Bayesian methods for interpretable and scalable modelling of population dynamics

Ioannis Rotous

School of Mathematics, Statistics and Actuarial Science

University of Kent

This dissertation is submitted for the degree of

Doctor of Philosophy

June 2024

This dissertation is dedicated to my family: my mother, Ino; my father, Giorgos; and my late brother, Konstantinos who left us early. They have been constant sources of inspiration and support throughout my life. Their love, compassion, and sacrifices have been invaluable throughout my academic journey. I would also like to dedicate this work to my uncle Antonis, who from a young age, has been my inspiration and tutor in pursuing and loving mathematics.

Taking a new step, uttering a new word, is what people fear most.

Fyodor Dostoevsky

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Eleni Matechou and Alex Diana, for their guidance, support, and encouragement throughout my Ph.D. journey. Their expertise, knowledge, and willingness to share their time and ideas have been instrumental in the successful completion of this thesis. I have learned a great deal from them, and I will always be thankful for their patience and kindness. I would also like to acknowledge the funding I received for my Ph.D. from Engineering & Physical Sciences Research Council.

On a personal note, to my parents, Ino and Giorgos, 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 other family members, thank you for all your unwavering love and support.

I am equally grateful to my friends in Greece and United Kingdom, for their unwavering support and love throughout my academic journey. Their, understanding, and constant encouragement have been the pillars of my success. They has been there for me in the good times and the bad, offering a listening ear, a shoulder to cry on, and cheer me when I needed it the most. Without you, this achievement would not have been possible.

Abstract

The work in this thesis presents novel Bayesian methods for enhancing wildlife monitoring, a critical component in addressing climate change due to the regulatory role of species and their habitats in the climate system. Given the consistent decline in biodiversity, marked by shifts in life history events, the importance of effective wildlife monitoring is underscored. The thesis highlights the significance of understanding species behavior for better management, utilizing time-series monitoring and species-borne devices like GPS and acoustic recorders to gather detailed behavioral data. These studies focus on latent behavioral states with Markovian dependence, for which Hidden Markov Models (HMMs) are frequently employed. Monitoring population dynamics is also emphasized, as population size is a key measure for assessing biodiversity loss and informing conservation policies. Depending on survey duration and species characteristics, populations can be closed or open. Common models like Jolly-Seber (JS) and Cormack-Jolly-Seber (CJS) are used, with temporary emigration (TE) models accommodating species with seasonal patterns. The thesis addresses the critical role of sampling schemes in population dynamics studies. Capture-Recapture (CR) is extensively used but when infeasible, Batch-Mark (BM) sampling serves as an alternative, providing accurate inference in open population models. Employing parametric and non-parametric hierarchical Bayesian models, the thesis uses Bayes' Theorem to update prior beliefs based on observed data. The non-parametric approach, particularly using the Pólya Tree prior, offers flexibility by allowing data to shape the distribution. Markov Chain Monte Carlo (MCMC) methods are used to sample from posterior distributions for behavioral states and population parameters, demonstrating the robust application of these Bayesian methods in ecological datasets.

Chapter 2 develops a parametric Bayesian model to infer species behavior over time using Bayesian HMMs. The model addresses the challenge of selecting latent states by employing a

Reversible Jump MCMC algorithm and repulsive priors to prevent overfitting. Extensive simulations and real case studies on muskox and Cape gannets demonstrate the model's utility.

Chapter 3 introduces a non-parametric Bayesian model for population dynamics using JS models for BM data, leveraging the Pólya Tree Prior. The model is computationally efficient and exact, accommodating various survey designs and incorporating capture losses. Robustness is demonstrated through simulations and case studies on weather loaches and golden mantella frogs.

Chapter 4 addresses the complexity of TE models with Approximate Bayesian Computation (ABC), using Sequential Monte Carlo (SMC) ABC for efficient posterior sampling. The method's flexibility is shown through simulations and a case study on alpine common toad.

Overall, this thesis advances Bayesian methods for wildlife monitoring, providing robust tools for understanding species behavior and population dynamics, ultimately contributing to biodiversity conservation and climate change mitigation.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

Ecological wildlife monitoring plays a pivotal role in addressing the pressing issue of climate change, as species and their habitats contribute to regulate the climate system, as stated in Shin et al. (2022). However, there has been a consistent decline in biodiversity, marked by shifts in the timing of life history events (Mawdsley, et al., 2009; Diaz, et al., 2022), enhancing the heightened significance of wildlife monitoring. This imperative monitoring serves to deepen our comprehension of biodiversity decline. Therefore, this thesis develops novel Bayesian methods and explores their applications in various ecological datasets to strengthen the machinery of wildlife monitoring inference.

First and foremost, understanding species behavior is of paramount importance as it provides insights for better management of species (Schmidt et al., 2016). Most behavioral studies base their analysis on monitoring species across time to infer their unobserved underlying behavior, for example different type of movements or activities of species. In recent years, with the accessibility of technology, many studies are more commonly conducted using species-borne devices (Lahoz-Monfort and Magrath, 2021). These devices, for instance, can measure location using the Global Positioning System (GPS) or record species acoustics.

Practitioners seeking to infer the behavioral states of species only have access to observed responses generated by the species, such as the location/coordinates of the species or sounds measured by species-borne devices at a certain time. However, the main focus of the inference is the underlying latent behavioral state that gives rise to the observed responses. Additionally, behavioral states change over time, thus exhibiting Markovian dependence, where a behavioral state depends on the

previous behavioral states that took place. Consequently, frequently used models to infer such data of Markovian-dependent latent states that give rise to observables are typically referred to as Hidden Markov Models (HMMs) (Rabiner and Juang, 1986).

Monitoring population dynamics of species, as stated in McComb et al. (2010), is pivotal for addressing the current loss of biodiversity. The most frequently used measure is considering the population size of species, which informs us about the number of individuals of a species existing in their habitat at a given time. Anthropogenic activities, for example, such as hunting, trapping, and urban expansion, have caused species that are unable to adapt to experience decreases in population size and, in many cases, become extinct, leading to biodiversity loss. A way to monitor such cases of endangered species, who are on the risk of extinction is through their population size across time as stated in (Matthies et al., 2004), Hence, it is evident that population size is strongly correlated with the loss of biodiversity (through extinction of species), as it provides information such as whether species are increasing or declining, which is then used to inform conservation policies for species (Moussy et al., 2022).

Depending on the survey design, study area and species, the population can be regarded as either closed or open. The population is considered closed when the number of individuals of a species are not expected to change throughout the duration of the survey. This could be due to the short duration of the survey or the phenology of the species. On the contrary, a population is regarded as open if individuals can enter or exit the study area during the survey, through, for example births, deaths, immigration, and emigration. The models often used to infer such populations are Jolly-Seber (JS; Jolly, 1965; Seber, 1965) and Cormack-Jolly-Seber (CJS; Cormack, 1964; Jolly, 1965; Seber, 1965), where it is assumed that the emigration of species individuals is permanent. Additionally, less restrictive models called temporary emigration (TE) models allow for individuals to leave the study area, and this can be either permanent or temporary. Hence, such a model can be used in cases where the species that we monitor exhibit, for example, seasonal migration, breeding, or even when individuals traverse multiple times through the study area.

To derive inference on the population dynamics of monitored species, we must choose from a range of different sampling schemes. The appropriate choice relies on factors such as the research objectives, the study area, and the characteristics of the species to determine the sampling method.

Capture-Recapture (CR) is one popular sampling method that has been extensively used in the literature. The observation procedure involves sampling the study area across multiple sampling occasions during which a portion of individuals from the population is captured. Newly captured individuals receive an individual unique mark and are later released back into the population, along with the individuals that had already been marked during previous sampling occasions. Therefore, CR data contain individual information, as unique marks are distributed among the newly captured individuals, which allows us to derive estimates for the population size, capture probability and survival probability. However, this is not always feasible due to cost, time and species size constrains (Rosser et al., 2022). In those cases, a an alternative sampling scheme can be employed, called Batch-Mark (BM), which is considered in two chapters of this thesis. The BM sampling method involves sampling a study area across multiple sampling occasions, where a proportion of individuals from the population are captured on each occasion. The newly captured individuals receive a marking that is unique per sampling occasion but not individuals, and are later released into the population. Therefore, the difference with CR is that the marking for newly caught individuals differs per sampling occasion, rather than per individual. Thus, it is evident that BM data can be viewed as the aggregated version of CR, aggregating the individual capture histories conditional on the sampling occasion that were first caught. With that said, CR is expected to be more informative than BM. This holds true only for open population models, where still the BM provides accurate inference; in closed population cases, both sampling schemes are equivalent, only when individual heterogeneity is regarded negligible.

The sampling schemes considered in this thesis accommodate nuisance parameters such as capture probabilities. These nuisance parameters are the result of our sampling efforts in the study area. The sampling effort is always constrained by time and financial limitations, which restricts our ability to capture individuals with certainty while in the area. Hence, the models created throughout this thesis make inferences about the population dynamics of species by taking into account the nuisance parameter of capture probability.

To perform inference, we use parametric and non-parametric hierarchical Bayesian models. Both the parametric and non-parametric approaches rely on the use of Bayes' Theorem for updating our prior beliefs for the parameters of the model conditional on the observed data. In the parametric case, we place prior distributions directly on the model parameters. These priors can be fully specified using

a finite set of parameters. Conversely, in the nonparametric case, we assume that the model parameters themselves are drawn from a distribution. We then place a prior on this underlying distribution. This means that our prior domain is the space of distributions, which is a functional space. This approach allows for greater flexibility as the data can directly influence the shape of the overall distribution. In particular for the non-parametric case, we use the Pólya Tree prior first introduced in Lavine (1994). In either case, we sample from the corresponding posterior distribution using Markov Chain Monte Carlo (MCMC) methods to obtain samples. These samples are then used to calculate posterior summary statistics for the behavioral states and population dynamic parameters of interest.

In Chapter 2, we develop a novel parametric Bayesian model for inferring the underlying latent states of species behavior across time. Since the states exhibit Markovian dependence over time, we employ a Bayesian HMM. However, one of the key challenges when applying HMM models is the appropriate selection of latent states, given that the true number of latent behavioral states is unknown in reality. Therefore, one must either pre-define the number of latent states based on prior knowledge, or compare models with predefined but different numbers of latent states, or alternatively, allow the number of latent states to be inferred from the model through the data. In this thesis, we use the latter approach, since in Bayesian inference, every unknown is regarded as a parameter to be inferred from the data. Hence, to accomplish this, we utilize a Reversible Jump Markov Chain Monte Carlo (RJMCMC) algorithm that allows us to explore different numbers of latent states while simultaneously making inferences about the parameters characterizing the behavioral states. HMMs often exhibit signs of overfitting and derivation of overly complex models. To tackle this issue, we introduce the use of repulsive priors (Natarajan et al., 2023; Petralia et al., 2012; Quinlan et al., 2021), which penalize the distance between behavioral states, leading to more parsimonious models. Overall, we demonstrate the usefulness of our model by an extensive simulation study and by analyzing real case studies of GPS data for muskox, *Ovibos moschatus* (Pohle et al., 2017) and acoustic data for Cape gannets, *Morus capensis* (Thiebault et al., 2021), with results displayed in Chapter 2.

Chapter 3 proposes a novel non-parametric Bayesian model for inferring the dynamics of a population using JS models for BM data. In this chapter, we utilize the Pólya Tree Prior, which aligns with the nature of the BM data, aggregated form of CR data, to infer the distribution of the number of individuals in any entry/exit configuration. Firstly, we created a model that is computationally efficient

and exact, in contrast to existing models for BM data, and readily extends to consider data collected under different survey designs. The chapter demonstrates how losses in capture can be incorporated within this new framework and how ignoring these losses can lead to biases in the estimated parameters of the underlying population. Secondly, we analysed BM data under the robust design (RD Pollock, 1982), by deriving an exact likelihood function for the RD observation process. When utilizing the RD method, it is assumed that the population remains open during primary periods, such as months, but closed within secondary periods, like days within a month. This means that individuals can leave the population between primary periods but not between secondary periods. We demonstrate that our model is consistent in inferring the true population dynamics through an extensive simulation study and the analysis of two real case studies involving data collected for weather loaches (*Misgurnus anguillicaudatus*) (Cowen et al., 2017) and golden mantella frogs (*Mantella aurantiaca*) (Zhang et al., 2023), with results displayed in Chapter 3.

Finally, in Chapter 4, we develop an approximate Bayesian computation methodology for addressing overly complicated models such as TE to infer the dynamics of a species population from BM data. In this case, the likelihood function of the model becomes intractable due to the complexity of the model's latent parameters. To overcome this problem, we use a method called Approximate Bayesian Computation (ABC) (Marin et al., 2012), a flexible approach that involves approximated repetitive sampling from regions of high posterior mass. In particular, we employ Sequential Monte Carlo (SMC) ABC (Del Moral et al., 2006), which avoids pitfalls of vanilla ABC and provides more efficient approximate posterior sampling. We demonstrate the flexibility and ease of use of these methods through a simulation study and the analysis of a real case study involving collected data of the common toad (*Bufo bufo*), with results displayed in Chapter 4.

Chapter 2

Hidden Markov models with an unknown number of states and a repulsive prior on the state parameters

Abstract

Hidden Markov models (HMMs) offer a robust and efficient framework for analyzing time series data, modelling both the underlying latent state progression over time and the observation process, conditional on the latent state. However, a critical challenge lies in determining the appropriate number of underlying states, often unknown in practice. In this chapter, we employ a Bayesian framework, treating the number of states as a random variable and employing Reversible Jump Markov Chain Monte Carlo to sample from the posterior distributions of all parameters, including the number of states. Additionally, we introduce repulsive priors for the state parameters in HMMs, and hence avoid overfitting issues and promote parsimonious models with dissimilar state components. We perform an extensive simulation study comparing performance of models with independent and repulsive prior distributions on the state parameters, and demonstrate our proposed framework on two ecological case studies: GPS tracking data on muskox in Antarctica and acoustic data on Cape gannets in South Africa. Our results highlight how our framework effectively explores the model space, defined by models with different latent state dimensions, while leading to latent states that are distinguished better and hence are more interpretable, enabling better understanding of complex dynamic systems.

2.1 Introduction

Hidden Markov models (HMMs) are a powerful and well-established framework for analyzing time series data in cases where the studied system transitions between a set of hidden states over time. HMMs jointly model two processes: the underlying latent process of the hidden states, and the observation process, conditional on the states, as shown in Figure 2.1 (Cappé et al., 2009; Zucchini and MacDonald, 2009). HMMs enable efficient modelling of the evolution of the latent states across time and, conditional on those latent states, explicit modelling of the data observation process, even in

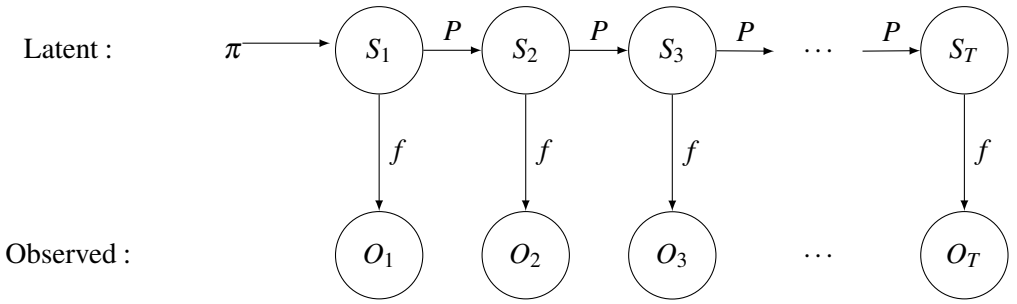

Fig. 2.1 Illustration of a hidden Markov model evolution across $t = 1, 2, ..., T$ time points, with latent states S_t and corresponding observations O_t, characterized by an initial latent distribution π, transition probabilities P, and emission distribution f.

complex systems and processes with multiple latent states and complicated observation processes (Popov et al., 2017).

HMMs commonly employ a first-order Markov chain, where the evolution of the latent states depends only on the previous time point. Additionally, conditional on the latent state, they emit observables at the current time point, independent of the rest of the observables. Further details can be found in Section 2.2.1, which describes the joint distribution of observables with latent states (Equation 2.2.1). The efficacy of HMMs relies on the separation of the latent and observation processes and the use of algorithms that efficiently marginalize over the latent states, such as the forward/backward algorithm for computing the likelihood function and the Viterbi algorithm for finding the most likely sequence of hidden states (Bartolucci et al., 2013; Zucchini and MacDonald, 2009). Hence, HMMs have proven to be powerful and easy-to-use tools, and are widely utilized in various fields, such as finance (Rydén et al., 1998), biology (Leroux and Puterman, 1992), social science (Rabiner, 1989; Zucchini and MacDonald, 2009), medicine (Farcomeni, 2017), and ecology (Patterson et al., 2017; Schmidt et al., 2016), among others.

One of the key challenges in employing HMMs for data analysis is the decision on the appropriate number of underlying states in the system. In practice, the true number of states is often unknown. It is standard practice to fix the number of latent states or fit models that consider different numbers of latent states (Chib, 1996; Robert and Titterington, 1998; Robert et al., 1993) in either a classical (Huang et al., 2017), or Bayesian framework (Berkhof et al., 2003), and subsequently compare them with appropriate criteria to select the number of states. However, in this case, the model needs to be

fitted multiple times, and in the end, a single model is used for interpretation, but without accounting for the uncertainty in the model selection process itself (McLachlan et al., 2019). Alternatively, in a Bayesian framework, the number of latent states can be treated as an additional random variable, and hence Reversible Jump Markov Chain Monte Carlo (RJMCMC) (Green, 1995) methods can be employed to sample from the posterior distribution in this case where the model dimension is not fixed; indeed RJMCMC has been used extensively within an HMM context (Cappé et al., 2003, 2009; Robert et al., 2000; Russo et al., 2022). We note that HMMs can equivalently be viewed from the perspective of dynamic mixture models (Spezia, 2020), both in discrete and continuous space (Bartolucci and Pandolfi, 2011; Reynolds et al., 2009), with the number of mixture components corresponding to the number of states. In this chapter, we refer to HMMs and corresponding states, but the concepts equally apply to dynamic mixture models and corresponding components.

Within a Bayesian framework, state allocation (S_t in Figure 2.1) can be sampled within the MCMC algorithm (Stephens and Phil, 1997), so that the complete data likelihood is used (King, 2014). This approach however can lead to a substantial number of sampled latent variables. Instead, state allocation can be marginalised out of the model, as is standard practice within the HMM machinery (Russo et al., 2022), so that the observed data likelihood is used for inference, and this is the approach we employ in this chapter. However, in either case, HMMs with a variable number of states are prone to overfitting, and hence such algorithms can lead to an unnecessarily large number of similar states (Duan and Dunson, 2018). Recent advancements in the field of mixture modelling have introduced the use of repulsive priors, which promote parsimony in the model (Natarajan et al., 2023; Petralia et al., 2012; Quinlan et al., 2021). These repulsive prior distributions serve to impose constraints on the proximity of state parameters, which discourages similar states from being created. Gladly, this particular form of penalty applied to the parameter space also affects the selection of the number of states (Natarajan et al., 2023). An additional advantage associated with incorporating a repulsive prior into HMMs, whether with fixed or variable dimensions, is the mitigation of overfitting. In certain instances, conventional mixture models, and hence HMMs, may excessively fine-tune their components (states) to capture noise in the data, resulting in poor generalization of the obtained results. Extensive research, as documented in the literature (Beraha et al., 2022; Petralia et al., 2012; Quinlan et al., 2021), has demonstrated that these issues can be effectively addressed by introducing

repulsion constraints among distribution parameters within the model, thereby promoting dissimilarity among states.

To introduce a repulsion constraint within an HMM framework we use interaction point processes, referred to as a repulsive prior in this chapter, which is a class of distributions on a set of points that actively penalises cases where points (parameters) are close together (similar). Specifically, we consider a distribution belonging to the family of pairwise interaction point processes, the Strauss point process prior, first described in Strauss (1975), as a prior distribution on the state parameters of the emission distribution f as illustrated in Figure 2.1. This approach enables us to achieve more effective and interpretable modelling without pre-specifying the number of states, thereby improving various aspects of analysis, including inference, model selection, and our overall understanding of system dynamics.

HMMs have been extensively used in the ecological literature (Gimenez et al., 2009; Langrock, 2012; Patterson et al., 2017; Schmidt et al., 2016), since they are well-suited to capturing the underlying latent state structure in ecological systems, enabling researchers to seamlessly integrate the observed data with the unobserved latent states (Glennie et al., 2023). In ecological systems, these latent states can correspond to life stages (McClintock et al., 2020) or behavioural states (Nicol et al., 2023; Schmidt et al., 2016). In this chapter, we demonstrate our approach using two ecological case studies: GPS data on muskox, *Ovibos moschatus*, in east Greenland, also analysed in Pohle et al. (2017), who used model selection criteria to select the number of states in their HMM, and acoustic data on Cape gannets, *Morus capensis*, in South Africa, analysed in Thiebault et al. (2021) where behavioural state classification was performed manually to train a subsequent model. Our results demonstrate how our framework effectively explores the model space, defined by models with different latent state dimensions, while leading simpler models with latent states that are distinguished better and hence are more interpretable. Finally, our extensive simulation study demonstrates that, when the true model is fitted to the data, then the repulsive prior leads to inference, in terms of density estimation and state classification, that is equivalent to, or marginally better than, that of independent priors. Therefore, for noisy, real data, where the model is typically a simplified version of the actual data-generating process, the repulsive prior avoids overfitting and leads to more parsimonious models, whilst in data

simulated from the fitted model, the repulsive prior does not overpenalise, and leads to results that are on par to those obtained from an independent prior.

The chapter is structured as follows: Section 2.2 introduces the general concepts of HMMs and repulsive priors, and gives a broad overview of the model-fitting approach developed in this chapter, with technical details provided in Appendix A (page 93). Section 2.3 discusses the results of our extensive simulation study, while Section 2.4 presents the results of modelling the muskox GPS data, and Section 2.5 the results of modelling the Cape gannet acoustic data. Finally, the chapter concludes with a discussion in Section 2.6.

2.2 Models

2.2.1 Hidden Markov Models

A first-order hidden Markov model is a stochastic process consisting of a set of hidden/latent states S and observations O. The state process is assumed to be an N-state Markov chain $P(S_t|S_1,S_2,...,S_{t-1}) = P(S_t|S_{t-1})$ with $S_t \in \{1,2,...,N\}$. The evolution of the hidden states across time is described by the transition matrix P, where P_{ij} is the probability of transitioning from state i to state j for all t, i.e.,

$$P(S_t = j|S_{t-1} = i) = P_{ij}.$$

The probability of being in a particular state at the first time point can be modeled using an initial state distribution π i.e. $P(S_1 = i) = \pi_i$. At each time step, we observe O_t, whose distribution only depends on the current value S_t,

$$f(O_t|O_1,...,O_{t-1},S_1,...,S_t) = f(O_t|S_t).$$

Therefore, the model for a particular sequence of observations given the hidden states, $f(O_1,O_2,...,O_T|S_1,S_2,...,S_T)$, can be factorised as $\prod_{t=1}^{T} f(O_t|S_t)$, and the joint model of a particular sequence of hidden states and observations is equal to

$$P(O_1,O_2,...,O_T,S_1,S_2,...,S_T) = \pi_{S_1} f(O_1|S_1) \prod_{t=2}^{T} P_{S_{t-1},S_t} f(O_t|S_t) \tag{2.2.1}$$

The emission distribution, f, which describes how the observations are generated conditional on the states, is a function of corresponding state-specific parameters, θ_i, where θ_i can be a scalar or a vector of parameters, and it is on these parameters that we place the repulsive prior distributions proposed in this chapter.

2.2.2 Repulsive prior

We assume a pairwise interaction point processes on the parameters θ_i described in Section 2.2.1. Pairwise interaction point processes can be constructed by defining a point process with density of the form

$$h(\theta_1,\theta_2,...,\theta_N|\xi_1,\xi_2) = \frac{1}{Z_\xi} g(\theta_1,\theta_2,...,\theta_N|\xi_1,\xi_2) = \frac{1}{Z_\xi}\prod_{i=1}^{N}\phi_1(\theta_i|\xi_1)\prod_{1\le i<j\le N}\phi_2(\theta_i,\theta_j|\xi_2) \quad (2.2.2)$$

where $g(\cdot)$ corresponds to the unnormalized density of $\theta_1,\cdots,\theta_N$, $\phi_1(\cdot)$ is a nonnegative function and $0 \le \phi_2 \le 1$ (repulsiveness) and $\xi = (\xi_1,\xi_2)$. More details on interaction point processes can be found in Moller and Waagepetersen (2003). It is convenient to take $\phi_1(\theta_i|\xi_1) = \xi_1\mathbb{I}[\theta_i \in R]$, where ξ_1 is the intensity of the points and R is the region where θ is defined. In our case, we use the Strauss process (Strauss, 1975), which assumes

$$\phi_2(\theta_i,\theta_j|\xi_2 = \{a,d\}) = a^{\mathbb{I}[\|\theta_i-\theta_j<d\|]}.$$

This term denotes the interaction term between the locations θ_i,θ_j for parameters a,d. Parameter a ranges from 0 to 1 and controls the penalty magnitude between the points θ_i and θ_j; the smaller the a the stronger the penalty. Parameter d is the threshold such that if the distance (typically the Euclidean distance) between two components is less than d, the penalty applies. Moreover, there exist alternatively interaction functions ϕ_2 which apply continuous penalty as the one used in Petralia et al. (2012). However, we used the Strauss process since it is the simplest non-trivial interaction function. Lastly, the normalizing constant Z_ξ of Equation (2.2.2) is intractable, which makes inference on parameter ξ challenging. Additionally, we adopt the approach described by Beraha et al. (2022), maintaining the penalty a and threshold d constant throughout the chapter. If $a = 1$ there is no penalty,

and we retrieve a point process (referred to as the independent point process for the rest of the chapter), with points being drawn from an independent Uniform distribution with $\phi_1(\theta_i|\xi_1) = \xi_1 \mathbb{I}[\theta_i \in R]$,

$$h(\theta_1, \theta_2, ..., \theta_N|N, \xi_1) = \frac{1}{Z_{\xi_1}} \prod_{i=1}^{N} \phi_1(\theta_i|\xi_1), \tag{2.2.3}$$

In this case, the normalizing constant is tractable and equal to $Z_{\xi_1} = \xi_1^N |R|^N$.

Finally, bringing together the concepts described in Sections 2.2.1 and 2.2.2, the hierarchical representation of an HMM model with a random number of states and a repulsive prior on the state parameters is:

$$\begin{aligned}
&N \sim \text{Uniform}\{1, 2, \cdots, N_{max}\} \\
&O_t|S_t, \theta \sim f(O_t|\theta_{S_t}),\ t = 1, 2, ..., T \\
&\underline{\theta} = (\theta_1, \theta_2, ..., \theta_N) \sim \text{StraussProcess}(\xi_1, \xi_2) \\
&P(S_1 = i) = \pi_i,\ i = 1, 2, ..., N \\
&P(S_t = j|S_{t-1} = i) = P_{ij},\ i, j = 1, 2, ..., N,\ \ t = 2, 3, ..., T \\
&\underline{P_{i.}} = (P_{i1}, P_{i2}, ..., P_{iN})|N \sim \text{Dirichlet}(a_1^P, a_2^P, ..., a_N^P),\ i = 1, 2, ..., N \\
&\underline{\pi} = (\pi_1, \pi_2, ..., \pi_N)|N \sim \text{Dirichlet}(a_1^\pi, a_2^\pi, ..., a_N^\pi)
\end{aligned} \tag{2.2.4}$$

2.2.3 Inference

Inference is made on the parameters $\underline{\theta}, \underline{\pi}, P, \xi$ and N. Since the dimension of $\underline{\theta}, P, \underline{\pi}$ changes according to N, we employ a RJMCMC sampling algorithm that allows us to move between models with different parameter dimensions. On each iteration of the algorithm, we implement a fixed and a variable dimension move. The fixed dimension move updates the model parameters $(\underline{\theta}, P, \underline{\pi})$ conditional on the number of states, and the variable dimension move updates the dimension of the model. Finally, we update ξ with the use of the exchange algorithm (Murray et al., 2012), described in Section 2.2.4.

• **Fixed dimension Moves**

We update the model parameters $\underline{\pi}, P, \underline{\theta}$, for a fixed value N, by sampling from the corresponding posterior distributions using a Metropolis Hastings algorithm (Hastings, 1970; Metropolis et al., 1953), since the HMM likelihood with state allocation marginalised out is not conjugate to the prior distribution(s).

- **Variable dimension moves**

With probability 0.5, we choose between the moves Split/Combine and Birth/Death.

Split/Combine moves

In each step, we choose whether to split or combine states with probability 0.5. In the split case, if we have a single state, with probability one we propose to split that state. If we have more than one states, we choose uniformly one of the N states, denoted by j_*, which we propose to split into j_1 and j_2, therefore proposing to split $\pi_{j_*}, P_{j_*,\cdot}, P_{\cdot j_*}, \theta_{j_*}$ into new model parameters $(\pi_{j_1}, P_{j_1\cdot}, P_{\cdot j_1}, \theta_{j_1})$ and $(\pi_{j_2}, P_{j_2\cdot}, P_{\cdot j_2}, \theta_{j_2})$.

The split move is accepted with probability $\min\{1, A\}$, where

$$A = \frac{f(\{O_t\}_{t=1}^T \,|\, \{\pi\}_{j=1}^{N+1}, \{P_{j\cdot}\}_{j=1}^{N+1}, \{\theta_j\}_{j=1}^{N+1})}{f(\{O_t\}_{t=1}^T \,|\, \{\pi\}_{j=1}^{N}, \{P_{j\cdot}\}_{j=1}^{N}, \{\theta_j\}_{j=1}^{N})} \frac{p(\{\pi\}_{j=1}^{N+1}, \{P_{j\cdot}\}_{j=1}^{N+1}, \{\theta_j\}_{j=1}^{N+1}, N+1)}{p(\{\pi\}_{j=1}^{N}, \{P_{j\cdot}\}_{j=1}^{N}, \{\theta_j\}_{j=1}^{N}, N)} \frac{q(N+1 \to N)}{q(N \to N+1)}$$

where $p(\cdot)$ is the joint prior distribution of all parameters, and $q(N+1 \to N)$ and $q(N \to N+1)$ are the proposal probabilities for the transdimensional moves with details given in Sections A.5.2 (page 102) and A.6.1 (page 109) in Appendix A.

In the combine case, we choose the two states j_1 and j_2 whose distance is the smallest, and we propose to combine them to j_*. The combine move is accepted with probability $\min\left\{1, A^{-1}\right\}$.

Birth/Death moves

The Birth/Death move is performed similarly to the Split/Combine move. Specifically, if we have N states, we choose with probability 0.5 to give birth to a new state or kill an existing one.

In the birth move, we propose a new state generated by sampling its parameters from the prior distribution. On the other hand, for the death move, we uniformly choose a state and propose to kill it. In this case, the acceptance probability of the birth move is again $\min\{1, A\}$ whereas for the death move it is $\left\{1, A^{-1}\right\}$ with

$$A = \frac{f(\{O_t\}_{t=1}^T \,|\, \{\pi\}_{j=1}^{N+1}, \{P_{j\cdot}\}_{j=1}^{N+1}, \{\theta_j\}_{j=1}^{N+1})}{f(\{O_t\}_{t=1}^T \,|\, \{\pi\}_{j=1}^{N}, \{P_{j\cdot}\}_{j=1}^{N}, \{\theta_j\}_{j=1}^{N})} \frac{p(\{\pi\}_{j=1}^{N+1}, \{P_{j\cdot}\}_{j=1}^{N+1}, \{\theta_j\}_{j=1}^{N+1}, N+1)}{p(\{\pi\}_{j=1}^{N}, \{P_{j\cdot}\}_{j=1}^{N}, \{\theta_j\}_{j=1}^{N}, N)} \frac{q(N+1 \to N)}{q(N \to N+1)}$$

where $p(\cdot)$ is the joint prior distribution of all parameters, and $q(N+1 \to N)$ and $q(N \to N+1)$ are the proposal probabilities for the transdimensional moves with details given in Sections A.5.2 (page 102) and A.6.1 (page 109) in Appendix A.

2.2.4 Update ξ

Parameter ξ (equivalently intensity ξ_1 described in Section 2.2.2) of the Strauss process prior is updated with a Metropolis Hastings algorithm. At each iteration, we propose ξ_* from

$$\xi_* \sim q(\xi_*|\xi) = \text{LogNormal}(\log(\xi), \tau_\xi)$$

However, calculation of the Metropolis Hastings acceptance ratio depends on the ratio of the corresponding densities (Equation 2.2.2) at ξ and ξ_*,

$$\frac{h(\theta_1,\theta_2,...,\theta_N|N,\xi_*,a,d)}{h(\theta_1,\theta_2,...,\theta_N|N,\xi,a,d)} = \frac{\frac{1}{Z_{\xi_*}}\prod_{i=1}^{N}\xi_*\mathbb{I}[\theta_i \in R]}{\frac{1}{Z_{\xi}}\prod_{i=1}^{N}\xi\mathbb{I}[\theta_i \in R]} = \frac{Z_\xi}{Z_{\xi_*}}\frac{\prod_{i=1}^{N}\xi_*\mathbb{I}[\theta_i \in R]}{\prod_{i=1}^{N}\xi\mathbb{I}[\theta_i \in R]} \tag{2.2.5}$$

which is intractable. Therefore, we employ the exchange algorithm of Murray et al. (2012), and simulate a new parameter, $\underline{\theta_{aux}}$, from the Strauss process (Equation 2.2.2) with parameter ξ_* using the birth and death algorithm (Møller and Sørensen, 1994) described in Section A.1 (page 93) of Appendix A, and accept ξ_* with probability $\min(1,A)$, where

$$A = \frac{q(\xi|\xi_*)p(\xi_*)g(\theta_1,\theta_2,\ldots,\theta_N|\xi_*,a,d)}{q(\xi_*|\xi)p(\xi)g(\theta_1,\theta_2,\ldots,\theta_N|\xi,a,d)}\frac{g(\underline{\theta_{aux}}||\theta_{aux}|,\xi,a,d)}{g(\underline{\theta_{aux}}||\theta_{aux}|,\xi_*,a,d)} \tag{2.2.6}$$

with $p(\xi)$ the prior distribution assigned on parameter ξ, $g(\cdot)$ the unormalised density of the Strauss process in Equation (2.2.2) and $\underline{\theta_{aux}}$ the simulated new parameters vector of size $|\theta_{aux}|$.

In this chapter, we choose a and d based on the recommendation of Beraha et al. (2022). The threshold is calculated as $d = \min_{r>0}\{r : \text{local minimum for } p(r)\}$. Then r corresponds to all pairwise distances between the observed data generated from the HMM across time, and since the calculation of those distances is symmetric we focus only on the $r > 0$. Also, $p(r)$ is the kernel density of all pairwise distances r between the observations in the sample. We present examples that demonstrate this process in Section A.2 (page 94) of Appendix A. The penalty a is set equal to $\exp(-n^* \log(k_s))$,

where n^* corresponds to the minimum acceptable size of a state, for example 5% of the sample size, and set $\log(k_s) = 1$, so that a is only a function of n^*. By regarding a as function of only n^*, we can explicitly tune the penalty dependent on our subjective knowledge of what size of clusters are inferentially meaningful for our case studies. Further details on the penalty choice are given in Section A.2 (page 94) of Appendix A.

2.2.5 Label Switching

Inference for mixture models is usually complicated by label switching, which occurs because the labels of states are interchangeable, since reordering the states has no effect on resulting inference. Addressing label switching is important for interpreting the resulting states and corresponding parameters.

In this chapter, we employ two approaches to choose an ordering of states and hence deal with label switching. The first, employed in the case study of Section 2.4, involves imposing an ordering on one of the state parameters in cases when such ordering is meaningful, such as $\theta_1 \leq \theta_2 \leq ... \leq \theta_N$, with θ_i the mean of the state i, as demonstrated by Russo et al. (2022). During each iteration of the RJMCMC, the parameters are rearranged according to this predefined order, ensuring that the same state in different iterations maintains its label. The second, employed in the case study of Section 2.5, is the post-processing method of Bartolucci and Pandolfi (2011), performed after the end of all MCMC iterations, which involves computing the posterior mode (MAP) and subsequently determining, for each distinct posterior sample, the permutation that minimizes the distance between the MAP estimate and the permuted posterior sample.

2.2.6 State Allocation

As discussed in Section 2.2, our approach relies on marginalising over the latent state allocation instead of sampling it as part of the MCMC inference. Therefore, when interest lies in interpreting state allocations, these can be obtained at a post-processing stage by sampling from their posterior

distribution

$$P(S_1 = j|\{O_t\}_{t=1}^T, \pi_j, \theta_j) \propto \pi_j f(O_1; \theta_j), \qquad t = 1$$
$$P(S_t = j|S_{t-1} = i, \{O_t\}_{t=1}^T, P_{i,j}, \theta_j) \propto P_{i,j} f(O_t; \theta_j), \qquad t > 2$$

2.3 Simulation Study

We conducted an extensive simulation study to compare a model with a repulsive prior with a model with an independent prior. We simulated data from a model with five states, each described by a normal distribution with mean locations $\underline{\mu} = (-10, -5, 0, 5, 10)$ and standard deviations $\underline{\sigma} = (\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5)$ chosen so that there is increasing amount of overlap between the state distributions. Specifically, we took $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \sigma_5$ and we chose them to have value 1.1408, 1.4726, 2.5709 and 4.2319, such the overlap between consecutive mixture is equal to 3%, 9%, 33% and 55%, respectively, which corresponds to 5%, 15%, 50%, and 75% overall overlap. The consecutive index overlap is calculated by integrating the area where the two density functions overlap, which is equal to $\int_{\mathbb{R}} \min\left[\text{Normal}(x; \mu_i, \sigma_i), \text{Normal}(x; \mu_{i+1}, \sigma_{i+1})\right] dx$ for $i = 1, 2, ..., 4$ whereas the overall overlap is based on the overlap index described in Pastore and Calcagnì (2019), We plot the corresponding state distributions in each case in Section A.4 (page 95) of Appendix A. For each degree of overlap, we varied the sample size $n = \{50, 100\}$ and number of time points $T = \{5, 10\}$. The initial probability distribution π and transition probability matrix P were chosen to have all elements equal to 1/5, ensuring equal state sizes across all time points, allowing us to focus on the effect of state overlap on inference.

The repulsive parameters, penalty a and threshold d, were chosen according to Beraha et al. (2022) as described in Section 2.2.4. We considered two values of n^*: $n^*_{2.5}$, corresponding to penalties that consider cluster of sizes less than 2.5% and n^*_5, corresponding to penalties that consider cluster of sizes less than 5%. The results are given in Tables A.1 and A.2 of Section A.4 (page 95) of Appendix A. For each simulation scenario, we placed repulsive (Equation 2.2.2) and independent (Equation 2.2.3) priors on the location parameters μ. Finally, we placed Dirichlet priors with parameters equal to

1 on the initial and transition probabilities π and P and a uniform distribution on standard deviations σ with lower and upper bound 0 and $2\times$ 90% quantile of the observations, respectively.

We report the Kullback-Liebler (KL) divergence between the true state distributions and the estimated distributions, together with the misclassification rate between the true state allocation of observations and the inferred allocation, averaged across MCMC iterations, time points, and 100 replications for each scenario. Details about the calculation of KL divergence and miscassification rate are given in Section A.4 (page 95) of Appendix A.

We employed the ordering constraint described in Section 2.2.5, with $\mu_i \leq \mu_{i+1}$, for $i = 1, 2, ..., N$ to deal with label switching and we calculate the misclassification rate in each case computing state allocation as described in Section 2.2.6. Finally, we summarise the posterior mode of the distribution for the number of states in each scenario, averaged across all replications.

The results for case with $n^*_{2.5}$ and n^*_5 are presented in Table A.1 and A.2, respectively, of Section A.4 (page 95) in Appendix A. The results demonstrate that the two models have very similar performance in terms of selected number of states, density estimation and misclassification but with a small advantage of the RP. For penalty $n^*_{2.5}$, 13/16 cases and 10/16, in Table A.1, KL and misclassification error is lower for RP compared to ID. On the other hand, for penalty n^*_5, 14/16 and 10/16, in Table A.2, KL and misclassification error is lower for RP compared to ID. As the number of time points T or sample size n increases, KL divergence and misclassification error decrease for both models in general. As expected, as the amount of overlap increases, the posterior mode of the number of states decreases for both models. In this case, there is no obvious differences in the effects of the two chosen penalties.

2.4 Case study 1: Muskox GPS data

We consider data on muskox movement in east Greenland analysed in Pohle et al. (2017). The data consist of 25103 hourly locations of a single adult female muskox collected with a GPS collar, covering a period of roughly three years, giving information on the step length, L_t, which represents the distance in meters between time points $t-1$ and t, and the turning angle between time points $t-2$ and t, A_t, as is standard practice in GPS tracking data (Langrock, 2012; Patterson et al., 2017;

Zucchini and MacDonald, 2009). The data are visualised in Figures A.4 and A.5 of Appendix A.5 (page 99).

We model the step-length at time t, L_t, using a 0-inflated Gamma distribution to account for the number of 0s in the data (0.58% or 145):

$$f(L_t|S_t) = z_{S_t}\delta_{L_t}(0) + (1 - z_{S_t})\text{Gamma}(L_t; \mu_{S_t}, \sigma_{S_t}) \tag{2.4.1}$$

where z_{S_t} represents the probability of individuals being stationary given their corresponding state at time t, with $\delta_{L_t}(0)$ being a Dirac measure at step-length 0, and μ_{S_t} and σ_{S_t} denote the mean and standard deviation of the Gamma distribution governing the step length, conditional on state.

We model the turning angle between time points $t-2$ and t, A_t, using a vonMises distribution with location and concentration deviation parameters m_{S_t} and k_{S_t} , respectively

$$f(A_t|S_t) = \text{vonMises}(A_t; m_{S_t}, k_{S_t}). \tag{2.4.2}$$

Therefore, the observation at time t, given state at time t, is modelled as

$$f(O_t|S_t) = f(L_t|S_t)f(A_t|S_t). \tag{2.4.3}$$

We choose to set a repulsive prior on the mean step length $\mu_1, \mu_2, ..., \mu_N$, in terms of interpretability since we want to distinguish between different types of length movements. If our goal was to distinguish between directed and undirected movements we could have alternatively applied repulsion to the parameters of the angle distribution.

$$\underline{\mu} = (\mu_1, \mu_2, ..., \mu_N) \sim \text{StraussProcess}(\mu_1, \mu_2, ..., \mu_N; \xi, a, d) \tag{2.4.4}$$

and for comparison purposes also present results considering an independent, prior distribution

$$\underline{\mu} = (\mu_1, \mu_2, ..., \mu_N) \sim \text{IndependentProcess}(\mu_1, \mu_2, ..., \mu_N; \xi) \tag{2.4.5}$$

as described in Section 2.2.2. Details on Equations (2.4.4) and (2.4.5) are given in Section A.5.1 (page 99) in Appendix A.

We also place the following prior distributions on the remaining model parameters (with more details on the prior distribution choices and inference given in Section A.5 (page 99) of Appendix A.)

$$
\begin{aligned}
N &\sim \text{Uniform}\{1,2,...,N_{max}\} \\
\pi = (\pi_1,\pi_2,...,\pi_N)|N &\sim \text{Dirichlet}(a_1^\pi, a_2^\pi, ..., a_N^\pi) \\
P_{i\cdot} = (P_{i,1},P_{i,2},...,P_{i,N})|N &\sim \text{Dirichlet}(a_1^P, a_2^P, ..., a_N^P), \;\; i = 1,2,...,N \\
z_i &\sim \text{Beta}(a^z, b^z), \;\; i = 1,2,...,N \\
k_i &\sim \text{Uniform}(a^k, b^k), \;\; i = 1,2,...,N \\
m_i &\sim \text{Uniform}(a^m, b^m), \;\; i = 1,2,...,N \\
\sigma_i &\sim \text{Uniform}(a^\sigma, b^\sigma), \;\; i = 1,2,...,N
\end{aligned}
$$

Hence, the model parameters are $(N, \{\mu_i\}_{i=1}^N, \{\pi_i\}_{i=1}^N, \{P_{i\cdot}\}_{i=1}^N, \{z_i\}_{i=1}^N, \{k_i\}_{i=1}^N, \{m_i\}_{i=1}^N, \{\sigma_i\}_{i=1}^N)$. We run a RJMCMC algorithm for 300,000 iterations with 50,000 burn-in iterations, imposing the ordering constraint $\mu_i \leq \mu_{i+1}$ for $i = 1,2,...,N$. We fit the model with the repulsive prior of Equation (2.4.4) and the independent prior of Equation (2.4.5) and compare our results to those obtained by Pohle et al. (2017). The penalty parameter a and threshold d were chosen as described in Section 2.2, with $a = \exp(-n^*_{2.5})$ and $n^*_{2.5} = 627$, with results presented in Figure 2.2, or $a = \exp(-n^*_5)$ with $n^*_5 = 1255$, with results presented in Figure 2 of Section A.5.3 (page 106) in Appendix A. To select which value of the penalty a is the most appropriate, we computed the Bayes factor between the two models using the two different settings. The log Bayes Factor is 1077 in favour of $a = \exp(-n^*_{2.5})$. Therefore, we conclude that the model with $a = \exp(-n^*_{2.5})$ better supports the data and the value $a = \exp(-n^*_5)$ overpenalizes the number of states. Moreover, the value $a = \exp(-n^*_5)$ tends to point towards two states, which also contradict the findings of Pohle et al. (2017).

We assess the convergence of our RJMCMC algorithm considering diagnostics such as Geweke's diagnostic (Geweke, 1991), effective sample size and Monte Carlo error of the parameters. The results are displayed in Table A.4 and FiguresA.8-A.16. The convergence analysis indicates no issues, as the

trace plots indicate good mixing, and the Monte Carlo error is sufficiently small, ensuring confidence in our posterior estimates. The algorithm took 2 days to run on a computer with CPU 11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80GHz.

The repulsive prior of Equation (2.4.4) leads to a posterior mode of four states, with posterior distribution on the number of explored states $p(2) = 0.009$, $p(3) = 0.211$, $p(4) = 0.480$, $p(5) = 0.295$, $p(6) = 0.004$, $p(7) = 0.001$ with $\sum_{i=2}^{8} p(i) = 1$, with maximum number of states considered equal to 80. Additionally, the independent prior of Equation (2.4.5) leads to a posterior mode of seven states. Pohle et al. (2017) considered models with up to five states and selected the model with four states according to the integrated completed likelihood (ICL Biernacki et al., 2000) criterion. However, we note that the model with seven states actually leads to a smaller ICL (see Table A.3 in Section A.5.3 (page 106) of Appendix A), agreeing with our results in the case of a independent prior.

We plot the resulting distributions of step length and angle of the last time point since we are interested in the long-term behaviour of our estimates, conditional on four states, from our model with a repulsive prior and those obtained by Pohle et al. (2017) in Figure 2.2. The corresponding distribution results for the independent prior are given in Section A.5.3 of Appendix A.

Similarly to Pohle et al. (2017), we identify four types of step length, corresponding to hardly any movement (state 1), small movement (state 2), moving (state 3), and traveling (state 4). Additionally, states 3 and 4 have a much more directed movement compared to states 1 and 2 as observed from Figure 2.2 and discussed in Pohle et al. (2017). In contrast, when we consider an independent prior on the mean step length parameters $\mu_1, \mu_2, ..., \mu_N$, we observe common issues with models of this type, namely state distributions that are almost completely overlapping and state distributions that are assigned very small weights (Figure A.6 in Section A.5.3 (page 106) of Appendix A). Overall, from our analysis alongside with the results displayed in Figure 2.2, we observe that our proposed repulsive model actually identifies the model chosen subjectively from Pohle et al. (2017). However, they should have chosen a model with more states based on the ICL criterion (see Table A.3 in Section A.5.3 (page 106) of Appendix A) if they had consider a larger number of states which introduces some unjustified subjectivity. Hence our model can actually work objectively and identify a good number of states in terms of interpretability.

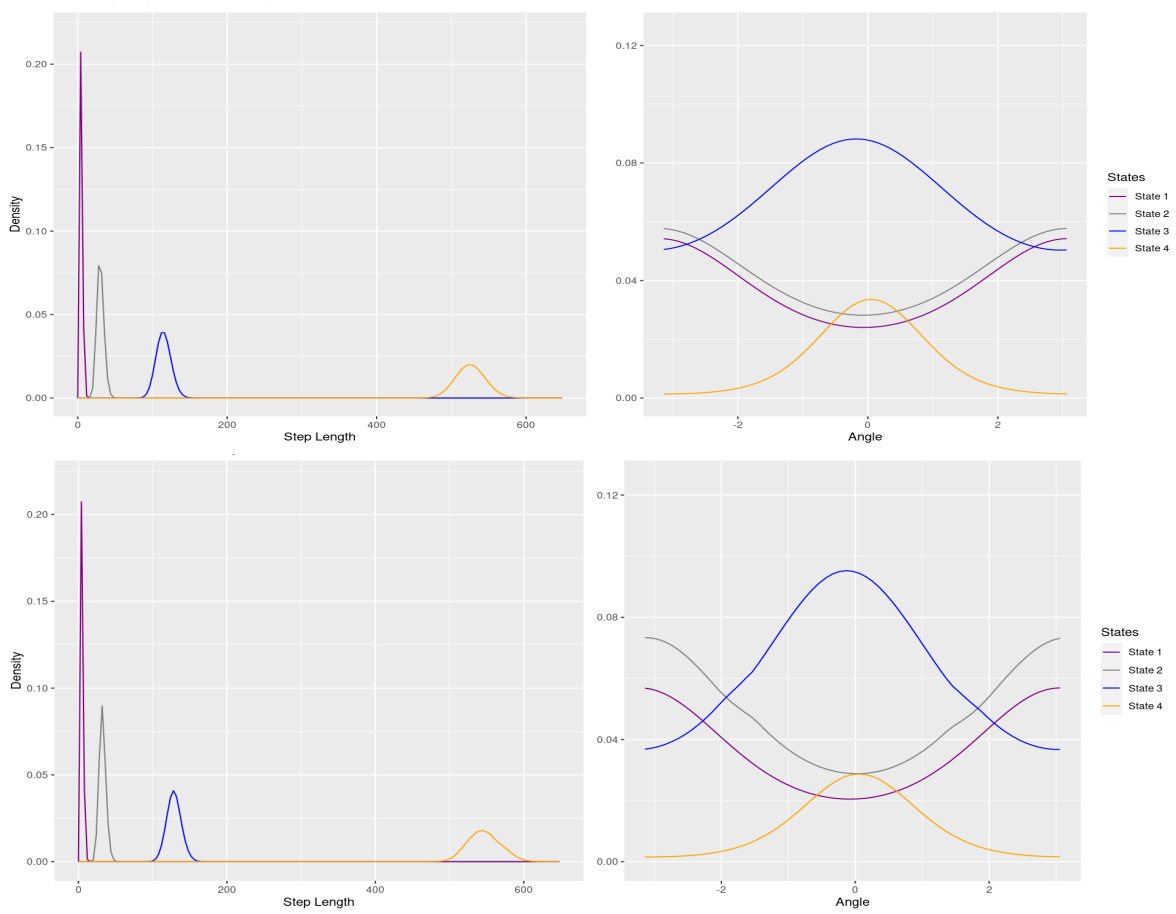

Fig. 2.2 The first row illustrates the distribution of step length (left) and angle (right) for the last time point as inferred by Pohle et al. (2017). The second row illustrates the posterior distribution of the step length (left) and angle (right) as inferred by our model, conditional on the posterior mode of four states, with a repulsive prior distribution.

2.5 Case study 2 : Cape gannet acoustic data

We consider data on Cape gannets (Morus capensis), which is an endangered seabird endemic to southern Africa. The data comprise of 3078.1 seconds of acoustic time points for a single chick-rearing Cape gannet in Algoa Bay, South Africa, using an animal-borne device, analyzed in Thiebault et al. (2021). The raw acoustic data are displayed in Figure A.17 in Appendix Section A.6 (page 110). The data were recorded at 22.05kHz sampling frequency and were pre-processed by downsampling the audio at 12 kHz and with a high-pass filter above 10 Hz before being segmented into 2179 intervals of 1.4 seconds (Thiebault et al., 2021). For each time segment of length 1.4 seconds we extracted 12 acoustic features based on the Mel-frequency cepstral coefficients with n measurements each (Cheng et al., 2010), which is standard practise in acoustic data analysis (Chalmers et al., 2021; Cheng et al., 2010; Noda et al., 2019; Ramirez et al., 2018). However, these 12 features are correlated with each other, and so we employ principal component analysis (PCA) to obtain a set of uncorrelated components as model inputs, instead of modelling the 12 features directly, as described in Trang et al. (2014). We consider the first two principal components (2-PC) that explain 70% of the variability of the original Mel-frequency cepstral coefficients.

We model the 2-PC at time t, $\underline{E_{t,c}} = (E_{t,c,1}, E_{t,c,2})$, with $c = 1, 2, ..., C$ the index of Mel-frequency cepstral coefficients measurements. using a Multivariate Normal distribution:

$$\underline{E_{t,c}} \sim \text{Normal}_2(\underline{\mu_{S_t}}, \Sigma_{S_t})$$

where $\underline{\mu_{S_t}}$ corresponds to the mean vector of the 2-PC for the latent state S_t and similarly Σ_{S_t} is the covariance matrix for the 2-PC under the latent state S_t.

We considered a repulsive prior on the mean parameters $\underline{\mu_1}, \ldots, \underline{\mu_N}$ and, for comparison, the independent prior described in Section 2.2.2. Details can be found in Section A.6.1 (page 106) of Appendix A.

The prior distributions placed on the rest of parameters of the model, such as initial probabilities, transition probabilities and covariance matrices are the following

$$
\begin{aligned}
N &\sim \text{Uniform}\{1,2,...,N_{max}\} \\
\pi = (\pi_1,\pi_2,...,\pi_N)|N &\sim \text{Dirichlet}(a_1^\pi, a_2^\pi, ..., a_N^\pi) \\
P_i = (P_{i,1},P_{i,2},...,P_{i,N})|N &\sim \text{Dirichlet}(a_1^P, a_2^P, ..., a_N^P), \;\; i = 1,2,...,N \\
\Sigma_i &\sim \text{Wishart}(n^\Sigma, \Sigma_0), \;\; i = 1,2,...,N
\end{aligned}
$$

Hence, the model parameters are $(N, \left\{\underline{\mu}_i\right\}_{i=1}^N, \{\pi_i\}_{i=1}^N, \{P_{i.}\}_{i=1}^N, \{\Sigma_i\}_{i=1}^N)$. The penalty parameter a was chosen as $a = \exp(-n_{2.5}^*)$ with $n_{2.5}^* = 54$, with results presented in Section 2.4, or $a = \exp(-n_5^*)$ with $n_5^* = 108$, with results presented in Section A.6.2 (page 114) of Appendix A, and $d = 21$. To select which value of the penalty is the most appropriate for the data, we used the Bayes factor between the two different settings for a. The value $a = \exp(-n_{2.5}^*)$ is supported with a log Bayes Factor of 20043. Hence, the model with $a = \exp(-n_{2.5}^*)$ is preferred from the data, compared to $a = \exp(-n_5^*)$ even though both models give similar results. We run a RJMCMC algorithm for 100,000 iterations with 10,000 burn-in iterations. Details of the prior distribution choices and inference are displayed in Sections A.6.1, A.6.2 (page 106, 114) of Appendix A. We use the post-processing technique described in Section 2.2.5 to label the states and the state allocation method described in Section 2.2.6 to obtain the posterior distribution of state allocation for each observation. We fit the model with the repulsive prior and the independent prior and show the posterior distributions in Section A.6.2 (page 114) in Appendix A.

For each model, we sample the allocation state of each observation at each time point as described in Section A.6.1 (page 106) of Appendix A. In Thiebault et al. (2021), state allocation is conducted manually with the assistance of experts by listening to the audio. The results of state allocation across time within our Bayesian framework are compared to the manual allocation of Thiebault et al. (2021) in Figure 2.3. We focus our interpretation on the model with three states, which is the posterior mode of the distribution on the number of states (the corresponding results for states such as two and four, for the repulsive case are given in Section A.6.2 (page 114) of Appendix A).

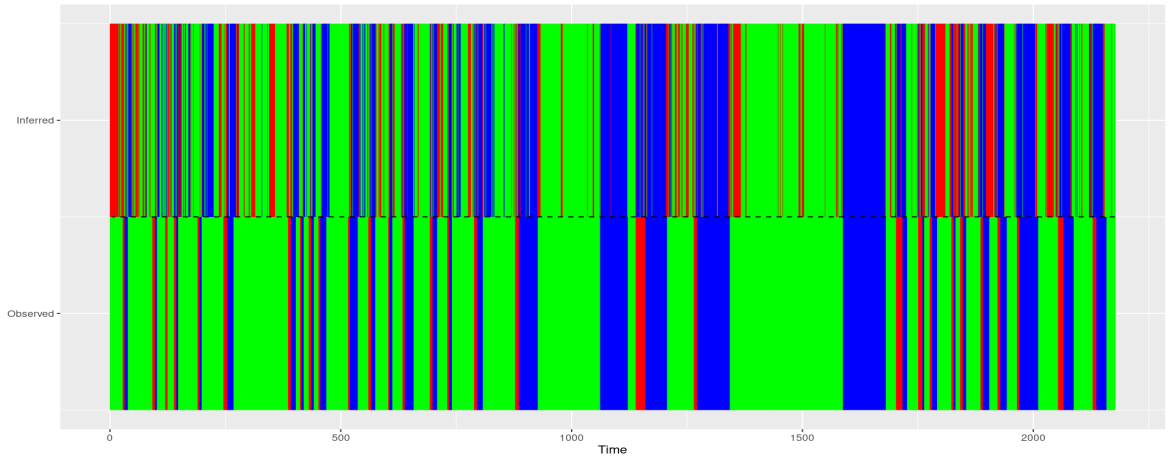

Fig. 2.3 Comparison of the posterior classification of our model (top half) with the manual classification of Thiebault et al. (2021) (bottom half). Based on the manual classification of Thiebault et al. (2021), the states are: floating on water (blue), flying (green) and diving (red).

Thiebault et al. (2021) identified three states, flying (green), floating on water (blue) and diving (red), as indicated in Figure 2.3. There is general agreement in state allocation between our model, which is completely unsupervised, and the manual allocation by Thiebault et al. (2021), in particular for the most common states of water (blue) and flying (green). However, our model classifies more time points to the third (red) state than the manual classification. According to expert knowledge and based on the recording, the time points allocated to this third state correspond to sounds made mainly after the take-off of the species with vigorous flapping, and general alterations in flying behavior, such as changes in flying direction or speed, coupled with variations in wind speed. Hence, in our three state model the third state corresponds to diving, alongside nuisance acoustic occurrences from the device or instantaneous changes in the individual's flying behaviour. This can potentially be mitigated if we use a second (or higher)-order HMM, which will help minimize this nuisance behaviour by discouraging states that are highly unlikely to occur between same state time points.

In Section A.6.2 (page 11) of Appendix A in Figure 4 we display the uncertainty of the classification inferred by our Bayesian framework, by plotting the posterior allocation probabilities for each state across time. All time points are in fairly high proportion of the time ($\sim 25\%$) allocated to states other than their modal state, indicating that state allocation has a high degree of uncertainty in this case, which is expected given the noisy and multivariate nature of the data and the very short time span of 1.4 seconds between time points. Nevertheless, the model successfully manages to allocate

the majority of points to a state that agrees with expert knowledge, demonstrating the potential of the approach even when modelling noisy and multidimensional data.

In Figure 2.4 we display the observations with points coloured according to their modal state allocation on the domain defined by the 2-PC. The first PC is dominated by the 1st Mel-frequency cepstral coefficient, which has a negative coefficient, whereas the second PC is a contrast between a weighted average of 5th, 8th, 10th and 11th Mel-frequency cepstral coefficients and 2nd Mel-frequency cepstral coefficient.

Based on Figure 2.4, flying (green) is characterized by large scores for the first PC, whereas floating on the water (blue) is characterized by small scores. Diving with the instantaneous nuisance acoustics (red) is associated with increasing PC2 scores, widening the range of PC1 scores, albeit always remaining mid-range. This suggests that when there is no strong support for allocating a time point to flying or floating on the water, then it is allocated to the third state.

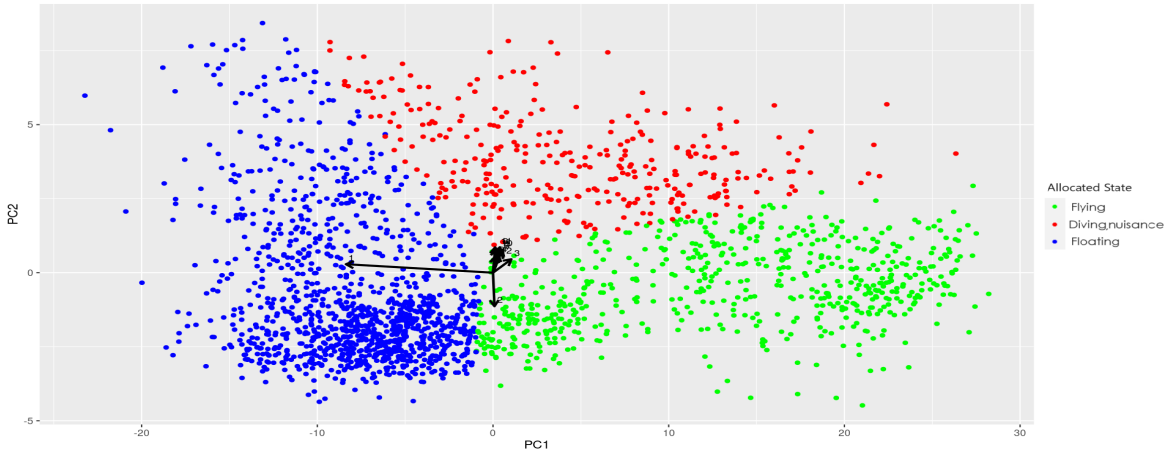

Fig. 2.4 Biplot, with observations coloured according to their modal state allocation, in the case of three states, plotted on the domain of the first two PC. Based on the manual classification of Thiebault et al. (2021) blue corresponds to floating on water, green to flying and red to diving.

Additionally, we assess the convergence of our RJMCMC algorithm considering diagnostics such as Geweke's diagnostic (Geweke, 1991), effective sample size and Monte Carlo error of the parameters. The results are displayed in Table A.5 and FiguresA.25-A.31. The convergence analysis indicates no issues, as the trace plots indicate good mixing, and the Monte Carlo error is sufficiently small, ensuring confidence in our posterior estimates. The algorithm took 2.5 days to run on a computer with CPU 11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80GHz.

Finally, when the independent prior is used, the posterior distribution of the number of states is very diffuse, with 2, 3, 4, ..., 42 states almost equally supported, with the results given in Figure 3, Section A.6.2 (page 114) of Appendix A, suggesting clear overfitting, particularly evident in this case because of the noisy nature of acoustic data.

2.6 Conclusion

In this chapter, we developed a new modeling framework for inferring the number of states and the corresponding distribution parameters in HMMs. In particular, we placed a repulsive prior on the location parameters of the HMM distributions in order to penalize small differences between the state parameters, and we treated the number of latent states as random and inferred it directly from the model without making any subjective decisions. We accomplished this using an RJMCMC algorithm, which samples the entire Bayesian model space.

We demonstrated the model using two interesting and challenging types of ecological applications using GPS and acoustic data. The results demonstrated the ability of our framework to yield parsimonious models with good state allocation ability in a completely unsupervised modelling framework. The case studies showcase the effectiveness and practicality of our framework, with the repulsive prior penalizing the number of underlying states, leading to simpler models, while effectively exploring the model sample space. Additionally, we conducted an extensive simulation study, which demonstrated that, when the generating model is fitted to the data, the repulsive prior model and the independent prior model yield practically indistinguishable results in terms of chosen number of states, density estimation and state allocation, with a small benefit in some cases when using the repulsive prior. Future work could explore simulation scenarios where the fitted model is only an approximation or simplification of the data-generating process, as is typically the case for real data.

The repulsive prior distribution is a function of a threshold and penalty parameter, which are not being inferred in the model and instead they have to be subjectively chosen. Our chosen approach, based on the work by Beraha et al. (2022), builds on a systematic way of choosing the penalty and threshold based on the minimum cluster size expected. Alternatively, different repulsive models that avoid the use of threshold and penalty parameters can be used, as discussed in Petralia et al. (2012).

In this case, we only placed repulsive prior distributions on one set of state parameters, the mean step length and mean vectors of the PC. Potential avenues for future research would be to apply joint repulsion priors on different parameters, such as step length and angle, or apply repulsion solely on the variance matrix for the acoustic coefficient. Such research might lead to interesting combinations of parameter components for the HMMs, giving a finer resolution of state dynamics.

The approach of fitting HMMs to ecological data within a dynamic mixture modelling framework with a repulsive prior on the latent number of components provide a valuable new point of view for this widely used class of models, which can be applied to a variety of disciplines such as in finance, biology, social science, medicine and ecology among others making it generic framework for statisticians and practitioners to use on their corresponding disciplines. Overall, we can get a better insight on the unknown underlying behaviour of species and infer more accurately how their behaviour, such as migration patterns, changes in response to changes in the environment.

Chapter 3

A Pólya Tree modelling framework for batch-mark data

Abstract

Wildlife population surveys typically consist of multiple sampling occasions, where individuals are followed over time, enabling estimation of population size and, in open populations, of entry and exit patterns. Batch-mark (BM) surveys, where newly sampled individuals are given the same marking, unique for each sampling occasion but not for each individual, provide the only viable monitoring tool for many species of amphibians, birds and fish. Modelling BM data for open populations has proven more challenging than modelling data where individuals are uniquely marked, and approaches proposed in the literature thus far rely on approximate inference or do not scale well with the number of individuals, and do not readily extend to the joint modelling of different observation processes often employed in practice. In this chapter, we propose a novel approach for modelling BM data, by defining a bivariate grid for modelling the latent entry and exit patterns, as well as population size. We employ the Bayesian nonparametric Pólya Tree (PT) prior for defining a model on the grid cells, which enables exact and highly efficient Bayesian inference on the number of individuals in each cell, and hence of the entry/exit pattern. Our approach scales with the number of sampling occasions, instead of the number of individuals, and allows us to easily write the likelihood function for BM data under different observation processes. We demonstrate our new Pólya Tree prior Batch Mark (PTBM) approach using extensive simulations and two case studies, comparing its performance with two recently proposed approaches.

3.1 Introduction

Ecological monitoring is of paramount importance in safeguarding our natural ecosystems. By studying wildlife populations and accurately estimating essential demographic parameters, such as population size, birth/death rates or phenological patterns, we can gain invaluable insights into their

dynamics. One approach that can be employed for population monitoring is batch-marking (BM), which involves repeated sampling occasions when individuals are physically caught, with newly caught individuals marked using a sampling occasion-specific mark, which is a mark that is typically different among sampling occasions but not among individuals. BM surveys are particularly useful in cases where individual marking cannot be employed, such as with species that are highly abundant or small in size, for example fish or insects, and have been extensively utilized providing valuable data for population monitoring (Cowen et al., 2017; Davidson et al., 2019; Doll et al., 2021; Rosser et al., 2022; Vavassori et al., 2019).

BM data on each sampling occasion consist of the number of individuals caught that are unmarked and the number of individuals caught that were first marked on previous sampling occasions. Therefore, as opposed to standard capture-recapture studies (King and McCrea, 2019; McCrea and Morgan, 2014; Seber and Schofield, 2019), where individuals are uniquely marked, it is not known how many times and on which sampling occasions each individual is re-caught. This aggregated nature of BM data means that likelihood-based inference is more challenging, and the literature on appropriate models for BM data is limited. To discuss existing approaches we need to introduce the concept of individual presence histories and capture histories. Both are vectors of length equal to the number of sampling occasions. The former indicates when the corresponding individual is present at the surveyed site, while the latter indicates when the individual has been caught. Individual presence histories are always latent in ecological data, while individual capture histories are observed in capture-recapture data but latent in BM data. Current approaches for modelling BM data include the work by Huggins et al. (2010), who derived estimating equations and closed-form solutions for survival and capture probabilities, along with abundance estimates using a Horvitz-Thompson-type estimator. Another important contribution came from Cowen et al. (2014), who developed a tractable likelihood function for marked individuals only, but with an associated computational cost that increases substantially with more sampling occasions, since the calculation of the likelihood involves nested summations of all possible individual latent presence histories. The Cowen et al. (2014) work was further extended by Cowen et al. (2017) who employed Hidden Markov Models (HMM) to model BM data for both marked and unmarked individuals, but this HMM approach does not scale well with the number of individuals, since it involves high dimensional state-transition and state-dependent probability

matrices. More recently, Zhang et al. (2023) introduced an innovative approach to approximating the likelihood function for BM data under the robust design (RD, Kendall and Pollock, 1992). When the RD is employed, it is assumed that the population is open between primary periods, e.g. months, but closed within secondary periods, e.g. days within a month, so that individuals can enter/exit between primary periods but not between secondary periods. Zhang et al. (2023) proposed an approach that relies on the saddlepoint approximation (SPA Butler, 2007) to the likelihood function, which requires reconstructing the latent capture history as well as the latent presence history for each individual, so the computational burden increases considerably with the number of sampling occasions.

In this chapter, we present a novel Bayesian nonparametric approach utilizing the Pólya Tree (PT) prior for BM data analysis. This approach, which builds on the work by Diana et al. (2023) for count and ring-recovery data, offers numerous benefits. The fundamental idea is that we build a lower-right triangular grid with $\frac{(K+1)(K+2)}{2}$ latent cells, as shown in Figure 3.1, where K is the number of sampling occasions. The grid cells represent the latent number of individuals with a specific combination of entry and exit intervals, where the intervals are defined by pairs of consecutive sampling occasions, so that $n_{i,j}$ is the number of individuals with entry between sampling occasions i and $i+1$ and exit between sampling occasions j and $j+1$. We note that entry and exit can correspond to birth/death, arrival/departure, or any other process that introduces/removes individuals from the population. By using our proposed grid approach and corresponding PT prior, as we discuss below, we naturally account for the aggregated nature of BM data because we do not need to infer, impute or marginalise over latent individual presence histories or capture histories, and instead only need to infer the latent cells of the grid. Consequently, within this framework, we overcome previous challenges related to BM model inference and establish a tractable likelihood function and a corresponding efficient model-fitting algorithm for standard BM data, as well as BM data collected under different sampling designs, as we discuss in the two case studies of the chapter.

Inference of the cells $n_{i,j}$ is efficient and flexible within a Bayesian framework based on the PT prior, which allows us to define and infer the grid probabilities. By relying on the PT prior, we can build a model directly on the distributions of entry and exit patterns, with minimal parametric assumptions on the shape of these distributions. Additionally, the replicate PT framework, first

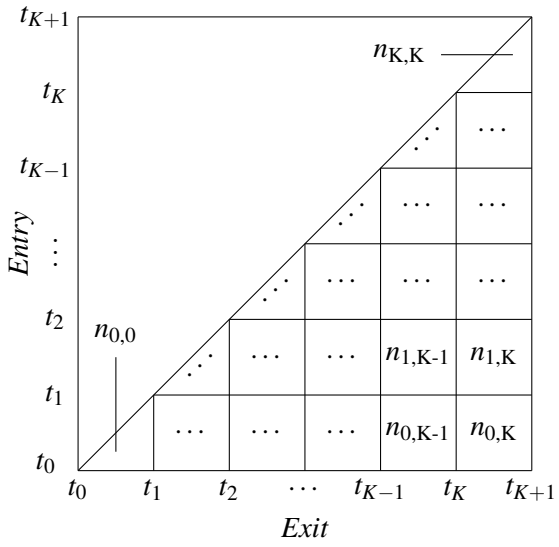

Fig. 3.1 Entry and exit sample space, for K sampling occasions, taking place at times $t_1, t_2, ..., t_K$ with convention that $t_0 = -\infty, t_{K+1} = \infty$. The latent number of individuals in cell (i, j), that is with entry between the ith and $(i+1)$th and exit between the jth and $(j+1)$th sampling occasions are denoted by $n_{i,j}$, with $i = 0, 1, ..., K$ and $j = i, ..., K$.

introduced in Diana et al. (2023), allows us to impose constraints on the entry and exit processes, leading to more parsimonious models, as we discuss in Section 3.3.

We present results for two case studies previously analysed in the literature. Both consider open populations and fit models of Jolly Seber (JS Jolly, 1965; Seber, 1965) type, which are models that allow direct estimation of population size while accounting for entry and exit from the population. The data from the two case studies result from different observation processes. The first case study considers data on weather loaches (*Misgurnus anguillicaudatus*), which are freshwater fish, monitored in South-Eastern Australia. In this case, during the BM survey, individuals caught as unmarked are either marked and returned to the population in the standard BM approach, or are removed from the population, akin to removal sampling (Matechou et al., 2016). We show how our framework can easily be adapted to model the two processes jointly, which was not possible using the approach proposed by Cowen et al. (2017). A corresponding simulation study demonstrates the resulting bias when the removal process is ignored. The second case study considers data on golden mantella frogs (*Mantella aurantiaca*), collected in Central Madagascar under the RD across six primary periods. In this case, we demonstrate how our framework can easily account for BM data collected under a RD

sampling, using the exact likelihood, and compare our results to those obtained by Zhang et al. (2023). A corresponding simulation study explores issues around allocation of effort within a RD framework for BM surveys.

The chapter is organized as follows: In Section 3.2 we introduce our modelling approach of the latent entry/exit pattern and population size from BM data, and in Section 3.3 we describe the PT prior, which provides the foundation of our modelling framework. We present simulation and real data results for each of the two case studies in Sections 3.4 and 3.5. The chapter concludes with a discussion of the findings and outlines potential avenues for future research in Section 3.6.

3.2 Model for batch-mark data

First, we introduce the notation and modelling of standard BM data, with K sampling occasions, taking place at times $t_1, ..., t_K$, with $t_0 = -\infty, t_{K+1} = \infty$, for convenience. Additional notation for the two different observation processes employed in the case studies is introduced in the corresponding sections. In what follows, we refer to individuals in (grid) cell (i, j) as those individuals that entered between sampling occasions i and $(i+1)$ and exited between sampling occasions j and $(j+1)$.

Data

u_t : observed number of unmarked individuals caught on sampling occasion t.

$m_{k,t}$: observed number of individuals that were marked on sampling occasion k and were recaptured on sampling occasion t.

Parameters

N : population size, with $\sum_{i=0}^{K}\sum_{j=i}^{K} n_{i,j} = N$ (see Figure 3.1). We place on N a Poisson prior distribution with mean ω, $N \sim \text{Poisson}(\omega)$.

$w_{i,j}$: probability of an individual belonging to cell (i, j) of matrix n.

p_t : probability of capturing an individual that is present on sampling occasion t.

Latent

$n : (K+1) \times (K+1)$ matrix where $n_{i,j}$ corresponds to the latent number of individuals in cell (i, j). The latent cell counts $n_{i,j}$, conditional on N are modelled as

$$n|N \sim \text{Multinomial}(N, \{w_{i,j}\}_{i=0,\ldots,K, j=i,\ldots,K})$$

so that, since $N \sim \text{Poisson}(\omega)$, the unconditional distribution of the counts in the cells of matrix n is given by

$$n_{i,j} \sim \text{Poisson}(\omega \times w_{i,j}), i = 0, \ldots K, j = i, \ldots K \tag{3.2.1}$$

$U^k : (K+1) \times (K+1)$ matrix where $U^k_{i,j}$ corresponds to the latent number of individuals in cell (i, j) of matrix n, that are present and unmarked on sampling occasion k.

$M^k : (K+1) \times (K+1)$ matrix where $M^k_{i,j}$ corresponds to the number of individuals in cell (i, j) of matrix n, that were caught as unmarked (and subsequently marked) on sampling occasion k. Each of the latent $U^k_{i,j}$ individuals can be caught with probability p_k, independent of other individuals, so that

$$M^k_{i,j} \sim \text{Binomial}(U^k_{i,j}, p_k).$$

We note that for $t \leq i$ or $t > j$, $U^t_{i,j} = 0$ (the latent number of individuals in cell (i, j) of matrix n, that are unmarked on sampling occasion t) by definition, since individuals from cell i, j are not present on sampling occasion t in this case. Otherwise, for $i < t \leq j$, $U^t_{i,j}$ is equal to the difference between the total latent number of individuals in that cell and the latent number of individuals in that cell that were marked before sampling occasion t, so that $U^t_{i,j} = n_{i,j} - \sum_{k=1}^{t-1} M^k_{i,j}$, with $i < t \leq j$. Note that when $t = 1$ there are no marked individuals yet and hence $U^1_{i,j} = n_{i,j}$ for $i < 1 \leq j$, and 0 otherwise.

Therefore, conditional on the latent matrices M^k, the observed number of unmarked individuals caught on sampling occasion t, u_t, is

$$u_t = \sum_{i=0}^{t-1} \sum_{j=t}^{K} M^t_{i,j}.$$

The number of individuals marked on sampling occasion k that are still present on sampling occasion t is equal to $\sum_{i=0}^{k-1}\sum_{j=t}^{K} M_{i,j}^{k}$ for $i<k<t\leq j$, that is the sum of $M_{i,j}^{k}$ terms for individuals that entered before k and have not yet exited at t and each of these individuals is recaptured on sampling occasion t with probability p_t, independent of other individuals, so that the observed number of individuals marked on sampling occasion k and recaptured on sampling occasion t is distributed as

$$m_{k,t}\sim \text{Binomial}\left(\sum_{i=0}^{k-1}\sum_{j=t}^{K} M_{i,j}^{k}, p_t\right),\ i<k<t\leq j$$

The latent model for the standard BM data is

• latent number of individuals in cell (i,j) of matrix n	$n_{i,j}\sim \text{Poisson}(\omega\times w_{i,j}),\ i=0,...,K,\ j=i,...,K$
• latent number of individuals in cell (i,j) of matrix n that are unmarked on sampling occasion t for $i<t\leq j$	$U_{i,j}^{t}=\begin{cases} n_{i,j}, & t=1,\\ n_{i,j}-\sum_{k=1}^{t-1} M_{i,j}^{k}, & t>1,\end{cases}$
• latent number of individuals in cell (i,j) of matrix n caught as unmarked and subsequently marked on sampling occasion t	$M_{i,j}^{t}\sim \text{Binomial}(U_{i,j}^{t}, p_t),$
• observed number of unmarked individuals caught on occasion t	$u_t=\sum_{i=0}^{t-1}\sum_{j=t}^{K} M_{i,j}^{t},\ t\geq 1$
• observed number of individuals that were marked on sampling occasion k and were recaptured on sampling occasion t, for $i<k<t\leq j$	$m_{k,t}\sim \text{Binomial}\left(\sum_{i=0}^{k-1}\sum_{j=t}^{K} M_{i,j}^{k}, p_t\right),$

(3.2.2)

We note that, in our model, N is not the equivalent of the super-population of individuals, N^S, that became available for capture at least once (Schwarz and Arnason, 1996) but instead also accounts for individuals that entered and exited without ever becoming available for capture (Matechou and Caron, 2017) that is individuals in cells with $i = j$. Inference on these individuals is possible within this framework thanks to the PT prior, as also discussed in Diana et al. (2019). However, inference on the super-population size is also readily available from the n matrix, since $N^S = N - \sum_{i=0}^{K} n_{i,i}$, i.e. by excluding individuals that never became available for capture, and in the case studies we report both quantities.

A simulation study for this standard BM model is presented in Appendix B, Section B.2 (page 127).

3.3 Pólya Tree Prior

Inferring the grid matrix n relies on inferring the grid probabilities $w_{i,j}$, through the application of the PT prior. The PT prior requires partitioning the sample space and specifying a sequence of positive numbers to assign probabilities to each set in these partitions.

In our case, following the methodology outlined in Section 3.2, we need to partition the entry and exit space, the latter conditional upon entry. This partitioning is achieved by first splitting the sample space into the sets $B_i = (t_i, t_{i+1}] \times (t_i, t_{K+1})$ for $i = 0, 1, 2, ..., K$, corresponding to the individuals entering in the interval between the ith and $(i+1)$th sampling occasions and exiting after entry. The probability of an individual belonging to B_i is V_i, where $(V_0, V_1, ..., V_K) \sim \text{Dirichlet}(\alpha_0, \alpha_1, ..., \alpha_K)$. The Dirichlet distribution on the Vs can be centred on any distribution G_0 that expresses our prior expectation for the entry pattern, by assuming $\mathbb{E}[V_i] = G_0(B_i)$, which can be achieved by defining the sequence of positive numbers such that $\alpha_i \propto G_0(B_i)$. If we do not have prior knowledge about the entry pattern we can choose a uniform distribution for G_0, which corresponds to setting $\alpha_0 = \alpha_1 = ... = \alpha_K = 1$. Further examples about partitions of the entry/exit space and PT prior centring can be found in Diana et al. (2023).

Next, we split each entry set, $B_i = (t_i, t_{i+1}) \times (t_i, t_{K+1})$ into the sets $B_{i,j} = (t_i, t_{i+1}) \times (t_j, t_{j+1}),\ j = i, ..., K$, where $(t_i, t_{i+1}) \times (t_j, t_{j+1})$ corresponds to the individuals entering between the ith and $(i+1)$th

sampling occasion and exiting between the jth and $(j+1)$th sampling occasion. The probability of an individual belonging to $B_{i,j}$ conditional on being in B_i is $V_{i,j}$, where $V_{i,i}, V_{i,i+1}, ..., V_{i,K} \sim$ Dirichlet$(\alpha_{i,i}, \alpha_{i,i+1}, ..., \alpha_{i,K})$, $i = 0, ..., K$. We employ the replicate PT as described in Diana et al. (2023), which imposes a structure on the exit probabilities, and hence builds a more parsimonious PT prior with a smaller number of parameters. In our case, we use a time-dependent constraint for the exit probabilities and assume that $V_{0,j} = V_{1,j} = ... = V_{j,j},\ j = 1, ..., K$, i.e. we replicate the exit probabilities across the entry intervals, so that the probability of exit in a given interval is the same for all individuals, regardless of entry.

Finally, the PT prior distribution of the probabilities $w_{i,j}$ is defined as

$$\left\{w_{i,j}\right\} = \left\{V_i V_{i,j}\right\} \sim PT(\{\alpha_i\}, \left\{\alpha_{i,j}\right\}),\ i = 0, 1, 2, ..., K \quad j = i, ..., K \tag{3.3.1}$$

3.3.1 Inference

We describe our employed Markov Chain Monte Carlo (MCMC), with corresponding conditional distributions, where appropriate, in Appendix B, Section B.1 (page 123). Briefly, the latent matrices $\left\{M^k\right\}_{k=1}^{K}$ and n are updated using a standard Metropolis Hastings (MH) random walk (Robert et al., 2004) whereas the $\left\{w_{i,j}\right\}_{i=0,j=i}^{K,K}$ parameters are updated using a Gibbs algorithm (Gelfand and Smith, 1990) since we exploit the conjugacy properties of the PT prior (Lavine, 1992). Parameters $\{p_t\}_{t=1}^{K}$ and ω are also updated with the use of a Gibbs sampler.

3.4 Case Study 1

BM data on weather-loch were collected across 11 sampling occasions, but in contrast to standard BM sampling, a proportion of unmarked individuals caught on each sampling occasion were removed from the population (Cowen et al., 2017). Here we demonstrate how our modelling framework introduced in Section 3.2 can naturally account for removals on captures in BM surveys. From Section 3.2, we make use of data $\{u_t\}_{t=1}^{K}, \{m_{k,t}\}_{k=1,t=k+1}^{K-1,K}$, parameters $N, \left\{w_{i,j}\right\}_{i=0,j=i}^{K,K}, \{p_t\}_{t=1}^{K}$, and latent $n, \left\{M^k\right\}_{k=1}^{K}$ and we introduce additional data, parameters and latent terms corresponding to removals per sampling occasion.

3.4.1 Data and Notation

First we introduce some case-study-specific notation.

Data

r_k : observed number of individuals that were removed from the population on sampling occasion k.

Parameters

l_k : probability of removing an unmarked individual caught on sampling occasion k.

Latent

Y^k : $(K+1)\times(K+1)$ matrix where $Y^k_{i,j}$ corresponds to the latent number of unmarked individuals in cell (i,j) of matrix n, caught on sampling occasion k. We note that this is not the same as $M^k_{i,j}$ introduced in Section 3.2 since not all newly caught individuals are marked in this case, as we discuss below. Similarly to the standard BM model described in Section 3.2,

$$Y^k_{i,j} \sim \text{Binomial}(U^k_{i,j}, p_k),\ i < k \leq j$$

R^k : $(K+1)\times(K+1)$ matrix where $R^k_{i,j}$ corresponds to the latent number of individuals in cell (i,j) of matrix n that were removed on sampling occasion k. Each of the $Y^k_{i,j}$ individuals caught as unmarked has the same probability, l_k, independent of other individuals, of being removed instead of being marked and returned to the population, so that

$$R^k_{i,j} \sim \text{Binomial}(Y^k_{i,j}, l_k)$$

We note that now the number of individuals in cell (i,j) of matrix n that are present and marked on sampling occasion t, for $i < t \leq j$, is $M^t_{i,j} = Y^t_{i,j} - R^t_{i,j}$, and 0 otherwise, and the latent number of individuals in cell (i,j) of matrix n that are available as unmarked on sampling occasion t is $U^t_{i,j} = n_{i,j} - \sum_{k=1}^{t-1} M^k_{i,j} - \sum_{k=1}^{t-1} R^k_{i,j},\ i < t \leq j$ since individuals in cell (i,j) of matrix n are no longer available as unmarked on sampling occasion t if they were previously marked or removed from the population.

Therefore it follows that, conditional on $R^k_{i,j}$, the observed number of individuals removed from the population on sampling occasion k are

$$r_k = \sum_{i=0}^{k-1} \sum_{j=k}^{K} R^k_{i,j}$$

3.4.2 Model

The model of Equation (3.2.2) is extended to account for the removal process, in addition to the standard BM process, as shown in Equation (3.4.1).

• latent number of individuals in cell (i,j) of matrix n	$n_{i,j} \sim \text{Poisson}(\omega \times w_{i,j}),\ i = 0,...,K,\ j = i,...,K$
• latent number of individuals in cell (i,j) of matrix n, that are unmarked on sampling occasion t for $i < t \leq j$	$U_{i,j}^t = \begin{cases} n_{i,j},\ t = 1, \\ n_{i,j} - \sum_{k=1}^{t-1} M_{i,j}^k - \sum_{k=1}^{t-1} R_{i,j}^k,\ t > 1, \end{cases}$
• latent number of individuals in cell (i,j) of matrix n caught as unmarked on sampling occasion t	$Y_{i,j}^t \sim \text{Binomial}(U_{i,j}^t, p_t)$
• observed number of unmarked individuals caught on sampling occasion t	$u_t = \sum_{i=0}^{t-1} \sum_{j=t}^{K} Y_{i,j}^t,\ t \geq 1$
• latent number of individuals in cell (i,j) of matrix n removed on sampling occasion t	$R_{i,j}^t \sim \text{Binomial}(Y_{i,j}^t, l_t),$
• observed number of individuals removed on sampling occasion t	$r_t = \sum_{i=0}^{t-1} \sum_{j=t}^{K} R_{i,j}^t\ ,\ t \geq 1$
• latent number of individuals in cell (i,j) of matrix n, marked on sampling occasion t	$M_{i,j}^t = Y_{i,j}^t - R_{i,j}^t$
• observed number of individuals that were marked on sampling occasion k and were re-captured on sampling occasion t for $i < k < t \leq j$	$m_{k,t} \sim \text{Binomial}\left(\sum_{i=0}^{k-1} \sum_{j=t}^{K} M_{i,j}^k, p_t\right)$

(3.4.1)

3.4.3 Case study 1: Inference

Inference for $\left\{M^k\right\}_{k=1}^{K}$, n, $\left\{w_{i,j}\right\}_{i=0,j=i}^{K,K}$, $\left\{p_t\right\}_{t=1}^{K}$ and ω is described in Section 3.2. Parameters l_k and latent matrix $\left\{R^k\right\}_{k=1}^{K}$, are updated using an MH random walk,as described in Appendix B, Section B.1.

3.4.4 Simulation study

We conducted a simulation study with the aim of exploring how ignoring removed individuals affects inference quality. For 100 replications, we simulated BM data for 11 sampling occasions, involving 6000 individuals and with removal probabilities of 0.05, 0.1, or 0.2. We used time varying entry/exit probabilities as well as capture probabilities. The chosen values are given in Appendix B, Section B.3, together with the prior distribution choices for all parameters. In each case, we ran an MCMC algorithm for 500,000 iterations, with a burn-in of 20,000 iterations.

We denote the true parameter value in each case by $\tilde{\theta}$ and the corresponding parameter sampled on the b-th iteration of replication m after burn-in by θ_b^m. We calculate and report the posterior bias across simulations, $\bar{B}_m$, as $\bar{B}_m = \frac{1}{B}\sum_{b=1}^{B}(\theta_b^m - \tilde{\theta}),\ m = 1,2,...,100$. In Figure 3.2 we display the posterior bias obtained by our model when removed individuals are accounted for (denoted by PTBM-R) and when they are not (denoted by PTBM). Inference for N is summarised in Figure 3.2 and for all other parameters in Appendix B, Section B.3 (page 129).

As expected, when the removed individuals are not accounted for, inference is biased, especially for parameters concerning the number of individuals in the survey such as N, U_t and N_t, which is defined as the number of individuals available on sampling occasion t (see in Appendix B, Section B.3 (page 129)). This bias becomes more severe as the proportion of unmarked individuals removed increases. The rest of the parameters, such as $V_i, V_{.,j}$ and p_t, exhibit less severe bias, especially when removal probability is low, since they are mainly informed by the marked individuals that are followed over time, and hence are not affected as much by the removal of unmarked individuals. Finally, as expected, when the correct model is fitted to the data, estimation is unbiased.

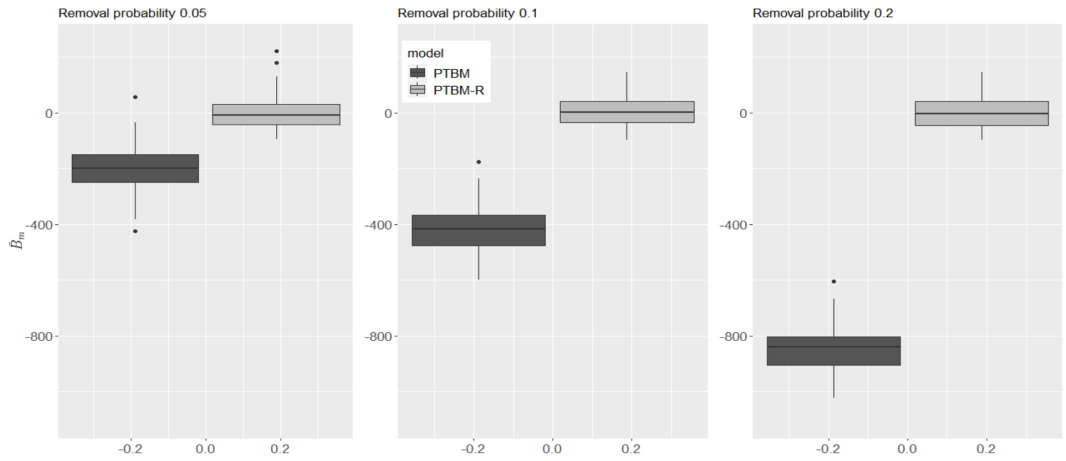

Fig. 3.2 Posterior bias of population size, N. Each column corresponds to a different removal probability. We compare the posterior bias across the models PTBM-R and PTBM, i.e. when we account for removals and when we do not.

3.4.5 Weather-loch data

We present results for this case study in Figures 3.3 and Table 3.1 and compare them to those obtained by Cowen et al. (2017). Following Cowen et al. (2017), we assume that entry probabilities and exit probabilities are constant, that is $V_i = \beta$, for all $i = 0, 1, ..., K$ and $V_{i,j} = 1 - \phi$ for all $j = i, ..., K$. The prior distribution choices for our parameters are described in Appendix B, Section B.4 (page 130). Hence, the model parameters are $(\omega, \beta, \phi, \{p_t\}_{t=1}^{T}, \{M^t\}_{t=1}^{T}, \{R^t\}_{t=1}^{T}, n)$.

The overall patterns in our findings agree with those in the simulation study. Namely, in Table 3.1, we observe that estimation of entry and exit probabilities for the PTBM-R, PTBM, and HMM models is, on average, aligned. Additionally, in Figure 3.3, we observe that, on average, our estimates for the number of unmarked individuals present on each sampling occasion, and for the population size are larger when accounting for removals compared to the other two models PTBM and HMM. Finally, the posterior median population and super-population for the PTBM-R model are 3280 with a 95% posterior credible interval of $(2892, 3771)$ and 2805 with a 95% posterior credible interval of $(2554, 3047)$, respectively, whereas for the PTBM model are 2198 with a 95% posterior credible interval of $(1958, 2419)$ and 2182 with a 95% posterior credible interval of $(2095, 2276)$ respectively. The super-population for the Cowen et al. (2017) model is estimated as 2242 (no interval was reported)

which is in alignment with out PTBM model, which is expected since neither model accounts for removals.

Figures 3.3 in this section and 8 in in Appendix B, Section B.4 (page 130) demonstrate that the number of available individuals increases and peaks on the 5th sampling occasion before gradually decreasing. This could potentially be interpreted as a result of seasonality. Notably, Huggins et al. (2010) mentioned that they estimated the smallest number of available individuals on the 7th sampling occasion, which they attributed to the winter season, with the population then increasing during the spring. In contrast, Cowen et al. (2017) stated that they estimated a decrease in the number of available individuals over time when using time-varying capture probabilities, as opposed to Huggins et al. (2010). Interestingly, our PTBM-R model manages to identify some seasonality, with population size peaking before the winter time and subsequently estimating a decrease in the population.

Additionally, we assess the convergence of our MCMC algorithm for the PTBM-R model by considering diagnostics such as Geweke's diagnostic (Geweke, 1991), effective sample size, Monte Carlo error of the parameters, and the $\hat{R}$ statistic (Vehtari et al., 2021), given that we ran four MCMC chains. The results are displayed in Table B.1 and Figures B.8, B.9 and B.10. The convergence analysis indicates no issues, as the $\hat{R}$ for all parameters is close to 1, and the Monte Carlo error is sufficiently small, ensuring confidence in our posterior estimates. The algorithm took 15 minutes to run on a computer with CPU 11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80GHz.

Finally, we conducted a goodness of fit assessment, by comparing the observed number of unmarked individuals $\{u_t\}_{t=1}^{K}$ caught on each sampling occasion with summaries of the corresponding data simulated from each model (see Appendix B, Section B.4 (page 130)). It is evident that the PTBM-R model, which better describes the data-generating process, outperforms the other two models in generalizing observed patterns in the data.

3.5 Case Study 2

Next, we consider the case study first published in Zhang et al. (2023) where sampling follows Pollock's robust design (RD Pollock, 1982). The sampling of golden mantella took place over 6 primary periods with $(3,3,3,4,4,4)$ secondary sampling occasions for each primary period. During

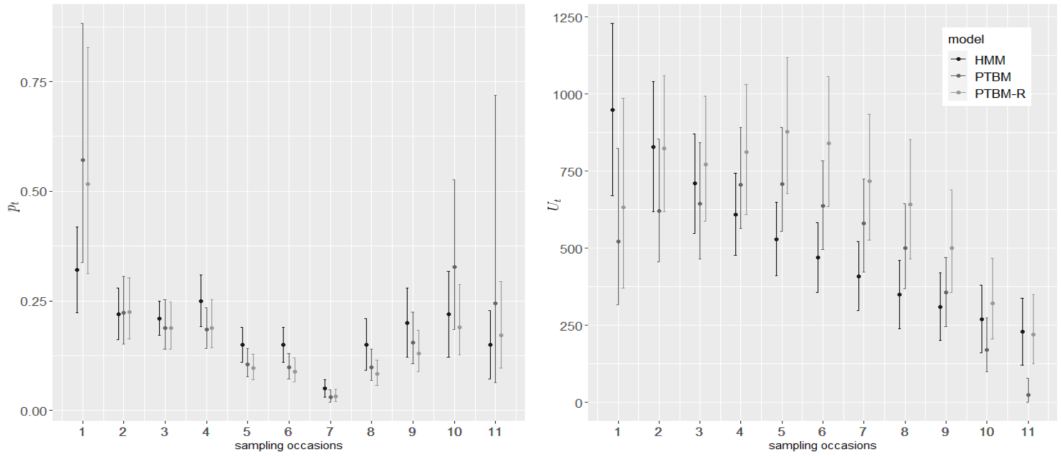

Fig. 3.3 Summaries of capture probabilities p_t, (a), and of the inferred number of unmarked individuals present on each sampling occasion U_t, $t = 1, 2, ..., K$, (b). The HMM model of Cowen et al. (2017) was fitted classically, so the summaries correspond to the maximum likelihood estimate and corresponding 95% confidence interval in each case, whereas our PTBM and PTBM-R models are fitted in a Bayesian framework, so the summaries correspond to the posterior mean and 95% posterior credible interval.

Table 3.1 Summaries of entry probability β, and exit probability $1-\phi$. In the HMM case, these correspond to the maximum likelihood estimate and corresponding 95% confidence interval, while in the PTBM and PTBM-R models, these correspond to the posterior median and corresponding 95% posterior credible interval.

	Parameter	
Model	β	ϕ
HMM	0.24 (0.18, 0.30)	0.63 (0.58,0.69)
PTBM	0.22 (0.14, 0.35)	0.61 (0.27,0.81)
PTBM-R	0.21 (0.14, 0.34)	0.63 (0.40,0.87)

each primary period, captured individuals receive a distinct mark and are then released so can be recaught on later secondary sampling occasions within the same or different primary periods.

3.5.1 Data and Notation

First we introduce some case-study-specific notation. In the case of the RD, there are K primary periods with T_k secondary sampling occasions within primary period k. The underlying latent process of entry and exit refers to intervals between primary periods, as described in Section 3.2, while the observation process has the same structure as in Equation (3.2.2), but in this case with multiple secondary sampling occasions per primary period, as we describe below.

Data

- $u_{k,t}$: observed number of unmarked individuals caught on secondary sampling occasion t of primary period k.
- $m_{k,\nu,t}$: observed number of individuals that were marked in primary period k and were recaptured on secondary sampling occasion t of primary period ν for $\nu \geq k$.

Parameters

$p_{k,t}$: probability of capturing an individual present on secondary sampling occasion t of primary period k.

Latent

- $U^{k,t}$: $(K+1)\times(K+1)$ matrix where $U^{k,t}_{i,j}$ corresponds to the latent number of unmarked individuals in cell (i,j) present on secondary sampling occasion t of primary period k for $i<k\leq j$.
- $M^{k,t}$: is a $(K+1)\times(K+1)$ matrix, where $M^{k,t}_{i,j}$ corresponds to the latent number of individuals in cell (i,j) caught as unmarked on secondary sampling occasion t of primary period k.

$$M^{k,t}_{i,j}\sim \text{Binomial}(U^{k,t}_{i,j},p_{k,t})$$

We note that now the latent number of individuals in cell (i,j) that are available as unmarked on secondary sampling occasion t of primary period k is $U^{t}_{i,j}=n_{i,j}-\left(\sum_{\nu=1}^{k-1}\sum_{\ell=1}^{T_\nu}M^{\nu,l}_{i,j}+\sum_{\ell=1}^{t-1}M^{k,l}_{i,j}\right)$ for $k\neq 1, t\geq 1$ and $U^{k,t}_{i,j}=n_{i,j}$ for $k=1,t=1$, where $\sum_{\nu=1}^{k-1}\sum_{\ell=1}^{T_\nu}M^{\nu,l}_{i,j}$ is the total number of individuals from cell (i,j) marked in any primary period before period k and $\sum_{\ell=1}^{t-1}M^{k,l}_{i,j}$ is the total number of individuals from cell (i,j) marked on any secondary sampling occasion before t within primary period k.

Therefore, similarly to Sections 3.2 and 3.4, it follows that the observed number of unmarked individuals caught on secondary sampling occasion t of primary period k is

$$u_{k,t}=\sum_{i=0}^{k-1}\sum_{j=k}^{K}M^{k,t}_{i,j}$$

while the total number of individuals that were marked in primary period k and recaptured on secondary sampling occasion t of primary period ν are modelled as

$$m_{k,\nu,t}\sim\begin{cases}\text{Binomial}\left(\sum_{i=0}^{k-1}\sum_{j=\nu}^{K}\sum_{\ell=1}^{t-1}M^{k,l}_{i,j},p_{k,t}\right), & \nu=k,\ t=2,3,...,T_\nu\\ \text{Binomial}\left(\sum_{i=0}^{k-1}\sum_{j=\nu}^{K}\sum_{\ell=1}^{T_k}M^{k,l}_{i,j},p_{k,t}\right), & \nu>k,\ t=1,2,...,T_\nu\end{cases}$$

In the above, we distinguish between two cases: recaptures taking place in the same primary period as the first capture ($\nu = k$), and recaptures taking place in subsequent primary periods ($\nu > k$), where the sum over ℓ in each case is used to calculate the total number of individuals from cell (i, j) that were marked in primary period k.

3.5.2 Model

The complete model for BM data collected under a RD is given in Equation 3.5.1.

• latent number individuals in cell (i,j) of matrix n	$n_{i,j} \sim \text{Poisson}(\omega \times w_{i,j}),\ i=0,...,K,\ j=i,...,K$
• latent number of individuals in cell (i,j) of matrix n, that are unmarked on secondary sampling occasion t of primary period k	$U_{i,j}^{k,t} = \begin{cases} n_{i,j},\ k=1,\ t=1 \\ n_{i,j} - \sum_{\nu=1}^{k-1}\sum_{l=1}^{T_\nu} M_{i,j}^{\nu,l} - \sum_{l=1}^{t-1} M_{i,j}^{k,l},\ k \neq 1, t \geq 1 \end{cases}$
• latent number of individuals in cell (i,j) of matrix n marked on secondary sampling occasion t of primary period k	$M_{i,j}^{k,t} \sim \text{Binomial}(U_{i,j}^{k,t}, p_{k,t})$
• observed number of unmarked individuals caught on secondary sampling occasion t of primary period k	$u_{k,t} = \sum_{i=0}^{k-1}\sum_{j=k}^{K} M_{i,j}^{k,t},\ k=1,...,K\ \ t=1,...,T_k$
• observed number of individuals that were marked in primary period k and were recaptured on secondary sampling occasion t of primary period ν	$m_{k,\nu,t} \sim \begin{cases} \text{Binomial}\left(\sum_{i=0}^{k-1}\sum_{j=\nu}^{K}\sum_{l=1}^{t-1} M_{i,j}^{k,l}, p_{k,t}\right), \\ \nu = k,\ t=2,3,...,T_\nu \\ \text{Binomial}\left(\sum_{i=0}^{k-1}\sum_{j=\nu}^{K}\sum_{l=1}^{T_k} M_{i,j}^{k,l}, p_{k,t}\right), \\ \nu > k,\ t=1,2,...,T_\nu \end{cases}$

(3.5.1)

3.5.3 Case Study 2: Inference

We employ an MCMC MH random walk for inferring the latent matrices $\left\{M^k\right\}_{k=1}^{K}$ and n, and a Gibbs sampler for the probabilities $\left\{w_{i,j}\right\}_{i=0,j=i}^{K,K}$, $\left\{p_{k,t}\right\}_{k=1,t=1}^{K,T_k}$ and for ω. Details, about the conditional distributions are given in Appendix B, Section B.1 (page 123).

3.5.4 Simulation study

We simulated a population size of 5000 individuals across 5 primary periods, considering time varying entry/exit probabilities and capture probabilities. Their values are given in Appendix B, Section B.5 (page 134), together with the prior distribution choices for the parameters. To assess the impact of the number of secondary sampling occasions on the quality of inference, we simulated data using 1, 2, 4, 8, and 16 secondary sampling occasions.

We ran the MCMC for 500,000 iterations, with the 100,000 iterations as burn-in. The results are presented in Table 3.2 and in Appendix B, Section B.5 (page 134). Here, we summarize the effect, in terms of percentage decrease in root mean squared error (RMSE), in the estimated entry and exit probabilities and population size when the number of secondary occasions increases (Table 3.2). Our findings indicate that, as expected, incorporating more sampling occasions within each primary period leads to smaller RMSE for all parameters. However, the benefit in terms of decrease in RMSE is not proportionate to the increase in effort for larger numbers of secondary periods, and the effects of the additional effort practically diminish for more than eight secondary periods, especially in the case of population size. These results highlight the benefit of the RD within BM studies in comparison to the standard BM without any periods of closure, and can support decisions around study design and allocation of effort in BM surveys.

Table 3.2 The median decrease change in RMSE as we increase the number of secondary sampling occasions from 1 to 2, 2 to 4, 4 to 8 and lastly from 8 to 16. The displayed parameters are the population size, N, and the entry and exit probabilities $V_i, V_{.,j}$ for $i = 1,2,...,K$ and $j = 1,2,...,K-1$, for K primary periods.

	Secondary sampling occasions					
Parameters				1→2	2→4	4→8	8→16
N				8%	5%	3%	0.3%
V_i				20%	19%	11%	4%
$V_{.,j}$				21%	20%	18%	14%

3.5.5 Golden mantella data

We apply our model to the data presented in Zhang et al. (2023) and compare the estimates of the entry/exit probabilities in Figure 3.4. Estimates of the capture probabilities and the number of

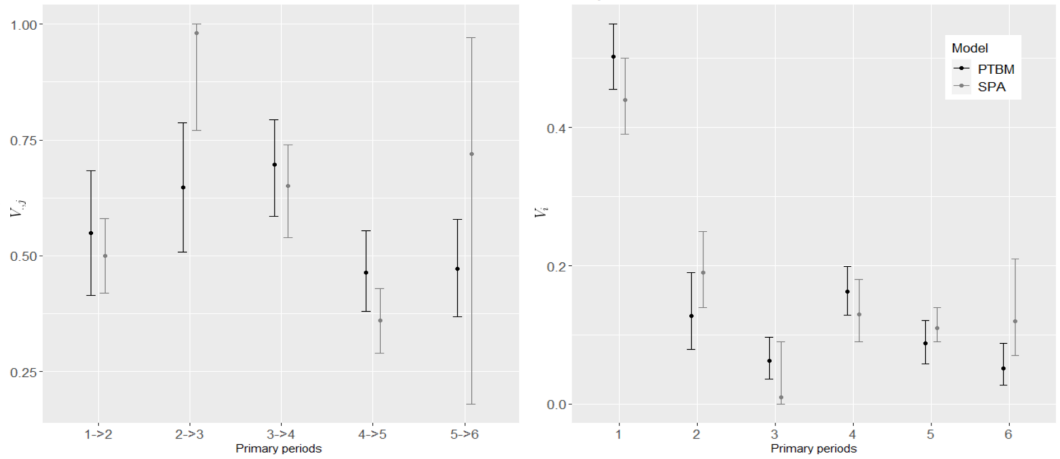

Fig. 3.4 Summaries of exit probabilities $V_{.,j}$ for $j = 1,2,...,K-1$, (a), and of entry probabilities V_i, $i = 1,2,...,K$, (b). The SPA model of Zhang et al. (2023) was fitted classically, so the summaries correspond to the maximum likelihood estimate and corresponding 95% confidence interval in each case, whereas our PTBM and PTBM-R models are fitted in a Bayesian framework, so the summaries correspond to the posterior mean and 95% posterior credible interval.

available individuals can be found in Appendix B, Section B.6 (page 138). We use the same prior distribution choices as the ones displayed in Appendix B, Section B.5 (page 134). Additionally, the model parameters are $(\omega, \{V_i\}_{i=1}^{K}, \{V_{.,j}\}_{j=1}^{K}, \{p_{k,t}\}_{k=1,t=1}^{K,T_k}, \{M^{k,t}\}_{k=1,t=1}^{K,T_k}, \{R^{k,t}\}_{k=1,t=1}^{K,T_k}, n)$.

The results demonstrate that the two have similar entry probabilities slightly different super-population, exit and capture probabilities, with PTBM consistently narrower posterior credible intervals than the corresponding confidence intervals obtained by Zhang et al. (2023) using SPA. We note here that we have chosen flat priors for all model parameters, to make our results as comparable as possible with the SPA results, and hence our interpretation of this difference in interval width is that it is due to the use of the exact, instead of approximate, likelihood in our case.

Figure 3.4 demonstrates that the exit probabilities are consistently larger for the midpoints of the breeding seasons, which is expected, as also reported by Zhang et al. (2023). Additionally, the entry probabilities, on average, are higher in the early primary periods within each breeding season. The estimates of the number of individuals present on each primary period are similar for SPA and PTBM for the early primary periods however, in our case we estimate a smaller number of individuals being present for the later primary periods, with the PTBM having narrower intervals as shown in Figure 13 in Appendix B, Section B.6 (page 138). Lastly, the super-population size is smaller for our

PTBM model with 95% posterior credible interval (4808, $(4371, 5310)$ by PTBM) compared to the corresponding 95% confidence interval (5567 $(5145, 6063)$ by SPA).

Lastly, we assess the convergence of our MCMC algorithm for the PTBM model by considering diagnostics such as Geweke's diagnostic (Geweke, 1991), effective sample size, Monte Carlo error of the parameters, and the $\hat{R}$ statistic (Vehtari et al., 2021), given that we ran four MCMC chains. The results are displayed in Table B.2 and Figures B.17 - B.28. The convergence analysis indicates no issues, as the $\hat{R}$ for all parameters is close to 1, and the Monte Carlo error is sufficiently small, ensuring confidence in our posterior estimates. In Figure B.29 we display a goodness of fit that enhances that our posterior estimates are reasonable for the given data since they can perfectly produce the number of captured individual across the secondary sampling occasions across the primary periods. The algorithm took 30 minutes to run on a computer with CPU 11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80GHz.

3.6 Discussion

BM surveys are an important monitoring approach for several species and our chapter contributes to the, fairly limited thus far, statistical literature for modelling the corresponding data. Our introduced latent grid for modelling the entry and exit pattern, together with the associated PT prior for the grid probabilities, provides a new, general, flexible and computationally efficient modelling framework for BM data. Our model scales with the number of sampling occasions, which is typically much smaller than the number of individuals, as it infers the number of individuals with a specific entry/exit interval, rather than inferring latent individual presence or capture histories, and gives us access to exact inference under different observation processes. The PT model described in this chapter may appear similar to a Dirichlet-Multinomial model (Royle et al., 2012), which is, in fact, a special case of the PT framework when considering the partition described in Section 3.3. However, the two methods are not identical since we utilize the PT machinery, such as the replicate PT, to impose restrictions on the dynamic parameters of the model. The replicate PT structure allows us to place constraints on the model parameters to build parsimonious models that are also ecologically meaningfully. In our case, we have assumed a replicate PT structure that leads to the assumption that the probability

of exit depends on the sampling occasion, and not on the time of entry (time-varying exit). The PT framework gives us access to efficient approaches for model comparisons of different constraints, such as constant, time, or age-varying probabilities (Holmes et al., 2015), and hence future work could explore establishing the required methodology for building and comparing models with different constraints for the model parameters.

Additionally, in the case studies and corresponding simulations, we have employed flat prior distributions for all parameters, to make our results as comparable as possible to the classical models considered thus far in the literature. However, the PT can be centred on parametric distributions, such as the normal, that express our prior expectation of the entry/exit pattern, and in that case inference benefits from both the smoothness of the parametric curve and the flexibility of the nonparametric PT prior.

We considered two case studies with different observation processes to demonstrate the flexibility of our proposed PT prior approach to accommodate different data-generating processes. Our framework can easily be extended to jointly model BM data with capture-recapture, count or other types of ecological data that are often collected in practice. However, an open challenge that remains is the incorporation of covariates in the model parameters, and in particular when these are measured at the individual level, as in that case the grid-approach does not apply, at least not in its current form.

Overall, we developed two models, in the first one we integrated essential information coming from individual removals and derived more accurate inference about population sizes, whereas in the second model we created a tractable model which produces results independent of any kind of approximation error. Hence, we created tools that can produce reliable results for practitioners that seek to study population size as indices of species dynamics, which is a crucial metric in conservation.

Chapter 4

An approximate Bayesian computation approach for modelling batch-mark data on populations exhibiting temporary emigration

Abstract

Batch-mark data are frequently collected and analysed for monitoring populations of different species, such as amphibians. Chapter 3 of the thesis introduced a grid-based approach for performing exact inference for batch-mark data collected under different study designs. However, this approach does not extend to the case of batch-mark data if the corresponding surveyed population exhibits temporary emigration, as is the case for the motivating case study considered in this chapter. Instead, to estimate population size and other demographic parameters of interest, such as the probability of survival, while accounting for temporary emigration, this chapter introduces an approximate Bayesian computation approach. The approach is demonstrated using simulation and for the motivating case study.

4.1 Introduction

Open population models have been extensively employed to infer population dynamics, such as entry/exit patterns and population size (Jolly, 1965; Seber, 1965; Seber and Schofield, 2019) of surveyed populations, where entry can correspond to birth or arrival, whereas exit to death or departure. Sampling of such populations occurs on a series of sampling occasions, during which newly caught individuals are marked and then returned to the population. If the assigned marks are unique for each individual, then the process is referred to as capture-recapture (CR) (Schwarz and Arnason, 1996), whereas if the marks are unique to each sampling occasion, then the process is referred to as batch-marking (BM) (Cowen et al., 2017). Individuals enter and exit between any pair of consecutive sampling occasions and can be caught if they are present at the time of sampling. It is typically assumed that emigration is permanent. This means each individual can only enter and exit the population once. They are also available for detection during all sampling occasions occurring between entry and exit times. However, in reality, this assumption is often

violated. For example, when individuals are sampled at a breeding site over different periods, they can temporarily become unavailable for sampling in periods where they have not visited the breeding site (Spendelow and Nichols, 1989). Similarly, individuals can temporarily move to inaccessible parts of the surveyed area and hence become unavailable for sampling during a period of time. In these cases, models should account for temporary emigration to reliably estimate demographic parameters. For example, the absence of accounting for temporary emigration can lead to positively biased estimates of the population size, as the fact that individuals may temporarily emigrate is ignored (Bird et al., 2014; Kendall et al., 1997). For CR data, the existing literature (Kendall et al., 1997; Schaub et al., 2004) mainly relies on Pollock's robust design (RD) (Pollock, 1982) and recently, there has been an advancement as described in Kendall and Bjorkland (2001); Matechou and Argiento (2023); Schwarz and Stobo (1997), involving the creation of a model that does not rely on the closure assumption of RD. However, we are not aware of any model incorporating temporary emigration for BM data.

We assume that the survey takes place over K periods with T_k sampling occasions within period $k = 1, 2, ..., K$. The population is considered open between and within the K periods, so we are not assuming a RD. Each individual has a corresponding unique pair of first entry and last exit time during the survey, which for example correspond to their birth and death, indicating their lifespan, or the time of their first and last breeding seasons. Within each period, individuals can migrate/travel/enter the survey area, which for example can be a breeding site or a stopover site, and hence can have a single corresponding pair of entry and exit times for that corresponding season, indicating the sampling occasions within the period where that individual is available for detection. However, individuals do not necessarily visit the study site in each period, and hence they become temporary unavailable in that case. Accounting for temporary emigration in this case is important for obtaining reliable inference on population parameters, such as population size and survival probability (between periods), (Bird et al., 2014; Kendall et al., 1997).

In Chapter 3, we defined and described a grid approach for modeling open population BM data using the Pólya Tree prior. However, in the case of BM data with temporary emigration, this approach does not lead to exact inference due to the complexity of the likelihood function, which cannot be expressed analytically. This is because the grid latent matrices described in Chapter 3 contain nested information that we cannot disentangle using only BM data, as individual-level information is

required. We provide details on why this is not feasible in Section C.1 (page 140) of the Appendix C. Instead, to address this challenge, in this chapter we adopt a likelihood-free inference approach: Approximate Bayesian Computation (ABC) (Marin et al., 2012; Sisson et al., 2018; Wilkinson, 2013). The idea of ABC is to generate data from the model using a proposed set of parameters and then accept those parameters if the generated data are close to the observed ones. However, ABC can sometimes be inefficient and time-consuming since building an approximation sufficiently close to the posterior distribution relies on randomly proposing parameters that generate data close to the observed ones . To mitigate this, we utilized the Approximate Bayesian Computation Sequential Monte Carlo (ABC-SMC) algorithm (Bonassi and West, 2015; Del Moral et al., 2006, 2012; Sisson et al., 2007) and in particular the ABC-SMC algorithm described in Sisson et al. (2007). The basic idea is grounded in Particle Filtering theory, where we consider a set of particles (in our case, a number of simulated parameters), each with a corresponding weight. These particles are iteratively propagated towards regions of higher mass in the posterior distribution. With each iteration, parameters that simulate data closer to the observed data have larger weights and are represented more frequently in each propagation step. This is performed by progressively decreasing a threshold that tunes the approximation to the true posterior distribution, in a similar fashion as simulated tempering (Marinari and Parisi, 1992).

We demonstrate the efficiency and novelty of our proposed method through a simulation study and by applying to a data-set of an alpine common toad (*Bufo bufo*) population, collected across a period of 8 years from 1986-1993.

The chapter is structured as follows: In Section 4.2, we describe the model and the generating process of the BM data. In Section 4.3, we describe the ABC-SMC algorithm that we use to approximate the posterior distribution. Additionally, in Section 4.4, we present the distributions used for data inference. In Section 4.5, we provide a simulation study demonstrating the performance of our algorithm. Lastly, in Section 4.6, we analyze the alpine common toad data, and in Section 4.7, we discuss the limitations, advantages, and further avenues for our model.

4.2 Data and Latent Notation

The data are expressed in terms of the sufficient statistics of batch mark data.

Data

K: total number of periods, taking place at times $t_1, ..., t_K$.

T_k: total number of sampling occasions in period k, taking place at times $t_1^k, ..., t_{T_k}^k$.

u_t^k: observed number of unmarked individuals caught in sampling occasion $t = 1, 2, ..., T_k$ in period $k = 1, 2, ..., K$.

$m_t^{k,s}$: observed number of individuals recaptured in sampling occasion $t = 1, 2, ..., T_s$ of period $s \geq k$, first caught in period $k = 1, 2, ..., K$, where if $s = k$ then $t > 1$.

Parameters

N : population size, corresponding to individuals that were alive regardless if they were present on any period. We model N as a Poisson(ω), where ω is the expected population size.

ψ: probability of an individual being present in a given period, conditional on being alive.

p: probability of capturing an individual that is available on a given sampling occasion, conditional on being alive and present.

η_i: the probability of an individual entering the population (start being alive) between periods t_{i-1} and t_i , $i = 1, 2, ..., K$.

ϕ: the probability of an individual exiting the population (stop being alive), conditional on being alive on the previous time period.

V_i^k: the probability of an individual present in period k entering the survey area (becoming available for capture) between the sampling occasions (t_{i-1}^k, t_i^k), $i = 1, 2, ..., T_k$.

L_i^k: the probability of an individual present in period k exiting the survey area (becoming unavailable for capture) between the sampling occasions (t_i^k, t_{i+1}^k) $i = 1, 2, ..., T_k$ conditional on being in the survey area on the previous sampling occasion.

Generating Process

Next, we describe the data generating process.

- For each individual in the population, we define the alive history $\underline{A}$, which is a vector of length K with component k equal to 1 if the individual is alive in period k and 0 otherwise. To sample the alive history, we first sample the period in which the individual enters the population for the first time from a categorical distribution with probabilities $(\eta_1, \eta_2, \ldots, \eta_K)$. Next, starting from the period in which the individual enters the population, we sample the alive history up to the point where they exit the population. If the individual enters the population on the kth period then we set $A_j = 0$ for $j < k$, and we simulate

$$A_{j+1} \sim \text{Bernoulli}(A_j(1-\phi)),\ j = k, k+1, ..., K-1$$

- We define the presence history $\underline{P}$, which is a vector of length K with component k equal to 1 if individual is present in period k and 0 otherwise, conditional on being alive. We sample the presence history by thinning the alive histories with probability of presence ψ.

$$P_k \sim \text{Bernoulli}(A_k \psi),\ k = 1, 2, ..., K$$

- For each individual and period k, we define the availability history $\underline{Z}^k$, which is a vector of length T_k with component l equal to 1 if the individual is available on sampling occasion l of period k and 0 otherwise. We sample the availability history conditional on the presence history $\underline{P}$ for each individual. For each period k, first we simulate on which sampling occasion the individual becomes first available from a categorical distribution with probabilities $(V_1^k, V_2^k, \ldots, V_{T_k}^k)$. Next, assuming that the individual becomes available in sampling occasion i, we sample the sampling occasion in which they exit the survey area from a categorical distribution with probabilities $(L_{i,k}, L_{i+1,k}, \ldots, L_{T_k,k})$ and $L_{i,j}$ corresponding to the probability that an individual that became available in sampling occasion i exits between the jth and $(j+1)$th sampling occasions for $j = i, \ldots, K$. Those probabilities will always sum up to 1, since they are calculated as normalized Dirichlet random variables.

- Finally, conditional on the availability history $\underline{Z}^k$, we simulate for each period k the capture history $\underline{C}^k$, which is a vector of length T_k with component l equal to 1 if the individual is captured on sampling occasion l of period k, as

$$C_l^k \sim \text{Bernoulli}(Z_l^k p),\ l = 1,2,\ldots,T_k.$$

After having simulated the capture histories for each individual and each period, we aggregate them and derive the sufficient statistics, which are the number of unmarked u_t^k and marked $m_t^{k,s}$ individuals caught on each sampling occasion, as introduced above.

4.3 Approximate Bayesian Computation Sequential Monte Carlo

Firstly, we provide in Algorithm 1 an outline of the vanilla ABC:

Algorithm 1 Approximate Bayesian Computation (ABC)

Require: Observations D, Model $p(D|\theta)$, Prior distribution $\pi(\theta)$
1: Initialize empty sample set Θ
2: Initialize tolerance threshold ε
3: **for** $i = 1$ to G **do** ▷ Generate G parameter samples
4: Sample θ_i from the prior distribution $\pi(\theta)$
5: Simulate data D_i from the model $p(D|\theta_i)$
6: Calculate distance measure $d(D, D_i)$ ▷ e.g., Euclidean distance
7: **if** $d(D, D_i) < \varepsilon$ **then**
8: Add θ_i to the sample set Θ
9: **else**
10: Go to step 4 and repeat the process
11: **end if**
12: **end for**
13: Approximate posterior distribution $p(\theta|D)$ with the distribution $p_\varepsilon(\theta|D)$ using the sample set Θ

The efficiency of the approximation depends on the choice of the threshold ε. It is easy to observe that if $\varepsilon = 0$ we would sample from the true posterior distribution, in cases where we use the raw generated data or sufficient statistics of them. However, this would require an unacceptably large number of iterations to accept any number of samples G, as the probability of generating data exactly equal to the observed data is extremely small. On the other hand, if ε is chosen too large then the simulated samples are not a good approximation of the posterior distribution, hence the threshold ε has to be chosen with care.

The Algorithm 1 exhibits inefficiency because of independently sampling from the prior, without accounting for information-sharing across samples. Consequently, with each draw, success relies on the chance of landing in a region of high posterior probability, in order to accept a parameter θ. This issue can be overcome by using ABC-SMC (Del Moral et al., 2012; Sisson et al., 2007) which allows to propagate particles through regions of large posterior mass via iteratively sampling parameters using a decreasing threshold.

Algorithm 2 Approximate Bayesian Computation Sequential Monte Carlo

Require: Observations D, Model $p(D|\theta)$, Prior distribution $\pi(\theta)$, Perturbation kernel $K_t(\cdot|\cdot)$, Backward Kernel $L_{t-1}(\cdot|\cdot)$, ESS threshold E

1: Initialize $\varepsilon_1, ..., \varepsilon_T$ ▷ decreasing tolerance thresholds sequence
2: Initialize a number of samples G ▷ occasionally referred as particles
3: Set the propagation indicator $t = 1$
4: **for** $t = 1, \ldots, T$ **do**
5: **for** $i = 1, \ldots, G$ **do**
6: **if** $t = 1$ **then**
7: Sample θ^{**} independently from $\pi(\theta)$.
8: **else**
9: Sample θ^* from the previous population $\left\{\theta_{t-1}^{(i)}\right\}_{i=1}^{G}$ with weights $\left\{\gamma_{t-1}^{(i)}\right\}_{i=1}^{G}$ and perturb the particle to obtain $\theta^{**} \sim K_t(\theta|\theta^*)$
10: **end if**
11: Sample $D^* \sim p(D|\theta^{**})$
12: **if** $d(D^*, D) \geq \varepsilon_t$ **then** ▷ e.g., Euclidean distance
13: return to 6.
14: **end if**
15: Set $\theta_t^i = \theta^{**}$ and calculate the weight for particle θ_t^i

$$\gamma_t^i = \begin{cases} 1, & \text{if } t = 0 \\ \frac{\pi(\theta_t^i) L_{t-1}(\theta^*|\theta_t^{(i)})}{\pi(\theta^*) K_t(\theta_t^{(i)}|\theta^*)}, & \text{if } t > 0 \end{cases}$$

16: **end for**
17: Normalize the weights, so that $\sum_{i=1}^{G} \gamma_t^{(i)} = 1$
18: **if** ESS$= \frac{1}{\sum_{i=1}^{G} (\gamma_t^{(i)})^2} < E$ **then**
19: resample, $\left\{\theta_t^{(i)}\right\}$ with $\left\{\gamma_t^{(i)}\right\}$ to obtain new particles and set weights $\left\{\gamma_t^{(i)} = 1/G\right\}$
20: **end if**
21: **end for**

Briefly, ABC-SMC works as follows. We start by selecting a number T of propagation steps, a number of a decreasing thresholds $\{\varepsilon_t\}_{t=1}^{T}$ and a number G of particles. At the first iteration $t = 1$, we sample the initial set of G particles $\left\{\theta_1^{(i)}\right\}_{i=1}^{G}$ from the prior distribution $\pi(\theta)$ of the parameters, and assign the weights $\left\{\gamma_1^{(i)} = 1/G\right\}_{i=1}^{G}$. Next, for each iteration $t > 1$, for each particle $i = 1, \ldots, G$, we

- Sample $\theta_t^{i,*}$ from the particles at the previous time point $\left\{\theta_{t-1}^{i}\right\}_{i=1}^{G}$ with probabilities $\left\{\gamma_{t-1}^{i}\right\}_{i=1}^{G}$.
- Perturbate the sampled particle $\theta_t^{i,*}$ using a transition kernel $K_t(\theta_t^{i,**}|\theta_t^{i,*})$.
- For each particle $\theta_t^{i,**}$ simulate data D^*. If the distance between the simulated data D^* and the real data D is less than the threshold ε_t, accept the particle. If the distance is greater than the threshold, we go back to the first step.
- Following Sisson et al. (2007), we calculate the weight $\gamma_t^{(i)}$ of each particle $\theta_t^{(i)}$ using a backward kernel L_t that minimize the variance of the weights induced through K_t. More details on L_t are presented in Sisson et al. (2007). Also, there exists a unique optimal kernel for L_t given the forward kernel K_t, defined in Proposition 1 in (Del Moral et al., 2006). However, in practise such calculation is not trivial, making practitioners choose their own backward kernel by trial and error based on how more information the particles are bringing for the different backward kernels.
- If the Effective Sample Size (ESS) is smaller than a threshold E then we resample the particles $\left\{\theta_t^{(i)}\right\}$ with weight $\left\{\gamma_t^{(i)}\right\}$ and we set the weights to $\left\{\gamma_t^{(i)} = 1/G\right\}$.

4.4 Inference

We apply Algorithm 2 to the model outlined in Section 4.2 . For the implementation of the algorithm, we require the following components: the observation D, a process to generate new data D conditional on the parameters θ, the prior $\pi(\theta)$, perturbation and backward kernels $K_t(\cdot|\cdot)$, $L_{t-1}(\cdot|\cdot)$, the threshold E, the tolerance sequence $\varepsilon_1,\ldots,\varepsilon_T$, and the number of particles G. The observations consist of the standard output from the realization of the BM sampling survey and are denoted as $D = \left(\left\{u_t^k\right\}_{k=1,t=1}^{K,T_k}, \left\{m_t^{k,s}\right\}_{k=1,s=k,t=1}^{K,K,T_s}\right)$. The parameters are

$$\theta = (\omega, \psi, p, (\eta_1, \eta_2, ..., \eta_K), \phi, \left\{(V_1^k, V_2^k, ..., V_{T_k}^k)\right\}_{k=1}^{K}, \left\{(L_1^k, L_2^k, ..., L_{T_k}^k)\right\}_{k=1}^{K}),$$

.

The prior distributions assigned to the parameter θ, are

$$\text{Expected population size: } \omega \sim \text{Uniform}\left(\sum_{k=1}^{K}\sum_{t=1}^{T_k} u_t^k, UB\right),\ UB \text{ sufficiently large upper bound}$$

$$\text{Presence probability: } \psi \sim \text{Uniform}(0,1)$$

$$\text{Capture probability: } p \sim \text{Uniform}(0,1)$$

$$\text{Population entry probabilities: } (\eta_1,\eta_2,\ldots,\eta_K) \sim \text{Dirichlet}(1,1,\ldots,1)$$

$$\text{Population exit probability: } \phi \sim \text{Uniform}(0,1)$$

$$\text{Survey area entry probabilities: } (V_1^k,V_2^k,\ldots,V_{T_k}^k) \sim \text{Dirichlet}(1,1,\ldots,1),\ k=1,2,\ldots K$$

$$\text{Survey area exit probabilities: } (L_1^k,L_2^k,\ldots,L_{T_k}^k) \sim \text{Dirichlet}(1,1,\ldots,1),\ k=1,2,\ldots,K$$

To achieve a more parsimonious model, we assume that the parameters $(L_1^k,L_2^k,\ldots,L_{T_k}^k)$ and $(V_1^k,V_2^k,\ldots,V_{T_k}^k)$ are the same across different periods. However, different periods might have different number of sampling occasions. To overcome this problem, we define the unnormalized probabilities $(\tilde{V}_1,\tilde{V}_2,\ldots,\tilde{V}_{\max\{T_k\}_{k=1}^K})$ and $(\tilde{L}_1,\tilde{L}_2,\ldots,\tilde{L}_{\max\{T_k\}_{k=1}^K})$, and assign on each a Gamma$(0.001,0.001)$ distribution. Then, conditional on the period and the number of sampling occasions within the period, we can truncate the vector $(\tilde{V}_1,\tilde{V}_2,\ldots,\tilde{V}_{\max\{T_k\}_{k=1}^K})$ and $(\tilde{L}_1,\tilde{L}_2,\ldots,\tilde{L}_{\max\{T_k\}_{k=1}^K})$ and normalise it to obtain the within-period entry probabilities by normalizing them. The process to generate new data is described in Section 2.

For the perturbation and backward kernel, we use, for every time point t, LogNormal distributions with scale 0.05 for the expected population size ω and the unnormalised probabilities $(\tilde{V}_1,\tilde{V}_2,\ldots,\tilde{V}_{\max\{T_k\}_{k=1}^K})$, and $(\tilde{L}_1,\tilde{L}_2,\ldots,\tilde{L}_{\max\{T_k\}_{k=1}^K})$ and a logistic normal with scale 0.5 for the other parameters. The threshold E value is taken to be equal to $G/2$ as explained in Sisson et al. (2007). The threshold sequence is chosen such that ε_T does not require more than a 1000000 iterations to accept a particle. In order to achieve that, we start from $\varepsilon_1 = 2000$ and progressively increase ε_t by 1 until the algorithm needs more than 1000000 iterations to accept a particle. Finally, we choose the distance $d(\cdot)$ as the Euclidean distance. The choice of distance metric in ABC-SMC is crucial, especially when dealing with a large number of summary statistics, which can lead to the curse of dimensionality. While Euclidean distance is a common choice, the weighted Euclidean distance can

often be more effective. By using the inverse of the standard deviation of the summary statistics as weights, this metric accounts for differences in scale between the statistics, preventing the distance from becoming excessively large. However, in our case we used the Euclidean distance because we did not have a large number of summary statistics and it behaved well in terms of the ESS and estimations that we received in our simulation study.

4.5 Simulations

We conducted a simulation study to explore the robustness of our approximation. For 100 replications, we simulated BM data with temporary emigration over a span of $K = 8$ periods, with a population size of $N = 5000$ individuals. The presence probability ψ is taken equal to 0.7 and the capture probabilities p across periods are taken equal to 0.5. Within each period, we considered $T_k = 5$ sampling occasions. The chosen values for the probabilities of entering the population across periods are $\eta_1 = 0.616, \eta_2 = 0.132, \eta_3 = 0.099, \eta_4 = 0.055, \eta_5 = 0.049, \eta_6 = 0.032, \eta_7 = 0.011, \eta_8 = 0.006$. We also set $\phi = 0.129$. For the within periods entry and exit probabilities, we take $V_1^k = 0.471, V_2^k = 0.257, V_3^k = 0.202, V_4^k = 0.044, V_5^k = 0.026$. For the probability of exiting within period, we chose $L_1^k = 0.056, L_2^k = 0.086, L_3^k = 0.135, L_4^k = 0.330, L_5^k = 0.391$. The parameter values were chosen to ensure that the population does not deplete, both across and within the years, allowing for accurate estimation of the parameters.

The priors assigned to the model parameters are

$$
\begin{aligned}
p &\sim \text{Uniform}(0,1) \\
\psi &\sim \text{Uniform}(0,1) \\
\omega &\sim \text{Uniform}\left(\sum_{k=1}^{K}\sum_{t=1}^{T_k} u_t^k, 15000\right) \\
\eta_1, \eta_2, \ldots, \eta_8 &\sim \text{Dirichlet}(1,1,\ldots,1) \\
\phi &\sim \text{Uniform}(0,1) \\
V_1^k, V_2^k, \ldots, V_5^k &\sim \text{Dirichlet}(1,1,\ldots,1) \\
L_1^k, L_2^k, \ldots, L_5^k &\sim \text{Dirichlet}(1,1,\ldots,1)
\end{aligned}
$$

In Table 4.1, we display summaries of the simulations. All parameters are estimated correctly up to 3 decimal places.

4.6 Case study

We applied our model to data on alpine common toad, gathered over a period of eight years, from 1986 to 1993. Alpine common toads were collected at Grosse Scheidegg near Grindelwald in the Alps, at an altitude of approximately 1,850 meters. Every spring, a mark-recapture study was conducted at their breeding ponds, where the toads were marked by toe clipping. Within each year, we had $T_k = (4,3,5,3,3,4,4,3)$ sampling occasions. The practitioners who collected the data informed us that they capture all toads with certainty. Since their survey areas are ponds, they conduct thorough searches and capture every toad present. Hence, we assume a capture probability equal to 1. The prior distributions used for all the parameters are:

$$\begin{aligned}
\omega &\sim \text{Uniform}(\sum_{k=1}^{8}\sum_{t=1}^{T_k} u_t^k, 4000) \\
\psi &\sim \text{Uniform}(0,1) \\
\eta_1, \eta_2, ..., \eta_8 &\sim \text{Dirichlet}(1,1,...,1) \\
\phi &\sim \text{Uniform}(0,1) \\
\tilde{V}_t &\sim \text{Gamma}(0.001, 0.001),\ t = 1,2,...,\max\{T_k\}_{k=1}^{8} \\
\tilde{L}_t &\sim \text{Gamma}(0.001, 0.001),\ t = 1,2,...,\max\{T_k\}_{k=1}^{8}
\end{aligned}$$

The parameters to be inferred are $(\omega, \psi, \eta_1, \eta_2, \ldots, \eta_8, \phi, \tilde{V}_1, \ldots, \tilde{V}_{\max\{T_k\}_{k=1}^{8}}, \tilde{L}_1, \ldots, \tilde{L}_{\max\{T_k\}_{k=1}^{8}}$, since the capture probability is taken equal to one based on the information given from the practitioners that they capture every individual in a pond.

We ran the ABC-SMC algorithm with $G = 500$ particles and display the results in Figures 4.1-4.4 and Table 4.2. Based on the algorithm configuration described in Section 4.4 the algorithm did 1.5 day to run on a computer with CPU 11th Gen Intel(R) Core(TM) i7-1165G7 @ 2.80GHz. We

experimented with different numbers of particles G, but ultimately chose $G = 500$ because it provided a reasonable effective sample size and ensured that the particles adequately covered the entire domain of the posterior distributions and it was fast compared to options of larger number of particles. The analysis used summary statistics, including the number of unmarked individuals caught during each secondary sampling occasion across the years, the number of individuals recaptured within each year, and the total number of recaptures between years, aggregated across all secondary sampling occasions. The posterior distribution on the inferred number of individuals alive and the inferred number of individuals present in each period, summarized in Figure 4.1, demonstrate that the population size increased substantially from year 1 during the study period. The population increased in the 4th year, remained stable between the 5-7th years, and experienced a second increase in the 8th year. Similarly, the number of individuals uniquely captured each year, i.e., the number of unmarked individuals captured each year and the maximum number of recaptures of individuals marked from previous years, follows a pattern with a large portion of individuals caught in 4th and 8th years.

In Figures 4.2 through 4.4, we display the observed captured individuals: unmarked individuals for each year and sampling occasion (Figure 4.2), recaptured marked individuals within each year (Figure 4.3), and recaptured marked individuals across years (Figure 4.4). We then compare these observations with the simulated captured individuals from our inferred model as a measure of goodness of fit. What we observe is that there is a lot of variability across the observed data. In Figures 4.2 - 4.4, 14 out of 29, 18 out of 21, and 26 out of 28 of the observed data points are within the 95% credible intervals. Overall, we observe that most of the summary statistics fall within the credible intervals, though some do fall outside. This could be due to the approximations made by the algorithm. Since we are not using the actual likelihood function, it is expected that not all statistics will lie within the credible intervals.We performed a goodness-of-fit test to check whether the observed summary statistics match the distribution predicted by the posterior model, (Lemaire et al., 2016). We compared the Z-score of the observed summary statistic with the Z-scores derived from the posterior predictive samples. The resulting p-value of 0.55 indicates that there is no evidence to reject the null hypothesis, meaning the observed summary statistics likely come from the same distribution as predicted by the posterior model. Additionally, there is no evidence of systematic bias in any of these figures. However, our simulated data attempt to capture the trends in the observed data across years and sampling occasions.

Table 4.2 reports posterior summaries of all model parameters, including breeding probability $(0.73(0.67, 0.80))$ and expected population size across years $(504(438, 577))$.

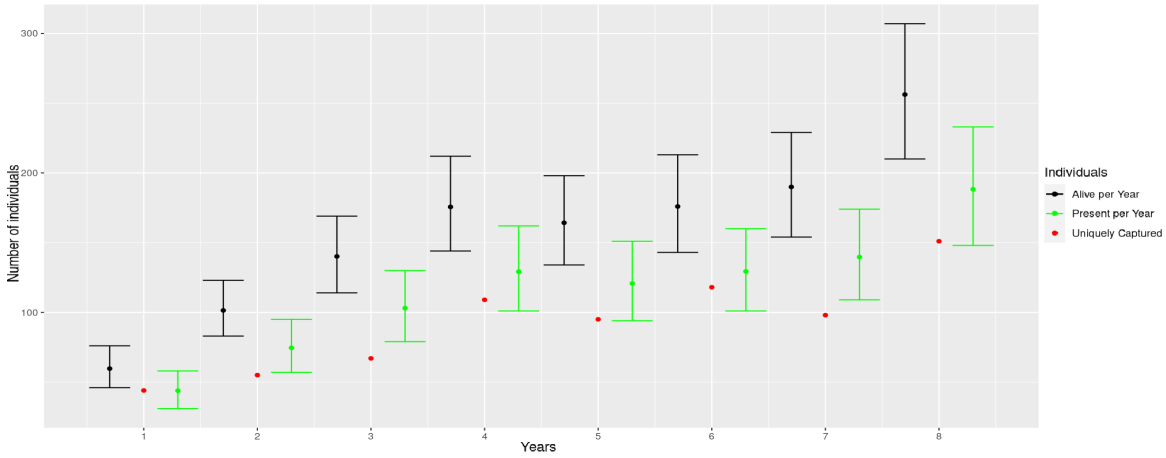

Fig. 4.1 Posterior inferred average number of individuals alive (black points), present (green points) across the years, 1986-1993, together with their 95% credible intervals and the unique number of captured individuals (red points), across the years, 1986-1993.

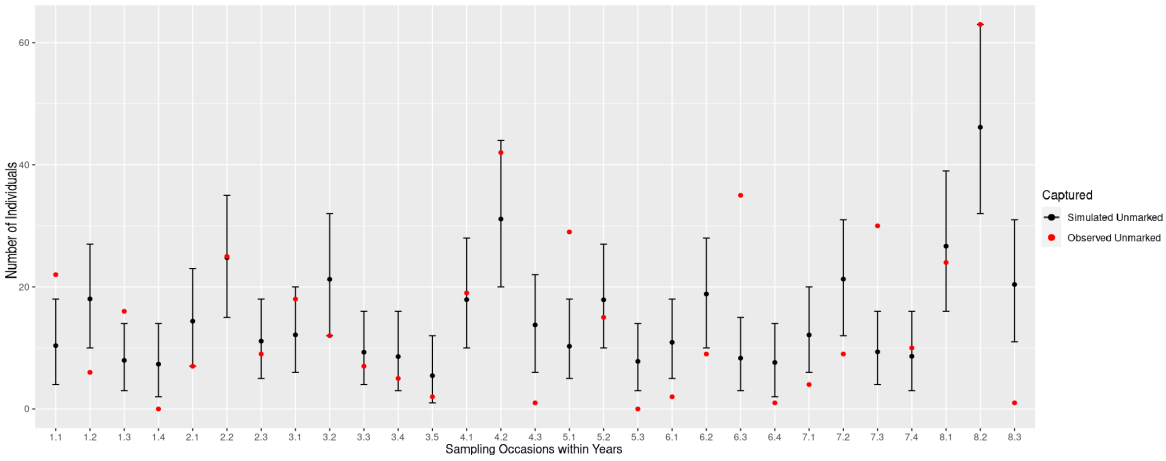

Fig. 4.2 Posterior simulated average and 95% credible interval of captured unmarked individuals (black points) compared with the observed unmarked individuals (red points) per sampling occasion across years 1986-1993.

Lastly, the effective sample size calculated is 489, which is almost equal to the number of particles used indicating a good exploration of the approximated posterior distributions.

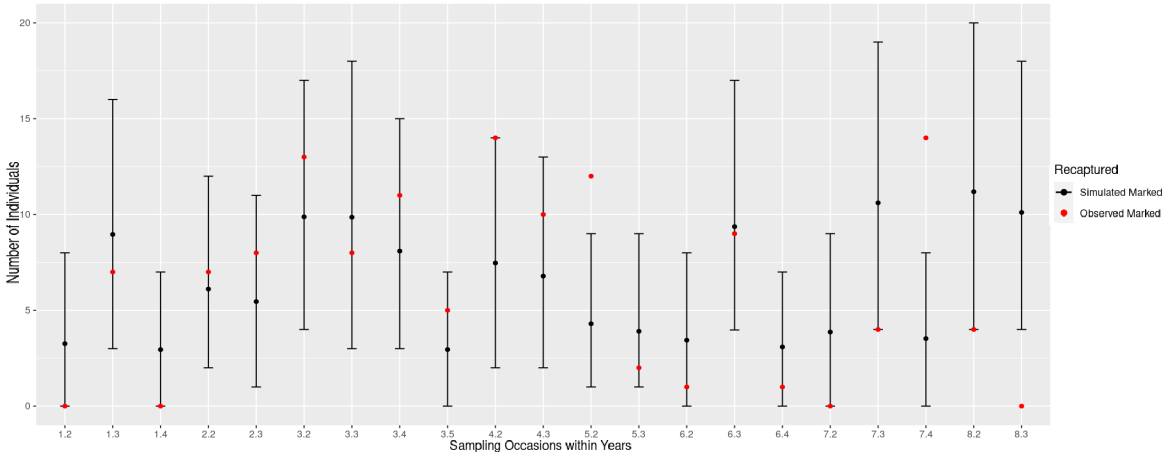

Fig. 4.3 Posterior simulated average and 95% credible interval of recaptured marked individuals (black points) within each year compared with the observed recaptured marked individuals (red points) per sampling occasion across years 1986-1993.

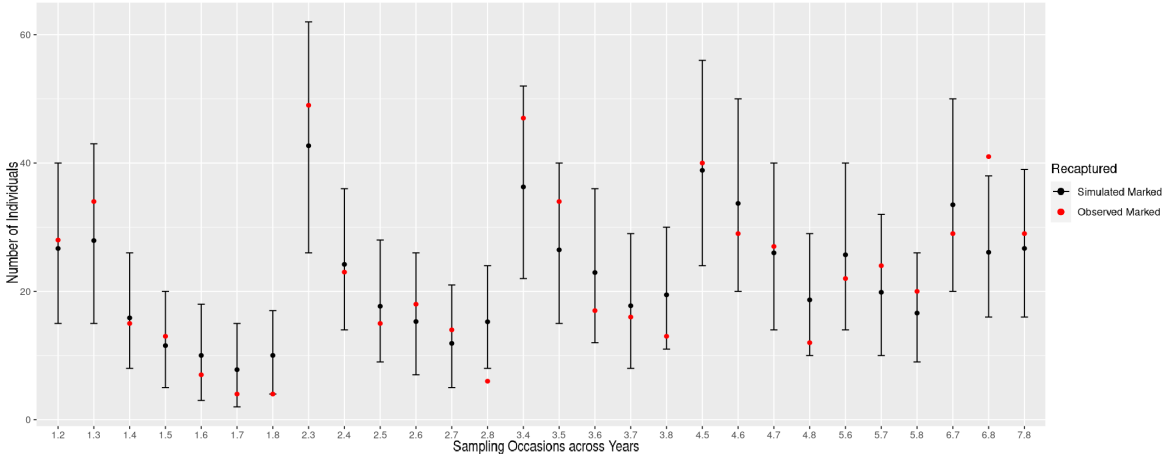

Fig. 4.4 Posterior simulated average and 95% credible interval of recaptured marked individuals (black points) between years compared with the observed recaptured marked individuals (red points) per sampling occasion across years 1986-1993.

4.7 Conclusion

In this chapter, we have developed for first time a model for complex observation processes in ecological surveys with intractable likelihoods, such as those resulting from surveying population that exhibit temporary emigrations with BM data across multiple years. We employed the ABC-SMC approach, which relies on being able to sample new data, instead of writing the likelihood, to approximate the posterior distribution of the parameters. Thanks to ABC-SMC, we developed a framework to address the intractability of BM data with temporary emigration, which was never done before. We believe that this can represent a fruitful research idea in ecology since complex observation processes are ubiquitous.

From our simulation study in Section 4.5, we observed that the model defined using ABC-SMC is able to capture the true values of parameters with high accuracy. Therefore, overall, our framework is fast and intuitive to implement and allows us to obtain new insights in cases where likelihood-based approaches are not accessible.

In the case of batch-mark data, the sufficient summary statistics are the full data, which means that we do not need to decide on how to summarise the data and introduce more approximations in inference. However, additional research could be done in the way the distance is chosen to compare summary statistics.

In the case study, we have assumed that all parameters are constant. This assumption can be relaxed to accommodate time-varying capture, presence, and entry/exit probabilities. Also, we have only considered eight years of data since after 1993 no batch marking took place. However, after 1993, the BM design was replaced by a capture-recapture design, and therefore the ABC approach could be extended to model the BM and CR data jointly.

Overall, the model presented in this section can provide insights into population dynamics, such as estimating species populations when breeding occurs during the sampling period. This can help practitioners better understand trends in species abundance and develop appropriate monitoring policies when necessary.

Table 4.1 Posterior summaries include the mean, standard deviation (SD), Bias, RMSE for parameters (p, ψ, ω, $\eta_1, \eta_2, \ldots, \eta_8$, ϕ, $V_1^k, V_2^k, \ldots, V_5^k$, $L_1^k, L_2^k, \ldots, L_5^k$)

Parameters	**Mean**	**Bias**	**RMSE**
p	0.49(0.02)	−0.010	0.092
ψ	0.72(0.02)	0.020	0.095
ω	5010(35)	10	36
η_1	0.60(0.01)	−0.013	0.020
η_2	0.13(0.01)	0.001	0.013
η_3	0.10(0.01)	0.002	0.010
η_4	0.05(0.01)	0.001	0.009
η_5	0.05(0.01)	0.001	0.008
η_6	0.03(0.01)	−0.024	0.006
η_7	0.01(0.01)	0.002	0.004
η_8	0.01(0.01)	0.007	0.008
ϕ	0.87(0.01)	0.002	0.006
V_1^k	0.52(0.01)	0.050	0.052
V_2^k	0.23(0.02)	−0.030	0.035
V_3^k	0.20(0.01)	−0.006	0.015
V_4^k	0.03(0.01)	−0.011	0.014
V_5^k	0.02(0.01)	0.003	0.018
L_1^k	0.08(0.001)	0.025	0.026
L_2^k	0.11(0.01)	0.029	0.031
L_3^k	0.17(0.01)	0.032	0.035
L_4^k	0.35(0.03)	0.028	0.001
L_5^k	0.28(0.04)	−0.115	0.123

Table 4.2 ABC-SMC posterior average parameter estimates together with their 95% credible interval.

Parameters							
ψ	ω	η_1	η_2	η_3	η_4	η_5	η_6
0.73(0.67,0.80)	504(438,577)	0.12(0.09,0.14)	0.11(0.08,0.13)	0.12(0.10,0.15)	0.14(0.11,0.16)	0.06(0.04,0.08)	0.10(0.07,0.12)
Parameters							
η_7	η_8	ϕ	V_1^k	V_2^k	V_3^k	V_4^k	V_5^k
0.11(0.09,0.14)	0.24(0.18,0.26)	0.75(0.72,0.78)	0.21(0.18,0.25)	0.38(0.32,0.42)	0.16(0.12,0.20)	0.15(0.11,0.20)	0.10(0.04,0.16)
Parameters							
L_1^k	L_2^k	L_3^k	L_4^k	L_5^k			
0.02(0.01,0.03)	0.09(0.04,0.13)	0.16(0.09,0.21)	0.55(0.44,0.68)	0.18(0.08,0.31)			

Chapter 5

Discussion

Throughout this thesis, we have presented three manuscripts in Chapters 2, 3, and 4. In this chapter, we discuss the main contributions and novelty of the methods described, and we provide ideas and describe avenues for future work that we have identified as important for a deeper understanding and coherent application of the methods outlined in this thesis.

Chapter 2: Hidden Markov models with an unknown number of states and a repulsive prior on the state parameters

In Chapter 2, we developed a Bayesian HMM using RJMCMC and a repulsive prior on the underlying latent state parameters for acoustic and GPS dynamic data. The research was motivated by Pohle et al. (2017), who created an HMM for GPS data. However, their approach had the limitation that the unknown number of states was considered known among a set of candidate state numbers. This approach is highly subjective and can potentially bias the inferred results. A method to overcome this problem and address subsequent concerns when the number of states is treated as completely unknown and random was presented in Chapter 2.

Dynamic data are most frequently modeled using HMMs, as they provide a flexible way to model the observed process conditional on the evolution of the latent state process over time. Usually, when such a model is defined, at least in a Bayesian setting, the state allocation parameters are sampled at each iteration of the MCMC algorithm. This requires substantial computational effort and storage

to sample the allocation parameters, even though we are primarily interested in the state-specific parameters. To avoid this, we employed the algorithm presented in Russo et al. (2022), where the likelihood function calculation is based on the marginalization of the state allocation parameters.

In a Bayesian setting, when treating the number of states as a random parameter, the standard MCMC algorithm cannot accommodate this type of model. This occurs because the dimension of the parameters of the states changes with the variation in the number of states, and the MCMC algorithm does not allow transdimensional moves. We address this issue by employing an RJMCMC algorithm, which is a standard tool for cases where we need to explore different dimensions of the model. To the best of our knowledge in statistical ecology, RJMCMC has only been used for variable selection cases and abundance estimation. However, it has never been used for inferring the number of states in the model. Therefore, we believe that demonstrating how RJMCMC is utilized in the latter case is important.

When treating the number of states as unknown, it is common for redundant states to be introduced, meaning they are identical or similar to already existing states. This is referred to in the literature as overfitting, where additional number of states do not contribute to the inference of the analysis, and may even worsen the interpretation. To overcome this problem, we employ what was introduced in Chapter 2 as the repulsive prior. Such a prior penalizes states that are close together, for instance, states with location means that are in proximity to each other but do not significantly contribute to the inference of the analysis. By employing such a prior, we can regulate the number of states, avoid the introduction of redundant states, and retain only those that can effectively distinguish between the data.

We demonstrated the robustness of our model through a comprehensive simulation study, comparing the model with the use of a repulsive prior to a model with the use of a standard prior across various scenarios. These scenarios included different sample sizes, time points, and overlap between the state distributions. We observed that regardless of the scenario, the introduction of a repulsive prior resulted in slightly better efficiency measurements, such as the KL divergence and the misclassification rate. Hence, this suggests that the use of a repulsive prior would be beneficial to introduce as a standard tool when applying HMM with an unknown number of states. Lastly, we analyzed two real datasets: one for muskox GPS data, presented in Pohle et al. (2017), and another for Cape gannets acoustic data,

presented in Thiebault et al. (2021). In the muskox case, we ran two models: one with a repulsive prior and one with a standard prior, where the latter resulted in a larger number of inferred states. In the former case, we compared our results with those derived in Pohle et al. (2017) and found that we made similar inferences for the underlying states. This indicates that with the use of the repulsive prior, we are able to obtain meaningful results when we allow the number of states to be random instead of fixed, potentially avoiding bias. In the Cape gannet case, inference was more complicated as we had to work with extracted acoustic data rather than directly inferable quantities. Similarly, we ran two models: one with a repulsive prior and one with a standard prior. We found that the repulsive prior managed to capture the underlying states as expected, based on the details presented in Thiebault et al. (2021), whereas the standard prior led to an excessive number of underlying states.

In Chapter 2, we demonstrated that we developed a flexible and intuitive model that can help practitioners better understand the behavior of species that serve as indicators of climate change. This model can guide policy-making, as it draws on existing knowledge of species' seasonal behaviors. Any significant deviations from these norms may signal the need for further monitoring and policy adjustments. Nevertheless, there is still a lot of research to be conducted on this topic such as: 1) Penalization can be applied to more than one parameter, examining how this alters the inference. In this case, states that are considered close for one state parameter might be considered far for another state parameter and this imbalance behavior might give better inference. For example, in the second case study of Chapter 2, we could introduce repulsion in the covariance matrices. This would lead the model to identify states where the correlations between the dimensions of the Multivariate Gaussian distributions differ significantly between the sates. 2) Also, as we have seen in the second case study there are cases where there is uncertainty between consecutive time points, i.e. within 1.4 second moving from flying to diving and then back to flying which is unrealistic. Hence, the presented HMM model can be extended to a second or third order Markov chain instead of first order and examine the changes on the classification performance. 3) Extend the model to include time-specific transition matrices to allow for dynamic changes in transition probabilities. This extension might better capture the true underlying evolution of the states. 4) Enhance the information of the state parameters by incorporating covariates.

Chapter 3: A Pólya Tree modelling framework for batch-mark data

In Chapter 3, we presented the development and application of a Pólya Tree prior model for Batch Mark data. The creation of such a model was motivated by Diana et al. (2023), where they introduced the idea of the Pólya Tree prior for ecological data. They developed a model for Capture-Recapture data; however, the acquisition of such data is not always possible due to time, financial, or species-specific constraints. In cases such as those described in Chapter 3, Batch Mark data is the way forward.

We developed a Pólya Tree prior model for Batch Mark data, which is a Bayesian nonparametric model. It is highly flexible since it can accommodate parsimonious models through the use of the idea of Replicated Pólya Tree prior, as described in Chapter 3. Furthermore, an essential benefit is that it scales with the number of sampling occasions instead of the number of individuals, as frequently encountered in most models in ecological literature. Moreover, Batch Mark data are an aggregated version of Capture-Recapture data, wherein we integrate out individual marking. Therefore, the Pólya Tree prior, which is built on the latent counts of individuals without requiring any individual manipulation, is a natural way of modeling such data. Hence, it is evident that the Pólya Tree prior is naturally the most suitable model for inferring such data.

Initially, we developed a model for the data used in Cowen et al. (2017). However, during their analysis, they ignored a portion of data corresponding to removed individuals per sampling occasion. In our thorough simulation study, we demonstrated that ignoring such data in the analysis can introduce bias, especially in parameters related to population sizes. By disregarding the information that additional individuals were present during the survey but eventually removed, we risk distorting the results. The proposed Pólya Tree prior model can extend to such data quite easily, since we do not have to make any significant changes to the model. To demonstrate this, we developed a model as described in Chapter 3, incorporating all available data from Cowen et al. (2017), and compared the results with theirs. The output of our analysis confirmed our simulation study expectations. We found that our model estimated more individuals during the survey compared to theirs. However, parameters such as survival and entry were estimated fairly similarly for both models. This underscores how ignoring removed individuals from the model can introduce bias into population size estimations.

To further demonstrate the versatility of the Pólya Tree prior across various data schemes, we developed a model for Robust Design Batch Mark data, inspired by Zhang et al. (2023). In their analysis, they cleverly approximate the likelihood function using Saddlepoint Approximation to infer the latent capture history of individuals. However, such an approach requires approximation instead of using the true likelihood function, and it involves an additional step of inferring the latent capture history, which ideally should be avoided given that our original data are Batch Mark. On the contrary, our Pólya Tree prior model can easily accommodate such Robust Design Batch Mark data, by changing only the observation process while keeping the latent process the same as for the non Robust Design Batch Mark data. Through a thorough simulation study, we demonstrated the robustness of our model, and then applied it to the data described in Zhang et al. (2023). From Chapter 3, it is evident that we obtained similar estimates to theirs for the parameters of the model. However, in our case, the estimates are less uncertain because we used the true likelihood function instead of an approximation, which reduces the uncertainty of the estimations by avoiding the inaccuracies introduced by approximation.

Lastly, the analysis of Batch Mark data is a growing area of interest, and much research still needs to be conducted. We developed a new model that helps ecologists and policymakers understand changes in population dynamics. It can identify increases or decreases in species populations, monitor changes in species survival, and protect those that might be rapidly depleting or growing too quickly, potentially threatening ecosystem balance. We have identified quite a few avenues for future research which we discuss here: 1) With regard to the Pólya Tree prior, it would be interesting to incorporate individual or batch specific covariates to the latent or the observation process. 2) In terms of the Pólya Tree prior, or would be useful to create ways of model comparison between time-dependent and age-dependent models, amongst others, since the Pólya Tree prior gives the flexibility to derive such model selection comparisons in closed mathematical formulas. 3) Since Batch Mark data are an aggregated version of Capture-Recapture data, we inevitably lose information. Therefore, it would be really interesting to quantify the benefits of incorporating Capture-Recapture data with Batch Mark data. 4) It would be really helpful to investigate the information loss from Capture-Recapture to Batch Mark data. We have conducted some initial investigations and found that Batch Mark and Capture-Recapture data are equivalent within closed populations. This finding could prove to be

financially and time-effective. For example, instead of using Robust Design Capture-Recapture, maybe using Robust Design Batch Mark could be considered, though this is an area that still requires further research.

Chapter 4: An approximate Bayesian computation approach for modelling batch-mark data on populations exhibiting temporary emigration

In Chapter 4, we extended the model developed in Chapter 3, for Batch Mark temporary emigration data, using a likelihood-free framework. There are many cases in ecology where the assumption of a permanent emigration is violated due to poorly defined surveys or species' breeding seasons. In such cases, if we do not account for these behaviors, our model may be biased. For Batch Mark data, a model that accounts for temporary emigration has not yet been developed. Therefore, it is highly important and urgent for such a model to be defined.

We developed a temporary emigration model for Batch Mark data in Chapter 4. However, the aggregated nature and the lack of individual information, as seen in cases of Capture-Recapture data, made inference on such data challenging. The problem arises from the fact that individuals can be present in multiple periods. However, in a Pólya tree prior framework, there is no possible way to distinguish and keep track of those individuals, as the Pólya tree prior operates at the sampling occasion level, with the use of grid probabilities, rather than the individual level. Hence, we are not able to write a tractable likelihood function.

To overcome this issue, we can work within a likelihood-free framework. In Bayesian statistics, a common tool used is ABC. However, in our case, it was proved to be inefficient. Therefore, we utilized ABC-SMC, which is more robust in identifying and gradually sampling from regions of higher posterior mass. We are unable to formulate the likelihood function for temporary emigration Batch Mark data. However, leveraging the ABC-SMC, we can simulate parameters and retain those that generate Batch Mark data close to the observed Batch Mark data. Sampling Batch Mark data is straightforward, fulfilling the primary requirement for the ABC-SMC algorithm. Since we cannot directly sample Batch Mark data, which would imply having a likelihood function, we resort to sampling them as Capture-Recapture data, a computationally efficient and easily implementable

approach. Subsequently, we aggregate these sampled data to form Batch Mark data. In this case, we can efficiently sample from the posterior distribution of the parameters for the temporary emigration Batch Mark data, indirectly.

To demonstrate the robustness of our model, we conducted a simulation study and identified that despite being an approximation, our model successfully captures the underlying true values that generated the simulation data. Additionally, we analyzed data collected for the alpine common toad *Bufo bufo*. We successfully inferred the breeding probability and population abundance, a task previously unattempted due to the absence of a temporary emigration model for Batch Mark data.

Overall, the development of this likelihood-free model has been an essential addition to the Batch Mark literature. This provided practitioners with tools to make inferences about population dynamics during breeding seasons, often based on more realistic assumptions, leading to a better understanding of long-term changes in species. In Chapter 4, we demonstrated the ease of conception and implementation, making it an important tool for the future development of Batch Mark sampling designs. Future work on this subject would be: 1) Integrate capture-recapture data, as its integration in the simulation step of the ABC-SMC is relatively straightforward. With even a limited amount of additional capture-recapture data, we could considerably reduce uncertainty in parameter estimation. This is because knowing exactly which years and sampling occasions individuals were captured would provide more precise information for estimating breeding and survival probabilities. 2) Explore different methods for choosing the threshold parameter, such as employing a full Bayesian framework where it is assumed to be random. This is crucial because the accuracy of the approximation to the true posterior distribution depends on how close the threshold is to zero. 3) Allow for the incorporation of batch covariates.

References

Raffaele Argiento and Maria De Iorio. Is infinity that far? a Bayesian nonparametric perspective of finite mixture models. The Annals of Statistics, 50(5):2641–2663, 2022.

Francesco Bartolucci and Silvia Pandolfi. Bayesian inference for a class of latent Markov models for categorical longitudinal data. arXiv preprint arXiv:1101.0391, 2011.

Francesco Bartolucci, Alessio Farcomeni, and Fulvia Pennoni. Latent Markov models for longitudinal data. CRC Press, 2013.

Mario Beraha, Raffaele Argiento, Jesper Møller, and Alessandra Guglielmi. MCMC computations for Bayesian mixture models using repulsive point processes. Journal of Computational and Graphical Statistics, 31(2):422–435, 2022.

Johannes Berkhof, Iven Van Mechelen, and Andrew Gelman. A Bayesian approach to the selection and testing of mixture models. Statistica Sinica, pages 423–442, 2003.

Christophe Biernacki, Gilles Celeux, and Gérard Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. IEEE transactions on pattern analysis and machine intelligence, 22(7):719–725, 2000.

Tomas Bird, Jarod Lyon, Simon Nicol, Michael McCarthy, and Richard Barker. Estimating population size in the presence of temporary migration using a joint analysis of telemetry and capture–recapture data. Methods in Ecology and Evolution, 5(7):615–625, 2014.

Fernando V Bonassi and Mike West. Sequential monte carlo with adaptive weights for approximate bayesian computation. 2015.

Ronald W. Butler. Saddlepoint Approximations with Applications. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2007.

Olivier Cappé, Christian P Robert, and Tobias Rydén. Reversible jump, birth-and-death and more general continuous time Markov chain Monte Carlo samplers. Journal of the Royal Statistical Society Series B: Statistical Methodology, 65(3):679–700, 2003.

Olivier Cappé, Eric Moulines, and Tobias Rydén. Inference in hidden Markov models. In Proceedings of EUSFLAT conference, pages 14–16, 2009.

Carl Chalmers, Paul Fergus, S Wich, and SN Longmore. Modelling animal biodiversity using acoustic monitoring and deep learning. In 2021 International Joint Conference on Neural Networks (IJCNN), pages 1–7. IEEE, 2021.

Jinkui Cheng, Yuehua Sun, and Liqiang Ji. A call-independent and automatic acoustic system for the individual recognition of animals: A novel model using four passerines. Pattern Recognition, 43 (11):3846–3852, 2010.

Siddhartha Chib. Calculating posterior distributions and modal estimates in Markov mixture models. Journal of Econometrics, 75(1):79–97, 1996.

Richard M Cormack. Estimates of survival from the sighting of marked animals. Biometrika, 51(3/4): 429–438, 1964.

Laura LE Cowen, Panagiotis Besbeas, Byron JT Morgan, and Carl J Schwarz. A comparison of abundance estimates from extended batch-marking and Jolly–Seber-type experiments. Ecology and Evolution, 4(2):210–218, 2014.

Laura LE Cowen, Panagiotis Besbeas, Byron JT Morgan, and Carl J Schwarz. Hidden Markov models for extended batch data. Biometrics, 73(4):1321–1331, 2017.

Jenna R Davidson, Rusdiyah Sudirman, Isra Wahid, Robert N Baskin, Hajar Hasan, Andi Muhammad Arfah, Nirwana Nur, Muhammad Yusuf Hidayat, Din Syafruddin, and Neil F Lobo. Mark-release-recapture studies reveal preferred spatial and temporal behaviors of Anopheles barbirostris in West Sulawesi, Indonesia. Parasites & vectors, 12:1–11, 2019.

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. Sequential monte carlo samplers. Journal of the Royal Statistical Society Series B: Statistical Methodology, 68(3):411–436, 2006.

Pierre Del Moral, Arnaud Doucet, and Ajay Jasra. An adaptive sequential monte carlo method for approximate bayesian computation. Statistics and computing, 22:1009–1020, 2012.

Alex Diana, Jim Griffin, and Eleni Matechou. A Polya tree based model for unmarked individuals in an open wildlife population. In Bayesian Statistics and New Generations: BAYSM 2018, Warwick, UK, July 2-3 Selected Contributions, pages 3–11. Springer, 2019.

Alex Diana, Eleni Matechou, Jim Griffin, Todd Arnold, Simone Tenan, and Stefano Volponi. A general modeling framework for open wildlife populations based on the polya tree prior. Biometrics, 2022.

Alex Diana, Eleni Matechou, Jim Griffin, Todd Arnold, Simone Tenan, and Stefano Volponi. A general modeling framework for open wildlife populations based on the Polya tree prior. Biometrics, 79(3): 2171–2183, 2023.

Jason C Doll, Chris J Wood, David W Goodfred, and Jacob M Rash. Incorporating batch mark–recapture data into an integrated population model of brown trout. North American Journal of Fisheries Management, 41(5):1390–1407, 2021.

Leo L Duan and David B Dunson. Bayesian distance clustering. arXiv preprint arXiv:1810.08537, 2018.

Alessio Farcomeni. Penalized estimation in latent Markov models, with application to monitoring serum calcium levels in end-stage kidney insufficiency. Biometrical Journal, 59(5):1035–1046, 2017.

Alan E Gelfand and Adrian FM Smith. Sampling-based approaches to calculating marginal densities. Journal of the American statistical association, 85(410):398–409, 1990.

John Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Technical report, Federal Reserve Bank of Minneapolis, 1991.

Olivier Gimenez, Simon J Bonner, Ruth King, Richard A Parker, Stephen P Brooks, Lara E Jamieson, Vladimir Grosbois, Byron JT Morgan, and Len Thomas. Winbugs for population ecologists: Bayesian modeling using Markov chain Monte Carlo methods. Modeling demographic processes in marked populations, pages 883–915, 2009.

Richard Glennie, Timo Adam, Vianey Leos-Barajas, Théo Michelot, Theoni Photopoulou, and Brett T McClintock. Hidden Markov models: Pitfalls and opportunities in ecology. Methods in Ecology and Evolution, 14(1):43–56, 2023.

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

W Keith Hastings. Monte carlo sampling methods using markov chains and their applications. 1970.

Chris C Holmes, François Caron, Jim E Griffin, and David A Stephens. Two-sample Bayesian nonparametric hypothesis testing. 2015.

Tao Huang, Heng Peng, and Kun Zhang. Model selection for gaussian mixture models. Statistica Sinica, pages 147–169, 2017.

Richard Huggins, Yan Wang, and Joanne Kearns. Analysis of an extended batch marking experiment using estimating equations. Journal of agricultural, biological, and environmental statistics, 15: 279–289, 2010.

George M Jolly. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. Biometrika, 52(1/2):225–247, 1965.

William L Kendall and Rhema Bjorkland. Using open robust design models to estimate temporary emigration from capture—recapture data. Biometrics, 57(4):1113–1122, 2001.

William L Kendall and Kenneth H Pollock. The robust design in capture-recapture studies: a review and evaluation by Monte Carlo simulation. Wildlife 2001: populations, pages 31–43, 1992.

William L Kendall, James D Nichols, and James E Hines. Estimating temporary emigration using capture–recapture data with pollock's robust design. Ecology, 78(2):563–578, 1997.

R. King. Statistical ecology. Annual Review of Statistics and Its Application, 1:401–426, 2014.

Ruth King and Rachel McCrea. Capture–Recapture methods and models: estimating population size. In Handbook of statistics, volume 40, pages 33–83. Elsevier, 2019.

José J Lahoz-Monfort and Michael JL Magrath. A comprehensive overview of technologies for species and habitat monitoring and conservation. BioScience, 71(10):1038–1062, 2021.

Roland Langrock. Flexible latent-state modelling of Old Faithful's eruption inter-arrival times in 2009. Australian & New Zealand Journal of Statistics, 54(3):261–279, 2012.

Michael Lavine. Some aspects of polya tree distributions for statistical modelling. The annals of statistics, pages 1222–1235, 1992.

Michael Lavine. More aspects of polya tree distributions for statistical modelling. The Annals of Statistics, 22(3):1161–1176, 1994.

Louisiane Lemaire, Flora Jay, I Lee, Katalin Csilléry, Michael GB Blum, et al. Goodness-of-fit statistics for approximate bayesian computation. arXiv preprint arXiv:1601.04096, 2016.

Brian G Leroux and Martin L Puterman. Maximum-penalized-likelihood estimation for independent and Markov-dependent mixture models. Biometrics, pages 545–558, 1992.

Jean-Michel Marin, Pierre Pudlo, Christian P Robert, and Robin J Ryder. Approximate bayesian computational methods. Statistics and computing, 22(6):1167–1180, 2012.

Enzo Marinari and Giorgio Parisi. Simulated tempering: a new monte carlo scheme. Europhysics letters, 19(6):451, 1992.

Eleni Matechou and Raffaele Argiento. Capture-recapture models with heterogeneous temporary emigration. Journal of the American Statistical Association, 118(541):56–69, 2023.

Eleni Matechou and François Caron. Modelling individual migration patterns using a bayesian nonparametric approach for capture–recapture data. 2017.

Eleni Matechou, Rachel S McCrea, Byron JT Morgan, Darryn J Nash, and Richard A Griffiths. Open models for removal data. The Annals of Applied Statistics, 10(3):1572–1589, 2016.

Diethart Matthies, Ingo Bräuer, Wiebke Maibom, and Teja Tscharntke. Population size and the risk of local extinction: empirical evidence from rare plants. Oikos, 105(3):481–488, 2004.

Brett T McClintock, Roland Langrock, Olivier Gimenez, Emmanuelle Cam, David L Borchers, Richard Glennie, and Toby A Patterson. Uncovering ecological state dynamics with hidden markov models. arXiv preprint arXiv:2002.10497, 2020. URL https://onlinelibrary.wiley.com/doi/10.1111/ele.13610.

Brenda McComb, Benjamin Zuckerberg, David Vesely, and Christopher Jordan. Monitoring animal populations and their habitats: a practitioner's guide. CRC Press, 2010.

Rachel S McCrea and Byron JT Morgan. Analysis of capture-recapture data. CRC Press, 2014.

Geoffrey J McLachlan, Sharon X Lee, and Suren I Rathnayake. Finite mixture models. Annual review of statistics and its application, 6:355–378, 2019.

Nicholas Metropolis, Arianna W Rosenbluth, Marshall N Rosenbluth, Augusta H Teller, and Edward Teller. Equation of state calculations by fast computing machines. The journal of chemical physics, 21(6):1087–1092, 1953.

Jesper Møller and Michael Sørensen. Statistical analysis of a spatial birth-and-death process model with a view to modelling linear dune fields. Scandinavian journal of statistics, pages 1–19, 1994.

Jesper Moller and Rasmus Plenge Waagepetersen. Statistical inference and simulation for spatial point processes. CRC press, 2003.

Caroline Moussy, Ian J Burfield, PJ Stephenson, Arabella FE Newton, Stuart HM Butchart, William J Sutherland, Richard D Gregory, Louise McRae, Philip Bubb, Ignacio Roesler, et al. A quantitative global review of species population monitoring. Conservation Biology, 36(1):e13721, 2022.

Iain Murray, Zoubin Ghahramani, and David MacKay. MCMC for doubly-intractable distributions. arXiv preprint arXiv:1206.6848, 2012.

Abhinav Natarajan, Maria De Iorio, Andreas Heinecke, Emanuel Mayer, and Simon Glenn. Cohesion and repulsion in Bayesian distance clustering. Journal of the American Statistical Association, pages 1–11, 2023.

Sam Nicol, Marie-Josée Cros, Nathalie Peyrard, Régis Sabbadin, Ronan Trépos, Richard A Fuller, and Bradley K Woodworth. Flywaynet: A hidden semi-Markov model for inferring the structure of migratory bird networks from count data. Methods in Ecology and Evolution, 14(1):265–279, 2023.

Juan J Noda, Carlos M Travieso-González, David Sanchez-Rodriguez, and Jesús B Alonso-Hernández. Acoustic classification of singing insects based on MFCC/LFCC fusion. Applied Sciences, 9(19): 4097, 2019.

Massimiliano Pastore and Antonio Calcagnì. Measuring distribution similarities between samples: a distribution-free overlapping index. Frontiers in psychology, 10:455421, 2019.

Toby A Patterson, Alison Parton, Roland Langrock, Paul G Blackwell, Len Thomas, and Ruth King. Statistical modelling of individual animal movement: an overview of key methods and a discussion of practical challenges. AStA Advances in Statistical Analysis, 101:399–438, 2017.

Francesca Petralia, Vinayak Rao, and David Dunson. Repulsive mixtures. Advances in neural information processing systems, 25, 2012.

Jennifer Pohle, Roland Langrock, Floris M Van Beest, and Niels Martin Schmidt. Selecting the number of states in hidden Markov models: pragmatic solutions illustrated using animal movement. Journal of Agricultural, Biological and Environmental Statistics, 22:270–293, 2017.

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

Valentin Popov, Roland Langrock, Stacy L DeRuiter, and Fleur Visser. An analysis of pilot whale vocalization activity using hidden Markov models. The Journal of the Acoustical Society of America, 141(1):159–171, 2017.

José J Quinlan, Fernando A Quintana, and Garritt L Page. On a class of repulsive mixture models. Test, 30:445–461, 2021.

Lawrence Rabiner and Biinghwang Juang. An introduction to hidden markov models. ieee assp magazine, 3(1):4–16, 1986.

Lawrence R Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2):257–286, 1989.

Angel David Pedroza Ramirez, Jose Ismael de la Rosa Vargas, Rogelio Rosas Valdez, and Aldonso Becerra. A comparative between mel frequency cepstral coefficients (MFCC) and inverse mel frequency cepstral coefficients (IMFCC) features for an automatic bird species recognition system. In 2018 IEEE Latin American Conference on Computational Intelligence (LA-CCI), pages 1–4. IEEE, 2018.

Douglas A Reynolds et al. Gaussian mixture models. Encyclopedia of biometrics, 741(659-663), 2009.

Christian P Robert and M Titterington. Resampling schemes for hidden Markov models and their application for maximum likelihood estimation. In Statistical Computing, volume 8, pages 145–158, 1998.

Christian P Robert, Gilles Celeux, and Jean Diebolt. Bayesian estimation of hidden Markov chains: a stochastic implementation. Statistics & Probability Letters, 16(1):77–83, 1993.

Christian P Robert, Tobias Ryden, and David M Titterington. Bayesian inference in hidden Markov models through the reversible jump Markov chain Monte Carlo method. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 62(1):57–75, 2000.

Christian P Robert, George Casella, Christian P Robert, and George Casella. The Metropolis—Hastings algorithm. Monte Carlo statistical methods, pages 267–320, 2004.

Emma Rosser, Samantha A Willden, and Gregory M Loeb. Effects of Smartwater, a fluorescent mark, on the dispersal, behavior, and biocontrol efficacy of Phytoseiulus persimilis. Experimental and Applied Acarology, 87(2-3):163–174, 2022.

J Andrew Royle, Sarah J Converse, and William A Link. Data augmentation for hierarchical capture-recapture models. arXiv preprint arXiv:1211.5706, 2012.

Alfonso Russo, Alessio Farcomeni, Maria Grazia Pittau, and Roberto Zelli. Covariate-modulated rectangular latent markov models with an unknown number of regime profiles. Statistical Modelling, page 1471082X221127732, 2022.

Tobias Rydén, Timo Teräsvirta, and Stefan Åsbrink. Stylized facts of daily return series and the hidden Markov model. Journal of applied econometrics, 13(3):217–244, 1998.

Michael Schaub, Olivier Gimenez, Benedikt R Schmidt, and Roger Pradel. Estimating survival and temporary emigration in the multistate capture–recapture framework. Ecology, 85(8):2107–2113, 2004.

Niels M Schmidt, Floris M van Beest, Jesper B Mosbacher, Mikkel Stelvig, Lars H Hansen, Jacob Nabe-Nielsen, and Carsten Gr ndahl. Ungulate movement in an extreme seasonal environment: Year-round movement patterns of high-arctic muskoxen. Wildlife Biology, 22(6):253–267, 2016.

Carl James Schwarz and A Neil Arnason. A general methodology for the analysis of capture-recapture experiments in open populations. Biometrics, pages 860–873, 1996.

Carl James Schwarz and Wayne T Stobo. Estimating temporary migration using the robust design. Biometrics, pages 178–194, 1997.

George AF Seber. A note on the multiple-recapture census. Biometrika, 52(1/2):249–259, 1965.

George AF Seber and Matthew R Schofield. Capture-recapture: Parameter Estimation for Open Animal Populations. Springer, 2019.

Yunne-Jai Shin, Guy F Midgley, Emma RM Archer, Almut Arneth, David KA Barnes, Lena Chan, Shizuka Hashimoto, Ove Hoegh-Guldberg, Gregory Insarov, Paul Leadley, et al. Actions to halt biodiversity loss generally benefit the climate. Global change biology, 28(9):2846–2874, 2022.

Scott A Sisson, Yanan Fan, and Mark M Tanaka. Sequential monte carlo without likelihoods. Proceedings of the National Academy of Sciences, 104(6):1760–1765, 2007.

Scott A Sisson, Yanan Fan, and Mark Beaumont. Handbook of approximate Bayesian computation. CRC Press, 2018.

Jeffrey A Spendelow and James D Nichols. Annual survival rates of breeding adult roseate terns. The Auk, 106(3):367–374, 1989.

Luigi Spezia. Bayesian variable selection in non-homogeneous hidden Markov models through an evolutionary Monte Carlo method. Computational Statistics & Data Analysis, 143:106840, 2020.

Matthew Stephens and D Phil. Bayesian methods for mixtures of normal distributions. 1997.

David J Strauss. A model for clustering. Biometrika, 62(2):467–475, 1975.

Andréa Thiebault, Chloé Huetz, Pierre Pistorius, Thierry Aubin, and Isabelle Charrier. Animal-borne acoustic data alone can provide high accuracy classification of activity budgets. Animal Biotelemetry, 9:1–16, 2021.

Hoang Trang, Tran Hoang Loc, and Huynh Bui Hoang Nam. Proposed combination of PCA and MFCC feature extraction in speech recognition system. In 2014 international conference on advanced technologies for communications (ATC 2014), pages 697–702. IEEE, 2014.

Laura Vavassori, Adam Saddler, and Pie Müller. Active dispersal of Aedes albopictus: a mark-release-recapture study using self-marking units. Parasites & vectors, 12(1):1–14, 2019.

Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved r for assessing convergence of mcmc (with discussion). Bayesian analysis, 16(2):667–718, 2021.

Richard David Wilkinson. Approximate bayesian computation (abc) gives exact results under the assumption of model error. Statistical applications in genetics and molecular biology, 12(2):129–141, 2013.

Wei Zhang, Simon J Bonner, and Rachel S McCrea. Latent multinomial models for extended batch-mark data. Biometrics, 79(3):2732–2742, 2023.

Zhihua Zhang, Kap Luk Chan, Yiming Wu, and Chibiao Chen. Learning a multivariate gaussian mixture model with the reversible jump MCMC algorithm. Statistics and Computing, 14:343–355, 2004.

Walter Zucchini and Iain L MacDonald. Hidden Markov models for time series: an introduction using R. Chapman and Hall/CRC, 2009.

Appendix A

Hidden Markov models with an unknown number of states and a repulsive prior on the state parameters

A.1 Birth and Death Algorithm

A brief description of the Birth hand Death algorithm is

1. We initialize a point pattern θ_1, choose birth and death probabilities q_{birth} and q_{death}, specify the number of iterations M, and proposal parameters η.

2. At iteration it, if the cardinality of θ_{it} is one, then with probability 1 we choose to give birth to a new point and add it to θ_{it} and form the θ^*, sampled from a proposal distribution with proposal parameters η. In any other case, we either propose to give birth to a new point and add it to θ_{it} to form θ^*, sampled from a proposal distribution with probability q_{birth} or give death to a randomly uniformly chosen point from θ_{it} with probability q_{death} and form θ^*.

3. The acceptance ratio is
$$A = \frac{g(\underline{\theta^*}|\,|\underline{\theta^*}|,\xi_*,a,d)}{g(\underline{\theta_{it}}|\,|\underline{\theta_{it}}|,\xi_*,a,d)}\frac{q(\theta_i \to \theta^*)}{q(\theta^* \to \theta_i)}$$
where $q(\cdot \to \cdot)$ corresponds to the proposal distribution times the death or birth probability.

4. We repeat this process for M iterations. The resulting $\underline{\theta}_{\text{aux}} = \underline{\theta}_M$.

A.2 Penalty & Threshold

The density over distances r is expected to have more than one mode, as the existence of only one mode suggests the presence of only one mixture component. Local minima of $p(r)$ can be likened to "valleys" between "peaks" of higher density, indicating distances where fewer observations are present. By choosing the minimum over the local minimum for $p(r)$, we select a distance that is not too large to affect the density estimation severely. Examples, can be found in Section A.3 of Appendix A.

The penalty a as explained in Beraha et al. (2022) is chosen as follows: suppose we have two components with location parameters θ_h and $\theta_{h'}$ such that their distance is $\left\|\theta_h - \theta_{h'}\right\| \leq d$, while the pairwise distances between $\theta_h, \theta_{h'}$ and the rest of the component parameters are larger than d. Hence, the conditional likelihood of the component θ_h is proportional to the penalty a (since we have only one pair of distance less than the threshold d) times the likelihood function for the observations assigned to the hth mixture component, denoted as F_h. Therefore, $p(\theta_h) \propto aF_h$. Now, in the case where the cardinality of the cluster h is small, we want the penalization to determine whether we keep the mixture component h or not. Therefore, a has to be chosen such that aF_h is small. In that case we define what we assume of being a small cluster size n^*, and then as they mention in Beraha et al. (2022) we take a "guess" of the value of F_h denoted as k_s. Then an estimate of $a = \exp(-n^* \log(k_s))$. This choice of a regulates how small aF_h is going to be since, we can rewrite it as $\exp(-n^* \log(k_s))F_h = k_s^{-n^*} F_h = F_h / k_s^{n^*}$ making the conditional likelihood of the component θ_h sufficiently small, hence in that case the repulsion prevails.

A.3 Threshold

To understand the derivation of the threshold d, we will provide a few examples. The first example is a mixture of two normal distributions with means 0 and 4, and standard deviations 1, respectively, each with mixture weights of 0.5. In Figure A.1 we display the kernel density estimate for the observations of the mixture, and the kernel density estimate of the distances between the observations. From Figure A.1, we observe that the kernel density estimation of the distances has only one local

minimum, corresponding to the value of 2.8304, which is an ideal threshold. Since the true means of the normal distributions are at 0 and 4, which have larger distance than 2.8304, hence mixture components with smaller distance will be penalized. The second example corresponds to a mixture of three normal distributions with means -5, 0, and 5, and a standard deviation equal to 1, respectively. Each distribution has a mixture weight of $1/3$. In Figure A.2, we display the kernel density estimate for the observations of the mixture and the kernel density estimate of the distances between the observations. As we can see, there are two local minimum. However, we consider the minimum local minimum indicated with a red dot, corresponding to the value 2.7279. The other local minimum has a value of 7.9296. Choosing the latter one would severely affect the estimation of the true mixture density of the observations, as the mixture components have distances less than 7.9296. Therefore, we choose the minimum local minimum to avoid overpenalization of the mixture components.

A.4 Simulations

We conducted a simulation study varying the sample size $n = \{50, 100\}$ and number of time points $T = \{5, 10\}$ for four cases ofoverlap between consecutive states distributions equal to 3%, 9%, 33% and 55% or equivalently this can be viewed as 5%, 15%, 50%, and 75% overall overlap. The results were averaged across 100 replications. The Normal distributions have mean $(-10, -5, 0, 5, 10)$ and standard deviations $\sigma_1 = \sigma_2 = ... = \sigma_5$ where for each degree of overlap are 1.1408, 1.4726, 2.5709 and 4.2319. The initial probability distribution π and transition probability matrix P were chosen to have all elements equal to 1/5, ensuring equal state sizes across all time points, allowing us to focus on the effect of state overlap on inference. The corresponding mixture in each case are displayed in Figure A.3.

To begin with, we provide details about the KL divergence statistics utilized in our simulation study. At each time point t, we have access to the true distribution p_t^0, from which we generated our observations. For a replication r we compute the KL divergence for each time point t and posterior sample $l = 1, 2, ..., L$, as we have estimated the posterior density $p_{l,t,r}$ as $\mathrm{KL}_t^{l,r}(p_t^0|p_{l,t,r}) = \int p_t^0(x) log \frac{p_t^0(x)}{p_{l,t,r}(x)}$, Subsequently, we average across posterior samples to obtain $\bar{\mathrm{KL}}_t^r = \frac{1}{L}\sum_{l=1}^{L} \mathrm{KL}_t^{l,r}(p_t^0|p_{l,t,r})$ and across time points to acquire $\bar{\mathrm{KL}}^r = \frac{1}{T}\sum_{t=1}^{T} \bar{\mathrm{KL}}_t^r$ and across replications $\bar{\mathrm{KL}} = \frac{1}{100}\sum_{r=1}^{100} \bar{\mathrm{KL}}^r$ for which we

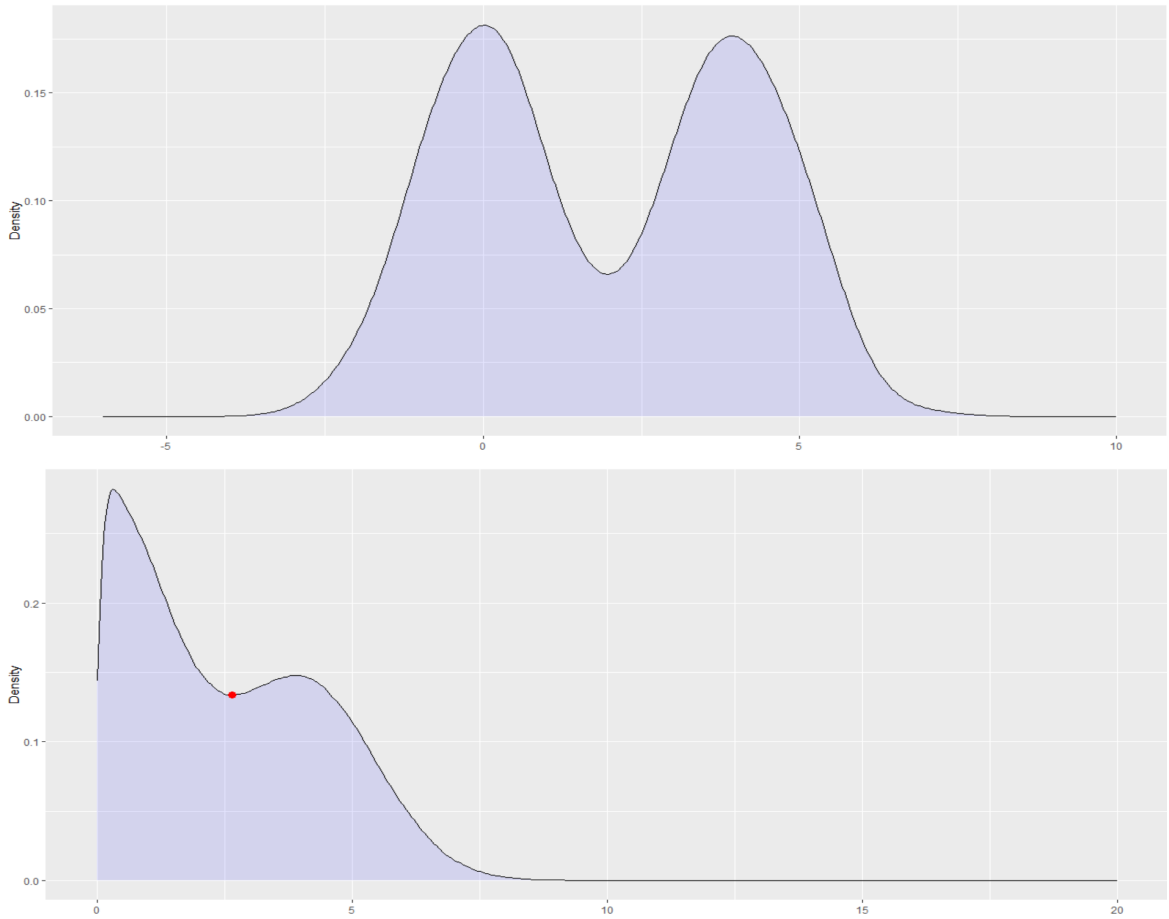

Fig. A.1 The top row corresponds to kernel density estimation of the observations, while the bottom row corresponds to the kernel density estimate of the distances. The red dot indicates the local minimum of the kernel density estimate of the distances..

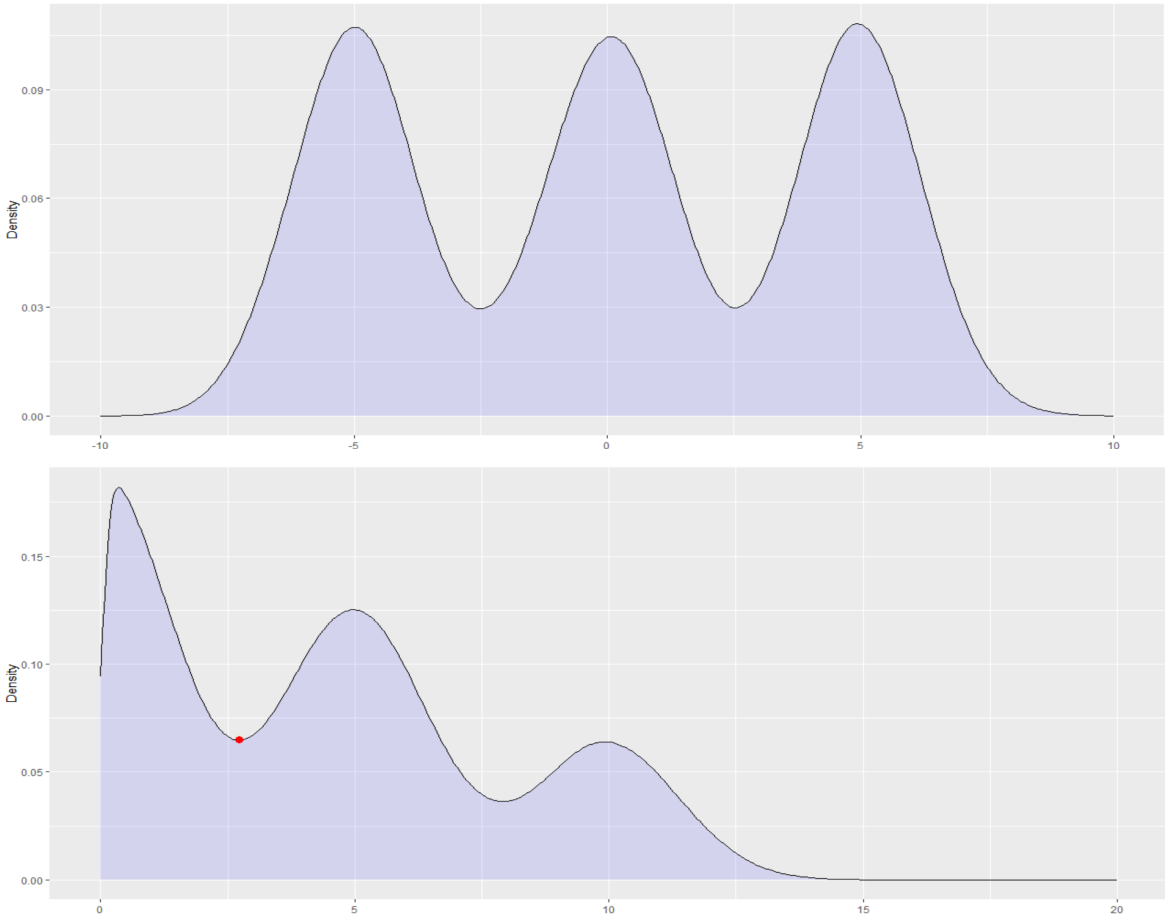

Fig. A.2 The top row corresponds to kernel density estimation of the observations, while the bottom row corresponds to the kernel density estimate of the distances. The red dot indicates the minimum of the local minimum of the kernel density estimate of the distances..

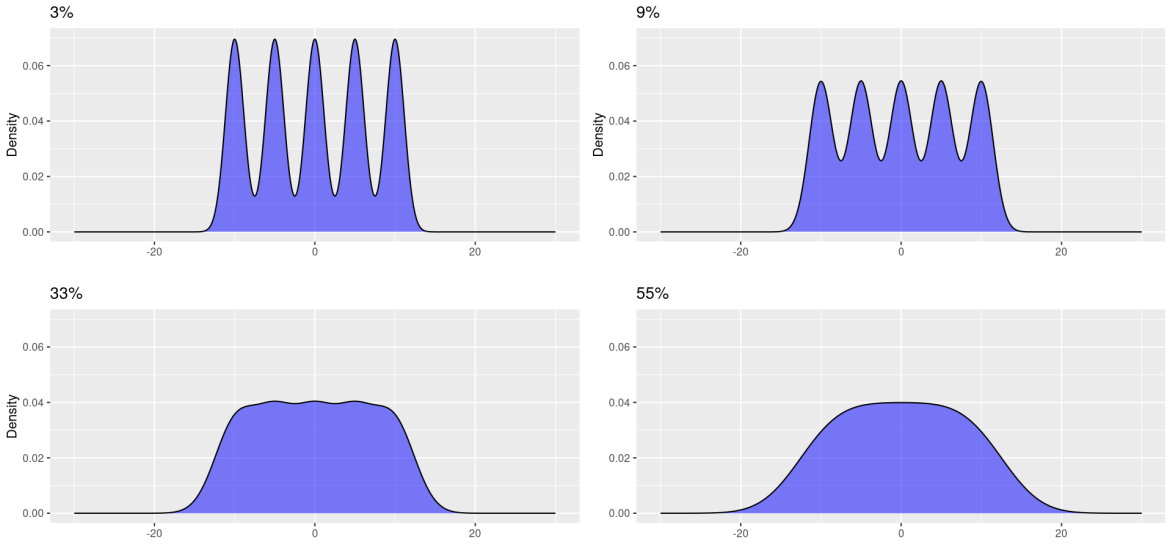

Fig. A.3 Density mixture distributions under the different consecutive overlaps of 3%, 9%, 33%, and 55%.

have calculated also their 95% credible intervals. Moreover, for the misclassification statistics we define the true similarity matrix for time point t and replication r, $S_{t,r}$ of dimensions $n \times n$, defined such that entries where (i, j) belong to the same component take the value 1, otherwise zero for time point t, for all $i, j = 1, 2, ..., n$ and we compare it with the posterior sample similarity matrix $\hat{S_{l,t,r}}$ for each posterior sample l, each time point t and replication r. We report the misclassification error as described in Petralia et al. (2012) averaged across time points for replication r. $\bar{MS}_r = \frac{1}{T}\sum_{t=1}^{T}\sum_{l=1}^{L}\frac{1}{\frac{n(n-1)}{2}}\sum_{i=1}^{n}\sum_{j=i+1}^{n} 1(\hat{S_{l,t,r}}(i,j) \neq S_{t,r}(i,j))$. Then we average across replications and derive $\bar{MS} = \frac{1}{100}\sum_{r=1}^{100}\bar{MS}_r$ alongside with their 95% credible intervals.

We employed RJMCMC for 10,000 iterations, of which we discarded the first 1000 as burn-in. To ensure identifiability, we use the ordering constraint $\mu_i \leq \mu_{i+1}$, for $i = 1, 2, ..., N$. For each scenario, we replicated and averaged our results over 100 iterations. Table A.1 displays the mean and 95% credible interval of the KL divergence and misclassification error for the independent (ID) and repulsive (RP) priors for the penalty case $n^*_{2.5}$.

Next, we display the simulation results for the penalty case n^*_5 for the mean and 95% credible interval of the KL divergence and misclassification error for the independent (ID) and repulsive (RP)

Table A.1 Comparison of different degrees of overlaps 3%, 9%, 33% and 55% between independent and repulsive priors based on measurements of Kullback–Leibler (KL) and misclassification error (Miscl) when considering $a = \exp(-n^*_{2.5})$ single.

n	T	ID:N	ID:KL	ID:Miscl	RP:N	RP:KL	RP:Miscl
			Overlap 3%				
50	5	2	0.2124(0.1402,0.2364)	0.5098(0.3229,0.6332)	2	0.1986(0.1451,0.2336)	0.5000(0.2975,0.6207)
50	10	3	0.1285(0.0535,0.2059)	0.3603(0.1047,0.5540)	4	0.1238(0.0449,0.2008)	0.3373(0.0424,0.5937)
100	5	4	0.1381(0.0489,0.2057)	0.3570(0.1095,0.6077)	4	0.1268(0.0376,0.2003)	0.3386(0.0359,0.5696)
100	10	5	0.0503(0.0146,0.1142)	0.1437(0.0278,0.3512)	4	0.0652(0.0161,0.1169)	0.1839(0.0282,0.3812)
			Overlap 9%				
50	5	2	0.1181(0.1026,0.1441)	0.4991(0.3450,0.6518)	2	0.1167(0.0843,0.1481)	0.5087(0.3573,0.6746)
50	10	3	0.0905(0.0571,0.1189)	0.4423(0.3058,0.6099)	3	0.0856(0.0563,0.1141)	0.3999(0.3086,0.5976)
100	5	3	0.0889(0.0547,0.1106)	0.4185(0.3031,0.5853)	3	0.0854(0.0536,0.1114)	0.4359(0.3107,0.5965)
100	10	3	0.0542(0.0304,0.0961)	0.3642(0.2741,0.5214)	3	0.0500(0.0310,0.0920)	0.3393(0.2462,0.5349)
			Overlap 33%				
50	5	2	0.0533(0.0319,0.0771)	0.5629(0.3682,0.7706)	2	0.0559(0.0303,0.0742)	0.6186(0.3785,0.8018)
50	10	2	0.0291(0.0210,0.0487)	0.4442(0.3699,0.6199)	2	0.0282(0.0198,0.0665)	0.4259(0.3704,0.8015)
100	5	2	0.0315(0.0208,0.0634)	0.4630(0.3700,0.7221)	2	0.0279(0.0202,0.0680)	0.4420(0.3670,0.8015)
100	10	2	0.0200(0.0120,0.0273)	0.3981(0.3327,0.4825)	2	0.0195(0.0118,0.0319)	0.3922(0.3671,0.5069)
			Overlap 55%				
50	5	1	0.0321(0.0269,0.0396)	0.7639(0.5412,0.7835)	1	0.0291(0.0259,0.0385)	0.7714(0.5861,0.8043)
50	10	1	0.0249(0.0110,0.0299)	0.7500(0.4294,0.7842)	1	0.0228(0.0110,0.0303)	0.7102(0.4237,0.8035)
100	5	1	0.0251(0.0123,0.0319)	0.7446(0.4304,0.7854)	1	0.0240(0.0129,0.0294)	0.7483(0.4479,0.8029)
100	10	2	0.0134(0.0057,0.0246)	0.5671(0.4108,0.7823)	2	0.0103(0.0054,0.0245)	0.5082(0.4058,0.8016)

priors in Table A.2. All the input and tuning parameters for the simulation and RJMCMC were kept the same.

A.5 GPS Muskox Application

In this section we present supplementary information for the first case study of Section 2.4. The length-step data are presented in Figure A.4 whereas the angle in Figure in A.5. Clearly, from Figure A.4, we have positive observations which can be modelled with the use of Gamma distribution whereas from Figure A.5, we observe three modes for the angle observations, which are bell-shaped, indicating that a von Mises distribution can be used to infer the underlying states that give rise to those modes.

A.5.1 Model

In our case study, GPS technology is used to monitor individual location over time. As a result, the data gathered include both step lengths, L_t and turning angles, A_t between time points. The model is

Table A.2 Comparison of different degrees of consecutive overlaps 3%, 9%, 33% and 55% between independent and repulsive priors based on measurements of Kullback–Leibler (KL) and misclassification error (Miscl) when considering $a = \exp(-n_5^*)$ single.

n	T	ID:N	ID:KL	ID:Miscl	RP:N	RP:KL	RP:Miscl
			Overlap 3%				
50	5	2	0.2124(0.1402,0.2364)	0.5098(0.3229,0.6332)	2	0.1902(0.1385,0.2298)	0.4784(0.2851,0.6215)
50	10	3	0.1285(0.0535,0.2059)	0.3603(0.1047,0.5540)	3	0.1280(0.0514,0.2032)	0.3579(0.1035,0.6044)
100	5	4	0.1381(0.0489,0.2057)	0.3570(0.1095,0.6077)	3	0.1319(0.0356,0.2039)	0.3635(0.0325,0.5808)
100	10	5	0.0503(0.0146,0.1142)	0.1437(0.0278,0.3512)	4	0.0626(0.0161,0.1334)	0.1710(0.0269,0.3778)
			Overlap 9%				
50	5	2	0.1181(0.1026,0.1441)	0.4991(0.3450,0.6518)	2	0.1163(0.1003,0.1681)	0.5069(0.3624,0.7920)
50	10	3	0.0905(0.0571,0.1189)	0.4423(0.3058,0.6099)	2	0.0888(0.0578,0.1156)	0.4058(0.2984,0.6142)
100	5	3	0.0889(0.0547,0.1106)	0.4185(0.3031,0.5853)	3	0.0822(0.0567,0.1169)	0.4383(0.3138,0.5970)
100	10	3	0.0542(0.0304,0.0961)	0.3642(0.2741,0.5214)	3	0.0514(0.0300,0.0971)	0.3488(0.2673,0.5213)
			Overlap 33%				
50	5	2	0.0533(0.0319,0.0771)	0.5629(0.3682,0.7706)	2	0.0576(0.0318,0.0737)	0.6433(0.3926,0.8058)
50	10	2	0.0291(0.0210,0.0487)	0.4442(0.3699,0.6199)	2	0.0271(0.0210,0.0683)	0.4302(0.3662,0.8016)
100	5	2	0.0315(0.0208,0.0634)	0.4630(0.3700,0.7221)	2	0.0308(0.0205,0.0688)	0.4513(0.3693,0.8027)
100	10	2	0.0200(0.0120,0.0273)	0.3981(0.3327,0.4825)	2	0.0188(0.0098,0.0280)	0.3901(0.3548,0.4988)
			Overlap 55%				
50	5	1	0.0321(0.0269,0.0396)	0.7639(0.5412,0.7835)	1	0.0283(0.0258,0.0387)	0.7815(0.6505,0.8080)
50	10	1	0.0249(0.0110,0.0299)	0.7500(0.4294,0.7842)	1	0.0224(0.0104,0.0295)	0.7033(0.4222,0.8034)
100	5	1	0.0251(0.0123,0.0319)	0.7446(0.4304,0.7854)	1	0.0229(0.0113,0.0279)	0.7222(0.4359,0.8037)
100	10	2	0.0134(0.0057,0.0246)	0.5671(0.4108,0.7823)	2	0.0103(0.0048,0.0246)	0.4980(0.4087,0.8020)

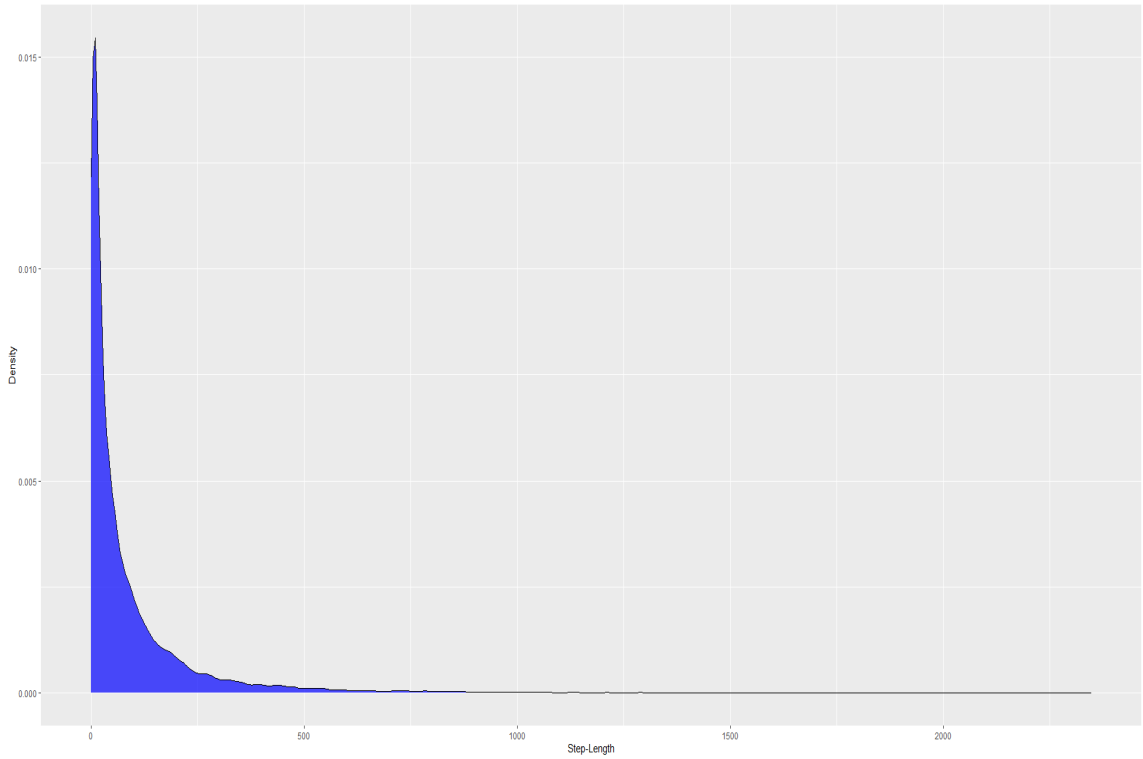

Fig. A.4 Observed step-length across time points.

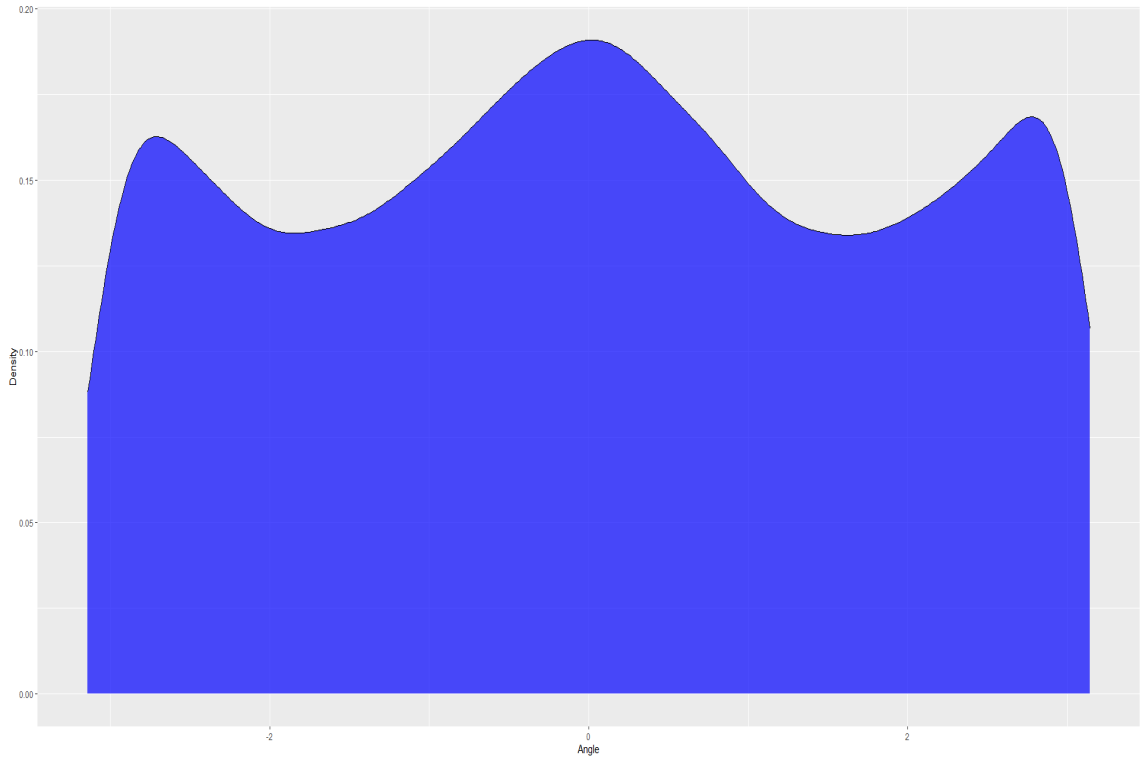

Fig. A.5 Observed angle across time points.

described as follows:

$$O_t = (L_t, A_t)$$

$$S_t \in \{1, 2, ..., N\}$$

$$f(O_t|S_t) = f(L_t|S_t) f(A_t|S_t)$$

$$f(L_t|S_t) = z_{S_t} \delta_{L_t}(0) + (1 - z_{S_t}) \text{Gamma}(L_t; \mu_{S_t}, \sigma_{S_t})$$

$$f(A_t|S_t) = \text{vonMises}(A_t; m_{S_t}, k_{S_t}) = \frac{e^{k_{S_t} cos(A_t - m_{S_t})}}{2\pi I_0(k_{S_t})},\ I_0 \text{ Bessel function of order 0}$$

for $t = 1,2,...,T$ and the specified priors are:

$$
\begin{aligned}
N &\sim \text{Uniform}\{1,2,...,80\} \\
\pi = (\pi_1,\pi_2,...,\pi_N) &= \left(\frac{\lambda_1}{\sum_{i=1}^N \lambda_i}, \frac{\lambda_2}{\sum_{i=1}^N \lambda_i}, ..., \frac{\lambda_N}{\sum_{i=1}^N \lambda_i}\right) \Rightarrow \lambda_i \sim \text{Gamma}(1,1), \;\; i = 1,2,...,N \\
P_{i\cdot} = (P_{i,1},P_{i,2},...,P_{i,N}) &= \left(\frac{\Lambda_{i1}}{\sum_{j=1}^N \Lambda_{ij}}, \frac{\Lambda_{i2}}{\sum_{j=1}^N \Lambda_{ij}}, ..., \frac{\Lambda_{iN}}{\sum_{j=1}^N \Lambda_{ij}}\right) \Rightarrow \Lambda_{ij} \sim \text{Gamma}(1,1), \;\; i,j = 1,2,...,N \\
z_i &\sim \text{Beta}(1,100), \;\; i = 1,2,...,N \\
k_i &\sim \text{Uniform}(0.5,2), \;\; i = 1,2,...,N \\
m_i &\sim \text{Uniform}(-\pi,\pi), \;\; i = 1,2,...,N \\
\sigma_i &\sim \text{Uniform}(0.5q_{0.1}, 2q_{0.9}), \;\; i = 1,2,...,N
\end{aligned}
$$

with $q_{0.1} = \{l_t : \mathbb{P}(L_t \leq l_t) = 0.1\}$ and $\{l_t : \mathbb{P}(L_t \leq l_t) = 0.9\}$ quantiles of the step length. The corresponding priors on the mean parameters of the step length for the repulsive case:

$$
\begin{aligned}
\underline{\mu} = (\mu_1,\mu_2,...,\mu_N) &\sim \text{StraussProcess}(\mu_1,\mu_2,...,\mu_N;\xi,a,d) = h(\mu_1,\mu_2,...,\mu_N|\xi,a,d) \\
&\propto \left[\prod_{i=1}^N \xi \mathbb{I}[\mu_i \in R]\right] a^{\sum_{1\leq i\leq j\leq N} \mathbb{I}[\|\mu_i-\mu_j\|<d]}
\end{aligned}
$$

for the independent case:

$$
\begin{aligned}
\underline{\mu} = (\mu_1,\mu_2,...,\mu_N) &\sim \text{IndependentProcess}(\mu_1,\mu_2,...,\mu_N;\xi) = h(\mu_1,\mu_2,...,\mu_N|\xi) \\
&= \frac{1}{\xi^N |R|^N}\left[\prod_{i=1}^N \xi \mathbb{I}[\mu_i \in R]\right] = \prod_{i=1}^N \frac{\mathbb{I}[\mu_i \in R]}{|R|}
\end{aligned}
$$

which corresponds to the product of N Uniform distributions in the region R. The point process parameter ξ for either cause of Strauss Process or Independent Process has

$$\xi \sim \text{Uniform}(|R|^{-1}, 80|R|^{-1})$$

The unknown number of states, denoted as N, follows a Uniform prior distribution with an upper bound of 80. This choice reflects our prior belief that there is not good reason to consider more than

80 behavioral states. It is important to note that this upper bound is subjective; although we could have opted for a smaller value, it should not be selected close to the expected number of behavioral states. This precaution is taken to allow the algorithm to independently identify the latent states without introducing bias to the results through the prior distribution.

For the initial and transition probability distribution we used a Dirichlet prior distribution, with parameters equal to 1, and use their Gamma decomposition equivalence explained in Argiento and De Iorio (2022) for obtaining a much more efficient mixing for the RJMCMC algorithm. Next, for the probability parameter z_i of the zero-inflated Gamma distribution, we selected a Beta prior distribution to express the proportion of zeros in the step length data. The prior distributions for the parameters k_i a Uniform within $[0.5, 2]$, m_i a Uniform within $[-\pi, \pi]$ and σ_i a Uniform within $[0.5\{l_t : \mathbb{P}(L_t \leq l_t) = 0.1\}, 2\{l_t : \mathbb{P}(L_t \leq l_t) = 0.9\}]$ which are the same distributions chosen in Pohle et al. (2017) for initializing the parameters values for their likelihood optimization. We made these choices for the prior distributions of the previously mentioned parameters, in order to be a ground of comparison between the methods proposed in Pohle et al. (2017) and our method, i.e. our goal was to observe how the RJMCMC alongside with the Strauss point process are affecting the inference, by minimizing the effect of prior distributions.

Moreover, for the repulsive prior we choose as norm measure $\|\cdot\|$ the Euclidean distance, for the penalty a and threshold d we can choose them based on the method explained in Beraha et al. (2022) which are equal to $a = \exp(-n^*_{2.5})$ with $n^*_{2.5} = 627$, $a = \exp(-n^*_5)$ with $n^*_5 = 1255$ and $d = 98$.

A.5.2 Inference

• Fixed dimension Moves

In the first step of the algorithm, we update the model parameters, for a fixed value N, by sampling from the corresponding posterior distributions. We sequentially update each parameter using a Metropolis Hastings algorithm. The proposal steps are of the following form

1. $\mu^*_{S_t} = \mu_{S_t} + \varepsilon_\mu, \quad \varepsilon_\mu \sim \text{LogNormal}(0, 0.01), \quad S_t = 1, 2, ..., N$

2. $\sigma^*_{S_t} = \sigma_{S_t} + \varepsilon_\sigma, \quad \varepsilon_\sigma \sim \text{LogNormal}(0, 0.03), \quad S_t = 1, 2, ..., N$

3. $k^*_{S_t} = k_{S_t} + \varepsilon_k, \quad \varepsilon_k \sim \text{LogNormal}(0, 0.08), \quad S_t = 1, 2, ..., N$

4. $m^*_{S_t} = m_{S_t} + \varepsilon_m, \quad \varepsilon_m \sim \text{Normal}(0, 0.08), \quad S_t = 1, 2, ..., N$

5. $\Lambda^*_{S_t S_{t+1}} = \Lambda_{S_t S_{t+1}} + \varepsilon_L, \quad \varepsilon_L \sim \text{LogNormal}(0, 0.05), \quad S_t, S_{t+1} = 1, 2, ..., N$

6. $\lambda^*_{S_1} = \lambda_{S_1} + \varepsilon_\lambda, \quad \varepsilon_\lambda \sim \text{LogNormal}(0, 0.07), \quad S_1 = 1, 2, ..., N$

7. $\text{logit}(z^*_{S_t}) = \text{logit}(z_{S_t}) + \varepsilon_z, \quad \varepsilon_z \sim \text{Normal}(0, 0.5), \quad S_t = 1, 2, ..., N \Leftrightarrow z^*_{S_t} \sim \text{LogNormal}(\text{logit}(z_{S_t}), \tau_z),$

8. ξ is sampled from each full conditional with a Metropolis Hastings algorithm and the steps are described in Section 2.2.4 in the main text.

Care must be taken when calculating the Metropolis-Hastings ratio since most of the proposed moves are not symmetric, which must be accounted for. The acceptance probabilities of the proposed values, for both versions, include the Jacobian that arises because we work with a logit and log scale transformation. The proposal distribution were chosen such the acceptance ratio were close to 0.25.

- **Variable dimension moves**

With probability 0.5, we choose between the moves Split/Combine and Birth/Death. The Split/Combine move splits or combines two existing components; in particular the choice of components to me combined is based on how similar they are, the similarity measure can be found later on. In the Birth/Death case, we kill or give birth to a new component by sampling from the corresponding proposal distributions.

Split/combine moves

In this step, we choose whether to split or combine components with probability 0.5. If we only have a single component, then with probability one, we split. In the split move, we choose uniformly one of the N components, denoted as j_* which we decide to split it to j_1 and j_2. Then the corresponding parameters split as follows

1. $\lambda_{j_1} = \rho \lambda_{j_*}, \ \lambda_{j_2} = (1-\rho)\lambda_{j_*}, \ \rho \sim \text{Beta}(2,2)$

2. $z_{j_1} = z_{j_*} - u_z, \ \ z_{j_2} = z_{j_*} + u_z, \ u_z \sim \text{Uniform}(0, \min(z_{j_*}, 1 - z_{j_*}))$

3. $\Lambda_{j j_1} = \Lambda_{j j_*} \rho_j, \ \Lambda_{j j_2} = \Lambda_{j j_*}(1-\rho_j), \ \rho_j \sim \text{Beta}(2,2), \ j \neq j_*$

4. $\Lambda_{j_1 j} = \Lambda_{j_* j}\theta_j,\ \Lambda_{j_2 j} = \Lambda_{j_* j}/\theta_j,\ \theta_j \sim \text{Gamma}(1,3),\ j \neq j_*$

5. $$\Lambda_{j_1 j_1} = \Lambda_{j_* j_*}\rho_{j_*}\theta_{j_1},\ \Lambda_{j_1 j_2} = \Lambda_{j_* j_*}(1-\rho_{j_*})\theta_{j_2}$$
$$\Lambda_{j_2 j_1} = \Lambda_{j_* j_*}\rho_{j_*}/\theta_{j_1},\ \Lambda_{j_2 j_2} = \Lambda_{j_* j_*}(1-\rho_{j_*})/\theta_{j_2}$$
with $\rho_{j_*} \sim \text{Beta}(2,2)$ and $\theta_{j_1}, \theta_{j_2} \sim \text{Gamma}(1,3)$

6. $\mu_{j_1} = \mu_{j_*} - \theta_\mu,\ \mu_{j_2} = \mu_{j_*} + \theta_\mu,\ \theta_\mu \sim \text{Uniform}(0, \mu_{j_*} - \mu_{j_*}/2)$

7. $\sigma_{j_1} = \sigma_{j_*} - \theta_\sigma,\ \sigma_{j_2} = \sigma_{j_*} + \theta_\sigma,\ \theta_\sigma \sim \text{Uniform}(0, \sigma_{j_*} - \sigma_{j_*}/2)$

8. $k_{j_1} = k_{j_*} - \theta_k,\ k_{j_2} = k_{j_*} + \theta_k,\ \theta_k \sim \text{Uniform}(0, k_{j_*} - k_{j_*}/2)$

9. $m_{j_1} = m_{j_*} + \varepsilon_m,\ m_{j_2} = m_{j_*} - \varepsilon_m,\ \varepsilon_m \sim \text{Normal}(0,2)$

In the reverse move, we merge the most similar components j_1 and j_2 and to j_*.

1. $\lambda_{j_*} = \lambda_{j_1} + \lambda_{j_2}$

2. $z_{j_*} = \frac{z_{j_1} + z_{j_2}}{2}$

3. $\Lambda_{j j_*} = \Lambda_{j j_1} + \Lambda_{j j_2}\ \ j \neq j_*$

4. $\Lambda_{j_* j} = (\Lambda_{j_1 j}\Lambda_{j_2 j})^{0.5},\ \ j \neq j_*$

5. $\Lambda_{j_* j_*} = (\Lambda_{j_1 j_1}\Lambda_{j_2 j_1})^{0.5} + (\Lambda_{j_1 j_2}\Lambda_{j_2 j_2})^{0.5}$

6. $\mu_{j_*} = \frac{\mu_{j_1} + \mu_{j_2}}{2}$

7. $\sigma_{j_*} = \frac{\sigma_{j_1} + \sigma_{j_2}}{2}$

8. $k_{j_*} = \frac{k_{j_1} + k_{j_2}}{2}$

9. $m_{j_*} = \frac{m_{j_1} + m_{j_2}}{2}$

The split is accepted with probability $\min\{1, A\}$ whereas in the combine move we accepted with $\min\left\{1, A^{-1}\right\}$.

$$A = \frac{f(\{O_t\}_{t=1}^T | \{\pi\}_{j=1}^{N+1}, \{P_j\}_{j=1}^{N+1}, \{\mu_j\}_{j=1}^{N+1}, \{\sigma_j\}_{j=1}^{N+1}, \{m_j\}_{j=1}^{N+1}, \{k_j\}_{j=1}^{N+1})}{f(\{O_t\}_{t=1}^T | \{\pi\}_{j=1}^{N}, \{P_j\}_{j=1}^{N}, \{\mu_j\}_{j=1}^{N}, \{\sigma_j\}_{j=1}^{N}, \{m_j\}_{j=1}^{N}, \{k_j\}_{j=1}^{N})}$$
$$\frac{p(\{\pi\}_{j=1}^{N+1}, \{P_j\}_{j=1}^{N+1}, \{\mu_j\}_{j=1}^{N+1}, \{\sigma_j\}_{j=1}^{N+1}, \{m_j\}_{j=1}^{N+1}, \{k_j\}_{j=1}^{N+1}) p(N+1)}{p(\{\pi\}_{j=1}^{N}, \{P_j\}_{j=1}^{N}, \{\mu_j\}_{j=1}^{N}, \{\sigma_j\}_{j=1}^{N}, \{m_j\}_{j=1}^{N}, \{k_j\}_{j=1}^{N}) p(N)}$$
$$\underbrace{\frac{(N+1)!}{N!} \frac{P_c(N+1)/[2(N+1)]}{P_s(N)/N} \frac{|J|}{p(\theta_{j_1}) p(\theta_{j_2}) \prod_j p(\theta_j) p(\theta_\mu) p(\theta_\sigma) p(\theta_k) p(\varepsilon_m) p(\rho) \prod_j p(\rho_j) p(z_{j_*})}}_{q(N+1\to N)/q(N\to N+1)}$$

The Jacobian, $|J|$ of the transformation from N to $N+1$, is equal to

$$|J| = 2\lambda_{j_*} k_{j_*} \mu_{j_*} \sigma_{j_*} 4\rho_{j_*}(1-\rho_{j_*}) \frac{L_{j_* j_*}^3}{\theta_{j_1} \theta_{j_2}} \prod_j \Lambda_{j j_*} 2^{N-1} \prod_j \frac{\Lambda_{j_* j}}{\theta_j}$$

Also, the probabilities P_c and P_s correspond to the probabilities of making a combine or split movie, which in our case will cancel out. The split and merge moves are the ones described in Bartolucci et al. (2013).

Lastly the similarity measure for combining two components, is calculated as

$$d_i = \sum_{j \neq i} \sqrt{(\mu_i - \mu_j)^2 + (\sigma_i - \sigma_j)^2 + (m_i - m_j)^2 + (k_i - k_j)^2}$$

for each component i. Then we choose the two components that have the smallest $d_i's$ values, and we combine them.

Birth/death moves

The Birth/Death move is performed similarly to the Split/Combine. Likewise, if we have N components, we choose with probability 0.5 to give birth to a new component or death to an existing one. In the birth move, we generate new component parameters from the prior distribution, and the rest of the components are simply copied. On the other hand, for the death move, we uniformly choose a component and kill it. In this case, the acceptance probability of birth move is again $\min\{1, A\}$

whereas for the death move $\min\left\{1, A^{-1}\right\}$ with

$$A = \frac{f(\{O_t\}_{t=1}^{T} | \{\pi\}_{j=1}^{N+1}, \{P_j\}_{j=1}^{N+1}, \{\mu_j\}_{j=1}^{N+1}, \{\sigma_j\}_{j=1}^{N+1}, \{m_j\}_{j=1}^{N+1}, \{k_j\}_{j=1}^{N+1})}{f(\{O_t\}_{t=1}^{T} | \{\pi\}_{j=1}^{N}, \{P_j\}_{j=1}^{N}, \{\mu_j\}_{j=1}^{N}, \{\sigma_j\}_{j=1}^{N}, \{m_j\}_{j=1}^{N}, \{k_j\}_{j=1}^{N})}$$
$$\frac{p(\{\pi\}_{j=1}^{N+1}, \{P_j\}_{j=1}^{N+1}, \{\mu_j\}_{j=1}^{N+1}, \{\sigma_j\}_{j=1}^{N+1}, \{m_j\}_{j=1}^{N+1}, \{k_j\}_{j=1}^{N+1}) p(N+1)}{p(\{\pi\}_{j=1}^{N}, \{P_j\}_{j=1}^{N}, \{\mu_j\}_{j=1}^{N}, \{\sigma_j\}_{j=1}^{N}, \{m_j\}_{j=1}^{N}, \{k_j\}_{j=1}^{N}) p(N)}$$
$$\underbrace{\frac{(N+1)!}{N!} \frac{P_d(N+1)/N}{P_b(N)/(N+1)} \frac{|J|}{p(\lambda_{j_*}) p(z_{j_*}) p(\mu_{j_*}) p(\sigma_{j_*}) p(k_{j_*}) p(m_{j_*}) p(\Lambda_{j_* j_*}) \prod_j p(\Lambda_{j j_*}) p(\Lambda_{j_* j})}}_{q(N+1 \to N)/q(N \to N+1)}$$

Since the parameters are drawn from their respective priors, the Jacobian term $|J|$ will equal one.

A.5.3 Results

We display the results of the ICL information criterion for model comparison as described in Pohle et al. (2017), in Table A.3.

Table A.3 Values of model selection criterion ICL for the different numbers of components 2, 3, 4, 5, 6 and 7, computed with the algorithm displayed in Pohle et al. (2017)

	Number of components					
	2	3	4	5	6	7
ICL	354,829	351,544	350,159	351,247	354,701	350,051

Then for the last time point of the time series we display the averaged posterior mixture distributions of the step length and angle for the independent prior model, in Figure A.6.

We also present the averaged posterior mixture distributions of the step length and angle for the repulsive prior model, with $a = \exp(-n_5^*)$ with $n_5^* = 1255$, displayed in Figure A.7.

Since, we have used a smaller a compared to the case of 2.5% we expect to infer a small number of clusters which is evident from the Figure A.7 where we identify two clusters, one corresponding to small undirected steps (State 1) and one corresponding to big directed steps (State 2). Interestingly, if we look at the posterior distribution on the number of states N, $p(2) = 0.4307, p(3) = 0.1265, p(4) = 0.2357, p(5) = 0.2045, p(6) = 0.0026$, we can observe that even though the mode of N is on 2 there

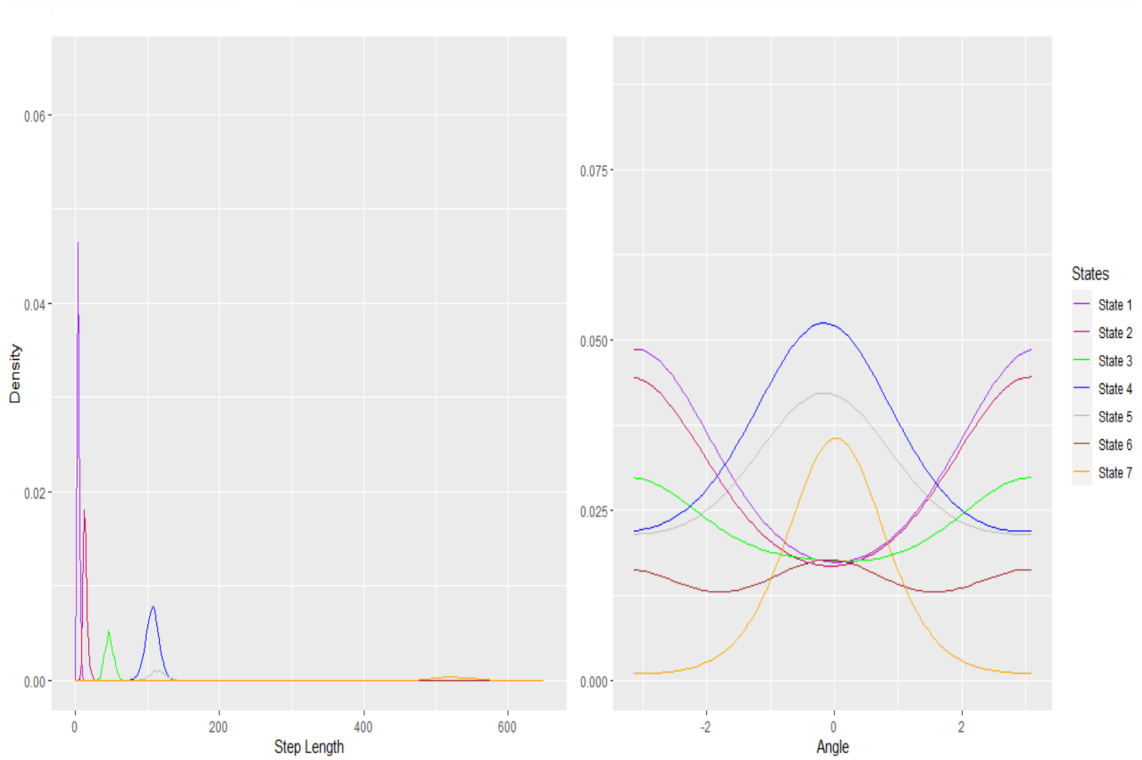

Fig. A.6 Averaged posterior mixture distribution of step length (left) and angle (right) for the last time point of the time series, for the independent prior model.

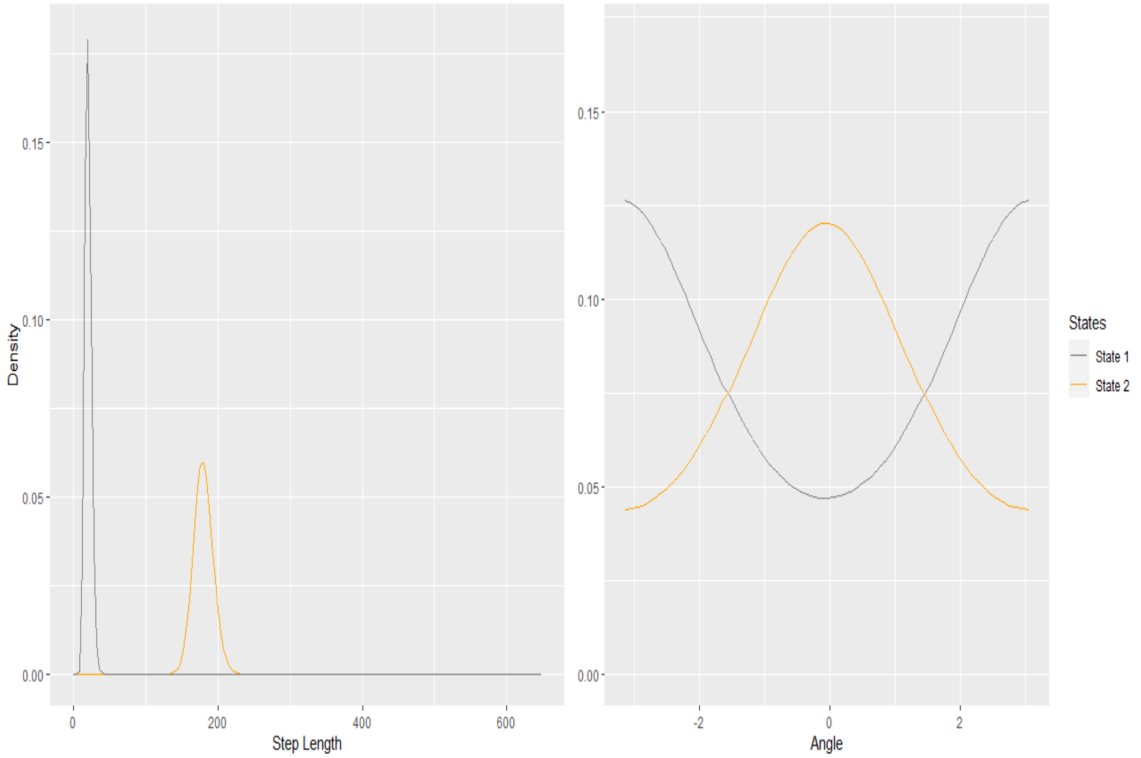

Fig. A.7 Averaged posterior mixture distribution of step length (left) and angle (right) for the last time point of the time series, for the repulsive prior model, when accounting for penalty $a = \exp(-n_5^*)$.

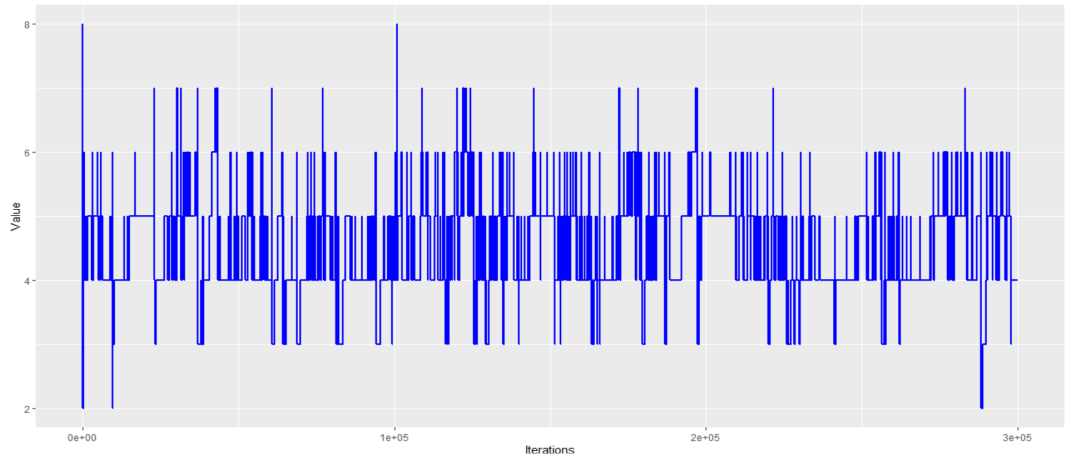

Fig. A.8 Iterations of number of components N, with burn-in iterations removed.

is some significant mass on the number of states 4 and 5 showing evidence that the choice of 5% might lead to overpenalization.

Lastly, we assess the convergence of our RJMCMC algorithm. We run a chain for 300000 iterations. The number of iterations was chosen such that the Monte Carlo error (MCE) for each parameter estimation is sufficiently small with respect to the scale of the parameter. In Table A.4 we display the Z-value of the Geweke statistic, the effective sample size (ESS) and the MCE.

Also, we display the trace plot for the parameters of the number of components N and means and standard deviation μ_i, σ_i, $i = 1,2,4$ in Figures A.8-A.16.

A.6 Acoustic Application

In this section we present supplementary information for the second case study of Section 2.5. The acoustic data presented in Figure A.17.

A.6.1 Model

In our case study, airborne devices are used to collect acoustic data over time. As a result, the data we gathered are acoustic features based on the Mel-frequency cepstral coefficient measured for each

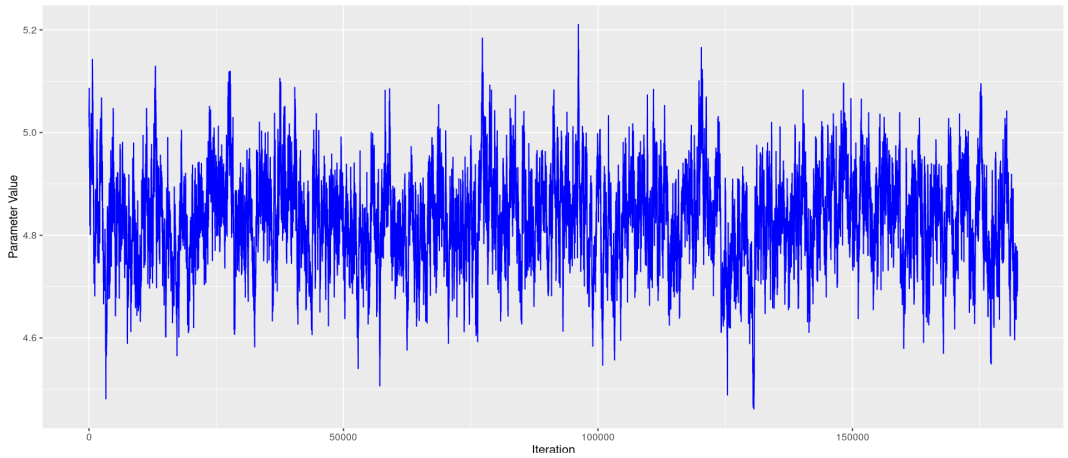

Fig. A.9 Iterations of mean μ_1, with burn-in iterations removed.

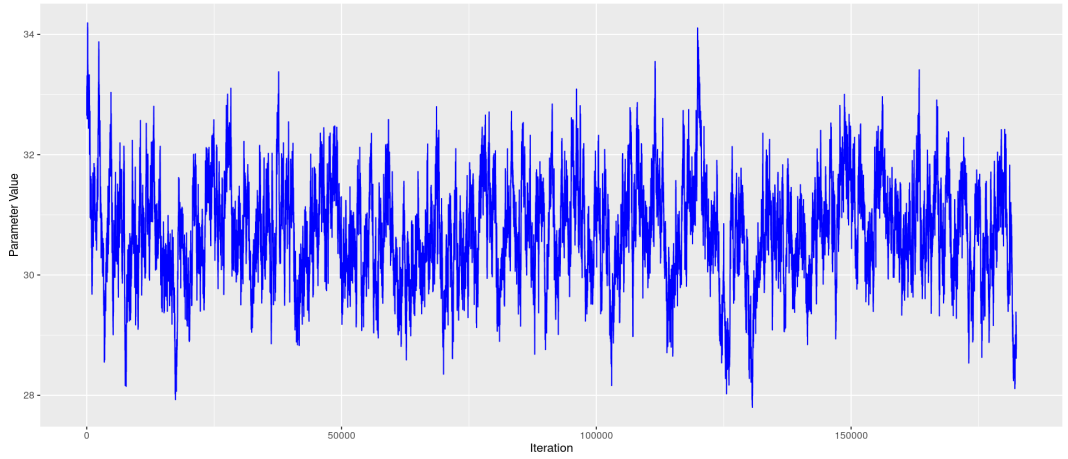

Fig. A.10 Iterations of mean μ_2, with burn-in iterations removed.

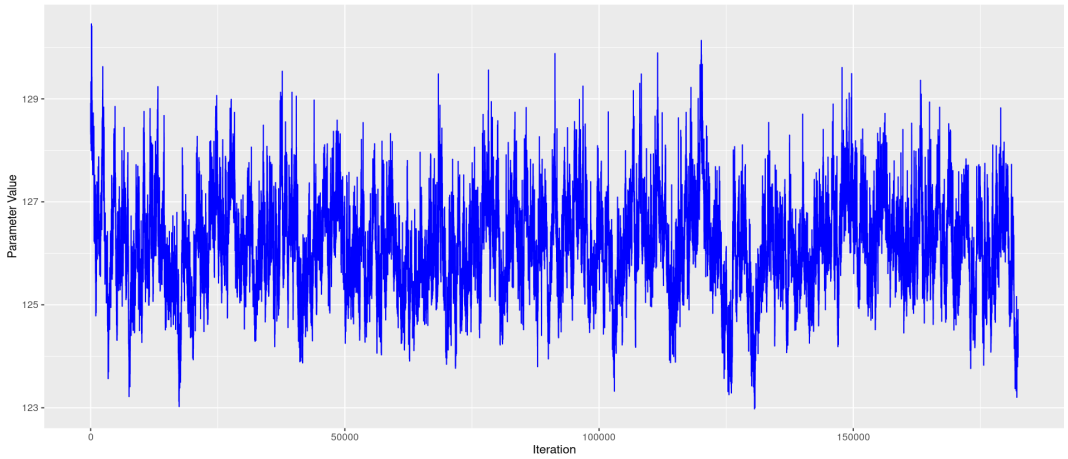

Fig. A.11 Iterations of mean μ_3, with burn-in iterations removed.

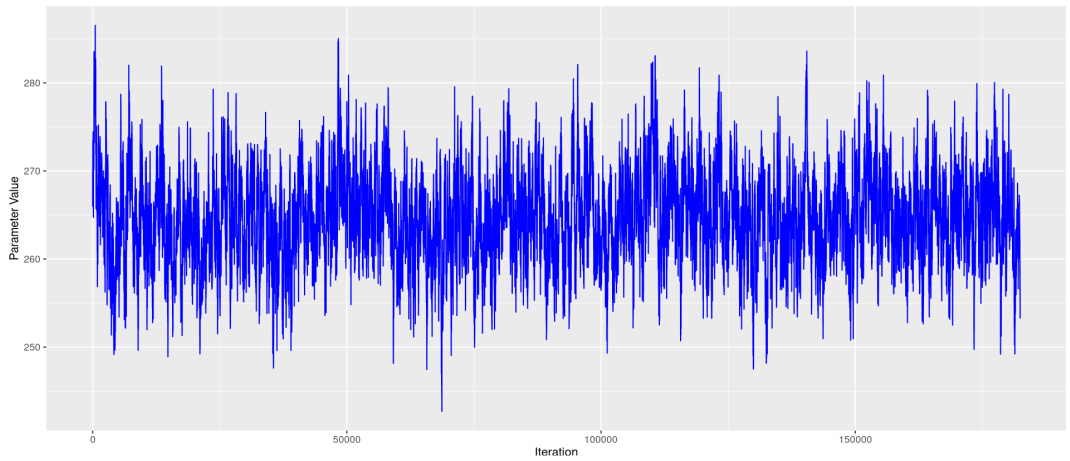

Fig. A.12 Iterations of mean μ_4, with burn-in iterations removed.

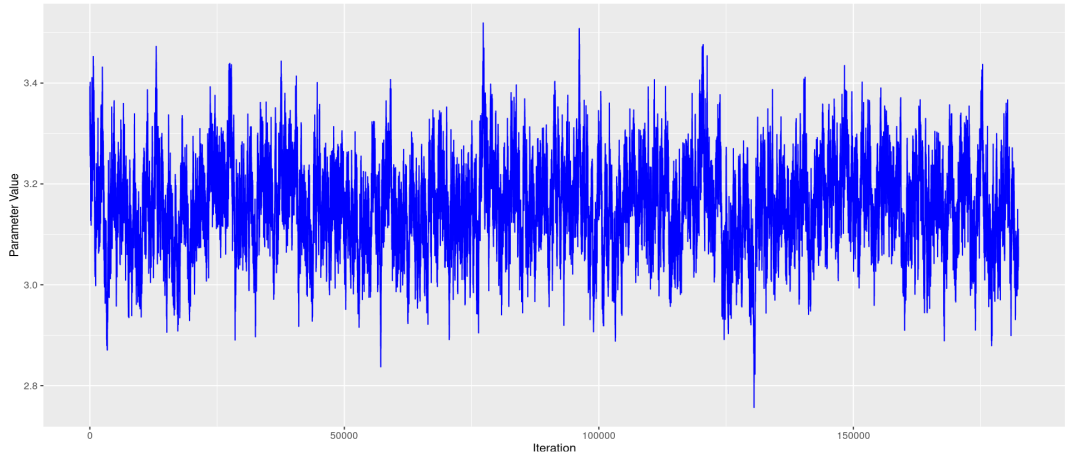

Fig. A.13 Iterations of mean σ_1, with burn-in iterations removed.

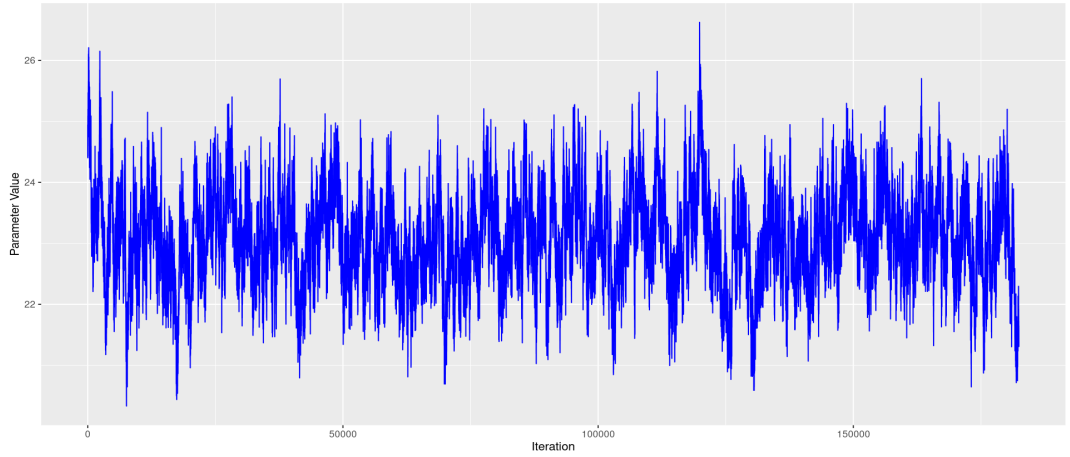

Fig. A.14 Iterations of mean σ_2, with burn-in iterations removed.

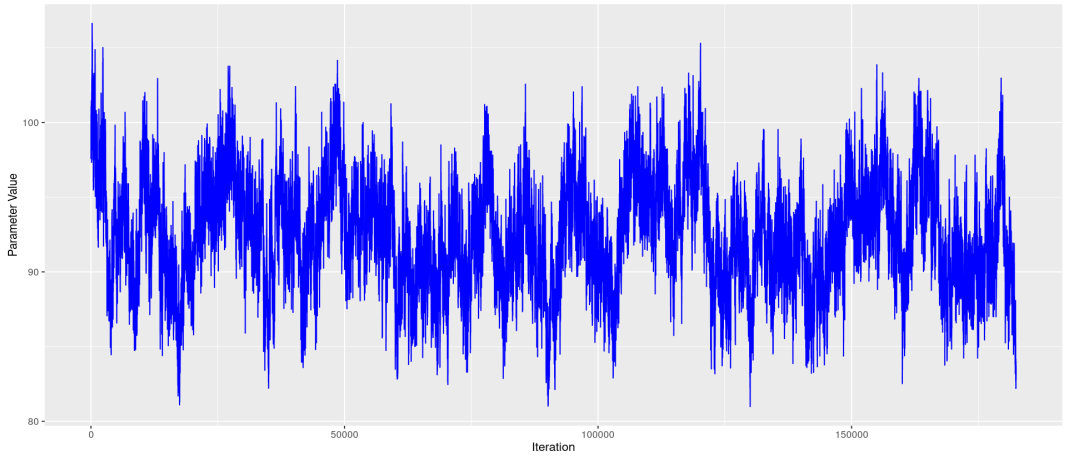

Fig. A.15 Iterations of mean σ_3, with burn-in iterations removed.

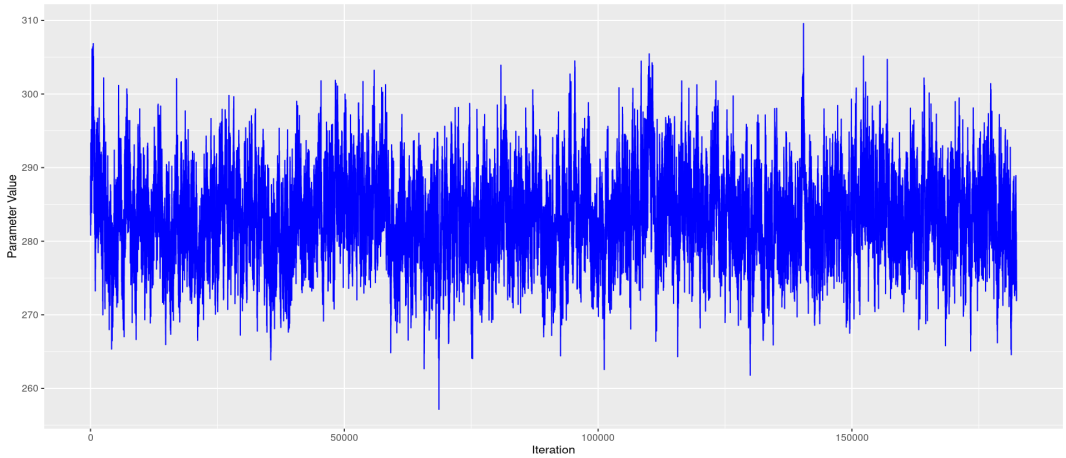

Fig. A.16 Iterations of mean σ_4, with burn-in iterations removed.

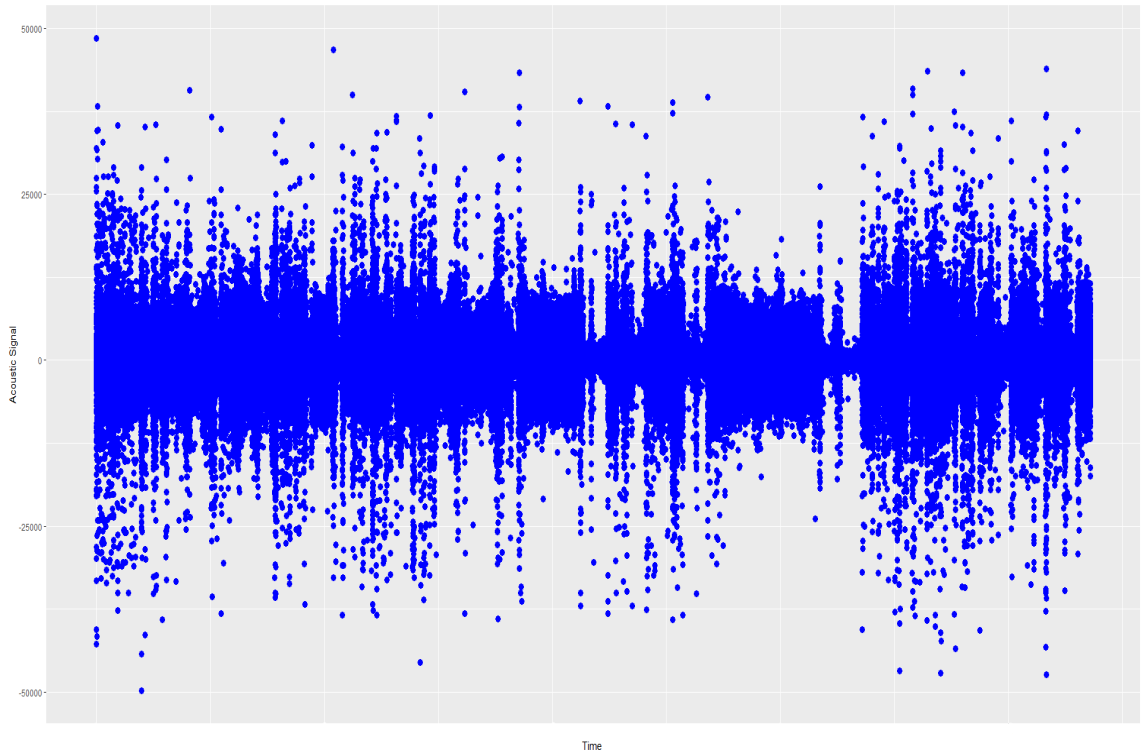

Fig. A.17 Observed acoustic data cross time.

segment time point. However, those features are highly correlated for which reason we applied a PCA and kelp only the first two PC. The model is described as follows:

$$O_{t,c} = \underline{E_{t,c}} = \{E_{c,t,1}, E_{c,t,1}\}, \; c = 1,2,...,C$$

$$S_t \in \{1,2,...,N\}$$

$$f(O_{t,c}|S_t) = f(\underline{E_{t,c}}|S_t)$$

$$f(\underline{E_{t,c}}|S_t) = \text{Normal}_2(\underline{E_{t,c}}; \underline{\mu_{S_t}}, \Sigma_{S_t})$$

the specified priors are:

$$
\begin{aligned}
N &\sim \text{Uniform}\{1,2,...,120\} \\
\pi = (\pi_1,\pi_2,...,\pi_N) &= \left(\frac{\lambda_1}{\sum_{i=1}^N \lambda_i}, \frac{\lambda_2}{\sum_{i=1}^N \lambda_i}, ..., \frac{\lambda_N}{\sum_{i=1}^N \lambda_i}\right) \Rightarrow \lambda_i \sim \text{Gamma}(1,1), \;\; i = 1,2,...,N \\
P_{i\cdot} = (P_{i,1},P_{i,2},...,P_{i,N}) &= \left(\frac{\Lambda_{i1}}{\sum_{j=1}^N \Lambda_{ij}}, \frac{\Lambda_{i2}}{\sum_{j=1}^N \Lambda_{ij}}, ..., \frac{\Lambda_{iN}}{\sum_{j=1}^N \Lambda_{ij}}\right) \Rightarrow \Lambda_{ij} \sim \text{Gamma}(1,1), \;\; i,j = 1,2,...,N \\
\Sigma_i &\sim \text{Wishart}(20, \hat{\Sigma}_0/10), \;\; i = 1,2,...,N
\end{aligned}
$$

and the corresponding priors on the vector means are:

$$
\begin{aligned}
\underline{\mu} = (\underline{\mu_1},\underline{\mu_2},...,\underline{\mu_N}) &\sim \text{StraussProcess}(\underline{\mu_1},\underline{\mu_2},...,\underline{\mu_N};\xi,a,d) = h(\underline{\mu_1},\underline{\mu_2},...,\underline{\mu_N}|\xi,a,d) \\
&\propto \left[\prod_{i=1}^N \xi \mathbb{I}[\underline{\mu_i} \in R^2]\right] a^{\sum_{1\le i\le j\le N} \mathbb{I}[\left\|\underline{\mu_i}-\underline{\mu_j}\right\|<d]}
\end{aligned}
$$

for the independent prior case:

$$
\begin{aligned}
\underline{\mu} = (\underline{\mu_1},\underline{\mu_2},...,\underline{\mu_N}) &\sim \text{IndependentProcess}(\underline{\mu_1},\underline{\mu_2},...,\underline{\mu_N};\xi) = h(\underline{\mu_1},\underline{\mu_2},...,\underline{\mu_N}|\xi) \\
&= \left[\prod_{i=1}^N \xi \mathbb{I}[\underline{\mu_i} \in R^2]\right] \frac{1}{\xi^N |R^2|^N} = \prod_{i=1}^N \frac{\mathbb{I}[\underline{\mu_i} \in R^2]}{|R^2|}
\end{aligned}
$$

which corresponds to the product of N Uniform distributions in the region R^2. The point process parameter ξ for either cause of Strauss Process or Strandard Process has

$$
\xi \sim \text{Uniform}(|R|^{-2}, 120|R|^{-2})
$$

The unknown number of states, denoted as N, follows a Uniform prior distribution with an upper bound of 120. This choice reflects our prior belief that there should be no more than 120 behavioral states. It is important to note that this upper bound is subjective; although we could have opted for a smaller value, it should not be selected close to the expected number of behavioral states. This precaution is taken to allow the algorithm to independently identify the latent states without introducing bias to the results through the prior distribution.

For the initial and transition probability distribution we used Dirichlet prior distributions with parameters equal to 1, and use their Gamma decomposition equivalence explained in Argiento and De Iorio (2022) for obtaining a much more efficient mixing for the RJMCMC algorithm. Next, for the covariance matrix we allow for a Wishart prior distribution with $\hat{\Sigma_0}$ the sampled covariance matrix of all segment time points cosidered together. Lastly, if R is the range of the transformed acoustic features on the first two PC, then we have the range of the product space R^2 as $|R|^2$.

The penalty parameters of the Strauss point process a, d are fixed and the prior distribution for ξ are defined, based on the rules outlined in Beraha et al. (2022). Moreover, for the repulsive prior we choose as norm measure $\|\cdot\|$ the Euclidean distance, for the penalty a and threshold d we can choose them based on the method explained in Beraha et al. (2022) which are equal to $a = \exp(-n^*_{2.5})$ with $n^*_{2.5} = 54$, $a = \exp(-n^*_5)$ with $n^*_5 = 108$ and $d = 21$.

• **Fixed dimension Moves**

In the first step of the algorithm, we update the model parameters, for a fixed value N, by sampling from the corresponding posterior distributions. We sequentially update each parameter using a Metropolis Hastings algorithm. The steps are of the following form

1. $\underline{\mu}^*_{S_t} = \mu_{S_t} + \varepsilon_\mu, \quad \varepsilon_\mu \sim \text{Normal}(0, 0.3), \quad S_t = 1, 2, ..., N$
2. $\Sigma^*_{S_t} \sim \text{Wishart}(1200, \Sigma_{S_t}/1200), \quad S_t = 1, 2, ..., N$
3. $\Lambda^*_{S_t S_{t+1}} = \Lambda_{S_t S_{t+1}} + \varepsilon_L, \quad \varepsilon_L \sim \text{LogNormal}(0, 1), \quad S_t, S_{t+1} = 1, 2, ..., N$
4. $\lambda^*_{S_1} = \lambda_{S_1} + \varepsilon_\lambda, \quad \varepsilon_\lambda \sim \text{LogNormal}(0, 1.5), \quad S_1 = 1, 2, ..., N$
5. ξ is sampled from each full conditional with a Metropolis Hastings algorithm and the steps are described in Section 2.4 in the main text.

Care must be taken when calculating the Metropolis-Hastings ratio since most of the proposed moves are not symmetric, which must be accounted for. The acceptance probabilities of the proposed values, for both versions, include the Jacobian that arises because we work with a logit and log scale transformation. The proposal distribution were chosen such the acceptance ratio were close to 0.25.

- **Variable dimension moves**

With probability 0.5, we choose between the moves Split/Combine and Birth/Death. The Split/Combine move splits or combines two existing components; in particular the choice of components to me combined is based on how similar they are, the similarity measure can be found later on. In the Birth/Death case, we kill or give birth to a new component by sampling from the corresponding proposal distributions.

Split/combine moves

In this step, we choose whether to split or combine components with probability 0.5. If we only have a single component, then with probability one, we split. In the split move, we choose uniformly one of the N components, denoted as j_* which we decide to split it to j_1 and j_2. Based on the split/combine moves described in Bartolucci et al. (2013); Zhang et al. (2004) the corresponding parameters split are as follows

1. $\lambda_{j_1} = \rho\lambda_{j_*},\ \lambda_{j_2} = (1-\rho)\lambda_{j_*},\ \rho$
2. $\Lambda_{jj_1} = \Lambda_{jj_*}\rho_j,\ \Lambda_{jj_2} = \Lambda_{jj_*}(1-\rho_j),\ \rho_j \sim \text{Beta}(2,2),\ j \neq j_*$
3. $\Lambda_{j_1 j} = \Lambda_{j_* j}\theta_j,\ \Lambda_{j_2 j} = \Lambda_{j_* j}/\theta_j,\ \theta_j \sim \text{Gamma}(1,3),\ j \neq j_*$
4. $$\Lambda_{j_1 j_1} = \Lambda_{j_* j_*}\rho_{j_*}\theta_{j_1},\ \Lambda_{j_1 j_2} = \Lambda_{j_* j_*}(1-\rho_{j_*})\theta_{j_2}$$
$$\Lambda_{j_2 j_1} = \Lambda_{j_* j_*}\rho_{j_*}/\theta_{j_1},\ \Lambda_{j_2 j_2} = \Lambda_{j_* j_*}(1-\rho_{j_*})/\theta_{j_2}$$
with $\rho_{j_*} \sim \text{Beta}(2,2)$ and $\theta_{j_1}, \theta_{j_2} \sim \text{Gamma}(1,3)$
5. $\underline{\mu_{j_1}} = \underline{\mu_*} - \sqrt{\frac{\pi_{j_2}}{\pi_{j_1}}}\sum_{d=1}^{12} r_{*d}^{\frac{1}{2}} u_d \underline{a_d},\ \ \underline{\mu_{j_2}} = \underline{\mu_*} + \sqrt{\frac{\pi_{j_1}}{\pi_{j_2}}}\sum_{d=1}^{12} r_{*d}^{\frac{1}{2}} u_d \underline{a_d},\ \ u_d \sim \text{Beta}(2,2)$
the r_{*d} is the dth eigenvalue of the covariance matrix Σ_* and the a_d is the dth eigenvector of the sampled covariance matrix $\hat{\Sigma_0}$.

6. $r_{j_1 d} = \beta_d(1-u_d^2)\frac{\pi_*}{\pi_{j_1}} r_{*d}, \quad r_{j_2 d} = (1-\beta_d)(1-u_d^2)\frac{\pi_*}{\pi_{j_2}} r_{*d}, \ \beta_d \sim \text{Beta}(1,1), \ d = 1,2,...,12$
   we employ this specific eigenvalue decomposition split move, as described in Zhang et al. (2004), to ensure that the resulting covariance matrices for components j_1 and j_2 are both positive-definite and symmetric. This guarantees that they meet the criteria for being valid covariance matrices.

In the reverse move, we merge the most similar components j_1 and j_2 and to j_*.

1. $\lambda_{j_*} = \lambda_{j_1} + \lambda_{j_2}$
2. $\Lambda_{jj_*} = \Lambda_{jj_1} + \Lambda_{jj_2} \quad j \neq j_*$
3. $\Lambda_{j_* j} = (\Lambda_{j_1 j}\Lambda_{j_2 j})^{0.5}, \ \ j \neq j_*$
4. $\Lambda_{j_* j_*} = (\Lambda_{j_1 j_1}\Lambda_{j_2 j_1})^{0.5} + (\Lambda_{j_1 j_2}\Lambda_{j_2 j_2})^{0.5}$
5. $\mu_{j_*} = \mu_{j_1}\frac{\pi_{j_1}}{\pi_{j_*}} + \mu_{j_2}\frac{\pi_{j_2}}{\pi_{j_*}}$
6. $\lambda_{j_*} = \frac{\pi_{j_1}}{\pi_{j_*}} r_{j_1 d} + \frac{\pi_{j_2}}{\pi_{j_*}} r_{j_2 d} + \frac{\pi_{j_1}\pi_{j_2}}{\pi_{j_*}}(\mu_{j_1 d} - \mu_{j_2 d})^2, \quad d = 1,2,...,d$

The split is accepted with probability $\min\{1, A\}$ whereas in the combine move we accepted with $\min\left\{1, A^{-1}\right\}$.

$$A = \frac{f(\{O_t\}_{t=1}^T \,|\, \{\pi\}_{j=1}^{N+1}, \{P_j\}_{j=1}^{N+1}, \left\{\underline{\mu_j}\right\}_{j=1}^{N+1}, \{\Sigma_j\}_{j=1}^{N+1})}{f(\{O_t\}_{t=1}^T \,|\, \{\pi\}_{j=1}^{N}, \{P_j\}_{j=1}^{N}, \left\{\underline{\mu_j}\right\}_{j=1}^{N}, \{\Sigma_j\}_{j=1}^{N})}$$
$$\frac{p(\{\pi\}_{j=1}^{N+1}, \{P_j\}_{j=1}^{N+1}, \left\{\underline{\mu_j}\right\}_{j=1}^{N+1}, \{\Sigma_j\}_{j=1}^{N+1})p(N+1)}{p(\{\pi\}_{j=1}^{N}, \{P_j\}_{j=1}^{N}, \left\{\underline{\mu_j}\right\}_{j=1}^{N}, \{\Sigma_j\}_{j=1}^{N})p(N)}$$
$$\underbrace{\frac{(N+1)!}{N!}\frac{P_c(N+1)/[2(N+1)]}{P_s(N)/N}\frac{|J|}{p(\theta_{j_1})p(\theta_{j_2})\prod_j p(\theta_j)p(\rho)\prod_j p(\rho_j)\prod_d p(\beta_d)p(u_d)}}_{q(N+1\to N)/q(N\to N+1)}$$

The Jacobian, $|J|$ of the transformation from N to $N+1$, is equal to

$$|J| = 4\rho_{j_*}(1-\rho_{j_*})\frac{L_{j_* j_*}^3}{\theta_{j_1}\theta_{j_2}}\prod_j \Lambda_{jj_*} 2^{N-1}\prod_j \frac{\Lambda_{j_* j}}{\theta_j}\frac{\pi_{j_*}^{3*12+1}}{(\pi_{j_1}\pi_{j_2})^{\frac{3*12}{2}}}\sum r_{kd}^{\frac{3}{2}}(1-u_d^2)$$

Also, the probabilities P_c and P_s correspond to the probabilities of making a combine or split movie, which in our case will cancel out. The split and merge moves are the ones described in Bartolucci et al. (2013).

Lastly the similarity measure for combining two components, is calculated as

$$d_i = \sum_{j \neq i} \left\| \underline{\mu_i} - \underline{\mu_j} \right\|_2$$

for each component i. Then we choose the two components that have the smallest $d_i's$ values, and we combine them. The $\|\cdot\|$ corresponds to the Euclidean distance.

Birth/death moves

The Birth/death move is performed similarly to the Split/Combine. Likewise, if we have N components, we choose with probability 0.5 to give birth to a new component or death to an existing one. In the birth move, we generate new component parameters from the prior distribution, and the rest of the components are simply copied. On the other hand, for the death move, we uniformly choose a component and kill it. In this case, the acceptance probability of birth move is again $\min\{1, A\}$ whereas for the death move $\min\left\{1, A^{-1}\right\}$ with

$$A = \frac{f(\{O_t\}_{t=1}^T \,|\, \{\pi\}_{j=1}^{N+1}, \{P_j\}_{j=1}^{N+1}, \left\{\underline{\mu_j}\right\}_{j=1}^{N+1}, \{\Sigma_j\}_{j=1}^{N+1})}{f(\{O_t\}_{t=1}^T \,|\, \{\pi\}_{j=1}^{N}, \{P_j\}_{j=1}^{N}, \left\{\underline{\mu_j}\right\}_{j=1}^{N}, \{\Sigma_j\}_{j=1}^{N})} \frac{p(\{\pi\}_{j=1}^{N+1}, \{P_j\}_{j=1}^{N+1}, \left\{\underline{\mu_j}\right\}_{j=1}^{N+1}, \{\Sigma_j\}_{j=1}^{N+1}) p(N+1)}{p(\{\pi\}_{j=1}^{N}, \{P_j\}_{j=1}^{N}, \left\{\underline{\mu_j}\right\}_{j=1}^{N}, \{\Sigma_j\}_{j=1}^{N}) p(N)} \underbrace{\frac{(N+1)!}{N!} \frac{P_d(N+1)/N}{P_b(N)/(N+1)} \frac{|J|}{p(\lambda_{j_*}) p(\Lambda_{j_* j_*}) \prod_j p(\Lambda_{j j_*}) p(\Lambda_{j_* j}) \prod_d p(\beta_d) p(u_d)}}_{q(N+1 \to N)/q(N \to N+1)}$$

Since the parameters are drawn from their respective priors, the Jacobian term $|J|$ will equal one.

A.6.2 Results

Posterior distribution on the number of states N with repulsive prior for $a = \exp(-n^*_{2.5})$, is $p(2) = 0.08764, p(3) = 0.12862, p(4) = 0.11209, p(5) = 0.10901, p(6) = 0.10076, p(7) = 0.09890, ..., p(25) = 0.00008$ with $\sum_{i=2}^{25} p(i) = 1$ and for the independent prior we display the posterior distribution of the allocation in Figure A.18.

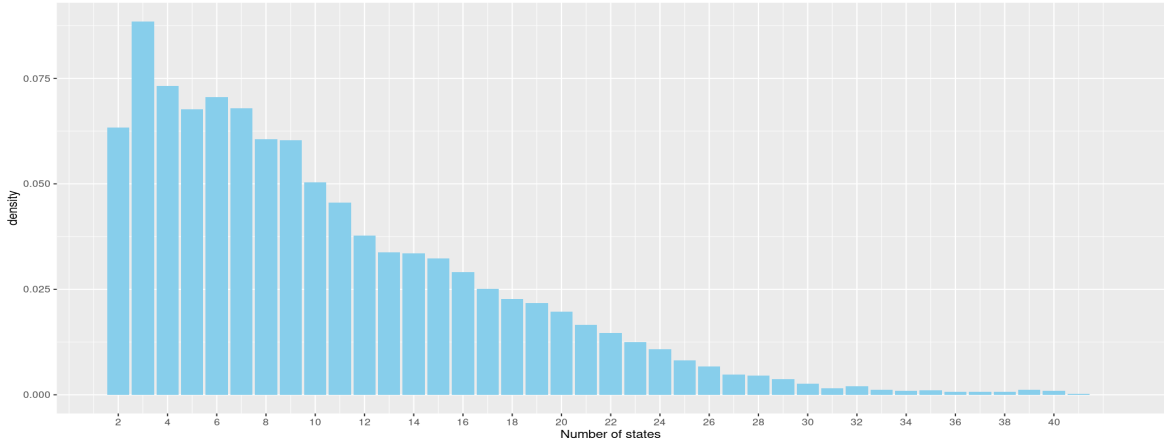

Fig. A.18 Posterior distribution across the states 2 to 41 for the independent prior model.

We give the posterior uncertainty probabilities of classification for the models with two and three and four mixture states, in Figure A.19.

Next, on Figure A.20 we display the observations with points coloured according to their modal state allocation on the domain defined by the 2-PC.

We also, present results for the case of $a = \exp(-n^*_5)$ in Figures A.22, A.23 and A.24.

For the case were we have $a = \exp(-n^*_5)$ and give rise to a posterior distribution for N for the repulsive prior, $p(2) = 0.11474$, $p(3) = 0.18027$ $p(4) = 0.15573$, $p(5) = 0.12996$, $p(6) = 0.11106$, ... $p(21) = 0.00003$, with $\sum_{i=2}^{21} p(i) = 1$. We observe that the posterior distributions for the parameter N are almost the same either we use percentage 2.5% or 5% with the latter one giving slightly more mass to slightly larger number of states which is expected since it places a smaller penalty.

Next, we assess the convergence of our RJMCMC algorithm. We run a chain for 150000 iterations. The number of iterations was chosen such that the Monte Carlo error (MCE) for each parameter estimation is sufficiently small with respect to the scale of the parameter. In Table A.5 we display the Z-value of the Geweke statistic, the effective sample size (ESS) and the MCE.

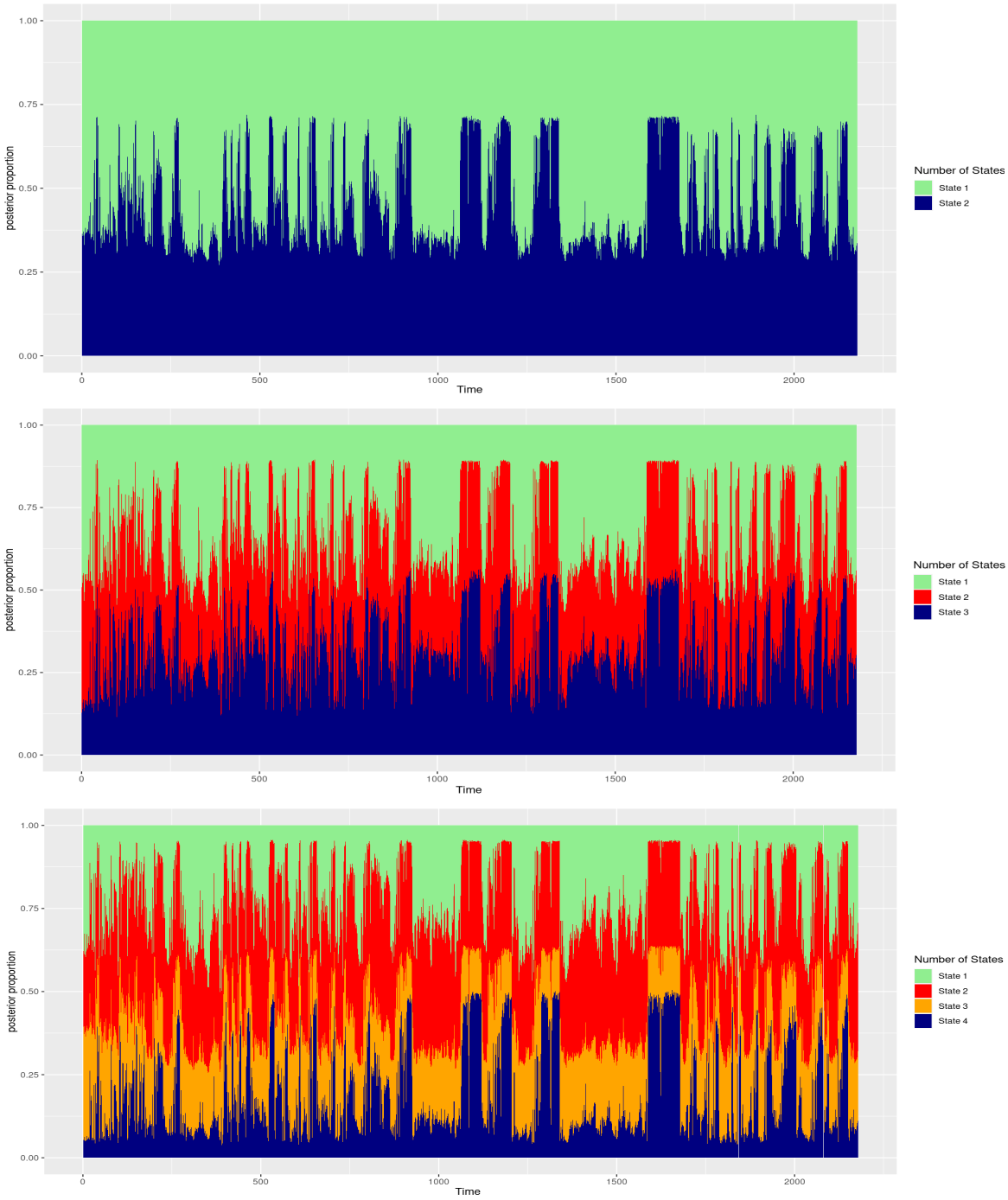

Fig. A.19 Posterior uncertainty probabilities of classification for the two, three and four state mixture models. On top row we have two states corresponding to (blue) floating on the water and (green) flying. On the middle row we have three states corresponding to (blue) floating on the water, (green) flying and (red) diving/nuisance. Lastly, on the bottom row, we have (blue) floating on the water, (green) flying, (red) diving/nuisance and (orange) floating/flying/nuisance.

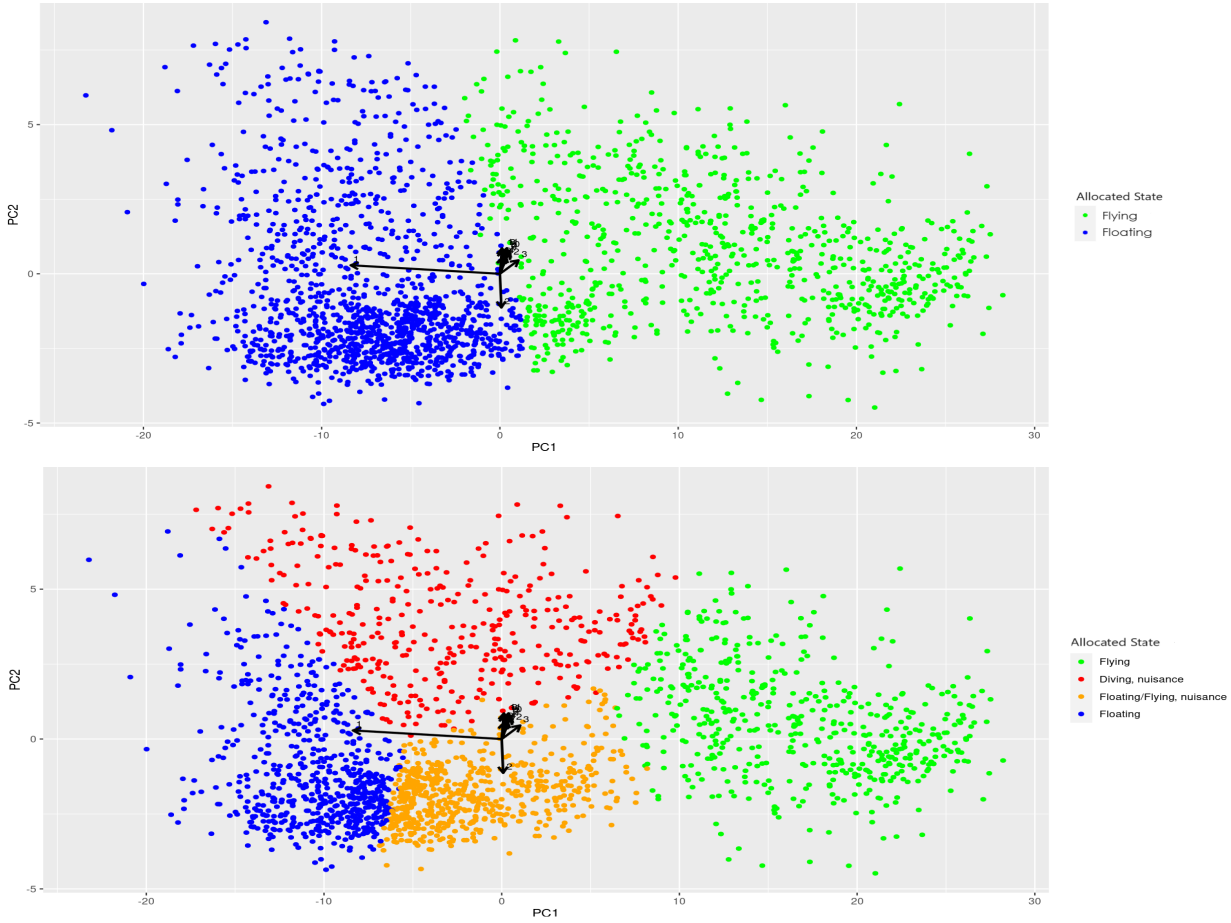

Fig. A.20 Biplot, with observations coloured according to their modal state allocation, in the case of two states (top row) and four states (bottom row), plotted on the domain of the first two PC.

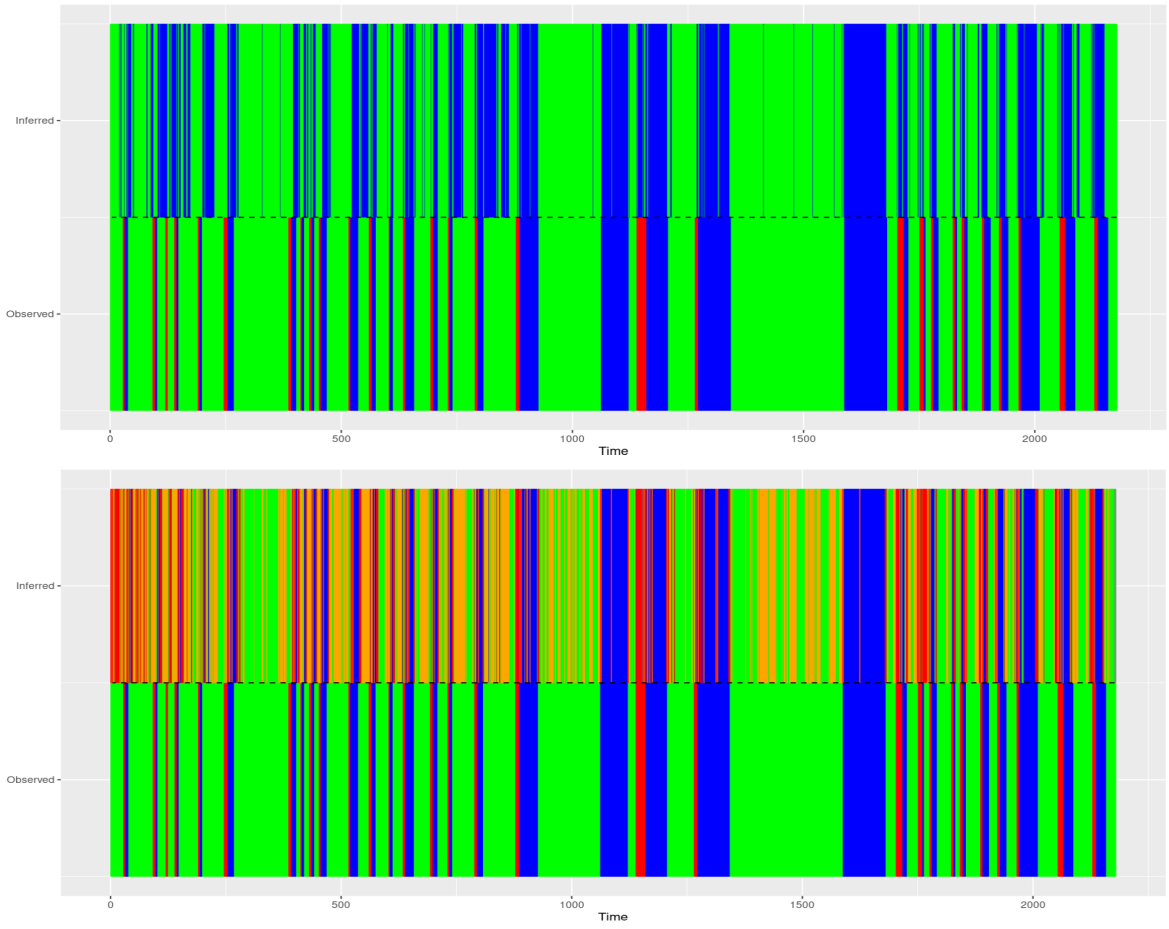

Fig. A.21 Comparison of the posterior classification of our model, for two states (top row) and four states (bottom row) with the manual classification of Thiebault et al. (2021) (on the bottom half of each row). Based on the manual classification of Thiebault et al. (2021), the states are: floating on water (blue), flying (green) and diving (red). In our model the states are: (blue) floating on the water, (green) flying, (red) diving/nuisance, (orange) floating/flying/nuisance.

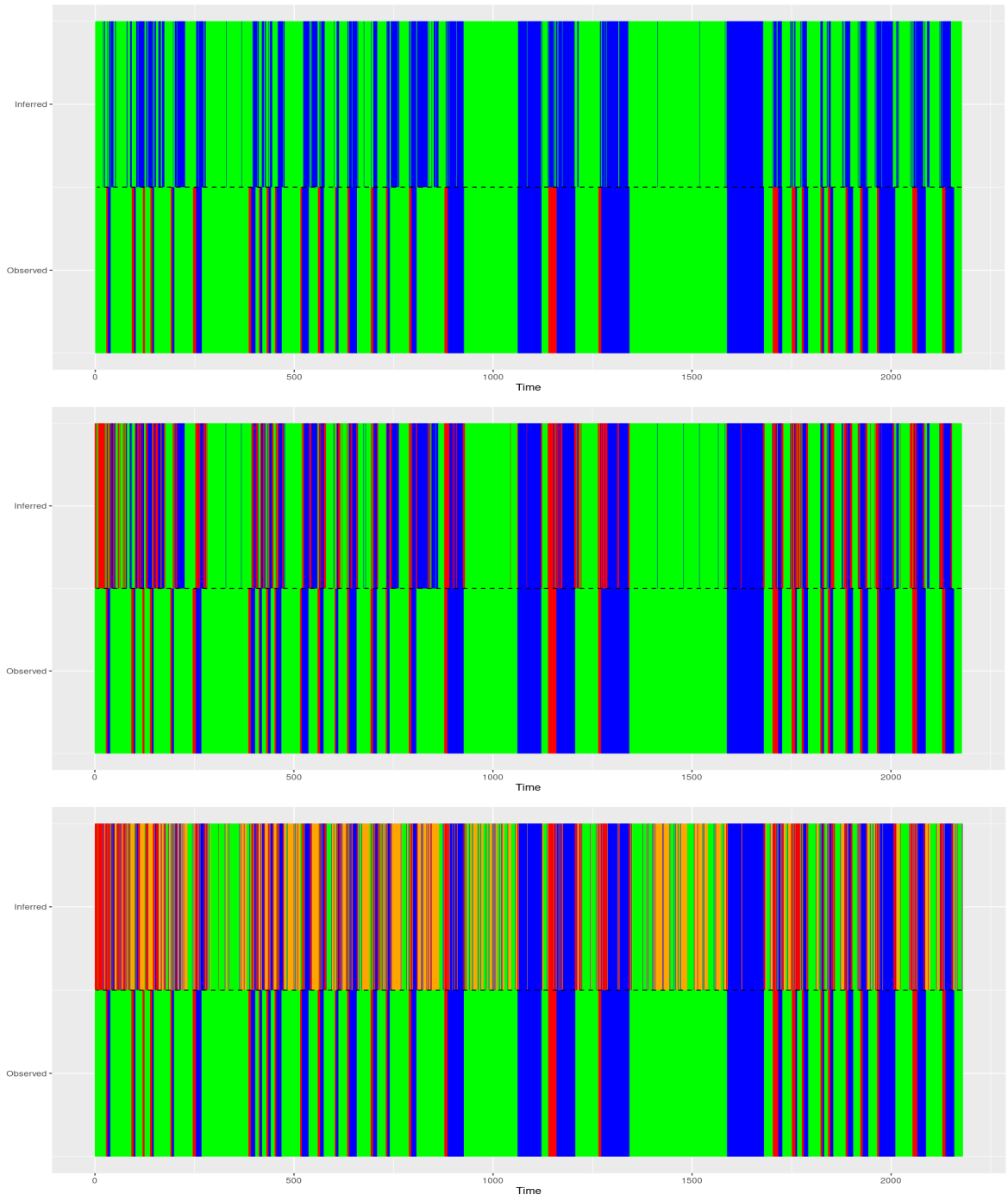

Fig. A.22 Comparison of the posterior classification of our model, for two states (top row) and four states (bottom row) with the manual classification of Thiebault et al. (2021) (on the bottom half of each row). Based on the manual classification of Thiebault et al. (2021), the states are: floating on water (blue), flying (green) and diving (red). In our model the states are: (blue) floating on the water, (green) flying, (red) diving/nuisance, (orange) floating/flying/nuisance.

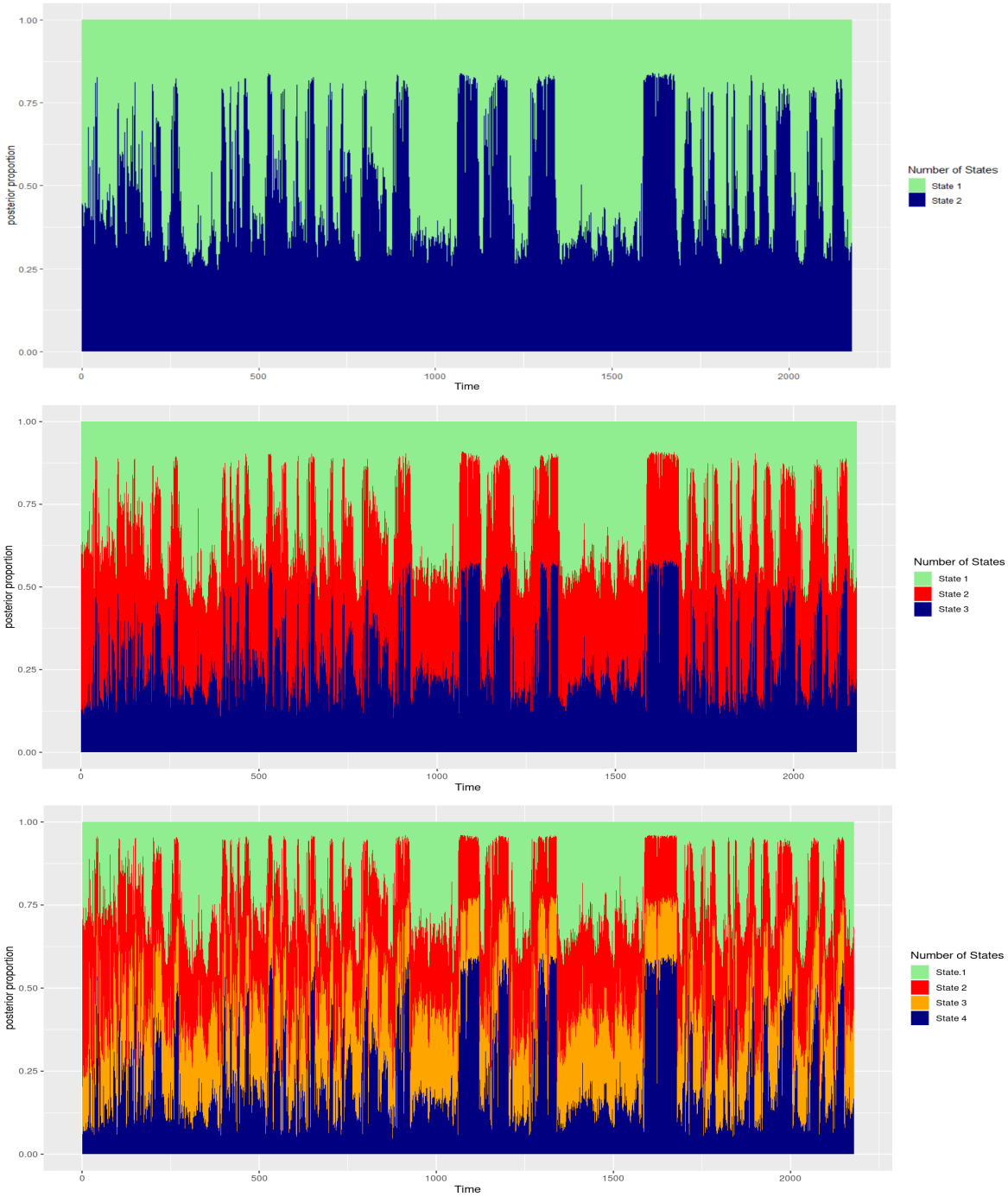

Fig. A.23 Posterior uncertainty probabilities of classification for the two, three and four state mixture models. On top row we have two states corresponding to (blue) floating on the water and (green) flying. On the middle row we have three states corresponding to (blue) floating on the water, (green) flying and (red) diving/nuisance. Lastly, on the bottom row, we have (blue) floating on the water, (green) flying, (red) diving/nuisance and (orange) floating/flying/nuisance.

Fig. A.24 Biplot, with observations coloured according to their modal state allocation, in the case of two (top row), three (middle row) and four states (bottom row), plotted on the domain of the first two PC.

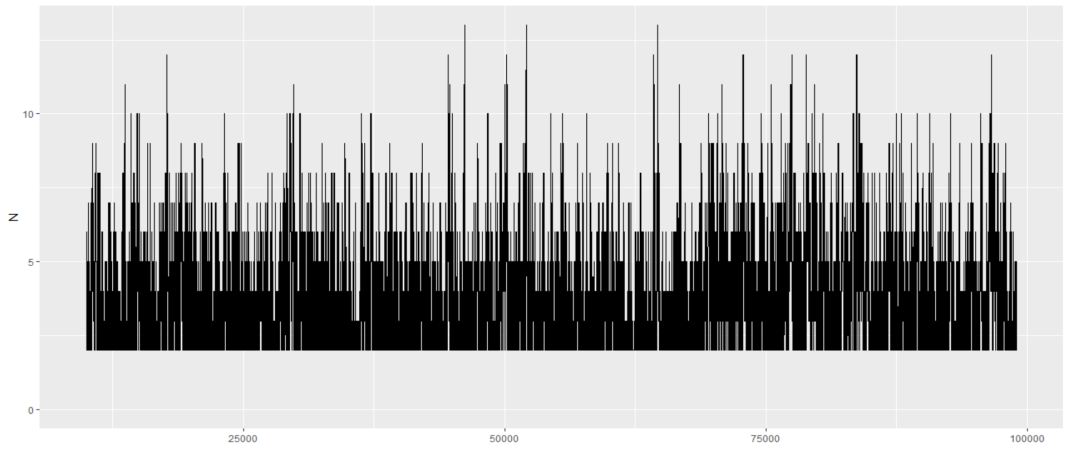

Fig. A.25 Iterations of number of components N, with burn-in iterations removed.

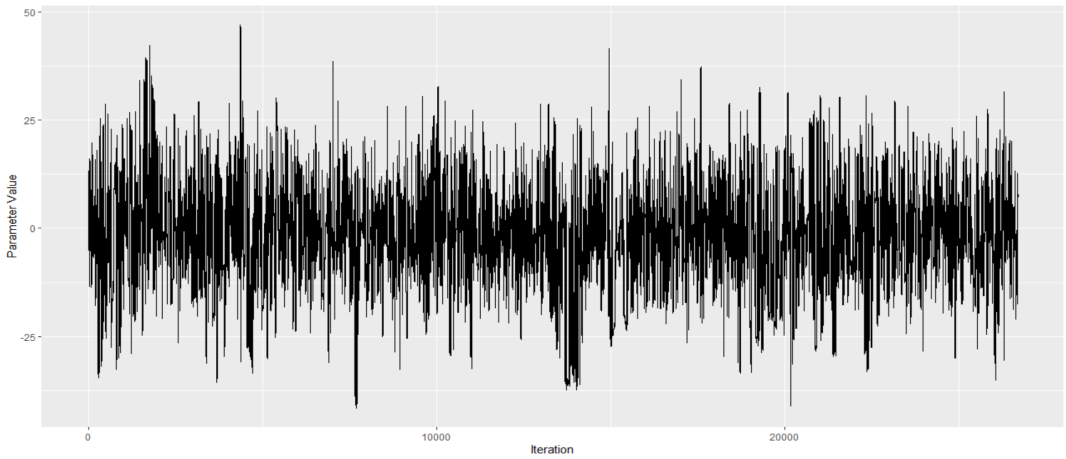

Fig. A.26 Iterations of mean $\mu_{1,1}$, with burn-in iterations removed.

Lasty, we display the trace plot for the parameters of the number of components N and means $\underline{\mu_i}$, $i = 1,2,3$ in Figures A.25-A.31.

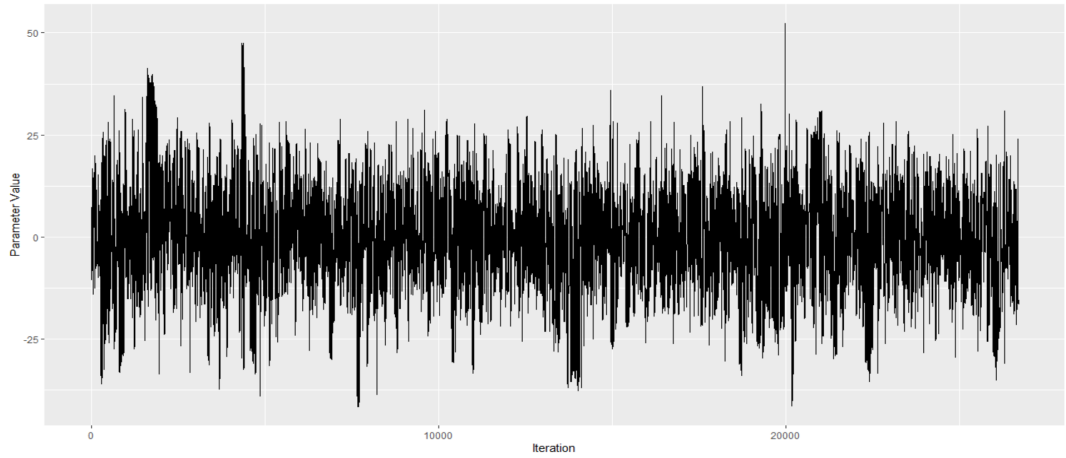

Fig. A.27 Iterations of mean $\mu_{2,1}$, with burn-in iterations removed.

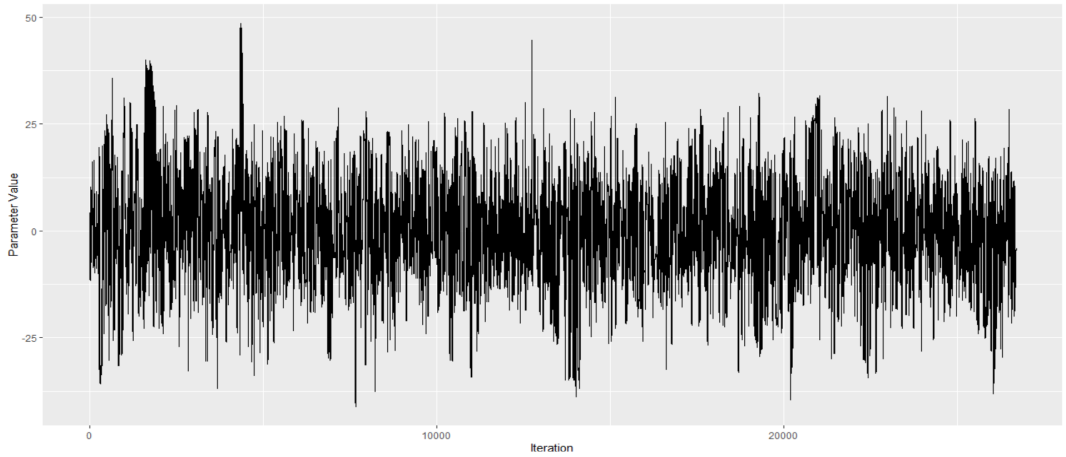

Fig. A.28 Iterations of mean $\mu_{3,1}$, with burn-in iterations removed.

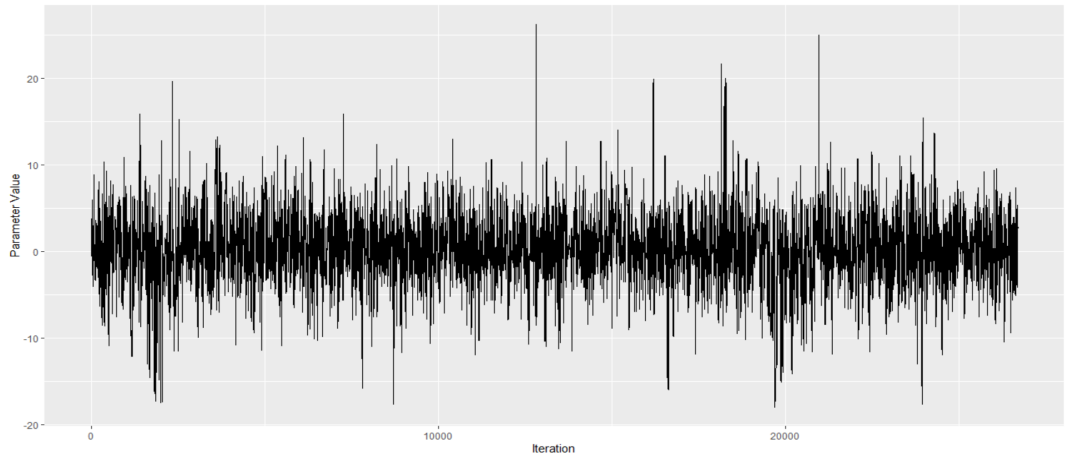

Fig. A.29 Iterations of mean $\mu_{1,2}$, with burn-in iterations removed.

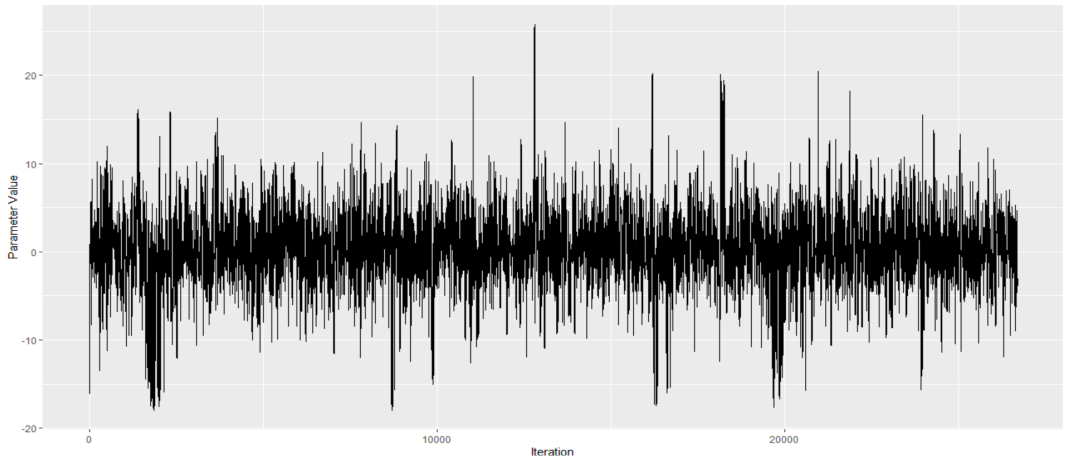

Fig. A.30 Iterations of mean $\mu_{2,2}$, with burn-in iterations removed.

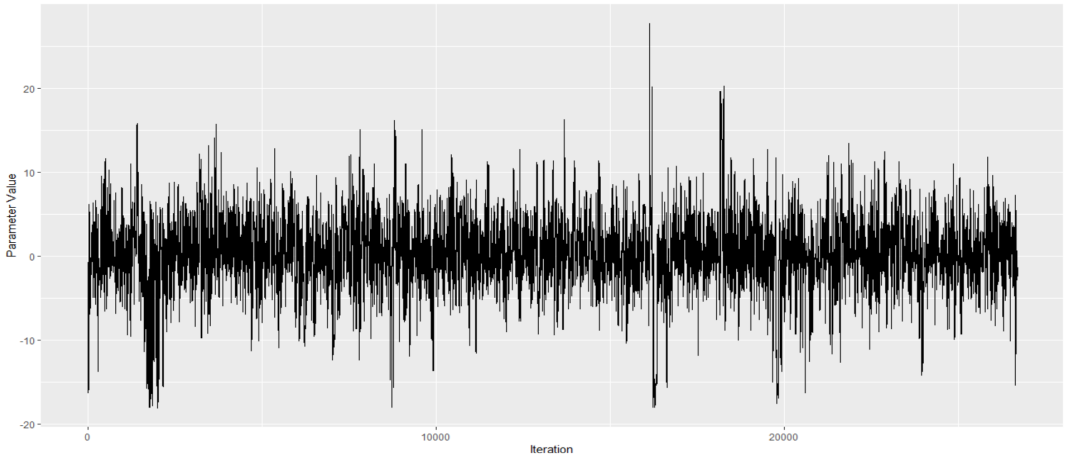

Fig. A.31 Iterations of mean $\mu_{3,2}$, with burn-in iterations removed.

Table A.4 The Geweke Z, ESS, and MCE statistics for the parameters mean μ_i, standard deviation σ_i, location m_i, concentration k_i and unormalized initial and transition probabilities λ_i and $\Lambda_{i,j}$ for the most frequently visited mode $N = 4$, i.e. , $i, j = 1, 2, 4$.

Parameters	**Geweke**	**ESS**	**MCE**
μ_1	-1.00	500	0.0040
μ_2	-0.24	300	0.0049
μ_3	-0.22	350	0.0490
μ_4	-0.90	658	0.2037
σ_1	-1.11	510	0.0038
σ_2	-0.22	355	0.0409
σ_3	0.29	325	0.2004
σ_4	-0.58	868	0.1934
k_1	0.11	1864	0.0005
k_2	0.52	5222	0.0004
k_3	-0.25	4318	0.0010
k_4	-0.43	300	0.0062
m_1	0.19	2000	0.0009
m_2	0.12	1894	0.0009
m_3	-1.45	1438	0.0010
m_4	-1.46	8208	0.0001
λ_1	0.73	16895	0.0084
λ_2	0.47	13747	0.0084
λ_3	-0.39	12603	0.0074
λ_4	0.50	13572	0.0073
$\Lambda_{1,2}$	1.49	250	0.0135
$\Lambda_{1,3}$	1.23	220	0.0180
$\Lambda_{1,4}$	0.93	286	0.0060
$\Lambda_{2,1}$	-0.61	180	0.1045
$\Lambda_{2,3}$	-0.36	210	0.0168
$\Lambda_{2,4}$	-0.17	173	0.0035
$\Lambda_{3,1}$	0.72	154	0.0102
$\Lambda_{3,2}$	1.20	166	0.0285
$\Lambda_{3,4}$	1.33	198	0.0996
$\Lambda_{4,1}$	-1.24	171	0.0204
$\Lambda_{4,2}$	-1.37	221	0.0181
$\Lambda_{4,3}$	-1.76	231	0.0207
z_1	0.94	3117	0.0004
z_2	-0.26	1132	0.0001
z_3	1.39	1669	0.0001
z_4	-0.52	3093	0.0001

Table A.5 The Geweke Z, ESS, and MCE statistics for the parameters mean $\underline{\mu_i}$, covariance matrix Σ_i, unormalized initial and transition probabilities λ_i and $\Lambda_{i,j}$ for the most frequently visited mode $N = 3$, i.e. , $i, j = 1, 2, 3$.

Parameters	Geweke Z	ESS	MCE
$\mu_{1,1}$	-1.34	1388	0.3393
$\mu_{2,1}$	-0.10	1343	0.3574
$\mu_{3,1}$	-0.28	1316	0.3740
$\mu_{1,2}$	-1.27	1920	0.0940
$\mu_{2,2}$	-1.02	1353	0.1186
$\mu_{3,2}$	-1.06	1100	0.1490
$\Sigma_1^{1,1}$	0.94	1290	7.0214
$\Sigma_1^{2,1}$	0.25	3249	0.1315
$\Sigma_1^{2,2}$	-0.21	1406	0.6328
$\Sigma_2^{1,1}$	0.43	2401	8.1524
$\Sigma_2^{2,1}$	-1.6	3985	0.1155
$\Sigma_2^{2,2}$	-0.89	2333	0.4816
$\Sigma_3^{1,1}$	-0.55	1716	10.641
$\Sigma_3^{2,1}$	-0.74	2507	0.1740
$\Sigma_3^{2,2}$	0.27	1833	0.5268
λ_1	1.03	6560	0.0767
λ_2	0.00	7757	0.0643
λ_3	-0.24	6129	0.0880
$\Lambda_{1,2}$	-1.36	3962	0.1572
$\Lambda_{1,3}$	-0.94	6154	0.1307
$\Lambda_{2,1}$	-0.36	7032	0.8717
$\Lambda_{2,3}$	0.80	9141	0.7749
$\Lambda_{3,1}$	-0.87	3326	1.1776
$\Lambda_{3,2}$	0.08	13051	4.4860

Appendix B

A Pólya Tree modelling framework for batch-mark data

B.1 Inference

We describe the updating steps of the Markov Chain Monte Carlo (MCMC) along with the corresponding posterior and proposal distributions.

B.1.1 Standard Batch Mark

Update of M^k

The update of the matrix M^k, $k = 1,2,...,K$ follows a random walk MH update. Given a matrix M^k, we propose a new matrix $M^{k,*}$, by proposing to remove λ individuals from a randomly selected cell, $(t_i,t_{i+1}) \times (t_j,t_{j+1})$ and add it to a different randomly selected cell $(t_{i^*},t_{i^*+1}) \times (t_{j^*},t_{j^*+1})$. The parameter λ, is sampled from a Discrete Uniform distribution, on the interval $[L_\lambda, U_\lambda]$ where L_λ, U_λ are chosen to achieve optimal mixing.

The posterior distribution can be written as

$$p(\left\{M_{i,j}^{k}\right\}_{i=0,j=i}^{k-1,K} | \left\{w_{i,j}\right\}_{i=0,j=i}^{k-1,K}, \{p_t\}_{t=1}^{K}, \{m_{k,t}\}_{t=k+1}^{K}, \{u_t\}_{t=1}^{k}, n) \propto$$
$$\text{Multinomial}(\left\{M_{i,j}^{k}\right\}_{i=0,j=i}^{k-1,K}; u_k, \left\{w_{i,j}\right\}_{i=0,j=i}^{k-1,K}) \prod_{t=k+1}^{K} \text{Binomial}(m_{k,t}; \sum_{i=0}^{k-1}\sum_{j=t}^{K} M_{i,j}^{k}, p_t) \prod_{t=1}^{k} \text{Binomial}(u_t; U_t, p_t)$$

$$p(\left\{M_{i,j}^{k,*}\right\}_{i=0,j=i}^{k-1,K} | \left\{w_{i,j}\right\}_{i=0,j=i}^{k-1,K}, \{p_t\}_{t=1}^{K}, \{m_{k,t}\}_{t=k+1}^{K}, \{u_t\}_{t=1}^{k}, n) \propto$$
$$\text{Multinomial}(\left\{M_{i,j}^{k,*}\right\}_{i=0,j=i}^{k-1,K}; u_k, \left\{w_{i,j}\right\}_{i=0,j=i}^{k-1,K}) \prod_{t=k+1}^{K} \text{Binomial}(m_{k,t}; \sum_{i=0}^{k-1}\sum_{j=t}^{K} M_{i,j}^{k,*}, p_t) \prod_{t=1}^{k} \text{Binomial}(u_t; U_t^*, p_t)$$

where U_t and U_t^*, are derived from the matrices $\left\{M^k\right\}_{k=1}^{K}, n$ and $\left\{M^{k,*}\right\}_{k=1}^{K}, n$ respectively.

Update of n

The update of the matrix n, follows a random walk MH update. Given a matrix n, we propose a new matrix n^*, by proposing to add or remove λ individuals from a randomly selected cell of the matrix, $n_{i,j}$ of matrix n, $(t_i, t_{i+1}) \times (t_j, t_{j+1})$ to a different randomly selected $(t_{i^*}, t_{i^*+1}) \times (t_{j^*}, t_{j^*+1})$ and derive the candidate matrix n^*. The parameter λ is sampled from a balanced mixture of two Discrete Uniform distributions, one on positive and one on negative numbers respectively.

The posterior distribution can be written as

$$p(\left\{n_{i,j}\right\}_{i=0,j=i}^{K,K} | \omega, \left\{w_{i,j}\right\}_{i=0,j=i}^{K,K}, \{p_t\}_{t=1}^{K}, \{m_{k,t}\}_{k=1,t=k+1}^{K,K}, \{u_t\}_{t=1}^{K}, \left\{M^k\right\}_{k=1}^{K}) \propto$$
$$\text{Poisson}(\left\{n_{i,j}\right\}_{i=0,j=i}^{K,K}; \omega \times \left\{w_{i,j}\right\}_{i=0,j=i}^{K,K}) \prod_{k=1}^{K} \prod_{t=k+1}^{K} \text{Binomial}(m_{k,t}; \sum_{i=0}^{k-1}\sum_{j=t}^{K} M_{i,j}^{k}, p_t) \prod_{t=1}^{K} \text{Binomial}(u_t; U_t, p_t)$$

$$p(\left\{n_{i,j}^{*}\right\}_{i=0,j=i}^{K,K} | \omega, \left\{w_{i,j}\right\}_{i=0,j=i}^{K,K}, \{p_t\}_{t=1}^{K}, \{m_{k,t}\}_{k=1,t=k+1}^{K,K}, \{u_t\}_{t=1}^{K}, \left\{M^k\right\}_{k=1}^{K}) \propto$$
$$\text{Poisson}(\left\{n_{i,j}^{*}\right\}_{i=0,j=i}^{K,K}; \omega \times \left\{w_{i,j}\right\}_{i=0,j=i}^{K,K}) \prod_{k=1}^{K} \prod_{t=k+1}^{K} \text{Binomial}(m_{k,t}; \sum_{i=0}^{k-1}\sum_{j=t}^{K} M_{i,j}^{k}, p_t) \prod_{t=1}^{K} \text{Binomial}(u_t; U_t^*, p_t)$$

Update of $\left\{w_{i,j}\right\}$

For the update of grid probabilities $\left\{w_{i,j}\right\}_{i=0,j=i}^{K,K}$, since the PT is conjugate, there is a closed expression

for the posterior distribution. Consequently, we utilize a Gibbs sampling algorithm. The posterior in that case can be calculated as

$$
\begin{aligned}
&p(\left\{w_{i,j}\right\}_{i=0,j=i}^{K,K}|n,\omega,\{a_i\}_{i=0}^{K},\left\{a_{i,j}\right\}_{i=0,j=i}^{K,K}) \propto \\
&\quad \text{Dirichlet}(V_0,V_1,...,V_K;a_0,a_1,...,a_K)\prod_{i=0}^{K}\text{Dirichlet}(V_{i,i},...,V_{i,K};a_{i,i},...,a_{i,K}) \\
&\quad \text{Poisson}(\left\{n_{i,j}\right\}_{i=0,j=i}^{K,K};\omega\times\left\{w_{i,j}\right\}_{i=0,j=i}^{K,K}) \propto \\
&\quad \text{Dirichlet}(V_0,V_1,...,V_K;a_0,a_1,...,a_K)\prod_{i=0}^{K}\text{Dirichlet}(V_{i,i},...,V_{i,K};a_{i,i},...,a_{i,K}) \\
&\quad \prod_{i=0}^{K}\prod_{j=i}^{K}w_{i,j}^{n_{i,j}} \propto \\
&\quad \text{Dirichlet}(V_0,V_1,...,V_K;a_0,a_1,...,a_K)\prod_{i=0}^{K}\text{Dirichlet}(V_{i,i},...,V_{i,K};a_{i,i},...,a_{i,K}) \\
&\quad \prod_{i=0}^{K}\prod_{j=i}^{K}(V_iV_{i,j})^{n_{i,j}}
\end{aligned}
$$

We can identify a conjugacy between the prior Dirichlet distribution and the product of elements $(V_iV_{i,j})^{n_{i,j}}$, which results again in a posterior Dirichlet distribution

$$
\begin{aligned}
(V_0,V_1,...,V_K) &\sim \text{Dirichlet}(a_0+n_0,a_1+n_1,...,a_K+n_K),\ \ n_i=\sum_{j=i}^{K}n_{i,j} \\
(V_{i,i},...,V_{i,K}) &\sim \text{Dirichlet}(a_{i,i}+n_{i,i},...,a_{i,K}+n_{i,K}),\ \ i=0,1,2,...,K
\end{aligned}
$$

Update ω

The posterior of ω can be written as

$$
\begin{aligned}
&p(\omega|N,\left\{w_{i,j}\right\}_{i=0,j=i}^{K,K},a_\omega,b_\omega) \propto \text{Gamma}(\omega;a_\omega,b_\omega)\text{Poisson}(N;\omega) \propto \\
&\quad \omega^{a_\omega}e^{-b_\omega\omega}\omega^{N}e^{-\omega} \propto \omega^{N+a_\omega}e^{-(b_\omega+1)\omega}
\end{aligned}
$$

Hence, we have the posterior distribution of ω is

$$\omega \sim \text{Gamma}(N + a_\omega + 1, b_\omega + 1)$$

where a_ω, b_ω are the prior distribution parameters.

Update of p_t

The posterior distribution of the capture probabilities p_t can be expressed as

$$\begin{aligned}
&p(p_t | \left\{M^k\right\}_{k=1}^K, n, \{m_{k,t}\}_{k=1}^K, u_t, a_p, b_p) \propto \\
&Beta(p_t; a_p, b_p)\text{Binomial}(u_t; U_t, p_t)\prod_{k=1}^{t-1}\text{Binomial}(m_{k,t}; \sum_{i=0}^{k-1}\sum_{j=t}^{K} M_{i,j}^k, p_t) \propto \\
&p_t^{a_p-1}(1-p_t)^{1-b_p} p_t^{u_t}(1-p_t)^{U_t-u_t}\prod_{k=1}^{t-1} p_t^{m_{k,t}}(1-p_t)^{\sum_{i=0}^{k-1}\sum_{j=t}^{K} M_{i,j}^k - m_{k,t}} \propto \\
&p_t^{u_t+\sum_{k=1}^{t-1} m_{k,t}+a_p-1}(1-p_t)^{U_t+\sum_{k=1}^{t-1}\sum_{i=0}^{k-1}\sum_{j=t}^{K} M_{i,j}^k + b_p - u_t - \sum_{k=1}^{t-1} m_{k,t} - 1}
\end{aligned}$$

Hence, the posterior distribution of p_t is

$$p_t \sim \text{Beta}(u_t + \sum_{k=1}^{t-1} m_{k,t} + a_p, U_t + \sum_{k=1}^{t-1}\sum_{i=0}^{k-1}\sum_{j=t}^{K} M_{i,j}^k + b_p), \;\; t = 1, 2, ..., K$$

where a_p, b_p are the prior distribution parameters.

B.1.2 Case Study 1

Update of R^k:

In case study 1, on which we remove individuals in each sampling occasion, we have to update the additional matrices R^k, $k = 1, 2, ..., K$ and for that we follow a random walk MH update. Given the current matrix R^k, then we propose to move to the matrix $R^{k,*}$, by proposing to remove λ individuals

from a randomly selected cell, $(t_i, t_{i+1}) \times (t_j, t_{j+1})$ and add it to a different randomly selected cell $(t_{i^*}, t_{i^*+1}) \times (t_{j^*}, t_{j^*+1})$ and derive the candidate matrix $R^{k,*}$. The parameter λ is sampled in similarly to the update of M^k.

The posterior distribution can be written as

$$
\begin{aligned}
&p(\left\{R^k_{i,j}\right\}^{K,k}_{i=0,j=k} \mid \left\{w_{i,j}\right\}^{K,k}_{i=0,j=k}, \{p_t\}^K_{t=1}, \{u_t\}^K_{t=1}, n, \left\{M^k\right\}^K_{k=1}) \propto \\
&\quad \text{Multinomial}(\left\{R^k_{i,j}\right\}^{K,k}_{i=0,j=k}; r_k, \left\{w_{i,j}\right\}^{K,k}_{i=0,j=k}) \prod_{t=1}^{k} \text{Binomial}(u_t; U_t, p_t) \\
&p(\left\{R^{k,*}_{i,j}\right\}^{K,k}_{i=0,j=k} \mid \left\{w_{i,j}\right\}^{K,k}_{i=0,j=k}, \{p_t\}^K_{t=1}, \{u_t\}^K_{t=1}, n, \left\{M^k\right\}^K_{k=1}) \propto \\
&\quad \text{Multinomial}(\left\{R^{k,*}_{i,j}\right\}^{K,k}_{i=0,j=k}; r_k, \left\{w_{i,j}\right\}^{K,k}_{i=0,j=k}) \prod_{t=1}^{k} \text{Binomial}(u_t; U^*_t, p_t)
\end{aligned}
$$

where $U_t = \sum_{i=0}^{k-1} \sum_{j=t}^{K} U^t_{i,j}$ similar for U^*_t, are the number of unmarked individuals on sampling occasion t. Those are calculated from the matrices $\left\{M^k\right\}^K_{k=1}, \left\{R^k\right\}^K_{k=1}, n$ and $\left\{M^k\right\}^K_{k=1}, \left\{R^{k,*}\right\}^K_{k=1}, n$ respectively.

The proposal distributions are $p(R^k|R^{k,*}) = p(R^{k,*}|R^k)$ and are exactly the same as the ones used for the updated of M^k.

B.1.3 Case Study 2

In case study 2, we adhere to the detailed instructions provided in Section B.1.1. The only modification made is in the likelihood functions, where they now accommodate multiple secondary sampling occasions. The priors and proposals, however, remain unchanged.

B.2 Standard Batch Mark Implementation

The parameters used for the generation of this simulation study are:

- $K = 5$ sampling occasions
- population size $N = 6000$

- time varying entry probabilities $V_0 = 0.12963746, V_1 = 0.26378673, V_2 = 0.17323397, V_3 = 0.34080943, V_4 = 0.07646343, V_5 = 0.01606898$
- time varying exit probabilities $V_{.,1} = 0.25619101, 0.36269471, V_{.,2} = 0.22757203, V_{.,3} = 0.06916488, V_{.,4} = 0.05717003, V_{.,5} = 0.02720733$, where $V_{.,j}$ indicates the exiting probability between the jth and $(j+1)th$ sampling occasion regardless where they entered
- time varying capture probabilities $p_1 = 0.4646893, p_2 = 0.3280358, p_3 = 0.4432990, p_4 = 0.2778812, p_5 = 0.5206719$

For the generation of the data, we first generate the matrix n of dimensions $(5+1) \times (5+1)$, based on the population size, and entry/exit probabilities. On the 1st sampling occasion there are $N_1 = U_1 = \sum_{i=0}^{1-1} \sum_{j=i}^{5} n_{i,j}$ available individuals, from which we capture u_1 unmarked with probability p_1. Those individuals u_1 form the matrix M^1 of dimensions $(5+1) \times (5+1)$. Then on the 2nd sampling occasion the available number of of unmarked individuals, since a portion has already been marked, is $U_2 = \sum_{i=0}^{2-1} \sum_{j=i}^{5} (n_{i,j} - M^1_{i,j})$ from which we capture u_2 unmarked individuals with probability p_2. On the 2nd sampling occasion we have the possibility of recapturing already marked individuals, such available individuals can be calculated as $\sum_{i=0}^{1-1} \sum_{j=i}^{5} M^1_{i,j}$, from which we recapture $m_{1,2}$ marked individuals with probability p_2. We repeat this process up and to the last sampling occasion. This leads to the generation of the following data

- $u_1, ..., u_5$ the number of captured unmarked individuals per sampling occasion
- $m_{1,2}, ..., m_{1,5}, m_{2,3}, ..., m_{2,5}, ..., m_{4,5}$ the number of recaptures of marked individuals

We used an MCMC algorithm, with 100,000 total iterations where the first 10,000 correspond to burn-in iterations. We followed the updating steps as described in Appendix B, Section B.1. We have chosen flat priors for all model parameters.

- $p_t \sim \text{Beta}(a_p = 1, b_p = 1)$
- $\omega \sim \text{Gamma}(a_\omega = 0.001, b_\omega = 0.001)$
- $(V_0, V_1, ..., V_5) \sim \text{Dirichlet}(a_0 = 1, a_1 = 1, ..., a_5 = 1)$

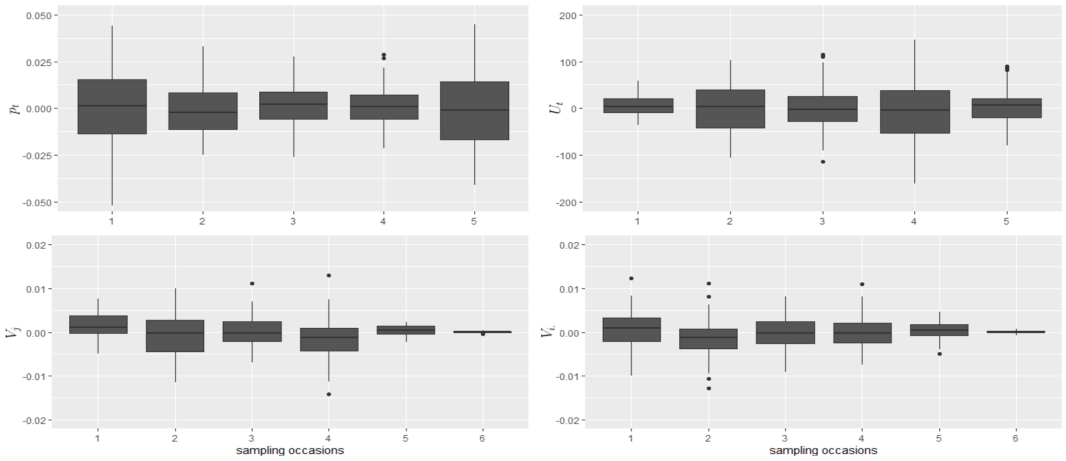

Fig. B.1 Posterior bias summaries of capture probabilities p_t, (a: top left), number of available unmarked individuals U_t, (b: top right), entry probabilities V_i, (c: bottom left), exit probabilities $V_{.,j}$, (d: bottom right), across sampling occasions $t = 1,2,...,5$ with $i = 0,1,...,5$ and $j = 0,...,5$.

- $(V_{.,0}, V_{.,1}, ..., V_{.,5}) \sim \text{Dirichlet}(a_{.,0} = 1, a_{.,1} = 1, ..., a_{.,5} = 1)$

We repeat the described simulation 100 times and we display the posterior bias, of the parameters in Figure B.1.

The results show that our framework for the standard batch mark data actually produces solid results since the bias of the parameters fluctuates around zero.

B.3 Case Study 1: Simulation

We ran 100 simulations for 200,000 Markov Chain Monte Carlo simulations each, across $K = 11$ sampling occasion for three scenarios. On each sampling occasion we removed a proportion of captured unmarked individual, the proportion was held constant across sampling occasions. The proportions of removed unmarked individuals for each scenario arose with removal probabilities 0.05, 0.1 and 0.2. The parameters used for the simulation study are the following, which are the same for each scenario:

- time varying entry probabilities $V_0 = 0.13, V_1 = 0.14, V_2 = 0.16, V_3 = 0.14, V_4 = 0.06, V_5 = 0.07, V_6 = 0.04, V_7 = 0.03, V_8 = 0.03, V_9 = 0.08, V_{10} = 0.10, V_{11} = 0.02$
- time varying exit probabilities $V_{.,0} = 0.02, V_{.,1} = 0.11, V_{.,2} = 0.01, V_{.,3} = 0.22, V_{.,4} = 0.12, V_{.,5} = 0.02, V_{.,6} = 0.01, V_{.,7} = 0.01, V_{.,8} = 0.17, V_{.,9} = 0.16, V_{.,10} = 0.12, V_{.,11} = 0.03$, where $V_{.,j}$ indi-

cates the exiting probability between the jth and $(j+1)th$ sampling occasion regardless where they entered

- time varying capture probabilities $p_1 = 0.51, p_2 = 0.24, p_3 = 0.21, p_4 = 0.25, p_5 = 0.15, p_6 = 0.15, p_7 = 0.05, p_8 = 0.16, p_9 = 0.23, p_{10} = 0.31, p_{11} = 0.13$
- population of $N = 6000$ individuals

The results are illustrated on Figures B.2, B.4, B.6. As we can observe the bias exists in all parameters $p_t, U_t, N_t, \phi_t, \beta_t$, when we do not account for the removed individuals. The bias is more significant for the population size parameters, since for the model PTBM we completely disregard a significant number of individuals that were removed from the population, $\{r_k\}_{k=1}^{K}$. Also, the bias becomes more prevalent as the number of removed individuals increases.

We used the following flat priors for the parameters:

- $p_t \sim \text{Beta}(a_p = 1, b_p = 1)$
- $\omega \sim \text{Gamma}(a_\omega = 0.001, b_\omega = 0.001)$
- $(V_0, V_1, ..., V_{11}) \sim \text{Dirichlet}(a_0 = 1, a_1 = 1, ..., a_{11} = 1)$
- $(V_{.,0}, V_{.,1}, ..., V_{.,11}) \sim \text{Dirichlet}(a_{.,0} = 1, a_{.,1} = 1, ..., a_{.,11} = 1)$

B.4 Weather-loach

In order to compare our model to the frequentist model, we choose flat prior for the parameters ω, p_t, ϕ, β.

- $p_t \sim Beta(a_p = 1, b_p = 1)$
- $\omega \sim \text{Gamma}(a_\omega = 0.001, \omega_b = 0.001)$
- $\phi \sim \text{Beta}(a_\phi = 1, b_\phi = 1)$
- $\beta \sim \text{Beta}(a_\beta = 1, b_\beta = 1)$

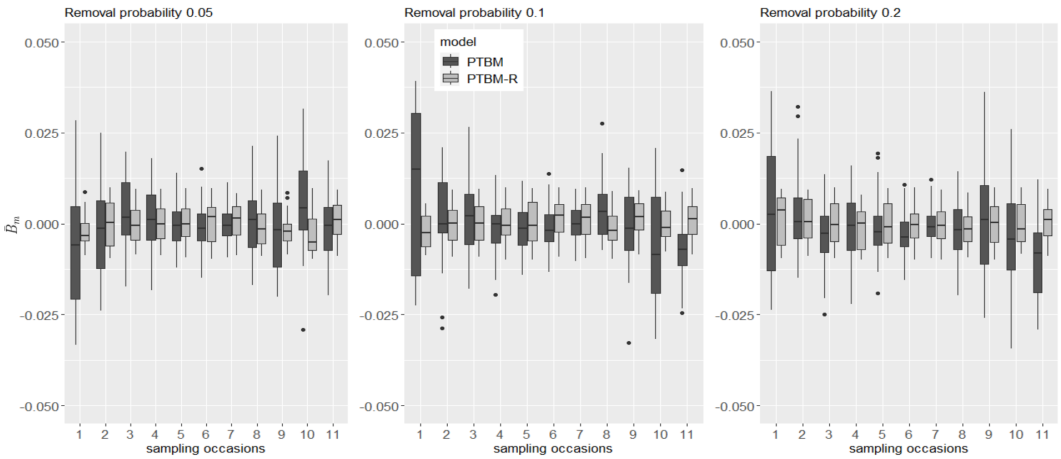

Fig. B.2 Posterior bias of capture probabilities, p_t, $t = 1,2,...,K$. Each column corresponds to a different removal probability. We compare the posterior bias across the models PTBM-R and PTBM, i.e. when we account for removals and when we do not.

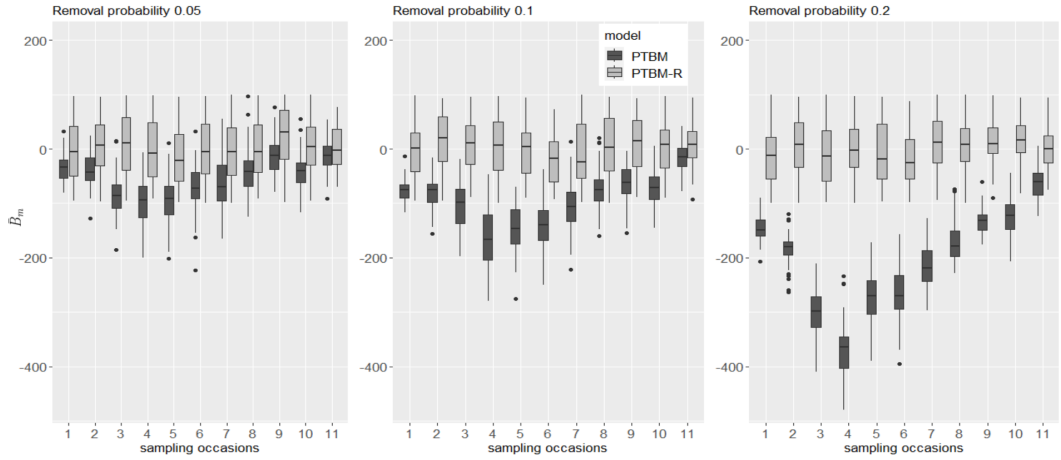

Fig. B.3 Posterior bias of available unmarked individuals, U_t, $t = 1,2,...,K$. Each column corresponds to a different removal probability. We compare the posterior bias across the models PTBM-R and PTBM, i.e. when we account for removals and when we do not.

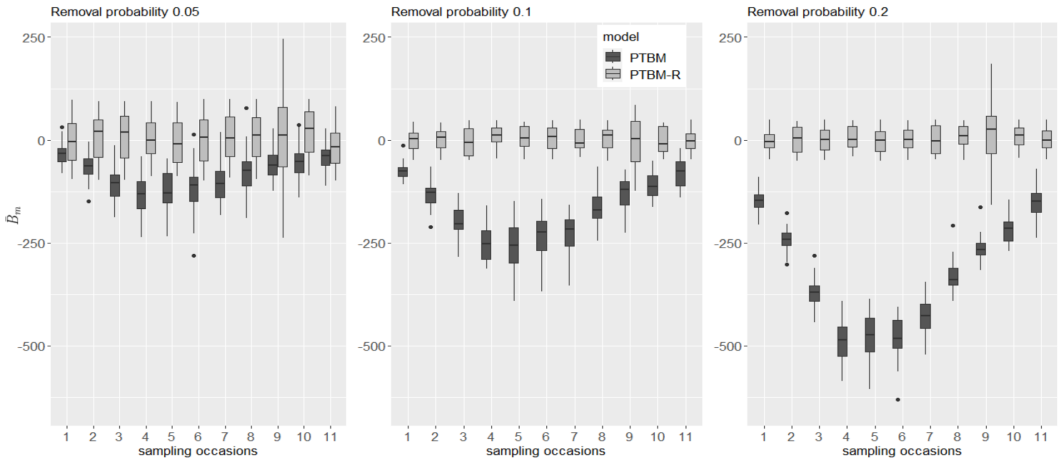

Fig. B.4 Posterior bias of available individuals, N_t, $t = 1,2,...,K$. Each column corresponds to a different removal probability. We compare the posterior bias across the models PTBM-R and PTBM, i.e. when we account for removals and when we do not.

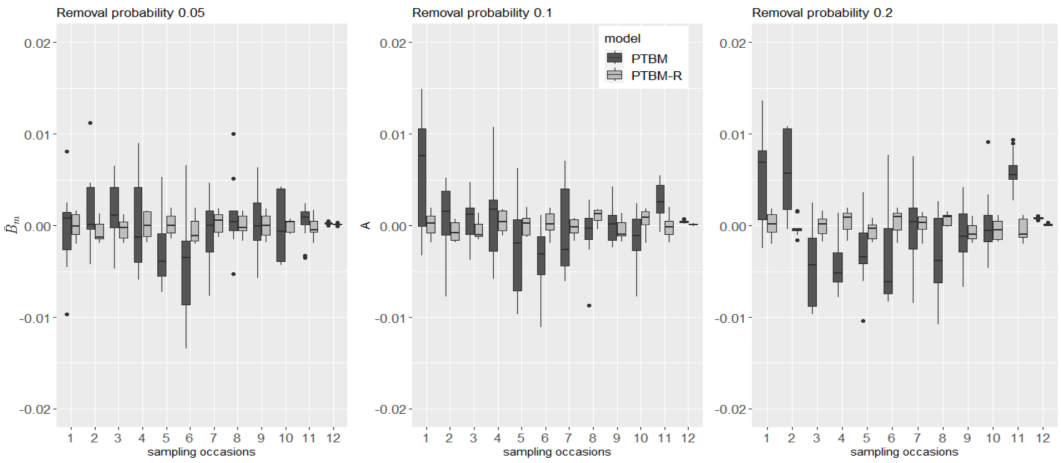

Fig. B.5 Posterior bias of exit probability, ϕ_t, $t = 1,2,...,K$. Each column corresponds to a different removal probability. We compare the posterior bias across the models PTBM-R and PTBM, i.e. when we account for removals and when we do not.

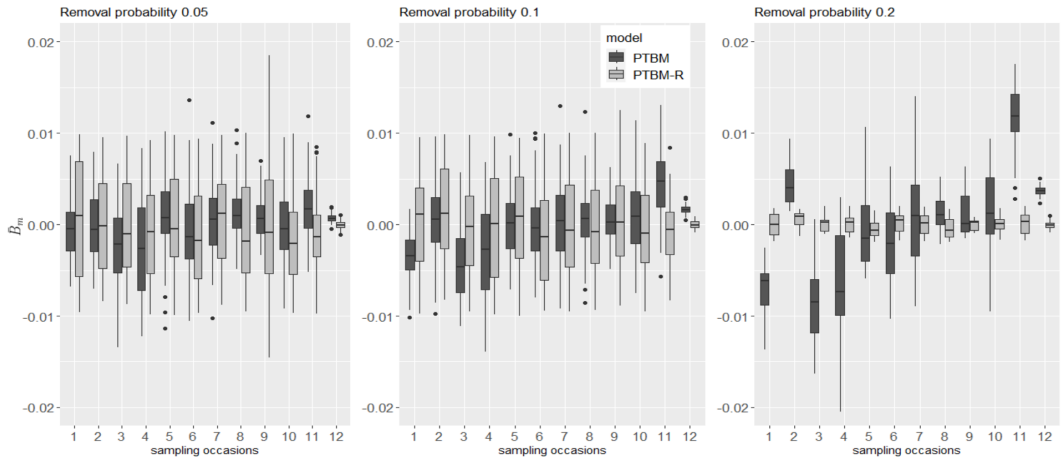

Fig. B.6 Posterior bias of entry probability, β_t, $t = 1,2,...,K$. Each column corresponds to a different removal probability. We compare the posterior bias across the models PTBM-R and PTBM, i.e. when we account for removals and when we do not.

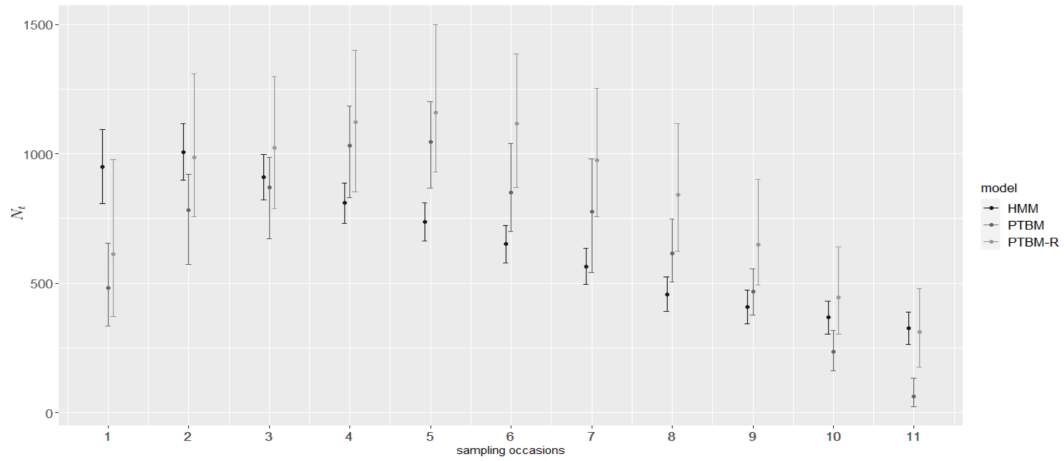

Fig. B.7 Number of available individual, N_t, on each sampling occasion $t = 1,2,...,11$, across each model.

In Figure B.7 we display the number of available individuals estimated under the PTBM-R, PTBM and HMM.

The goodness of fit check is displayed on Figure B.11. We can clearly observe that on average the model that accounts for removals PTBM-R manages to generalize much better the observed pattern compared to PTBM and HMM.

Next, we assess the convergence of our MCMC algorithm for the PTBM-R model. We considered 4 chains across 150000 iterations. The number of iterations was chosen such that the Monte Carlo error (MCE) for each parameter estimation is sufficiently small with respect to the scale of the parameter. In Table B.1 we display the Z-value of the Geweke statistic, the effective sample size (ESS), the MCE and the $\hat{R}$ statistic across the 4 chains.

Table B.1 The Geweke Z, ESS, MCE, and $\hat{R}$ statistics for the parameters population size N, entry/exit parameters β, ϕ and capture probabilities $p_1, p_2, ..., p_{11}$.

Parameters	**Geweke Z**	**ESS**	**MCE**	$\hat{R}$
N	(-0.59,0.58,0.96,-0.73)	(1488,1108,1542,1317)	(6.9381,8.3313,6.7276,7.6722)	1.008
$1-\phi$	(1.24,1.61,1.82,1.62)	(96,80,110,65)	(0.0027,0.0035,0.0024,0.0039)	1.009
β	(-1.22,-0.45,-0.18,-1.48)	(665,706,842,776)	(0.0005,0.0005,0.0005,0.0005)	0.9975
p_1	(0.22,-0.17,0.01,-0.21)	(763,614,750,655)	(0.0015,0.0018,0.0015,0.0017)	1.005
p_2	(-0.20,0.84,-0.33,0.24)	(793,556,689,571)	(0.0008,0.0010,0.0009,0.0010)	1.008
p_3	(0.12,0.84,-0.45,0.04)	(580,380,474,390)	(0.0009,0.0012,0.0010,0.0012)	1.006
p_4	(0.05,0.73,-0.78,0.35)	(384,244,334,242)	(0.0013,0.0019,0.0015,0.0019)	1.004
p_5	(0.06,0.71,-0.82,0.24)	(436,294,394,285)	(0.0009,0.0012,0.0010,0.0012)	1.000
p_6	(0.06,0.55,-0.93,0.38)	(415,285,381,266)	(0.0010,0.0013,0.0011,0.0014)	1.001
p_7	(0.08,0.55,-1.12,0.38)	(945,652,861,597)	(0.0003,0.0004,0.0004,0.0005)	1.000
p_8	(0.28,0.48,-1.13,0.33)	(462,298,403,288)	(0.0012,0.0016,0.0012,0.0016)	1.005
p_9	(0.26,0.56,-0.86,0.41)	(381,273,349,265)	(0.0019,0.0024,0.0020,0.0025)	1.006
p_{10}	(0.20,0.34,-1.00,0.30)	(390,278,365,277)	(0.0024,0.0031,0.0025,0.0032)	1.007
p_{11}	(-0.12,0.31,-1.30,0.37)	(555,422,544,389)	(0.0017,0.0020,0.0017,0.0022)	1.001

Lasty, we display the traceplot for the parameter N, β and ϕ in Figures B.8, B.9, B.10.

B.5 Case Study 2: Simulation

We ran 100 simulations for 500,000 Markov Chain Monte Carlo simulations. The parameters used for the simulation study are the following:

- $K = 5$ primary periods

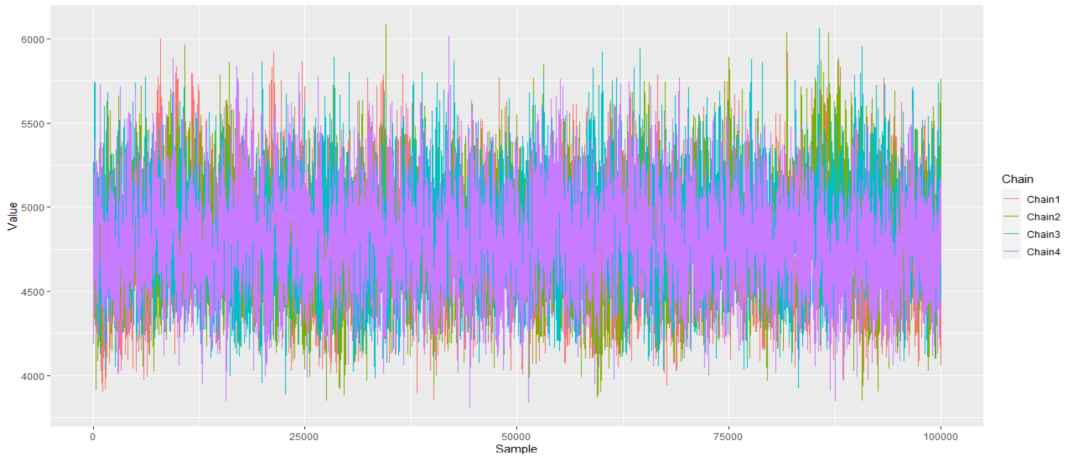

Fig. B.8 Iterations of abundance N across the four chains, with burn-in iterations removed.

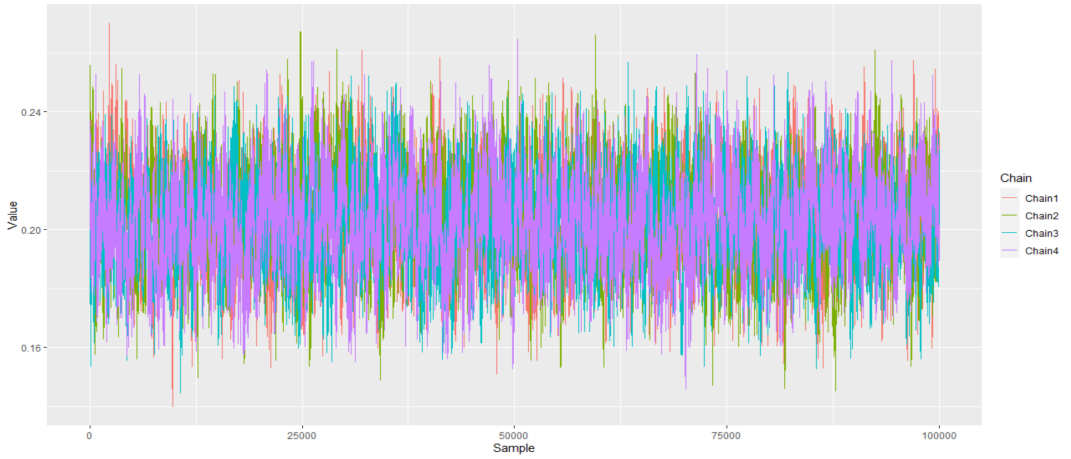

Fig. B.9 Iterations of entry probability β across the four chains, with burn-in iterations removed.

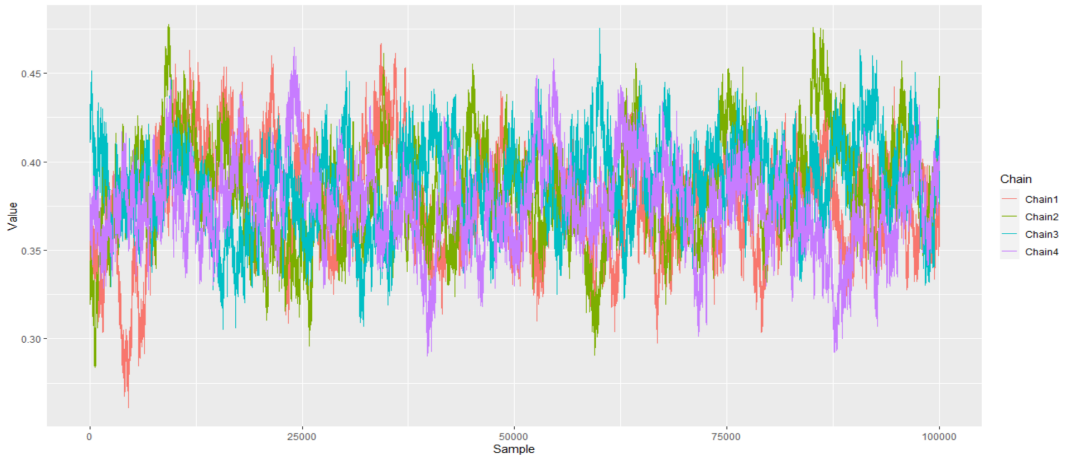

Fig. B.10 Iterations of exit probability ϕ across the four chains, with burn-in iterations removed.

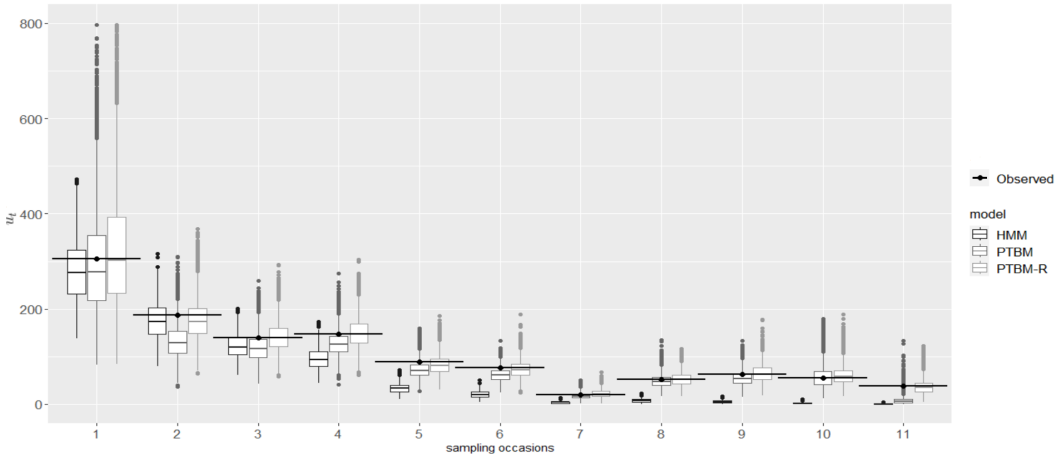

Fig. B.11 Number of unmarked individuals, u_t caught on each sampling occasion $t = 1,2,...,11$, together with the corresponding simulated values for each model.

- the entry probabilities are $V_0 = 0.41, V_1 = 0.53, V_2 = 0.49, V_3 = 0.57, V_4 = 0.31, V_5 = 0.48$
- the exit probabilities are $V_{.,0} = 0.34, V_{.,1} = 0.03, V_{.,2} = 0.28, V_{.,3} = 0.23, V_{.,4} = 0.09, V_{.,5} = 0.38$, where $V_{.,j}$ indicates the exiting probability between the jth and $(j+1)th$ sampling occasion regardless where they entered
- the capture probabilities are $p_1 = 0.21, p_2 = 0.54, p_3 = 0.48, p_4 = 0.51, p_5 = 0.37, p_6 = 0.26$

We replicate each value of the capture probabilities depending on the number of primary occasions that we use i.e. T_k equal to $1,2,4,8$ and 16. The results are illustrated in Figures B.12-B.15. It can be observed that with more secondary sampling occasions the RMSE for all the parameters decreases, especially for the capture probabilities in Figure B.15. However, the percentage of decrease can vary based on the parameter as well as on the number of secondary sampling occasions that we are comparing.

We have used the following flat priors:

- $p_{k,t} \sim \text{Beta}(a_p = 1, b_p = 1)$
- $\omega \sim \text{Gamma}(a_\omega = 0.001, b_\omega = 0.001)$
- $(V_0, V_1, ..., V_5) \sim \text{Dirichlet}(a_0 = 1, a_1 = 1, ..., a_5 = 1)$
- $(V_{.,0}, V_{.,1}, ..., V_{.,5}) \sim \text{Dirichlet}(a_{.,0} = 1, a_{.,1} = 1, ..., a_{.,5} = 1)$

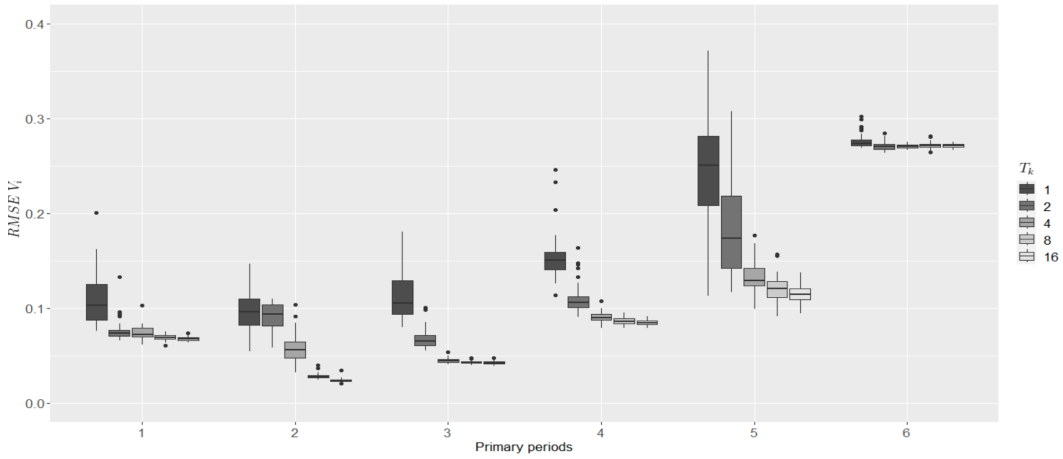

Fig. B.12 RMSE calculated based on posterior samples for the entry probabilities, V_i, $i = 1, 2, ..., K$, with K the primary periods, across different number of secondary sampling occasions, 1, 2, 4, 8, 16.

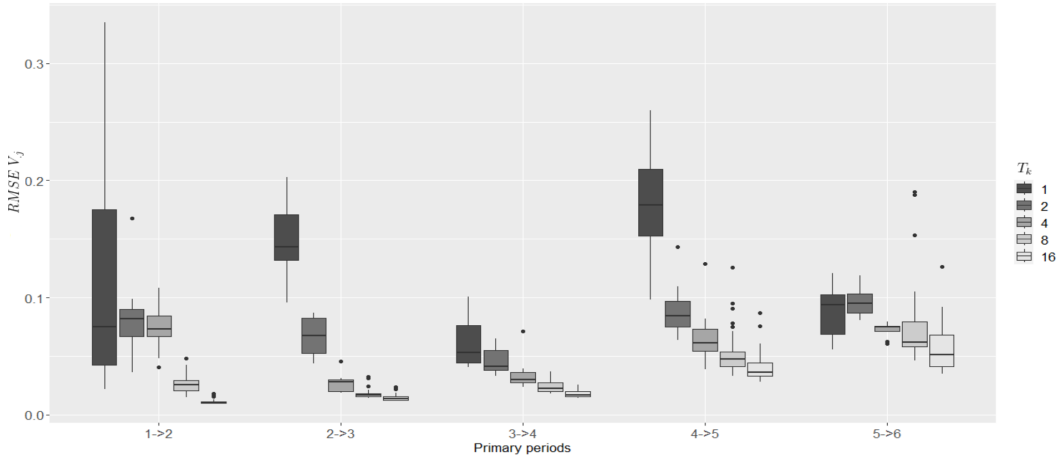

Fig. B.13 RMSE calculated based on posterior samples for the exit probabilities, regardless of the entry interval, $V_{.,j}$ $j = 1, 2, ..., K-1$ with K the primary periods, across different number of secondary sampling occasions 1, 2, 4, 8, 16.

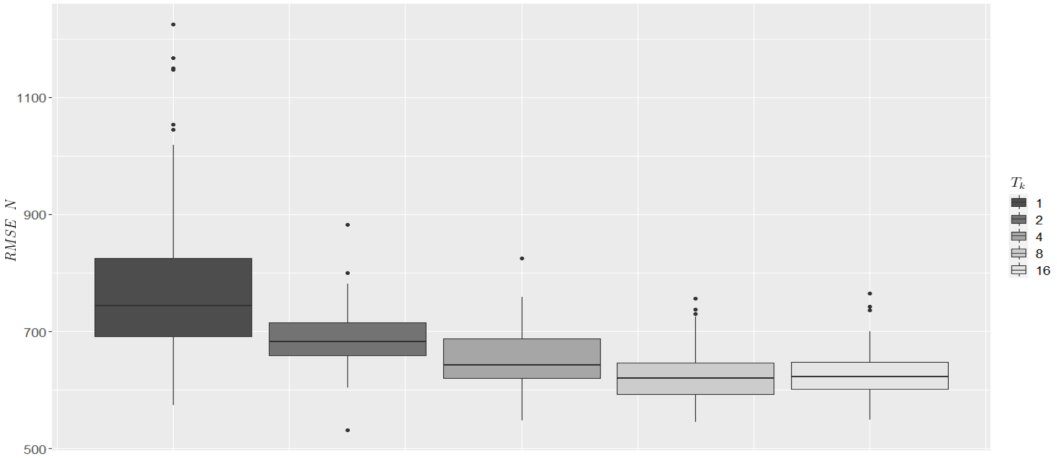

Fig. B.14 RMSE calculated based on posterior samples for the population size, N, across different number of secondary sampling occasions 1, 2, 4, 8, 16.

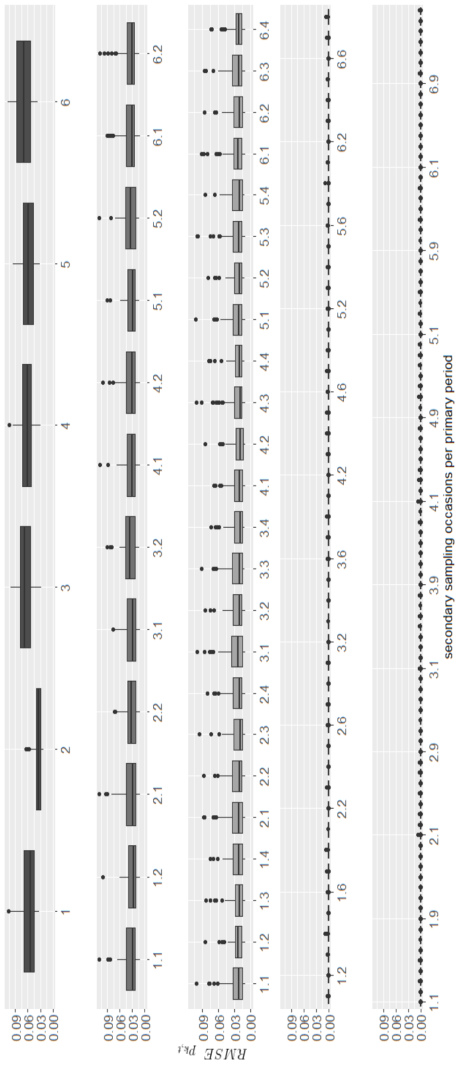

Fig. B.15 RMSE calculated based on posterior samples for the capture probabilities, $p_{k,t}$ fir different primary period $k = 1, 2, ..., K$ and $t = 1, 2, ..., T_k$ across different number of secondary sampling occasions 1, 2, 4, 8, 16.

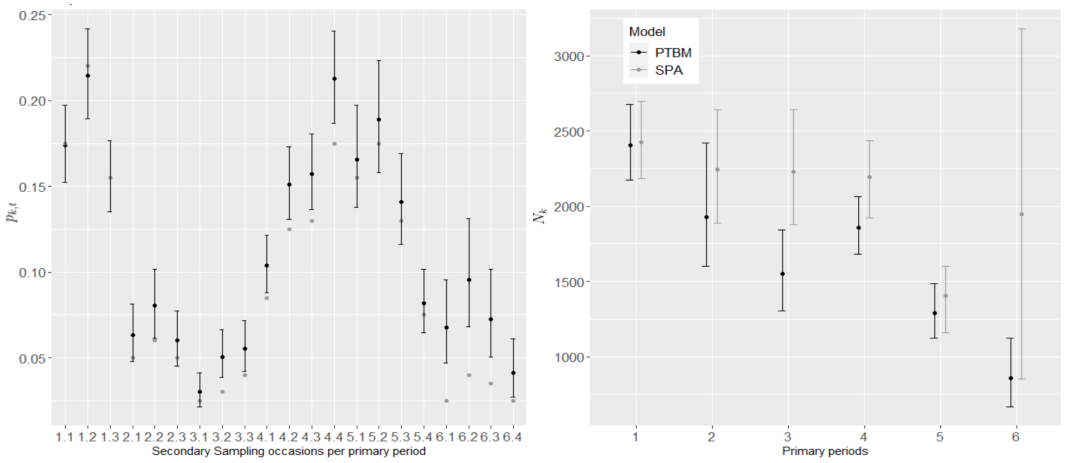

Fig. B.16 Summaries of capture probabilities $p_{k,t}$, for $t = 1,2,...,T_k$ (a), and of number of available individuals on each primary period N_k, $k = 1,2,...,K$, (b) under the models PTBM and SPA. The SPA was fitted classically, so the summaries correspond to the maximum likelihood estimate and corresponding 95% confidence interval in each case, whereas our PTBM and model are fitted in a Bayesian framework, so the summaries correspond to the posterior mean and 95% posterior credible interval.

B.6 Golden mantella

We illustrate the rest of the results in Figure B.16. Those are the capture probabilities across secondary sampling occasions per primary period as well as the available number of individuals per primary period. We were unable to reproduce the same inference using the code provided in Zhang et al. (2023). However, we can observe that our estimates are close to the ones provided in Zhang et al. (2023), but in general we have smaller uncertainty around the estimates of the parameters since we use the exact likelihood.

Next, we assess the convergence of our MCMC algorithm for the PTBM model. We considered 4 chains across 150000 iterations. The number of iterations was chosen such that the Monte Carlo error (MCE) for each parameter estimation is sufficiently small with respect to the scale of the parameter. In Table B.2 we display the Z-value of the Geweke statistic, the effective sample size (ESS), the MCE and the $\hat{R}$ statistic across the 4 chains.

Lasty, we display the trace plot for the parameter N, β and ϕ in Figures B.17, B.18, B.19, B.20, B.21, B.22, B.23, B.24, B.25, B.26, B.27, B.28.

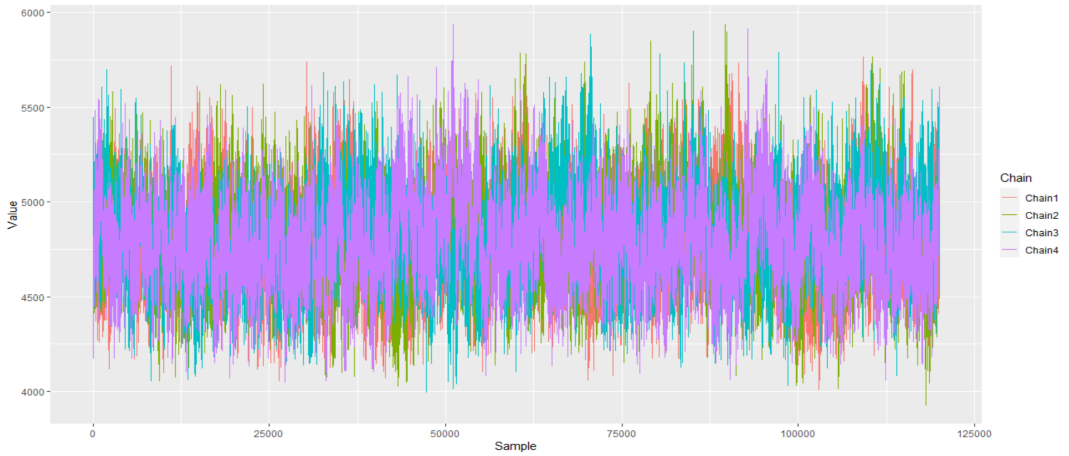

Fig. B.17 Iterations of super-population N^S across the four chains, with burn-in iterations removed.

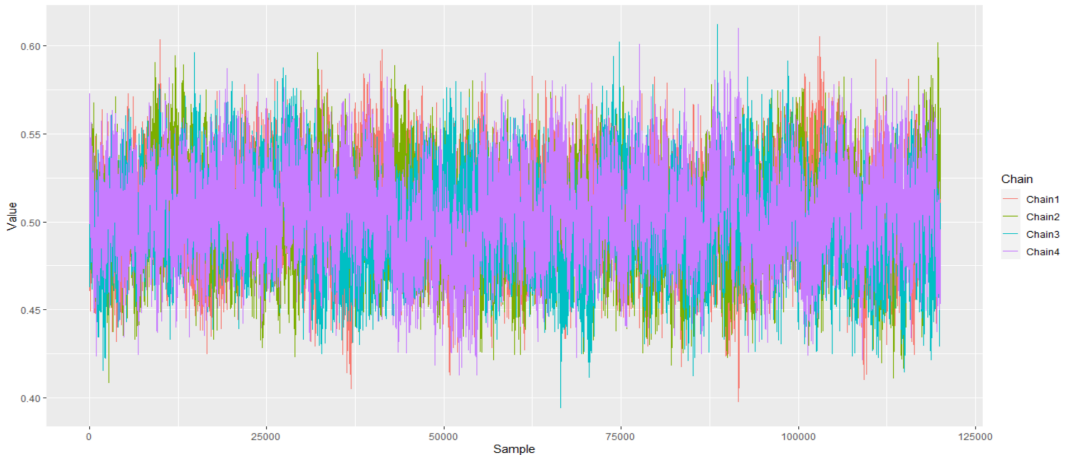

Fig. B.18 Iterations of entry probability V_1 across the four chains, with burn-in iterations removed.

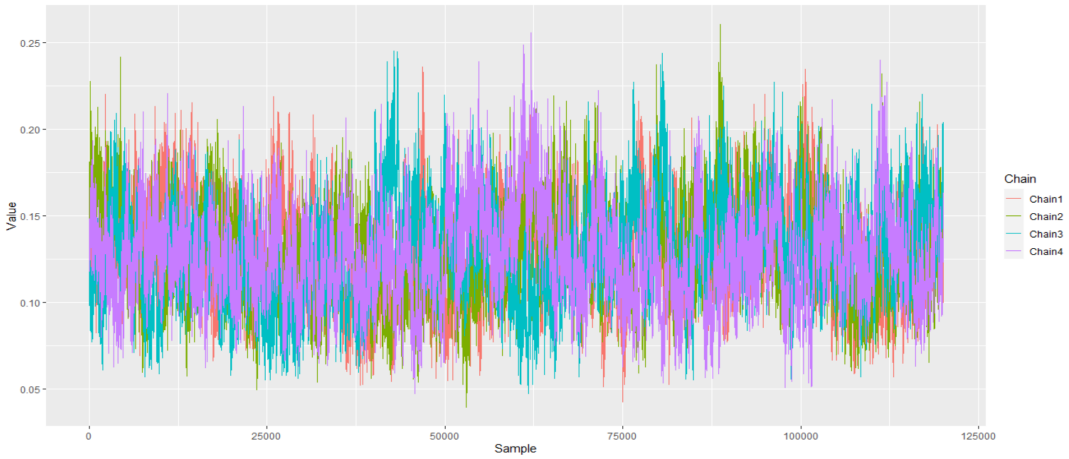

Fig. B.19 Iterations of entry probability V_2 across the four chains, with burn-in iterations removed.

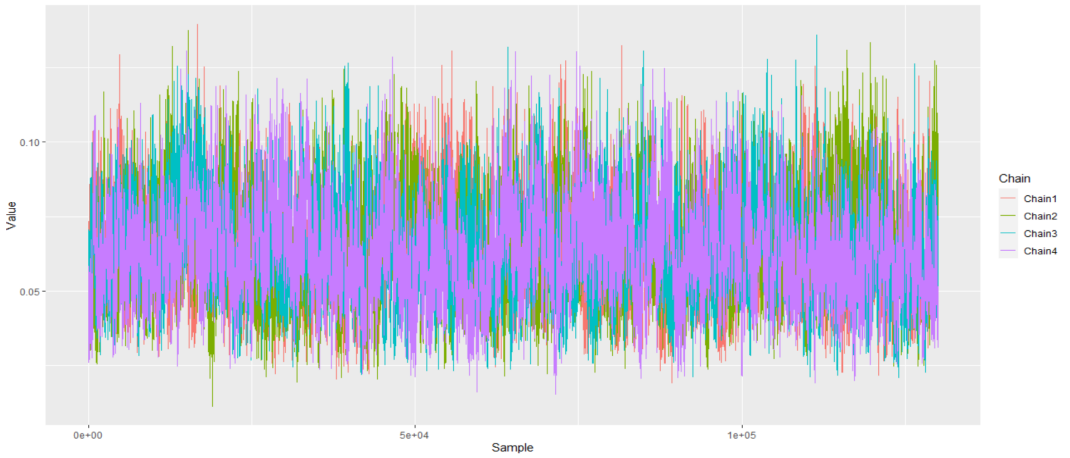

Fig. B.20 Iterations of entry probability V_3 across the four chains, with burn-in iterations removed.

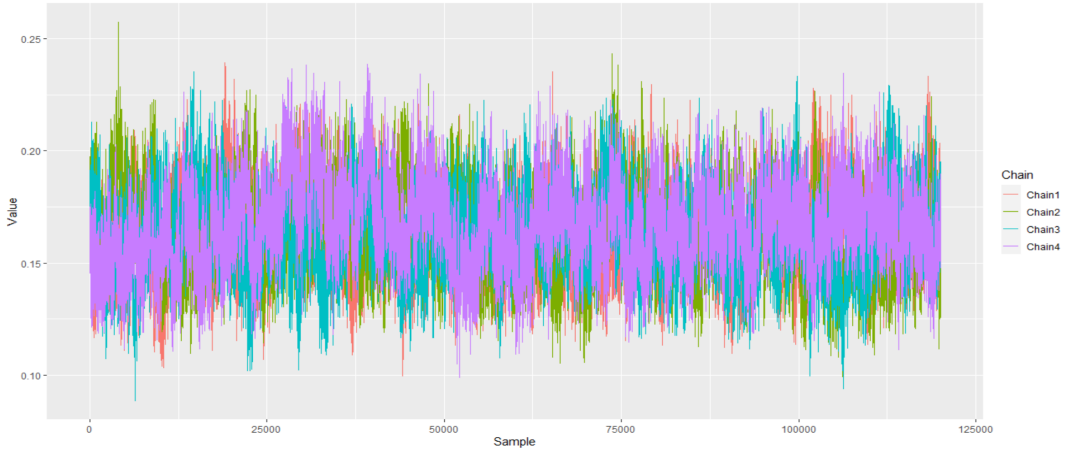

Fig. B.21 Iterations of entry probability V_4 across the four chains, with burn-in iterations removed.

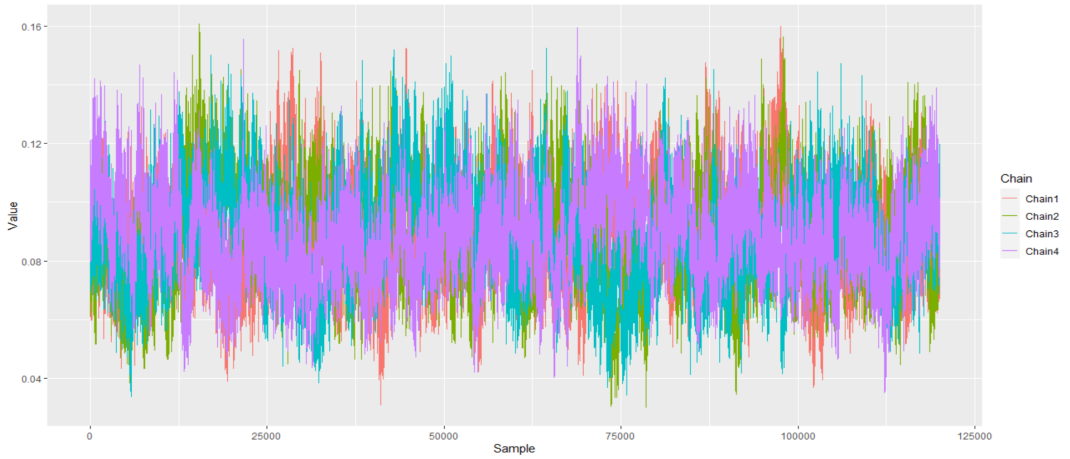

Fig. B.22 Iterations of entry probability V_5 across the four chains, with burn-in iterations removed.

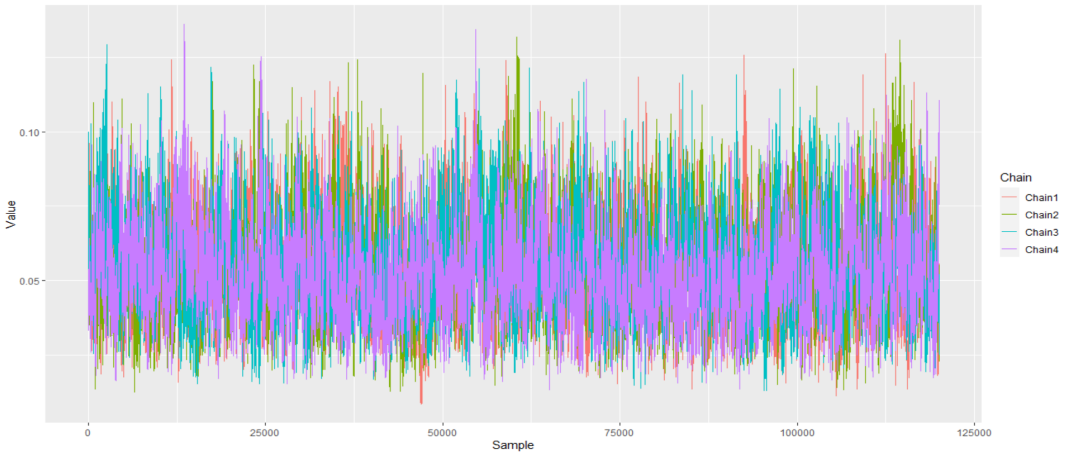

Fig. B.23 Iterations of entry probability V_6 across the four chains, with burn-in iterations removed.

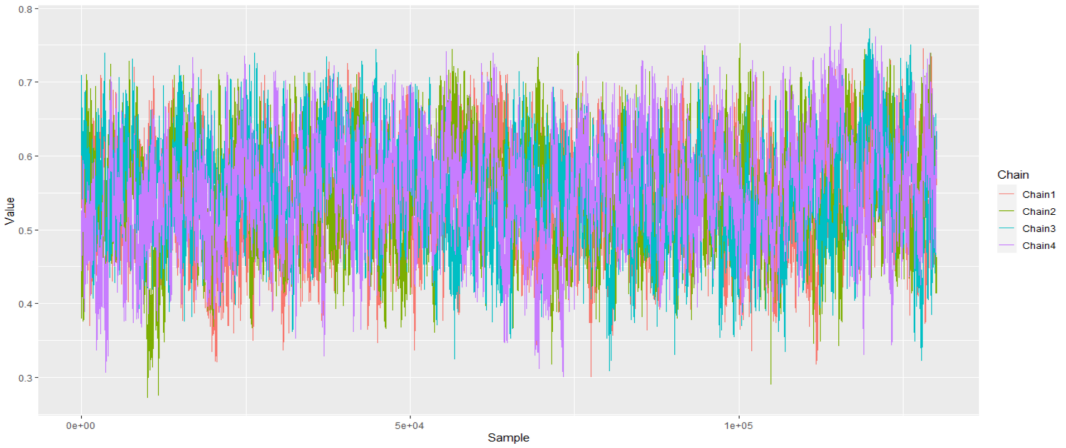

Fig. B.24 Iterations of exit probability $V_{\cdot,1}$ across the four chains, with burn-in iterations removed.

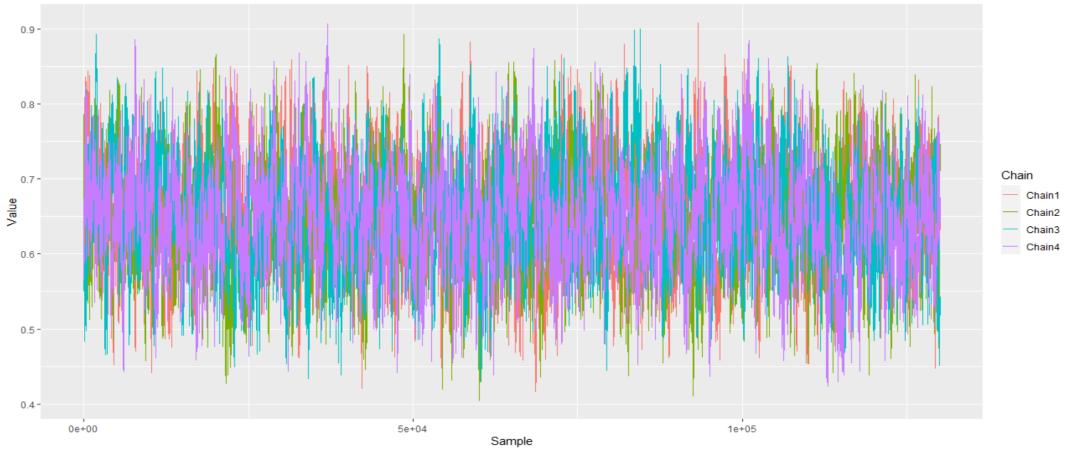

Fig. B.25 Iterations of exit probability $V_{\cdot,2}$ across the four chains, with burn-in iterations removed.

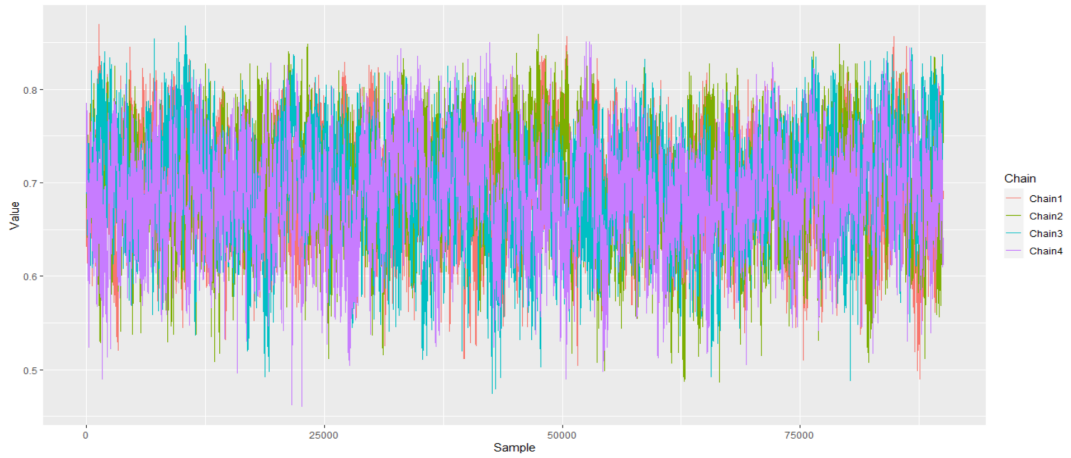

Fig. B.26 Iterations of exit probability $V_{.,3}$ across the four chains, with burn-in iterations removed.

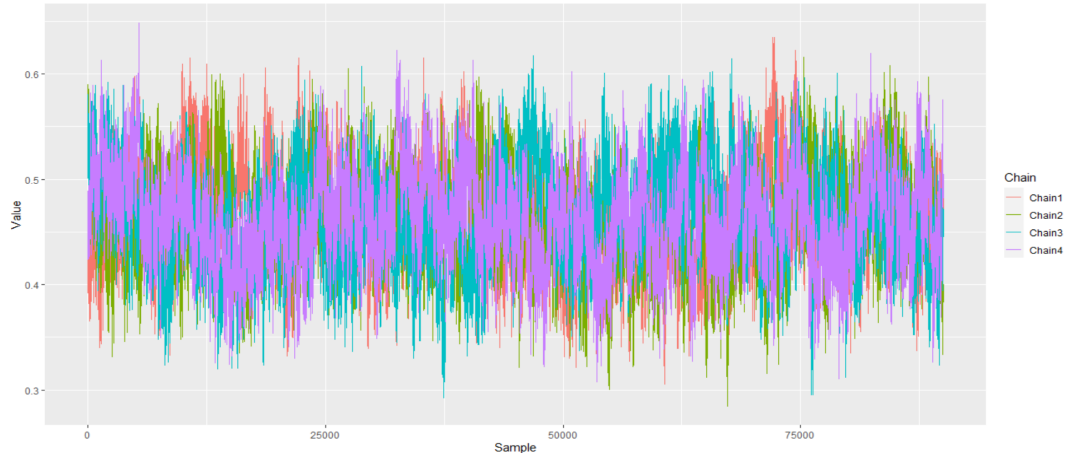

Fig. B.27 Iterations of exit probability $V_{.,4}$ across the four chains, with burn-in iterations removed.

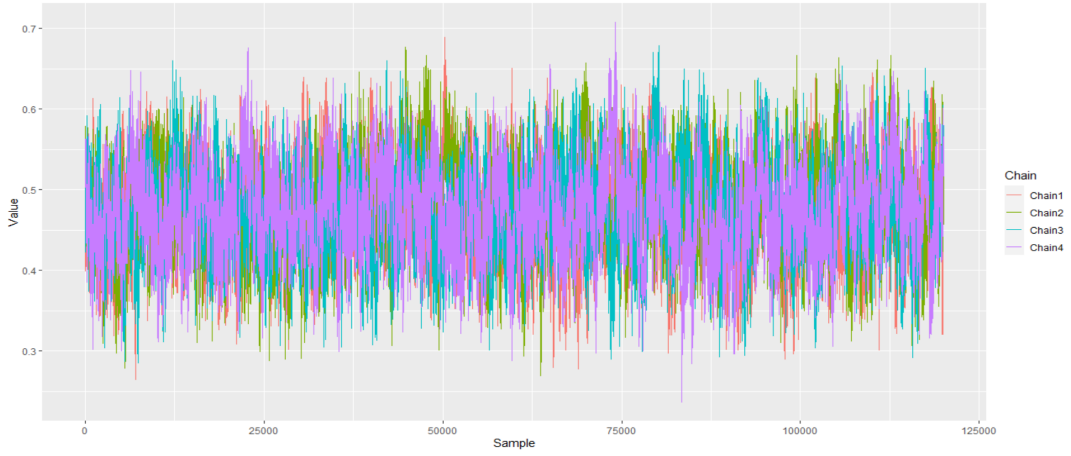

Fig. B.28 Iterations of exit probability $V_{.,5}$ across the four chains, with burn-in iterations removed.

Table B.2 The Geweke Z, ESS, MCE, and $\hat{R}$ statistics for the parameters super-population size N^S, entry/exit parameters $V_1, V_2, \cdots, V_6, V_{.,1}, V_{.,2}, \cdots, V_{.6}$ and capture probabilities $p_{1,1}, p_{1,2}, p_{1,3}, \cdots, p_{6,4}$.

Parameters	**Geweke Z**	**ESS**	**MCE**	$\hat{R}$
N^S	(-1.90,-0.58,-1.26,0.54)	(664,662,659,661)	(8.99,9.10,9.13,8.86)	1.000
$V_{.,1}$	(-0.89,-0.97,1.75,-0.93)	(453,350,372,343)	(0.0030,0.0036,0.0033,0.0037)	0.995
$V_{.,2}$	(0.37,0.39,0.27,-0.68)	(471,474,428,453)	(0.0031,0.0030,0.0032,0.0031)	1.000
$V_{.,3}$	(1.87,-0.98,0.07,-0.12)	(796,627,604,761)	(0.0017,0.0021,0.0022,0.0018)	1.018
$V_{.,4}$	(1.90,0.00,0.13,1.43)	(430,489,419,433)	(0.0021,0.0011,0.0021,0.0020)	0.994
$V_{.,5}$	(-0.65,-1.28,-1.90,-0.26)	(549,456,484,482)	(0.0022,0.0025,0.0024,0.0021)	1.011
V_1	(-0.12,-1.86,1.13,-0.29)	(480,379,396,398)	(0.0009,0.0013,0.0011,0.0011)	1.027
V_2	(1.45,1.20,-0.95,1.20)	(520,338,504,454)	(0.0007,0.0009,0.0007,0.0008)	1.006
V_3	(0.00,1.43,-1.40,-1.25)	(618,687,601,586)	(0.0008,0.0007,0.0008,0.0007)	1.007
V_4	(-1.07,0.38,-1.75,-1.28)	(677,718,575,791)	(0.0007,0.0008,0.0008,0.0007)	1.008
V_5	(-0.73,-0.08,-0.56,-1.89)	(429,436,399,410)	(0.0016,0.0016,0.0017,0.0017)	1.000
V_6	(0.27,1.28,-1.66,-1.50)	(551,439,504,407)	(0.0014,0.0016,0.0015,0.0018)	1.000
$p_{1,1}$	(0.25,-0.01,1.16,-1.54)	(4796,4095,5034,4252)	(0.0001,0.0001,0.0001,0.0001)	1.000
$p_{1,2}$	(0.20,0.09,-1.90,-1.50)	(4209,4016,4546,3977)	(0.0002,0.0002,0.0001,0.0002)	1.000
$p_{1,3}$	(0.19,0.19,-1.02,-1.31)	(5167,4177,5002,4520)	(0.0001,0.0001,0.0001,0.0001)	1.000
$p_{2,1}$	(-1.32,1.24,1.05,-1.14)	(1159,951,932,725)	(0.0002,0.0002,0.0002,0.0003)	1.003
$p_{2,2}$	(-1.13,-1.00,-1.79,-1.28)	(966,830,787,637)	(0.0003,0.0003,0.0003,0.0003)	1.004
$p_{2,3}$	(-1.51,1.21,-1.71,-1.05)	(1170,988,939,768)	(0.0002,0.0002,0.0002,0.0002)	1.003
$p_{3,1}$	(-1.19,1.93,0.06,-1.14)	(2788,2270,2944,2473)	(0.0000,0.0001,0.0000,0.0001)	1.002
$p_{3,2}$	(-1.21,1.73,0.15,-1.22)	(1693,1407,1924,1525)	(0.0001,0.0001,0.0001,0.0001)	1.004
$p_{3,3}$	(-1.19,1.83,0.23,-1.11)	(1651,1314,1813,1437)	(0.0001,0.0002,0.0001,0.0001)	1.004
$p_{4,1}$	(-1.22,1.44,-0.45,0.02)	(2056,2033,2074,1681)	(0.0001,0.0001,0.0001,0.0002)	1.001
$p_{4,2}$	(-1.38,1.50,-0.48,0.07)	(1512,1393,1425,1203)	(0.0002,0.0002,0.0002,0.0003)	1.001
$p_{4,3}$	(-1.29,1.54,-0.54,0.11)	(1415,1331,1364,1129)	(0.0002,0.0003,0.0003,0.0003)	1.001
$p_{4,4}$	(-1.39,1.55,-0.52,-0.24)	(1112,1047,1110,916)	(0.0004,0.0004,0.0004,0.0004)	1.002
$p_{5,1}$	(1.87,-1.51,-0.39,-1.75)	(999,749,911,778)	(0.0004,0.0005,0.0004,0.0005)	1.001
$p_{5,2}$	(-1.93,-1.49,-0.41,-1.74)	(927,664,835,712)	(0.0005,0.0006,0.0005,0.0006)	1.002
$p_{5,3}$	(-1.82,-1.55,-0.38,-1.62)	(1116,807,970,895)	(0.0003,0.0004,0.0004,0.0004)	1.001
$p_{5,4}$	(-1.05,-1.56,-0.23,-1.90)	(1835,1264,1510,1419)	(0.0002,0.0002,0.0002,0.0002)	1.001
$p_{6,1}$	(0.65,1.81,0.54,-1.05)	(954,774,854,863)	(0.0003,0.0004,0.0004,0.0004)	1.003
$p_{6,2}$	(0.72,1.86,0.51,-1.26)	(800,638,677,700)	(0.0005,0.0006,0.0006,0.0006)	1.003
$p_{6,3}$	(0.74,1.53,0.55,-1.18)	(963,733,809,845)	(0.0004,0.0004,0.0004,0.0004)	1.003
$p_{6,4}$	(0.74,1.85,0.60,-1.21)	(1384,1106,1106,1278)	(0.0002,0.0002,0.0002,0.0002)	1.003

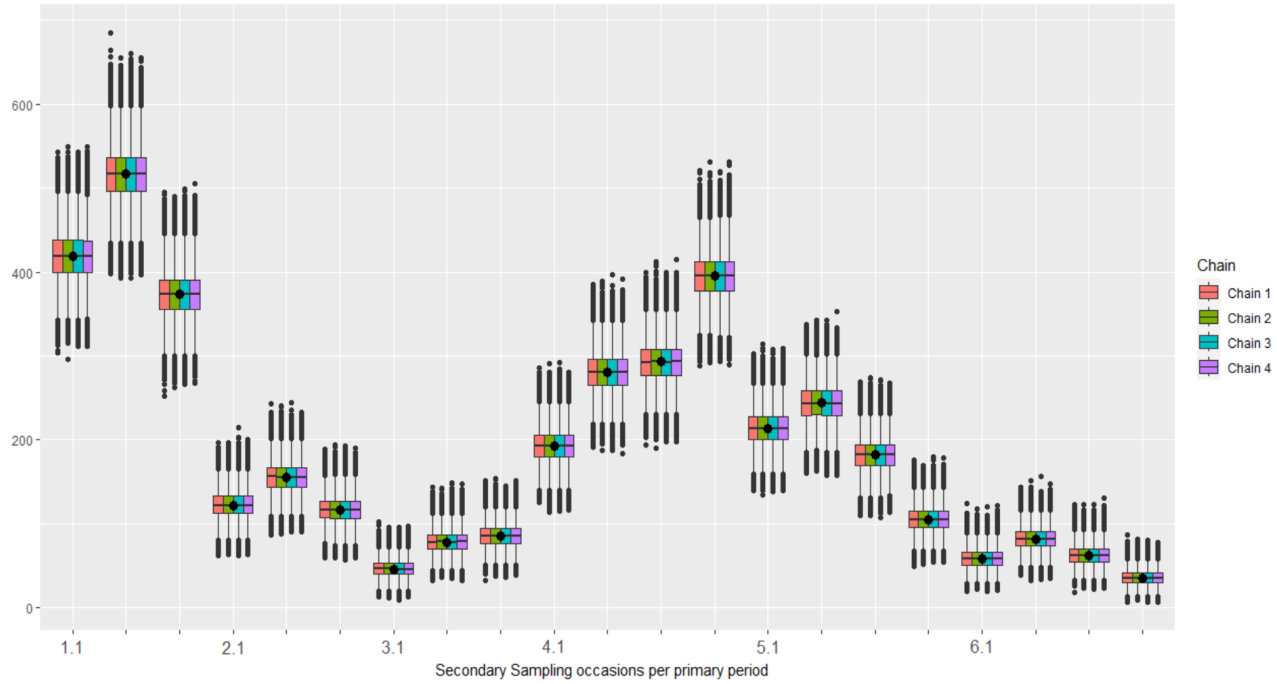

Fig. B.29 Boxplot of the posterior predicted number of individuals captured on each secondary sampling occasion across primary period for the four chains, compared to the observed number of captured individuals.

Appendix C

Approximate Bayesian Computation Sequential Monte Carlo for multi-year Batch Mark data

C.1 Intractability of the standard grid-based approach

In this thesis, we adopted the grid-based approach introduced by Diana et al. (2022) and further extended in chapter 3 for modeling the entry and exit patterns, together with the Pólya Tree (PT) prior for inferring the probabilities of the grid. This approach allows us to efficiently model BM data. This approach has been used to derive exact and efficient inference for different types of ecological data, such as CR and BM data under the assumption of permanent emigration. We tried to extend these models to consider temporary emigration; however, the Pólya Tree prior fails to achieve this.

As in Diana et al. (2022) and Chapter 3, we build a lower-triangular grid with $\frac{K(K+1)}{2}$ latent cells to describe the entry and exit pattern across periods as shown in Figure C.1. The cells represent the latent number of individuals that are alive with a particular entry and exit pattern across periods i.e. latent cells correspond to the number of individuals that entered the population, for example by becoming of breeding age by being born, between periods i and $(i+1)$ and exited, by stopping to breed or by death , between periods j and $(j+1)$. Additionally, for each primary period k we consider a $T_k \times T_k$ grid, with $\frac{T_k(T_k+1)}{2}$ latent cells, describing the entry and exit pattern of individuals that are

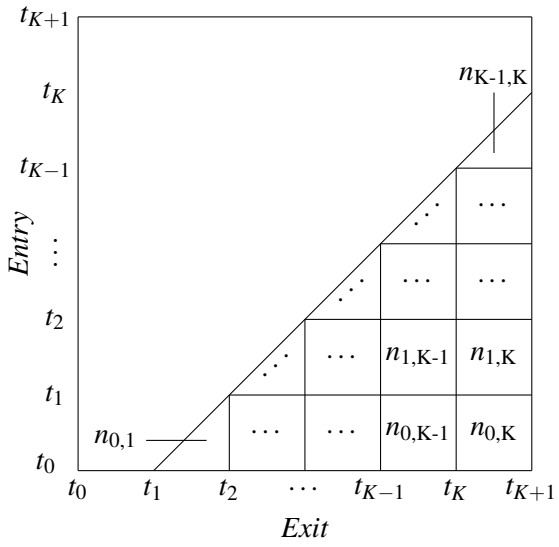

Fig. C.1 Entry and exit sample space, for K periods, taking place at times $t_1, t_2, ..., t_K$ with the convention that $t_0 = -\infty, t_{K+1} = \infty$. The latent number of individuals in cell (i, j) are the individuals with entry between the ith and $(i+1)$th period and exit between the jth and $(j+1)$th period and are denoted by $n_{i,j}$, with $i = 0, 1, ..., K-1$ and $j = i+1, ..., K$.

alive and present in period k and available for sampling, as shown in Figure C.2. The latent cells represent the latent number of individuals with a particular configuration of entry/exit within period k conditional on being alive and present that period, i.e. latent cells correspond to the number of individuals alive and present in period k, that entered between the i and $(i+1)$ sampling occasions and exited between the j and $(j+1)$ sampling occasions.

It follows that the latent information held in matrices n and $\left\{n^k\right\}_{k=1}^{K}$ is nested, in the sense that an individual in cell (i, j) in matrix n belongs to matrices n^k for $k \in \{i+1, ..., j\}$ or non. This implies that an individual in cell (i, j) in matrix n is alive to visit the study area in periods $i+1, i+2, ..., j-1, j$. Therefore, an individual can be present in any or all of periods $i+1, i+2, ..., j-1, j$, or not be present at all. This indicates that a individual can be part of many of the matrices $n^{i+1}, n^{i+2}, ..., n^{j-1}, n^j$, which denote the entry/exit configurations of the individual across sampling occasions within each corresponding period. Hence, since an individual in matrix n can correspond to many matrices $\left\{n^k\right\}_{k=1}^{K}$ we cannot analytically derive a likelihood function within a Pólya tree pior framework, since there is no possible mathematical way to keep track those nested relations.

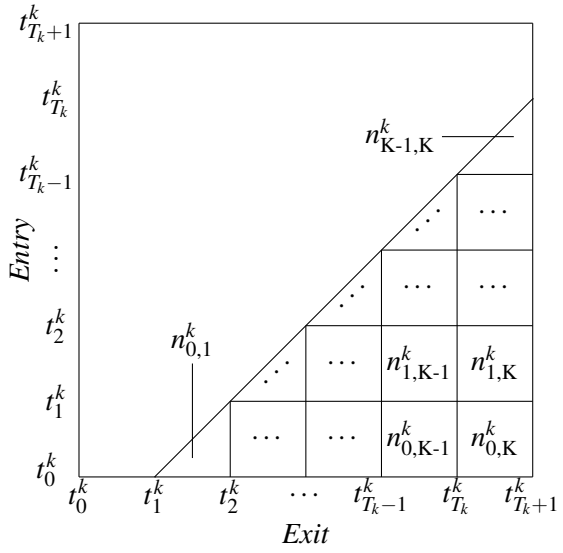

Fig. C.2 Entry and exit space in period k across T_k sampling occasions, taking place at times $t^k_1, t^k_2, ..., t^k_{T_k}$ with convention that $t^k_0 = -\infty, t^k_{K+1} = \infty$. The latent number of individuals in cell (i, j), that is with entry between the ith and $(i+1)$th sampling occasions and exit between the jth and $(j+1)$th sampling occasions in period k are denoted by $n^k_{i,j}$, with $i = 0, 1, ..., K-1$ and $j = i+1, ..., K$.