DISSERTATION

BAYESIAN METHODS FOR SPATIO-TEMPORAL ECOLOGICAL PROCESSES USING IMAGERY DATA

Submitted by
Xinyi (Lucy) Lu
Department of Statistics

In partial fulfillment of the requirements
For the Degree of Doctor of Philosophy
Colorado State University
Fort Collins, Colorado
Summer 2021

Doctoral Committee:

Advisor: Mevin Hooten

Andee Kaplan
Bailey Fosdick
David Koons

ABSTRACT

BAYESIAN METHODS FOR SPATIO-TEMPORAL ECOLOGICAL PROCESSES USING IMAGERY DATA

In this dissertation, I present novel Bayesian hierarchical models to statistically characterize spatio-temporal ecological processes. I am motivated by the volatility of Alaskan ecosystems in the face of global climate change and I demonstrate methods for emerging imagery data as survey technologies advance. For the nearshore marine ecosystem, I developed a model that combines ecological diffusion and logistic growth to quantify colonization dynamics of a population that establishes long-term equilibrium over a heterogeneous environment. I also unified modeling concepts from entity resolution and capture-recapture to identify unique individuals of the population from overlapping images and infer total abundance. For the terrestrial ecosystem, I developed a stochastic state-space model to quantify the impact of climate change on the structural transformation of land cover types. The methods presented in this dissertation provide interpretable inference and employ statistical computing strategies to achieve scalability.

ACKNOWLEDGEMENTS

I would like to thank my committee, Bailey Fosdick, David Koons, Andee Kaplan, and Ander Wilson, for devoting your precious time and effort to my professional development.

I would like to thank my cohort, especially Alex Fout, Dan Mork, Lauren Hoskovec, and Ian Taylor. I would not have made through the courses without your generous help, and it is wonderful to see all of us researching on diverse topics for the culmination of our degrees.

I would like to thank the Hooten Lab, whose current members, Ann Raiho, Clint Leach, Hanna McCaslin, Abbey Feuka, Justin Van Ee, Wilson Wright, and Michael Schwob, are both an inspiration and a good time to be around with. I have also had the fortune to cross paths with previous members Henry Scharf and Perry Williams. You have been great mentors to me and I wish you much success in your academic careers.

I would like to thank my collaborators, Jim Powell, Marti Garlick, Jamie Womble, Michael Bower, Dave Swanson, Carl Roland, and Sarah Stehn. Working with you has been a true pleasure. It is your support and guidance that has made this dissertation possible.

I would like to thank my family and friends, whose companionship, near or far, has sustained me through challenging times. You are my pillars of strength, and I am honored to share with you my joy upon arriving at this milestone.

Finally, I would like to thank my advisor, Mevin Hooten. Your enthusiasm for research and dedication to students has shaped me both professionally and personally. I am forever grateful for the opportunities you have led me to.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Recent years have seen a rapid growth in the popularity of Bayesian methods. In this dissertation, I explore statistical solutions to an array of ecological problems from population estimation to ecosystem transformation. I begin by introducing the fundamental components of a Bayesian hierarchical model, the common framework utilized throughout my dissertation. I demonstrate modeling techniques specific to the statistical challenges in Chapters 2 – 4. I also introduce the marine and terrestrial ecosystems of Alaska that motivated my dissertation. I explain briefly the focus of this dissertation on imagery data and conclude this chapter with an overview to Chapters 2 – 4.

1.1 Bayesian hierarchical modeling

The fundamental Bayesian concept is based on conditional probability often expressed as

$$\underbrace{[A|B]}_{\textit{Posterior}} = \frac{\overbrace{[B|A]}^{\textit{Likelihood}} \times \overbrace{[A]}^{\textit{Prior}}}{\underbrace{[B]}_{\textit{Marginal}}},$$

where the brackets represent probability distributions for the rest of this dissertation (Gelfand and Smith, 1990). In the context of statistical modeling, A usually represents an unknown quantity of interest, such as a parameter to be estimated, and B usually represents the observed data. The conditional relationship between the known and the unknown can be applied iteratively, and the result is the following hierarchical framework (Berliner, 1996; Wikle et al., 1998; Royle and Berliner, 1999),

$$\underbrace{[\text{Parameter}|\text{Data}]}_{\textit{Posterior}} \propto \underbrace{[\text{Data}|\text{Process}] \times [\text{Process}|\text{Parameter}]}_{\textit{Likelihood}} \times \underbrace{[\text{Parameter}]}_{\textit{Prior}}.$$

The hierarchical framework decomposes sources of variance into three stochastic models: Data, Process, and Parameter. The Data model accounts for uncertainty in data collection, the Process model accounts for uncertainty in our understanding of the true underlying process, and the Parameter model accounts for uncertainty in our prior knowledge of the parameters. The hierarchical framework provides a systematic approach to construct complex statistical models, and is of particular ecological relevance as exemplified in Chapters 2 – 4 of this dissertation.

1.1.1 Spatio-temporal stochasticity

Spatio-temporal stochasticity arises for diverse data types among various ecological processes explored in this dissertation. Separable spatio-temporal covariance functions are conveniently constructed by multiplying a spatially-explicit covariance function (e.g., Matérn covariance functions) and a temporally-explicit covariance function (e.g., autocorrelation functions). On the other hand, non-separable spatio-temporal covariance functions that satisfy admissibility are more complicated to construct (Stein, 2005). One way to specify joint spatio-temporal stochasticity makes use of partial differential equations (PDEs) (Wikle and Hooten, 2010; Hefley et al., 2017; Williams et al., 2017; Lu et al., 2020).

In Chapter 2, I developed a mechanistic process model based on the ecological diffusion equation and logistic growth. The diffusion equation characterizes spatio-temporal stochasticity of a system through the first principles of movement. The Lagrangian interpretation of the first principles states that, in a 1-D spatial domain, an individual at location s and time t can either take a step Δs to the left, to the right, or stay with probabilities $\phi_L(s,t)$, $\phi_R(s,t)$, and $\phi_N(s,t)$, respectively. The probability that the individual is at location s at time t, $p(s,t)$, conditional on the individual's location at time $t-\Delta t$, is

$$\begin{aligned}p(s,t) = \phi_N(s,t-\Delta t)p(s,t-\Delta t) + \phi_R(s-\Delta s,t-\Delta t)p(s-\Delta s,t-\Delta t) + \\ \phi_L(s+\Delta s,t-\Delta t)p(s+\Delta s,t-\Delta t).\end{aligned}$$

A Taylor series expansion of the state probabilities can be expressed as

$$\begin{aligned}
p(s, t-\Delta t) &= p - \Delta t \frac{\partial p}{\partial t} + \mathcal{O}\left(\Delta t^2\right), \\
p(s-\Delta s, t-\Delta t) &= p - \Delta t \frac{\partial p}{\partial t} - \Delta s \frac{\partial p}{\partial s} + \frac{\Delta s^2}{2} \frac{\partial^2 p}{\partial s^2} + \mathcal{O}\left(\Delta t^2\right) + \mathcal{O}\left(\Delta s^3\right), \\
p(s+\Delta s, t-\Delta t) &= p - \Delta t \frac{\partial p}{\partial t} + \Delta s \frac{\partial p}{\partial s} + \frac{\Delta s^2}{2} \frac{\partial^2 p}{\partial s^2} + \mathcal{O}\left(\Delta t^2\right) + \mathcal{O}\left(\Delta s^3\right).
\end{aligned}$$

The movement probabilities can be expanded similarly. Substituting the state and movement probabilities with their Taylor series expansions in the expression of $p(s,t)$ gives,

$$\begin{aligned}
p = p(\phi_N + \phi_R + \phi_L) &- \Delta t \frac{\partial p}{\partial t}(\phi_N + \phi_R + \phi_L) - \Delta t p \frac{\partial}{\partial t}(\phi_N + \phi_R + \phi_L) - \\
&\Delta s \frac{\partial p}{\partial s}(\phi_R - \phi_L) - \Delta s p \frac{\partial}{\partial s}(\phi_R - \phi_L) + \frac{\Delta s^2}{2} \frac{\partial^2 p}{\partial s^2}(\phi_R + \phi_L) + \Delta s^2 \frac{\partial p}{\partial s} \frac{\partial}{\partial s}(\phi_R + \phi_L) + \\
&\frac{\Delta s^2}{2} p \frac{\partial^2}{\partial s^2}(\phi_R + \phi_L) + \mathcal{O}\left(\Delta t^2\right) + \mathcal{O}\left(\Delta s^3\right).
\end{aligned}$$

The spatio-temporal indices of the state and movement probabilities are omitted for the ease of illustration. Gathering like terms and discarding high-order terms yields the Fokker-Planck equation,

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial s} \beta p + \frac{\partial^2}{\partial s^2} \delta p,$$

where $\beta = \frac{\Delta s(\phi_R - \phi_L)}{\Delta t}$ is the advection coefficient and $\delta = \frac{\Delta s^2(\phi_R + \phi_L)}{2\Delta t}$ is the diffusion coefficient. The advection component vanishes under the assumption of isotropic movement ($\phi_R = \phi_L$), and the heat equation, $\frac{\partial p}{\partial t} = \delta \frac{\partial^2}{\partial s^2} p$, results from homogeneous diffusion. However, I let the diffusion coefficients vary in space to investigate the environmental drivers to different behaviors of the species. The ecological diffusion equation on population intensity, $\lambda(s,t)$, results from multiplying the individual state probabilities, $p(s,t)$, and the total number of animals in the study domain, N, as

$$\frac{\partial}{\partial} \lambda(s,t) = \frac{\partial^2}{\partial s^2} \delta(s) \lambda(s,t),$$

where $\lambda(s,t) \equiv p(s,t)N$. When ecological diffusion is combined with a reaction component (e.g., Fisher's equation), the resulting reaction-diffusion equation characterizes the spread and growth of a colonizing population.

1.1.2 Latent processes

Ecological studies commonly involve random variables with discrete support (e.g., binary data from capture-recapture studies and count data from population surveys). The hierarchy between the Data and the Process models allows me to specify probabilistic models that conform to the data support and quantify underlying dynamics using latent processes. As such, I am able to assemble model components most suitable for their respective inferential tasks.

In Chapter 3, I unified modeling concepts from entity resolution and capture-recapture to detect duplicate individuals from overlapping aerial images and estimate population abundance. Entity resolution refers to the problem of identifying statistical units across multiple files when lacking unique identifiers. The problem proves challenging in many cases because of distortion and missingness in data. Traditional approaches express the likelihood in terms of pairwise comparison vectors. For example, suppose $\mathcal{A}$ and $\mathcal{B}$ are two files to be linked, with n_A and n_B records, respectively. Each record has H fields and is represented by a length H vector. For a record pair $\boldsymbol{x}_a^A$ from file $\mathcal{A}$ and $\boldsymbol{x}_b^B$ from file $\mathcal{B}$, $a \in \{1,\ldots,n_A\}$ and $b \in \{1,\ldots,n_B\}$, define the hth element of comparison vector, y_{ab}^h, as

$$y_{ab}^h = \begin{cases} 1, \text{ if } x_a^{A_h} = x_b^{B_h}; \\ 0, \text{ if } x_a^{A_h} \neq x_b^{B_h} \end{cases}, \; h = 1, \ldots, H.$$

The likelihood of $\boldsymbol{y}_{ab}$ can be expressed as a mixture of distributions between the set of matching pairs, $\mathcal{M}$, and the set of unmatching pairs, $\mathcal{U}$, so that

$$[\boldsymbol{y}_{ab}|w, \boldsymbol{m}, \boldsymbol{u}] = w \prod_{h=1}^{H} m_h^{y_{ab}^h}(1-m_h)^{1-y_{ab}^h} + (1-w) \prod_{h=1}^{H} u_h^{y_{ab}^h}(1-u_h)^{1-y_{ab}^h},$$

where w is the mixture probability and $\boldsymbol{m}$ and $\boldsymbol{u}$ are probability vectors characterizing the distributions of $\mathcal{M}$ and $\mathcal{U}$. Despite being an intuitive approach, comparison-based methods are criticized for unrealistic independence assumptions (between comparison vectors as well as elements within a comparison vector), arbitrary decision rules for declaring a match, and inability to produce multiple matches (i.e., when a record in A is linked to more than one record in B). An alternative to pairwise comparison, graphical record linkage methods (Steorts et al., 2015) account for distortion by comparing an observed record, $\boldsymbol{x}_a^A$, to a latent identity, λ_a^A, as

$$\left[x_a^{A_h} \middle| \lambda_a^A, y^h_{\lambda_a^A}, z^h_{\lambda_a^A}\right] = \begin{cases} \delta\left(y^h_{\lambda_a^A}\right), & \text{if } z^h_{\lambda_a^A} = 0; \\ \left[y^h_{\lambda_a^A}\right], & \text{if } z^h_{\lambda_a^A} = 1 \end{cases},$$

where $y^h_{\lambda_a^A}$ denotes the true value of field h associated with λ_a^A. The distortion indicator, $z^h_{\lambda_a^A}$, has a Bernoulli distribution and requires the observed value of field h to match its true value if $z^h_{\lambda_a^A} = 0$; otherwise, the observed value of field h is modeled as a function of its truth. The latent identity, λ_a^A, has a multinomial distribution on the set of all possible identities, $\boldsymbol{\Lambda}$. Without further information, $\boldsymbol{\Lambda}$ is commonly defined as the set of positive integers whose size equals the total number of records. The extensive use of subscripts and superscripts is not uncommon in the record linkage literature (Tancredi and Liseo, 2011; Liseo and Tancredi, 2011).

Capture-recapture studies, on the other hand, estimate population abundance by recording encounter histories of captured and marked individuals. A challenge in abundance estimation with these studies is to account for undetected individuals that belong to the population, and one statistical interpretation uses data augmentation. This interpretation augments the observed encounter history with auxiliary records of zeros and specifies a latent membership process to distinguish which records belong to undetected individuals and which belong to individuals alien to the population. The encounter outcome, $y_{i,j}$, of the ith individual on the jth visit is modeled as

$$[y_{i,j}|z_i, p] = \begin{cases} 0, & \text{if } z_i = 0; \\ \text{Bern}(p), & \text{if } z_i = 1 \end{cases}, \ i = 1, \ldots, M, \ j = 1, \ldots, J,$$

$$z_i \sim \text{Bern}(\psi),$$

where p denotes the detection probability, ψ denotes the membership probability, and z_i denotes the membership of individual i to the population. The size of the augmented data set M is defined as the upper limit of the total abundance, N, which is derived as $N = \sum_{i=1}^{M} z_i$.

Capture mechanisms vary depending on the design of the study. Spatial capture-recapture models were developed to account for temporary emigration during area searches that lack geographical closure. In a spatial capture-recapture model, the latent membership indicator is defined as $z_i = \mathbb{I}(\boldsymbol{u}_{i,j} \in \mathcal{D})$, where $\boldsymbol{u}_{i,j}$ denotes the observed location of individual i at time j and $\mathcal{D}$ denotes the spatial domain. Further, $\boldsymbol{u}_{i,j} \sim \text{N}(\boldsymbol{\mu}_i, \sigma_u^2 \boldsymbol{I})$ is modeled with a latent movement center, $\boldsymbol{\mu}_i$, and movement uncertainty, σ_u^2. Latent processes are a common strategy employed by both the graphical record linkage model and the spatial capture-recapture model. Therefore, fusing the two models by their relevant latent quantities (e.g., latent identities and their attributes including movement centers and membership indicators; Chapter 3) is natural under a hierarchical framework.

1.1.3 Data augmentation

Bayesian methods can be computationally challenging to implement due to the iterative sampling of MCMC algorithms. Two computing strategies to reduce the run time of an MCMC algorithm without utilizing extra resources are: 1. decreasing the computation complexity per iteration; 2. increasing the overall sampling efficiency. The first strategy is commonly performed by replacing expensive operations with cheap approximations (e.g., homogenization; Chapter 2) and reducing the model dimension by exploiting its dependence structure (e.g., graphical record linkage; Chapter 3). The second strategy is commonly performed by sampling from high density regions of the target distribution and proposing samples with low autocorrelation. Data augmentation is aligned with the second computing strategy and often improves sampling efficiency by inducing conjugacy.

One well-known data augmentation scheme was proposed by Albert and Chib (1993) for probit regression. Albert and Chib (1993) modeled binary outcomes, y_i, using auxiliary Gaussian random variables, z_i, characterized by a set of linear predictors, so that

$$\begin{aligned} y_i &= \begin{cases} 1, \text{ if } z_i \geq 0; \\ 0, \text{ if } z_i < 0 \end{cases}, \\ z_i &= \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i, \\ \epsilon_i &\sim \mathrm{N}(0,1),\ i = 1, \ldots, n. \end{aligned}$$

The marginal likelihood of y_i is equivalent to the cumulative density function (CDF) of a standard normal random variable evaluated at $\boldsymbol{x}_i'\boldsymbol{\beta}$. Moreover, the posterior distribution of $\boldsymbol{\beta}$ conditional on z_i, for $i = 1, \ldots, n$, can be analytically determined under a normal prior, thereby facilitating a fully Gibbs algorithm. Augmentation schemes similar to Albert and Chib (1993) have been proposed for logistic regression by specifying that $\epsilon_i \sim \text{Logistic}(1)$ follows a standard logistic distribution. However, these schemes do not directly lead to conjugacy in $\boldsymbol{\beta}$, and approximations to the logistic distribution by a mixture of normal distributions increase the complexity of the algorithm. In comparison, Polson et al. (2013) proposed an augmentation strategy using Pólya-Gamma (PG) random variables that provides direct conjugacy (i.e., using a single layer of auxiliary variables) to linear estimators in logistic regression. Polson et al. (2013) demonstrated that the PG data augmentation strategy is superior to other existing data augmentation schemes and Metropolis-Hastings algorithms in a variety of model settings. I discuss the strategy in detail in Chapter 4.

I explored the utility of PG data augmentation in modeling state (transition) probabilities of multinomial random variables (Chapter 4). Multinomial logistic regression characterizes a K-dimensional categorical variable using a set of $K - 1$ regression functions because all state probabilities must sum to one. Multinomial logistic regression may be represented by a bijective map on $(\eta_1, \ldots, \eta_K)'$ from $\mathbb{R}^K$ to $[0,1]^K$, as

$$\text{softmax}(k, \eta_1, \ldots, \eta_K) = \frac{\exp(\eta_k)}{\sum_{j=1}^{K} \exp(\eta_j)},$$

Denoting K as the reference category, the state probabilities implied by the softmax function are

$$[y_i = k|\boldsymbol{\eta}] = \frac{\exp\left(\eta'_k\right)}{1 + \sum_{j=1}^{K-1} \exp\left(\eta'_j\right)},$$
$$[y_i = K|\boldsymbol{\eta}] = \frac{1}{1 + \sum_{j=1}^{K-1} \exp\left(\eta'_j\right)},$$

for $i = 1, \ldots, n$. The set of identifiable parameters is comprised of $\eta'_k \equiv \eta_k - \eta_K$, for $k = 1, \ldots, K-1$. Holmes and Held (2006) proposed a Bayesian implementation of multinomial logistic regression by recognizing that the conditional likelihood of η_k given $\eta_1, \ldots, \eta_{k-1}, \eta_{k+1}, \ldots, \eta_{K-1}$ has the form of a logistic regression as

$$[\eta_k, \boldsymbol{y}|\boldsymbol{\eta}_{-k}] \propto \prod_{i=1}^{n} \left(\frac{\exp(\eta_k - C_k)}{1 + \exp(\eta_k - C_k)} \right)^{\mathbb{I}(y_i = k)} \left(1 - \frac{\exp(\eta_k - C_k)}{1 + \exp(\eta_k - C_k)} \right)^{\mathbb{I}(y_i \neq k)},$$

where $C_k = \log \sum_{j \neq k, j < K} \exp(\eta_j)$. Consequently, data augmentation schemes to facilitate Gibbs sampling in logistic regression may be employed iteratively over η_k, $k = 1, \ldots, K-1$. The multinomial likelihood may be alternatively represented as a product of $K-1$ conditional binomial likelihoods. I introduce this representation and the corresponding sampling algorithm in Chapter 4.

1.2 Alaskan ecosystems and imagery data

From temperate rainforests in the southeast to the Arctic tundra on the North Slope, Alaska is home to a wealth of diverse and pristine ecosystems. However, recent years have witnessed a growing volatility in ecosystems across Alaska due to changing climates. In this dissertation, I investigated the marine and the terrestrial ecosystems of Alaska by developing statistical models to quantify population dynamics of sea otter colonization in Glacier Bay, and to characterize the transition dynamics of land cover types in boreal Alaska.

Sea otters are native to the coasts of the northern and eastern North Pacific Ocean. They are often considered a keystone species because of their role in maintaining kelp ecosystems by limiting sea urchin populations. In the 18th and 19th centuries, sea otters were hunted extensively during the multi-national commercial fur trade (Kenyon, 1969; Bodkin, 2015). In the years that followed, conservation legislation and translocation efforts enabled the return of sea otters to their former distribution (Kenyon, 1969; Bodkin, 2015). In southeastern Alaska, the tidewater glaciers in Glacier Bay have been retreating since 1750 at a speed and scale unparalleled in modern times (Lawrence, 1958; Chapin et al., 1994). The retreating glaciers resulted in a diverse and abundant prey source of marine invertebrates, and sea otters quickly thrived in the environment as an apex predator since they were first documented at the mouth of Glacier Bay in 1988 (Weitzman, 2013). The study of sea otter colonization in Glacier Bay provides a rare opportunity to investigate the recovery of a keystone species under a changing climate, and to reassess the focus of management and conservation decisions (Williams et al., 2019).

Temperature in Alaska has increased twice as much as the rest of the continental U.S. in recent decades (Stewart et al., 2013). Deglaciation is one manifestation of a warming climate; meanwhile, another prominent manifestation is the growing presence of trees and shrubs in boreal Alaska, known colloquially as “Arctic greening" or “shrubification" (Tape et al., 2006). The expansion of woody landscapes impacts local ecosystems by favoring browsing species (Tape et al., 2010), reducing erosion, and increasing fire frequency (Frost and Epstein, 2014). Woody encroachment also impacts the climate globally by changing snow deposition (Sturm et al., 2001) and decreasing albedo (Chapin III et al., 2005), which consequently amplify warming. The combination of these effects influences biological and physical processes of the ecosystems at multiple scales (Hinzman et al., 2013; Pastick et al., 2019). The study of land cover response to climate changes over a heterogeneous spatial domain provides insight into the trajectories of future landscape transformations under various climate scenarios.

Ecological studies are presented with increasingly complex data types such as sound, images, and videos as technological advancements made automated data collection possible. In this dissertation, I focus on the analysis of aerial imagery data. Aerial surveys are commonly used to gather information about species abundance (Caughley, 1974; Ver Hoef, 2014). Imagery surveys reduce the risks and costs associated with manual operation compared to traditional observer-based surveys (Buckland et al., 2012). Imagery data provide high-resolution information that can be independently verified (Svenningsen et al., 2015). Being able to process these data efficiently while preserving as much information as possible will be the goal of future data analysis. I developed statistical methods for lab-processed imagery data to meet the specific challenges of this data type, such as accounting for uncertainty due to the distortion of image footprints and spatial correlation among pixelated data. I also developed a framework to integrate contemporary imagery data with historic observer-based data that were collected at a different spatial scale.

1.3 Overview

In Chapter 2, I developed a model that combines an ecological diffusion equation and logistic growth to characterize colonization processes of a population that establishes long-term equilibrium over a heterogeneous environment. Partial differential equations (PDEs) are a useful tool for modeling spatio-temporal dynamics of ecological processes. I developed a homogenization strategy to statistically upscale the PDEs for faster computation and adopted a hierarchical framework to accommodate multiple data sources (observational surveys and aerial imagery surveys) collected at different spatial scales. As a case study, I demonstrated that my model improves spatio-temporal abundance forecasts of sea otters in Glacier Bay, Alaska. Further, I predicted spatially-varying local equilibrium abundances as a result of environmentally-driven diffusion and density-regulated growth. Inference on the overall carrying capacity was enabled by integrating local equilibrium abundances over the study area.

In Chapter 3, I developed a method that leverages aerial imagery data for population modeling through entity resolution. Resolving duplicate individuals in overlapping images that are distorted requires realigning observed point patterns optimally; however, popular machine learning algorithms for image stitching do not often account for alignment uncertainty. Moreover, duplicated individuals can provide insight about detection probability when overlaps are viewed as replicate surveys. My model resolves individual identities by linking observed locations to latent activity centers and estimates total population as informed by the linkage structure. I developed a hierarchical framework to achieve entity resolution and abundance estimation cohesively, thereby avoiding single-direction error propagation that is common in two-stage models. I illustrated my method through simulation and a case study using aerial images of sea otters in Glacier Bay, Alaska.

In Chapter 4, I developed a dynamic statistical model to quantify the impact of climate change on the structural transformation of terrestrial ecosystems in Alaska using remotely sensed imagery. My model accommodates changes in temperature and precipitation to infer and predict rates of land cover transitions while accounting for spatio-temporal heterogeneity. Transition types are highly correlated at both plot and subplot levels in the study system, therefore I characterized multi-scale spatial correlation using Gaussian processes. Because imagery pairs were collected at irregular time intervals, I also modeled dynamic state probabilities that evolve annually using a hierarchical framework. I developed a Pólya-Gamma representation of the model to improve computation. My model facilitates inference on the response of ecosystem state probabilities to shifts in climate and can be used to predict future land cover transitions under various climate scenarios.

Chapter 2

Nonlinear Reaction-Diffusion Process Models for Population Dynamics

2.1 Introduction

The dynamics of ecological systems are complicated because the interaction between organisms and their host environment may vary in space and time. Traditionally, generalized linear mixed models (GLMM) have been favored to describe spatio-temporal ecological processes (e.g., Banerjee et al., 2014). While such models enjoy a relatively large degree of flexibility, they lack an explicit mechanistic interpretation of the underlying process. Moreover, conventional GLMMs are often incapable of capturing the joint spatio-temporal dependence that is characteristic of ecological processes (Wikle and Hooten, 2010). On the other hand, mechanistic statistical models are increasingly popular because they allow us to formally incorporate our knowledge of the system we seek to understand in latent processes (Hilborn and Mangel, 1997). In particular, partial differential equations (PDEs) have been commonly used to represent ecological processes due to their connections to physical laws (Wikle, 2003; Cressie and Wikle, 2011). By embedding the PDEs in a hierarchical framework, we can appropriately account for uncertainty in the data, our prior understanding of the process, and parameters that influence the process (Berliner, 1996; Hobbs and Hooten, 2015).

In what follows, we present a hierarchical reaction-diffusion model that was motivated by the case study of sea otter colonization in Glacier Bay, Alaska. Across their North Pacific range, sea otter populations have undergone significant fluctuations over the past two centuries. After being hunted to near extirpation during the maritime fur trade, 13 remnant colonies remained, and sea otter populations have subsequently recovered in many areas due to a combination of conservation efforts and environmental changes (Larson et al., 2014). In 1988, sea otters were first documented at the mouth of Glacier Bay, and have expanded throughout much of the bay. The study of sea otter colonization in Glacier Bay provides important insight into the ability of a species to recover from near extirpation, as well as the impact of recent deglaciation and a changing climate on their recovery (Williams et al., 2019).

Reaction-diffusion models have long been used to describe the colonization or invasion of a species (e.g., Holmes et al., 1994). Past studies of reaction-diffusion models in a statistical framework have focused on development of a spatially dynamic diffusion component, while relying on a relatively simple reaction term (Wikle, 2003; Hooten and Wikle, 2008; Zheng and Aukema, 2010; Williams et al., 2017). A commonly used model for reaction, the Malthusian growth model, assumes that the per capita growth rate remains the same regardless of population size (Turchin, 2003). This may be reasonable at initial stages of a colonization, but as the species expands into the environment, the population can become resource-limited, resulting in a decline in growth rate. Recently collected data indicate a slowing of sea otter expansion in Glacier Bay; thus, we adopted a reaction model based on logistic growth that assumes the per capita growth rate declines as the population size approaches a maximum (i.e., carrying capacity) regulated by the amount of available resources. Logistic growth is more realistic in characterizing population growth during colonization, and allows us to gain insight about the system at its equilibrium. For example, we may learn about the spatially-varying equilibrium abundances over a heterogeneous environment, as well as the overall carrying capacity of the environment with associated uncertainty.

Fitting statistical reaction-diffusion models can be computationally challenging when the scale of the process is fine in space and/or time. The approach we present induces computational economy via the mathematical technique of homogenization. Using the “method of multiple scales" (Holmes, 2013; Garlick et al., 2011), our implementation relies on a solution to the PDE at a larger spatial scale, while maintaining the inference on parameters at the original small scale (Hooten et al., 2013).

Furthermore, our method is useful for reconciling multiple data sources collected at different spatial scales with varying degrees of accuracy. In our application, inconsistency in spatial scales comes partly from improvements in aerial survey technology over time. Our modeling framework is compatible with both the past and the more recent survey methodology for monitoring sea otters in Glacier Bay, and is therefore useful for inference and forecasting based on ongoing data collection efforts.

The rest of the paper proceeds as follows. In Section 2.2.1, we develop a hierarchical model and demonstrate that a homogenized reaction-diffusion process reduces computational complexity. In Section 2.3, we illustrate the model through simulation and the sea otter case study, thereby showing that the logistic reaction component improves parameter inference and population forecast compared to the Malthusian reaction component. Finally, in Section 2.4, we conclude the paper with a discussion of possible extensions and broader applications of our model.

2.2 Methods

2.2.1 Hierarchical Model

Data Model

The goal of our model is to infer sea otter abundance in continuous space and time within our study area given observed data on relative abundances at a subset of locations and time points, and true abundances observed at a subsequent subset of locations and time points. We let $y_{i,t}$ denote the observed relative abundance of sea otters at a site $\mathcal{S}_i$ in year t. Following the N-mixture framework (Royle, 2004), we modeled the relative abundance using a binomial distribution conditioned on (latent) site-specific true abundance, $N_{i,t}$, and detection probability, p_t, as follows,

$$y_{i,t} \sim \text{Binom}\left(N_{i,t},\ p_t\right). \tag{2.1}$$

Process Model

We modeled the latent, true abundance, $N_{i,t}$, using a negative binomial distribution conditioned on a dynamically evolving mean (population intensity), $\lambda_{i,t}$, and dispersion parameter τ, thereby providing the process model with more flexibility than other commonly used count models, such as the Poisson model (Ver Hoef and Boveng, 2007). Population intensity $\lambda(\boldsymbol{s}, t)$ with $\boldsymbol{s} \equiv (s_1, s_2)'$ is modeled in continuous space and time, where integration over a site $\mathcal{S}_i$ results in the mean abundance $\lambda_{i,t}$. Assuming conditional independence of latent, true abundances given population intensities, we have

$$N_{i,t} \sim \text{NB}\left(\lambda_{i,t},\ \tau\right), \tag{2.2}$$

$$\lambda_{i,t} = \int_{\mathcal{S}_i} \lambda(\boldsymbol{s}, t) d\boldsymbol{s}. \tag{2.3}$$

We then modeled the spatio-temporal dynamics of population intensities with the following reaction-diffusion equation,

$$\frac{\partial}{\partial t}\lambda(\boldsymbol{s}, t) = \underbrace{\left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}\right)\delta(\boldsymbol{s})\lambda(\boldsymbol{s}, t)}_{\textbf{(i)}} + \underbrace{\gamma\lambda(\boldsymbol{s}, t)\left(1 - \frac{\lambda(\boldsymbol{s}, t)}{K}\right)}_{\textbf{(ii)}}. \tag{2.4}$$

The diffusion component in **(i)** of (2.4) is known as a Fokker-Planck equation (Risken, 1989), and can be derived from individual movement processes following the convention of Turchin (1998). The diffusion coefficients, $\delta(\boldsymbol{s})$, also known as motility coefficients, are inversely related to residence time (Turchin, 1998; Hooten et al., 2013). With $\delta(\boldsymbol{s})$ inside the second derivative, the Fokker-Planck equation allows population intensity to vary sharply between neighboring locations at the transition of habitat types (Garlick et al., 2011; Hooten et al., 2013; Hefley et al., 2017), which is useful for capturing the variability in sea otter intensity due to their resting and foraging behaviors in different environments. We modeled heterogeneity in diffusion coefficients as a log-linear function of a set of environmental covariates such that $\log(\delta(\boldsymbol{s})) = \boldsymbol{x}(\boldsymbol{s})'\boldsymbol{\beta}$, where $\boldsymbol{x}(\boldsymbol{s})$

is a vector of ocean depth (indicator of < 40 m), distance to shore, slope of the ocean floor, and shoreline complexity (Williams et al., 2017).

The reaction component in **(ii)** of (2.4) is modeled after logistic growth, where γ is the parameter for intrinsic growth rate and K is the local density-dependent parameter regulating growth. We let the parameters γ and K be constant in space and time. Although it is possible to model them as variable in space, heterogeneity in both the diffusion and the reaction components may be unidentifiable. In what follows, we illustrate that although a single parameter K is used to regulate growth, the resulting equilibrium abundances in our study system are spatially heterogeneous due to the changing balance between ecological diffusion and density dependence.

We used a scaled Gaussian kernel for the initial conditions of (2.4),

$$\lambda(\boldsymbol{s}, t_0) = \frac{\theta \exp\left(\frac{-|\boldsymbol{s}-\boldsymbol{s}_d|^2}{\kappa^2}\right)}{\int_{\mathcal{S}} \exp\left(\frac{-|\boldsymbol{s}-\boldsymbol{s}_d|^2}{\kappa^2}\right) d\boldsymbol{s}}, \tag{2.5}$$

where θ controls the magnitude of initial population intensity, and κ controls the initial population range. The location $\boldsymbol{s}_d = (s_{1d}, s_{2d})'$ is an epicenter fixed to be near the mouth of Glacier Bay, where sea otters were observed before the colonization initiated. The starting time t_0 was chosen to be year 1993 when the earliest data were collected. Following the example by Williams et al. (2017), we used a no-flux spatial boundary condition (Cantrell and Cosner, 2004) at locations adjacent to land, so that diffusive movement onto land will be reflected back to water at such boundaries. Rigorous survey data outside the study area that could be useful to aid in the estimation of flux at the mouth of Glacier Bay are not available. Thus, we assumed a boundary condition at the mouth of Glacier Bay that allows for emigration of sea otters only because they are protected inside the National Park, but not outside.

Parameter Model

To complete the model hierarchy, we specified prior distributions for the data and process model parameters. We used a uniform prior from 0 to 0.5 for intrinsic growth rate, γ, because we had sufficient evidence that the population was expanding during the study period, and the study by Estes (1990) estimated the maximum reproductive rate of sea otters in south-east Alaska to range from 0.196 to 0.237. We used a beta prior centered at 0.75 for the detection probabilities, p_t, as informed by previous studies (Williams et al., 2017). The rest of the parameters were given vague priors. A full description of prior specifications can be found in Section A.1.

The joint posterior distribution associated with our model is

$$\begin{aligned}[\boldsymbol{N}_u, \boldsymbol{p}, \tau, \boldsymbol{\beta}, \theta, \kappa, \gamma, K | \boldsymbol{Y}, \boldsymbol{N}_o] \propto & \left\{ \prod_{i=1}^{n_t} \prod_{t=1}^{T} [y_{i,t} | N_{i,t}, p_t] \right\} \\ & \times \left\{ \prod_{j=1}^{n_t - n_{o,t}} \prod_{t=1}^{T} [N_{j,t} | \boldsymbol{\beta}, \theta, \kappa, \gamma, K, \tau] \right\} \\ & \times [\boldsymbol{p}][\tau][\boldsymbol{\beta}][\theta][\kappa][\gamma][K][\tau], \end{aligned} \tag{2.6}$$

where $\boldsymbol{N}_o$ denotes the vector of observed true abundances, and $\boldsymbol{N}_u$ denotes the vector of unobserved true abundances that are modeled as latent variables. We let n_t represent the total number of sites where relative abundance was observed in year t, and $n_{o,t}$ represent the number of sites where true abundance was observed in year t.

2.2.2 Homogenization

When the spatial domain is large and the spatial resolution is fine, solving (2.4) repeatedly can be computationally demanding. The concept of homogenization is to rewrite (2.4) in terms of both large and small spatial scales, so that, under certain approximation conditions, we can solve the PDE numerically at the large scale and recover the small scale solutions through a downscaling transformation.

Suppose the diffusion coefficient, δ, depends on two spatial scales, varying quickly on a small spatial scale, and much more slowly on a large spatial scale. We let $\boldsymbol{s}$ denote the fine grain spatial variable in two-dimensions, and introduce the coarse grain spatial variable $\boldsymbol{\omega} \equiv (\omega_1, \omega_2)'$. Suppose $\boldsymbol{\omega} = \boldsymbol{s}\epsilon$, where $0 < \epsilon \ll 1$ is the ratio between the two scales, such that changes on the order of $\mathcal{O}(\epsilon)$ in $\boldsymbol{\omega}$ become changes on the order of $\mathcal{O}(1)$ in $\boldsymbol{s}$ (Powell and Zimmermann, 2004). Although there have not been individual-level movement studies of sea otters in Glacier Bay, in our application, we assumed sea otters exhibit relatively high site-fidelity and daily movements on the scale of hundreds of meters based on studies from other areas (e.g., Jameson, 1989), whereas the available environmental covariates in Glacier Bay vary on the scale of kilometers, which indicates that $\epsilon \approx 1/10$. In addition, we let t denote the temporal variable associated with $\boldsymbol{\omega}$. We consider (2.4), which we outline below as a reminder,

$$\frac{\partial \lambda}{\partial t} = \left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}\right) \delta \lambda + \gamma \lambda \left(1 - \frac{\lambda}{K}\right).$$

By transforming derivatives on the spatial variables in each dimension, $\frac{\partial^2}{\partial s_i^2} \to \frac{1}{\epsilon^2}\frac{\partial^2}{\partial s_i^2} + \frac{2}{\epsilon}\frac{\partial^2}{\partial s_i \partial \omega_i} + \frac{\partial^2}{\partial \omega_i^2}$, $i = 1, 2$, and writing λ as a power series in ϵ, $\lambda = \lambda_0 + \epsilon\lambda_1 + \epsilon^2\lambda_2 + \cdots$, we obtain the following PDE,

$$\epsilon^2 \frac{\partial}{\partial t}\left(\lambda_0 + \epsilon\lambda_1 + \epsilon^2\lambda_2 + \cdots\right) = \left[\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2} + 2\epsilon\left(\frac{\partial^2}{\partial s_1 \partial \omega_1} + \frac{\partial^2}{\partial s_2 \partial \omega_2}\right) + \epsilon^2\left(\frac{\partial^2}{\partial \omega_1^2} + \frac{\partial^2}{\partial \omega_2^2}\right)\right]$$
$$\times \left[\delta\left(\lambda_0 + \epsilon\lambda_1 + \epsilon^2\lambda_2 + \cdots\right)\right] + \epsilon^2\gamma\left(\lambda_0 + \epsilon\lambda_1 + \epsilon^2\lambda_2 + \cdots\right)\left[1 - \frac{\lambda_0 + \epsilon\lambda_1 + \epsilon^2\lambda_2 + \cdots}{K}\right].$$

Gathering terms of $\mathcal{O}(\epsilon^0)$, we have $0 = \left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}\right)\delta\lambda_0$, which implies $\delta\lambda_0 = C_0(\boldsymbol{\omega}, t)$. Gathering terms of $\mathcal{O}(\epsilon^1)$, we have $0 = \left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}\right)\delta\lambda_1$. Because λ_1 satisfies the same equation as λ_0, no new information is provided, and without loss of generality we let $\lambda_1 = 0$ (Garlick et al., 2011). Finally, gathering terms of $\mathcal{O}(\epsilon^2)$, we have

$$\frac{1}{\delta}\frac{\partial}{\partial t}C_0(\boldsymbol{\omega}, t) = \left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}\right)(\delta\lambda_2) + \left(\frac{\partial^2}{\partial \omega_1^2} + \frac{\partial^2}{\partial \omega_2^2}\right)C_0(\boldsymbol{\omega}, t) + \gamma\frac{C_0(\boldsymbol{\omega}, t)}{\delta}\left(1 - \frac{C_0(\boldsymbol{\omega}, t)}{\delta K}\right).$$

A solvability condition for λ_2 (Garlick et al., 2011) requires the non-homogeneous terms (terms *not* involving λ_2) integrate to zero on scales larger than $\boldsymbol{s}$. Following Yurk and Cobbold (2018) and Maciel and Lutscher (2018), we integrate over a region, Ω, which is intermediate in scale between

$\boldsymbol{s}$ and $\boldsymbol{\omega}$. This integration leads to the *homogenized* equation for C_0,

$$\frac{\partial}{\partial t}C_0(\boldsymbol{\omega},t) = D(\boldsymbol{\omega})\left(\frac{\partial^2}{\partial \omega_1^2} + \frac{\partial^2}{\partial \omega_2^2}\right)C_0(\boldsymbol{\omega},t) + \gamma C_0(\boldsymbol{\omega},t)\left(1 - \frac{C_0(\boldsymbol{\omega},t)}{\tilde{K}(\boldsymbol{\omega})}\right), \tag{2.7}$$

where the *homogenized* diffusion coefficients are

$$D(\boldsymbol{\omega}) = \frac{|\Omega|}{\int_\Omega \frac{1}{\delta} d\boldsymbol{s}},$$

and the *homogenized* density-dependence parameters are

$$\tilde{K}(\boldsymbol{\omega}) = \frac{K|\Omega|}{D\int_\Omega \frac{1}{\delta^2} d\boldsymbol{s}},$$

with $|\Omega| = \int_\Omega 1 d\boldsymbol{s}$.

To obtain numerical solutions, we discretized (2.7) using first-order forward differences in time, and centered differences in space (Wikle, 2003; Zheng and Aukema, 2010; Hooten and Hefley, 2019), such that

$$\begin{aligned}
\frac{\partial}{\partial t}C_0(\boldsymbol{\omega},t) &\approx \frac{C_0(\boldsymbol{\omega},t) - C_0(\boldsymbol{\omega},t-\Delta t)}{\Delta t}, \\
\frac{\partial^2}{\partial \omega_1^2}C_0(\boldsymbol{\omega},t) &\approx \frac{C_0(\omega_1+\Delta\omega_1,\omega_2,t) - 2C_0(\boldsymbol{\omega},t) + C_0(\omega_1-\Delta\omega_1,\omega_2,t)}{\Delta\omega_1^2}, \\
\frac{\partial^2}{\partial \omega_2^2}C_0(\boldsymbol{\omega},t) &\approx \frac{C_0(\omega_1,\omega_2+\Delta\omega_2,t) - 2C_0(\boldsymbol{\omega},t) + C_0(\omega_1,\omega_2-\Delta\omega_2,t)}{\Delta\omega_2^2}.
\end{aligned}$$

As a result of applying the above differences, we have

$$\begin{aligned}
C_0(\boldsymbol{\omega}, t) \approx\, & C_0(\boldsymbol{\omega}, t-\Delta t)\left[1 - 2D(\boldsymbol{\omega})\left(\frac{\Delta t}{\Delta\omega_1^2} + \frac{\Delta t}{\Delta\omega_2^2}\right) + \gamma\Delta t\right] \\
& + C_0(\omega_1 - \Delta\omega_1, \omega_2, t-\Delta t)\left[\frac{\Delta t}{\Delta\omega_1^2}D(\boldsymbol{\omega})\right] \\
& + C_0(\omega_1 + \Delta\omega_1, \omega_2, t-\Delta t)\left[\frac{\Delta t}{\Delta\omega_1^2}D(\boldsymbol{\omega})\right] \\
& + C_0(\omega_1, \omega_2 - \Delta\omega_2, t-\Delta t)\left[\frac{\Delta t}{\Delta\omega_2^2}D(\boldsymbol{\omega})\right] \\
& + C_0(\omega_1, \omega_2 + \Delta\omega_2, t-\Delta t)\left[\frac{\Delta t}{\Delta\omega_2^2}D(\boldsymbol{\omega})\right] \\
& - C_0(\boldsymbol{\omega}, t-\Delta t)^2\left(\frac{\gamma}{\tilde{K}(\boldsymbol{\omega})}\Delta t\right). \qquad (2.8)
\end{aligned}$$

We rewrite (2.8) using matrix notation as

$$\boldsymbol{C}_0(t) \approx \boldsymbol{H}\boldsymbol{C}_0(t-\Delta t) - \boldsymbol{C}_0(t-\Delta t)^2\left(\frac{\gamma}{\tilde{\boldsymbol{K}}}\Delta t\right), \qquad (2.9)$$

where $\boldsymbol{H}$ is a propagator matrix with five non-zero entries row-wise, except those related to boundary conditions. The model described in (2.9) fits into the class of general quadratic nonlinear models developed by Wikle and Hooten (2010).

Solving (2.7) numerically yields approximate solutions for $\lambda(\boldsymbol{s}, t)$ at the large spatial scale. To retrieve the approximate small scale solutions, we use $\lambda(\boldsymbol{s}, t) \approx C_0(\boldsymbol{\omega}, t)/\delta(\boldsymbol{s})$. Graphical illustrations of the homogenization procedure in our application can be found in Section A.3. The homogenized solution we derived will only apply exactly when consistent initial conditions are given. However, as shown by Garlick et al. (2011), solutions with components not precisely aligned with the homogenization assumptions will decay exponentially rapidly to the homogenized solution, and consequently, associated errors can be safely neglected on the slow temporal scale. Homogenization in two-dimensional space reduces computation complexity by $\mathcal{O}(\epsilon)$ in each spatial dimension, and relaxation of the numerical stability requirement that temporal discretization scales with the square of spatial discretization further reduces complexity by $\mathcal{O}\left(\epsilon^2\right)$. Because $\epsilon \approx 0.1$, our algorithm for solving (2.4) using homogenization is about 10^4 times faster than solving it without using homogenization.

2.3 Application

2.3.1 Data

To estimate spatio-temporal sea otter abundance in Glacier Bay, and to understand the effect of environmental factors on their population dynamics, we fit our model using data from three different sources of sea otter counts: aerial surveys and intensive survey units (ISU) using visual observers, as well as aerial photographic images.

The aerial survey data were collected during 1993, 1996-2006, 2009, 2010, and 2012, by observers flying in an aircraft at an elevation of 300ft over 400m-wide transects systematically placed across Glacier Bay. Sea otters were counted from the aircraft over contiguous 400m×400m regions.

The ISU data were collected during 1999-2004, 2006, and 2012, using the method developed by Bodkin and Udevitz (1999). During an aerial survey, intensive searches were initiated upon detection of sea otters, by observers flying repeatedly along the circumference of a 400m×400m region until no additional individuals were observed. The ISU data serve as a direct observation of true abundances $\boldsymbol{N}_o$.

The data collected in 2017 and 2018 reflected recent advancements in survey technology, which include using aerial photographic surveys instead of observer-based methods. After the survey, the sea otters in each image were counted by a trained observer. Each image covers a 60m×90m region, with overlap between two consecutive images. To reconcile the spatial scales of data collected during photographic surveys and observer-based surveys, we assumed homogeneous population intensity within a 400m×400m region, so that the intensity over any 60m×90m sub-region is proportional to the intensity over the 400m×400m region. In addition, we used only non-overlapping images that are conditionally independent samples. Modeling dependence among overlapping images is beyond the scope of this paper; however, see Williams et al. (2017a) for a detailed discussion on using image overlap to estimate detection probability. Lastly, we aggregated counts in images belonging to the same 400m×400m region, so that

$$y_{i,t} = \sum_{j=1}^{n_{i,t}} y_{i,j,t} \sim \text{Binom}(N_{i,t},\ p_t),$$

$$N_{i,t} \sim \text{NB}(n_{i,t} A \lambda_{i,t},\ n_{i,t}\tau),$$

where $y_{i,j,t}$ denotes the observed relative abundance in the jth image at site i in year t, $n_{i,t}$ denotes the number of non-overlapping images, and $A = \frac{60\times 90}{400\times 400}$ denotes the ratio between the two survey spatial scales.

2.3.2 Simulation

We conducted a simulation study to compare our model that includes the logistic reaction component to a model with the Malthusian reaction component, when population dynamics follow density-regulated growth. We denote our hierarchical model outlined in Section 2.2.1 as the "logistic model," and the model with the same hierarchy but the following process,

$$\frac{\partial}{\partial t}\lambda(\boldsymbol{s},t) = \left(\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2}\right)\delta(\boldsymbol{s})\lambda(\boldsymbol{s},t) + \gamma\lambda(\boldsymbol{s},t),$$

as the "Malthusian model."

We simulated sea otter population intensities at 400m×400m spatial resolution over Glacier Bay from 1993 to 2018 using the process model in Section 2.2.1. Then, we generated true abundances and relative abundances using the data model in Section 2.2.1. We sampled relative and true abundances similar to the actual data collection procedure. That is, in each year from 1993 to 2018, we first randomly sampled horizontal strips across Glacier Bay as our transects, where we recorded observed relative abundances, $\boldsymbol{Y}$. Then, we randomly sampled ISU locations from these transects, where we recorded observed true abundances, $\boldsymbol{N}_o$. We summarized the true parameter values and their posterior distributions resulting from the two model fits in Table 2.1.

Table 2.1: True parameter values and estimated posterior means (95% credible intervals) from the logistic and the Malthusian models, for the simulated data.

Parameter	True	Logistic	Malthusian
τ (dispersion)	0.5	0.49 (0.47, 0.51)	0.48 (0.46, 0.50)
β_0 (intercept)	18	17.95 (17.59, 18.29)	18.12 (17.61, 18.55)
β_1 (depth)	-1.5	-1.49 (-1.54, -1.43)	-1.52 (-1.57, -1.47)
β_2 (distance to shore)	0.8	0.76 (0.72, 0.80)	0.75 (0.72, 0.79)
β_3 (bottom slope)	-0.3	-0.29 (-0.34, -0.24)	-0.29 (-0.34, -0.23)
β_4 (shoreline complexity)	1	0.97 (0.93, 1.01)	0.98 (0.94, 1.01)
θ (magnitude)	500	467 (412, 517)	786 (728, 845)
κ (range)	60	60 (47, 74)	81 (67, 98)
γ (growth rate)	0.25	0.25 (0.24, 0.26)	0.18 (0.17, 0.19)
K (density dependence)	5	4.79 (4.25, 5.38)	-

Table 2.1 shows that the logistic model was able to capture all true parameter values in their respective 95% credible intervals. On the other hand, the Malthusian model overestimated initial conditions for magnitude and range, and underestimated intrinsic growth rate based on simulated data. The Malthusian model was also unable to provide inference on the density-dependence parameter, K, due to misspecification.

An important quantity derived from spatio-temporal forecasts of sea otter abundance in Glacier Bay is the total abundance through time, $N(t) = \int_{\mathcal{S}} N(\boldsymbol{s}, t) d\boldsymbol{s}$. A sample of total abundance in year t is obtained by

$$N^{(r)}(t) = \sum_{i=1}^{n_{o,t}} N_{i,t} + \sum_{j=1}^{n_t - n_{o,t}} \hat{N}_{j,t}^{(r)} + \sum_{k=1}^{n-n_t} \tilde{N}_{k,t}^{(r)},$$

where $N_{i,t}$ is an observation of true abundance, $\hat{N}_{j,t}^{(r)}$ is a posterior sample of true abundance where relative abundance was observed, and $\tilde{N}_{k,t}^{(r)}$ is a posterior predictive sample of true abundance where no data were observed. Figure 2.1 indicates that the logistic model was able to capture true total abundances in their respective 95% credible intervals; however, the Malthusian model tended to overestimate abundance after year 2015. This demonstrates the limitation of the Malthusian model when population growth is density regulated. Although the exponential growth curve may mimic the behavior of the logistic growth curve before total abundance reaches the inflection point (possibly by overestimating initial abundance and underestimating growth rate), it will nonetheless deviate from the truth as population size approaches the asymptote.

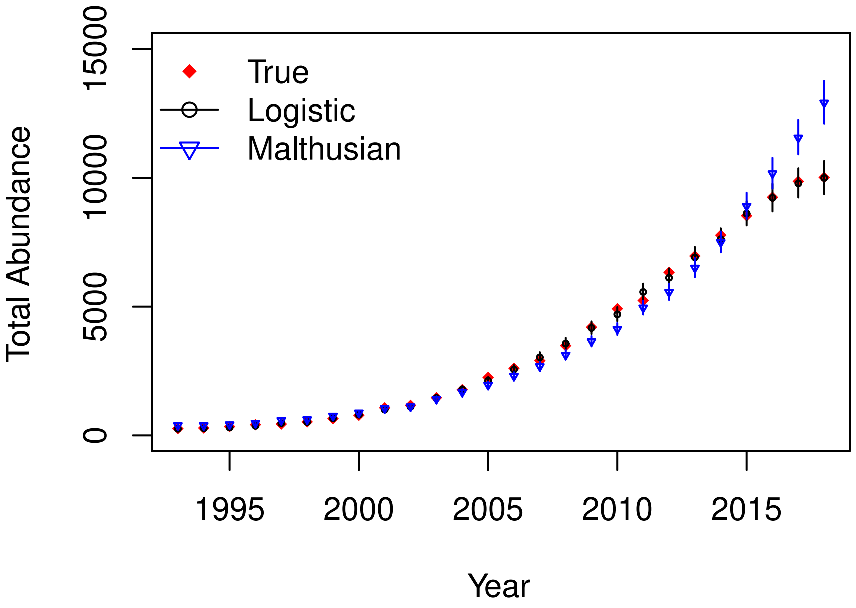


Figure 2.1: Estimated posterior predictive means and 95% credible intervals for total abundances, $N(t)$, from the logistic and the Malthusian models, overlaid with true total abundances, for the simulated data.

2.3.3 Case Study

We fit our model to the data described in Section 2.3.1. For homogenization, we defined the small computational scale in space to be 400m×400m, and the large computational scale in space to be 4000m×4000m, at which we solved the discretized PDE in (2.8). We calculated the homogenized coefficients over areas $\Omega = 6000\text{m} \times 6000\text{m}$ centered on each large-scale cell. The homogenization scale and the small and large computational scales were selected based on previous implementations to balance between desired accuracy and available computational resources. We ran the MCMC algorithm in R version 3.0.2 (R Core Team, 2019) with 15,000 iterations, and used a burn-in of 7,500 and a thinning rate of 1/10. Table 2.2 summarizes posterior distributions of model parameters. All four coefficients for environmental covariates have 95% credible intervals that did not include zero, suggesting that sea otter diffusion is significantly influenced by habitat. Specifically, high motility is related to deep water, areas away from the shore, steep bottom slopes, and complex shorelines. The posterior mean intrinsic growth rate of 0.24 is close to the estimate reported by Estes (1990), which is reasonable to expect during colonization. Figure 2.2 shows that the estimated total abundances from our model agreed with the design-based estimates available during 1999-2004, 2006, and 2012 (Bodkin and Udevitz, 1999). Maps of the log of posterior predictive mean abundances and the table summarizing posterior predictive total abundances can be found in Section A.2.

Table 2.2: Estimated posterior means and 95% credible intervals of model parameters, for the sea otter case study.

Parameter	Posterior Mean	95% CI
τ (dispersion)	0.032	(0.030, 0.034)
β_0 (intercept)	16.09	(15.91, 16.32)
β_1 (depth)	-1.12	(-1.30, -0.95)
β_2 (distance to shore)	0.18	(0.08, 0.27)
β_3 (bottom slope)	-0.79	(-0.91, -0.64)
β_4 (shoreline complexity)	0.81	(0.71, 0.91)
θ (magnitude)	649	(515, 801)
κ (range)	8.07	(7.07, 9.19)
γ (growth rate)	0.25	(0.23, 0.27)
K (carrying capacity)	6.12	(4.41, 8.74)

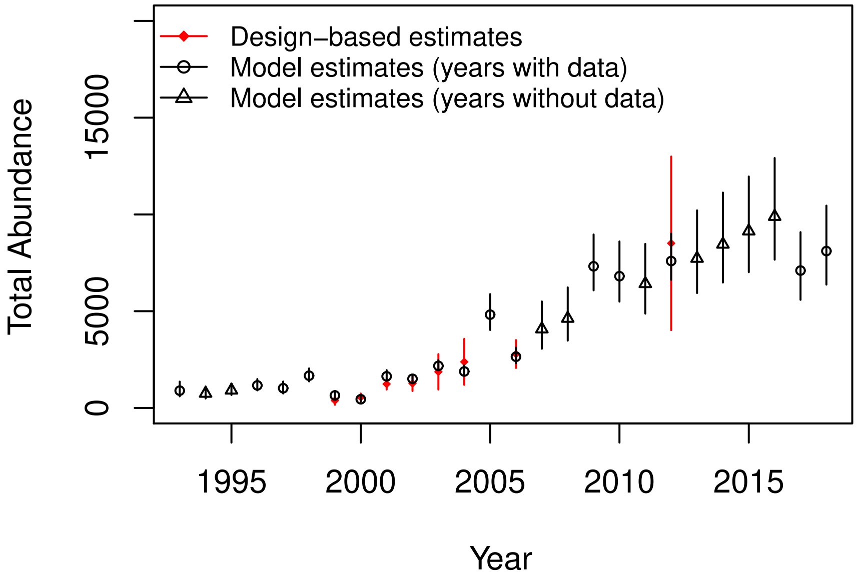


Figure 2.2: Estimated posterior predictive mean total abundances, $N(t)$, and their 95% credible intervals, overlaid with design-based estimates and their uncertainties, for the sea otter case study.

To demonstrate that the logistic model improves forecasts of sea otter abundance in Glacier Bay, we conducted a 5-fold cross-validation using the posterior predictive score (Gelman et al., 2014; Hooten and Hobbs, 2015)

$$\sum_{m=1}^{M} \log \left(\frac{\sum_{r=1}^{R} [\boldsymbol{N}_o^m | \boldsymbol{Y}, \boldsymbol{N}_o^{-m}, \boldsymbol{\theta}^{(r)}]}{R} \right),$$

where $\boldsymbol{N}_o^m$ and $\boldsymbol{N}_o^{-m}$ are the observed true abundances for validation and training in the mth fold, respectively. The vector $\boldsymbol{\theta}^{(r)} = \left(\boldsymbol{p}^{(r)}, \tau^{(r)}, \boldsymbol{\beta}^{(r)}, \theta^{(r)}, \kappa^{(r)}, \gamma^{(r)}, K^{(r)}\right)$ is the rth posterior sample of parameters. The score for the logistic model (-2880) showed an improvement in forecast ability over the score for the Malthusian model (-2896).

The logistic model also allowed us to investigate the equilibrium abundance of sea otters in Glacier Bay. Population dynamics at equilibrium satisfy $\frac{\partial}{\partial t}\lambda(\boldsymbol{s}, t) = 0$, and we can determine the state of equilibrium numerically using posterior predictive samples of abundances, such that for $0 < u \ll 1$,

$$\left| \frac{1}{R} \sum_{r=1}^{R} \left(\tilde{N}_{k,t_e}^{(r)} - \tilde{N}_{k,t_e - \Delta t}^{(r)} \right) \right| < u, \text{ for } k = 1, \ldots, n.$$

We denote T_e, the smallest t_e that satistfies the above condition, as the time of equilibrium. In our case study, we found $T_e = 2050$ to provide an acceptable approximation. Alternatively, because analytical solutions to (2.7), $C_0(\boldsymbol{\omega}, t)$, converges to $\tilde{K}(\boldsymbol{\omega})$ away from boundaries as t goes to infinity, the homogenization procedure suggests that $\lambda\left(\boldsymbol{s}, T_e\right) \approx \tilde{K}(\boldsymbol{\omega})/\delta(\boldsymbol{s})$ at system equilibrium. Therefore, we obtained a posterior predictive realization of local equilibrium abundance by sampling from the predictive full-conditional distribution,

$$\tilde{N}_{k,T_e}^{(r)} \sim \text{NB}\left(\frac{\tilde{K}^{(r)}(\boldsymbol{\omega})}{\delta^{(r)}(\boldsymbol{s})}, \tau^{(r)} \right).$$

We mapped mean equilibrium abundances with associated uncertainties in Figure 2.3. Our analysis shows that, while K may be perceived as a parameter that regulates local abundance through intraspecific competition, local abundance at equilibrium is not bounded by K (Yurk and

Cobbold, 2018). The solution to $\frac{\partial}{\partial t}\lambda\left(\boldsymbol{s},T_e\right)=0$ leads to the equation $\frac{\gamma}{K}\lambda\left(\boldsymbol{s},T_e\right)\left(K-\lambda\left(\boldsymbol{s},T_e\right)\right)=\left(\frac{\partial^2}{\partial s_1^2}+\frac{\partial^2}{\partial s_2^2}\right)\delta(\boldsymbol{s})\lambda\left(\boldsymbol{s},T_e\right)$, and $\lambda\left(\boldsymbol{s},T_e\right)$ will exceed K when the second derivatives on the right hand side are negative.

Further, we refer to the total abundance at equilibrium as the "effective carrying capacity," and it is unlikely to reach the "nominal carrying capacity" obtained by integrating K over the study area (Figure 2.4). When diffusion coefficients are constant in space, the local equilibrium intensities will approach K asymptotically, and the effective carrying capacity will converge to the nominal carrying capacity. However, when diffusion coefficients are spatially heterogeneous, Jensen's inequality implies that

$$\frac{\tilde{K}(\boldsymbol{\omega})}{D(\boldsymbol{\omega})}=\frac{K|\Omega|}{\int_\Omega \frac{1}{\delta^2}d\boldsymbol{s}}\cdot\left(\frac{1}{|\Omega|}\int_\Omega \frac{1}{\delta}d\boldsymbol{s}\right)^2<\frac{K}{\left(\int_\Omega \frac{1}{\delta}d\boldsymbol{s}\right)^2}\cdot\left(\int_\Omega \frac{1}{\delta}d\boldsymbol{s}\right)^2=K,$$

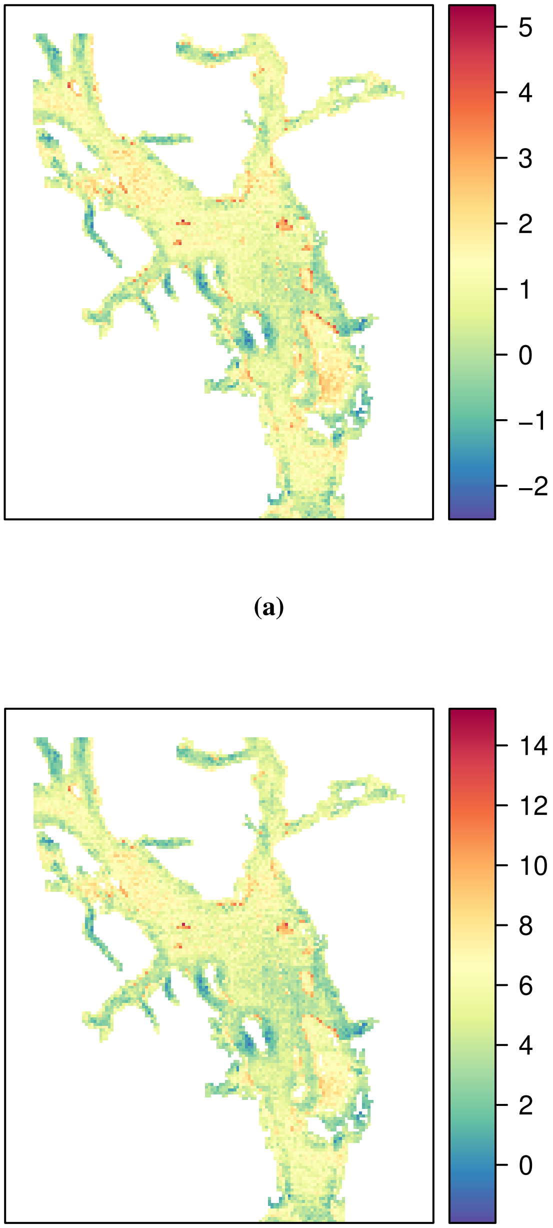
and because $C_0(\boldsymbol{\omega},t)\rightarrow\tilde{K}(\boldsymbol{\omega})$, we can infer an upper bound on population intensities,

$$\lambda(\boldsymbol{s},t)\approx\frac{C_0(\boldsymbol{\omega},t)}{\delta(\boldsymbol{s})}\rightarrow\frac{\tilde{K}(\boldsymbol{\omega})}{\delta(\boldsymbol{s})}<\frac{D(\boldsymbol{\omega})K}{\delta(\boldsymbol{s})}.$$

Integrating the above inequality over the entire study domain results in

$$\int_{\mathcal{S}}\lambda\left(\boldsymbol{s},T_e\right)d\boldsymbol{s}<K\int_{\mathcal{S}}\frac{D(\boldsymbol{\omega})}{\delta(\boldsymbol{s})}d\boldsymbol{s}=K|\mathcal{S}|,$$

which indicates that the mean effective carrying capacity is bounded above by the nominal carrying capacity. In fact, the more spatial variability there exists in $\delta(\boldsymbol{s})$, the further the mean effective carrying capacity will be bounded away from the nominal carrying capacity.



(a)

(b)

Figure 2.3: (a) Log of estimated posterior predictive mean equilibrium abundances, $\tilde{N}_{k,T_e}$, for the sea otter case study. (b) Log of estimated posterior predictive equilibrium variance for the sea otter case study.

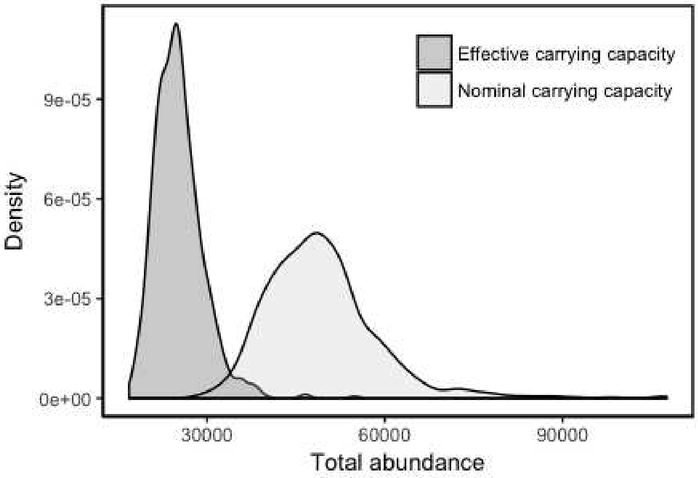


Figure 2.4: Posterior predictive distributions of the effective carrying capacity, $N(T_e)$, versus the nominal carrying capacity, $K|\mathcal{S}|$, for the sea otter case study.

2.4 Discussion

We demonstrated that using logistic growth in the reaction-diffusion model improved forecast of sea otter abundance in Glacier Bay. The logistic reaction component allowed us to infer spatially-varying local equilibrium abundances, and it also enabled us to study the effect of heterogeneous diffusion on the carrying capacity of the system. Logistic growth is a relatively simple model for population growth that demonstrate long-term equilibrium (Turchin, 2003). One way to extend our model is to modify the reaction component by allowing more complicated population dynamics, such as an Allee effect or multiple population equilibria (Estes, 1990). We are exploring such extensions in ongoing research.

Our model is helpful for understanding the impact of preferential dispersion on system equilibrium, and can be applied to ecological processes beyond colonization. For example, our model can be extended to study dispersal-mediated coexistence of multiple species, where population diffusion and growth are driven by predator-prey interactions (Holmes et al., 1994). Further, learning about spatially-varying local equilibirum abundances over a heterogeneous environment will be important for developing future ecological models of the nearshore benthic food web in Glacier

Bay, particularly given the relatively small-scales at which sea otters move, their high site-fidelity, and small home-ranges.

The formulation of a homogenized PDE was essential for model implementation in the Glacier Bay study system associated with our example because it enabled us to obtain inference at a fine spatial resolution with feasible computation time. The homogenized coefficients in the form of harmonic means also provided an alternative way to consider dimension reduction. Although homogenization theory suggests that the small and large spatial scales associated with implementation should be set based on empirical properties (periodicities) of the covariates, in practice, the study system may not be perfectly periodic and the coefficients associated with the influence of the covariates on diffusion are unknown. Thus, the implementation scales are often driven by the availability of data, the amount of computation resources, and the requirements for inference. Because homogenization uses approximation by power series and the order of approximation error is the same as the ratio of small-to-large spatial scales, faster computation will come at the cost of less accuracy.

Finally, the ongoing collection of aerial imagery provides an incentive for developing statistical sampling methods to optimally combine supervised and unsupervised object classification approaches (Seymour et al., 2017). It also motivates the development of a statistically rigorous georectification procedure, whose uncertainty will be measured in a hierarchical framework, so that we can better account for replications and detectabilities using image overlaps.

Chapter 3

Improving Wildlife Population Inference using Aerial Imagery and Entity Resolution

3.1 Introduction

Aerial surveys are widely used to provide abundance information about various terrestrial and marine species (Caughley, 1974; Ver Hoef, 2014). Compared to traditional observer-based surveys, imagery surveys have the advantage of reducing risk for observers and providing a permanent record that can be independently verified (Buckland et al., 2012). In addition to population counts, the imagery data (often referred to as photographs; Figure B.3) provide individual-level information such as color, size, and location, which can be leveraged to identify animals without marking them (Williams et al., 2020). This gain in information leads to more reliable modeling of population abundance than using count data only (Dennis et al., 2015; Barker et al., 2018; Ketz et al., 2019). In what follows, we describe a Bayesian hierarchical model to identify unique individuals in overlapping images and estimate population size under a unified framework. We apply our model to analyze aerial imagery data of sea otters (*Enhydra lutris kenyoni*) in Glacier Bay, Alaska. During a survey, images are acquired at a regular time interval with overlapping regions in the direction of aircraft movement as it flies along transects systematically placed across the Glacier Bay. Sea otters in the images are located and counted by trained observers after the survey. Past studies using these data have either discarded overlapping images to meet the independent count assumption of binomial models (Lu et al., 2019), or treated counts from overlapping regions as temporal replicates in N-mixture models (Williams et al., 2017b). We demonstrate that our method reduces bias in abundance estimation and improves previously described methods.

The information we use to resolve individual identities are the observed locations of individuals in a sequence of images. However, distortion of individual positions may be present when the aircraft deviates from its scheduled trajectory due to a variety of reasons that can influence altitude and aircraft position, resulting in an artificial transformation of the image footprints. Further, micro-movement of sea otters and pinpoint uncertainty by observers make exact matching of observed locations in overlapping regions nearly impossible. There exists a rich literature on image stitching where the common objective is to optimally combine a sequence of overlapping images into a composite image by minimizing a loss function (Levin et al., 2004; Szeliski, 2006; Brown and Lowe, 2015; Gross and Heumann, 2016). However, optimization-based image stitching algorithms do not usually provide uncertainty about the stitching process and are seldom integrated into other models to provide additional learning about the system. On the other hand, the statistical literature associated with entity resolution, also known as record linkage when the objective is to merge multiple data files (in our case, images) in the absence of unique identifiers (in our case, individual tags, for example), may provide a theoretical basis for uncertainty quantification. We incorporate uncertainty in the record linkage process into a capture-recapture model for abundance estimation.

Traditional approaches to record linkage compare similarities between pairs of records from which matching decisions are made (Fellegi and Sunter, 1969; Jaro, 1989; Winkler, 1995). Larsen and Rubin (2001) presented record linkage in terms of mixture models that mix linkage probability of record pairs between a model for probable links and a model for probable nonlinks. Bayesian approaches based on the same idea were developed by Fortini et al. (2001), McGlincy (2004), and Larsen (2004). However, comparison-based approaches are largely infeasible computationally even when the number of possible pairs is moderately large (Winkler, 2006). One way to reduce the computational cost of record linkage is by “blocking," where records partitioned into different blocks are considered nonlinks *a priori* (Christen, 2011; Steorts et al., 2014). Alternatively, record linkage can be presented as the clustering of observed records by unobserved identities (Copas and Hilton, 1990; Tancredi and Liseo, 2011; Liseo and Tancredi, 2011; Steorts et al., 2015; Tancredi et al., 2018). Each latent identity has a “true" value and the associated records are modeled as stochastic distortions from the truth. Steorts et al. (2015) introduced the graphical record linkage model by representing the linkage structure as a bipartite graph between observed records and latent identities. By comparing records to latent identities instead of each other, the computation time to link d data files with a maximum of n records per file can be substantially reduced from $\mathcal{O}\left(n^d\right)$ to $\mathcal{O}(dn)$. One distinction between the graphical record linkage model and other non-parametric clustering methods such as Dirichlet process models and Pitman-Yor process models is that the latter often assume linear growth of cluster size with the size of data (Wallach et al., 2010; Betancourt et al., 2016), whereas in record linkage problems, co-referent clusters tend to stay small even when the number of records grows. Following Liseo and Tancredi (2011) and Steorts et al. (2015), we made use of multivariate normal models on the latent truths to identify unique individuals in imagery data.

The output of a record linkage model can be used to learn about population size. When uncertainty exists in linkage structure, record linkage and size estimation are often regarded as two separate stages. Sadinle (2018) proposed using “linkage-averaging" to transfer linkage uncertainty as quantified by Bayesian posterior samples into the subsequent stage of population size estimation. Although linkage-averaging facilitates model exploration by allowing the combination of different record linkage models with population models, any bias in the record linkage stage will propagate into the size estimation stage regardless of model choice (Tancredi and Liseo, 2011). Our hierarchical framework naturally relates entity resolution and abundance estimation as one generative process, thereby allowing information exchange and feedback between these two model objectives. Other unified modeling approaches exist, including those presented by Link et al. (2009) and Wright et al. (2009) that incorporate misidentification into capture-recapture models by sampling from latent multinomial distributions, the hierarchical record linkage models proposed by Tancredi and Liseo (2011) and Liseo and Tancredi (2011) that reflect capture-recapture dynamics through latent matching matrices, and the latent Poisson process model proposed by Green and Mardia (2005) to align partially labeled protein structures. We propose a novel framework that combines a record linkage model that aligns distorted animal locations and a spatial capture-recapture model

(Royle and Young, 2008) that accounts for heterogeneity in detection probability due to temporally changing survey units.

We present our hierarchical record linkage model in Section 3.2. In Section 3.3, we illustrate the model through simulation and a case study using aerial photographs of sea otters in Glacier Bay, Alaska. Finally in Section 3.4, we discuss possible extensions and broader applications of our model.

3.2 Model

3.2.1 Data model

Consider a sequence of T images with n_t observed individuals in image t, for $t = 1, \ldots, T$ (see Figure B.2, for example). Let $\boldsymbol{y}_{i,t}$ denote the observed Universal Transverse Mercator (UTM) coordinates of the ith individual in image t, and let $\boldsymbol{u}_{i,t}$ denote the true location of that individual. Distortion in $\boldsymbol{y}_{i,t}$ occurs in laboratory processing when the image footprint, $\mathcal{F}_t$, is artificially scaled and rotated to fit in a template, $\mathcal{Q}_t$, assuming the aircraft trajectory follows a fixed height and orientation. Using the known image center $\boldsymbol{\mu}_t$ (recorded during the aerial survey) as a reference point, we connect the distorted displacements of the observed locations with the true locations $\boldsymbol{u}_{i,t}$ and the image center as

$$\boldsymbol{y}_{i,t} - \boldsymbol{\mu}_t = (1 + c_t)\, \boldsymbol{R}\left(\theta_t\right) \left(\boldsymbol{u}_{i,t} - \boldsymbol{\mu}_t\right), \tag{3.1}$$

where the counterclockwise rotation matrix is given by

$$\boldsymbol{R}\left(\theta_t\right) = \begin{pmatrix} \cos\theta_t & -\sin\theta_t \\ \sin\theta_t & \cos\theta_t \end{pmatrix}.$$

The scaling parameter, c_t, and the rotation parameter, θ_t, are modeled using basis function regression (Hefley et al., 2017). To ensure smoothness and flexibility in the aircraft trajectory, we specify

$$\begin{aligned} c_t &= \boldsymbol{w}(t)'\boldsymbol{\alpha}, \\ \theta_t &= \boldsymbol{v}(t)'\boldsymbol{\beta}, \end{aligned} \tag{3.2}$$

where $\boldsymbol{w}(t)$ and $\boldsymbol{v}(t)$ are the basis functions evaluated at time t for scaling and rotation, respectively. Due to unknown distortion, the true image footprints are also unknown, and we model the four vertices of the rectangular image footprint $\mathcal{F}_t$ through a georectification process from the known template $\mathcal{Q}_t$,

$$\boldsymbol{\nu}_{j,t} - \boldsymbol{\mu}_t = \frac{1}{(1+c_t)} \boldsymbol{R}\left(-\theta_t\right) \left(\boldsymbol{\nu}^*_{j,t} - \boldsymbol{\mu}_t\right), \; j = 1,2,3,4, \tag{3.3}$$

where $\boldsymbol{\nu}_{j,t}$ denote the vertices of $\mathcal{F}_t$ and $\boldsymbol{\nu}^*_{j,t}$ denote the vertices of $\mathcal{Q}_t$.

We assume every observed individual has a latent identity, $\lambda_{i,t}$, that may be shared across images but not within the same image. The true locations, $\boldsymbol{u}_{i,t}$, are modeled as Gaussian conditioned on a transient activity center associated with the latent identity, $\boldsymbol{s}_{\lambda_{i,t}}$, and movement uncertainty $\sigma_u^2 \boldsymbol{I}$, such that

$$\boldsymbol{u}_{i,t} | \boldsymbol{s}_{\lambda_{i,t}}, \sigma_u^2 \sim \mathrm{N}\left(\boldsymbol{s}_{\lambda_{i,t}}, \sigma_u^2 \boldsymbol{I}\right).$$

We marginalize over $\boldsymbol{u}_{i,t}$ to obtain the integrated conditional distributions of the observed locations as follows,

$$\boldsymbol{y}_{i,t} | \boldsymbol{s}_{\lambda_{i,t}}, \sigma_u^2, c_t, \theta_t \sim \mathrm{N}\left(\boldsymbol{\mu}_t + (1+c_t)\boldsymbol{R}\left(\theta_t\right)\left(\boldsymbol{s}_{\lambda_{i,t}} - \boldsymbol{\mu}_t\right), \sigma_u^2 \left(1+c_t\right)^2 \boldsymbol{R}\left(\theta_t\right) \boldsymbol{R}^T\left(\theta_t\right)\right). \tag{3.4}$$

Based on the latent identities, the data model in (3.4) allows us to minimize the Procrustes distance (Dryden and Mardia, 1998) between configurations of points in the overlapping regions, and the process model that we describe in what follows enables inference about the latent identities.

3.2.2 Process model

We adopt a parameter expanded data augmentation approach (Royle, 2009; Royle and Dorazio, 2012) and assume there is a super-population of size M much greater than the total number of observations in a study domain $\mathcal{D}$ that contains the union of all image footprints. Each individual in the super-population has a binary variable z_m representing whether the individual belongs to the population being sampled, where $z_m \sim \text{Bern}(\psi)$ for $m = 1, \ldots, M$. Conditional on the latent identities of the observed individuals, the augmented data are a zero-inflated version of the capture history. The prior specification on the zero-inflation parameter ψ along with the super-population size M implicitly suggest a prior for the unknown population size N (Royle et al., 2007).

We let $\boldsymbol{\lambda}_t$ denote the vector of latent identities indexed by m for the observed individuals in image t. A plausible configuration of $\boldsymbol{\lambda}_t$ must satisfy two conditions: (a) there are no duplicate identities, and (b) any identity in $\boldsymbol{\lambda}_t$ must be detectable at time t. Otherwise the probability of observing $\boldsymbol{\lambda}_t$ is zero. Each individual in the super-population is associated with an activity center $\boldsymbol{s}_m$. We let the activity centers be uniformly distributed in the study domain *a priori*. We require that an individual is detectable at time t if and only if it is a member of the population being sampled ($z_m = 1$) and its realized location is inside the image footprint at time t ($\boldsymbol{u}_{m,t} \in \mathcal{F}_t$). In the spatial capture-recapture model by Royle and Young (2008), realized locations are fully augmented for all individuals in the super-population and unobserved $\boldsymbol{u}_{m,t}$ are treated as missing data (the model does not account for measurement error so the observations are the realized locations). However, when the observed individuals are unidentified, accounting for missingness becomes challenging. Therefore, we integrate $\boldsymbol{u}_{m,t}$ from the process model by letting $p_{m,t}$ denote the probability that $\boldsymbol{u}_{m,t}$ falls in $\mathcal{F}_t$, conditional on $\mathcal{F}_t$, the activity center $\boldsymbol{s}_m$, and the movement process variance σ_u^2, such that

$$p_{m,t} = \mathbb{P}\left(\boldsymbol{u}_{m,t} \in \mathcal{F}_t \middle| \mathcal{F}_t, \boldsymbol{s}_m, \sigma_u^2\right) = \int_{\mathcal{F}_t} \frac{1}{2\pi\sigma_u^2} \exp\left(-\frac{\left(\boldsymbol{u} - \boldsymbol{s}_m\right)' \left(\boldsymbol{u} - \boldsymbol{s}_m\right)}{2\sigma_u^2}\right) d\boldsymbol{u}. \qquad (3.5)$$

Let p_0 denote the baseline detection probability (e.g., sea otter detectability due to diving behaviors). Then we have

$$\mathbb{P}\left(\lambda_{i,t} = m \middle| z_m, \boldsymbol{s}_m, \mathcal{F}_t, \sigma_u^2, p_0\right) = \begin{cases} p_0 \times p_{m,t}, \text{ if } z_m = 1; \\ 0, \text{ otherwise.} \end{cases}$$

Assuming the individuals are independently detected, the probability of observing $\boldsymbol{\lambda}_t$ is as follows,

$$\mathbb{P}\left(\boldsymbol{\lambda}_t \middle| \left\{z_m\right\}_m, \left\{\boldsymbol{s}_m\right\}_m, \sigma_u^2, \mathcal{F}_t, p_0\right) = \frac{1}{n_t!} \prod_{m:z_m=1} \left\{p_0 p_{m,t} \mathbb{I}\left(m \in \boldsymbol{\lambda_t}\right) + \left(1 - p_0 p_{m,t}\right) \mathbb{I}\left(m \notin \boldsymbol{\lambda}_t\right)\right\}, \tag{3.6}$$

where the factor of $\frac{1}{n_t!}$ indicates that all permutations of $\boldsymbol{\lambda}_t$ are equally likely *a priori*. The process model induces regularization on the number of unique latent identities by controlling the number of activity centers in an image that belong to the population being sampled. When the super-population is much larger than the total number of observed individuals, under-linkage is likely when each observation seeks its own activity center. However, the Bernoulli model on $\lambda_{i,t}$ penalizes “extra" activity centers in the image that are not observed, thereby motivating linkage between observed locations that are spatially proximal.

3.2.3 Parameter model

We used an informative inverse-gamma prior on σ_u^2 because we have specific knowledge about the extent of sea otter movement that is physically possible between consecutive images (Williams, 1989). We imposed a penalization on the second derivatives of the fitted B-splines through the prior variances of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. The penalty parameters were selected by cross-validation (Wahba, 1978; Wood et al., 2016). We specified $\psi \sim \text{Beta}(0.001, 1)$ to approximate a scale prior for N ($[N] \propto 1/N$, Link, 2013), and we centered the prior for p_0 at 0.75 based on a prior data analysis and as suggested in past studies (Williams et al., 2017b; Lu et al., 2019). A full description of prior distributions can be found in Section B.1.

The joint posterior distribution associated with our model is

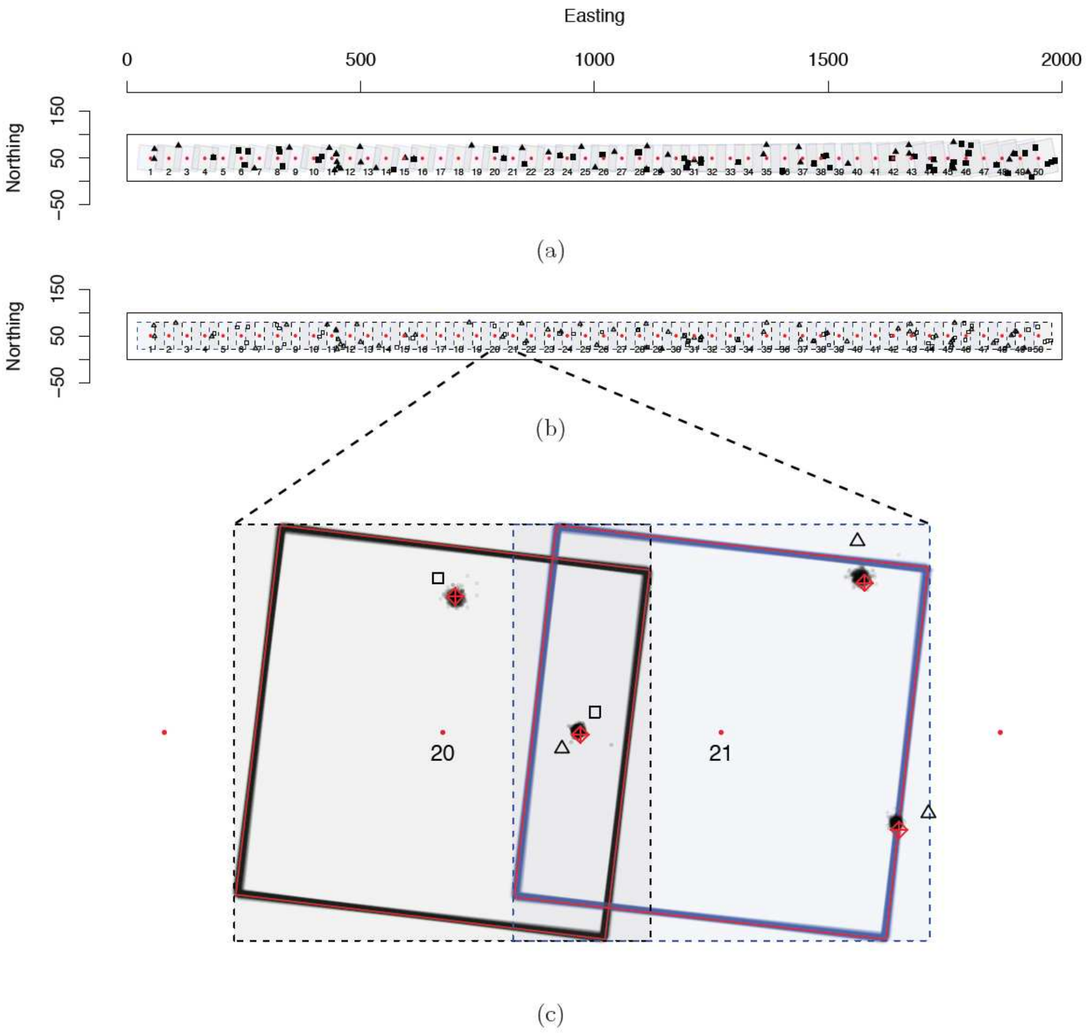
$$\begin{aligned}\left[\{\boldsymbol{\lambda}_t\}_{t=1}^{T}, \{\boldsymbol{s}_m\}_{m=1}^{M}, \sigma_u^2, \boldsymbol{\alpha}, \boldsymbol{\beta}, \{z_m\}_{m=1}^{M}, p_0, \psi \middle| \boldsymbol{Y}\right] \propto & \prod_{t=1}^{T}\prod_{i=1}^{n_t}\left[\boldsymbol{y}_{i,t}\middle|\boldsymbol{s}_{\lambda_{i,t}}, \sigma_u^2, \boldsymbol{\alpha}, \boldsymbol{\beta}\right] \times \prod_{m=1}^{M}\left[\boldsymbol{s}_m\right] \\ & \times \prod_{t=1}^{T}\left[\boldsymbol{\lambda}_t \middle| \{z_m\}_{m=1}^{M}, \{\boldsymbol{s}_m\}_{m=1}^{M}, \sigma_u^2, \boldsymbol{\alpha}, \boldsymbol{\beta}, p_0\right] \\ & \times [p_0] \times \prod_{m=1}^{M}[z_m|\psi] \times [\psi] \\ & \times \left[\sigma_u^2\right] \times [\boldsymbol{\alpha}] \times [\boldsymbol{\beta}].\end{aligned}$$

We implemented our model using MCMC and provide a full description of the algorithm in Section B.4.

3.3 Application

3.3.1 Simulation

We simulated a population of $N = 200$ individuals and sampled their activity centers $\boldsymbol{s}_m$, for $m = 1, \ldots, N$, uniformly from a 100m×2000m study domain, $\mathcal{D}$. For $t = 1, \ldots, T$, $T = 50$, we sampled realized locations $\boldsymbol{u}_{m,t} \sim \mathrm{N}\left(\boldsymbol{s}_m, \sigma_u^2 \boldsymbol{I}\right)$ with $\sigma_u^2 = 0.25$. For the measurement process, we set the image centers to be equally spaced between $\boldsymbol{\mu}_1 = (50, 50)'$ and $\boldsymbol{\mu}_{50} = (1950, 50)'$ and let the footprint template at time t, $\mathcal{Q}_t$, be a 58m×58m square centered at $\boldsymbol{\mu_t}$ and parallel to the horizontal axis. We generated distortion parameters from cubic B-splines with coefficients $\boldsymbol{\alpha} = \boldsymbol{\beta} = (0.1, 0.2, 0.1, -0.1, -0.2)'$ and obtained the image footprint $\mathcal{F}_t$ using (3.3). When $\boldsymbol{u}_{m,t} \in \mathcal{F}_t$, individual m was detected at time t with probability $p_0 = 0.75$. We recorded the distorted locations of the detected individuals by (3.1). Figure 3.1 illustrates the simulated image footprints and the true locations as well as the corresponding footprint templates and the observed locations.



(a)

(b)

(c)

Figure 3.1: (a) Simulated image footprints $\mathcal{F}_t$ overlaid with true locations $\boldsymbol{u}_{i,t}$. The time-indexed points represent image centers $\boldsymbol{\mu}_t$. True locations are marked with "■" in even images and "▲" in odd images. The largest rectangle containing all images is the study domain $\mathcal{D}$; (b) simulated footprint templates $\mathcal{Q}_t$ (dashed rectangles) overlaid with observed locations that are marked with "□" in even images and "△" in odd images; (c) a focused illustration on observed images 20 and 21 (dashed rectangles), overlaid with posterior samples of image footprints (solid rectangles) and activity centers along with their truth (points overlaid with crossed diamonds).

In our implementation of the model, we let the super-population be of size $M = 3000$. We ran the MCMC algorithm in R version 3.0.2 (R Core Team, 2019) for 15000 iterations and used a burn-in of 5000 iterations. For the remaining $K = 10000$ iterations, we obtained posterior realizations of population size as a derived quantity using the posterior sample of z_m as

$$N^{(k)} = \sum_{m=1}^{M} z_m^{(k)},\ k = 1, \ldots, K. \tag{3.7}$$

Table 3.1 summarizes the marginal posterior distributions from the simulated data. Our model captured the true parameters within their respective 95% credible intervals.

Table 3.1: True parameter values and marginal posterior means (95% credible intervals) for the simulated data.

Parameter	True	Posterior mean (95% CI)
p_0	0.75	0.70 (0.58, 0.81)
ψ	0.07	0.06 (0.05, 0.08)
N	200	192 (160, 233)
σ_u^2	0.25	0.26 (0.22, 0.31)

To our knowledge, there have been two studies using aerial photographs to estimate sea otter population in Glacier Bay, Alaska. Lu et al. (2019) proposed a nonlinear reaction-diffusion process model for population intensity, but used only every other image in accordance with the assumptions of their model. An arbitrary selection of images to use for data analysis may lead to bias in abundance estimation, especially if population intensity is spatially heterogeneous. Although our method is not directly comparable to that of Lu et al. (2019), we can compare the estimated number of unique individuals in all images because it refers to the observed abundance. In our simulation, the total number of observations from all images was 111, which correspond to 90 simulated individuals. By counting the number of unique labels in the kth posterior sample of latent identities for $k = 1, \ldots, K$, we estimated a posterior mean of 90 unique individuals. However, counting observations in every other image (to avoid overlapping images) starting from the first image resulted in 60 individuals, while counting from the second image resulted in 51. Discarding half of the images led to inconsistent and insufficient counts, whereas we improved abundance estimates by accounting for duplicated individuals in overlapping regions.

By comparison, Williams et al. (2017b) proposed an N-mixture model where counts in overlapping regions are considered temporal replicates. The model divides images into mutually exclusive regions of overlap and non-overlap, and denotes $y\left(A_i, j\right)$ as the count from the jth overlap of region A_i, such that $\cup_{i=1}^{n} A_i = \cup_{t=1}^{T} \mathcal{F}_t$ and $A_i \cap A_j = \emptyset,\ i \neq j$. Under the assumption of homogeneous detection probability p_0 and population intensity η, the counts are modeled by

$$\begin{aligned} y\left(A_i, j\right) &\sim \text{Binom}\left(N\left(A_i\right), p_0\right) \\ N\left(A_i\right) &\sim \text{Pois}\left(\eta |A_i|\right), \end{aligned} \tag{3.8}$$

where $|A_i|$ is the area of A_i. Although Williams et al. (2017b) accounted for heterogeneity in p_0 and η based on spatial covariates, for illustration, we fit the homogeneous version of their model in (3.8) to our simulated data using an MCMC algorithm. A posterior realization of population size is obtained as a derived quantity by $N^{(k)} = \eta^{(k)} |\mathcal{D}|$, for $k = 1, \ldots, K$ MCMC iterations. The estimated posterior mean abundance was 178 with a 95% credible interval (142, 216).

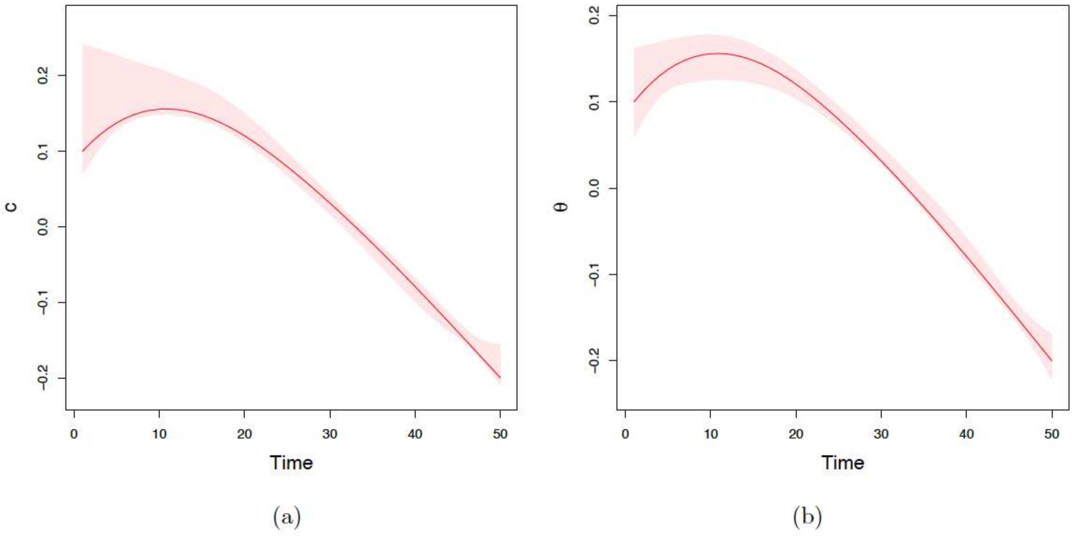
To evaluate our model performance in linkage estimation, we used false discovery rate (FDR) and false negative rate (FNR) as recommended by Steorts (2015) to account for the large number of non-links in our application. There are four possible results when comparing the estimated linkage and the truth:

1. True positive (TP): two individuals have the same latent identity in both estimation and the truth;
2. False positive (FP): two individuals are estimated to have the same latent identity when they are actually different;
3. True negative (TN): two individuals have different latent identities in both estimation and the truth;

4. False negative (FN): two individuals are estimated to have different latent identities when they are actually the same.

We computed the posterior mean FDR and FNR as $\mathbb{E}[\text{FDR}|\boldsymbol{Y}] = \frac{1}{K}\sum_{k=1}^{K} \frac{\text{FP}^{(k)}}{\text{FP}^{(k)}+\text{TP}^{(k)}} = 0.000001$ and $\mathbb{E}[\text{FNR}|\boldsymbol{Y}] = \frac{1}{K}\sum_{k=1}^{K} \frac{\text{FN}^{(k)}}{\text{FN}^{(k)}+\text{TP}^{(k)}} = 0.001$ based on the model fit to simulated data.

Posterior realizations of the scaling and rotation parameters, $\boldsymbol{c}$ and $\boldsymbol{\theta}$, were obtained as derived quantities using (3.2), and posterior realizations of image footprints $\mathcal{F}_t$ were obtained as derived quantities using (3.3). Figure 3.1 illustrates posterior samples of image footprints and activity centers overlaid with the truth for a subset of images (images 20 and 21). Our model performed well, linking observations that correspond to the same individual in the overlapping region and correctly estimating their activity centers despite distortion. Figure 3.2 demonstrates the point-wise 95% credible intervals for $\boldsymbol{c}$ and $\boldsymbol{\theta}$. The point-wise 95% credible intervals contained the true simulated values for both parameters.



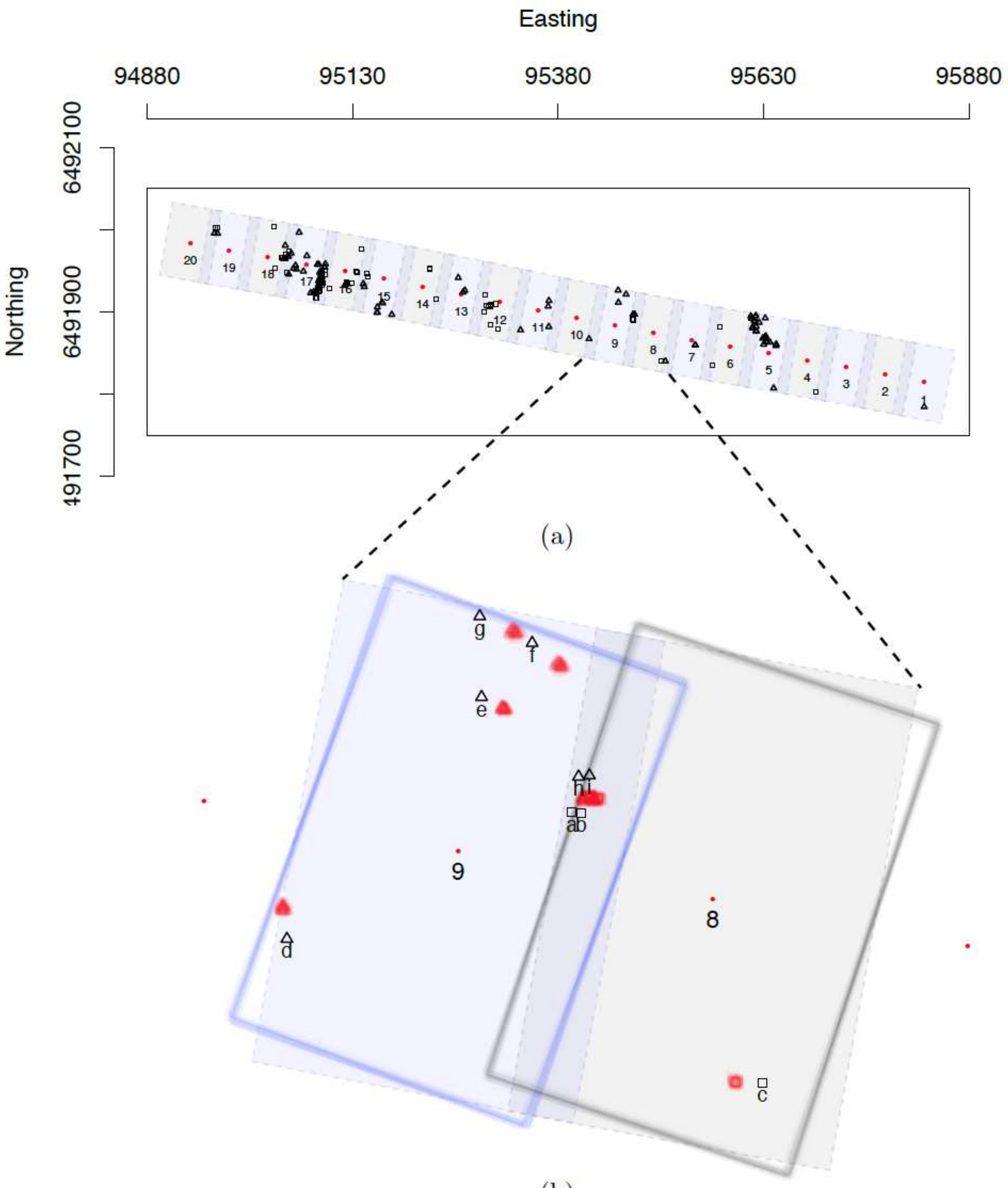
(a)

(b)

Figure 3.2: (a) Estimated point-wise 95% credible interval overlaid with the truth for the scaling parameter $\boldsymbol{c}$; (b) estimated point-wise 95% credible interval overlaid with the truth for the rotation parameter $\boldsymbol{\theta}$.

3.3.2 Case Study

Sea otter populations have undergone significant fluctuations throughout their range over the past two centuries (Jameson et al., 1982). After being hunted to near extinction during the maritime fur trade, sea otter populations have recovered in many areas due to a combination of conservation efforts including the International Fur Seal Treaty Act, translocations, and environmental changes (Larson et al., 2014). Monitoring sea otter colonization in Glacier Bay provides important insight into the ability of a keystone species to recover from near extirpation and to understand their role in structuring the nearshore food web in Glacier Bay (Williams et al., 2019). From 1993 to 2012, surveys were conducted by observers flying in an aircraft over systematically placed transects and counting sea otters (Esslinger et al., 2015). Beginning in 2017, aerial imagery survey methods were implemented where trained observers post-processed images acquired during the model-based optimized surveys to obtain counts of sea otters. For our case study, we analyzed a sequence of 20 consecutive images with a total of 151 observations (see Figure B.3 for an example of real images). Sea otter locations were recorded for 60m×90m footprint templates with long sides perpendicular to the direction of aircraft movement (vectors connecting consecutive image centers). Figure 3.3 illustrates the observed locations overlaid with footprint templates. Our study domain was a 300m×1000m rectangular region containing all image footprints as shown in Figure 3.3.



(a)

(b)

Figure 3.3: (a) Observations from the case study. Footprint templates $\mathcal{Q}_t$ (dashed rectangles) are overlaid with observed locations that are marked with "$\square$" in even images and "$\triangle$" in odd images. The time-indexed points represent image centers $\boldsymbol{\mu}_t$. The largest rectangle containing all templates is the study domain $\mathcal{D}$; (b) a focused illustration on observed images 8 and 9 (dashed rectangles), overlaid with posterior samples of image footprints $\mathcal{F}_t$ (solid rectangles). Observed locations are indexed by letters, and posterior samples of true locations $\boldsymbol{u}_{i,t}$ are shown in solid squares (image 8) and triangles (image 9).

We let the super-population be of size $M = 3000$ and ran the MCMC algorithm for 15000 iterations using a burn-in of 5000 iterations. Table 3.2 summarizes the marginal posterior distributions from the case study data. Our estimated detection probability (95% credible interval (0.52, 0.75)) agrees with the estimates from previous studies (Williams et al., 2017b; Lu et al., 2019). Posterior realizations of population size were obtained as derived quantities by (3.7). By counting unique labels in posterior samples of latent identities, our model resulted in a posterior mean of 125 unique individuals among the 151 observations. Posterior samples of image footprints $\mathcal{F}_t$ obtained as derived quantities using (3.3) and posterior samples of true locations $\boldsymbol{u}_{i,t}$ obtained as derived quantities using (3.1) are illustrated for a subset of images ($t = 8, 9$) in Figure 3.3. Using our model, we estimated counterclockwise rotation as the distortion process for both images, thereby linking the two pairs of observations in the overlapping region. The posterior mean linkage probability was 0.98 for observation pairs (a, h) and (b, i), and all other observation pairs have less than 0.02 posterior mean linkage probabilities. Figure 3.4 demonstrates the point-wise 95% credible intervals overlaid with the posterior means for scaling and rotation, respectively.

Table 3.2: Marginal posterior means and 95% credible intervals for the case study.

Parameter	Posterior mean (95% CI)
p_0	0.64 (0.52, 0.75)
ψ	0.19 (0.15, 0.24)
N	566 (453, 704)
σ_u^2	0.45 (0.34, 0.59)

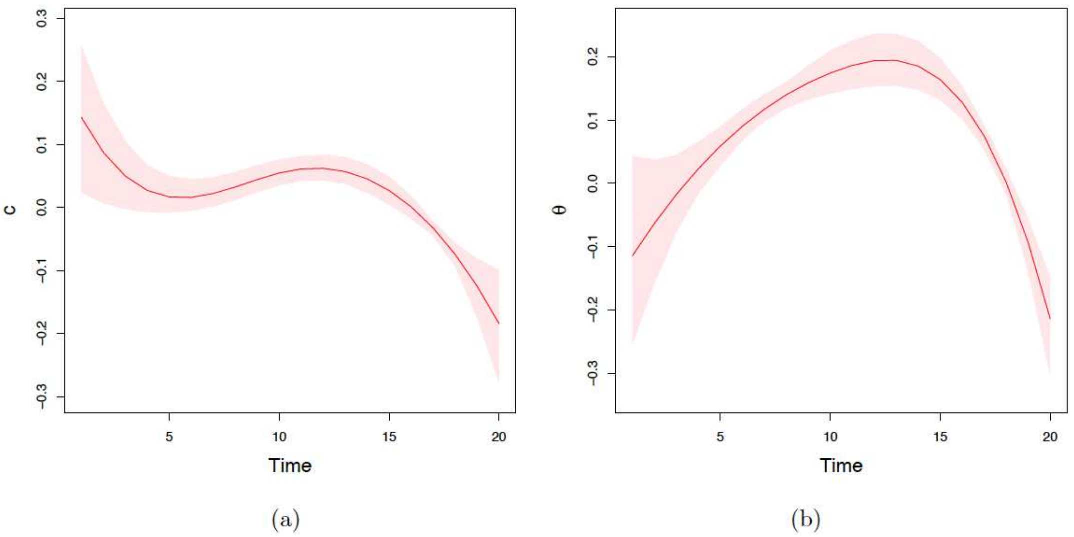


Figure 3.4: (a) Estimated posterior mean and point-wise 95% credible interval for the scaling parameter $\boldsymbol{c}$; (b) estimated posterior mean and point-wise 95% credible interval for the rotation parameter $\boldsymbol{\theta}$.

3.4 Discussion

We presented a novel method to perform entity resolution and population size estimation using individual locations obtained from aerial imagery data. We coupled record linkage and capture-recapture models to accommodate important features of aerial imagery data. Our unified framework allows information exchange and uncertainty propagation between the estimation of linkage structure and abundance, and our model is adequate for both inferential tasks.

Record linkage models are often sensitive to parameters that control linkage probability. In a sensitivity analysis for the graphical record linkage model, Steorts (2015) showed that linkage inference is only reliable when a very precise prior is used on the parameter for distortion probability. In the Bayesian alignment model, Green and Mardia (2005) advised that informative priors be used for parameters that dictate matching tendency. In our model, linkage of observed locations is motivated by their proximity in Euclidean distance to latent activity centers. Therefore, as expected, our model is sensitive to σ_u^2, the parameter controlling movement. Reliable inference requires that animal movement between consecutive detections be small relative to distortion, otherwise the model would struggle to identify unique individuals using locations only. Fortunately, much is known about movement characteristics of various species and this information can be used to specify an informative prior for σ_u^2. During aerial surveys in Glacier Bay, Alaska, the time lapse between consecutive images is so brief (1 second) that sea otter movement is significantly limited by their physical capability, thus we specified the prior for σ_u^2 such that movement distance between consecutive images was less than a meter (Williams, 1989). We provide a sensitivity analysis of prior distributions on σ_u^2 in Section B.2.

Although our method is designed to link observed individuals and estimate population size simultaneously, it can be useful even when the objective is only one of the two. The output of a record linkage model provides insight about the number of unique individuals observed at least once, and abundance estimation requires only the additional subset of population that is not observed. Population models can be used to provide prior information about the total number of latent individuals in a graphical record linkage model (Tancredi et al., 2018), a parameter that has also proven to be influential for inference (Steorts, 2015). Detection mechanisms can guide learning about the number of times an individual's record is observed. Our model assumptions can be generalized to account for more complicated monitoring situations. For example, hypergeometric models may be used in place of binomial models in capture-recapture studies when individual detections are correlated due to sampling without replacement from a finite population (Darroch, 1958; Link et al., 2009; Tancredi and Liseo, 2011). We may also model heterogeneity in p_0 to account for factors such as animal diving in response to aircraft disturbance and survey conditions that affect the backdrop (e.g., kelp, sun angle, sea state).

The use of observed locations in our model helped us better understand spatial heterogeneity in population intensity. Under the uniform prior on $\boldsymbol{s}_m$, the variation in population intensity is implicitly reflected through the estimated activity centers. A natural extension to our method is to model the spatial distribution of activity centers explicitly (Efford, 2004, 2011; Brost et al., 2017, 2020). We could account for heterogeneity in the distribution of activity centers using a species distribution model (SDM; e.g., Hefley and Hooten, 2016). An SDM is often specified as a spatial point process model, which, in our case, could take the form

$$[\boldsymbol{s}_m|\boldsymbol{x}(\boldsymbol{s}_m)] = \frac{\exp\left(\boldsymbol{x}(\boldsymbol{s}_m)'\boldsymbol{\beta}\right)}{\int_{\mathcal{D}} \exp\left(\boldsymbol{x}(\boldsymbol{s})'\boldsymbol{\beta}\right) d\boldsymbol{s}},$$

for $m = 1, \ldots, M$, where $\boldsymbol{x}(\boldsymbol{s}_m)$ denotes the vector of spatial covariates at $\boldsymbol{s}_m$ and $\boldsymbol{\beta}$ denotes the associated coefficients. Alternatively, we could attribute heterogeneity in the distribution of activity centers to the interaction among individuals which could be modeled mechanistically (e.g., Scharf et al., 2016).

Although our model is designed for aerial imagery data from sea otter population surveys in Glacier Bay, Alaska, our framework can be adapted for a variety of applications that involve aligning unlabeled point patterns with consistent measurement error, such as reconstruction of a 3-dimensional object from 2-dimensional views (Ourselin et al., 2001; Rezende et al., 2016) and reconstruction of a movement trajectory using multiple snapshots (Ando, 1991; Du et al., 2016).

Chapter 4

Multivariate Spatio-Temporal Models for Landscape Change using Aerial Imagery

4.1 Introduction

Climate change can impact ecosystems by altering vegetation composition. A prominent example due to global warming is the expansion of shrubs, coined “shrubification", in northern Alaska (Tape et al., 2006; Swanson, 2013; Brodie et al., 2019). The encroachment of woody plants could reduce erosion along rivers (Tape et al., 2010), increase fire frequency by providing fuels (Higuera et al., 2008), and further warming by decreasing albedo (Chapin III et al., 2005). Remotely sensed imagery provides readily available high-resolution information about land cover, thereby aiding in the understanding of landscape changes (Svenningsen et al., 2015; Pastick et al., 2019). We developed a statistical model to provide inference about Alaskan land cover transformation by comparing ecotypes between historic and contemporary aerial images. Our model is multivariate and dynamic to account for spatio-temporal dependence in landscape transitions.

Markov chains are commonly used to model a sequence of events by characterizing the probabilities associated with state changes (Holsclaw et al., 2017; Schafer et al., 2020). When the state space is finite (or countable) and the sequence temporally indexed, the stochastic process is known as a discrete-time Markov chain (DTMC) satisfying the Markov property, $\mathbb{P}\left(X_{t+1}|X_t, \ldots, X_1\right) = \mathbb{P}\left(X_{t+1}|X_t\right)$, where $X_1, \ldots, X_{t+1}$ belong to the state space, $\mathcal{X}$. Our study fits the profile of such state-space modeling: the states are characterized by ecotypes, and changes occur in discrete time because of distinct growth seasons in Alaska. However, in the presence of temporal irregularity, modeling transition probabilities with a DTMC would require imputation of “missing" states at a constant temporal resolution. Without regularization on the transition mechanism or model averaging techniques such as multiple imputation (Scharf et al., 2017, 2019) there is no guarantee of consistency for the imputed states. Our objective is to learn about transition probabilities at 30-year intervals (i.e., climate scale) from imagery data collected 25-32 years apart. A naive DTMC would be annually indexed to account for sampling discrepancy and is likely to produce a sequence of unrealistic ecotypes when plant communities undergo ecological succession at climate scales that are orders of magnitude greater.

We model state transitions under a hierarchical framework, where we account for spatio-temporal dynamics using a latent trajectory model in the logit-transformed probability space. We eliminate the need to impute missing states by adding depth to the model structure, and redirect the dependence between states to dependence between state probabilities represented by latent locations. Our approach can be related to spatial process models for non-Gaussian data through generalized linear modeling (Diggle et al., 1998; Finley et al., 2009), and we induce temporal dynamics by letting state probabilities vary over time. Dependence in state probabilities is often explained by exogenous input variables (Holsclaw et al., 2017) or varying parameter models (Tipton et al., 2019); however, we characterize spatio-temporal dependence through movement trajectories in the latent space. Different trajectory models reflect various growth mechanisms that we are able to accommodate. For example, vegetation succession is represented by an autoregressive model through a drift component toward regions with high probabilities in the dominant state (community). We illustrate this concept in Section 4.2.2.

Our model reduces computational complexity compared to a DTMC. For a state space $\mathcal{X}$ with size K, a conventional DTMC characterizes transitions between all pairs of states; therefore the model complexity is at minimum $\mathcal{O}\left(K^2\right)$. Our model characterizes displacements in the $(K-1)$-dimensional latent space; therefore the complexity is at minimum $\mathcal{O}(K)$. The dimension-reduction facilitates implementation and is advantageous with large K. In addition, our model achieves sampling efficiency through a Pólya-Gamma data augmentation strategy (Polson et al., 2013). Similar to the Albert-Chib data augmentation for probit regression (Albert and Chib, 1993), sampling auxiliary Pólya-Gamma random variables in a logistic regression model promotes conjugacy of linear predictors. We extend the Pólya-Gamma approach to multinomial logistic regression by representing the multinomial distribution as a product of conditional binomial distributions (Tipton et al., 2019). We introduce multi-scale spatial correlation and develop a Gibbs sampling algorithm that we apply in the case study.

4.2 Model

4.2.1 Data Model

We let $\boldsymbol{y}_{i,s,t}$ be a K-dimensional vector with all zeros and a single one denoting the observed state (ecotype) of plot i, subplot s, at time t. For $i = 1, \ldots, n_I$ (number of plots), $s = 1, \ldots, n_S$ (number of subplots per plot), and $t \in \mathbb{N}$ (years), we model $\boldsymbol{y}_{i,s,t}$ as

$$\boldsymbol{y}_{i,s,t} \sim \text{MN}\left(1, \boldsymbol{p}_{i,s,t}\right), \tag{4.1}$$

$$\boldsymbol{p}_{i,s,t} = \pi_{\text{SB}}^{-1}(\tilde{\boldsymbol{p}}_{i,s,t}), \tag{4.2}$$

where $\pi_{SB}(\cdot) : [0,1]^K \to [0,1]^{K-1}$ is a bijective mapping with

$$\begin{aligned} \tilde{p}_{i,s,t,1} &= p_{i,s,t,1}, \\ \tilde{p}_{i,s,t,k} &= \frac{p_{i,s,t,k}}{1 - \sum_{r<k} p_{i,s,t,r}}, \; k = 2, \ldots, K-1. \end{aligned}$$

The mapping is known as a stick-breaking transformation similar to that used in the construction of Dirichlet distributions (Ishwaran and James, 2001). The kth element of $\tilde{\boldsymbol{p}}_{i,s,t}$ represents the conditional probability that $\boldsymbol{y}_{i,s,t}$ is in state k given it is not in any of the states $1, \ldots, k-1$. This allows us to express the probability mass function of $\boldsymbol{y}_{i,s,t}$ as a product of conditional binomials,

$$[\boldsymbol{y}_{i,s,t} | \tilde{\boldsymbol{p}}_{i,s,t}] = \prod_{k=1}^{K-1} \text{Binom}(y_{i,s,t,k}; N_{i,s,t,k}, \tilde{p}_{i,s,t,k}), \tag{4.3}$$

where $N_{i,s,t,1} = 1$ and $N_{i,s,t,k} = 1 - \sum_{r<k} y_{i,s,t,r}$ for $k = 2, \ldots, K-1$. The stick-breaking transformation exploits conjugacy in each of the binomial models through Pólya-Gamma data augmen-

tation, and we describe this strategy in Section 4.2.3. Although the stick-breaking process implies a prior stochastic ordering, it is not of practical concern when the model is data driven and the state space finite. In addition, the stick-breaking representation of multinomial logistic regression can be more economical than the alternative representation in Holmes and Held (2006) because it does not require us to evaluate a proportionality constant for every state-specific parameter.

4.2.2 Process Model

We specify a logit function that maps $\tilde{\boldsymbol{p}}_{i,s,t}$ to $\boldsymbol{\eta}_{i,s,t}$ for plot i, subplot s, at time t in the $(K-1)$-dimensional real space as logit $(\tilde{\boldsymbol{p}}_{i,s,t}) = \boldsymbol{\eta}_{i,s,t}$. We model the dynamics in $\tilde{\boldsymbol{p}}_{i,s,t}$ through a trajectory in the logit-transformed probability space ($\boldsymbol{\eta}$-space). For illustration, Figure 4.1 shows a 2D-plane where each point is associated with a 3D probability vector and colored by the most probable state. The ecosystem started at $\boldsymbol{\eta}_0 = (-0.4, 0.8)'$ (black point), which corresponds to the state probability vector $\boldsymbol{p}_0 = (0.40, 0.41, 0.19)'$, indicating that the ecosystem is likely to be in state 1 (blue) or 2 (green) and unlikely to be in state 3 (yellow). At the end of a simulated trajectory (red point), the ecosystem is at $\boldsymbol{\eta}_T = (-0.2, -2)'$, which corresponds to the state probability vector $\boldsymbol{p}_T = (0.45, 0.07, 0.48)'$, indicating that after time T the ecosystem is likely to be in state 1 or 3 and unlikely to be in state 2. An ecosystem in the $\boldsymbol{\eta}$-space will always have non-zero probabilities in all states; however, state probabilities can become highly concentrated as the ecosystem departs from the origin (e.g., at $\boldsymbol{\eta}_t = (7, 0)'$, the probability of the ecosystem being in state 1 is greater than 0.99).

Among an array of discrete-time continuous-space trajectory models, we specify an autoregressive model with drift because it is a simple model that accommodates a temporal trend. A temporal trend in the trajectory process indicates momentum associated with ecological succession. For example, post-fire colonization of forests is represented by a drift vector directing from regions with high probabilities in the barren state to regions with high probabilities in the forest

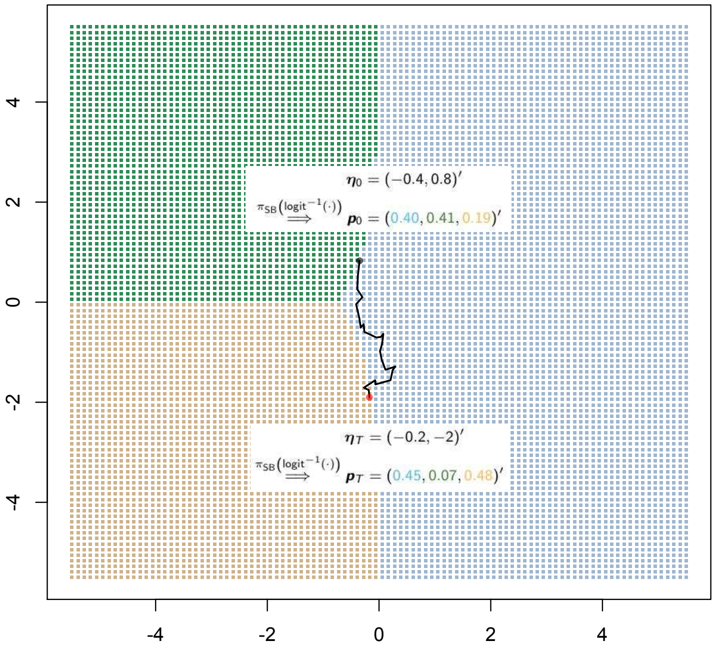


Figure 4.1: A simulated trajectory in a 2D logit-transformed probability space. Each location is colored by the most probable state. A simulated trajectory shows how the state probabilities change as an ecosystem travels across the latent space.

state, and the magnitude of drift indicates the rate of succession. We have, for $k = 1, \ldots, K-1$,

$$\eta_{i,s,t,k} = \eta_{i,s,t-1,k} + \delta_k + e_{i,s,t,k},$$

which implies the latent location of the ecosystem at time T is

$$\begin{aligned}\eta_{i,s,T,k} &= \sum_{t=1}^{T} \eta_{i,s,t,k}, \\ &= \eta_{i,s,0,k} + T\delta_k + \sum_{t=1}^{T} e_{i,s,t,k}.\end{aligned}$$

We model the initial conditions with landscape covariates as follows

$$\eta_{i,s,0,k} = \boldsymbol{h}'_{i,s}\boldsymbol{\alpha}_k + \zeta_{i,s,k}, \tag{4.4}$$

where the Gaussian random effects, $\zeta_{i,s,k}$, provide additional flexibility. We introduce spatio-temporal heterogeneity to the drift with climate covariates as follows

$$\delta_{i,s,t,k} = \boldsymbol{x}'_{i,s,t}\boldsymbol{\beta}_k. \tag{4.5}$$

We decompose $e_{i,s,t,k}$ into two-levels of spatial random effects by accounting for correlation at the plot level, $\boldsymbol{\xi}_{t,k}$, using a geostatistical model and correlation at the subplot level, $\boldsymbol{\epsilon}_{i,t,k}$, using an intrinsic conditional autoregressive (ICAR) model, so that

$$\boldsymbol{\xi}_{t,k} \sim \mathrm{N}\left(\boldsymbol{0}, \sigma^2_\xi \exp(-\boldsymbol{D}/\phi)\right), \tag{4.6}$$

$$\boldsymbol{\epsilon}_{i,t,k} \sim \mathrm{N}\left(\boldsymbol{0}, \sigma^2_\epsilon (\boldsymbol{R} - \boldsymbol{W})^{-1}\right), \tag{4.7}$$

where $\boldsymbol{D}$ defines geodesic distances (in kilometers) between plots, $\boldsymbol{W}$ is the adjacency matrix for subplots within a plot, and $\boldsymbol{R}$ is the diagonal matrix of row sums of $\boldsymbol{W}$ (Ver Hoef et al., 2018). The ecosystem trajectory model then becomes

$$\begin{aligned} \eta_{i,s,T,k} &= \boldsymbol{h}'_{i,s}\boldsymbol{\alpha}_k + \zeta_{i,s,k} + \sum_{t=1}^{T}\left\{\boldsymbol{x}'_{i,s,t}\boldsymbol{\beta}_k + \xi_{i,t,k} + \epsilon_{i,s,t,k}\right\}, \\ &= \boldsymbol{h}'_{i,s}\boldsymbol{\alpha}_k + \zeta_{i,s,k} + \underbrace{\boldsymbol{\beta}'_k \sum_{t=1}^{T} \boldsymbol{x}_{i,s,t} + \tilde{\xi}_{i,T,k} + \tilde{\epsilon}_{i,s,T,k}}_{\Delta_{i,s,T,k}}. \end{aligned} \tag{4.8}$$

We denote $\mathcal{T}_i$ as the sequence of years spanned at plot i. The covariance function of $\tilde{\boldsymbol{\xi}}_k$ accounts for spatio-temporal dependence due to overlapping time intervals as follows,

$$\mathrm{Cov}\left(\tilde{\xi}_{i,k}, \tilde{\xi}_{j,k}\right) = \begin{cases} |\mathcal{T}_i|\sigma^2_\xi, \text{ if } i = j; \\ |\mathcal{T}_i \cap \mathcal{T}_j|\sigma^2_\xi \exp(-D_{ij}/\phi), \text{ if } i \neq j. \end{cases} \tag{4.9}$$

All subplots within a plot share the same time interval, therefore the covariance function of $\tilde{\boldsymbol{\epsilon}}_{i,k}$ is that of $\boldsymbol{\epsilon}_{i,t,k}$ scaled by time as

$$\text{Cov}\left(\tilde{\boldsymbol{\epsilon}}_{i,k}\right) = |\mathcal{T}_i|\sigma_\epsilon^2(\boldsymbol{R} - \boldsymbol{W})^{-1}. \tag{4.10}$$

To maintain propriety in the ICAR model, we constrain the subplot level random effects to sum to zero by sampling $\tilde{\boldsymbol{\epsilon}}_{i,k}$ using conditioning by kriging (Rue and Held, 2005).

4.2.3 Pólya-Gamma Data Augmentation

Every binomial component of the multinomial likelihood in (4.3) can be expressed as $y_{i,s,t,k} \sim \text{Binom}(N_{i,s,t,k}, \text{logit}^{-1}(\eta_{i,s,t,k}))$ for $i = 1, \ldots, n$. By Theorem 1 in Polson et al. (2013), the following integral identity holds for the binomial likelihood $[y_{i,s,t,k}|\eta_{i,s,t,k}]$,

$$\frac{\{\exp(\eta_{i,s,t,k})\}^{y_{i,s,t,k}}}{\{1+\exp(\eta_{i,s,t,k})\}^{N_{i,s,t,k}}} = 2^{-N_{i,s,t,k}} \exp(\kappa_i \eta_i) \int_0^\infty \exp\left(-\omega\eta_i^2/2\right) p(\omega) d\omega, \tag{4.11}$$

where $\kappa_{i,s,t,k} = y_{i,s,t,k} - N_{i,s,t,k}$ and $\omega \sim \text{PG}(1,0)$. Because the right hand side of (4.11) contains a Gaussian kernel when conditioned on the Pólya-Gamma random variable ω, normal conjugacy holds for $\boldsymbol{\eta}_{i,s,t}$ under a normal prior, $\boldsymbol{\eta}_{i,s,t} \sim \text{N}\left(\boldsymbol{\mu}_{i,s,t}, \boldsymbol{\Sigma}_{i,s,t}\right)$. The posterior distribution of $\boldsymbol{\eta}_{i,s,t}$ is

$$[\boldsymbol{\eta}_{i,s,t}|\boldsymbol{y}_{i,s,t}, \boldsymbol{\omega}_{i,s,t}] = \text{N}\left(\boldsymbol{m}_{i,s,t}, \boldsymbol{V}_{i,s,t}\right),$$

where

$$\begin{aligned}
&\boldsymbol{V}_{i,s,t} = \left(\boldsymbol{\Sigma}_{i,s,t}^{-1} + \boldsymbol{\Omega}_{i,s,t}\right)^{-1}, \ \boldsymbol{m}_{i,s,t} = \boldsymbol{V}_{i,s,t}\left(\boldsymbol{\Sigma}_{i,s,t}^{-1}\boldsymbol{\mu}_{i,s,t} + \boldsymbol{\kappa}_{i,s,t}\right), \\
&\boldsymbol{\Omega}_{i,s,t} = \text{diag}(\boldsymbol{\omega}_{i,s,t}), \ \boldsymbol{\kappa}_{i,s,t} = \boldsymbol{y}_{i,s,t} - \boldsymbol{N}_{i,s,t}/2.
\end{aligned}$$

Conjugacy also holds for the Pólya-Gamma random variables. The posterior distribution of $\omega_{i,s,t,k}$ is

$$[\omega_{i,s,t,k}|\eta_{i,s,t,k}] = \text{PG}\left(N_{i,s,t,k}, \eta_{i,s,t,k}\right),$$

for $k = 1, \dots, K-1$. The Pólya-Gamma approach aligns well with our stick-breaking representation of the multinomial likelihood by facilitating conjugacy in the linear predictors and spatial random effects of the latent trajectory model. Although Johndrow et al. (2018) suggested that Metropolis-Hastings algorithms may be more efficient than data augmentation schemes including Pólya-Gamma with imbalanced categorical data, empirical examination of the trace plots in our study did not suggest significant autocorrelation. The data augmentation approach circumvents tuning, and is therefore particularly appealing to complex models such as ours. We provide a detailed description of the MCMC algorithm developed for this study in Appendix C.4.

4.2.4 Parameter Model

We specify exchangeable Gaussian priors for the covariate coefficients, $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$, for $k = 1, \dots, K-1$, and we specify Inverse-Gamma priors for the variance parameters, σ^2_ζ, σ^2_ϵ, and σ^2_ξ. The range parameter ϕ is given a uniform prior bounded above by 1/3 of the maximum distance between plots. We provide a full description of the priors used for the case study in Appendix C.1.

4.2.5 Transition Matrix

Estimated transition matrices can be used to compare our model inference to the observed transition frequencies and illustrate land cover changes under future climate scenarios. Our model allows us to infer directly on instantaneous state probabilities from latent locations in the $\boldsymbol{\eta}$-space, and we obtain elements of the transition matrix as a derived quantity. A posterior predictive realization of the transition matrix given the qth posterior sample $\boldsymbol{\eta}^{(q)}_{i,s,0}$ and $\boldsymbol{\Delta}^{(q)}_{i,s,T}$ (see (4.8) for definition), for $q = 1, \dots, Q$, is derived as follows,

1. Sample $\boldsymbol{y}^{(q)}_{i,s,0} \sim \text{MN}\left(1, \pi^{-1}_{\text{SB}}\left(\text{logit}^{-1}\left(\boldsymbol{\eta}^{(q)}_{i,s,0}\right)\right)\right)$ for $i = 1, \dots, n_I$, $s = 1, \dots, n_S$;
2. Sample $\boldsymbol{y}^{(q)}_{i,s,T} \sim \text{MN}\left(1, \pi^{-1}_{\text{SB}}\left(\text{logit}^{-1}\left(\boldsymbol{\eta}^{(q)}_{i,s,0} + \boldsymbol{\Delta}^{(q)}_{i,s,T}\right)\right)\right)$ for $i = 1, \dots, n_I$, $s = 1, \dots, n_S$;

3. Calculate the qth realization of posterior predictive transition probability from state k_0 to k_1 as
$$M_{k_0,k_1}^{(q)} = \frac{\sum_{i=1}^{n_I} \sum_{s=1}^{n_S} \mathbb{I}\left(y_{i,s,0,k_0} = 1, y_{i,s,T,k_1} = 1\right)}{\sum_{i=1}^{n_I} \sum_{s=1}^{n_S} \mathbb{I}\left(y_{i,s,0,k_0} = 1\right)},$$
where $\boldsymbol{M}^{(q)}$ is the qth posterior predictive transition matrix, $k_0 = 1, \ldots, K$, $k_1 = 1, \ldots, K$.

The posterior mean predictive transition matrix is evaluated as $\widehat{\boldsymbol{M}} = \frac{1}{Q} \sum_{q=1}^{Q} \boldsymbol{M}^{(q)}$, and elementwise credible intervals can be constructed by applying quantile functions to the Q realizations of $\boldsymbol{M}$.

We use simulation to illustrate that we are able to recover the covariate coefficients and spatial parameters in Appendix C.2.

4.3 Case Study

We analysed 209 aerial imagery pairs acquired across the NPS Arctic Inventory and Monitoring Network (ARCN) from 1977 to 2010. A pair of aerial images consists of a georeferenced high-resolution color digital aerial image taken between 2008-2010 and a scanned and georeferenced color-infrared aerial image taken between 1977-1985 of the same plot on a systematic grid over ARCN (for example, see Figure C.1; Appendix C.3). Each image was further divided into 37 hexagonal subplots during processing (Figure C.2; Appendix C.3) and their dominant ecotypes were visually determined according to the scheme developed for the ARCN Ecological Land Survey and Land Cover Map (Jorgenson et al., 2009). We omitted 9 image pairs because of incomplete ecotype or covariate information. For a preliminary analysis, we summarized the 44 ecotypes into 5 categories and classified each subplot into one of these categories:

- **Forest**: Lowland Black Spruce Forest, Riverine Poplar Forest, Riverine White Spruce-Poplar Forest, Riverine White Spruce-Willow Forest, Upland Birch Forest, Upland Spruce-Birch Forest, Upland White Spruce Forest, Upland White Spruce-Lichen Woodland;
- **Tall Shrub**: Lowland Alder Tall Shrub, Riverine Alder or Willow Tall Shrub, Upland Alder-Willow Tall Shrub;

- **Low Shrub**: Alpine Dryas Dwarf Shrub, Alpine Ericaceous Dwarf Shrub, Coastal Crowberry Dwarf Shrub, Lowland Birch-Ericaceous-Willow Low Shrub, Lowland Ericaceous Shrub Bog, Lowland Willow Low Shrub, Riverine Birch-Willow Low Shrub, Riverine Dryas Dwarf Shrub, Riverine Willow Low Shrub, Upland Birch-Ericaceous-Willow Low Shrub, Upland Dwarf Birch-Tussock Shrub, Upland Willow Low Shrub;
- **Barren**: Alpine Acidic Barrens, Alpine Acidic Barrens, Alpine Mafic Barrens, Coastal Barrens, Human Modified Barrens, Riverine Barrens, Upland Mafic Barrens, Upland Sandy Barrens;
- **Other**: Alpine Lake, Alpine Wet Sedge Meadow, Coastal Brackish Sedge–Grass Meadow, Coastal Dunegrass Meadow, Coastal Water, Lowland Lake, Lowland Sedge-Dryas Meadow, Lowland Sedge Fen, Riverine Water, Riverine Wet Sedge Meadow, Shadow/ Indeterminate, Snow, Upland Sedge-Dryas Meadow.

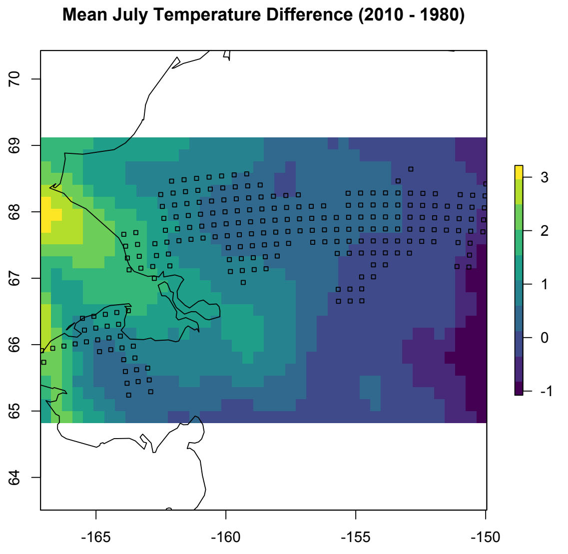
This classification accounted for 90% of the total ecotype transitions and Table 4.1 summarizes the observed transition frequencies in subplots.

Table 4.1: Summary of transition frequencies in subplots. Rows indicate categorization in c. 1980 and columns indicate categorization in c. 2010.

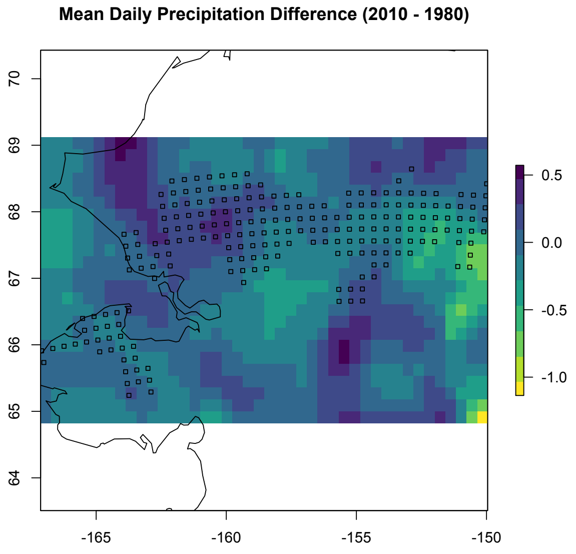
		(c. 2010)					
		Forest	Tall	Low	Barren	Other	Sum
(c. 1980)	Other	4	4	17	26	944	995
	Barren	0	0	17	1107	4	1128
	Low	117	66	4043	0	4	4230
	Tall	15	440	0	0	0	455
	Forest	587	0	0	0	5	592
	Sum	723	510	4077	1133	957	7400

The majority of subplots experienced no change over the study period. Among the 5 categories used for this analysis, the most frequent transitions were from Low Shrub to Forest, followed by Low Shrub to Tall Shrub and Other to Barren. Empirical studies suggested that most Shrub to Forest transitions occurred by post-fire succession, and most Low Shrub to Tall Shrub transitions occurred by tundra shrub increase; most transitions between Barren and Other occurred by fluvial processes in riverine environments, and a few other transitions occurred by thermokarst (Swanson, 2013).

In our application, the landscape and the climate variables only vary at the plot level due to their large scale spatial resolutions. We used a covariate vector ($\boldsymbol{h}_i$) including an intercept, mean July temperature in the first year (Celcius), aspect degree (azimuth clockwise from north), slope degree, interaction between slope and aspect, and elevation (feet) for the initial condition model in (4.4). The interaction term quantifies potential insolation in addition to what is accounted for by temperature. We mapped aspect onto a linear spectrum using Beers' transformation (Beers et al., 1966) and scaled all covariates (besides the intercept) by the standard normal distribution. We used a covariate vector ($\boldsymbol{x}_{i,t}$) including an intercept, change in mean July temperature from year $t-1$ to t (Celcius), and change in mean daily precipitation from year $t-1$ to t (mm) for the drift model in (4.5). We obtained daily temperature and precipitation from the downscaled European Centre for Medium-Range Weather Forecasts Re-Analysis (ERA)-Interim historical reanalysis data (SNAP) at 20km spatial resolution over the state of Alaska. Because these environmental data were unavailable prior to 1979, we used the 1979 data as a proxy for images collected in 1977 and 1978 (78 out of 200). We then aggregated daily measurements at each plot to obtain their mean July temperature and mean daily precipitation in the corresponding years. For illustration, Figure 4.2 shows the difference in mean July temperature and mean daily precipitation between 1980 and 2010 over the study area, although the actual values used for model fitting differed depending on the years of image collection.



(a)



(b)

Figure 4.2: Maps of north-western Alaska (black contour) and difference in (a) mean July temperature and (b) mean daily precipitation between 1980 and 2010 over the study area. Black boxes represent the plots.

We ran the MCMC algorithm for 10,000 iterations and used a burn-in of 2,000 iterations. Table 4.2 summarizes the posterior distributions of covariate coefficients and spatial parameters.

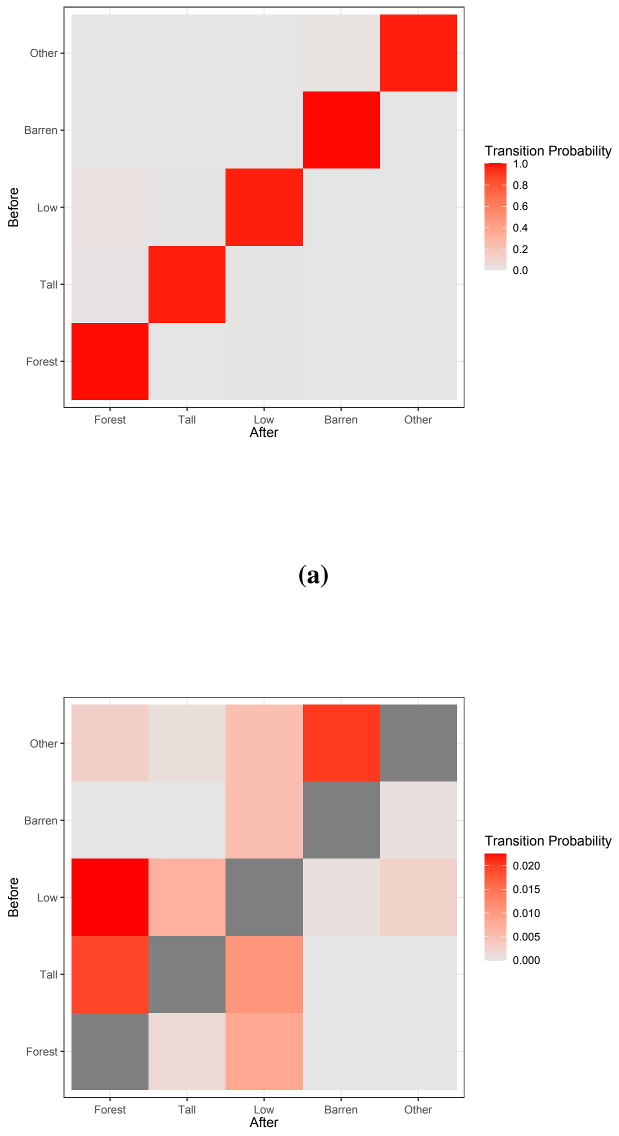
Table 4.2: Estimated posterior means (95% credible intervals) for the case study.

Parameter	Posterior	Parameter	Posterior
α_{01} (intercept)	-20.2 (-21.3, -18.9)	α_{51} (elv)	-9.5 (-10.5, -8.7)
α_{02}	-17.6 (-18.6, -16.5)	α_{52}	-1.4 (-2.2, -0.7)
α_{03}	5.6 (5.1, 6.1)	α_{53}	-1.3 (-1.8, -0.7)
α_{04}	-2.3 (-3.4, -1.2)	α_{54}	6.7 (5.4, 7.8)
α_{11} (temp)	2.6 (2.1, 3.1)	β_{01} (time)	0.05 (0.04, 0.07)
α_{12}	0.2 (-0.3, 0.7)	β_{02}	0.02 (0.00, 0.04)
α_{13}	0.9 (0.5, 1.3)	β_{03}	0.00 (-0.01, 0.01)
α_{14}	3.1 (2.1, 4.2)	β_{04}	0.02 (-0.01, 0.05)
α_{21} (asp)	-1.4 (-2.0, -0.8)	β_{11} (temp)	1.30 (1.06, 1.54)
α_{22}	-1.5 (-2.2, -0.8)	β_{12}	0.01 (-0.16, 0.19)
α_{23}	-3.3 (-3.9, -2.8)	β_{13}	-0.03 (-0.16, 0.09)
α_{24}	1.4 (0.4, 2.5)	β_{14}	-0.53 (-0.87, -0.22)
α_{31} (slope)	4.4 (3.4, 5.3)	β_{21} (pcpt)	1.18 (-0.37, 2.85)
α_{32}	-0.3 (-1.4, 0.7)	β_{22}	-1.94 (-3.34, -0.60)
α_{33}	-3.5 (-4.1, -2.7)	β_{23}	-1.69 (-2.78, -0.62)
α_{34}	11.0 (9.6, 12.5)	β_{24}	1.07 (-1.19, 3.40)
α_{41} (asp × slope)	-2.3 (-3.5, -1.1)	σ^2_ζ	137.8 (119.8, 152.3)
α_{42}	1.5 (0.4, 2.6)	σ^2_ξ	0.003 (0.003, 0.004)
α_{43}	2.1 (1.3, 2.8)	ϕ	85.7 (2.5, 239.9)
α_{44}	2.6 (0.7, 4.3)	σ^2_ϵ	0.002 (0.002, 0.003)

Each covariate coefficient is first indexed by the explanatory variable and then by the state (Forest = 1, Tall Shrub = 2, Low Shrub = 3, Barren = 4, and Other is the reference category). Due to the stick-breaking transformation, the kth element of a coefficient vector represents the change in *conditional* log odds with a unit increment in the corresponding covariate given that a subplot is not classified as any of the previous categories $1, \ldots, k-1$. The α coefficients explain the association between the landscape covariates and the distribution of ecotypes in c. 1980. Inference on the intercept vector ($\boldsymbol{\alpha}_0$) conforms with the empirical frequencies in c. 1980 (i.e., row sums in Table 4.2). The negative estimated intercepts for Forest (α_{01}) and Tall Shrub (α_{02}) indicate low initial probabilities in these states, and the positive estimated intercept for Low Shrub (α_{03}) indicate a high initial probability in the state. We estimated a positive temperature coefficient (α_{11}) and a negative elevation coefficient (α_{51}) for Forest, suggesting that forest ecotypes are likely to occupy plots at low elevations with warm growing seasons. The model fit also suggested that barren ecotypes are likely to occupy plots at high elevations (α_{54}) with steep slopes (α_{34}).

The β coefficients explain the temporal dynamics of, and the effect of climate change on, the land cover state probabilities. The time coefficients for Forest (β_{01}) and Tall Shrub (β_{02}) have positive estimated 95% credible intervals, suggesting that probability mass will accumulate in these states over time. The coefficients for Low Shrub (β_{03}) and Barren (β_{04}) both have zero-overlapping estimated 95% credible intervals; however, as the probabilities in Forest and Tall Shrub grow, probabilities in other states will likely decline. We estimated a positive temperature coefficient for Forest (β_{11}) and a negative temperature coefficient for Barren (β_{14}), suggesting that warmer climates will lead to more frequent forest ecotypes and cooler climates will lead to more frequent barren ecotypes. We estimated negative precipitation coefficients for Tall Shrub (β_{22}) and Low Shrub (β_{23}), suggesting that drier climates are conducive to more frequent shrub ecotypes.

We obtained posterior predictive realizations of transition probabilities following Section 4.2.5 for a variety of climate scenarios. Figure 4.3a shows the posterior mean predictive transition matrix for the case study, which is validated by the empirical transition frequencies in Table 4.1.



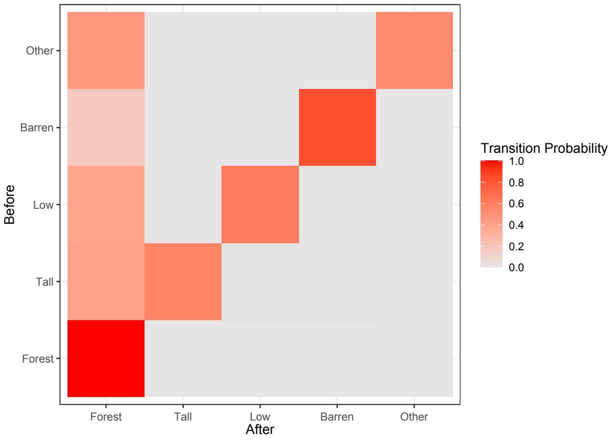
(a)

(b)

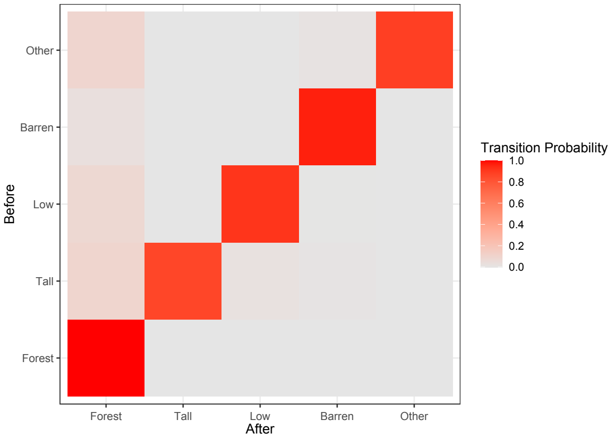
Figure 4.3: Illustrations of (a) the posterior mean predictive transition matrix and (b) the off-diagonal elements of the posterior mean predictive transition matrix for the case study. Rows indicate categorization in c. 1980 and columns indicate categorization in c. 2010.

Temperature and precipitation are highly variable across Alaska; nonetheless, in recent decades the state has experienced an overall warming more than twice as fast as the contiguous U.S., with most dramatic changes in spring and winter (Stewart et al., 2013). Studies project that the annual mean temperature will increase from 4 to 10 degrees Celsius by the end of this century under higher emission scenarios or from 2 to 6 degrees Celsius under lower emission scenarios (Stewart et al., 2017). The annual precipitation in Alaska is also projected to increase by 10% or more by mid-century (Stewart et al., 2017). We projected land cover transitions under naive scenarios for illustration. Our model inference can be used in conjunction with spatially detailed climate forecasts to obtain more realistic predictions on landscape transformation.

Figure 4.4a shows the posterior mean predictive transition matrix under a high emission scenario where we assume that July temperature increases by 8 degrees Celsius and daily precipitation increases by 2mm (equivalent to a 730mm increment in annual precipitation) uniformly in the study region from c. 1980 to 2100. As suggested by the β estimates (Table 4.2), in c. 120 years probabilities will accumulate in Forest from all other states. Most predicted transitions into Forest are from Other, Tall Shrub, and Low Shrub, possibly due to succession facilitated by warmer climates. In comparison, there are fewer transitions from Barren to Forest, possibly due to landscape factors (e.g., temperature and elevation) that limit forest expansion. Figure 4.4b shows the posterior mean predictive transition matrix under a low emission scenario where we assume that July temperature increases by 4 degrees Celsius and daily precipitation increases by 2mm uniformly in the study region from c. 1980 to 2100. There are significantly fewer transitions of all types than those under the high emission scenario. Figure 4.4b amplifies the pattern in Figure 4.3a, where the most frequent transitions are those to Forest, with some transitions from Other to Barren. There are fewer transitions from Low Shrub to Tall Shrub in Figure 4.4b than in Figure 4.3a, possibly because low shrubs undergoing succession have passed the state of Tall Shrub and reached the state of climax community (Forest) after over a century.



(a)



(b)

Figure 4.4: The posterior mean predictive transition matrices under (a) a high emission scenario (time = 120, temp = 8, pcpt = 2); and (b) a low emission scenario (time = 120, temp = 4, pcpt = 2).

4.4 Discussion

We presented a method for modeling state transitions while accounting for spatio-temporal heterogeneity. Our model characterizes state dependence in the logit-transformed probability space and leverages computational efficiency through Pólya-Gamma data augmentation. We demonstrated dependence structures that manifest evolutionary mechanisms of ecosystems. Unlike in a DTMC, our model bypasses the imputation of intermediate states when presented with temporal irregularity, and therefore it is most suitable when the true transition process takes place at a lower frequency compared to that of data collection.

We decomposed the variance in the latent trajectory process into three sources: the uncertainty in the initial conditions, σ_ζ^2, the plot level uncertainty, σ_ξ^2, and the subplot level uncertainty, σ_ϵ^2. The three parameters are identifiable because σ_ζ^2 is informed by both the historic and the contemporary images, whereas σ_ξ^2 and σ_ϵ^2 are informed by the contemporary images only and occur at different spatial scales. The estimated σ_ζ^2 is larger than the other two variance parameters (Table 4.2) because the effects of σ_ξ^2 and σ_ϵ^2 are scaled by time. However, the large initial variance due to the imbalanced transition types in our case study may limit our ability to predict transitions at unobserved locations (but not future transitions at observed locations). This challenge may be overcome with covariates that further explain heterogeneity in the initial conditions and more temporal replicates at each plot. The $\boldsymbol{\alpha}$ estimates are substantial in magnitudes. They produce initial conditions with highly concentrated probabilities, which demonstrates our learning about the temporal dynamics of the transition process. An uneven initial condition (with a significant probability in one state and negligible probabilities in the other states) indicates that the ecosystem is unlikely to experience any systematic change in the immediate future, possibly due to limiting environmental factors explained by the covariate vector $\boldsymbol{h}_i$. On the other hand, the $\boldsymbol{\beta}$ estimates representing movement along the latent trajectories are smaller, because fundamental changes in ecosystems usually take place over the course of centuries. As such, our model was able to account for small-scale temporal irregularity within the large-scale ecological process.

There are several ways to extend our model. A constant time coefficient (β_0) implies that over time, probability mass will likely converge in one state as an ecosystem moves toward regions with concentrated probabilities in that state. Such behaviors may be appropriate for some types of succession (e.g., primary/secondary successions where the climax community is stable in the “final" state), but not otherwise (e.g., seasonal successions where the dominant species alternates frequently). A more flexible process model could use higher order terms, interactions, or basis functions to represent the temporal trend. For succession mechanisms that are characterized by the coexistence or competition between species, we can explicitly model interactions between states through the covariance of (the $K-1$ dimensions of) $\boldsymbol{\eta}$. Lastly, climate events such as fire and flooding often result in substantial changes of the ecosystem. These events can be important predictors of transformation; however, only 2 out of our 200 plots had a record of fire prior to the collection of historic images, and a spatio-temporal domain larger than that in our case study is required to capture their effects. Incorporating these events into dynamic models is a potential area of future research.

Our hierarchical framework aids in integrating various data sources, some of which may be outputs from other studies on Alaskan vegetation change. For example, Raiho et al. (*In Revision*) and Scharf et al. (*In Revision*) developed a model to identify landscape factors that resist climate-driven vegetation change. Although limited by the absence of temporal replicates, Raiho et al. (*In Revision*) made inference over a greater spatial domain and performed statistical model selection to distinguish key variables. Their analysis outcome — a measurement of robustness of an ecosystem to climate change given its landscape covariates — could be used in our latent trajectory model because “robust" plots should move less in the $\boldsymbol{\eta}$-space due to climate change. On the other hand, our model is based on temporal replicates, and we could compare our inference with the results in Raiho et al. (*In Revision*) in a future study to visualize how our estimated climate effects relate to the robustness scores.

Chapter 5

Conclusion

5.1 Overview

In this dissertation, I presented novel statistical methods to characterize spatio-temporal dynamics of Alaskan ecosystems using aerial imagery data. In Chapter 2, I developed a nonlinear reaction-diffusion process model to represent the environmentally-driven diffusion and the density-regulated growth of sea otters in Glacier Bay, Alaska. Using a hierarchical framework, I synthesized multiple data sources (observer-based survey and aerial imagery data) collected at different spatial scales. I developed a homogenization strategy to statistically upscale the PDE for faster computation. My model improved population estimation and acquired inference on the local equilibrium abundances and the overall carrying capacity of sea otters in Glacier Bay, Alaska.

In Chapter 3, I developed a hierarchical model to identify unique individuals from overlapping aerial images and to estimate population abundance in the study region. My cohesive framework combined modeling techniques from entity resolution and capture-recapture fields, thereby improving inference by promoting information exchange between individual identification and population estimation.

In Chapter 4, I developed a dynamic statistical model to quantify Alaskan landscape transformation using remotely sensed imagery. I accounted for multi-level spatial dependence and temporal irregularity and improved computation by extending the Pólya-Gamma data augmentation strategy to multinomial logistic regression. My model inferred the effect of climate change on the ecosystem state probabilities and projected future transitions under various climate scenarios.

The methods I developed in this dissertation have been adopted in population surveys by conservation agencies and applied in ecological studies over larger spatial (temporal) domains. Although these methods were motivated by ecological problems, they are relevant for a wide range of applications and a diversity of topics.

5.2 Practical issues

Certain model assumptions were essential to improve the stability of algorithms and the consistency of estimators in this dissertation. For example, I demonstrated in Appendix A.2 that the record linkage model may not provide useful inference when movement uncertainty is large. I specified a precise prior for movement uncertainty (σ_u^2) because of information obtained from empirical studies on the species. However, further investigation into the sensitivity of this parameter is necessary to understand when the proposed method is suitable. A secondary analysis could consider various settings of individual movement, population density, and image overlaps to answer relevant ecological questions. Similarly, the boundary conditions of the reaction-diffusion equation in Chapter 2 were specific to the geological features of, and the conservation actions for, Glacier Bay, Alaska. These assumptions could be updated if the method is applied to other ecosystems for which the inferred population dynamics do not directly translate.

The appeal of methods in Chapters 2 – 4 is their mechanistic interpretation regarding ecological processes, rather than their superiority to existing methods. Meaningful conclusions may be derived by comparing the proposed methods to their alternatives under a statistically rigorous framework. In Chapter 3, I developed a cohesive approach to achieve entity resolution and abundance estimation. Information exchange between these model components may facilitate inference; it may also introduce additional uncertainty to either component. The advantages and disadvantages of the cohesive approach compared to a two-stage approach may be illustrated empirically by comparing their predictive performance under different simulation settings. In Chapter 4, I modeled state transitions using a latent trajectory model instead of a conventional DTMC. The original motivation was to address irregular sampling intervals; however, simulation studies can investigate whether the latent trajectory model performs comparably to the DTMC without temporal irregularity so as to determine its potential to replace DTMCs when modeling a sequence of stochastic events.

5.3 Future directions

The increasing prevalence of automated data collection in ecological studies demands the corresponding analytical tools to process more data, faster. The methods presented in this dissertation may benefit computationally from parallel algorithms in situations such as big data and streaming data in addition to the proposed statistical techniques. Among a variety of computing strategies, Recursive Bayesian (RB) methods enables parallelization by fitting multi-stage MCMC algorithms to partitions of observed data and exploits the conditional relationships of a hierarchical model to synthesize inference (Särkkä, 2013). Existing RB methods include: Prior-RB, Proposal-RB, Prior-Proposal-RB (PP-RB), and transformation assisted RB (TARB). Prior-RB manifests the recursive nature of Bayesian methods by sequentially updating parameters using posterior distributions from the previous partition to inform priors for the next partition. Prior-RB naturally appeals to inherently ordered data such as time series. On the other hand, Proposal-RB improves computation for a hierarchical model by fitting the Data model to data partitions in parallel in the first stage (Lunn et al., 2013). The transient posterior samples from the first stage are used as proposals to fit the full model in the second stage, thereby simplifying the Metropolis-Hastings (M-H) ratio. Proposal-RB does not rely on an inherent order between the partitions; however, the partitions need to be conditionally independent for the first stage parallelization. PP-RB combines concepts from Prior-RB and Proposal-RB to expedite computation even when the partitions are conditionally dependent (Hooten et al., 2021). PP-RB employs posterior samples from the first stage to calculate both the prior and the proposal in the second stage, so that the M-H ratio only involves conditional data likelihoods that can be evaluated in parallel in advance. Lastly, TARB reconciles distributional discrepancies between the first and the second stages due to parameter transformation (McCaslin et al., 2021). I illustrate the relevance of RB methods to the hierarchical models in Chapters 2 – 4 in what follows.

As a reminder, the Data and the Process models in Chapter 2 are summarized as follows,

$$
\begin{aligned}
y_{i,t} &\sim \text{Binom}\left(N_{i,t}, p_t\right), \\
N_{i,t} &\sim \text{NB}\left(\lambda_{i,t}, \tau\right), \\
\lambda_{i,t} &= f(\boldsymbol{\psi}),
\end{aligned}
$$

for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, where f represents the reaction-diffusion equation and $\boldsymbol{\psi} = (\boldsymbol{\beta}, \theta, \kappa, \gamma, K)'$. The computational crux is evaluating high-dimensional matrix operations in f over the finely resolved spatio-temporal domain. Sea otter population surveys are conducted annually in Glacier Bay, Alaska, and a PP-RB approach provides a computational advantage (in addition to homogenization) by updating model inference without fitting the historic data every year. I outline a possible updating procedure upon the arrival of $\boldsymbol{y}_{T+1}$ based on available results from a first stage analysis of $\boldsymbol{y}_{1:T}$:

1. Obtain transient posterior samples $\boldsymbol{\psi}^{(r)}$ and $\tau^{(r)}$ for $r = 1, \ldots, R$ from the first stage.
2. For $r = 1, \ldots, R$, calculate $\tilde{\boldsymbol{\lambda}}_{T+1}^{(*)} = f\left(\boldsymbol{\psi}^{(r)}\right)$ in parallel (crux) and sample transient predictive realizations $\tilde{\boldsymbol{N}}_{T+1}^{(*)} \sim \text{NB}\left(\tilde{\boldsymbol{\lambda}}_{T+1}^{(*)}, \tau^{(r)}\right)$.
3. In the second stage, calculate the rth M-H ratio to update all parameters (i.e., $\boldsymbol{N}_{T+1}$, $\boldsymbol{\psi}$, and τ) conditional on $p_t^{(r)}$ as follows,

$$
\begin{aligned}
mh &= \frac{\left[\boldsymbol{y}_{T+1} \middle| \boldsymbol{N}_{T+1}^{(*)}, p_t^{(r)}\right] \left[\boldsymbol{N}_{T+1}^{(*)}, \boldsymbol{\psi}^{(*)}, \tau^{(*)} \middle| \boldsymbol{y}_{1:T}\right] \left[\boldsymbol{N}_{T+1}^{(r-1)}, \boldsymbol{\psi}^{(r-1)}, \tau^{(r-1)} \middle| \boldsymbol{y}_{1:T}\right]}{\left[\boldsymbol{y}_{T+1} \middle| \boldsymbol{N}_{T+1}^{(r-1)}, p_t^{(r)}\right] \left[\boldsymbol{N}_{T+1}^{(r-1)}, \boldsymbol{\psi}^{(r-1)}, \tau^{(r-1)} \middle| \boldsymbol{y}_{1:T}\right] \left[\boldsymbol{N}_{T+1}^{(*)}, \boldsymbol{\psi}^{(*)}, \tau^{(*)} \middle| \boldsymbol{y}_{1:T}\right]} \\
&= \frac{\left[\boldsymbol{y}_{T+1} \middle| \boldsymbol{N}_{T+1}^{(*)}, p_t^{(r)}\right]}{\left[\boldsymbol{y}_{T+1} \middle| \boldsymbol{N}_{T+1}^{(r-1)}, p_t^{(r)}\right]}.
\end{aligned}
$$

The posterior distribution of p_t is conjugate with a beta prior and can be directly sampled from.

The cohesive approach in Chapter 3 becomes computationally challenging when applied to the entire sequence of aerial images collected during a survey (on the order of 10^4 images). Consecutive images that contained any individual often occurred in short segments separated by longer segments with no individuals. The segments of species occurrence comprise a natural data partition that may be characterized by a random effect model. For example, suppose the observed locations, $\boldsymbol{Y}$, are partitioned into J segments, $\boldsymbol{Y}_1, \dots, \boldsymbol{Y}_J$, so that $\cup_j \boldsymbol{Y}_j = \boldsymbol{Y}$ and $\boldsymbol{Y}_i \cap \boldsymbol{Y}_j = \emptyset, \ i \neq j$. I extend the method in Chapter 3 to model heterogeneous baseline detection probabilities using a linear combination of explanatory variables (e.g., average kelp cover, wind speed, daylight, etc.), so that

$$\operatorname{logit}\left(p_{0j}\right) = \boldsymbol{x}_j' \boldsymbol{\gamma} + \epsilon_j, \ \epsilon_j \sim \mathrm{N}\left(0, \sigma_p^2\right),$$

for $j = 1, \dots, J$. A Proposal-RB approach acquires population inference on $\boldsymbol{\gamma}$ and σ_p^2 from parallelized model fits to the J-partition as follows,

1. In the first stage, specify a transient prior $p_{0j} \sim [p_{0j}]$ and fit the full model in parallel to the J-partition. Obtain transient posterior samples $p_{0j}^{(k)}$ for $k = 1, \dots, K$ and $j = 1, \dots, J$.

2. In the second stage, denote $p_{0j}^{(*)} = p_{0j}^{(k)}$ and calculate the kth M-H ratio to update p_{0j} conditional on $\boldsymbol{\gamma}^{(k)}$, $\sigma_p^{2(k)}$, and $\boldsymbol{\phi}_j^{(k)} = \left(\boldsymbol{\Lambda}_j^{(k)}, \boldsymbol{S}_j^{(k)}, \sigma_{uj}^{2(k)}, \boldsymbol{\alpha}_j^{(k)}, \boldsymbol{\beta}_j^{(k)}, \boldsymbol{Z}_j^{(k)}, \psi_j^{(k)}\right)$ as follows,

$$\begin{aligned} mh &= \frac{\left[\boldsymbol{Y}_j \middle| p_{0j}^{(*)}, \boldsymbol{\phi}_j^{(k)}\right] \left[p_{0j}^{(*)} \middle| \boldsymbol{\gamma}^{(k)}, \sigma_p^{2(k)}\right] \left[\boldsymbol{Y}_j \middle| p_{0j}^{(k-1)}, \boldsymbol{\phi}^{(k)}\right] \left[p_{0j}^{(k-1)}\right]}{\left[\boldsymbol{Y}_j \middle| p_{0j}^{(k-1)}, \boldsymbol{\phi}_j^{(k)}\right] \left[p_{0j}^{(k-1)} \middle| \boldsymbol{\gamma}^{(k)}, \sigma_p^{2(k)}\right] \left[\boldsymbol{Y}_j \middle| p_{0j}^{(*)}, \boldsymbol{\phi}^{(k)}\right] \left[p_{0j}^{(*)}\right]} \\ &= \frac{\left[p_{0j}^{(*)} \middle| \boldsymbol{\gamma}^{(k)}, \sigma_p^{2(k)}\right] \left[p_{0j}^{(k-1)}\right]}{\left[p_{0j}^{(k-1)} \middle| \boldsymbol{\gamma}^{(k)}, \sigma_p^{2(k)}\right] \left[p_{0j}^{(*)}\right]}. \end{aligned}$$

3. Update $\boldsymbol{\gamma}$ and σ_p^2 based on their respective full-conditional distributions,

$$\begin{aligned} [\boldsymbol{\gamma}|\cdot] &\propto \left(\prod_{j=1}^{J} \left[p_{0j}|\boldsymbol{\gamma}, \sigma_p^2\right]\right) [\boldsymbol{\gamma}], \\ \left[\sigma_p^2|\cdot\right] &\propto \left(\prod_{j=1}^{J} \left[p_{0j}|\boldsymbol{\gamma}, \sigma_p^2\right]\right) \left[\sigma_p^2\right]. \end{aligned}$$

Pólya-Gamma data augmentation can be used in conjunction with a normal prior on γ and an inverse-Gamma prior on σ_p^2 to facilitate sampling.

Similar to the procedure described above, Proposal-RB may benefit a random coefficient extension to the method developed in Chapter 4 when inferring global covariate effects based on multiple geographically distinct study systems. That is, partition the full dataset into J conditionally independent subsets and let $\boldsymbol{\beta}_{kj}$ vary by subset. Specify priors $\boldsymbol{\beta}_{kj} \sim \mathrm{N}\left(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta\right)$ for $k = 1, \ldots, K$ and $j = 1, \ldots, J$, and specify a hyperprior $\boldsymbol{\mu}_\beta \sim \mathrm{N}(\mathbf{0}, 10\boldsymbol{I})$. Under transient priors $[\boldsymbol{\beta}_{kj}]$, the algorithm fits the model to the partitions in parallel in the first stage and acquires posterior inference on $\boldsymbol{\mu}_\beta$ in the second stage based on updated $\boldsymbol{\beta}_{kj}$ samples.

A variety of computing strategies related to RB have become popular, including Consensus MC (Scott et al., 2016), embarrassingly parallel MCMC (Neiswanger et al., 2013), and Modular Bayes (Jacob et al., 2017). Bayesian hierarchical models are a powerful analytical tool because they can solve any complicated joint problem by decomposing it into simpler conditional components. Following the divide-and-conquer principle, Bayesian computing strategies can promote the development of Bayesian models in the Age of Information that aim to quantify progressively complicated dynamics of the increasingly high-dimensional data.

Bibliography

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88,** 669–679.

Ando, H. (1991). Dynamic reconstruction of 3d structure and 3d motion. In *Proceedings of the IEEE Workshop on Visual Motion*, pages 101–102.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical Modeling and Analysis for Spatial Data*. CRC Press.

Barker, R. J., Schofield, M. R., Link, W. A., and Sauer, J. R. (2018). On the reliability of N-mixture models for count data. *Biometrics* **74,** 369–377.

Beers, T. W., Dress, P. E., and Wensel, L. C. (1966). Notes and observations: Aspect transformation in site productivity research. *Journal of Forestry* **64,** 691–692.

Berliner, L. M. (1996). Hierarchical Bayesian time series models. In *Maximum Entropy and Bayesian Methods*, pages 15–22. Springer.

Betancourt, B., Zanella, G., Miller, J. W., Wallach, H., Zaidi, A., and Steorts, R. C. (2016). Flexible models for microclustering with application to entity resolution. In *Advances in Neural Information Processing Systems*, pages 1417–1425.

Bodkin, J. L. (2015). Historic and contemporary status of sea otters in the North Pacific. In *Sea Otter Conservation*, pages 43–61. Elsevier.

Bodkin, J. L. and Udevitz, M. S. (1999). An aerial survey method to estimate sea otter abundance. *Marine Mammal Survey and Assessment Methods. Balekema, Rotterdam, Netherlands* pages 13–26.

Brodie, J. F., Roland, C. A., Stehn, S. E., and Smirnova, E. (2019). Variability in the expansion of trees and shrubs in boreal Alaska. *Ecology* **100,** e02660.

Brost, B. M., Hooten, M. B., and Small, R. J. (2017). Leveraging constraints and biotelemetry data to pinpoint repetitively used spatial features. *Ecology* **98(1),** 12–20.

Brost, B. M., Hooten, M. B., and Small, R. J. (2020). Model-based clustering reveals patterns in central place use of a marine top predator. *Ecosphere* **11,** e03123.

Brown, M. and Lowe, D. G. (2015). Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision* **74,** 59–73.

Buckland, S. T., Burt, M. L., Rexstad, E. A., Mellor, M., Williams, A. E., and Woodward, R. (2012). Aerial surveys of seabirds: The advent of digital methods. *Journal of Applied Ecology* **49,** 960–967.

Cantrell, R. S. and Cosner, C. (2004). *Spatial Ecology via Reaction-Diffusion Equations*. John Wiley & Sons, Chichester, UK.

Caughley, G. (1974). Bias in aerial survey. *The Journal of Wildlife Management* pages 921–933.

Chapin, F. S., Walker, L. R., Fastie, C. L., and Sharman, L. C. (1994). Mechanisms of primary succession following deglaciation at Glacier Bay, Alaska. *Ecological Monographs* **64,** 149–175.

Chapin III, F. S., Sturm, M., Serreze, M. C., McFadden, J. P., Key, J. R., Lloyd, A. H., McGuire, A. D., Rupp, T. S., Lynch, A. H., Schimel, J. P., Beringer, J., Chapman, W. L., Epstein, H. E., Euskirchen, E. S., Hinzman, L. D., Jia, G., Ping, C. L., Tape, K. D., Thompson, C. D. C., Walker, D. A., and Welker, J. M. (2005). Role of land-surface changes in Arctic summer warming. *Science* **310,** 657–660.

Christen, P. (2011). A survey of indexing techniques for scalable record linkage and deduplication. *IEEE Transactions on Knowledge and Data Engineering* **24,** 1537–1555.

Copas, J. and Hilton, F. (1990). Record linkage: statistical models for matching computer records. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **153,** 287–312.

Cressie, N. and Wikle, C. K. (2011). *Statistics for Spatio-Temporal Data*. John Wiley & Sons, Hoboken, New Jersey, USA.

Darroch, J. N. (1958). The multiple capture census i. Estimation of a closed population. *Biometrika* **45,** 343–358.

Dennis, E. B., Morgan, B. J. T., and Ridout, M. S. (2015). Computational aspects of N-mixture models. *Biometrics* **71,** 237–246.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998). Model-based geostatistics (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **47,** 299–350.

Dryden, I. L. and Mardia, K. V. (1998). *Statistical Analysis of Shape*. Wiley.

Du, Y., Wong, Y., Liu, Y., Han, F., Gui, Y., Wang, Z., Kankanhalli, M., and Geng, W. (2016). Marker-less 3d human motion capture with monocular image sequence and height-maps. In *European Conference on Computer Vision*, pages 20–36. Springer.

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

Efford, M. G. (2011). Estimation of population density by spatially explicit capture–recapture analysis of data from area searches. *Ecology* **92,** 2202–2207.

Esslinger, G. G., Esler, D., Howlin, S., and Starcevich, L. (2015). *Monitoring population status of sea otters (Enhydra lutris) in Glacier Bay National Park and Preserve, Alaska: options and considerations*. US Department of the Interior, US Geological Survey.

Estes, J. A. (1990). Growth and equilibrium in sea otter populations. *Journal of Animal Ecology* **59,** 385–401.

Fellegi, I. P. and Sunter, A. B. (1969). A theory for record linkage. *Journal of the American Statistical Association* **64,** 1183–1210.

Finley, A. O., Banerjee, S., and McRoberts, R. E. (2009). Hierarchical spatial models for predicting tree species assemblages across large domains. *Annals of Applied Statistics* **3,** 1052–1079.

Fortini, M., Liseo, B., Nuccitelli, A., and Scanu, M. (2001). On Bayesian record linkage. *Research in Official Statistics* **4,** 185–198.

Frost, G. V. and Epstein, H. E. (2014). Tall shrub and tree expansion in Siberian tundra ecotones since the 1960s. *Global Change Biology* **20,** 1264–1277.

Garlick, M. J., Powell, J. A., Hooten, M. B., and McFarlane, L. R. (2011). Homogenization of large-scale movement models in ecology. *Bulletin of Mathematical Biology* **73,** 2088–2108.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85,** 398–409.

Gelman, A., Huang, J., and Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing* **24,** 997–1016.

Green, P. J. and Mardia, K. (2005). Bayesian alignment using hierarchical models, with application in protein bioinformatics. *Biometrika* **93(2),** 235–254.

Gross, J. W. and Heumann, B. W. (2016). A statistical examination of image stitching software packages for use with unmanned aerial systems. *Photogrammetric Engineering & Remote Sensing* **82(6),** 419–425.

Hefley, T. J. and Hooten, M. B. (2016). Hierarchical species distribution models. *Current Landscape Ecological Reports* **1,** 87–97.

Hefley, T. J., Hooten, M. B., Russell, R. E., Walsh, D. P., and Powell, J. A. (2017). When mechanism matters: Bayesian forecasting using models of ecological diffusion. *Ecology Letters* **20,** 640–650.

Higuera, P. E., Brubaker, L. B., Anderson, P. M., Brown, T. A., Kennedy, A. T., and Hu, F. S. (2008). Frequent fires in ancient shrub tundra: implications of paleorecords for arctic environmental change. *PLoS ONE* **3,** e0001744.

Hilborn, R. and Mangel, M. (1997). *The Ecological Detective: Confronting Models with Data*. Princeton University Press, Princeton, New Jersey, USA.

Hinzman, L. D., Deal, C. J., McGuire, A. D., Mernild, S. H., Polyakov, I. V., and Walsh, J. E. (2013). Trajectory of the Arctic as an integrated system. *Ecological Applications* **23,** 1837–1868.

Hobbs, N. T. and Hooten, M. B. (2015). *Bayesian Models: A Statistical Primer for Ecologists*. Princeton University Press, Princeton, New Jersey, USA.

Holmes, C. C. and Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis* **1,** 145–168.

Holmes, E. E., Lewis, M. A., Banks, J., and Veit, R. (1994). Partial differential equations in ecology: spatial interactions and population dynamics. *Ecology* **75,** 17–29.

Holmes, M. H. (2013). *Introduction to Perturbation Methods (Second Edition)*. Springer, New York, New York, USA.

Holsclaw, T., Greene, A. M., Robertson, A. W., and Smyth, P. (2017). Bayesian nonhomogeneous Markov models via Pólya-Gamma data augmentation with applications to rainfall modeling. *Annals of Applied Statistics* **11,** 393–426.

Hooten, M. B., Garlick, M. J., and Powell, J. A. (2013). Computationally efficient statistical differential equation modeling using homogenization. *Journal of Agricultural, Biological, and Environmental Statistics* **18,** 405–428.

Hooten, M. B. and Hefley, T. J. (2019). *Bringing Bayesian Models to Life*. Chapman and Hall/CRC.

Hooten, M. B. and Hobbs, N. (2015). A guide to Bayesian model selection for ecologists. *Ecological Monographs* **85,** 3–28.

Hooten, M. B., Johnson, D. S., and Brost, B. M. (2021). Making recursive Bayesian inference accessible. *The American Statistician* **75,** 185–194.

Hooten, M. B. and Wikle, C. K. (2008). A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collard-Dove. *Environmental and Ecological Statistics* **15,** 59–70.

Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* **96,** 161–173.

Jacob, P. E., Murray, L. M., Holmes, C. C., and Robert, C. P. (2017). Better together? Statistical learning in models made of modules. *arXiv preprint arXiv:1708.08719* .

Jameson, R. J. (1989). Movements, home range, and territories of male sea otters off central California. *Marine Mammal Science* **5,** 159–172.

Jameson, R. J., Kenyon, K. W., Johnson, A. M., and Wight, H. M. (1982). History and status of translocated sea otter populations in North America. *Wildlife Society Bulletin* **10(2),** 100–107.

Jaro, M. A. (1989). Advances in record-linkage methodology as applied to matching the 1985 census of Tampa, Florida. *Journal of the American Statistical Association* **84,** 414–420.

Johndrow, J. E., Smith, A., Pillai, N., and Dunson, D. B. (2018). MCMC for imbalanced categorical data. *Journal of the American Statistical Association* **114,** 1394–1403.

Jorgenson, M. T., Roth, J. E., Miller, P. F., Macander, M. J., Duffy, M. S., Wells, A. F., Frost, G. V., and Pullman, E. R. (2009). An ecological land survey and landcover map of the Arctic Network. Technical report, National Park Service.

Kenyon, K. W. (1969). *The sea otter in the eastern Pacific Ocean.* No. 68. US Bureau of Sport Fisheries and Wildlife.

Ketz, A. C., Johnson, T. L., Hooten, M. B., and Hobbs, N. T. (2019). A hierarchical Bayesian approach for handling missing classification data. *Ecology and Evolution* **9(6),** 3130–3140.

Larsen, M. D. (2004). Record linkage using finite mixture models. *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family* pages 309–318.

Larsen, M. D. and Rubin, D. B. (2001). Iterative automated record linkage using mixture models. *Journal of the American Statistical Association* **96,** 32–41.

Larson, S. E., Bodkin, J. L., and VanBlaricom, G. R. (2014). *Sea otter conservation*. Academic Press.

Lawrence, D. B. (1958). Glaciers and vegetation in south-eastern Alaska. *American Scientist* **46,** 138A–122.

Levin, A., Zomet, A., Peleg, S., and Weiss, Y. (2004). Seamless image stitching in the gradient domain. In *European Conference on Computer Vision*, pages 377–389. Springer.

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

Link, W. A., Yoshizaki, J., Bailey, L. L., and Pollock, K. H. (2009). Uncovering a latent multinomial: Analysis of mark-recapture data with misidentification. *Biometrics* **66,** 178–185.

Liseo, B. and Tancredi, A. (2011). Bayesian estimation of population size via linkage of multivariate normal data sets. *Journal of Official Statistics* **27,** 491.

Lu, X., Williams, P. J., Hooten, M. B., Powell, J. A., Womble, J. N., and Bower, M. R. (2019). Nonlinear reaction–diffusion process models improve inference for population dynamics. *Environmetrics* **31(3),** e2604.

Lu, X., Williams, P. J., Hooten, M. B., Powell, J. A., Womble, J. N., and Bower, M. R. (2020). Nonlinear reaction–diffusion process models improve inference for population dynamics. *Environmetrics* **31,** e2604.

Lunn, D., Barrett, J., Sweeting, M., and Thompson, S. (2013). Fully Bayesian hierarchical modelling in two stages, with application to meta-analysis. *Journal of the Royal Statistical Society. Series C, Applied Statistics* **62,** 551–572.

Maciel, G. A. and Lutscher, F. (2018). Movement behavior determines competitive outcome and spread rates in strongly heterogeneous landscapes. *Theoretical Ecology* **11,** 351–365.

McCaslin, H. M., Feuka, A. B., and Hooten, M. B. (2021). Hierarchical computing for hierarchical models in ecology. *Methods in Ecology and Evolution* **12,** 245–254.

McGlincy, M. H. (2004). A Bayesian record linkage methodology for multiple imputation of missing links. In *ASA Proceedings of the Joint Statistical Meetings*, pages 4001–4008. Alexandria, VA: American Statistical Association.

Neiswanger, W., Wang, C., and Xing, E. (2013). Asymptotically exact, embarrassingly parallel MCMC. *arXiv preprint arXiv:1311.4780* .

Ourselin, S., Roche, A., Subsol, G., Pennec, X., and Ayache, N. (2001). Reconstructing a 3D structure from serial histological sections. *Image and Vision Computing* **19,** 25–31.

Pastick, N. J., Jorgenson, M. T., Goetz, S. J., Jones, B. M., Wylie, B. K., Minsley, B. J., Genet, H., Knight, J. F., Swanson, D. K., and Jorgenson, J. C. (2019). Spatiotemporal remote sensing of ecosystem change and causation across Alaska. *Global Change Biology* **25,** 1171–1189.

Polson, N. G., Scott, J. G., and Windle, J. (2013). Bayesian inference for logistic models using Pólya-Gamma latent variables. *Journal of the American Statistical Association* **108,** 1339–1349.

Powell, J. A. and Zimmermann, N. E. (2004). Multiscale analysis of active seed dispersal contributes to resolving Reid's paradox. *Ecology* **85,** 490–506.

Raiho, A. M., Scharf, H. R., Roland, C. A., Swanson, D. K., Stehn, S. E., and Hooten, M. B. (*In Revision*). Searching for a refuge: a framework for identifying site factors conferring resistance to climate-driven vegetation change. *Ecological Applications* .

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Rezende, D. J., Eslami, S. A., Mohamed, S., Battaglia, P., Jaderberg, M., and Heess, N. (2016). Unsupervised learning of 3d structure from images. In *Advances in Neural Information Processing Systems*, pages 4996–5004.

Risken, H. (1989). *The Fokker-Planck Equation: Methods of Solution and Applications*. Springer, New York, New York, USA.

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

Royle, A. J. and Berliner, L. M. (1999). A hierarchical approach to multivariate spatial modeling and prediction. *Journal of Agricultural, Biological, and Environmental Statistics* **4,** 29–56.

Royle, J. A. (2009). Analysis of capture–recapture models with individual covariates using data augmentation. *Biometrics* **65,** 267–274.

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

Royle, J. A., Dorazio, R. M., and Link, W. A. (2007). Analysis of multinomial models with unknown index using data augmentation. *Journal of Computational and Graphical Statistics* **16,** 67–85.

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

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. CRC Press.

Sadinle, M. (2018). Bayesian propagation of record linkage uncertainty into population size estimation of human rights violations. *The Annals of Applied Statistics* **12(2),** 1013–1038.

Särkkä, S. (2013). *Bayesian Filtering and Smoothing*. Cambridge University Press.

Schafer, T. L. J., Wikle, C. K., VonBank, J. A., Ballard, B. M., and Weegman, M. D. (2020). A Bayesian Markov model with Pólya-Gamma sampling for estimating individual behavior transition probabilities from accelerometer classifications. *Journal of Agricultural, Biological and Environmental Statistics* **25,** 365–382.

Scharf, H. R., Hooten, M. B., Fosdick, B. K., Johnson, D. S., London, J. M., Durban, J. W., et al. (2016). Dynamic social networks based on movement. *The Annals of Applied Statistics* **10,** 2182–2202.

Scharf, H. R., Hooten, M. B., and Johnson, D. S. (2017). Imputation approaches for animal movement modeling. *Journal of Agricultural, Biological and Environmental Statistics* **22,** 335–352.

Scharf, H. R., Hooten, M. B., Wilson, R. R., Durner, G. M., and Atwood, T. C. (2019). Accounting for phenology in the analysis of animal movement. *Biometrics* **75,** 810–820.

Scharf, H. R., Raiho, A. M., Pugh, S., Roland, C. A., Swanson, D. K., Stehn, S. E., and Hooten, M. B. (*In Press*). Multivariate Bayesian clustering using covariate-informed components with application to boreal vegetation sensitivity. *Biometrics* .

Scott, S. L., Blocker, A. W., Bonassi, F. V., Chipman, H. A., George, E. I., and McCulloch, R. E. (2016). Bayes and big data: The consensus Monte Carlo algorithm. *International Journal of Management Science and Engineering Management* **11,** 78–88.

Seymour, A. C., Dale, J., Hammill, M., Halpin, P. N., and Johnston, D. W. (2017). Automated detection and enumeration of marine wildlife using unmanned aircraft systems (UAS) and thermal imagery. *Scientific Reports* **7,** 45127.

Stein, M. L. (2005). Space–time covariance functions. *Journal of the American Statistical Association* **100,** 310–321.

Steorts, R. C. (2015). Entity resolution with empirically motivated priors. *Bayesian Analysis* **10(4),** 849–875.

Steorts, R. C., Hall, R., and Fienberg, S. E. (2015). A Bayesian approach to graphical record linkage and deduplication. *Journal of the American Statistical Association* **111,** 1660–1672.

Steorts, R. C., Ventura, S. L., Sadinle, M., and Fienberg, S. E. (2014). A comparison of blocking methods for record linkage. In *International Conference on Privacy in Statistical Databases*, pages 253–268. Springer, Cham.

Stewart, B., Kunkel, K., Champion, S., Frankson, R., Stevens, L., and Wendler, G. (2017). Alaska state climate summary. Technical report, National Oceanic and Atmospheric Association.

Stewart, B., Kunkel, K., Stevens, L., Sun, L., and Walsh, J. (2013). Regional climate trends and scenarios for the us national climate assessment: Part 7. Climate of Alaska. Technical report, National Oceanic and Atmospheric Association.

Sturm, M., Holmgren, J., McFadden, J. P., Liston, G. E., Chapin III, F. S., and Racine, C. H. (2001). Snow–shrub interactions in Arctic tundra: A hypothesis with climatic implications. *Journal of Climate* **14,** 336–344.

Svenningsen, S. R., Brandt, J., Christensen, A. A., Dahl, M. C., and Dupont, H. (2015). Historical oblique aerial photographs as a powerful tool for communicating landscape changes. *Land Use Policy* **43,** 82–95.

Swanson, D. K. (2013). Three decades of landscape change in Alaska's Arctic national parks. Technical report, National Park Service.

Szeliski, R. (2006). Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision* **2,** 1–104.

Tancredi, A. and Liseo, B. (2011). A hierarchical Bayesian approach to record linkage and population size problems. *The Annals of Applied Statistics* **5(2B),** 1553–1585.

Tancredi, A., Steorts, R., Liseo, B., et al. (2018). A unified framework for de-duplication and population size estimation. *Bayesian Analysis* **15(2),** 633–682.

Tape, K., Lord, R., Marshall, H.-P., and Ruess, R. W. (2010). Snow-mediated ptarmigan browsing and shrub expansion in Arctic Alaska. *Ecoscience* **17,** 186–193.

Tape, K. D., Sturm, M., and Racine, C. (2006). The evidence for shrub expansion in northern Alaska and the Pan-Arctic. *Global Change Biology* **12,** 686–702.

Tipton, J. R., Hooten, M. B., Nolan, C., Booth, R. K., and McLachlan, J. (2019). Preicting paleoclimate from compositional data using multivariate Gaussian process inverse prediction. *Annals of Applied Statistics* **13,** 2363–2388.

Turchin, P. (1998). *Quantitative Analysis of Movement*. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, USA.

Turchin, P. (2003). *Complex Population Dynamics: A Theoretical/Empirical Synthesis*. Princeton University Press, Princeton, New Jersey, USA.

Ver Hoef, J. M. (2014). Aerial survey data. *Wiley StatsRef: Statistics Reference Online* .

Ver Hoef, J. M. and Boveng, P. L. (2007). Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data? *Ecology* **88,** 2766–2772.

Ver Hoef, J. M., Peterson, E. E., Hooten, M. B., Hanks, E. M., and Fortin, M.-J. (2018). Spatial autoregressive models for statistical inference from ecological data. *Ecological Monographs* **88,** 36–59.

Wahba, G. (1978). Improper priors, spline smoothing and the problem of guarding against model errors in regression. *Journal of the Royal Statistical Society, Series B* **40(3),** 364–372.

Wallach, H., Jensen, S., Dicker, L., and Heller, K. (2010). An alternative prior process for nonparametric Bayesian clustering. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 892–899.

Weitzman, B. P. (2013). *Effects of sea otter colonization on soft-sediment intertidal prey assemblages in Glacier Bay, Alaska*. PhD thesis, UC Santa Cruz.

Wikle, C. K. (2003). Hierarchical Bayesian models for predicting the spread of ecological processes. *Ecology* **84,** 1382–1394.

Wikle, C. K., Berliner, L. M., and Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics* **5,** 117–154.

Wikle, C. K. and Hooten, M. B. (2010). A general science-based framework for predicting the spread of ecological processes. *Test* **19,** 417–451.

Williams, P. J., Hooten, M. B., Esslinger, G. G., Womble, J. N., Bodkin, J. L., and Bower, M. R. (2019). The rise of an apex predator following deglaciation. *Diversity and Distributions* **25,** 895–908.

Williams, P. J., Hooten, M. B., Womble, J. N., and Bower, M. R. (2017a). Estimating occupancy and abundance using aerial images with imperfect detection. *Methods in Ecology and Evolution* **8,** 1679–1689.

Williams, P. J., Hooten, M. B., Womble, J. N., and Bower, M. R. (2017b). Estimating occupancy and abundance using aerial images with imperfect detection. *Methods in Ecology and Evolution* **8,** 1679–1689.

Williams, P. J., Hooten, M. B., Womble, J. N., Esslinger, G. G., Bower, M. R., and Hefley, T. J. (2017). An integrated data model to estimate spatiotemporal occupancy, abundance, and colonization dynamics. *Ecology* **98,** 328–336.

Williams, P. J., Schroeder, C., and Jackson, P. (2020). Estimating reproduction and survival of unmarked juveniles using aerial images and marked adults. *Journal of Agricultural, Biological and Environmental Statistics* **25,** 133–147.

Williams, T. M. (1989). Swimming by sea otters: adaptations for low energetic cost locomotion. *Journal of Comparative Physiology A* **164(6),** 815–824.

Winkler, W. E. (1995). Matching and record linkage. *Business Survey Methods* **1,** 355–384.

Winkler, W. E. (2006). Overview of record linkage and current research directions. In *Bureau of the Census*.

Wood, S. N., Pya, N., and Safken, B. (2016). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association* **111(516),** 1548–1563.

Wright, J. A., Baker, R. J., Schofield, M. R., Frantz, A. C., Byrom, A. E., and Gleeson, D. M. (2009). Incorporating genotype uncertainty into mark-recapture-type models for estimating abundance using DNA samples. *Biometrics* **65,** 833–840.

Yurk, B. P. and Cobbold, C. A. (2018). Homogenization techniques for population dynamics in strongly heterogeneous landscapes. *Journal of Biological Dynamics* **12,** 171–193.

Zheng, Y. and Aukema, B. H. (2010). Hierarchical dynamic modeling of outbreaks of mountain pine beetle using partial differential equations. *Environmetrics* **21,** 801–816.

Appendix A

Supplemental Material for Chapter 2

A.1 Prior distributions

$$
\begin{aligned}
[p_t] &= \text{Beta}(30, 10) \text{ for } t = 1, \dots, T, \\
[\boldsymbol{\beta}] &= \text{N}(\mathbf{0}, 10^2 \boldsymbol{I}), \\
[\theta] &= \text{TN}(100, 200^2)_0^\infty, \\
[\kappa] &= \text{TN}(10, 100^2)_0^\infty, \\
[\gamma] &= \text{Unif}(0, 0.5), \\
[K] &= \text{Unif}(0, 100), \\
[\tau] &= \text{Unif}(0, 1).
\end{aligned}
$$

A.2 Summary of posterior predictive abundances

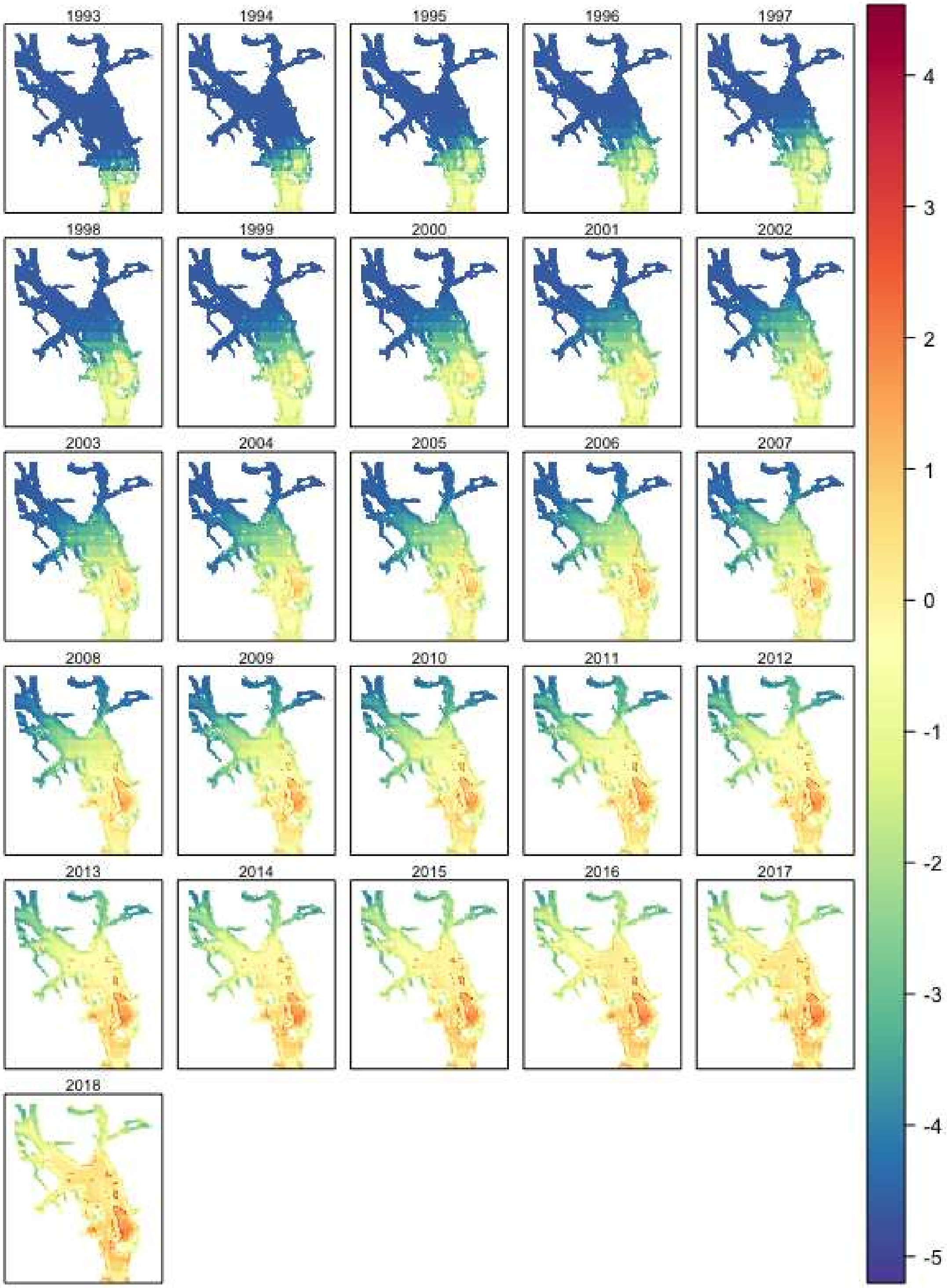


Figure A.1: Log of estimated posterior predictive mean sea otter abundances, $N_{i,t}$, in Glacier Bay, from 1993 to 2018.

A.3 Graphical illustrations of homogenization

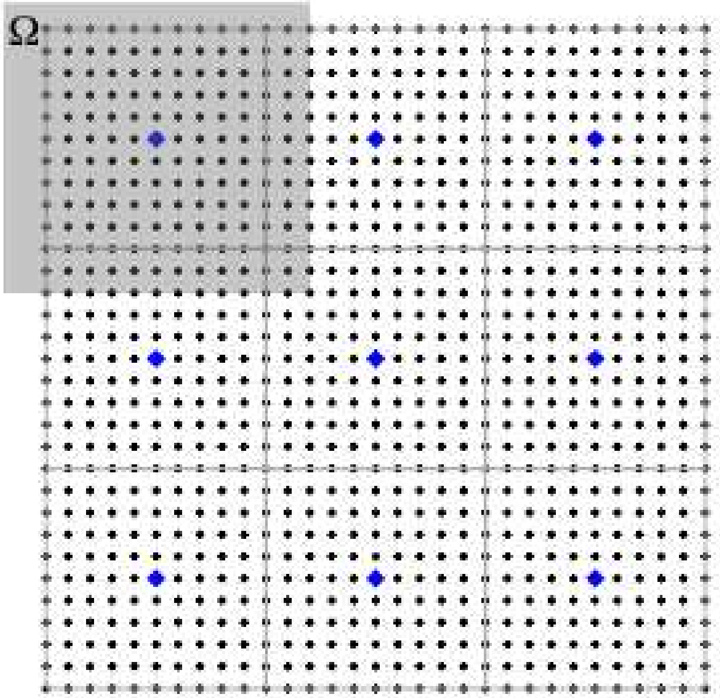
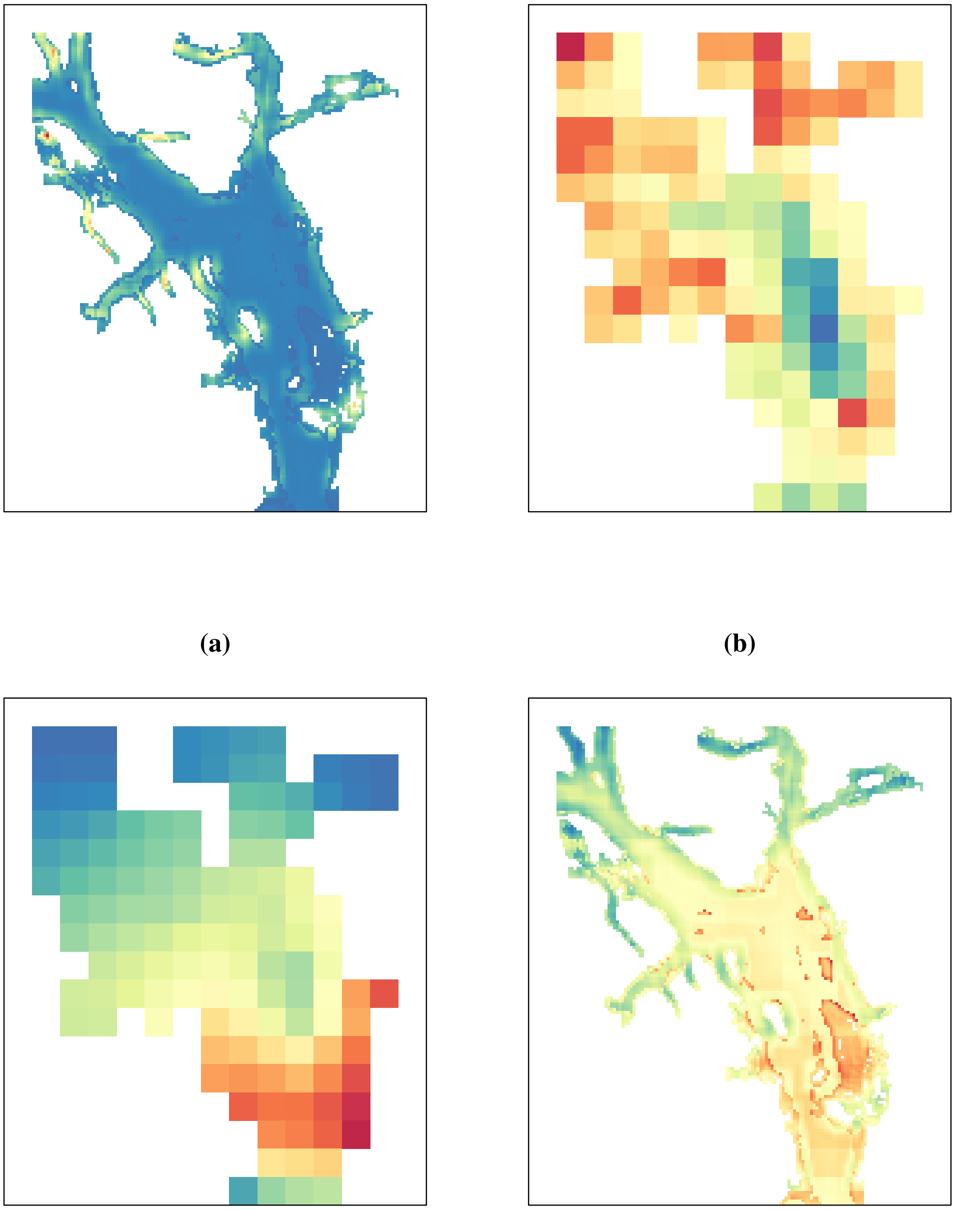


Figure A.2: Different scales in the application of homogenization. Small dots represent the centers of computational grids used at the small scale, s. Diamonds represent the centers of computational grids used at the large scale, ω. The homogenized coefficients are computed over a domain, Ω, which is larger than the large computational scale but not large enough to cross multiple large scale grids.



(a)

(b)

(c)

(d)

Figure A.3: This series of figures from left to right conceptually illustrate the homogenization surfaces at one time point, t, in one MCMC iteration. In (a) we have the small scale diffusion coefficients, $\delta(\boldsymbol{s})$, from which we obtain the homogenized diffusion coefficients, $D(\boldsymbol{\omega})$, in (b). We then solve the homogenized PDE and map the large scale solutions, $C_0(\boldsymbol{\omega}, t)$, in (c), and finally in (d) we obtain the small scale solutions, $\lambda(\boldsymbol{s}, t)$, to the original PDE by $C_0(\boldsymbol{\omega}, t)/\delta(\boldsymbol{s})$.

A.4 Markov chain Monte Carlo algorithm

1. Define initial values for $\boldsymbol{\beta}^{(0)}$, $\gamma^{(0)}$, $K^{(0)}$, $\theta^{(0)}$, $\kappa^{(0)}$, $\tau^{(0)}$, and $\boldsymbol{p}^{(0)}$.

2. Initialize population intensity surface $\boldsymbol{\lambda}^{(0)}$ and true abundance $N_{i,t}^{(0)}$ where not observed:

    (a) Calculate the diffusion surface, $\boldsymbol{\delta}^{(0)} = \exp\left(\boldsymbol{X}\boldsymbol{\beta}^{(0)}\right)$.

    (b) Calculate the homogenized diffusion coefficients, $D(\boldsymbol{\omega})$. From this point on, spatial indices will be used with subscripts to imply discretization. For each grid cell $\boldsymbol{\omega}_j$ at the large computation scale, let the homogenization region Ω_j be centered at $\boldsymbol{\omega}_j$ and consist of c^2 grid cells, $\boldsymbol{s}_i$, at the small computation scale. Let c be slightly larger than $1/\epsilon$. Then, $D(\boldsymbol{\omega}_j)^{(0)} = \frac{c^2}{\sum_{\boldsymbol{s}_i \in \Omega_j} 1/\delta(\boldsymbol{s}_i)^{(0)}}$.

    (c) Calculate the homogenized density dependence parameters,

    $$\tilde{K}(\boldsymbol{\omega}_j)^{(0)} = \frac{K^{(0)}c^2}{D(\boldsymbol{\omega}_j)^{(0)} \sum_{\boldsymbol{s}_i \in \Omega_j} 1/\delta^2(\boldsymbol{s}_i)^{(0)}}.$$

    (d) Calculate the propagator matrix $\boldsymbol{H}^{(0)}$ as described in (2.8).

    (e) Calculate $C_0(\boldsymbol{\omega}_j, 0)^{(0)} = \frac{\theta^{(0)} \exp\left(-\frac{|\boldsymbol{\omega}_j - \boldsymbol{s}_d|^2}{\kappa^{2(0)}}\right)}{\sum_{\boldsymbol{s}_i \in \mathcal{S}} \exp\left(-\frac{|\boldsymbol{s}_i - \boldsymbol{s}_d|^2}{\kappa^{2(0)}}\right)}$, and propagate the homogenized solutions, $\boldsymbol{C}_0(t)^{(0)} = \boldsymbol{H}^{(0)}\boldsymbol{C}_0(t-1)^{(0)} - \boldsymbol{C}_0^2(t-1)^{(0)}\frac{\gamma^{(0)}}{\tilde{\boldsymbol{K}}^{(0)}}$ for $t = 2, \ldots, T$.

    (f) Retrieve the original solutions, $\lambda_{i,t}^{(0)} = C_0(\boldsymbol{\omega}_j, t)/\delta(\boldsymbol{s}_i)$, where $\boldsymbol{s}_i \in \mathcal{A}(\boldsymbol{\omega}_j)$.

    (g) If true abundance is not observed at $\boldsymbol{s}_i$ in year t, sample $N_{i,t}^{(0)} \sim \text{NB}\left(\lambda_{i,t}^{(0)}, \tau^{(0)}\right)$; otherwise, fix $N_{i,t}$ to be the observed value.

3. Set $k = 1$

4. Update $\beta_r^{(k-1)}$ for r = 0, …, 4 using Metropolis-Hastings. Sample $\beta_r^{(*)} \sim \text{N}\left(\beta_r^{(k-1)}, \sigma^2_{\beta,\text{tune}}\right)$. Calculate the new propagator matrix $\boldsymbol{H}^{(*)}$ following Steps 2(a) - (d), and then calculate

$$mh_{\beta_r} = \left(\frac{\prod_{i=1}^{n_t}\prod_{t=1}^{T} \text{NB}\left(N_{i,t}^{(k-1)}; \lambda_{i,t}^{(*)}, \tau^{(k-1)}\right)}{\prod_{i=1}^{n_t}\prod_{t=1}^{T} \text{NB}\left(N_{i,t}^{(k-1)}; \lambda_{i,t}^{(k-1)}, \tau^{(k-1)}\right)}\right) \times \left(\frac{\text{N}\left(\beta_r^{(*)}; \mu_\beta, \sigma_\beta^2\right)}{\text{N}\left(\beta_r^{(k-1)}; \mu_\beta, \sigma_\beta^2\right)}\right).$$

If $mh_{\beta_r} > u$, where $u \sim \text{Unif}(0,1)$, let $\beta_r^{(k)} = \beta_r^{(*)}$ and update $\boldsymbol{\lambda}^{(k-1)} = \boldsymbol{\lambda}^{(*)}$; otherwise, let $\beta_r^{(k)} = \beta_r^{(k-1)}$.

5. Update $\gamma^{(k-1)}$ using Metropolis-Hastings. Sample $\gamma^{(*)} \sim \text{N}\left(\gamma^{(k-1)}, \sigma^2_{\gamma,\text{tune}}\right)$. If $\gamma^{(*)} \in [0, 0.5]$, calculate the new propagator matrix and solve for $\boldsymbol{\lambda}^{(*)}$ following Steps 2(d) - (f). Calculate

$$mh_\gamma = \frac{\prod_{i=1}^{n_t} \prod_{t=1}^{T} \text{NB}\left(N_{i,t}^{(k-1)}; \lambda_{i,t}^{(*)}, \tau^{(k-1)}\right)}{\prod_{i=1}^{n_t} \prod_{t=1}^{T} \text{NB}\left(N_{i,t}^{(k-1)}; \lambda_{i,t}^{(k-1)}, \tau^{(k-1)}\right)}.$$

If $mh_\gamma > u$, where $u \sim \text{Unif}(0,1)$, let $\gamma^{(k)} = \gamma^{(*)}$ and update $\boldsymbol{\lambda}^{(k-1)} = \boldsymbol{\lambda}^{(*)}$; otherwise, let $\gamma^{(k)} = \gamma^{(k-1)}$.

6. Update $K^{(k-1)}$ using Metropolis-Hastings. Sample $K^{(*)} \sim \text{N}\left(K^{(k-1)}, \sigma^2_{K,\text{tune}}\right)$. If $K^{(*)} \in [0, 100]$, calculate the new homogenized density dependence parameters and solve for $\boldsymbol{\lambda}^{(*)}$ following Steps 2(c) - (f). Calculate

$$mh_K = \frac{\prod_{i=1}^{n_t} \prod_{t=1}^{T} \text{NB}\left(N_{i,t}^{(k-1)}; \lambda_{i,t}^{(*)}, \tau^{(k-1)}\right)}{\prod_{i=1}^{n_t} \prod_{t=1}^{T} \text{NB}\left(N_{i,t}^{(k-1)}; \lambda_{i,t}^{(k-1)}, \tau^{(k-1)}\right)}.$$

If $mh_K > u$, where $u \sim \text{Unif}(0,1)$, let $K^{(k)} = K^{(*)}$ and update $\boldsymbol{\lambda}^{(k-1)} = \boldsymbol{\lambda}^{(*)}$; otherwise, let $K^{(k)} = K^{(k-1)}$.

7. Update $\theta^{(k-1)}$ using Metropolis-Hastings. Sample $\theta^{(*)} \sim \text{N}\left(\theta^{(k-1)}, \sigma^2_{\theta,\text{tune}}\right)$. If $\theta^{(*)} > 0$, calculate the new initial conditions and solve for $\boldsymbol{\lambda}^{(*)}$ following Step 2(e). Calculate

$$mh_\theta = \left(\frac{\prod_{i=1}^{n_t} \prod_{t=1}^{T} \text{NB}\left(N_{i,t}^{(k-1)}; \lambda_{i,t}^{(*)}, \tau^{(k-1)}\right)}{\prod_{i=1}^{n_t} \prod_{t=1}^{T} \text{NB}\left(N_{i,t}^{(k-1)}; \lambda_{i,t}^{(k-1)}, \tau^{(k-1)}\right)}\right) \times \left(\frac{\text{TN}\left(\theta^{(*)}; \mu_\theta, \sigma^2_\theta\right)_0^\infty}{\text{TN}\left(\theta^{(k-1)}; \mu_\theta, \sigma^2_\theta\right)_0^\infty}\right).$$

If $mh_\theta > u$, where $u \sim \text{Unif}(0,1)$, let $\theta^{(k)} = \theta^{(*)}$ and update $\boldsymbol{\lambda}^{(k-1)} = \boldsymbol{\lambda}^{(*)}$; otherwise, let $\theta^{(k)} = \theta^{(k-1)}$.

8. Update $\kappa^{(k-1)}$ using Metropolis-Hastings. Sample $\kappa^{(*)} \sim \text{N}\left(\kappa^{(k-1)}, \sigma^2_{\kappa,\text{tune}}\right)$. If $\kappa^{(*)} > 0$, calculate the new initial conditions and solve for $\boldsymbol{\lambda}^{(*)}$ following Step 2(e). Calculate

$$mh_\kappa = \left(\frac{\prod_{i=1}^{n_t}\prod_{t=1}^{T} \text{NB}\left(N_{i,t}^{(k-1)}; \lambda_{i,t}^{(*)}, \tau^{(k-1)}\right)}{\prod_{i=1}^{n_t}\prod_{t=1}^{T} \text{NB}\left(N_{i,t}^{(k-1)}; \lambda_{i,t}^{(k-1)}, \tau^{(k-1)}\right)}\right) \times \left(\frac{\text{TN}\left(\kappa^{(*)}; \mu_\kappa, \sigma^2_\kappa\right)_0^\infty}{\text{TN}\left(\kappa^{(k-1)}; \mu_\kappa, \sigma^2_\kappa\right)_0^\infty}\right).$$

   If $mh_\kappa > u$, where $u \sim \text{Unif}(0,1)$, let $\kappa^{(k)} = \kappa^{(*)}$ and update $\boldsymbol{\lambda}^{(k)} = \boldsymbol{\lambda}^{(*)}$; otherwise, let $\kappa^{(k)} = \kappa^{(k-1)}$.

9. Update $\tau^{(k-1)}$ using Metropolis-Hastings. Sample $\tau^{(*)} \sim \text{N}\left(\tau^{(k-1)}, \sigma^2_{\tau,\text{tune}}\right)$. If $\tau^{(*)} \in [0,1]$, calculate

$$mh_\tau = \frac{\prod_{i=1}^{n_t}\prod_{t=1}^{T} \text{NB}\left(N_{i,t}^{(k-1)}; \lambda_{i,t}^{(k)}, \tau^{(*)}\right)}{\prod_{i=1}^{n_t}\prod_{t=1}^{T} \text{NB}\left(N_{i,t}^{(k-1)}; \lambda_{i,t}^{(k)}, \tau^{(k-1)}\right)}.$$

   If $mh_\tau > u$, where $u \sim \text{Unif}(0,1)$, let $\tau^{(k)} = \tau^{(*)}$; otherwise, let $\tau^{(k)} = \tau^{(k-1)}$.

10. Update $p_t^{(k-1)}$ using Gibbs sampling for years when true abundances were observed:

$$p_t^{(k)} \sim \text{Beta}\left(\sum_{i \in \boldsymbol{n}_{o,t}} y_{i,t} + a_p, \sum_{i \in \boldsymbol{n}_{o,t}} (N_{i,t} - y_{i,t}) + b_p\right),$$

   where $\boldsymbol{n}_{o,t}$ is a vector of cell indices where true abundances were observed in year t. If no true abundance was observed in year t, let $p_t^{(k)} = p_t^{(k-1)}$.

11. Update $\boldsymbol{N}^{(k-1)}$:

(a) If relative abundance $y_{i,t}$ was observed, update $N_{i,t}^{(k-1)}$ using Metropolis-Hastings. Sample $N_{i,t}^{(*)} \sim \text{Pois}\left(N_{i,t}^{(k-1)} + 0.5\right)$. Calculate

$$mh_n = \left(\frac{\text{Binom}\left(y_{i,t}; N_{i,t}^{(*)}, p_t^{(k)}\right)}{\text{Binom}\left(y_{i,t}; N_{i,t}^{(k-1)}, p_t^{(k)}\right)} \right) \times \left(\frac{\text{NB}\left(N_{i,t}^{(*)}; \lambda_{i,t}^{(k)}, \tau^{(k)}\right)}{\text{NB}\left(N_{i,t}^{(k-1)}; \lambda_{i,t}^{(k)}, \tau^{(k)}\right)} \right) \\ \times \left(\frac{\text{Pois}\left(N_{i,t}^{(k-1)}; N_{i,t}^{(*)} + 0.5\right)}{\text{Pois}\left(N_{i,t}^{(*)}; N_{i,t}^{(k-1)} + 0.5\right)} \right).$$

If $mh_n > u$, where $u \sim \text{Unif}(0,1)$, let $N_{i,t}^{(k)} = N_{i,t}^{(*)}$; otherwise, let $N_{i,t}^{(k)} = N_{i,t}^{(k-1)}$.

(b) If relative abundance $y_{i,t}$ was not observed, sample $N_{i,t}^{(k)}$ from its posterior predictive distribution NB $\left(\lambda_{i,t}^{(k)}, \tau^{(k)}\right)$.

12. Save $\boldsymbol{\beta}^{(k)}$, $\gamma^{(k)}$, $K^{(k)}$, $\theta^{(k)}$, $\kappa^{(k)}$, $\tau^{(k)}$, $\boldsymbol{p}^{(k)}$, $\boldsymbol{\lambda}^{(k)}$, and $\boldsymbol{N}^{(k)}$.

13. Set $k = k + 1$ and return to Step 4. Iterate the algorithm through Steps 4 to 12 until the sample size is large enough to approximate the posterior distributions.

Appendix B

Supplemental Material for Chapter 3

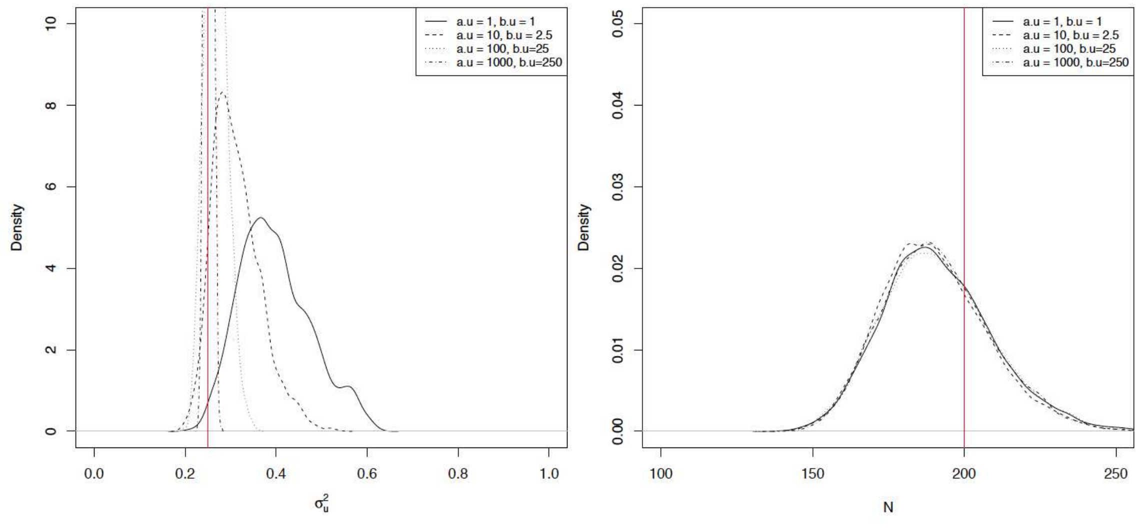
B.1 Prior distributions

$$\begin{aligned}
p_0 &\sim \text{Beta}\left(3, 1\right), \\
\psi &\sim \text{Beta}\left(0.001, 1\right), \\
\sigma_u^2 &\sim \text{IG}\left(100, 25\right), \\
\boldsymbol{s}_m &\sim \text{Unif}(\mathcal{D}),\ m = 1, \ldots, M, \\
\boldsymbol{\alpha} &\sim \text{N}\left(\boldsymbol{0}, 0.001\boldsymbol{R} + 0.01\boldsymbol{I}\right), \\
\boldsymbol{\beta} &\sim \text{N}\left(\boldsymbol{0}, 0.001\boldsymbol{R} + 0.01\boldsymbol{I}\right),
\end{aligned}$$

where $\boldsymbol{R} = \left(\boldsymbol{D}_2^-\right)' \boldsymbol{D}_2$ and $\boldsymbol{D}_2 = \begin{bmatrix} 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \end{bmatrix}$.

B.2 Sensitivity analysis

To study the sensitivity of our model to various prior specifications on σ_u^2, we used simulated data to compare posterior inference on σ_u^2 and population size under an objective prior and three subjective priors centered at the true value with increasing precision. Figure B.1 illustrates that our model tends to overestimate σ_u^2 as prior variance increases; however, our model inference on population size remained robust to bias in σ_u^2, possibly because the amount of bias is negligible compared to the amount of simulated distortion.



(a) (b)

Figure B.1: (a) Posterior densities of σ_u^2. Red line represents the truth; (b) posterior densities of derived population size. Red line represents the truth.

B.3 Supplemental figures

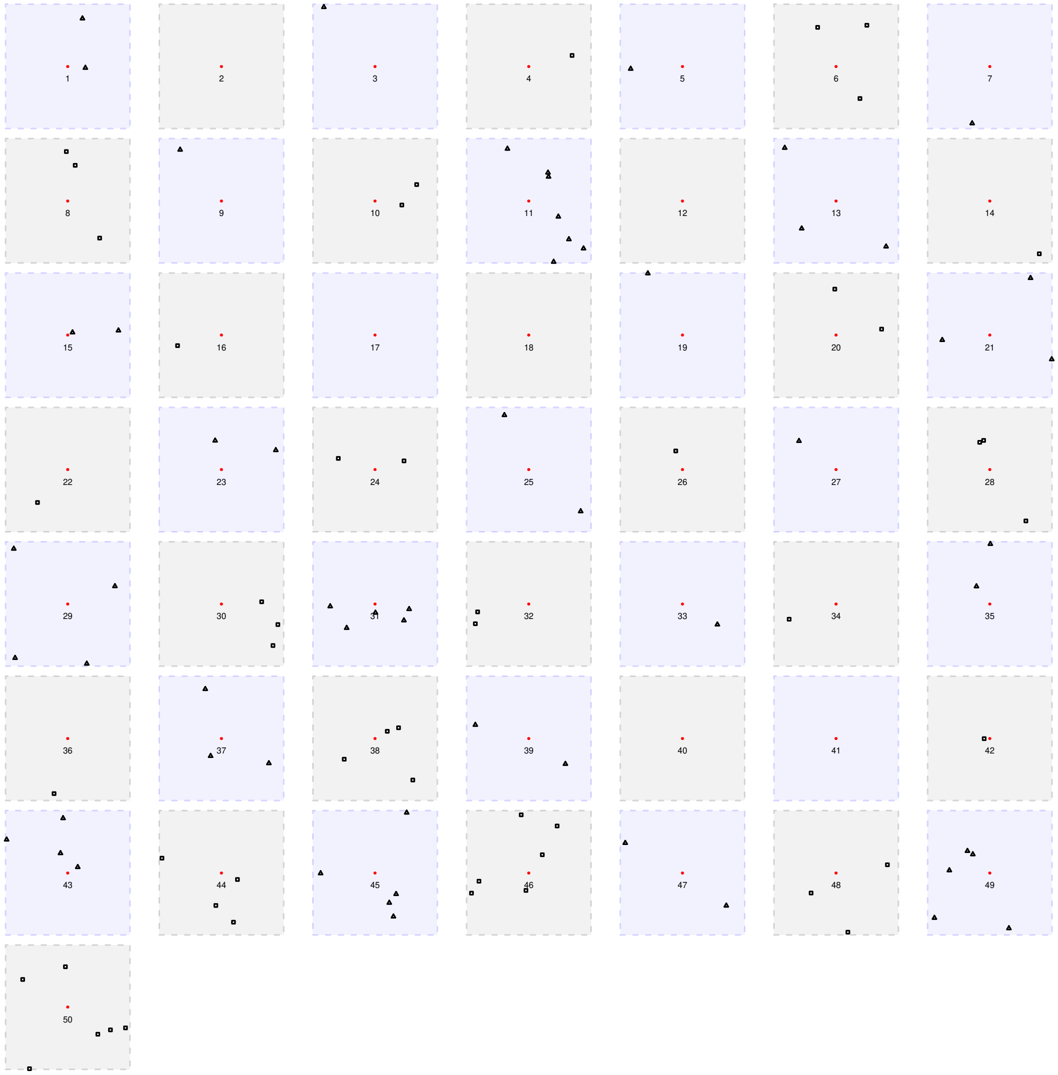


Figure B.2: The sequence of simulated observations. Footprint templates $\mathcal{Q}_t$ are shown in dashed boxes. Image centers $\boldsymbol{\mu}_t$ are represented by red points indexed by time stamp. Observed locations $\boldsymbol{y}_{i,t}$ are marked with "□" in even images and "△" in odd images.

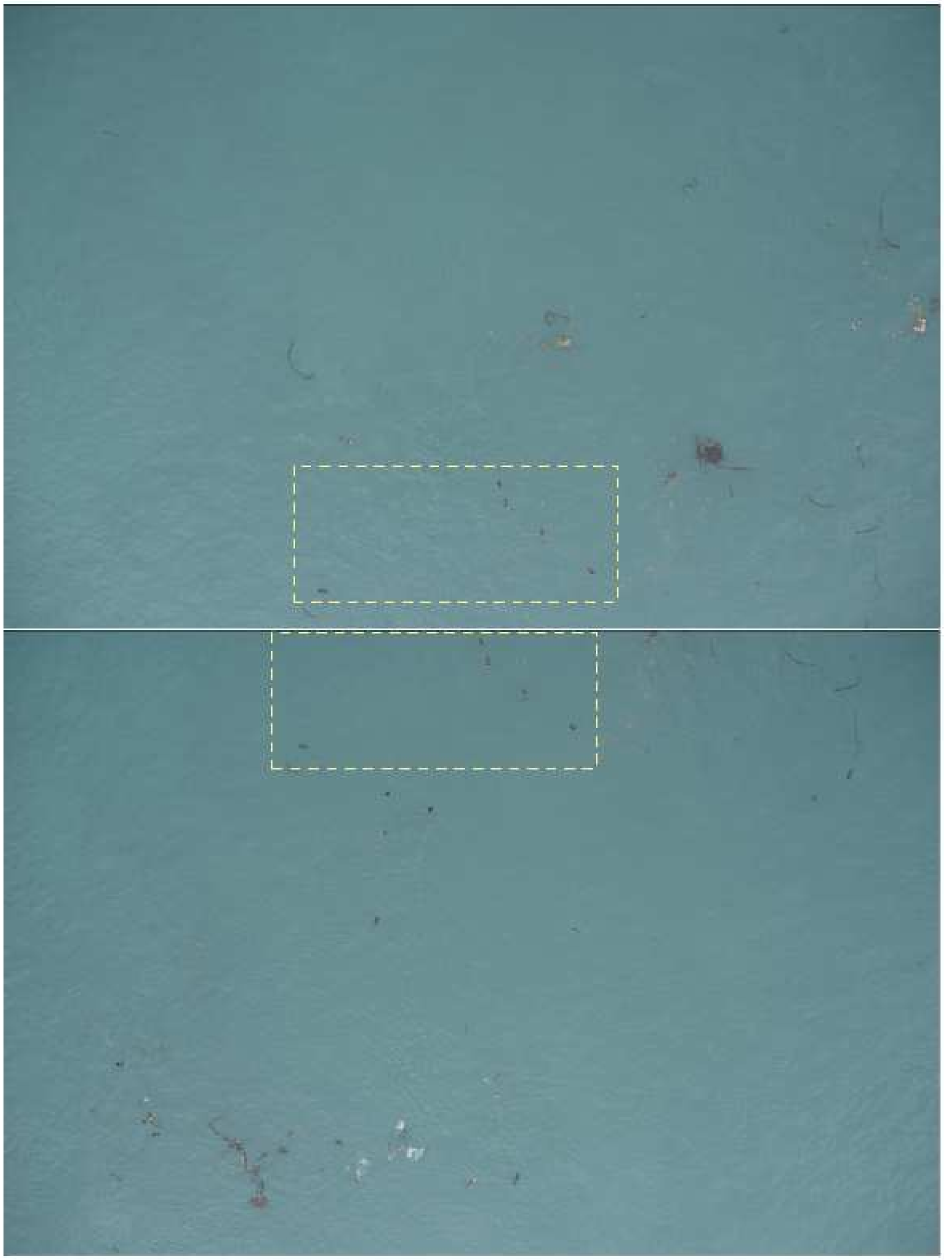


Figure B.3: Two consecutive images from the 2017 aerial survey of sea otter population in Glacier Bay, Alaska. Dashed rectangles show similar patterns of sea otter configurations—evidence of overlap between the two images.

B.4 Markov chain Monte Carlo algorithm

1. Define initial values for $\boldsymbol{\alpha}^{(0)}$ and $\boldsymbol{\beta}^{(0)}$ and calculate $\boldsymbol{c}^{(0)}$ and $\boldsymbol{\theta}^{(0)}$ by (3.2). Calculate $\left\{\mathcal{F}_t^{(0)}\right\}_t$ by (3.3). Sample $\boldsymbol{s}_m^{(0)}$ uniformly from $\mathcal{D}$ for $m = 1, \ldots, M$. Calculate $p_{m,t}^{(0)}$ by (3.5) for

$m = 1, \ldots, M$ and $t = 1, \ldots, T$. Let $z_m^{(0)} = 1$ for $m = 1, \ldots, M$. Sample $\boldsymbol{\lambda}_t^{(0)}$ from a multinomial distribution of plausible configurations at time t with probability vector given by (3.6) for $t = 1, \ldots, T$. Define initial values for $\sigma_u^{2(0)}$, $p_0^{(0)}$, and $\psi^{(0)}$.

2. Set $k = 1$.
3. For $m = 1, \ldots, M$, if $m \in \left\{\boldsymbol{\lambda}_t^{(k-1)}\right\}_{t=1}^T$, update $\boldsymbol{s}_m^{(k-1)}$ using Metropolis-Hastings. Propose $\boldsymbol{s}_m^* \sim \mathrm{N}\left(\boldsymbol{s}_m^{(k-1)}, \sigma^2_{\text{s, tune}}\boldsymbol{I}\right)$. Calculate $p^*_{m,t}$ given $\boldsymbol{s}_m^*$ by (3.5) for $t = 1, \ldots, T$. If $\prod_{t:m\in\boldsymbol{\lambda}_t} p^*_{m,t} > 0$ numerically, calculate M-H ratio as follows

$$mh_{\boldsymbol{s}_m} = \frac{\prod_{\lambda_{i,t}^{(k-1)}=m}\left[\boldsymbol{y}_{i,t}\middle|\boldsymbol{s}_m^*, \sigma_u^{2(k-1)}, \boldsymbol{\alpha}^{(k-1)}, \boldsymbol{\theta}^{(k-1)}\right] \times \left[\boldsymbol{s}_m^*\right]}{\prod_{\lambda_{i,t}^{(k-1)}=m}\left[\boldsymbol{y}_{i,t}\middle|\boldsymbol{s}_m^{(k-1)}, \sigma_u^{2(k-1)}, \boldsymbol{\alpha}^{(k-1)}, \boldsymbol{\beta}^{(k-1)}\right] \times \left[\boldsymbol{s}_m^{(k-1)}\right]} \times \frac{\prod_{t=1}^T\left\{p_0^{(k-1)}p^*_{m,t}\mathbb{I}\left(m \in \boldsymbol{\lambda}_t^{(k-1)}\right) + \left(1 - p_0^{(k-1)}p^*_{m,t}\right)\mathbb{I}\left(m \notin \boldsymbol{\lambda}_t^{(k-1)}\right)\right\}}{\prod_{t=1}^T\left\{p_0^{(k-1)}p^{(k-1)}_{m,t}\mathbb{I}\left(m \in \boldsymbol{\lambda}_t^{(k-1)}\right) + \left(1 - p_0^{(k-1)}p^{(k-1)}_{m,t}\right)\mathbb{I}\left(m \notin \boldsymbol{\lambda}_t^{(k-1)}\right)\right\}}.$$

   Set $\boldsymbol{s}_m^{(k)} = \boldsymbol{s}_m^*$ with probability $\min(1, mh_{\boldsymbol{s}_m})$; otherwise, set $\boldsymbol{s}_m^{(k)} = \boldsymbol{s}_m^{(k-1)}$. Update $p_{m,t}^{(k-1)}$ accordingly by (3.5) for $t = 1, \ldots, T$.
4. For $m = 1, \ldots, M$, if $m \in \left\{\boldsymbol{\lambda}_t^{(k-1)}\right\}_{t=1}^T$, let $z_m^{(k)} = 1$. Otherwise, update $z_m^{(k-1)}$ using Gibbs sampling,

$$z_m^{(k)} \sim \text{Bern}\left(\frac{\psi^{(k-1)}\prod_{t=1}^T\left(1 - p_0^{(k-1)}p_{m,t}^{(k)}\right)}{\psi^{(k-1)}\prod_{t=1}^T\left(1 - p_0^{(k-1)}p_{m,t}^{(k)}\right) + 1 - \psi^{(k-1)}}\right).$$

5. Update $\boldsymbol{\alpha}^{(k-1)}$ using Metropolis-Hastings. Propose $\boldsymbol{\alpha}^* \sim \mathrm{N}\left(\boldsymbol{\alpha}^{(k-1)}, \boldsymbol{\Sigma}_{\alpha,\text{tune}}\right)$. Calculate $\boldsymbol{c}^*$ by (3.2) and calculate $\{\mathcal{F}_t^*\}_{t=1}^T$ given $\boldsymbol{c}^*$ by (3.3). Calculate $p^*_{m,t}$ given $\boldsymbol{c}^*$ by (3.5) for $m = 1, \ldots, M$ and $t = 1, \ldots, T$. If $\prod_{(m,t):m\in\boldsymbol{\lambda}_t} p^*_{m,t} > 0$ numerically, calculate M-H ratio

as follows

$$mh_{\boldsymbol{\alpha}} = \frac{\prod_{t=1}^{T} \prod_{i=1}^{n_t} \left[\boldsymbol{y}_{i,t} \middle| \boldsymbol{s}^{(k)}_{\lambda_{i,t}^{(k-1)}}, \sigma_u^{2(k-1)}, \boldsymbol{\alpha}^*, \boldsymbol{\beta}^{(k-1)} \right] \times [\boldsymbol{\alpha}^*]}{\prod_{t=1}^{T} \prod_{i=1}^{n_t} \left[\boldsymbol{y}_{i,t} \middle| \boldsymbol{s}^{(k)}_{\lambda_{i,t}^{(k-1)}}, \sigma_u^{2(k-1)}, \boldsymbol{\alpha}^{(k-1)}, \boldsymbol{\beta}^{(k-1)} \right] \times [\boldsymbol{\alpha}^{(k-1)}]} \times \frac{\prod_{t=1}^{T} \prod_{m:z_m=1} \left\{ p_0^{(k-1)} p^*_{m,t} \mathbb{I}\left(m \in \boldsymbol{\lambda}_t^{(k-1)} \right) + \left(1 - p_0^{(k-1)} p^*_{m,t} \right) \mathbb{I}\left(m \notin \boldsymbol{\lambda}_t^{(k-1)} \right) \right\}}{\prod_{t=1}^{T} \prod_{m:z_m=1} \left\{ p_0^{(k-1)} p^{(k)}_{m,t} \mathbb{I}\left(m \in \boldsymbol{\lambda}_t^{(k-1)} \right) + \left(1 - p_0^{(k-1)} p^{(k)}_{m,t} \right) \mathbb{I}\left(m \notin \boldsymbol{\lambda}_t^{(k-1)} \right) \right\}}.$$

Set $\boldsymbol{\alpha}^{(k)} = \boldsymbol{\alpha}^*$ with probability $\min(1, mh_{\boldsymbol{\alpha}})$; otherwise, set $\boldsymbol{\alpha}^{(k)} = \boldsymbol{\alpha}^{(k-1)}$. Update $\boldsymbol{c}^{(k-1)}$ accordingly by (3.2). Update $p^{(k)}_{m,t}$ accordingly by (3.5) for $m = 1, \ldots, M$ and $t = 1, \ldots, T$.

6. Update $\boldsymbol{\beta}^{(k-1)}$ using Metropolis-Hastings. Propose $\boldsymbol{\beta}^* \sim \mathrm{N}\left(\boldsymbol{\beta}^{(k-1)}, \boldsymbol{\Sigma}_{\beta,\text{tune}}\right)$. Calculate $\boldsymbol{\theta}^*$ by (3.2) and calculate $\{\mathcal{F}_t^*\}_{t=1}^{T}$ given $\boldsymbol{\theta}^*$ by (3.3). Calculate $p^*_{m,t}$ given $\boldsymbol{\theta}^*$ by (3.5) for $m = 1, \ldots, M$ and $t = 1, \ldots, T$. If $\prod_{(m,t):m \in \boldsymbol{\lambda}_t} p^*_{m,t} > 0$ numerically, calculate M-H ratio as follows

$$mh_{\boldsymbol{\beta}} = \frac{\prod_{t=1}^{T} \prod_{i=1}^{n_t} \left[\boldsymbol{y}_{i,t} \middle| \boldsymbol{s}^{(k)}_{\lambda_{i,t}^{(k-1)}}, \sigma_u^{2(k-1)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^* \right] \times [\boldsymbol{\beta}^*]}{\prod_{t=1}^{T} \prod_{i=1}^{n_t} \left[\boldsymbol{y}_{i,t} \middle| \boldsymbol{s}^{(k)}_{\lambda_{i,t}^{(k-1)}}, \sigma_u^{2(k-1)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k-1)} \right] \times [\boldsymbol{\beta}^{(k-1)}]} \times \frac{\prod_{t=1}^{T} \prod_{m:z_m=1} \left\{ p_0^{(k-1)} p^*_{m,t} \mathbb{I}\left(m \in \boldsymbol{\lambda}_t^{(k-1)} \right) + \left(1 - p_0^{(k-1)} p^*_{m,t} \right) \mathbb{I}\left(m \notin \boldsymbol{\lambda}_t^{(k-1)} \right) \right\}}{\prod_{t=1}^{T} \prod_{m:z_m=1} \left\{ p_0^{(k-1)} p^{(k)}_{m,t} \mathbb{I}\left(m \in \boldsymbol{\lambda}_t^{(k-1)} \right) + \left(1 - p_0^{(k-1)} p^{(k)}_{m,t} \right) \mathbb{I}\left(m \notin \boldsymbol{\lambda}_t^{(k-1)} \right) \right\}}.$$

Set $\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^*$ with probability $\min(1, mh_{\boldsymbol{\beta}})$; otherwise, set $\boldsymbol{\beta}^{(k)} = \boldsymbol{\beta}^{(k-1)}$. Update $\boldsymbol{\theta}^{(k-1)}$ accordingly by (3.2). Update $p^{(k)}_{m,t}$ accordingly by (3.5) for $m = 1, \ldots, M$ and $t = 1, \ldots, T$.

7. Update $\psi^{(k-1)}$ by Gibbs sampling,

$$\psi^{(k)} \sim \text{Beta}\left(0.001 + \sum_{m=1}^{M} z_m^{(k)}, 1 + \sum_{m=1}^{M} \left(1 - z_m^{(k)} \right) \right).$$

8. Update $p_0^{(k-1)}$ using Metropolis-Hastings. Propose $p_0^* \sim \mathrm{N}\left(p_0^{(k-1)}, \sigma^2_{p0,\text{tune}}\right)$. If $p_0^* \in [0,1]$, calculate M-H ratios as follows

$$mh_{p_0} = \frac{\prod_{t=1}^{T} \prod_{m:z_m=1} \left\{p_0^* p_{m,t}^{(k)} \mathbb{I}\left(m \in \boldsymbol{\lambda}_t^{(k-1)}\right) + \left(1 - p_0^* p_{m,t}^{(k)}\right) \mathbb{I}\left(m \notin \boldsymbol{\lambda}_t^{(k-1)}\right)\right\}}{\prod_{t=1}^{T} \prod_{m:z_m=1} \left\{p_0^{(k-1)} p_{m,t}^{(k)} \mathbb{I}\left(m \in \boldsymbol{\lambda}_t^{(k-1)}\right) + \left(1 - p_0^{(k-1)} p_{m,t}^{(k)}\right) \mathbb{I}\left(m \notin \boldsymbol{\lambda}_t^{(k-1)}\right)\right\}} \times \frac{[p_0^*]}{\left[p_0^{(k-1)}\right]}$$

Set $p_0^{(k)} = p_0^*$ with probability $\min(1, mh_{p_0})$; otherwise, set $p_0^{(k)} = p_0^{(k-1)}$.

9. For $t = 1, \ldots, T$, update $\boldsymbol{\lambda}_t^{(k-1)}$ using Gibbs sampling. Let $\mathcal{L}_t^{(k-1)}$ be the set of plausible configurations of latent identities in image t such that any $\boldsymbol{\lambda}_t^j \in \mathcal{L}_t^{(k-1)}$ satisfies

$$\mathbb{P}\left(\boldsymbol{\lambda}_t^j \middle| \left\{z_m^{(k)}\right\}_{m=1}^{M}, \left\{\boldsymbol{s}_m^{(k)}\right\}_{m=1}^{M}, \sigma_u^{2(k-1)}, \mathcal{F}_t^{(k)}, p_0^{(k)}\right) > 0$$

numerically. Sample

$$\boldsymbol{\lambda}_t^{(k)} \sim \mathrm{MN}(1, \boldsymbol{p}\left(\mathcal{L}_t^{(k-1)}\right)),$$

where

$$p_j\left(\mathcal{L}_t^{(k-1)}\right) = \mathbb{P}\left(\boldsymbol{\lambda}_t^{(k)} = \boldsymbol{\lambda}_t^j\right) \propto \prod_{i=1}^{n_t} \left[\boldsymbol{y}_{i,t} \middle| \boldsymbol{s}^{(k)}_{\lambda^j_{i,t}}, \sigma_u^{2(k-1)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}\right] \times \prod_{m:z_m=1} \left\{p_0^{(k)} p_{m,t}^{(k)} \mathbb{I}\left(m \in \boldsymbol{\lambda}_t^j\right) + \left(1 - p_0^{(k)} p_{m,t}^{(k)}\right) \mathbb{I}\left(m \notin \boldsymbol{\lambda}_t^j\right)\right\}.$$

10. Update $\sigma_u^{2(k-1)}$ using Metropolis-Hastings. Propose $\sigma_u^{2*} \sim \mathrm{N}\left(\sigma_u^{2(k-1)}, \sigma^2_{u,\text{tune}}\right)$. Calculate $p^*_{m,t}$ given σ_u^{2*} by (3.5) for $m = 1, \ldots, M$ and $t = 1, \ldots, T$. If $\prod_{(m,t):m\in\boldsymbol{\lambda}_t} p^*_{m,t} > 0$ and $\sigma_u^{2*} > 0$, calculate M-H ratio as follows

$$mh_{\sigma_u^2} = \frac{\prod_{t=1}^{T}\prod_{i=1}^{n_t}\left[\boldsymbol{y}_{i,t}\middle|\boldsymbol{s}^{(k)}_{\lambda^{(k)}_{i,t}}\sigma_u^{2*}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}\right] \times [\sigma_u^{2*}]}{\prod_{t=1}^{T}\prod_{i=1}^{n_t}\left[\boldsymbol{y}_{i,t}\middle|\boldsymbol{s}^{(k)}_{\lambda^{(k)}_{i,t}}, \sigma_u^{2(k-1)}, \boldsymbol{\alpha}^{(k)}, \boldsymbol{\beta}^{(k)}\right] \times \left[\sigma_u^{2(k-1)}\right]} \times$$
$$\frac{\prod_{t=1}^{T}\prod_{m:z_m=1}\left\{p_0^{(k)}p^*_{m,t}\mathbb{I}\left(m \in \boldsymbol{\lambda}_t^{(k)}\right) + \left(1 - p_0^{(k)}p^*_{m,t}\right)\mathbb{I}\left(m \notin \boldsymbol{\lambda}_t^{(k)}\right)\right\}}{\prod_{t=1}^{T}\prod_{m:z_m=1}\left\{p_0^{(k)}p^{(k)}_{m,t}\mathbb{I}\left(m \in \boldsymbol{\lambda}_t^{(k)}\right) + \left(1 - p_0^{(k)}p^{(k)}_{m,t}\right)\mathbb{I}\left(m \notin \boldsymbol{\lambda}_t^{(k)}\right)\right\}}.$$

Set $\sigma_u^{2(k)} = \sigma_u^{2*}$ with probability $\min(1, mh_{\sigma_u^2})$; otherwise, set $\sigma_u^{2(k)} = \sigma_u^{2(k-1)}$. Update $p^{(k)}_{m,t}$ accordingly by (3.5) for $m = 1, \ldots, M$ and $t = 1, \ldots, T$.

11. Save $\left\{\boldsymbol{s}_m^{(k)}\right\}_{m=1}^{M}$, $\left\{z_m^{(k)}\right\}_{m=1}^{M}$, $\boldsymbol{\alpha}^{(k)}$, $\boldsymbol{\beta}^{(k)}$, $\psi^{(k)}$, $p_0^{(k)}$, $\left\{\boldsymbol{\lambda}_t^{(k)}\right\}_{t=1}^{T}$, and $\sigma_u^{2(k)}$.
12. Set $k = k + 1$ and return to Step 3. Iterate the algorithm through Steps 3 to 10 until the sample size is large enough to approximate the posterior distribution.

Appendix C

Supplemental Material for Chapter 4

C.1 Prior distributions

$$\begin{aligned}
\boldsymbol{\alpha}_k &\sim \mathrm{N}\left(\boldsymbol{\mu}_\alpha = \mathbf{0}, \boldsymbol{\Sigma}_\alpha = 10\boldsymbol{I}\right), \\
\boldsymbol{\beta}_k &\sim \mathrm{N}\left(\boldsymbol{\mu}_\beta = \mathbf{0}, \boldsymbol{\Sigma}_\beta = 10\boldsymbol{I}\right), \\
\sigma^2_\zeta &\sim \text{IG}(a_\zeta = 10, b_\zeta = 1), \\
\sigma^2_\epsilon &\sim \text{IG}(a_\epsilon = 10, b_\epsilon = 1), \\
\sigma^2_\xi &\sim \text{IG}(a_\xi = 10, b_\xi = 1), \\
\phi &\sim \text{Unif}(a_\phi = 0, b_\phi = 252).
\end{aligned}$$

C.2 Simulation

We simulated $n_I = 100$ plots, each containing a square lattice of $n_S = 36$ subplots. We let the state space, $\mathcal{X}$, be of size $K = 4$, and observed the states at each plot twice to emulate the data collection procedure in our case study. For spatial correlation, we sampled plot locations (in 2-D) uniformly from a unit square to obtain the image level covariance matrix, $\boldsymbol{D}$, in (4.9). We defined a Moore neighborhood to obtain the subplot level adjacency matrix, $\boldsymbol{W}$, in (4.10). For temporal heterogeneity, we sampled an integer uniformly between 25 and 35 as the time between the two observations at each plot. We simulated additional covariates at plot level by sampling from the standard normal distribution.

We generated initial conditions using (4.4) and simulated trajectories using (4.8). Then, we transformed $\boldsymbol{\eta}_{i,s,0}$ and $\boldsymbol{\eta}_{i,s,T}$ through a bijective mapping, $\pi_{\text{SB}}^{-1}\left(\text{logit}^{-1}(\cdot)\right)$, to obtain their corresponding probability vectors in the state space. We sampled the first observed states from multinomial distributions informed by the initial conditions as $\boldsymbol{y}_{i,s,0} \sim \text{MN}\left(1; \pi_{\text{SB}}^{-1}\left(\text{logit}^{-1}(\boldsymbol{\eta}_{i,s,0})\right)\right)$, and we sampled the second observed states from multinomial distributions informed by the simulated trajectories as $\boldsymbol{y}_{i,s,0} \sim \text{MN}\left(1; \pi_{\text{SB}}^{-1}\left(\text{logit}^{-1}(\boldsymbol{\eta}_{i,s,T})\right)\right)$.

We ran the MCMC algorithm for 10,000 iterations and used a burn-in of 2,000 iterations. Table C.1 summarizes the posterior distributions of the parameters and their truth in simulation. We obtained posterior predictive realizations of transition probabilities following Section 4.2.5. Table C.2 summarizes the empirical and the element-wise posterior predictive distributions of transition probabilities in simulation. Our model captured all illustrated parameters by their respective 95% credible intervals.

Table C.1: True parameters and their estimated posterior means (95% credible intervals) for the simulation.

Parameter	Truth	Posterior
α_{01} (intercept)	-2.5	-2.55 (-2.69, -2.43)
α_{02}	0.8	0.77 (0.69, 0.85)
α_{03}	0.2	0.26 (0.13, 0.39)
α_{11} (sim. landscape)	1	1.06 (0.97, 1.15)
α_{12}	0.5	0.49 (0.41, 0.56)
α_{13}	-0.5	-0.45 (-0.56, -0.35)
β_{01} (time)	0.04	0.06 (0.04, 0.07)
β_{02}	-0.06	-0.06 (-0.07, -0.05)
β_{03}	-0.02	-0.01 (-0.03, 0.00)
β_{11} (sim. climate)	1.5	1.54 (1.42, 1.66)
β_{12}	0.8	0.78 (0.64, 0.91)
β_{13}	0.3	0.34 (0.19, 0.49)
σ_{ζ}^2	0.1	0.11 (0.06, 0.20)
σ_{ξ}^2	0.005	0.005 (0.004, 0.005)
ϕ	0.37	0.33 (0.13, 0.49)
σ_{ϵ}^2	0.005	0.004 (0.004, 0.005)

Table C.2: Empirical transition probabilities and their estimated posterior means (95% credible intervals) for the simulation.

From	To	Empirical Probability	Posterior
1	1	0.53	0.55 (0.50, 0.60)
1	2	0.17	0.14 (0.11, 0.18)
1	3	0.11	0.12 (0.08, 0.15)
1	4	0.19	0.20 (0.16, 0.24)
2	1	0.43	0.43 (0.41, 0.45)
2	2	0.17	0.17 (0.15, 0.19)
2	3	0.17	0.18 (0.16, 0.20)
2	4	0.23	0.22 (0.20, 0.25)
3	1	0.36	0.35 (0.31, 0.39)
3	2	0.15	0.16 (0.13, 0.19)
3	3	0.27	0.25 (0.21, 0.29)
3	4	0.22	0.24 (0.20, 0.27)
4	1	0.42	0.40 (0.36, 0.45)
4	2	0.14	0.16 (0.12, 0.20)
4	3	0.19	0.20 (0.16, 0.24)
4	4	0.25	0.24 (0.20, 0.28)

C.3 Supplemental figures

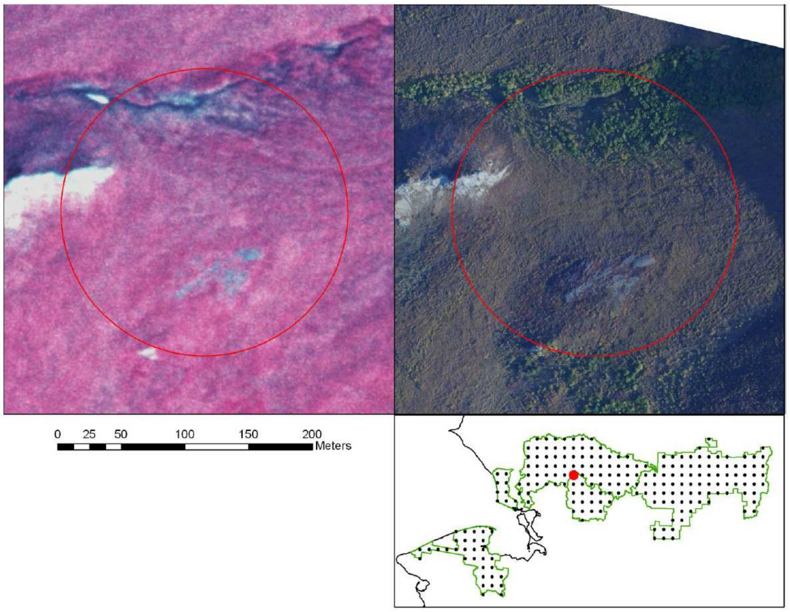


Figure C.1: A pair of images collected in 1979 and 2008, respectively, on a plot in the Noatak national Preserve, Alaska (red dot on the inset map). The historic image (AHAP color-infrared photo, left) consists mostly of herbaceous and low shrub vegetation. The contemporary image (small-format true color photo, right) consists mostly of tall shrubs. Source: Swanson (2013).

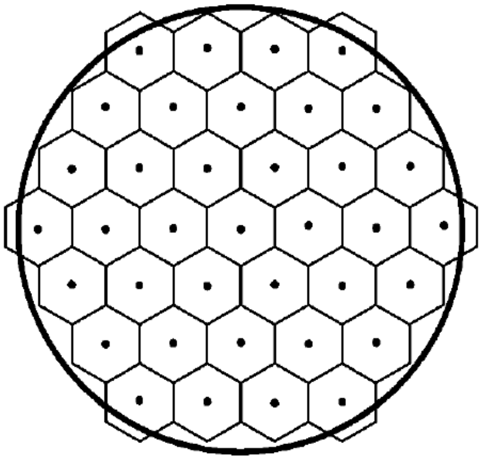


Figure C.2: A four-hectare sample plot with 37 hexagonal subplots for ecotype classification. Source: Swanson (2013).

C.4 Markov chain Monte Carlo algorithm

1. Define initial values $\left\{\boldsymbol{\eta}_{i,s,0}^{(0)}\right\}_{i,s}$, $\left\{\boldsymbol{\alpha}_{k}^{(0)}\right\}_{k}$, $\left\{\boldsymbol{\beta}_{k}^{(0)}\right\}_{k}$, $\left\{\tilde{\boldsymbol{\xi}}_{k}^{(0)}\right\}_{k}$, $\left\{\tilde{\boldsymbol{\epsilon}}_{i,k}^{(0)}\right\}_{i,k}$, $\sigma_{\zeta}^{2(0)}$, $\sigma_{\xi}^{2(0)}$, $\sigma_{\epsilon}^{2(0)}$, and $\phi^{(0)}$ by sampling from their respective prior distributions. For simplicity, we denote quantities relative to images in c. 1980 with temporal subscript 0 ($\boldsymbol{y}_{i,s,0}$, $\boldsymbol{N}_{i,s,0}$, $\boldsymbol{\kappa}_{i,s,0}$, and $\boldsymbol{\omega}_{i,s,0}$) and quantities relative to images in c. 2010 with temporal subscript 1 ($\boldsymbol{y}_{i,s,1}$, $\boldsymbol{N}_{i,s,1}$, $\boldsymbol{\kappa}_{i,s,1}$, $\boldsymbol{\omega}_{i,s,1}$, $\tilde{\boldsymbol{\epsilon}}_{i,s,1}$, $\tilde{\boldsymbol{\xi}}_{i,s,1}$, and $\boldsymbol{\Delta}_{i,s,1}$). Let $\boldsymbol{A}$ be the matrix comprised of $\boldsymbol{\alpha}_k$ and $\boldsymbol{B}$ be the matrix comprised of $\boldsymbol{\beta}_k$ for $k = 1, \ldots, K-1$. Calculate $\{\boldsymbol{N}_{i,s,0}\}_{i,s}$, $\{\boldsymbol{N}_{i,s,1}\}_{i,s}$, $\{\boldsymbol{\kappa}_{i,s,0}\}_{i,s}$, and $\{\boldsymbol{\kappa}_{i,s,1}\}_{i,s}$ as described in Section 4.2. Sample $\left\{\omega_{i,s,0,k}^{(0)}\right\}_{i,s,k}$ and $\left\{\omega_{i,s,1,k}^{(0)}\right\}_{i,s,k}$ from PG$(0,1)$.
2. Set $q = 1$.
3. Sample $\boldsymbol{\eta}_{i,s,0}^{(q)}$ for $i = 1, \ldots, n_I$ and $s = 1, \ldots, n_S$ from its posterior distribution

$$\left[\boldsymbol{\eta}_{i,s,0}^{(q)}\middle|\cdot\right] = \mathrm{N}\left(\boldsymbol{m}_{i,s,0}, \boldsymbol{V}_{i,s,0}\right),$$

where

$$\begin{aligned}
\boldsymbol{V}_{i,s,0} &= \left(\sigma_{\zeta}^{-2(q-1)}\boldsymbol{I} + \boldsymbol{\Omega}_{i,s,0}^{(q-1)} + \boldsymbol{\Omega}_{i,s,1}^{(q-1)}\right)^{-1}, \\
\boldsymbol{m}_{i,s,0} &= \boldsymbol{V}_{i,s,0}\left(\sigma_{\zeta}^{-2(q-1)}\boldsymbol{A}^{(q-1)\prime}\boldsymbol{h}_{i,s} + \boldsymbol{\kappa}_{i,s,0} + \boldsymbol{\kappa}_{i,s,1} - \boldsymbol{\Omega}_{i,s,1}^{(q-1)}\boldsymbol{\Delta}_{i,s,1}^{(q-1)}\right), \\
\boldsymbol{\Omega}_{i,s,0}^{(q-1)} &= \mathrm{diag}\left(\left\{\omega_{i,s,0,k}^{(q-1)}\right\}_{k}\right), \text{ and } \boldsymbol{\Omega}_{i,s,1}^{(q-1)} = \mathrm{diag}\left(\left\{\omega_{i,s,1,k}^{(q-1)}\right\}_{k}\right).
\end{aligned}$$

4. Sample $\boldsymbol{\alpha}_{k}^{(q)}$ for $k = 1, \ldots, K-1$ from its posterior distribution

$$\left[\boldsymbol{\alpha}_{k}^{(q)}\middle|\cdot\right] = \mathrm{N}\left(\boldsymbol{m}_{\alpha}^{k}, \boldsymbol{V}_{\alpha}^{k}\right),$$

where

$$\boldsymbol{V}_{\alpha}^{k} = \left(\sum_{i} \sum_{s} \boldsymbol{h}_{i,s} \boldsymbol{h}_{i,s}' / \sigma_{\zeta}^{(q-1)} + \boldsymbol{\Sigma}_{\alpha}^{-1} \right)^{-1} \text{ and}$$
$$\boldsymbol{m}_{\alpha}^{k} = \boldsymbol{V}_{\alpha}^{k} \left(\sum_{i} \sum_{s} \eta_{i,s,0,k}^{(q)} \boldsymbol{h}_{i,s} / \sigma_{\zeta}^{2(q)} + \boldsymbol{\Sigma}_{\alpha}^{-1} \boldsymbol{\mu}_{\alpha} \right).$$

5. Sample $\boldsymbol{\beta}_{k}^{(q)}$ for $k = 1, \dots, K-1$ from its posterior distribution

$$\left[\boldsymbol{\beta}_{k}^{(q)} \middle| \cdot \right] = \mathrm{N}\left(\boldsymbol{m}_{\beta}^{k}, \boldsymbol{V}_{\beta}^{k} \right),$$

where

$$\boldsymbol{V}_{\beta}^{k} = \left(\boldsymbol{\Sigma}_{\beta}^{-1} + \tilde{\boldsymbol{\Omega}}_{\beta}^{k} \right), \ \boldsymbol{m}_{\beta}^{k} = \boldsymbol{V}_{\beta}^{k} \left(\boldsymbol{\Sigma}_{\beta}^{-1} \boldsymbol{\mu}_{\beta} + \tilde{\boldsymbol{\kappa}}_{\beta}^{k} \right),$$
$$\tilde{\boldsymbol{\Omega}}_{\beta}^{k} = \sum_{i} \sum_{s} \omega_{i,s,1,k}^{(q-1)} \boldsymbol{x}_{i,s} \boldsymbol{x}_{i,s}', \text{ and}$$
$$\tilde{\boldsymbol{\kappa}}_{\beta}^{k} = \sum_{i} \sum_{s} \left(\kappa_{i,s,1,k} - \omega_{i,s,1,k}^{(q-1)} \left(\eta_{i,s,0,k}^{(q)} + \tilde{\epsilon}_{i,s,1,k}^{(q-1)} + \tilde{\xi}_{i,s,1,k}^{(q-1)} \right) \right) \boldsymbol{x}_{i,s}'.$$

6. Sample $\tilde{\boldsymbol{\epsilon}}_{i,k}^{(q)}$ from its posterior distribution

$$\left[\tilde{\boldsymbol{\epsilon}}_{i,k}^{(q)} \middle| \cdot \right] = \mathrm{N}\left(\boldsymbol{m}_{i,k}, \boldsymbol{V}_{i,k} \right),$$

where

$$\boldsymbol{V}_{i,k} = \left((\boldsymbol{R} - \boldsymbol{W}) / \left(|\mathcal{T}_i| \sigma_{\epsilon}^{2(q-1)} \right) + \tilde{\boldsymbol{\Omega}}_{i,k} \right)^{-1}, \ \boldsymbol{m}_{i,k} = \boldsymbol{V}_{i,k} \tilde{\boldsymbol{\kappa}}_{i,k},$$
$$\tilde{\boldsymbol{\Omega}}_{i,k} = \mathrm{diag}\left(\left\{ \omega_{i,s,1,k}^{(q-1)} \right\}_{s} \right), \text{ and}$$
$$\tilde{\boldsymbol{\kappa}}_{i,k} = \left(\left\{ \kappa_{i,s,1,k} - \omega_{i,s,1,k}^{(q-1)} \left(\eta_{i,s,0,k}^{(q)} + \boldsymbol{x}_{i,s}' \boldsymbol{\beta}_{k}^{(q)} + \tilde{\xi}_{i,s,1,k}^{(q-1)} \right) \right\}_{s} \right)'.$$

7. Sample $\tilde{\boldsymbol{\xi}}_k^{(q)}$ for $k = 1, \ldots, K-1$ from its posterior distribution

$$\left[\tilde{\boldsymbol{\xi}}_k^{(q)} \middle| \cdot\right] = \mathrm{N}\left(\boldsymbol{m}_\xi^k, \boldsymbol{V}_\xi^k\right),$$

   where

$$\begin{aligned}
\boldsymbol{V}_\xi^k &= \left(\boldsymbol{\Sigma}_\xi^{-1} + \tilde{\boldsymbol{\Omega}}_\xi^k\right)^{-1}, \ \boldsymbol{m}_\xi^k = \boldsymbol{V}_\xi^k \tilde{\boldsymbol{\kappa}}_\xi^k, \\
\tilde{\boldsymbol{\Omega}}_\xi^k &= \text{diag}\left(\sum_s \omega_{i,s,1,k}^{(q-1)}\right), \text{ and} \\
\tilde{\boldsymbol{\kappa}}_\xi^k &= \left(\left\{\sum_s \kappa_{i,s,1,k} - \omega_{i,s,1,k}^{(q-1)}\left(\eta_{i,s,0,k}^{(q)} + \boldsymbol{x}_{i,s}'\boldsymbol{\beta}_k^{(q)} + \tilde{\epsilon}_{i,s,1,k}^{(q)}\right)\right\}_i\right)'.
\end{aligned}$$

   Update $\Delta_{i,s,1,k}^{(q)} = \boldsymbol{x}_{i,s}'\boldsymbol{\beta}_k^{(q)} + \tilde{\epsilon}_{i,s,1,k}^{(q)} + \tilde{\xi}_{i,s,1,k}^{(q)}$.

8. Sample $\omega_{i,s,0,k}^{(q)}$ and $\omega_{i,s,1,k}^{(q)}$ for $i = 1, \ldots, n_I$, $s = 1, \ldots, n_S$, and $k = 1, \ldots, K-1$ from their respective posterior distributions

$$\begin{aligned}
\left[\omega_{i,s,0k}^{(q)} \middle| \cdot\right] &= \text{PG}\left(N_{i,s,0,k}, \eta_{i,s,0,k}^{(q)}\right), \text{ and} \\
\left[\omega_{i,s,1,k}^{(q)} \middle| \cdot\right] &= \text{PG}\left(N_{i,s,1,k}, \eta_{i,s,0,k}^{(q)} + \Delta_{i,s,1,k}^{(q)}\right).
\end{aligned}$$

9. Sample $\sigma_\zeta^{2(q)}$ from its posterior distribution

$$\left[\sigma_\zeta^{2(q)} \middle| \cdot\right] = \text{IG}\left(\tilde{a}_\zeta, \tilde{b}_\zeta\right),$$

   where

$$\begin{aligned}
\tilde{a}_\zeta &= a_\zeta + n_I \times n_S \times (K-1)/2 \text{ and} \\
\tilde{b}_\zeta &= \left(b_\zeta^{-1} + 2^{-1}\sum_i\sum_s\sum_k \left(\eta_{i,s,0,k}^{(q)} - \boldsymbol{h}_{i,s}'\boldsymbol{\alpha}_k^{(q)}\right)'\left(\eta_{i,s,0,k}^{(q)} - \boldsymbol{h}_{i,s}'\boldsymbol{\alpha}_k^{(q)}\right)\right)^{-1}.
\end{aligned}$$

10. Sample $\sigma_\epsilon^{2(q)}$ from its posterior distribution

$$\left[\sigma_\epsilon^{2(q)} \middle| \cdot\right] = \text{IG}\left(\tilde{a}_\epsilon, \tilde{b}_\epsilon\right),$$

where

$$\tilde{a}_\epsilon = a_\epsilon + n_I \times n_S \times (K-1)/2 \text{ and}$$
$$\tilde{b}_\epsilon = \left(b_\epsilon^{-1} + 2^{-1}\sum_i \sum_k \tilde{\boldsymbol{\epsilon}}_{i,k}^{(q)\prime} |\mathcal{T}_i|^{-1}(\boldsymbol{R} - \boldsymbol{W})\tilde{\boldsymbol{\epsilon}}_{i,k}^{(q)}\right)^{-1}.$$

11. Sample $\sigma_\xi^{2(q)}$ from its posterior distribution

$$\left[\sigma_\xi^2 \middle| \cdot\right] = \text{IG}\left(\tilde{a}_\xi, \tilde{b}_\xi\right),$$

where

$$\tilde{a}_\xi = a_\xi + n_I \times (K-1)/2,$$
$$\tilde{b}_\xi = \left(b_\xi^{-1} + 2^{-1}\sum_k \tilde{\boldsymbol{\xi}}_k^{(q)\prime}\left(\exp\left(-\boldsymbol{D}/\phi^{(q-1)}\right)\boldsymbol{S}\right)^{-1}\tilde{\boldsymbol{\xi}}_k^{(q)}\right)^{-1}, \text{ and}$$
$$\boldsymbol{S}_{ij} = \begin{cases} |\mathcal{T}_i|, \text{ if } i = j; \\ |\mathcal{T}_i \cap \mathcal{T}_j|, \text{ o.w.} \end{cases}$$

12. Update $\phi^{(q)}$ using M-H. Propose ϕ^* from N $\left(\phi^{(q-1)}, \sigma^2_{\phi,\text{tune}}\right)$. If $\phi^* \in (a_\phi, b_\phi)$, then calculate

$$\text{mh}_\phi = \frac{\prod_k \left[\tilde{\boldsymbol{\xi}}_k^{(q)} \middle| \sigma_\xi^{2(q)}, \phi^*\right]}{\prod_k \left[\tilde{\boldsymbol{\xi}}_k^{(q)} \middle| \sigma_\xi^{2(q)}, \phi^{(q-1)}\right]}.$$

Let $\phi^{(q)} = \phi^*$ with probability $\min(1, \text{mh}_\phi)$. Otherwise, let $\phi^{(q)} = \phi^{(q-1)}$.

13. Set $q = q + 1$ and return to Step 3. Iterate the algorithm through Steps 3 to 10 until the sample size is large enough to approximate the posterior distribution.