | $(0.251, 0.283)$ |
| $\sigma_{d_i=1}$ | 0.581 | 0.038 | $(0.512, 0.664)$ |
| | | | |
| $\delta$ | 0.210 | 0.043 | $(0.134, 0.298)$ |
| $\beta_0$ | $-3.712$ | 0.573 | $(-5.474, -2.925)$ |
| $\beta_1$ | 0.065 | 0.195 | $(-0.286, 0.508)$ |
| | | | |
| $q_{11}^{\text{DPP}}$ | 0.520 | 0.026 | $(0.471, 0.573)$ |
| $q_{11}^{\text{Ifn}}$ | 0.656 | 0.038 | $(0.587, 0.730)$ |
| $q_{11}^{\text{Cul}}$ | 0.182 | 0.031 | $(0.126, 0.243)$ |
| $q_{00}^{\text{DPP}}$ | 0.978 | 0.007 | $(0.961, 0.990)$ |
| $q_{00}^{\text{Ifn}}$ | 0.907 | 0.013 | $(0.879, 0.931)$ |
| $q_{00}^{\text{Cul}}$ | 0.991 | 0.005 | $(0.978, 0.998)$ |

Overall, the survival probability of infected individuals was approximately 10% lower than uninfected individuals, albeit with more error around this estimate - potentially caused by the smaller sample size of infected individuals or greater variability in their survival. The sensitivity ($q_{11}$) and specificity ($q_{00}$) estimates of each test also provide valuable information on test performance. Specifically, Cul was found to have low sensitivity (18.2%) but had the best specificity (99.1%), Ifn showed the best sensitivity (65.6%) and good specificity (90.7%), and DPP had good sensitivity (52.0%) and very high specificity (97.8%). These estimates are similar to those obtained by Ashford et al. (2020) for DPP and Drewe et al. (2010) for Ifn and Cul and reflect known differences in the performance and

purposes of the tests used (e.g. Ifn detecting an initial response to infection while Cul only detects only infectious individuals).

Fig. 4.5 displays the population density maps for infected and uninfected individuals across years at the first sampling occasion (spring). These plots are standardized across years with the black dots representing the setts trapped. These outputs reveal spatio-temporal variation in the density of uninfected and infected individuals across the population. High densities of infected badgers were concentrated in the central (and northern) and western area of the study site at the start of our study period, becoming more diffuse over time. The eastern parts of the study site maintained consistently higher densities of uninfected badgers throughout the period.

Finally, density is estimated to have a weak positive effect on the probability of disease transmission that is not clearly different from zero (wide 95% PCI that includes 0). Naively this could be interpreted as transmission being independent of local population density during our study period. However, as discussed in Section 4.4, density is a latent variable with an unknown effect, and hence the power to detect small effects relies heavily on the number of sampling occasions and the number of individuals. Consequently, for now we can interpret this finding as indicating that strong density-dependence of transmission is highly unlikely, but that transmission could instead either be weakly density-dependent or close to frequency-dependent.

(a) Total population size

(b) Population size of uninfected individuals

(c) Population size of infected individuals

(d) Disease prevalence
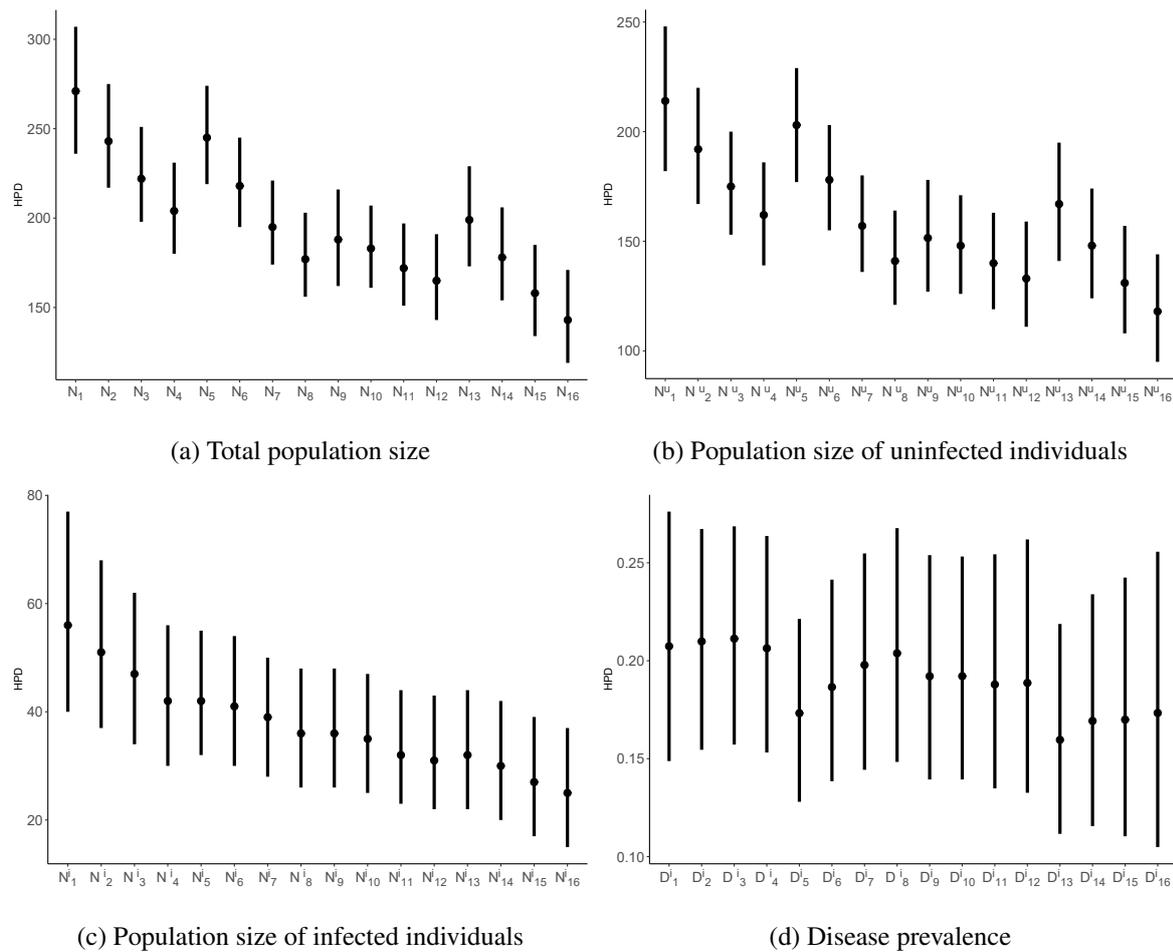
Fig. 4.4 Caterpillar plots of posterior samples of total (a), uninfected, (b), infected, (c) population size and, (d), disease prevalence at each sampling occasion. The black dot in each case represents the posterior mean and the bands represent the 95% posterior credible interval.
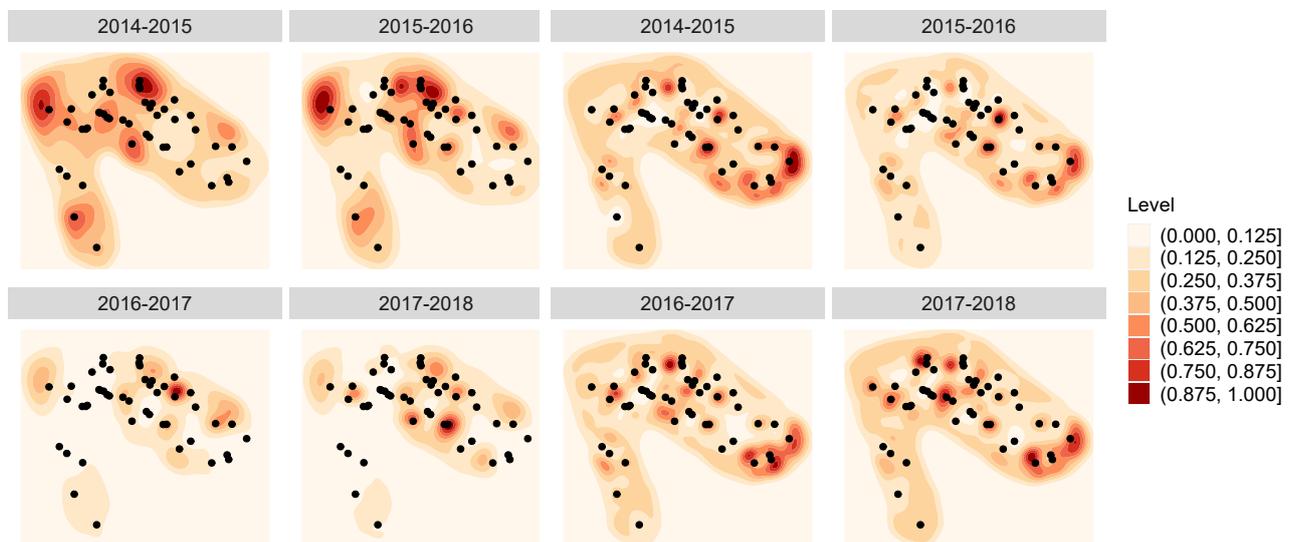
(a) Infected Density Maps                    (b) Uninfected Density Maps

Fig. 4.5 Standardized Density maps for infected, (a), and uninfected, (b), individuals across years in Spring each year. Black dots represent the setts trapped and higher level values indicate higher density and vice-versa.

## 4.6   Discussion

We have developed a novel SCR model that uses disease data from multiple imperfect tests together with SCR data to simultaneously model population demographics and disease dynamics within a spatiotemporal context. Accounting for observation error in both the individual detection process and the disease testing process, our modelling approach accounts for spatial variation in survival and individual detection probabilities as a function of individual (latent) disease state as well as variation in disease transmission as a function of population density. This allows for a better understanding of disease dynamics related to population demographics in a spatiotemporal context at the individual level.

We also conducted an extensive simulation study to assess model performance for a number of scenarios. Our simulations generated encouraging results for our modelling approach and highlighted that, if spatial variation in disease transmission and heterogeneity in demographic rates (detection and survival) induced by individual disease status are not accounted for, biased estimates of population density can be produced. Notably, there are existing models that use finite mixtures to model heterogeneity in these demographic rates (Pledger et al., 2010). We have not considered these models but it is likely that such models might return similar inference on population density to our proposed model. However, such models do not provide information on density-dependent disease transmission and information on individual disease status effect on such demographic parameters.

Applying this new model to a dataset on European badgers, naturally infected with bTB, our model provided novel insights into badger-bTB ecology with broader implications for wildlife disease ecology in general. Our model results agreed closely with previous findings from the Woodchester Park study system. Estimates of population size align with those from the long-term study (Delahay et al., 2013; McDonald et al., 2018) and support the idea that the population has recently declined. Our estimate of home range area for individual badgers is also similar to previous studies from Woodchester Park (Tuyttens et al., 2000). Finally, our estimates of disease prevalence and incidence, as well as disease-associated changes in survival are similar to those found in previous work in this study population (prevalence: Delahay et al. (2013); incidence and changes in survival: Graham et al.

(2013). However, the approach considered in this paper is the first to simultaneously model all these processes from the available SCR data.

By separately estimating the density of uninfected and infected badgers in the population, our SCR approach provides an intuitive approach to analyse spatiotemporal variation in bTB epidemiology in the population while accounting for uncertainty generated from the use of capture-recapture data (i.e. imperfect detection of badgers) and limitations in diagnostic testing data (i.e. imperfect knowledge of disease state). The local density maps generated (Fig. 4.5) can provide a useful tool for stakeholders keen on identifying hotspots of disease (i.e. areas with a high prevalence of infected individuals) or guiding surveillance (e.g. by revealing areas with rapidly increasing or decreasing prevalence). Consequently, our study highlights the value of integrating disease status within an SCR framework for applied disease ecology more generally.

Another key advantage of our modelling framework is that by using longitudinal diagnostic test results to infer the (unobservable) disease status of an individual we are able to gain insight into how the movement of individuals changes on infection. Specifically, we showed that infected badgers ranged over larger areas than uninfected badgers. Previous research has detected a tendency for test-positive individuals (i.e. those likely to be infected) to make greater use of outlying setts (Weber et al., 2013a) and more between-group contacts (Silk et al., 2018; Weber et al., 2013b), which are both traits expected to be linked with greater ranging behaviour. The tendency for infected badgers to start ranging further likely has important implications for the epidemiology of the badger-*Mycobacterium bovis* host-pathogen system. Due to the modular nature of badger contact networks (Rozins et al., 2018) movements between groups offer important opportunities for transmission between groups that enable wider pathogen spread. Therefore, changes in the behaviour of infected badgers could play an important role in the longer-term persistence of the disease. Previous research in the Woodchester system has revealed a positive association between new individuals arriving in a group and the incidence of disease (Vicente et al., 2007). Infected badgers ranging over larger areas than uninfected badgers which is larger than typically social group territory size would provide a mechanism to explain these findings; it is disproportionately likely that an individual moving between groups is infected relative to the overall population.

Our results also indicate no clear relationship between local population density and the incidence of infection in the Woodchester Park badger population. Historically, many infections spread by (non-sexual) close contact were assumed to display density-dependent (as opposed to frequency-dependent) transmission, and this principle of density-dependence underlies multiple interventions in wildlife disease management (McCallum, 2016). More recently, studies have more commonly considered a continuum between frequency-dependent and density-dependent transmission driven by changes in individual behaviour (Hopkins et al., 2020). For example, this result suggested that it may be best to consider the transmission of many infectious diseases including tuberculosis to be a function of population density at low population densities (i.e. density-dependent) but independent of it (i.e. frequency-dependent) at high population densities (Hu et al., 2013). Our results fit nicely with this general pattern given that the Woodchester Park badger population is high density compared with badger populations in other areas. At this population scale, it is likely that the social structure of the population plays an important role and it would be valuable for future research to focus on this question at finer social and spatial scales.

The choice of grid size is a crucial factor in this modelling framework as it can impact accuracy and computational efficiency. A smaller grid size provides finer resolution, capturing intricate details and small-scale patterns in density. However, using a smaller grid size comes at the cost of increased computational complexity and memory requirements. On the other hand, a larger grid size provides a coarser resolution that can overlook smaller-scale patterns but reduces the computational burden. Thus, it is important to strike a balance between accuracy and computational efficiency when selecting grid size. To achieve this, a sensitivity analysis can be carried out. Varying the grid size helps determine the most appropriate grid size and also helps ensure that the chosen grid size does not unduly influence the results, while at the same time since disease transmission is dependent on latent density, the grid size needs to be chosen such that there are adequate latent density points to serve as a covariate on disease transmission.

One caveat to our model is that we assumed individual activity centres are independent and do not change over the period of study. This is due to limited badger movement as highlighted by Rogers et al. (1998) and in C.2.1 of the Supplementary material. However, this assumption will be violated for species that change activity centres frequently. Our modelling approach can be extended

to accommodate such movement by using different state models such as the independent (Royle et al., 2014) and the Markovian random walk models (Raabe et al., 2014). Equally important, this assumption assumes local density, and in turn, disease transmission probability does not change dramatically between grids and sampling occasions. Thus, the state model definition plays a big role in modelling disease transmission. Other populations can exhibit attraction or repulsion, which leads to substantial changes in local density. However, this violates the assumption of independence between individuals and hence the point pattern of ACs being described by a homogeneous Poisson process. In these cases, our model would need to be extended, using models such as the repulsion (Diana et al., 2022) and attraction (McLaughlin and Bar, 2021) models.

Another avenue for future work is the modelling of disease transmission. Here we assumed that once infected an individual remained so for the rest of its life, consistent with the normal approach when modelling *Mycobacterium bovis* transmission in badgers. However, in other cases it may be important to introduce further states to the disease model, such as a recovered state to represent individuals that have been infected and are now immune to re-infection or a vaccinated state for populations currently experiencing management interventions. Such multistate disease models have been fitted to capture-recapture data (e.g. Marescot et al. (2018)), and could be easily incorporated within our modelling framework.

It could also be possible to vary how disease transmission probability is associated with the spatiotemporal distribution of infected and uninfected individuals. In this paper, we have used a logistic regression to investigate the effect of (latent) local density on disease transmission probability and induce heterogeneity. However, other latent variables of this type could also be considered. For example, we roughly introduce the idea of two alternative models: the half-normal model and the overlap model. The half-normal model assumes that an uninfected individual is more likely to become infected the closer it is to infected individuals. The overlap model considers disease transmission probability as a function of the overlap of home range areas between an uninfected individual and surrounding infected individuals. Thus, future work can be done to investigate these models.

In conclusion, our SCR model provides a novel tool to investigate the relationship between population demographics, spatial behaviour, and infectious disease dynamics in imperfectly sampled systems. By applying it in new contexts, it will be possible to gain valuable insight into how spatial

behaviour and pathogen epidemiology are interwoven with important implications for wildlife disease ecology and management.

# Chapter 5

# Discussion

Three manuscripts, described in Chapters 2, 3 and 4, have been presented in this thesis. In this chapter, the work presented in the thesis is discussed and ideas for future work, in some cases with preliminary results, are presented. The chapter closes with a short conclusion.

## Chapter 2: Specifying and selecting N-mixture models in a Bayesian framework.

Chapter 2 presented an evaluation of N-mixture models in a Bayesian framework, specifically prior specification and model selection. We implemented and tested a novel objective prior that is proper, the OB prior. This objective prior performed similarly in terms of inference to popular approximations of the Jeffreys priors. Importantly, as is well known in the classical setting (Dennis et al., 2015), we observed very large estimates of expected population size when the detection probability is small. However, in this case, we also found that expected population size can be severely underestimated when using priors that are concentrated at zero with a long tail, a finding we believe to be previously largely unknown.

We further investigated model selection via WAIC on an extensive class of N-mixture models. We considered both the conditional and marginal WAIC criteria, advocating the use of the marginal WAIC to select between N-mixture models as it was observed that conditional WAIC can lead to misleading results that favour the complicated models rather than the true, while the marginal WAIC selected

the true model with a high probability. To improve this research, the stability of both conditional and marginal WAIC can be investigated using guidelines by Vehtari et al. (2017). WAIC sensitivity to prior choice in N-mixture models can also be pursued, as Ariyo et al. (2022b) highlighted WAIC prior sensitivity in Bayesian linear mixed models for longitudinal data.

The identifiability of N-mixture models in a Bayesian framework is another important avenue for future work. Non-identifiability is the scenario where models can be fitted to data without all model parameters being estimable. Identifiability issues have been found with N-mixture models in a classical setting. Dennis et al. (2015) showed that when the probability of detection and the number of sampling occasions are small, infinite estimates of population size can be obtained. Barker et al. (2018) highlighted that compared to capture-recapture surveys, the loss of individual information resulting from count surveys is critical and causes problems in estimated parameters in Binomial N-mixture models. Kéry (2018) responded to some of these problems of parameter identifiability in a classical framework and called for more research to be done on the parameter identifiability of N-mixture models.

Thus, we investigated parameter identifiability of the set of N-mixture models considered in this thesis using data cloning (DC) (Lele et al., 2007). DC is a statistical computing method introduced by Lele et al. (2007). Cloning the data $K$ times, DC takes advantage of the computational simplicity of the MCMC algorithms that are used in a Bayesian framework to provide maximum likelihood point estimates and their standard errors for complex hierarchical models. Importantly, Lele et al. (2010) proved that for estimable parameters in the model, the scale posterior variance should be approximately $1/K$. If parameters do not follow this trend then parameters are non-identifiable. This is primarily a method of detecting extrinsic parameter identifiability, that is, this method is used to detect parameter identifiability for a specific data set.

Consequently, an important component in using DC to investigate parameter identifiability is the choice of $K$. Ponciano et al. (2012) showed that if parameters are weakly estimable, a large number of clones is needed as the parameters mean and variance may increase at the beginning but as the number of clones increases, the variance will converge to zero. Parameters that are weakly estimable produce likelihoods that are relatively flat resulting in parameter estimation with large variance.

To determine whether DC can be used to assess parameter identifiability in the P-B N-mixture model we compare DC to the covariance diagnostic proposed by Dennis et al. (2015). We simulate data with $p = 0.1$, $\lambda = 5$, $M = 20$, $J = 3$ and select data sets such that the P-B model is identifiable for 10 data sets ("identifiable cases") and non-identifiable for 10 data sets ("non-identifiable case") according to the covariance diagnostic. At the same time, we also investigate the prior effects on the performance of DC. Three types of priors were investigated: the OB prior, an approximation to Jeffreys prior (Gamma(0.5,0.00001)), and an informative prior (Gamma(5,1)). For $p$, a Uniform(0,1) prior was assigned. We focus on the identifiability of $\lambda$ and set $K = 10$. For $K = 1$, we run 505000 MCMC iterations with a burn-in of 40000 and thinning of 5 for 2 chains. For $K \geq 1$, we run 705000 MCMC iterations with a burn-in of 50000 and thinning of 5 for 2 chains.

For the "identifiable cases", DC indicated the identifiability of $\lambda$ in all data sets. Fig. 5.1 displays 4 such DC plots indicating parameter identifiability. For the "non-identifiable cases", DC indicated the non-identifiability of $\lambda$ in all data sets. Fig. 5.2 displays DC plots for 4 data sets indicating non-identifiability. In this case, $\lambda$ was severely overestimated giving unrealistic estimates of population size. Additionally, from these Figs., it can be seen that DC results are similar for the different types of prior considered for both "identifiable cases" and "non-identifiable cases", indicating DC is not sensitive to prior specification in this scenario.
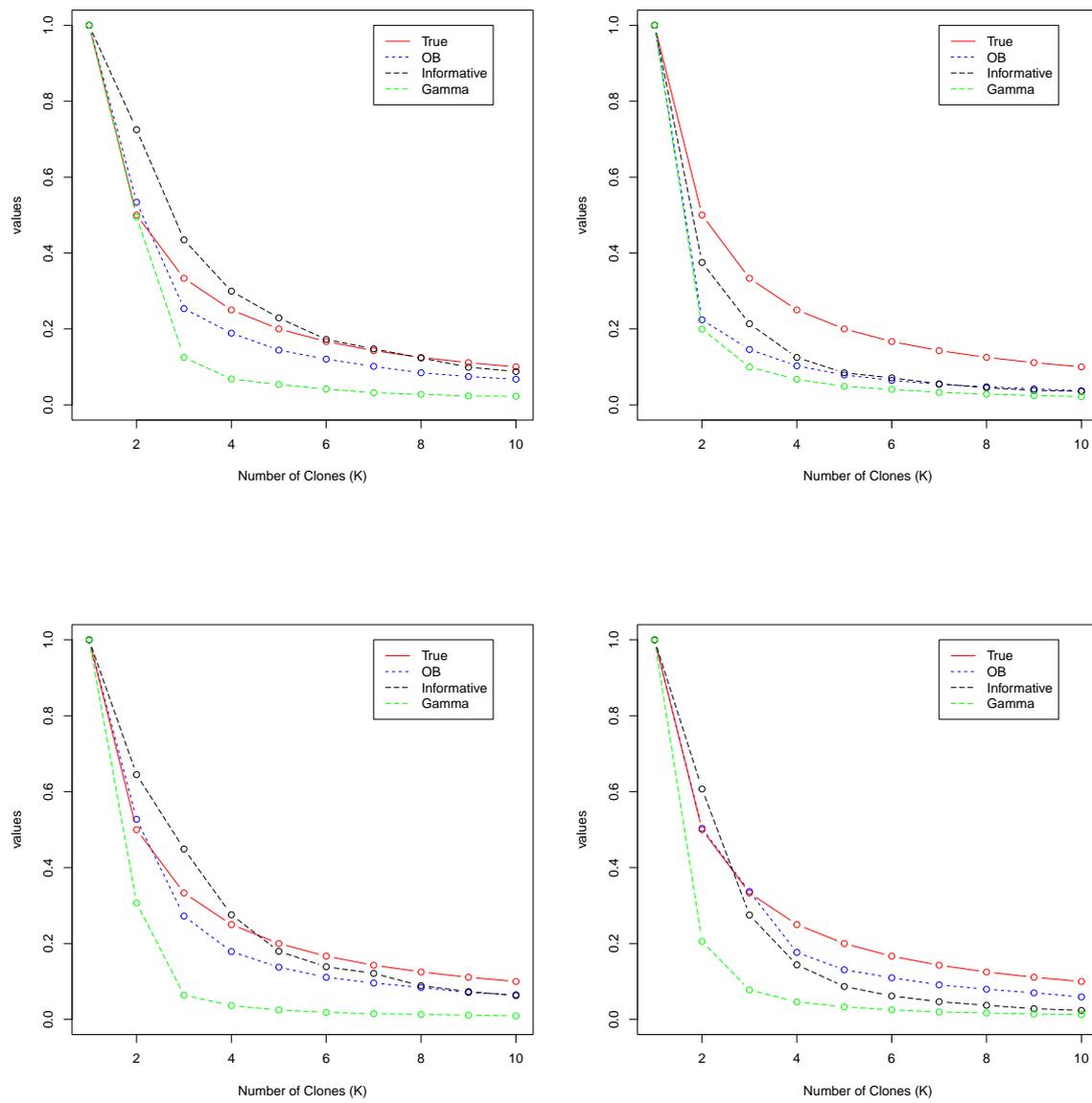
Fig. 5.1 Data cloning identifiability diagnostic plots for 4 "identifiable cases".
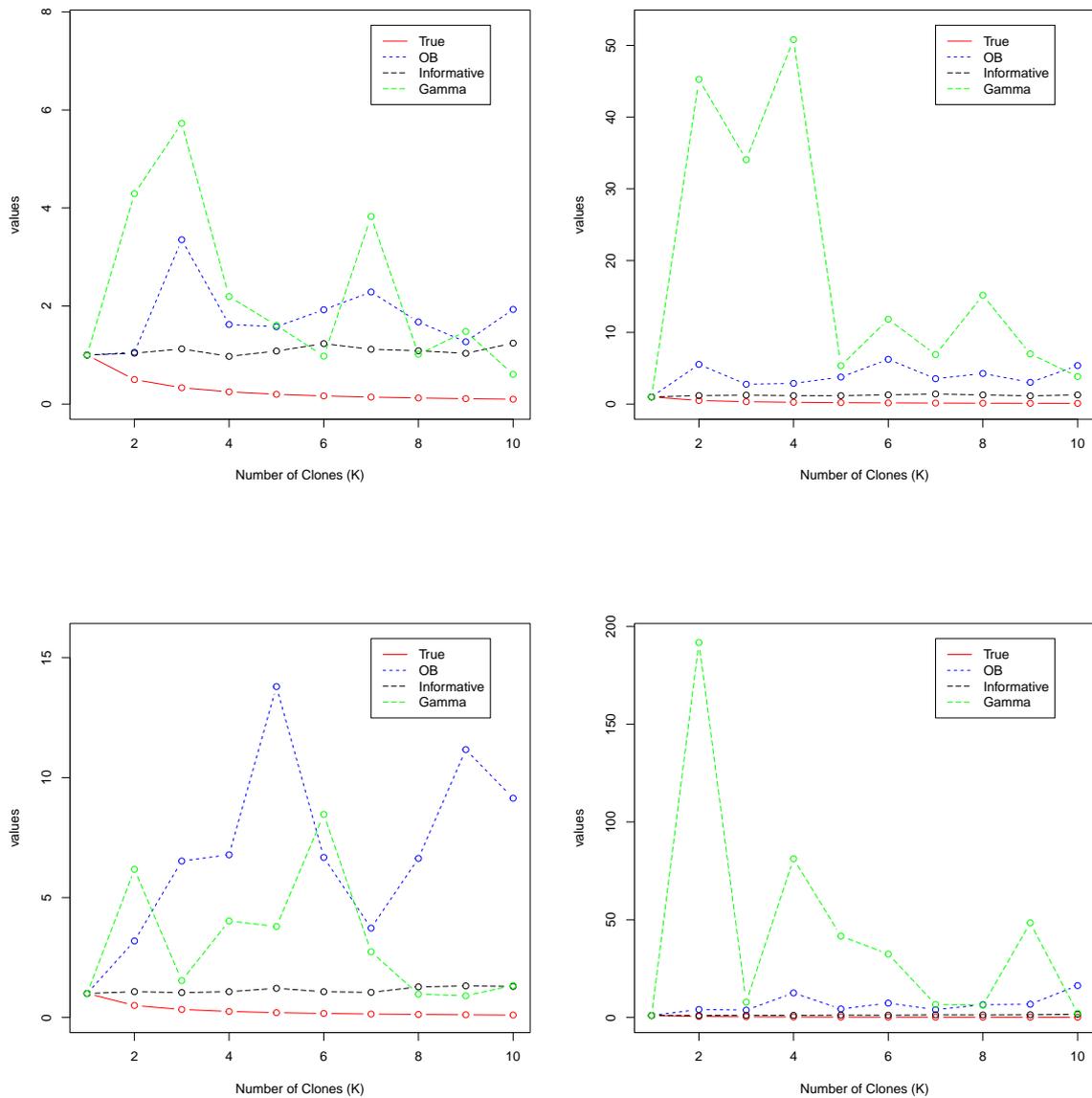
Fig. 5.2 Data cloning identifiability diagnostic plots for 4 " non-identifiable cases".

We further investigated the identifiability of over-dispersion N-mixture models using DC. We perform 10 simulation runs for each N-mixture model: DW-B, NB-B, P-BB and DW-BB for $p = 0.1$, $\lambda = 20$, $M = 20, J = 3, K = 20$. The OB prior was assigned to parameters in the parameter space $(0, \infty)$, and a Uniform$(0, 1)$ prior was assigned to parameters in the parameter space $(0, 1)$.

For the NB-B model, 8/10 datasets DC indicated non-identifiability issues for the size parameter of the NB distribution. These estimates of the size parameter were unrealistic large estimates but estimates of expected population size and $p$ were realistic indicating identifiability. For the DW-B model, 10/10 datasets DC indicated parameter identifiability with realistic inference. For the P-BB model, 8/10 datasets DC indicated the identifiability of all parameters. Two datasets indicated the non-identifiability of $\lambda$ and $\beta$, where these were over-estimated and the mean detection probability and $\rho$ were underestimated suggesting non-identifiability. For the DW-BB model, 6/10 datasets DC indicated the identifiability of all parameters. In the other 4 datasets, there were identifiability issues for $\beta$ as it was severely underestimated. However, there were no obvious signs of non-identifiability as inference on mean detection probability and expected population size was not unrealistic.

All in all, these results show that DC can be a valuable tool for investigating the identifiability of the P-B N-mixture model in a Bayesian setting. However, for over-dispersion N-mixture models, parameter identifiability via DC was not straightforward as in this case DC can indicate that either one or both parameters of the distribution for $N$ are non-identifiable, but inference on $N$ itself is reliable, suggesting that perhaps there exist several combinations of values or ranges of values for these parameters that yield similar inference for $N$. Dennis et al. (2015) also proposed two diagnostics to identify identifiability issues in the NB-B N-mixture model but these were found to be unreliable when used singly or in combination. Hence, future work is needed to investigate parameter identifiability in N-mixture models.


## Chapter 3: A new modelling framework for roost count data

In Chapter 3, a novel Bayesian modelling framework for the estimation of population size from roost count data is presented that accounts for both temporary emigration (TE) and observation error. We model TE using two classes of models, a parametric one, which is based on temporal models, and a non-parametric one, which relies on Dirichlet process (DP) mixture models. The DP mixture model provides information on temporary emigration cyclical patterns that can give new insights into the behaviors of the species, such as breeding patterns and seasonal availability of foods. The

temporal models increase the strength of estimation by sharing information across primary periods by accounting for temporal auto-correlation, enabling an intuitive ecological interpretation of TE.

We also highlight the importance of a mixed-effects model for the observation error when the observation error is misspecified as significant positive bias in the estimation of the population size can be obtained when not accounted for. To identify important predictors of observation error, we implemented an efficient Bayesian variable selection algorithm, BGLSS. This method has the advantage of identifying both categorical and continuous predictors to reduce uncertainty around predictor effects and population size.

An important constraint to consider in this framework is model selection. We use the well-established model selection tool, cross-validation, to select between competing models. However, cross-validation can be computationally intensive as it requires multiple fitting of the data. Other model selection methods such as the Watanable-Akaike information criterion (WAIC) (Watanabe and Opper, 2010) and the log pseudo marginal likelihood (LPML) criterion (Geisser and Eddy, 1979) only require fitting the data once and can be easily computed using MCMC samples from popular software such as NIMBLE and Stan (Carpenter et al., 2017). Consequently, we investigate model selection via WAIC and LPML in this framework. We consider these criteria at both the conditional (cLPML,cWAIC) and marginal (mLPML,mWAIC) levels as Ariyo et al. (2022a) and Millar (2018) advocated that the marginal criterion should be used instead of the conditional criterion. We define $\Delta cLPML, \Delta cWAIC$, $\Delta mLPML$, and $\Delta mWAIC$ to be the difference between the true and competing model at the conditional and marginal level for LPML and WAIC respectively. We set the DP as the true model and compared it to the RW1 model. We perform 10 simulation runs using the mixed effect detection model settings defined in the simulation section of Chapter 3. We fit the conditional model as it was found to be more computationally efficient than the marginal model and use MCMC samples to compute WAIC and LPML.

At the high detection probability setting, model selection via cLPML selected the true model 40% of the time, and the RW1 model 60%, with $\Delta cLPML = -1.711$. cWAIC selected the true model 30% of the time, and the RW1 model 70%, with $\Delta cWAIC = 0.227$. Finally, mLPML and mWAIC selected the true model 20% of the time, and the RW1 model 80%, with $\Delta mLPML = -5.636, \Delta mWAIC =$

78.235 respectively. Both models gave similar estimates of population size, with median relative bias $(-0.005, -0.004)$ and median coverage $(1, 1)$ for the DP and RW1 model respectively.

At the low detection probability setting, model selection via cLPML selected the true model 60% of the time, and the RW1 model 40%, with $\Delta cLPML = 2.937$. cWAIC selected the true model 80% of the time, and the RW1 20%, with $\Delta cWAIC = -1.269$. Finally, mLPML selected the true 0% and the RW1 100% with $\Delta mLPML = -342.668$ and mWAIC selected the true model 20% of the time, and the RW1 80%, with $\Delta mWAIC = 2866.5$. In addition, both models gave similar estimates of population size, with median relative bias $(0.020, 0.055)$ and median coverage $(1, 1)$ for the DP and RW1 model respectively. Additionally, at the low detection probability level for some data sets, the computation of mLPML and mWAIC led to numerical issues, with probabilities rounded to 0, and hence returning $-\infty$ value on the log scale. These datasets were discarded from the simulation study.

Hence, this small simulation study seems to suggest that WAIC and LPML are not viable model selection tools in this temporally correlated framework. The poor performance of WAIC and LPML may be due to the independence assumption of data given the parameters needed for computation. This assumption is often violated in temporal models where dependence among the data is a key modelling feature. Hence, future work can be focused on investigating/developing efficient model selection methods for temporally correlated data.

## Chapter 4: Modelling disease dynamics from spatially explicit capture-recapture data

Chapter 4 presents a novel spatially explicit capture-recapture (SCR) model that models disease dynamics using disease data from multiple imperfect tests together with SCR data with a focus on population density effect on disease transmission probability. Accounting for observation errors in both detection and disease tests, this model simultaneously models population demographics and disease dynamics in a spatial context enabling the estimation of critical parameters, such as population size, local and global density by disease status, disease prevalence, and demographic rates as functions of disease status, etc. Notably, this framework models disease transmission probability as a function of local density, enabling inference on the impact of density on disease transmission. Thus, this

model provides a better understanding of how disease dynamics relate to population demographics in spatiotemporal contexts at an individual level.

In our simulation study and real data analysis of chapter 4, we assumed a constant state model, that is, individual activity centres (ACs) are independent and do not change over the period of study due to species behaviour. This assumption implies local density does not change dramatically between grids and sampling occasions. Thus, the state model definition plays a big role in modelling disease transmission and future work can be focused to assess model sensitivity to different state models. Other state models such as the independent (Royle et al., 2014) and the Markovian random walk models (Raabe et al., 2014) can be easily adopted. Repulsion (Diana et al., 2022) and attraction (McLaughlin and Bar, 2021) models can also be considered to relax the independence assumption among individual ACs.

We perform a short simulation study to assess our model sensitivity when 1/3 of the population at the start of the study is clustered into one group as depicted in Fig. 5.3. This cluster will result in local density, and in turn disease transmission probability, varying substantially between grids. We perform 5 simulation runs using the simulation settings of high population size and density effect given in Chapter 4. We run 15000 MCMC iterations, burn-in of 10000 iterations, and thinning of 5 for 2 chains.

As can be seen from Table 5.1, when individual ACs are not independently distributed, as the model of Chapter 4 assumes, inference is biased, particularly for the population size ($N_{1:T}$) and the population size of uninfected individuals ($N_{1:T}^{\mathrm{u}}$). Hence, future work can be focused on extending this primary study to investigate model sensitivity to state models.
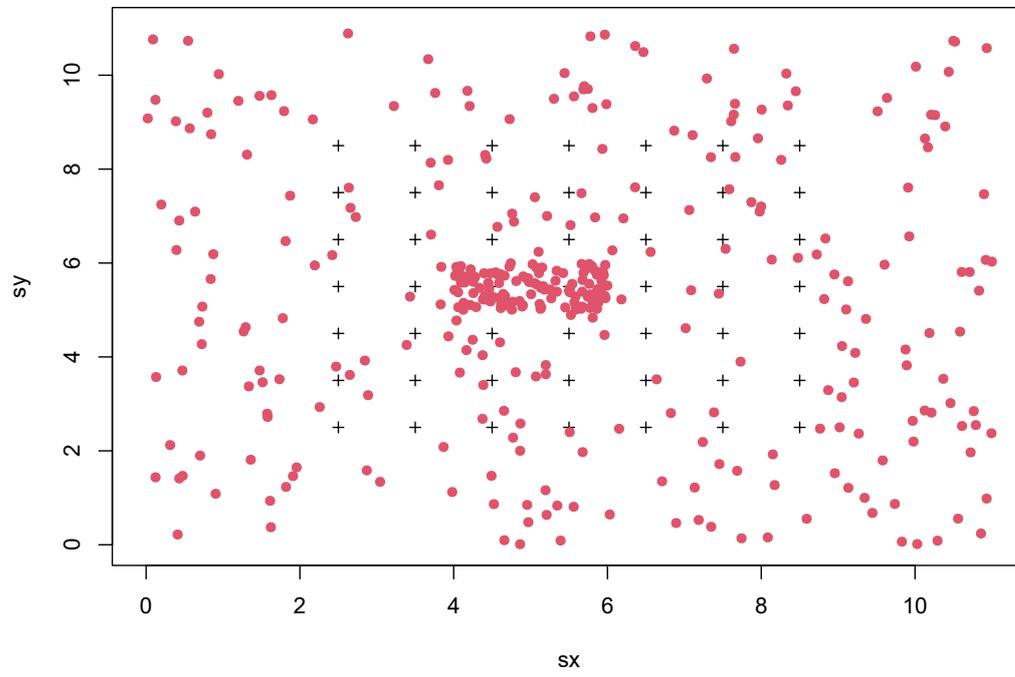
Fig. 5.3 Individual ACs at the start of the study

Table 5.1 Model simulation results for clustered population. The table shows the relative bias (RB), coefficient of variation (CV), and coverage with values within brackets showing the 95% quantiles.

| Parameters | RB | CV | Coverage |
|---|---|---|---|
| $\gamma_1$ | $0.418(0.163, 0.548)$ | $0.069(0.058, 0.069)$ | 20 |
| $\gamma_2$ | $-0.029(-0.624, 0.197)$ | $0.242(0.213, 0.676)$ | 80 |
| $\gamma_3$ | $0.177(-0.054, 0.662)$ | $0.249(0.204, 0.357)$ | 80 |
| $\gamma_4$ | $0.613(0.018, 0.850)$ | $0.258(0.223, 0.291)$ | 60 |
| $\phi_{d_i=0}$ | $0.011(-0.008, 0.025)$ | $0.019(0.018, 0.025)$ | 100 |
| $\phi_{d_i=1}$ | $-0.030(-0.042, -0.014)$ | $0.022(0.020, 0.024)$ | 100 |
| $p0_{d_i=0}$ | $-0.119(-0.165, 0.082)$ | $0.071(0.065, 0.086)$ | 80 |
| $p0_{d_i=1}$ | $0.020(-0.009, 0.047)$ | $0.048(0.046, 0.053)$ | 100 |
| $\sigma_{d_i=0}$ | $-0.027(-0.072, 0.020)$ | $0.028(0.027, 0.036)$ | 80 |
| $\sigma_{d_i=1}$ | $0.007(-0.013, 0.018)$ | $0.018(0.017, 0.019)$ | 100 |
| $\delta$ | $-0.121(-0.357, 0.156)$ | $0.212(0.203, 0.239)$ | 80 |
| $\beta_0$ | $-0.079(-0.208, -0.013)$ | $0.109(0.103, 0.141)$ | 60 |
| $\beta_1$ | $0.831(0.561, 1.715)$ | $0.238(0.226, 0.274)$ | 20 |
| $q_{11}^1$ | $0.012(-0.024, 0.052)$ | $0.033(0.032, 0.034)$ | 100 |
| $q_{11}^2$ | $-0.015(-0.025, 0.043)$ | $0.021(0.019, 0.022)$ | 80 |
| $q_{11}^3$ | $-0.144(-0.202, 0.232)$ | $0.123(0.106, 0.129)$ | 80 |
| $q_{00}^1$ | $0.001(-0.022, 0.0203)$ | $0.015(0.013, 0.017)$ | 100 |
| $q_{00}^2$ | $-0.001(-0.005, 0.008)$ | $0.017(0.014, 0.018)$ | 100 |
| $q_{00}^3$ | $-0.001(-0.003, -0.001)$ | $0.002(0.002, 0.004)$ | 100 |
| $N_{1:T}$ | $0.127(0.051, 0.396)$ | $0.048(0.041, 0.063)$ | 0 |
| $N_{1:T}^{\mathrm{u}}$ | $0.202(0.004, 0.519)$ | $0.083(0.065, 0.101)$ | 0 |
| $N_{1:T}^{\mathrm{i}}$ | $0.058(-0.065, 0.313)$ | $0.070(0.063, 0.197)$ | 100 |

Alternative models for disease transmission probability can also be considered. For example, we define and test two alternative models: the half-normal (HN) model and the overlap (OV) model. Spatial covariates can also be used to model disease transmission probability.

## Half Normal (HN) model

Let $\theta_{i,l}$ be the probability that an uninfected individual $i$ becomes infected by an infected individual $l$. We let

$$\theta_{i,l} = \exp\left(-\frac{1}{2\sigma^2_{d_{i=0}}} \parallel s_i - s_l \parallel^2\right)$$

Then the disease transmission probability of an uninfected individual $i$ being infected at occasion $t$ ($\psi_{i,t}$) can be written as

$$\psi_{i,t} = 1 - \left(\prod_{l=1}^{L} 1 - \theta_{i,l}\right)$$

for $l = 1,\ldots,L$ infected individuals. Thus, the HN model assumes that an uninfected individual is more likely to become infected the closer it is to infected individuals conditional on the movement of uninfected individuals.

To test the validity of the model proposed, we perform a simulation study of 15 runs with the following parameter values: $M = 500, T = 8, \phi_{d_{i=0}} = 0.9, \phi_{d_{i=1}} = 0.8, p0_{d_{i=0}} = 0.5, p0_{d_{i=1}} = 0.2, \sigma_{d_{i=0}} = 0.3, \sigma_{d_{i=1}} = 1, \delta = 0.2, q^1_{11} = 0.492, q^2_{11} = 0.809, q^3_{11} = 0.1, q^1_{00} = 0.931, q^2_{00} = 0.936, q^3_{00} = 0.999,$ $\gamma_1 = 0.4, \gamma_2 = \gamma_{2:4} = 0.05, \gamma_3 = 0.2, \gamma_4 = \gamma_{6:T} = 0.05$. We run 10000 MCMC iterations, burn-in of 4000 iterations, and thinning of 5 for 2 chains.

As can be seen from Table 5.2, the HN model performs well. However, the limitation of this approach is that it only considers movement in one direction, the direction of uninfected to infected individuals. However, infected individuals also move and, as shown in Chapter 4, in some cases their home range can be wider than that of uninfected individuals.

Table 5.2 HN model simulation results. The table shows the median of posterior means, coverage, median values of the residual mean square, and relative bias (RB).

| Parameters | Mean | Coverage | Residual Mean Square | Relative Bias |
|---|---|---|---|---|
| $\gamma_1$ | 0.409 | 100 | 0.086 | 0.023 |
| $\gamma_2$ | 0.063 | 100 | 0.478 | 0.270 |
| $\gamma_3$ | 0.149 | 93 | 0.440 | $-0.252$ |
| $\gamma_4$ | 0.056 | 100 | 0.518 | 0.126 |
| | | | | |
| $\phi_{d_{i=0}}$ | 0.920 | 100 | 0.049 | 0.023 |
| $\phi_{d_{i=1}}$ | 0.789 | 93 | 0.045 | $-0.018$ |
| $p0_{d_{i=0}}$ | 0.472 | 100 | 0.114 | $-0.051$ |
| $p0_{d_{i=1}}$ | 0.191 | 93 | 0.092 | $-0.045$ |
| $\sigma_{d_{i=0}}$ | 0.300 | 93 | 0.065 | 0.002 |
| $\sigma_{d_{i=1}}$ | 1.003 | 100 | 0.033 | 0.003 |
| | | | | |
| $\delta$ | 0.198 | 100 | 0.237 | $-0.009$ |
| $q_{11}^1$ | 0.491 | 100 | 0.023 | $-0.003$ |
| $q_{11}^2$ | 0.800 | 100 | 0.033 | $-0.003$ |
| $q_{11}^3$ | 0.098 | 100 | 0.169 | $-0.012$ |
| $q_{00}^1$ | 0.927 | 87 | 0.032 | $-0.003$ |
| $q_{00}^2$ | 0.933 | 100 | 0.025 | $-0.003$ |
| $q_{00}^3$ | 0.994 | 100 | 0.005 | $-0.004$ |

**Overlap (OV) Model**
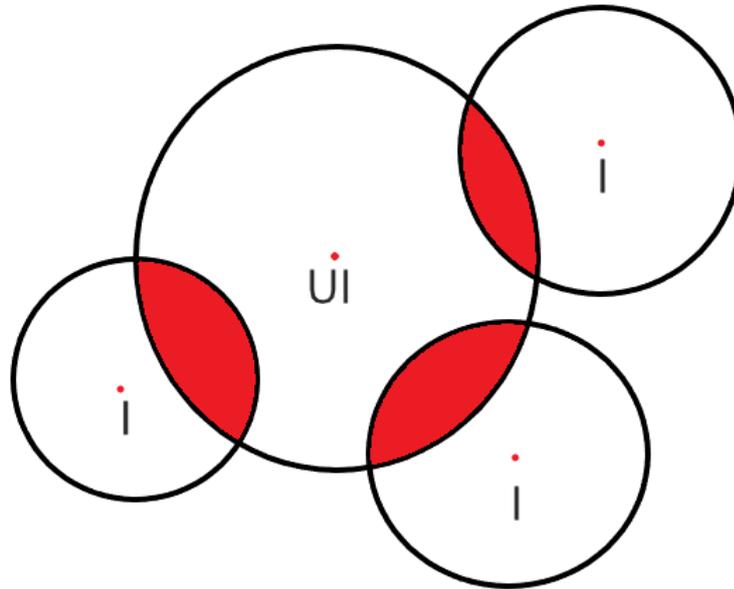


Fig. 5.4 An uninfected (UI) individual surrounded by three infected individuals (I) with the intersection of home range areas highlighted in red.

Another approach considered is to model disease transmission probability as a function of the overlap of home range areas between an uninfected individual and surrounding infected individuals. Fig. 5.4, shows an uninfected individual being surrounded by 3 infected individuals with home range overlap shown in red. Consequently, we model disease transmission as

$$\text{logit}(\psi_{i,t}) = \beta_0 + \beta_1 \cdot A_{i,t-1} \tag{5.0.1}$$

where $A_{i,t-1}$ is the sum of the overlap area between the uninfected individual $i$ and all infected individuals at the previous time point and $\beta_1$ is the coefficient describing the relationship between the overlap of home range areas and disease transmission probability.

Assuming circular individual home range areas, to compute the overlap of home range area between an uninfected and an infected individual, we let $C_1$ and $C_2$ be the two circular individual home ranges

with radii $r_1$ and $r_2$ respectively whose centers are at a distance $d$ from each other as depicted in Fig. 5.5. Assume, that $r_1 \geq r_2$, the overlap of home range area between an uninfected and an infected individual can be computed using equation (5.0.2). This approach enables movement to be considered in both directions and the individual with the larger home range area is assigned to $C_1$.
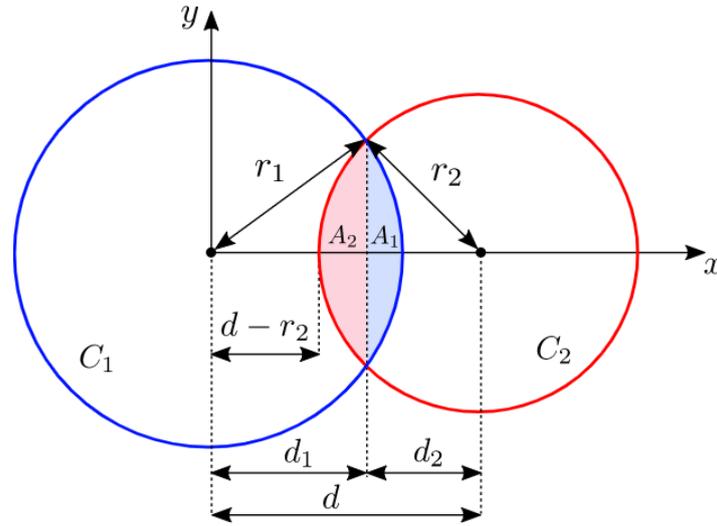


Fig. 5.5 Overlap of two circular individual home ranges, $C_1$ and $C_2$ with radii $r_1$ and $r_2$ respectively whose centers are at a distance $d$ from each other.

$$\text{Overlap area} = r_1^2 cos^{-1}\left(\frac{d_1}{r_1}\right) - d_1\sqrt{r_1^2 - d_1^2} + r_2^2 cos^{-1}\left(\frac{d_2}{r_2}\right) - d_2\sqrt{r_2^2 - d_2^2}, \qquad (5.0.2)$$

where

$$d_1 = \frac{r_1^2 - r_2^2 + d^2}{2d} \text{ and } d_2 = d - d_1 = \frac{r_2^2 - r_1^2 + d^2}{2d}$$

Consequently, the overlap of the home range area of an uninfected and infected individual is:

- zero, if $d \geq r_1 + r_2$, since in this case the circles intersect at most up to a point.

- $\pi r_2^2$, if $d \leq r_1 - r_2$, since in this case $C_2$ is entirely contained within $C_1$.

- equation (5.0.2) in all other cases.

To assess the overlap model performance a short simulation of 10 runs was carried out with parameter values: $M = 500, T = 8, \phi_{d_i=0} = 0.9, \phi_{d_i=1} = 0.8, p_{0_{d_i=0}} = 0.65, p_{0_{d_i=1}} = 0.75, \sigma_{d_i=0} = 0.75, \sigma_{d_i=1} = 1,$ $\delta = 0.2, \beta_0 = -5.5, \beta_1 = 0.1, q_{11}^1 = 0.492, q_{11}^2 = 0.809, q_{11}^3 = 0.1, q_{00}^1 = 0.931, q_{00}^2 = 0.936, q_{00}^3 = 0.999, \gamma_1 = 0.4, \gamma_2 = \gamma_{2:4} = 0.05, \gamma_3 = 0.2, \gamma_4 = \gamma_{6:T} = 0.05.$ We run 10000 MCMC iterations, burn-in of 4000 iterations, and thinning of 5 for 2 chains.

Table 5.3 OV model simulation results. The table shows the median of posterior means, coverage, median values of the residual mean square, and relative bias (RB).

| Parameters | Mean | Coverage | Residual Mean Square | Relative Bias |
|---|---|---|---|---|
| $\gamma_1$ | 0.400 | 90 | 0.060 | 0.001 |
| $\gamma_2$ | 0.053 | 100 | 0.226 | 0.071 |
| $\gamma_3$ | 0.202 | 100 | 0.158 | 0.012 |
| $\gamma_4$ | 0.046 | 90 | 0.282 | $-0.085$ |
| $\phi_{d_{i=0}}$ | 0.936 | 40 | 0.040 | 0.040 |
| $\phi_{d_{i=1}}$ | 0.767 | 50 | 0.068 | $-0.041$ |
| $p0_{d_{i=0}}$ | 0.626 | 100 | 0.051 | $-0.044$ |
| $p0_{d_{i=1}}$ | 0.734 | 100 | 0.026 | $-0.017$ |
| $\sigma_{d_{i=0}}$ | 0.745 | 90 | 0.026 | $-0.006$ |
| $\sigma_{d_{i=1}}$ | 0.993 | 100 | 0.016 | $-0.007$ |
| $\delta$ | 0.226 | 70 | 0.298 | 0.130 |
| $\beta_0$ | $-4.783$ | 100 | $-0.160$ | $-0.139$ |
| $\beta_1$ | 0.083 | 60 | 0.230 | $-0.167$ |
| $q_{11}^1$ | 0.471 | 100 | 0.048 | $-0.041$ |
| $q_{11}^2$ | 0.761 | 40 | 0.061 | $-0.059$ |
| $q_{11}^3$ | 0.093 | 100 | 0.106 | $-0.069$ |
| $q_{00}^1$ | 0.931 | 90 | 0.013 | 0.002 |
| $q_{00}^2$ | 0.940 | 100 | 0.007 | 0.004 |
| $q_{00}^3$ | 0.998 | 90 | 0.002 | $-0.001$ |

Looking at Table 5.3, it can be seen that the overlap model estimated parameters well where $\phi_{d_{i=0}}, \phi_{d_{i=1}}$ and $q_{11}^2$ were estimated with narrow credible intervals but still estimated well, $\beta_1$ and $\delta$ were also estimated well even with higher bias and lower coverage. It is highly plausible that these

biases may be due to the low population size simulated. Future work can be focused to investigate the robustness of this model especially with a larger population size.

Notably, all three models considered for disease transmission probability are very computationally demanding and resulted in convergence issues and poor mixing, requiring long run times. Consequently, future work can be done to improve computational efficiency.

In this thesis, 100, 50, and 10 simulation runs were performed in the simulation studies for chapters 2, 3, and 4 respectively. These numbers of runs were chosen due to model complexity in combination with the computational burden of Bayesian inference and the range of simulation scenarios considered. For example, the proposed model in chapter 4 took approximately 12 hours for 1 simulation run. Consequently, Monte Carlo error is only averaged across the simulation runs for each chapter. Hence, care needs to be taken on how the results are interpreted, especially in cases where simulation runs are low and where models are performing similarly in different scenarios.

## Conclusion

Ecological data of the type considered in this thesis will continue to be key for monitoring wildlife populations, and this thesis has contributed to the statistical ecology literature by developing a suite of BHM that provide valuable information on some of the key topics in ecology and population dynamics, including observation error, temporary emigration and the complex relationships between disease dynamics and population demographics. These models can be further extended in a number of directions as described. Additionally, future work can be focused on dealing with the challenges of increasingly large data sets, or complex observation processes that arise due to advances in technology for collecting data, such as remote sensing, DNA sequencing, and autonomous systems.

# References

Rosamund EA Almond, Monique Grooten, and T Peterson. *Living Planet Report 2020-Bending the curve of biodiversity loss*. World Wildlife Fund, 2020.

Oludare Ariyo, Adrian Quintero, Johanna Muñoz, Geert Verbeke, and Emmanuel Lesaffre. Bayesian model selection in linear mixed models for longitudinal data. *Journal of Applied Statistics*, 47(5): 890–913, 2020.

Oludare Ariyo, Emmanuel Lesaffre, Geert Verbeke, and Adrian Quintero. Bayesian model selection for longitudinal count data. *Sankhya B*, 84(2):516–547, 2022a.

Oludare Ariyo, Emmanuel Lesaffre, Geert Verbeke, and Adrian Quintero. Model selection for Bayesian linear mixed models with longitudinal data: Sensitivity to the choice of priors. *Communications in statistics-simulation and computation*, 51(4):1591–1615, 2022b.

Roland T Ashford, Paul Anderson, Laura Waring, Dipesh Davé, Freya Smith, Richard J Delahay, Eamonn Gormley, Mark A Chambers, Jason Sawyer, and Sandrine Lesellier. Evaluation of the dual path platform (DPP) VetTB assay for the detection of Mycobacterium bovis infection in badgers. *Preventive veterinary medicine*, 180:105005, 2020.

Ray Bai, Gemma E Moran, Joseph L Antonelli, Yong Chen, and Mary R Boland. Spike-and-slab group lassos for grouped regression and sparse generalized additive models. *Journal of the American Statistical Association*, pages 1–14, 2020.

Ray Bai, Veronika Ročková, and Edward I George. Spike-and-slab meets lasso: A review of the spike-and-slab lasso. *Handbook of Bayesian Variable Selection*, pages 81–108, 2021.

Katharine M Banner, Kathryn M Irvine, and Thomas J Rodhouse. The use of Bayesian priors in Ecology: The good, the bad and the not great. *Methods in Ecology and Evolution*, 11(8):882–889, 2020.

Maria Maddalena Barbieri and James O Berger. Optimal predictive model selection. *The Annals of Statistics*, 32(3):870–897, 2004.

Richard J Barker, Matthew R Schofield, William A Link, and John R Sauer. On the reliability of N-mixture models for count data. *Biometrics*, 74(1):369–377, 2018.

Guy Beauchamp. The evolution of communal roosting in birds: origin and secondary losses. *Behavioral Ecology*, 10(6):675–687, 1999.

Rudolf Beran. Minimum Hellinger distance estimates for parametric models. *The Annals of Statistics*, pages 445–463, 1977.

Karl S Berg and Rafael R Angel. Seasonal roosts of red-lored Amazons in Ecuador provide information about population size and structure. *Journal of Field Ornithology*, 77(2):95–103, 2006.

James O Berger, José M Bernardo, and Dongchu Sun. The formal definition of reference priors. *The Annals of Statistics*, 37(2):905–938, 2009.

Julian Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225, 1974.

Julian Besag, Jeremy York, and Annie Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–20, 1991.

Rebecca Biddle, Ivette Solis Ponce, Paul Cun, Simon Tollington, Martin Jones, Stuart Marsden, Christian Devenish, Eric Horstman, Karl Berg, and Mark Pilgrim. Conservation status of the recently described Ecuadorian Amazon parrot Amazona lilacina. *Bird Conservation International*, 30(4):586–598, 2020.

Rebecca Biddle, Ivette Solis-Ponce, Martin Jones, Stuart Marsden, Mark Pilgrim, and Christian Devenish. The value of local community knowledge in species distribution modelling for a threatened neotropical parrot. *Biodiversity and Conservation*, 30(6):1803–1823, 2021a.

Rebecca Biddle, Ivette Solis-Ponce, Martin Jones, Mark Pilgrim, and Stuart Marsden. Parrot ownership and capture in coastal ecuador: Developing a trapping pressure index. *Diversity*, 13(1):15, 2021b.

Richard Bischof, Daniel Turek, Cyril Milleret, Torbjørn Ergon, Pierre Dupont, and Perry de Valpine. nimblescr: spatial capture-recapture (SCR) methods using 'nimble'. *R package version 0.1. 0*, 2020.

David Blackwell and James B MacQueen. Ferguson distributions via pólya urn schemes. *The Annals of Statistics*, 1(2):353–355, 1973.

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

Svetlana N Buzdugan, Timothée Vergne, Vladimir Grosbois, Richard J Delahay, and Julian A Drewe. Inference of the infection status of individuals using longitudinal testing data from cryptic populations: Towards a probabilistic approach to diagnosis. *Scientific Reports*, 7(1):1–11, 2017.

Bradley J Cardinale, J Emmett Duffy, Andrew Gonzalez, David U Hooper, Charles Perrings, Patrick Venail, Anita Narwani, Georgina M Mace, David Tilman, David A Wardle, et al. Biodiversity loss and its impact on humanity. *Nature*, 486(7401):59–67, 2012.

Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of Statistical Software*, 76(1), 2017.

John P Carroll and Michael J Conroy. *Quantitative conservation of vertebrates*. John Wiley & Sons, 2011.

George Casella, Malay Ghosh, Jeff Gill, and Minjung Kyung. Penalized regression, standard errors, and Bayesian lassos. *Bayesian analysis*, 5(2):369–411, 2010.

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

CL Cheeseman, CHEESEMAN CL, et al. Methods of marking badgers (meles meles). 1982.

Ray-Bing Chen, Chi-Hsiang Chu, Shinsheng Yuan, and Ying Nian Wu. Bayesian sparse group selection. *Journal of Computational and Graphical Statistics*, 25(3):665–683, 2016.

Rémi Choquet, Cécile Carrié, Thierry Chambert, and Thierry Boulinier. Estimating transitions between states using measurements with imperfect detection: application to serological data. *Ecology*, 94(10):2160–2165, 2013.

Stacey Cougill and Stuart J Marsden. Variability in roost size in an amazona parrot: implications for roost monitoring. *Journal of Field Ornithology*, 75(1):67–73, 2004.

Deanna Dalley, Dipesh Davé, Sandrine Lesellier, Simonette Palmer, Timothy Crawshaw, R Glyn Hewinson, and Mark Chambers. Development and evaluation of a gamma-interferon assay for tuberculosis in badgers (meles meles). *Tuberculosis*, 88(3):235–243, 2008.

Francisco De Castro and Benjamin Bolker. Mechanisms of disease-induced extinction. *Ecology letters*, 8(1):117–126, 2005.

Leiliany Negrao De Moura, Jacques ME Vielliard, and Maria Luisa Da Silva. Seasonal fluctuation of the Orange-winged Amazon at a roosting site in Amazonia. *The Wilson Journal of Ornithology*, 122(1):88–94, 2010.

Perry de Valpine, Daniel Turek, Christopher J Paciorek, Clifford Anderson-Bergman, Duncan Temple Lang, and Rastislav Bodik. Programming with models: writing statistical algorithms for general model structures with NIMBLE. *Journal of Computational and Graphical Statistics*, 26(2):403–413, 2017.

Richard J Delahay, N Walker, GS Smith, D Wilkinson, RS Clifton-Hadley, CL Cheeseman, AJ Tomlinson, and MA Chambers. Long-term temporal trends and estimated transmission rates for mycobacterium bovis infection in an undisturbed high-density badger (meles meles) population. *Epidemiology & Infection*, 141(7):1445–1456, 2013.

RJ Delahay, SP Carter, GJ Forrester, A Mitchell, and CL Cheeseman. Habitat correlates of group size, bodyweight and reproductive performance in a high-density eurasian badger (meles meles) population. *Journal of Zoology*, 270(3):437–447, 2006.

Francisco V Dénes, José L Tella, and Steven R Beissinger. Revisiting methods for estimating parrot abundance and population size. *Emu-Austral Ornithology*, 118(1):67–79, 2018.

Emily B Dennis, Byron J T Morgan, and Martin S Ridout. Computational aspects of N-mixture models. *Biometrics*, 71(1):237–246, 2015.

Alex Diana, Eleni Matechou, Jim E Griffin, Yadvendradev Jhala, and Qamar Qureshi. A vector of point processes for modeling interactions between and within species using capture-recapture data. *Environmetrics*, page e2781, 2022.

Julian A Drewe, Alexandra J Tomlinson, Neil J Walker, and Richard J Delahay. Diagnostic accuracy and optimal use of three tests for tuberculosis in live badgers. *PloS one*, 5(6):e11196, 2010.

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

Claes Enøe, Marios P Georgiadis, and Wesley O Johnson. Estimation of sensitivity and specificity of diagnostic tests and disease prevalence when the true disease state is unknown. *Preventive veterinary medicine*, 45(1-2):61–81, 2000.

Michael D Escobar and Mike West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430):577–588, 1995.

Ludwig Fahrmeir and Stefan Lang. Bayesian inference for generalized additive mixed models based on Markov random field priors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 50(2):201–220, 2001.

Thomas S Ferguson. A Bayesian analysis of some nonparametric problems. *The Annals of Statistics*, pages 209–230, 1973.

J Gallagher and DM Horwill. A selective oleic acid albumin agar medium for the cultivation of Mycobacterium bovis. *Epidemiology & Infection*, 79(1):155–160, 1977.

Seymour Geisser and William F Eddy. A predictive approach to model selection. *Journal of the American Statistical Association*, 74(365):153–160, 1979.

Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical science*, pages 457–472, 1992.

Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for Bayesian models. *Statistics and computing*, 24(6):997–1016, 2014.

Edward I George and Robert E McCulloch. Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.

J Graham, GC Smith, RJ Delahay, T Bailey, RA McDonald, and D Hodgson. Multi-state modelling reveals sex-dependent transmission, progression and severity of tuberculosis in wild badgers. *Epidemiology & Infection*, 141(7):1429–1436, 2013.

Jim E Griffin, Eleni Matechou, Andrew S Buxton, Dimitrios Bormpoudakis, and Richard A Griffiths. Modelling environmental DNA data; Bayesian variable selection accounting for false positive and false negative errors. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69(2): 377–392, 2020.

Skylar R Hopkins, Arietta E Fleming-Davies, Lisa K Belden, and Jeremy M Wojdak. Systematic review of modelling assumptions and empirical evidence: does parasite transmission increase nonlinearly with host density? *Methods in Ecology and Evolution*, 11(4):476–486, 2020.

Hao Hu, Karima Nigmatulina, and Philip Eckhoff. The scaling of contact rates with population density for the infectious disease models. *Mathematical biosciences*, 244(2):125–134, 2013.

EA Hunter, NP Nibbelink, and RJ Cooper. Divergent forecasts for two salt marsh specialists in response to sea level rise. *Animal Conservation*, 20(1):20–28, 2017.

Harold Jeffreys. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 186(1007):453–461, 1946.

Walter Jetz, Melodie A McGeoch, Robert Guralnick, Simon Ferrier, Jan Beck, Mark J Costello, Miguel Fernandez, Gary N Geller, Petr Keil, Cory Merow, et al. Essential biodiversity variables for mapping and monitoring species populations. *Nature ecology & evolution*, 3(4):539–551, 2019.

Rana Jreich, Christine Hatte, and Eric Parent. Review of Bayesian selection methods for categorical predictors using JAGS. *Journal of Applied Statistics*, pages 1–19, 2021.

Hadeel Saleh Kalktawi. *Discrete Weibull regression model for count data*. PhD thesis, Brunel University London, 2017.

Robert E Kass and Adrian E Raftery. Bayes factors. *Journal of the American Statistical Association*, 90(430):773–795, 1995.

Marc Kéry. Identifiability in N-mixture models: A large-scale screening test with bird data. *Ecology*, 99(2):281–288, 2018.

Marc Kéry and J Andrew Royle. Applied Hierarchical Modeling in Ecology: Analysis of Distribution, Abundance and Species Richness in R and BUGS. 2015.

Marc Kéry and J Andrew Royle. Applied hierarchical modeling in ecology: analysis of distribution, abundance and species richness in R and BUGS. *(No Title)*, 2017.

Marc Kéry and J Andrew Royle. *Applied Hierarchical Modeling in Ecology: Analysis of distribution, abundance and species richness in R and BUGS: Volume 2: Dynamic and Advanced Models.* Academic Press, 2020.

Fabian R. Ketwaroo. N-mixture models from a Bayesian perspective. *Unpublished MSc. Thesis*, 2019.

Jonas Knape, Debora Arlt, Frédéric Barraquand, Åke Berg, Mathieu Chevalier, Tomas Pärt, Alejandro Ruete, and Michał Żmihorski. Sensitivity of binomial N-mixture models to overdispersion: The importance of assessing model fit. *Methods in Ecology and Evolution*, 9(10):2102–2114, 2018.

Athanasios Kottas. Dirichlet process mixtures of beta distributions, with applications to density and intensity estimation. In *Workshop on Learning with Nonparametric Bayesian Methods, 23rd International Conference on Machine Learning (ICML)*, volume 47, 2006.

Solomon Kullback and Richard A Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

Thomas H Kunz. Roosting ecology of bats. *Ecology of bats*, pages 1–55, 1982.

Zachary S Ladin, Vincent D'Amico, Jan M Baetens, Roland R Roth, and W Gregory Shriver. Predicting metapopulation responses to conservation in human-dominated landscapes. *Frontiers in Ecology and Evolution*, 4:122, 2016.

Fabrizio Leisen, Cristiano Villa, Stephen G Walker, et al. On a class of objective priors from scoring rules. *Bayesian Analysis*, 2018.

Subhash R Lele, Brian Dennis, and Frithjof Lutscher. Data cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology letters*, 10(7):551–563, 2007.

Subhash R Lele, Khurram Nadeem, and Byron Schmuland. Estimability and likelihood inference for generalized linear mixed models using data cloning. *Journal of the American Statistical Association*, 105(492):1617–1625, 2010.

William A Link, Matthew R Schofield, Richard J Barker, and John R Sauer. On the robustness of N-mixture models. *Ecology*, 99(7):1547–1551, 2018.

Benoit Liquet, Kerrie Mengersen, AN Pettitt, and Matt Sutton. Bayesian variable selection regression of multivariate responses for group data. *Bayesian Analysis*, 12(4):1039–1067, 2017.

Kezia Manlove, E Frances Cassirer, Paul C Cross, Raina K Plowright, and Peter J Hudson. Disease introduction is associated with a phase transition in bighorn sheep demographics. *Ecology*, 97(10): 2593–2602, 2016.

Lucile Marescot, Sarah Benhaiem, Olivier Gimenez, Heribert Hofer, Jean-Dominique Lebreton, Ximena A Olarte-Castillo, Stephanie Kramer-Schadt, and Marion L East. Social status mediates the fitness costs of infection with canine distemper virus in serengeti spotted hyenas. *Functional Ecology*, 32(5):1237–1250, 2018.

Julien Martin, J Andrew Royle, Darryl I Mackenzie, Holly H Edwards, Marc Kery, and Beth Gardner. Accounting for non-independent detection when estimating abundance of organisms with a Bayesian approach. *Methods in Ecology and Evolution*, 2(6):595–601, 2011.

Greg D Matuzak and Donald J Brightsmith. Roosting of yellow-naped parrots in costa rica: Estimating the size and recruitment of threatened populations. *Journal of Field Ornithology*, 78(2):159–169, 2007.

Rebecca McCaffery, J Joshua Nowak, and Paul M Lukacs. Improved analysis of lek count data using N-mixture models. *The Journal of Wildlife Management*, 80(6):1011–1021, 2016.

Hamish McCallum. Disease and the dynamics of extinction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1604):2828–2839, 2012.

Hamish McCallum. Models for managing wildlife disease. *Parasitology*, 143(7):805–820, 2016.

Hamish McCallum, Daniel M Tompkins, Menna Jones, Shelly Lachish, Steve Marvanek, Billie Lazenby, Greg Hocking, Jason Wiersma, and Clare E Hawkins. Distribution and impacts of tasmanian devil facial tumor disease. *EcoHealth*, 4:318–325, 2007.

Jenni L McDonald, Andrew Robertson, and Matthew J Silk. Wildlife disease ecology from the individual to the population: Insights from a long-term study of a naturally infected European badger population. *Journal of Animal Ecology*, 87(1):101–112, 2018.

Paul McLaughlin and Haim Bar. A spatial capture–recapture model with attractions between individuals. *Environmetrics*, 32(1):e2653, 2021.

Russell B Millar. Conditional vs marginal estimation of the predictive loss of hierarchical models using WAIC and cross-validation. *Statistics and Computing*, 28(2):375–385, 2018.

Cyril Milleret, Soumen Dey, Pierre Dupont, Henrik Brøseth, Daniel Turek, Perry de Valpine, and Richard Bischof. Estimating spatially variable and density-dependent survival using open-population spatial capture–recapture models. *Ecology*, 104(2):e3934, 2023.

Toby J Mitchell and John J Beauchamp. Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032, 1988.

William K Morris, Peter A Vesk, Michael A McCarthy, Sarayudh Bunyavejchewin, and Patrick J Baker. The neglected tool in the bayesian ecologist's shed: A case study testing informative priors' effect on model accuracy. *Ecology and Evolution*, 5(1):102–108, 2015.

Toshio Nakagawa and Shunji Osaki. The discrete Weibull distribution. *IEEE Transactions on Reliability*, 24(5):300–301, 1975.

Radford M Neal. Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9(2):249–265, 2000.

Grzegorz Neubauer, Alicja Wolska, Patryk Rowiński, and Tomasz Wesołowski. N-mixture models estimate abundance reliably: A field test on marsh tit using time-for-space substitution. *The Condor*, 124(1):duab054, 2022.

James D Nichols. Capture-recapture models. *BioScience*, 42(2):94–102, 1992.

Robert B O'hara and Mikko J Sillanpää. A review of Bayesian variable selection methods: what, how and which.

Christopher Paciorek. Technical vignette 5: Understanding intrinsic Gaussian Markov random field spatial models, including intrinsic conditional autoregressive models. *Technical report*, 2009.

Alina Peluso, Veronica Vinciotti, and Keming Yu. Discrete Weibull generalized additive model: an application to count fertility data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 68(3):565–583, 2019.

Shirley Pledger, Kenneth H Pollock, and James L Norris. Open capture–recapture models with heterogeneity: Ii. jolly–seber model. *Biometrics*, 66(3):883–890, 2010.

Kenneth H Pollock. A capture-recapture design robust to unequal probability of capture. *The Journal of Wildlife Management*, 46(3):752–757, 1982.

José Miguel Ponciano, J Gordon Burleigh, Edward L Braun, and Mark L Taper. Assessing parameter identifiability in phylogenetic models using data cloning. *Systematic biology*, 61(6):955–972, 2012.

Lauren C Ponisio, Perry de Valpine, Nicholas Michaud, and Daniel Turek. One size does not fit all: Customizing mcmc methods for hierarchical models using NIMBLE. *Ecology and evolution*, 10 (5):2385–2416, 2020.

Joshua K Raabe, Beth Gardner, and Joseph E Hightower. A spatial capture–recapture model to estimate fish survival and location from linear continuous monitoring arrays. *Canadian Journal of Fisheries and Aquatic Sciences*, 71(1):120–130, 2014.

LM Rogers, CL Cheeseman, PJ Mallinson, and R Clifton-Hadley. The demography of a high-density badger (meles meles) population in the west of England. *Journal of Zoology*, 242(4):705–728, 1997.

LM Rogers, R Delahay, CL Cheeseman, S Langton, GC Smith, and RS Clifton-Hadley. Movement of badgers (meles meles) in a high–density population: individual, population and disease effects. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1403):1269–1276, 1998.

Antonio Romano, Andrea Costa, Marco Basile, Ranieri Raimondi, Mario Posillico, Daniele Scinti Roger, Aldo Crisci, Riccardo Piraccini, Pasquale Raia, Giorgio Matteucci, et al. Conservation of salamanders in managed forests: Methods and costs of monitoring abundance and habitat selection. *Forest ecology and management*, 400:12–18, 2017.

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

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

J Andrew Royle and Robert M Dorazio. Parameter-expanded data augmentation for Bayesian analysis of capture–recapture models. *Journal of Ornithology*, 152(2):521–537, 2012.

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

J Andrew Royle, Angela K Fuller, and Christopher Sutherland. Unifying population and landscape ecology with spatial capture–recapture. *Ecography*, 41(3):444–456, 2018.

JA Royle, RB Chandler, R Sollmann, and B Gardner. Spatial capture-recapture. waltham, massachusetts, 2014.

Carly Rozins, Matthew J Silk, Darren P Croft, Richard J Delahay, Dave J Hodgson, Robbie A McDonald, Nicola Weber, and Mike Boots. Social structure contains epidemics and regulates individual roles in disease transmission in a group-living mammal. *Ecology and Evolution*, 8(23): 12044–12055, 2018.

Alejandro Salinas-Melgoza, Vicente Salinas-Melgoza, and Timothy F Wright. Behavioral plasticity of a threatened parrot in human-modified landscapes. *Biological Conservation*, 159:303–312, 2013.

Gláucia Helena Fernandes Seixas and Guilherme Mourao. Communal roosts of the blue-fronted amazons (amazona aestiva) in a large tropical wetland: Are they of different types? *PloS one*, 13 (10):e0204824, 2018.

Jayaram Sethuraman. A constructive definition of Dirichlet priors. *Statistica sinica*, pages 639–650, 1994.

Matthew J Silk, Nicola L Weber, Lucy C Steward, David J Hodgson, Mike Boots, Darren P Croft, Richard J Delahay, and Robbie A McDonald. Contact networks structured by sex underpin sex-specific epidemiology of infection. *Ecology letters*, 21(2):309–318, 2018.

Matthew J Silk, David J Hodgson, Carly Rozins, Darren P Croft, Richard J Delahay, Mike Boots, and Robbie A McDonald. Integrating social behaviour, demography and disease dynamics in network models: applications to disease management in declining wildlife populations. *Philosophical Transactions of the Royal Society B*, 374(1781):20180211, 2019.

David Spiegelhalter, Andrew Thomas, Nicky Best, and Dave Lunn. WinBUGS user manual, 2003.

David J Spiegelhalter, Nicola G Best, Bradley P Carlin, and Angelika Van Der Linde. Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4):583–639, 2002.

Colin E Studds, Bruce E Kendall, Nicholas J Murray, Howard B Wilson, Danny I Rogers, Robert S Clemens, Ken Gosbell, Chris J Hassell, Rosalind Jessop, David S Melville, et al. Rapid population decline in migratory shorebirds relying on yellow sea tidal mudflats as stopover sites. *Nature communications*, 8:14895, 2017.

Chris Sutherland, J Andrew Royle, and Daniel W Linden. oscr: A Spatial Capture-Recapture R package for inference about spatial ecological processes. *Ecography*, 2019.

Yee Teh, Michael Jordan, Matthew Beal, and David Blei. Sharing clusters among related groups: Hierarchical dirichlet processes. *Advances in neural information processing systems*, 17, 2004.

Chris D Thomas. Local diversity stays about the same, regional diversity increases, and global diversity declines. *Proceedings of the National Academy of Sciences*, 110(48):19187–19188, 2013.

Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288, 1996.

SG Toribio, BR Gray, and S Liang. An evaluation of the Bayesian approach to fitting the N-mixture model for use with pseudo-replicated count data. *Journal of Statistical Computation and Simulation*, 82(8):1135–1143, 2012.

Daniel Turek, Cyril Milleret, Torbjørn Ergon, Henrik Brøseth, Pierre Dupont, Richard Bischof, and Perry De Valpine. Efficient estimation of large-scale spatial capture–recapture models. *Ecosphere*, 12(2):e03385, 2021.

FAM Tuyttens, RJ Delahay, DW Macdonald, CL Cheeseman, B Long, and CA Donnelly. Spatial perturbation caused by a badger (Meles meles) culling operation: implications for the function of territoriality and the control of bovine tuberculosis (Mycobacterium bovis). *Journal of Animal Ecology*, 69(5):815–828, 2000.

Aki Vehtari, Andrew Gelman, and Jonah Gabry. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and computing*, 27:1413–1432, 2017.

J Vicente, RJ Delahay, NJ Walker, and CL Cheeseman. Social organization and movement influence the incidence of bovine tuberculosis in an undisturbed high-density badger meles meles population. *Journal of Animal Ecology*, 76(2):348–360, 2007.

Vance T Vredenburg, Roland A Knapp, Tate S Tunstall, and Cheryl J Briggs. Dynamics of an emerging disease drive large-scale amphibian population extinctions. *Proceedings of the National Academy of Sciences*, 107(21):9689–9694, 2010.

Sara Wade and Zoubin Ghahramani. Bayesian cluster analysis: Point estimation and credible balls (with discussion). *Bayesian Analysis*, 13(2):559–626, 2018.

Stephen Walker and Cristiano Villa. An objective prior from a scoring rule. *Entropy*, 23:833, 06 2021. doi: 10.3390/e23070833.

Sumio Watanabe and Manfred Opper. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of machine learning research*, 11(12), 2010.

Nicola Weber, Stuart Bearhop, Sasha RX Dall, Richard J Delahay, Robbie A McDonald, and Stephen P Carter. Denning behaviour of the European badger (Meles meles) correlates with bovine tuberculosis infection status. *Behavioral Ecology and Sociobiology*, 67(3):471–479, 2013a.

Nicola Weber, Stephen P Carter, Sasha RX Dall, Richard J Delahay, Jennifer L McDonald, Stuart Bearhop, and Robbie A McDonald. Badger social networks correlate with tuberculosis infection. *Current Biology*, 23(20):R915–R916, 2013b.

Rosie Woodroffe, Christl A Donnelly, DR Cox, F John Bourne, CL Cheeseman, RJ Delahay, George Gettinby, John P Mcinerney, and W Ivan Morrison. Effects of culling on badger Meles meles spatial organization: implications for the control of bovine tuberculosis. *Journal of Applied Ecology*, 43 (1):1–10, 2006.

Xiaofan Xu and Malay Ghosh. Bayesian variable selection and estimation for group lasso. *Bayesian Analysis*, 10(4):909–936, 2015.

Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67, 2006.

# Appendix A

# Supplementary material for Specifying and selecting N-mixture models in a Bayesian framework.

## A.1   Case 2-Model selection via WAIC MCMC settings

Table A.1 MCMC settings for scenario 1.

| $p$ | Model | MCMC iterations | Burn-in | Thinning | Chains |
|-----|-------|-----------------|---------|----------|--------|
| 0.6 | P-B | 25000 | 500 | 5 | 1 |
| | DW-B | 65000 | 5000 | 15 | 1 |
| | NB-B | 125000, | 5000 | 15 | 1 |
| | P-BB | 155000 | 5000 | 30 | 1 |
| | DW-BB | 155000 | 5000 | 30 | 1 |
| 0.25 | P-B | 225000 | 5000 | 30 | 1 |
| | DW-B | 205000 | 5000 | 50 | 1 |
| | NB-B | 405000 | 5000 | 100 | 1 |
| | P-BB | 125000 | 5000 | 10 | 1 |
| | DW-BB | 205000 | 5000 | 100 | 1 |

Table A.2 MCMC settings for scenario 2.

| $p$ | Model | MCMC iterations | Burn-in | Thinning | Chains |
|-----|-------|-----------------|---------|----------|--------|
| 0.6 | P-B | 125000 | 5000 | 30 | 1 |
| | DW-B | 205000 | 5000 | 50 | 1 |
| | NB-B | 405000 | 5000 | 100 | 1 |
| | P-BB | 125000 | 5000 | 10 | 1 |
| | DW-BB | 205000 | 5000 | 100 | 1 |
| 0.25 | P-B | 225000 | 5000 | 30 | 1 |
| | DW-B | 205000 | 5000 | 50 | 1 |
| | NB-B | 405000 | 5000 | 100 | 1 |
| | P-BB | 125000 | 5000 | 10 | 1 |
| | DW-BB | 205000 | 5000 | 100 | 1 |

Table A.3 MCMC settings for scenario 3.

| $p$ | Model | MCMC iterations | Burn-in | Thinning | Chains |
|-----|-------|-----------------|---------|----------|--------|
| 0.6 | P-B | 25000 | 5000 | 5 | 1 |
| | DW-B | 85000 | 5000 | 20 | 1 |
| | NB-B | 805000 | 5000 | 200 | 1 |
| | P-BB | 205000 | 5000 | 100 | 1 |
| | DW-BB | 205000 | 5000 | 100 | 1 |
| 0.25 | P-B | 85000 | 5000 | 20 | 1 |
| | DW-B | 85000 | 5000 | 20 | 1 |
| | NB-B | 805000 | 5000 | 200 | 1 |
| | P-BB | 505000 | 5000 | 100 | 1 |
| | DW-BB | 505000 | 5000 | 100 | 1 |

Table A.4 MCMC settings for scenario 4.

| $p$ | Model | MCMC iterations | Burn-in | Thinning | Chains |
|---|---|---|---|---|---|
| 0.6 | P-B | 125000 | 5000 | 30 | 1 |
| | DW-B | 405000 | 5000 | 100 | 1 |
| | NB-B | 805000 | 5000 | 200 | 1 |
| | P-BB | 805000 | 5000 | 200 | 1 |
| | DW-BB | 805000 | 5000 | 200 | 1 |
| 0.25 | P-B | 125000 | 5000 | 30 | 1 |
| | DW-B | 405000 | 5000 | 100 | 1 |
| | NB-B | 805000 | 5000 | 200 | 1 |
| | P-BB | 805000 | 5000 | 200 | 1 |
| | DW-BB | 805000 | 5000 | 200 | 1 |

## A.2   Case Studies

Table A.5 MCMC settings used for analyzing yellow-bellied toads.

| Model | MCMC iterations | Burn-in | Thinning | Chains |
|---|---|---|---|---|
| P-B | 25000 | 5000 | 5 | 2 |
| DW-B | 900000 | 40000 | 25 | 2 |
| NB-B | 700000 | 50000 | 15 | 2 |
| P-BB | 605000 | 50000 | 15 | 2 |
| DW-BB | 605000 | 50000 | 15 | 2 |

Table A.6 MCMC settings used for analyzing Swiss great tits.

| Model | MCMC iterations | Burn-in | Thinning | chains |
|---|---|---|---|---|
| P-B | 25000 | 5000 | 5 | 2 |
| DW-B | 900000 | 40000 | 25 | 2 |
| NB-B | 350000 | 50000 | 15 | 2 |
| P-BB | 605000 | 50000 | 15 | 2 |
| DW-BB | 305000 | 60000 | 15 | 2 |

# Appendix B

# Supplementary material for A new modelling framework for roost count data.

## B.1    Dirichlet Process (DP) Mixture Model

The Dirichlet process (DP) mixture model can be viewed as an infinite dimensional mixture model which represents an unknown density $F(y)$ taking the form

$$F(y) = \int k(y|\theta)p(\theta|G)d\theta,$$

$$G|\alpha, G_0 \sim DP(\alpha, G_0)$$

where $k(.|\theta)$ is a suitable mixture kernel with parameter vector $\theta$. The mixing distribution/random measure $G$ is a Dirichlet process prior with $\alpha$ and $G_0$ as the concentration parameter and baseline distribution respectively. Any prior knowledge about $F(y)$ is summarised in the baseline distribution while the concentration parameter specifies the prior variance and controls relatively how much the prior and the data contribute to the posterior. More information on the DP can be found in Escobar and West (1995).

One approach to implement the DP mixture model is to integrate out the random measure $G$ from the model. This allows the DP mixture model to be written using latent or membership variables, $z_i$, following a Chinese Restaurant Process (CRP) distribution (Blackwell and MacQueen, 1973). As a result, for $i = 1, \ldots, N$ data points, the DP mixture model takes the form

$$F(y)|\tilde{\theta}, z_i \sim k(.|\tilde{\theta}_{z_i}),$$

$$z|\alpha \sim \mathrm{CRP}(\alpha), \quad \tilde{\theta}_j \sim G_0,$$

where $\mathrm{CRP}(\alpha)$ denotes the CRP distribution with concentration parameter $\alpha$. The CRP can be described as a stochastic process. Customers arrive at a restaurant one at a time with a potentially infinite number of tables and are seated in the order in which they arrive. The first customer will sit at the first table. The second customer will be seated at the first table with a probability of $1/(1+\alpha)$ or at the second (empty) table with a probability of $\alpha/(1+\alpha)$ where $\alpha \in \mathbb{R}^+$. Following customers can sit at any of the occupied tables with a probability that is proportional to the number of customers who are already sitting there or with a probability proportional to $\alpha$ at the next unoccupied table. In this way, the CRP partitions the set of customers by assigning them to the tables in the restaurant. The stick-breaking representation (Sethuraman, 1994) is another approach that can be used to implement DP mixture models.

Another advantage of DP mixture models is that different choices for kernel $k$ and baseline distribution $G_0$ can be chosen. In a real space, a univariate or multivariate Gaussian mixture is often used. However, this kernel is not an appropriate application when considering distributions with bounded intervals $[0, T]$ or survival analysis. Consequently, Kottas (2006) considered Dirichlet process mixtures of Beta distributions for the non-parametric estimation of continuous distributions that are defined on a bounded interval, $[0, T]$.

## B.2 Bayesian Group Lasso Spike and Slab (BGLSS)

Consider the simple model

$$Y_s \sim \text{Binomial}(K, p_s)$$

$$\text{logit}(p_s) = \eta = \mu + \sum_{g=1}^{G} X_{s,g} \beta_g \tag{B.2.1}$$

for $s = 1, \ldots, S$ data points, $g = 1, \ldots, G$ are continuous/categorical variables, such that variable $g$ requires $C_g$ coefficients to model its effect, so that if $g$ is a continuous variable, $C_g = 1$, and if $g$ is a categorical variable, $C_g$ is its number of levels (excluding baseline). Finally, $\beta_g$ is the $(C_g \times 1)$ vector corresponding to the logistic regression coefficients for variable $g$, $X_{s,g}$ is an $(S \times G)$ dimensional design matrix relative to variable $g$.

Within such a logistic regression setting, Bayesian variable selection (BVS) methods can be used to understand and identify significant covariate effects on observation error. Bayesian Group Lasso Spike and Slab (BGLSS) is one BVS technique that can be used within a logistic regression to understand covariate effects. Bayesian group lasso spike and slab (BGLSS) is a BVS method specifically developed to perform variable selection on categorical variables, but it can also perform variable selection on continuous variables. Essentially, it employs a spike and slab prior where the slab density is the Bayesian group lasso density (Casella et al., 2010).

The spike and slab prior introduced by Mitchell and Beauchamp (1988), typically in a logistic regression setting has the form

$$p(\beta|\gamma) = \prod_{g=1}^{G} [(1 - \gamma_g)\delta_0(\beta_g) + \gamma_g p(\beta_g)],$$

$$p(\gamma|\phi) = \prod_{g=1}^{G} \phi^{\gamma_g}(1 - \phi)^{1 - \gamma_g},$$

$$\phi \sim p(\phi)$$

where the slab, $p(\beta_g)$, is a diffuse and/or heavy-tailed density to model the distribution of coefficients of covariates included in the model, the spike, $\delta_0$, is a point mass at zero used to model excluded covariates, $\gamma$ is a binary vector that indexes the $2^G$ possible models, and $\phi \in (0,1)$ is a mixing proportion. Importantly, with a well-chosen prior on $\phi$, this prior avoids the curse of dimensionality

by automatically favouring parsimonious models in high dimensions (Bai et al., 2021). The mixing proportion can be fixed to 0.5 or can be assigned a uniform or Beta prior distribution. However, in high dimensions, this point-mass spike and slab prior can be computationally expensive to explore the full posterior mainly due to the combinatorial challenge of updating $\gamma$.

The Bayesian group lasso is the Bayesian formulation of the group lasso (Yuan and Lin, 2006) which was developed to facilitate group-level selection in a frequentist setting. This Bayesian formulation allows the shrinkage parameter ($\psi$) to be estimated simultaneously with model parameters, possibly reducing computational cost by rather testing several values of $\psi$ via cross-validation as done in a classical setting. For $G$ categorical predictors with each categorical covariate $g$ contains $m_g$ levels such that $\sum_{g=1}^{G} m_g = \rho$, The Bayesian group lasso density can be defined as

$$\pi(\beta_g | \psi) = m_g \psi^{C_g} \exp(-\psi \parallel \beta_g \parallel_2)$$

where $m_g = 2^{-C_g} \pi^{-(C_g - 1)/2} [\Gamma((C_g + 1)/2)]^{-1}$, $\beta_g$ is the vector regression effects relating to covariate $g$, and for $v \in \mathbb{R}^\rho$, $||v||_2 = \sqrt{\sum_{i=1}^{\rho} v_i^2}$ (Bai et al., 2020). Casella et al. (2010) and Xu and Ghosh (2015) both used this prior for Bayesian inference in a grouped regression model. Importantly, Casella et al. (2010) showed that the Bayesian group lasso in a linear regression framework can be written as a scaled mixture of a Normal distribution with a Gamma distribution on the hyper parameter. In a logistic regression framework this can be expressed as:

$$\beta_g | \tau_g \sim N_{m_g}(0, \tau_g^2 I_{C_g})$$
$$\tau_g^2 \sim \text{Gamma}\left( \frac{C_g + 1}{2}, \frac{\psi^2}{2} \right)$$

Unfortunately, this prior does not provide null regression effects. Consequently, Xu and Ghosh (2015) proposed a point mass spike and slab prior:

$$\beta_g | \tau_g^2 \sim \gamma_g N(0, \tau_g^2 I_{m_g}) + (1 - \gamma_g)\delta_0(\beta_g)$$

$$\tau_g^2 \sim \text{Gamma}\left(\frac{C_g + 1}{2}, \frac{\psi^2}{2}\right)$$

$$\gamma_g \sim \text{Bernoulli}(\phi_g)$$

$$\psi \sim \text{Gamma}(a, b)$$

where $\delta_0$ is the Dirac distribution on 0, $\gamma_g$ is a binary variable which indicates whether factor $g$ was included or not, and $I_{C_g}$ is the identity matrix ($C_g \times C_g$). This approach can easily be implemented using MCMC methods. The prior inclusion probability $\phi_g$ can be fixed to 0.5 or can be assigned a uniform or Beta prior distribution. The advantage of BGLSS is that it requires little tuning by the user. In addition, Jreich et al. (2021) showed the BGLSS preformed better than the Bayesian Group Lasso. The disadvantage is that BGLSS can only identify significant categorical covariates not significant levels of categorical variables. In addition, Xu and Ghosh (2015) reported poor selection performance for high-dimensional predictors, in particular when most levels of categorical predictors have no effect.

## B.3 Bayesian Sparse Group Selection (BSGS)

Bayesian Sparse Group selection (BSGS) developed by Chen et al. (2016) enables variable selection of both continuous and categorical variables. It has the advantage of identifying both significant categorical covariates and their relative levels. BSGS models the joint behaviour of binary indicators for grouped levels by adding a hierarchical layer to the stochastic search variable selection (SSVS) method by George and McCulloch (1993). These two levels of binary indicators are nested so that if a categorical variable is removed from the model, all its levels are also removed. Hierarchically, for each level $l$, $l = 1, \ldots, C_g$ of a categorical covariate $g$, $g = 1, \ldots, G$ the prior distribution of $v_{lg}$ (the binary indicator relative to level l within the group $g$) and $\gamma_g$ (the binary indicator of covariate $g$) are

defined as follows:

$$\gamma_g \sim \text{Bernoulli}(p_g)$$

$$v_{lg}|\gamma_g \sim (1-\gamma_g)\delta_0 + \gamma_g\text{Bernoulli}(p_{lg})$$

where $\delta_0$ is the Dirac distribution on 0, $p_g$ and $p_{lg}$ are the prior inclusion probabilities for categorical covariate $g$ and level $l$ within covariate $g$, respectively. For the regression effects, a zero inflated mixture Gaussian prior is proposed:

$$\beta_{lg}|v_{lg}\gamma_g \sim (1-v_{lg}\gamma_g)\delta_0 + v_{lg}\gamma_g N(0, \tau_{lg}^2)$$

This approach requires user tuning on three parameters: $p_g, p_{lg}$ and $\tau_{lg}$. For the inclusion probabilities, they can be set to 0.5 or a Beta or Uniform distribution can be assigned. A prior can be assigned to $\tau_{lg}$ or cross validation techniques can be employed to select $\tau_{lg}$. Importantly, the performance of the BSGS is sensitive to $\tau_{lg}$ and the total number of levels considered within a categorical covariate (Jreich et al., 2021). As $\beta_{lg}$ is inversely proportional to $\tau_{lg}$ and the number of levels, a large value of $\tau_{lg}$ or an important number of levels can decrease the prior inclusion probability. Barbieri and Berger (2004) posterior median model is often used as the model selection criterion in BSGS.

We investigate BSGS performance using the same simulation settings as case 2 in Chapter 3 but with $\beta = (1, 1.25, 0.2, 2, 0, -0.6, -0.5, -0.3, 0)$ used for $p \approx 0.6$ and $\beta = (-1.5, -0.2, -0.6, 0, -0.5, -0.8, 0.8, 0.3, 0)$ used for $p \approx 0.3$. Also, 30 simulation runs were conducted for $p \approx 0.6$ and 50 for $p \approx 0.3$. To introduce misspecification in the model for detection, variables $x_5$ and $x_6$ were not included in the model

As can be seen from Tables B.1 and B.2, $\lambda$ and $\sigma$ are estimated well in both cases at both levels of detection with good coverage, low RMSE and relative bias. In addition, weak effects are estimated poorly in case b) for both levels of detection with low coverage, high RMSE and relative bias. Contrary, in case a), these weak effects are estimated well with high coverage, low RMSE and relative bias. Notably, as can be seen in Fig. B.1, population size is estimated well with relatively low

relative bias in both cases for both levels of detection. In these cases, population size coverage was $\geq 80$.

In case a) and b), for both levels of detection, all strong significant variables were correctly identified by BSGS within the global covariate ($\gamma_g$) and local covariate level ($v_{lg}$) in all runs. In addition, in case a), 1/30 and 2/50 simulation runs a non-significant variable was identified as significant for $p \approx (0.6, 0.3)$ respectively within the global covariate level. In case b), 0/30 and 1/50 runs a non-significant variable was identified as significant for $p \approx (0.6, 0.3)$ respectively with the global covariate level. At the local covariate level, in case a), 2/30, 1/50 runs a non-significant variable was identified as significant for $p \approx 0.6, 0.3$ respectively. In case b), at the local covariate level, 3/30 and 4/50 runs a non-significant variable was identified as significant for $p \approx 0.6, 0.3$ respectively.

For $p \approx 0.6$, $\beta = 0.2$ was correctly identified in 30/30 runs at the global and local level in case a) whilst, in case b), this weak effect was correctly identified in 22/30, 13/30 runs at the global and local level respectively. For the weak categorical covariate, $\beta = (-0.3, 0)$, $\beta = -0.3$ was correctly identified 29/30, 28/30 runs at the global and local level in case a) whilst, in case b), 11/30, 10/30 runs were correctly identified at the global and local level. For the non-significant categorical level, 28/30 and 27/30 runs were correctly identified as non-significant in cases a) and b) respectively.

For $p \approx 0.3$, $\beta = -0.2$ was correctly identified in 50/50 runs at the global and local level in case a) whilst, in case b), this weak effect was correctly identified in 38/50, 22/50 runs at the global and local level respectively. The weak categorical covariate effect $\beta = 0.3$ was correctly identified 50/50, 48/50 runs at the global and local level in case a) whilst, in case b), 24/50, 15/50 runs were correctly identified at the global and local level. The non-significant categorical level was correctly identified as non-significant in case a) and b), 49/50 and 46/50 runs respectively.

This simulation study demonstrated that BSGS can be used as a BVS technique when covariates fully explain detection. Notably, like the BGLSS, with the addition of random effect, the strength to identify weak effects in both cases of $p$ decreased, while strong effects are correctly identified irrespective of random effects. This simulation seems to suggest that BGLSS performs better than BSGS as BGLSS estimated weak effects better. The main reason why BSGS performed worse is most likely due to the sensitivity present with the slab standard deviation (Jreich et al., 2021). A prior was used in this simulation but it was found that reducing the standard deviation by using fixed

values improved the ability of BSGS to identify weak effects but also increased opportunities for non-significant variables to be identified as significant. Hence, it was hard to select fixed values of the slab standard deviation.

Table B.1 Simulation results for BSGS for $p \approx 0.6$.

| Case | Parameters | Median | Coverage | Residual Mean Square | Relative Bias |
|------|-----------|--------|----------|---------------------|---------------|
| a) | $\beta_0$ | 1.011 | 87 | 0.044 | 0.011 |
|    | $\beta_1$ | 1.257 | 93 | 0.026 | 0.006 |
|    | $\beta_2$ | 0.203 | 93 | 0.126 | 0.015 |
|    | $\beta_3$ | 2.003 | 100 | 0.015 | 0.002 |
|    | $\beta_4$ | 0 | 0 | – | – |
|    | $\beta_5$ | $-0.603$ | 93 | $-0.054$ | 0.005 |
|    | $\beta_6$ | $-0.502$ | 93 | $-0.052$ | 0.004 |
|    | $\beta_7$ | $-0.306$ | 87 | $-0.301$ | 0.02 |
|    | $\beta_8$ | 0 | 0 | – | – |
|    | $\lambda$ | 93.381 | 90 | 0.105 | $-0.066$ |
|    | $\sigma$ | 0.982 | 100 | 0.143 | $-0.018$ |
| b) | $\beta_0$ | 0.911 | 97 | 0.104 | $-0.089$ |
|    | $\beta_1$ | 1.247 | 97 | 0.051 | $-0.002$ |
|    | $\beta_2$ | 0 | 70 | 0.788 | $-1$ |
|    | $\beta_3$ | 1.994 | 93 | 0.044 | $-0.003$ |
|    | $\beta_4$ | 0 | 0 | – | – |
|    | $\beta_7$ | 0 | 63 | $-0.877$ | $-1$ |
|    | $\beta_8$ | 0 | 6.7 | – | – |
|    | $\lambda$ | 93.971 | 90 | 0.103 | $-0.060$ |
|    | $\sigma$ | 0.986 | 100 | 0.156 | $-0.014$ |

Table B.2 Simulation results for BSGS for $p \approx 0.3$.

| Case | Parameters | Median | Coverage | Residual Mean Square | Relative Bias |
|------|-----------|--------|----------|---------------------|---------------|
| a) | $\beta_0$ | $-1.506$ | 96 | $-0.039$ | 0.004 |
| | $\beta_1$ | $-0.195$ | 100 | $-0.103$ | $-0.023$ |
| | $\beta_2$ | $-0.594$ | 98 | $-0.049$ | $-0.009$ |
| | $\beta_3$ | 0 | 0 | – | – |
| | $\beta_4$ | $-0.499$ | 92 | $-0.062$ | $-0.005$ |
| | $\beta_5$ | $-0.802$ | 92 | $-0.046$ | 0.003 |
| | $\beta_6$ | 0.799 | 94 | 0.038 | $-0.003$ |
| | $\beta_7$ | 0.302 | 94 | 0.256 | 0.008 |
| | $\beta_8$ | 0 | 42 | – | – |
| | $\lambda$ | 91.606 | 94 | 0.135 | $-0.84$ |
| | $\sigma$ | 1.016 | 100 | 0.242 | 0.016 |
| b) | $\beta_0$ | $-1.496$ | 94 | $-0.096$ | $-0.002$ |
| | $\beta_1$ | 0 | 82 | $-0.787$ | $-1$ |
| | $\beta_2$ | $-0.586$ | 96 | $-0.132$ | $-0.023$ |
| | $\beta_3$ | 0 | 0 | – | – |
| | $\beta_4$ | $-0.472$ | 96 | $-0.170$ | $-0.056$ |
| | $\beta_7$ | 0 | 72 | 0.891 | $-1$ |
| | $\beta_8$ | 0 | 56 | – | – |
| | $\lambda$ | 90.325 | 88 | 0.153 | $-0.097$ |
| | $\sigma$ | 1.164 | 94 | 2.369 | 0.164 |

(a) Case a) $p \approx 0.6$



(b) Case b) $p \approx 0.6$



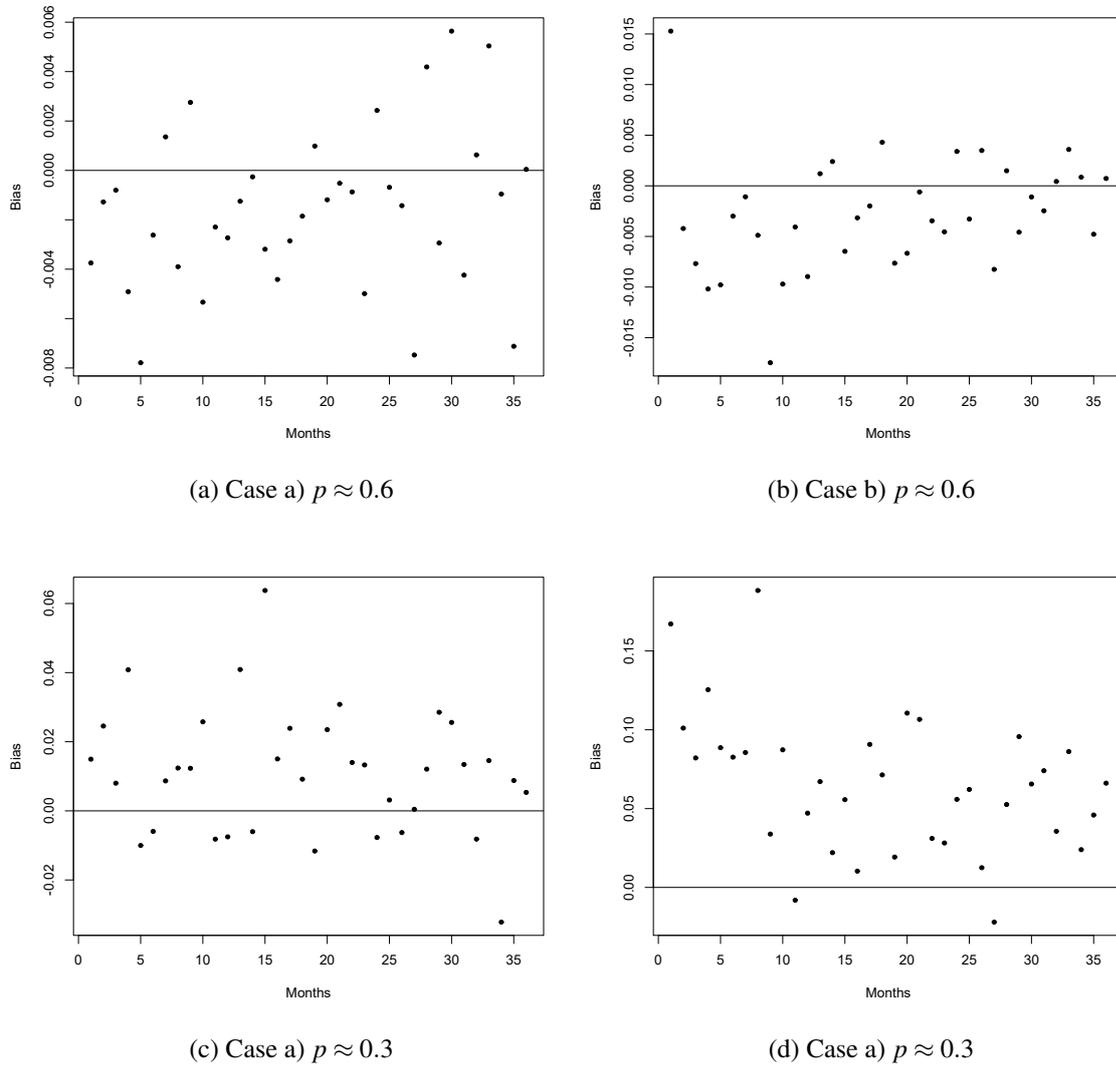(c) Case a) $p \approx 0.3$



(d) Case a) $p \approx 0.3$

Fig. B.1 Population size median relative bias obtained for BSGS at case a) and b) when $p \approx (0.6, 0.3)$.

## B.4    Temporal Models

### B.4.1    Random Walk Order 1 (RW1)

In this case, $\theta_\ell$, $\ell = 2, \ldots T \cdot Y$ can be defined as

$$\theta_\ell = \theta_{\ell-1} + \varepsilon_\ell \tag{B.4.1}$$

where $\varepsilon_\ell \sim N(0, \sigma^2)$ are independent and identically distributed random noise effect terms. In a Bayesian framework, $\theta_1$ can be assigned a vague prior (Fahrmeir and Lang, 2001). For $T \cdot Y \geq 3$, the conditional distributions of equation (B.4.1) can be written as

$$
\theta_\ell | \theta_{-\ell}, \sigma^2 = \begin{cases} N(\theta_{\ell+1}, \sigma^2) & t = 1 \\[2em] N\left(\frac{\theta_{\ell-1} + \theta_{\ell+1}}{2}, \frac{\sigma^2}{2}\right) & t = 2, \ldots, (T \cdot Y) - 1 \\[2em] N(\theta_{\ell-1}, \sigma^2) & t = T \cdot Y \end{cases}
$$

## B.4.2   Random Walk Order 2 (RW2)

In this case, $\theta_\ell$, $T \cdot Y \geq 3$, $\ell = 3, \ldots, T \cdot Y$ can be defined as

$$
\theta_\ell = 2\theta_{\ell-1} - \theta_{\ell-2} + \varepsilon_\ell \tag{B.4.2}
$$

where $\varepsilon_\ell \sim N(0, \sigma^2)$ are iid noise effect terms. Vague priors can be assigned to $\theta_1$ and $\theta_2$. Like the RW1 model, the RW2 model (equation (B.4.2)) can be defined as a set of conditional probability distributions

$$\theta_\ell | \theta_{-\ell}, \sigma^2 = \begin{cases} N(2\theta_{\ell+1} - \theta_{\ell+2}, \sigma^2) & t = 1 \\[2em] N\left( \frac{2\theta_{\ell-1} + 4\theta_{\ell+1} - \theta_{\ell+2}}{5}, \frac{\sigma^2}{5} \right) & t = 2 \\[2em] N\left( \frac{-\theta_{\ell-2} + 4\theta_{\ell-1} + 4\theta_{\ell+1} - \theta_{\ell+2}}{6}, \frac{\sigma^2}{6} \right) & t = 3, \ldots, (T \cdot Y) - 2 \\[2em] N\left( \frac{-\theta_{\ell-2} + 4\theta_{\ell-1} + 2\theta_{\ell+1}}{5}, \frac{\sigma^2}{5} \right) & t = (T \cdot Y) - 2 \\[2em] N(-\theta_{\ell-2} + 2\theta_{\ell-1}, \sigma^2) & t = T \cdot Y \end{cases}$$

## B.5   Simulation Study

For case 1. a) 55000 MCMC iterations, burn-in of 10000 and thinning of 20 was used at both levels of detection, and for case 1. b) 55000 MCMC iterations, burn-in of 10000 and thinning of 20 was used at high detection level while 150000 MCMC iteration, burn-in of 35000, thinning of 20 was used at low detection level. For case 2. a) and b) 55000 MCMC iterations, burn-in of 10000 and thinning of 20 was used at the high detection level while 150000 MCMC iterations, burn-in of 35000, thinning of 20 was used at the low detection level. In both cases, 2 MCMC chains were run and convergence was assessed using the Gelman-Rubin statistic (Rhat) which compares within-chain variance to between-chain variance (Gelman and Rubin, 1992). Rhat value below 1.1 indicated convergence.

## B.5.1   Case 1. a)

Table B.3 Case 1. a) simulation results for $p \approx 0.6$.
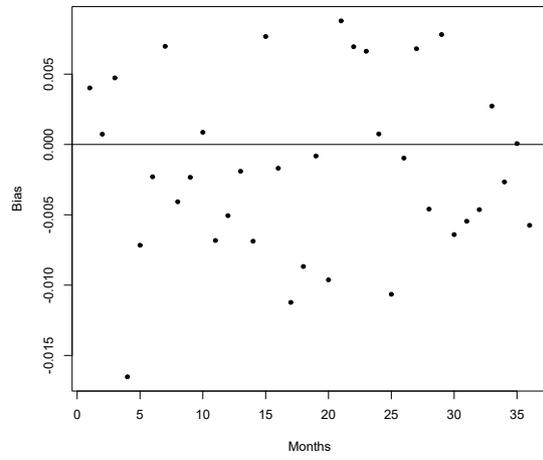
| Model | Parameters | Median | Coverage | Residual Mean Square | Relative Bias |
|---|---|---|---|---|---|
| DP | $\beta_0$ | 0.754 | 92 | 0.069 | 0.005 |
| | $\beta_1$ | 1.260 | 94 | 0.024 | 0.008 |
| | $\beta_2$ | 0.199 | 96 | 0.109 | −0.001 |
| | $\beta_3$ | 2.005 | 92 | 0.022 | 0.003 |
| | $\beta_4$ | −0.005 | 90 | – | – |
| | $\beta_5$ | −0.597 | 88 | −0.046 | −0.004 |
| | $\beta_6$ | 0.494 | 96 | 0.079 | −0.012 |
| | $\beta_7$ | −0.999 | 92 | −0.054 | −0.001 |
| | $\beta_8$ | −0.005 | 94 | – | – |
| | $\lambda$ | 104.511 | 94 | 0.104 | 0.045 |
| RW1 | $\beta_0$ | 0.744 | 92 | 0.076 | −0.008 |
| | $\beta_1$ | 1.254 | 96 | 0.028 | 0.003 |
| | $\beta_2$ | 0.204 | 94 | 0.118 | 0.021 |
| | $\beta_3$ | 1.999 | 98 | 0.021 | −0.000 |
| | $\beta_4$ | −0.002 | 96 | – | – |
| | $\beta_5$ | −0.602 | 100 | −0.040 | 0.002 |
| | $\beta_6$ | 0.509 | 98 | 0.100 | 0.018 |
| | $\beta_7$ | −0.986 | 96 | −0.062 | −0.014 |
| | $\beta_8$ | 0.008 | 88 | – | – |
| | $\lambda$ | 91.542 | 92 | 0.1222 | −0.084 |
| | $\sigma$ | 1.011 | 98 | 0.158 | 0.012 |

Table B.4 Case 1. a) simulation results for $p \approx 0.3$.

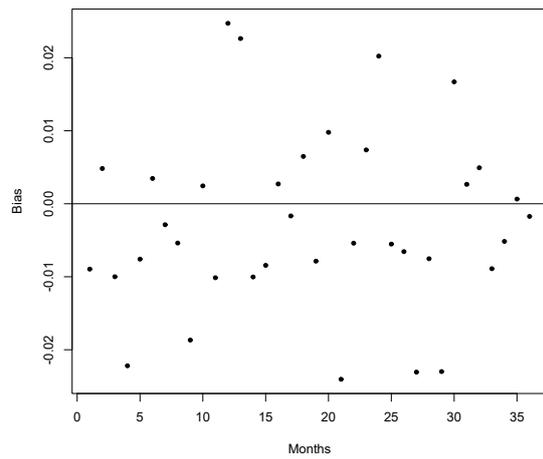| Model | Parameters | Median | Coverage | Residual Mean Square | Relative Bias |
|---|---|---|---|---|---|
| DP | $\beta_0$ | $-1.493$ | 96 | $-0.030$ | $-0.004$ |
| | $\beta_1$ | 1.252 | 98 | 0.024 | 0.002 |
| | $\beta_2$ | 0.203 | 98 | 0.114 | 0.015 |
| | $\beta_3$ | 2.017 | 96 | 0.020 | 0.008 |
| | $\beta_4$ | 0.003 | 94 | – | – |
| | $\beta_5$ | $-0.602$ | 98 | $-0.042$ | 0.004 |
| | $\beta_6$ | 0.493 | 100 | 0.067 | $-0.012$ |
| | $\beta_7$ | $-1.009$ | 98 | $-0.060$ | 0.009 |
| | $\beta_8$ | $-0.004$ | 96 | – | – |
| | $\lambda$ | 106.546 | 94 | 0.121 | 0.065 |
| RW1 | $\beta_0$ | $-1.485$ | 96 | $-0.048$ | $-0.009$ |
| | $\beta_1$ | 1.246 | 96 | 0.030 | $-0.003$ |
| | $\beta_2$ | 0.202 | 96 | 0.122 | 0.013 |
| | $\beta_3$ | 2.002 | 98 | 0.024 | 0.001 |
| | $\beta_4$ | $-0.004$ | 94 | – | – |
| | $\beta_5$ | $-0.603$ | 94 | $-0.056$ | 0.004 |
| | $\beta_6$ | 0.483 | 98 | 0.103 | $-0.034$ |
| | $\beta_7$ | $-1.011$ | 96 | $-0.076$ | 0.011 |
| | $\beta_8$ | 0.009 | 90 | – | – |
| | $\lambda$ | 93.803 | 92 | 0.123 | $-0.062$ |
| | $\sigma$ | 0.964 | 100 | 0.161 | $-0.035$ |
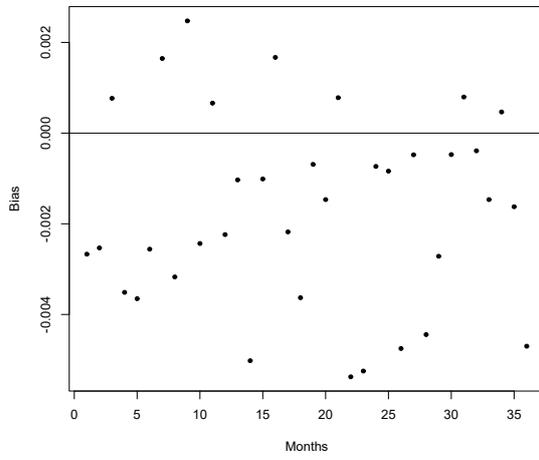
(a) DP $p \approx 0.6$

(b) DP $p \approx 0.3$
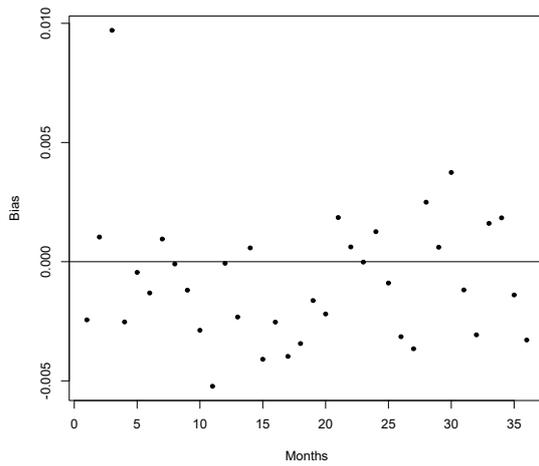
(c) RW1 $p \approx 0.6$

(d) RW1 $p \approx 0.3$

Fig. B.2 Case 1. a) Population size (N) median relative bias obtained from DP and RW1 model when $p \approx (0.6, 0.3)$.

## B.5.2   Case 1. b)

Table B.5 Case 1. b) Simulation results for $p \approx 0.6$ for the DP model.

| Model | Parameters | Median | Coverage | Residual Mean Square | Relative Bias |
|---|---|---|---|---|---|
| Fixed | $\beta_0$ | $-2.735$ | 0 | 4.607 | $-4.647$ |
| | $\beta_2$ | 0.042 | 2 | 0.776 | $-0.787$ |
| | $\beta_3$ | 0.492 | 0 | 0.747 | $-0.754$ |
| | $\beta_4$ | 0.001 | 60 | – | – |
| | $\beta_5$ | $-0.151$ | 0 | $-0.747$ | $-0.747$ |
| | $\beta_6$ | $-0.251$ | 0 | $-0.735$ | $-0.749$ |
| | $\beta_8$ | 0.027 | 46 | – | – |
| | $\lambda$ | 1162.675 | 0 | 11.093 | 10.626 |
| Mixed | $\beta_0$ | 0.890 | 78 | 0.294 | 0.187 |
| | $\beta_2$ | 0.207 | 94 | 0.445 | 0.036 |
| | $\beta_3$ | 1.993 | 96 | 0.062 | $-0.003$ |
| | $\beta_4$ | $-0.004$ | 100 | – | – |
| | $\beta_5$ | $-0.593$ | 96 | $-0.145$ | $-0.012$ |
| | $\beta_6$ | $-0.917$ | 96 | $-0.222$ | $-0.083$ |
| | $\beta_8$ | 0.103 | 86 | – | – |
| | $\lambda$ | 106.501 | 94 | 0.117 | 0.065 |

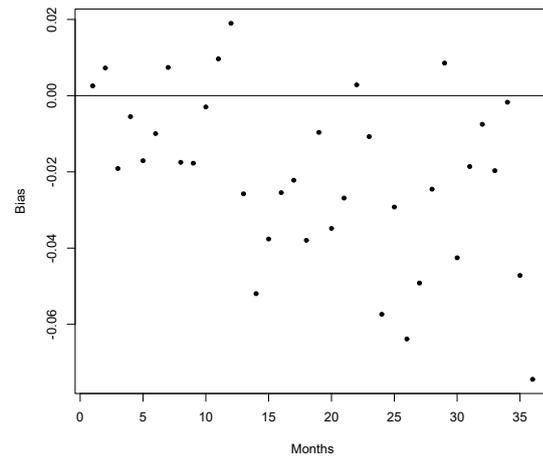Table B.6 Case 1. b) Simulation results for $p \approx 0.3$ for the DP model.

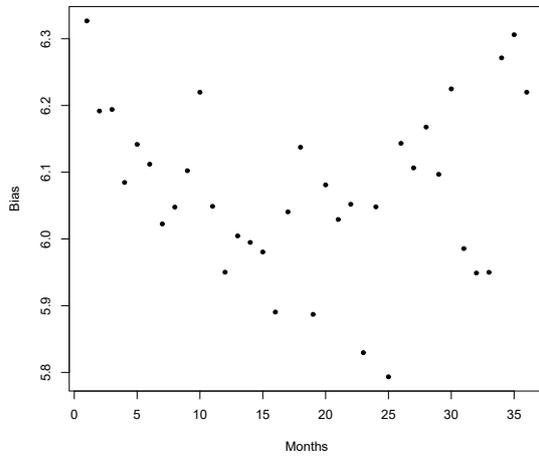| Model | Parameters | Median | Coverage | Residual Mean Square | Relative Bias |
|-------|-----------|--------|----------|---------------------|---------------|
| Fixed | $\beta_0$ | −3.085 | 2 | −1.017 | 1.057 |
|       | $\beta_2$ | 0.082 | 4 | 0.647 | −0.544 |
|       | $\beta_3$ | 0.912 | 0 | 0.542 | −0.544 |
|       | $\beta_4$ | −0.003 | 46 | – | – |
|       | $\beta_5$ | −0.268 | 0 | −0.541 | −0.554 |
|       | $\beta_6$ | −0.457 | 0 | −0.532 | −0.543 |
|       | $\beta_8$ | 0.008 | 44 | – | – |
|       | $\lambda$ | 800.167 | 0 | 7.553 | 7.002 |
| Mixed | $\beta_0$ | −1.108 | 34 | −0.273 | −0.261 |
|       | $\beta_2$ | 0.202 | 90 | 0.506 | 0.011 |
|       | $\beta_3$ | 2.032 | 92 | 0.065 | 0.016 |
|       | $\beta_4$ | −0.008 | 100 | – | – |
|       | $\beta_5$ | −0.609 | 98 | −0.148 | 0.016 |
|       | $\beta_6$ | −1.088 | 94 | −0.213 | 0.088 |
|       | $\beta_8$ | 0.090 | 92 | – | – |
|       | $\lambda$ | 112.898 | 92 | 0.188 | 0.129 |

Table B.7 Case 1. b) Simulation results for $p \approx 0.6$ for the RW1 model.

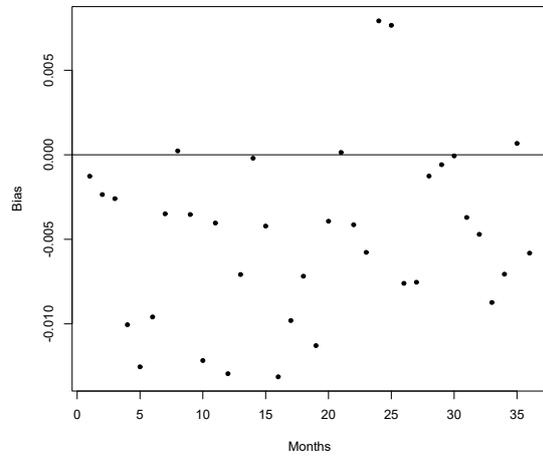| Model | Parameters | Median | Coverage | Residual Mean Square | Relative Bias |
|-------|-----------|--------|----------|----------------------|---------------|
| Fixed | $\beta_0$ | $-2.432$ | 0 | 4.105 | $-4.243$ |
| | $\beta_2$ | 0.051 | 4 | 0.728 | $-0.744$ |
| | $\beta_3$ | 0.508 | 0 | 0.723 | $-0.745$ |
| | $\beta_4$ | 0.009 | 56 | - | $-$ |
| | $\beta_5$ | $-0.156$ | 2 | $-0.731$ | $-0.740$ |
| | $\beta_6$ | $-0.286$ | 0 | $-0.708$ | $-0.713$ |
| | $\beta_8$ | 0.002 | 62 | $-$ | $-$ |
| | $\lambda$ | 815.879 | 6 | 7.709 | 7.158 |
| | $\sigma$ | 0.984 | 94 | 0.232 | $-0.016$ |
| | | | | | |
| Mixed | $\beta_0$ | 0.888 | 90 | 0.248 | 0.018 |
| | $\beta_2$ | 0.209 | 98 | 0.451 | 0.046 |
| | $\beta_3$ | 1.952 | 96 | 0.060 | $-0.023$ |
| | $\beta_4$ | $-0.018$ | 90 | $-$ | $-$ |
| | $\beta_5$ | $-0.574$ | 98 | $-0.147$ | $-0.043$ |
| | $\beta_6$ | $-0.880$ | 86 | $-0.257$ | $-0.120$ |
| | $\beta_8$ | 0.089 | 88 | $-$ | $-$ |
| | $\lambda$ | 91.641 | 90 | 0.122 | $-0.084$ |
| | $\sigma$ | 1.002 | 96 | 0.171 | 0.002 |

Table B.8 Case 1. b) Simulation results for $p \approx 0.3$ for the RW1 model.

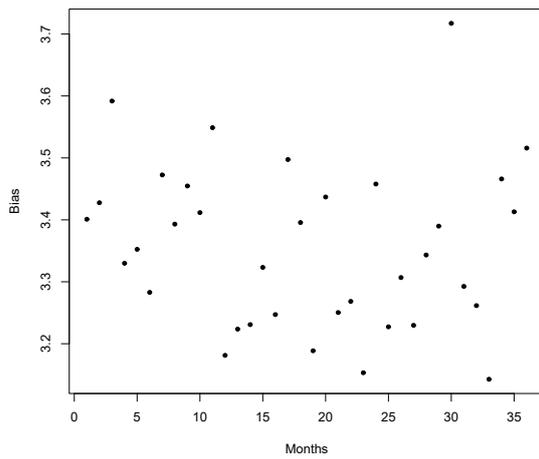| Model | Parameters | Median | Coverage | Residual Mean Square | Relative Bias |
|---|---|---|---|---|---|
| Fixed | $\beta_0$ | $-2.789$ | 0.04 | $-0.872$ | 0.859 |
| | $\beta_2$ | 0.010 | 12 | 0.593 | $-0.502$ |
| | $\beta_3$ | 0.971 | 0 | 0.509 | $-0.514$ |
| | $\beta_4$ | 0.012 | 50 | – | – |
| | $\beta_5$ | $-0.292$ | 2 | $-0.523$ | $-0.512$ |
| | $\beta_6$ | $-0.507$ | 0 | $-0.522$ | $-0.493$ |
| | $\beta_8$ | $-0.003$ | 50 | – | – |
| | $\lambda$ | 537.499 | 0 | 5.519 | 4.375 |
| | $\sigma$ | 0.982 | 90 | 0.238 | $-0.017$ |
| Mixed | $\beta_0$ | $-1.169$ | 42 | $-0.231$ | $-0.220$ |
| | $\beta_2$ | 0.208 | 96 | 0.507 | 0.043 |
| | $\beta_3$ | 1.953 | 80 | 0.078 | $-0.023$ |
| | $\beta_4$ | 0.008 | 94 | – | – |
| | $\beta_5$ | $-0.574$ | 92 | $-0.166$ | $-0.042$ |
| | $\beta_6$ | $-1.042$ | 94 | $-0.207$ | 0.042 |
| | $\beta_8$ | $-0.129$ | 98 | – | – |
| | $\lambda$ | 90.324 | 90 | 0.139 | $-0.097$ |
| | $\sigma$ | 1.012 | 94 | 1.532 | 0.012 |
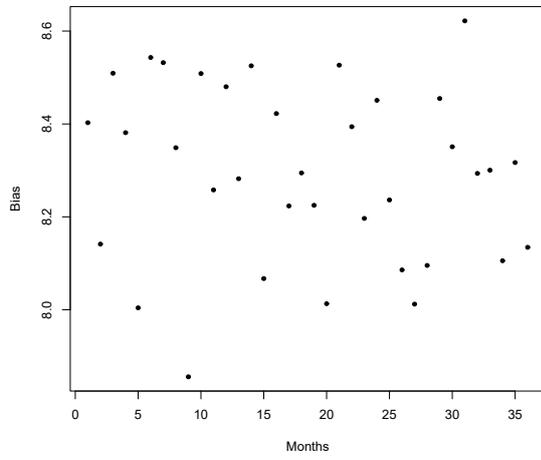
(a) DP Fixed $p \approx 0.6$
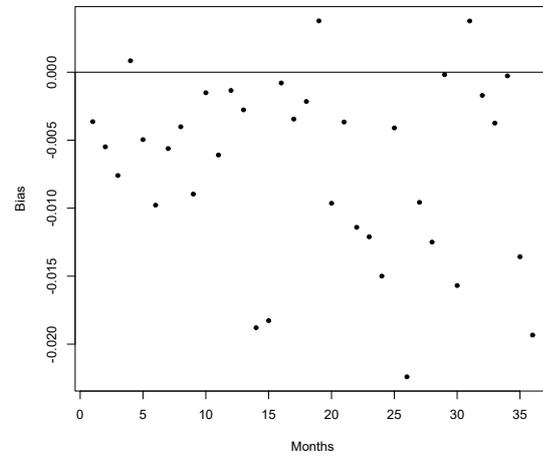
(b) DP Mixed $p \approx 0.6$
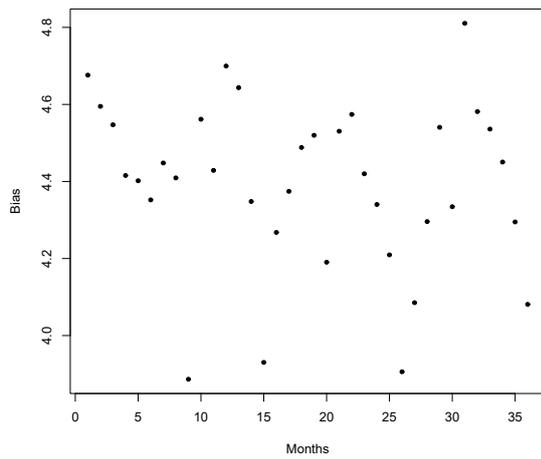
(c) DP Fixed $p \approx 0.3$

(d) DP Mixed $p \approx 0.3$

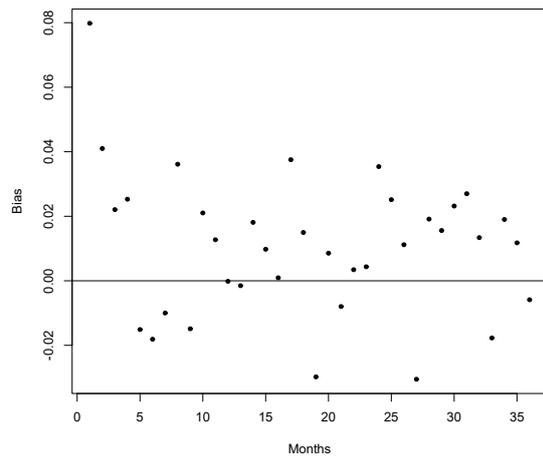Fig. B.3 Case 1. b) population size median relative bias obtained from DP model when $p \approx (0.6, 0.3)$.

(a) RW1 Fixed $p \approx 0.6$

(b) RW1 Mixed $p \approx 0.6$

(c) RW1 Fixed $p \approx 0.3$

(d) RW1 Mixed $p \approx 0.3$

Fig. B.4 Case 1. b) population size median relative bias obtained from the RW1 model when $p \approx (0.6, 0.3)$.

## B.5.3 Case 2

Table B.9 Case 2 simulation results for BGLSS for $p \approx 0.6$.

| Case | Parameters | Median | Coverage | Residual Mean Square | Relative Bias |
|------|-----------|--------|----------|----------------------|---------------|
| a) | $\beta_0$ | 0.748 | 96 | 0.075 | −0.003 |
|  | $\beta_1$ | 1.251 | 96 | 0.028 | 0.001 |
|  | $\beta_2$ | 0.203 | 92 | 0.119 | 0.016 |
|  | $\beta_3$ | 1.998 | 96 | 0.021 | −0.001 |
|  | $\beta_4$ | 0 | 2 | – | – |
|  | $\beta_5$ | −0.601 | 98 | −0.040 | 0.002 |
|  | $\beta_6$ | 0497 | 100 | 0.102 | −0.006 |
|  | $\beta_7$ | −0.980 | 96 | −0.063 | −0.019 |
|  | $\beta_8$ | 0.003 | 90 | – | – |
|  | $\lambda$ | 91.873 | 90 | 0.126 | −0.081 |
|  | $\sigma$ | 1.011 | 98 | 0.159 | 0.011 |
| b) | $\beta_0$ | 0.923 | 90 | 0.284 | 0.231 |
|  | $\beta_2$ | 0.120 | 74 | 0.711 | −0.399 |
|  | $\beta_3$ | 1.954 | 96 | 0.059 | −0.023 |
|  | $\beta_4$ | 0 | 70 | – | – |
|  | $\beta_5$ | −0.563 | 96 | −0.156 | −0.062 |
|  | $\beta_6$ | −0.872 | 86 | −0.289 | −0.128 |
|  | $\beta_8$ | 0.067 | 98 | – | – |
|  | $\lambda$ | 91.789 | 90 | 0.125 | −0.082 |
|  | $\sigma$ | 1.003 | 98 | 0.170 | 0.003 |

Table B.10 Case 2 simulation results for BGLSS for $p \approx 0.3$.

| Case | Parameters | Median | Coverage | Residual Mean Square | Relative Bias |
|---|---|---|---|---|---|
| a) | $\beta_0$ | $-1.504$ | 96 | $-0.043$ | 0.003 |
| | $\beta_1$ | 1.242 | 96 | 0.031 | $-0.006$ |
| | $\beta_2$ | 0.199 | 96 | 0.118 | $-0.001$ |
| | $\beta_3$ | 1.996 | 96 | 0.024 | $-0.002$ |
| | $\beta_4$ | 0 | 4 | – | – |
| | $\beta_5$ | $-0.597$ | 94 | $-0.054$ | -0.004 |
| | $\beta_6$ | 0.484 | 98 | 0.105 | $-0.032$ |
| | $\beta_7$ | $-0.981$ | 90 | $-0.076$ | $-0.019$ |
| | $\beta_8$ | 0.021 | 90 | – | – |
| | $\lambda$ | 94.010 | 92 | 0.122 | $-0.059$ |
| | $\sigma$ | 0.971 | 100 | 0.169 | $-0.028$ |
| b) | $\beta_0$ | $-1.351$ | 94 | $-0.126$ | $-0.099$ |
| | $\beta_2$ | 0.098 | 76 | 0.746 | $-0.505$ |
| | $\beta_3$ | 1.918 | 80 | 0.084 | $-0.040$ |
| | $\beta_4$ | 0 | 72 | – | – |
| | $\beta_5$ | $-0.555$ | 94 | $-0.175$ | $-0.074$ |
| | $\beta_6$ | $-0.846$ | 90 | $-0.296$ | $-0.153$ |
| | $\beta_8$ | 0.018 | 98 | – | – |
| | $\lambda$ | 90.558 | 86 | 0.142 | $-0.094$ |
| | $\sigma$ | 1.027 | 100 | 1.562 | 0.027 |
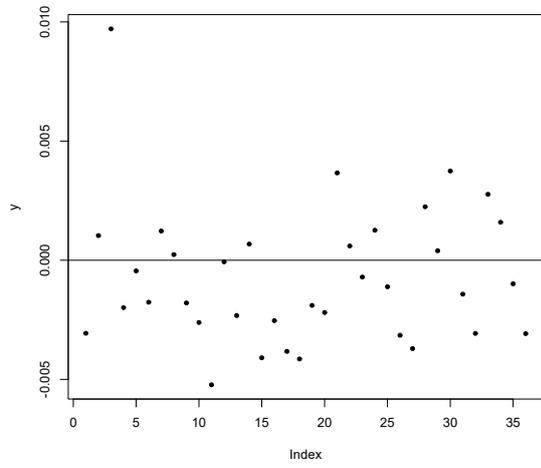
(a) Case 2. a) $p \approx 0.6$
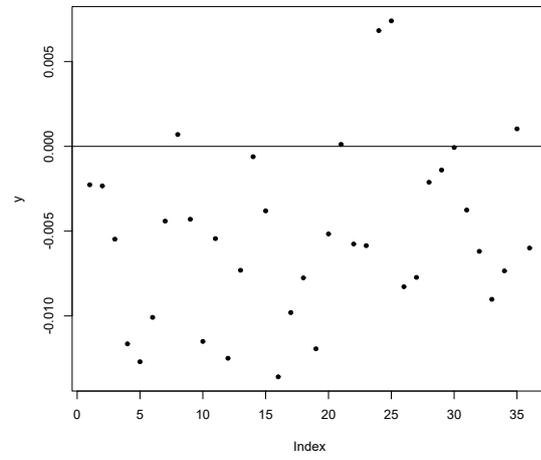


(b) Case 2. b) $p \approx 0.6$
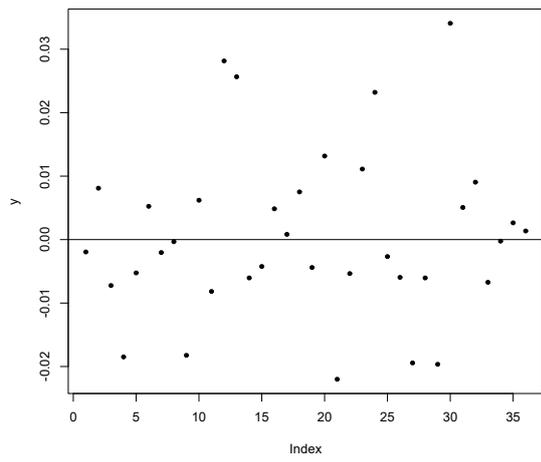


(c) Case 2. a) $p \approx 0.3$
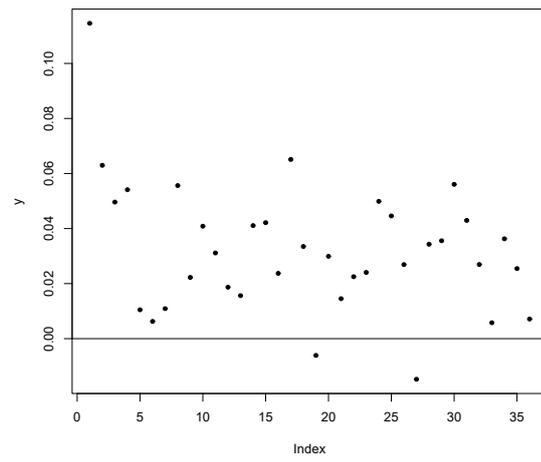


(d) Case 2. b) $p \approx 0.3$

Fig. B.5 Population size median relative bias obtained for case 2 when $p \approx (0.6, 0.3)$.

## B.6   Case Studies
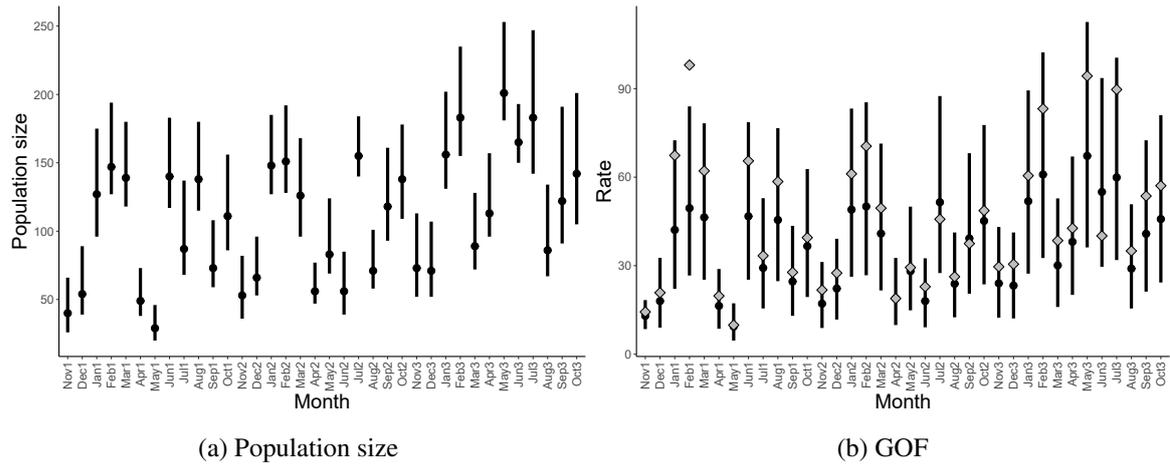
### B.6.1   Ecuadorian Amazon Parrots

**DP**



(a) Population size                                    (b) GOF

Fig. B.6 Amazon parrots case study. (a) The black dots represent the posterior mean population size each month and the thick bands represent the corresponding 95% posterior credible interval. (b) The diamonds are the observed monthly rates and the thick bands represent the 95% intervals of simulated monthly rates. In both cases, the x-axis represents the months in each year with months ending in 1, 2, and 3 denoting months in the 1st, 2nd, and 3rd year, respectively.

Rain, storm, and time of sampling are identified as important predictors for observation error with posterior inclusion probabilities (PIP): $0.587, 0.734$ and $0.7548$ respectively. Rain, storm, and surveying in PM instead of AM have an estimated positive effect on the probability of detection.

**RW2**



(a) Population size                                      (b) GOF
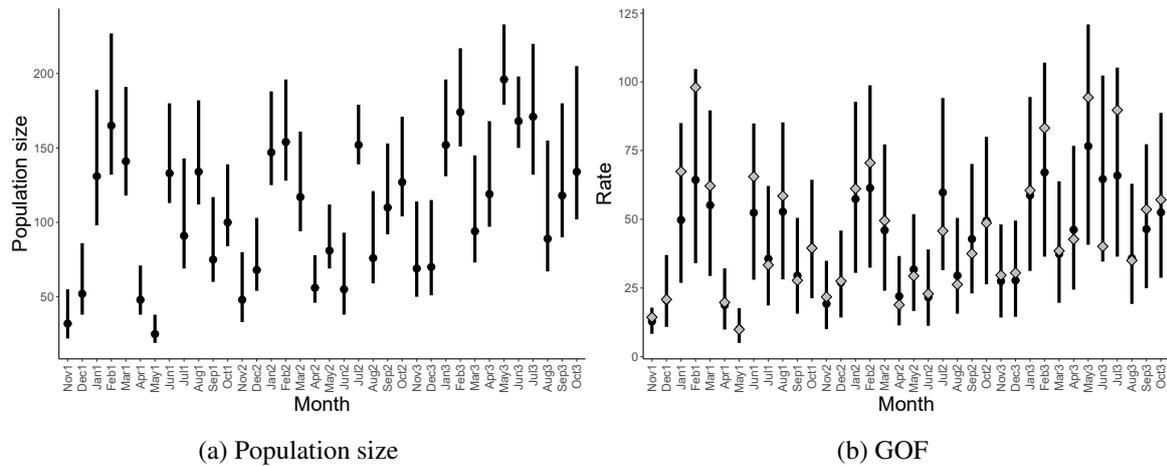
Fig. B.7 Amazon parrots case study. (a) The black dots represent the posterior mean population size each month and the thick bands represent the corresponding 95% posterior credible interval. (b) The diamonds are the observed monthly rates and the thick bands represent the 95% intervals of simulated monthly rates. In both cases, the x-axis represents the months in each year with months ending in 1, 2, and 3 denoting months in the 1st, 2nd, and 3rd year, respectively.

Rain, storm, and time of sampling are identified as important predictors for observation error with posterior inclusion probabilities (PIP): $0.531, 0.573$ and $0.637$ respectively. Rain, storm, and surveying in PM instead of AM have an estimated positive effect on the probability of detection.

**Cor**



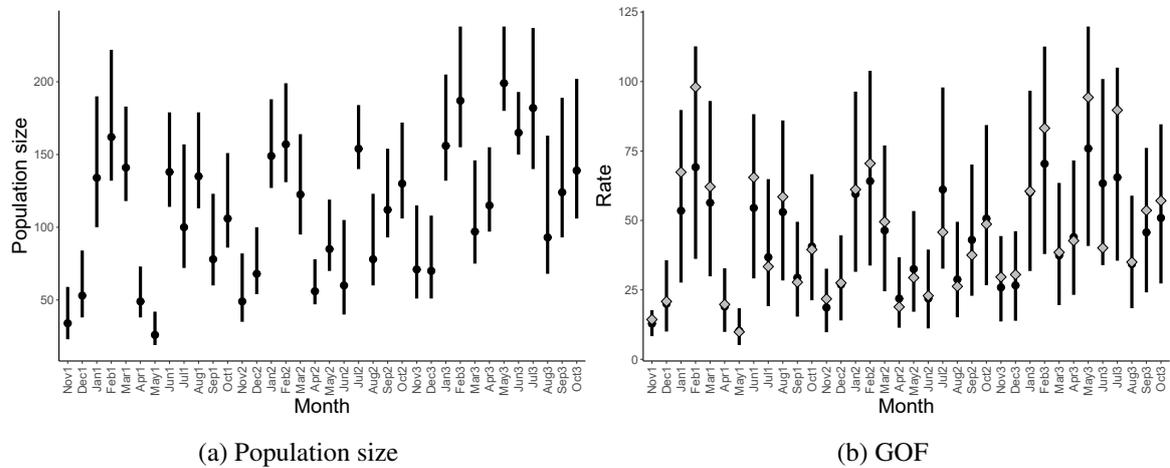(a) Population size                    (b) GOF

Fig. B.8 Amazon parrots case study. (a) The black dots represent the posterior mean population size each month and the thick bands represent the corresponding 95% posterior credible interval. The x-axis represents the months in each year with months ending in 1, 2 and 3 denoting months in the 1st, 2nd and 3rd year, respectively. (b) The diamonds are the observed monthly rates and the thick bands represent the 95% intervals of simulated monthly rates. In both cases, the x-axis represents the months in each year with months ending in 1, 2, and 3 denoting months in the 1st, 2nd, and 3rd year, respectively.

Rain, storm, and time of sampling are identified as important predictors for observation error with posterior inclusion probabilities (PIP): $0.519, 0.651$ and $0.7214$ respectively. Storm and surveying in PM instead of AM have an estimated positive effect on the probability of detection while rain has an estimated negative effect.

**AR1**
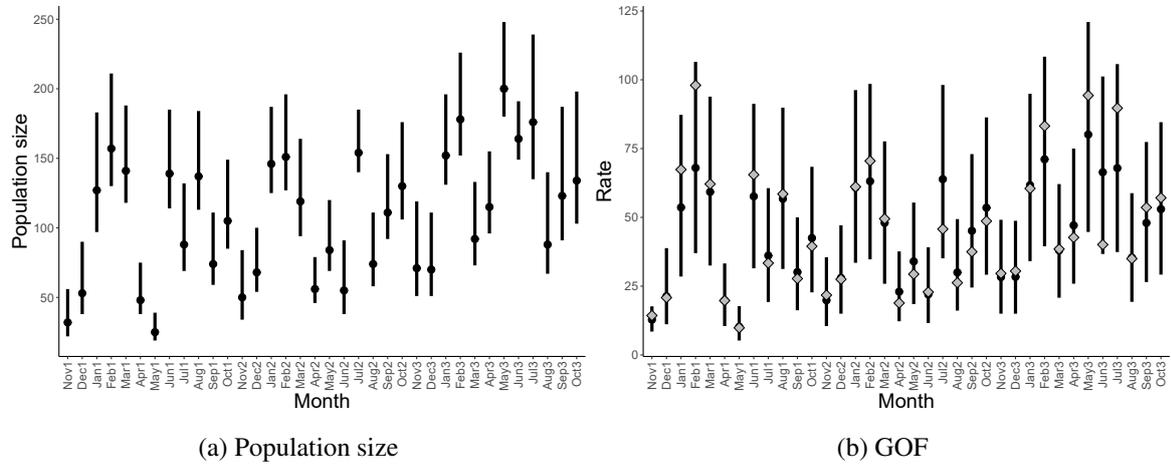


(a) Population size                    (b) GOF

Fig. B.9 Amazon parrots case study-AR1. (a) The black dots represent the posterior mean population size each month and the thick bands represent the corresponding 95% posterior credible interval. The x-axis represents the months in each year with months ending in 1, 2 and 3 denoting months in the 1st, 2nd and 3rd year, respectively. (b) The diamonds are the observed monthly rates and the thick bands represent the 95% intervals of simulated monthly rates. In both cases, the x-axis represents the months in each year with months ending in 1, 2, and 3 denoting months in the 1st, 2nd, and 3rd year, respectively.

Rain, storm, and time of sampling are identified as important predictors for observation error with posterior inclusion probabilities (PIP): $0.566, 0.622$ and $0.693$ respectively. Storm and surveying in PM instead of AM have an estimated positive effect on the probability of detection while rain has an estimated negative effect.

## B.6.2   Orange-Winged Amazon Parrots

**DP**



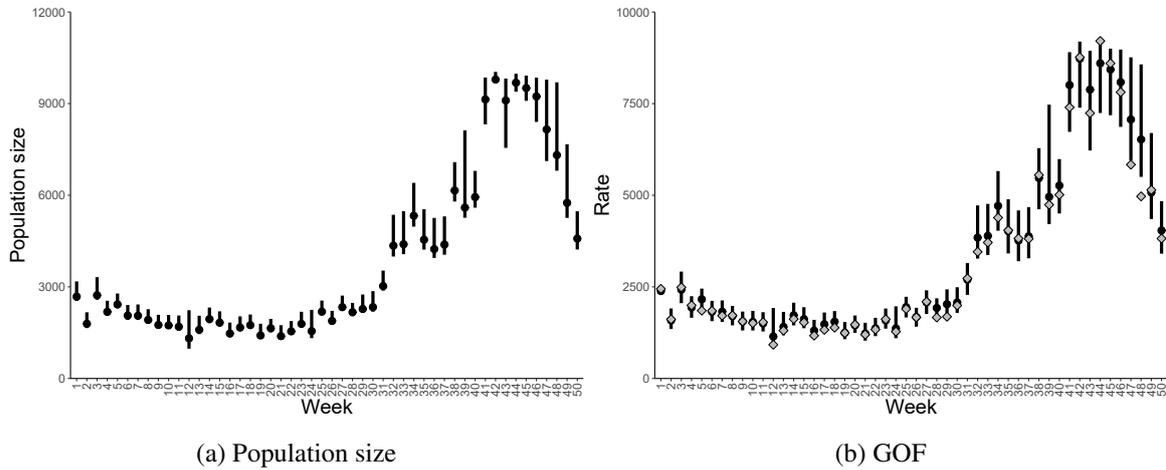(a) Population size                                          (b) GOF

Fig. B.10 Orange-winged Amazon parrots case study, DP model results. (a) The black dots represent the posterior mean population size each week and the thick bands represent the corresponding 95% posterior credible interval. (b) The diamonds are the observed weekly rates and the thick bands represent the 95% intervals of simulated weekly rates.

Three clusters weeks (1-31), (32-40, 49-50) and (41-48) were found. For variable selection, only rain was found to be a significant predictor of observation error with a PIP of 0.511.
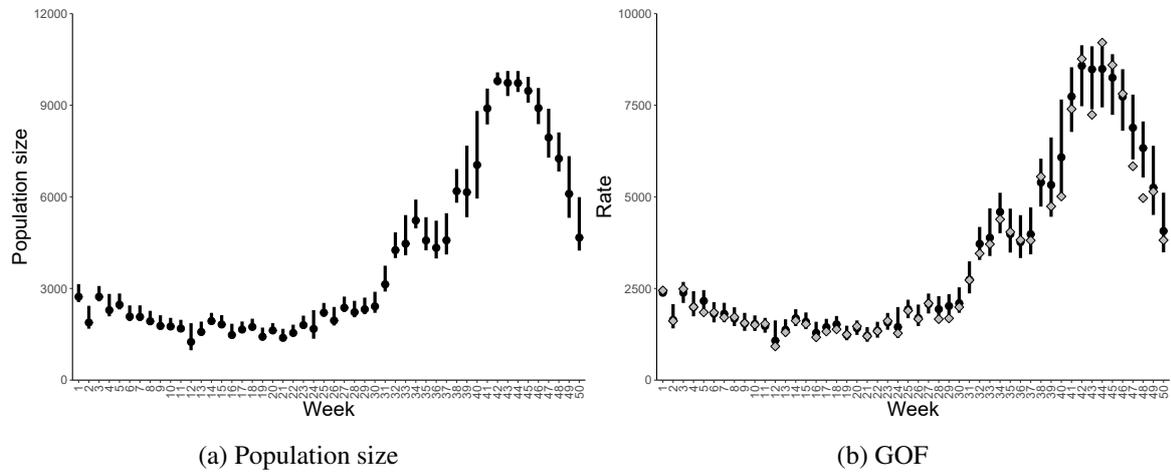
**RW2**



(a) Population size



(b) GOF

Fig. B.11 Orange-winged Amazon parrots case study, RW2 model results. (a) The black dots represent the posterior mean population size each week and the thick bands represent the corresponding 95% posterior credible interval. (b) The diamonds are the observed weekly rates and the thick bands represent the 95% intervals of simulated weekly rates.

Cloud, wind, rain, and time of sampling were found to be significant predictors of observation error with PIP of $0.533, 0.514, 0.504$, and $0.528$ respectively.

**AR1**
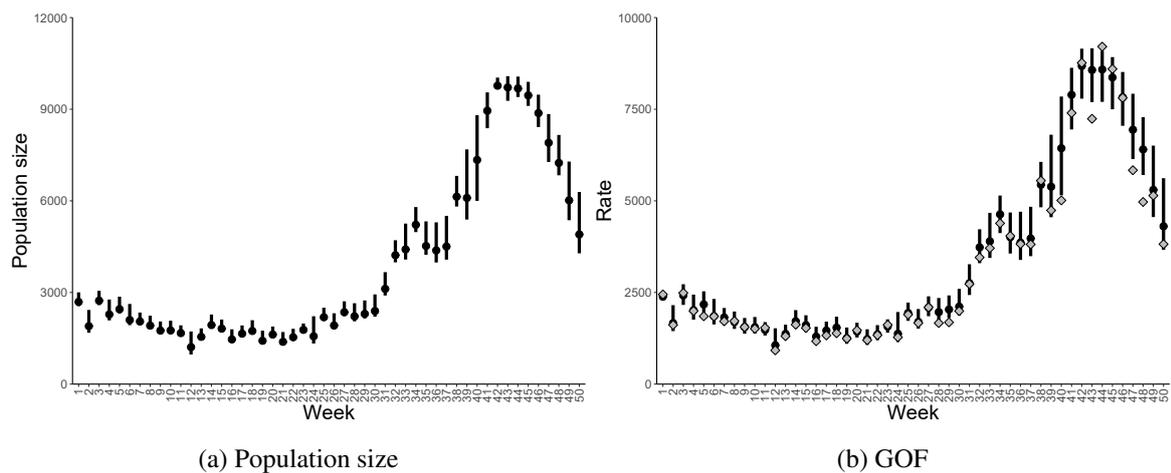


(a) Population size



(b) GOF

Fig. B.12 Orange-winged Amazon parrots case study, AR1 model results. (a) The black dots represent the posterior mean population size each week and the thick bands represent the corresponding 95% posterior credible interval. (b) The diamonds are the observed weekly rates and the thick bands represent the 95% intervals of simulated weekly rates.

For variable selection, only rain was found to be a significant predictor of observation error with PIP of 0.508.

# Appendix C

# Supplementary material for Modelling disease dynamics from spatially explicit capture-recapture data
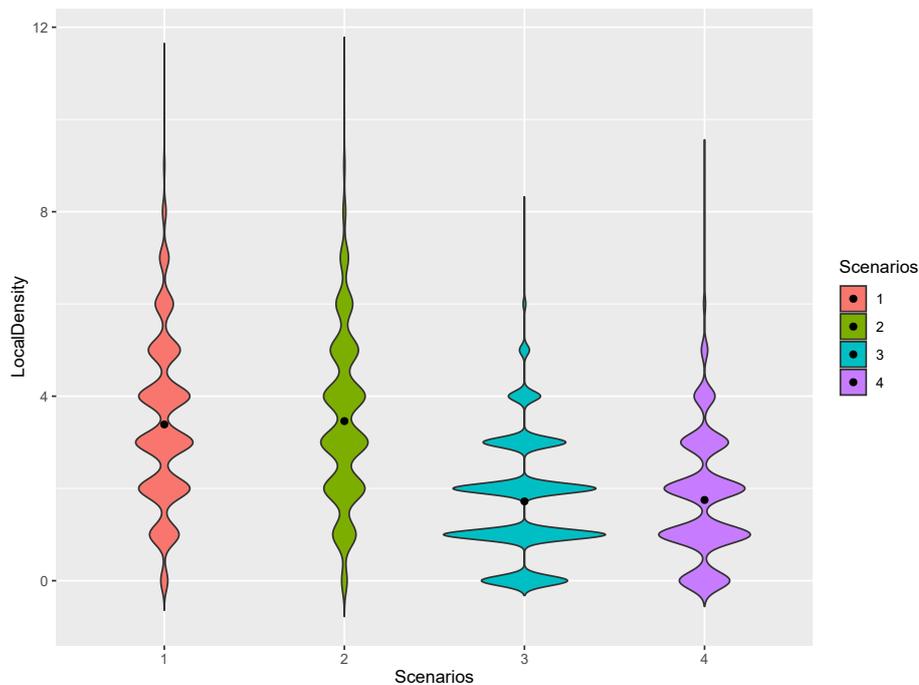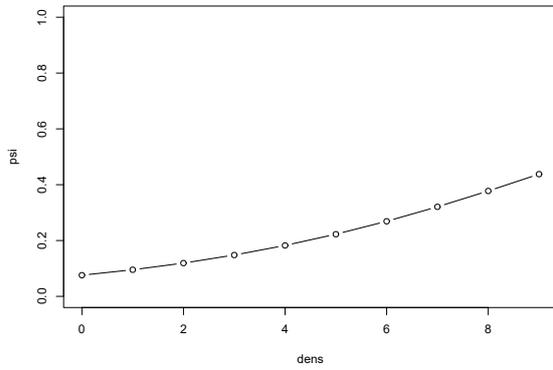
## C.1 Simulation Study



Fig. C.1 Local density violin plot for each scenario considered: (1- high N, high density effect, 2- high N, low density effect, 3- low N, high density effect,4- low N, low density effect). Dots are the median values
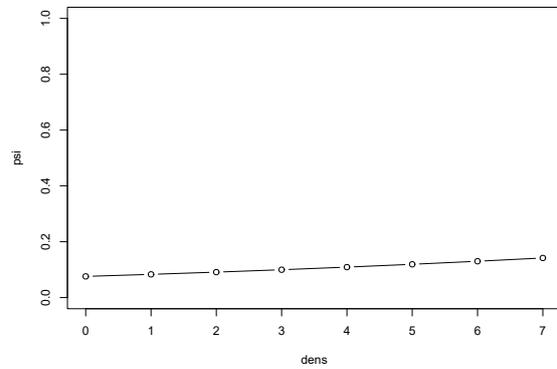
(a) High population size high density effect.
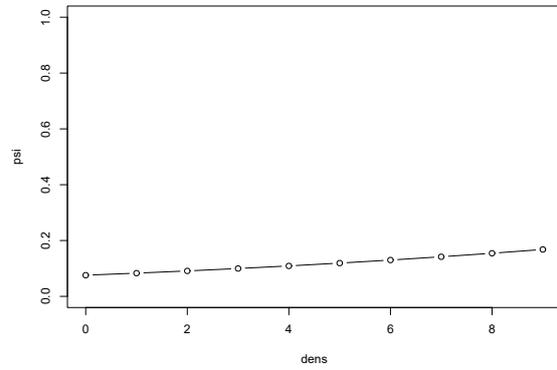
(b) High population size low density effect.
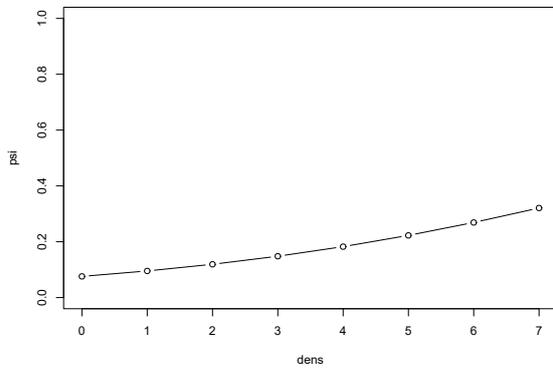
(c) Low population size high density effect.

(d) Low population size low density effect.

Fig. C.2 Expected $\psi_{it}$ for each case considered.

Table C.1 High population size. Posterior median with values within brackets showing the 95% quantiles.

| Parameter | Density effect | | | | | |
|---|---|---|---|---|---|---|
| | High | | | Low | | |
| | RB | CV | Coverage | RB | CV | Coverage |
| $\gamma_1$ | $0.04(-0.08, 0.190)$ | $0.08(0.07, 0.09)$ | 90 | $0.02(-0.16, 0.13)$ | $0.08(0.07, 0.09)$ | 90 |
| $\gamma_2$ | $0.03(-0.24, 0.28)$ | $0.19(0.14, 0.26)$ | 100 | $0.08(-0.17, 0.34)$ | $0.18(0.15, 0.23)$ | 100 |
| $\gamma_3$ | $-0.05(-0.56, 0.31)$ | $0.25(0.19, 0.48)$ | 90 | $-0.01(-045, 0.42)$ | $0.25(0.2, 0.43)$ | 90 |
| $\gamma_4$ | $-0.07(-0.19, 0.11)$ | $0.21(0.19, 0.26)$ | 100 | $-0.03(-0.31, 0.42)$ | $0.22(0.19, 0.27)$ | 90 |
| $\phi_{d_i=0}$ | $-0.01(-0.04, 0.02)$ | $0.20(0.01, 0.2)$ | 80 | $-0.00(-0.03, 0.04)$ | $0.02(0.01, 0.02)$ | 90 |
| $\phi_{d_i=1}$ | $0.02(-0.02, 0.04)$ | $0.03(0.02, 0.03)$ | 100 | $-0.03(0.09, 0.02)$ | $0.03(0.03, 0.05)$ | 80 |
| $p0_{d_i=0}$ | $-0.04(-0.18, 0.08)$ | $0.08(0.07, 0.09)$ | 90 | $0.02(-0.10, 0.09)$ | $0.06(0.5, 0.07)$ | 90 |
| $p0_{d_i=1}$ | $-0.03(-0.08, 0.07)$ | $0.06(0.05, 0.07)$ | 100 | $0.00(-0.08, 0.08)$ | $0.08(0.07, 0.09)$ | 100 |
| $\sigma_{d_i=0}$ | $0.02(-0.04, 0.06)$ | $0.03(0.03, 0.03)$ | 100 | $-0.00(-0.04, 0.05)$ | $0.03(0.02, 0.03)$ | 90 |
| $\sigma_{d_i=1}$ | $0.01(-0.05, 0.03)$ | $0.02(0.02, 0.03)$ | 90 | $0.00(-0.03, 0.01)$ | $0.03(0.03, 0.04)$ | 100 |
| $\delta$ | $0.02(-0.42, 0.44)$ | $0.25(0.18, 0.34)$ | 80 | $0.15(-0.38, 0.52)$ | $0.23(0.17, 0.31)$ | 90 |
| $\beta_0$ | $-0.36(-0.49, -0.19)$ | $0.13(0.11, 0.14)$ | 30 | $-0.18(-0.28, 0.01)$ | $0.09(0.08, 0.13)$ | 40 |
| $\beta_1$ | $0.21(-0.61, 0.94)$ | $0.47(0.38, 1.73)$ | 100 | $-0.61(-1.88, 1.69)$ | $3.15(0.69, 9.57)$ | 100 |
| $q_{11}^{DPP}$ | $-0.01(-0.05, 0.04)$ | $0.04(0.04, 0.04)$ | 100 | $-0.00(-0.04, 0.05)$ | $0.04(0.04, 0.05)$ | 100 |
| $q_{11}^{Ifn}$ | $-0.01(-0.04, 0.03)$ | $0.03(0.02, 0.3)$ | 100 | $-0.01(-0.05, 0.02)$ | $0.03(0.02, 0.04)$ | 100 |
| $q_{11}^{Cul}$ | $-0.03(0.27, 0.12)$ | $0.14(0.12, 0.16)$ | 90 | $-0.02(-0.27, 0.25)$ | $0.16(0.14, 0.19)$ | 90 |
| $q_{00}^{DPP}$ | $-0.01(-0.02, 0.02)$ | $0.01(0.01, 0.02)$ | 100 | $0.01(-0.00, 0.02)$ | $0.01(0.01, 0.01)$ | 100 |
| $q_{00}^{Ifn}$ | $0.00(-0.03, 0.02)$ | $0.02(0.01, 0.02)$ | 100 | $0.00(-0.03, 0.01)$ | $0.01(0.01, 0.02)$ | 90 |
| $q_{00}^{Cul}$ | $-0.00(-0.00, -0.00)$ | $0.00(0.00, 0.00)$ | 100 | $-0.00(-0.00, -0.00)$ | $0.00(0.01, 0.01)$ | 100 |
| $N_{1:T}$ | $0.01(-0.08, 0.09)$ | $0.05(0.04, 0.08)$ | 100 | $-0.00(0.09, 0.07)$ | $0.06(0.05, 0.08)$ | 100 |
| $N_{1:T}^{u}$ | $-0.05(-0.17, 0.11)$ | $0.09(0.07, 0.11)$ | 100 | $0.00(-0.13, 0.11)$ | $0.08(0.06, 0.10)$ | 100 |
| $N_{1:T}^{i}$ | $0.04(-0.19, 0.32)$ | $0.08(0.06, 0.23)$ | 100 | $0.00(-0.23, 0.16)$ | $0.10(0.08, 0.23)$ | 100 |

Table C.2 Low population size. Posterior median with values within brackets showing the 95% quantiles.

| | Density effect | | | | | |
|---|---|---|---|---|---|---|
| | High | | | Low | | |
| Parameter | RB | CV | Coverage | RB | CV | Coverage |
| $\gamma_1$ | $-0.02(-0.26, 0.12)$ | $0.12(0.11, 0.14)$ | 90 | $-0.04(-0.19, 0.24)$ | $0.12(0.11, 0.14)$ | 90 |
| $\gamma_2$ | $0.18(-0.10, 0.61)$ | $0.26(0.18, 0.30)$ | 90 | $0.11(-0.21, 0.33)$ | $0.26(0.22, 0.43)$ | 100 |
| $\gamma_3$ | $-0.04(-0.48, 0.44)$ | $0.38(0.27, 0.63)$ | 100 | $-0.09(-0.27, 0.93)$ | $0.38(0.27, 0.46)$ | 90 |
| $\gamma_4$ | $0.01(-0.31, 0.52)$ | $0.35(0.31, 0.42)$ | 100 | $-0.14(0.43, 0.65)$ | $0.36(0.29, 0.55)$ | 90 |
| $\phi_{d_i=0}$ | $-0.01(-0.02, 0.02)$ | $0.03(0.02, 0.03)$ | 100 | $0.01(-0.04, 0.03)$ | $0.02(0.02, 0.03)$ | 100 |
| $\phi_{d_i=1}$ | $0.02(-0.06, 0.04)$ | $0.04(0.03, 0.05)$ | 100 | $0.01(-0.07, 0.07)$ | $0.05(0.04, 0.06)$ | 100 |
| $p_{0_{d_i=0}}$ | $-0.04(-0.12, 0.16)$ | $0.09(0.08, 0.10)$ | 100 | $-0.02(-0.18, 0.08)$ | $0.08(0.08, 0.09)$ | 90 |
| $p_{0_{d_i=1}}$ | $-0.07(-0.18, 0.13)$ | $0.11(0.09, 0.13)$ | 100 | $0.00(-0.19, 0.24)$ | $0.11(0.10, 0.13)$ | 90 |
| $\sigma_{d_i=0}$ | $0.01(-0.05, 0.06)$ | $0.04(0.03, 0.05)$ | 100 | $0.01(-0.04, 0.06)$ | $0.03(0.03, 0.04)$ | 90 |
| $\sigma_{d_i=1}$ | $0.01(-0.05, 0.07)$ | $0.04(0.04, 0.05)$ | 100 | $0.01(-0.03, 0.09)$ | $0.05(0.04, 0.05)$ | 100 |
| $\delta$ | $0.11(-0.24, 0.72)$ | $0.33(0.23, 0.38)$ | 90 | $-0.16(-0.31, 0.43)$ | $0.34(0.28, 0.44)$ | 100 |
| $\beta_0$ | $-0.12(-0.26, 0.16)$ | $0.22(0.17, 0.33)$ | 90 | $-0.03(-0.28, 0.41)$ | $0.21(0.14, 0.31)$ | 100 |
| $\beta_1$ | $0.64(-0.86, 2.24)$ | $1.07(0.71, 8.52)$ | 100 | $1.09(-8.64, 11.57)$ | $1.58(0.58, 12.78)$ | 90 |
| $q_{11}^{\text{DPP}}$ | $-0.01(-0.03, 0.04)$ | $0.05(0.04, 0.05)$ | 100 | $-0.00(-0.03, 0.04)$ | $0.05(0.05, 0.05)$ | 100 |
| $q_{11}^{\text{Ifn}}$ | $0.01(-0.07, 0.04)$ | $0.04(0.034, 0.05)$ | 100 | $0.00(-0.07, 0.03)$ | $0.04(0.04, 0.05)$ | 90 |
| $q_{11}^{\text{Cul}}$ | $-0.07(-0.36, 0.25)$ | $0.22(0.19, 0.27)$ | 90 | $-0.08(-0.32, 0.58)$ | $0.24(0.18, 0.30)$ | 90 |
| $q_{00}^{\text{DPP}}$ | $0.00(-0.01, 0.02)$ | $0.02(0.01, 0.02)$ | 100 | $0.00(-0.02, 0.03)$ | $0.02(0.01, 0.02)$ | 90 |
| $q_{00}^{\text{Ifn}}$ | $-0.01(-0.04, 0.04)$ | $0.02(0.01, 0.024)$ | 80 | $-0.00(-0.03, 0.01)$ | $0.02(0.01, 0.02)$ | 100 |
| $q_{00}^{\text{Cul}}$ | $-0.00(-0.01, -0.00)$ | $0.00(0.00, 0.01)$ | 90 | $-0.00(-0.01, -0.00)$ | $0.00(0.00, 0.00)$ | 90 |
| $N_{1:T}$ | $0.02(-0.11, 0.13)$ | $0.08(0.06, 0.12)$ | 100 | $-0.01(-0.14, 0.16)$ | $0.08(0.07, 0.12)$ | 100 |
| $N_{1:T}^{\text{u}}$ | $-0.02(-0.18, 0.16)$ | $0.11(0.09, 0.15)$ | 100 | $-0.01(-0.19, 0.20)$ | $0.11(0.09, 0.14)$ | 100 |
| $N_{1:T}^{\text{i}}$ | $0.05(-0.17, 0.41)$ | $0.13(0.10, 0.31)$ | 100 | $0.02(-0.21, 0.37)$ | $0.14(0.11, 0.33)$ | 100 |

(a) High Density effect RB

(b) High Density effect CV
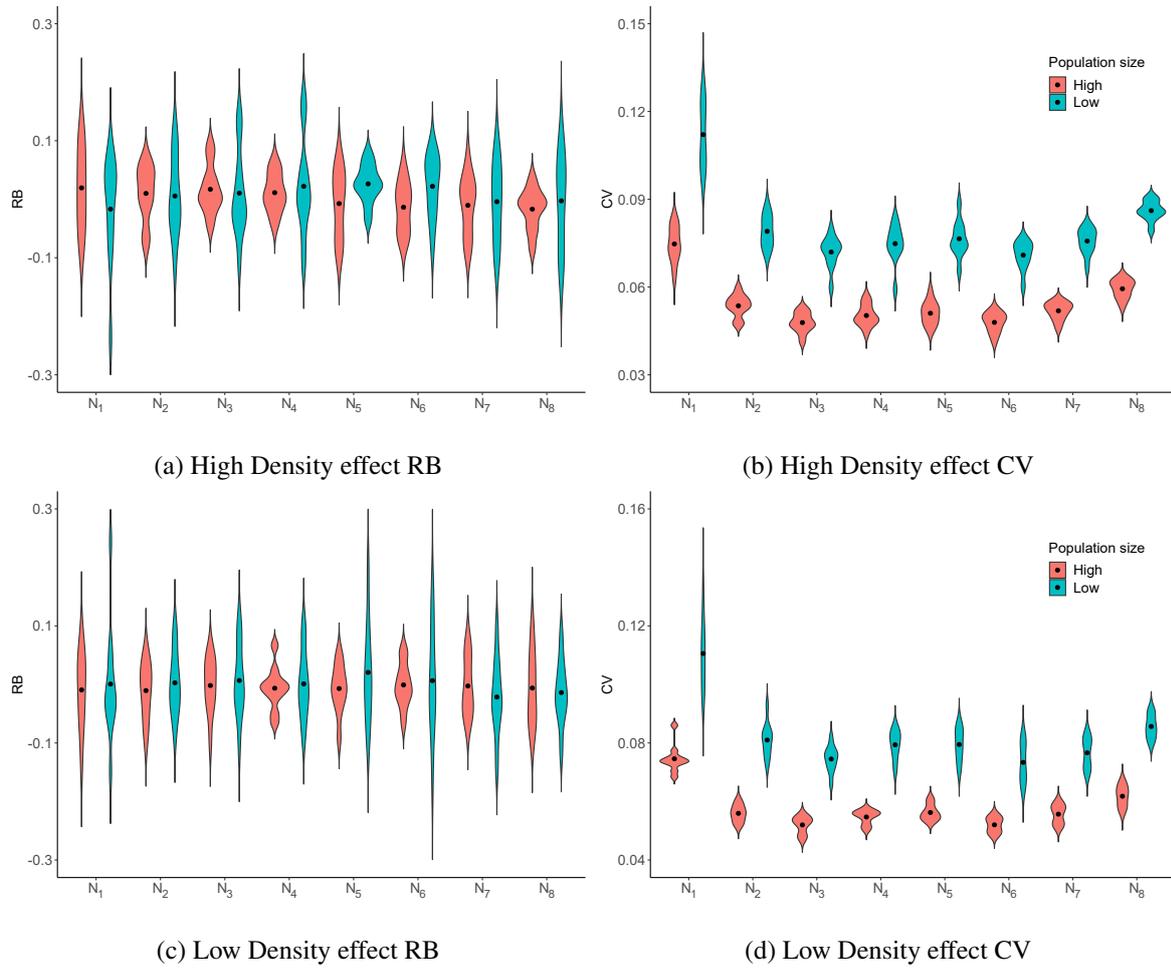
(c) Low Density effect RB

(d) Low Density effect CV

Fig. C.3 RB and CV of population size, $N$, at high and low density effect levels ($\beta_1 = 0.25$, and $\beta_1 = 0.1$, respectively) at high and low levels of population size ($N \approx (400, 200)$ respectively).
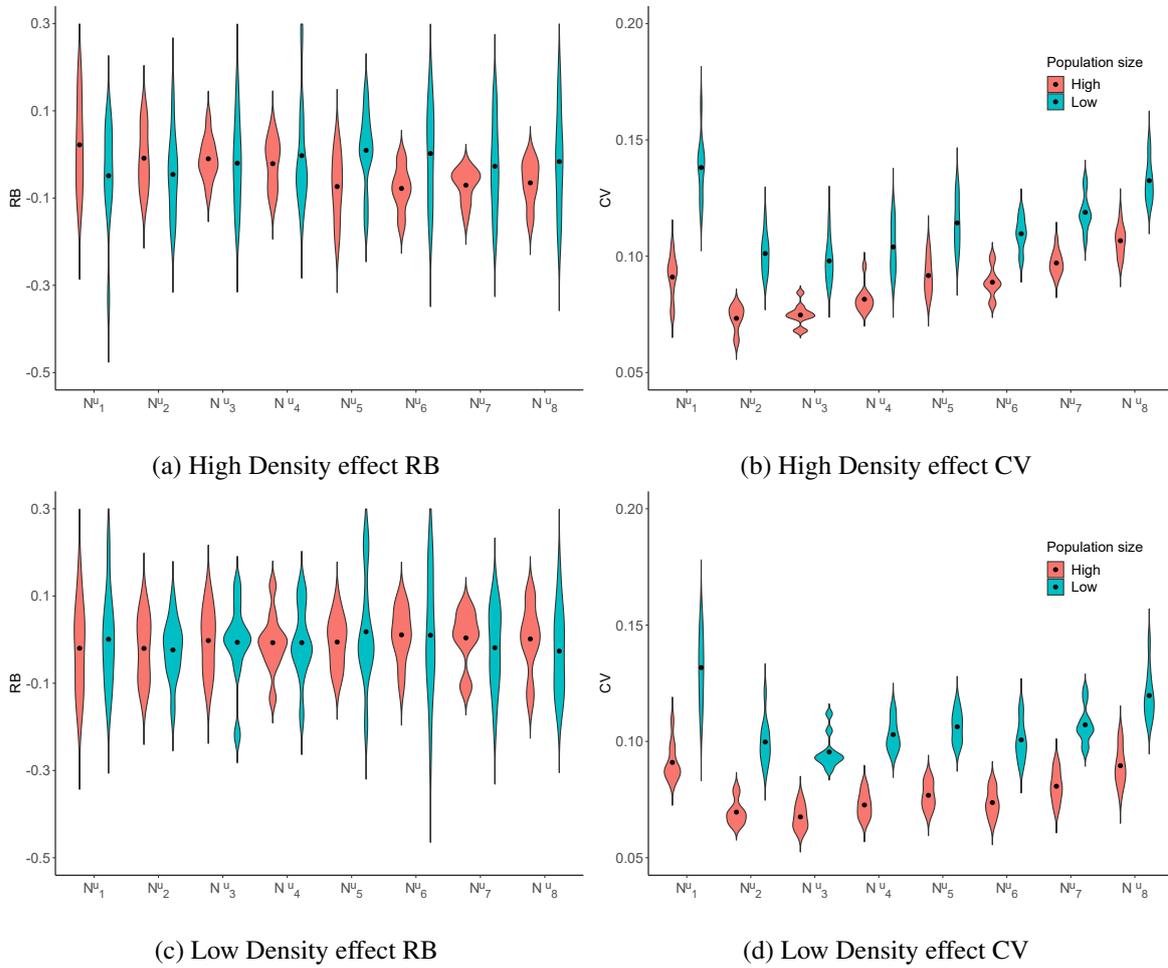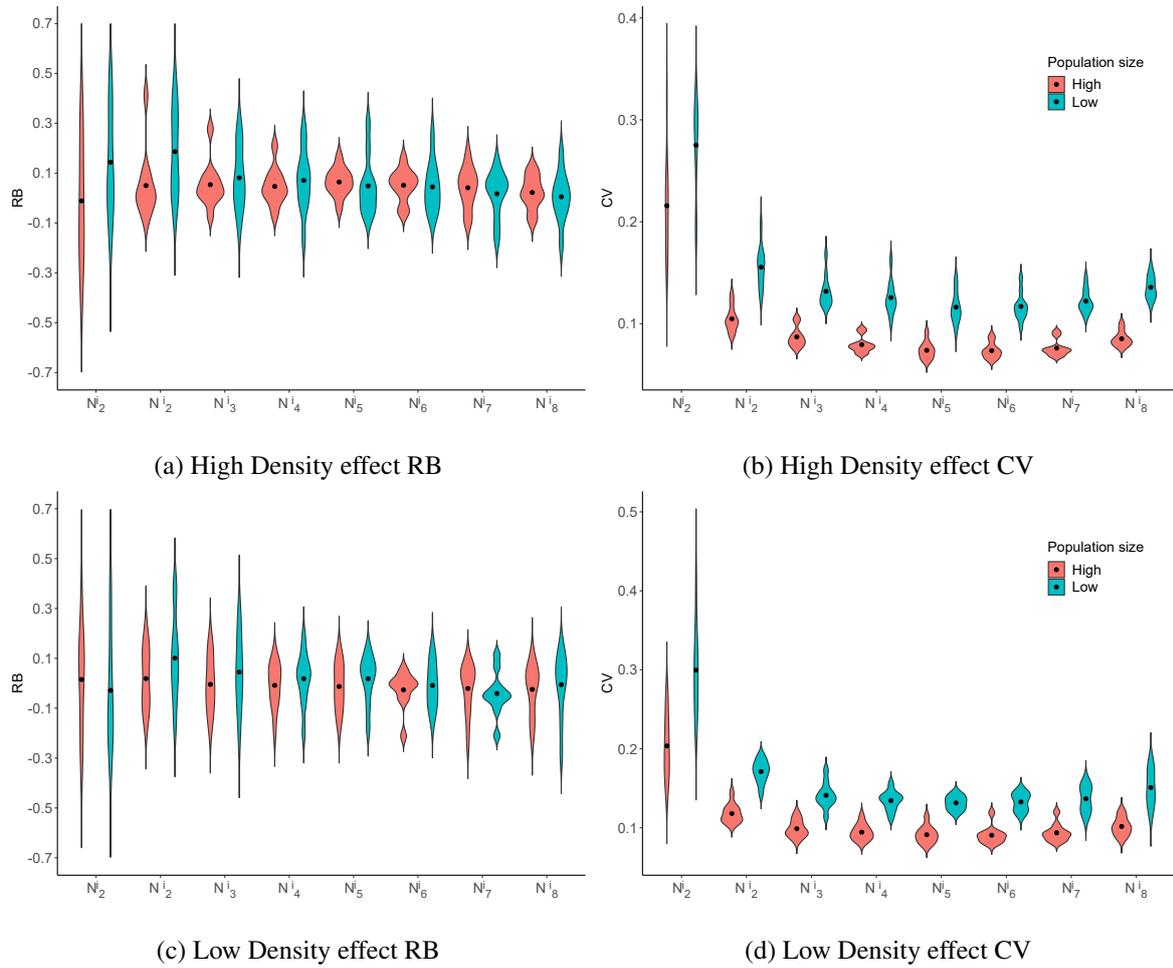
(a) High Density effect RB

(b) High Density effect CV

(c) Low Density effect RB

(d) Low Density effect CV

Fig. C.4 RB and CV of Uninfected population size, $N^U$, at high and low density effect levels ($\beta_1 = 0.25$, and $\beta_1 = 0.1$, respectively) at high and low levels of population size ($N \approx (400, 200)$ respectively).

(a) High Density effect RB

(b) High Density effect CV

(c) Low Density effect RB

(d) Low Density effect CV

Fig. C.5 RB and CV of Infected population size, $N^I$, at high and low density effect levels ($\beta_1 = 0.25$, and $\beta_1 = 0.1$, respectively) at high and low levels of population size ($N \approx (400, 200)$ respectively).

**Ignoring Disease transmission heterogeneity**



(a) High Density effect RB

(b) High Density effect CV

(c) Low Density effect RB

(d) Low Density effect CV

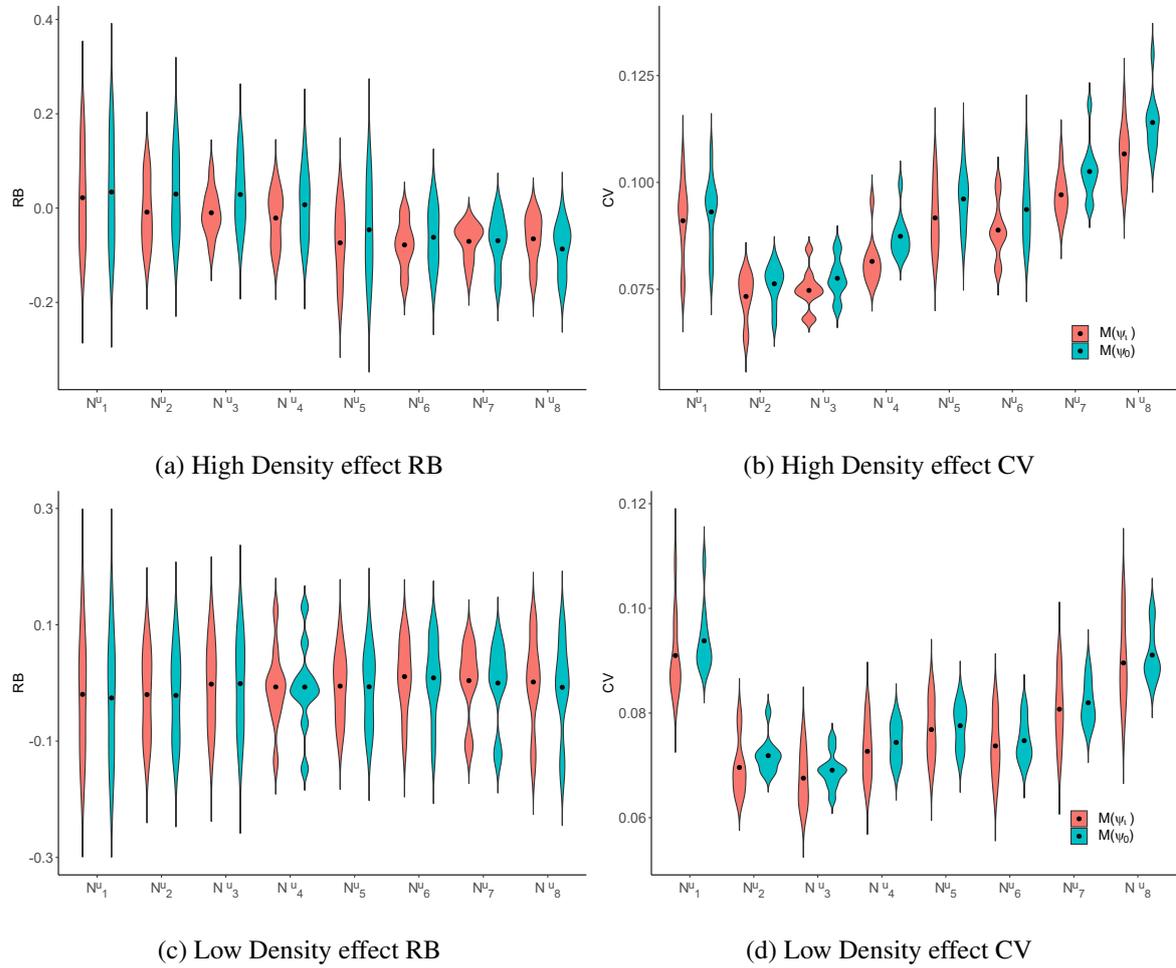Fig. C.6 RB and CV of Uninfected population size, $N^U$, at high population size data ($N \approx 400$), high and low density effect levels ($\beta_1 = 0.25$, and $\beta_1 = 0.1$, respectively) for two models: our proposed model, $M(\psi_\ell)$ and the model that does not account for density-dependence in disease transmission, $M(\psi_0)$.
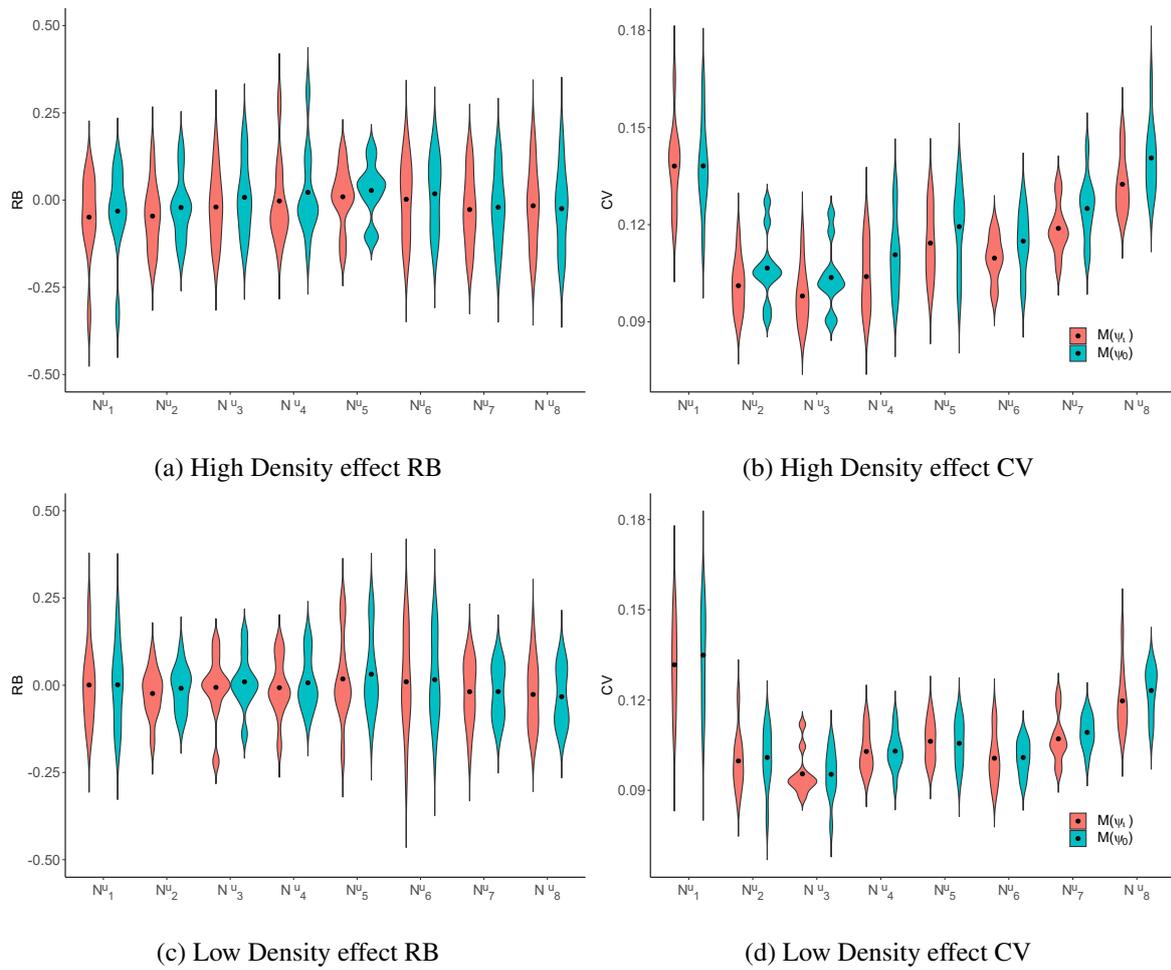
(a) High Density effect RB

(b) High Density effect CV

(c) Low Density effect RB

(d) Low Density effect CV

Fig. C.7 RB and CV of Uninfected population size, $N^U$, at low population size data ($N \approx 200$), high and low density effect levels ($\beta_1 = 0.25$, and $\beta_1 = 0.1$, respectively) for two models: our proposed model, $M(\psi_\ell)$ and the model that does not account for density-dependence in disease transmission, $M(\psi_0)$.
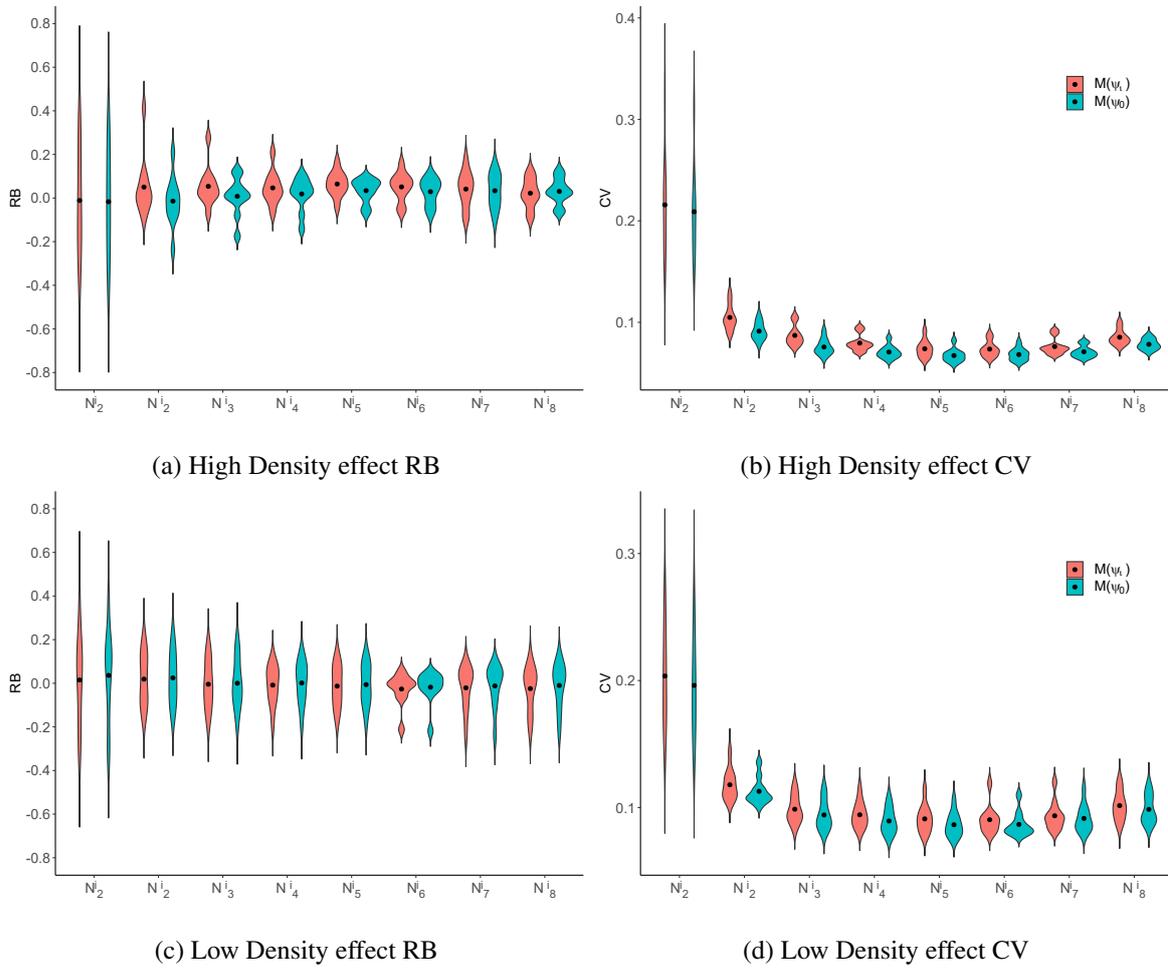
(a) High Density effect RB

(b) High Density effect CV

(c) Low Density effect RB

(d) Low Density effect CV

Fig. C.8 RB and CV of Infected population size, $N^I$, at high population size data ($N \approx 400$), high and low density effect levels ($\beta_1 = 0.25$, and $\beta_1 = 0.1$, respectively) for two models: our proposed model, $M(\psi_\ell)$ and the model that does not account for density-dependence in disease transmission, $M(\psi_0)$.
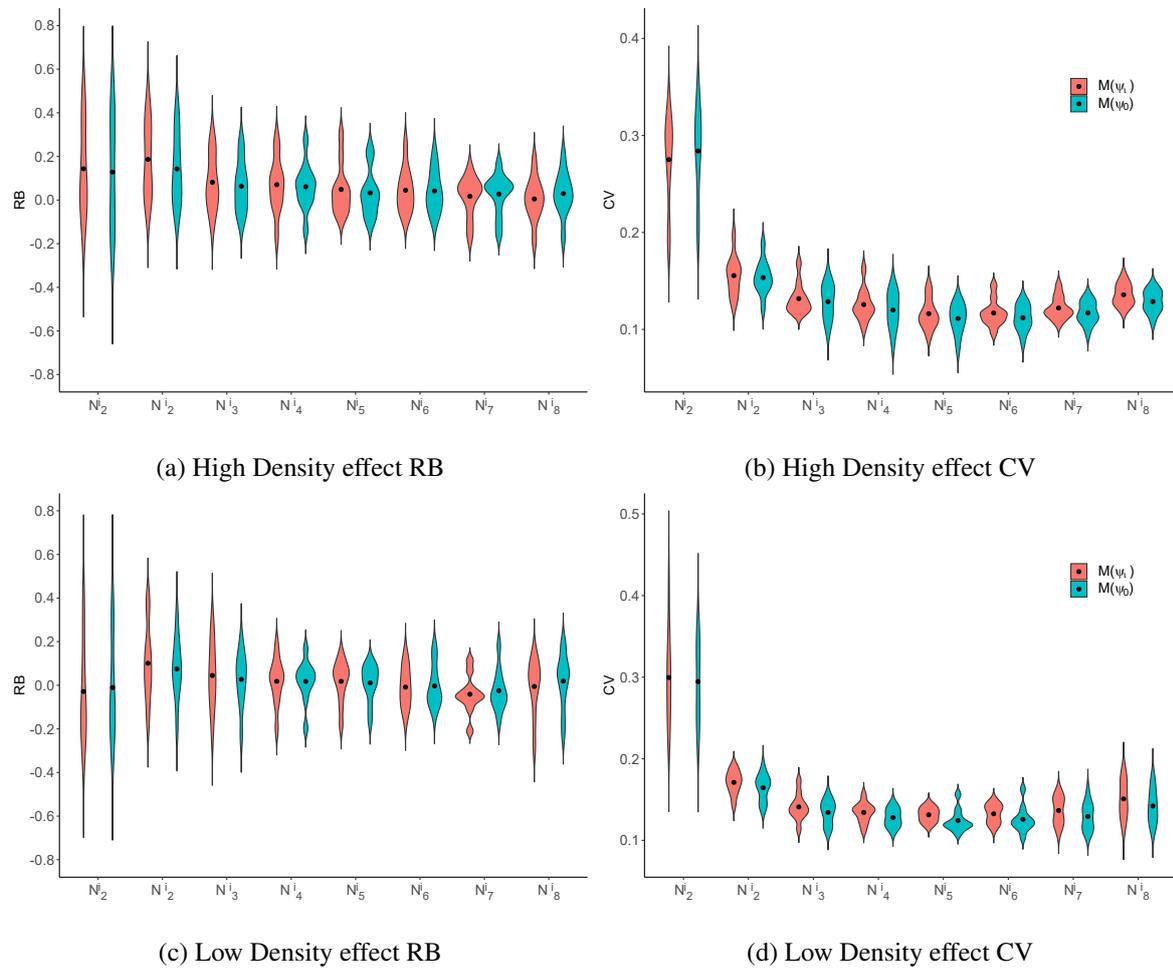
(a) High Density effect RB

(b) High Density effect CV

(c) Low Density effect RB

(d) Low Density effect CV

Fig. C.9 RB and CV of Infected population size, $N^I$, at low population size data ($N \approx 200$), high and low density effect levels ($\beta_1 = 0.25$, and $\beta_1 = 0.1$, respectively) for two models: our proposed model, $M(\psi_\ell)$ and the model that does not account for density-dependence in disease transmission, $M(\psi_0)$.

## C.2 Real Data Analysis

### C.2.1 Badger Movement

To better understand the movement of badgers between setts, we use static directed network graphs. Networks are made up of nodes that are connected by edges, where nodes can represent individuals, groups, classes of individuals, or other entities and edges generally represent the relationship between two nodes and can be used to describe how often they connect or interact. In a directed network, edges represent the direction of the relationship between nodes. In a static network, nodes and edges

do not vary temporally. Hence, static directed networks are networks where edges are directional and the network structure does not change dynamically over time.

We define the nodes as the setts and edges representing the frequency of movement of a badger from one sett to another. We produce static directed networks for each badger year from 2010-2018.

Badger year 2010-2011

Badger year 2011-2012

Badger year 2012-2013

Badger year 2013-2014

Badger year 2014-2015

Badger year 2015-2016

Badger year 2016-2017

Badger year 2017-2018

Fig. C.12 Static directed Network graph plot for badger years 2010-2011, 2011-2012, 2012-2013, 2013-2014 ,2014-2015, 2015-2016, 2016-2017, and 2017-2018 respectively.

In Fig. C.12 the nodes/setts are representative of their true geographical locations. It should also be noted that there are different nodes in some sub-Figs. This is due to the fact that some setts were not active in a particular badger year. The edges represent the frequency/strength of movement between setts, for example, there is edge weight of 1 connecting Old Oak and Nettle. This represents that a badger was originally caught at Old Oak and then caught at Nettle. Hence, these network plots indicated that badgers seldom move between setts.

## C.2.2 Priors Settings

$\delta \sim \text{Uniform}(0,1), \beta_0 \sim \text{Normal}(0,0.1), \beta_1 \sim \text{Normal}(0,0.1), q_{11}^{\text{DPP}} \sim \text{Beta}(127.02, 131.12), q_{00}^{\text{DPP}} \sim \text{Beta}(10.22, 1.68), q_{11}^{\text{Ifn}} \sim \text{Beta}(26.41, 7), q_{00}^{\text{Ifn}} \sim \text{Beta}(9.95, 1.6), q_{11}^{\text{Cul}} \sim \text{Cul}(2.25, 12.26), q_{00}^{\text{Cul}} \sim \text{Beta}(60.61, 1.06), \gamma_t \sim \text{Uniform}(0,1)$ for $t = 1, \ldots, T$, $\text{logit}(\phi_{d_i}) = \alpha_{d_i}, \alpha_{d_i} \sim \text{Normal}(0,0.1), \text{logit}(p_{0_{d_i}}) = \alpha_{1_{d_i}}, \alpha_{1_{d_i}} \sim \text{Normal}(0,0.1), \sigma_{d_i} \sim \text{Uniform}(0,15).$

## C.2.3 Demographic Parameters

In Figs. C.13, C.14, C.15, C.16, the black dot is the posterior median and the thick line represents the 95% credible interval. Fig. C.17 displays the standardized density maps for population density across years in spring each year, where the black dots represent the setts trapped.
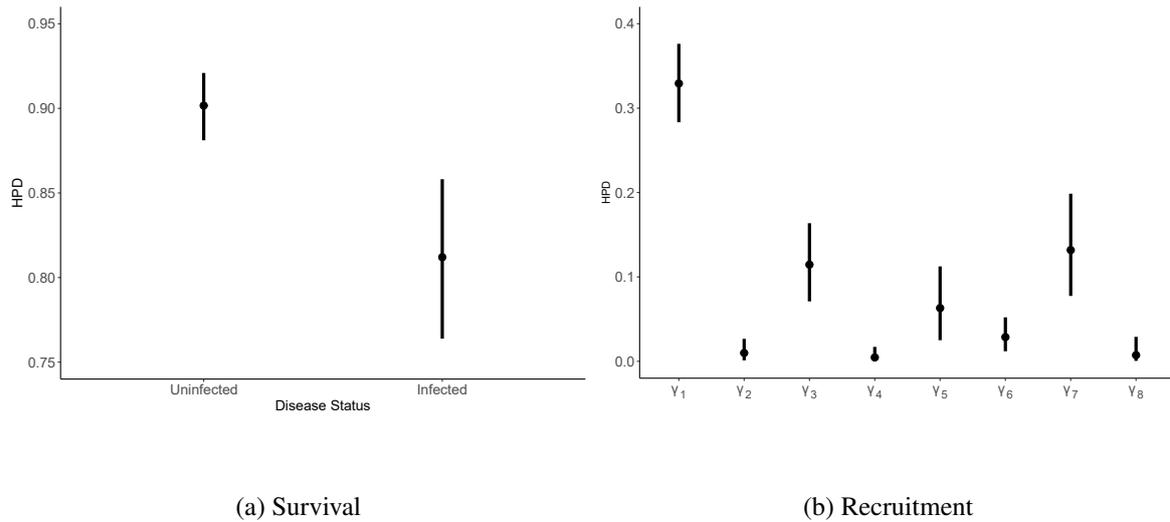


(a) Survival

(b) Recruitment

Fig. C.13 Caterpillar plots of posterior samples for survival probabilities and recruitment probabilities.

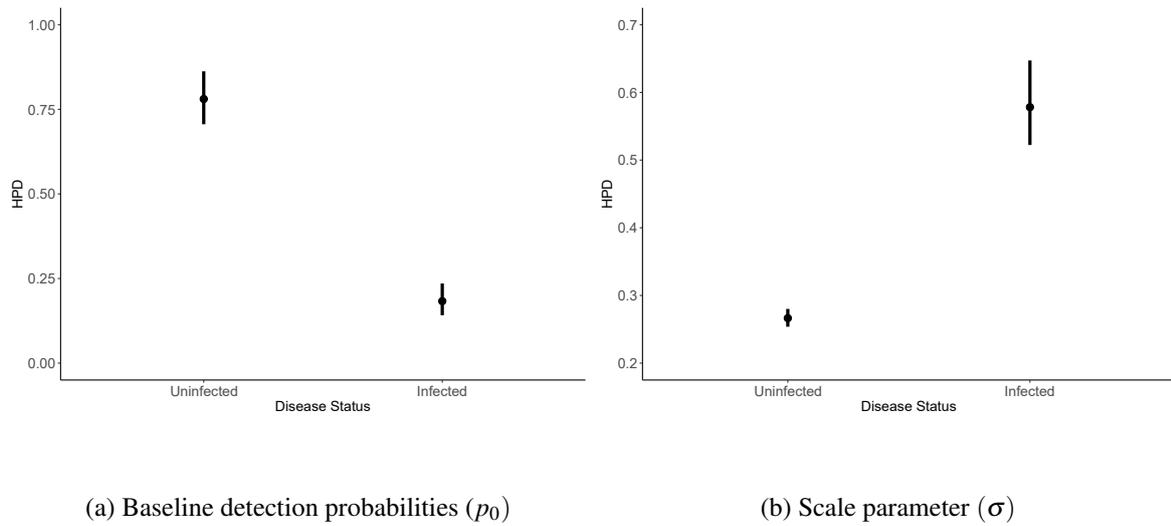(a) Baseline detection probabilities ($p_0$)

(b) Scale parameter ($\sigma$)

Fig. C.14 Caterpillar plots of posterior samples for baseline detection probabilities and for scale parameter ($\sigma$).



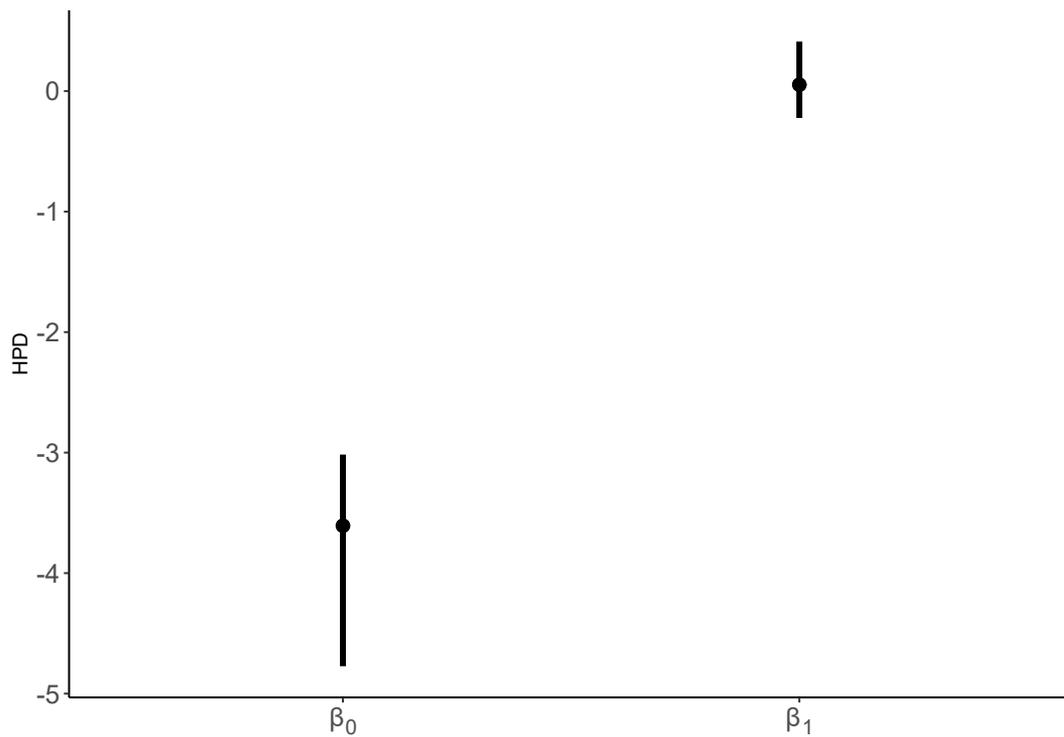Fig. C.15 Caterpillar plots of posterior samples for $\beta_0$ and $\beta_1$

(a) Sensitivity                                             (b) Specificity
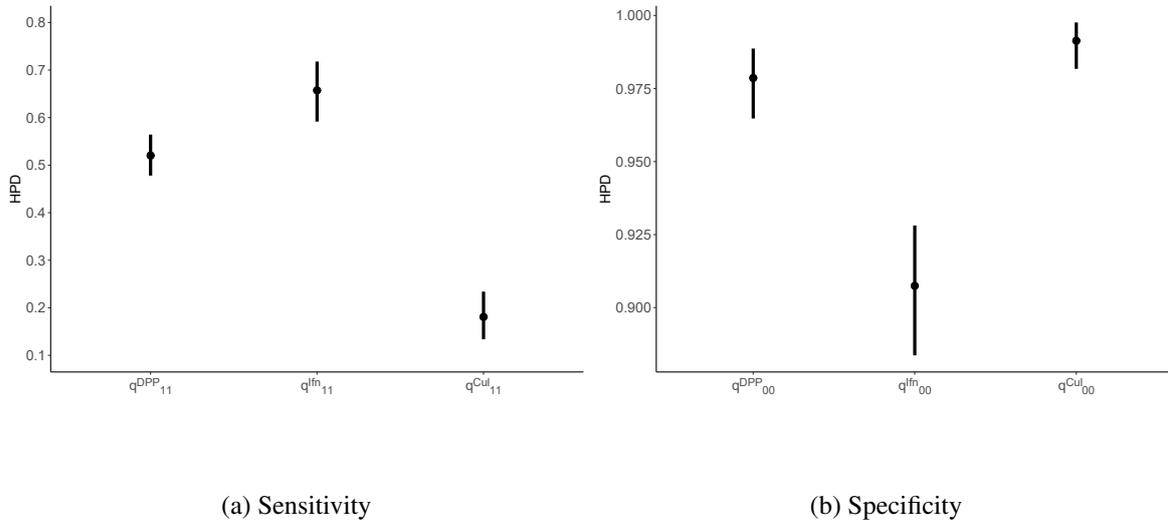
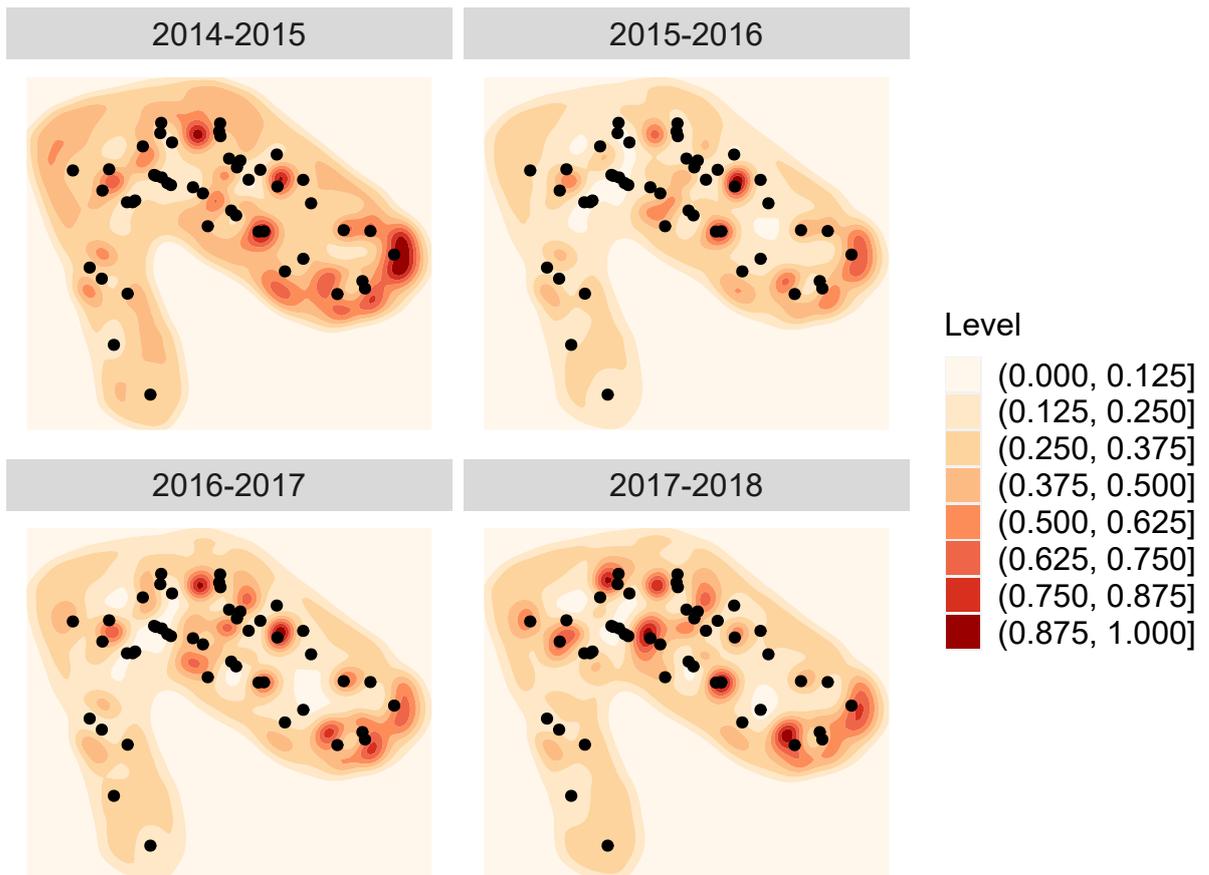Fig. C.16 Standardized Density maps for density across years in Spring each year.



Fig. C.17 Standardized Density maps for density across years in Spring each year, where the black dots represent the trapping location.