HIERARCHICAL SPATIO-TEMPORAL MODELS FOR ECOLOGICAL PROCESSES

A Dissertation

presented to

the Faculty of the Graduate School

University of Missouri-Columbia

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

---

by

MEVIN B. HOOTEN

Dr. Christopher K. Wikle, Dissertation Supervisor

AUGUST 2006

The undersigned, appointed by the Dean of the Graduate School, have examined the dissertation entitled:

HIERARCHICAL SPATIO-TEMPORAL MODELS FOR ECOLOGICAL PROCESSES

Presented by Mevin B. Hooten

a candidate for the degree of Doctor of Philosophy, and hereby certify that in their opinion it is worthy of acceptance.

_____

_____

_____

_____

_____

ACKNOWLEDGEMENTS

I would like to thank my wife (Gina Hooten), advisor (Chris Wikle), doctoral committee (Larry Ries, Sakis Micheas, Dave Larsen, and Josh Millspaugh), family, and friends. Also a special thanks to Joe Cavanaugh, Rose-Marie Muzika, Judy Dooley, Tracy Pickens, Richard Smemoe, Ali Arab, Andy Royle, Bob Dorazio, Lance Waller, and Noel Cressie for providing helpful comments, data, advice, and assistance.

Finally, a very special thanks to my parents for providing me with the two simple rules by which I live:

1.) The path in life is chosen by he who lives.

2.) Success is only limited by desire.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 GENERAL INTRODUCTION

1.1 Ecology in Space and Time

Ecology has been defined numerous times since its introduction in the last half of the nineteenth century (Krebs 1978). Perhaps the best concise definition comes from Elton (1927) where he claims ecology is, "scientific natural history." Krebs (1978), seeking a more descriptive definition, stated that, "ecology is the scientific study of the interactions that determine the distribution and abundance of organisms." Studying the distribution and abundance of organisms generally involves the investigation of physical and biotic processes. Such processes naturally occur and vary over space and time, and though often over-simplified in experimental settings, are generally non-linear and dynamic (Cushing et al. 2003).

The fields of landscape ecology, metapopulation ecology, macroecology, and biogeography are all concerned with analyzing ecological processes that vary in space and (or) time (MacArthur and Wilson 1967; Forman and Godron 1986; Brown and Maurer 1989; Hanski 1999). Each discipline places emphasis on different aspects of ecosystem function and variation, yet they are all rapidly growing areas of interest, linked by a common thread. That is, they are linked by a large-scale spatial perspective. Additionally, imminent rapid global climate change and the ever-increasing human impact on natural ecosystems have brought temporal analysis of ecological processes to the forefront in terms of scientific priority (Clark et al. 2001).

In order to effectively study ecological phenomena on large spatial and temporal domains, extensive observation of the relevant process is required. Data collection on these scales is intensive and often subject to spatial and (or) temporal irregularity (Renshaw

1991). Rarely are spatio-temporal processes observed with replication, and, in fact, it could be argued that it is impossible to truly measure a natural time-evolving process repeatedly (e.g., Reckhow 1990). Thus scientific examination of such processes has conventionally consisted of inappropriately specified statistical models or non-statistical deterministic models (Clark 2005).

1.2 Ecological Modeling

It is sensible to employ deterministic models for studying systems that follow fixed rules and are observed with absolute precision (e.g., Zhao 2003). Similarly specified models are even reasonable in cases where the system may not be observed perfectly but is under the ultimate control of the observer and can be designed to attain a given level of precision (e.g., as is done in engineering and simulation modeling). In settings where a naturally varying system is imprecisely observed, probability models (i.e., statistical models) can be useful for characterizing the processes of interest as well as their inherent variability, given, potentially, multiple sources of uncertainty (Hilborn and Mangel 1997).

Much statistical data analysis used in ecology still involves ANOVA and Regression methods applied in experimental settings. These proven methods have retained their utility for testing statistical hypotheses in carefully controlled or designed experimental settings. The analysis of processes that evolve in space and time cannot always be approached in the same manner, however (at least partially due to complicated correlation structures in the data, Banerjee et al. (2004)). Scientific disciplines that do not have the luxury of a controlled experimental setting and can only observe constantly evolving non-linear dynamical systems (e.g., atmospheric and astonomical sciences), have developed extensive methodologies for

studying such systems (Wikle 2002a). Though not always directly transferable to other disciplines, many existing approaches can serve as motivation for the development of new statistical methodologies catered more specifically to ecological problems.

1.2.1 Hierarchical Models

A main goal in the rigorous characterization of ecological phenomena is estimation and prediction of processes as well as the parameters governing processes. Thus a flexible framework capable of accomodating complex relationships between data and process models while incorporating various sources of uncertainty is necessary. Traditional likelihood based approaches to modeling have allowed for ecologically meaningful data structures, though, in complicated situations with heavily parameterized models and limited or missing data, estimation by likelihood maximization is often problematic or infeasible (Hilborn and Mangel 1997). Developments in numerical approximation methods have been useful in many cases, especially high-dimensional parameter spaces (e.g., Newton and E-M methods, Givens and Hoeting 2005), though can still be difficult or impossible to implement and have no provision for accomodating uncertainty at multiple levels (traditional implementation in spatio-temporal settings is discussed in Section 1.2.2).

Hierarchical models, whereby a problem is decomposed into a series of levels, assume a very flexible framework capable of accomodating uncertainty and potential *a priori* scientific knowledge while retaining many advantages of a strict likelihood approach (e.g., multiple sources of data and scientifically meaningful structure). Gelfand and Smith (1990) illustrated the accessibility of hierarchical statistical models and their implementation via computation, while, Berliner (1996) provided the simple, yet powerful, representation of

the complicated and often intractable joint Bayesian posterior distribution in terms of a product of simpler and often analytically tractable conditional distributions.

The years after introduction of the Bayesian hierarchical model and development of Markov Chain Monte Carlo (MCMC) have brought on an explosion of research, both theoretical and applied, utilizing and (or) developing hierarchical models. Specifically, ecology has witnessed a dramatic increase in the complexity of models implemented as well as an increase in the scope and scale of ecological problems addressed (Clark 2005). In fact, one could argue that ecology, as with many other disciplines, has witnessed a paradigm shift in ways of thinking about ecological problems. No longer restricted (or confused) by Analysis of Variance and Regression models, ecologists, with minimal training (e.g., calculus and probability theory) can employ simple yet effective and powerful models to statistically characterize natural phenomena using previously existing data, regardless of its intended use. The present state of Bayesian hierarchical modeling in ecology, the philosophy of which was once hotly debated (as in statistics), has reached a point where *Ecology*, the premier journal in the field, contains at least one hierarchical Bayes paper in nearly every issue.

The developers of the automated Gibbs sampler program, BUGS (Thomas 1994), have been quick to point out the caveat that comes with misuse of the software (and Bayes methods in general):

> There is . . . a need for caution. A knowledge of Bayesian statistics is assumed, including recognition of the potential importance of prior distributions, and MCMC is inherently less robust than analytic statistical methods. There is no in-built protection against misuse.

With this warning in mind, consider a brief overview of the Bayesian hierarchical framework as detailed by Wikle et al. (1998), Calder et al. (2003), Wikle (2003), and others. Hierarchical modeling is based on a simple fact from probability that the joint distribution of a collection of random variables can be decomposed into a series of conditional models. For example, if a, b, c are random variables, then basic probability allows the factorization $[a, b, c] = [a|b, c][b|c][c]$. In the case of spatio-temporal models, the joint distribution describes the behavior of the process at all spatial locations and all times. This is difficult to specify for complicated processes. Typically, it is much easier to specify the distribution of the conditional models. In that case, the product of the series of relatively simple conditional models gives a joint distribution that can be quite complex.

When modeling complicated processes in the presence of data, it is helpful to write the hierarchical model in three basic stages:

Stage 1. Data Model: $[data|process, data\ parameters]$

Stage 2. Process Model: $[process|process\ parameters]$

Stage 3. Parameter Model: $[data\ parameters\ and\ process\ parameters]$.

The basic idea is to approach the complex problem by breaking it into subproblems. Although hierarchical modeling has been around a long time in statistics (Gelman et al. 2004), this basic formulation for modeling complicated temporal and spatio-temporal processes in the environmental sciences is relatively new (e.g., Berliner 1996; Wikle et al. 1998). The first stage is concerned with the observational process or "data model," which specifies the distribution of the data given the fundamental process of interest and parameters that describe the data model. The second stage then describes the process, conditional on other

process parameters. Finally, the last stage models the uncertainty in the parameters, from both the data and process stages. Note that each of these stages can have many sub-stages (e.g., Wikle et al. 1998, 2001).

The goal is to estimate the distribution of the process and parameters updated by the data. This posterior distribution is obtained via Bayes' Theorem:

$$[process, parameters|data] \propto [data|process, parameters][process|parameters][parameters]. \tag{1.1}$$

Bayesian statistics involves drawing statistical conclusions from the "posterior" which is proportional to the data model (i.e., likelihood) times the *a priori* knowledge (i.e., the prior). Bayes' theorem (in general probability form: [posterior] $\propto$ [likelihood][prior]) provides access to the posterior. Although simple in principle, the implementation of Bayes' theorem for complicated models can be challenging. One challenge concerns the specification of the parameterized component distributions on the right-hand side of (1.1). Although there has long been a debate in the statistics community concerning the appropriateness of "subjective" specification of such distributions, such choices are a natural part of scientific-based modeling. In fact, the incorporation of scientific-based information into these prior distributions provides a coherent mechanism by which to incorporate the uncertainty related to these specifications explicitly in the model. Perhaps more importantly from a practical perspective is the calculation of the posterior. The complex and high-dimensional nature of ecological models (and indeed, most spatio-temporal models) prohibits the direct evaluation of the posterior. However, MCMC approaches can be utilized to draw samples from the posterior distribution. As previously mentioned, the use of MCMC in Bayesian

hierarchical models has led to a revolution in that realistic (i.e., complicated) models can be considered; especially in the analysis of spatio-temporal processes. Yet, typically the computational burden must be considered when formulating the conditional models in such problems. Thus, the model building phase requires not only scientific understanding of the problem, but in what ways that understanding can be modified to fit into the MCMC computational framework.

1.2.2 Spatio-temporal Models

Consider the arbitrary spatio-temporal process $Y(s;t)$ where s is a spatial location in some spatial domain D (typically in two-dimensional Euclidean space, but not restricted to that case) and t denotes time, $t = \{t_1, \ldots, t_T\}$. Most processes in the physical, environmental and ecological sciences behave in such a way that the process at the current time is related to the process at a previous time (or times). Such processes are generally termed "dynamic" or "dynamical." Assuming the process cannot be completely described by deterministic rules, it would be ideal to characterize the joint distribution of this process for all times and spatial locations. Typically, this is not possible without some significant restrictions on the distribution. A common restriction is to assume the process behaves in a Markovian fashion; that is, the process at the current time, conditioned on all of the past, can be expressed completely by conditioning only on the most recent past. For example, consider the case with a finite number of spatial locations $\{s_1, \ldots, s_n\}$ and discrete times $t = \{0, 1, 2, \ldots, T\}$. Let $\mathbf{Y}_t \equiv (Y(s_1;t), \ldots, Y(s_n;t))'$. Then, the joint distribution of the

spatio-temporal process can be factored as follows:

$$[\mathbf{Y}_0, \ldots, \mathbf{Y}_T] = [\mathbf{Y}_T|\mathbf{Y}_{T-1}, \ldots, \mathbf{Y}_0][\mathbf{Y}_{T-1}|\mathbf{Y}_{T-2}, \ldots, \mathbf{Y}_0] \ldots [\mathbf{Y}_2|\mathbf{Y}_1, \mathbf{Y}_0][\mathbf{Y}_1|\mathbf{Y}_0][\mathbf{Y}_0], \tag{1.2}$$

where the use of brackets [ ] denotes distribution and $[a|b]$ to denote the conditional distribution of a given b. With the first-order Markov assumption, (1.2) can be written,

$$[\mathbf{Y}_0, \ldots, \mathbf{Y}_T] = [\mathbf{Y}_T|\mathbf{Y}_{T-1}][\mathbf{Y}_{T-1}|\mathbf{Y}_{T-2}] \ldots [\mathbf{Y}_2|\mathbf{Y}_1][\mathbf{Y}_1|\mathbf{Y}_0][\mathbf{Y}_0]. \tag{1.3}$$

This Markovian assumption is a dramatic simplification of (1.2), yet one that is very often realistic for dynamical processes. From a modeling perspective, the component distribution $[\mathbf{Y}_t|\mathbf{Y}_{t-1}]$ must then be specified. In general, this is written in terms of some function $\mathbf{Y}_t = h(\mathbf{Y}_{t-1}; \boldsymbol{\delta})$, where the parameters $\boldsymbol{\delta}$ describe the dynamical propagation. This function can be non-linear, and the associated distribution can be Gaussian or non-Gaussian. For illustration, consider the linear evolution equation with Gaussian errors,

$$\mathbf{Y}_t = \mathbf{H}\mathbf{Y}_{t-1} + \boldsymbol{\eta}_t, \;\; \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \tag{1.4}$$

where the propagator matrix $\mathbf{H}$ is an $n \times n$ matrix of typically unknown parameters. Consider the i-th element of $\mathbf{Y}_t$ and the associated evolution equation implied by (1.4),

$$Y(s_i; t) = \sum_{k=1}^{n} h(i, k) Y(s_k; t-1) + \eta(s_i; t), \tag{1.5}$$

where $h(i, k)$ refers to the i-th row and k-th column of $\mathbf{H}$. Thus, (1.5) shows that the process value at location s_i at time t is a linear combination of all the process values at the previous time, with the relative contribution given by the "redistribution" weights $h(i, k)$, and the addition of possibly correlated noise $\eta(s_i; t)$. The additive noise ($\boldsymbol{\eta}$) may not always be

necessary, as the general state equation (i.e., $\mathbf{Y}_t = \mathbf{H}\mathbf{Y}_{t-1}$) can be quite flexible depending on specification of $\mathbf{H}$. The additive noise ($\boldsymbol{\eta}$) is often used to account for misspecification of the process and (or) to allow for MCMC implementation and estimation of $\mathbf{Y}_t$. That is, the process, $\mathbf{Y}_t$, given $\mathbf{Y}_{t-1}$ and $\boldsymbol{\eta}_t$ can be sampled as a multivariate normal random variable.

The difficulty with such formulations in practice is that for most environmental and ecological processes the number of spatial locations of interest, n, is quite large, and there is simply not enough information to obtain reliable estimates of all parameters $h(i,k), i,k = 1,\ldots,n$. Thus, typically the propagator matrix $\mathbf{H}$ must be parameterized in terms of some parameters $\boldsymbol{\delta}$, whose dimensionality is significantly less than the n^2 required to estimate $\mathbf{H}$ directly.

Perhaps the simplest statistical parameterization for $\mathbf{H}$ is to assume $\mathbf{H} = \mathbf{I}$, a multivariate random-walk. Although advantageous from the perspective of having the fewest parameters in $\mathbf{H}$ (zero), this model is non-stationary in time. More importantly, such a structure is not able to capture complex interaction across space and time, and is not realistic for most physical, environmental, and ecological processes. A natural modification is to allow $\mathbf{H} = diag(\mathbf{h})$, a diagonal matrix with elements on the diagonal potentially varying with spatial location. Such a model is non-separable in space-time, yet it still does not account for realistic interactions between multiple spatial locations across time.

1.2.3 IDE-Based Dynamics

Now consider an alternative, yet related, approach for parameterizing $\mathbf{H}$. To capture dynamical interactions in space-time that are realistic for ecological processes, the propagator matrix $\mathbf{H}$ must contain non-zero off-diagonal elements. This can be seen clearly from the

IDE (i.e., integro-difference equation) perspective. Consider the linear stochastic IDE equation,

$$Y(s;t) = \int h(s,r)Y(r;t-1)dr + \eta(s;t), \tag{1.6}$$

where the error process $\eta(s;t)$ is correlated in space, but not time, and the redistribution kernel $h(s,r)$ describes how the process at the previous time is redistributed to the current time. Although similar to equation (1.5), the IDE equation considers continuous space rather than discrete space. General IDE equations are quite powerful for describing ecological processes (e.g., Kot 2001); the dynamics are controlled by the properties of the redistribution kernel. For example, the dilation of the kernel controls the rate of diffusion, and advection can be controlled by the skewness of the kernel (Wikle 2002b). In addition, the characteristics of the dynamics that can be explained are affected by the kernel tail thickness and modality. Although such models are rich in describing complicated ecological processes, they have not often been "fit" to data in a rigorous statistical framework. Wikle (2002b) and Xu et al. (2005) show that such models can be fit to data and that allowing the kernels to vary with spatial location can dramatically increase the complexity of the dynamics modeled. From an implementation perspective, a discretization of (1.6) suggests potential parameterizations of $\mathbf{H}$ as a function of the kernel parameters, $\boldsymbol{\delta}$. Such parameterizations include non-zero off-diagonal elements, and typically the propagator is non-symmetric (i.e., $h(i,k) \neq h(k,i)$) allowing for complicated interactions in time and space while using relatively few kernel parameters.

Disadvantages of using IDE models in this setting can be implementation within a statistical framework, parameter estimation (although hierarchical Bayes helps), choice of

an appropriate kernel, allowing for spatially varying parameters, and reduced computational efficiency due to non-sparse $\mathbf{H}$ matrix.

1.2.4 PDE-Based Dynamics

The IDE-based dynamics of the previous section suggest that the simplest, realistic statistical parameterization of $\mathbf{H}$ would have diagonal and non-symmetric non-diagonal elements. One could simply parameterize such a model statistically (e.g., see Wikle, Berliner and Cressie, 1998). However, in the case of physical and ecological processes, there is often a significant *a priori* understanding of the theory of the underlying dynamical process through differential equations (e.g., Holmes et al. 1994). In the case of linear PDEs (i.e., partial differential equations), standard finite differencing implies equations such as (1.4). More importantly, such discretizations imply parameterizations of $\mathbf{H}$ in terms of important parameters of the PDE, as well as the finite-difference discretization parameters (e.g., Wikle 2003).

Consider the general diffusion PDE,

$$\frac{\partial u}{\partial t} = \mathcal{H}(u, w, \delta), \tag{1.7}$$

where $\mathcal{H}$ is some operator on the variable of interest, u, other potential variables, w, and parameters δ. Note that the variable of interest (u) could simply be a function of the previously considered spatio-temporal process (y). Simple finite difference representations (e.g., Haberman 1987) suggest an approximate difference equation model,

$$\mathbf{u}_t = h(\mathbf{u}_{t-\Delta_t}, \mathbf{w}, \boldsymbol{\delta}) + \boldsymbol{\eta}_t, \tag{1.8}$$

where again the noise term $\boldsymbol{\eta}_t$ is added to account for the error of model misspecification.

Now, for illustration, consider the simple diffusion equation,

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\delta(x,y)\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(\delta(x,y)\frac{\partial u}{\partial y}\right), \tag{1.9}$$

where $u_t(x,y)$ is a spatio-temporal process at spatial location $\mathbf{s} = (x,y)$ in two-dimensional Euclidean space at time t and $\delta(x,y)$ is a spatially varying diffusion coefficient. Though many possible discretizations exist, forward differences in time and centered differences in space (e.g., Haberman 1987) give the difference equation representation of (1.9),

$$\begin{aligned}
u_t(x,y) \;=\; & u_{t-\Delta_t}(x,y)\left[1 - 2\delta(x,y)\left(\frac{\Delta_t}{\Delta_x^2} + \frac{\Delta_t}{\Delta_y^2}\right)\right] \\
& + \; u_{t-\Delta_t}(x-\Delta_x,y)\left[\frac{\Delta_t}{\Delta_x^2}\left\{\delta(x,y) - (\delta(x+\Delta_x,y) - \delta(x-\Delta_x,y))/4\right\}\right] \\
& + \; u_{t-\Delta_t}(x+\Delta_x,y)\left[\frac{\Delta_t}{\Delta_x^2}\left\{\delta(x,y) + (\delta(x+\Delta_x,y) - \delta(x-\Delta_x,y))/4\right\}\right] \\
& + \; u_{t-\Delta_t}(x,y+\Delta_y)\left[\frac{\Delta_t}{\Delta_y^2}\left\{\delta(x,y) + (\delta(x,y+\Delta_y) - \delta(x,y-\Delta_y))/4\right\}\right] \\
& + \; u_{t-\Delta_t}(x,y-\Delta_y)\left[\frac{\Delta_t}{\Delta_y^2}\left\{\delta(x,y) - (\delta(x,y+\Delta_y) - \delta(x,y-\Delta_y))/4\right\}\right] \\
& + \; \eta_t(x,y),
\end{aligned} \tag{1.10}$$

where it is assumed that the discrete u-process is on a rectangular grid with spacing Δ_x and Δ_y in the longitudinal and latitudinal directions, respectively, and with time spacing Δ_t. Again, the error term $\eta_t(x,y)$ has been added to (1.10) to account for the uncertainties due to the discretization as well as other model misspecifications.

From (1.10) it can be seen that the discretization can be written as (1.5) or (1.4) where the propagator (redistribution) matrix $\mathbf{H}$ depends upon the diffusion coefficients $\boldsymbol{\delta} = [\delta(s_1), \ldots, \delta(s_n)]'$ and the discretization parameters Δ_t, Δ_x, and Δ_y,

$$\mathbf{u}_t = \mathbf{H}(\boldsymbol{\delta}, \Delta_t, \Delta_x, \Delta_y)\mathbf{u}_{t-\Delta_t} + \mathbf{H}_B(\boldsymbol{\delta}, \Delta_t, \Delta_x, \Delta_y)\mathbf{u}^B_{t-\Delta_t} + \boldsymbol{\eta}_t, \tag{1.11}$$

where again, $\mathbf{u}_t$ corresponds to an arbitrary vectorization of the gridded u-process at time t, $\mathbf{H}(\boldsymbol{\delta}, \Delta_t, \Delta_x, \Delta_y)$ is a sparse $n \times n$ matrix with essentially five non-zero diagonals corresponding to the bracket coefficients in (1.10), hence its dependence on $\boldsymbol{\delta}$. Note also that a separate boundary specification is included, in that $\mathbf{u}^B_{t-\Delta_t}$ is an $n_B \times 1$ vector of boundary values for the u-process, and $\mathbf{H}_B(\boldsymbol{\delta}, \Delta_t, \Delta_x, \Delta_y)$ is an $n \times n_B$ sparse matrix with elements corresponding to the appropriate coefficients from (1.10). Thus, the product $\mathbf{H}_B(\boldsymbol{\delta}, \Delta_t, \Delta_x, \Delta_y)\mathbf{u}^B_{t-\Delta_y}$ is simply the specification of model edge effects.

Such diffusion models can be improved upon to better model "growing" ecological processes. This is the topic of Chapter 4. It is worth pointing out here, for the purposes of foreshadowing, that both IDE and PDE representations of dynamical systems suggest a Markovian structured form when discretized for use in an MCMC setting. The ecological literature commonly refers to such models as "matrix models" (e.g., Caswell 2001). This concept will be discussed in detail in Section 4.2.

1.2.5 CA-Based Dynamics

A dynamical system known as a cellular automaton (CA) can also be used to model naturally propagating phenomena. In such systems, a model defining the dynamics of the process can still be formulated with the state-space representation in (1.8), the primary difference being that the CA model is not based on a continuous operator such as the case with differential equation operators. Rather, the propagating operator h, is now defined by a neighborhood $\mathcal{N}$ and a set of rules $\mathcal{R}$ that map the past state of the neighborhood to the current state. In this way, an automaton (i.e., the entire partitioned domain made up of individual automata or cells) contains numerous interacting components, the state of which is governed by a set of

rules based on the previous state. These characteristics of automatous systems make them especially suited to high performance computing environments (e.g., parallel processing and cluster computing settings; Wolfram 1988).

In a deterministic CA, the current state ($\mathbf{u}_t$) is exactly mapped to the previous state ($\mathbf{u}_{t-1}$) by a function h, so that the state of the automaton $\mathbf{u}_t$ is unique given $\mathbf{u}_{t-1}$ and h. More specifically, each element $u_{i,t}$ (the automata) is exactly mapped to the previous set of neighboring automata $u_{\mathcal{N}_i,t-1}$ by the function h. The function h is often made up of a set of rules, defining the map from all possible states of the set $u_{\mathcal{N}_i,t-1}$ to the possible states of $u_{i,t}$. For binary valued automata, a one-dimensional spatial domain, and lag-1 neighborhood structure, the function h might be:

$$h = \begin{cases} u_{\mathcal{N}_i,t-1} = (0,0,0) \Rightarrow u_{i,t} = 0 \\ u_{\mathcal{N}_i,t-1} = (1,0,0) \Rightarrow u_{i,t} = 0 \\ u_{\mathcal{N}_i,t-1} = (0,1,0) \Rightarrow u_{i,t} = 0 \\ u_{\mathcal{N}_i,t-1} = (0,0,1) \Rightarrow u_{i,t} = 0 \\ u_{\mathcal{N}_i,t-1} = (1,1,0) \Rightarrow u_{i,t} = 1 \\ u_{\mathcal{N}_i,t-1} = (1,0,1) \Rightarrow u_{i,t} = 1 \\ u_{\mathcal{N}_i,t-1} = (0,1,1) \Rightarrow u_{i,t} = 1 \\ u_{\mathcal{N}_i,t-1} = (1,1,1) \Rightarrow u_{i,t} = 0 \end{cases}$$

The intuitive meaning of this function is based on several simple biological principles:

- If the neighborhood of a given automata has no active cells, including itself, then the given automata will remain inactive at the next time. ("active" here taken to mean occupied in the ecological sense).

- If an automata has exactly one member active in its neighborhood then it will become inactive if it was the active member and remain inactive if it wasn't the active member.

- If an automata has exactly two members active in its neighborhood, then it will either

become active if it was not one of the active members, or remain active if it was one of the active members.

- If an automata has all active neighbors, including itself, it will become inactive.

An automaton defined with this simple set of rules can exhibit strikingly complex behavior as a dynamical system. A deterministic automaton on a two-dimensional spatial domain can be defined similarly. CA systems are capable of exhibiting very complex behavior and from a modeling perspective are very powerful tools. The utilization of deterministic automatons from a statistical modeling perspective is challenging, however, because the rule space quickly becomes cumbersome as the size of the neighborhood inceases, the dimensionality of the system increases, and (or) the support of the automata is extended. For example, in the mapping function h above, there are only $2^3 = 8$ possible neighborhood structures creating a rule that is 8 dimensional; however, using the same neighborhood and one-dimensional spatial domain, there are $2^{(2^3)}$ possible rules. In general, for a neighborhood of dimension $d_{\mathcal{N}}$, there will be $2^{(2^{d_{\mathcal{N}}})}$ possible rules. For a two-dimensional spatial domain, with a Queen's neighborhood (i.e., $d_{\mathcal{N}} = 9$) the number of possible rules is 1.34×10^{154}. Searching the space of rules for CA systems with large neighborhoods is not feasible, therefore it is necessary to either simplify the space of rules using assumptions, or take an alternative approach.

A stochastic or probabilistic automata can be thought of as a random variable conditioned on the neighborhood structure at the previous time. For example, the evolution

equation of a probabilistic automata can be written as:

$$u_{i,t}|u_{\mathcal{N}_i,t-1} \sim h(u_{\mathcal{N}_i,t-1}), \tag{1.12}$$

whereas, the evolution equation of a deterministic automata is written:

$$u_{i,t} = h(u_{\mathcal{N}_i,t-1}). \tag{1.13}$$

In the case of the probabilistic automata, the function h is a probability distribution. For the binary automaton (i.e., $u_{i,t} \in \{0, 1\}$), the probability distribution h could be a Bernoulli distribution with the Bernoulli probability a function of the neighborhood at the previous time and some parameters. Notice that the specification in (1.12) is only different from the Gaussian state-space models in (1.5) and (1.8) in that the source of variability in the probabilistic CA (1.12) is more general than the additive Gaussian error term in the probabilistic PDE.

The advantage of the probabilistic specification (1.12) over the determinstic specification (1.13) is that, from a statistical perspective, it is easier to search the space of rules in terms of probability than it would be to do so exactly. The disadvantage is that the map from the previous state to the current state is no longer unique.

1.3 Overview of Chapters

In the chapters that follow, several hierarchical temporal and spatio-temporal models are developed and employed to characterize a variety of ecological processes.

Chapter 2 illustrates the utility of a somewhat conventional model specification for assessing the dynamics of a spatio-temporal process in ecology. It serves as a good starting place for the development of more sophisticated models and utilizes a naive, but effective,

parameterization of the propagator matrix ($\mathbf{H}$). More specifically, a regime-dependent state-space model is employed that allows for the detection and characterization of the changes in tree growth dynamics over space and time using readily available dendrochronological and climatic data in the presence of various sources of uncertainty. Methods common in atmospheric sciences but relatively unknown in ecology and forestry are employed to develop a hierarchical model for tree growth and describe the growth dynamics.

Chapter 3 introduces the basic concept of using a scientifically meaningful latent dynamical system within a hierarchical model. The models developed in this chapter only consider a dynamic ecological process over time (not in space). However, the methods developed therein are easily extended to the spatial setting (as discussed in Chapter 4). Specifically, in Chapter 3, scientifically based data and process models are applied to forecast the population size of endangered species for purposes of population viability analysis.

Chapter 4 develops methods specifically catered for (but not necessarily limited to) characterizing and predicting biological invasions. The spread of invasive species is a long studied subject that garners much interest in the ecological research community. Historically the phenomenon has been approached using a purely deterministic mathematical framework (usually involving differential equations of some form). Such deterministic methods, while scientifically meaningful, are generally highly simplified and fail to account for uncertainty in the data and process, of which observational knowledge could not possibly exist without error. Models developed in Chapter 4 address these issues and are applied specifically to the ongoing North American invasion of the Eurasian Collared-Dove.

In Section 4.1, a hierarchical Bayesian model is proposed, for modeling population

spread, that accomodates data sources with errors, dependence structures between population dynamics parameters, and takes into account prior scientific understanding via non-linear relationships between model parameters and space-time response variables. The process (i.e., relative abundance) is modeled as a Poisson response with spatially varying diffusion coefficients as well as a logistic population growth term using a common reaction-diffusion partial differential equation that realistically mimics the ecological process.

Section 4.2 approaches the problem from a different perspective. Models of relative abundance (such as those presented in 4.1) are informative and useful, however a better understanding of the dynamics of actual invasive population size would be beneficial to all branches of ecology (and other biological sciences). A simple and intuitive hierarchical matrix model, motivated by both IDE and PDE approaches, is proposed for modeling invasions while addressing the discrete nature of existing data as well as uncertainty associated with the probability of detection. The non-linear dynamics between discrete time points are modeled using separable density-dependent growth and dispersal components. Additionally, the importance of accomodating potentially differing dispersal rates is illustrated.

Finally, Chapter 5 addresses the problem of modeling invasions and other propagating phenomena (e.g., epidemics) on partitioned domains when available data has only binary support. The dynamics of a binary spatio-temporal process are estimated in terms of probabilistic cellular automata. A simple set of probabilistic neighborhood-based transition rules allows for very flexible space-time behavior including anisotropy and non-stationarity. Additionally, a model is proposed that utilizes covariates to create an environmental preference field, throughout which the phenomenon will propagate from areas of low preference to ar-

eas of high preference. An attraction model, motivated by a partial differential equation in the form of a set of directional gradient fields, transforms the covariate-based preference field into parameters governing the direction and magnitude of dispersal.

2 SPATIO-TEMPORAL FOREST GROWTH DYNAMICS

2.1 Introduction

Due to a steady decline in shortleaf pine (*Pinus echinata*) prevalence since European settlement along the Northwest border of its natural range (Southern Missouri, Arkansas, and Eastern Oklahoma), the species has garnered much attention as it is important both economically and ecologically (Stambaugh 2001; Hamilton 2003). It has been demonstrated that growth of shortleaf pine in this region is associated with climatic variables, especially Palmer Drought Severity Index (PDSI; Stambaugh and Guyette 2004). This does not come as a surprise since historical values of PDSI are reconstructed partially through the knowledge of tree growth during those periods (Fritts 1976). It is particularly interesting however, that Stambaugh and Guyette (2004) found the relationship between shortleaf pine growth rates and PDSI changing over time. Due to major land use changes and fire disturbance regimes in the region before and after 1880 (Cunningham and Hauser 1989; Guyette et al. 2002), it is speculated that the trees may be responding to some interaction between climatic and non-climatic factors. Information about the dynamics of growth in each one of those settings could be useful for managers wishing to promote active management of this species in the area. Shortleaf pine tree-ring data can offer insight into the changes in growth dynamics but presents numerous challenges for statistical modeling.

Dendrochronology is by no means a new science. Researchers have long been interested in the growth rates of trees over time and their relationship with other environmental and biotic variables (Fritts 1976). Dendrochronologists learn about tree growth mainly by collecting, examining, and analyzing cores from trees (Stokes and Smiley 1996). These cores

are quantitatively measured based on the amount of seasonal growth evidenced by sections of denser wood (i.e., tree-rings).

Data constructed from such measurements can be thought of statistically as an individual time-series for each tree, however, generally such data are combined into one time-series representing multiple trees of the same species in a given area. These time-series are termed "chronologies." Chronologies are thought to represent the general growth dynamics of the species in a given area over very large time domains (generally hundreds of years). Therefore, they make excellent sources of data for analyses relating tree growth to climate and other environmental factors. Of course, the chronologies are constructed in the presence of many different sources of error. Measurement error in the field, measurement error in the laboratory, preprocessing error (construction of the chronologies from individual measurements), as well as spatial auto-correlation are all possible sources of uncertainty affecting the quantitative analysis of these data.

In this chapter, a hierarchical Bayesian state-space model is implemented to estimate and assess the changes in spatio-temporal growth dynamics of shortleaf pine over a large regional area. Readily (and publicly) available dendrochronological and climatic data is used to inform the model. The hierarchical framework appropriately accounts for the numerous sources of uncertainty at various levels and provides a natural way to intuitively model the process in a non-linear dynamical fashion. It is important to note two things; first, that this method is quite general and could be applied to other ecological and environmental processes (Beliner et al. 2000; Wikle and Royle 2005) and, secondly, that other approaches could be taken to address this problem. The methods presented here were employed because

they are suited to the specific scientific application and more generally because they are underutilized in an ecological setting (owing much more popularity to atmospheric and environmental sciences).

2.2 Material and Methods

Eight shortleaf pine chronologies were selected from 3 states (Figure 1) based on the location, time period they cover, and the type of standardization applied. Location selection was limited mostly to the Ozark central hardwoods region because of the implications for specific changes in growth and management there (Stambaugh and Guyette 2004). Temporal selection was based on both the length of time period and coverage of the anthropogenically induced changepoint (year 1880). Both time criteria are critical because limited data will fail to adequately describe the dynamics in this setting. Specifically, tree-ring data from 1779 to 1977 were extracted from the 8 chronologies to yield a vector time-series of dimension $n \times T$ where $n = 8$ and $T = 199$.

Standard chronologies were selected because the standardization process involves detrending, indexing, and removal of effects from endogenous stand disturbances although they are not processed using autoregressive modeling (Cook and Holmes 1986) as this would interfere with the estimation of dynamics over time. Generally the raw data are such that values (unitless tree-ring indices) over 1,000 refer to years where trees grew more than average and under 1,000 corresponds to years where tree growth was below average. Since the dynamics are of primary interest, the tree-ring data were arbitrarily standardized for convenience (i.e., subtracted the mean and divided by the standard deviation). Table 1 provides the location names and data contributors of the selected shortleaf pine chronologies

meeting the above criteria.

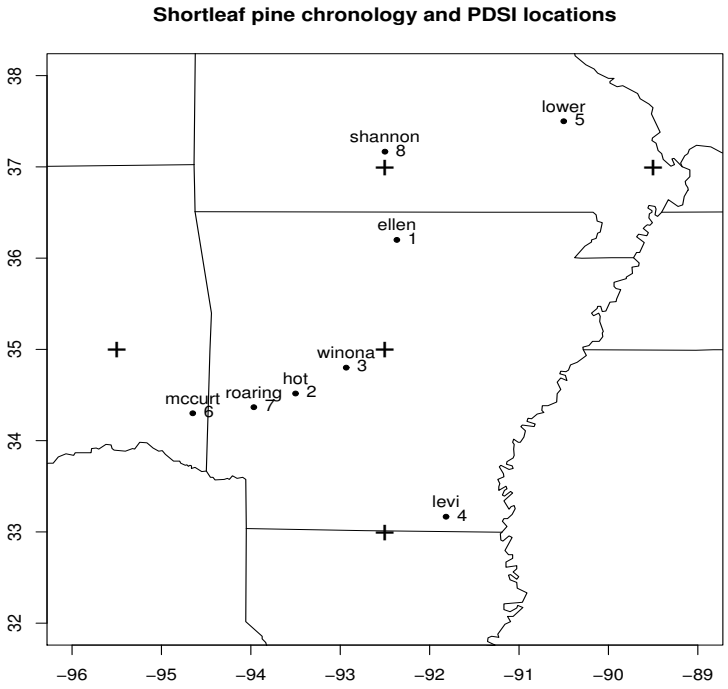


Figure 1: Locations of *Pinus echinata* stands from which chronologies were constructed (points). Numeric values represent location in state vector. Pluses (+) denote PDSI locations.

Table 1: Shortleaf pine chronology locations and contributors; order corresponds to state vector location. Compliments of the International Tree-Ring Data Bank (see references).

Name	**label**	**State**	**Contributor(s)**
Ellen Cockran Hollow	ellen	Arkansas	D. Stahle
Hot Springs	hot	Arkansas	D. Stahle et al.
Lake Winona Natural Area	winona	Arkansas	D. Stahle and G. Hawks
Levi Wilcoxon	levi	Arkansas	D. Stahle and G. Hawks
Lower Rock Creek	lower	Missouri	D. Duvick
McCurtain County Wilderness	mccurt	Oklahoma	D. Stahle et al.
Roaring Branch	roaring	Arkansas	D. Stahle et al.
Shannon County	shannon	Missouri	R. Guyette

Palmer Drought Severity Index (PDSI, Cook et al. 1999) was used as a covariate in the model because of the demonstrated correlation at various times with shortleaf pine growth. A total of 5 PDSI sites spanning the region occupied by the 8 chronology sites were selected (Figure 1) for the same time period. Values were averaged over space yielding a regional PDSI covariate.

Empirical Orthogonal Function (EOF) analysis (or Principal Components Analysis in the discrete setting considered here) has been used extensively in meteorological and atmospheric science to assess dominant spatial structures and their time-varying nature as well as to reduce the dimensionality of large space-time data sets common to the discipline (Preisendorfer 1988). In this setting EOF analysis can be employed to determine how the principal spatial structures in shortleaf pine growth evolve through time. This approach is well known for its optimality properties related to minimization of truncation error (Papoulis and Pillai 2002) and could be motivated more directly as a model for the spatial component of shortleaf pine growth. However, it is somewhat arbitrarily employed here as a means with which to define a lower-dimensional latent process within the ecological system of interest and therefore the interested reader is referred to Ash (1975) for details.

Consider the following matrix formulation: Let $\mathbf{Z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_t, \ldots, \mathbf{z}_T\}$ be the data matrix where each $\mathbf{z}_t$ is an $n \times 1$ vector, made up of tree-ring indices at time t for each spatial location i, where $i = 1, \ldots, n$. Then let the $n \times n$ sample lag-zero (in time) spatial covariance matrix, $\mathbf{C}_0^Z$, be decomposed using the spectral decomposition theorem:

$$\mathbf{C}_0^Z = \tilde{\mathbf{\Phi}} \tilde{\mathbf{\Lambda}} \tilde{\mathbf{\Phi}}', \tag{2.1}$$

where $\tilde{\mathbf{\Phi}}$ is composed of the eigenvectors ($\tilde{\boldsymbol{\phi}}_i$) and $\tilde{\mathbf{\Lambda}}$ is a diagonal matrix with the eigenvalues on the diagonal (i.e., $\tilde{\mathbf{\Lambda}} = diag(\tilde{\lambda}_i)$ for $i = 1, \ldots, n$ and $\tilde{\lambda}_1 > \cdots > \tilde{\lambda}_n$).

The eigenvectors (or principal component loadings) can then each be viewed as spatial maps representing spatial structures of the system in decreasing order of importance corresponding to the magnitude of the eigenvalues (which are decreasing and real since $\mathbf{C}_0^Z$ is positive definite). The i^{th} spatial structure's ($\tilde{\boldsymbol{\phi}}_i$) relative importance to the system at time t can be examined by assessing the time-series resulting from a projection of the data onto the set of basis functions. In essence, EOF analysis is about assessing dominant spatial patterns in the data as well as the magnitude and frequency of their reoccurrence over time.

The expression in (2.1) also provides a means of reducing the dimensionality of the data. By definition, the first l eigenvectors, eigenvalues, and PC score time-series are guaranteed to capture $100 \times (P_l)$ percent of the variability in the system, where $P_l = \frac{\sum_{i=1}^{l} \lambda_i}{\sum_{j=1}^{n} \lambda_j}$. Therefore, the sample covariance matrix in (2.1) can be approximated by,

$$\mathbf{C}_0^Z \approx \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}', \tag{2.2}$$

where $\mathbf{\Phi}$ is $n \times l$ and $\mathbf{\Lambda} = diag(\lambda_i)$, for $i = 1, \ldots, l$.

Thus, an EOF decomposition of the tree-ring data would allow dominant components of variation in the space-time system to be assessed as well as allow a smaller dimensional latent dynamical system to be modeled. This can greatly increase computational efficiency while retaining the information contributing significantly to the variability in the system (Wikle and Cressie 1999; Beliner et al. 2000). It is important to note that atmospheric datasets are often very high dimensional, whereas most ecological datasets (due to lack of scope or inaccessibility) are possibly orders of magnitude lower dimensional. In these cases,

the resulting lower dimensional system can still be advantageous as it is often decorrelated in addition to acting as a filter due to the exclusion of components contributing little to the overall variability.

The data consist of an $n \times T$ matrix $\mathbf{Z}$ with columns corresponding to time and rows corresponding to spatial locations (i.e., different chronologies) and is modeled as a latent dynamical process ($\mathbf{y}_t$) plus error, yielding the following data model:

$$\begin{aligned} \mathbf{z}_t &= \mathbf{y}_t + \tilde{\boldsymbol{\nu}}_t + \boldsymbol{\nu}_t, \\ &= \boldsymbol{\Phi}'\mathbf{a}_t + \boldsymbol{\varepsilon}_t, \text{ where } \boldsymbol{\varepsilon}_t \sim N(0, \mathbf{R}), \text{ for } t = 1, \ldots, T, \end{aligned} \tag{2.3}$$

where $\mathbf{y}_t \equiv \boldsymbol{\Phi}'\mathbf{a}_t$, $\tilde{\boldsymbol{\nu}}_t$ is a non-dynamical spatio-temporal process, and $\boldsymbol{\nu}_t$ corresponds to measurement error. The compact matrix notation as in (2.3) is used for convenience throughout the remainder of the chapter, however, summation notation illustrates the simplicity of the model:

$$y_{j,t} = \sum_{i=1}^{l} \phi_{i,j} a_{i,t} + \varepsilon_{j,t} \text{ for } t = 1, \ldots, T \text{ and } j = 1, \ldots, n,$$

where $\mathbf{y}_t = [y_{1,t}, \ldots, y_{n,t}]'$, $\boldsymbol{\Phi}' = [(\phi_{i,j})]_{n \times l}, \forall i, j$, and $\mathbf{a}_t = [a_{1,t}, \ldots, a_{l,t}]'$. The $\phi_{i,j}$ are assumed known and the $a_{i,t}$ will be modeled hierarchically. This clarifies the separation between temporal and spatial components (in $\mathbf{y}_t$).

The underlying (i.e., latent) dynamical system ($\mathbf{y}_t$) is assumed to lie on some lower-dimensional manifold, relative to the observed data. That is, the dynamics of shortleaf pine growth in this region can be adequately described by fewer than $n = 8$ components, say l components, where $l < n$, given by $\mathbf{a}_t$. Therefore, $\tilde{\boldsymbol{\nu}}_t$ can be thought of as an error associated with the truncation of basis functions from the spectral decomposition (e.g.,

Wikle and Cressie 1999) and $\boldsymbol{\nu}_t$ corresponds to measurement error. Inference on the non-dynamical error components is not of primary interest here, thus a general specification of $\boldsymbol{\varepsilon}_t$ can be used to absorb both types of error and simplify the model. The process model in the reduced-dimensional space can then be written as,

$$\mathbf{a}_t = \mathbf{H}_t \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, \text{ where } \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}), \text{ for } t = 1, \ldots, T. \tag{2.4}$$

The $l \times l$ matrix $\mathbf{H}_t$ can be thought of as a propagator matrix for the dynamical system at time t. Although this VAR (vector autoregressive) process is generally considered to be linear, in the case where the propagator matrix is time-varying it provides a simple way to build non-linear dynamics into the model. The process model in (2.4) is very general and would be difficult to estimate if $\mathbf{H}$ is allowed to vary at all times. However, to study the dynamics of the system before and after year 1880, as well as how the dynamics change with drought conditions, the model can be simplified as follows (Wikle and Royle 2005):

$$\mathbf{a}_t = \begin{cases} \mathbf{H}_1 \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, & \text{if } t \leq t^*, x_t < x^* \\ \mathbf{H}_2 \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, & \text{if } t \leq t^*, x_t \geq x^* \\ \mathbf{H}_3 \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, & \text{if } t > t^*, x_t < x^* \\ \mathbf{H}_4 \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, & \text{if } t > t^*, x_t \geq x^*, \end{cases} \tag{2.5}$$

where $\mathbf{x} = \{x_1, \ldots, x_T\}'$ is the PDSI covariate (averaged over space) and x^* is a threshold above which indicates wet years and below which indicates dry years (in this case, $x^* = 0$). Recall that the year 1880 is a temporal changepoint, thus for the data considered here: $t^* = 101$. In principle, one could allow t^* and x^* to be random variables in the hierarhical model, but that is not necessary here since inference is concerned with specific, known, values for these parameters.

The model in (2.5), which involves estimation of four fully random propagator matrices (along with all the other random parameters), is intuitive and easy to implement in a fully Bayesian setting. However, in this case where only a total of $T = 199$ times exist from which to estimate $4 \times l \times l$ parameters, may lead to an overfitted model. Since the focus here is examination of the propagator matrices, the highest reasonable data to parameter ratio (i.e., parsimonious model) is desirable. Therefore, consider the following reparameterization of the data and process models:

$$\mathbf{z}_t = \mathbf{y}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \mathbf{R}), \tag{2.6}$$

$$\mathbf{y}_t = \begin{cases} \boldsymbol{\Phi}_1 \mathbf{a}_t, & \text{if } \ t \leq t^* \\ \boldsymbol{\Phi}_2 \mathbf{a}_t, & \text{if } \ t > t^*, \end{cases} \tag{2.7}$$

$$\mathbf{a}_t = \begin{cases} \mathbf{H}_1 \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, & \text{if, } \ x_t < x^*, \ \ \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t) \\ \mathbf{H}_2 \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, & \text{if, } \ x_t \geq x^*, \ \ \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t), \end{cases} \tag{2.8}$$

where the basis functions $\boldsymbol{\Phi}_1$ and $\boldsymbol{\Phi}_2$ are found from the same decomposition as in (2.2) using $\mathbf{C}_0^1$ and $\mathbf{C}_0^2$, the sample lag-zero spatial covariance matrices from $\mathbf{z}_t$ with $t = 1, \ldots, t^* - 1$ and $t = t^*, \ldots, T$ respectively. This allows for incorporation of the temporal changepoint in the model while only having to estimate half as many propagator parameters. It is easily shown that model (2.7) and (2.8) results in an implicit model similar to that of (2.5) in the space of $\mathbf{y}_t$.

Consider the process model (2.8) and recall that,

$$\mathbf{y}_t = \begin{cases} \boldsymbol{\Phi}_1 \mathbf{a}_t, & \text{if } \ t \leq t^* \\ \boldsymbol{\Phi}_2 \mathbf{a}_t, & \text{if } \ t > t^*, \end{cases}$$

thus implying,

$$\Rightarrow \mathbf{a}_t = \begin{cases} \boldsymbol{\Phi}_1' \mathbf{y}_t, & \text{if } \ t \leq t^* \\ \boldsymbol{\Phi}_2' \mathbf{y}_t, & \text{if } \ t > t^*. \end{cases}$$

Then, substitution of $\boldsymbol{\Phi}_i' \mathbf{y}_t$ into the LHS of model 2.8 yields,

$$\Rightarrow \begin{cases} \boldsymbol{\Phi}_1' \mathbf{y}_t = \begin{cases} \mathbf{H}_1 \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, & \text{if}, \ \ x_t < x^*, \ \ t \leq t^* \\ \mathbf{H}_2 \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, & \text{if}, \ \ x_t \geq x^*, \ \ t \leq t^* \end{cases} \\ \boldsymbol{\Phi}_2' \mathbf{y}_t = \begin{cases} \mathbf{H}_1 \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, & \text{if}, \ \ x_t < x^*, \ \ t > t^* \\ \mathbf{H}_2 \mathbf{a}_{t-1} + \boldsymbol{\eta}_t, & \text{if}, \ \ x_t \geq x^*, \ \ t > t^*, \end{cases} \end{cases}$$

and pre-multiplication of both sides by Φ_i as well as substitution of $\Phi_i' \mathbf{y}_t$ into the RHS gives,

$$\Rightarrow \mathbf{y}_t = \begin{cases} \boldsymbol{\Phi}_1 \mathbf{H}_1 \boldsymbol{\Phi}_1' \mathbf{y}_{t-1} + \boldsymbol{\Phi}_1 \boldsymbol{\eta}_t, & \text{if}, \ \ x_t < x^*, \ \ t \leq t^* \\ \boldsymbol{\Phi}_1 \mathbf{H}_2 \boldsymbol{\Phi}_1' \mathbf{y}_{t-1} + \boldsymbol{\Phi}_1 \boldsymbol{\eta}_t, & \text{if}, \ \ x_t \geq x^*, \ \ t \leq t^* \\ \boldsymbol{\Phi}_2 \mathbf{H}_1 \boldsymbol{\Phi}_1' \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{\eta}_t, & \text{if}, \ \ x_t < x^*, \ \ t = t^* + 1 \\ \boldsymbol{\Phi}_2 \mathbf{H}_2 \boldsymbol{\Phi}_1' \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{\eta}_t, & \text{if}, \ \ x_t \geq x^*, \ \ t = t^* + 1 \\ \boldsymbol{\Phi}_2 \mathbf{H}_1 \boldsymbol{\Phi}_2' \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{\eta}_t, & \text{if}, \ \ x_t < x^*, \ \ t > t^* + 1 \\ \boldsymbol{\Phi}_2 \mathbf{H}_2 \boldsymbol{\Phi}_2' \mathbf{y}_{t-1} + \boldsymbol{\Phi}_2 \boldsymbol{\eta}_t, & \text{if}, \ \ x_t \geq x^*, \ \ t > t^* + 1. \end{cases}$$

Now, redefinition of the implicit propagator matrices of the system yields,

$$\Rightarrow \mathbf{y}_t = \begin{cases} \mathbf{B}_1 \mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if}, \ \ x_t < x^*, \ \ t \leq t^* \\ \mathbf{B}_2 \mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if}, \ \ x_t \geq x^*, \ \ t \leq t^* \\ \mathbf{B}_3^* \mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if}, \ \ x_t < x^*, \ \ t = t^* + 1 \\ \mathbf{B}_4^* \mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if}, \ \ x_t \geq x^*, \ \ t = t^* + 1 \\ \mathbf{B}_3 \mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if}, \ \ x_t < x^*, \ \ t > t^* + 1 \\ \mathbf{B}_4 \mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if}, \ \ x_t \geq x^*, \ \ t > t^* + 1. \end{cases}$$

where (recall that $\mathbf{Q}_i = \sigma_Q^2 \boldsymbol{\Lambda}_i$),

$$\boldsymbol{\xi}_t = \begin{cases} \boldsymbol{\Phi}_1 \boldsymbol{\eta}_t, & \text{if}, \ \ t \leq t^* \\ \boldsymbol{\Phi}_2 \boldsymbol{\eta}_t, & \text{if}, \ \ t > t^* \end{cases} \Rightarrow \boldsymbol{\xi}_t \sim \begin{cases} N(\mathbf{0}, \boldsymbol{\Phi}_1 \mathbf{Q}_1 \boldsymbol{\Phi}_1'), & \text{if}, \ \ t \leq t^* \\ N(\mathbf{0}, \boldsymbol{\Phi}_2 \mathbf{Q}_2 \boldsymbol{\Phi}_2'), & \text{if}, \ \ t > t^* \end{cases} \Rightarrow$$

$$\Rightarrow \boldsymbol{\xi}_t = \begin{cases} N(\mathbf{0}, \sigma_Q^2 \boldsymbol{\Phi}_\mathbf{1} \boldsymbol{\Lambda}_\mathbf{1} \boldsymbol{\Phi}_\mathbf{1}'), & \text{if}, \ \ t \leq t^* \\ N(\mathbf{0}, \sigma_Q^2 \boldsymbol{\Phi}_\mathbf{2} \boldsymbol{\Lambda}_\mathbf{2} \boldsymbol{\Phi}_\mathbf{2}'), & \text{if}, \ \ t > t^* \end{cases} \Rightarrow \boldsymbol{\xi}_t \sim \begin{cases} N(\mathbf{0}, \sigma_Q^2 \mathbf{C}_0^1), & \text{if}, \ \ t \leq t^* \\ N(\mathbf{0}, \sigma_Q^2 \mathbf{C}_0^2), & \text{if}, \ \ t > t^*. \end{cases}$$

Since only the dynamics before and after the change point are of interest, then by allowing for two spatio-temporal covariance matrices in the data model, an implicit model on the $\mathbf{y}_t$'s (for $t = 1, \ldots, t^*, t^* + 2, \ldots, T$) with 4 (or more) propagator matrices results:

$$\Rightarrow \mathbf{y}_t = \begin{cases} \mathbf{B}_1\mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if, } \; x_t < x^*, \;\; t \leq t^* \\ \mathbf{B}_2\mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if, } \; x_t \geq x^*, \;\; t \leq t^* \\ \mathbf{B}_3\mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if, } \; x_t < x^*, \;\; t > t^* \\ \mathbf{B}_4\mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if, } \; x_t \geq x^*, \;\; t > t^*. \end{cases} \qquad \square$$

Now, consider the parameter models,

$$\mathbf{h}_1 = vec(\mathbf{H}_1) \sim N(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1), \tag{2.9}$$

$$\mathbf{h}_2 = vec(\mathbf{H}_2) \sim N(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_2), \tag{2.10}$$

$$\mathbf{R}^{-1} \sim Wish((v_R\mathbf{C}_R)^{-1}, v_R), \tag{2.11}$$

and let $\mathbf{Q}_t$, the latent process error covariance, be defined as,

$$\mathbf{Q}_t = \sigma_Q^2\boldsymbol{\Lambda}_t, \tag{2.12}$$

$$\boldsymbol{\Lambda}_t = \begin{cases} \boldsymbol{\Lambda}_1, & \text{if } \; t \leq t^* \\ \boldsymbol{\Lambda}_2, & \text{if } \; t > t^*, \end{cases} \tag{2.13}$$

$$\sigma_Q^2 \sim IG(r_Q, q_Q), \tag{2.14}$$

where $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ are the diagonal eigenvalue matrices resulting from the spectral decomposition in (2.2). The vector $\mathbf{a}_0$ is the prior for the state vector at the time just before the data were observed and is distributed as $\mathbf{a}_0 \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0 = \sigma_0^2\boldsymbol{\Lambda}_1)$. Having little *a priori* intuition about the structure of the data model covariance ($\mathbf{R}$) indicates the need for a general prior such as the Wishart offers. Rather than impose spatial structure on the model, this specification accounts for limited knowledge about the covariance while allowing for the

possibility of spatial error structure. Additionally, if there was reason to expect a different error structure before and after t^*, $\mathbf{R}$ and $\mathbf{R}^*$ could be specified separately. Although here the structure in $\mathbf{R}$ is assumed to be similar for the different time periods, a large discrepancy could affect the estimation of other model parameters. This issue would become especially evident in situations where the dimensionality of the latent system was much less than that of the data. In such cases, it is possible that $\mathbf{R}$ could contain significantly more structure resulting from the truncation of the latent system for the two time periods and would likely be different. In the specific application considered here, the trucation is less dramatic and thus the effects on $\mathbf{R}$ are assumed to be minimal.

Ultimately, inference on the posterior distributions of the random parameters is desired. i.e.,

$$[\mathbf{a}_{t=0,\ldots,T}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{R}, \sigma_Q^2 | \mathbf{Z}, \mathbf{x}] \propto \prod_{t=1}^{T} [\mathbf{z}_t | \mathbf{a}_t, \mathbf{R}] \prod_{t \in \{t | x_t < x^*\}} [\mathbf{a}_t | \mathbf{a}_{t-1}, \mathbf{H}_1, \sigma_Q^2] \times \\ \times \prod_{t \in \{t | x_t \geq x^*\}} [\mathbf{a}_t | \mathbf{a}_{t-1}, \mathbf{H}_2, \sigma_Q^2] [\mathbf{a}_0, \mathbf{H}_1, \mathbf{H}_2, \mathbf{R}, \sigma_Q^2]. \quad (2.15)$$

Making the assumption that the priors are independent,

$$[\mathbf{a}_0, \mathbf{H}_1, \mathbf{H}_2, \mathbf{R}, \sigma_Q^2] \propto [\mathbf{a}_0][\mathbf{H}_1][\mathbf{H}_2][\mathbf{R}][\sigma_Q^2].$$

Conjugate model specification and the assumption of independent priors allow for very efficient MCMC sampling. Although it may not be appropriate in all situations, the manner in which it is used here allows for possible model misspecifications. More complex non-linear and non-conjugate hierarchical models are the focus of ongoing research.

Note that the joint posterior distribution and the relevant marginal distributions of interest are analytically intractable. Consequently, a Gibbs sampler is used to obtain samples

from the posterior distribution. The full-conditional distributions used in the Gibbs sampler are given in the Appendix (Section 2.6).

A useful method for assessing the structure of VAR dynamical systems is principal oscillation pattern (POP) analysis (Von Storch et al. 1995). As with EOF analysis, there are other ways to introduce POP analysis (i.e., as a model for the phenomena of interest) and although the continued use of compact matrix notation in what follows may seem terse, POP analysis in this setting is simply a matrix decomposition yielding very rich information about the evolution of ecological spatial patterns through time.

POP analysis involves a similar spectral decomposition as in (2.1) except now the propagator matrix (say $\mathbf{B}_k$, where k corresponds to different regimes) is decomposed rather than the covariance matrix ($\mathbf{C}_0$). This cannot be approached in the same manner however, because $\mathbf{B}_k$, although square, may be non-symmetric ($\mathbf{B}_k \neq \mathbf{B}_k'$). In this setting, the characteristic equation $|\mathbf{B}_k - \lambda\mathbf{I}|$=0 will have n roots (assuming $\mathbf{B}_k$ is $n \times n$), some of which may be complex. Consider the spectral decomposition:

$$\mathbf{B}_k = \mathbf{WDG}^*$$

where * denotes the Hermitian transpose (i.e., $\mathbf{G}^* = \bar{\mathbf{W}}^{-1}$, the inverse complex conjugate as in Caswell (2001)) and $\mathbf{W}$ is comprised of the right singular vectors $\{\mathbf{w}_1, \ldots, \mathbf{w}_n\}$ and $\mathbf{G}^*$ is comprised of the left singular vectors $\{\mathbf{g}_1^*, \ldots, \mathbf{g}_n^*\}$, $\mathbf{D} = diag(\delta_i)$ is diagonal, and $\mathbf{B}_k$ is the propagator matrix of the process: $\mathbf{y}_t = \mathbf{B}_k\mathbf{y}_{t-1} + \boldsymbol{\xi}_t$.

Inspection of the decomposition components can reveal much information about the stability and dynamical structure of the system (Von Storch et al. 1995). The vectors $\mathbf{w}_i$

are called the principal oscillation patterns (or system normal modes), while the elements of the vector $\boldsymbol{\alpha}_t$ are called the POP coefficients, where $\boldsymbol{\alpha}_t = \mathbf{G}^* \mathbf{y}_t$. In the spatio-temporal setting, $\mathbf{w}_j$ is a spatial map. The importance of this map to $\mathbf{y}_t$ is described by the magnitude of $\alpha_{j,t}$ when δ_j is real. In the event that δ_j is complex (i.e., $\delta_j = \delta_j^{\Re} + i\delta_j^{\Im}$, where $\Re$ and $\Im$ denote the real and imaginary components of the complex number respectively), δ_j has a sister root that is the complex conjugate of δ_j, say δ_k. Also note the identity: $\delta_j = \gamma_j e^{i\omega_j}$, where γ and ω are such that $\delta_j^{\Re} = \gamma_j \cos(\omega_j)$ and $\delta_j^{\Im} = \gamma_j \sin(\omega_j)$. Thus $\alpha_{j,t}$ evolves as a damped spiral in the complex plane with damping rate γ_j and frequency ω_j (when δ_j is complex and less than one in modulus).

The POP coefficients for the complex conjugate pair can be combined together and thus if $\mathbf{w}_j \equiv \mathbf{w}_j^{\Re} + i\mathbf{w}_j^{\Im}$, then the spatial pattern evolves throughout the period of oscillation as: $\ldots \rightarrow \mathbf{w}_j^{\Re} \rightarrow -\mathbf{w}_j^{\Im} \rightarrow -\mathbf{w}_j^{\Re} \rightarrow \mathbf{w}_j^{\Im} \rightarrow \mathbf{w}_j^{\Re} \rightarrow \cdots$ with a time interval of $\frac{\pi}{2\theta}$ between each successive pattern, where $\theta = \tan^{-1}(\frac{\delta_j^{\Re}}{\delta_j^{\Im}})$. Additionally, if $|\delta_j| < 1$, the amplitude decreases exponentially and can be characterized by the "e-folding time" (the time needed to reduce the initial amplitude by e^1), $\tau_j = -1/\log(\gamma_j)$. Intuitively, when there exists a propagating wave in the state process, $\mathbf{w}^{\Im}$ is just a translated (in space) version of $\mathbf{w}^{\Re}$.

In the specific application considered here, there are two propagator matrices in the reduced dimension process model (2.8). A POP analysis on these propagator matrices would yield information about the dynamics in EOF space, not in a spatial context. To assess the changes in dynamics from a more intuitive perspective, consider a POP analysis on the implicit (full dimensional) process model:

$$\mathbf{y}_t = \begin{cases} \mathbf{B}_1\mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if, } x_t < x^*, \ t \leq t^* \\ \mathbf{B}_2\mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if, } x_t \geq x^*, \ t \leq t^* \\ \mathbf{B}_3\mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if, } x_t < x^*, \ t > t^* \\ \mathbf{B}_4\mathbf{y}_{t-1} + \boldsymbol{\xi}_t, & \text{if, } x_t \geq x^*, \ t > t^*, \end{cases}$$

where, $\boldsymbol{\xi}_t = \boldsymbol{\Phi}_i\boldsymbol{\eta}_t$ and the $n \times n$ propagator matrices are defined as,

$$\mathbf{B}_1 \equiv \boldsymbol{\Phi}_1\mathbf{H}_1\boldsymbol{\Phi}_1',$$

$$\mathbf{B}_2 \equiv \boldsymbol{\Phi}_1\mathbf{H}_2\boldsymbol{\Phi}_1',$$

$$\mathbf{B}_3 \equiv \boldsymbol{\Phi}_2\mathbf{H}_1\boldsymbol{\Phi}_2',$$

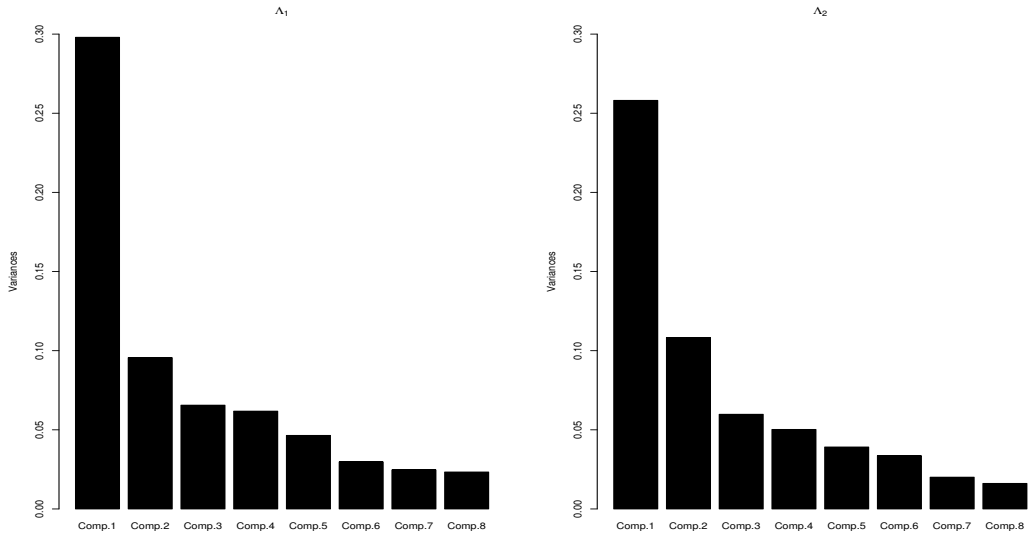
$$\mathbf{B}_4 \equiv \boldsymbol{\Phi}_2\mathbf{H}_2\boldsymbol{\Phi}_2'.$$

Now, each propagator $\mathbf{B}_i$ will only yield l significant POP's due to the structure induced by the pre- and post-multiplication of the orthogonal basis functions. That is, the underlying dynamics exist on an l-dimensional manifold, but the expression of these dynamics in physical space is different for each period due to the different basis functions (projections).

2.3 Results

Spectral decomposition of the spatio-temporal covariance matrix (2.2) from the data prior to and after the changepoint suggested that the correlation structure was indeed different for the two time periods (using $l = 4$, accounting for $> 80\%$ of the variability in data; scree plots of eigenvalues are in Figure 2). This can be visually assessed by looking at the distribution of off-diagonal elements in the absolute difference of correlation matrices (corresponding to $\mathbf{C}_0^1$ and $\mathbf{C}_0^2$) for the two time periods (Figure 3). In addition, a standard Bartlett's test for differences in covariance matrices confirmed that the two covariance structures are likely different, although this was only used as an exploratory measure for justifying the

need for using two different sets of basis functions. Also of note is the correlation between the PDSI covariate and the principal components resulting from the EOF decompositions. The correlation between the first component and PDSI was approximately 0.5 while the correlation with the other components was less than 0.01 for both early (pre-1880) and late (post-1880) years.



(a) (b)

Figure 2: Scree plots showing the proportion of variance (eigenvalues) explained by the orthogonal components of each decomposition. $\mathbf{\Lambda}_1$ from early years (pre-1880) and $\mathbf{\Lambda}_2$ from later years (post-1880).

The Gibbs sampler was run for 10,000 iterations with a burn-in period of 1,000 iterations and MCMC-based posterior statistics were calculated. Convergence was assessed visually using trace plots and occurred rapidly due in part to multivariate conjugate updates, identifiable model parameters, and relatively simple structure. The error term corresponding to the process model, σ_Q^2, has posterior mean and standard deviation of 0.224 and 0.048,

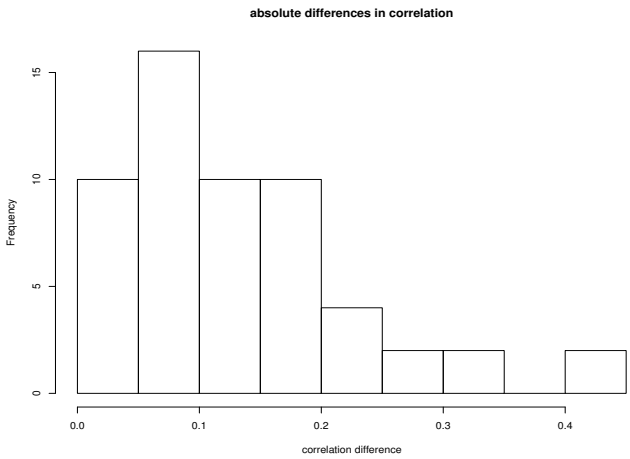


Figure 3: Absolute differences in off-diagonal elements of the correlation matrices pre and post 1880.

respectively.

The posterior distributions for the multivariate and matrix parameters in the model can be displayed in several ways. The posterior distributions for $\mathbf{H}_1$ and $\mathbf{H}_2$ can be viewed simultaneously to assess if (and how) they differ (Figure 4). Consider also an image plot representing the posterior mean of the data model error covariance matrix $\mathbf{R}$ (Figure 5). Also, since $\mathbf{a}_t, t = 1, \ldots, T$ consists of 199 times, it is best viewed as $l = 4$ individual time-series. The temporal dimension ($T = 199$) is too large to provide useful visual information here, thus plots from only a portion of each time-series (years 1827-1927) as a posterior mean with 95% credible intervals are given in Figure 6. Additionally, the posterior means of the implicit propagator matrices ($\mathbf{B}_i$) are depicted as images in Figure 7.

The dynamics of the system ($\mathbf{a}_t$, and hence the implicit process $\mathbf{y}_t$) can be partially assessed by examining the POP's of the posterior mean $\mathbf{B}$ matrices. Table 2 summarizes the resulting eigenvalues from the POP decomposition (note, the eigenvalues for $\mathbf{B}_1, \mathbf{B}_3$ and $\mathbf{B}_2, \mathbf{B}_4$ must be the same since they correspond to the same reduced dimension propagators

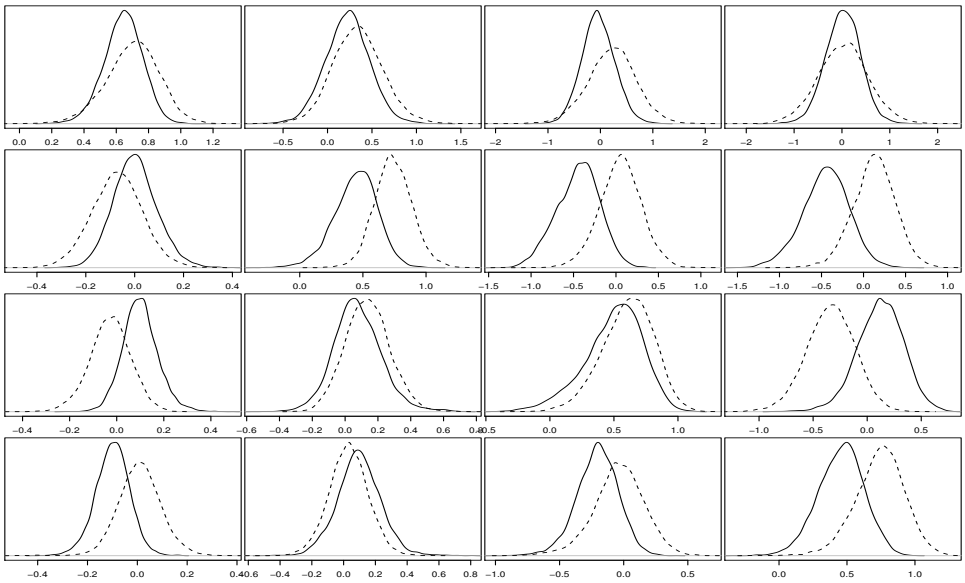


Figure 4: Posterior distributions on an element by element basis for $\mathbf{H}_1$ (solid line, drought) and $\mathbf{H}_2$ (dashed line, non-drought). The arrangement of plots corresponds to the elements of the 4×4 propagator matrices from the process model (e.g., upper left plot represents the posterior distributions of element (1,1) from $\mathbf{H}_1$ and $\mathbf{H}_2$, etc).

$\mathbf{H}_1$ and $\mathbf{H}_2$, respectively). Recall that the POP's can be viewed as spatial maps. Consider first the maps for the two POP's corresponding to real eigenvalues, POP_1 (Figure 8) and POP_4 (Figure 9).

Table 2: Significant eigenvalue summary from $\mathbf{B}$ matrices (complex eigenvalues are in the form $\delta = \delta^{\Re} + i\delta^{\Im}$).

$\mathbf{B}_1, \mathbf{B}_3$			$\mathbf{B}_2, \mathbf{B}_4$		
$\delta^{\Re}$	$\delta^{\Im}$	$\lvert\delta\rvert$	$\delta^{\Re}$	$\delta^{\Im}$	$\lvert\delta\rvert$
0.70	0	0.70	0.81	0	0.81
0.47	0.35	0.59	0.75	0.18	0.77
0.47	-0.35	0.59	0.75	-0.18	0.77
0.39	0	0.39	0.49	0	0.49

Additionally, the complex POP's (corresponding to complex eigenvalues) are presented

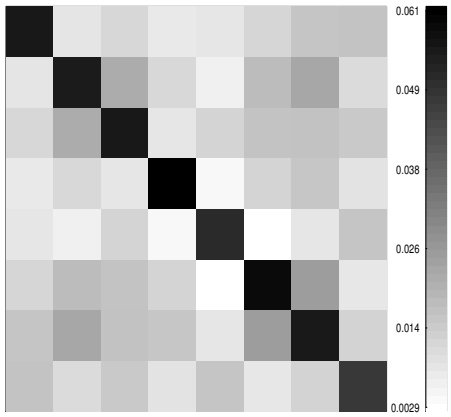


Figure 5: Image representing the posterior mean of the error covariance matrix from the data model. Pixels in the image correspond to elements of the posterior mean matrix (i.e., $E([\mathbf{R}|\mathbf{Z}, \mathbf{x}])$.

as spatial maps and provide some idea of the oscillating patterns. Figure 10 illustrates the oscillating nature of the system through the decomposition of the complex conjugate pair of POP's for each of the propagator matrices ($\mathbf{B}_i$).

The POP decomposition can also yield useful diagnostic measures. Table 3 presents the posterior estimates of the frequencies, periods, and e-folding times. Note that since the eigenvalues are the same, as in Table 2, it is only necessary to present two sets of estimates.

Table 3: Frequencies, periods, and e-folding times from POP analysis of $\mathbf{B}$ matrices. Values correspond to the complex pair of POP's (i.e., POP_2, POP_3).

$\mathbf{B}_1, \mathbf{B}_3$			$\mathbf{B}_2, \mathbf{B}_4$		
$\omega_j \left(\frac{rad}{yr}\right)$	$\frac{2\pi}{\omega_j} \left(\frac{yrs}{cycle}\right)$	τ_j (yr)	$\omega_j \left(\frac{rad}{yr}\right)$	$\frac{2\pi}{\omega_j} \left(\frac{yrs}{cycle}\right)$	τ_j (yr)
0.639	9.83	1.86	0.232	27.1	3.77

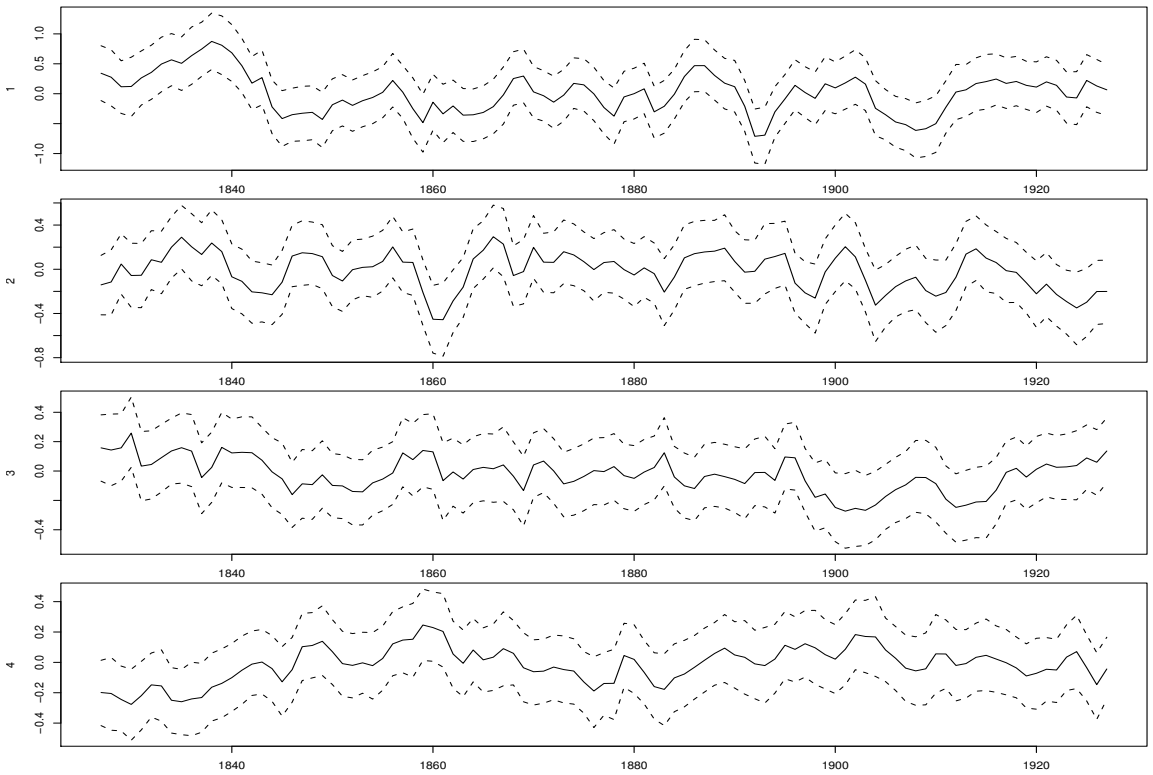
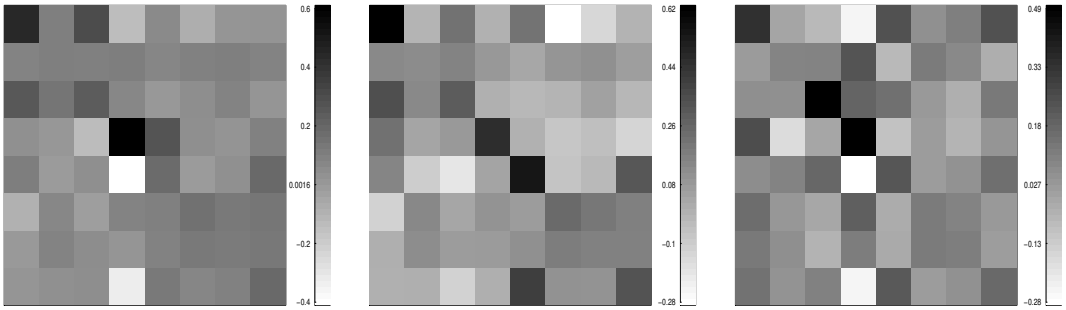


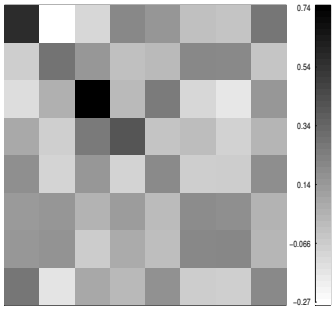
Figure 6: Posterior means of the $\mathbf{a}_t$ process and 95% credible intervals for years 1827-1927.

2.4 Discussion

Initial spectral decomposition of the data matrix ($\mathbf{Z}$) as in (2.2), revealed that $l = 4$ of the EOF's contributed greater than 80% of the variability in the data and provided a sufficient reduction in dimension to accomodate the dynamics of the system. Additionally, it was found that both of the matrices of basis functions in the data model (2.7) had dominant EOF's that were negatively associated with the covariate (PDSI) whereas EOF_2, EOF_3, and EOF_4 had little association with PDSI. This suggests that the most dominant component of the reduced latent process may be somewhat related to a response in growth due to drought. Based on the minimal associations of the remaining EOF's with PDSI these



(a) $\mathbf{B}_1$ (early, dry) (b) $\mathbf{B}_2$ (early, wet) (c) $\mathbf{B}_3$ (late, dry)



(d) $\mathbf{B}_4$ (late, wet)

Figure 7: Images representing the posterior mean of the implicit propagator matrices. Pixels in the images correspond to elements of the posterior mean matrices. Recall that the values in the i^{th} row are weights relating the process at the previous time to the process in the i^{th} location at the current time.

EOF's could correspond more with non-drought related factors (perhaps anthropogenic, biotic, and (or) other climatic factors). Despite any direct associations between the EOF's and PDSI covariate, it is possible for all elements of the process ($\mathbf{a}_t$) to evolve differently under different drought regimes. While other studies have focused on the more direct correlations of climate and tree growth, the emphasis here is on the dynamics of the tree growth over time as it relates to PDSI and anthropogenic influence. Figure 3 illustrates

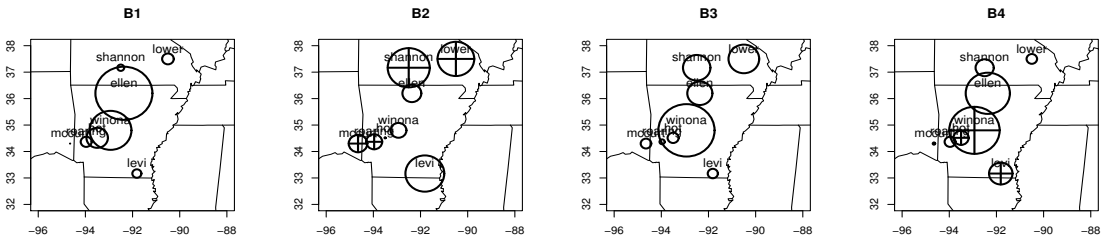


Figure 8: The first normal mode plotted spatially for each of the $\mathbf{B}$ matrices. Open circles denote negative values, while circles with pluses represent positive values; larger circles imply larger magnitude, thus small circles are values near zero. From left to right: (early, dry), (early,wet), (late, dry), and (late, wet) years, respectively.

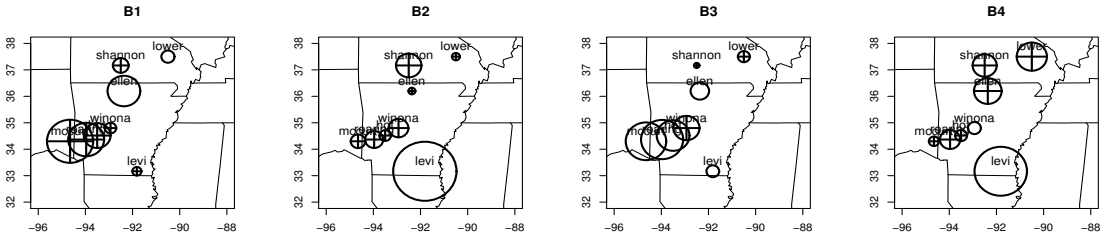


Figure 9: The fourth normal mode plotted spatially for each of the $\mathbf{B}$ matrices. Open circles denote negative values, while circles with pluses represent positive values; larger circles imply larger magnitude, thus small circles are values near zero. From left to right: (early, dry), (early,wet), (late, dry), and (late, wet) years, respectively.

that regardless of EOF interpretation, differences in the lag-zero (in time) spatial correlation structure for the two time periods (i.e., before and after $t^* = 1880$) do exist. This implies that the EOF's will exhibit different behavior for the two time periods as well and indicates a need for two different sets of basis functions.

It is important to note that these sets of basis functions $(\mathbf{\Phi}_1, \mathbf{\Phi}_2)$ are not modeled in this setting. They only provide access to one of many latent systems and other sets of basis functions could be alternatively considered. In this case, they provide not only a means of dimension reduction, but also, basis functions typically act as decorrolaters, therefore a

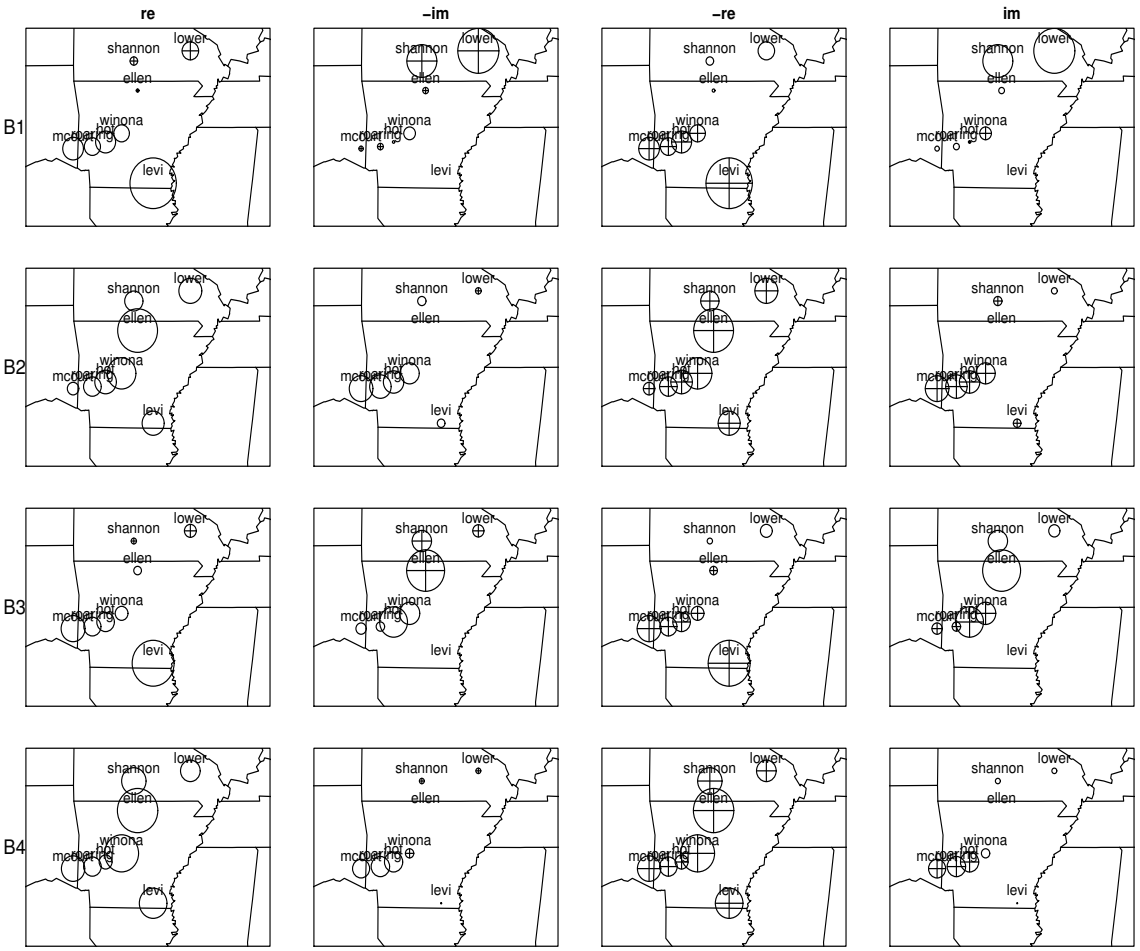


Figure 10: The second (and third) normal mode plotted spatially for each of the **B** matrices. Open circles denote negative values, while circles with pluses represent positive values; larger circles imply larger magnitude, thus small circles are values near zero. Patterns evolve from left to right.

simple error structure ($\boldsymbol{\eta}_t$) on the latent process is reasonable. In general, projection of the data on a truncated EOF expansion acts as a filter and removes some of the noise in the data thereby magnifying the POP signal in the process model making it more detectable (Von Storch et al. 1995).

Inspection of the posterior mean of the error covariance matrix ($\mathbf{R}$) from the data model (Figure 5) indicates that spatial auto-correlation in the data is being accounted for,

as the posterior mean of ($\mathbf{R}$) contains spatial structure, some of which is likely due to the truncation error from (2.3) (Wikle and Cressie 1999; Beliner et al. 2000). Notice that data locations near each other (e.g., locations 2 and 3) correspond to covariances (in $\mathbf{R}$, Figure 5) larger in magnitude than covariances corresponding to locations that are farther apart (e.g., locations 4 and 5). The spatial structure allowed by the Wishart on $\mathbf{R}$ suggests two possible alternate specifications for the data model error covariance (to be considered elsewhere).

1. Parameterize $\mathbf{R}$ such that, $\mathbf{R} = \sigma_{\varepsilon}^2 R(\theta)$ where $R(\theta) = [(r_{i,j})] = exp(-||d_{i,j}||/\theta)$, a typical spatial covariance specification, requiring the estimation of much fewer parameters, hence saving degrees of freedom for estimation of the propagator matrices.

2. Partition the error covariance from the data model into two components, one spatial and informative, the other general and vague. This would allow estimations of the process to be made at spatial locations other than where the data were collected.

The posterior mean and 95% credible interval for the $\mathbf{a}_t$ process in Figure 6 illustrates the uncertainty associated with the latent dynamical system. Figure 4 illustrates the differences between the estimated propagator matrices that control the evolution of the latent process ($\mathbf{a}_t$). The differences in the posterior distributions of these propagators (i.e., $\mathbf{H}_1$ and $\mathbf{H}_2$) implies that the process evolves differently under drought conditions than it does during wetter years. A closer examination of the elements in $\mathbf{H}_1$ and $\mathbf{H}_2$ can reveal similarities and differences in the way the process evolves under varying conditions. Notice that the first row of plots in Figure 4 corresponds to the coefficients that form the linear combination of the elements in $\mathbf{a}_{t-1}$ to produce $a_{1,t}$. The first row suggests that there is little difference

between propagator matrices in their influence on the first EOF, and really only the first two elements of $\mathbf{a}$ at the previous time $(t-1)$ are important (i.e., non-zero) in this regard. However, the second row of plots highlights key differences in the propagator matrices. The second element of the process at the previous time is important in both drought and non-drought conditions, however the third and fourth elements at the previous time are likely only important to the process at the current time $(a_{2,t})$ during drought years. In the third row of plots, it is evident that $a_{1,t-1}$ is important to $a_{3,t}$ during drought years, while $a_{4,t-1}$ is important to $a_{3,t}$ only during non-drought years. A similar interpretation can be made with the fourth row of plots.

The disadvantage to interpreting these differences in the dynamics of the latent process is that the components of this process themselves are not intuitive. It appears that the first component (i.e., $a_{1,t}$) is somewhat related to the effect of PDSI on tree growth and that there are differences in the way the latent process evolves under varying drought regimes, but beyond that the components (EOF's or $\mathbf{a}_t$'s) are difficult to interpret. They have the implicit spatial interpretation given by the eigenvectors of the EOF decomposition, however they lack any obvious anthropogenic, environmental, or biological meaning. The dynamics of the physical process (i.e., $\mathbf{y}_t$, tree growth) are of interest here rather than the underlying process $(\mathbf{a}_t)$ which evolves in a different and not always intuitive space.

Visualizing dynamical structure in the implicit process $\mathbf{y}_t$ may provide more insight on the scale of the observed process (i.e., the data). The posterior mean propagator matrices of the implicit process (Figure 7) illustrates how shortleaf pine growth at one location may be associated with growth at another location at the next time. Individually, these propagator

matrices represent the implicit Markovian dynamics within each portion of the process subset by the time-period / covariate combinations. Thus changes in growth dynamics between drought and non-drought years before and after 1880 are more easily assessed.

Notice in Figure 7 that the implicit propagator matrices corresponding to the covariate effect are similar. That is, $\mathbf{B}_1$ is similar to $\mathbf{B}_3$ (drought years) and $\mathbf{B}_2$ is similar to $\mathbf{B}_4$ (non-drought years), although not identical. This implies that the effect of PDSI on growth dynamics is more significant than change in time period. Subtle differences do exist however, between these similar propagator matrices. For example, the elements in the first row and third column of the matrices $\mathbf{B}_1$ and $\mathbf{B}_3$ are significantly different with a posterior mean difference of 0.37 and 95% credible interval of $(0.08, 0.65)$. This suggests that during early years of drought ($\mathbf{B}_1$), shortleaf pine growth at the third location (winona) is positively associated with growth in the first location (ellen) at the next time. While after 1880, in drought years ($\mathbf{B}_3$), the eighth location (with posterior difference credible interval: $(-0.42, -0.01)$), rather than the third location, is more associated with the first location at the next time. Similar spatio-temporal associations can be drawn between growth at various locations based on the interpretation of the different implicit propagator matrices. Such information is useful in cases where there may be missing data and one might want to estimate the growth at a certain location and time given the rest of the data. Additionally, this can be useful in forcasting growth at various locations although the focus here is assessing the changes in growth dynamics.

Consider now a more formal method for evaluating the dynamics of the process. Spectral decomposition of the propagator matrices (non-symmetric) yields the principal oscillation

patterns (or normal modes of the system). These POP's can be used to examine the oscillation properties and spatial structure of the implicit dynamical process ($\mathbf{y}_t$). Figure 8 illustrates that although the relative importance of the POP's themselves are the same for covariate-based groups (see Table 2), the normal modes are different, representing the varying dynamic structure in the process. This same information can be viewed spatially (Figure 8). Notice that the first oscillation pattern for years when $x_t < x^*$ ($\mathbf{B}_1, \mathbf{B}_3$; drought years) is similar before and after 1880, but differs notably in the dynamics at two sites, ellen and winona. This implies that during drought years, the difference in dynamical structure accounted for by the first normal mode occurs at those sites. Likewise, for non-drought years, the dynamical structure evident in the first POP occurs between the two sites in southern Missouri (shannon and lower) and the site in southeast Arkansas (levi). Also of interest is the fact that the patterns corresponding to covariate effect are similar, suggesting that the response in growth to PDSI is more influential to the system (at least in the first normal mode of the system). As a contrast, consider the fourth POP (Figure 9). In this normal mode, $\mathbf{B}_1$ and $\mathbf{B}_3$ are quite different from each other and different from $\mathbf{B}_2$ and $\mathbf{B}_4$ as well, although $\mathbf{B}_2$ and $\mathbf{B}_4$ are very similar. This implies that during drought years, the dynamics evident in the fourth mode are very different before and after 1880, however during non-drought years, there is little difference in the dynamics before and after 1880.

In cases where there are complex eigenvalues (POP_2 and POP_3), even more information is available about the oscillatory behavior of the system. Recall that the sequence of real and imaginary components from the complex pair of POP's corresponds to an evolving spatial pattern (Figure 10) with frequency and period given in Table 3. These measures

suggest that during drought years ($\mathbf{B}_1, \mathbf{B}_3$) the pattern sequence in the first and third row of Figure 10 evolves more rapidly than the pattern in rows two and four (since the e-folding times are smaller in drought years). Notice again that the primary difference between time periods in drought years occurs at the ellen site, while in non-drought years the biggest difference in pattern occurs at the winona site. Another interesting feature is that in drought years the pattern evolves in a North / South fashion, while in wetter years, the pattern appears to evolve more in an East / West orientation.

2.5 Conclusion

The characterization of dynamical shifts in annual growth of shortleaf pine as a result of biotic and abiotic environmental factors as well as the biology of the species can be ecological and economically important from a management perspective. In order to thoroughly assess such shifts in vegetative dynamics in a given region, spatially explicit landscape scale data should be considered.

Such data, though readily available as dendrochronological (tree-ring) data, is subject to many sources of error. Additionally, the true process that generates the data (shortleaf pine growth) is likely correlated with changes in climate and other external influences as well as internal mechanisms that regulate growth over time. Such relationships coupled with a latent state-space process model allows the decomposition and examination of the underlying components driving the process. Here, a model and model-output analysis methods have been adopted that allow for the detection of whether shifts in the system dynamics exist and characterization of them in the presence of various sources of uncertainty (e.g., measurement and preprocessing error, spatial random effects, and possible model

misspecifications).

Specifically, it is shown that by projecting the data (and implicit process in physical space) onto two reduced sets of basis functions corresponding to a changepoint in time (related to anthropogenic influence) implementation of a parsimonious model focused on accurate parameter estimation while implying a more complicated latent process model is feasible. Differences in the dynamical structure of shortleaf pine growth in the northwest portion of its natural range corresponding to levels of PDSI and time-period were found to be significant and therefore described in detail. Ecologists may find the spatially explicit properties of the posterior system dynamics useful for managing the species, especially in times of rapid climate change and (or) increasing anthropogenic influence.

The model and methods presented here, although new in this setting, have received much more popularity in fields where more obvious dynamical systems are present (e.g., fluid dynamics in atmospheric science and oceanography). High-dimensional spatio-temporal research is less common in ecological sciences due mostly to the difficulty in data collection. However, as more and more large scale datasets become available (such as those available from the International Tree-Ring Data Bank and the World Data Center for Paleoclimatology, as used here) more advanced methods common to other disciplines grow in utility for addressing ecological problems. Concerning the methods presented here, it is important to note that many other model specifications and parameterizations are possible and may be of varying utility in other applications.

2.6 Appendix: Full-Conditionals

To streamline the notation let,

$$\mathbf{\Phi}_t = \begin{cases} \mathbf{\Phi}_1, & \text{if } \ t = 1, \ldots, t^* - 1 \\ \mathbf{\Phi}_2, & \text{if } \ t = t^*, \ldots, T, \end{cases}$$

$$\mathbf{H}_t = \begin{cases} \mathbf{H}_1, & \text{if } \ t \in \{t | x_t < x^*\} \\ \mathbf{H}_2, & \text{if } \ t \in \{t | x_t \geq x^*\}, \end{cases}$$

This allows the posterior distribution to be written as:

$$[\mathbf{a}_{t=0,\ldots,T}, \mathbf{H}_1, \mathbf{H}_2, \mathbf{R}, \sigma_Q^2 | \mathbf{Z}, \mathbf{x}] \propto \prod_{t=1}^{T} [\mathbf{z}_t | \mathbf{a}_t, \mathbf{R}] \prod_{t=1}^{T} [\mathbf{a}_t | \mathbf{a}_{t-1}, \mathbf{H}_t, \sigma_Q^2] \times$$
$$\times [\mathbf{a}_0][\mathbf{H}_1][\mathbf{H}_2][\mathbf{R}][\sigma_Q^2].$$

Thus the full-conditional distributions can be written as:

$$\mathbf{a}_0 | \cdot \sim N((\mathbf{H}_1' \mathbf{Q}_1^{-1} \mathbf{H}_1 + \mathbf{\Sigma}_0^{-1})^{-1} (\mathbf{H}_1' \mathbf{Q}_1^{-1} \mathbf{a}_1 + \mathbf{\Sigma}_0^{-1} \boldsymbol{\mu}_0), (\mathbf{H}_1' \mathbf{Q}_1^{-1} \mathbf{H}_1 + \mathbf{\Sigma}_0^{-1})^{-1}),$$

$$\mathbf{a}_t | \cdot \sim N((\mathbf{\Phi}_t' \mathbf{R}^{-1} \mathbf{\Phi}_t + \mathbf{Q}_t^{-1} + \mathbf{H}_{t+1}' \mathbf{Q}_{t+1}^{-1} \mathbf{H}_{t+1})^{-1} (\mathbf{\Phi}_t' \mathbf{R}^{-1} \mathbf{z}_t + \mathbf{Q}_t^{-1} \mathbf{H}_t \mathbf{a}_{t-1} +$$
$$+ \mathbf{H}_{t+1}' \mathbf{Q}_{t+1}^{-1} \mathbf{a}_{t+1}), (\mathbf{\Phi}_t' \mathbf{R}^{-1} \mathbf{\Phi}_t + \mathbf{Q}_t^{-1} + \mathbf{H}_{t+1}' \mathbf{Q}_{t+1}^{-1} \mathbf{H}_{t+1})^{-1}),$$
$$\text{for } t = 1, \ldots, T-1,$$

$$\mathbf{a}_T | \cdot \sim N((\mathbf{\Phi}_T' \mathbf{R}^{-1} \mathbf{\Phi}_T + \mathbf{Q}_T^{-1})^{-1} (\mathbf{\Phi}_T' \mathbf{R}^{-1} \mathbf{z}_T + \mathbf{Q}_T^{-1} \mathbf{H}_T \mathbf{a}_{T-1}),$$
$$(\mathbf{\Phi}_T' \mathbf{R}^{-1} \mathbf{\Phi}_T + \mathbf{Q}_T^{-1})^{-1}),$$

$$\mathbf{h}_i | \cdot \sim N(\mathbf{V}_i^{-1} \mathbf{b}_i, \mathbf{V}_i^{-1}), \quad \text{for } \ i = 1, 2,$$

$$\mathbf{R}^{-1} | \cdot \sim Wish \left(\left(\sum_{t=1}^{T} (\mathbf{z}_t - \mathbf{\Phi}_t \mathbf{a}_t)(\mathbf{z}_t - \mathbf{\Phi}_t \mathbf{a}_t)' + v_R \mathbf{C}_R \right)^{-1}, v_R + T \right),$$

$$\sigma_Q^2 | \cdot \sim IG \left(\left(\frac{1}{2} \sum_{t=1}^{T} (\mathbf{a}_t - \mathbf{H}_t \mathbf{a}_{t-1})' \mathbf{\Lambda}_t^{-1} (\mathbf{a}_t - \mathbf{H}_t \mathbf{a}_{t-1}) + \frac{1}{r_Q} \right)^{-1}, \frac{lT}{2} + q_Q \right).$$

Definition of index sets $M_1 = \{t|x_t < x^*\}$ and $M_1 = \{t|x_t \geq x^*\}$ with $m_1 = dim(M_1)$ and $m_2 = dim(M_2)$, yields:

$$\mathbf{V}_i = (\mathbf{A}'_{m_i-1} \otimes \mathbf{I}_l)' \tilde{\mathbf{Q}}_i^{-1} (\mathbf{A}'_{m_i-1} \otimes \mathbf{I}_l) + \mathbf{\Sigma}_i^{-1}$$

$$\mathbf{b}_i = (\mathbf{A}'_{m_i-1} \otimes \mathbf{I}_l)' \tilde{\mathbf{Q}}_i^{-1} vec(\mathbf{A}_{m_i}) + \mathbf{\Sigma}_i^{-1} \boldsymbol{\mu}_i$$

where,

$$\mathbf{A}_{m_i} \equiv [(\mathbf{a}_t)]_{l \times m_i}, \quad t \in M_i$$

$$\mathbf{A}_{m_i-1} \equiv [(\mathbf{a}_{t-1})]_{l \times m_i}, \quad t \in M_i$$

$$\tilde{\mathbf{Q}}_i \equiv \text{blockdiag}(\mathbf{Q}_t), \quad t \in M_i.$$

3 POPULATION VIABILITY OF ENDANGERED SPECIES

3.1 Introduction

The Red-Cockaded Woodpecker (RCW) is a federally listed endangered species (35 Federal Register 1970) native to open, fire-maintained pine ecosystems and is endemic to the Southeastern United States. Historically, southern pine ecosystems, particularly longleaf pine – wiregrass ecosystems, were fire-prone due to lightning strikes from frequent thunderstorms (Frost 1993). The RCW is a cooperatively breeding species that lives in family groups (Walters 1990) and requires stands of mature living pines with large heartwood/sapwood ratios in which to excavate roosting and nesting cavities (Jackson et al. 1979). Additionally, the aggregate of cavity trees used by a group (cluster) must occur in stands with little or no hardwood midstory or overstory. The RCW also requires an abundant foraging habitat of large pines, low densities of small pines, little or no hardwood midstory, and an abundant groundcover of native bunchgrasses and forbs (Conner et al. 2001). Major factors limiting the growth of RCW populations are a lack of suitable cavities and fire suppression, which allows for hardwood encroachment (USFWS 2003). Managers can mitigate for a shortage of suitable cavities by installing artificial cavities in younger pines with smaller heartwood/sapwood ratios (Allen 1991). A potential breeding group (PBG) is an adult male and adult female that occupy the same cluster with or without one or more adult helpers (usually males from previous nesting seasons).

Clusters of cavity trees that are occupied by a family group of RCWs are termed "active clusters." Because some clusters are occupied by single birds, the number of active clusters is not a complete measure of a successful RCW population. The ratio of PBGs to active

clusters is a better of RCW population vigor, therefore a population model that accomodates both active clusters and PBGs is necessary.

3.2 Methods

RCW data were collected from Ft. Stewart, Georgia for the years 1994 through 2004. At this site, there are inactive clusters (clusters of cavity trees containing no RCWs), that have been installed by managers to provide for RCW population expansion. The focus of model 1 is on two measures of RCW population status: number of active clusters and number of PBGs, while model 2 accommodates total clusters as well (i.e., active and inactive clusters). The number of PBGs is an estimate of the effective breeding size of the population at each time. Here, it is assumed that the number of active clusters is known at all times when data were collected at Ft. Stewart. Additionally, PBGs were not observed in every cluster in every year (Figure 11).

One way to model the growth in population over time while taking into consideration the limits imposed by density-dependence is to assume the true (underlying) growth in the population to follow a conventional population biology growth equation (e.g., a Ricker growth curve). Such equations allow for ecologically realistic growth and carrying capacities while being flexible enough to allow for chaotic population behavior and possible limit-cycles.

A hierarchical framework allows for the incorporation of multiple sources of information about the growth of the RCW population and also accommodates the assumption of the existence of an underlying dynamical system. Also, as in previous chapters, a Bayesian implementation of the model allows various sources of uncertainty and previously existing

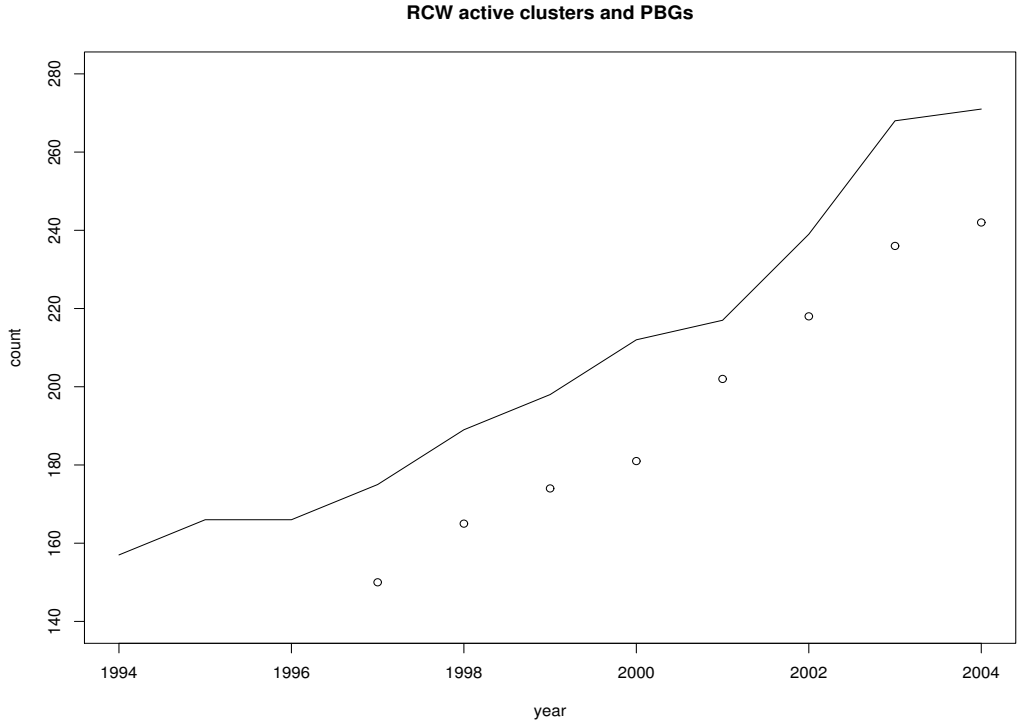


Figure 11: RCW active clusters (solid line) and PBGs (points) at Ft. Stewart for 1994 through 2004.

scientific information about the RCW biology to be accounted for.

3.2.1 Model 1

First, consider a model for the number of active clusters and PBGs. Let N_t denote the number of active clusters at time t for $t = 1, \ldots, T$. This parameter is assumed to be known at each time for which data exists, although it need not be (see Model 2). Additionally, it is not known for times $t > T$ and is of interest at those future times.

Let n_t denote the number of PBGs at time t for $t \in S$, where S is the set of times where PBG data were collected. Note that this may not be at every time for which N_t is known.

Now consider the following data model, in the usual hierarchical fashion:

$$n_t|N_t, \theta \sim Binom(N_t, \theta), \text{ for } t = 1, \ldots, T, \tag{3.1}$$

where θ is the true proportion of active clusters that contain a PBG. Conventional scientific wisdom suggests that there always exists a portion of active clusters not containing a breeding group. This proportion may be of use to managers, thus it is estimated using this model. Although we have little prior information about what it should be, and it may vary from one population to another, therefore a non-informative Jeffreys' prior is indicated. This allows us to account for lack of specific and precise information about this parameter while still being able to estimate it.

Assuming that the dynamics of the ecological process (i.e., RCW population growth) occur on some latent system, the following process model can be specified.

$$N_t|\lambda_t \sim Pois(\lambda_t), \text{ for } t = 1, \ldots, T, \tag{3.2}$$

where λ_t is the Poisson intensity (the mean number of active clusters at time t) and evolves according to the following density dependent population model (i.e., Ricker equation).

$$\lambda_t = \lambda_{t-1} \exp(\beta_1(1 - \frac{\lambda_{t-1}}{\beta_2})), \text{for } t = 1, \ldots, T, \tag{3.3}$$

where β_1 and β_2 are parameters controlling the growth rate and carrying capacity respectively. In this case, quite specific prior information about the carrying capacity (β_2) exists due to the geographic extent of the study area and the cluster size constraints. Therefore *a priori* information in the form of a subjective prior can be used to help inform this

parameter.

$$\beta_2 \sim Gamma(\text{mean} = 625, \text{var} = 400) \tag{3.4}$$

Conversely, relatively little is known about the growth rate parameter (β_1) and, in fact, can implicitly attribute it to active management of the species. Therefore a vague prior distribution over a scientifically reasonable range of values is specified for this parameter.

$$\beta_1 \sim Gamma(\text{mean} = 1, \text{var} = 4) \tag{3.5}$$

Obviously, this growth rate varies according to varying management practices and habitat change. One could implement a more complex time-varying growth rate model given additional covariate information. In the absence of such information the covariate effect is assumed to be implicit.

Finally, the true population size at the first time is assumed to be random with a fairly informative prior around the "known" parameter N_1.

$$\lambda_1 \sim Gamma(\text{mean} = N_1, \text{var} = 4) \tag{3.6}$$

3.2.2 Model 2

Now, consider a model that accounts for a known number of total clusters (i.e., active and inactive clusters) and observed number of active clusters . Let the number of total clusters at time t be denoted as C_t and observed number of active clusters and PBGs be denoted as N_t and n_t, respectively, as in model 1. All of n_t, N_t, and C_t are assumed to be known at times for which data exist and have the following constraint: $n_t \leq N_t \leq C_t$. In addition, it is believed that while the average ratio of PBGs to active clusters remains constant over time, the number of active clusters actually approaches the number of total clusters over time.

This is partially due to the age of the forest becoming more suitable for RCW success, but may also be due to the utilization of all available space as resources becoming limiting when population size approaches carrying capacity. This constraint poses significant challenges for predicting active clusters and PBGs at future times. Therefore, if total clusters is to be accommodated, a non-trivial extension to model 1 must be made. The added benefit is a richer probabilistic characterization of all components of RCW population growth.

Consider the following updated data model specification:

$$[n_t, N_t|C_t, \theta_t, \theta] \propto [n_t|N_t, \theta][N_t|C_t, \theta_t], \quad t = 1, \ldots, T, \tag{3.7}$$

where, $n_t|N_t, \theta \sim Binom(N_t, \theta)$ and $N_t|C_t, \theta_t \sim Binom(C_t, \theta_t)$. Now, the dynamical process model for population growth can be specified with respect to the total clusters (C_t):

$$C_t|\lambda_{C,t} \sim Pois(\lambda_{C,t}), \tag{3.8}$$

where, $\lambda_{C,t}$ evolves according to a Ricker growth equation with random growth rate ($\beta_{C,1}$) and carrying capacity (β_2) parameters (i.e., $\lambda_{C,t} = \lambda_{C,t-1}\exp(\beta_{C,1}(1 - \lambda_{C,t-1}/\beta_2))$). In order for N_t to approach C_t as C_t approaches the carrying capacity (β_2), θ_t must approach 1. Since the rate at which $\theta_t \to 1$ is unknown, it must be estimated from the data. Thus, consider a second dynamical growth process ($\lambda_{N,t}$) that represents the "average" number of active clusters over time. Let $\lambda_{N,t}$ evolve over time according to the same growth equation as $\lambda_{C,t}$ with the same carrying capacity, but different growth rate ($\beta_{N,1}$) (i.e., $\lambda_{N,t} = \lambda_{N,t-1}\exp(\beta_{N,1}(1 - \lambda_{N,t-1}/\beta_2))$). Using matrix notation, and letting $\boldsymbol{\lambda}_t = [\lambda_{C,t}, \lambda_{N,t}]'$ and $\boldsymbol{\beta}_1 = [\beta_{C,1}, \beta_{N,1}]'$ yields the following underlying dynamical system:

$$\boldsymbol{\lambda}_t = \boldsymbol{\lambda}_{t-1}\exp(\boldsymbol{\beta}_1(1 - \frac{\boldsymbol{\lambda}_{t-1}}{\beta_2})), \tag{3.9}$$

where all multiplication is scalar. Now, the definition of $\theta_t = \frac{\lambda_{N,t}}{\lambda_{C,t}}$ allows $\theta_t \to 1$.

The following prior distributions were specified in a similar fashion to those in Model 1, where again a non-informative Jeffreys prior is utilized for θ and,

$$\lambda_{C,1} \sim Gamma(\text{mean} = C_1, \text{var} = 4),$$

$$\lambda_{N,1} \sim Gamma(\text{mean} = N_1, \text{var} = 4),$$

$$\beta_{C,1} \sim Gamma(\text{mean} = 1, \text{var} = 4),$$

$$\beta_{N,1} \sim Gamma(\text{mean} = 1, \text{var} = 4),$$

$$\beta_2 \sim Gamma(\text{mean} = 625, \text{var} = 400).$$

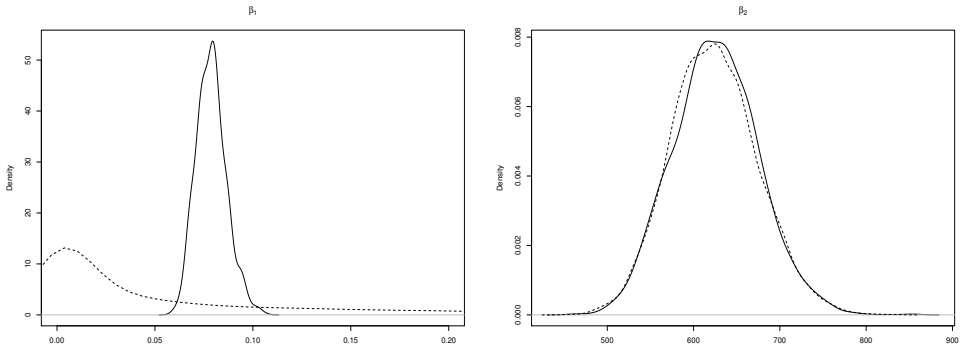
As in Model 1, posterior predictions of active clusters and PBGs, for $t > T$, are the primary focus for inference and can be found readily by integrating over the model parameters in the usual fashion.

3.3 Results

Models 1 and 2 are non-conjugate in all parameters with the exception of θ thus a Gibbs sampler with Metropolis-Hastings updates was necessary to sample from the posterior distribution. Convergence was reached almost immediately for all paremeters (using 100,000 gibbs iterations with a 10000 iteration burn-in period). Gibbs samples were systematically resampled to remove any MCMC induced correlation.

3.3.1 Model 1 Results

First, for Model 1, consider the parameter estimates for the population growth (β_1, β_2 in Figure 12). Their corresponding prior distributions are given as well. The posterior distribution for parameter relating to the proportion of PBGs to active clusters (θ) is given



(a) (b)

Figure 12: Prior (dashed) and posterior (solid) distributions of the growth rate and carrying capacity parameters respectively.

in Figure 13. The posterior means and 95% credible intervals are provided for the PBGs at

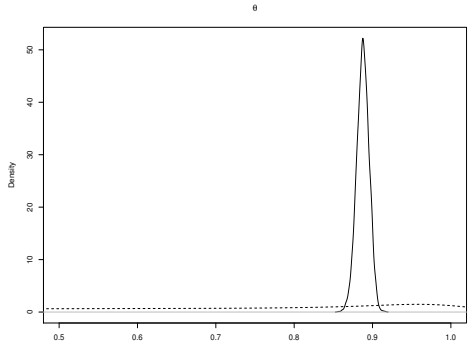


Figure 13: Prior (dashed) and Posterior distribution (solid) for the probability (θ) of an active cluster containing a PBG.

all times and for the posterior predictions of active clusters and PBGs for 50 years into the future in Figure 14. Additionally, in Figure 14, is the current PBG management objective of 350 and its intersection with the posterior predictive distribution for n_t.

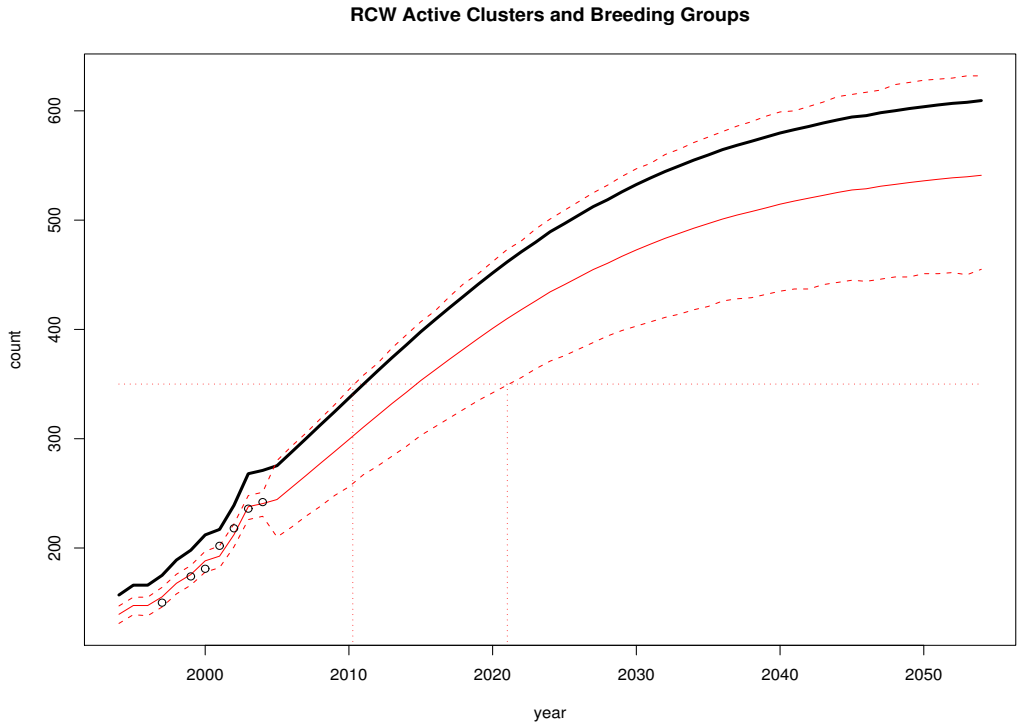
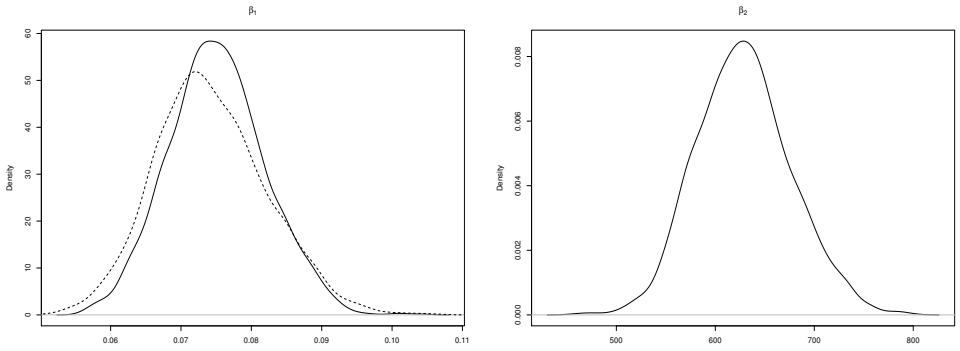


Figure 14: Predictions for active clusters (bold line) and posterior mean (solid, thin line) and 95% credible interval (dashed, thin line) for the number of PBGs (thin) at Ft. Stewart for 1994 through 2004. Also shown are the PBG recovery goal (horizontal dotted line) and its intersection with the PBG credible interval (vertical dotted lines). PBG data are shown as points.

3.3.2 Model 2 Results

For Model 2, there are two growth rate parameters ($\beta_{N,1}$ and $\beta_{C,1}$) that control the speed at which the number of active clusters and total clusters converge (Figure 15a) to the carrying capacity (β_2), (Figure 15b). The posterior distribution for parameter relating to the proportion of PBGs to active clusters (θ) is given in Figure 16. The 95% posterior predictive credible intervals are provided for the PBGs at all times and for the total and active clusters 50 years into the future in Figure 17. Additionally, in Figure 17, is the current PBG management objective of 350 and its intersection with the posterior predictive



(a) (b)

Figure 15: Posterior distributions of the growth rates ($\beta_{C,1}$=solid and $\beta_{N,1}$=dashed) as well as carrying capacity (β_2), respectively.

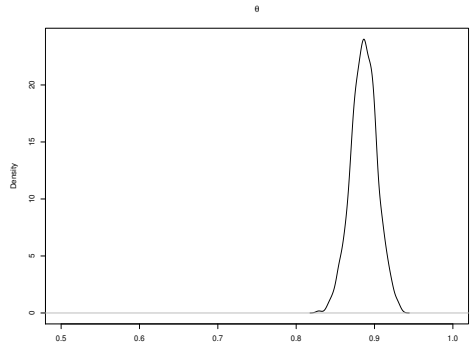


Figure 16: Posterior distribution for the probability (θ) of an active cluster containing a PBG.

distribution for n_t.

3.4 Discussion

These models highlight the advantages of incorporating known sources of uncertainty into an ecological model as well as the effect of excluding potentially important information (i.e., number of total clusters). For Model 1, the fact that the prior and posterior distribution for the carrying capacity parameter (β_2, Figure 12b) are quite similar suggests that the data

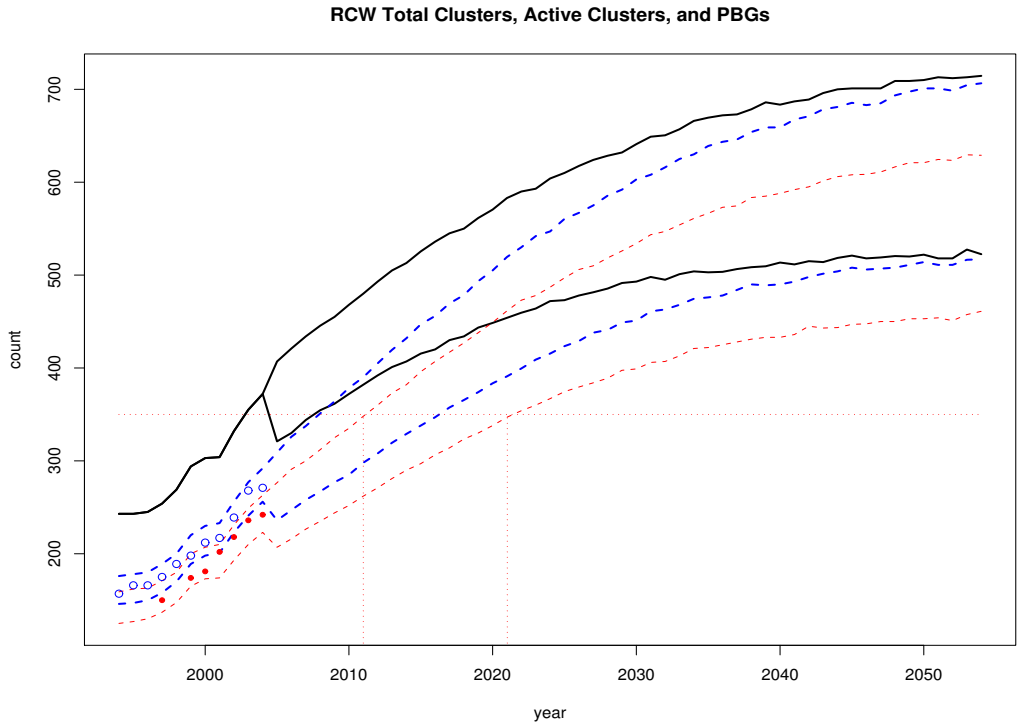


Figure 17: 95% predictive credible intervals for total clusters (bold lines), active clusters (dashed lines), and for the number of PBGs (thin dotted lines) at Ft. Stewart for 1994 through 2004. Also shown are the PBG recovery goal (horizontal dotted line) and its intersection with the PBG credible interval (vertical dotted lines). PBG data are shown as points.

contains limited information about the RCW saturation for this study area. This is likely due to the lack of an asymptote in the data (Figure 11). However, since specific knowledge about this parameter exists, it allows the model to focus on the estimation of the growth rate parameter (β_1). In this case, as mentioned earlier, a very naive model is assumed for β_1. Though, in the presence of covariate information, this need not be the case. Also, notice how precise the posterior estimate of β_1 is, given the data, versus its prior (Figure 12a).

The probability of an active cluster containing a PBG is also a parameter which is estimated precisely despite its non-informative prior (Figure 13). The posterior mean of

θ (i.e., $E(\theta|N,n) = 0.888$) is quite similar to point estimates from previous studies ($\hat{\theta} = 0.89$). The uncertainty in this parameter, however, represents a little studied, but much speculated about characteristic of RCW biology and may provide useful information for setting management objectives for this species.

One of the primary interests here is the evolution of the true population as a dynamical system. Of specific utility, are the trajectories and forecasts for critical population measures. From Figure 14 it is apparent that for years where data exist there is little uncertainty about the number of PBGs and the observed data fall into the expected range of values. However, for times without data, especially where the number of active clusters is unknown, there exists substantial uncertainty associated with these population measures. Over the course of 50 years, the posterior predicted population levels approach saturation. The level of saturation, though, varies by about 200 clusters. Additionally, the posterior predicted date at which the current management goal of 350 PBGs will be reached is between the years 2010 and 2021 with probability 0.95 (the posterior predictive mean date at which 350 PBGs will be reached is year 2015). Model 2 produces similar results with subtle but possibly important differences. Based on the posterior distributions for the growth rate parameters ($\beta_{N,1}$ and $\beta_{C,1}$), it is clear that although $\beta_{C,1}$ is more precise, the two estimated distributions are not significantly different. This suggests that based on the data, the initial state and carrying capacity have more influence on the rate of convergence between N_t and C_t than do separate growth rates. Obviously, with more data, this influence could change.

As with Model 1, the posterior distribution of θ is estimated using Model 2 (Figure 16) and is similar but slightly less precise than that of Model 1. This suggests that there is

slightly more uncertainty associated with the probability of an active cluster containing a PBG when information exists about the total number of clusters.

The 95% posterior predictive credible envelopes for C_t, N_t, and n_t from Model 2 (Figure 17) appear similar to those from Model 1 (Figure 14). Notice, however, that the predicted date to attain the management goal of 350 PBGs is shifted back to between 2011 and 2021 with a probability of 0.95 using Model 2. Thus the extra information related to total clusters increases the precision of the posterior predictions but delays the earliest probable date by approximately 1 year.

3.5 Conclusion

In summary, simple but ecologically meaningful models are proposed for estimating population growth parameters and forecasting critical population measures via a hierarchical Bayesian framework with a latent non-linear dynamical system.

Again both process growth models are very naive, in that, they do not explicitly consider management practices and (or) habitat changes over time. Thus the parameter estimates are only valid under the assumption that the model is implicitly accounting for the covariate effects. Additionally, the current models have no mechanism to account for possible future population collapses (e.g., major disturbances such as hurricanes or military activities). However, based on information available and strong but scientifically reasonable assumptions, they allow for long-range population projections and provide a means for true probabilistic inference.

4 POPULATION GROWTH OF INVASIVE SPECIES IN SPACE AND TIME

4.1 Models for Relative Abundance

4.1.1 Introduction

Differential equation based advection-diffusion models have long been used in atmospheric science to mimic complex processes such as weather and climate. Differential and partial differential equations (PDE's) have become popular in biological and ecological fields as well. In many cases these models are considered in a strictly deterministic framework even though many sources of uncertainty in the process, the model, and the measurements may exist. Specifically in the ecology realm, Clark et al. 2001 allude to a problem that arises when various sources of uncertainty are not appropriately accounted for. That is, inferences resulting from such methods will be misleading.

The spread of invasive species is one phenomenon in particular for which there is much to be (and has already been) learned about through the use of PDE models (Holmes et al. 1994, Shigesada and Kawasaki 2002). Such phenomena usually are based on complicated dynamic processes, many of which are nonlinear and are potentially correlated with and/or controlled by other environmental features (e.g., climate patterns, biotic interactions, physiological characteristics, human population density, etc).

Many deterministic PDE models are well-equiped to represent the theoretical spread of organisms, but have no mechanism to account for the various sources of uncertainty related to the inadequacies of the model as well as the process itself and knowledge of it. However the use of a PDE within the framework of a hierarchical Bayesian model can provide a useful link between scientifically based deterministic models and statistical models

that accurately portray variability (Wikle 2003). Characterizing, and more importantly, predicting realistic levels of abundance over time depends on the models ability to recognize biologically meaningful limitations to growth, in addition to the growth rates themselves. In the early stages of an invasion, a linear growth model and density dependent growth model may behave similarly. As the carrying capacity is appoached, however, a model with the flexibility to exhibit density dependence will provide more realistic results. Therefore, an invasive species model with the ability to account for non-linear growth is desirable.

Specifically, the spread of Eurasian Collared-Dove (ECD; *Streptopelia decaocto*) is modeled in terms of "relative abundance" throughout the United States using a reaction-diffusion PDE (Fisher 1937, Skellam 1951) within a hierarchical model. Note that, modeling of relative abundance does not account for the detectability of organisms (nor uncertainty related to it). This assumption implies that model results will portray general patterns of growth and spread throughout a given area but not actual numbers of organisms expected at each location. Section 4.2 details a method for simultaneously modeling growth and spread while accounting for the probability of detection.

The North American Breeding Bird Survey (BBS; Robbins et al. 1986) monitors many birds across the United States and Canada, and provides a major source of data for studying invasive bird species. One such species is the Eurasian Collared-Dove. The ECD is an often misidentified species in North America (due to its similarity to the Ringed Turtle-Dove) and is considered invasive and a potential threat to indigenous ecosystems (Hengeveld 1993). After moving into Europe in the 1930's, the ECD was introduced in the Bahamas and was observed in the mid-1980's in Florida (Hudson 1965, Romagosa and Labisky 2000).

Not only has its range increased as it has spread through the Southeastern United States (Figure 18), but it has experienced growth in population as well (Figure 19).

BBS data were collected by volunteer observers each breeding season along specified routes (Robbins et al. 1986). The sampling units are roadside routes of length approximately 39.2 km, along which an observer makes 50 stops and counts birds by sight and sound for a period of three minutes. There are over 4000 routes in the survey, but not all are sampled each year. It should be noted that these data are subject to various types of uncertainty, including observer error, spatial location error, and error related to change of spatial support (Sauer et al. 1994). Additionally, Hengeveld (1993) and Romagosa and Labisky (2000) suggest that the rate of ECD spread may vary spatially and temporally and may be correlated with other environmental covariates (e.g., human population density). Furthermore, the ECD is an important candidate for study in the United States because of the quantity of data and knowledge accumulated since its introduction combined with historical information gathered from its invasion of Europe (Hengeveld 1993).

A model that incorporates these factors and responsibly accounts for the various sources of uncertainty could be a useful tool for ecologists wanting to monitor the spread and growth in population as well as identify if factors influence the rate of spread in order to make management decisions to help protect native species.

4.1.2 Methods

In order to account for the many sources of uncertainty, spatial effects, temporal effects, population growth, and spread of the invasive species, a 3-stage hierarchical model is adopted (as discussed in Chapter 1), and has the following factorization (Berliner 1996):

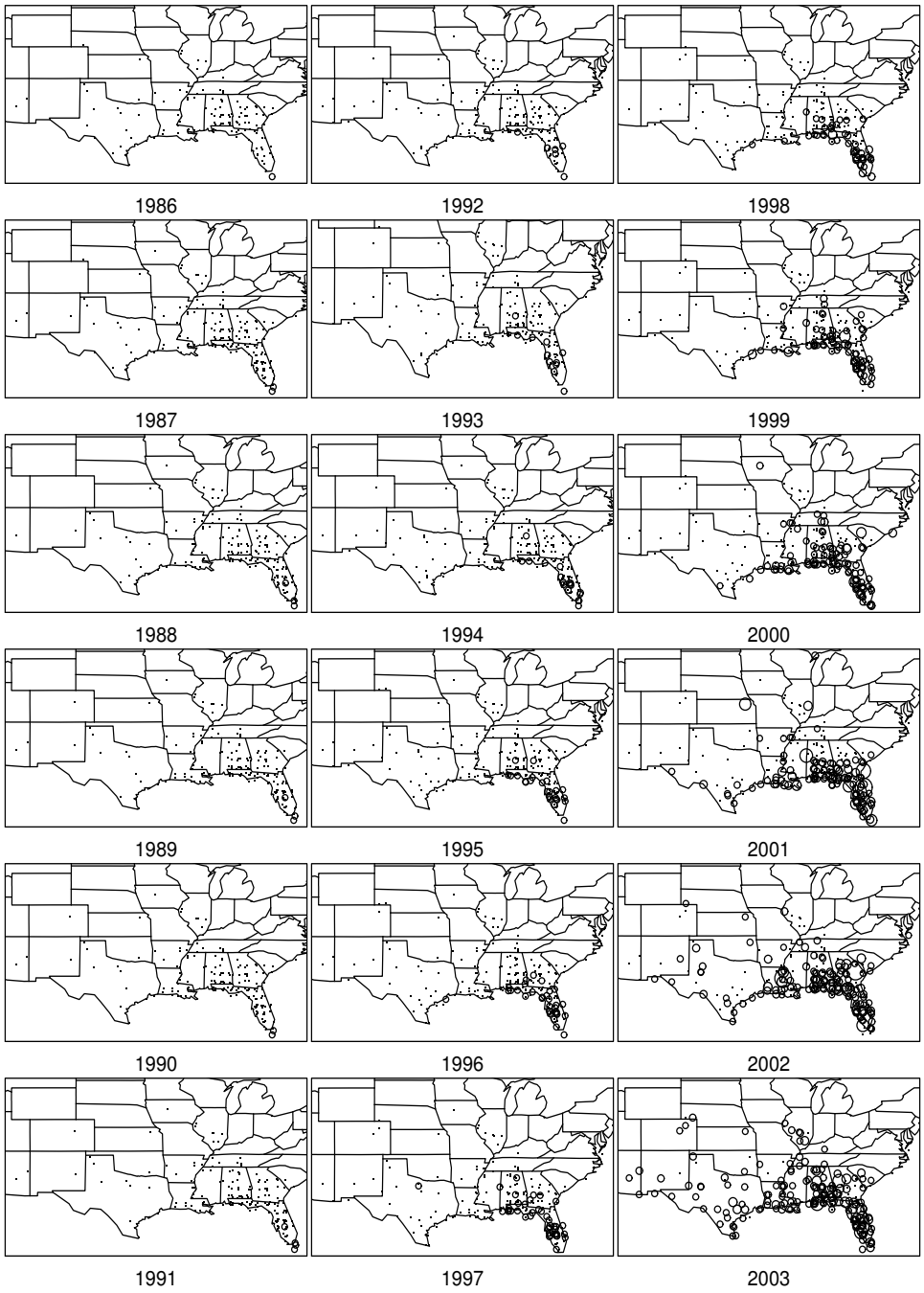


Figure 18: Spread of ECD throughout the United States from 1986 through 2003 (points represent zero counts at sampled location, while circle size corresponds to non-zero count magnitude).

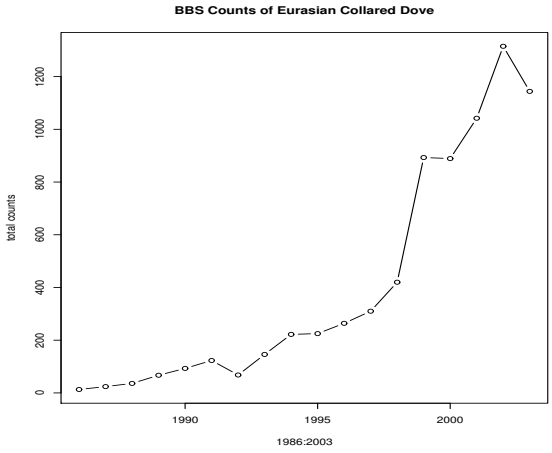


Figure 19: Population growth of ECD in the United States from 1986 through 2003 (total counts over time)

[data, process, parameters] $\propto$ [data | process, parameters][process | parameters][parameters].

Consider the relative abundances of birds ($\mathbf{n}_t = [n_{1,t}, \ldots, n_{i,t}, \ldots, n_{m,t}]'$) at spatial locations $i = 1, \ldots, m$ and times $t = 1, \ldots, T$ given an intensity process as Poisson where m is the number of spatial locations and T the number of years:

$$n_{i,t}|\lambda_{i,t} \sim Pois(\lambda_{i,t}), \qquad i = 1, \ldots, m \qquad t = 1, \ldots, T. \tag{4.1}$$

In this case, since the counts are originally recorded as observed along routes in the spatial domain. Here, the route center is used as the point location for count data at location i. Since primary concern about the process is continental in scale, route length (tens of kilometers) and orientation have minimal overall influence.

The Poisson intensity process is controlled by a latent spatio-temporal process ($\mathbf{u}_t$) at gridded locations linked by an incidence matrix ($\mathbf{K}_t$). The number of gridded locations (i.e., $dim(\mathbf{u}_t) = N$) need not equal the number of observation locations; in this case $N <$

m increases computational efficiency. It may be possible to address some issues related to differences in biological processes at various scales by proposing different sets of basis functions for $\mathbf{K}_t$ but this will be left as the focus of another study. For the model proposed here, $\mathbf{K}_t$ is simply a matrix with an element in the i^{th} row associating the i^{th} measurement location with a grid location at time t. An additive independent noise term is added to account for observer error and small scale spatio-temporal variability:

$$\log(\boldsymbol{\lambda}_t) = \mathbf{K}_t\mathbf{u}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon_t} \sim N(\mathbf{0}, \sigma_\varepsilon^2\mathbf{I}), \qquad t = 1, \ldots, T. \tag{4.2}$$

The latent process ($\mathbf{u}_t$) is motivated via a reaction-diffusion equation (Fisher 1937, Skellam 1951),

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(\delta(x,y)\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(\delta(x,y)\frac{\partial u}{\partial y}\right) + \gamma_0 u\left(1 - \frac{u}{\gamma_1}\right), \tag{4.3}$$

which can be discretized as (Haberman 1987),

$$\begin{aligned}
u_t(x,y) \;=\; & u_{t-\Delta_t}(x,y)\left[1 - 2\delta(x,y)\left(\frac{\Delta_t}{\Delta_x^2} + \frac{\Delta_t}{\Delta_y^2}\right)\right] \\
& + \; u_{t-\Delta_t}(x-\Delta_x,y)\left[\frac{\Delta_t}{\Delta_x^2}\left\{\delta(x,y) - (\delta(x+\Delta_x,y) - \delta(x-\Delta_x,y))/4\right\}\right] \\
& + \; u_{t-\Delta_t}(x+\Delta_x,y)\left[\frac{\Delta_t}{\Delta_x^2}\left\{\delta(x,y) + (\delta(x+\Delta_x,y) - \delta(x-\Delta_x,y))/4\right\}\right] \\
& + \; u_{t-\Delta_t}(x,y+\Delta_y)\left[\frac{\Delta_t}{\Delta_y^2}\left\{\delta(x,y) + (\delta(x,y+\Delta_y) - \delta(x,y-\Delta_y))/4\right\}\right] \\
& + \; u_{t-\Delta_t}(x,y-\Delta_y)\left[\frac{\Delta_t}{\Delta_y^2}\left\{\delta(x,y) - (\delta(x,y+\Delta_y) - \delta(x,y-\Delta_y))/4\right\}\right] \\
& + \; u_{t-\Delta_t}(x,y)\gamma_0 - u_{t-\Delta_t}^2(x,y)\left[\frac{\gamma_0}{\gamma_1}\right],
\end{aligned} \tag{4.4}$$

where x and y correspond to the spatial location in two dimensions. In matrix notation

with additive error and letting $\Delta_t = 1$, this can be written:

$$\mathbf{u}_t = \mathbf{H}(\boldsymbol{\delta})\mathbf{u}_{t-1} + \alpha_0\mathbf{u}_{t-1} - \alpha_1 diag(\mathbf{u}_{t-1})\mathbf{u}_{t-1} + \boldsymbol{\eta}_t, \tag{4.5}$$

$$\boldsymbol{\eta}_t \sim N(\mathbf{0}, \sigma_\eta^2\mathbf{I}),$$

where $diag(\mathbf{u}_{t-1})$ is an $N \times N$ square matrix with the elements of $\mathbf{u}_{t-1}$ on the diagonal and zeros elsewhere and γ_0 and γ_1 in (4.3) represent the intrinsic population growth rate and carrying capacity, respectively, and have the following relationships with the growth parameters in the vector form of the model (4.5):

$$\alpha_0 = \gamma_0, \qquad \alpha_1 = \frac{\gamma_0}{\gamma_1}.$$

If the estimate of the carrying capacity parameter (γ_1) is greater than $\mathbf{u}_t$ for all t, this would imply that the process has not stopped growing yet and estimation or prediction of abundance at times $\{t|t < T + 1\}$ would likely not be different than those resulting from a simpler model.

From (4.4) it can be seen how this model differs from that of Wikle (2003). Here the growth in the process is allowed to be non-linear (from the squared term) whereas previously only a linear (Malthusian growth) term was implemented and may not have been appropriate (or sufficient) (Wikle 2003). Although on the log-scale this growth term may not be ideal either, it does allow for a much more flexible and generalized model in that it can always assume the linear form if indeed α_1 is not influential. Additionally, if the estimate of the carrying capacity parameter (γ_1) is greater than $\mathbf{u}_t$ for all t, this would imply that the process has not stopped growing yet and estimation or prediction of abundance at times $\{t|t < T+1\}$ would likely not be different than those resulting from a simpler model.

Also, the discrete **u**-process is modeled on a rectangular grid with spacing Δ_x and Δ_y in the longitudinal and latitudinal directions, respectively, and with time spacing Δ_t.

The propagator matrix ($\mathbf{H}$) in (4.5) depends upon the diffusion coefficients $\boldsymbol{\delta}$ which are allowed to vary in space. Construction of $\mathbf{H}$ depends on the arrangement of grid locations in the spatial domain and is thus quite project specific. For the irregularly shaped spatial domain considered here the construction of $\mathbf{H}$ is similar to that used in Wikle and Hooten (2006).

Such a non-linear model (4.5) is relatively difficult to implement in an MCMC (Markov Chain Monte Carlo) algorithm due to the high dimensional Metropolis-Hastings updates. Therefore, consider the following modification (thereby also inducing a 2nd-order Markov structure):

$$\mathbf{u}_t = \mathbf{H}(\boldsymbol{\delta})\mathbf{u}_{t-1} + \alpha_0\mathbf{u}_{t-1} - \alpha_1 diag(\mathbf{u}_{t-1})\mathbf{u}_{t-2} + \boldsymbol{\eta}_t. \tag{4.6}$$

The $\mathbf{u}_{t-2}$ term in (4.6) allows for a complete derivation of the full-conditional distribution for $\mathbf{u}_t$, thereby improving computational efficiency. For a growing process such as the one considered here, the effect of using the discretization (4.6) of the PDE in (3) only differs from that of (4.5) by $\alpha_1\mathbf{u}_{t-1}|\mathbf{u}_{t-1} - \mathbf{u}_{t-2}|$. In this case, this difference is negligible and was verified by simulation. It may be possible to make exact inference on the model in (4.5) using (4.6) as a proposal in a Metropolis-Hastings step. It is reasonable to implement (4.6 however, since the discretized PDE is only an approximation intended to allow for density dependence. An alternative non-conjugate approach is detailed in Section 4.2.

Let $\mathbf{X}$ be an $N \times 2$ covariate matrix made up of an intercept and human population density on the gridded spatial domain and let $\boldsymbol{\beta}$ be a 2-dimensional vector of regression

coefficients with $\mathbf{R}$ a spatial correlation matrix depending on parameter θ. Then, $\boldsymbol{\delta}$ can be defined by a linear model,

$$\boldsymbol{\delta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\xi}, \quad \boldsymbol{\xi} \sim N(\mathbf{0}, \sigma_\delta^2 \mathbf{R}(\theta)). \tag{4.7}$$

Let the spatial correlation matrix ($\mathbf{R}$) be defined by the Euclidean distance ($||d||$; the Euclidian distance between points) and the spatial range parameter (θ) in the standard stationary and isotropic exponential covariogram model:

$$R(\theta, d) = exp(-\theta||d||). \tag{4.8}$$

This allows $\boldsymbol{\delta}$ to utilize not only the human population covariate but the (random) effects of another spatial covariate possibly based on an unknown environmental factor. Also, let the ECD population growth parameters have a lognormal distribution :

$$\log(\boldsymbol{\alpha}) \equiv \log\left(\begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix}\right) \sim N(\tilde{\boldsymbol{\alpha}}, \boldsymbol{\Sigma}_\alpha). \tag{4.9}$$

The remaining parameters have the following priors:

$$\boldsymbol{\beta} \sim N(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_\beta), \tag{4.10}$$

$$\sigma_\varepsilon^2 \sim IG(q_\varepsilon, r_\varepsilon), \tag{4.11}$$

$$\sigma_\eta^2 \sim IG(q_\eta, r_\eta), \text{ and} \tag{4.12}$$

$$\sigma_\delta^2 \sim IG(q_\delta, r_\delta). \tag{4.13}$$

A prior distribution for the $\mathbf{u}_t$-process is specified at time zero, $\mathbf{u}_0 \sim N(\tilde{\mathbf{u}}_0, \boldsymbol{\Sigma}_0)$ and $\mathbf{u}_{-1} = -10 \times \mathbf{1}$ is fixed (where $\mathbf{1}$ is a vector of ones). It is reasonable to fix $\mathbf{u}_{-1}$ at a small value because it represents the process at a time before the invasion began; about which

there is little uncertainty. Finally, a reference prior is adopted for the spatial parameter θ as derived by Berger et al. (2002). Such a reference prior is non-informative and more importantly allows the posterior to be proper. Berger et al. (2002) argue that, in cases where little *a priori* information is available for such spatial parameters, it is better to use a reference prior (as derived in their paper) than a Jeffreys prior (which may not yield a proper posterior) or a vague proper prior (which may only hide propriety issues). Prior distributions for other parameters (4.9-4.13) were chosen to ensure model conjugacy for purposes of computational and analytical efficiency, while hyperparameters were either chosen to ensure vague priors or, if sufficient information was available, based on previous studies (e.g., rates of ECD spread in Europe). Table 4 shows the numerical values used for the hyperparameters.

Table 4: Hyperparameters used in MCMC

Hyperparameter	**Value**
q_ε	2.3
r_ε	0.4
q_η	2.0
r_η	2.0
q_δ	2
r_δ	1000
$\tilde{\alpha}_0, \tilde{\alpha}_1$	0.001
$(\mathbf{\Sigma}_\alpha)_{11}, (\mathbf{\Sigma}_\alpha)_{22}$	10
$(\mathbf{\Sigma}_\alpha)_{12}, (\mathbf{\Sigma}_\alpha)_{21}$	2
$\tilde{\mathbf{u}}_0$	$\mathbf{0}$
$\boldsymbol{\beta}_0$	$\mathbf{0}$
$\mathbf{\Sigma}_0, \mathbf{\Sigma}_\beta$	$10 \times \mathbf{I}$

The Bayesian formulation of the hierarchical model is summarized by the following posterior distribution and does not have an analytical representation.

$$\begin{aligned}
&[\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_T, \mathbf{u}_0, \ldots, \mathbf{u}_T, \boldsymbol{\delta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2_\epsilon, \sigma^2_\eta, \sigma^2_\delta, \theta | \mathbf{n}_1, \ldots, \mathbf{n}_T] \\
&\quad \propto \quad \left\{ \prod_{t=1}^{T} [\mathbf{n}_t | \boldsymbol{\lambda}_t] [\boldsymbol{\lambda}_t | \mathbf{u}_t, \sigma^2_\epsilon] \right\} \left\{ \prod_{t=1}^{T} [\mathbf{u}_t | \boldsymbol{\delta}, \mathbf{u}_{t-1}, \mathbf{u}_{t-2}, \boldsymbol{\alpha}, \sigma^2_\eta] [\mathbf{u}_0] \right\} \\
&\qquad \times \quad [\boldsymbol{\delta} | \boldsymbol{\beta}, \sigma^2_\delta, \theta] [\boldsymbol{\alpha}] [\boldsymbol{\beta}] [\sigma^2_\epsilon] [\sigma^2_\eta] [\sigma^2_\delta] [\theta]
\end{aligned} \tag{4.14}$$

Therefore, a Gibbs sampler was used to sample from the posterior using the relevant full-conditional distributions. Additionally, non-conjugate parameters $\{\boldsymbol{\lambda}_t : t = 1, \ldots, T\}$ and θ, are sampled via Metropolis-Hastings updates within the Gibbs sampler.

4.1.3 Results

This section contains results of the MCMC output and analysis; a discussion of the results follows in the next section. Additionally, for the purpose of illustration only, all remaining maps based on model output are shown as images, where the grid locations correspond to pixel centers. The Gibbs sampler was run for 200,000 iterations with a burn-in of 20,000 iterations to ensure convergence of the parameter chains. The marginal posterior distribution of model parameters by is displayed in histograms constructed from the MCMC samples (Figure 20).

Figure 21 shows the posterior mean and standard deviation of the spatial diffusion coefficient ($\boldsymbol{\delta}$) in the form of a map on the spatial domain. The human population covariate (in $\mathbf{X}$) is also shown in Figure 21 for comparison.

One way to view the change over time in the Poisson intensity parameter is to plot the posterior mean and 95% credible interval at a location of interest. Consider ECD growth at

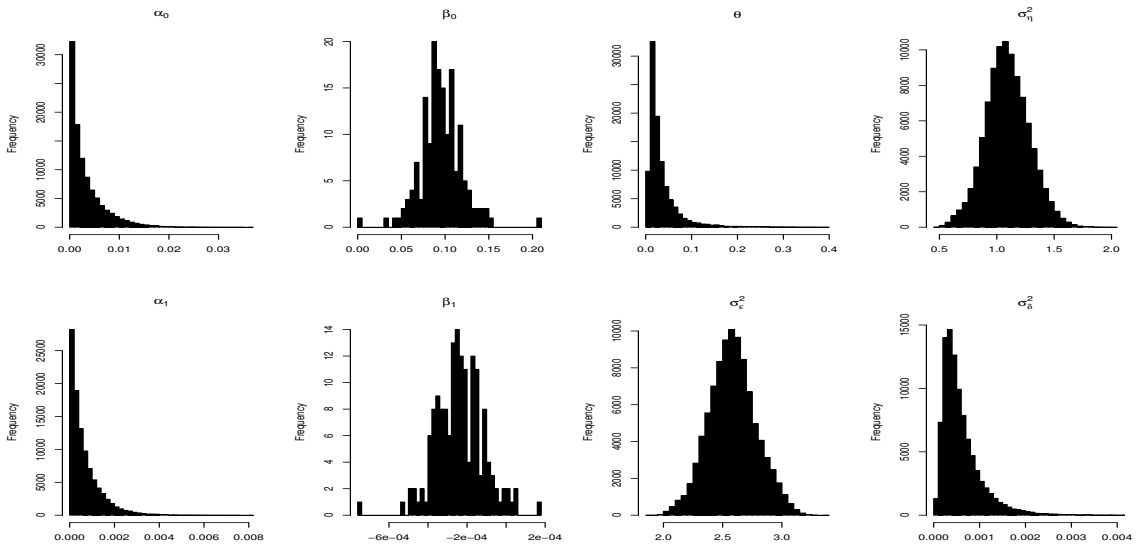
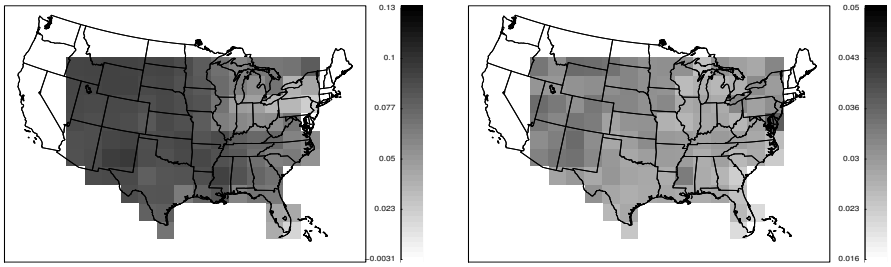
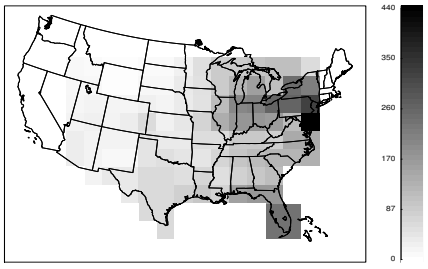


Figure 20: Posterior distributions of bivariate parameters $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and univariate $(\theta, \sigma^2_\varepsilon, \sigma^2_\eta, \sigma^2_\delta)$.



(a) Posterior mean of $\boldsymbol{\delta}$

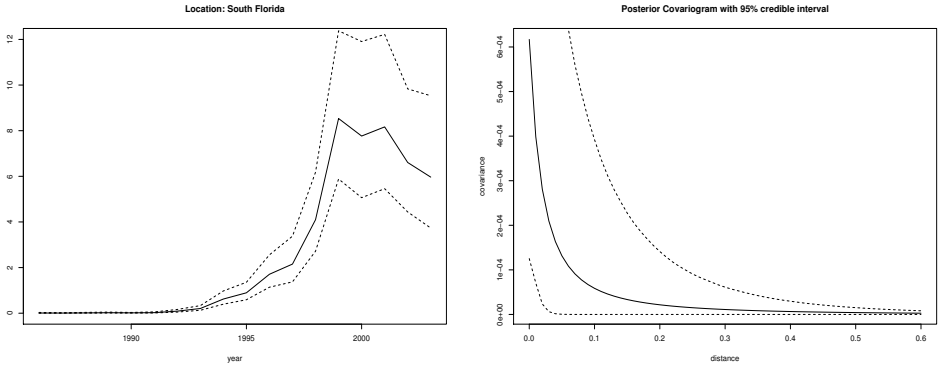
(b) Posterior standard dev. of $\boldsymbol{\delta}$



(c) Human population covariate map (i.e., 2nd column of $\mathbf{X}$).

Figure 21: Posterior summary of $\boldsymbol{\delta}$ and covariate.

a South Florida location (Figure 22). Additionally, by utilizing the MCMC samples from the spatial parameters $(\sigma_\delta^2, \theta)$, a posterior covariogram can be constructed (assuming the covariogram model in (4.8)) with 95% credible interval to provide some idea of the spatial structure in $\boldsymbol{\delta}$ beyond that available in the covariate matrix ($\mathbf{X}$) (Figure 22). Also, the intensity parameters ($\boldsymbol{\lambda}$) can be viewed on the spatial domain as a series of maps over time (Figure 23). Additionally, consider the posterior prediction for $\boldsymbol{\lambda}$ (Figure 24).



(a) Posterior mean (and 95% credible interval) of the Poisson intensity process at a South Florida location through time (represents ECD population growth).

(b) Posterior covariogram with 95% credible intervals (i.e., covariance structure of $\boldsymbol{\xi}$). Maximum distance in the covariogram equals maximum map distance.

Figure 22: Posterior summary of $\boldsymbol{\lambda}$ and $\boldsymbol{\xi}$.

4.1.4 Discussion

It appears that human population density is negatively associated with the rate of diffusion, however the association is only borderline significant (Figures 20, 21), although the intercept coefficient in $\boldsymbol{\beta}$ is clearly significant, suggesting that at least a mean term is important in the model. The negative association is somewhat contrary to popular scientific opinion, although it could be that ECD relative abundance, rather than rate of diffusion, is more

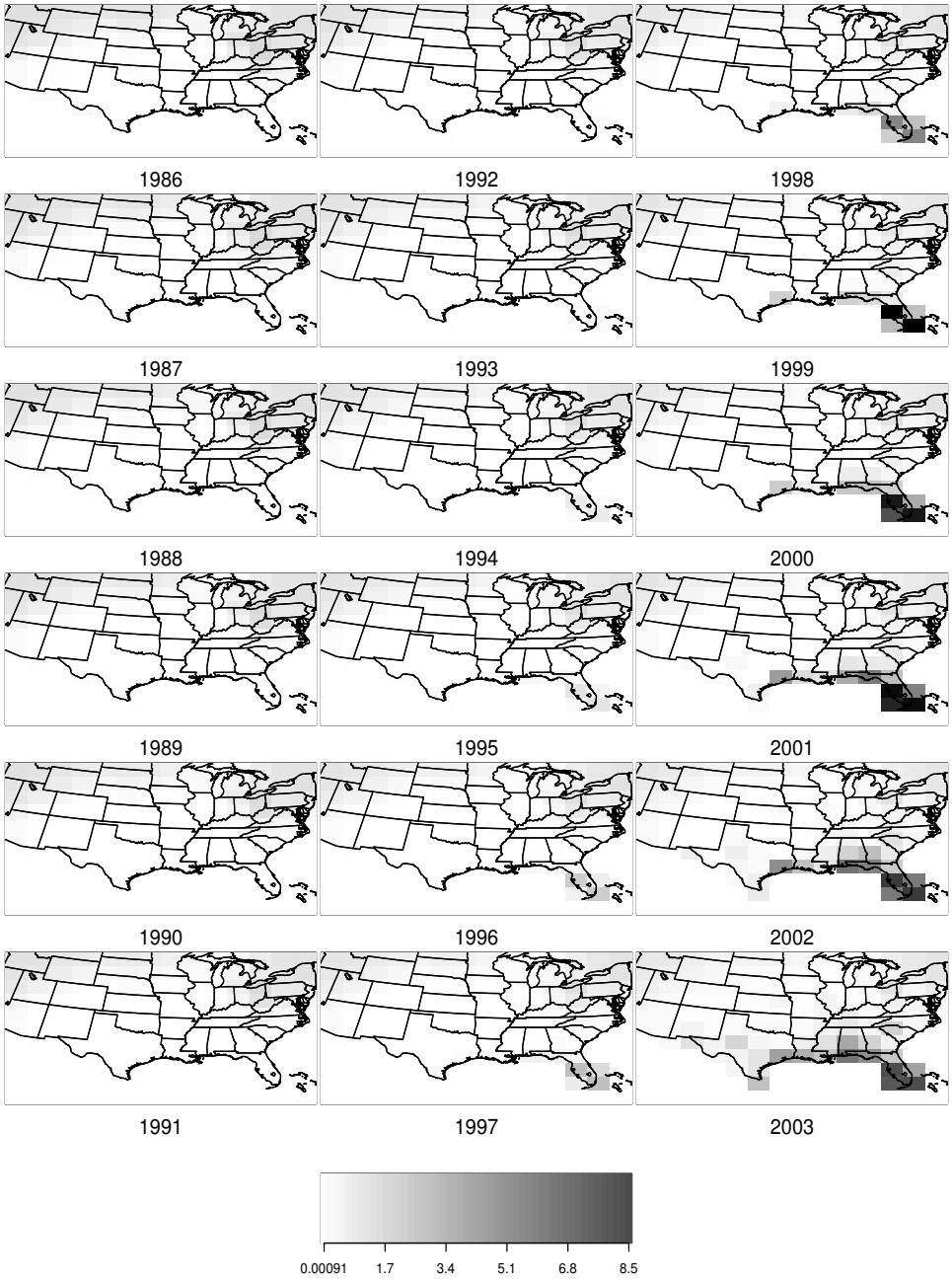
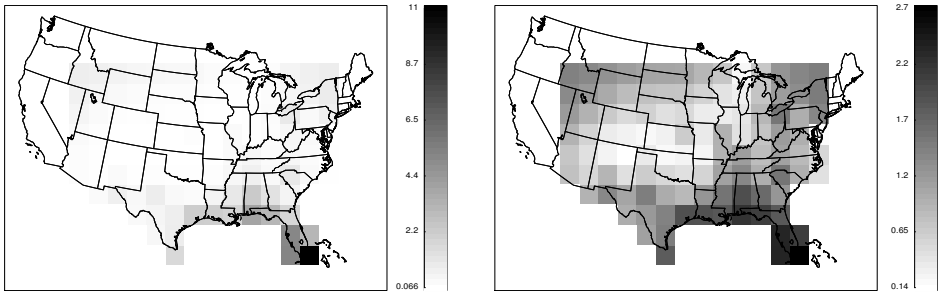


Figure 23: The posterior mean of the Poisson intensity process in time and space.



(a) Posterior mean for Poisson intensity prediction.

(b) Posterior standard deviation for Poisson intensity prediction.

Figure 24: Posterior prediction of $\boldsymbol{\lambda}$ for 2004.

strongly correlated with human proximity. Note also that a few samples of $\boldsymbol{\delta}$ were negative (Figure 21). Technically, $\boldsymbol{\delta}$ should not take negative values, however in this case the model is flexible enough to allow $\boldsymbol{\delta}$ to be negative to make up for other possible deficiencies or misspecifications. Specifically, the human population covariate (Figure 21) contains an outlier and its linear relationship with $\boldsymbol{\delta}$ forces it to be negative. Due to its magnitude, the effect of such an artifact is minimal.

On the other hand, the spatial parameters $(\sigma^2_\delta, \theta)$ influencing the varibility in rate of diffusion are significant (Figure 20) suggesting that diffusion is affected by some underlying process (e.g., an unknown latent spatial covariate such as land type, temperature, or precipitation) and are useful in modeling ECD relative abundance in time and space. The covariogram in Figure 22 illustrates that significant spatial structure exists within approximately one third of the maximum map distance; in fact, most (i.e., 95%) of the spatial structure lies within $1/6^{th}$ of the maximum map distance (i.e., $1/6^{th}$ the width of the United States). This implies that the variability associated with the rate of diffusion that is not

attributed to human population is likely due to some unknown spatial covariate with a range of spatial dependence equal to 0.1 in map distance and (or) due to the biological characteristics of ECD dispersal.

The posterior mean (at most locations) of the Poisson intensity process (e.g., Figure 22) generally shows an increasing trend over time as expected and could indeed be the driving force behind the measurements ($\mathbf{n}_t$). Some locations have an initial decrease in intensity that could be due to an Allee effect (Shigesada and Kawasaki 2002) not present in the original data. Recall that this intensity process is controlled by an underlying latent process ($\mathbf{u}_t$) and thus appears similar (Figure 23) over space and time as σ^2_η is quite small and the growth parameters in the PDE are only weakly influential (Figure 20). Moreover, as with similar models (e.g., Wikle 2003), the process is mostly data-driven and the growth component (although possibly misspecified) allows ample flexibility for modeling the relative abundance as well as non-linear population growth and diffusion.

The prediction for the Poisson intensity process (i.e., mean relative ECD abundance) may be useful for managers wishing to control future range expansion of this species and compare with future monitoring efforts. Notice however that there exists a substantial amount of uncertainty in the prediction (Figure 24), accurate knowledge of which may not be available through other conventional prediction methods. The uncertainty associated with the prediction here is prohibitive of long range predictions. Stonger assumptions in long-range stability and imperfect detection can improve long range detectability. This requires a completely different model specification however, and comes at the cost of computational efficiency (details are provided in Section 4.2).

4.1.5 Conclusion

The ecology of invasive species is complicated, with non-linear non-separable spatio-temporal variation influencing the amount of growth and rate of spread of which observational knowledge is affected by measurement error and various other sources of uncertainty. Implemented in this section is an ecologically meaningful PDE within a hierarchical Bayesian framework as a latent dynamical system to help manage such uncertainty and account for complicated dependence structures in parameters.

Specifically, inference based on this model suggests that not only does correlation exist between ECD abundance in space and time but the growth and spread of this species across North America can be described (even at locations without data) while accurately and responsibly accounting for uncertainty. It is important to note that the reaction-diffusion model for the latent process here is one that represents scientific opinion. However, it is by no means the only way to incorporate scientific knowledge into the model. In fact, matrix models (Caswell 2001) have proven to be a promising (and alternative) tool in invasive species research efforts. Such settings allow for the inclusion of well-known non-linear growth and dispersal equations acting upon a more intuitive latent process. Implementation of these models, however, presents a variety of challenges as discussed in the next section.

4.2 Models for Population Size

4.2.1 Introduction

The growth in population and dispersal of biotic organisms as a function of time has been recognized as an important subject throughout the relatively short history of ecology as a science (Bullock et al. 2002). Ecologists have long been able to accurately describe survival and fecundity in plant and animal populations (Renshaw 1991, chapter 1; Bullock et al. 2002) and have also developed quantitative model-based approaches to study the dynamics of dispersal (Clark et al. 2003; Nathan et al. 2003). Of particular interest are the dynamics of invasive species (Kolar and Lodge 2001). Such nonindigenous animals (and plants) can levy significant impacts on native biotic communities (Elton 1958; Shigesada and Kawasaki 1997, chapter 1). Multiple approaches have been proposed to model invasions, many of which stem from early work by Fisher (1937) and Skellam (1951) in diffusion models and biological waves (Shigesada and Kawasaki 1997, chapter 3). These led ultimately to more complex metapopulation models, matrix models, and state-space models (e.g., Hanski 1999, chapter 12; Caswell 2001, chapter 4; Borchers et al. 2002, chapter 13). Often such models have been considered in the deterministic sense or considered with known stochastic components for purposes of simulation (Renshaw 1991, chapter 2; Kot 2001, chapter A; Turchin 2003, chapter 5; Hastings et al. 2005). Statistical estimation of parameters in such models has conventionally consisted of likelihood-based methods, linear regressions, time-series methods, and more recently, hierarchical approaches (Shigesada and Kawasaki 1997; Caswell 2001; Calder et al. 2003; Buckland et al. 2004; Thomas et al. 2005).

Recent advances in computational efficiency have allowed for the implementation of so-

phisticated hierarchical Bayesian models (e.g., Calder et al. 2003; Clark 2003; Wikle 2003; Su et al. 2004; Buckland et al. 2004; Thomas et al. 2005; Wikle and Hooten 2006; Hooten and Wikle 2006). It has been shown that such models have much to offer, including more precise and less biased parameter estimation (Calder et al. 2003), accounting for multiple sources of uncertainty (Wikle 2003), and the ability to describe non-linear spatio-temporal dispersal and growth in relative abundance (Hooten and Wikle 2006), as well as dynamics of population demographics (Buckland et al. 2004). Many previous studies have been concerned with the rate of spread only (e.g., Arim et al. 2006). However, models specifically formulated for describing the spread of invasive species explicitly on spatio-temporal domains have relied on a state-space framework for implementing discretized partial differential equations to model a log-transformed Poisson intensity which represents "relative" abundance (Wikle 2003; Wikle and Hooten 2006; Hooten and Wikle 2006). Although proven to be informative and useful, these formulations lack an intuitive appeal, as parameter estimates arise from a latent, log-transformed process which may also induce a growth model misspecification (Hooten and Wikle 2006). Conjugate full-conditional distributions highlight the primary mechanical advantage of these models in that large multivariate parameter updates can easily be obtained from an MCMC algorithm. The combination of non-conjugacy, non-Gaussianity, and high-dimensional parameter spaces presents significant implementation challenges. At the same time, such model specifications allow for a more intuitive setting where identification and estimation of posterior parameter distributions comes more naturally.

So-called matrix models (i.e., first-order Markov models) offer a very flexible framework

for describing population dynamics in time, space, and age classes (Caswell 2001). Much can be learned about population dynamics through matrix model simulations. Neubert and Caswell (2000) provide an extensive discussion of demographic matrix models and the effects of demographics on dispersal, however they do not estimate parameters on large spatio-temporal domains explicitly from a model-based statistical perspective. Conventional parameter estimation in such models requires extensive datasets from well-designed experiments (Caswell 2001, chapter 6). Although large-scale spatio-temporal data are becoming more readily available, they are often subject to multiple sources of uncertainty (e.g., observer bias and irregularly sampled locations and times) and not specifically collected to accommodate complicated space-time demographic models (Hastings et al. 2005).

Although much more general as a class of models, matrix models can be specifically employed to model the dynamics of invasive species in space and time and, therefore, such models are the focus of this chapter. Again, the problem is approached hierarchically, with an intuitive data model, flexible matrix process model comprised of growth and dispersal components, and parameter model consisting mostly of non-conjugate priors. This approach has the advantage of appearing less complex than those proposed previously (e.g., Wikle and Hooten 2006; Hooten and Wikle 2006). The general form of the spatio-temporal matrix model, however, is suggested by discretized partial differential (Hooten and Wikle 2006) and integro-difference equations (Wikle 2002b), yet the direct matrix model approach proposed here allows for more general dispersal.

In most ecological studies, the state variable of interest cannot be observed directly because of bias related to imperfect detection. In fact, studies ignoring such sampling as-

sumptions generally focus on modeling "relative abundance" only (e.g., Wikle 2003, Hooten and Wikle 2006). While models for relative abundance are useful, it is often necessary to make direct inference about the size of a population (i.e., numbers of organisms) at a given set of locations. Monitoring efforts for invasive species often result in count data, that is, the number of individuals observed at each location (or area). In cases where less than the total number of individuals is observed (i.e., imperfect detection), a binomial data model is suggested. In general, non-identifiability issues arise when considering such models (Olkin et al. 1981; Carroll and Lombard 1985; Raftery 1988). For cases where prior information about the probability of detection is available, a data model is proposed with identifiable population size parameters that accounts for inherent uncertainty associated with non-detection bias.

As previously mentioned, large scale spatio-temporal ecological datasets consisting of quantitative population information are not common, especially for invasive species (Caswell 2001; Hastings et al. 2005). Fortunately, long-term monitoring efforts such as the North American Breeding Bird Survey (BBS; Robbins et al. 1986) exist and can provide high-dimensional space-time invasive species data (North American BBS data can be obtained at the following website: http://www.pwrc.usgs.gov/bbs/).

As discussed in Section 4.1, the Eurasian Collared-Dove (ECD, *Streptopelia decaocto*) is a non-indigenous species in North America for which data have been collected since its introduction in the early 1980's (Romagosa and Labisky 2000). Although similar in appearance to the Ringed Turtle-Dove, ECD presence is not so benign as it rapidly spread through Europe in the 1900's and poses a possible threat to native ecosystems in North

America (Hengeveld 1993).

The BBS data are collected in a series of routes, at irregular locations, along which the number of birds (of various species) are counted. The linear nature of the routes along which data are collected presents a significant challenge to spatial modeling. Herein it is assumed that the data occur as counts located at the route centers; at a continental scale, it is assumed that the effect of aggregating over a 30-40 km route will be minimal. Additionally, the BBS data are subject to multiple sources of uncertainty, especially within-site variability (Link et al. 1994; Sauer et al. 1994). Knowledge of such intricacies in the data provides a significant motivation for the development of models that explicitly account for known sources of uncertainty, particularly at the data level.

ECD currently occupies the Eastern United States although the invasion is ongoing. Models for relative abundance have suggested that the ECD displays typical invasive behavior (Hooten and Wikle 2006). That is, the underlying dynamics of the invasion can be characterized by common density-dependent population growth and dispersal processes. Romagosa and Labisky (2000) speculated that the ECD would likely colonize the entire United States in a few decades. The approach that follows allows for the rigorous probabilistic evaluation of such claims.

4.2.2 Methods

The general hierarchical model framework for spatio-temporal processes as in Wikle et al. (1998), Calder et al. (2003), Wikle (2003), and Hooten and Wikle (2006), as proposed for time-series by Berliner (1996), is adopted here. The model is constructed in 3 stages: data, process, and parameters. The hierarchical approach allows the for the estimation

of a complex joint distribution via a sequence of simpler, and more intuitive, conditional distributions. This allows for the incorportation of different forms of uncertainty within each level of the hierarchy. The use of square bracket and conditional notation (i.e., $[a|b]$) is used hereafter for simplicity and denotes a conditional probability distribution. That is, $[a|b]$ implies the conditional probability distribution of a given b.

In general, recording counts of individuals (or groups) at various locations over time is a common data collection scheme for monitoring plant and animal abundance. Because animals are usually detected imperfectly in most animal surveys, such data often represents a subset of the true, unknown, number of organisms at each time and location. If the assumption is made that each organism at a given time and location is observed independently *given* that the organism is present, then a binomial data model would be reasonable. Let $n_{i,t}$ represent the observed number of organisms at location i and time t for a total of m spatial locations and T times, while the true, unknown number of organisms at that location and time is denoted $N_{i,t}$. If the probability of observing (i.e., detecting) each organism is θ, then we would have (temporarily) the following data model specification:

$$n_{i,t}|N_{i,t},\theta \sim Binom(N_{i,t},\theta) \qquad i = 1,\ldots,m; \qquad t = 1,\ldots,T.$$

It would be preferable to use heterogeneous detection probabilities, yielding the data model: $n_{i,t} \sim \text{Binom}(N_{i,t},\theta_{i,t})$. However, substantial data would be required to estimate such highly parameterized models in the absence of additional information. One feasible approach is to absorb much of the uncertainty regarding the probability of detection with a distribution on θ.

Although simple point counts are perhaps the most common sampling protocol employed

in avian surveys, they present a problem in the estimation of $N_{i,t}$ when the detection probability (θ) is unknown and must be estimated as well (Royle 2004). The two parameters are not identifiable in the absence of additional information. A minor extension of the sampling protocol can provide information on both θ and $N_{i,t}$, though its implementation specifically involving the ECD is not feasible using BBS observations due to the lack of replicate data. Thus, in order to estimate the true population size ($N_{i,t}$) it is necessary to gain some *a priori* understanding of the probability of detection. Consider separately, a sampling protocol wherein each local population (i.e., each sample unit) is visited on k separate occasions, yielding a sequence of independent binomial counts $n_{i,t}$ (for $t = 1, \ldots, k$). Assuming the true population size, N_i, remains fixed during the period required to collect k observations, the likelihood is the product-binomial,

$$f(n_{i,1}, \ldots, n_{i,k}|N_i, \theta) = \prod_t \frac{N_i!}{(N_i - n_{i,t})! n_{i,t}!} \; \theta^{n_{i,t}} \; (1-\theta)^{N_i - n_{i,t}}. \tag{4.15}$$

The general instability of the MLE for N_i based on (4.15) is well-known (Olkin et al. 1981; Carroll and Lombard 1985; Raftery 1988), and this deficiency may explain why the sampling protocol is not widely used in practice. However, when this sampling protocol is applied at replicate spatial locations and additional model structure is imposed on the location-specific abundance parameters, the estimation problem becomes more tractable (Royle 2004). For example, if $N_i \sim \text{Poisson}(\lambda)$, then λ and θ are well-identified.

This approach can be used in the presence of replicate data to obtain a posterior distribution for θ under a Bayesian implementation of (4.15) (Royle and Dorazio 2006). As previously mentioned, in the case of the ECD, the operational BBS survey does not yield replicate counts at each BBS sample location. However, a study conducted by Link et al.

(1994) obtained replicate counts on a number of BBS routes within the ECD range. Specifically, we used replicate data from a different species (Mourning Dove, *Zenaida macroura*) that shares similar detectability characteristics with ECD to inform θ. The posterior for the detection probability of the Mourning Dove is the standard form of a beta distribution with parameters α_θ and β_θ. The estimate of θ (in terms of α_θ and β_θ) is not expected to represent the detectability of the ECD exactly, however, based on similarities in dove behavior it is expected to be a reasonable estimate; especially considering that the uncertainty in θ is allowed through the beta distribution.

The information about detection probability based on the Mourning Dove data provides the best estimate of θ attainable given available data and methods. In fact, given the identifiability issues between $N_{i,t}$ and θ nothing further can be learned about the probability of detection in a hierarchical spatio-temporal model without additional replicate data. However, by allowing the Mourning Dove estimate to be the prior for θ and then treating it as a nuisance parameter, it can be integrated out of the data model while accounting for its uncertanty (Berger et al. 1999). This suggests the following modified data model,

$$[n_{i,t}|N_{i,t}] = \int_0^1 [n_{i,t}|N_{i,t},\theta][\theta]d\theta \qquad i = 1,\ldots,m; \qquad t = 1,\ldots,T, \tag{4.16}$$

which analytically yields a beta-binomial data model when the prior distribution for θ is indeed specified as a beta distribution (i.e., $\theta \sim Beta(\alpha_\theta, \beta_\theta)$):

$$n_{i,t}|N_{i,t},\alpha_\theta,\beta_\theta \sim \text{Beta-Binom}(N_{i,t},\alpha_\theta,\beta_\theta) \qquad i = 1,\ldots,m; \qquad t = 1,\ldots,T. \tag{4.17}$$

Hence, this beta-binomial model is used hereafter as the data model component in the hierarchical matrix model.

Although the estimation, spatial prediction, and temporal forecasting of population size are of primary concern in this setting, the underlying ecological process is also of interest. That is, knowledge about the true process from which the data arise is useful for characterizing and studying the invasion. Many previously proposed methods split the process model into two stages where the first stage is a non-linear (and hence non-conjugate) transformation of the second stage (a Gaussian state equation). This provides a computational advantage yet detracts from the intuitive nature of the process itself. A process model with intuitive construction while retaining scientifically meaningful dynamic behavior is desirable.

First to define the process, a natural and common model for the true number of individuals at a given location and time is adopted. That is, the true population ($N_{i,t}$) is assumed to be conditionally Poisson with intensity ($\lambda_{i,t}$) depending hierarchically on population growth and dispersal parameters.

$$N_{i,t}|\lambda_{i,t} \sim Pois(\lambda_{i,t}) \qquad i = 1, \ldots, m; \qquad t = 1, \ldots, T.$$

Assuming the Poisson intensity ($\boldsymbol{\lambda}_t \equiv [\lambda_{1,t}, \ldots, \lambda_{m,t}]'$) evolves according to first order Markovian dynamics, allows for the utilization of a matrix model framework (Caswell 2001, chapter 4). In the specific setting where information about the age structure of a species is lacking, $\boldsymbol{\lambda}_t$ can be defined as the state vector, being propagated by a transition matrix $\mathbf{H}$

in the following evolution equation.

$$\begin{aligned} \boldsymbol{\lambda}_t &= \mathbf{H}\boldsymbol{\lambda}_{t-1} \\ &= \mathbf{MG}\boldsymbol{\lambda}_{t-1} \\ &= \mathbf{M}(\boldsymbol{\tau})\mathbf{G}(K, r, \boldsymbol{\lambda}_{t-1})\boldsymbol{\lambda}_{t-1} \qquad t = 1, \ldots, T. \end{aligned} \tag{4.18}$$

Notice in (4.18) that the transition matrix $\mathbf{H}$ is decomposed into a product of two distinct $m \times m$ matrices corresponding to growth ($\mathbf{G}$) and movement ($\mathbf{M}$) of the population. This formulation allows the process, which is random through its parameters (i.e., K, r, $\boldsymbol{\tau}$, and $\boldsymbol{\lambda}_1$ are all probabilistically specified in the next section) and quite flexible, to evolve in a conventional dynamical system. Additionally, it provides a convenient and ecologically meaningful specification of the growth and dispersal parameters. Similar decompositions of the transition matrix H in settings where population demographics is also being studied have been discussed (e.g., Neubert and Caswell 2000; Buckland et al. 2004), though the specification presented here differs from those used previously in order to accommodate the specific ECD dataset as well as prediction of a high-dimensional spatially explicit process.

The matrix controlling population growth ($\mathbf{G}$) can be parameterized many ways. A diagonal form is proposed here, where population growth at each location is described by a common density dependent formulation given by the Ricker growth equation (Turchin 2003, chapter 3). Although other growth equations exist and are commonly used (e.g., Beverton-Holt, Malthusian, Gompertz), the Ricker form of growth is flexible and allows for potentially chaotic behavior and oscillatory characteristics in population stability (Kot 2001, chapter A). Additionally, investigators wishing to perform inference on the form of growth model

specifically could implement various process models separately and compare their effect using any number of model selection criteria (e.g., Deviance Information Criterion or Bayes factors). Another approach would be to consider multiple growth models simultaneously using a mixture-based specification for the process model and reversable jump MCMC. In simulation, the Ricker growth model with random parameters, has proved to be very flexible and capable of exhibiting behavior similar to that of other forms of theoretical population growth. Thus, here the focus is on Ricker growth and the i^{th} diagonal element of $\mathbf{G}$ can be defined as,

$$G(r, K, \lambda_{i,t}) = \exp\{r\ (1 - \lambda_{i,t}/K)\}, \qquad t = 1, \ldots, T, \tag{4.19}$$

where K and r correspond to carrying capacity and growth rate, respectively. In cases where there are significant differences in the carrying capacity (K) and growth rate (r) across spatial locations, one could consider parameterizations where K and r are allowed to vary in space. Effective estimation of spatially varying growth rates may be difficult in models already heavily parameterized without informative prior knowledge. In the presence of data describing a non-transient process (i.e., a process where population size has reached carrying capacity at all locations) it may be possible to estimate spatially varying carrying capacities (K); in cases where a population is still growing, however, these parameters will be difficult to estimate.

The dispersal of organisms in space is described in (4.18) by the matrix $\mathbf{M}$. Dispersal in this case is assumed to be characterized by a spatially varying Gaussian kernel. That is, the population in a given location at the current time is a spatially weighted average of the

population at surrounding locations depending on the Euclidean distances ($d_{i,j}$) between locations and dispersal parameters, $\boldsymbol{\tau}$. We let $\mathbf{M}$ be defined as:

$$\mathbf{M}(\boldsymbol{\tau}) = [(M_{i,j})]_{m\times m}, \tag{4.20}$$
$$M_{i,j} \propto \exp\{-\frac{d_{i,j}^2}{\tau_j}\},$$

where each row of $\mathbf{M}$ is constrained to sum to one with each value between 0 and 1, as motivated by discretized diffusion equations (e.g., Wikle 2002b, Hooten and Wikle 2006).

Other parameterizations of this dispersal matrix could be considered. For example, an alternative specification of the dispersal matrix, say $\mathbf{M}^{(\text{alt})}$, is one where each column (rather than row) of $\mathbf{M}^{(\text{alt})}$ is constrained to sum to one with $0 \leq M_{i,j}^{(\text{alt})} \leq 1$ for all i and j. This dispersal matrix is especially suited to the "multisite" situation where the $M_{i,j}^{(\text{alt})}$ refer to transition probabilities. In such a situation, where individual animals are moving from one distinct "population" to another with a certain probability, it is sensible to think of the transition probabilities as: $M_{i,j}^{(\text{alt})}$ = P(animal moves to location i | animal is currently in location j). In a spatial setting, with data pertaining to irregularly observed locations on a landscape where there are open boundaries (as is typically the case in invasion problems), a dispersal model needs to have the flexibility that the quantity being dispersed can move out of the domain through the boundary. To help visualize the difference between $\mathbf{M}$ and $\mathbf{M}^{(\text{alt})}$, consider the simplest case, where $\mathbf{G} = \mathbf{I}$ (i.e., no growth in population); here $\mathbf{M}^{(\text{alt})}$ preserves the number of total organisms over time and only allows them to occur in distinct areas, whereas $\mathbf{M}$ does not account for every organism at each time and thus the total number of organisms may fluctuate over time even though there is no growth component in the model. This effect of parameterization implies that when $\mathbf{G} \neq \mathbf{I}$, the parameters in

$\mathbf{M}$ and $\mathbf{G}$ may not be completely identifiable since they could both be contributing to the overall growth in population. However, in the situation considered here, with the Eurasian Collared-Dove and BBS count data, the preservation of population size resulting from the "probability parameterization" of $\mathbf{M}^{(\text{alt})}$ is not as important as the dispersal behavior, thus the parameterization of $\mathbf{M}$ as in (4.20) is retained. Though the chosen specification for $\mathbf{M}$ implies that the growth (i.e., K and r) and dispersal (i.e., $\boldsymbol{\tau}$) parameters may not be separable, simulations verify that for the spatial and temporal domains considered in this application, the contribution from $\mathbf{M}$ to overall growth is minimal. Still, rigorous interpretation of growth and dispersal parameters, separately, could be misleading, and thus it is best to interpret the effect of K, r, and $\boldsymbol{\tau}$ simultaneously. Here, the emphasis is on the prediction of true population size ($N_{i,t}$) in time and space and thus there is no need to interpret the parameters separately, however, the posterior results of these parameters are provided for the sake of completeness.

As with the growth parameters, one could specify this Gaussian kernel to be spatially homogeneous in cases where limited data exist. For the application considered here, a multivariate parameterization has been shown to accommodate regional differences in dispersal (Hooten and Wikle 2006) when adequately estimated. The dispersal mechanism employed here is more flexible than those previously proposed, thus the patterns of dispersal are expected to be different. In fact, since the emphasis is on prediction and estimation of the true population size ($N_{i,t}$), by specifying a simpler model with the constraint that all τ_i are equal, the risk of assuming homogeneous dispersal can be illustrated. Additionally, model complexity and effectiveness can be compared using DIC (i.e., Deviance Information

Criteria; see Gelman et al. 2004, chapter 6). Though no methods of model comparison and selection are without criticism, DIC focused on the most important level in complex hierarchical models (such as the Data Model in this case) can provide a means of comparing the overall effect of different parameterizations while taking into account model complexity.

The underlying process (4.18) in this hierarchical model is a dynamical system with random components. Conventionally, latent processes are formulated with additive error to account for misspecification of the model. Such specifications are very general and often result in non-identifiable variance components (e.g., Hooten and Wikle 2006) and prohibit long-range prediction due to additive error propagating through the predictive forward model. Accounting for the process model error through probabilistic growth and dispersal parameter specifications alleviates such complications while offering sensible flexibility in the latent process.

With the nuisance parameter, θ, integrated out, a very simple specification is proposed for prior distributions of the remaining model parameters. Assuming independence between priors (i.e., $[K, r, \boldsymbol{\tau}, \boldsymbol{\lambda}_1] \propto [K][r][\boldsymbol{\tau}][\boldsymbol{\lambda}_1]]$), let:

$$K \sim \text{Gamma}(\alpha_K = 4, \beta_K = 50),$$

$$r \sim \text{N}(\mu_r = 0.5, \sigma_r^2 = 0.25),$$

$$\log \begin{pmatrix} \boldsymbol{\tau} \\ \tilde{\boldsymbol{\tau}} \end{pmatrix} \sim \text{N}\left(\begin{pmatrix} \boldsymbol{\mu}_\tau \\ \tilde{\boldsymbol{\mu}}_\tau \end{pmatrix}, \boldsymbol{\Sigma}_\tau = \begin{pmatrix} \boldsymbol{\Sigma}_\tau^{1,1} \boldsymbol{\Sigma}_\tau^{1,2} \\ \boldsymbol{\Sigma}_\tau^{2,1} \boldsymbol{\Sigma}_\tau^{2,2} \end{pmatrix} \right), \tag{4.21}$$

$$\log \begin{pmatrix} \boldsymbol{\lambda}_1 \\ \tilde{\boldsymbol{\lambda}}_1 \end{pmatrix} \sim \text{N}\left(\begin{pmatrix} \boldsymbol{\mu}_\lambda \\ \tilde{\boldsymbol{\mu}}_\lambda \end{pmatrix}, \boldsymbol{\Sigma}_\lambda = \begin{pmatrix} \boldsymbol{\Sigma}_\lambda^{1,1} \boldsymbol{\Sigma}_\lambda^{1,2} \\ \boldsymbol{\Sigma}_\lambda^{2,1} \boldsymbol{\Sigma}_\lambda^{2,2} \end{pmatrix} \right), \tag{4.22}$$

where $\tilde{\boldsymbol{\tau}}$ and $\tilde{\boldsymbol{\lambda}}_1$ are the parameter values at a set of locations where predictions are de-

sired. Also, the Gamma distribution on K has a mean 200 and variance of 10,000 and is parameterized such that β_K is the scale parameter. $\mathbf{\Sigma}_\tau$ and $\mathbf{\Sigma}_\lambda$ are characterized by exponential spatial hyperpriors on the covariance structure of the dispersal and intensity process at the initial time (i.e., $t = 1$), respectively. Specifically, $[\mu_\tau]_i = [\tilde{\mu}_\tau]_j = \log(2), \forall i, j$ and $[\mathbf{\Sigma}_\tau]_{i,j} = \log(1.1)[\exp(-2||d_{i,j}||)]$ represent the hyperpriors for $\boldsymbol{\tau}$ while $\mu_\lambda = \tilde{\mu}_\lambda = \log(15)$ for locations in extreme southern Florida (i.e., the location where ECD was introduced into the United States) with zero elsewhere and $[\mathbf{\Sigma}_\lambda]_{i,j} = \log(1.1)[\exp(-2||d_{i,j}||)]$ represent the hyperpriors for $\boldsymbol{\lambda}_1$. The Euclidean distance between locations i and j is represented by $||d_{i,j}||$. The spatial specification of $\boldsymbol{\tau}$ and $\boldsymbol{\lambda}_1$ allows for posterior predictive estimation at locations and times without data as well as likely autocorrelation in the parameters (see Appendix). Hyperpriors are specified based on their ability to exhibit reasonable growth and dispersal behavior in simulations. For example, it is known that the population size is generally growing over time and thus μ_r was specified to be positive, although r could still be negative depending on σ_r^2 and the influence of the data. Recall that prior specification for θ (i.e., $\theta \sim Beta(1.7, 7.7)$, implying a mean of 0.18 and standard deviation of 0.12) was based on posterior estimates of detection probability from a separate model and dataset with repeated measurements and then integrated out, as discussed earlier.

4.2.3 Simulation

To partially evaluate the effectiveness of the proposed hierarchical model, realizations can easily be obtained and then utilized as simulated data with which to fit the model. Here, due to the limiting computational requirements, focus is placed on simulating data for a similar spatio-temporal domain with similar growth rates and hypothesized carrying capacity as

well as probability of detection for the ECD. Consider a one-dimensional spatial domain where upon introduction of a number of organisms to the center of the domain at the initial time, the process is then allowed to evolve forward 16 time units based on random variates from the prior distributions for the ECD population growth and dispersal parameters. Observational data ($\mathbf{n}_t$) is then sampled according to the prior detection probability for ECD (θ) at 50 spatial locations in the 1-D domain at each time iteration for use in estimating the true population sizes ($\mathbf{N}_t$) at those locations as well as predicting $\mathbf{N}_t$ at a set of 50 new locations in the spatial domain at which the "true" population size is known. The one-dimensional spatial domain can be thought of here as some linear environment such as a river segment or sampled transect across a landscape.

Figure 25 shows both the "true" population sizes at all simulated locations for times $t = \{3, 6, 9\}$ as well as the posterior predictive 95% credible envelopes across all locations. The growth curves are displayed as time-series in Figure 26 for two locations (one in the middle of the spatial domain and one near the edge) for times $t = 1, \ldots, 16$.

The credible intervals in Figures 25 and 26 illustrate that given relatively limited data, the estimation of $\mathbf{N}_t$ is quite good at all spatial locations and times. It appears that the model is underestimating the population sizes near carrying capacity (Figure 25) however this is due to the lack of precise prior information about the probability of detection and was verified by fixing θ at a known quantity in other simulations (not included). In general, the simulations provide evidence that the model is able to probabilistically characterize "invading" phenomena in the presence of uncertainty. This is by no means an exhaustive simulation study, though it does illustrate that the model is useful in at least some settings

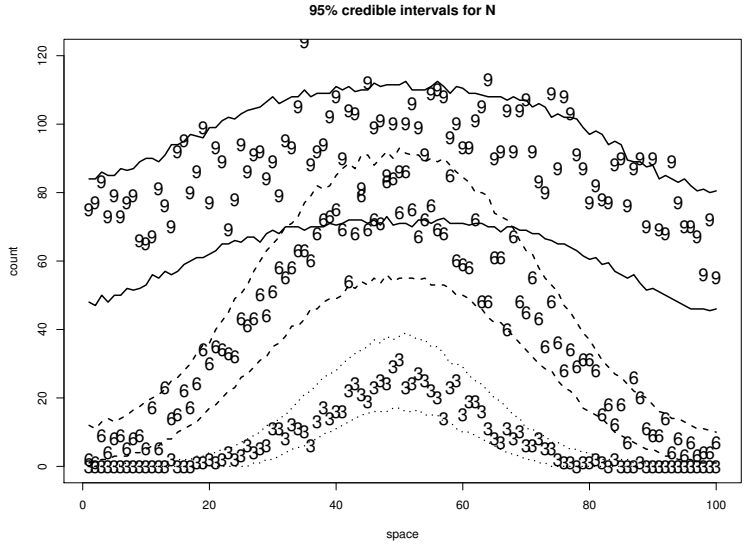


Figure 25: 95% Credible envelopes for posterior simulated population size ($N_{i,t}$) for all locations at times $t = 3$ (dotted lines), $t = 6$ (dashed lines), and $t = 9$ (solid lines). The "true" population sizes (unknown in the model) are displayed as numeric values representing their time unit where height on the y-axis represents magnitude of population size (in counts) and placement on the x-axis denotes spatial location.

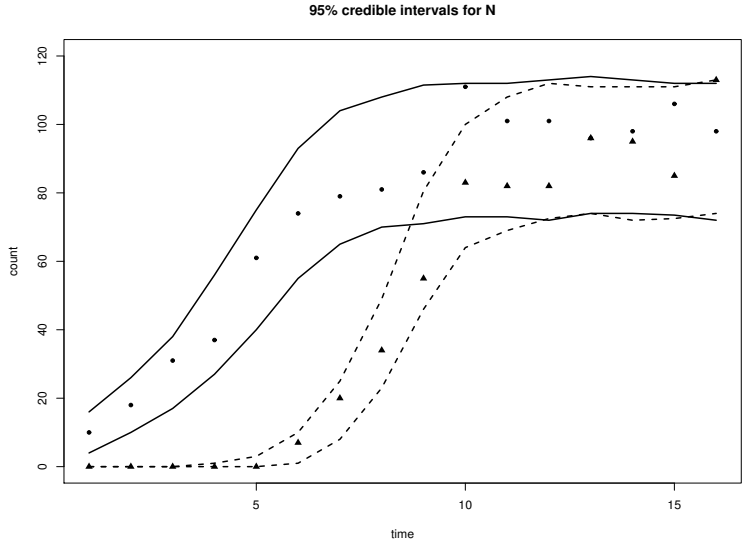


Figure 26: 96% Credible envelopes for posterior simulated population growth ($N_{i,t}$) from times $t = 1, \ldots, 16$ at locations in the middle (solid lines) and near the edge (dashed lines) of the spatial domain, simultaneously. "True" population sizes ($N_{i,t}$) are denoted by circles and triangles.

(especially in settings similar to that of the ECD).

4.2.4 Results and Discussion

The previous section suggests that the model performs relatively well for estimating phenomena with similar characteristics and on similar spatio-temporal domains as the ECD. Now consider the results of fitting the model to the ECD data specifically.

Methods for estimating the posterior distribution for the models considered here include Markov Chain Monte Carlo (MCMC) and various forms of importance sampling (Gelman et al. 2004, chapter 11). Importance sampling is especially desirable in settings with non-conjugate full-conditional distributions and non-linear relationships between parameters, however suffers from degeneracy problems, especially in situations with high-dimensional parameter spaces, as is the case here. Metropolis-Hastings updates within a Gibbs Sampler (one form of MCMC) is an alternative approach for sampling, but requires careful selection of proposal distributions that can sufficiently explore the parameter space. An MCMC approach for estimation is applied here although other sampling algorithms may also provide adequate results.

Several multivariate Metropolis-Hastings parameter updates slow the mixing of MCMC samples, thus the Gibbs sampler was allowed to run for 200,000 iterations to insure that the posterior parameter space was explored sufficiently. Resampling was used to remove correlation between realizations and convergence, which was assessed by visually inspecting parameter chains, occurred before the burn-in period of 20,000 iterations which were used for calculation of posterior summary statistics. Metropolis-Hastings acceptance rates for the parameter updates varied between 20% and 40%. For additional details about the sampling

algorithm and predictions, see the Appendix to this chapter.

Estimation of the posterior distributions of model parameters provides useful information about the underlying dynamics of the ECD invasion. Although the focus here is on the estimation and prediction of the population size at various locations and times, knowledge about the process driving the population size can also be informative. In this setting population dispersal can be evaluated as well as growth parametrically, though, inference on dispersal and growth parameters, separately, should only serve as a general guideline since the degree of separation is unknown. However, the combined effect of growth and dispersal is valid and can be assessed by interpreting the process itself (i.e., $\boldsymbol{\lambda}_t$). Simulated processes on similar spatial and temporal domains suggested that the shared contribution from K, r, and $\boldsymbol{\tau}$ to overall population size is negligible (as evidenced by only a slight bias in the posterior estimates of parameters based on simulated data), however, the following posterior summaries are provided with a cautionary warning.

Posterior distributions for the univariate parameters K and r yielded 95% credible intervals of $[103, 201]$ and $[0.26, 0.32]$, respectively. Posterior standard deviations for K and r (i.e., $\sqrt{V(K|n)} = 24.3$, $\sqrt{V(r|n)} = 0.02$) were very precise compared to their prior standard deviations (i.e., $\sqrt{V(K)} = 100$, $\sqrt{V(r)} = 0.5$), thus implying that there was sufficient data to inform the growth of the process.

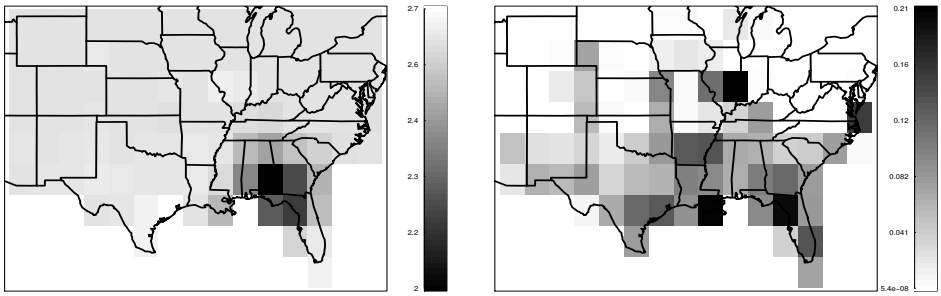
Consider, first, the posterior 95% credible interval for the Ricker growth rate (r). In situations where there is no change in population size due to dispersal, a growth rate of $r < 0$ implies a decrease in population while $r > 0$ implies an increase. Here r does not overlap zero, implying that there is indeed growth in the population, in addition to dispersal. In

this case, where dispersal contributes to population size, and since an invasion generally constitutes a growing population, we can expect r to be positive but smaller than it would be in a population without dispersal. Also, recall from (4.19) that when $r \to 1$ it implies that $G(K, r, \lambda_{i,t}) \to \exp\{1 - \lambda_{i,t}/K\}$; a much simpler growth model. Here, the posterior credible interval of r suggests that it is significantly different from 1, thus the assumption of a two parameter growth equation is reasonable.

The posterior credible interval for the carrying capacity (K) suggests that the number of individuals a given location can support is quite variable. This is not surprising given the lack of prior information about this parameter and the fact that the invasion is still ongoing and has not reached saturation at all locations. The variability in the posterior distribution for K accommodates the variability induced by possibly heterogeneous (i.e., spatially explicit) carrying capacities. For situations where the population size at more locations had reached carrying capacity, the variability in K could likely be reduced by conditioning on spatial covariates.

Statistics from the posterior distribution of multivariate parameters can be expressed as spatial maps. Prediction and estimation at a set of regularly spaced locations provides a convenient way to view the posterior parameter spaces. Note that all figures containing maps herein are displayed as images with grid cell intensity representing parameter magnitude at the location of the grid cell center. The image format is for visualization only (as an alternative to variable sized points) and should not be misinterpreted as "areal" support. That is, each pixel in the images represents a point at the pixel center, rather than the area of the pixel itself.

Consider the posterior mean and standard deviation maps of dispersal $\boldsymbol{\tau}$ (Figure 27). As with the growth parameters, the posterior standard deviations for dispersal were much smaller than the prior standard deviation (i.e., $\max(V(\tau_i|n)) << V(\tau_i)$). From the parameter estimates in Figure 27 it is evident that there is a clear and significant area of low dispersal in the region of northern Florida (in the southeast portion of the map in Figure 27 a). The effect of this will slow the spread of ECD westward out of Florida while northward, along the east coast, ECD dispersal remains uninhibited. Note that the values for the mean parameter estimates of $\boldsymbol{\tau}$ are directly related to the scale of the map distance units, and therefore the general pattern rather than the magnitude should be interpreted here. That is, the use of a different coordinate system would result in different parameter values, but retain the same pattern.



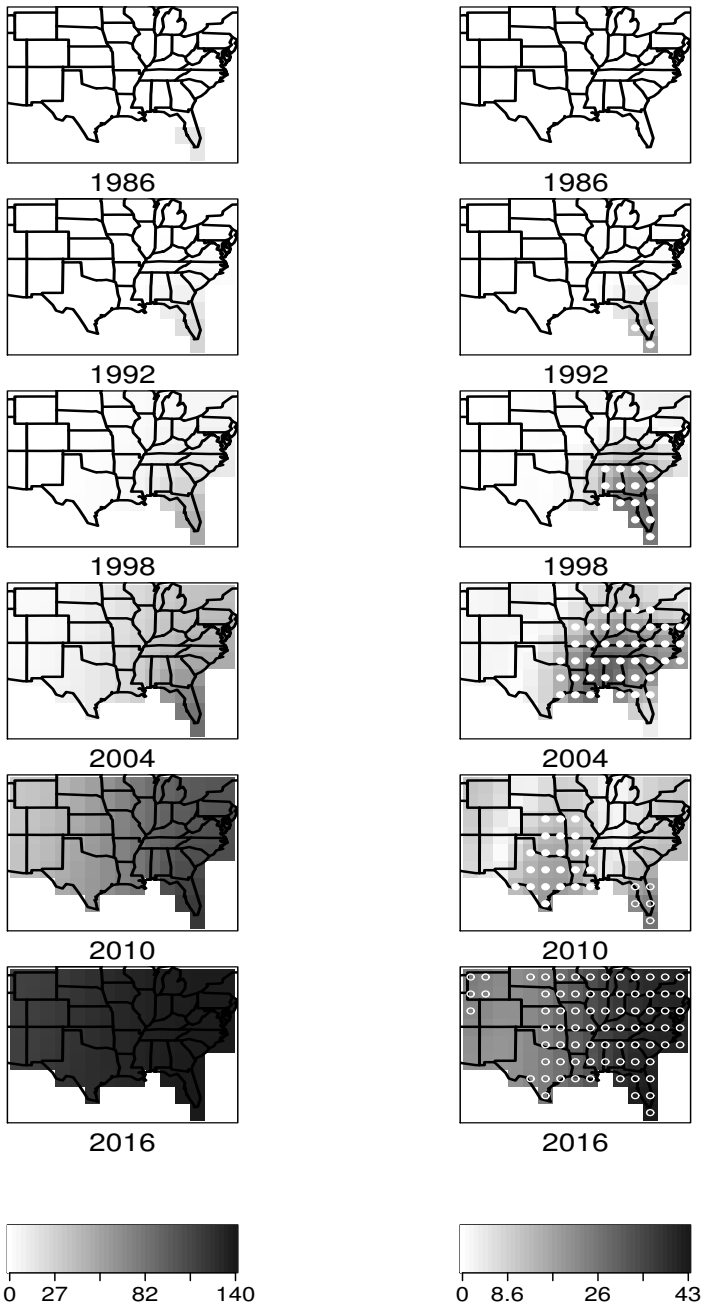
(a) $E(\boldsymbol{\tau}|n_{i,t}), \forall i, t$ (b) $\sqrt{V(\boldsymbol{\tau}|n_{i,t})}, \forall i, t$

Figure 27: Maps representing the posterior mean ($E(\boldsymbol{\tau}|n_{i,t}), \forall i, t$) and standard deviation ($\sqrt{V(\boldsymbol{\tau}|n_{i,t}}, \forall i, t$) of dispersal.

A primary concern of ecologists and resource managers related to invasive species is the extent and magnitude of the invasion at a given time (including future times). In the case of the ECD, the mean posterior population size presented in Figure 28a illustrates the

estimated magnitude and extent of the invasion at a set of times (including predictions) given the sample data. By the year 2016, the ECD population is predicted to be at or near carrying capacity for the portion of the United States shown in the figures. Posterior credible interval growth curves for individual locations can reveal at what time the population is projected to reach its carrying capacity. Consider 95% credible intervals for a location in south Florida and a location in northwest Wyoming (i.e., Northwest United States) simultaneously (Figure 29). At the time of this writing, very few, if any, ECDs have likely dispersed as far as Wyoming, whereas the population sizes in most of Florida are nearly at carrying capacity. Although, by the year 2020, Wyoming will be near carrying capacity with high probability as well, as speculated by Romagosa and Labisky (2000).

By implementing a simpler form of the model with homogeneous dispersal rates (i.e., $\tau_i = \tau, \forall i$) any advantage spatially varying dispersal contributes to estimation and prediction can be assessed. Using DIC (i.e., Deviance Information Criterion; e.g., Gelman et al. 2004, chapter 6) focused on the data model, the two models can be compared in terms of their predictive capacity while penalizing for model complexity. Denote model 1 as the proposed model allowing for heterogeneous dispersal and model 2 as the simplified model with homogeneous dispersal. Then, the resulting number of effective parameters for each model is: $pD_1 = 133$ and $pD_2 = 130$. The DIC for each model is: $DIC_1 = 13052$ and $DIC_2 = 13107$. This suggests that after correcting for the number of unconstrained parameters, model 1 still provides better predictions due to the lower DIC value. Additionally, Figure 28b illustrates the relative differences between models and shows how assuming homogenous dispersal can lead to inflated and deflated population size estimates in both time



(a) $E(\mathbf{N}_t^{(1)}|\mathbf{n})$

(b) $|E(\mathbf{N}_t^{(2)}|\mathbf{n}) - E(\mathbf{N}_t^{(1)}|\mathbf{n})|$

Figure 28: (a) Posterior (and posterior predicted) mean of ECD population size throughout the United States from 1986 through 2016. (b) Also, the absolute difference between posterior means for $\mathbf{N}_t^{(2)}$ and $\mathbf{N}_t^{(1)}$ (i.e., the difference in estimates of population size for the model with homogeneous dispersal versus the model with heterogeneous dispersal). Overall shade corresponds to absolute difference and solid white circles represent large positive differences while empty white circles represent large negative differences.

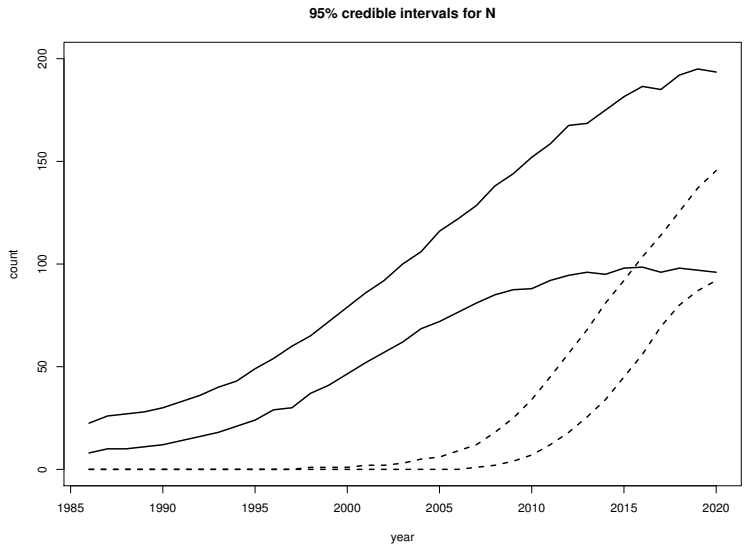


Figure 29: Credible intervals for posterior population growth of ECD ($N_{i,t}$) from 1986 through 2020 for locations in south Florida (solid lines) and northwest Wyoming (dashed lines) simultaneously.

and space. In Figure 28b the areas with solid white circles imply that model 2 (homogeneous dispersal) is overestimating the population size relative to the predictions resulting from model 1, while the areas with empty white circles indicate that model 2 is underestimating population size relative to the predictions resulting from model 1. These areas of mis-estimation clearly evolve over time and can be quite misleading. Overall, the DIC analysis and difference maps suggest that a spatially varying dispersal parameter is indeed important in the model.

4.2.5 Conclusion

The estimation and prediction of population size (i.e., the N parameter of the binomial distribution, Olkin et al. (1981); Carroll and Lombard (1985); Raftery (1988)) is a long studied problem. By using *a priori* knowledge gained from models proposed by Royle (2004)

and Royle and Dorazio (2006), it is possible to estimate parameters that characterize a non-linear and non-Gaussian dynamical system underlying the dispersal of invasive species. This use of subjective probability illustrates a strength of Bayesian models in complicated hierarchical settings where conventional likelihood based parameter estimation would not be feasible. From a scientific perspective, it is important to note that previous dynamical spatio-temporal population models have largely ignored the detectability of the species, whereas here we are directly accounting for detection as well as uncertainty associated with detection. In the specific case of the ECD, this represents the only way to gain an understanding of the spatio-temporal dynamics in true population size (i.e., $N_{i,t}$).

The model proposed in this paper adopts the intuitive and simple yet powerful form of a matrix model (e.g., Caswell 2001). Although information about the age distribution of ECD over time and space are not available, this model framework could easily accommodate such data. In fact, several approaches have been taken to implement the demographic version of matrix models in a hierarchical framework (e.g., Buckland et al. 2004, Thomas et al. 2005). The dynamical system underlying the population size in this space-time setting is specified in two components, a density-dependent growth component and a dispersal component and operates on an untransformed Poisson intensity that is random in its parameters. This provides a very flexible non-linear dynamical process model capable of chaotic and oscillatory behavior while sensible prior distributions can be chosen to account for uncertainty in the parameters. Also, as with most hierarchical models, uncertainty associated with predictions can easily be obtained and interpreted and model-based validation methods are readily available. It is important to note that this form of the proposed model relies heavily

on the estimated dynamics of the invasion in order to predict population size at future times and unsampled locations. For cases where there exists more complete data for less of a generalist species, the model could easily be extended to accommodate covariate effects on model parameters, such as growth rate ($\mathbf{r}$), carrying capacity ($\mathbf{K}$), and dispersal ($\boldsymbol{\tau}$). In cases where such information is available, predictions could also serve as a way to evaluate scenarios on how to control the spread of invasions.

The main advantage of this model over other similar invasive species models is that it is intuitive and simple yet flexible enough to accommodate various specifications and sources of uncertainty (e.g., such as the uncertainty in detection). It provides tangible graphical output (in addition to numerical output) and is based on rigorous data-driven statistical methodology, the basic principles of which ecologists and resource managers can understand and utilize for scientific endeavor and decision making.

4.2.6 Appendix: Full-Conditionals

One approach to estimation of the joint posterior distribution in this hierarchical setting requires iterative sampling from the full-conditional distributions. That is, sample K, r, $\boldsymbol{\tau}$, $\boldsymbol{\lambda}_1$, and $N_{i,t}$ for $i = 1, \ldots, m$ and $t = 1, \ldots, T$, sequentially, from their respective conditional distributions given all other parameters. Here, the full-conditionals are non-conjugate and thus must be sampled via Metropolis-Hastings updates, where parameter values are proposed and then accepted with a probability based on a conditional density ratio. Normalizing constants cancel out in this ratio and thus can be omitted without loss of generality. Consider the proportional full-conditionals for the model parameters (where the distribution

name represents the respective density function):

$$
\begin{aligned}
[\boldsymbol{\lambda}_1|\cdot] &\propto \prod_i \prod_t \text{Pois}(N_{i,t}|\lambda_{i,t}) \times \text{N}(log(\boldsymbol{\lambda}_1)|\boldsymbol{\mu}_\lambda, \boldsymbol{\Sigma}_\lambda^{1,1}), \\
[N_{i,t}|\cdot] &\propto \text{Beta-Binom}(n_{i,t}|N_{i,t}) \times \text{Pois}(N_{i,t}|\lambda_{i,t}), \text{ for } i = 1, \ldots, m, \text{and } t = 1, \ldots, T, \\
[\boldsymbol{\tau}|\cdot] &\propto \prod_i \prod_t \text{Pois}(N_{i,t}|\lambda_{i,t}) \times \text{N}(log(\boldsymbol{\tau})|\boldsymbol{\mu}_\tau, \boldsymbol{\Sigma}_\tau^{1,1}), \\
[K|\cdot] &\propto \prod_i \prod_t \text{Pois}(N_{i,t}|\lambda_{i,t}) \times \text{Gamma}(K|\alpha_K, \beta_K), \\
[r|\cdot] &\propto \prod_i \prod_t \text{Pois}(N_{i,t}|\lambda_{i,t}) \times \text{N}(r|\mu_r, \sigma_r^2),
\end{aligned}
$$

where, the product Poisson is included in the full conditionals for K, r, $\boldsymbol{\tau}$, and $\boldsymbol{\lambda}_1$ because $\boldsymbol{\lambda}_t$ is a direct function of those parameters (i.e., (4.18)), thus they appear in each $(i,t)^{th}$ density. Upon each iterative sampling of K, r, $\boldsymbol{\tau}$, and $\boldsymbol{\lambda}_1$, the process $(\boldsymbol{\lambda}_t)$ is updated using (4.18). The following steps can be taken to implement the necessary sampling for this model:

1. Define initial values for model parameters: $r^{(0)}$, $K^{(0)}$, $\boldsymbol{\tau}^{(0)}$, and $\boldsymbol{\lambda}_1^{(0)}$. Run forward process model (4.18) given initial parameter values to obtain intial values for the process $\boldsymbol{\lambda}_t^{(0)}$ for $t = 2, \ldots, T$. Sample the starting values $N_{i,t}^{(0)}$ from a Poisson distribution with intensity $\lambda_{i,t}^{(0)}$. Set $j = 0$.

2. Set $j = j + 1$ and sample $\boldsymbol{\lambda}_1^{(\star)}$ from a proposal distribution, say $[\boldsymbol{\lambda}_1^{(*)}|\boldsymbol{\lambda}_1^{(j-1)}]$. Run forward process model (4.18) given $r^{(0)}$, $K^{(0)}$, $\boldsymbol{\tau}^{(0)}$, and $\boldsymbol{\lambda}_1^{(\star)}$ to obtain proposal values for $\boldsymbol{\lambda}_t^{(\star)}$ for $t = 2, \ldots, T$. If the proposal distribution is symmetric with respect to $\boldsymbol{\lambda}_1^{(*)}$ and $\boldsymbol{\lambda}_1^{(j-1)}$ (e.g., $\text{N}(log(\boldsymbol{\lambda}_1^{(\star)})|\log(\boldsymbol{\lambda}_1^{(j-1)}), a_\lambda \mathbf{I})$, with tuning parameter a_λ), then let $\boldsymbol{\lambda}_1^{(j)} = \boldsymbol{\lambda}_1^{(*)}$ (and hence $\boldsymbol{\lambda}_t^{(j)} = \boldsymbol{\lambda}_t^{(*)}, \forall t$) if $p_\lambda^\star > u$; otherwise let $\boldsymbol{\lambda}_1^{(j)} = \boldsymbol{\lambda}_1^{(j-1)}$. Where, $u \sim \text{Uniform}(0,1)$ and

$$p^{\star}_{\lambda} = \frac{\prod_i \prod_t \text{Pois}(N^{(j-1)}_{i,t}|\lambda^{(\star)}_{i,t}) \times \text{N}(log(\boldsymbol{\lambda}^{(\star)}_1)|\boldsymbol{\mu}_{\lambda}, \boldsymbol{\Sigma}_{\lambda})}{\prod_i \prod_t \text{Pois}(N^{(j-1)}_{i,t}|\lambda^{(j-1)}_{i,t}) \times \text{N}(log(\boldsymbol{\lambda}^{(j-1)}_1)|\boldsymbol{\mu}_{\lambda}, \boldsymbol{\Sigma}_{\lambda})}.$$

3. For each i and t where data exists, sample $N^{(\star)}_{i,t}$ from proposal distribution $[N^{(\star)}_{i,t}|N^{(j-1)}_{i,t}]$. When the proposal is symmetric with respect to $N^{(\star)}_{i,t}$ and $N^{(j-1)}_{i,t}$ (e.g., a discrete uniform with support: $\{(N^{(j-1)}_{i,t} - a_N), \ldots, (N^{(j-1)}_{i,t} + a_N)\}$ with tuning parameter a_N), let $N^{(j)}_{i,t} = N^{(\star)}_{i,t}$ if $p^{\star}_N > u$; otherwise let $N^{(j)}_{i,t} = N^{(j-1)}_{i,t}$, where $u \sim \text{Uniform}(0,1)$ and

$$p^{\star}_N = \frac{\text{Beta-Binom}(n_{i,t}|N^{(\star)}_{i,t}) \times \text{Pois}(N^{(\star)}_{i,t}|\lambda^{(j)}_{i,t})}{\text{Beta-Binom}(n_{i,t}|N^{(j-1)}_{i,t}) \times \text{Pois}(N^{(j-1)}_{i,t}|\lambda^{(j)}_{i,t})}.$$

4. Sample $\boldsymbol{\tau}^{(\star)}$ from proposal distribution $[\boldsymbol{\tau}^{(\star)}|\boldsymbol{\tau}^{(j-1)}]$. Run forward process model (4.18) given $r^{(j-1)}$, $K^{(j-1)}$, $\boldsymbol{\tau}^{(\star)}$, and $\boldsymbol{\lambda}^{(j)}_1$ to obtain new proposal values for $\boldsymbol{\lambda}^{(\star)}_t$ for $t = 2, \ldots, T$. When the proposal is symmetric with respect to $\boldsymbol{\tau}^{(\star)}$ and $\boldsymbol{\tau}^{(j-1)}$ (e.g., $\text{N}(log(\boldsymbol{\tau}^{(\star)}_1)|\log(\boldsymbol{\tau}^{(j-1)}_1), a_{\tau}\mathbf{I})$, with tuning parameter a_{τ}), let $\boldsymbol{\tau}^{(j)} = \boldsymbol{\tau}^{(\star)}$ (and hence $\boldsymbol{\lambda}^{(j)}_t = \boldsymbol{\lambda}^{(*)}_t, \forall t$) if $p^{\star}_{\tau} > u$; otherwise let $\boldsymbol{\tau}^{(j)} = \boldsymbol{\tau}^{(j-1)}$, where $u \sim \text{Uniform}(0,1)$ and

$$p^{\star}_{\tau} = \frac{\prod_i \prod_t \text{Pois}(N^{(j)}_{i,t}|\lambda^{(\star)}_{i,t}) \times \text{N}(log(\boldsymbol{\tau}^{(\star)})|\boldsymbol{\mu}_{\tau}, \boldsymbol{\Sigma}_{\tau})}{\prod_i \prod_t \text{Pois}(N^{(j)}_{i,t}|\lambda^{(j)}_{i,t}) \times \text{N}(log(\boldsymbol{\tau}^{(j-1)})|\boldsymbol{\mu}_{\tau}, \boldsymbol{\Sigma}_{\tau})}.$$

5. Sample $K^{(\star)}$ from proposal distribution $[K^{(\star)}|K^{(j-1)}]$. Run forward process model (4.18) given $r^{(j-1)}$, $K^{(\star)}$, $\boldsymbol{\tau}^{(j)}$, and $\boldsymbol{\lambda}^{(j)}_1$ to obtain new proposal values for $\boldsymbol{\lambda}^{(\star)}_t$ for $t = 2, \ldots, T$. When the proposal is symmetric with respect to $K^{(\star)}$ and $K^{(j-1)}$ (e.g., $\text{N}(K^{(\star)}|K^{(j-1)}, a_K)$, with tuning parameter a_K), let $K^{(j)} = K^{(\star)}$ (and hence $\boldsymbol{\lambda}^{(j)}_t = \boldsymbol{\lambda}^{(*)}_t, \forall t$) if $p^{\star}_K > u$; otherwise let $K^{(j)} = K^{(j-1)}$, where $u \sim \text{Uniform}(0,1)$ and

$$p^{\star}_K = \frac{\prod_i \prod_t \text{Pois}(N^{(j)}_{i,t}|\lambda^{(\star)}_{i,t}) \times \text{Gamma}(K^{(\star)}|\alpha_K, \beta_K)}{\prod_i \prod_t \text{Pois}(N^{(j)}_{i,t}|\lambda^{(j)}_{i,t}) \times \text{Gamma}(K^{(j-1)}|\alpha_K, \beta_K)}.$$

6. Sample $r^{(\star)}$ from proposal distribution $[r^{(\star)}|r^{(j-1)}]$. Run forward process model (4.18) given $r^{(\star)}$, $K^{(j)}$, $\boldsymbol{\tau}^{(j)}$, and $\boldsymbol{\lambda}^{(j)}_1$ to obtain new proposal values for $\boldsymbol{\lambda}^{(\star)}_t$ for $t = 2, \ldots, T$.

When the proposal is symmetric with respect to $r^{(\star)}$ and $r^{(j-1)}$ (e.g., $\mathrm{N}(r^{(\star)}|r^{(j-1)}, a_r)$, with tuning parameter a_r), let $r^{(j)} = r^{(\star)}$ (and hence $\boldsymbol{\lambda}_t^{(j)} = \boldsymbol{\lambda}_t^{(*)}, \forall t$) if $p_r^\star > u$; otherwise let $r^{(j)} = r^{(j-1)}$, where $u \sim \mathrm{Uniform}(0,1)$ and

$$p_K^\star = \frac{\prod_i \prod_t \mathrm{Pois}(N_{i,t}^{(j)}|\lambda_{i,t}^{(\star)}) \times \mathrm{N}(r^{(\star)}|\sigma_r^2)}{\prod_i \prod_t \mathrm{Pois}(N_{i,t}^{(j)}|\lambda_{i,t}^{(j)}) \times \mathrm{N}(r^{(j-1)}|\sigma_r^2)}.$$

7. Calculate predictions of the state vector at future times ($\lambda_{i,T+\Delta}, \forall i$ and $\Delta \in \{1, 2, \ldots, T_2\}$) using the forward process model (4.18) given $r^{(j)}$, $K^{(j)}$, $\boldsymbol{\tau}^{(j)}$, and $\boldsymbol{\lambda}_1^{(j)}$. Posterior predictive samples for future population sizes (i.e., $N_{i,T+\Delta}^{(j)}$, $\forall i$ and $\Delta \in \{1, 2, \ldots, T_2\}$) can be obtained by then sampling $N_{i,T+\Delta}^{(j)} \sim \mathrm{Pois}(\lambda_{i,T+\Delta}^{(j)})$.

8. If predictions are desired for locations other than where data exist, the spatial structure on $\tilde{\boldsymbol{\tau}}$ and $\tilde{\boldsymbol{\lambda}}_1$ (i.e., from (4.21) and (4.22)) can be used to perform posterior predictive interpolations. That is, the posterior predictions for $\tilde{\boldsymbol{\tau}}^{(j)}$ can be calculated by: $\log(\tilde{\boldsymbol{\tau}}^{(j)}) = \tilde{\boldsymbol{\mu}}_\tau + \boldsymbol{\Sigma}_\tau^{2,1}[\boldsymbol{\Sigma}_\tau^{1,1}]^{-1}(\log(\boldsymbol{\tau}^{(j)}) - \boldsymbol{\mu}_\tau)$, which is a result of the multivariate normal specification. $\tilde{\boldsymbol{\lambda}}_1^{(j)}$ can be predicted in a similar fashion; then by using $r^{(j)}$, $K^{(j)}$, $\tilde{\boldsymbol{\tau}}^{(j)}$, and $\tilde{\boldsymbol{\lambda}}_1^{(j)}$ in the process model (4.18), we can calculate $\tilde{\boldsymbol{\lambda}}_t^{(j)}, \forall t$ (the posterior predicted process) and then sample $\tilde{N}_{i,t}^{(j)} \sim \mathrm{Pois}(\tilde{\lambda}_{i,t}^{(j)})$ (the posterior predicted population size) for any spatial location i and time t.

9. After convergence is attained (i.e., $j > b$, for some burn-in period b), sampled parameter values can be stored, resampled to remove auto-correlation, and used to calculate posterior summary statistics. The algorithm can be iterated by repeating steps 2 through 9 until a sufficiently large sample has been obtained from which to approximate the posterior distribution.

5 PRESENCE / ABSENCE OF INVASIVE SPECIES AND EPIDEMICS IN SPACE AND TIME

5.1 Introduction

Many types of ecological, environmental, and epidemiological data are collected over discrete spatial and temporal domains (Krebs 1978; Hanski 1999; Waller and Gotway 2004). Moreover, such data are often binary valued. For example, consider a natural process $\mathcal{Z}$ that evolves in some space ($\mathcal{S}$) over a set of time ($\mathcal{T}$) continuously. It is rarely feasible to observe such a process in a continuous fashion; in fact, it is impossible to do so if the process is being observed and (or) recorded digitally. Therefore data collection schemes are often designed for convenience, and the process $\mathcal{Z}$ is observed in discrete snapshots, say, $y_{i,t}$, where $t \in \mathbf{T} \subset \mathcal{T}$ and $i \in \mathbf{S} \subseteq \mathcal{S}$ for $\dim(\mathbf{S}) < \infty$, where $\mathbf{S}$ and $\mathbf{T}$ are finite sets of spatial locations (or areas) and times (or periods), respectively.

Models such as those proposed in Section 4.1 can be employed in cases where the data ($y_{i,t}$) are counts on domains with areal spatial support; whereas those proposed in 4.2 can be employed in cases where count data occur in continuous space. Both types of models previously discussed for characterizing spreading phenomena involve population growth and dispersal processes. It is often the case that data reflecting population growth are not available and only the occurrence of a phenomena is observed. In such cases where only binary data exist and without repeated measurements, population growth parameters are not identifiable (without strong priors). The focus in such situations shifts to the estimation of dispersal-based dynamics.

Conventional methods for modeling spatial processes on partitioned domains include

conditional and simultaneously specified spatial Gaussian models (also known as conditional autoregressive models and simultaneous autoregressive models) which have now been extended to the spatio-temporal setting (see Cressie 1993 and Banerjee et al. 2004 for a complete discussion). The "autologistic" model has served as the dominant spatial model for binary data (Besag 1974; Cressie 1993; Heikkinen and Hogmander 1994; Hogmander and Moller 1995; Hoeting et al. 2000), where the probability of presence is conditioned on its neighbors. More recently, methods have been developed that extend the auto-logistic model, as well as other models using a logit transform of presence/absence probability, to spatio-temporal settings (e.g., Zhu et al. 2005; Royle and Kery 2006; Royle 2006).

Many previously developed methods address the problem from an autoregressive modeling perspective (e.g., Pace et al. 2000), where a state of the system at the previous time is related to the state at the current time through (possibly time varying) spatial and temporal autoregressive coefficients. This is a very powerful approach when considering the process as a whole and in settings where a Gaussian autoregressive model can be employed. Despite its ubiquitous appearance in the literature, the logit transform is not the most intuitive or wieldy way to characterize changes in probability. In fact, the main reason for its use is the ability to use Gaussian model error (e.g., logistic regression). Addressing the problem more directly and modeling the untransformed probability components of presence and absence allows for the characterization of the dynamics of a propagating phenomenon over partitioned spatial and temporal domains.

The methods that follow introduce a class of models that are motivated by viewing the movement (i.e., dispersal) of a phenomenon from the perspective of the phenomenon itself,

rather than the system as a whole. For example, a species of organism, such as an exotic invasive species or a pathogenic species, in a new environment will often move from areas of lower quality to areas of higher quality (quality defined in terms of many possible factors, from environmental suitability to overpopulation to availability of hosts). In this way, the system as a whole (i.e., the "automaton") can be thought of as a process with numerous automatous components. Such "automata" can act in their own manner given the state of their surrounding environment.

The properties and behavior of automata have been studied in nearly every field and are so powerful, they have been hypothesized to form the underpinnings of life itself (Wolfram 1984). In essence, an automata is defined as an entity, whose movement (movement can taken to be as abstract as necessary in this context) is governed by a set of simple rules, and whose interaction with its environment can result in extremely complex "life-like" systematic behavior. In other words, the combined behavior of numerous individual automata results in very complicated dynamical system behavior incapable of being described in any other way (Wolfram 1983).

Automata are most often defined in a deterministic dynamical system framework where the state of the "neighborhood" of an entity (if an areal unit of space, called a "cell") determines the future state of the entity. Another type of automata can be formulated probabilistically, where movement of an entity into its neighborhood is defined by parametric probability distributions and the behavior of the system as a whole (i.e., the "automaton") given the probability rule is not unique, but can be expressed in terms of likelihood (Lee et al. 1990). A propagating automatous system defined in either manner is capable

of exhibiting spatially irregular wave-like behavior commonly found in natural phenomena (Hogeweg 1988). Using more traditional terminology found in the spatial statistics literature, "spatially irregular" refers to spatial structure that is either anisotropic (i.e., varying directionally) or non-stationary (i.e., varying locationally) or both (e.g., Cressie 1993).

This fundamental difference in the construction of models is known as the "top-down" versus "bottom-up" approach (Grimm et al. 2005). Traditionally, bottom-up approaches to studying ecosystem function and ecological processes have been used in simulation settings whereas top-down approaches have been taken using statistical methodology. Both have contributed much to ecology in theory and application, though they rarely intermingle.

5.1.1 Application: Rabies Epidemics in Raccoon Populations

An example of a natural spatio-temporal process where binary data is available on a partitioned spatial domain and exhibiting irregular spreading behavior is the ongoing rabies epidemic on the east coast of the United States of America that began in the mid-1970's (Nettles et al. 1979). Specifically, space-time data documenting the spread of rabies through the raccoon population in Connecticut is available as "time since arrival in township" format (e.g., Smith et al. 2002). That is, after appearing first in 1991 in the western townships of Connecticut (those bordering New York), rabies spread throughout the state over the following five year period. Although epidemics are not typically thought of as invasive species, they propagate through populations of organisms in space and time in a similar manner as non-epidemiological invasions propagate in space and time.

The dataset shown in Figure 30 illustrates the spatio-temporal behavior of rabies epidemic in Connecticut from 1991–1995, where several long-distance dispersal events occurred

in addition to anisotropic and non-stationary neighborhood-based dispersal patterns. Smith et al. (2002) employed a bottom-up simulation model in the spirit of an automatous system, though focusing on the temporal domain by characterization of rates of spread from neighboring townships, and found that such models can be useful in studying this particular natural phenomenon. Moreover, they allowed for long-distance dispersal, spatially varying rates of spread between neighboring townships, and found a correlation with environmental covariates such as proximity of a large river (i.e., the Connecticut River) and human population density.

The models developed in the next section approach the problem from the same perspective as Smith et al. (2002), but in a rigorous probability framework where inference on the dispersal parameters can be made from a statistical perspective given the available data.

5.2 Methods

5.2.1 Anisotropic Stationary Dispersal on Homogeneous Discrete Domains

One approach to modeling such a spatio-temporal process is to estimate the dynamics of spread in terms of probability of presence. That is, assume the phenomenon in question is present ($y_{i,t} = 1$) at the spatial location s_i and time t with probability $\theta_{i,t}$ and absent ($y_{i,t} = 0$) with probability $1 - \theta_{i,t}$. This is common form for the classes of models known as occurrence and occupancy models (e.g., MacKenzie et al. 2002, 2003; Royle and Dorazio 2006). A more convenient notation for this data model is: $y_{i,t}|\theta_{i,t} \sim \text{Bern}(\theta_{i,t})$. Rather than model the transformed probability (either as a logit or probit transform) at the next level in the hierarchy, consider the following specification for $\theta_{i,t}$:

$$\theta_{i,t} = y_{i,t-1}\phi + (1 - y_{i,t-1})(I_{\mathcal{N}_{i,t-1}})\bar{p}_{i,t} + (1 - I_{\mathcal{N}_{i,t-1}})\psi, \tag{5.1}$$

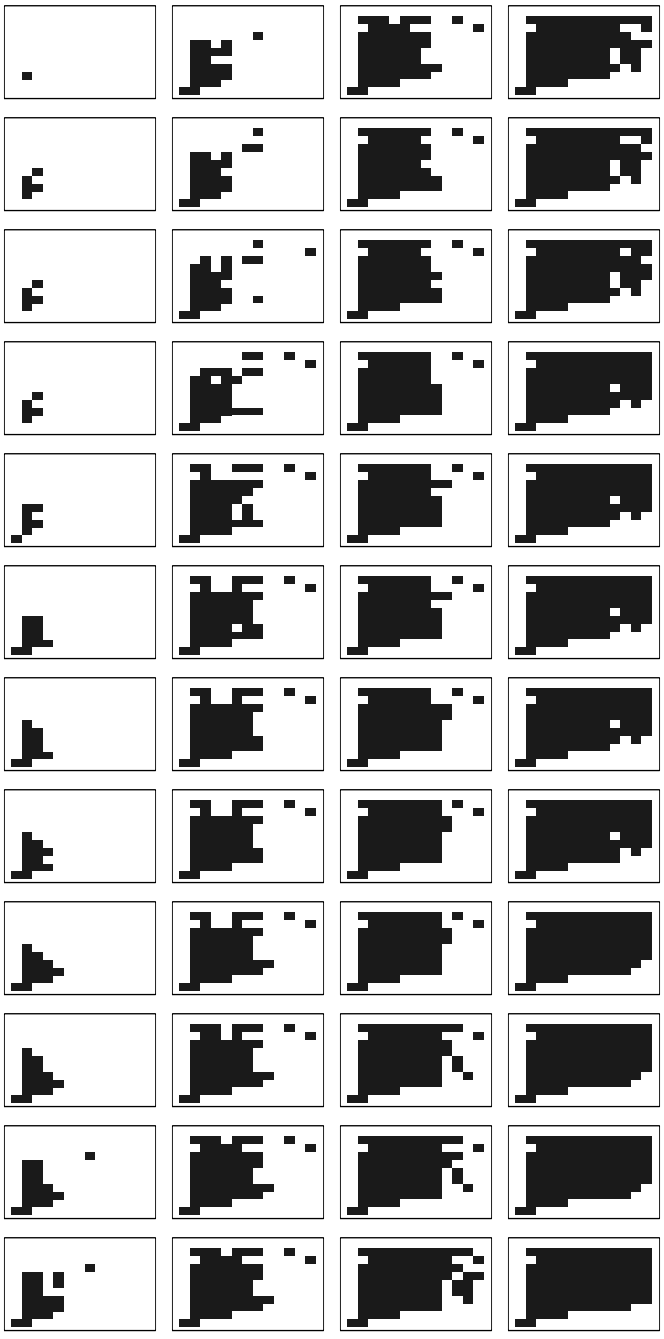


Figure 30: Presence / Absence of rabies in raccoon populations in Connecticut over 48 regularly spaced time periods beginning in 1991 (top left) and ending in 1995 (bottom right). Black cells denote presence while white cells denote absence of rabies.

where,

$$I_{\mathcal{N}_{i,t-1}} = \begin{cases} 1, \sum_{j \in \mathcal{N}_i} y_{j,t-1} > 0 \\ 0, \sum_{j \in \mathcal{N}_i} y_{j,t-1} = 0 \end{cases}$$

and, $\mathcal{N}_i$ is the spatial neighborhood of area i. In a regularly gridded 2-dimensional domain with a "Queen's neighborhood,"

$$\mathcal{N}_i = \begin{bmatrix} \mathrm{N}_{j=i-4} & \mathrm{N}_{j=i-1} & \mathrm{N}_{j=i+2} \\ \mathrm{N}_{j=i-3} & \mathrm{N}_{j=i} & \mathrm{N}_{j=i+3} \\ \mathrm{N}_{j=i-2} & \mathrm{N}_{j=i+1} & \mathrm{N}_{j=i+4} \end{bmatrix} = \begin{bmatrix} \mathrm{N}_1 & \mathrm{N}_4 & \mathrm{N}_7 \\ \mathrm{N}_2 & \mathrm{N}_5 & \mathrm{N}_8 \\ \mathrm{N}_3 & \mathrm{N}_6 & \mathrm{N}_9 \end{bmatrix}. \tag{5.2}$$

In (5.1) the probability ϕ corresponds to persistence of the phenomena; that is, once the phenomena is present in spatial area i, the probability it will be present at the next time step is ϕ. The probability ψ corresponds to out of neighborhood dispersal; that is, the phenomena will occur in an area outside of those occupied with probability ψ. This is a somewhat naive specification for what is referred to in the ecological literature as long-distance dispersal (e.g., Clark 2003). In some cases it would be more realistic to let ψ vary with distance from source or other environmental covariates. Without appropriate *a priori* knowledge of the mechanisms of long-distance dispersal and data to support this notion, a naive specification (such as in (5.1)) at least allows for its effect.

The probability $\bar{p}_{i,t}$ is the component of the process that handles the neighborhood-based dispersal in this model. Many possible specifications for $\bar{p}_{i,t}$ are possible depending on the spatial domain, level of realistic detail desired in the model, and estimability of $\bar{p}_{i,t}$. Assuming a regularly gridded spatial domain and Queen's neighborhood dispersal, let,

$$\bar{p}_{i,t} = 1 - \exp((\mathbf{y}_{\mathcal{N}_i,t-1})' \log(\mathbf{1} - \mathbf{p})), \tag{5.3}$$

where, $\mathbf{y}_{\mathcal{N}_i,t-1} = [y_{N_{i+4},t-1}, \ldots, y_{i,t-1}, \ldots, y_{N_{i-4},t-1}]'$ is a vector of the data (i.e., ones and zeros) corresponding to the neighborhood of the i^{th} spatial area. The probabilities in $\mathbf{p}$

($\mathbf{p} = [p_1, p_2, \ldots, p_{d_{\mathcal{N}}})]$'), where $d_{\mathcal{N}}$ is the dimensionality of the neighborhood, refer to the transition probability from spatial area i to its j^{th} neighbor. This specification allows for the probability of presence at time t, in an area that has been previously unoccupied, to be the union of transition probabilities from occupied neighbors. To show this result (i.e., (5.3)), denote the event that the phenomenon propagates from neighboring area N_j into area i at time t as event $E_{N_j,t}$, then the probability that area i becomes occupied at time t, given its neighborhood at time $t-1$ is:

$$\begin{aligned}
\mathrm{P}(\cup_j E_{N_j,t};\ \text{for all occupied neighbors } j) &= 1 - \mathrm{P}(\cap_j E^c_{N_j,t};\ \text{for all occupied neighbors } j) \\
&= 1 - \prod_j \mathrm{P}(E^c_{N_j,t}) \\
&= 1 - \prod_j (1 - \mathrm{P}(E_{N_j,t})) \\
&= 1 - \exp(\sum_j \log(1 - \mathrm{P}(E_{N_j,t}))),
\end{aligned}$$

where the last equality is shown only for computational efficiency and the assumption is made that the phenomenon moves from neighboring area j into area i independently of the phenomenon moving from neighbor k into area i at time t for all occupied areas j and k in the neighborhood of area i. This assumption of independent propagation, conditional on the parameters $\mathbf{p}$, is strong, and implies that phenomena entering a new area from two (or more) neighboring areas will not interact as they do so. This assumption does, however, allow the analytical calculation of $\bar{p}_{i,t}$ in the absence of significant information about the numerous possible intersections of events ($E_{N_j,t}$) which would be required to calculate the probability of the union.

Assuming the transition probabilities in $\mathbf{p}$ sum to one and the propagating process

can be modeled by stationary dynamics, a Dirichlet probability model for $\mathbf{p}$ is the natural choice (i.e., $\mathbf{p}|\mathbf{a} \sim \text{Dir}(\mathbf{a})$). The Dirichlet parameters ($\mathbf{a}$) can then be modeled as $\log(\mathbf{a}) \sim \text{N}(\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a)$.

It is important to point out several consequences concerning the effect of probabilities, $\mathbf{p}$, on the propagating phenomenon:

- This specification assumes that the neighborhood-based dispersal of the process is stationary. However, because the process also depends on non-neighborhood dispersal and persistence, it appears non-stationary; though, because of the exchangeable specification for ϕ and ψ in space and time, it could be argued that the overall process model is still stationary in the sense that given the neighborhood-based dispersal, the probabilities for non-neighborhood dispersal and persistence are uniformly distributed in space and time.

- By allowing $\mathbf{p}$ to be a vector, the neighborhood-based dispersal model is explicitly anisotropic. A simpler form of dispersal would have a phenomena propagating to all neighboring areas with an equal probability. Thus, in the case with a regularly gridded spatial domain and Queen's neighborhood, the multivariate nature of $\mathbf{p}$ allows for differing rates of dispersal in different directions (8 directions in this case).

- In the case with a regularly gridded spatial domain and Queen's neighborhood, allowing the dimension of $\mathbf{p}$ (i.e., $d_{\mathcal{N}}$) to be equal to eight assumes that the propagating phenomena will indeed propagate to a neighboring area. Letting $d_{\mathcal{N}} = 9$ in this setting allows the phenomena to remain in the current area with some probability.

However, since the probability of persistence is being modeled in θ, by ϕ, the addition of p_i seems redundant. It is important to note that in this setting, p_i will not affect the persistence, due to the $(y_{i,t-1}\phi)$ term in (5.1), but it will affect the overall rate of dispersal. A larger p_i will slow down the propagating phenomena without affecting the anisotropy of the process, hence providing the model with additional flexibility.

Priors for ϕ and ψ can be specified in terms of Beta distributions and result in conjugate full-conditional distributions. Thus, the priors $[\phi] = \text{Beta}(\alpha_\phi, \beta_\phi)$ and $[\psi] = \text{Beta}(\alpha_\psi, \beta_\psi)$ as well as the hyperpriors α_ϕ, β_ϕ, α_ψ, β_ψ, $\boldsymbol{\mu}_a$, and $\boldsymbol{\Sigma}_a = \sigma_a^2\mathbf{I}$ result in the joint posterior for this model:

$$[\mathbf{p}, \mathbf{a}, \phi, \psi | \{y_{i,t}, \text{ for } t = 1, \ldots, T \text{ and } i = 1, ..., m\}] \propto \prod_{t=1}^{T}\prod_{i=1}^{m}[y_{i,t}|\theta(\mathbf{p}, \phi, \psi)_{i,t}] \times \\ \times [\mathbf{p}|\mathbf{a}][\mathbf{a}][\phi][\psi] \tag{5.4}$$

It can be shown (see the Appendix to this chapter for details) that the full-conditional distributions for the model parameters, necessary for implementation in an MCMC setting, are:

$$\phi| \cdot \sim \text{Beta}(\sum_{t=2}^{T}\sum_{i \in \mathcal{A}_{\phi_{t-1}}} y_{i,t} + \alpha_\phi, \sum_{t=2}^{T}\sum_{i \in \mathcal{A}_{\phi_{t-1}}} (1 - y_{i,t}) + \beta_\phi),$$

$$\psi| \cdot \sim \text{Beta}(\sum_{t=2}^{T}\sum_{i \in \mathcal{A}_{\psi_{t-1}}} y_{i,t} + \alpha_\phi, \sum_{t=2}^{T}\sum_{i \in \mathcal{A}_{\psi_{t-1}}} (1 - y_{i,t}) + \beta_\psi),$$

$$\mathbf{p}| \cdot \sim [\mathbf{p}|\cdot] \propto \prod_{t=2}^{T}\prod_{i \in \mathcal{A}_{\bar{p}_i}} \text{Bern}(y_{i,t}|\bar{p}_{i,t}) \times \text{Dir}(\mathbf{p}|\mathbf{a}),$$

$$\mathbf{a}| \cdot \sim [\mathbf{a}|\cdot] \propto \text{Dir}(\mathbf{p}|\mathbf{a}) \times \text{N}(\log(\mathbf{a})|\boldsymbol{\mu}_a, \sigma_a^2\mathbf{I}),$$

where, the $\mathcal{A}$ are sets of indices defined by:

$$\mathcal{A}_{\phi_{t-1}} = \{j \mid \forall j \text{ such that } y_{j,t-1} = 1\},$$

$$\mathcal{A}_{\psi_{t-1}} = \{j \mid \forall j \text{ such that } (1 - I_{\mathcal{N}_{i,t-1}})) = 1\},$$

$$\mathcal{A}_{\bar{p}_i} = \{j \mid \forall j \text{ such that } (1 - y_{j,t-1})(I_{\mathcal{N}_{i,t-1}})) = 1\}.$$

In cases where $y_{i,t}$ is not observed for all i in the spatial domain (i.e., missing data $\tilde{\mathbf{y}}_t$), the posterior predictive distribution can be estimated for the missing data:

$$[\tilde{\mathbf{y}}_t|\mathbf{y}_t, \forall t] = \int \cdots \int [\tilde{\mathbf{y}}_t|\mathbf{p}, \mathbf{a}, \phi, \psi][\mathbf{p}, \mathbf{a}, \phi, \psi|\mathbf{y}_t, \forall t] d\mathbf{p} d\mathbf{a} d\phi d\psi.$$

5.2.2 Anisotropic Non-Stationary Dispersal on Heterogeneous Discrete Domains

The model in the previous section is very robust in that it allows for anisotropic dynamic behavior in the process which is capable of dispersing at varying rates. Like the matrix model introduced in section 4.2, it relies heavily on the dynamics of the process to model the data. From a scientific perspective, it is often of interest to make inference about the propagating nature of the process not only in the dynamics, but also in the covariance structure regarding the likely heterogeneous environment. In the situation where a phenomenon is propagating in space over time, it is reasonable to think that certain areas of the spatial domain are more suitable than others. The ecological literature refers to this notion as "habitat suitability" or "habitat preference," depending on the residence of the phenomenon in question relative to the surrounding environment (e.g., Hirzel et al. 2002). From a dynamic modeling perspective it seems natural to think that a phenomenon will most likely move from areas of undesirable habitat to areas of desirable habitat. Note that the term "habitat" is used very generally here, and in regards to the successful propagation

of the phenomenon under study. For example, if the phenomenon were a wildfire, it might "prefer" an exposed topography with steep slopes; whereas if it were a songbird, it might "prefer" forest edges.

Distinguishing areas of suitable or desirable habitat from those that are unsuitable is a long studied subject in ecology. One of the simplest and most common approaches being taken to map habitat suitability is to find the associations between organism abundance (or presence) and environmental covariates (e.g., Hooten et al. 2003). Due to limitations imposed though data collection schemes, such analyses are often focused in a static temporal domain over space. The study of phenomena that are actively propagating into new areas or reoccupying previous areas must involve some dynamic component; the challenge is in accounting for covariate effects on the dynamics.

In what follows, the terms "suitability" and "preference" are used interchangeably. Consider the suitability ($\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_m]'$) of a spatial domain, partitioned into m areas, in terms of phenomenon preference. Scientifically, a large portion of the variability in $\boldsymbol{\alpha}$ is often thought to be explained by a set of environmental covariates $\mathbf{X}$ that may be linked to $\boldsymbol{\alpha}$ by some function f, parameters $\boldsymbol{\beta}$, and some error structure $\boldsymbol{\varepsilon}$ accounting for any unknown covariates and (or) error related to the choice of f. Assuming a linear function f, and Gaussian error $\boldsymbol{\varepsilon}$, the standard linear model follows:

$$\boldsymbol{\alpha} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{5.5}$$

Making the assumption that the error contains residual spatial dependence, and letting $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\alpha)$, where $\boldsymbol{\Sigma}_\alpha = \sigma^2_\alpha \exp(-\frac{\mathbf{D}}{\theta_\alpha})$, and $\mathbf{D}$ is an $m \times m$ matrix of Euclidean distances between areas, it is then possible to estimate the residual spatial structure as well as the

covariate effects ($\boldsymbol{\beta}$) on habitat suitability. Note that if warranted, other suitable spatial covariance models could be used to define $\boldsymbol{\Sigma}_\alpha$.

Since the propagating phenomena are attracted to areas of higher suitability, an "attraction" model is useful to explain probability. One such model is based on partial differential equations and can be used in this context. Consider the spatial field, $\boldsymbol{\alpha}$, exhibiting the preferred habitat of an organism. Given that the phenomena is present at location i, the probability that it will propagate to location j depends, at least partially, on the quality of habitat relative to other surrounding habitats. The continuous version of this type of attraction can be represented as a partial differential equation,

$$\tilde{\mathbf{A}} = \nabla_{\mathcal{S}}(\boldsymbol{\alpha}) = \frac{\partial \boldsymbol{\alpha}}{\partial \mathcal{S}}, \tag{5.6}$$

where the partial derivative of the environmental preference field, $\boldsymbol{\alpha}$, is taken with regards to space ($\mathcal{S}$) and thus the gradient operator ($\nabla_{\mathcal{S}}$) will be infinite dimensional in the sense that there will be a different tangent for every directional derivative. In a discrete context however, the gradient in (5.6) can be approximated by a finite number of directional difference equations. In the specific case of the regular 2-D gridded spatial domain and Queen's neighborhood, $\tilde{\mathbf{A}}$ is a matrix and the directional derivative at location i can be approximated by a vector of differences $\tilde{\mathbf{a}}_i$, where $\tilde{\mathbf{A}} = [\tilde{\mathbf{a}}_1, \ldots, \tilde{\mathbf{a}}_m]'$. Each element of $\tilde{\mathbf{a}}_i$ is then a function of the spatial field of habitat suitability ($\boldsymbol{\alpha}$):

$$\tilde{\mathbf{a}}_i = \begin{cases} \tilde{a}_{i,1} = \frac{\alpha_{N_1} - \alpha_i}{d_{N_1,i}} \\ \tilde{a}_{i,2} = \frac{\alpha_{N_2} - \alpha_i}{d_{N_2,i}} \\ \vdots \\ \tilde{a}_{i,9} = \frac{\alpha_{N_9} - \alpha_i}{d_{N_9,i}} \end{cases}$$

where, N_j refers to the j^{th} neighbor of i in the Queen's neighborhood ($\mathcal{N}_i$ in (5.2)) and

$d_{N_j,i}$ is the Euclidean distance between area N_j and i.

The utility of $\tilde{\mathbf{A}}$ is as a likelihood of movement from one area to another for a propagating phenomena. A transformation of $\tilde{\mathbf{A}}$ allows for its inclusion as parameters in a model similar to that of Section 5.1. Since the transition probabilities $\mathbf{p}$ are the focal point of the neighborhood-based dynamics in this class of models, allowing them to depend hierarchically upon the approximate gradients $\tilde{\mathbf{A}}$ is critical to the utilization of covariate information and non-stationary anisotropic dynamics in the process. Therefore, let $\mathbf{A}$ be a fuction of the directional gradient fields in $\tilde{\mathbf{A}}$ (e.g., $\mathbf{A} = g(\tilde{\mathbf{A}})$) such that they meet the requirements for hyperparameters of a Dirichlet distribution (i.e., positive, finite support). The choice of function g is subjective and may be somewhat arbitrary. Consider, for example, the probit or standard normal CDF function (the exponential is another option) which maps a set of real numbers to the set of positive real numbers: $\mathbf{a}_i = g(\tilde{\mathbf{a}}_i) = \Phi(\tilde{\mathbf{a}}_i)$. A more flexible form of such a transformation is one where the parameters $\mathbf{A}$ are only proportional to the transforming function of $\tilde{\mathbf{A}}$:

$$\mathbf{a}_i = g(\tilde{\mathbf{a}}_i) = c\ \Phi(\tilde{\mathbf{a}}_i) \tag{5.7}$$

where c is a multiplicative scalar modeled at a lower level (or user defined, serving as a tuning parameter). Letting the vector of transition probabilities vary for each spatial location i, and depend on the parameters $\mathbf{a}_i$ through the Dirichlet distribution, gives the probability model that characterizes the neighborhood-based dynamics of the propagating phenomena:

$$\mathbf{p}_i|\mathbf{a}_i \sim \text{Dir}(\mathbf{a}_i),\ \text{for } i = 1, \ldots, m. \tag{5.8}$$

Recall, (from (5.1) that the actual probability of presence in area i when area i has

been previously unoccupied and at least one neighbor at the previous time was occupied, is dictated by the parameter $\bar{p}_{i,t}$. The calculation of $\bar{p}_{i,t}$ in the stationary model was fairly straightforward because $\mathbf{p}$ did not vary with space. Utilizing the same data model from the previous section (i.e., $y_{i,t}|\theta_{i,t} \sim \text{Bern}(\theta_{i,t})$), and specification for the probability of presence ($\theta_{i,t}$) as in (5.1), a different function mapping $\mathbf{p}_{\mathcal{N}_i}$ to $\bar{p}_{i,t}$ must be used. For this non-stationary model, let $\bar{p}_{i,t}$ be defined as:

$$\bar{p}_{i,t} = 1 - \exp((\mathbf{y}_{\mathcal{N}_i,t-1})' \log(\mathbf{1} - \mathbf{p}_{\mathcal{N}_i})), \tag{5.9}$$

where, $\mathbf{y}_{\mathcal{N}_i,t-1}$ is defined as in the previous section, $\mathbf{p}_{\mathcal{N}_i} = [p_{N_9,1}, p_{N_8,2}, \ldots, p_{N_1,9}]'$, and N_j refers to the j^{th} neighboring area or location of i. In this way, the probability of presence in area i at the current time is the probability of the unions of transitions into area i from occupied neighboring areas at the previous time.

To complete the hierarchical model, let $\boldsymbol{\beta} \sim \text{N}(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta)$, where $\boldsymbol{\mu}_\beta$ is specified as a hyperprior and $\boldsymbol{\Sigma}_\beta$ is modeled at a lower level (i.e., $\boldsymbol{\Sigma}_\beta^{-1} \sim \text{Wish}((\nu \mathbf{S}_\beta)^{-1}, \nu)$) to handle any multicollinearity between covariates. Let the persistence and non-neighborhood based probability of presence be defined as before, $\phi \sim \text{Beta}(\alpha_\phi, \beta_\phi)$ and $\psi \sim \text{Beta}(\alpha_\psi, \beta_\psi)$. The spatial covariance parameters on the habitat preference spatial field, σ^2_α and θ_α, can be modeled in the conventional manner where $\sigma^2_\alpha \sim \text{InvGamma}(q, r)$ and $\theta_\alpha \sim \text{Gamma}(\alpha_\theta, \beta_\theta)$. The spatial range parameter θ_α may be difficult to identify if there is minimal residual spatial structure in the $\boldsymbol{\alpha}$ process. Additionally, due to the formulation of the model, a large θ_α (implying greater spatial structure) may be counterproductive considering that the model is trying to find sharp differences between neighboring areas (i.e., due to the directional gradient fields in $\tilde{\mathbf{A}}$).

The posterior distribution of interest can now be written as:

$$[\mathbf{p}_1,\ldots,\mathbf{p}_m,\mathbf{a}_1,\ldots,\mathbf{a}_m,\phi,\psi,\boldsymbol{\beta},\boldsymbol{\Sigma}_\beta,\sigma^2_\alpha,\theta_\alpha|y_{i,t},\forall i,t] \propto$$
$$\propto \prod_{t=1}^{T}\prod_{i=1}^{m}[y_{i,t}|\theta(\mathbf{p},\phi,\psi)_{i,t}]\prod_{i=1}^{m}[\mathbf{p}_i|\mathbf{a}(\boldsymbol{\alpha})_i][\boldsymbol{\alpha}|\boldsymbol{\beta},\sigma^2_\alpha,\theta_\alpha][\boldsymbol{\beta}|\boldsymbol{\Sigma}_\beta][\boldsymbol{\Sigma}_\beta][\phi][\psi][\sigma^2_\alpha][\theta_\alpha]\ .$$

Similar to the previous section (and detailed in the Appendix), the full-conditional distributions for the model parameters, necessary for implementation of this non-stationary anisotropic model, are:

$$\phi|\cdot \sim \text{Beta}(\sum_{t=2}^{T}\sum_{i\in\mathcal{A}_{\phi_{t-1}}} y_{i,t}+\alpha_\phi, \sum_{t=2}^{T}\sum_{i\in\mathcal{A}_{\phi_{t-1}}}(1-y_{i,t})+\beta_\phi)\ ,$$
$$\psi|\cdot \sim \text{Beta}(\sum_{t=2}^{T}\sum_{i\in\mathcal{A}_{\psi_{t-1}}} y_{i,t}+\alpha_\phi, \sum_{t=2}^{T}\sum_{i\in\mathcal{A}_{\psi_{t-1}}}(1-y_{i,t})+\beta_\psi)\ ,$$
$$\mathbf{P}|\cdot \sim [\mathbf{P}|\cdot] \propto \prod_{t=2}^{T}\prod_{i\in\mathcal{A}_{\bar{p}_i}}\text{Bern}(y_{i,t}|\bar{p}_{i,t}) \times \prod_{i=1}^{m}\text{Dir}(\mathbf{p}_i|\mathbf{a}_i)\ ,$$
$$\boldsymbol{\alpha}|\cdot \sim [\boldsymbol{\alpha}|\cdot] \propto \prod_{i=1}^{m}\text{Dir}(\mathbf{p}_i|\mathbf{a}(\boldsymbol{\alpha})_i) \times \text{N}(\boldsymbol{\alpha}|\boldsymbol{\beta},\sigma^2_\alpha,\theta_\alpha)\ ,$$
$$\boldsymbol{\beta}|\cdot \sim \text{N}((\mathbf{X}'\boldsymbol{\Sigma}_\alpha^{-1}\mathbf{X}+\boldsymbol{\Sigma}_\beta^{-1})^{-1}(\mathbf{X}'\boldsymbol{\Sigma}_\alpha^{-1}\boldsymbol{\alpha}+\boldsymbol{\Sigma}_\beta^{-1}\boldsymbol{\mu}_\beta),(\mathbf{X}'\boldsymbol{\Sigma}_\alpha^{2}\mathbf{X}+\boldsymbol{\Sigma}_\beta^{-1})^{-1})\ ,$$

$$\boldsymbol{\Sigma}_\beta|\cdot \sim \text{Wish}((\boldsymbol{\beta}-\boldsymbol{\mu}_\beta)(\boldsymbol{\beta}-\boldsymbol{\mu}_\beta)'+\nu\mathbf{S})^{-1}, 1+\nu)\ ,$$
$$\sigma^2_\alpha|\cdot \sim \text{InvGamma}\left(\left(\frac{(\boldsymbol{\alpha}-\mathbf{X}\boldsymbol{\beta})'[\exp(-\mathbf{D}/\theta_\alpha)]^{-1}(\boldsymbol{\alpha}-\mathbf{X}\boldsymbol{\beta})}{2}+\frac{1}{r}\right)^{-1}, \frac{m}{2}+q\right)\ ,$$
$$\theta_\alpha|\cdot \sim [\theta_\alpha|\cdot] \propto \text{N}(\boldsymbol{\alpha}|\boldsymbol{\beta},\sigma^2_\alpha,\theta_\alpha) \times \text{Gamma}(\theta_\alpha|\alpha_\theta,\beta_\theta)\ ,$$

where, $\mathbf{P}$ is the matrix containing the transition probability vectors $\mathbf{p}_i$ and the index sets $\mathcal{A}$ are defined as before:

$$\mathcal{A}_{\phi_{t-1}} = \{j \mid \forall j \text{ such that } y_{j,t-1}=1\},$$
$$\mathcal{A}_{\psi_{t-1}} = \{j \mid \forall j \text{ such that } (1-I_{\mathcal{N}_{i,t-1}}))=1\},$$
$$\mathcal{A}_{\bar{p}_i} = \{j \mid \forall j \text{ such that } (1-y_{j,t-1})(I_{\mathcal{N}_{i,t-1}}))=1\}.$$

5.3 Results

This section contains a description of the results of fitting the two previously introduced models (i.e., (5.2.1) and (5.2.2)) to the Connecticut rabies dataset (Figure 30); a full discussion of the results follows in the next section (5.4).

The rabies dataset introduced in Section (5.1.1) was converted to a gridded areal unit format from a township areal unit format (as in Smith et al. 2002) for use with the proposed models. The models proposed were presented with sufficient generality to model the data in their original format however, and the implementation of such generalizations is the focus of ongoing research. Moreover, as previously discussed, the rabies data were originally available in "time since first appearance" in each township (i.e., spatial area), and were converted into presence / absence data for use with this model. This form of data is more common in epidemiological and studies where more traditional biostatistical models are utilized (e.g., longitudinal data analysis and survival analysis). The models developed here, though not specifically developed to address such data, could be modified to allow time-delay data to enter through a different data model (e.g., an exponential or Wiebull distribution). Such generalizations to other data structures are also the focus of ongoing research.

Another important note, regarding the persistence of the process, is that epidemic behavior is often characterized by a travelling epidemic "wave," where after infection of the local population, local extinction occurs and the propagating disease, having no more hosts in that area, subsides. To effectively capture the behavior in such cases, the persistence term (ϕ) becomes critical. In the case of the rabies data, without knowledge of subsidence, the data were transformed such that each infected area was assumed to persist. In this

case, with rabies in a wild animal population where the disease can be spread readily and is largely unmanaged, the assumption is not unreasonable; however, such an assumption would not be appropriate in all epidemiological situations. In either event, the persistence term (ϕ) should accommodate differing levels of dispersal.

The models from Sections (5.2.1) and (5.2.2) were implemented by sampling from the posterior distributions via the full conditional distributions. Non-conjugate full-conditional distributions were sampled via Metropolis-Hastings updates (further details in the Appendix to this chapter). Convergence in the MCMC samples was reached relatively quickly (i.e., $<$ 1000 iterations) and the remaining 19,000 realizations were resampled to remove autocorrelation induced by the algorithm and then used to calculate posterior statistics. Tables 5 and 6 provide the hyperprior specifications used in fitting these models respectively:

Table 5: Hyperparameters used in stationary model from Section 5.2.1

Hyperparameter	**Value**
$\mu_{a,i}$	1
$\mathbf{\Sigma}_a$	$1 \times \mathbf{I}$
α_ϕ	10
β_ϕ	1
α_ψ	1
β_ψ	10

Table 6: Hyperparameters used in non-stationary model from Section 5.2.2

Hyperparameter	**Value**
$\mu_{\beta,i}$	0
$\mathbf{S}_\beta$	$0.5 \times \mathbf{I}$
ν	50
r	0.8
q	2.25
μ_θ	1
σ^2_θ	1
α_ϕ	10
β_ϕ	1
α_ψ	1
β_ψ	10

5.3.1 Stationary Model

This section contains the results of fitting the anisotropic stationary model introduced in Section 5.2.1 to the Connecticut rabies data (Figure 30). The estimated posterior distributions for the neighborhood-based transition probabilities $\mathbf{p}$ are presented as boxplots in Figure 31. Recall that $\mathbf{p}$ is comprised of 9 transition probabilities, one representing neighborhood based dispersal in each of the first 8 cardinal directions (i.e., Northwest, West, Southwest, North, South, Northeast, East, Southeast), also p_5 corresponds to the probability of no transition. Also shown in Figure 31 is the expected posterior 95% credible interval of $\mathbf{p}$ given only the information from the prior distribution of $\mathbf{a}$ (also the expected posterior mean). Due to the hierarchical structure of the model, the neighborhood-based transition probabilities $\mathbf{p}$ have no prior distribution, however, if the data were uninformative, the

posterior distribution for **p** would resemble the credible interval shown in dashed lines in Figure 31. The estimated posterior distributions for persistence (ϕ) and non-neighborhood

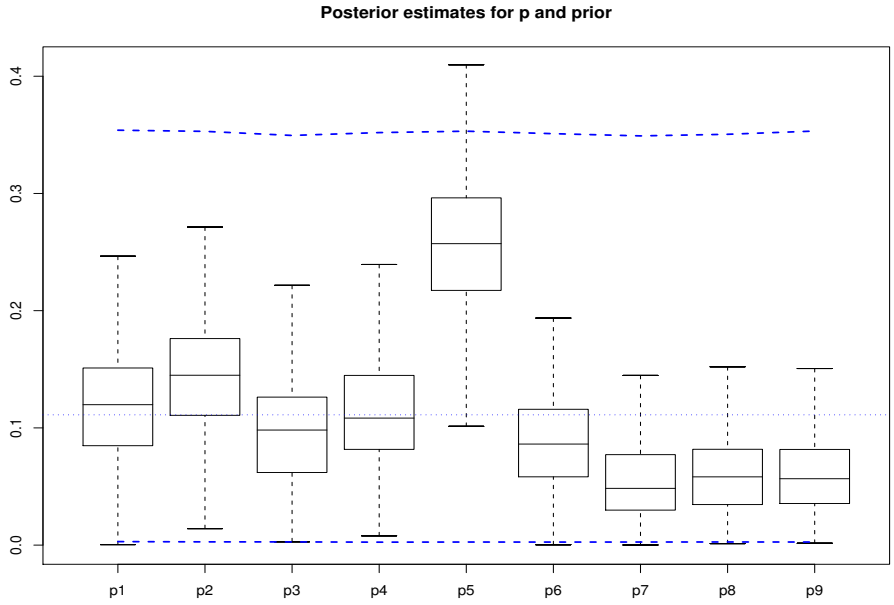
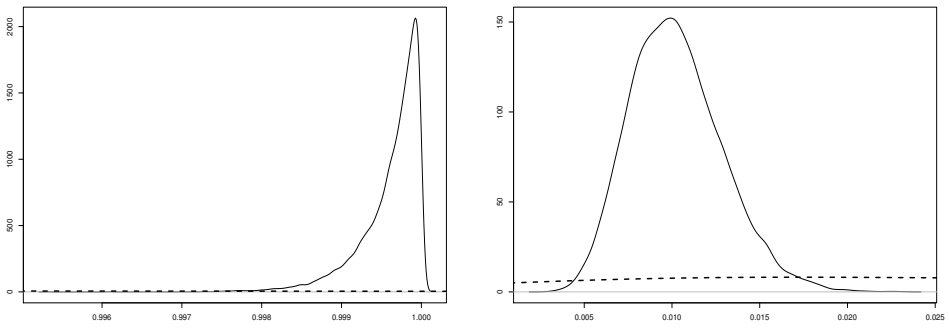


Figure 31: Posterior estimates for the parameters in **p** corresponding to: NW, W, SW, N, None, S, NE, E, SE in terms of directional probability of neighborhood-based propagation. Dashed lines correspond to the 95% credible interval based on only the prior distribution of **a** in the case where no information is gained from the data about **p**. Dotted line corresponds to the mean directional propagation probability under the prior.

dispersal (ψ) are given in Figures 32a and 32b. The prior distributions for ϕ and ψ are provided for comparison as dashed lines in each figure. The posterior mean of the probability of presence ($\theta_{i,t}$) is shown for each time step t and spatial area i in Figure 33. The sequence of maps is shown in the same order as the data in Figure 30, where the first time step appears in the upper left corner of the figure and downward, then to the next column from top to bottom, and so on until reaching the final time step in the lower right corner of the figure. The posterior standard deviation of the probability of presence ($\theta_{i,t}$) is shown



(a) $\phi|\mathbf{y}$ (b) $\psi|\mathbf{y}$

Figure 32: Posterior distributions of ϕ and ψ from the stationary model, respectively. Dashed lines represent the prior distribution.

in the same manner as the posterior mean in Figure 34. Deviance Information Criterion (DIC) was used as a form of model comparison for the stationary versus non-stationary models implemented here is applicable especially in this case because the data models used are specified the same. DIC, in this case, provides a measure of how well the top level parameters in the hierarchy fit the data. Here, the top level parameters (i.e., $\mathbf{p}$, ϕ, and ψ) are the most important because they provide an insight into the dynamics of dispersal in spreading phenomena. DIC also penalizes (likewise for AIC and BIC) for the number of effective parameters in the model so that overfit models can be recognized and inference can be made cautiously. The effective number of parameters for the stationary model was estimated to be $p_{D_1} = 36.5$, while the DIC value was 723.9 (after correcting for the effective number of parameters).

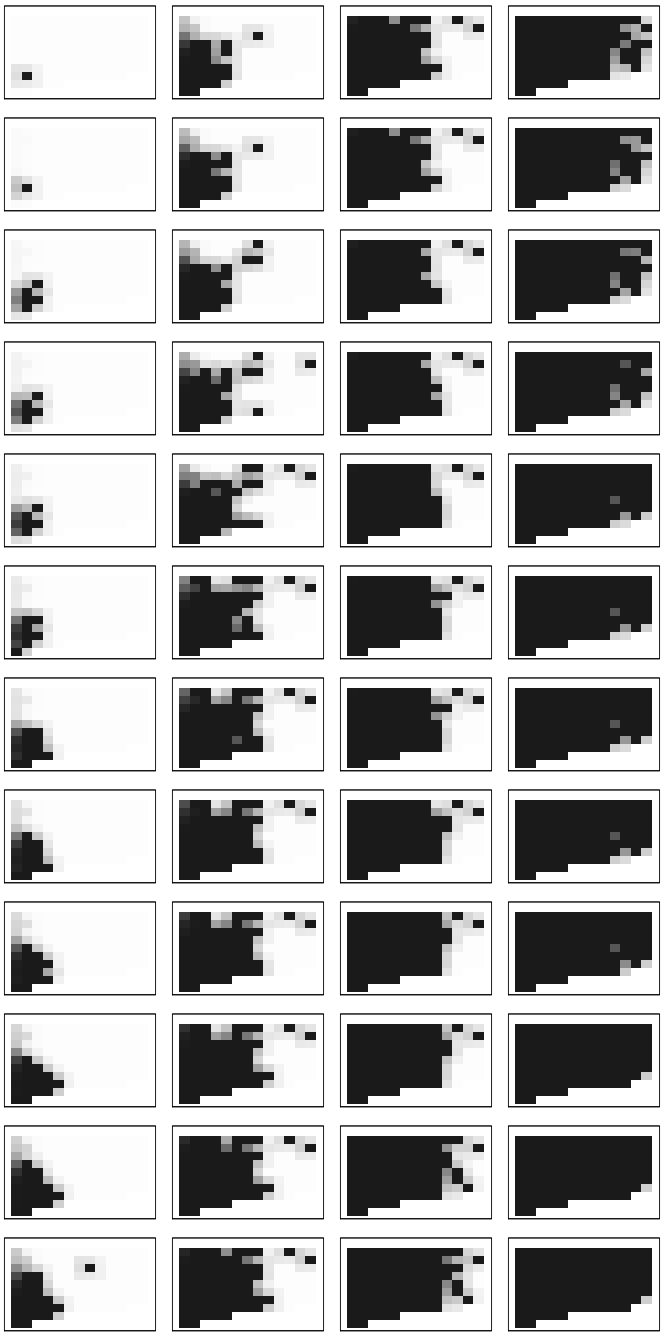


Figure 33: Posterior mean of $[\theta_{i,t}|\mathbf{y}]$ from stationary model in the same sequence as the data in Figure 30. Values of intensity range between zero (white) and one (black).

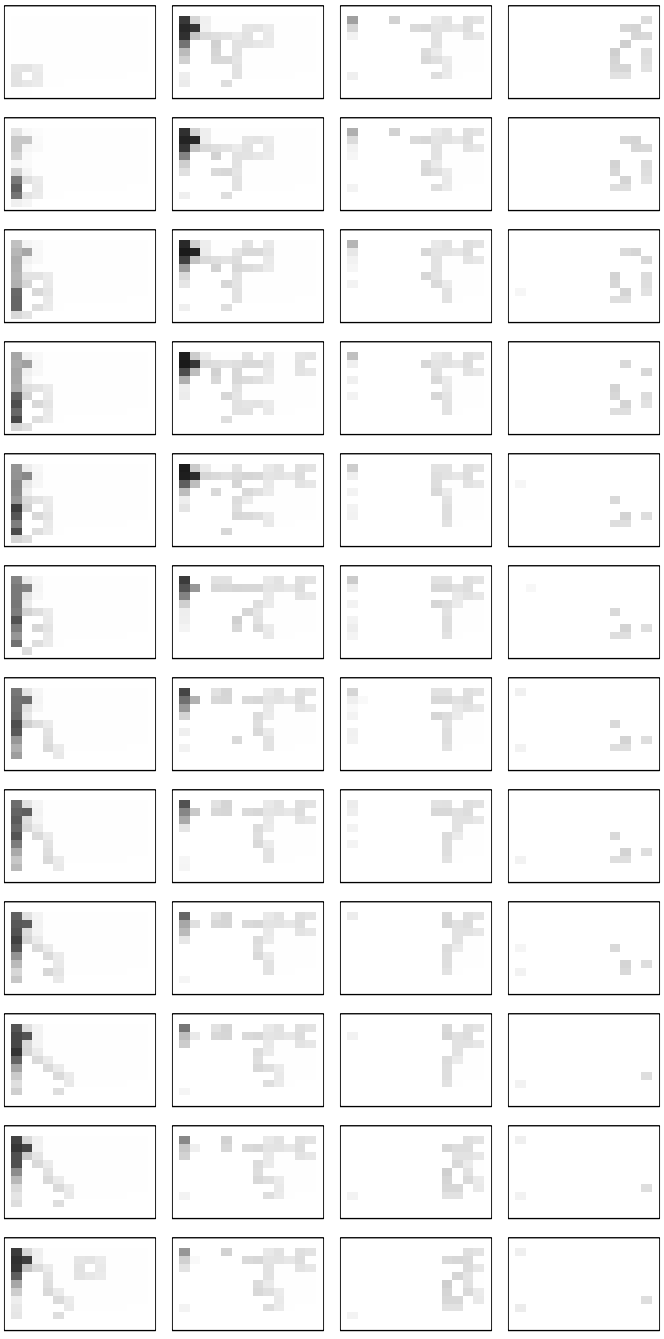


Figure 34: Posterior standard deviation of $[\theta_{i,t}|\mathbf{y}]$ from stationary model in the same sequence as the data in Figure 30. Values of intensity range between zero (white) and one (black).

5.3.2 Non-Stationary Model

Figure 35 shows the covariates ($\mathbf{X}$) used in fitting the anisotropic non-stationary model from Section 5.2.2 to the Connecticut rabies dataset given in Figure 30. All covariates used in this analysis were binary with the black regions in Figure 35 corresponding to the occurrence of the environmental characteristic. Specifically, the X_1 and X_2 covariates correspond the west and east sides of the Connecticut River, respectively. The X_3 covariate corresponds to land area in Connecticut (and some in New York on the western edge of the maps) which is not adjacent to the Connecticut River. The X_4 covariate corresponds to the Connecticut coastline where the state meets the Long Island Sound. A covariate corresponding to human population density was omitted from this analysis because of lack of availability. It is important to note, however, that major population centers in Connecticut coincide with the Connecticut River (X_1 and X_2) and coastline (X_4) covariates and is therefore implicitly accounted for in $\mathbf{X}$. An explicit continuous human population covariate may prove to be helpful in future analyses.

The posterior mean and standard deviation of the probability of presence parameters ($\theta_{i,t}$) are provided as a sequence of maps in Figures 36 and 37. The posterior distributions of the neighborhood-based dispersal for the anisotropic non-stationary model could be displayed as in Figure 31, however they would have to shown for each spatial location i since $\mathbf{p}_i$ now varies in space. Instead, consider the posterior mean of each spatial field of the neighborhood-based dispersal probabilities in Figure 38. The j^{th} map in Figure 38 corresponds to the intensity in probability of the phenomenon (rabies in this case) dispersing in that direction, where the center plot (i.e., $j = 5$) is the probability of non-dispersal. For

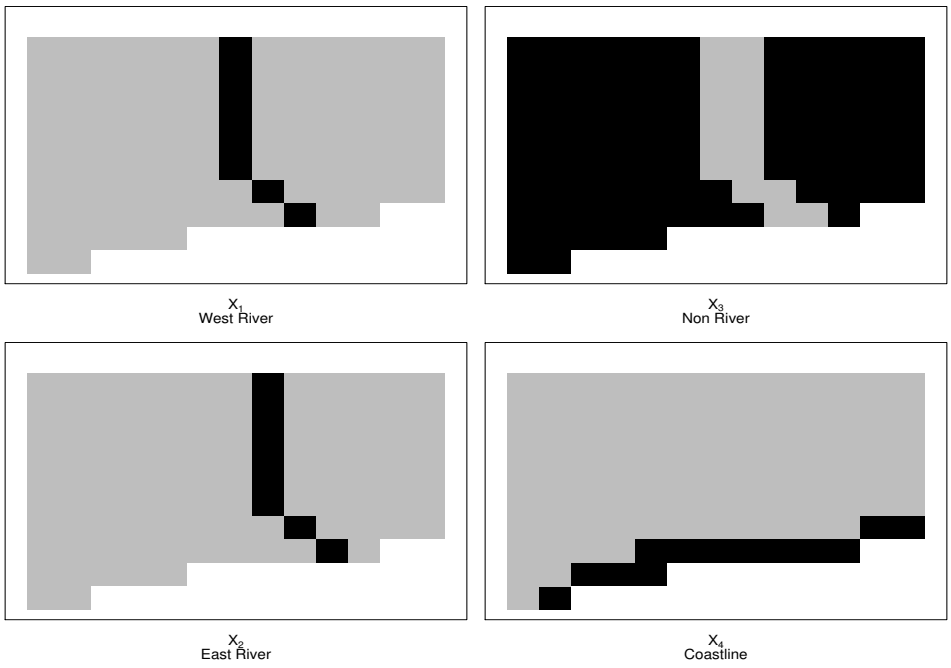


Figure 35: The covariates used in the non-stationary model: $\mathbf{x}_1$ and $\mathbf{x}_2$ represent the west and east sides of the Connecticut River, respectively, while $\mathbf{x}_3$ represents all land area not adjacent to the Connecticut River, and $\mathbf{x}_3$ represents the Connecticut coastline.

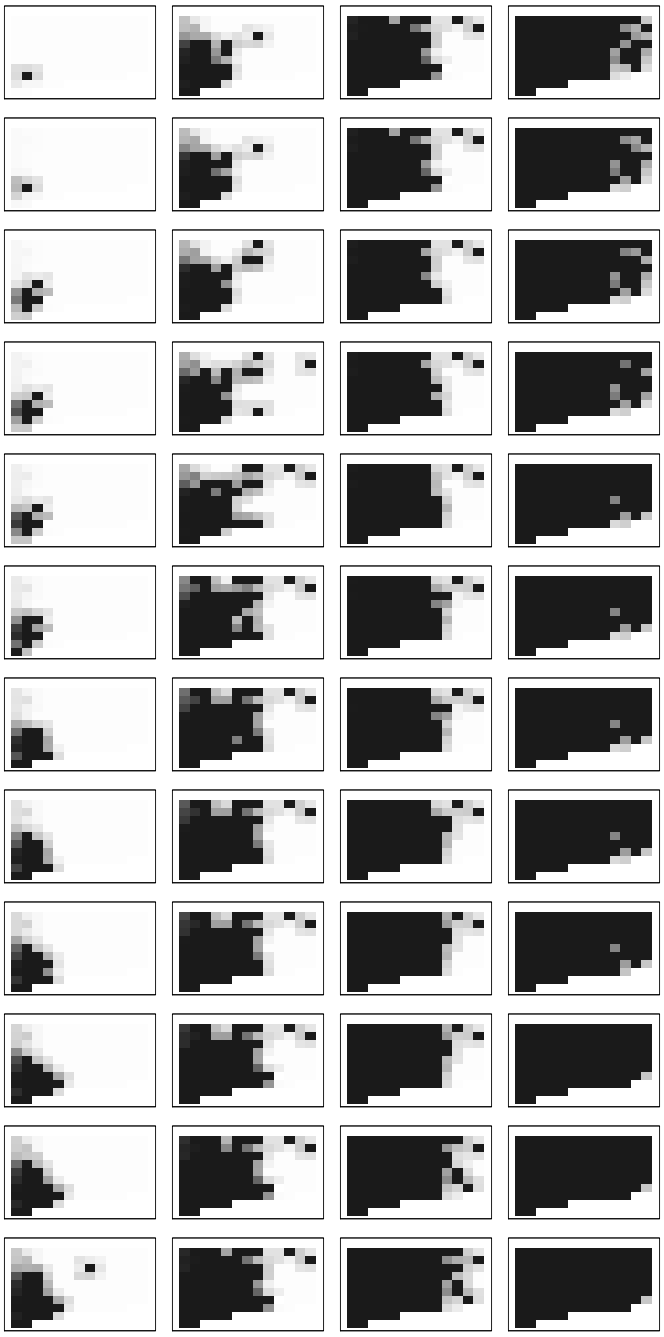


Figure 36: Posterior mean of $[\theta_{i,t}|\mathbf{y}]$ from non-stationary model in the same sequence as the data in Figure 30. Scale ranges between zero (white) and one (black).

Figure 37: Posterior standard deviation of $[\theta_{i,t}|\mathbf{y}]$ from non-stationary model in the same sequence as the data in Figure 30. Scale ranges between zero (white) and one (black).

example, the map in Figure 38 labeled '3' corresponds to the probability of neighborhood-based dispersal in the Southwest direction given that the phenomenon is currently in the grid cell of interest. The posterior standard deviations of the neighborhood-based prob-

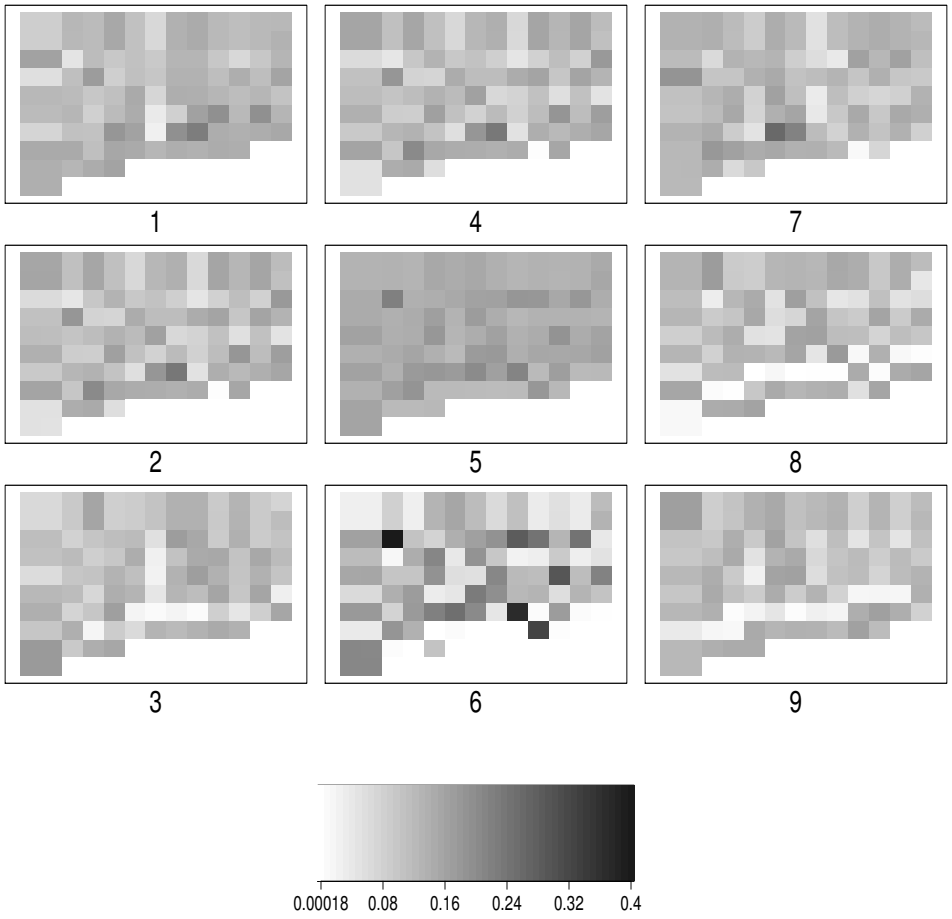


Figure 38: Posterior mean of $[\mathbf{P}|\mathbf{y}]$ from non-stationary model with respect to the Queen's neighborhood arrangement. The j^{th} map refers to the posterior mean spatial field of the neighborhood-based transition probabilities in the j^{th} direction.

abilities of dispersal ($\mathbf{p}_i$) can be shown in a similar manner (Figure 39), and provide an idea of the uncertainty in neighborhood-based dispersal in each direction for a given location. The map of habitat preference ($\boldsymbol{\alpha}$) given in Figure 40a, in this case corresponds to areas (shown as darker grid cells) where the rabies epidemic spreads most quickly to-

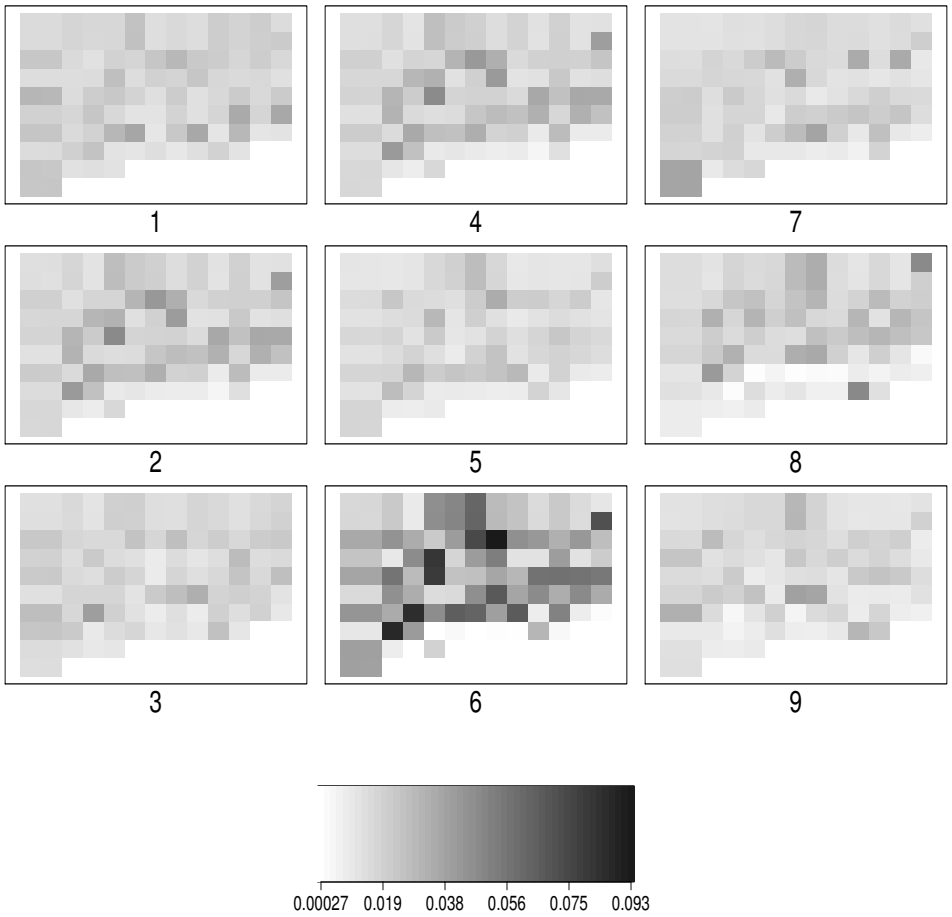
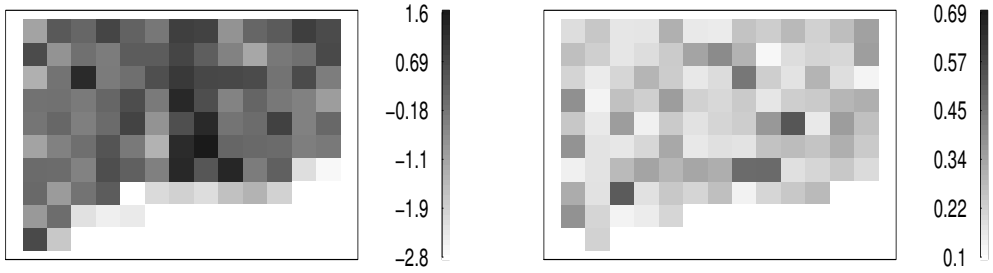


Figure 39: Posterior standard deviation of $[\mathbf{P}|\mathbf{y}]$ from non-stationary model with respect to the Queen's neighborhood arrangement. The j^{th} map refers to the posterior standard deviation spatial field of the neighborhood-based transition probabilities in the j^{th} direction.

wards. The posterior standard deviation corresponding to habitat preference is provided in Figure 40b and illustrates where areas of high uncertainty regarding habitat preference are over the spatial domain. The posterior distributions for the covariate effect parameters



(a) $\mathrm{E}(\boldsymbol{\alpha}|\mathbf{y})$ (b) $\sqrt{V(\boldsymbol{\alpha}|\mathbf{y})}$

Figure 40: Posterior mean and standard deviation of habitat preference ($\boldsymbol{\alpha}$), respectively. Large values in the $\mathrm{E}(\boldsymbol{\alpha}|\mathbf{y})$ correspond to areas of higher suitability.

($\boldsymbol{\beta}$) are provided in Figure 41 and, when premultiplied by the covariates (X), correspond to the mean of habitat preference ($\boldsymbol{\alpha}$). The posterior distributions for persistence (ϕ) and non-neighborhood dispersal (ψ) for the anisotropic non-stationary model are provided in Figures 42a and 42b. Additionally, their corresponding posterior distributions are shown as dashed lines for comparison.

The posterior mean of the scale parameter (σ^2_α) corresponding to the spatial covariance of $\boldsymbol{\alpha}$ was found to be 0.489 with a posterior standard deviation of 0.068, as compared to the prior mean and standard deviation of 1 and 4 respectively. The posterior distribution of the spatial range parameter (θ_α) in the covariance of $\boldsymbol{\alpha}$ was found to be very close to zero,

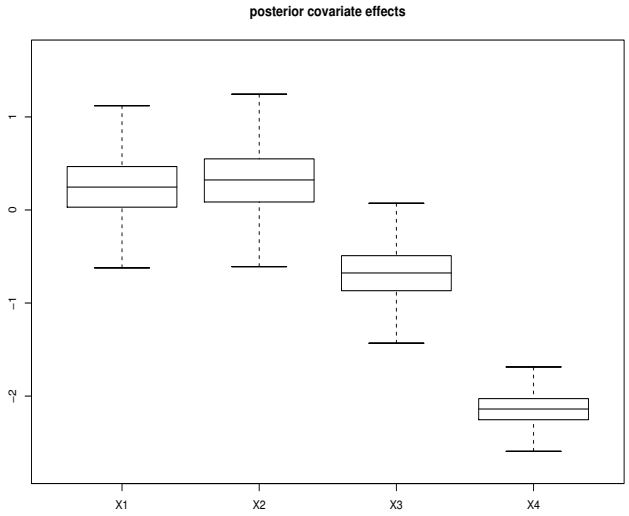
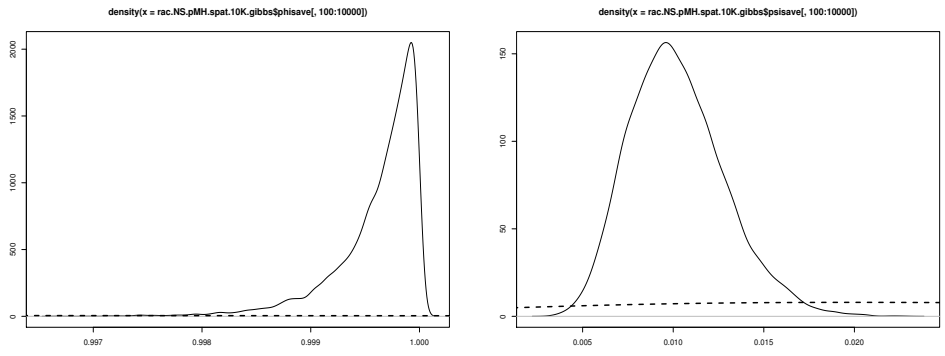


Figure 41: Posterior distributions of the covariate effects ($\boldsymbol{\beta}$), corresponding to the covariates in Figure 35.



(a) $\phi|\mathbf{y}$ (b) $\psi|\mathbf{y}$

Figure 42: Posterior distributions of ϕ and ψ from the non-stationary model, respectively. Dashed lines represent the prior distribution.

suggesting that there is very little, if any, residual spatial structure in the habitat preference covariance and was hence removed from the analysis. Note that the full-conditional distributions for all other parameters retain the same form, only with θ_α set to be zero.

Deviance Information Criterion was also calculated for this anisotropic non-stationary model and had a value of 702.6, while the effective number of parameters was estimated as $p_{D_2} = 1.92$.

5.4 Discussion

The results of fitting the stationary and non-stationary models to the Connecticut rabies dataset indicate that both methods have strengths and weaknesses. Beginning with the results from the anisotropic stationary model: Figure 31 provides a description of the neighborhood-based transition probabilities in each direction. More specifically, Figure 31 shows that in addition to all of the parameter estimates being much more precise than they would have been had the data been uninformative, they also suggest that neighborhood based dispersal is overall more likely in the North, Northwest, and West directions than in other directions due to the larger posterior values for p_1, p_2, and p_4 and smaller p_7, p_8, and p_9 values. Another notable feature of the posterior distributions for $\mathbf{p}$ is that p_5 is significantly greater than the other transition probabilities. Recall that p_5 is the parameter that controls the rate of spread. The fact that p_5 is greater suggests that rabies is propagating across Connecticut at a slower rate than it is capable of, given the time step. In sum, the neighborhood dynamics, when estimated with a stationary model, suggest that the spread of rabies in Connecticut during 1991–1995 was likely an anisotropic spatio-temporal process given the degree of separation in the posterior distributions of $\mathbf{p}$.

The posterior estimates of the persistence and non-neighborhood dispersal parameters ϕ and ψ (Figure 32) suggest that the process is extremely persistent due to the fact that ϕ is very close to one, while non-neighborhood dispersal (or long-distance dispersal) does indeed occur and has a mean near 0.01. Recall that in the transformation of the dataset for use with these models, the process was forced to persist, thus the posterior of ϕ is only indicating a known characteristic of the process. Had there not been persistence in the dataset, the ϕ parameter would have indicated this with a posterior significantly less than one. The posterior mean for ψ indicates that dispersal events out of the neighborhood of areas with rabies present occurs only 1% of the time. It may be that if ψ were allowed to vary with distance from active areas, it may be possible to make more specific inference regarding the likely distance of non-neighborhood dispersal. For example, perhaps dispersal events at 3–4 cells away from active cells are much more probable than other distances. This type of extension to the model is the focus of ongoing research.

Close inspection of the sequence of maps (Figure 33) showing the posterior mean of the probability of dispersal ($\theta_{i,t}$) in space and time illustrates the effect of the neighborhood-based dispersal parameters ($\mathbf{p}$) from Figure 31 on the overall propagation of the process. Specifically, the first map in the sequence (top left, Figure 33) shows that in the neighborhood of the active cell, the neighbors in the North, Northwest, and West directions are slightly darker than the neighbors in the other directions. This directly corresponds to the posterior mean values of p_1, p_2, and p_4 being greater than the remaining parameter values in $\mathbf{p}$. Notice that rabies propagates first up the left hand side of the map (i.e., western edge of Connecticut) and then eastward into the rest of the state. Again, this is the effect

suggested by the posterior distribution of $\mathbf{p}$.

The posterior standard deviation of the probability of presence (Figure 34) provides information about the uncertainty regarding the propagation of rabies throughout Connecticut. Notice first that in the first column of maps in Figure 34 there are large values of standard deviation along the western edge of Connecticut. This is due to the lack of data from New York (Connecticut's western neighbor). The uncertainty is higher in New York because the parameters are being estimated in those locations without the aid of data. This high uncertainty in areas of limited data is a common artifact in spatial and spatio-temporal statistical models. Another feature evident in the sequence of maps in Figure 34 is the traveling wave of high uncertainty moving from the Southwest corner of Connecticut to the Northeast corner and then finally to the Southeast region. Comparing this figure with with Figure 33 indicates that the propagating uncertainty corresponds to the wave front of the propagating rabies epidemic and suggests that there is less certainty regarding those areas in flux than in the persisting areas and yet to be infected areas.

The results pertaining to the anisotropic non-stationary model provide some additional insight into the dynamics of the rabies epidemic in Connecticut. Notice first, in Figure 36, that the posterior mean propagation of probability of presence ($\theta_{i,t}$) looks similar to that of the stationary model (Figure 33). However, when looking at the posterior standard deviation for probability of presence ($\theta_{i,t}$), the uncertainty in areas without data remains but the uncertainty corresponding to the wave front of the epidemic is less than that resulting from the stationary model (intensity scale is the same between figures; zero is shown in white and one in black). This suggests that even though the pattern of uncertainty is similar,

the non-stationary model is more precisely estimating the neighborhood-based component of the probability of presence ($\theta_{i,t}$).

In fact, the maps corresponding to the spatial fields of the directional neighborhood-based dispersal parameters ($\mathbf{p}_j$) in Figure 38 indicate that there are indeed differences in directional dispersal depending on location. For example, in map 1 (Northwest direction) and 3 (Southwest direction) of Figure 38 there is a distinct column of low dispersal corresponding to the west side of the Connecticut River covariate (X_1 in Figure 35). From Figure 40a, it is evident that the Connecticut river is an area of high habitat preference; this is also substantiated by the posterior distribution of $\boldsymbol{\beta}$ in Figure 41. Thus, as the rabies epidemic spreads from the western region of Connecticut it exhibits a different pattern of neighborhood-based dispersal than in the other other areas. Specifically, it shows a very low probability of dispersal in the Northwest and Southwest directions as it approaches the western edge of the Connecticut River, and a higher probability of dispersal in the North, South, Northeast, and Southeast directions. This characteristic in dispersal is also evident in Figures 30 and 36, where the second and third columns of maps in the sequence show the presence of rabies slowing in the eastward direction and instead spreading out to the North and South as the epidemic approaches the Connecticut River. Also evident from Figure 38 is that in areas just North of the Connecticut coastline, rabies is dispersing very slowly in the Southwest, Southeast, and East directions. One interpretation of this result is that the rabies epidemic in areas near the coast is propagating more inward (toward land) rather than outward (toward the Long Island Sound). This seems like an obvious characteristic of a land based phenomenon, yet such differences in neighborhood-based dispersal proba-

bilities over a heterogeneous environment provides the added flexibility needed for a better model fit.

The posterior probability of persistence (Figure 42) is again very close to one, suggesting that, given the available data (which were constructed to be persistent), the rabies epidemic in Connecticut is very persistent. The ramifications of this inference were previously discussed. The posterior mean probability of non-neighborhood dispersal is again estimated to be near 0.01, suggesting a 1% chance of long-distance rabies dispersal for non-neighboring areas.

One way to compare, *a posteriori*, the stationary model with the non-stationary model is to assess the difference in effective number of parameters as well as DIC for the two models. The effective number of parameters provides an indication of the degree to which parameters in the model are informed by the data (as opposed to informed by the prior). DIC provides a measure of the likelihood given the posterior estimates and corrects for the number of effective parameters. Therefore, given models with equal numbers of parameters, a higher p_D value and a lower DIC value is desirable. In models with differing numbers of parameters (as with the two models considered here), the effective number of parameters are difficult to compare; however, the DIC value can still serve as a basis for model comparison. Based on the results of the two models fitted here, the non-stationary model has a lower DIC and thus suggests a better fitting model. On the other hand, parsimony is an important factor in model comparison so there is certainly still value in the stationary model as it is more computationally efficient and retains more degrees of freedom. The non-stationary model, like the stationary model has a dynamical component but also has an explicit connection

to scientifically meaningful environmental covariates and is thus capable of fitting the data better while performing inference on the connection between different types of dispersal and the heterogeneous environment over which the epidemic is dispersing.

5.5 Conclusion

In this chapter, two methods for modeling a spreading phenomena over a partitioned landscape, given only binary data, were developed and applied to the rabies epidemic in raccoon populations in Connecticut during 1991–1995. Both models take a probabilistic cellular automata approach, while the first model was specified so that the focus was on the estimation of the dynamics governing the process in terms of stationary neighborhood-based dispersal. The second model utilized environmental covariates to help determine how patterns of dispersal might differ given heterogeneous environmental conditions.

Although model 1 assumed stationary dynamics, it was able to accommodate anisotropic dispersal behavior based on transition probabilities into neighboring areas. Model 2 allowed for anisotropic and non-stationary dynamics while accounting for associations between dispersal and environmental factors. Both models allowed for the possibility of long-distance dispersal and varying levels of persistence, although in this case the persistence parameter was not informative.

The models presented could be extended to address other issues as well. For example, an assumption one must make when applying these models is that of perfect detection. That is, there may be some probability associated with the actual detection of rabies in a given area, suggesting that perhaps rabies is present in an area earlier than it was recorded. One approach to addressing such a situation is to allow for a detection probability parameter in

the data model and then a latent process model would be employed to model the "true" presence or absence of rabies in an area. Royle and Kery (2006) and Royle (2006) have taken this approach in a logit-transformed framework and are able to make use of repeated measurements to estimate such parameters. In the situation considered here, much like the situation in Chapter 4 with the Eurasian Collared-Dove, repeated measurements are not available and without another means to estimate detectability, the employment of such methods are futile. It could be argued, however, that by modeling the probability of presence in the manner adopted by this chapter, the posterior distribution of $\theta_{i,t}$ (Figures 33 and 36) absorbs the detection probability in $\theta_{i,t}$ at locations where rabies has yet to be observed.

Overall, the methods presented here place the simulation models presented by Smith et al. (2002) into a rigorous statistical framework where the parameters governing dispersal can be estimated and inference can be made about the effects of environmental covariates on dispersal. Models such as these could be used to identify important factors in the spread of epidemics (e.g., river corridors, shorelines, and population centers) and ultimately utilized in policy making and management decisions.

5.6 Appendix: Full-Conditionals

In the anisotropic stationary model proposed in Section 5.2.1 the full-conditional distributions were reported without proof. Consider now the derivation of $[\phi|\cdot]$:

$$
\begin{aligned}
[\phi|\cdot] &\propto \prod_{t=2}^{T} \prod_{i \in \mathcal{A}_{\phi_{t-1}}} \text{Bern}(y_{i,t}|\theta(\phi)_{i,t}) \times \text{Beta}(\alpha_\phi, \beta_\phi), \\
&\propto \left[\prod_{t=2}^{T} \prod_{i \in \mathcal{A}_{\phi_{t-1}}} \theta(\phi)_{i,t}^{y_{i,t}} (1-\theta(\phi)_{i,t})^{1-y_{i,t}} \right] \phi^{\alpha_\phi - 1}(1-\phi)^{\beta_\phi - 1}, \\
&\propto \left[\prod_{t=2}^{T} \prod_{i \in \mathcal{A}_{\phi_{t-1}}} \phi^{y_{i,t}} (1-\phi)^{1-y_{i,t}} \right] \phi^{\alpha_\phi - 1}(1-\phi)^{\beta_\phi - 1}, \\
&\propto \phi^{\sum_{t=2}^{T} \sum_{i \in \mathcal{A}_{\phi_{t-1}}} y_{i,t}} (1-\phi)^{\sum_{t=2}^{T} \sum_{i \in \mathcal{A}_{\phi_{t-1}}} (1-y_{i,t})} \phi^{\alpha_\phi - 1}(1-\phi)^{\beta_\phi - 1}, \\
&\propto \phi^{\sum_{t=2}^{T} \sum_{i \in \mathcal{A}_{\phi_{t-1}}} y_{i,t} + \alpha_\phi - 1} (1-\phi)^{\sum_{t=2}^{T} \sum_{i \in \mathcal{A}_{\phi_{t-1}}} (1-y_{i,t}) + \beta_\phi - 1}, \\
&\Rightarrow \phi|\cdot \sim \text{Beta}(\sum_{t=2}^{T} \sum_{i \in \mathcal{A}_{\phi_{t-1}}} y_{i,t} + \alpha_\phi, \sum_{t=2}^{T} \sum_{i \in \mathcal{A}_{\phi_{t-1}}} (1-y_{i,t}) + \beta_\phi),
\end{aligned}
$$

where, the index set $\mathcal{A}_{\phi_{t-1}}$ is defined as it was in Section 5.2.1. The full-conditional for ψ can be derived in the same manner where ϕ in the above equations is replaced with ψ and the index set $\mathcal{A}_{\psi_{t-1}}$ is used instead of $\mathcal{A}_{\phi_{t-1}}$.

The parameters $\mathbf{p}$ and $\mathbf{a}$ are sampled from the full-conditionals via Metropolis-Hastings:

$$
\begin{aligned}
\mathbf{p}|\cdot \sim [\mathbf{p}|\cdot] &\propto \prod_{t=2}^{T} \prod_{i \in \mathcal{A}_{\bar{p}_t}} \text{Bern}(y_{i,t}|\bar{p}_{i,t}) \times \text{Dir}(\mathbf{p}|\mathbf{a}), \\
\mathbf{a}|\cdot \sim [\mathbf{a}|\cdot] &\propto \text{Dir}(\mathbf{p}|\mathbf{a}) \times \text{N}(\log(\mathbf{a})|\boldsymbol{\mu}_a, \sigma_a^2 \mathbf{I}),
\end{aligned}
$$

where the proposal $\mathbf{p}^*$ is sampled from a Dirichlet distribution given the current Gibbs sample $\mathbf{a}^{(k)}$. The proposed parameter vector is then accepted with the probability:

$$
\frac{\prod_{t=2}^{T} \prod_{i \in \mathcal{A}_{\bar{p}_i}} \text{Bern}(y_{i,t}|\bar{p}^*_{i,t})}{\prod_{t=2}^{T} \prod_{i \in \mathcal{A}_{\bar{p}_i}} \text{Bern}(y_{i,t}|\bar{p}^{(k-1)}_{i,t})},
$$

where the index set $\mathcal{A}_{\bar{p}_t}$ is defined as it was in Section 5.2.1. The proposal $\mathbf{a}^*$ is sampled via a Gaussian random walk (i.e., $\mathbf{a}^* \sim \mathrm{N}(\mathbf{a}^{(k-1)}, \sigma^2_{\text{tune}}\mathbf{I})$). Since the proposal distribution is symmetric with respect to $\mathbf{a}^{(k-1)}$ and $\mathbf{a}^*$, the proposed vector can be accepted with the probability:

$$\frac{\mathrm{Dir}(\mathbf{p}^{(k)}|\mathbf{a}^*) \times \mathrm{N}(\log(\mathbf{a}^*)|\boldsymbol{\mu}_a, \sigma^2_a\mathbf{I})}{\mathrm{Dir}(\mathbf{p}^{(k)}|\mathbf{a}^{(k-1)}) \times \mathrm{N}(\log(\mathbf{a}^{(k-1)})|\boldsymbol{\mu}_a, \sigma^2_a\mathbf{I})}.$$

In the Gibbs sampler for the anisotropic non-stationary model proposed in Section 5.2.2, the full-conditional distributions for ϕ and ψ can be derived in the same manner as above. Then, a proposal for all $\mathbf{p}_i$ simultaneously is obtained by sampling $\mathbf{p}^*_i \sim \mathrm{Dir}(\mathbf{a}^{(k)}_i)$ for all $i = 1, \ldots, m$, thus yielding $\mathbf{P}^* = [\mathbf{p}^*_1, \ldots, \mathbf{p}^*_m]$, and accepted with probability:

$$\frac{\prod_{t=2}^{T}\prod_{i\in\mathcal{A}_{\bar{p}_i}} \mathrm{Bern}(y_{i,t}|\bar{p}^*_{i,t})}{\prod_{t=2}^{T}\prod_{i\in\mathcal{A}_{\bar{p}_i}} \mathrm{Bern}(y_{i,t}|\bar{p}^{(k-1)}_{i,t})}.$$

Proposals for $\boldsymbol{\alpha}$ can be taken from a random walk, $\boldsymbol{\alpha}^* \sim \mathrm{N}(\boldsymbol{\alpha}^{(k-1)}, \sigma^2_{\text{tune}}\mathbf{I})$ and accepted with probability:

$$\frac{\prod_{i=1}^{m} \mathrm{Dir}(\mathbf{p}^{(k)}_i|\mathbf{a}(\boldsymbol{\alpha})^*_i) \times \mathrm{N}(\boldsymbol{\alpha}^*|\boldsymbol{\beta}^{(k)}, (\sigma^2_\alpha)^{(k)}, \theta^{(k)}_\alpha)}{\prod_{i=1}^{m} \mathrm{Dir}(\mathbf{p}^{(k)}_i|\mathbf{a}(\boldsymbol{\alpha})^{(k-1)}_i) \times \mathrm{N}(\boldsymbol{\alpha}^{(k-1)}|\boldsymbol{\beta}^{(k)}, (\sigma^2_\alpha)^{(k)}, \theta^{(k)}_\alpha)}.$$

The parameter θ_α can also be sampled via Metropolis-Hastings with the random walk proposal, $\theta^*_\alpha \sim \mathrm{N}(\theta^{(k-1)}_\alpha, \sigma^2_{\text{tune}})$ and accepted with probability:

$$\frac{\mathrm{N}(\boldsymbol{\alpha}^{(k)}|\boldsymbol{\beta}^{(k)}, (\sigma^2_\alpha)^{(k)}, \theta^*_\alpha) \times \mathrm{Gamma}(\theta^*_\alpha|\alpha_\theta, \beta_\theta)}{\mathrm{N}(\boldsymbol{\alpha}^{(k)}|\boldsymbol{\beta}^{(k)}, (\sigma^2_\alpha)^{(k)}, \theta^{(k-1)}_\alpha) \times \mathrm{Gamma}(\theta^{(k-1)}_\alpha|\alpha_\theta, \beta_\theta)}.$$

Full-conditional distributions for the remaining parameters (i.e., $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_\beta$, and σ^2_α) are conjugate and can be trivially derived in the fashion described in Gelman et al. 2004.

6 GENERAL CONCLUSIONS

After introduction and motivation in Chapter 1, several hierarhical statistical models were developed and utilized in Chapters 2 through 5 for characterizing non-linear ecological processes that vary in time and often space. Though applied to a sweeping variety of ecological phenomena (from forest growth dynamics and invasive species to endangered species and epidemics), the models developed here all share several common themes.

Each model presented is hierarchical, containing a scientifically meaningful data model (i.e., likelihood), a process model containing a non-linear dynamical component, and scientifically based *a priori* information related to the probabilistic behavior of model parameters. Additionally, in all cases, the models are heavily parameterized. Though in most cases, effort was made to reduce the dimensionality of parameter spaces, effective estimation in models with such parameterizations illustrates a primary advantage of adopting a Bayesian hierarchical framework. Conventional (i.e., frequentist) estimation in similar settings would be impossible due to over-parameterization and non-linearity; additionally, there would be no mechanism to account for prior understanding of the phenomena. Concerning the specific models described in previous chapters, implementation of the dendrochonology model in Chapter 2 may be possible using Kalman filtering and E-M algorithmic methods, but frequentist implementation of the models in Chapters 3, 4, and 5 would likely be hopeless or *ad hoc* at best.

Another key advantage of hierarchical Bayesian methods is the ability to account for uncertainty at multiple levels. These models allow uncertainty to propagate throughout the model space and ultimately into the process or parameters of interest. Inference on

such model components can then be achieved responsibly. Illustration of this can be seen in Section 4.2, where the true population sizes of the invading Eurasian Collared-Dove are estimated and predicted when only a single, imperfectly detected, space-time observation is available.

Clearly there are scientific situations where conventional statistical methods are most appropriate (e.g., carefully designed experiments). For many cases, however, where only sparse observational data are available and significant *a priori* scientific knowledge exists, hierarchical methods are often the only option. Despite the recent adoption of Bayesian statistics in ecology (as well as nearly all applied sciences), much of the field still associates statistical analysis with objectivity. Early scientists were slow to acknowledge the need for statistics, and now with frequentist ideals firmly in place, it seems that ecology is slow to accept the validity of appropriately utilizing subjective probability. It is likely the case, however, that the slow adoption has less to do with ingrained ideals and more to do with vast expanse of quantitative knowledge necessary to implement the more complicated hierarchical models. Although, the same could be said for pure likelihood-based approaches and thus perhaps a more subtle issue should be addressed.

Significant advances in statistical methodology are made regularly and with the field of statistics growing each year, theoretical contributions are allowing for more complex natural systems to be studied. In fact, as new methods of analysis emerge, a separation between statistics and science is inevitable. Thus the challenge in modern statistics is in the development of methodology that exhibits a relatively intuitive appeal and demonstrated utility. Likewise, the challenge in modern ecology is in making a commitment to the adoption of a

more formal quantitative curriculum. Statistics and science are naturally linked and that is why a collaborative effort is critical to insure the success of both disciplines.

LITERATURE CITED

Allen, D. Constructing artificial red-cockaded woodpecker cavities. Forest Service Technical Report SE-73, USDA, 1991.

Arim, M., S. Abades, P. Neill, M. Lima, and P. Marquet. 2006. Spread dynamics of invasive species. *Proceedings of the National Academy of Sciences, USA*, 103:374–378.

Ash, R. 1975. *Topics in Stochastic Processes*. Academic Press.

Banerjee, S., B. Carlin, and A. Gelfand. 2004. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, New York.

Beliner, L., C. Wikle, and N. Cressie. 2000. Long-lead prediction of pacific ssts via bayesian dynamic modeling. *Journal of Climate*, 13:3953–3968.

Berger, J., V. De Oliveira, and B. Sanso. 2002. Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96:1361–1374.

Berger, J., B. Liseo, and R. Wolpert. 1999. Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 14:1–28.

Berliner, L. *In: Maximum Entropy and Bayesian Methods*, chapter Hierarchical Bayesian time series models, pages 15–22. Kluwer Academic Publishers, 1996.

Besag, J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistics Society*, 36:192–236.

Borchers, D., S. Buckland, and W. Zucchini. 2002. *Estimating Animal Abundance: Closed Populations*. Springer-Verlag, London, England.

Brown, J. and B. Maurer. 1989. Macroecology: The division of food and space among species on continents. *Science*, 243:1145–1150.

Buckland, S., K. Newman, L. Thomas, and N. Koesters. 2004. State-space models for the dynamics of wild animal populations. *Ecological Modelling*, 171:157–175.

Bullock, J., R. Kenward, and R. Hails, editors. 2002. *Dispersal Ecology*. Blackwell Science Ltd, Malden, MA.

Calder, C., M. Lavine, P. Muller, and J. Clark. 2003. Incorporating multiple sources of stochasticity into dynamic population models. *Ecology*, 84(6):1395–1402.

Carroll, R. and F. Lombard. 1985. A note on N estimators for the binomial distribution. *Journal of the American Statistical Association*, 80:423–426.

Caswell, H. 2001. *Matrix Population Models*. Sinauer Associates, Inc., Sunderland, MA.

Clark, J. 2003. Uncertainty and variability in demography and population growth: A hierarchical approach. *Ecology*, 84(6):1370–1381.

Clark, J. 2005. Why environmental scientists are becoming Bayesians. *Ecology Letters*, 8: 2–14.

Clark, J., S. Carpenter, M. Barber, S. Collins, A. Dobson, J. Foley, D. Lodge, M. Pascual, R. Pielke Jr., W. Pizer, C. Pringle, W. Reid, K. Rose, O. Sala, W. Schlesinger, D. Wall, and D. Wear. 2001. Ecological Forecasts: An emerging imperative. *Science*, 293(5530): 657–660.

Clark, J., M. Lewis, J. McLachlan, and J. HilleRisLambers. 2003. Estimating population spread: what can we forecast and how well? *Ecology*, 84(8):1979–1988.

Conner, R., D. Rudolph, and J. Walters. 2001. *The red-cockaded woodpecker: surviving in a fire-maintained ecosystem*. University of Texas Press, Austin, Texas.

Contributors of the International Tree-Ring Data Bank. Igbp pages/world data center for paleoclimatology, noaa/ncdc paleoclimatology program; boulder, colorado, 2005.

Cook, E. and R. Holmes. *Users Manual for Program ARSTAN*. Laboratory of Tree-Ring Research, University of Arizona, Tuscon, Arizona, 1986.

Cook, E., D. Meko, and D. Stahle. 1999. Drought reconstructions for the continental United States. *Journal of Climate*, 12:1145–1162.

Cressie, N. 1993. *Statistics for Spatial Data: Revised Edition*. John Wiley and Sons, New York, New York, USA.

Cunningham, R. and C. Hauser. The decline of Missouri forests between 1880 and 1920. In Walldrop, T., editor, *Proceedings of pine-hardwood mixtures: a symposium on the management and ecology of the type*, volume GTR, pages 34–37, Asheville, North Carolina, 1989. U.S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station.

Cushing, J., R. Constantino, B. Dennis, R. Desharnais, and S. Henson. 2003. *Chaos in Ecology: experimental nonlinear dynamics*. Academic Press, New York.

Elton, C. 1927. *Animal Ecology*. Sidgwick and Jackson, London, England.

Elton, C. 1958. *The Ecology of Invasions by Animals and Plants*. Methuen, London, England.

Fisher, R. 1937. The wave of advance of advantageous genes. *Annals of Eugenics*, 7: 355–369.

Forman, R. and M. Godron. 1986. *Landscape Ecology*. John Wiley and Sons, New York, New York, USA.

Fritts, H. 1976. *Tree Rings and Climate*. Academic Press, New York.

Frost, C. Four centuries of changing landscape patterns in the longleaf pine ecosystem. In Herman, S., editor, *The longleaf pine ecosystem: ecology, restoration, and management*, number 18, pages 17–44, Tallahassee, Florida, 1993. Tall Timbers Research Station.

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

Gelman, A., J. Carlin, H. Stern, and D. Rubin. 2004. *Bayesian Data Analysis, Second Edition*. Chapman & Hall/CRC, Boca Raton.

Givens, G. and J. Hoeting. 2005. *Computational Statistics*. Wiley, New Jersey.

Grimm, V., E. Revilla, U. Berger, F. Jeltsch, W. Mooij, S. Railsback, H.-H. Thulke, J. Weiner, T. Wiegand, and D. DeAngelis. 2005. Pattern-oriented modeling of agent-based complex systems: Lessons from ecology. *Science*, 310:987–991.

Guyette, R., R. Muzika, and D. Dey. 2002. Dynamics of an anthropogenic fire regime. *Ecosystems*, 5:472–486.

Haberman, R. 1987. *Elementary Applied Partial Differential Equations*. Prentice Hall, Inc., Englewood Cliffs, New Jersey, USA.

Hamilton, J. Historic and contemporary vegetation / environment relationships in the shortleaf pine region of the Missouri Ozarks. Master's thesis, University of Missouri, Columbia, Missouri, USA, 2003.

Hanski, I. 1999. *Metapopulation Ecology*. Oxford University Press, New York.

Hastings, A., K. Cuddington, K. Davies, C. Dugaw, S. Elmendorf, A. Freestone, S. Harrison, M. Holland, J. Lambrinos, U. Malvadkar, B. Melbourne, K. Moore, C. Taylor, and D. Thomson. 2005. The spatial spread of invasions: new developments in theory and practice. *Ecology Letters*, 8:91–101.

Heikkinen, J. and H. Hogmander. 1994. Fully bayesian approach to image restoration with an application in biogeography. *Applied Statistics*, 43(4):569–582.

Hengeveld, R. 1993. What to do about the North American invasion by the Collared Dove? *Journal of Field Ornithology*, 64:477–489.

Hilborn, R. and M. Mangel. 1997. *The Ecological Detective*. Princeton University Press, Princeton, New Jersey.

Hirzel, A., J. Hausser, D. Chessel, and N. Perrin. 2002. Ecological-niche factor analysis: How to compute habitat-suitability maps without absence data? *Ecology*, 83:2027–2036.

Hoeting, J., M. Leecaster, and D. Bowden. 2000. An improved model for spatially correlated binary responses. *Journal of Agricultural, Biological, and Environmental Statistics*, 5(1): 102–114.

Hogeweg, P. 1988. Cellular automata as a paradigm for ecological modeling. *Applied Mathematics and Computation*, 27:81–100.

Hogmander, H. and J. Moller. 1995. Estimating distribution maps from atlas data using methods of statistical image analysis. *Biometrics*, 51:393–404.

Holmes, E., M. Lewis, J. Banks, and R. Veit. 1994. Partial differential equations in ecology: Spatial interations and population dynamics. *Ecology*, 75:17–29.

Hooten, M., D. Larsen, and C. Wikle. 2003. Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landscape Ecology*, 18:487–502.

Hooten, M. and C. Wikle. 2006. A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. *Ecological and Environmental Statistics*, In Press.

Hudson, R. 1965. The spread of the Collared Dove in Britain and Ireland. *British Birds*, 58:105–139.

Jackson, J., M. Lennartz, and R. Hooper. 1979. Tree age and cavity initiation by red-cockaded woodpeckers. *Journal of Forestry*, 77:59–63.

Kolar, C. and D. Lodge. 2001. Progress in invasion biology: predicting invaders. *Trends in Ecology & Evolution*, 16:199–204.

Kot, M. 2001. *Elements of Mathematical Ecology*. Cambridge University Press, Cambridge, UK.

Krebs, C. 1978. *Ecology: The Experimental Analysis of Distribution and Abundance*. Harper & Row Publishers Inc., New York, New York.

Lee, Y., S. Qian, R. Jones, C. Barnes, G. Flake, M. O'Rourke, K. Lee, H. Chen, G. Sun, Y. Zhang, and D. Chen. 1990. Adaptive stochastic cellular automata: Theory. *Physica D*, 45:159–180.

Link, W., R. Barker, J. Sauer, and S. Droege. 1994. Within-site variability in surveys of wildlife populations. *Ecology*, 74:1097–1108.

MacArthur, R. and E. Wilson. 1967. *The Theory of Island Biogeography*. Princeton University Press, New Jersey.

MacKenzie, D., J. Nichols, J. Hines, M. Knutson, and A. Franklin. 2003. Estimating site occupancy, colonization, and local extinction when a species is detected imperfectly. *Ecology*, 84:2200–2207.

MacKenzie, D., J. Nichols, G. Lachman, S. Droege, J. Royle, and C. Langtimm. 2002. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83:2248–2255.

Nathan, R., G. Perry, J. Cronin, A. Strand, and M. Cain. 2003. Methods for estimating long-distance dispersal. *Oikos*, 103:261–273.

Nettles, V., J. Shaddock, R. Sikes, and C. Reyes. 1979. Rabies in translocated raccoons. *American Journal of Public Health*, 69:601–602.

Neubert, M. and H. Caswell. 2000. Demography and dispersal: Calculation and sensitivity analysis of invasion speed for structured populations. *Ecology*, 81:1613–1628.

Olkin, I., A. Petkau, and J. Zidek. 1981. A comparison of N estimators for the binomial distribution. *Journal of the American Statistical Association*, 76:637–642.

Pace, R., R. Barry, O. Gilley, and C. Sirmans. 2000. A method for spatial-temporal forecasting with an application to real estate prices. *International Journal of Forecasting*, 16:229–246.

Papoulis, R. and S. Pillai. 2002. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill.

Preisendorfer, R. 1988. *Principal Component Analysis in Meteorology and Oceanography*. Elsevier.

Raftery, A. 1988. Inference for the binomial N parameter: a hierarchical Bayes approach. *Biometrika*, 75:223–228.

Reckhow, K. 1990. Bayesian inference in non-replicated ecological studies. *Ecology*, 71: 2053–2059.

Renshaw, E. 1991. *Modeling biological populations in space and time*. Cambridge University Press, Cambridge, UK.

Robbins, C., D. Bystrak, and P. Geissler. The Breeding Bird Survey: its first fifteen years, 1965-1979. Fish and Wildlife Service Resource Publication 157, USDOI, Washington, DC. USA, 1986.

Romagosa, C. and R. Labisky. 2000. Establishment and dispersal of the Eurasian Collared-Dove in Florida. *Journal of Field Ornithology*, 71:159–166.

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

Royle, J. 2006. Modeling spatio-temporal dynamics in occupancy: Nest site occupancy in a Kittiwake colony. *In Review*.

Royle, J. and R. Dorazio. 2006. Hierarchical models of animal abundance and occurrence. *Journal of Agricultural, Biological, and Environmental Statistics*, In Press.

Royle, J. and M. Kery. 2006. A Bayesian state-space formulation of dynamic occupancy models. *In Review*.

Sauer, J., B. Peterjohn, and W. Link. 1994. Observer differences in the North American Breeding Bird Survey. *Auk*, 111:50–62.

Shigesada, N. and K. Kawasaki. 1997. *Biological Invasions: Theory and Practice*. Oxford University Press, New York.

Shigesada, N. and K. Kawasaki. *Dispersal Ecology*, chapter Invasion and the long range expansion of species: Effects of long-distance dispersal. Blackwell Publishing, Malden, Massachusetts, 2002.

Skellam, J. 1951. Random dispersal in theoretical populations. *Biometrika*, 38:196–218.

Smith, D., B. Lucey, L. Waller, J. Childs, and L. Real. 2002. Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *Proceedings of the National Academy of Sciences of the United States of America*, 99:3668–3672.

Stambaugh, M. Forest canopy gap disturbances in shortleaf pine forests of the Ozark highlands. Master's thesis, University of Missouri, Columbia, Missouri, USA, 2001.

Stambaugh, M. and R. Guyette. 2004. Long-term growth and climate response of Shortleaf pine at the Missouri Ozark Forest Ecosystem Project. *Proceedings of the 14th Central Hardwood Forest Conference*, (GTR-NE-316):448–458.

Stokes, M. and T. Smiley. 1996. *An Introduction to Tree Ring Dating*. University of Arizona Press, Tuscon, Arizona.

Su, Z., R. Peterman, and S. Haeseker. 2004. Spatial hierarchical Bayesian models for stock-recruitment analysis of pink salmon (*Oncorhynchus gorbuscha*). *Canadian Journal of Fisheries and Aquatic Sciences*, 61:2471–2486.

Thomas, A. 1994. Bugs: A statistical modelling package. *RTA/BCS Modular Languages Newsletter*, 2:36–38.

Thomas, L., S. Buckland, K. Newman, and J. Harwood. 2005. A unified framework for modelling wildlife population dynamics. *Australian and New Zealand Journal of Statistics*, 47:19–34.

Turchin, P. 2003. *Complex Population Dynamics*. Princeton University Press, Princeton, NJ.

USFWS. 2003. *Recovery plan for the red-cockaded woodpecker (*Picoides borealis*)*. U.S. Fish and Wildlife Service, Atlanta, Georgia.

Von Storch, H., G. Burger, R. Schnur, and J. Von Storch. 1995. Principal oscillation patterns: A review. *Journal of Climate*, 8:377–400.

Waller, L. and C. Gotway. 2004. *Applied Spatial Statistics for Public Health Data*. John Wiley & Sons, New Jersey.

Walters, J. *Cooperative breeding in birds*, chapter Red-cockaded woodpeckers: a primitive cooperative breeder. Cambridge University Press, London, UK, 1990.

Wikle, C. *Encyclopedia of Life Support Systems*, chapter Spatio-temporal models in climatology. EOLSS Publishers Co. Ltd., 2002a.

Wikle, C. 2002b. A kernel-based spectral model for non-gaussian spatio-temporal processes. *Statistical Modelling: An International Journal*, 2:299–314.

Wikle, C. 2003. Hierarchical Bayesian methods for predicting the spread of ecological processes. *Ecology*, 84:1382–1394.

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

Wikle, C. and N. Cressie. 1999. A dimension reduced approach to space-time Kalman filtering. *Biometrika*, 86:815–829.

Wikle, C. and M. Hooten. *Applications of Computational Statistics in the Environmental Sciences: Hierarchical Bayes and MCMC Methods*, chapter Hierarchical Bayesian Spatio-Temporal Models for Population Spread. Oxford University Press, 2006.

Wikle, C., R. Milliff, D. Nychka, and L. Berliner. 2001. Spatio-temporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of the American Statistical Association*, 96:382–397.

Wikle, C. and J. Royle. 2005. Predicting migratory bird settling patterns with hierarchical bayesian spatio-temporal models. *In Review*.

Wolfram, S. 1983. Statistical mechanics of cellular automata. *Reviews of Modern Physics*, 55:601–644.

Wolfram, S. 1984. Cellular automata as models for complexity. *Nature*, 311:419.

Wolfram, S. *In: High-Speed Computing: Scientific Applications and Algorithm Design*, chapter Cellular Automaton Supercomputing, pages 40–48. University of Illinois Press, 1988.

Xu, K., C. Wikle, and N. Fox. 2005. A kernel based spatio temporal dynamical model for nowcasting radar precipitation. *Journal of the American Statistical Association*, To Appear.

Zhao, X. 2003. *Dynamical Systems in Population Biology*. Springer-Verlag, New York.

Zhu, J., H.-C. Huang, and C.-T. Wu. 2005. Modeling spatial-temporal binary data using Markov random fields. *Journal of Agricultural, Biological, and Environmental Statistics*, 10:212–225.

VITA

Mevin Brice Hooten was born in Bogota, Colombia, on May 13, 1976. Son of Julie and Andrew Hooten, he was raised in Stilwell, Kansas, and graduated from Blue Valley High School. He earned his B.S. from Kansas State University (1999) in Natural Resource Management with minor in Wildlife Biology. He moved to Columbia, Missouri, married Gina Groat, and pursued a M.S. in Forestry at the University of Missouri under David R. Larsen. Upon completion of his Masters degree (2001), Mevin entered the Ph.D. program in Statistics at the University of Missouri under the direction of Christopher K. Wikle. He has accepted a position as Assistant Professor of Statistics in the Department of Mathematics and Statistics at Utah State University in Logan, Utah, and will begin in August of 2006.