HIERARCHICAL SPATIO-TEMPORAL MODELS FOR ENVIRONMENTAL 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
ALI ARAB
Dr. Christopher K. Wikle, Dissertation Supervisor
AUGUST 2007

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

HIERARCHICAL SPATIO-TEMPORAL MODELS FOR ENVIRONMENTAL PROCESSES

Presented by Ali Arab

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

Dr. Christopher K. Wikle

Dr. Scott Holan

Dr. Mark L. Wildhaber

Dr. Neil I. Fox

Dr. Athanasios Micheas

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Chris Wikle, for all the help, support and education I have received from him in the past few years. His incredible mentorship, creative ideas, and positive outlook are truly inspiring.

I would also like to thank my doctoral committee members: Dr. Sakis Micheas, Dr. Mark Wildhaber, Dr. Scott Holan, and Dr. Neil Fox for their help and support. Also a special thanks to Mevin Hooten, the hardworking and kind staff at the MU Statistics department (Tracy Pickens, Judy Dooley, Margie Gurwit, Clay Pickens, Ray Bacon), all my fellow graduate students at the Department of Statistics and my colleagues at the U.S. Geological Survey- Columbia Environmental Research Center (with special thanks to Janice Bryan for help with data) for all their help and support throughout all these years.

I dedicate this work to my parents, Elaheh and Hossein for all the sacrifices they made throughout their life to help me accomplish my dreams.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 General Introduction

1.1 Environmental Processes

The processes governing environmental systems are often complex, involving different interacting scales of variability in space and time. Scientists and engineers are often interested in making predictions regarding consequences of changes in environmental systems, whether such changes are caused by natural phenomena (e.g., floods, hurricanes), engineering designs (e.g., water transfer designs, construction of dams), or other human-related changes (e.g., air pollution, water pollution, animal harvest). Such prediction problems are often carried out using numerical models which are based on basic laws of physics (Abbott and Basco 1989). For example, in problems involving incompressible fluid flow, the mathematical models describing the process are often based on governing equations of fluid dynamics (e.g., mass conservation, momentum conservation, and energy conservation) (e.g., Zienkiewicz and Taylor 2000). The deterministic nature of these models limits their ability to effectively describe "real-world" situations. Another important shortcoming of using such models is the inability of the deterministic modeling framework to take advantage of useful and often easy to access data obtained from the system under study. The limitations of such modeling approaches are intensified by elements such as uncertainties involved in choices for parameters governing the dynamics, and more importantly, the uncertainty in appropriateness of the model considered to describe the underlying process.

Engineers and scientists, knowing the limitations of the deterministic models, have made attempts to address these shortcomings in the recent decades. The development of inverse modeling, data assimilation procedures, and stochastic differential equations are instances

of such attempts to account for uncertainties in modeling environmental and physical processes (e.g., Lynch 2005). However, these methods are often very complicated and unable to effectively account for different sources of uncertainty as well as assimilation of different sources of data into the model.

Statistical modeling provides a more realistic framework to address uncertainties involved in the process under study. However, issues such as complexity of model structure, high dimensionality, and non-identifiability of model parameters, present challenges which necessitate the utilization of *a priori* information based on scientific knowledge and expert opinion. The use of solutions to partial diffusion equations (PDEs), and also stochastically-forced differential equations, in developing stochastic models have been studied in the recent years (e.g., Liebelt 1967; Wunsch 1996; Lynch and McGillicuddy 2001; Lynch 2005). However, efficient implementation of such modeling frameworks is often not extendable to "real-world" spatio-temporal processes, due to inappropriateness or computational inefficiency of covariance-based models. Moreover, conventional covariance-based methods are often incapable of allowing the researcher to quantify uncertainties corresponding to the model parameters since the parameter space of most complex spatial and spatio-temporal models is very large. Hierarchical Bayesian modeling has attracted researchers in recent years, providing a flexible modeling alternative to address model and parameter uncertainties while accounting for mathematical models describing the spatio-temporal processes in the physical and environmental sciences (e.g., Wikle et al. 2001; Wikle 2003).

1.2 Modeling Environmental Spatio-Temporal Processes

Spatio-temporal processes are often complex, exhibiting different scales of spatial and temporal variability. Such processes are typically characterized by a large number of observations and prediction locations in space and time, differing spatial and temporal support, orientation and alignment (relative to the process of interest), and complicated underlying dynamics. The complexity of such processes in "real-world" situations is often intensified due to the difficulty of applying simplifying assumptions such as Gaussianity, spatial and temporal stationarity, linearity, and space-time separability of the covariance function. Thus, a joint perspective for modeling spatio-temporal processes, although relatively easy to formulate, is challenging to implement. On the contrary, a hierarchical formulation allows the modeling of complicated spatial and temporal structures by decomposing an intricate joint spatio-temporal process into relatively simple conditional models. The main advantage of the Bayesian hierarchical model over traditional covariance-based methods is that it allows the complicated structure to be modeled at a lower level in the hierarchy, rather than attempting to model the complex joint dependencies.

1.2.1 Hierarchical Models

The issues involved in modeling complex spatio-temporal processes necessitate a flexible modeling framework which allows for estimation and prediction of such processes, as well as the parameters governing the dynamics. A joint modeling approach to such complicated processes often fails due to inability to accommodate an efficient parametrization of the model and lack of an effective means to address spatial and temporal dependency structures.

However, a hierarchical modeling perspective allows for modeling complex processes by decomposing the problem into a series of simple sub-problems linked by simple rules of probability.

A hierarchical spatio-temporal model is the product of conditional distributions for data conditioned on a spatio-temporal process and parameters, the spatial process conditioned on the parameters defining the spatial dependencies between process locations, and the parameters themselves. Hierarchical modeling has many advantages such as allowing for multiple sources of data, and the ability to consider scientifically meaningful structures in the model, *a priori*.

In modeling complex processes in the presence of data, the hierarchical model is described in terms of three basic stages (e.g., Berliner 1996):

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

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

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

The notation $[x]$ refers to the probability distribution of x, and the notation $[x|y]$ refers to the conditional probability distribution of x given y. The basic idea is to approach the complex problem by breaking it into simpler sub-problems. Although hierarchical modeling is not new to statistics (Lindley and Smith 1972), this basic formulation for modeling complicated spatial and spatio-temporal processes in the environmental sciences is a relatively new development (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; Wikle et al. 2001).

Hierarchical modeling can be computationally expensive and challenging. However, a Bayesian implementation of such models is often easy to implement. In the Bayesian framework, one must assign prior probability densities to the unknown parameters of interest and obtain the distribution of the process and parameters updated by the data via Bayes' rule (i.e. posterior distributions):

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

In a Bayesian modeling framework, inference is based on the posterior distribution. The posterior distribution is often very complex and the normalizing constant integral cannot be analytically solved. Instead, one can simulate from the posterior distribution in a Monte Carlo framework to do inference. One popular approach used for simulating from the posterior distributions is called Markov Chain Monte Carlo (MCMC). MCMC methods are a class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its stationary distribution (Gelfand and Smith 1990) and includes such algorithms as Metropolis-Hastings and the Gibbs sampler (Casella and George 1992; Robert and Casella 2005). Gibbs Sampling is the primary algorithm of the freely-distributed software WinBUGS (Spiegelhalter et al. 2003; http://www.mrc-bsu.cam.ac.uk/bugs). A key issue in implementation of the Gibbs sampler (or any other MCMC sampler) is that the number of iterations of the algorithm should be

large enough to guarantee that the chain approaches stationarity (i.e., convergence to the target density). Typically an initial period of iterations are considered as the "burn-in" period of the chain and are thrown-out (Congdon 2001). The number of burn-in iterations required can be influenced by the choice of starting value as well as dependencies in model parameters.

1.2.2 General Spatio-Temporal Models

Let $Y(s,t)$ be a spatio-temporal process where $s \in D_s$, and D_s is a continuous or discrete spatial domain and $t \in D_t$ is a discrete temporal domain. The generality of the definition of the spatial domain allows for the spatio-temporal process to be applicable to both cases of continuous data and areal data. A general model relating data $Z(s,t)$ to the "unobserved" process $Y(s,t)$ (in the special case where $Z(s,t)$ and $Y(s,t)$ have the same spatial support and there is no missing data) can be written as:

$$Z(s,t) = Y(s,t) + \varepsilon(s,t), \tag{1.1}$$

where $\varepsilon(s,t)$ is a zero-mean measurement error process. The underlying process $Y(s,t)$ can be further decomposed into a mean process, additive error process, and spatial or temporal random effects (e.g., Stein 1986).

Recent approaches to spatio-temporal modeling have focused on the specification of joint space-time covariance structures (e.g., Cressie and Huang 1999; Stein 2005). However, in high-dimensional settings with complicated non-linear spatio-temporal behavior, such covariance structures are very difficult to formulate and/or implement. An alternative approach to modeling such complicated processes is to use spatio-temporal dynamic models

in a hierarchical fashion.

1.2.3 Spatio-Temporal Dynamic Models

Many spatio-temporal processes are dynamic in the sense that the current state of the process is a function of the previous states. There are many examples of spatio-temporal models with dynamic components in the literature (e.g.,West and Harrison 1989; Wikle et al. 1998; Wikle and Cressie 1999; Stroud et al. 2001; Wikle 2003). The joint spatio-temporal process can be factored into conditional models based on a Markovian assumption:

$$[\mathbf{Y} \mid \{\boldsymbol{\theta}_t, t = 1, \ldots, T\}] = [\mathbf{y}_0] \prod_{t=1}^{T} [\mathbf{y}_t | \mathbf{y}_{t-1}, \boldsymbol{\theta}_t], \tag{1.2}$$

where $\boldsymbol{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_T)$, $\mathbf{y}_t = (y(s_1, t), \ldots, y(s_n, t))'$ and the conditional distribution $[\mathbf{y}_t | \mathbf{y}_{t-1}, \boldsymbol{\theta}_t]$ depends on a vector of parameters $\boldsymbol{\theta}_t$ which govern the dynamics of the spatio-temporal process of interest. Note that the are no observations available for the initial state $\mathbf{y}_0$. However in a Bayesian modeling framework this initial state of the dynamical spatio-temporal model can be estimated through the estimated dynamical relationship between the model states (i.e., propagator matrix) and use of a prior distribution (e.g., $\mathbf{y}_0 \sim N(\tilde{\mathbf{y}}_0, \boldsymbol{\Sigma}_0)$ where $\tilde{\mathbf{y}}_0$ and $\boldsymbol{\Sigma}_0$ are known). An example of such spatio-temporal dynamic models is when the process has a first-order Markovian structure:

$$\mathbf{y}_t = \mathbf{H}_{\theta_t} \mathbf{y}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \tag{1.3}$$

where $\boldsymbol{\eta}_t$ is a spatial error process, and $\mathbf{H}_{\theta_t}$ is a "propagator matrix" (sometimes called a transition or evolution matrix) which includes the parameters that govern the dynamics of the process. If these parameters are known or easy to estimate, an implementation of the

model through Kalman filtering is possible (e.g., Shumway and Stoffer 2005 ; for spatio-temporal problems see Wikle and Cressie 1999). If the parameters are unknown, $\mathbf{H}_{\theta_t}$ can be modeled in a hierarchical fashion by specifying prior distributions for $\mathbf{H}_{\theta_t}$ or its parameters $\boldsymbol{\theta}_t$. The hierarchical form for spatio-temporal dynamic models is sometimes motivated by partial differential equations (PDEs) that describe the approximate behavior of underlying physical processes (e.g., Wikle 2003).

1.2.4 PDE-based Dynamics

Spatio-temporal dynamical processes in the physical and environmental sciences are often described by partial differential equations (PDEs). The inherent complexity of such processes due to high-dimensionality and multiple scales of spatial and temporal variability is often intensified by characteristics such as sparsity of data, complicated boundaries and irregular geometrical spatial domains, among others. In addition, uncertainties in the appropriateness of any given PDE for a "real-world" process, as well as uncertainties in the parameters associated with the PDEs are typically present. These issues necessitate the incorporation of efficient parameterizations of spatio-temporal models that are capable of addressing such characteristics.

Finite difference methods are intuitive and easy to implement methods to discretize/solve PDEs which result in a simple system of multiple equations. The symmetric and sparse dynamical structure obtained by finite difference methods often provides an efficient parametrization of the PDE. Such a framework can be effectively utilized in the context of hierarchical Bayesian models (e.g., Wikle 2003; Wikle et al. 2003). However, the disadvantages and limitations of the finite difference method, such as its difficulty in dealing with irregular ge-

ometrical spatial domains and natural boundary conditions, necessitate the use of methods that are able to address such issues.

The Galerkin finite element method (FEM) is another popular numerical method which provides a flexible tool to discretize/solve PDEs. Finite elements provide an approximation to the solution of the PDE, by means of compact support basis functions and thus, it is applicable to complex geometries, as well as, regular geometrical domains. Using finite element methods to discretize a PDE, one has to first consider a variational formulation of the PDE and then discretize the problem by choosing appropriate compactly supported basis functions (e.g., Zienkiewicz and Taylor 2000). The basis functions are low-order polynomial functions which are defined to take value 1 on a single node and value 0 everywhere else. The accuracy of solutions of the PDEs as interpolated between the grid points (nodes) in the finite element method is usually much higher than finite difference methods. The compact support basis functions used in the finite element method applied to transient PDEs provide a sparse structure for the dynamical model. Such attractive attributes of the finite element method motivate a hybrid statistical-physical setting to model environmental spatio-temporal processes. In Chapter 2, hierarchical Bayesian modeling of PDE-based dynamics for spatio-temporal processes motivated by their Galerkin finite element method representations are considered and simple implementations based on advection-diffusion equations are discussed.

1.2.5 Count Processes: Multivariate and Zero-Inflated Extensions

Environmental studies often include count processes that interact on different scales. Although the study of such interactions is of great interest to researchers, modeling approaches

required for conducting such analyses are not well-developed. Some of the factors contributing to such sparsity of methods in the literature include lack of multivariate distributions that could provide an effective dependence structure, as well as the usual complexities involved in joint modeling of problems in high dimensions and/or complicated covariance structures. The analysis of count processes is often considered with simplifying assumptions to reduce the problem to univariate cases for which an extensive methodology exists.

Multivariate count processes: Accounting for dependencies among different count processes (e.g., abundance of species) is often necessary for environmental and ecological scientists due to complex relationships among the processes. However, such complex dependency structures are often hard or impossible to model using conventional methods (e.g., ANOVA). Hierarchical Bayesian modeling provides a more straightforward and easy to implement modeling framework for modeling multivariate counts. For example, in the literature, there are several approaches described for implementing multivariate Poisson regression (Tsionas 2001; Congdon 2001; Congdon 2005). We consider a semiparametric hierarchical Bayesian modeling framework for modeling count processes and study the description of dependency structures in the model from two different perspectives, each providing a subtle, yet scientifically different setting. Such different approaches for accounting for dependence between count processes is necessitated by the diverse nature of such dependencies in "real-world" problems.

Zero-inflated Poisson processes: Zero-inflated discrete models and, in particular, zero-inflated Poisson (ZIP) models, have gained popularity among researchers in the recent years due to their ability to account for excess zeros, a problem ubiquitous in environmental

studies. The presence of excess zeros in environmental studies and ecological surveys often includes two different (and theoretically distinguishable) types of zero values; structural zeros and sampling zeros. Structural zeros are the zero values that correspond to a true absence of the process (e.g., species not present) at that particular site. Sampling zeros correspond to sites were the process was present (e.g., species occurred) but was not detected (Royle 2006). Sampling zeros are unavoidable in habitat analysis due to sensitivity of observations to habitat conditions and gear detectability issues. For example, in ecological habitat management, sampling zeros can generate serious implications in the analysis which can influence the ability to obtain accurate inference from the data (MacKenzie et al. 2003; Moilanen 2002). Martin et al. (2005) make recommendations on the choice of appropriate modeling approaches to model the source of zeros. In reality, often there exists a mixture of sampling zeros and structural zeros which can be addressed using a zero-inflated model if we have information on the detection probabilities. A so-called "hurdle" model or a two-stage modeling approach is common for modeling heavy-zero data when structural zeros are believed to be the only source of zeros (Greene 2003; Lambert 1992). For example, zero-inflated models provide a natural setting for modeling the abundance of rare species (Welsh et al. 1996).

Lambert (1992) employs an Expectation-Maximization (EM) algorithm (Hartley 1958; Dempster et al. 1977) to obtain the maximum likelihood estimates for the ZIP parameters. Hall (2000) adapts Lambert's methodology to a ZIP model with random effects. Maximum likelihood estimation is also possible by using non-linear mixed model estimation method such as PROC NLMIXED in SAS (e.g., Littell et al. 2006). Although the optimization

techniques used by most of these methods (such as PROC NLMIXED) are some of the best ones available, convergence problems are often observed for complex models. Here, we consider a Bayesian implementation for the ZIP models with random effects (e.g., Wikle and Anderson 2003; Martin et al. 2005). The Bayesian implementation provides a more flexible and reliable estimation tool.

The Bayesian modeling framework for zero-inflated models is a flexible modeling approach which not only provides a tool for the researcher to simultaneously model data with a high percentage of zeros but also enables him/her to include scientific knowledge and/or beliefs into the model by the means of assigning prior probabilities to the unknown variables and using data to update these beliefs (Wikle and Anderson 2003). The coefficients in the model are random effects. Furthermore, in the Bayesian framework, inferential statements on model parameters (i.e., credible intervals) and p-values on hypotheses are more in line with common sense interpretations (Congdon 2001).

In Chapter 3, semiparametric hierarchical Bayesian modeling of multivariate zero-inflated Poisson processes are presented. Two different modeling approaches based on a multivariate normal distribution and a multivariate Poisson distribution will be considered and results of applying these two different approaches to an application for modeling of benthic fishes abundance in the Missouri River will be discussed. This modeling approach can easily be extended to model spatio-temporal count processes as discussed in chapter 4.

1.2.6 Multiresolution and Dynamical-Resolution methods for Spatio- Temporal Processes

Spatio-temporal processes are often high-dimensional and include interactions across various spatial scales. This fact makes multiresolution approaches to modeling such processes in a way that allows for effective dimension reduction, both appealing and necessary. One approach to conduct modeling for such high-dimensional processes is through dimension-reduced spectral representations of such processes. For example, let $\mathbf{u}_t = (u_1(s_1), \ldots, u_T(s_n))'$ indicate a process at time t for locations $s_1, \ldots, s_n$. A general spectral representation of $\mathbf{u}_t$ (vector of length n) can be written as

$$\mathbf{u}_t = \Phi \mathbf{a}_t,$$

where Φ is an $n \times k$ matrix of spectral basis functions of choice (e.g., wavelets, EOFs, spline bases) and $\mathbf{a}_t$ is a vector of length k ($k \ll n$). Thus, the modeling procedure can be done using $\mathbf{a}_t$ which has a substantially lower dimension compared to $\mathbf{u}_t$. The choice of basis functions in Φ can accommodate various methods for efficient modeling of spatio-temporal process. For example, using a discrete wavelet transform (DWT) the process can be modeled in separate, but related, resolutions (several levels of coarse and fine resolutions) resulting in a multiresolution modeling approach. This approach is discussed in Chapter 5 and examples are presented.

A useful extension of the described modeling approaches in the spectral domain (rather than in the physical/spatial domain) is to let matrix Φ be time-varying (i.e., Φ_t). This assumption, although being more intuitive and resulting in a more flexible modeling framework, can be very difficult to implement and requires efficient model parameterizations in

order to be able to estimate the model parameters (note that the number of parameters can increase dramatically in such settings). In Chapter 5, a dynamical-resolution model for spatio-temporal processes is discussed where a spline representation of the model based on thin-plate spline functions with both fixed and moveable knots is employed. In this modeling approach, the dynamical-resolution is accommodated in a hierarchical Bayesian modeling framework using a moveable knot for which the location is estimated dynamically, at every step of the MCMC algorithm.

1.3 Overview of Chapters

Chapter 2 describes an efficient parameterization of PDE-based dynamics motivated by a Galerkin finite element representation. This approach, implemented within a hierarchical Bayesian framework, allows for modeling the complex spatio-temporal dynamical processes within both regular and irregular geometric domains. Problems such as open boundary condition PDEs can be considered and treated stochastically within this modeling framework. An application to the modeling of Sturgeon migration patterns based on data obtained from radio-tracking tags is discussed. The approach discussed in this chapter provides a novel method for implementing an efficiently-parameterized physical/statistical model using a Galerkin finite element representation for PDE-based dynamical spatio-temporal processes.

Chapter 3 considers the modeling of multivariate spatio-temporal zero-inflated count processes with applications to fisheries community-level modeling, accounting for dependencies among species. Two different alternative methods for building dependency structures into the model based on the multivariate normal distribution and multivariate Poisson distribution are presented and the applicability of each of these approaches is discussed. Also,

a semiparametric modeling approach is considered in modeling the abundance of benthic fishes of the Missouri River which allows for non-linearities involved in the problem by considering non-linear functions of the effects of physical variables such as temperature and turbidity. Model comparison and selection is conducted to identify the necessity of such non-linear terms in the models. The methodology described in this chapter is a novel approach for modeling multivariate count processes with excess zeros for which no previous examples are found in the literature.

Chapter 4 describes an application of a hierarchical Bayesian multivariate zero-inflated modeling approach for two possibly correlated spatio-temporal count processes. The application consists of modeling tornado report counts for tornadoes as classified by the "Fujita" scale. In the model described in this chapter, the effect of the El Niño/Southern Oscillation (ENSO) on F0-F1 ("less damaging") and F2-F5 ("more damaging") tornadoes is considered and spatial patterns of this effect are obtained. Also, a spatially-varying temporal trend for these two categories of tornadoes is derived. The effect of population intensity on the identification of tornadoes is modeled as a covariate related to the zero-inflation probability. This chapter describes a spatio-temporal extension of the method described in Chapter 3 and provides a new method to model multivariate spatio-temporal count processes with excess zeros.

Chapter 5 describes two different approaches for efficiently modeling high-dimensional spatio-temporal processes by means of conducting dynamical modeling in much lower dimensional setting. The first method consists of developing a multiresolution hierarchical model by taking advantage of the multiresolution structure of Discrete Wavelet Transform (DWT).

Using this approach first a multiresolution wavelet transform of the high-dimensional dynamical spatio-temporal process is obtained, then the modeling procedure is implemented within each resolution while accounting for proper interactions between these resolutions. Such implementation of multiresolution methods in statistical modeling is novel and no previous examples can be found in the literature. The second approach considers a low-dimensional spline-based model using thin-plate spline bases with both fixed and moveable knots where the location of the moveable knot is dynamically estimated at every iteration of the MCMC simulation. The novelty of this approach is in using moveable knots in the context of spline-based models where one can estimate the location of the moveable knots, while accounting for a dynamical model for the fixed knots.

2 Galerkin Finite Element Approaches for Efficiently Parameterizing Spatio-Temporal Dynamical Processes

2.1 Introduction

The application of spatio-temporal methods in the physical and environmental sciences has become increasingly popular with technological advances in remote sensing, monitoring networks and other methods of collecting spatial and temporal data. Parallel advancements in computational techniques in the last decades have made such efforts possible. However, the complexity of such processes requires the development of more advanced methods to enable scientists to efficiently take advantage of the information available to them. Statistical models play an essential role in such research efforts and hierarchical Bayesian models offer flexible tools for researchers to combine scientific knowledge (e.g. mathematical equations based on laws of physics or representation of dynamics) and data from various spatial and temporal scales (Wikle et al. 2001; Wikle 2003).

In this chapter, new methods for the efficient incorporation of PDE-based spatio-temporal processes in hierarchical statistical models are described. The hierarchical Bayesian Galerkin finite element-based spatio-temporal models considered in this chapter potentially can be applied to dynamical processes with irregular domains, random, spatially-varying parameters and complicated (stochastic or deterministic) boundary conditions. The next section describes the Galerkin finite element method in general, and then relative to one-dimensional and two-dimensional advection-diffusion PDEs. Section 3 describes the hierarchical Bayesian modeling approach to finite element-based dynamical processes described

in Section 2. Section 4 describes simulations based on one-dimensional and two-dimensional advection-diffusion PDEs and simulation results are presented and discussed in Section 5. Section 6 discusses the application of a one-dimensional advection-diffusion model to the problem of estimating migration (advection) and diffusion rates for shovelnose sturgeon in the Missouri River and the results are presented in Section 7 followed by a discussion in Section 8. Finally, Section 9 contains a conclusion.

2.2 Galerkin Finite Element Method

The Galerkin procedure is a method for solving PDEs in which the dependent variables are represented by a sum of functions that have a prescribed spatial structure and an associated coefficient. Such methods transform a PDE into a set of ordinary differential equations (ODEs). The two most common Galerkin methods are spectral methods and Finite Element Methods (FEM). Spectral methods employ orthogonal basis functions and the FEM usually employs low-order polynomials with compact support as basis functions. This feature of the Galerkin FEM motivates efficient parameterization of spatio-temporal dynamical models. The Galerkin methods, and particularly the FEM, also provide modeling tools to address issues such as flexible resolution and complicated boundaries, both of special interest in modeling physical and environmental processes.

To illustrate the Galerkin procedure, consider the deterministic differential equation:

$$L(U_t(x)) = f(x_t),$$

where L is the differential operator, $U_t(x)$ is the dependent variable and $f(x_t)$ is the specified forcing function. Thus, $U_t(x)$ can be approximated using a series of linearly independent

basis functions, $\phi_i(x)$:

$$U_t(x) \approx \sum_{j=1}^{N} u_{t,j}\phi_j(x).$$

The error in satisfying the differential equation is:

$$e_N = L\left(\sum_{j=1}^{N} u_{t,j}\phi_j(x)\right) - f(x_t).$$

The Galerkin procedure requires that the error be orthogonal to each basis function:

$$e_N = \int_\Omega e_N\phi_j(x)dx = 0, \qquad \text{for } j = 1, \ldots, N,$$

where Ω represents the domain of the study (e.g., open region in the (x, y) plane). This requirement guarantees the Galerkin method to find the best approximation in the solution space (i.e., the solution that is closest to the actual solution).

2.2.1 Choice of Basis Functions

In principle, any set of compactly-supported basis functions can be used in the Galerkin FEM. However, low-order polynomials are often used as basis functions for simplicity (Burnett 1987). The basis functions act as local interpolators within the elements. Often, simple piecewise linear basis functions perform well for such interpolations (e.g., Burnett 1987; Kwon and Bang 2000; Lynch 2005). In the following examples, simple low-order polynomial basis functions will be considered.

2.2.2 Example 1: Advection-Diffusion Equation (1-D)

Consider the following one-dimensional advection-diffusion equation:

$$\frac{\partial U(x,t)}{\partial t} + \alpha(x)\frac{\partial}{\partial x}U(x,t) - \frac{\partial}{\partial x}\left(\beta\frac{\partial U(x,t)}{\partial x}\right) = 0, \tag{2.4}$$

where $\alpha(x)$ corresponds to the spatially-varying advection parameter and β corresponds to the diffusion parameter. Note that the diffusion parameter, β, can also be spatially-varying, but for illustration is considered constant here. The Galerkin procedure requires

$$\int_\Omega \left[\frac{\partial U(x,t)}{\partial t} + \alpha(x) \frac{\partial}{\partial x} U(x,t) - \frac{\partial}{\partial x} \left(\beta \frac{\partial U(x,t)}{\partial x} \right) \right] \phi_i(x) dx = 0, \quad \text{for } i = 1, \ldots, N \quad (2.5)$$

where,

$$U(x,t) = \sum_{j=1}^{N} u_j(t) \phi_j(x). \quad (2.6)$$

To proceed, one must specify the basis functions, $\phi_j(x), j = 1, \ldots, N$ (Figure 1). As an example, the "hat" function which is a compactly supported basis function (i.e., it is non-zero only in elements containing node j; Kwon and Bang 2000), is considered and is defined as

$$\phi_j(x) = \begin{cases} \frac{x - x_{j-1}}{\Delta x} & \text{if } x_{j-1} \le x \le x_j \\ \frac{x_{j+1} - x}{\Delta x} & \text{if } x_j \le x \le x_{j+1} \\ 0 & \text{otherwise,} \end{cases}$$

where $\Delta x = x_j - x_{j-1} = x_{j+1} - x_j$. A generalization of the basis functions to the case where $x_j - x_{j-1} \neq x_{j+1} - x_j$ is also possible. Now, the derivative of $\phi_j(x)$ with respect to x is defined as

$$\frac{d\phi_j(x)}{dx} = \begin{cases} \frac{1}{\Delta x} & \text{if } x_{j-1} \le x \le x_j \\ \frac{-1}{\Delta x} & \text{if } x_j \le x \le x_{j+1} \\ 0 & \text{otherwise.} \end{cases}$$

Equation (2.5) involves a second derivative of the piecewise linear basis function so integration by parts can be applied to the term involving the second derivative to avoid

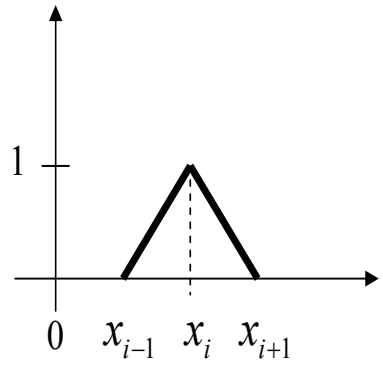


Figure 1: Schematic for the "hat" basis function.

misrepresentation of the boundary conditions (Burnett 1987):

$$\int_\Omega \frac{\partial U(x,t)}{\partial t}\phi_i(x)dx + \int_\Omega \boldsymbol{\alpha}(x)\frac{\partial U(x,t)}{\partial x}\phi_i(x)dx - \left[-\beta\frac{\partial U(x,t)}{\partial x}\phi_i(x)\right]_\Omega + \int_\Omega \frac{\partial \phi_i(x)}{\partial x}\beta\frac{\partial U(x,t)}{\partial x}dx = 0, \qquad \text{for } i = 1,\ldots,N, \tag{2.7}$$

where Ω denotes the 1-D domain of the study (i.e., Ω is the set of all the elements). Note that the boundary term $[-\beta\frac{\partial U(x,t)}{\partial x}\phi_i(x)]_\Omega$ contains the *flux*, the quantity $\beta\frac{\partial U(x,t)}{\partial x}$, which usually has a physical interpretation and can take boundary values (zero in this case).

Now, using the definition of the basis function and the separation of variables in (2.6) yields:

$$\frac{\partial U(x,t)}{\partial x} = \sum_{j=1}^{N} u_j(t)\frac{d\phi_j(x)}{dx},$$

$$\frac{\partial U(x,t)}{\partial t} = \sum_{j=1}^{N} \frac{du_j(t)}{dt}\phi_j(x).$$

Substituting these in Equation (2.7) yields:

$$\sum_{j=1}^{N}\left(\int_{\Omega}\phi_i(x)\phi_j(x)dx\right)\frac{du_j(t)}{dt}+\sum_{j=1}^{N}\left(\int_{\Omega}\frac{d\phi_i(x)}{dx}\boldsymbol{\alpha}(x)\phi_j(x)dx\right)u_j(t)$$
$$+\sum_{j=1}^{N}\left(\int_{\Omega}\frac{d\phi_i(x)}{dx}\beta\frac{d\phi_j(x)}{dx}dx\right)u_j(t)=0, \qquad \text{for } i=1,\ldots,N. \tag{2.8}$$

Using these basis functions, for each element (e) (notation (i) corresponds to parameters associated with the ith element), analytical derivation of the integrals involved in (2.8) is possible (Lynch 2005):

$$\int^{(e)}\phi_i(x)\phi_j(x)dx=\begin{cases}\frac{\Delta x}{3}, & i=j\\ \\ \frac{\Delta x}{6}, & i\neq j\end{cases}$$

$$\int^{(e)}\frac{d\phi_i(x)}{dx}\frac{d\phi_j(x)}{dx}dx=\frac{d\phi_i(x)}{dx}\frac{d\phi_j(x)}{dx}\int^{(e)}1dx=\begin{cases}\frac{-1}{\Delta x}, & i=j\\ \\ \frac{1}{\Delta x}, & i\neq j\end{cases}$$

$$\int^{(e)}\frac{d\phi_i(x)}{dx}\phi_j(x)dx=\frac{d\phi_i(x)}{dx}\int^{(e)}\phi_j(x)dx=\begin{cases}\frac{-1}{2}, & i=j\\ \\ \frac{1}{2}, & i\neq j\end{cases}$$

Note that $\int^{(e)}\phi_j(x)dx=\frac{\Delta x}{2}$ and $\int^{(e)}1dx=\Delta x$. Substituting these values in the element level equation (2.8), and rewriting the equation for the complete domain of the problem (Ω), the system of equations can be written as:

$$\mathbf{G}\frac{d\mathbf{u}_t}{dt}=\mathbf{H}\mathbf{u}_{t-1}, \tag{2.9}$$

where $\mathbf{H}$ is a sparse matrix of $\boldsymbol{\alpha}(x)$, β, and dx, and $\mathbf{G}$ is a sparse matrix of Δx:

$$\mathbf{G}=\begin{pmatrix}g_{0,1} & g_{2,2} & 0 & 0 & \ldots & 0\\ g_{1,1} & g_{0,2} & g_{2,3} & 0 & \ddots & 0\\ 0 & g_{1,2} & \ddots & \ddots & \ddots & \vdots\\ \vdots & \ddots & \ddots & \ddots & g_{0,N-1} & g_{2,N}\\ 0 & \ldots & \ldots & 0 & g_{1,N-1} & g_{0,N}\end{pmatrix},$$

with:

$$
\begin{aligned}
g_{0,1} &= \frac{\Delta x^{(1)}}{3\Delta t}, \\
g_{0,N} &= \frac{\Delta x^{(N)}}{3\Delta t}, \\
g_{0,i} &= \frac{2\Delta x^{(i)}}{3\Delta t}, && \text{for } i = 2, \ldots, N-1, \\
g_{1,i} &= \frac{\Delta x^{(i)}}{6\Delta t}, && \text{for } i = 1, \ldots, N, \\
g_{2,i} &= \frac{\Delta x^{(i)}}{6\Delta t}, && \text{for } i = 1, \ldots, N,
\end{aligned}
$$

and,

$$
\mathbf{H} = \begin{pmatrix}
h_{0,1} & h_{2,2} & 0 & 0 & \ldots & 0 \\
h_{1,1} & h_{0,2} & h_{2,3} & 0 & \ddots & 0 \\
0 & h_{1,2} & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & h_{0,N-1} & h_{2,N} \\
0 & \ldots & \ldots & 0 & h_{1,N-1} & h_{0,N}
\end{pmatrix},
$$

with:

$$
\begin{aligned}
h_{0,1} &= \frac{-\beta^{(1)}}{(\Delta x^{(1)})^2} + \frac{\alpha^{(1)}}{2}, \\
h_{0,N} &= \frac{-\beta^{(N)}}{(\Delta x^{(N)})^2} + \frac{\alpha^{(N)}}{2}, \\
h_{0,i} &= \frac{-2\beta^{(i)}}{(\Delta x^{(i)})^2}, && \text{for } i = 2, \ldots, N-1, \\
h_{1,i} &= \frac{-2\beta^{(i)}}{(\Delta x^{(i)})^2} - \frac{\alpha^{(i)}}{2}, && \text{for } i = 1, \ldots, N, \\
h_{2,i} &= \frac{-2\beta^{(i)}}{(\Delta x^{(i)})^2} + \frac{\alpha^{(i)}}{2}, && \text{for } i = 1, \ldots, N.
\end{aligned}
$$

2.2.3 Discretization in Time

Here, a forward differencing method for discretizing the equation in time is utilized:

$$
\frac{d\mathbf{u}_t}{dt} \approx \frac{\mathbf{u}_t - \mathbf{u}_{t-1}}{\Delta t},
$$

where, each of the terms in the equation is considered at the forward end of the time step (e.g., Burnett 1987; Carey and Oden 1984). Forward differencing time discretization yields

$$\mathbf{G}\left(\frac{\mathbf{u}_t - \mathbf{u}_{t-1}}{\Delta t}\right) = \mathbf{H}\mathbf{u}_{t-1}. \tag{2.10}$$

Thus, the dynamical model can be written as

$$\mathbf{G}\mathbf{u}_t = (\mathbf{G} + \Delta t\mathbf{H})\mathbf{u}_{t-1} \tag{2.11}$$

or,

$$\mathbf{u}_t = (\mathbf{I} + \Delta t\mathbf{G}^{-1}\mathbf{H})\mathbf{u}_{t-1}. \tag{2.12}$$

Equation (2.12) is the desired recurrence relation.

2.2.4 Example 2: Advection-Diffusion Equation (2-D)

Consider the two-dimensional advection-diffusion equation:

$$\frac{\partial U_t(x,y)}{\partial t} + \frac{\partial \alpha_x(x,y)U_t(x,y)}{\partial x} + \frac{\partial \alpha_y(x,y)U_t(x,y)}{\partial y} - \beta\frac{\partial^2 U_t(x,y)}{\partial x^2} - \beta\frac{\partial^2 U_t(x,y)}{\partial y^2} = 0 \tag{2.13}$$

Similar to the 1-D case, $u_t(x,y)$ can be approximated using a series of linearly independent basis functions, $\phi_i(x,y)$:

$$U_t(x,y) \approx \sum_{j=1}^{N} u_{t,j}\phi_j(x,y),$$

where,

$$\phi_j(x,y) = \frac{1}{2A}(a_j + b_j x + c_j y), j = 1,2,3,$$

are triangular basis functions (Figure 2) and

$$a_j = x_k y_l - x_l y_k,$$

$$b_j = y_k - y_l,$$

$$c_j = x_l - x_k,$$

with the derivatives with respect to x and y are defined as

$$\frac{d\phi_j(x,y)}{dx} = \frac{(y_k - y_l)}{2A},$$
$$\frac{d\phi_j(x,y)}{dy} = \frac{(x_l - x_k)}{2A},$$

where j, k, and l take values 1, 2, and 3 (the nodes of the triangular element) permuted cyclically, and A represents the area of the triangular element:

$$A = \frac{1}{2} \begin{vmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ 1 & x_3 & y_3 \end{vmatrix}.$$

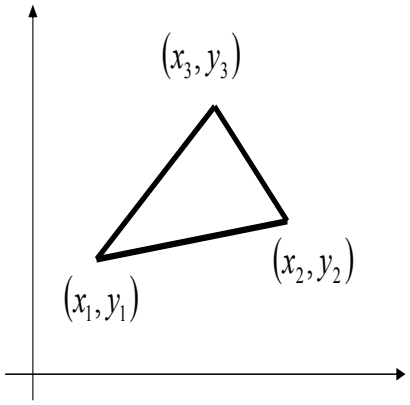


Figure 2: Schematic for the triangular basis function.

The Galerkin procedure requires

$$\iint_{\Omega}[\frac{\partial U_t(x,y)}{\partial t}+\frac{\partial \alpha_x(x,y)U_t(x,y)}{\partial x}+\frac{\partial \alpha_y(x,y)U_t(x,y)}{\partial y}-\beta\frac{\partial^2 U_t(x,y)}{\partial x^2} - \beta\frac{\partial^2 U_t(x,y)}{\partial y^2}]\phi_i(x,y)dxdy=0, \qquad \text{for } i=1,\ldots,N \tag{2.14}$$

Note that equation (2.14) involves a second derivative of the piecewise linear basis function. Applying integration by parts to the terms involving the second derivative yields:

$$\begin{aligned}&\iint_{\Omega}\frac{\partial U_t(x,y)}{\partial t}\phi_i(x,y)dxdy+\iint_{\Omega}\frac{\partial \alpha_x(x,y)U_t(x,y)}{\partial x}\phi_i(x,y)dxdy\\&+\iint_{\Omega}\frac{\partial \alpha_y(x,y)U_t(x,y)}{\partial y}\phi_i(x,y)dxdy-\left[-\beta\left(\frac{\partial U_t(x,y)}{\partial x}+\frac{\partial U_t(x,y)}{\partial y}\right)\phi_i(x,y)\right]_{\Omega}\\&+\iint_{\Omega}\frac{\partial \phi_i(x,y)}{\partial x}\beta\frac{\partial U_t(x,y)}{\partial x}dxdy+\iint_{\Omega}\frac{\partial \phi_i(x,y)}{\partial y}\beta\frac{\partial U_t(x,y)}{\partial y}dxdy=0,\\&\qquad \text{for } i=1,\ldots,N.\end{aligned} \tag{2.15}$$

Substituting these values into equation (2.14), the element-level equation can be written as

$$\begin{aligned}&\sum_{j=1}^{N}\left(\iint_{\Omega}\phi_j(x,y)\phi_i(x,y)dxdy\right)\frac{du_j(t)}{dt}+\sum_{j=1}^{N}\left(\iint_{\Omega}\frac{\partial \alpha_x(x,y)\phi_j(x,y)}{\partial x}\phi_i(x,y)dxdy\right)u_j(t)\\&+\sum_{j=1}^{N}\left(\iint_{\Omega}\frac{\partial \alpha_y(x,y)\phi_j(x,y)}{\partial y}\phi_i(x,y)dxdy\right)u_j(t)+\sum_{j=1}^{N}\left(\iint_{\Omega}\frac{\partial \phi_i(x,y)}{\partial x}\beta\frac{\partial \phi_j(x,y)}{\partial x}dxdy\right)u_j(t)\\&+\sum_{j=1}^{N}\left(\iint_{\Omega}\frac{\partial \phi_i(x,y)}{\partial y}\beta\frac{\partial \phi_j(x,y)}{\partial y}dxdy\right)u_j(t)=0, \qquad \text{for } i=1,\ldots,N,\end{aligned} \tag{2.16}$$

where Ω is the 2-D domain of study that includes all the elements. Note that the boundary term $[-\beta(\frac{\partial U_t(x,y)}{\partial x}+\frac{\partial U_t(x,y)}{\partial y})\phi_i(x,y)]_{\Omega}$ contains the flux, which is zero in this case.

The general form of the dynamical model for the 2-D case is similar to the one for the 1-D case described in (2.11). However, the values of the integrals involving the basis functions are different due to the use of triangular elements in the 2-D problem. Similar to

the 1-D case, analytical derivation of these integrals is possible (Lynch 2005). For example, for the triangular element e, the integrals using the triangle integration formula can be evaluated (Burnett 1987),

$$\iint \zeta_1^m \zeta_2^n \zeta_3^p dxdy = \frac{m!n!p!}{(m+n+p+2)!} 2A,$$

where ζ_1, ζ_2, ζ_3 represent the area coordinates for the triangular elements (i.e., $\zeta_i = \phi_i, i = 1, 2, 3$) to the arbitrary powers m, n, p, respectively. The obtained integrals are as follows.

$$\iint^{(e)} \phi_i(x,y)\phi_j(x,y)dxdy = \begin{cases} \frac{A^{(e)}}{6}, & i = j \\ \\ \frac{A^{(e)}}{12}, & i \neq j, \end{cases}$$

$$\iint^{(e)} \frac{d\phi_i(x,y)}{dx}\frac{d\phi_j(x,y)}{dx}dxdy = \iint^{(e)} \frac{(y_k^{(e)} - y_l^{(e)})}{2A^{(e)}}\frac{(y_k^{(e)} - y_l^{(e)})}{2A^{(e)}}dxdy$$
$$= \begin{cases} \frac{(y_k^{(e)} - y_l^{(e)})^2}{4A^{(e)}}, & i = j \\ \\ \frac{(y_k^{(e)} - y_l^{(e)})(y_j^{(e)} - y_k^{(e)})}{4A^{(e)}}, & i = l \\ \\ \frac{(y_k^{(e)} - y_l^{(e)})(y_l^{(e)} - y_j^{(e)})}{4A^{(e)}}, & i = k, \end{cases}$$

$$\iint^{(e)} \frac{d\phi_i(x,y)}{dy}\frac{d\phi_j(x,y)}{dy}dxdy = \iint^{(e)} \frac{(x_k^{(e)} - x_l^{(e)})}{2A^{(e)}}\frac{(x_k^{(e)} - x_l^{(e)})}{2A^{(e)}}dxdy$$
$$= \begin{cases} \frac{(x_k^{(e)} - x_l^{(e)})^2}{4A^{(e)}}, & i = j \\ \\ \frac{(x_k^{(e)} - x_l^{(e)})(x_j^{(e)} - x_k^{(e)})}{4A^{(e)}}, & i = l \\ \\ \frac{(x_k^{(e)} - x_l^{(e)})(x_l^{(e)} - x_j^{(e)})}{4A^{(e)}}, & i = k, \end{cases}$$

$$\iint^{(e)} \frac{d\phi_i(x,y)}{dx}\phi_j(x,y)dxdy = \iint^{(e)} \frac{(y_k^{(e)} - y_l^{(e)})}{2}\phi_j(x,y)dxdy$$
$$= \begin{cases} \frac{(y_k^{(e)} - y_l^{(e)})}{6}, & i = j \\ \frac{(y_j^{(e)} - y_k^{(e)})}{6}, & i = l \\ \frac{(y_l^{(e)} - y_j^{(e)})}{6}, & i = k, \end{cases}$$

$$\iint^{(e)} \frac{d\phi_i(x,y)}{dy}\phi_j(x,y)dxdy = \iint^{(e)} \frac{-(x_k^{(e)} - x_l^{(e)})}{2}\phi_j(x,y)dxdy$$
$$= \begin{cases} \frac{-(x_k^{(e)} - x_l^{(e)})}{6}, & i = j \\ \frac{-(x_k^{(e)} - x_l^{(e)})}{6}, & i = l \\ \frac{-(x_l^{(e)} - x_j^{(e)})}{6}, & i = k. \end{cases}$$

Note that $\iint^{(e)} \phi_j(x,y)dxdy = \frac{A^{(e)}}{3}$ and $\iint^{(e)} 1dxdxy = A^{(e)}$. Substituting the values for the integrals in equation (2.16) and rewriting the equation for the complete domain of the problem (Ω), the general form of the dynamical equation can be written as

$$\mathbf{G}\frac{d\mathbf{u}_t}{dt} = \mathbf{H}\mathbf{u}_{t-1}, \tag{2.17}$$

where $\mathbf{H}$ is a sparse matrix of $\alpha(x)$, β, and $\mathbf{A}$, and $\mathbf{G}$ is a sparse matrix of $\mathbf{A}$, and $\mathbf{A} = [A^{(e)}]$ includes the areas for the triangular elements. However, writing a general structure for matrices $\mathbf{H}$ and $\mathbf{G}$ is not possible in the 2-D case due to various possible arrangements of the triangular elements. Note that the sparsity of these matrices is conserved under such circumstances.

2.3 Hierarchical Bayesian Modeling Approach

The hierarchical spatio-temporal model for the PDE-based process described in equations (2.11) and (2.17) can be written as

$$\mathbf{Z}_t = \mathbf{K}_t \mathbf{u}_t + \boldsymbol{\varepsilon}_t, \qquad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \sigma_\varepsilon^2 \mathbf{I}), \tag{2.18}$$

$$\mathbf{G}\mathbf{u}_t = (\mathbf{G} + \triangle t \mathbf{H})\mathbf{u}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \sigma_\eta^2 \mathbf{R}(\theta)), \tag{2.19}$$

where $\mathbf{Z}_t = (z_{1t}, z_{2t}, \ldots, z_{mt})'$ denotes observational data, $\mathbf{u}_t = (u_{1t}, u_{2t}, \ldots, u_{nt})'$ denote the true process at all locations of interest at time t, $\mathbf{K}_t$ denote the m by n "incidence" or "mapping" matrices that for each time t map the process locations to the observation locations, $\boldsymbol{\varepsilon}_t$ is the observational error, $\boldsymbol{\eta}_t$ represents the error for the process, and $\mathbf{H}$ and $\mathbf{G}$ are sparse matrices as described in the previous section. Matrix $\mathbf{R}(\theta)$ represents the spatial correlation based on the Euclidean distance between locations ($||d||$) and the spatial range parameter (θ) in the standard stationary and isotropic exponential covariogram model:

$$\mathbf{R}(\theta) = \exp(-\theta||d||). \tag{2.20}$$

Note that this is just an example and more complicated spatial models could be considered.

The hierarchical Bayesian model considers prior distributions for the random and unknown parameters, along with the data, to update the posterior distribution of the parameters. For example, for the 1-D case advection-diffusion process, both the advection and diffusion parameters are considered as random and unknown parameters and take the following normal prior distributions:

$$\boldsymbol{\beta} \sim N(\tilde{\beta}, \sigma_\beta^2),$$

$$\boldsymbol{\alpha} \sim N(\tilde{\boldsymbol{\alpha}}, \sigma_\alpha^2 \mathbf{R}_\alpha(\theta_\alpha)),$$

where $\tilde{\boldsymbol{\alpha}}$, σ_α^2, $\tilde{\beta}$, σ_β^2 are the hyper-parameters which could be considered as random and having prior distributions or might be assumed as fixed and known, and $\mathbf{R}_\alpha$ is the exponential covariogram model defined in 2.20. Prior distributions are also needed for the variance of the error process, σ_η^2 and the variance of the measurement error, σ_ε^2, as follows:

$$\sigma_\varepsilon^2 \sim IG(q_\varepsilon, r_\varepsilon)$$

$$\sigma_\eta^2 \sim IG(q_\eta, r_\eta),$$

where IG represents the pdf of an inverse gamma distribution. The hyper-parameters, q_ε, r_ε, q_η, and r_η can be considered as fixed and known or random and having a prior distribution. Thus, the joint posterior distribution of interest can be written as

$$\begin{aligned}[\mathbf{u}_1, \ldots, \mathbf{u}_T, \boldsymbol{\beta}, \sigma_\eta^2, \sigma_\varepsilon^2 | \mathbf{Z}_1, \ldots, \mathbf{Z}_T] \propto & \prod_{t=1}^{T} [\mathbf{Z}_t | \mathbf{u}_t, \sigma_\varepsilon^2] \prod_{t=1}^{T} [\mathbf{u}_t | \mathbf{u}_{t-1}, \sigma_\eta^2, \boldsymbol{\beta}] \\ & \times [\mathbf{u}_0][\boldsymbol{\beta}][\sigma_\varepsilon^2][\sigma_\eta^2].\end{aligned}$$

Using the Gibbs sampler, the estimation can be conducted based on conditional distributions of each of the unknown parameters given all other parameters and data (i.e., full-conditional distributions). See Appendix A for derivation of all full-conditional distributions.

2.4 Simulations

In this section, simulation experiments for the 1-D and 2-D advection-diffusion process are considered and discussed. The results for the simulation experiments are presented in the next section.

2.4.1 Advection-Diffusion (1-D)

In the following simulation, $\Delta x = \Delta t = 1$ for 50 spatial locations and 101 time points. Note that the FEM allows for irregular domains in which Δx can vary over the domain. One has to be careful in extending the model to such cases so that the stability conditions of the PDE are not violated (Lynch 2005).

For simulations, the process is initialized using randomly generated normal variates. The spatially-varying advection parameter, $\boldsymbol{\alpha}(x)$, is generated by drawing from a normal density with mean $\boldsymbol{\xi}$ and variance $0.005 \times \mathbf{R}(0.005)$, where $\boldsymbol{\xi} = \frac{1}{3}\sin(\mathbf{x}), \quad x_i = 2\pi - (i-1)\frac{2\pi}{n}, \quad \text{for } i = 1, \ldots, n$. The diffusion parameter, β, is set to be 0.1. The following relatively vague prior distributions were assigned to the error variance parameters:

$$\tau_\varepsilon \equiv 1/\sigma_\varepsilon^2 \sim Gamma(\text{mean} = 0.1, \text{var} = 10)$$

$$\tau_\eta \equiv 1/\sigma_\eta^2 \sim Gamma(\text{mean} = 0.1, \text{var} = 10),$$

and the following relatively vague prior distributions were considered for the advection and diffusion parameters:

$$\beta \sim N(1, 1000),$$

$$\boldsymbol{\alpha} \sim N(\mathbf{0}, 0.1 \times \mathbf{R}_\alpha(0.005)),$$

where $\mathbf{R}_\alpha$ is the exponential covariagram model defined in (2.20).

Computations were conducted using MCMC methods (Gibbs sampler) with 50000 iterations to ensure convergence of the algorithm. The first 5000 iterations were considered as the "burn-in" period, although the convergence was achieved during the first 1000 iter-

ations, and posterior inference was conducted using the last 45000 iterations. Convergence was assessed visually using the MCMC chains for the parameters.

2.4.2 Advection-Diffusion (2-D)

A simulation experiment based on a 2-D advection-diffusion model is considered and a hierarchical Bayesian approach is presented to estimate the states of the spatio-temporal process as well as the parameters governing the dynamics (treated as unknown parameters), in the presence of missing observations. The 2-D advection-diffusion equation is considered with both advection and diffusion parameters in x and y directions, as described in (2.13). Such a parameterization enables angular outward directional movement of the flow in irregularly defined boundaries of the model domain (Burnett 1987).

The simulation experiment is considered for a 9 by 10 rectangular shaped domain, resulting in 90 nodes and a regular mesh constructed with triangular elements with equal areas (3). Other characteristics of the model include: time step $\triangle t = 2$, and regularly spaced nodes with $\triangle x = 1$, $\triangle y = 1$. The process was initialized with a surface generated from $f(x) = 10/(1 + (\frac{x}{.5})^2 + (\frac{y}{.5})^2)$ centered at a point with coordinates (5, 8) . The diffusion and advection parameters (constant over the spatial domain) are set as: advection in x-direction $(\alpha_x) = -0.05$, advection in y-direction $(\alpha_y) = 0.01$), diffusion in x-direction $(\beta_x) = 0.01$, and diffusion in y-direction $(\beta_y) = 0.01$. The values assigned to the parameters governing the dynamics results in a movement along the negative x-axis with slight upward movement along the positive y-axis, while diffusing equally in both directions. The simulation was evolved for 50 time steps.

The following relatively vague prior distributions were assigned to the error variance

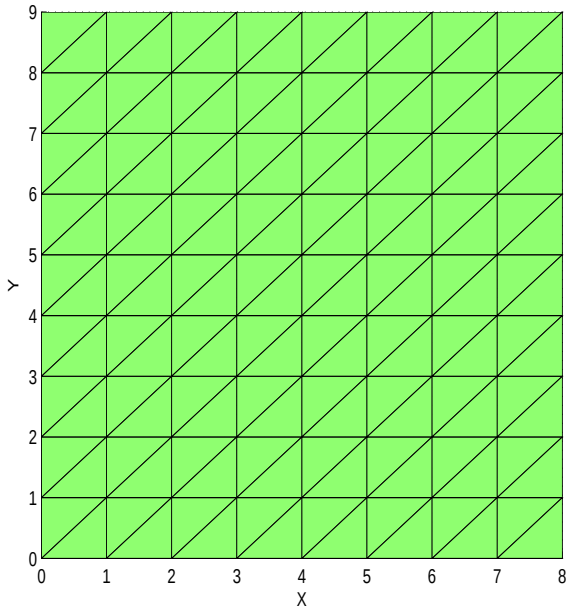


Figure 3: Regular mesh for 2-D advection-diffusion equation constructed using triangular elements.

parameters:

$$\tau_\varepsilon \equiv 1/\sigma_\varepsilon^2 \sim Gamma(\text{mean} = 1, \text{var} = 10)$$

$$\tau_\eta \equiv 1/\sigma_\eta^2 \sim Gamma(\text{mean} = 1, \text{var} = 10),$$

and the following prior distributions were considered for the advection and diffusion parameters:

$$\beta_x \sim N(0.05, 100),$$

$$\beta_y \sim N(0.05, 100),$$

$$\alpha_x \sim N(0, 100),$$

$$\alpha_y \sim N(0, 100).$$

The observations used in the hierarchical Bayesian model were obtained after randomly

removing 50% of the points in each simulated time. The hierarchical model is similar to the model described for the 1-D case simulation (after vectorizing the 2-D surface). Here, due to the presence of missing values, K_t incidence matrices are used to map the observations to the process, providing a tool to account for missing values in the estimation and prediction process.

Similar to the 1-D case, computations were conducted using a Gibbs sampler with 50000 iterations to guarantee convergence of the algorithm. The first 5000 iterations were considered as the "burn-in" period and posterior inference was conducted using the last 45000 iterations.

2.5 Simulation Results

The results for the hierarchical Bayesian model for the simulations described in the previous section are presented and discussed in this section. The inference is based on samples from the posterior density of the unknown parameters. Also a prior sensitivity analysis was conducted in each case and the results were found to be robust to the choice of hyper-parameters for prior distributions.

The results for the 1-D advection-diffusion simulation based on a constant diffusion parameter and spatially-varying advection parameter are encouraging. Plots for the simulated "truth", data, the estimated posterior mean for the experiment, and process regenerated using the posterior mean results for the advection and diffusion parameters (using the error process and initial values for the process as used in the original simulation) are shown in Figure 4. Figure 5 shows the histogram of the posterior density for the diffusion parameter β. The posterior mean of β is 0.1067 and the 95% credible intervals (CIs) is (0.0636,

0.1503), which considering that a vague prior is used for this parameter, is an acceptable estimate (the 95% CIs cover the true value,0.1). Figure 6 shows the posterior mean and 95% credible intervals (CIs) for the posterior density of $\boldsymbol{\alpha}(x)$, as well as the true values of the spatially-varying advection parameter and the lower and upper bounds for the prior density considered in estimation of this vector of parameters. The results for the estimation of $\boldsymbol{\alpha}(x)$ considering the relatively vague prior density considered for this parameter, are very encouraging and indicate that Bayesian learning is occurring regarding the advection parameters.

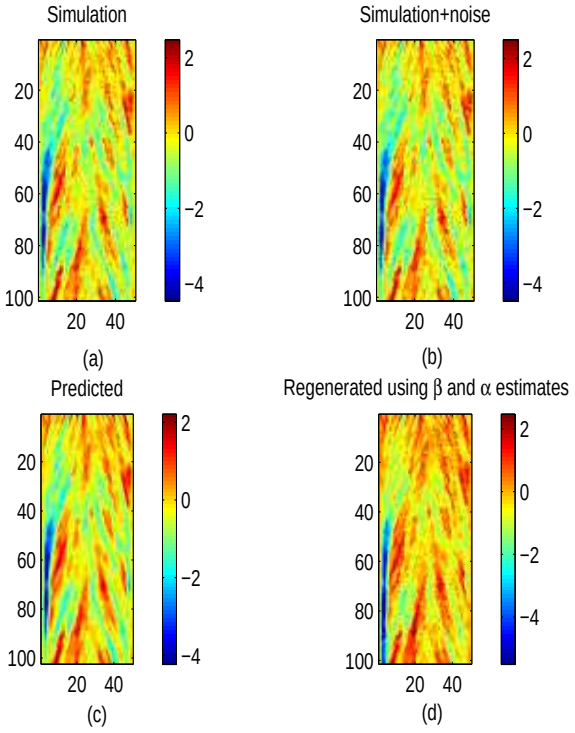


Figure 4: Simulation of 1-D linear advection-diffusion equation: plots for (a) simulated process ("truth"), (b) data (simulation with additive noise), (c) the posterior mean estimated using the finite element-based process model, and (d) regenerated process (regenerated using the posterior mean estimates of advection and diffusion parameters.).

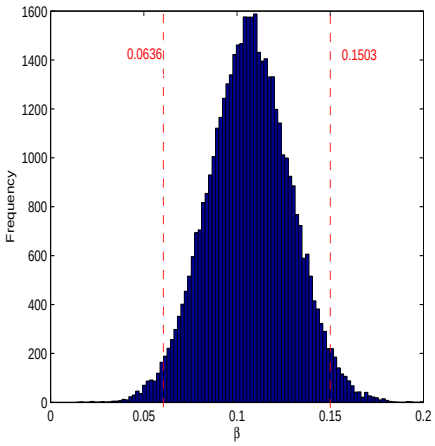


Figure 5: Histogram for the posterior density for the diffusion parameter (β) for the 1-D simulation. The dashed lines indicate the 95% CIs (0.0636, 0.1503).

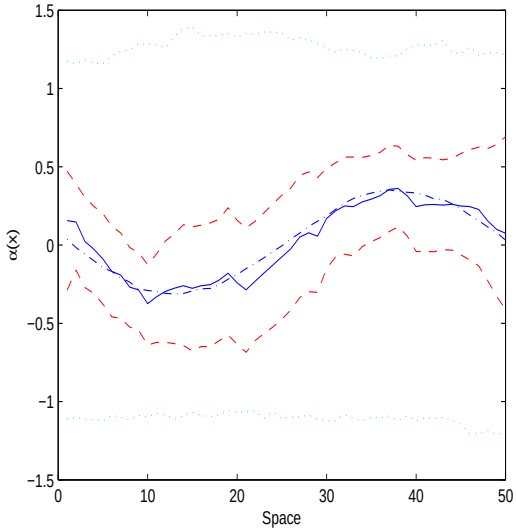


Figure 6: Plot for the spatially-varying advection parameter (1-D simulation), $\boldsymbol{\alpha}(x)$: posterior mean (solid line), 95% credible intervals for the posterior (dashed lines), true spatially-varying advection parameter (dash-dotted line), lower and upper bounds for the prior distribution (dotted lines)

The results for the 2-D advection-diffusion equation with unknown constant advection and diffusion parameters in both x and y directions and 50% missing values are very encouraging. Figures 7-8 show the true process, posterior mean of the estimated process, as well as, observations (containing missing values). The prediction of the missing values using the posterior mean of the process states is notable, considering only 50% of the observations were used in the prediction process.

The posterior means and 95% credible intervals (CIs) for all the parameters are shown in Table 1. Note that the posterior results for all the parameters are acceptable (within the 95% CIs) and the posterior means are very close to the true values.

Histograms for the posterior density of the advection and diffusion parameters are shown in Figures 9-12. Although the posterior mean values for the y-directional parameters of advection and diffusion over-estimate the true parameter values, the 95% CIs (as shown in Figures 10 and 12) contain the true values.

The results for the simulation studies encourage the application of this modeling approach to "real-world" problems. The characteristics of the finite element method provides a natural framework for considering problems with irregular domains, stochastic boundary conditions, and processes described by coupled PDEs (e.g., shallow-water equations). The ability to consider heterogeneously distributed parameters that govern the dynamics of the spatio-temporal process enables us to more effectively model complex processes with realistic assumptions.

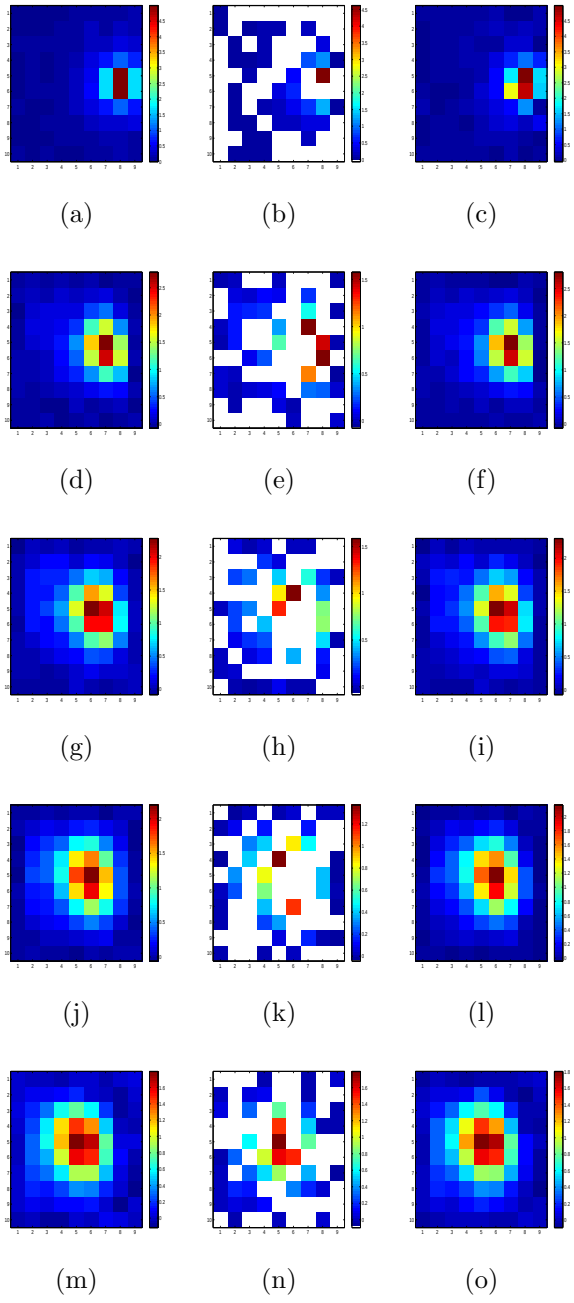


Figure 7: Truth (left column), data (middle column), and posterior mean values (right column) for the advection-diffusion simulation (2-D) for time steps 1-25; (a)-(c) t=1, (d)-(f) t=10, (g)-(i) t=15, (j)-(l) t=20, (m)-(o) t=25. Note that in the representation of the observations (figures in the second column), white colored cells indicate missing values.

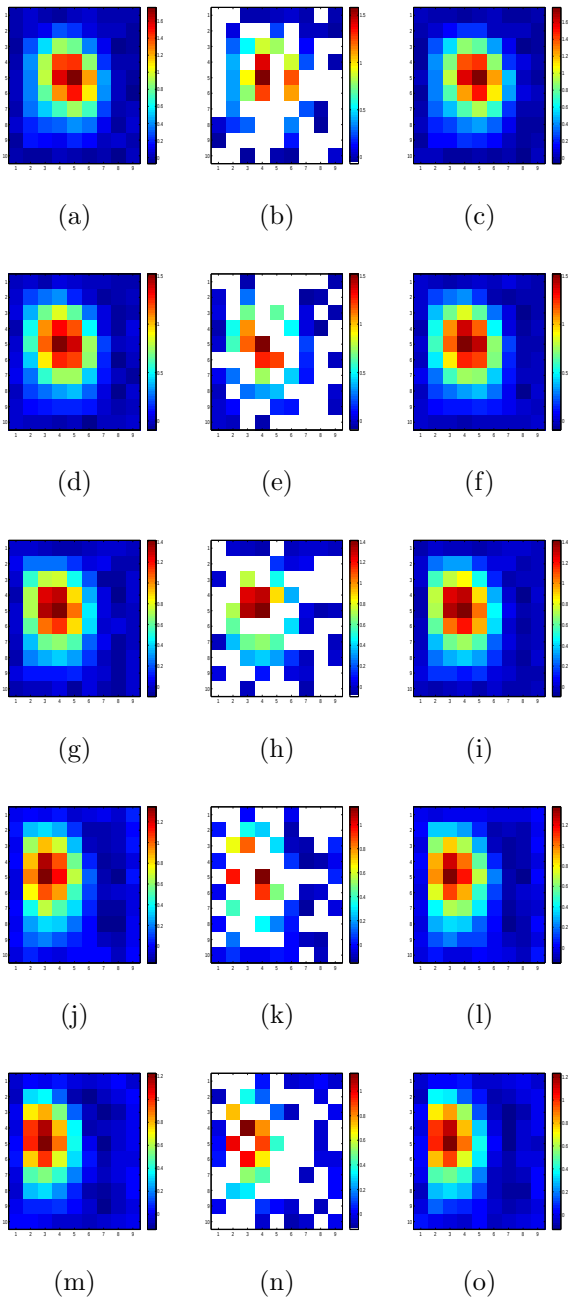


Figure 8: Truth (left column), data (middle column), and posterior mean values (right column) for the advection-diffusion simulation (2-D) for time steps 30-50; (a)-(c) t=30, (d)-(f) t=35, (g)-(i) t=40, (j)-(l) t=45, (m)-(o) t=50. Note that in the representation of the observations (figures in the second column), white colored cells indicate missing values.

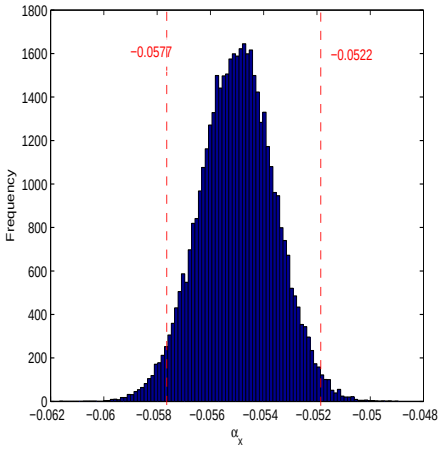


Figure 9: Histogram for the posterior density for the advection parameter in x-direction (α_x) for the 2-D simulation. The dashed lines indicate the 95% CIs (-0.0577, -0.0522).

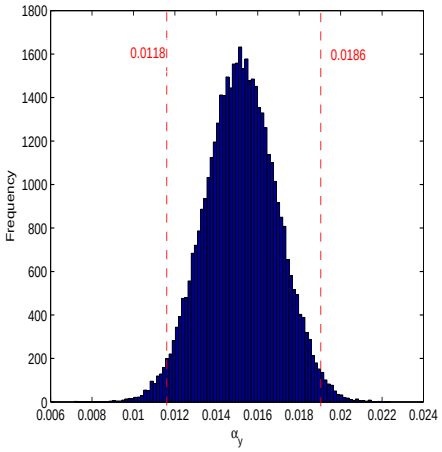


Figure 10: Histogram for the posterior density for the advection parameter in y-direction (α_y) for the 2-D simulation. The dashed lines indicate the 95% CIs (0.0118, 0.0186).

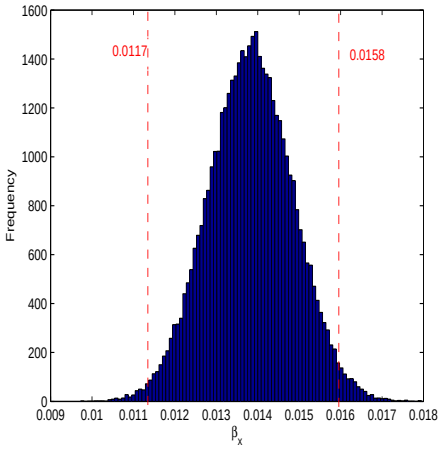


Figure 11: Histogram for the posterior density for the diffusion parameter in x-direction (β_x) for the 2-D simulation. The dashed lines indicate the 95% CIs (0.0117, 0.0158).

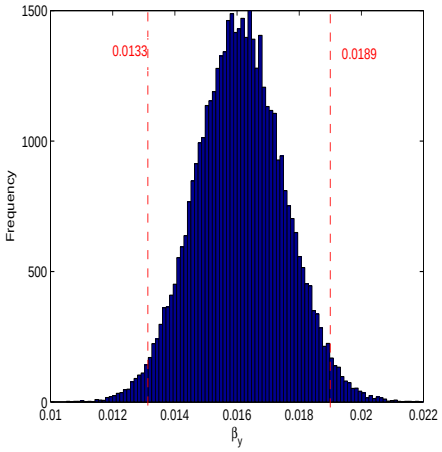


Figure 12: Histogram for the posterior density for the diffusion parameter in y-direction (β_y) for the 2-D simulation. The dashed lines indicate the 95% CIs (0.0133, 0.0189).

2.6 Application: Estimating Migration rates for Shovelnose Sturgeon in the Missouri River

In the period between 2004 and 2006, a study was conducted by the U.S. Geological Survey (USGS) to understand the attributes related to spawning of the shovelnose sturgeon (*Scaphirhynchus platorynchus*) in the Missouri River. This study was conducted with the goal of obtaining information to better understand the temporal and spatial attributes and physical conditions of successful spawning for the sturgeon species in the Lower Missouri River (Figure 13). The Lower Missouri River is defined as the 840 miles of river from Yankton, South Dakota to St. Louis, Missouri. It starts downstream from the lower-most mainstem dam (i.e., Gavins Point Dam) and ends at the confluence with the Mississippi River (USGS website; URL: http://infolink.cr.usgs.gov/RSB/Hab/).

The shovelnose sturgeon is of extreme importance in the fish assemblage of the Missouri River, especially since it is closely related to the endangered pallid sturgeon (*Scaphirhynchus albus*). The shovelnose sturgeon is often considered as a surrogate of the pallid sturgeon (Ruelle and Keenlyne 1994; Bramblett and White 2001). The management efforts conducted to benefit the sturgeon species in the Missouri River suffer from a lack of specific information on the biology and ecology of declining *Scaphirhynchus* sturgeon.

In the study conducted by the USGS, physiological measurements associated with readiness to spawn as well as behavioral data from telemetry and data storage devices (DST) were collected. The information obtained includes time and spatial location (based on river miles) of relocation for each tagged fish. Telemetry and DST sensor devices were placed in 50 gravid female shovelnose sturgeon in the Lower Missouri River that were expected

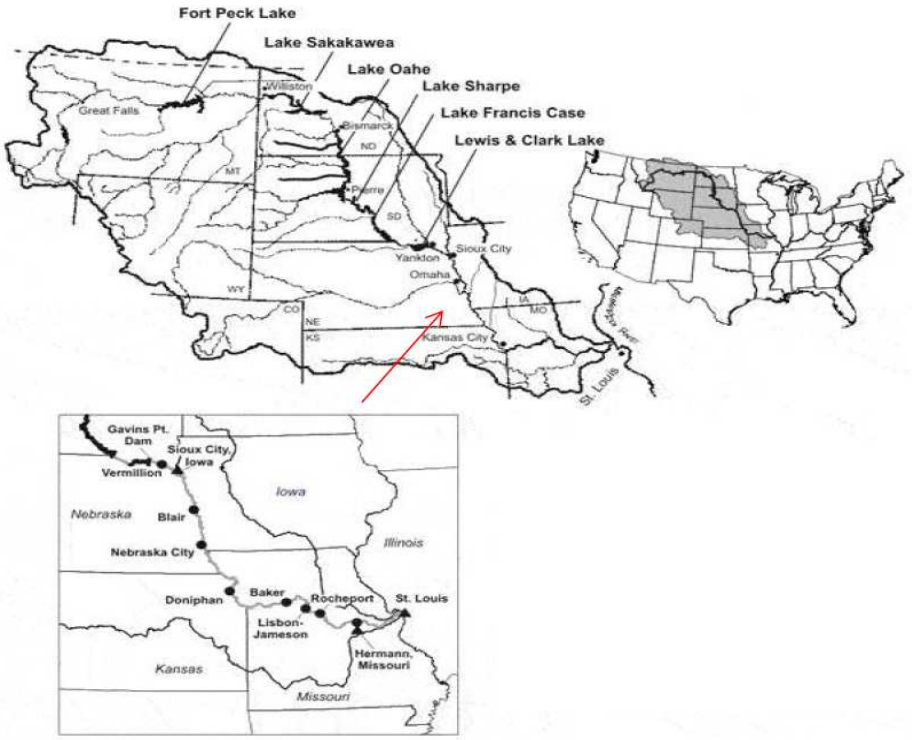


Figure 13: Map of the Lower Missouri River (*Source*: USGS website; URL: http:// infolink.cr.usgs.gov/ RSB/Hab/).

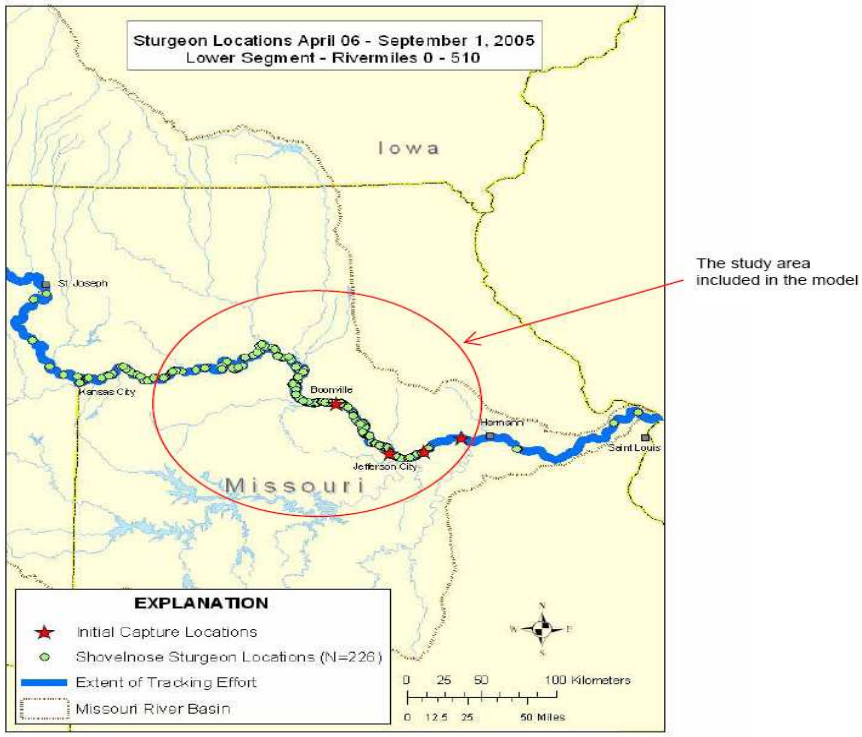


Figure 14: The section of the Lower Missouri River included in the model for 2005 shovelnose sturgeon (*Source*: USGS website; URL: http://infolink.cr.usgs.gov/ Science/PallidSturgeon/CERC/CERCUpdate.pdf).

to spawn within months. The data collected using the DST sensors was retrieved after recapture of the fish.

One of the main hypotheses considered in the design of this study was that the shovelnose sturgeon spawning happens in a period during or after an upstream movement of the fish (Berg 1981). The downstream dispersal follows spawning. The physical process underlying the migration of the shovelnose sturgeon, as common in modeling animal migrations, is considered to be based on a 1-D advection-diffusion PDE describing the spatial distribution of a population through time (Okubo and Levin 2002). The utilization of the advection-diffusion equation to characterize the dispersal and migration movements of fish populations is popular among fisheries biologists (Zabel and Anderson 1997; Skalski and Gilliam 2000; Sparrevohn et al. 2002; Zabel 2002). A hierarchical Bayesian model for estimating the migration (advection) and diffusion rates, is developed and the performance of the model is demonstrated using data collected in 2005 in the Lower Missouri River.

The area considered for the present model covers a section of the lower Missouri River between river miles 110 and 410 (Figure 14). A spatio-temporal grid is developed based on 32-mile spatial increments and 4-day temporal increments. These increments are decided based on the maximum range a boat involved in the sampling effort could travel and the length of time in which the boats cover the length of the river. Let $N(s_i;t)$ be the total number of shovelnose sturgeons relocated in a grid cell s_i at time t, and $\mathbf{N}_t = (N(s_1;t),\ldots,N(s_n;t))$. Thus, the data model can be written as

$$\log(\mathbf{N}_t) \equiv \mathbf{Z}_t = \mathbf{K}_t\mathbf{u}_t + \boldsymbol{\varepsilon_t}, \qquad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\sigma}_\varepsilon^2\mathbf{I}),$$

where $\mathbf{N}_t$ corresponds to the number of fish relocated and is strictly positive (there are no

zero values for $\mathbf{N}_t$), $\mathbf{K}_t$ matrices are "mapping" matrices (as described previously) which provide a convenient framework to account for missing data in the discussed spatio-temporal setting. The latent spatio-temporal process can be constructed based on the discretized advection-diffusion equation as described previously. Thus, the process model can be written as

$$\mathbf{G}\mathbf{u}_t = \mathbf{M}\mathbf{u}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \sigma_\eta^2 \mathbf{R}(\theta)),$$

Due to presence of uncertainty about the underlying physical process, several different models were fitted and model comparison was performed to determine which model (or models) provides a better fit for the data. The basis of such a modeling approach has roots in the uncertainty about the true structure of the parameters governing the dynamics (i.e., homogeneous or heterogeneous over the spatial domain). Also, there is uncertainty whether the pre-spawning upstream movement of the shovelnose sturgeon is more effectively characterized by a diffusion equation or an advection-diffusion equation. Although an advection-diffusion equation is considered as an appropriate model to mimic animal movement, in this case, since the upstream movement of fish is against the flow of water, a diffusion equation might be a more appropriate choice than an advection-diffusion equation. We consider five different models with constant and/or spatially-varying parameters, as described below. For all of the following models, the hierarchical Bayesian framework described in Section 2.3 applies and only the process models and the associated parameter models (priors) differ.

2.6.1 Model 1: Constant Diffusion

In model 1, the underlying physical process is characterized by a 1-D diffusion model with a constant diffusion parameter over the spatial domain:

$$\frac{\partial u(x,t)}{\partial t} - \frac{\partial}{\partial x}\left(\beta \frac{\partial u(x,t)}{\partial x}\right) = 0. \tag{2.21}$$

The hyperparameters used in the model are shown in Table 2.

2.6.2 Model 2: Spatially-Varying Diffusion

Model 1 can be extended to include spatially-varying diffusion parameters, providing a more flexible structure which allows for heterogeneity of the diffusion process over the spatial domain. Thus, the diffusion equation for Model 2 can be written as

$$\frac{\partial u(x,t)}{\partial t} - \frac{\partial}{\partial x}\left(\beta(x) \frac{\partial u(x,t)}{\partial x}\right) = 0, \tag{2.22}$$

Table 3 shows the hyperparameters specified for Model 2.

2.6.3 Model 3: Constant Advection and Diffusion

Model 3 comprises of an advection-diffusion equation with constant advection and diffusion parameters. Although the assumption of advection and diffusion parameters being constant over the spatial domain seems unrealistic, it can provide an insight into the large scale components of dynamics. Thus, the advection-diffusion equation for Model 3 is

$$\frac{\partial U(x,t)}{\partial t} + \alpha \frac{\partial}{\partial x} U(x,t) - \frac{\partial}{\partial x}\left(\beta \frac{\partial U(x,t)}{\partial x}\right) = 0. \tag{2.23}$$

Table 4 shows the hyperparameters specified for Model 3.

2.6.4 Model 4: Spatially-Varying Advection and Constant Diffusion

Model 3 can be extended to include a spatially-varying advection parameter, yielding the following advection-diffusion equation for Model 4

$$\frac{\partial U(x,t)}{\partial t} + \alpha(x)\frac{\partial}{\partial x}U(x,t) - \frac{\partial}{\partial x}\left(\beta\frac{\partial U(x,t)}{\partial x}\right) = 0. \tag{2.24}$$

Table 6 shows the hyperparameters specified for Model 4.

2.6.5 Model 5: Spatially-Varying Advection and Diffusion

Model 5 includes the most flexible version of the advection-diffusion equation, where both the advection and diffusion parameters are allowed to vary over the spatial domain. In this case the advection-diffusion equation can be written as

$$\frac{\partial U(x,t)}{\partial t} + \alpha(x)\frac{\partial}{\partial x}U(x,t) - \frac{\partial}{\partial x}(\beta(x)\frac{\partial U(x,t)}{\partial x}) = 0. \tag{2.25}$$

Table 6 shows the hyperparameters specified for Model 5.

The computations were carried out using a Gibbs sampler with 10000 iterations, where the first 1000 iterations were discarded as the "burn-in" period for the chain to achieve convergence.

2.7 Results

The results for all five models are presented in this section. The inference for the model parameters is based on the posterior density of the parameters. Specifically, the posterior mean and 95% CIs for each of the parameters are presented. Convergence was assessed by visual inspection of the MCMC chains for the parameters and was achieved very quickly (i.e., < 1000 iterations). The results for all the models discussed above are presented in

the next section, followed by model comparison results and discussion. The results for the parameters were found to be robust to the choice of hyper-parameters for prior distributions.

Model 1 Results:

The posterior mean values along with the 95% CIs for the parameters of Model 1 are shown in Table 7.

Figure 15 shows the posterior mean for the estimated process, as well as the observations. The histogram for the posterior mean of the constant diffusion parameter is shown in Figure 16.

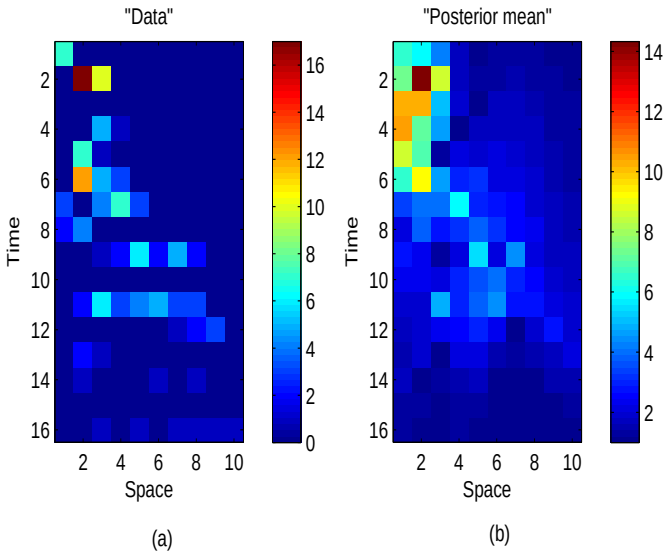


Figure 15: The shovelnose sturgeon tracking diffusion model with constant diffusion parameter (Model 1): (a) Data, (b) Posterior mean ($e^{E(\mathbf{u}|\mathbf{Z})}$).

Model 2 Results:

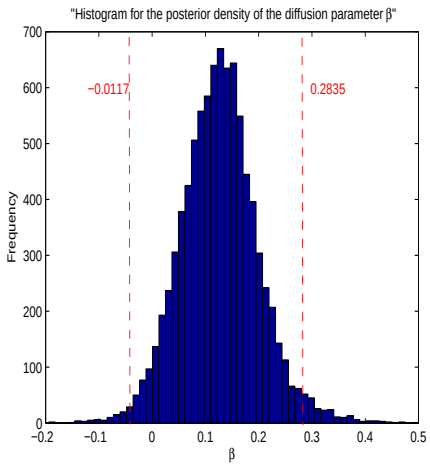


Figure 16: Histogram for the posterior density for the diffusion parameter (β) for the shovelnose sturgeon tracking diffusion model (Model 1).

The posterior mean values along with the 95% CIs for the parameters of Model 2 are shown in Table 8.

Figure 17 shows the posterior mean for the estimated process, as well as the observations and the plot of the posterior mean of the spatially-varying diffusion parameter and its 95% CIs are shown in Figure 18.

Model 3 Results:

The posterior mean values along with the 95% CIs for the parameters of Model 3 are shown in Table 9.

Figure 19 shows the posterior mean for the estimated process, as well as the observations. Figure 20 shows the histogram for the constant advection parameter. The histogram for the posterior mean of the constant diffusion parameter is shown in Figure 21.

Model 4 Results:

The posterior mean values along with the 95% CIs for the parameters of Model 4 are shown

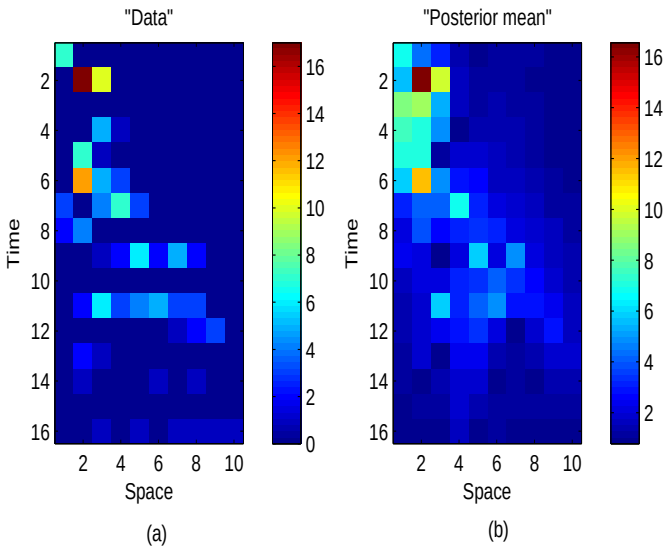


Figure 17: The shovelnose sturgeon tracking diffusion model with spatially-varying diffusion parameter (Model 2): (a) Data, (b) Posterior mean ($e^{E(\mathbf{u}|\mathbf{Z})}$).

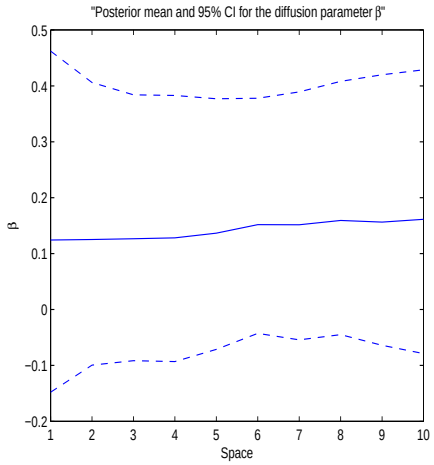


Figure 18: Plot of the posterior mean (Solid line) and 95% CIs (Dashed lines) for the spatially-varying diffusion parameter ($\boldsymbol{\beta}$) for the shovelnose sturgeon tracking diffusion model (Model 2).

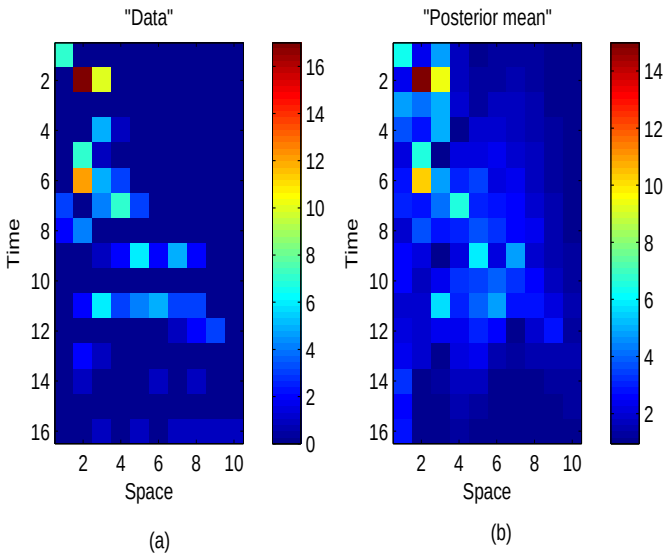


Figure 19: The shovelnose sturgeon tracking advection-diffusion model with constant advection and diffusion parameters (Model 3): (a) Data, (b) Posterior mean ($e^{E(\mathbf{u}|\mathbf{Z})}$).

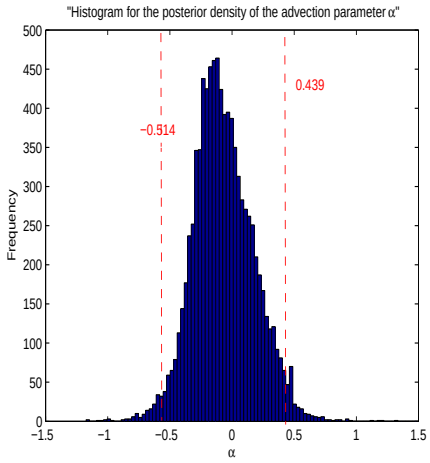


Figure 20: Histogram for the posterior density for the advection parameter (α) for the shovelnose sturgeon tracking advection-diffusion model with constant advection and diffusion parameters (Model 3).

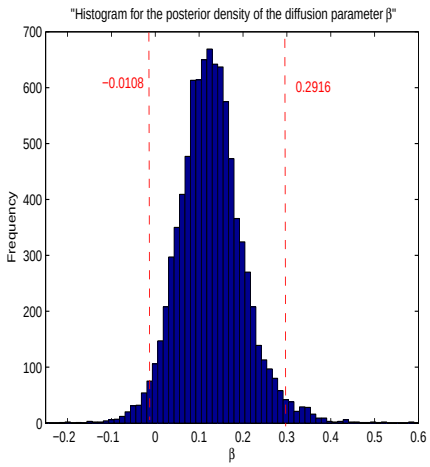


Figure 21: Histogram for the posterior density for the diffusion parameter (β) for the shovelnose sturgeon tracking advection-diffusion model with constant advection and diffusion parameters (Model 3).

in Table 10.

Figure 22 shows the posterior mean for the estimated process, as well as the observations. Figure 23 shows the posterior mean and 95% CIs for the spatially-varying advection parameters. The histogram for the posterior mean of the constant diffusion parameter is shown in Figure 24.

Model 5 Results:

The posterior mean values along with the 95% CIs for the parameters of Model 5 are shown in Table 11.

Figure 25 shows the posterior mean for the estimated process, as well as the observations. Figure 26 shows the posterior mean and 95% CIs for the spatially-varying advection parameter. The plot of the posterior mean and 95% CIs for the spatially-varying diffusion parameters are shown in Figure 27.

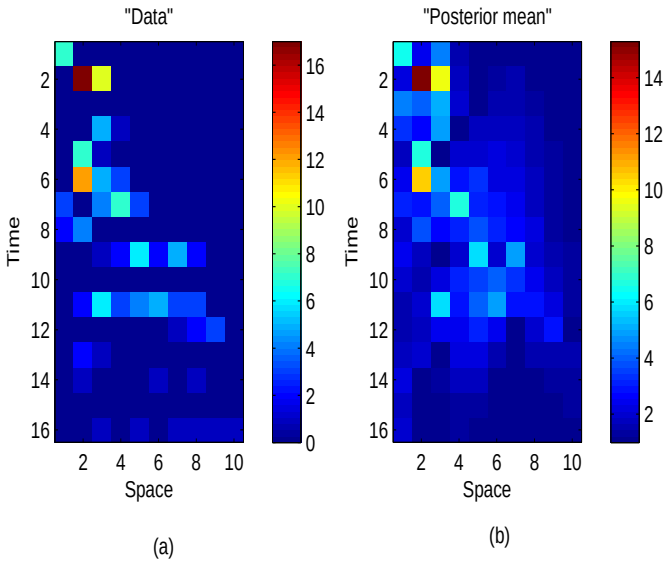


Figure 22: The shovelnose sturgeon tracking advection-diffusion model with spatially-varying advection and constant diffusion parameters (Model 4): (a) Data, (b) Posterior mean ($e^{E(\mathbf{u}|\mathbf{Z})}$).

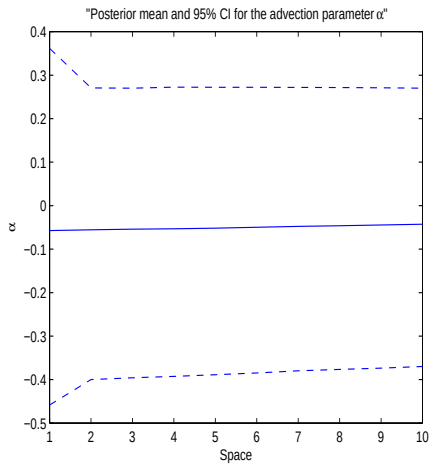


Figure 23: Plot of the posterior mean (Solid line) and 95% CIs (Dashed lines) for the spatially-varying advection parameter ($\boldsymbol{\alpha}$) for the shovelnose sturgeon tracking advection-diffusion model with spatially-varying advection and constant diffusion parameters (Model 4).

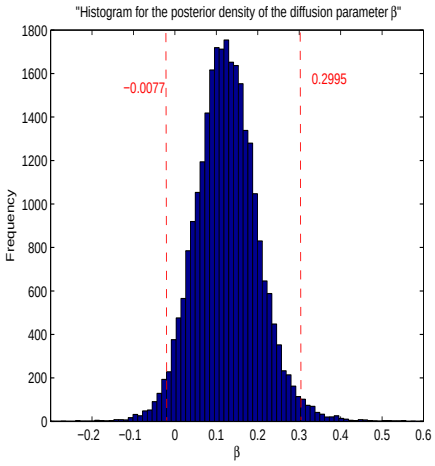


Figure 24: Histogram for the posterior density for the diffusion parameter (β) for the shovelnose sturgeon tracking advection-diffusion model with spatially-varying advection and constant diffusion parameters (Model 4).

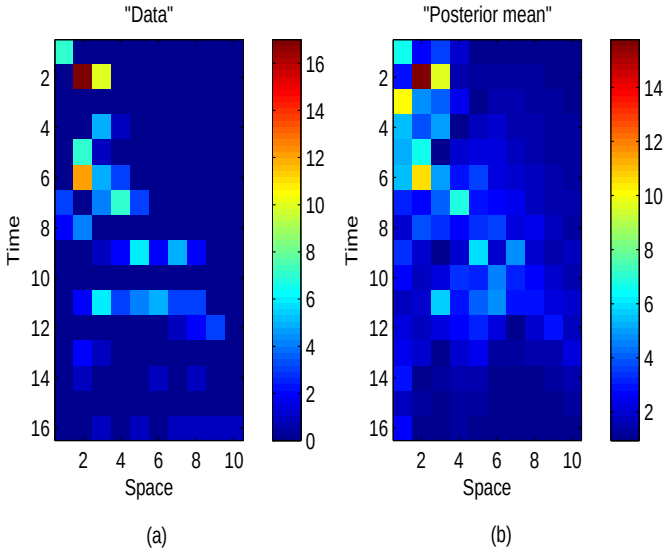


Figure 25: The shovelnose sturgeon tracking advection-diffusion model with spatially-varying advection and diffusion parameters (Model 5): (a) Data, (b) Posterior mean ($e^{E(\mathbf{u}|\mathbf{Z})}$).

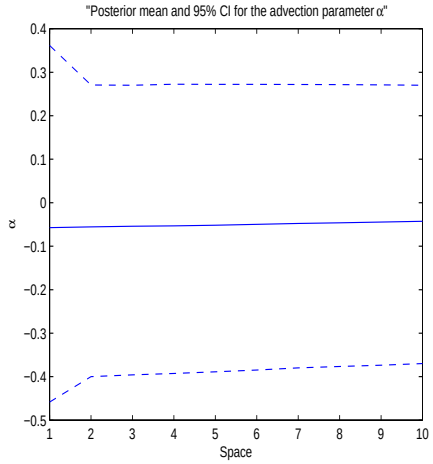


Figure 26: Plot of the posterior mean (Solid line) and 95% CIs (Dashed lines) for the spatially-varying advection parameter ($\boldsymbol{\alpha}$) for the shovelnose sturgeon tracking advection-diffusion model with spatially-varying advection and diffusion parameters (Model 5).

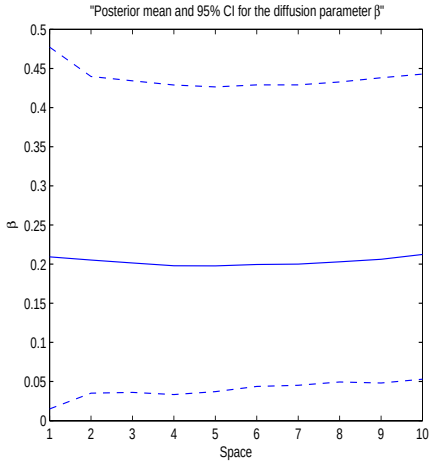


Figure 27: Plot of the posterior mean (Solid line) and 95% CIs (Dashed lines) for the spatially-varying diffusion parameter ($\boldsymbol{\beta}$) for the shovelnose sturgeon tracking advection-diffusion model with spatially-varying advection and diffusion parameters (Model 5).

2.8 Model Selection

A common and easy to implement method for model selection in hierarchical Bayesian modeling is based on the so-called deviance information criterion (DIC) introduced by Spiegelhalter et al. (2002), which is a generalization of Akaike's information criterion (AIC). DIC is a penalized likelihood method based on the posterior distribution of the deviance statistic. Based on the DIC criterion, models with relatively lower DIC values indicate a better fit to the data compared to models with higher DIC values. DIC is defined as

$$\text{DIC} = 2\bar{D} - D(\bar{\boldsymbol{\theta}}),$$

where $\bar{D}$ is the posterior mean of the deviance, and $D(\bar{\boldsymbol{\theta}})$ is the deviance of the vector of the posterior mean values for the model parameter vector ($\boldsymbol{\theta}$).

Due to the presence of issues discussed previously related to the choice of determining the underlying physical process, several different models were considered and fitted to the data. The problem of determining the appropriate underlying physical process is a very intricate and often impossible task. However, using model selection methods, and in particular, DIC, we will be able to draw conclusions on which models provide a better fit to the data based on their statistical significance. Table 12, shows the DIC values calculated for the five models fitted to the data. Model 2 has the lowest DIC value, indicating the increasing predictive power of the model by the use of spatially-varying diffusion parameter. Another important conclusion based on the DIC values is that the models with spatially-varying parameters are preferred, which is an indication of the possible heterogeneous nature of the underlying process.

2.9 Discussion

Modeling of spatio-temporal processes with realistic assumptions considered in the hierarchical Bayesian finite element-based simulation experiment described in this chapter, such as the spatially-varying advection parameter, are often intractable using other techniques. The hierarchical Bayesian finite element-based model, discussed for the sturgeon tracking problem offers a useful platform to estimate parameters governing the dynamics while providing predictive tools which could benefit the management of the Missouri River. Although in the application described in the previous section, the capability of the finite element method in discretizing the PDE-based process model on irregular spatial grids was not utilized, such modeling framework could provide a powerful tool for applications where considering an irregular spatial grid improves the accuracy and interpretability of the model (e.g., irregularly-spaced grid cells defined based on distinguished habitats in the river or clusters of migrating fish population).

An important advantage of considering models with spatially-varying parameters, other than providing more realistic modeling assumptions, is the interpretability of such models which helps maintain useful evidence about the behavior of the shovelnose sturgeon during the spawning process for both the fisheries biologists studying the behavior of the shovelnose sturgeon, and the managers attempting to make decisions to benefit this species and increase their spawning success rates.

An overall conclusion about the five different models fitted to the data reveals that the estimates for the constant diffusion parameters is somewhat invariant to the choice of the model for the advection parameters (i.e., whether constant or spatially-varying advection

was included or not). The comparison of the estimates of the diffusion parameters and the advection parameters in Models 3-5, indicates that the diffusion component is dominant, yielding smaller values for the advection parameters. Interestingly, the plots for spatially-varying advection parameters in Models 4 and 5 show a slight increase in the advection parameter over the spatial domain, indicating a relatively constant behavior of this parameter.

The estimates of the model parameters can provide useful information for fisheries biologists in studying the pre-spawning behavior of shovelnose sturgeon. Such information can help determine the effect of river management actions (e.g., possible effect of dams blocking the upstream migration of fish, or possible effects of spring-rise) on the success of the spawning process. For example, in Model 2, which was selected as the "best" model considering its DIC value compared to other models, the spatially-varying diffusion parameter, as shown in Figure 18, suggests that the increasing rate of diffusion slows down at the 6th grid node (around river mile 270) and increases with a slower rate up to the 8th grid cell (around river mile 334) and the rate of increase for the diffusion parameter is slower for the river miles higher than 334. Such a change in the rate of increase and value of the diffusion parameter could provide insight about a location where the fish migration rate decreases (e.g., due to a change in the environmental and physical conditions, or establishment of territory in a specific segment of the river) which could be a possible indication to where and when the spawning happened. However, since the same pattern is not apparent for all the realizations of the MCMC algorithm, the statistical significance of this result is not evident and no definitive conclusions can be made. Finally, making definitive conclusions

about the spawning behavior of fish based on such indications is not recommended until more data has been acquired.

2.10 Conclusion

A PDE-based hierarchical Bayesian model to address the modeling issues related to an uncertain diffusion process on regular or irregular domains was developed. The methods described in this chapter provide an effective approach to address the modeling of complex PDE-based spatio-temporal processes in the presence of uncertainty. Critically, the finite element method can be employed to allow for modeling the process in its natural and realistic domain, which often includes irregular geometries for environmental and physical processes.

The hierarchical Bayesian modeling approach presented in this chapter also allows for considering efficient parameterization of the PDE-based dynamical model, possibly including spatially-varying parameters, which yields flexible models with strong interpretive power.

Table 1: Posterior mean values for the parameters of the 2-D advection-diffusion simulation

Parameter	**Posterior mean**	**95 % CIs**
α_x	-0.055	(-0.0577, -0.0522)
α_y	0.0152	(0.0118, 0.0186)
β_x	0.0138	(0.0117, 0.0158)
β_y	0.0161	(0.0133, 0.0189)
σ^2_ε	0.0017	(.0016, .0019)
σ^2_η	0.0053	(.0048, .0059)

Table 2: Hyperparameters used in the constant diffusion model for the shovelnose sturgeon tracking problem (Model 1)

Hyperparameter	**Value**
$\tilde{\beta}$	1
σ^2_β	0.1
q_ε	2.1
r_ε	9.091
q_η	2.1
r_η	9.091

Table 3: Hyperparameters used in the spatially-varying diffusion model for the shovelnose sturgeon tracking problem (Model 2)

Hyperparameter	**Value**
$\tilde{\boldsymbol{\beta}}$	$\mathbf{1}$
$\boldsymbol{\Sigma}_\beta$	$0.01 \times \mathbf{R}(0.01)$
q_ε	2.1
r_ε	9.091
q_η	2.1
r_η	9.091

Table 4: Hyperparameters used in the advection-diffusion model for the shovelnose sturgeon tracking problem (Model 3)

Hyperparameter	**Value**
$\tilde{\beta}$	1
σ^2_β	0.01
$\tilde{\alpha}$	0
σ^2_α	1
q_ε	2.1
r_ε	9.091
q_η	2.1
r_η	9.091

Table 5: Hyperparameters used in the advection-diffusion model for the shovelnose sturgeon tracking problem (Model 4)

Hyperparameter	**Value**
$\tilde{\beta}$	1
σ^2_β	0.01
$\tilde{\boldsymbol{\alpha}}$	$\mathbf{0}$
$\boldsymbol{\Sigma}_\alpha$	$0.01 \times \mathbf{R}(0.01)$
q_ε	2.1
r_ε	9.091
q_η	2.1
r_η	9.091

Table 6: Hyperparameters used in the advection-diffusion model for the shovelnose sturgeon tracking problem (Model 5)

Hyperparameter	**Value**
$\tilde{\boldsymbol{\beta}}$	$\mathbf{1}$
$\boldsymbol{\Sigma}_{\beta}$	$0.01 \times \mathbf{R}(0.01)$
$\tilde{\boldsymbol{\alpha}}$	$\mathbf{0}$
$\boldsymbol{\Sigma}_{\alpha}$	$0.01 \times \mathbf{R}(0.01)$
q_{ε}	2.1
r_{ε}	9.091
q_{η}	2.1
r_{η}	9.091

Table 7: Posterior mean and 95% CIs for the univariate model parameters in the shovelnose sturgeon migration problem (Model 1)

Parameter	**Posterior Mean**	95% **CI**
β	0.127	(-0.0117, 0.2835)
σ^2_{ε}	0.0903	(0.0203, 0.2697)
σ^2_{η}	0.6215	(0.2211, 1.5108)

Table 8: Posterior mean and 95% CIs for the univariate model parameters in the shovelnose sturgeon migration problem (Model 2)

Parameter	**Posterior Mean**	95% **CI**
σ^2_{ε}	0.0789	(0.018, 0.2851)
σ^2_{η}	8.811	(1.3914, 35.8683)

Table 9: Posterior mean and 95% CIs for the univariate model parameters in the shovelnose sturgeon migration problem (Model 3)

Parameter	**Posterior Mean**	95% **CI**
α	-0.0668	(-0.5139, 0.4394)
β	0.1275	(-0.0108, 0.2916)
σ^2_ε	0.08131	(0.0185, 0.2791)
σ^2_η	2.0183	(0.5408, 7.1612)

Table 10: Posterior mean and 95% CIs for the univariate model parameters in the shovelnose sturgeon migration problem (Model 4)

Parameter	**Posterior Mean**	95% **CI**
β	0.1299	(-0.0077, 0.2995)
σ^2_ε	0.0773	(0.0188, 0.258)
σ^2_η	2.6446	(0.6527, 9.363)

Table 11: Posterior mean and 95% CIs for the univariate model parameters in the shovelnose sturgeon migration problem (Model 5)

Parameter	**Posterior Mean**	95% **CI**
σ^2_ε	0.0763	(0.0189, 0.2501)
σ^2_η	2.7248	(0.7703, 8.7204)

Table 12: DIC values for all five different models for the shovelnose sturgeon tracking problem

Model (parameters)	**DIC Value**
Model 1: Diffusion Eq.(β)	498.7
Model 2: Diffusion Eq.(β_{sp})	**245.5**
Model 3: Advection-Diffusion Eq.(α, β)	425.5
Model 4: Advection-Diffusion Eq.(α_{sp}, β)	385.0
Model 5: Advection-Diffusion Eq.(α_{sp}, β_{sp})	358.5

3 Semiparametric Zero-Inflated Models for Multivariate Count Processes

3.1 Introduction

The problem of having a large proportion of zero values is a common characteristic of data obtained from environmental and ecological studies involving counts of abundance, presence-absence or occupancy rates (Clarke and Green 1988; Welsh et al. 1996; Martin et al. 2005; Berry et al. 2005). Ignoring and excluding zero values from the analysis of data obtained from field studies can result in loss of important information. For example, when studying abundance or presence-absence of species in ecological studies, having a large proportion of zero values might be an indication of the following: (1) the species is rare or endangered, (2) the species is hard to detect, or (3) a combination of cases (1) and (2), a rare species that is hard to detect. The problem of dealing with rare species and low probability of detection is very common in ecological and biological studies and so zero-heavy data is often obtained from such studies. Thus, standard distributions such as Poisson, binomial and negative-binomial do not provide a good fit. Zero-inflated modeling is an appropriate approach to modeling zero-heavy data which allows the model to account for the large proportion of zero values (e,g,. Lambert 1992; Hall 2000).

Two popular models that account for data with excess zeros are the zero-inflated Poisson (ZIP) and the zero-inflated negative binomial (ZINB). The ZIP model is especially useful in analyzing count data with a large number of zero observations, whereas the ZINB model is more appropriate for cases where an upper bound exists for the response. The ZIP model has been applied to horticulture (Hall 2000), manufacturing (Lambert 1992), and various

other fields of study including health operations (Wang et al. 2002), meteorology (Wikle and Anderson 2003), and ecology (Welsh et al. 1996; Martin et al. 2005). Given the large number of zeros that often occur in data obtained from environmental studies and possible non-linear covariate effects in such problems, a semiparametric ZIP modeling approach for multivariate count processes is proposed. An application of the modeling approach to fisheries data is discussed with the goal of determining which factors are related to zero-inflation probability for a fish species, and which factors are related to catch rates of fish, while accounting for dependence among species. The modeling approaches for the species considered in this problem provides a practical example for a general modeling technique to analyze fisheries data with similar characteristics (i.e., excess zeros and dependence structure among certain species).

The models discussed in this chapter are based on a novel statistical approach for implementation of semiparametric zero-inflated models for multivariate count processes in hierarchical Bayesian framework. Although the modeling approach discussed in this chapter is intuitive and very useful for environmental and ecological applications, no examples of such modeling approaches were found in the literature.

3.2 Modeling Dependencies Among Species

The problem of modeling dependencies among species is of particular interest to ecologists and biologists since there is little information about the interactions among species in a community (e.g., predator-prey relationships), as well as between species and the environment (i.e., natural habitat). The fact that such dependencies existing in natural communities could possibly originate from several different sources transforms the task of modeling such

phenomena into a challenging scientific problem. In this chapter, two different modeling schemes are considered to address different types of dependencies.

For example, let $\mathbf{Y}_1 = (y_{11}, y_{12}, \ldots, y_{1n})'$ and $\mathbf{Y}_2 = (y_{21}, y_{22}, \ldots, y_{2n})'$ where y_{ji} indicates the ith observation for species j ($i = 1, \ldots, n$ and $j = 1, \ldots, J$), denote correlated counts with excess zeros, assumed to have zero-inflated Poisson distributions, marginally. Thus,

$$\mathbf{Y}_1 \sim \begin{cases} \mathbf{0} & \text{w.p. } \mathbf{p}_{1,0}, \\ \text{Poisson}(\boldsymbol{\mu}_1) & \text{w.p. } 1 - \mathbf{p}_{1,0}, \end{cases}$$

and

$$\mathbf{Y}_2 \sim \begin{cases} \mathbf{0} & \text{w.p. } \mathbf{p}_{2,0}, \\ \text{Poisson}(\boldsymbol{\mu}_2) & \text{w.p. } 1 - \mathbf{p}_{2,0}, \end{cases}$$

and $\text{Cov}(\mathbf{Y}_1, \mathbf{Y}_2) > 0$. The covariance structure can be defined in different fashions. The most popular approach to modeling correlated counts is to characterize a covariance structure through a multivariate distribution for the log of the Poisson intensities (i.e. $\boldsymbol{\mu}_i$'s for $i = 1, 2$.). Another approach to modeling correlated counts defines a covariance structure using a common latent Poisson process, yielding a multivariate Poisson distribution (Johnson et al. 1997; Kocherlakota and Kocherlakota 1992). In this chapter, two different models based on these different approaches for modeling dependencies among species with application to a fisheries problem will be discussed.

3.3 Semiparametric Hierarchical Bayesian Modeling Approach to Modeling Multivariate Count Processes

Semiparametric models are useful tools that allow for non-linear relationships between the response variable and covariates through functions of continuous variables (covariates). Semiparametric approaches to modeling zero-inflated count data, although providing a nat-

ural setting for modeling "real-world" phenomena where non-linear relationships among variables often exist, are rare in the literature (e.g., Fahrmeir and Echavarria 2006; Lam et al. 2006).

Penalized spline regression is considered as one of the most attractive and a relatively straightforward method for implementing scatterplot smoothing (Ruppert et al. 2003). Following Ruppert et al. (2003) a semiparametric regression model can be written as

$$y_i = f(x_i) + \epsilon_i,$$

where ϵ_i are i.i.d. $\mathrm{N}(0, \sigma_\epsilon^2)$, y_i and x_i denote the response and independent variables, respectively, and $f(\cdot)$ is a smooth function. There are various choices for the smooth function including smooth functions constructed using cubic splines, B-splines, truncated polynomials, radial splines among other basis functions.

Generalized additive models (GAMs) and in particular, the generalized P-spline model, an extension of penalized spline regression to generalized linear models (GLMs), provides a flexible framework for scatterplot smoothing in the context of GLMs. The hierarchical Bayesian treatment of penalized spline regression models is very easy to implement using WinBUGS (Crainiceanu et al. 2006).

In Bayesian analysis, the use of low-rank thin-plate splines is more appealing than other bases functions, due to mixing properties of the MCMC chains. In particular, when compared to other bases (e.g., truncated polynomials), the parameters of the low-rank thin-plate splines have smaller posterior correlation which improves mixing in the context of MCMC (Crainiceanu et al. 2006). The low-rank thin plate representation of the smooth

function, $f(\cdot)$, is

$$f(x, \boldsymbol{\theta}) = \beta_0 + \beta_1 x + \sum_{k=1}^{K} u_k |x - \kappa_k|^3, \tag{3.26}$$

where $\boldsymbol{\theta} = (\beta_0, \beta_1, u_1, \ldots, u_K)'$ is the vector of regression coefficients, and $\kappa_1 < \kappa_2 < \ldots < \kappa_K$ denote fixed knots. The number of knot locations, K, should be considered large enough to ensure the desired flexibility. A common choice for knots is the sample quantiles of the x's (i.e. unique predictor variables); for example, κ_k is the sample quantile corresponding to probability $\frac{k}{K+1}$. To penalize for overfitting,

$$\sum_{i=1}^{n} [y_i - f(x_i, \boldsymbol{\theta})]^2 + \frac{1}{\tau} \boldsymbol{\theta}' \mathbf{D} \boldsymbol{\theta},$$

is minimized, where τ is the smoothing parameter and $\mathbf{D}$ is a known positive semi-definite *penalty* matrix. The penalty matrix for the thin-plate splines is the $(K+2) \times (K+2)$ matrix

$$\mathbf{D} = \begin{bmatrix} \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times K} \\ \mathbf{0}_{K\times 2} & \boldsymbol{\Omega}_K \end{bmatrix},$$

where the entries of the sub-matrix $\boldsymbol{\Omega}_K$ penalize coefficients of $|x - \kappa_k|^3$. The penalized spline regression model can be written as a linear mixed model (LMM)(Brumback et al. 1999)

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_K \mathbf{u} + \boldsymbol{\epsilon}, \tag{3.27}$$

where $\mathbf{X}$ is the matrix of covariates, $\mathbf{Z}_K$ is an $n \times K$ matrix with the element (i, k) as $|x_i - \kappa_k|^3$, $\boldsymbol{\beta}$ and $\mathbf{u}$ are the regression coefficients for the non-penalized and penalized covariates, respectively, and $\text{cov}\left(\begin{bmatrix} \mathbf{u} \\ \boldsymbol{\epsilon} \end{bmatrix}\right) = \begin{bmatrix} \sigma_u^2 \boldsymbol{\Omega}_K^{-1} & 0 \\ 0 & \sigma_\epsilon^2 \mathbf{I}_n \end{bmatrix}$ (e.g., Crainiceanu et al.

2006). Using the reparameterizations $\mathbf{b} = \mathbf{\Omega}_K^{1/2}\mathbf{u}$ and $\mathbf{Z} = \mathbf{Z}_K\mathbf{\Omega}_K^{1/2}$, model (3.27) can be rewritten as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon}, \tag{3.28}$$

where $\text{cov}\left(\begin{bmatrix} \mathbf{b} \\ \boldsymbol{\epsilon} \end{bmatrix}\right) = \begin{bmatrix} \sigma_b^2\mathbf{I}_K & 0 \\ 0 & \sigma_\epsilon^2\mathbf{I}_n \end{bmatrix}$ (Crainiceanu et al. 2006). The mixed model (3.28) is a flexible form for the MCMC-based computations to estimate the smoothing parameter τ since we have $\tau = \frac{\sigma_b^2}{\sigma_\epsilon^2}$. The estimation can also be conducted in a frequentist setting using Best Linear Unbiased Predictor (BLUP) or Penalized Quasi-Likelihood (PQL) estimation.

3.3.1 Multivariate Normal Distribution Modeling Approach

The multivariate normal distribution is a popular choice for accounting for dependencies among multiple response variables. Multivariate normal distributions have also been used in modeling multivariate count data. For example, a popular method to analyze correlated count data is to utilize a multivariate normal correlation structure in a Poisson-lognormal regression model (Aitchison and Ho 1989; Chib and Winkelmann 2001). The semiparametric zero-inflated model based on a multivariate normal distribution for the log-linear components of the model can be written in general form as,

$$\mathbf{Y} \sim \begin{cases} \mathbf{0} & \text{w.p. } \mathbf{p}, \\ \text{Poisson}(\boldsymbol{\mu}) & \text{w.p. } 1-\mathbf{p}, \end{cases} \tag{3.29}$$

$$\log(\boldsymbol{\mu}) = \boldsymbol{\beta}_0 + \mathbf{X}\boldsymbol{\beta} + f(\mathbf{X}^s) + \boldsymbol{\epsilon}, \tag{3.30}$$

$$\text{logit}(\mathbf{p}) = \boldsymbol{\gamma}_0 + \mathbf{U}\boldsymbol{\gamma}, \tag{3.31}$$

where $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix}$ denotes the observations, $\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}$ denotes the Poisson intensities, $\mathbf{p} = \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{bmatrix}$ denotes the zero-inflation probability, and $\boldsymbol{\epsilon} \sim \text{N}(\mathbf{0}, \mathbf{\Sigma})$. The vectors of regression coefficients denoted by $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ are random intercepts for the models,

and $\mathbf{X}_{n\times p_X}$ and $\mathbf{U}_{n\times p_U}$ (p_X and p_U denote the number of covariates in the log-linear and logistic models, respectively) are matrices of covariates of interest with elements representing indicator variables corresponding to the covariates, as well as the continuous variables modeled parametrically. The continuous variables to be modeled non-parametrically are placed in matrix $\mathbf{X}^s$. The smooth function $f(.)$ in (3.30) is based on thin-plate splines as defined by (3.26), and can be written as

$$f(x^s, \boldsymbol{\theta}) = \alpha_0 + \alpha_1 x^s + \sum_{m=1}^{K} b_m |x_m^s - \kappa_m|^3,$$

with,

$$b_m \sim \mathrm{N}(0, \sigma_b^2).$$

The logit function, commonly used in generalized linear models (GLMs) is defined as (McCullagh and Nelder 1989):

$$\mathrm{logit}(\mathbf{p}) = \log\left(\frac{\mathbf{p}}{1-\mathbf{p}}\right).$$

Note that the intercepts are considered random to help account for uncertainties such as sampling errors and possible covariates that were excluded from the analysis. Also, note that no additive error term is included in the logistic regression model since inclusion of such term could possibly create identifiability issues in the estimation procedure (Hall 2000). However, no such identifiability problems were detected regarding to the error term included in the log-linear portion of the model.

3.3.2 Multivariate Poisson Distribution Modeling Approach

Perhaps a more natural choice for modeling correlated counts is using multivariate discrete distributions such as a multivariate Poisson (MVPoisson) distribution. However, the compu-

tational difficulties involved in fitting such models have traditionally prohibited researchers from using such an approach. Recent advances in hierarchical Bayesian modeling and development of computational methods such as MCMC, have provided easy implementation of multivariate discrete distributions (e.g., MVPoisson).

Definition: Let $\mathbf{Y}_1$ and $\mathbf{Y}_2$ denote variables from a bivariate Poisson distribution. The bivariate Poisson distribution is defined as

$$(\mathbf{Y}_1, \mathbf{Y}_2) \sim \text{BivPoisson}(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \boldsymbol{\lambda}_3), \tag{3.32}$$

which is based on the joint distribution of the variables

$$\mathbf{Y}_1 = \mathbf{Z}_1 + \mathbf{Z}_3,$$

and,

$$\mathbf{Y}_2 = \mathbf{Z}_2 + \mathbf{Z}_3,$$

where $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}_3$ are mutually independent Poisson random variables with mean and variances $\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$, and $\boldsymbol{\lambda}_3$, respectively (Kocherlakota and Kocherlakota 1992; Johnson et al. 1997).

Assuming $\mathbf{Y}_1$ and $\mathbf{Y}_2$ denote variables from a bivariate Poisson distribution, the covariance between $\mathbf{Y}_1$ and $\mathbf{Y}_2$ is

$$\text{cov}(\mathbf{Y}_1, \mathbf{Y}_2) = \text{cov}(\mathbf{Z}_1 + \mathbf{Z}_3, \mathbf{Z}_2 + \mathbf{Z}_3) = \text{var}(\mathbf{Z}_3) = \boldsymbol{\lambda}_3,$$

and the correlation coefficient between $\mathbf{Y}_1$ and $\mathbf{Y}_2$ is

$$\text{corr}(\mathbf{Y}_1, \mathbf{Y}_2) = \frac{\boldsymbol{\lambda}_3}{\sqrt{(\boldsymbol{\lambda}_1 + \boldsymbol{\lambda}_3)(\boldsymbol{\lambda}_2 + \boldsymbol{\lambda}_3)}}.$$

The joint probability mass function derived by Campbell (1938) is

$$\mathrm{P}(\mathbf{Y}_1 = \mathbf{y}_1, \mathbf{Y}_2 = \mathbf{y}_2) = e^{-(\boldsymbol{\lambda}_1+\boldsymbol{\lambda}_2+\boldsymbol{\lambda}_3)} \sum_{i=0}^{\min(\mathbf{y}_1,\mathbf{y}_2)} \frac{\boldsymbol{\lambda}_1^{\mathbf{y}_1-i}\boldsymbol{\lambda}_2^{\mathbf{y}_2-i}\boldsymbol{\lambda}_3^{i}}{(\mathbf{y}_1-i)!(\mathbf{y}_2-i)!i!}. \tag{3.33}$$

The bivariate Poisson distribution motivates the construction of a simple, yet intuitive, bivariate zero-inflated model which can be easily generalized to multivariate cases. The existing literature on models based on multivariate zero-inflated distributions is sparse and available models are often over-parameterized (e.g., Li et al. 1999). For the bivariate Poisson variables described previously, let $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}_3$ denote mutually independent random variables from a zero-inflated Poisson distribution. Thus,

$$\mathbf{Z}_1 \sim \begin{cases} \mathbf{0} & \text{w.p. } \mathbf{p}_1, \\ \text{Poisson}(\boldsymbol{\lambda}_1) & \text{w.p. } 1-\mathbf{p}_1, \end{cases}$$

$$\mathbf{Z}_2 \sim \begin{cases} \mathbf{0} & \text{w.p. } \mathbf{p}_2, \\ \text{Poisson}(\boldsymbol{\lambda}_2) & \text{w.p. } 1-\mathbf{p}_2, \end{cases}$$

and,

$$\mathbf{Z}_3 \sim \begin{cases} \mathbf{0} & \text{w.p. } \mathbf{p}_3, \\ \text{Poisson}(\boldsymbol{\lambda}_3) & \text{w.p. } 1-\mathbf{p}_3, \end{cases}$$

where $\mathbf{p}_1$, $\mathbf{p}_2$, and $\mathbf{p}_3$ denote the zero-inflation probability for $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}_3$, respectively. Thus, $\mathbf{Y}_1 = \mathbf{Z}_1 + \mathbf{Z}_3$ and $\mathbf{Y}_2 = \mathbf{Z}_2 + \mathbf{Z}_3$ follow a bivariate Poisson distribution with excess zeros (i.e., bivariate ZIP). The distributional assumption for $\mathbf{Z}_3$ in cases where there is no information available to inform the zero-inflation probability could be modified to a regular Poisson distribution to avoid identifiability issues in the modeling procedure.

A semiparametric hierarchical Bayesian model based on the discussed multivariate ZIP

setting can be written as

$$\mathbf{Y}_1 = \mathbf{Z}_1 + \mathbf{Z}_3 \quad \text{and} \quad \mathbf{Y}_2 = \mathbf{Z}_2 + \mathbf{Z}_3,$$

$$\mathbf{Z}_j \sim \begin{cases} \mathbf{0} & \text{w.p. } \mathbf{p}_j, \\ \text{Poisson}(\boldsymbol{\lambda}_j) & \text{w.p. } 1 - \mathbf{p}_j, \end{cases} \tag{3.34}$$

$$\log(\boldsymbol{\lambda}_j) = \beta_{j,0} + \mathbf{X}\boldsymbol{\beta}_j + f(\mathbf{X}^s), \tag{3.35}$$

$$\text{logit}(\mathbf{p}_j) = \gamma_{j,0} + \mathbf{U}\boldsymbol{\gamma}_j, \tag{3.36}$$

for $j = 1, 2, 3$ which denotes species 1, 2 and the latent process, respectively. This indicates a simple model for multiple species, assuming only one overall common latent process but more complex structures are possible. Other notations are similar to the multivariate normal case described previously.

The next section describes an application of the hierarchical Bayesian framework for two different semiparametric zero-inflated Poisson models for multivariate count data, as discussed in this section, implemented using WinBUGS.

3.4 Application: Modeling Catch Per Unit Area (CPUA)

In this section, models of fish catch per unit area resulting from multiple gears for two benthic fish species in the Missouri River are considered and the application of a hierarchical Bayesian modeling framework is discussed.

3.4.1 Study Background

In 1995, the United States Geological Survey (USGS) and the Montana Department of Fish, Wildlife, and Parks commenced a study to look at benthic fishes in the warm-water portion of the Missouri River system (Berry and Young 2001; Berry et al. 2005). The Missouri River extends 2,339 miles from southwest Montana to the Mississippi River. Benthic fish

are fish that live or feed on the bottom of the river and are of particular interest because of their sensitivity to changes in habitat. The main goal of the study was to evaluate the status, distribution, and habitat associations of these fishes in the Missouri River to provide data necessary for improvement in their management.

Included in the Missouri River benthic fishes study were 26 different species of benthic fishes (Berry et al. 2005). To analyze the data, researchers divided the Missouri River into three zones; the upper zone or "least-altered zone" (LA), the middle or "inter-reservoir" zone (IR), and the lower or "channelized zone" (CH) (see Figure 28). The LA zone included the lower Yellowstone River. The IR zone was characterized by short riverine segments between the six large mainstream reservoirs. The CH zone was channelized for navigation and flows are controlled by discharges from upstream dams and by inputs from tributaries. Each zone was then divided into segments creating a total of twenty seven segments for the entire river. The river was partitioned into 10-100 km long segments based on geomorphic (e.g., tributaries, geology) and constructed features (e.g., impoundments, channelization, urban areas). Although there were twenty seven segments included in the study design, there were only fifteen segments sampled during the three years of the study considered here (Figure 28). The LA zone includes segments 3, 5 and 9, the IR zone includes segments 7, 8, 10, 12, 14 and 15, and the CH zone includes segments 17, 19, 22, 23, 25 and 27.

The six primary macrohabitats found in the river were identified. These macrohabitats were defined within segments to be "distinctive, repeatable natural and man-made physical features" (Berry and Young 2001). In particular, macrohabitats were the inside bend (ISB), outside bend (OSB), channel cross-over (CHXO), tributary mouth (TRM),

secondary channel-connected (SCC), and secondary channel not-connected (SCN) as shown in Figure 29. The three macrohabitats associated with bends (ISB, OSB and CHXO) were considered under the label BEND, resulting in a total of four different macrohabitats for statistical purposes. The averaging of the three BEND macrohabitats was necessary because they were not selected independently (i.e., all three were sampled at each bend).

Common to fisheries field studies, the data obtained are based on multiple gears and include a large proportion of zeros which makes the analysis of this data complicated. Using standard parametric statistical methods on data from each gear separately, Berry et al. (2005) excluded several river segments and macrohabitats from the analysis due to high percentage of zero observations (i.e., violation of normality and homogeneity of variance assumptions). This resulted in limitations on making comprehensive conclusions about the complete domain of the study. In particular, the analyses conducted by Berry et al. (2005) were limited by: 1) a large percentage of zero observations in the data set that caused loss of power due to having to combine data at larger spatial and temporal scales, and 2) separate analyses for each gear. These problems created issues and constraints on the usage of the standard classic parametric statistical methods (such as ANOVA). Our goal was to develop and implement a modeling framework which would allow us to give meaningful ecological interpretations based on the model results and raise the predictive precision of the model in the presence of the realistic constraints as described above. The hope is that the type of gear, macrohabitat, segment and year will help identify characteristics that explain where certain species are most likely to populate. Also of interest is the large number of zeros in the data that cannot justifiably be deleted from the analysis. Therefore, these zeros must

somehow be accounted for in the modeling process. The chosen model for this data is the ZIP model, which will account for those parameters that help explain the mean fish count as well as those that explain the zero-inflation probability (i.e. excess zero observations).

3.4.2 Data Collection

For each sampling effort, a simple collect and count method was used for catching the fish where the number of fish, the fish species, and size of area sampled (estimated by width of gear and distance covered as described in Berry et al. 2005) were recorded. Each year, for each segment, researchers fished using as many as four of the five different gears chosen for this study to collect fish within each of the four randomly chosen macrohabitats. Not all gears were used in all macrohabitats because no gear was considered effective at sampling them all. Out of these five gears, four are active gears (benthic trawl, beach seine, drifting trammel net and electrofishing) and one is a passive gear (stationary gillnet). Active gears are the gears that are moved over the sampling area to collect fish. Passive gears are stationary in that they are located within the sampling area for a specific amount of time to collect fish. In our analysis, due to uncertainty about transformation of sampled area for the passive gear into a scale comparable to the active gears, only the active gears were considered. Finally, not only do the different gears cover different areas, but each is designed differently making each more prone to catch different species and sizes of fish. All of these factors combined affect the number of fish caught in any particular sample by any particular gear.

The process of collecting and counting the fish was repeated for three years from 1996 to 1998 (see Berry et al. 2005 for details). The process of data collection included multiple

sub-samples of fish using different gears within each segment over randomly chosen macrohabitats. The data used in this analysis and the analysis done by Berry et al. (2005) was obtained after combining data in the mesohabitat (smaller scale habitat within a macrohabitat) and the sub-sample levels resulting in data at the macrohabitat level. This was necessitated by the varying numbers of sub-samples collected at the mesohabitat level within the macrohabitat level. An average of the data at each sub-sample level was used to obtain data at the macrohabitat level, so that each sub-sample had an equal level of influence on the resulting means.

Smallmouth buffalo (*Ictiobus bubalus*) and bigmouth buffalo (*Ictiobus cyprinellus*) are two commercially important species that can be found in many habitats in rivers. These two species were collected and counted as part of the benthic fishes study in the Missouri River. The resulting data contains observations greater than zero and a large portion of zero observations (see Figure 30). There are 1477 observations for both smallmouth buffalo and bigmouth buffalo collected over the three-year period of interest. The observations for smallmouth buffalo included 73% zeros and the observations for bigmouth buffalo included 77% zeros. The correlation coefficient for these two species is 0.63 (based on the Pearson product-moment correlation) which implies a strong dependence among the two species. Thus, there is sufficient motivation for considering a zero-inflated modeling approach, specifically a ZIP model for multivariate counts, to analyze these data where the zero-inflated model accounts for the high percentage of zero observations and the multivariate distributional assumption accounts for the dependence among these two species.

In this section, semiparametric ZIP models are considered which allow for modeling

both sources of zeros (i.e., sampling and structural zeros) simultaneously by using the indicator variables corresponding to gears, macrohabitats and segments as covariates for modeling the zero-inflation probability. The covariates include data on: four different gears including electrofishing (EF), benthic trawl (BT), beach seine (BS), and drifting trammel net (DTN) where BS is considered as a baseline (i.e., set to zero); four macrohabitats including tributary mouth (TRM), secondary channel-connected (SCC), secondary channel not-connected (SCN), and Bend with TRM as a baseline; fifteen segments (3, 5, 7, 8, 9, 10, 12, 14, 15, 17, 19, 22, 23, 25, 27) with segment 25 set as a baseline; and three years (1996-1998) with 1998 set as baseline. Another categorical variable used in the model describes the substrate composition (sand, gravel, and silt) where silt is considered as a baseline. Note that the choice of baseline is arbitrary and has no effect on the analysis. There are also continuous variables that are used as covariates in the modeling process. These covariates include depth, water temperature, conductivity, turbidity (log(turbidity) is used), and velocity. In the models described in the next section, water temperature, depth and log(turbidity) were chosen to be modeled non-parametrically based on exploratory analysis.

3.4.3 Model 1: Multivariate Normal Distribution

Model 1 is based on a multivariate normal distribution for Poisson log-linear regression components of the model as described in (3.29- 3.31). The model developed in this section

considers this approach in a semiparametric zero-inflated setting as,

$$\mathbf{Y} \sim \begin{cases} \mathbf{0} & \text{w.p. } \mathbf{p}, \\ \text{Poisson}(\boldsymbol{\lambda}' \text{ diag}(\mathbf{a})) & \text{w.p. } 1-\mathbf{p}, \end{cases}$$

$$\log(\boldsymbol{\lambda}) = \boldsymbol{\beta}_0 + \mathbf{X}\boldsymbol{\beta} + f(\mathbf{X}^s) + \boldsymbol{\varepsilon}, \tag{3.37}$$

$$\text{logit}(\mathbf{p}) = \boldsymbol{\gamma}_0 + \mathbf{U}\boldsymbol{\gamma}, \tag{3.38}$$

where $\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \end{bmatrix}$ is an $(n_1+n_2)\times 1$ stacked vector of observations (n_i denotes the length of vector $\mathbf{Y}_i$ for $i = 1, 2$), $\boldsymbol{\lambda} = \begin{bmatrix} \boldsymbol{\lambda}_1 \\ \boldsymbol{\lambda}_2 \end{bmatrix}$, $\mathbf{p} = \begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \end{bmatrix}$ and $\mathbf{p}_j$ denotes the zero-inflation probability for $j = 1, 2$ (species 1 is smallmouth buffalo and species 2 is bigmouth buffalo), $\mathbf{a}$ accounts for the different areas (or "level of effort") involved in each separate measurement. The vectors of regression coefficients denoted by $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, $\boldsymbol{\beta}_0$ and $\boldsymbol{\gamma}_0$ are random intercepts for the models, and $\mathbf{X}$ and $\mathbf{U}$ are $(n_1 + n_2) \times n_p$ matrices of covariates of interest with elements representing indicator variables corresponding to the variables gear, segment, macrohabitat, and year, as well as the continuous variables modeled parametrically (n_p denotes the number of covariates). The continuous variables to be modeled non-parametrically are placed in matrix $\mathbf{X}^s$. The smooth function $f(\cdot)$ in (3.37) is based on thin-plate splines as defined by (3.26), and can be written as

$$f(x^s, \boldsymbol{\theta}) = \alpha_0 + \alpha_1 x^s + \sum_{m=1}^{K} b_m |x^s - \kappa_m|^3,$$

with,

$$b_m \sim \text{N}(0, \sigma_b^2).$$

Note that the covariates in the model correspond to levels of several categorical variables which make the interpretation of the intercepts impractical. However the intercepts are

considered random to help account for uncertainties such as sampling errors and possible covariates that were excluded from the analysis. The coefficients of the model corresponding to a specific level of a categorical variable are interpreted as the mean fish CPUA in that level relative to the baseline level (the level set to zero). The choice of baseline level for a categorical variable is arbitrary.

The hierarchical Bayesian implementation of the model requires the assignment of prior densities to unknown parameters. Here, vague but proper prior densities are considered as follows,

$$\begin{aligned}
\alpha_0 &\sim \mathrm{N}(0, 100), \\
\alpha_1 &\sim \mathrm{N}(0, 100), \\
\beta_0 &\sim \mathrm{N}(0, 100), \\
\beta_i &\sim \mathrm{N}(0, 100), \qquad \text{for} \quad i = 1, \ldots, p_X, \\
\gamma_0 &\sim \mathrm{N}(0, 100), \\
\gamma_i &\sim \mathrm{N}(0, 100), \qquad \text{for} \quad i = 1, \ldots, p_U, \\
\boldsymbol{\Sigma}^{-1} &\sim \mathrm{Wishart}(\mathbf{I}_{2\times 2}, 2), \\
\tau_b \equiv 1/\sigma^2 &\sim \mathrm{Gamma}(\mathrm{mean} = 0.01, \mathrm{var} = 100),
\end{aligned}$$

where p_X and p_U denote the number of covariates included in vectors $\mathbf{X}$ and $\mathbf{U}$, respectively.

3.4.4 Model 2: Multivariate Poisson Distribution

In Model 2 the fish counts for smallmouth buffalo and bigmouth buffalo are assumed to follow a bivariate Poisson distribution with excess zeros as described in previous section.

Thus, a semiparametric hierarchical Bayesian model based on this assumption is

$$(\mathbf{Y}_1, \mathbf{Y}_2), \sim \text{BivPoisson}(\boldsymbol{\lambda}_1' \text{diag}(\boldsymbol{a}), \boldsymbol{\lambda}_2' \text{diag}(\boldsymbol{a}), \boldsymbol{\lambda}_3' \text{diag}(\boldsymbol{a})), \tag{3.39}$$

$$\log(\boldsymbol{\lambda}_j) = \beta_{j,0} + \mathbf{X}\boldsymbol{\beta}_j + f_j(\mathbf{X}^s), \tag{3.40}$$

$$\text{logit}(\mathbf{p}_j) = \gamma_{j,0} + \mathbf{U}\boldsymbol{\gamma}_j, \tag{3.41}$$

where $j = 1, 2, 3$, correspond to parameters of independent Poisson variables $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}_3$, respectively. The other parameters and notations are defined similarly as described for Model 1.

The prior densities required for unknown parameters in a hierarchical Bayesian setting, are assigned using vague but proper prior distributions, similar to model 1.

3.5 Results

The inference is based on the posterior distribution of unknown parameters, as described in previous sections. The freeware WinBUGS (Spiegelhalter et al. 2003) is used to conduct MCMC computations. The results are obtained from WinBUGS after accounting for a "burn-in" period for the MCMC chains to guarantee convergence to the target posterior distribution (i.e., chain achieves stationarity). The number of iterations for the Gibbs sampler was 30000, where the first 10000 iterations were considered as "burn-in" and posterior inference was conducted using the last 20000 realizations obtained from the MCMC. Convergence of the Gibbs sampler was determined based on visual inspection of three MCMC chains with different initial values for each.

3.5.1 Model 1 Results

The results for parameters of the log-linear models are shown in Tables 13 and 15 for the log-linear model (3.37) for smallmouth buffalo and bigmouth buffalo. The results for parameters of the log-linear model for smallmouth buffalo (Table 13) imply significant effects for log(turbidity) (posterior mean=-0.424). For the categorical variables, bend and SCN have significantly lower mean CPUA compared to SCC and TRM (baseline), segment 14 has a significantly lower mean CPUA compared to all other segments (including the baseline, segment 25), the mean CPUA for years 1996 and 1997 is significantly lower than the baseline (1998). The mean CPUAs for BT, DTN and EF are significantly lower than BS (baseline).

The results for bigmouth buffalo (Table 15) show significant effects for conductivity (posterior mean= -9.991). The significant patterns based on the categorical variables include significantly low mean CPUA in segments 5, 12, 19 and 23 compared to all other segments including segment 25 (baseline), and significantly low mean CPUA for BT compared to all other gears including BS (baseline). The mean CPUA for gravel is significantly lower than sand and silt (baseline).

The only parameters that significantly impact the zero-inflation probability of smallmouth buffalo are segment 9 with a significantly decreasing effect (compared to the other segments including the baseline, segment 25), electrofishing with a significantly decreasing effect (compared to the other gears including the baseline, beach seine), depth (posterior mean= -6.404), conductivity (posterior mean= -11.889), water temperature (posterior mean= 1.184), and gravel with a significantly decreasing effect (compared to sand and silt) (Table 14). For bigmouth buffalo, the only parameters that significantly impact the zero-

inflation probability is depth with a significantly increasing effect (posterior mean= -16.652) (Table 16). The posterior mean values for the correlation matrix of the multivariate normal effects is 0.92 with 95% credible intervals (CIs) (0.82, 0.975) suggesting a very strong relationship between the two species under study.

Figure 31 shows the thin-plate spline fit as well as 95% CIs for water temperature, log(turbidity) and depth for both species. The non-linear effect of water temperature is strong for both species with peaks around 15 to 25 degrees for smallmouth buffalo and below 15 degrees for bigmouth buffalo, indicating the habitat choice of these two species. The effect of log(turbidity) has a slight departure from linearity for both species, showing a decreasing trend for smallmouth buffalo and an increasing trend for bigmouth buffalo. The fitted functions for depth show a slightly non-linear effect for both species with peaks around 5 to 8 meters.

3.5.2 Model 2 Results

The results for parameters of the log-linear models are shown in Table 17-19 for the log-linear model (3.40) for smallmouth buffalo and bigmouth buffalo. The results for parameters of the log-linear model for smallmouth buffalo (Table 17) imply significant effects for conductivity (posterior mean=-1.481) and water temperature (posterior mean= 0.327). For the categorical variables, bend has significantly lower mean CPUA compared to other macrohabitats including TRM (baseline) and SCN has a significantly higher mean CPUA compared to other macrohabitats including TRM (baseline), segments 3, 10, 17, 19 and 27 have significantly higher mean CPUA compared to all other segments (including the baseline, segment 25), the mean CPUA for year 1996 is significantly lower than the 1997 and the baseline

(1998). The mean CPUAs for BT and DTN are significantly lower than BS (baseline). Gravel has a significantly lower mean CPUA compared to sand and silt (baseline).

The results for bigmouth buffalo (Table 19) show significant effects for water temperature (posterior mean= 20.077). The significant patterns based on the categorical variables include significantly low mean CPUA for bend compared to all other macrohabitats, significantly lower mean CPUA for segments 3, 9 and 12 compared to all other segments including segment 25 (baseline), significantly higher mean CPUA for segments 8 and 14 compared to all other segments including segment 25 (baseline), and significantly lower mean CPUA for gravel and sand compared to silt (baseline).

The only parameters that significantly impact the zero-inflation probability of smallmouth buffalo are bend and SCC with a significantly decreasing effect (compared to SCN and TRM), SCN with a significantly increasing effect (compared to bend, SCC and TRM), electrofishing with a significantly decreasing effect (compared to the other gears including the baseline, beach seine), segments 7 and 14 with a significantly increasing effect (compared to all other segments including the baseline, segment 25), depth (posterior mean= -3.5), conductivity (posterior mean= -7.8), velocity (posterior mean= -14.5), water temperature (posterior mean= 1.5), and gravel with a significantly decreasing effect (compared to sand and silt)(Table 18). For bigmouth buffalo, the only parameters that significantly impact the zero-inflation probability are bend and SCC (compared to SCN and TRM), and segments 3 and 7 with a significantly decreasing effect (compared to all other segments including the baseline, segment 25), segment 5 with a significantly increasing effect (compared to all other segments including the baseline, segment 25)(Table 20).

Figure 32 shows the thin-plate spline fit as well as 95% CIs for water temperature, log(turbidity) and depth for both species. The non-linear effect of water temperature is strong for both species with peaks around 15 to 25 degrees for smallmouth buffalo and below 15 degrees for bigmouth buffalo, indicating the habitat choice of these two species. The effect of log(turbidity) has a slight departure from linearity for both species, showing a decreasing trend for smallmouth buffalo and an increasing trend for bigmouth buffalo. The fitted functions for depth show a non-linear effect for smallmouth buffalo with peak around 5 to 8 meters. However, the effect of depth for bigmouth buffalo is only slight non-linear, showing a strong decreasing trend.

3.5.3 Model Comparison

There are several approaches for conducting model comparison and selection in hierarchical Bayesian settings. Approaches based on Bayes factors, and also measures such as deviance information criterion (DIC), although gaining popularity among researchers, are either difficult to implement or difficult to interpret. In the context of zero-inflated modeling, the use of traditional DIC formulation for mixture models is not recommended (Spiegelhalter et al. 2002). However, there are alternative formulations of DIC available that are appropriate for cases of mixture models and missing data models (Celeux et al. 2003).

The model comparison approach used in this chapter is based on posterior predictive loss criteria (Gelfand and Ghosh 1998), which is easy to implement using output from MCMC posterior simulation. Another advantage of this approach is its powerful and practical interpretability, specifically in problems where predictive power of the analysis is of interest. The posterior predictive loss approach considered here is based on a squared error

loss (Gelfand and Ghosh 1998; Banerjee et al. 2004), where the predictive power of the model is appraised under a squared error loss function using predictions of replicates of the observed data, $Y_{l,rep}$, $l = 1, \ldots, n$.

Under a squared error loss, the posterior predictive loss criterion is

$$D_k = \frac{k}{k+1} G + P, \tag{3.42}$$

where

$$G = \sum_{l=1}^{n} (\mu_l - y_{l,obs})^2, \text{ and } P = \sum_{l=1}^{n} \sigma_l^2,$$

with $\mu_l = E(Y_{l,rep}|\mathbf{y})$ and $\sigma_l^2 = \text{Var}(Y_{l,rep}|\mathbf{y})$ representing the mean and the variance of the predictive distribution of $Y_{l,rep}$ given the observed data $\mathbf{y}$, respectively. In (3.42), G is a measure of goodness-of-fit and P is a penalty term to avoid overfitting. The model selection is insensitive to the choice of k (weights on the departure from the observed data) in (3.42). Finally, the models with better predictive power are the models with lower D_k values. In this chapter, value of $k = 1$ is considered to conduct model comparison based on 10000 replicated predictions for each model and for each species. The results based on posterior predictive loss criteria for both models are presented in Table 21. The results show that Model 2 (multivariate Poisson) is a better choice for modeling smallmouth buffalo due to lower D_1 value for Model 2 compared to Model 1. The results for bigmouth buffalo indicate a lower value for D_1 for Model 1 (multivariate normal) however a more careful assessment of the values of G and P for both models indicates that Model 1 (multivariate normal) tends to "overfit" the data. In general, models that overfit do better on the goodness of fit (i.e., smaller value for G) but tend to inflate the variance (i.e., larger value for P) (Banerjee

et al. 2004).

3.6 Discussion

The smallmouth buffalo and the bigmouth buffalo are closely related species. The results for Model 1 show a high estimated correlation (obtained using the posterior mean estimates of the variance components associated with the multivariate normal error) for these two species with the posterior mean of 0.92. The high correlation could be due to ecological and biological similarities among the species, as well as similar habitat preference and physical attributes of these habitats. The results also present descriptive patterns of abundance of fish, as well as effectiveness of certain gears. The results show that both species were caught more in deeper water, water temperatures between 15°C to 20°C, and lower log(turbidity). The abundance of both species is significantly lower in some inter-reservoir segments compared to most other segments. Note that segment 12 includes the Garrison dam and segment 14 includes the Fort Randall Dam. The bigmouth buffalo is also less abundant in most channelized segments near Kansas City. The least effective gear for the bigmouth buffalo is benthic trawl with a significantly lower mean CPUA for this species and the most effective gear for smallmouth buffalo is beach seine. Both species were caught less in bend macrohabitats. Smallmouth buffalo are also less abundant in secondary channel non-connected macrohabitats. Bigmouth buffalo are also less abundant in secondary channel connected macrohabitats. The temporal pattern for smallmouth buffalo indicates that the number of smallmouth buffalo has increased in 1998 compared to 1996 and 1997. Bigmouth buffalo are less abundant in gravel.

The results for Model 2 show mostly similar patterns. However, a higher number of

variables become significant based on this model. The results show significant effect for water temperature with more fish caught at temperatures between 15°C to 20°C. The results also show more smallmouth buffalo were caught at lower log(turbidity), higher depth and lower conductivity while more bigmouth buffalo were caught at higher log(turbidity), and in shallower depths. Both species are less abundant in bend macrohabitats. Smallmouth buffalo are more abundant in secondary channel non-connected. Smallmouth buffalo are more abundant in most channelized segments and some inter-reservoir segments while bigmouth buffalo are more abundant in some inter-reservoir segments and are less abundant in some least-altered segments. Smallmouth buffalo had the lowest mean CPUA in 1996 while bigmouth buffalo were more abundant in 1996. Smallmouth buffalo are less abundant in gravel. The least effective gears for smallmouth buffalo are BT and DTN while there are no significant results on the performance of gears for bigmouth buffalo.

Berry and Young (2001) indicate that the spawning of smallmouth buffalo initiates at 13°C by rising water levels and increasing temperatures. The results based on the estimated function fit for both models, as described above, could possibly be related to a preferred temperature range for spawning, resulting in a higher abundance of fish in water temperatures between 15°C and 20°C. The estimated function for water temperature for bigmouth buffalo shows a similar pattern to the ones discussed for smallmouth buffalo for Model 2, although the results for bigmouth buffalo in Model 1 are different. Berry and Young (2001) indicate that bigmouth buffalo spawning occurs in spring with water temperature of 14°C. The estimated function using Model 2 shows a clear change of pattern around 20°C which could be due to spawning of the species. Berry and Young (2001) indicate that

smallmouth buffalo typically inhabits large rivers and prefer deep, clear, warm waters. The results from Model 2 are in agreement with these characteristics. Berry and Young (2001) also indicate that bigmouth buffalo spawn in spring when water temperatures reach 14°C in shallow areas. The results from Model 2 are in complete agreement with their findings.

Berry et al. (2005) indicate that smallmouth buffalo was rarely found in high velocity habitats (e.g., channel cross-over, and outside bend). Models 1 and 2 results are in agreement with their findings, showing lower catch in bends. Berry et al. (2005) also showed that smallmouth buffalo were usually found in TRMs and SCNs with low turbidity, low percentage of slit, and high temperatures. Model 1 results are in partial agreement with their findings, showing high mean CPUA for TRMs and low log(turbidity). However, Model 2 results show higher mean CPUA for both TRMs and SCNs and high temperatures. Berry et al. (2005) did not find any significant segment contrasts for smallmouth buffalo. However, they reported that catches were lower in the IR segments than the LA and CH segments, and lower in segment 12 downstream from Garrison Dam than upstream. Model 1 results show segment 14 has a significantly lower mean CPUA compared to all other segments, and Model 2 results show that segments 3, 10, 17, 19 and 27 have significantly higher mean CPUA compared to all other segments.

Berry et al. (2005) show that bigmouth buffalo was abundant in all segments with highest total catch in the IR zone. Models 1 and 2 results partially support their findings (some of the IR zone segments have higher mean CPUA). Berry et al. (2005) show that most of the bigmouth buffalo catch was done using beach seine or by electrofishing. However, they do not show statistical analysis due to low catch. Model 1 results show that more

fish were caught using beach seine than benthic trawl (Model 2 results show no significant differences among gears). Berry et al. (2005) also state that bigmouth buffalo were more present in low velocity, high temperatures and finer substrates. However, they mention this finding is based on visual inspection of the catch per effort plots. Model 2 results show higher catch in high temperatures and finer substrates (significantly higher mean CPUA for slit compared to gravel and sand).

A comparison between the results obtained from both models show very similar results for the parameters of the logistic regression models. However, Model 2 results show a higher number of significant parameters and also the results are in agreement with findings from previous studies. Considering this and also based on the model comparison results discussed in the previous sections, there is strong evidence that Model 2 provides a more suitable and accurate modeling framework to study these species. However, care should be given to the generalization of such a statement when analyzing data from other species. Here, there is a very strong dependence among the two species which can be modeled better through a joint distribution assumption.

3.7 Conclusion

In this chapter, semiparametric hierarchical Bayesian models for multivariate count processes were discussed and an application to modeling fish catch per unit area based on observations on two closely relates species of benthic fish in the Missouri River was presented. Two different modeling approaches to analyze these data were considered and results were discussed. The modeling frameworks described in this chapter allow for several issues such as correlated counts, possible non-linear relationships, and parameter uncertainties,

as well as excess zeros that are common problems in environmental studies to be modeled simultaneously.

Table 13: Model 1- Posterior mean and 95% Credible Intervals (CIs) for the log-linear model parameters for smallmouth buffalo

Parameter	**Posterior Mean**	95% **CI**
intercept	-5.744	(-8.584, -2.978)
Bend	-1.849	(-2.879, -0.381)
SCC	-0.734	(-1.919, 0.686)
SCN	-2.109	(-3.899,-0.414)
Seg3	1.597	(-0.498,3.615)
Seg5	-0.618	(-2.426,1.177)
Seg7	-0.129	(-2.008,1.451)
Seg8	-1.329	(-3.038,0.318)
Seg9	-1.289	(-3.304,0.321)
Seg10	-0.034	(-2.051,1.934)
Seg12	-1.657	(-4.539,0.847)
Seg14	-1.955	(-4.070,-0.246)
Seg15	-1.147	(-2.757,0.386)
Seg17	-0.558	(-2.049,0.990)
Seg19	-0.338	(-2.130,1.237)
Seg22	-0.029	(-1.604,1.567)
Seg23	-0.932	(-2.480,0.423)
Seg27	-0.685	(-2.637,1.281)
Year96	-1.701	(-3.094,-0.508)
Year97	-0.986	(-1.812,-0.191)
BT	-13.039	(-17.200,-9.889)
DTN	-11.023	(-12.790,-9.043)
EF	-4.816	(-5.991,-3.463)
Depth	0.335	(-0.024,0.694)
Cond	-0.029	(-0.464,0.346)
LogTurb	-0.424	(-0.624,-0.090)
Gravel	-0.986	(-2.401,0.539)
Sand	0.5696	(-0.148 , 1.363)
WaterTemp	0.112	(-0.028 ,0.228)
Velocity	0.881	(-0.793,2.288)

Table 14: Model 1- Posterior mean and 95% Credible Intervals (CIs) for the logistic regression model parameters for smallmouth buffalo

Parameter	**Posterior Mean**	95% **CI**
intercept	-15.178	(-28.640, -3.525)
Bend	-3.856	(-11.630, 2.265)
SCC	-2.440	(-11.570, 5.331)
SCN	-5.544	(-18.390, 4.250)
Seg3	-2.502	(-15.260, 7.856)
Seg5	-7.950	(-18.010, 1.972)
Seg7	7.023	(-2.981, 17.030)
Seg8	-3.973	(-17.233, 6.935)
Seg9	-16.891	(-31.201, -3.562)
Seg10	0.079	(-10.530, 13.611)
Seg12	4.658	(-15.786, 20.131)
Seg14	-9.293	(-22.880, 0.873)
Seg15	8.244	(-2.979, 18.891)
Seg17	-5.787	(-22.030, 5.646)
Seg19	4.708	(-12.620, 15.160)
Seg22	-7.927	(-19.340, 1.554)
Seg23	-3.574	(-16.840, 7.198)
Seg27	1.317	(-13.001, 11.420)
BT	-0.201	(-19.381, 21.411)
DTN	-2.940	(-19.244, 9.639)
EF	-13.347	(-21.762, -6.591)
Depth	-6.404	(-16.260, -0.360)
Cond	-11.889	(-15.980, -7.370)
LogTurb	-0.031	(-1.662, 1.566)
Gravel	-13.390	(-23.930, -5.369)
Sand	1.691	(-6.980, 8.896)
WaterTemp	1.184	(0.606, 1.835)
Velocity	-12.336	(-23.901, 0.817)

Table 15: Model 1- Posterior mean and 95% Credible Intervals (CIs) for the log-linear model parameters for bigmouth buffalo

Parameter	**Posterior Mean**	95% **CI**
intercept	-13.039	(-17.200, -9.889)
Bend	-11.023	(-12.790, -9.043)
SCC	-4.816	(-5.991, -3.462)
SCN	0.335	(-0.024, 0.694)
Seg3	-0.029	(-0.464, 0.346)
Seg5	-0.424	(-0.625, -0.090)
Seg7	0.881	(-0.792, 2.288)
Seg8	0.112	(-0.028, 0.228)
Seg9	-0.986	(-2.401, 0.539)
Seg10	0.569	(-0.148, 1.363)
Seg12	-7.371	(-11.170, -4.682)
Seg14	-1.999	(-3.467, 0.116)
Seg15	-1.044	(-2.409, 0.109)
Seg17	-0.993	(-3.647, 1.263)
Seg19	-9.579	(-24.416, -0.539)
Seg22	-1.697	(-5.863, 1.753)
Seg23	-5.238	(-9.652, -1.452)
Seg27	-1.057	(-3.672, 1.614)
Year96	-0.092	(-2.761, 2.521)
Year97	-0.064	(-4.104, 3.782)
BT	-10.147	(-20.560, -1.724)
DTN	-0.775	(-4.269, 2.797)
EF	-1.541	(-4.103, 1.532)
Depth	-1.660	(-5.168, 1.385)
Cond	-9.991	(-23.760, -1.397)
LogTurb	-1.601	(-6.652, 2.142)
Gravel	-5.299	(-8.657, -2.061)
Sand	-1.088	(-2.422 , 0.457)
WaterTemp	2.038	(-0.369 , 4.013)
Velocity	-0.843	(-4.528, 2.362)

Table 16: Model 1- Posterior mean and 95% Credible Intervals (CIs) for the logistic regression model parameters for bigmouth buffalo

Parameter	**Posterior Mean**	95% **CI**
intercept	2.871	(-7.518 , 11.340)
Bend	9.084	(-5.818 , 21.414)
SCC	-13.507	(-27.340 , 3.998)
SCN	0.596	(-15.791 , 24.480)
Seg3	2.763	(-17.230 , 22.291)
Seg5	-5.217	(-22.290 , 10.640)
Seg7	-1.355	(-19.870 , 13.170)
Seg8	2.160	(-13.831 , 17.720)
Seg9	-3.176	(-20.611 , 12.970)
Seg10	-3.250	(-20.370 , 12.320)
Seg12	1.229	(-19.650 , 16.120)
Seg14	-2.199	(-18.780 , 11.920)
Seg15	0.020	(-17.450 , 16.450)
Seg17	-2.871	(-17.390 , 14.100)
Seg19	-0.138	(-19.190 , 17.290)
Seg22	1.859	(-16.790 , 23.860)
Seg23	2.167	(-18.861 , 20.140)
Seg27	1.673	(-15.432 , 23.481)
BT	3.685	(-17.661 , 22.831)
DTN	-0.096	(-19.730 , 23.640)
EF	-7.463	(-27.730 , 9.381)
Depth	-16.652	(-27.850 , -8.463)
Cond	-0.566	(-5.526 , 3.706)
LogTurb	-2.354	(-5.725 , 0.441)
Sand	3.484	(-15.260 , 20.561)
Gravel	7.154	(-10.230 , 21.960)
WaterTemp	0.229	(-0.942 , 0.960)
Velocity	-3.732	(-19.371 , 13.660)

Table 17: Model 2- Posterior mean and 95% Credible Intervals (CIs) for the log-linear model parameters for smallmouth buffalo

Parameter	**Posterior Mean**	95% **CI**
intercept	-15.648	(-18.200, -12.989)
Bend	-2.616	(-3.934, -1.093)
SCC	-2.246	(-3.865, 0.017)
SCN	8.926	(4.704, 12.430)
Seg3	4.865	(2.268, 7.361)
Seg5	-4.452	(-21.060, 6.750)
Seg7	0.378	(-2.954, 3.800)
Seg8	-8.314	(-23.190, 3.791)
Seg9	-1.183	(-4.066, 0.844)
Seg10	5.389	(1.774, 9.450)
Seg12	6.969	(-3.421, 10.920)
Seg14	1.502	(-0.705, 3.792)
Seg15	-6.254	(-20.900, 5.283)
Seg17	4.622	(1.257, 7.741)
Seg19	5.208	(2.143, 8.143)
Seg22	-2.759	(-20.101, 4.991)
Seg23	-4.818	(-19.831, 5.833)
Seg27	5.269	(1.675, 9.080)
Year96	-4.359	(-7.707, -1.797)
Year97	-0.153	(-1.272, 0.944)
BT	-10.583	(-19.980, -4.895)
DTN	-8.892	(-14.180, -4.228)
EF	-2.752	(-5.894, 0.507)
Depth	0.085	(-0.693, 0.729)
Cond	-1.481	(-2.198, -0.735)
LogTurb	-0.044	(-0.536, 0.317)
Gravel	-4.106	(-6.515, -1.616)
Sand	0.075	(-1.489, 1.377)
WaterTemp	0.327	(0.260, 0.406)
Velocity	-0.047	(-2.642, 2.039)

Table 18: Model 2- Posterior mean and 95% Credible Intervals (CIs) for the logistic regression model parameters for smallmouth buffalo

Parameter	**Posterior Mean**	95% **CI**
intercept	-14.145	(-31.660, -0.640)
Bend	-10.059	(-17.920, -0.882)
SCC	-6.679	(-12.740, -0.705)
SCN	13.331	(3.909, 24.352)
Seg3	-5.972	(-13.961, 7.425)
Seg5	4.798	(-10.301, 21.520)
Seg7	-11.494	(-23.131, -1.246)
Seg8	3.471	(-17.522, 18.920)
Seg9	-6.301	(-18.860, 1.844)
Seg10	0.057	(-8.201, 9.428)
Seg12	12.152	(-0.025, 23.142)
Seg14	-8.902	(-20.050, -0.786)
Seg15	5.305	(-12.841, 28.511)
Seg17	-5.699	(-14.900, 5.379)
Seg19	-2.383	(-10.210, 7.527)
Seg22	2.458	(-9.583, 15.181)
Seg23	1.544	(-18.021, 23.951)
Seg27	3.895	(-3.636, 14.610)
BT	0.429	(-21.880, 16.450)
DTN	-5.635	(-20.251, 6.897)
EF	-6.609	(-12.850, -2.517)
Depth	-3.510	(-6.428, -1.031)
Cond	-7.841	(-13.080, -4.984)
LogTurb	-0.257	(-1.247, 1.057)
Gravel	-7.417	(-19.051, -0.486)
Sand	1.859	(-3.067, 7.889)
WaterTemp	1.488	(1.009, 2.280)
Velocity	-14.524	(-31.303, -5.544)

Table 19: Model 2- Posterior mean and 95% Credible Intervals (CIs) for the log-linear model parameters for bigmouth buffalo

Parameter	**Posterior Mean**	95% **CI**
intercept	-10.583	(-19.980 , -4.895)
Bend	-8.892	(-14.180 , -4.228)
SCC	-2.752	(-5.894 , 0.507)
SCN	0.085	(-0.693 , 0.729)
Seg3	-1.481	(-2.198 , -0.735)
Seg5	-0.044	(-0.536 , 0.317)
Seg7	-0.047	(-2.642 , 2.039)
Seg8	0.327	(0.260 , 0.406)
Seg9	-4.106	(-6.515 , -1.616)
Seg10	0.075	(-1.489 , 1.377)
Seg12	-20.624	(-22.860 , -18.520)
Seg14	8.277	(4.690 , 11.750)
Seg15	0.355	(-1.928 , 3.106)
Seg17	1.307	(-2.105 , 4.985)
Seg19	-1.405	(-19.800 , 14.940)
Seg22	-1.625	(-19.860 , 14.050)
Seg23	-3.771	(-20.462 , 11.180)
Seg27	-3.033	(-15.490 , 7.375)
Year96	12.781	(10.549 , 14.820)
Year97	11.941	(-3.020 , 21.290)
BT	-6.221	(-18.680 , 6.693)
DTN	-0.338	(-1.472 , 1.148)
EF	-4.661	(-9.448 , 1.808)
Depth	-1.257	(-19.752 , 13.850)
Cond	-3.508	(-21.830 , 13.690)
LogTurb	-2.086	(-19.731 , 14.672)
Gravel	-14.579	(-20.732 , -8.054)
Sand	-2.712	(-4.665 , -0.695)
WaterTemp	20.077	(17.640 , 22.090)
Velocity	-3.012	(-20.990 , 12.250)

Table 20: Model 2- Posterior mean and 95% Credible Intervals (CIs) for the logistic regression model parameters for bigmouth buffalo

Parameter	**Posterior Mean**	95% **CI**
intercept	-6.609	(-12.850, -2.516)
Bend	-3.510	(-6.428, -1.032)
SCC	-7.841	(-13.080, -4.984)
SCN	-0.257	(-1.247, 1.057)
Seg3	-14.524	(-31.303, -5.544)
Seg5	1.488	(1.009, 2.280)
Seg7	-7.418	(-19.051, -0.486)
Seg8	1.859	(-3.067, 7.889)
Seg9	-4.701	(-19.391, 4.042)
Seg10	10.077	(-1.339, 19.870)
Seg12	-5.858	(-21.461, 3.244)
Seg14	2.294	(-9.110, 11.850)
Seg15	0.498	(-22.740, 20.490)
Seg17	-1.236	(-18.910, 16.910)
Seg19	2.306	(-13.320, 16.792)
Seg22	-4.685	(-25.412, 6.871)
Seg23	7.529	(-1.760, 16.090)
Seg27	-0.053	(-14.070, 12.900)
BT	-3.849	(-20.311, 7.681)
DTN	-10.416	(-26.470, 2.595)
EF	-8.460	(-19.390, 2.992)
Depth	3.473	(-9.566, 22.280)
Cond	5.229	(-12.400, 21.240)
LogTurb	2.899	(-15.182, 21.630)
Gravel	4.741	(-19.410, 21.740)
Sand	12.471	(-12.320, 29.081)
WaterTemp	3.842	(-3.549, 14.270)
Velocity	3.308	(-15.641, 26.590)

Table 21: Model comparison results for Model 1 (multivariate normal) and Model 2 (multivariate Poisson) based on posterior predictive loss criteria

Species	Model	G	P	D_1
Smallmouth buffalo	Multivariate Poisson	108.65	415.17	469.49
Bigmouth buffalo	Multivariate Poisson	77.56	585.46	624.24
Smallmouth buffalo	Multivariate Normal	48.58	465.84	490.13
Bigmouth buffalo	Multivariate Normal	12.69	590.21	596.55

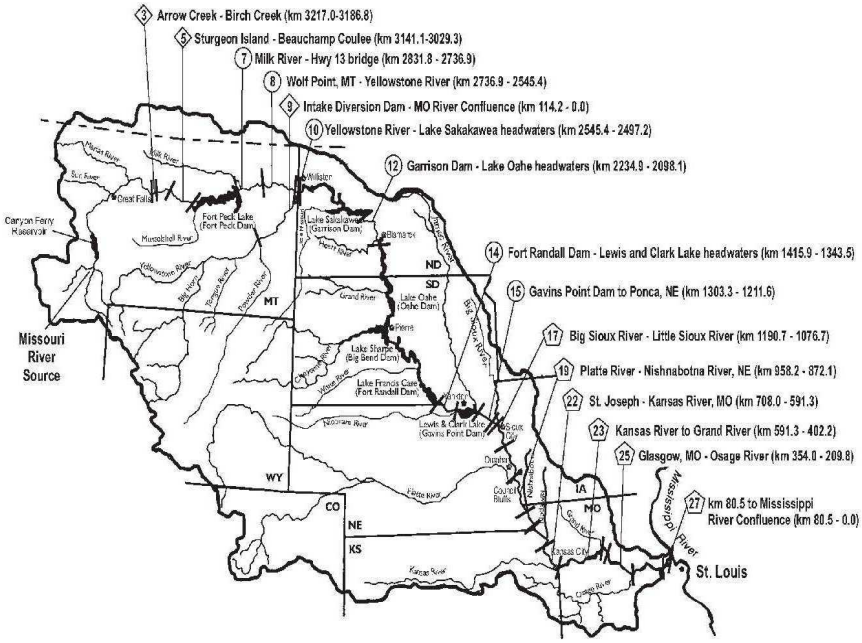


Figure 28: Missouri River Benthic Fishes Study area from Montana to its confluence with the Mississippi River in Missouri ($\diamondsuit$ = Least-Altered (LA), $\bigcirc$ = Inter-Reservoir (IR), Channelized (CH) Segments).

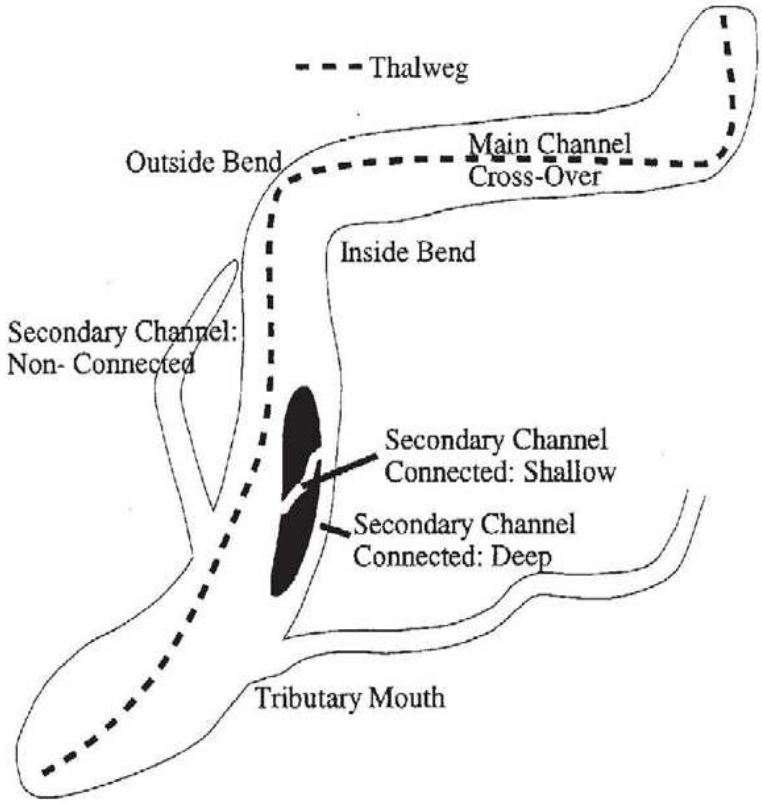


Figure 29: Schematic showing macrohabitats sampled during the Benthic Fishes Study.

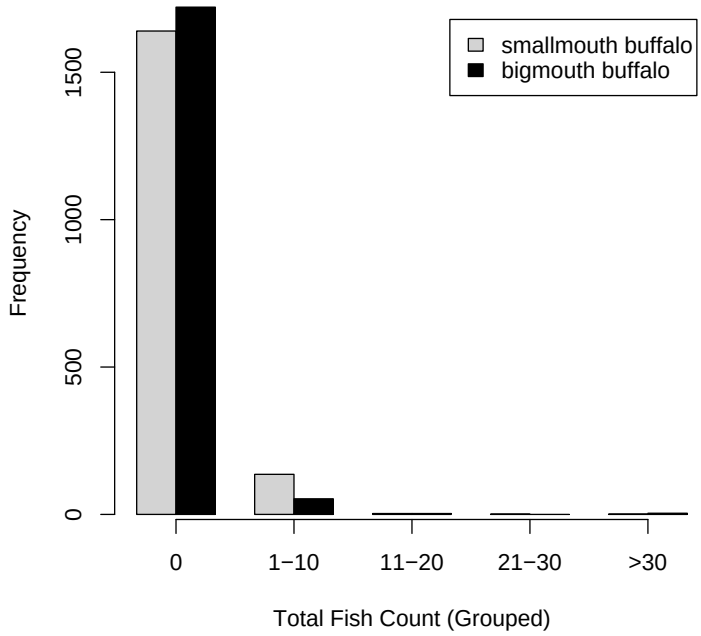
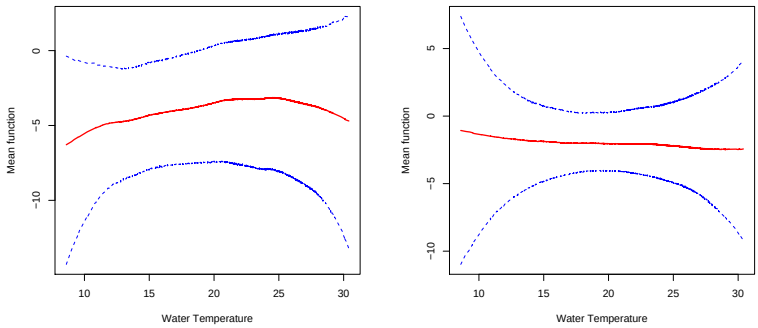
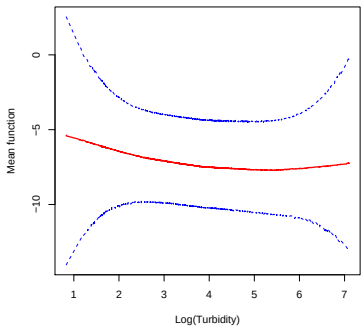


Figure 30: Frequency of total fish count for smallmouth buffalo and bigmouth buffalo (grouped).

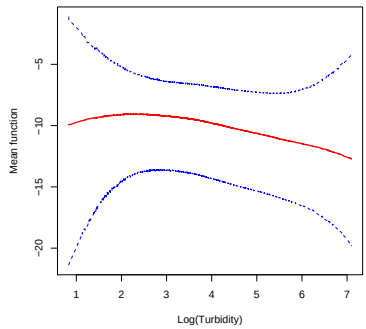


(a)

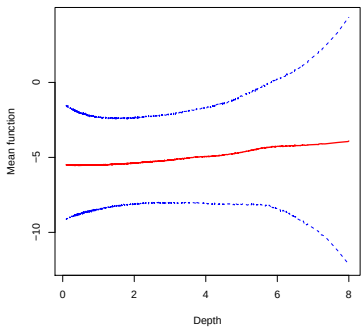
(b)



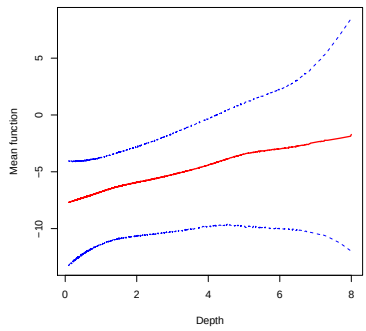
(c)



(d)

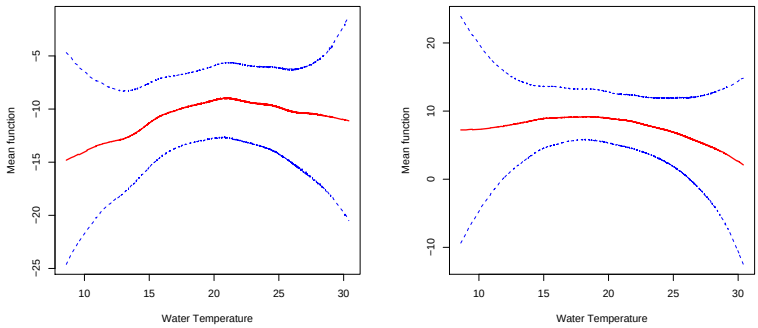


(e)



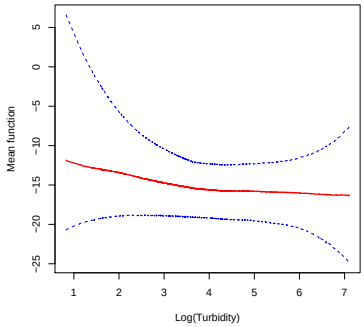
(f)

Figure 31: Model 1 (Multivariate Log-Normal)-Thin-plate spline smooth function fit (solid line) and 95% CIs (dashed lines) for water temperature for both species: (a) Water temperature (smallmouth buffalo), (b) Water temperature (bigmouth buffalo), (c) Log(Turbidity) (smallmouth buffalo), (d) Log(Turbidity) (bigmouth buffalo), (e) Depth (smallmouth buffalo), and (f) Depth (bigmouth buffalo).

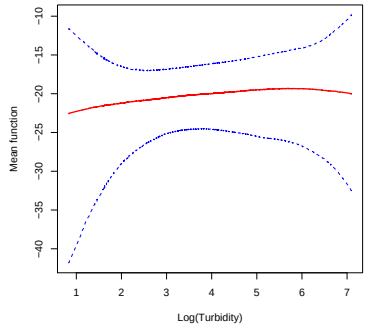


(a)

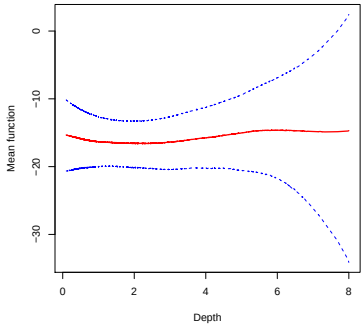
(b)



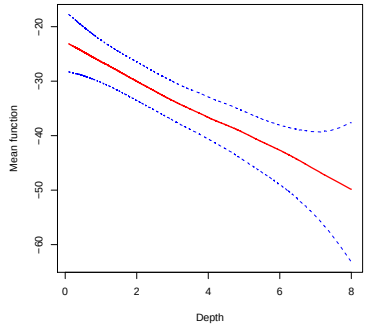
(c)



(d)



(e)



(f)

Figure 32: Model 2 (Multivariate Poisson)-Thin-plate spline smooth function fit (solid line) and 95% CIs (dashed lines) for water temperature for both species: (a) Water temperature (smallmouth buffalo), (b) Water temperature (bigmouth buffalo), (c) Log(Turbidity) (smallmouth buffalo), (d) Log(Turbidity) (bigmouth buffalo), (e) Depth (smallmouth buffalo), and (f) Depth (bigmouth buffalo).

4 Semiparametric Spatio-Temporal Zero-Inflated Models for Multivariate Count Processes

4.1 Introduction

Spatio-temporal count observations obtained from environmental and ecological processes are often characterized by zero-inflation due to sampling complexities involved in such surveys. The application of zero-inflated modeling for spatio-temporal processes is recent and there are only a few examples of spatio-temporal extensions of zero-inflated modeling available in the literature (e.g., Wikle and Anderson 2003; Fernandes et al. 2006).

Another important characteristic of environmental and ecological problems is that often interaction and intricate relationships among different sub-processes (i.e., correlated zero-inflated count processes) exist. These interdependencies among the sub-processes are often ignored in most modeling attempts. To address these issues, in this chapter, the semi-parametric zero-inflated modeling of multivariate count processes described in the previous chapter, is extended to spatio-temporal count processes. Finally, an application to jointly modeling "less damaging" and "more damaging" tornado counts is presented, followed by a discussion of the results and general conclusions.

The method discussed in this chapter is a novel method for modeling of multivariate spatio-temporal count processes with excess zeros while accounting for non-linear effects of covariate information. The methodology presented in this chapter is new and no other examples of such modeling attempts can be found in the statistical modeling literature. Such a modeling approach can be very useful for environmental and ecological applications where count processes often exhibit characteristics such as excess zeros and non-linear

relationships.

4.2 Modeling Approach

In this section, a spatio-temporal extension of a multivariate zero-inflated Poisson distribution, as described in the previous chapter, is considered. Thus,

$$Y_{1t} = Z_{1t} + Z_{3t},$$

and,

$$Y_{2t} = Z_{2t} + Z_{3t},$$

for $t = 1, \ldots, T$ where $(Y_{1t}, Y_{2t}) \sim \text{BivPoisson}(\lambda_{1t}, \lambda_{2t}, \lambda_{3t})$, Z_{1t}, Z_{2t}, and Z_{3t} are mutually independent Poisson random variables with mean and variances λ_{1t}, λ_{2t}, and λ_{3t}, respectively. Thus, the covariance between Y_{1t} and Y_{2t} is

$$\text{cov}(Y_{1t}, Y_{2t}) = \text{cov}(Z_{1t} + Z_{3t}, Z_{2t} + Z_{3t}) = \text{var}(Z_{3t}) = \lambda_{3t},$$

for $t = 1, \ldots, T$.

To construct a multivariate spatio-temporal zero-inflated Poisson model let Z_{1t}, Z_{2t}, and (possibly) Z_{3t} denote mutually independent random variables from a zero-inflated Poisson distribution. Thus,

$$Z_{1t} \sim \begin{cases} 0 & \text{w.p. } 1 - p_{1t}, \\ \text{Poisson}(\lambda_{1t}) & \text{w.p. } p_{1t}, \end{cases}$$

$$Z_{2t} \sim \begin{cases} 0 & \text{w.p. } 1 - p_{2t}, \\ \text{Poisson}(\lambda_{2t}) & \text{w.p. } p_{2t}, \end{cases}$$

and,

$$Z_{3t} \sim \begin{cases} 0 & \text{w.p. } 1 - p_{3t}, \\ \text{Poisson}(\lambda_{3t}) & \text{w.p. } p_{3t}, \end{cases}$$

where p_{1t}, p_{2t}, and p_{3t} denote the zero-inflation probability for Z_{1t}, Z_{2t}, and Z_{3t}, respectively. Thus, $Y_{1t} = Z_{1t} + Z_{3t}$ and $Y_{2t} = Z_{2t} + Z_{3t}$ follow a bivariate Poisson distribution with excess zeros. The common latent process, Z_{3t}, can be assumed to follow either a standard Poisson distribution or a zero-inflated Poisson (ZIP) distribution. However, if one considers the case where Z_{3t} is assumed to follow a ZIP distribution, care should be given in parameterizing the model as the estimation of the zero-inflation probability might be impossible where there are no covariates that can be linked to the common zero-inflation process. A semiparametric approach to modeling multivariate ZIP as described in the previous chapter can be easily extended to the spatio-temporal case. Another useful semiparametric approach in spatio-temporal settings is to take advantage of nonparametric techniques in a dimension reduction capacity. This idea has been implemented for an application of the semiparametric spatio-temporal multivariate ZIP model which will be discussed in the next section.

4.3 Application: Modeling Tornado Counts

In this section, an application of semiparametric spatio-temporal zero-inflated modeling of multivariate count processes is considered for a "real world" problem in climatological analysis of tornado report counts. The analysis is based on tornado reports submitted to the National Weather Service (NWS) over years 1953-1995. These data are archived at the National Oceanic and Atmospheric Administration (NOAA) Storm Prediction Center (SPC). Each tornado report includes the time and location of the initial and final observation of the tornado, estimated tornado path width, damage rating, estimated dollar amount of damage produces by the tornado, and other forms of severe weather observed in the immediate vicinity of the tornado. A 2400 km $\times$ 1700 km grid of 50-km boxes was overlaid

on the continental U.S. with the upper left grid box centered at approximately 45.5° N, 105° W, resulting in 2160 grid boxes. Only tornado report counts east of 105° W were considered, due to very infrequent occurrence of tornadoes within the Rocky Mountains and along the west coast. For every year, the number of observations at the initial positions of all tornadoes (F0-F5) that occurred within each grid box were tallied. The "F", which stands for "Fujita scale", is a rating system for tornado damage where F0 represents "minimal damage" and F5 represents "complete destruction".

Wikle and Anderson (2003) analyzed these data using a spatio-temporal zero-inflated Poisson model, due to a significant number of zero counts over the domain (approximately 90% of the observations are zero). Using this modeling approach, they found evidence of a significant temporal trend with spatial variation in the data and a significant association of an index of El Niño activity with the tornado count reports over the continental U.S., with substantial regional variability. Figure 33, shows a time series for the annual index of the El Niño/Southern Oscillation (ENSO) phenomenon, as given by the Niño3.4 index (Trenberth 1997), for years 1953-2001. The approach utilized in Wikle and Anderson (2003) is extended to the case of multivariate count processes in order to conduct the analysis based on two separate zero-inflated count processes for "less damaging" tornadoes (F0-F1) and "more damaging" tornadoes (F2-F5), while accounting for possible correlation among these processes. Accounting for such a correlation seems both necessary and realistic since the presence of common climatological phenomena and processes that impact the development of all (less damaging and more damaging) tornadoes is very possible (however the nature of such associations might differ). The assumption of zero-inflation is necessary since both F0-

F1 and F2-F5 tornado report counts include a high percentage of zero values (approximately 83% zeros for F0-F1 tornado counts and approximately 93% zeros for F2-F5 tornado counts).

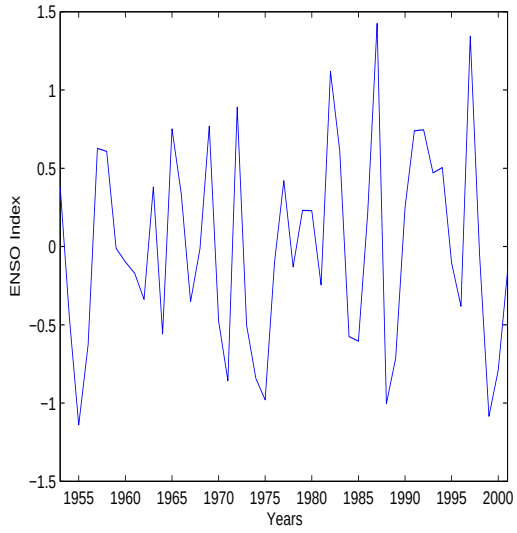


Figure 33: Time series for the annual index of the El Niño/Southern Oscillation (ENSO) phenomenon for years 1953-2001.

4.4 Model

In this section a semiparametric hierarchical Bayesian modeling framework will be considered to model the two groups of tornado counts (F0-F1 and F2-F5) jointly. In order to model the effect of ENSO on these tornado counts, a threshold modeling approach (Tong 1993) will be considered for the log-linear regression component of the model. In the logistic regression part of the model for the zero-inflation probability, population for each observation is used as a covariate. This was motivated by a previous study (Anderson et al. 2007) where the effect of population on identifying tornadoes was found to be influential.

4.4.1 Data Model

Let $\mathbf{Y}_{1,t} = \mathbf{Z}_{1,t} + \mathbf{Z}_{3,t}$ indicate F0-F1 tornado counts at time t and $\mathbf{Y}_{2,t} = \mathbf{Z}_{2,t} + \mathbf{Z}_{3,t}$ indicate F2-F5 tornado counts at time t where $\mathbf{Y}_{j,t} = (Y_{j,1}(s_1), \ldots, Y_{j,T}(s_n))'$ for $j = 1, 2$, $\mathbf{Z}_{k,t} = (Z_{k,1}(s_1), \ldots, Z_{k,T}(s_n))'$ for $k = 1, 2, 3$, for locations $s_1, \ldots, s_n$ and times $t = 1, \ldots, T$. Here, $\mathbf{Z}_{1,t}$ and $\mathbf{Z}_{2,t}$ are assumed to follow a zero-inflated distribution with intensity $\boldsymbol{\lambda}_{j,t} = (\lambda_{j,1}(s_1), \ldots, \lambda_{j,T}(s_n))'$, and zero-inflation probability $\mathbf{p}_{j,t} = (p_{j,1}(s_1), \ldots, p_{j,T}(s_n))'$, $j = 1, 2$ (i.e., $\mathbf{Z}_{j,t} \sim ZIP(\boldsymbol{\lambda}_{j,t}, \mathbf{p}_{j,t})$). In this problem, $\mathbf{Z}_{3,t}$ is assumed to follow a Poisson distribution with intensity $\boldsymbol{\lambda}_{3,t} = (\lambda_{3,1}(s_1), \ldots, \lambda_{3,T}(s_n))'$. Thus,

$$Z_{j,it}|\lambda_{j,it}, p_{j,it} \sim \begin{cases} \text{Poisson}(\lambda_{j,it}), & \text{w.p.} \quad p_{j,it}, \\ 0 & \text{w.p.} \quad 1 - p_{j,it}, \end{cases} \tag{4.43}$$

and,

$$Z_{3,it}|\lambda_{3,it} \sim \text{Poisson}(\lambda_{3,it}) \tag{4.44}$$

for $i = 1, \ldots, n$, $j = 1, 2$, and $t = 1, \ldots, T$.

4.4.2 Process Models

The process models include log-linear threshold regression model for Poisson intensity and logistic regression for the zero-inflation probability. The log-linear regression model for

$$\log(\boldsymbol{\lambda}_{j,t}) = \boldsymbol{\mu}_j + \boldsymbol{\beta}_{j,0}t + \begin{cases} \boldsymbol{\beta}_{j,1}x_t + \boldsymbol{\varepsilon}_{j,1t}, & x_t \geq c1 \\ \boldsymbol{\beta}_{j,2}x_t + \boldsymbol{\varepsilon}_{j,2t}, & c1 < x_t \leq c2 \\ \boldsymbol{\beta}_{j,3}x_t + \boldsymbol{\varepsilon}_{j,3t}, & x_t > c2 \end{cases} \tag{4.45}$$

for $t = 1, \ldots, T$, $\boldsymbol{\beta}_{j,i} = (\beta_{j,i}(s_1), \ldots, \beta_{j,i}(s_n))'$ for $i = 0, \ldots, 3$, and $\boldsymbol{\varepsilon}_{j,kt} \sim N(0, \sigma^2_{\varepsilon}\mathbf{I})$, for $j = 1, 2, 3$ and $k = 1, 2, 3$. Constants c_1 and c_2 are 33.3rd and 66.7th percentiles of the ENSO time series.

The threshold modeling approach is considered due to known non-linear behavior of the atmospheric response to ENSO which consists of the episodic warming and cooling of ocean waters with periods of approximately 3-5 years (Hannachi 2001; DeWeaver and Nigam 2002). Such time-varying behavior of ENSO phenomenon could effectively be addressed using a threshold modeling approach as considered in this section. The ENSO time series shown in Figure 33 consists of three different states of the ENSO phenomenon with highest values corresponding to El Niño, the lowest values corresponding to La Niña, and the middle range of values corresponding to normal seasons where neither El Niño nor La Ninña were occurring.

A semiparametric regression modeling approach based on thin-plate spline functions is considered in order to significantly reduce dimensionality for the model. Thus, the spatially-indexed regression parameters, $\boldsymbol{\beta}_{j,i}$ ($j = 1, 2, 3$ and $i = 0, \ldots, 3$) can be written as

$$\boldsymbol{\beta}_{j,i} = \Phi \mathbf{b}_{j,i}, \tag{4.46}$$

where Φ is an $n \times k$ matrix of thin plate spline bases, and k is chosen to be much smaller than the model dimensionality ($k \ll n = 2160$) which provides a method to reduce dimensionality in the model. In this problem $k = 20$ is used, but $k = 50$ was also considered to assess sensitivity analysis of the model to the choice of k. Knot selection was implemented using the *Clara* algorithm in the R *SemiPar* package.

The logistic regression for the zero-inflation probability can be written as

$$\text{logit}(\mathbf{p}_{j,t}) = \alpha_{j,0} + \mathbf{x}_p\alpha_j + \boldsymbol{\eta}_j, \qquad \boldsymbol{\eta}_j \sim N(0, \sigma^2_{j,\eta}\mathbf{I}) \tag{4.47}$$

where $\mathbf{x}_p$ denotes population (for each grid box) and $j = 1, 2$. Note that there is a known identifiability issue with considering a semiparametric model for the zero-inflation probability (see Fahrmeir and Echavarria 2006).

4.4.3 Parameter Models

The hierarchical Bayesian framework considered requires the assignment of prior distribution for all the unknown parameters. The following relatively vague prior distributions are considered for the unknown parameters:

$$\begin{aligned}
\mathbf{b}_{j,i} &\sim N(\mathbf{0}, 10 \times \mathbf{I}),\\
\boldsymbol{\mu}_j &\sim N(\mathbf{0}, 100 \times \mathbf{I}),\\
\alpha_{k,0} &\sim N(0, 100),\\
\alpha_{k,1} &\sim N(0, 100),\\
\tau_{j,\varepsilon_i} \equiv 1/\sigma^2_{j,\varepsilon_i} &\sim Gamma(\text{mean} = 1, \text{var} = 10),\\
\tau_{k,\eta} \equiv 1/\sigma^2_{k,\eta} &\sim Gamma(\text{mean} = 1, \text{var} = 10),
\end{aligned}$$

for $j = 0, \ldots, 3$ and $k = 1, 2$. All the hyperparameters are considered to be known as indicated above.

4.4.4 Model Implementation

The joint posterior distribution of the hierarchical model described above, cannot be evaluated analytically and numerical simulation methods must be used. This numerical simula-

tion can be done using MCMC methods and in particular Gibbs sampler with Metropolis-Hastings steps can be employed to generate samples from the posterior distribution. The posterior distribution of all the processes and parameters given the observations, and full-conditional distributions are shown in Appendix B. The MCMC strategy for threshold model is similar to the method described in Carlin et al. (1992). Metropolis-Hastings steps are needed within the Gibbs algorithm to generate samples from the conditional distributions for $\boldsymbol{\lambda}$'s and $\mathbf{p}$'s.

The MCMC simulations were run for 100,000 iterations from which the first 20,000 iterations were ignored as "burn-in". The convergence of the MCMC algorithm was checked by visual inspection of the chains.

4.5 Results

The results for spatially-varying coefficients of the log-linear models are shown in Figures 34-37. These results are based on MCMC posterior means. Figure 37 shows the posterior means and standard deviations for the $\boldsymbol{\mu}$ spatial process for both F0-F1 and F2-F5 tornadoes. Figure 38 shows the results for the parameters of the log-linear model for the common Poisson latent process. Figure 39 shows the spatial pattern of zero-inflation probability as presence (probability > 0.5)-absence (probability< 0.5)) over the domain of the study.

The results for the log-linear models for the Poisson intensities show substantially different spatial patterns for both categories. For example, the normal and El Niño seasons seem to have reverse effects on the spatial variability of tornado counts, while the effect of the La Niña seasons show similar spatial patterns for both categories of tornadoes. Table 22 shows the posterior means for the univariate parameters. The effect of population does not

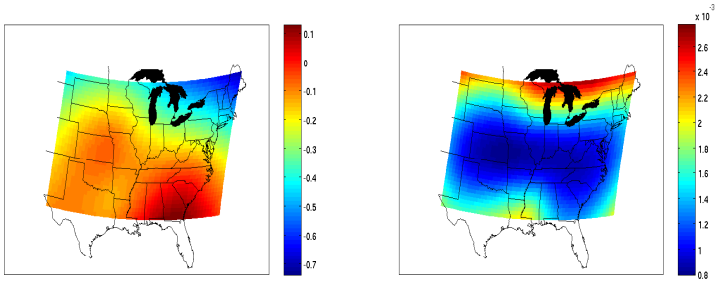
seem to be significant for either of the tornado categories.

Table 22: Posterior mean and 95% Credible Intervals (CIs) for the univariate model parameters

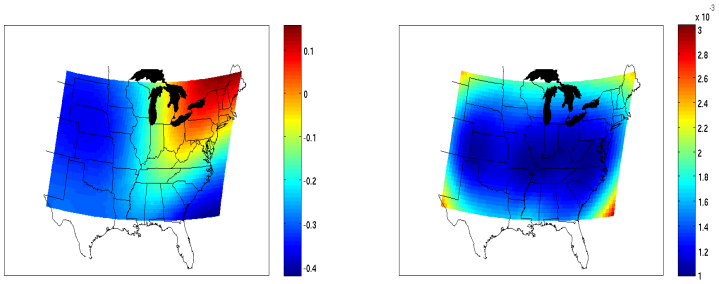
Parameter	**Posterior Mean**	95% **CI**
$\alpha_{1,0}$	-0.46	(-1.9,0.56)
$\alpha_{1,1}$	-0.04	(-0.79, 0.67)
$\alpha_{2,0}$	-0.85	(-2.65, 0.43)
$\alpha_{2,1}$	0.24	(-0.52, 0.98)
$\sigma^2_{1,1,\varepsilon}$	0.874	(0.79, 0.96)
$\sigma^2_{1,2,\varepsilon}$	1.004	(0.87, 1.16)
$\sigma^2_{1,3,\varepsilon}$	1.006	(0.89, 1.12)
$\sigma^2_{2,1,\varepsilon}$	0.82	(0.75, 0.88)
$\sigma^2_{2,2,\varepsilon}$	1.8	(1.64, 1.96)
$\sigma^2_{2,3,\varepsilon}$	1.94	(1.76, 2.12)
$\sigma^2_{1,\eta}$	1.93	(1.23, 2.68)
$\sigma^2_{2,\eta}$	1.68	(1.01, 2.43)

4.6 Discussion

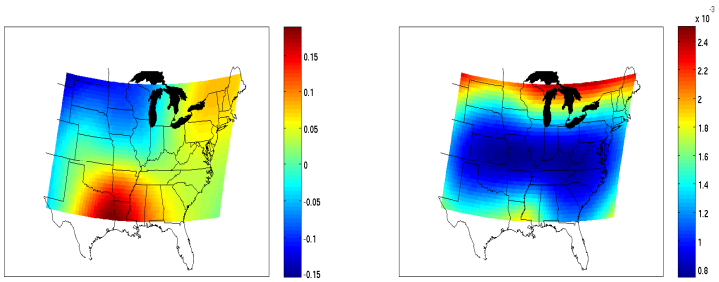
The results for the tornado climatological effects are very smooth due to the low number of spline knot points selected for the model. However, these estimates enable us to get a sense of the large scale spatial patterns of tornado counts with regards to these effects. For example, Figures 36(a) and (c) show substantially different temporal patterns for the F0-F1 and F2-F5 tornado counts. The spatial field for the posterior means of the temporal trend of F0-F1 tornado counts show an increasing trend in the west plains, slight increase in the Great Lakes area and nearly constant everywhere else (Figure 36(a)). The spatial field for the posterior means of the temporal trend of F2-F5 tornado counts shows increasing trend



(a) (b)



(c) (d)



(e) (f)

Figure 34: Posterior means and standard deviations for F0-F1 tornadoes: (a) La Niña seasons ($\boldsymbol{\beta}_1$) (posterior mean), (b) La Niña seasons ($\boldsymbol{\beta}_1$) (posterior standard deviation), (c) Normal seasons ($\boldsymbol{\beta}_2$) (posterior mean), (d) Normal seasons ($\boldsymbol{\beta}_2$) (posterior standard deviation), (e) El Niño seasons ($\boldsymbol{\beta}_3$) (posterior mean), (f) El Niño seasons ($\boldsymbol{\beta}_3$) (posterior standard deviation).

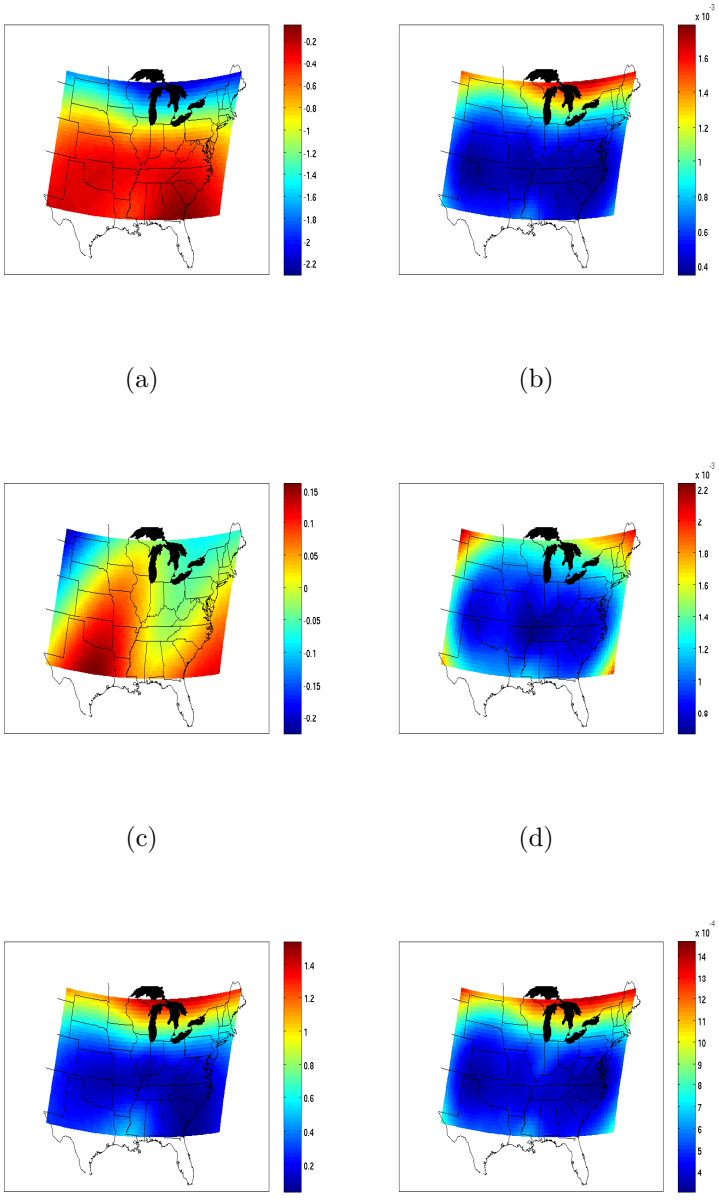
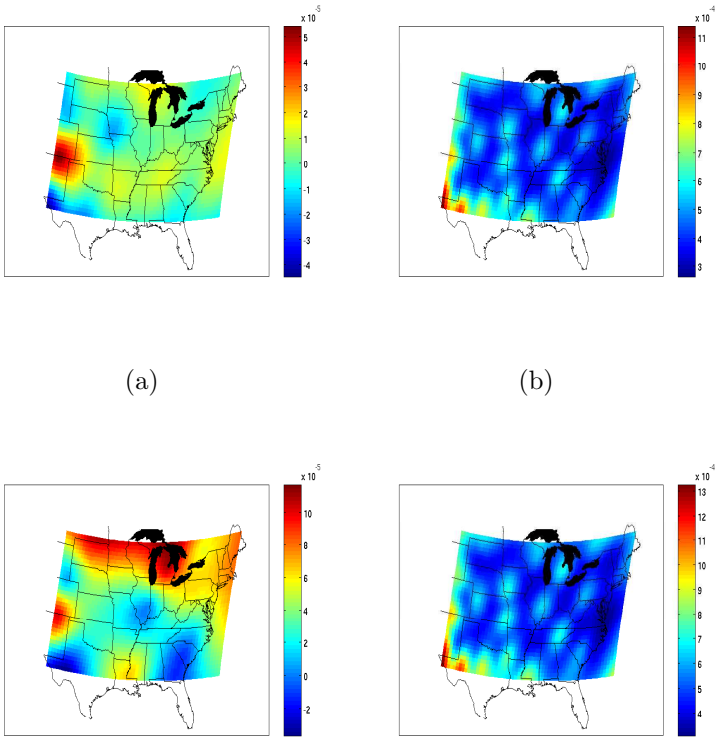


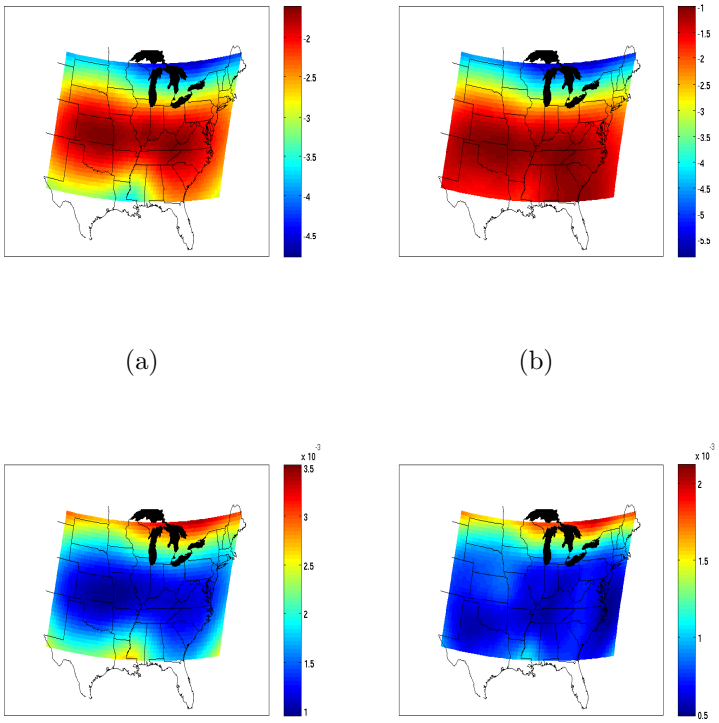
Figure 35: Posterior means and standard deviations for F2-F5 tornadoes: (a) La Niña seasons ($\boldsymbol{\beta}_1$) (posterior mean), (b) La Niña seasons ($\boldsymbol{\beta}_1$) (posterior standard deviation), (c) Normal seasons ($\boldsymbol{\beta}_2$) (posterior mean), (d) Normal seasons ($\boldsymbol{\beta}_2$) (posterior standard deviation), (e) El Niño seasons ($\boldsymbol{\beta}_3$) (posterior mean), (f) El Niño seasons ($\boldsymbol{\beta}_3$) (posterior standard deviation).



(a) (b)

(c) (d)

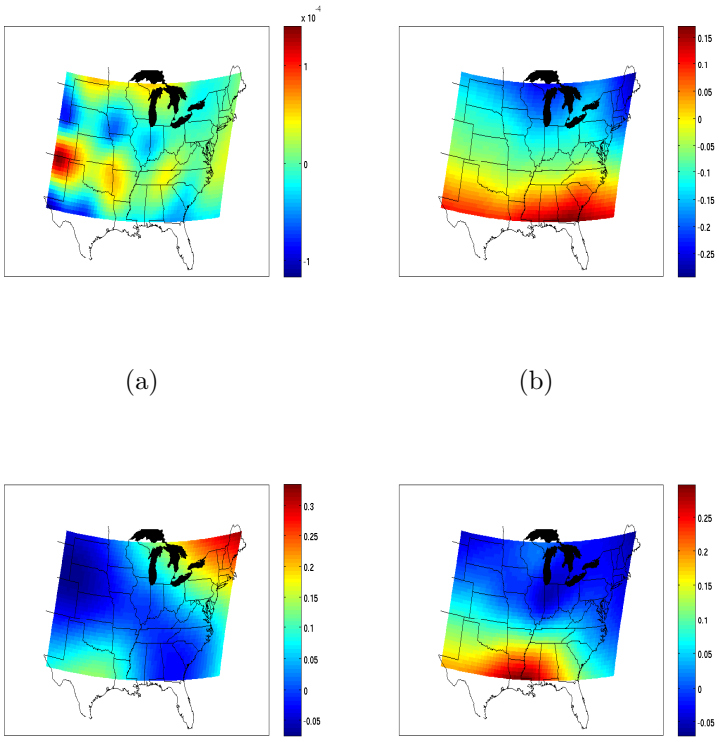
Figure 36: Posterior means and standard deviations for linear trend : (a) F0-F1 ($\boldsymbol{\beta}_0$) (posterior mean), (b) F0-F1 ($\boldsymbol{\beta}_0$) (posterior standard deviation), (c) F2-F5 ($\boldsymbol{\beta}_0$) (posterior mean), (d) F2-F5 ($\boldsymbol{\beta}_0$) (posterior standard deviation).



(a) (b)

(c) (d)

Figure 37: MCMC results for $\boldsymbol{\mu}$ spatial process for: (a) F0-F1 ($\boldsymbol{\mu}_1$) (posterior mean), (b) F2-F5 ($\boldsymbol{\mu}_2$) (posterior mean), (c) F0-F1 ($\boldsymbol{\mu}_1$) (posterior standard deviation), (d) F2-F5 ($\boldsymbol{\mu}_2$) (posterior standard deviation)

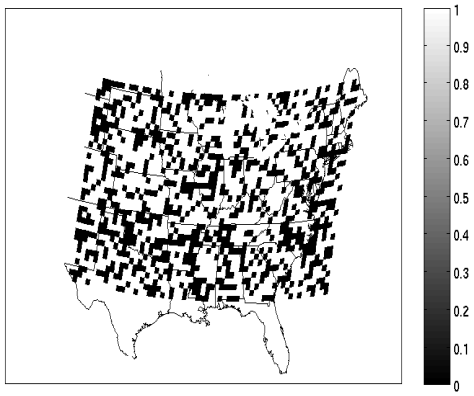


(a) (b)

(c) (d)

Figure 38: Posterior means for the common Poisson latent process ($\mathbf{Z}_3$): (a) Linear trend ($\boldsymbol{\beta}_0$), (b) La Niña seasons ($\boldsymbol{\beta}_1$), (c) Normal seasons ($\boldsymbol{\beta}_2$),(d) El Niño seasons ($\boldsymbol{\beta}_3$).

(a)



(b)

Figure 39: Presence-absence map of zero-inflation probabilities for years 1953-2001. Value 1 indicates zero-inflation probability higher than 0.5 and value 0 indicates zero-inflation probability lower than 0.5: (a) F0-F1, (b) F2-F5.

in the west plains (similar to the trend for F0-F1), increasing trend in the northern part of the domain, and also slight increase in the southern part of the domain (Figure 36(c)).

Figures 34 and 35 show climatological effects of ENSO on tornado counts for F0-F1 and F2-F5 tornadoes, respectively. Comparison of Figures 34(a) and 35 (a) show a similar spatial pattern of increasing tornado counts for both groups during La Niña seasons with a general north to south variation. However the intensity of F0-F1 tornadoes is substantially higher. This indicates that overall, during La Niña seasons, higher numbers of tornadoes occur in the southern part of the U.S. Figures 34(e) and 35 (e) indicate substantially different patterns for F0-F1 and F2-F5 tornadoes during El Niño seasons. For F0-F1 tornadoes, during El Niño seasons, more tornadoes occur in the south (Texas and Louisiana) and the northeast U.S. For F2-F5 tornadoes, more tornadoes are observed in the northern part of the U.S. during El Niño seasons. The spatial patterns of tornado counts during normal seasons (i.e., no El Niño or La Niña) is substantially different for F0-F1 and F2-F5 tornadoes with higher F2-F5 tornado counts in the south, southeast and midwest, and higher F0-F1 tornado counts in the northeast. Most of the general results are in agreement with previous studies (Wikle and Anderson 2003). However, such modeling perspective for studying F0-F1 and F2-F5 tornadoes separately, while accounting for correlation, has not been done previously, which makes these results unique and not directly comparable to results of previous studies.

Figure 38 shows the posterior means for the parameters of the log-linear model for the Poisson intensity of the common latent process. Note that the intensity of the common latent process accounts for the correlation between the F0-F1 and F2-F5 tornadoes processes. Figure 38(a) shows the linear temporal trend for the latent processes which shows

an increasing trend for the correlation between the F0-F1 and F2-F5 counts in the west plains (New Mexico and Colorado), slight increase in the midwest areas (intersection of Missouri, Kansas, Oklahoma and Arkansas; Tennessee and Kentucky), and slight increase in the northern part of the U.S. (Great Lakes area and South Dakota). Figure 38(b)-(d) show the climatological effects of ENSO on the intensity of the common latent process. Figure 38(b) shows increasing correlation between F0-F1 and F2-F5 tornadoes during La Niña seasons in the southern U.S. with the highest values around the Florida panhandle. Figure 38(c) shows increasing correlation between F0-F1 and F2-F5 tornadoes during normal seasons in the northeast U.S. and slight increase for the southwest. Figure 38(d) shows increasing correlation between F0-F1 and F2-F5 tornadoes during El Niño seasons in the southern part of the U.S., with the highest values in Louisiana and Mississippi.

Figure 39 shows a presence-absence of the zero-inflation probability based on the mean of the zero-inflation probability estimates for 1953-2001. This map could be considered as an indicator of the areas with excess zero tornado counts. For example, there is a large area in northern Mississippi with high zero-inflation probability of F2-F5 tornados. However the same area has low zero-inflation probabilities for F0-F1 tornadoes. This could indicate the higher number of occurrences of F0-F1 than F2-F5 tornadoes in that area or could possibly indicate misclassification of tornadoes resulting in under counts of F2-F5 tornadoes. The results for the population effects do not show a significant effect of population density on the zero-inflation probabilities. These results are not necessarily in disagreement with a previous study where such influence was found to be important (Anderson et al. 2007). Anderson et al. (2007) showed that population influence on tornado detectability is significant in

most urban and highly populated areas they studied. The model described in this chapter, as it is, does not have the capability to model the effect of population in such a fashion (i.e., distinction between high and low-populated areas). However, the discussed results can motivate future studies to investigate this hypothesis.

4.7 Conclusion

In this chapter, extension of semiparametric hierarchical Bayesian modeling to multivariate zero-inflated spatio-temporal count processes was considered and an application was discussed. Modeling such processes is a complex problem. The methodology discussed in this chapter, although based on a joint distributional perspective, is easy to implement in a hierarchical modeling framework. The application discussed can motivate studies of different environmental processes with similar characteristics (correlated zero-inflated count processes).

5 Multiresolution and Dynamic Resolution Hierarchical Models for Spatio-Temporal Dynamical Processes

5.1 Introduction

The complexity of spatio-temporal processes and large number of observations common in environmental and physical sciences necessitates the use of hierarchical models with sparse structures and efficient parameterizations. In this chapter two different dimension-reduced modeling approaches for such spatio-temporal processes are described. These approaches take advantage of multiresolution and spline-based modeling methods in a Bayesian hierarchical framework and use the minimum number of effective parameters possible to allow sub-processes at different scales to interact. The computational efficiency obtained by applying these multiresolution methods to such hierarchical models make these methods preferable to traditional covariance-based space-time statistical methods for very large data sets.

Using hierarchical Bayesian approaches, efficient modeling of complex spatio-temporal processes is possible via mapping the dynamics to hidden processes which are substantially lower in dimensionality. Such approaches are especially useful for cases where an efficient parameterization of state transition functions (i.e., propagation matrix) is difficult (or impossible), and/or the underlying dynamics of the spatio-temporal process in not known. Only a few examples of multiresolution methods for spatial and spatio-temporal processes were found in the statistical modeling literature (e.g., Nowak 1998; Nychka et al. 2002; Ferreira et al. 2006; for spatio-temporal models see Berliner et al. 1999; Wikle et al. 2001;

Johannesson et al. 2007).

The first method discussed in this chapter is designed to utilize multiresolution bases (e.g., wavelets) in order to model a spatio-temporal process by accounting for the dynamical relationships between a low-order subset of the multiresolution coefficients. The flexibility that this approach provides allows for rigorous modeling of the process at any level of detail that is of interest (i.e., coarse or fine scales of information). Such a framework provides a flexible tool for modeling dynamical spatio-temporal processes that accommodate computational efficiency (e.g., modeling large scale dynamics using coarse scale information) and resolution (modeling small scale dynamics using fine scale information). However, it should be noted that often there is a trade-off between computational efficiency and resolution. By using this approach, one has the option to emphasize one or the other as desired based on the application for which it is being used.

The second method described in this chapter describes spline-based models for dynamical spatio-temporal processes. Traditionally, spline-based models with fixed knot locations have been used to model spatial and spatio-temporal processes (e.g., van der Linde et al. 1995; O'Connell and Wolfinger 1997; Luo and Wahba 1998; Guttorp et al. 2006). However, there is uncertainty in defining the effective number of knots as well as the location of knots. In the method described in this chapter, spline-based models are considered with both fixed and moveable knots where the location of the moveable knots can be estimated dynamically. Such complex structure can be easily implemented in a hierarchical Bayesian modeling framework. Similar to the first method, using this approach, the modeling of the dynamical spatio-temporal processes can be implemented using a low-dimensional subset of

spline coefficients.

5.2 Methods

This section describes two different multiresolution methods for modeling spatio-temporal processes. The models constructed using these two methods are fundamentally different and thus the performance of these models are not necessarily comparable. However, the general modeling framework is very similar in the sense that for both methods the modeling of the spatio-temporal processes can be implemented using a low number of coefficients in the spectral space.

In the hierarchical modeling framework, the data model can be written as

$$\mathbf{Z}_t = \mathbf{K}_t \mathbf{u}_t + \boldsymbol{\varepsilon}_t, \tag{5.48}$$

where $\mathbf{Z}_t = (z_{1t}, \ldots, z_{mt})'$ denotes observations at time t from an unobserved spatio-temporal process $\mathbf{u}_t = (u_{1t}, \ldots, u_{nt})'$, $\mathbf{K}_t$ represent an incidence (or mapping) matrix (with 0 or 1 elements) for time t , and $\boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \sigma^2_{\varepsilon}\mathbf{I})$ represents measurement error. Note that the dimensionality of $\mathbf{Z}_t$ and $\mathbf{u}_t$ can be different and this is allowed by the incidence matrices ($\mathbf{K}_t$'s).

Often the process $\mathbf{u}_t$ is high-dimensional (i.e., large n) with substantially lower number of time points (i.e., small T). A first-order spatio-temporal dynamic model for this process can be defined as

$$\mathbf{u}_t = \mathbf{H}\mathbf{u}_{t-1} + \boldsymbol{\nu}_t, \tag{5.49}$$

where $\mathbf{H}$ is an $n \times n$ propagator matrix and $\boldsymbol{\nu}_t$ is a Gaussian noise process which is independent in time and correlated in space (i.e., $\boldsymbol{\nu}_t \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\nu)$). However, for environmental studies

where often the number of locations n is much larger than the number of times T for which observations are available, it is difficult or impossible to estimate $\mathbf{H}$ and $\mathbf{\Sigma}_\nu$. Alternatively, using spectral methods, the modeling procedure can be conducted on a dimension-reduced subset or transformation of the process in cases where parameterization of these matrices ($\mathbf{H}$ and $\mathbf{\Sigma}_\nu$) is not possible. In general, a spectral representation of the process at time t can be written as

$$\mathbf{u}_t = \Psi \mathbf{a}_t, \tag{5.50}$$

where Ψ is an $n \times k$ ($n \ll k$) matrix of basis functions and $\mathbf{a}_t$ is a $k \times 1$ vector of spectral coefficients. The modeling of the dynamical relationships will be implemented on the hidden process ($\mathbf{a}_t$) which has a much lower dimension than the original process ($\mathbf{u}_t$). Thus, the dynamical model can be written as,

$$\mathbf{a}_t = \mathbf{M}\mathbf{a}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim \mathrm{N}(\mathbf{0}, \mathbf{\Sigma}_\eta), \tag{5.51}$$

where $\mathbf{M} = \Psi' \mathbf{H} \Psi$ and $\mathbf{\Sigma}_\eta = \Psi' \mathbf{\Sigma}_\nu \Psi$. However, this is not helpful since the estimation of $\mathbf{H}$ and $\mathbf{\Sigma}_\nu$ is difficult, if not impossible, as discussed previously. Rather, the estimation of low-dimensional matrices, $\mathbf{M}$ and $\mathbf{\Sigma}_\eta$, can be done in the linear dynamical space. In this section, two different methods for implementing such a dimension-reduced dynamical modeling approach are discussed.

5.2.1 Method 1: A Multiresolution Hierarchical Model with a Hidden Dynamic Process

In this section a hierarchical model based on multiresolution bases is considered.

The process model can be written by first considering the multiresolution transformation

$$\mathbf{u}_t = \mathbf{W}\boldsymbol{\alpha}_t, \tag{5.52}$$

where $\mathbf{W}$, similar to Ψ in (5.50), is an $n \times n$ matrix of multiresolution bases (e.g., wavelets) and $\boldsymbol{\alpha}_t$'s are time-varying multiresolution coefficients. For each resolution, j, $\boldsymbol{\alpha}_t$ is conditioned on a common underlying dynamical process, $\mathbf{b}_t$,

$$\boldsymbol{\alpha}_t^j = \boldsymbol{\Phi}^j \mathbf{b}_t + \boldsymbol{\gamma}_t^j, \qquad \boldsymbol{\gamma}_t^j \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\gamma^j), \tag{5.53}$$

where $\Sigma_\gamma^j = \sigma_{\gamma^j}^2$ for simplicity, for each multiresolution scale, $j = 0, \ldots, J$, and the controlling dynamics are given by

$$\mathbf{b}_t = \mathbf{H}_b \mathbf{b}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta), \tag{5.54}$$

where $\boldsymbol{\Phi}^j$ is known with dimensionality $n_j \times p$ ($p \ll \Sigma_j n_j$) and the dimensionality of $\mathbf{b}_t$ is much lower than $\boldsymbol{\alpha}_t$. Alternatively, model 5.54 can applied in each resolution separately (i.e., $\mathbf{b}_t^j = \mathbf{H}_b^j \mathbf{b}_{t-1}^j + \boldsymbol{\eta}_t^j$, $\eta^j \sim N(\mathbf{0}, \Sigma_{\eta^j})$ for $j = 1, \ldots, J$)). The matrices $\mathbf{H}_b$ and $\boldsymbol{\Sigma}_\eta$ are then given prior distributions

$$\begin{aligned} \text{vec}(\mathbf{H}_b) &\sim N(\tilde{\mathbf{h}}_b, \boldsymbol{\Sigma}_h), \\ \boldsymbol{\Sigma}_{\eta^j}^{-1} &\sim \text{Wishart}((\mathbf{S}\nu)^{-1}, \nu), \quad \text{for} \quad j = 1, \ldots, J, \end{aligned}$$

where vec$(\cdot)$ denotes the matrix vectorizing operator (i.e., stacks the columns of a matrix on top of each other).

The dimension reduction discussed previously is implemented through the Φ^j matrices. Using these matrices, a subset of the multiresolution coefficients can be obtained and a

dynamical model can be considered as described in (5.54). There are many choices for Φ^j matrices including mapping matrices which map large-scale or small-scale coefficients to some or one of multiresolution coefficients (e.g., wavelet coefficients **b**'s), empirical orthogonal functions (**EOF**$'s$) in spectral space, and other analytical basis functions of choice.

Figure 40 shows a typical representation of the coefficients of a 2-D wavelet transform. This representation allows for modeling the dynamics on each resolution separately while allowing for interactions between scales. For example, the dynamics can be modeled using the coarsest scale (S^J) which has substantially lower number of coefficients compared to the original image. This is possible by using Φ^j matrices as mapping matrices (i.e., matrices with 0's and 1's as elements) which map subsets of **a**'s to a few number of **b** coefficients.

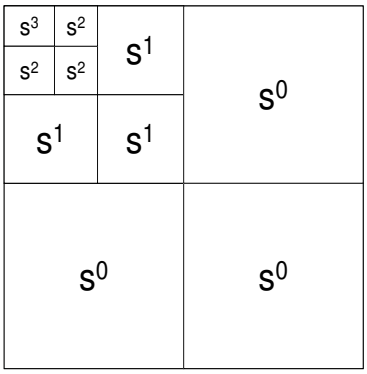


Figure 40: Schematic for coefficients of a discrete wavelet transform (DWT) of an image for three levels of resolution. Each level of resolution is shown as S^j where $j = 0, \ldots, 3$.

Using a hierarchical Bayesian modeling framework, one needs to define prior distributions for all the unknown parameters. Here, the following relatively vague prior distributions

are defined for all the unknown parameters

$$\mathbf{b}_0 \sim N(\mathbf{0}, 100 \times \mathbf{I}),$$

$$\mathbf{h} \sim N(\mathbf{0}, 10 \times \mathbf{I}), \quad \text{where} \quad \mathbf{h} \equiv \text{vec}(\mathbf{H}_b),$$

$$\sigma^2_\varepsilon \sim IG(q_\varepsilon = 2, r_\varepsilon = 9.99),$$

$$\sigma^2_{\gamma^j} \sim IG(q_{\gamma^j} = 2, r_{\gamma^j} = 9.99),$$

$$\mathbf{\Sigma}^{-1}_{\eta^j} \sim \text{Wishart}(S = 5 \times \mathbf{I}, \nu = 20), \quad \text{for} \quad j = 1, \ldots, J,$$

Thus, the estimation of model parameters can be carried out using Gibbs sampling. Appendix C shows all the full-conditional distributions required for the Gibbs sampling algorithm.

5.2.2 Method 2: A Dynamic Resolution Hierarchical Model with Moveable Knots

A very important task in modeling with basis functions (e.g., spline bases) is determining the number and topology of the bases (e.g., location of spline knots). Choosing pre-determined number and location of the knots may yield inflexible and/or ineffective models. This is particularly important in modeling high-dimensional spatio-temporal processes where obtaining a low number of effective bases is important to increase computational efficiency while maintaining resolution. In the context of spline-based models, often a predetermined number of fixed knot points are used (e.g., Ruppert et al. 2003). Alternatively, adaptive spline methods have been proposed to introduce more flexibility into this modeling framework. Some of these methods include time-varying splines (e.g., Harvey and Koopman 1993), Bayesian subset selection for spline models (e.g., Smith and Kohn 1997); Denison

et al. 1998), Bayesian varying-coefficient spline models (e.g., Biller and Fahrmeir 2001; Eubank et al. 2004), and free-knot spline models in a Bayesian setting (e.g., Holmes and Mallick 2003).

In this section, a spline-based hierarchical Bayesian model with dynamical resolution (i.e. moveable spline knots) is considered. This approach allows for effective modeling of spatio-temporal processes using a small number of spline bases which are dynamically chosen. Here, the spectral transformation matrix (Ψ) in (5.50), is assumed to be a time-varying matrix of thin-plate spline basis functions. This assumption, which introduces extra flexibility in the model, is accommodated by considering both fixed and moveable knots for the spline bases. Note that careful treatment of elements of the dynamical model is required as a consequence of considering moveable knots. For example, let Ψ be an $n \times k$ matrix of spline bases with $k - b$ fixed knots and b moveable knots. Thus, $\Psi = \Psi(\mathbf{x}_t^m, \mathbf{y}_t^m)$ where $(\mathbf{x}_t^m, \mathbf{y}_t^m)$ denote the coordinates for the moveable knots (superscript m indicates that the coordinates belong to a moveable knot). The spectral representation of the process at time t can be written as

$$\mathbf{u}_t = \Psi(\mathbf{x}_t^m, \mathbf{y}_t^m)\mathbf{a}_t, \tag{5.55}$$

where the (i, k) element of matrix Ψ is $\mathbf{s}_{i,k}^2 \log(\mathbf{s}_{i,k})$, for $i = 1, \ldots, n$, and $k = 1, \ldots, K$, with $\mathbf{s}_{i,k} = \sqrt{(x_i - \kappa_k)^2 + (y_i - \kappa_k)^2}$ (κ denotes the set of coordinates of the knots).

In this case $\mathbf{a}_t = \begin{bmatrix} \mathbf{a}_t^f \\ \mathbf{a}_t^m \end{bmatrix}$ where $\mathbf{a}_t^f$ indicates spectral coefficients related to the fixed knots, and $\mathbf{a}_t^m$ spectral coefficients related to the moveable knot(s). Thus, the dynamical

model can be defined for coefficients of the fixed knots

$$\mathbf{a}_t^f = \mathbf{G}\mathbf{a}_{t-1}^f + \boldsymbol{\eta}_t^f, \quad \boldsymbol{\eta}_t^f \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta^f). \tag{5.56}$$

Note that keeping track of these moveable coefficients in the MCMC algorithm is difficult, if not impossible (e.g., two moveable knots could possibly switch locations and the MCMC algorithm will not be able to recognize such an effect). Moreover, at each time step, these coefficients could possibly correspond to different knot locations with differing dynamics. Thus, dynamical treatment of these coefficients is not feasible. However, the coefficients corresponding to the moveable knots can be estimated non-dynamically by drawing samples from the full-conditional distributions of these coefficients for each time t.

The estimation of the location of the moveable knots can be conducted by estimating the coordinates of these knots. Here, a simple random walk model for the coordinates is considered. Thus,

$$\mathbf{x}_t^m = \mathbf{x}_{t-1}^m + \nu_t^x, \quad \nu_t^x \sim \mathrm{N}(0, \boldsymbol{\Sigma}_\nu^x),$$

$$\mathbf{y}_t^m = \mathbf{y}_{t-1}^m + \nu_t^y, \quad \nu_t^y \sim \mathrm{N}(0, \boldsymbol{\Sigma}_\nu^y),$$

where it is assumed $\boldsymbol{\Sigma}_\nu^x = \sigma_{\nu^x}^2\mathbf{I}$ and $\boldsymbol{\Sigma}_\nu^y = \sigma_{\nu^y}^2\mathbf{I}$, for simplicity.

Using a hierarchical Bayesian framework, the estimation of the unknown parameters can be conducted, after assigning prior distributions to all the unknown parameters, based on Gibbs sampling with Metropolis-Hastings steps for the estimation of coordinates ($\mathbf{x}^m$ and

$\mathbf{y}^m$) for the moveable knot. The following relatively vague prior distributions are considered:

$$\mathbf{a}_t^m \sim N(\mathbf{0}, 10 \times \mathbf{I}), \quad \text{for } t = 1, \ldots, T,$$

$$\mathbf{a}_0^f \sim N(\mathbf{0}, 10 \times \mathbf{I}),$$

$$\mathbf{g} \sim N(\mathbf{0}, 100 \times \mathbf{I}), \quad \text{where } \mathbf{g} \equiv \text{vec}(\mathbf{G}),$$

$$\sigma_\varepsilon^2 \sim IG(q_\varepsilon = 2, r_\varepsilon = 9.99),$$

$$(\mathbf{\Sigma}_\eta^f)^{-1} \sim \text{Wishart}(S = 5 \times \mathbf{I}, \nu = 20),$$

$$\sigma_{\nu_x}^2 \sim IG(q_\eta = 2, r_{k,\eta} = 9.99),$$

$$\sigma_{\nu_y}^2 \sim IG(q_\eta = 2, r_{k,\eta} = 9.99).$$

The details for the derivation of the full-conditional distributions are shown in Appendix D.

5.3 Example: Advection-Diffusion Simulation

In this section a simulation based on a 2-D advection-diffusion equation is considered. The 2-D advection-diffusion equation used to generate the process is similar to that in equation (2.23) described in Chapter 2. The simulation experiment is considered for a 16×16 rectangular shaped domain ($\triangle x = \triangle y = 1$ for 50 time steps with $\triangle t = 2$. Finally, 50% of the observations were removed at random for each time t ($t = 1, \ldots, 50$).

The methods described in the previous section are employed to estimate the spatio-temporal process with "missing data". Method 1 was used based on a 2-D Daubechies-4 (Daubechies 1988) wavelet transform and the modeling of the dynamics was conducted using three levels of resolution where only six elements for the dynamical model (i.e., $\mathbf{b}_t = \mathbf{H}_b \mathbf{b}_{t-1} + \boldsymbol{\eta}_t$) were considered. These six elements were chosen such that for the vectorized

wavelet transform (i.e., $\mathbf{a}_t$ which is a vector of length 256) at each time step, four elements (the first four elements in vector $\mathbf{b}_t$) correspond to the first four wavelet coefficients of the coarsest scale, S^3, one element corresponds to 5th-128th wavelet coefficients (the fifth element in vector $\mathbf{b}_t$), and one element corresponds to the remaining wavelet coefficients, coefficients 129-256 (the sixth element in vector $\mathbf{b}_t$). This was done by using ϕ^j mapping matrices with 0's and 1's as elements. Method 2 was employed based on a thin-plate spline representation with twenty eight fixed knots and four moveable knots (i.e., dimensionality of $\Psi = 256 \times 32$) which were initialized at points with coordinates (5,3), (6,8), (14,4), and (13,10) and were allowed to move independently.

The number of iterations for the MCMC was 25000, where the first 5000 iterations were discarded as "burn-in". For both models, convergence was achieved relatively quickly (i.e., <2000 iterations). Convergence was assessed visually, using the MCMC trace plots for the parameters.

5.3.1 Simulation Results

Figures 41 and 42 show the results for both methods along with observations and the true process. The model results for the simulated data are difficult to analyze without a quantitative measure of model performance. Here, Root Mean Squared Error (RMSE) which is a simple method for assessing the fit of the model to the data, is considered. Thus, the model which minimizes

$$RMSE(M_j) = \sqrt{\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1:}^{n}(\mu_t(s_i) - u_t(s_i))^2}, \tag{5.57}$$

for model M_j $(j = 1, 2)$ is regarded as the model which fits better to the data. Note that $\boldsymbol{\mu}_t$ denotes the posterior mean of the estimated process. The calculated RMSE value for M_1 (wavelet-based multiresolution model) is 0.1367, while the RMSE for M_2 (spline-based dynamic resolution model) is 0.1325. The values for RMSEs suggest that Model 2 fits better to the data compared to Model 1. However, a more careful assessment of the plots in Figures 41 and 42 shows that Model 1 does a better job in capturing the diffusive behavior (i.e., dynamics) of the simulated spatio-temporal process, while Model 2 seems to interpolate the missing observations better (note that Model 2 overestimates the diffusive behavior of the process). As stated earlier, these two models are based on completely different methods and not necessarily comparable. However, the results for the simulation experiment could provide more insight on the strengths and weaknesses of each of these modeling approaches.

5.4 Application: Nowcasting Radar Reflectivities

Nowcasts are short-period forecasts of radar reflectivities to predict heavy rainfall. Most common nowcasting methods are based on either deterministic models such as "Gandolf" (Pierce et al. 2000) and "Nimrod" (Golding 2000), or extrapolation methods such as the Thunderstorm Identification, Tracking, Analysis and Nowcasting (TITAN) system (Dixon and Wiener 1993), the Storm Cell Identification and Tracking (SCIT) system (Johnson et al. 1993). These methods, although being used widely, have limitations (e.g., see Xu et al. 2005 for details) and more importantly do not realistically account for prediction uncertainty.

Recent developments in weather radar nowcasting include methods that consider prediction uncertainty. Examples of such methods include Spectral Prognosis (S-PROG) system (Seed 2003) and a kernel-based physical/statistical model developed by Xu et al. (2005).

Xu et al. (2005) considered a hierarchical Bayesian modeling approach for a spatio-temporal dynamical model which is efficiently parameterized within a kernel-based integro-difference equation framework. In this section the multiresolution and dynamic resolution methods described previously, are employed for a problem in weather radar nowcasting.

5.4.1 Data and Methods

The data considered in this section includes 11 time steps of observed radar reflectivities at 10 minute intervals from San Antonio, TX on July 5, 2002 which contained a series of intense convective storm cells that moved over the area resulting in some flash flooding. The observations are NEXRAD LEVEL II data reduced to a 16×32 grid of 4-km resolution pixels obtained from a mosaic of three radars including Corpus Christi, TX (KCRP), Brownsville, TX (KBRO) and San Antonio, TX (KEWX).

In this section an application of the methods described for nowcasting radar reflectivities is considered. Two models are considered to conduct nowcasting for radar reflectivities using the first 10 time steps as "data" in order to forecast the radar reflectivities for the 11th and 12th time steps. Model 1 is a wavelet-based multiresolution model using Method 1, where the dynamics are modeled rigorously on the coarsest scales (i.e., $\mathbf{b}_t^j = \mathbf{H}_b^j \mathbf{b}_{t-1}^j + \boldsymbol{\eta}_t^j$, $\eta^j \sim N(\mathbf{0}, \Sigma_{\eta^j})$ for $j = 2, 3$) where $\mathbf{b}^j$'s are vectors of length 8, and simple autoregressive models for the finer scales ($j = 0, 1$) where $\mathbf{b}^j$'s are vectors of length 32 and 128. Model 2 is a dynamic resolution model described in Method 2, with 24 fixed knots and 5 moveable knots where the coordinates for one the moveable knots are initialized at points with coordinates (6,6), (11,6), (16,6), (21,6), and (26,6). These moveable knots are allowed to move independently of each other.

The number of MCMC iterations for Model 1 was considered as 25000 iterations with the first 5000 discarded as the "burn-in" period. For Model 2, 50000 iterations with 10000 iterations for the "burn-in" period was used. For each model, convergence was assessed by visual check of the MCMC chains for the parameters.

5.5 Results and Discussion

Figures 43 and 44 show the results for both models for time steps 1-10. The smoothness of the estimated radar reflectivities is due to the coarse level modeling in Model 1 (wavelet-based multiresolution model) and the low number of spline knots in Model 2 (spline-based model). Nowcasting results for Model 1 for time steps 11 and 12 are shown in Figure 45. No reasonable nowcasts for time steps 11 and 12 were obtained using Model 2. This is due to the fact that modeling such an intensifying storm system is only possible through linear dynamics with explosive behavior (i.e., as defined through the eigenvalues of the propagator matrix; the absolute values of the eigenvalues must be smaller than 1 in order for the dynamical model to be non-explosive).

The nowcasting results for time step 11 and 12 obtained from Model 1 provides a large scale forecast of the radar reflectivities for this time step. Considering that the modeling was conducted using a low number of wavelet coefficients, with focus on modeling on the coarse level, this nowcasting result is promising. The forecast for time step 12 is less accurate and shows explosive behavior. Model 2 results for radar reflectivities for time steps for which observation were available are very smooth due to the low number of spline knots and the smoothness can be reduced by adding more knots. However, other methods to increase the nowcasting ability of this modeling approach should be considered as increasing the

number of spline knots will result in dramatic increase in the computational time required for the MCMC algorithm. Other possible methods to consider in this framework include more effective methods of propagating the moveable knots (i.e., more complicated models for the coordinates of the moveable knots), and also incorporating meteorological variables to inform the evolution of the coordinates of the moveable knots.

Plots of the moveable and fixed knots are shown in Figures 43 and 44 which show the adaptation of the moveable knots as the dynamics changes. Note that the moveable knots slightly move during the first five time steps, but as the complexity of the process increases during time steps 6-10, the moveable knots significantly change locations. As illustrated in Figure44, the relocation of the moveable knots corresponds to the intensifying parts of the process.

Reduced computational time is essential in nowcasting weather radar images. The main advantage of the wavelet-based multiresolution model (Model 1) described in this chapter is to provide a setting for conducting such reduced-dimension modeling. However, the ability of the model to conduct nowcasts beyond one time step should be increased which could possibly be accommodated by considering more complicated methods of modeling the dynamics in the wavelet space (e.g., more complex interaction between the multiresolution scales).

Finally, the main disadvantage of using linear dynamics to model intensifying storm processes is the possible explosive behavior (as discussed previously) which could limit the ability of the model in providing useful nowcasts. Alternatively, one could impose restrictions to force the system to have a non-explosive behavior and control for explosive growth using

"growth" processes such as suggested by density dependent processes in ecology (Hilborn and Mangel 1997).

5.6 Conclusion

In this chapter, two different dimension-reduced modeling approaches for effective modeling of high-dimensional dynamical spatio-temporal processes were considered and implementation of these models for a simulated example and an application to weather radar nowcasting was discussed. Spatio-temporal processes are often high-dimensional and thus efficient approaches to reduce dimensionality while accounting for the underlying dynamical relationships among the sub-processes seem necessary. The multiresolution and dynamic resolution methods discussed in this chapter provide for such effective modeling of dynamical spatio-temporal processes, specifically when the underlying dynamics is not known.

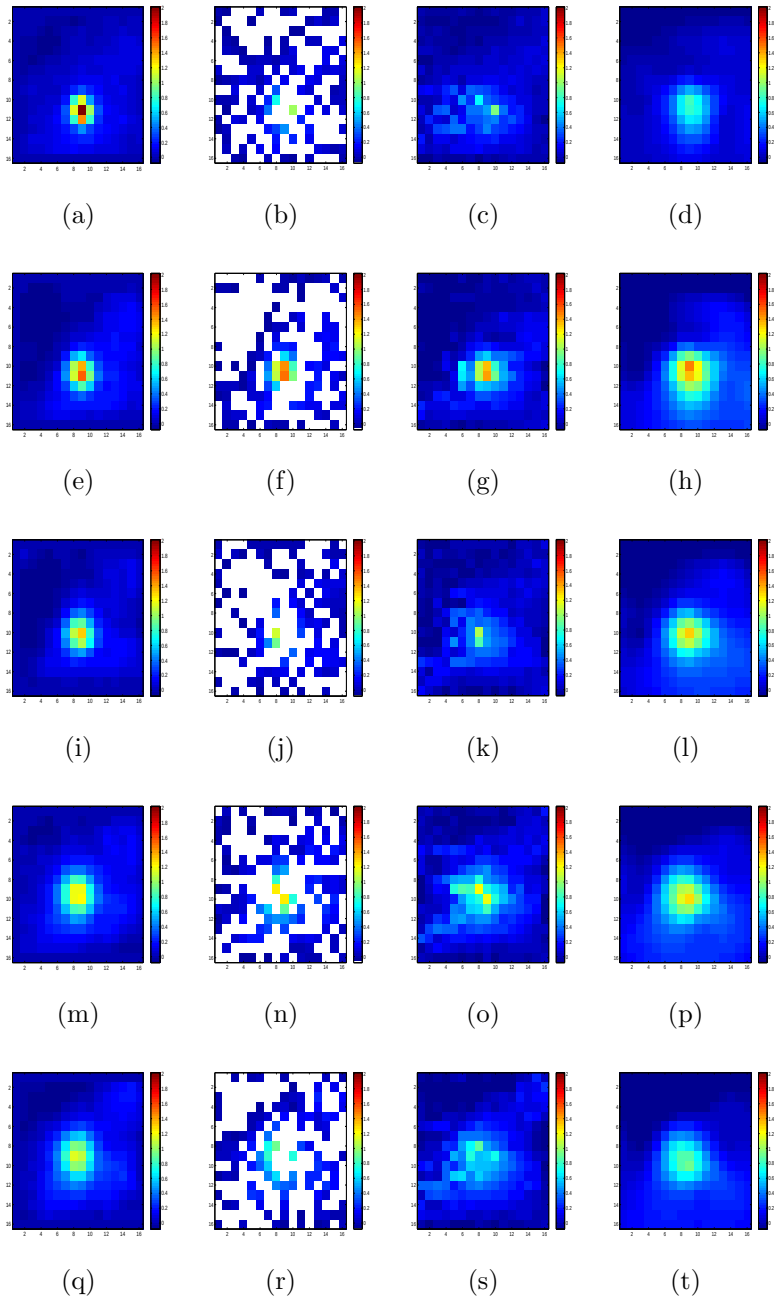


Figure 41: Truth (left column), data (second column), Method 1 posterior means (third column), and Method 2 posterior means (right column) for the advection-diffusion simulation (2-D) for time steps 5-25; (a)-(d) t=5, (e)-(h) t=10, (i)-(l) t=15, (m)-(p) t=20, (q)-(t) t=25. Note that in the representation of the observations (figures in the second column), white colored cells indicate missing values.

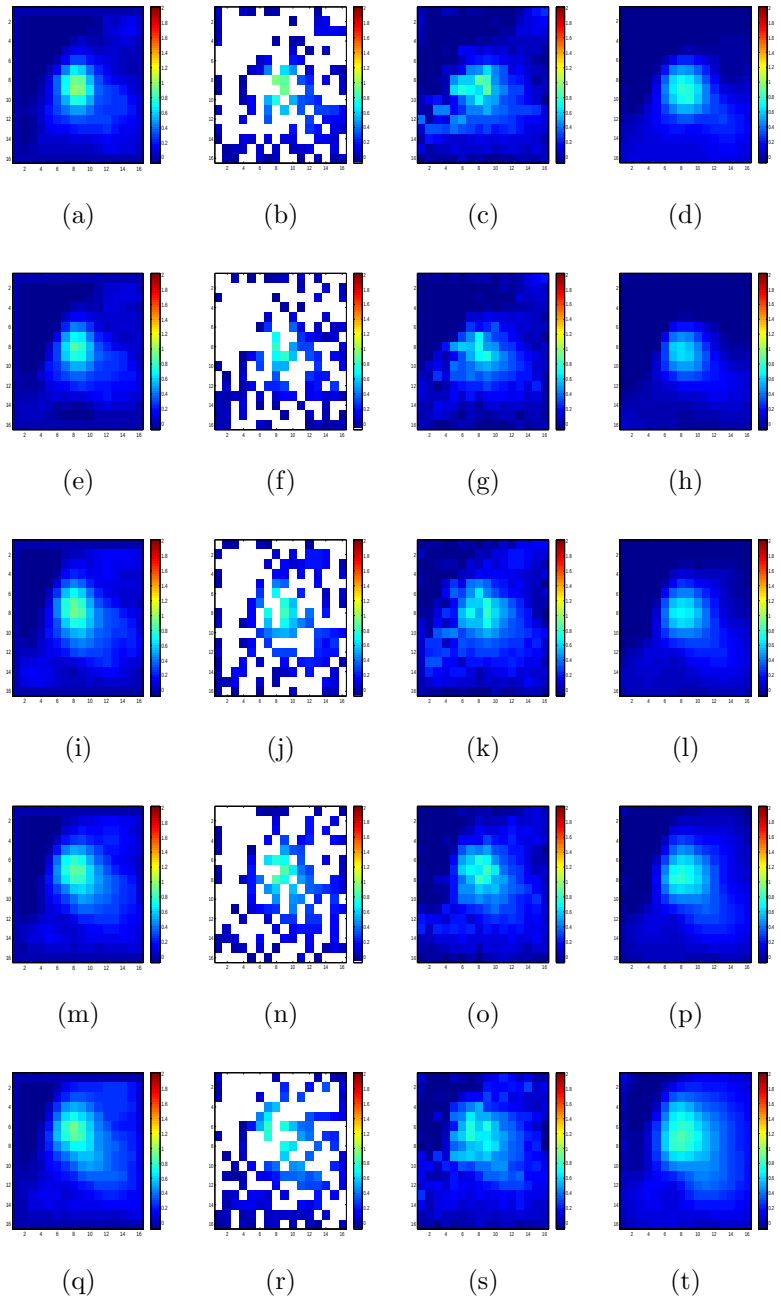


Figure 42: Truth (left column), data (second column), Method 1 posterior means (third column), and Method 2 posterior means (right column) for the advection-diffusion simulation (2-D) for time steps 30-50; (a)-(d) t=30, (e)-(h) t=35, (i)-(l) t=40, (m)-(p) t=45, (q)-(t) t=50. Note that in the representation of the observations (figures in the second column), white colored cells indicate missing values.

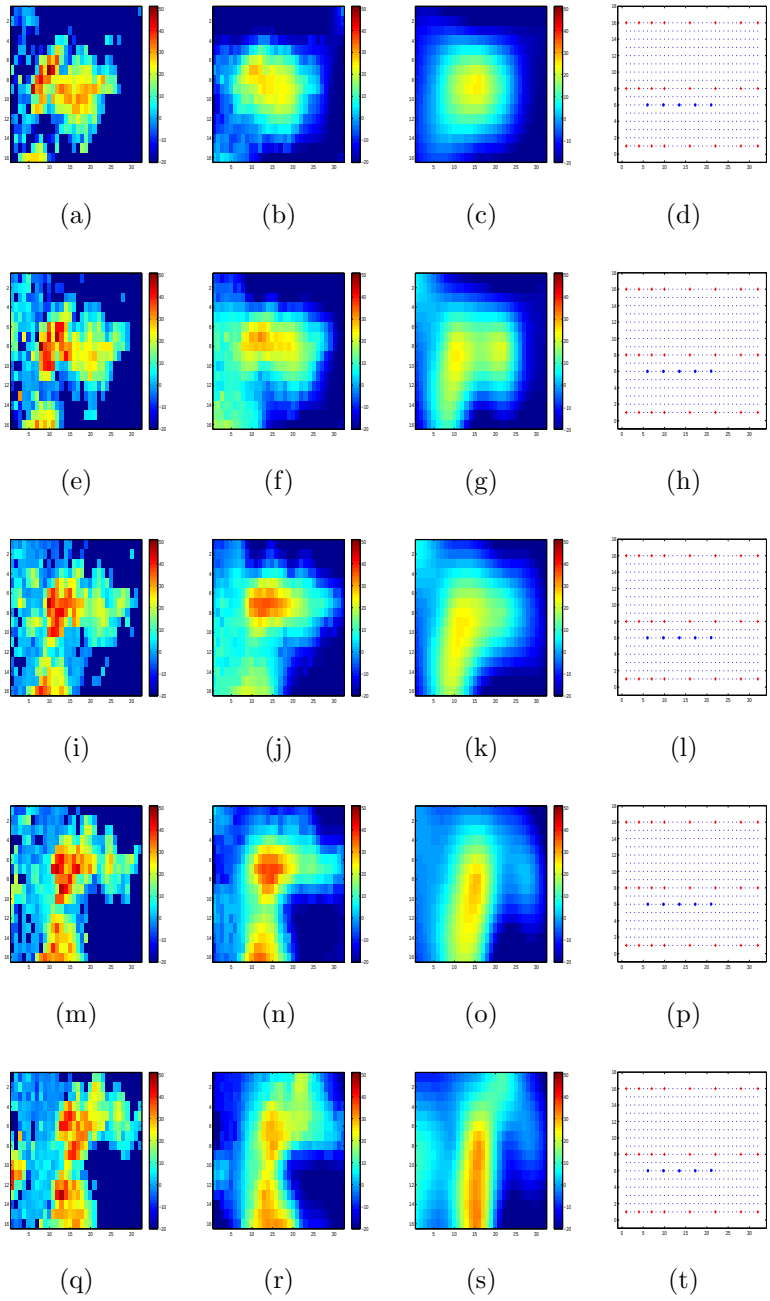


Figure 43: Data (left column), Model 1 (multiresolution model) posterior means (second column), Model 2 (dynamic resolution model) posterior means (third column), and knot locations for Model 2 where $(\cdot)$ indicates the grid coordinates, $(+)$ indicates the fixed knots, and $(*)$ indicates the moveable knots (right column) for radar reflectivities for time steps 1-5; (a)-(d) t=1, (e)-(h) t=2, (i)-(l) t=3, (m)-(p) t=4, (q)-(t) t=5.

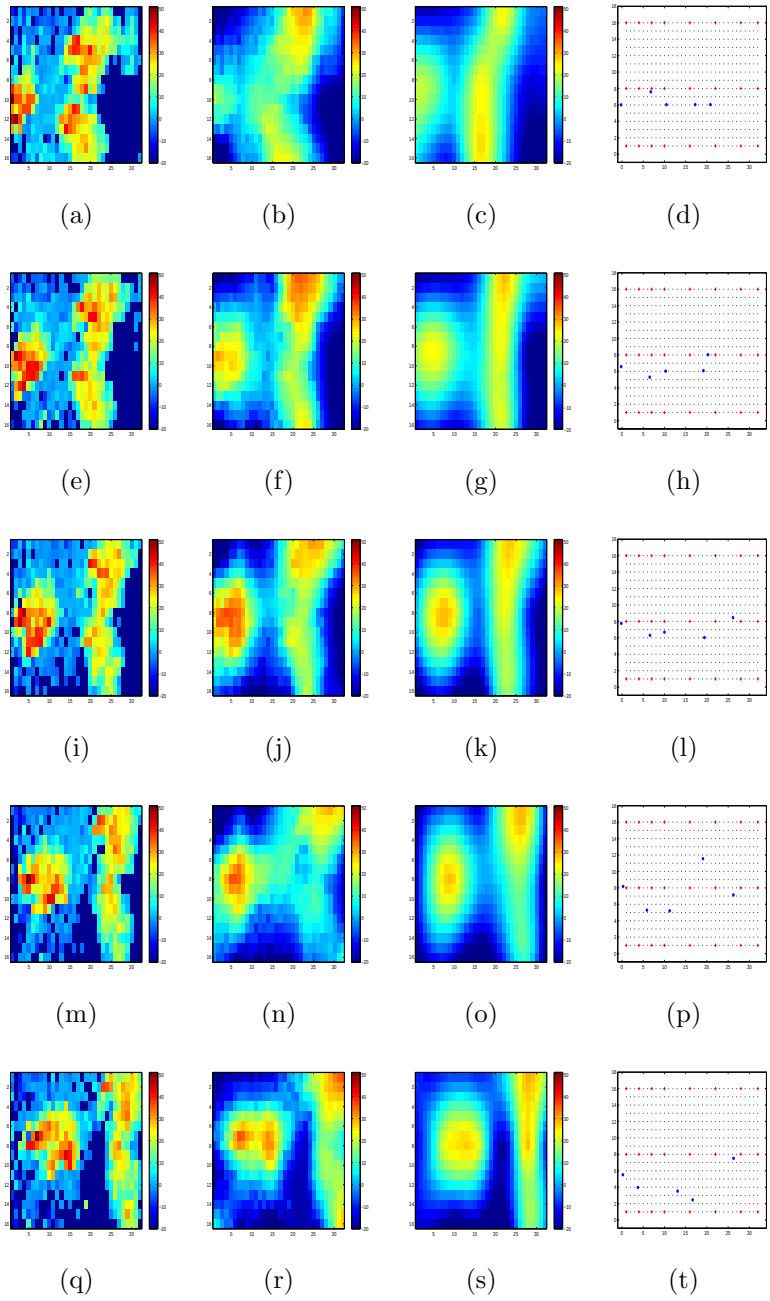
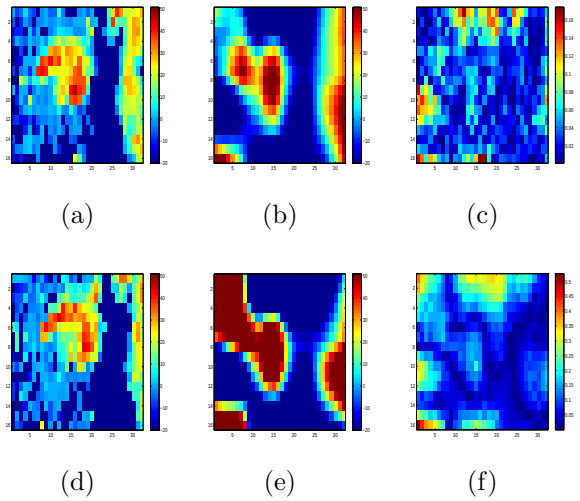


Figure 44: Data (left column), Model 1 (multiresolution model) posterior means (second column), Model 2 (dynamic resolution model) posterior means (third column), and knot locations for Model 2 where $(\cdot)$ indicates the grid coordinates, $(+)$ indicates the fixed knots, and $(*)$ indicates the moveable knots (right column) for radar reflectivities for time steps 6-10; (a)-(d) t=6, (e)-(h) t=7, (i)-(l) t=8, (m)-(p) t=9, (q)-(t) t=10.



(a) (b) (c)

(d) (e) (f)

Figure 45: Nowcasting results for Model 1 (multiresolution model) for time steps 11 and 12; (a) Data ("Truth") (t=11) (b) Forecast (t=11): posterior mean, (c) Standard deviation of forecast (t=11), (d) Data ("Truth") (t=12) (e) Forecast (t=12): posterior mean, (f) Standard deviation of forecast (t=12).

6 General Conclusions

Models discussed in the previous chapters illustrate the utilization of observational data to characterize spatio-temporal environmental processes. The increasing availability of data from different sources (e.g., remote sensing, monitoring networks) and different scales (i.e., spatial, spatio-temporal) present modeling challenges which can be effectively described using efficient parameterizations of spatio-temporal processes. Future research directions in this area require attempts to effectively model complicated processes while retaining the efficiency of the models. Hierarchical Bayesian models present a flexible modeling tool that allows for improvements in both of these areas simultaneously while enabling the use of data obtained from multiple sources and scales.

The methods discussed in this dissertation include a wide scope of problems related to the modeling of spatio-temporal environmental processes. Specifically, methods were described for efficient modeling of spatio-temporal processes using discrete- and continuous-valued data.

The methods described for continuous data included the following cases for dynamical spatio-temporal processes:

(a) Partially known dynamics which can be described (partially) by deterministic PDE models (e.g., PDE models based on physics, biology) with random parameters, expressed hierarchically.

(b) Unknown dynamics which can be modeled using multiresolution hidden processes.

Future research efforts in these areas should be focused on increasing model efficiency while maintaining model accuracy. PDE-based models can be used as prior information in

estimating the process, along with data, in the context of hierarchical Bayesian modeling. Such coupling of data and scientific knowledge can accommodate modeling of complicated processes. The method described in Chapter 2 provides simple examples of such a modeling technique. For example, using data and a simple linear PDE, the modeling of many non-linear processes can become feasible. The advantages of such modeling approaches go beyond computational efficiency and make the task of modeling of many complex processes possible.

Multiresolution methods have been under-utilized in spatio-temporal modeling. As methods of collecting data become more advanced, researchers become more interested in the modeling of very high-dimensional spatio-temporal processes and multiresolution methods can play a significant role in modeling such complex processes. The methods described in Chapter 5 motivate the use of such multiresolution methods for modeling spatio-temporal processes.

The methods described for discrete data include methodology for semiparametrically modeling multivariate count processes with excess zeros. This modeling approach was discussed in general and was extended to spatio-temporal processes. This novel methodology can be used to model many environmental and ecological count processes which often have complicated dependence structures, non-linear covariate effects, and excess zeros for the observations (due to inefficient sampling and/or heterogeneous patterns of presence/absence of the process). Often, the statistical models used for modeling such processes are incapable of addressing these common issues. Thus, the development of new methodology to account for these issues is both necessary and appealing.

Modeling dependencies among environmental count processes is complicated. The modeling approaches described in Chapters 3 and 4, address the development of such dependencies for certain cases where the dependencies among count processes are described by environmental factors (e.g., common habitat use for certain species; common climatological effects for tornadoes). Another important issue regarding dependencies among count processes is based on conditional dependence among two or more count processes due to presence of one or more common factor which is also a count process. Examples of such dependency structures are ubiquitous in ecology and include multi-species predator-prey models where two (or more) different species have the same species as prey. The inherent conditional structure of hierarchical Bayesian modeling framework provides a natural setting to address such conditional dependence structures. Future extensions of the methodology discussed in Chapters 3 and 4 include (but are not limited to) the development of models accounting for conditional dependencies among multivariate count processes.

Appendix A: Full-Conditionals

Consider the following hierarchical model for the 1-D advection-diffusion:

$$\begin{aligned}\mathbf{Z}_t =& \mathbf{K}_t\mathbf{u}_t + \boldsymbol{\varepsilon_t}, \qquad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\sigma}_{\varepsilon}^2\mathbf{I}),\\ \mathbf{G}\mathbf{u}_t =& \mathbf{M}\mathbf{u}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \sigma_{\eta}^2\mathbf{R}(\theta)),\end{aligned}$$

where $\mathbf{M} \equiv \mathbf{G} + \triangle t\mathbf{H}(\boldsymbol{\beta}, \triangle x, \triangle y)$ and $\mathbf{G} = \mathbf{G}(\triangle x, \triangle y)$. Let:

$$\begin{aligned}\mathbf{u}_0 &\sim N(\tilde{\mathbf{u}}_0, \boldsymbol{\Sigma}_0^{-1}),\\ \boldsymbol{\beta} &\sim N(\tilde{\beta}, \sigma_{\beta}^2),\\ \boldsymbol{\alpha} &\sim N(\tilde{\boldsymbol{\alpha}}, \sigma_{\alpha}^2\mathbf{R}(\theta)).\end{aligned}$$

The notation [x] is used to specify the probability distribution of x. The posterior distribution of interest can now be written as:

$$\begin{aligned}[\mathbf{u}_1, \ldots, \mathbf{u}_T, \boldsymbol{\beta}, \sigma_{\eta}^2, \sigma_{\varepsilon}^2 | \mathbf{Z}_1, \ldots, \mathbf{Z}_T] \propto & \prod_{t=1}^{T}[Z_t|\mathbf{u}_t, \sigma_{\varepsilon}^2] \prod_{t=1}^{T}[\mathbf{u}_t|\mathbf{u}_{t-1}, \sigma_{\eta}^2, \boldsymbol{\beta}]\\ & \times [\mathbf{u}_0][\boldsymbol{\beta}][\sigma_{\varepsilon}^2][\sigma_{\eta}^2].\end{aligned}$$

The full-conditional distributions are:

- $[\mathbf{u}_0|\cdot] \propto [\mathbf{u}_1|\mathbf{u}_0][\mathbf{u}_0]$

$$\mathbf{u}_0|\cdot \sim N(\mathbf{A}_0\mathbf{b}_0, \mathbf{A}_0),$$

where,

$$\begin{aligned}\mathbf{A}_0 &= (\mathbf{M}'\mathbf{R}(\theta)^{-1}\mathbf{M}/\sigma_{\eta}^2 + \boldsymbol{\Sigma}_0^{-1})^{-1},\\ \mathbf{b}_0 &= (\mathbf{u}_1'\mathbf{G}'\mathbf{R}(\theta)^{-1}\mathbf{M}/\sigma_{\eta}^2 + \tilde{\mathbf{u}}_0'\boldsymbol{\Sigma}_0^{-1})'.\end{aligned}$$

- $[\mathbf{u}_t|\cdot] \propto \prod_{t=1}^{T-1}[\mathbf{Z}_t|\mathbf{u}_t][\mathbf{u}_t|\mathbf{u}_{t-1}][\mathbf{u}_{t+1}|\mathbf{u}_t]$

$$\mathbf{u}_t|\cdot \sim N(\mathbf{A}_t\mathbf{b}_t, \mathbf{A}_t),$$

where,

$$\mathbf{A}_t = (\mathbf{K}_t'\mathbf{K}_t/\sigma_\varepsilon^2 + (\mathbf{M}'\mathbf{R}(\theta)^{-1}\mathbf{M} + \mathbf{G}'\mathbf{R}(\theta)^{-1}\mathbf{G})/\sigma_\eta^2)^{-1},$$

$$\mathbf{b}_t = (\mathbf{Z}_t'\mathbf{K}_t/\sigma_\varepsilon^2 + (\mathbf{u}_{t-1}'\mathbf{M}'\mathbf{R}(\theta)^{-1}\mathbf{G} + \mathbf{u}_{t+1}'\mathbf{G}'\mathbf{R}(\theta)^{-1}\mathbf{M})/\sigma_\eta^2)'.$$

- $[\mathbf{u}_T|\cdot] \propto [\mathbf{Z}_T|\mathbf{u}_T][\mathbf{u}_T|\mathbf{u}_{T-1}]$

$$\mathbf{u}_T|\cdot \sim N(\mathbf{A}_T\mathbf{b}_T, \mathbf{A}_T),$$

where,

$$\mathbf{A}_T = (\mathbf{K}_T'\mathbf{K}_T/\sigma_\varepsilon^2 + \mathbf{G}'\mathbf{R}(\theta)^{-1}\mathbf{G}/\sigma_\eta^2)^{-1},$$

$$\mathbf{b}_T = (\mathbf{Z}_T'\mathbf{K}_T/\sigma_\varepsilon^2 + \mathbf{u}_{T-1}'\mathbf{M}'\mathbf{R}(\theta)^{-1}\mathbf{G}/\sigma_\eta^2)'.$$

- $[\beta|\cdot] \propto \prod_{t=1}^{T}[\mathbf{u}_t|\mathbf{u}_{t-1}, \beta][\beta]$

$$\beta|\cdot \sim N(\mathbf{A}\mathbf{b}, \mathbf{A})$$

where,

$$\mathbf{A} = ((\triangle t)^2 \sum_{t=1}^{T}(\ \mathrm{diag}(\mathbf{u}_{t-1})'\mathbf{F}1'\mathbf{R}(\theta)^{-1}\mathbf{F}1\ \mathrm{diag}(\mathbf{u}_{t-1})/\sigma_\eta^2) + \mathbf{R}(\theta)^{-1}/\sigma_\beta^2)^{-1},$$

$$\mathbf{b} = (\triangle t \sum_{t=1}^{T}((-\mathbf{u}_t'\mathbf{P}_t' + \mathbf{u}_{t-1}'\mathbf{P}_{t-1}')\mathbf{R}(\theta)^{-1}\mathbf{F}1\ \mathrm{diag}(\mathbf{u}_{t-1})/\sigma_\eta^2) + \tilde{\beta}\mathbf{R}(\theta)^{-1}/\sigma_\beta^2)'.$$

(Note: $\mathbf{P}_t \equiv \mathbf{G} + \mathbf{F}2\ \mathrm{diag}(\mathbf{u}_t)\,\boldsymbol{\alpha}$ and $\mathbf{F}1$ is a sparse matrix such that $\mathbf{H}\mathbf{u}_t \equiv \mathbf{F}1\ \mathrm{diag}(\mathbf{u}_t)\,\beta + \mathbf{F}2\ \mathrm{diag}(\mathbf{u}_t)\boldsymbol{\alpha}$.)

- $[\boldsymbol{\alpha}|\cdot] \propto \prod_{t=1}^{T}[\mathbf{u}_t|\mathbf{u}_{t-1}, \boldsymbol{\alpha}][\boldsymbol{\alpha}]$

$$\boldsymbol{\alpha}|\cdot \sim N(\mathbf{Ab}, \mathbf{A})$$

where,

$$\mathbf{A} = ((\triangle t)^2 \sum_{t=1}^{T}(\ \mathrm{diag}(\mathbf{u}_{t-1})'\mathbf{F}2'\mathbf{R}(\theta)^{-1}\mathbf{F}2\ \mathrm{diag}(\mathbf{u}_{t-1})/\sigma_\eta^2) + \mathbf{R}(\theta)^{-1}/\sigma_\alpha^2)^{-1},$$
$$\mathbf{b} = (\triangle t \sum_{t=1}^{T}((-\mathbf{u}_t'\mathbf{Q}_t' + \mathbf{u}_{t-1}'\mathbf{Q}_{t-1}')\mathbf{R}(\theta)^{-1}\mathbf{F}2\ \mathrm{diag}(\mathbf{u}_{t-1})/\sigma_\eta^2) + \tilde{\boldsymbol{\alpha}}\mathbf{R}(\theta)^{-1}/\sigma_\alpha^2)'.$$

(Note: $\mathbf{Q}_t \equiv \mathbf{G} + \mathbf{F}1\ \mathrm{diag}(\mathbf{u}_t)\,\beta$ and $\mathbf{F}2$ is a sparse matrix such that $\mathbf{Hu}_t \equiv \mathbf{F}1\ \mathrm{diag}(\mathbf{u}_t)\,\beta + \mathbf{F}2\ \mathrm{diag}(\mathbf{u}_t)\boldsymbol{\alpha}$.)

- $[\sigma_\varepsilon^2|\cdot] \propto \prod_{t=1}^{T}[\mathbf{Z}_t|\mathbf{u}_t, \sigma_\varepsilon^2][\sigma_\varepsilon^2]$

$$\sigma_\varepsilon^2|\cdot \sim IG(q, r)$$

where,

$$q = q_\varepsilon + nT/2,$$
$$r = (1/r_\varepsilon + 0.5\sum_{t=1}^{T}((\mathbf{Z}_t - \mathbf{K}_t\mathbf{u}_t)'(\mathbf{Z}_t - \mathbf{K}_t\mathbf{u}_t)))^{-1}$$

- $[\sigma_\eta^2|\cdot] \propto \prod_{t=1}^{T}[\mathbf{u}_t|\mathbf{u}_{t-1}, \sigma_\eta^2][\sigma_\eta^2]$

$$\sigma_\eta^2|\cdot \sim IG(q, r)$$

where,

$$q = q_\eta + nT/2,$$
$$r = (1/r_\eta + 0.5\sum_{t=1}^{T}((\mathbf{Gu}_t - \mathbf{Mu}_{t-1})'\mathbf{R}(\theta)^{-1}(\mathbf{Gu}_t - \mathbf{Mu}_{t-1})))^{-1}.$$

Appendix B: Full-Conditionals

The posterior distribution of all the processes and parameters given the observations can be written as

$$[\mathbf{Z}_{1,t},\ldots,\mathbf{Z}_{1,T},\mathbf{Z}_{2,t},\ldots,\mathbf{Z}_{2,T},\mathbf{Z}_{3,t},\ldots,\mathbf{Z}_{3,T},\boldsymbol{\lambda}_{1,1},\ldots,\boldsymbol{\lambda}_{1,T},\boldsymbol{\lambda}_{2,1},\ldots,\boldsymbol{\lambda}_{2,T},\boldsymbol{\lambda}_{3,1},\ldots,\boldsymbol{\lambda}_{3,T},\mathbf{p}_{1,1},$$

$$\ldots,\mathbf{p}_{1,T},\mathbf{p}_{2,1},\ldots,\mathbf{p}_{2,T},\mathbf{b}_{1,0},\mathbf{b}_{1,1},\mathbf{b}_{1,2},\mathbf{b}_{1,3},\mathbf{b}_{2,0},\mathbf{b}_{2,1},\mathbf{b}_{2,2},\mathbf{b}_{2,3},\mathbf{b}_{3,0},\mathbf{b}_{3,1},\mathbf{b}_{3,2},\mathbf{b}_{3,3},\mu_1,\mu_2,$$

$$\alpha_{1,0},\alpha_{1,1},\alpha_{2,0},\alpha_{2,1},\sigma^2_{1,\eta},\sigma^2_{2,\eta},\sigma^2_{1,1,\varepsilon},\sigma^2_{1,2,\varepsilon},\sigma^2_{1,3,\varepsilon},\sigma^2_{2,1,\varepsilon},\sigma^2_{2,2,\varepsilon},\sigma^2_{2,3,\varepsilon},\sigma^2_{3,1,\varepsilon},\sigma^2_{3,2,\varepsilon},\sigma^2_{3,3,\varepsilon}|\mathbf{Y}_{1,1},$$

$$\ldots,\mathbf{Y}_{1,T},\mathbf{Y}_{2,1},\ldots,\mathbf{Y}_{2,T}] \propto \{\prod_{t=1}^{T}[\mathbf{Y}_{1,t}|\mathbf{Z}_{1,t},\mathbf{Z}_{3,t}][\mathbf{Y}_{2,t}|\mathbf{Z}_{2,t},\mathbf{Z}_{3,t}][\mathbf{Z}_{1,t}|\boldsymbol{\lambda}_{1,t},\mathbf{p}_{1,t}][\mathbf{Z}_{2,t}|\boldsymbol{\lambda}_{2,t},\mathbf{p}_{2,t}]$$

$$\times[\mathbf{Z}_{3,t}|\boldsymbol{\lambda}_{3,t}][\boldsymbol{\lambda}_{1,t}|\mu_1,\mathbf{b}_{1,0},\mathbf{b}_{1,1},\mathbf{b}_{1,2},\mathbf{b}_{1,3},\sigma^2_{1,1,\varepsilon},\sigma^2_{1,2,\varepsilon},\sigma^2_{1,3,\varepsilon}][\boldsymbol{\lambda}_{2,t}|\mu_2,\mathbf{b}_{2,0},\mathbf{b}_{2,1},\mathbf{b}_{2,2},\mathbf{b}_{2,3},\sigma^2_{2,1,\varepsilon},$$

$$\sigma^2_{2,2,\varepsilon},\sigma^2_{2,3,\varepsilon}][\boldsymbol{\lambda}_{3,t}|\mu_3,\mathbf{b}_{3,0},\mathbf{b}_{3,1},\mathbf{b}_{3,2},\mathbf{b}_{3,3},\sigma^2_{3,1,\varepsilon},\sigma^2_{3,2,\varepsilon},\sigma^2_{3,3,\varepsilon}][\mathbf{p}_{1,t}|\alpha_{1,0},\alpha_{1,1},\sigma^2_{1,\eta}][\mathbf{p}_{2,t}|\alpha_{2,0},\alpha_{2,1},$$

$$\sigma^2_{2,\eta}]\}[\alpha_{1,0}][\alpha_{1,1}][\alpha_{2,0}][\alpha_{2,1}][\mu_1][\mu_2][\mathbf{b}_{1,0}][\mathbf{b}_{1,1}][\mathbf{b}_{1,2}][\mathbf{b}_{1,3}][\mathbf{b}_{2,0}][\mathbf{b}_{2,1}][\mathbf{b}_{2,2}][\mathbf{b}_{2,3}][\mathbf{b}_{3,0}][\mathbf{b}_{3,1}][\mathbf{b}_{3,2}]$$

$$\times[\mathbf{b}_{3,3}][\sigma^2_{1,1,\varepsilon}][\sigma^2_{1,2,\varepsilon}][\sigma^2_{1,3,\varepsilon}][\sigma^2_{2,1,\varepsilon}][\sigma^2_{2,2,\varepsilon}][\sigma^2_{2,3,\varepsilon}][\sigma^2_{3,1,\varepsilon}][\sigma^2_{3,2,\varepsilon}][\sigma^2_{3,3,\varepsilon}][\sigma^2_{1,\eta}][\sigma^2_{2,\eta}].$$

This complicated posterior distribution can be numerically evaluated using MCMC methods and in particular the Gibbs sampler based on full-conditional distributions of the unknown processes and parameters. Also, Metropolis-Hastings steps within the Gibbs algorithm are required for the simulation of $\boldsymbol{\lambda}$'s and $\mathbf{p}$'s.

The full-conditional distributions are:

- $[v_{m,it} = \log(\lambda_{m,t}(s_i))|\cdot]$ for t=1,...,T, $m = 1, 2,$ and $i = 1,\ldots,n.$

  M-H step:

  1. Generate $v^*_{m,it} \sim N(v^{(j-1)}_{m,it},\theta)$ at the jth MCMC iteration, and compute ratio:

$$r = \frac{[Y_{m,t}(s_i)|v^*_{m,it}][v^*_{m,it}|\mu^{(j-1)}_m,\mathbf{b}^{(j-1)}_{m,0},\mathbf{b}^{(j-1)}_{m,1},\mathbf{b}^{(j-1)}_{m,2},\mathbf{b}^{(j-1)}_{m,3},\sigma^{2,(j-1)}_{m,1,\varepsilon},\sigma^{2,(j-1)}_{m,2,\varepsilon},\sigma^{2,(j-1)}_{m,3,\varepsilon}]}{[Y_{m,t}(s_i)|v^{(j-1)}_{m,it}][v^{(j-1)}_{m,it}|\mu^{(j-1)}_m,\mathbf{b}^{(j-1)}_{m,0},\mathbf{b}^{(j-1)}_{m,1},\mathbf{b}^{(j-1)}_{m,2},\mathbf{b}^{(j-1)}_{m,3},\sigma^{2,(j-1)}_{m,1,\varepsilon},\sigma^{2,(j-1)}_{m,2,\varepsilon},\sigma^{2,(j-1)}_{m,3,\varepsilon}]}.$$

2. Set $v_{m,it}^{(j)} = v_{m,it}^{*}$ with probability min(r,1); otherwise set $v_{m,it}^{(j)} = v_{m,it}^{(j-1)}$

- $[v_{3,it} = \log(\lambda_{3,t}(s_i))|\cdot]$ for t=1,...,T, and $i = 1, \ldots, n$.

  M-H step:

  1. Generate $v_{3,it}^{*} \sim N(v_{3,it}^{(j-1)}, \theta)$ at the jth MCMC iteration, and compute ratio:

$$r = \frac{[Y_{1,t}(s_i)|v_{3,it}^{*}][Y_{2,t}(s_i)|v_{3,it}^{*}][v_{3,it}^{*}|w^{(j-1)}]}{[Y_{1,t}(s_i)|v_{3,it}^{(j-1)}][Y_{2,t}(s_i)|v_{3,it}^{(j-1)}][v_{3,it}^{(j-1)}|w^{(j-1)}]}.$$

  where $w^{(j-1)} \equiv \{\mu_3^{(j-1)}, \mathbf{b}_{3,0}^{(j-1)}, \mathbf{b}_{3,1}^{(j-1)}, \mathbf{b}_{3,2}^{(j-1)}, \mathbf{b}_{3,3}^{(j-1)}, \sigma_{3,1,\varepsilon}^{2,(j-1)}, \sigma_{3,2,\varepsilon}^{2,(j-1)}, \sigma_{3,3,\varepsilon}^{2,(j-1)}\}$

  2. Set $v_{3,it}^{(j)} = v_{3,it}^{*}$ with probability min(r,1); otherwise set $v_{3,it}^{(j)} = v_{3,it}^{(j-1)}$

- $[P_{m,it} = \text{logit}(p_{m,t}(s_i))|\cdot]$ for t=1,...,T, $m = 1, 2$, and $i = 1, \ldots, n$.

  M-H step:

  1. Generate $P_{m,it}^{*} \sim N(P_{m,it}^{(j-1)}, \theta)$ at the jth MCMC iteration, and compute ratio:

$$r = \frac{[Y_{m,t}(s_i)|P_{m,it}^{*}][p_{m,it}^{*}|\alpha_{m,0}^{(j-1)}, \alpha_{m,1}^{(j-1)}\sigma_{m,\eta}^{2,(j-1)}]}{[Y_{m,t}(s_i)|P_{m,it}^{(j-1)}][p_{m,it}^{(j-1)}|\alpha_{m,0}^{(j-1)}, \alpha_{m,1}^{(j-1)}\sigma_{m,\eta}^{2,(j-1)}]}.$$

  2. Set $P_{m,it}^{(j)} = P_{m,it}^{*}$ with probability min(r,1); otherwise set $P_{m,it}^{(j)} = P_{m,it}^{(j-1)}$

- $[\mathbf{b}_{k,0}|\cdot] \propto \{\prod_{t=1}^{T}[v_{k,it}|\mathbf{b}_{k,0}][\mathbf{b}_{k,0}]$

$$\mathbf{b}_{k,0}|\cdot \sim N(\mathbf{Ab}, \mathbf{A}),$$

  where,

$$\mathbf{A} = (\sum_{t=1}^{T} t^2 \Phi' \Sigma_{\varepsilon}^{-1} \Phi + \Sigma_{b0}^{-1})^{-1},$$
$$\mathbf{b} = \sum_{t=1}^{T} \Phi'(v_{k,t} - \Phi\mu - \Phi B^{*}) + \Sigma_{b0}^{-1}\tilde{\mathbf{b}}_0,$$

where B^* is structured using elements of $\mathbf{b}_k, ix_t$ $(t = 1, \ldots, T)$ and reordered according to the association of each element to one of the ENSO categories (El Niño, La Niña or regular seasons). Σ_ε is constructed similarly using $\sigma^2_{i,\varepsilon}$'s $(i = 1, 2, 3)$.

- $[\mathbf{b}_{k,i}|\cdot] \propto \{\prod_{t=1}^{T}[v_{k,it}|\mathbf{b}_{k,i}][\mathbf{b}_{k,i}]$

$$\mathbf{b}_{k,i}|\cdot \sim N(\mathbf{Ab}, \mathbf{A}),$$

where,

$$\mathbf{A} = ((\mathbf{x}_i \otimes \Phi)'(\mathbf{x}_i \otimes \Phi)/\sigma^2_{k,i,\varepsilon} + \Sigma^{-1}_{bi})^{-1},$$

$$\mathbf{b} = (\text{vec}(v_{k,i,t}) - (\mathbf{1} \otimes \Phi)\mu - (\mathbf{t} \otimes \Phi)\mathbf{b}_{k,0})/\sigma^2_{k,i,\varepsilon}) + \Sigma^{-1}_{bi}\tilde{\mathbf{b}}_i,$$

where $i = 1, 2, 3$.

- $[\alpha_{k,0}|\cdot] \propto \{\prod_{t=1}^{T}[P_{k,it}|\alpha_{k,0}][\alpha_{k,0}]$

$$\alpha_{k,0}|\cdot \sim N(\mathbf{Ab}, \mathbf{A}),$$

where,

$$\mathbf{A} = 1/(T/\sigma^2_\eta + \sigma^2_{\alpha 0}),$$
$$\mathbf{b} = \sum_{t=1}^{T}(P_{k,t} - \alpha_{k,1}x_p) + \tilde{\alpha}_0/\sigma^2_{\alpha 0},$$

- $[\alpha_{k,1}|\cdot] \propto \{\prod_{t=1}^{T}[P_{k,it}|\alpha_{k,1}][\alpha_{k,1}]$

$$\alpha_{k,0}|\cdot \sim N(\mathbf{Ab}, \mathbf{A}),$$

where,

$$\mathbf{A} = 1/(x_p' x_p/\sigma_\eta^2 + \sigma_{\alpha 1}^2),$$
$$\mathbf{b} = \sum_{t=1}^{T} x_p'(P_{k,t} - \alpha_{k,0}) + \tilde{\alpha}_1/\sigma_{\alpha 1}^2,$$

- $[\sigma_{k,i,\varepsilon}^2|\cdot] \propto \prod_{t=1}^{T}[v_{k,it}|\sigma_{k,i,\varepsilon}^2][\sigma_{k,i,\varepsilon}^2]$

$$\sigma_{k,i,\varepsilon}^2|\cdot \sim IG(q,r)$$

where,

$$q = q_{k,i,\varepsilon} + nT_i/2,$$
$$r = (1/r_{k,i,\varepsilon} + 0.5\sum_{t=1}^{T}(v_{k,it} - \Phi\mu_k - \Phi\mathbf{b}_{k,0}t_i - \Phi\mathbf{b}_{k,i}x_{t_i}))^{-1}$$

where $k = 1, 2, 3$, $i = 1, 2, 3$ and t_i indicates the time components associated with the ith ENSO category.

- $[\sigma_{k,\eta}^2|\cdot] \propto \prod_{t=1}^{T}[P_{k,t}|\sigma_{k,\eta}^2][\sigma_{k,\eta}^2]$

$$\sigma_{k,\eta}^2|\cdot \sim IG(q,r)$$

where,

$$q = q_{k,\eta} + nT/2,$$
$$r = (1/r_{k,\eta} + 0.5\sum_{t=1}^{T}(P_{k,t} - \alpha_{k,0} - \alpha_{k,1}X_p))^{-1}$$

where $k = 1, 2$.

Appendix C: Full-Conditionals

Consider the following hierarchical model for the wavelet-based multiresolution model described in Chapter 5 (Method 1):

$$\begin{aligned}
\mathbf{Z}_t &= \mathbf{K}_t\mathbf{W}\mathbf{a}_t + \boldsymbol{\varepsilon_t}, \qquad \varepsilon_t \sim N(\mathbf{0}, \boldsymbol{\sigma}_\varepsilon^2\mathbf{I}), \\
\mathbf{a}_t^j &= \boldsymbol{\Phi}^j\mathbf{b}_t + \boldsymbol{\gamma}_t^j, \qquad \boldsymbol{\gamma}_t^j \sim N(\mathbf{0}, \sigma_{\gamma^j}^2\mathbf{I}), \\
\mathbf{b}_t &= \mathbf{H}_b\mathbf{b}_{t-1} + \boldsymbol{\eta}_t, \qquad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma}_\eta),
\end{aligned}$$

where $\mathbf{W}$ (note that $\mathbf{Y}_t = \mathbf{W}\mathbf{a}_t$) is a matrix of multiresolution wavelet bases, and j denotes resolution scales ($j = 0, \ldots, J$). Let:

$$\mathbf{b}_0 \sim N(\tilde{\mathbf{b}}_0, \boldsymbol{\Sigma}_0^{-1}).$$

Thus, the posterior distribution of interest can now be written as:

$$\begin{aligned}
[\mathbf{a}_1, \ldots, \mathbf{a}_T, \mathbf{b}_0, \ldots, \mathbf{b}_T, \mathbf{H}_b, \boldsymbol{\Sigma}_\eta, \sigma_\varepsilon^2 | \mathbf{Z}_1, \ldots, \mathbf{Z}_T] \propto &\prod_{t=1}^{T}[Z_t|\mathbf{a}_t, \sigma_\varepsilon^2]\prod_{t=1}^{T}[\mathbf{a}_t^j|\mathbf{b}_t, \sigma_{\gamma^j}]\prod_{t=1}^{T}[\mathbf{b}_t|\mathbf{b}_{t-1}, \boldsymbol{\Sigma}_\eta] \\
&\times [\mathbf{b}_0][\sigma_\varepsilon^2][\sigma_{\gamma^j}^2][\boldsymbol{\Sigma}_\eta].
\end{aligned}$$

The full-conditional distributions are:

- $[\mathbf{a}_t|\cdot] \propto \prod_{t=1}^{T-1}[\mathbf{Z}_t|\mathbf{a}_t][\mathbf{a}_t|\mathbf{b}_t]$

$$\mathbf{a}_t|\cdot \sim N(\mathbf{A}_t\mathbf{b}_t, \mathbf{A}_t),$$

where,

$$\mathbf{A}_t = (\mathbf{K}_t'\mathbf{K}_t/\sigma_\varepsilon^2 + \mathbf{I}/\sigma_\gamma^2)^{-1},$$

$$\mathbf{b}_t = (\mathbf{Z}_t'\mathbf{K}_t\mathbf{W}/\sigma_\varepsilon^2 + (\mathbf{b}_t'\boldsymbol{\Phi}')/\sigma_\gamma^2)'.$$

- $[\mathbf{b}_0|\cdot] \propto [\mathbf{b}_1|\mathbf{b}_0][\mathbf{b}_0]$

$$\mathbf{b}_0|\cdot \sim N(\mathbf{A}_0\mathbf{b}_0, \mathbf{A}_0),$$

where,

$$\mathbf{A}_0 = (\mathbf{H}_b'\boldsymbol{\Sigma}_\eta^{-1}\mathbf{H}_b + \boldsymbol{\Sigma}_0^{-1})^{-1},$$

$$\mathbf{b}_0 = (\mathbf{b}_1'\boldsymbol{\Sigma}_\eta^{-1}\mathbf{H}_b + \tilde{\mathbf{b}}_0'\boldsymbol{\Sigma}_0^{-1})'.$$

- $[\mathbf{b}_t|\cdot] \propto \prod_{t=1}^{T-1}[\mathbf{a}_t|\mathbf{b}_t][\mathbf{b}_t|\mathbf{b}_{t-1}][\mathbf{b}_{t+1}|\mathbf{b}_t]$

$$\mathbf{b}_t|\cdot \sim N(\mathbf{A}\mathbf{b}_t, \mathbf{A}),$$

where,

$$\mathbf{A} = (\boldsymbol{\Phi}'\boldsymbol{\Phi}/\sigma_\varepsilon^2 + \boldsymbol{\Sigma}_\eta^{-1} + \mathbf{H}_b'\boldsymbol{\Sigma}_\eta^{-1}\mathbf{H}_b))^{-1},$$

$$\mathbf{b}_t = (\mathbf{a}_t'\boldsymbol{\Phi}/\sigma_\varepsilon^2 + \mathbf{b}_{t-1}'\mathbf{H}_b'\boldsymbol{\Sigma}_\eta^{-1} + \mathbf{b}_{t+1}'\boldsymbol{\Sigma}_\eta^{-1}\mathbf{H}_b)'.$$

- $[\mathbf{u}_T|\cdot] \propto [\mathbf{Z}_T|\mathbf{u}_T][\mathbf{u}_T|\mathbf{u}_{T-1}]$

$$\mathbf{u}_T|\cdot \sim N(\mathbf{A}\mathbf{b}_T, \mathbf{A}),$$

where,

$$\mathbf{A} = (\boldsymbol{\Phi}'\boldsymbol{\Phi}/\sigma_\varepsilon^2 + \boldsymbol{\Sigma}_\eta^{-1}\mathbf{H}_b))^{-1},$$

$$\mathbf{b}_T = (\mathbf{a}_T'\boldsymbol{\Phi}/\sigma_\varepsilon^2 + \mathbf{b}_{T-1}'\mathbf{H}_b'\boldsymbol{\Sigma}_\eta^{-1})'.$$

- $[\sigma_\varepsilon^2|\cdot] \propto \prod_{t=1}^{T}[\mathbf{a}_t|\mathbf{b}_t, \sigma_\varepsilon^2][\sigma_\varepsilon^2]$

$$\sigma_\varepsilon^2|\cdot \sim IG(q, r)$$

where,

$$q = q_\eta + nT/2,$$
$$r = (1/r_\eta + 0.5\sum_{t=1}^{T}((\mathbf{a}_t - \mathbf{\Phi}\mathbf{b}_t)'(\mathbf{G}\mathbf{a}_t - \mathbf{\Phi}\mathbf{b}_t)))^{-1}.$$

- $[\mathbf{\Sigma}_\eta^- 1|\cdot] \propto \prod_{t=1}^{T}[\mathbf{b}_t|\mathbf{b}_{t-1}, \mathbf{\Sigma}_\eta][\mathbf{\Sigma}_\eta^2]$

$$\mathbf{\Sigma}_\eta^2|\cdot \sim \text{Wishart}\left(\left(\sum_{t=1}^{T}((\mathbf{a}_t - \mathbf{\Phi}\mathbf{b}_t)'(\mathbf{G}\mathbf{a}_t - \mathbf{\Phi}\mathbf{b}_t)) + \mathbf{S}\nu\right)^{-1}, \nu + T\right).$$

- $[\mathbf{h}_b \equiv \text{vec}(\boldsymbol{H}_b)|\cdot] \propto [\mathbf{B}_1|\mathbf{B}_0, \mathbf{\Sigma}_\eta][\mathbf{\Sigma}_\eta]$

$$\mathbf{h}_b|\cdot \sim N(\mathbf{A}\mathbf{b}, \mathbf{A}),$$

where,

$$\mathbf{A} = (\mathbf{B}_0 \otimes \mathbf{I})'\mathbf{\Sigma}_\eta^{-1}(\mathbf{B}_0 \otimes \mathbf{I}) + \mathbf{\Sigma}_h^{-1},$$
$$\mathbf{b} = (\mathbf{B}_1 \otimes \mathbf{I})'\mathbf{\Sigma}_\eta^{-1}\text{vec}(\mathbf{B}_0) + \mathbf{\Sigma}_h^{-1}\tilde{\mathbf{h}}_b).$$

and

$$\mathbf{B}_1 \equiv (\mathbf{b}_1, \ldots, \mathbf{b}_T)',$$
$$\mathbf{B}_0 \equiv (\mathbf{b}_0, \ldots, \mathbf{b}_{T-1})'.$$

Appendix D: Full-Conditionals

Consider the following hierarchical model for the spline-based dynamic resolution model described in Chapter 5 (Method 2):

$$\begin{aligned}\mathbf{Z}_t =& \mathbf{K}_t\mathbf{\Phi}(x_t^m, y_t^m)\mathbf{a}_t + \boldsymbol{\varepsilon_t}, \qquad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \boldsymbol{\sigma}_\varepsilon^2\mathbf{I}),\\ \mathbf{a}_t^f =& \mathbf{G}\mathbf{a}_{t-1}^f + \boldsymbol{\eta}_t^f, \qquad \boldsymbol{\eta}_t^f \sim N(\mathbf{0}, \Sigma_{\eta^f}),\end{aligned}$$

where $\mathbf{\Phi} = \mathbf{\Phi}(x_t^m, y_t^m)$ (note that $\mathbf{u} = \mathbf{\Phi}(x_t^m, y_t^m)\mathbf{a}_t$) is a matrix of spline bases (e.g., thin-plate spline bases), and f and m denote fixed and moveable knots, respectively. Let:

$$\mathbf{a}_0^f \sim N(\tilde{\mathbf{a}}_0^f, \mathbf{\Sigma}_0^{-1}).$$

The spline coefficients corresponding to the moveable knots will be sampled non-dynamically. Thus, the following prior distribution is considered

$$\mathbf{a}_t^m \sim N(\tilde{\mathbf{a}}_t^m, \mathbf{\Sigma}_0^{-1}),$$

for $t = 1, \ldots, T$.

The posterior distribution of interest can now be written as:

$$\begin{aligned}[\mathbf{a}_1, \ldots, \mathbf{a}_T, \mathbf{G}, \mathbf{\Sigma}_\eta, \sigma_\varepsilon^2, x_1^m, \ldots, x_T^m, y_1^m, \ldots, y_T^m | \mathbf{Z}_1, \ldots, \mathbf{Z}_T] \propto& \prod_{t=1}^T [Z_t|\mathbf{a}_t, \sigma_\varepsilon^2] \prod_{t=1}^T [\mathbf{a}_t^f | \mathbf{a}_{t-1}^f, \sigma_{\eta^f}]\\ &\times [\mathbf{a}_0^f] \prod_{t=1}^T [\mathbf{a}_t^m] \prod_{t=1}^T [x_t^m | x_{t-1}^m, \sigma_{\nu_x}^2] \prod_{t=1}^T [y_t^m | y_{t-1}^m, \sigma_{\nu_y}^2][x_0^m][y_0^m][\sigma_\varepsilon^2][\sigma_{\eta^f}^2].\end{aligned}$$

The full-conditional distributions are:

- $[\mathbf{a}_0^f|\cdot] \propto [\mathbf{a}_1^f|\mathbf{a}_0^f][\mathbf{a}_0^f]$

$$\mathbf{a}_0^f|\cdot \sim N(\mathbf{A}_0\mathbf{b}_0, \mathbf{A}_0),$$

where,

$$\mathbf{A}_0 = (\mathbf{G}'\mathbf{G}/\sigma_\varepsilon^2 + \mathbf{\Sigma}_0^{-1})^{-1},$$

$$\mathbf{b}_0 = (\mathbf{a}_1^{f'}\mathbf{G}/\sigma_\varepsilon^2 + \tilde{\mathbf{a}}_0^{f'}\mathbf{\Sigma}_0^{-1})'.$$

- $[\mathbf{a}_t^f|\cdot] \propto \prod_{t=1}^{T-1}[\mathbf{Z}_t|\mathbf{a}_t^f][\mathbf{a}_t^f|\mathbf{a}_{t-1}^f][\mathbf{a}_{t+1}^f|\mathbf{a}_t^f]$

$$\mathbf{a}_t^f|\cdot \sim N(\mathbf{A}\mathbf{b}_t, \mathbf{A}),$$

where,

$$\mathbf{A} = (\mathbf{\Phi}'\mathbf{\Phi}/\sigma_\varepsilon^2 + \mathbf{\Sigma}_\eta^{-1} + \mathbf{G}'\mathbf{\Sigma}_\eta^{-1}\mathbf{G}))^{-1},$$

$$\mathbf{b}_t = (\mathbf{Z}_t'\mathbf{\Phi}/\sigma_\varepsilon^2 + \mathbf{a}_{t-1}^{f'}\mathbf{G}'\mathbf{\Sigma}_\eta^{-1} + \mathbf{a}_{t+1}^{f'}\mathbf{\Sigma}_\eta^{-1}\mathbf{G})'.$$

- $[\mathbf{a}_T^f|\cdot] \propto [\mathbf{Z}_T|\mathbf{a}_T][\mathbf{a}_T|\mathbf{a}_{T-1}]$

$$\mathbf{a}_T|\cdot \sim N(\mathbf{A}\mathbf{b}_T, \mathbf{A}),$$

where,

$$\mathbf{A} = (\mathbf{\Phi}'\mathbf{\Phi}/\sigma_\varepsilon^2 + \mathbf{\Sigma}_\eta^{-1}\mathbf{G}))^{-1},$$

$$\mathbf{b}_T = (\mathbf{Z}_T'\mathbf{\Phi}/\sigma_\varepsilon^2 + \mathbf{a}_{T-1}^{f'}\mathbf{G}'\mathbf{\Sigma}_\eta^{-1})'.$$

- $[\sigma_\varepsilon^2|\cdot] \propto \prod_{t=1}^{T}[\mathbf{Z}_t|\mathbf{a}_t, \sigma_\varepsilon^2][\sigma_\varepsilon^2]$

$$\sigma_\varepsilon^2|\cdot \sim IG(q, r)$$

where,

$$q = q_\eta + nT/2,$$
$$r = (1/r_\eta + 0.5\sum_{t=1}^{T}((\mathbf{Z}_t - \mathbf{\Phi}\mathbf{a}_t^f)'(\mathbf{Z}_t - \mathbf{\Phi}\mathbf{a}_t^f)))^{-1}.$$

- $[\boldsymbol{\Sigma}_\eta^{-1}|\cdot] \propto \prod_{t=1}^{T}[\mathbf{a}_t^f|\mathbf{a}_{t-1}^f, \boldsymbol{\Sigma}_\eta][\boldsymbol{\Sigma}_\eta^2]$

$$\boldsymbol{\Sigma}_\eta^{-1}|\cdot \sim \text{Wishart}\left(\left(\sum_{t=1}^{T}((\mathbf{Z}_t - \boldsymbol{\Phi}\mathbf{a}_t)'(\mathbf{Z}_t - \boldsymbol{\Phi}\mathbf{a}_t)) + \mathbf{S}\nu\right)^{-1}, \nu + T\right).$$

- $[\mathbf{g} \equiv \text{vec}(\boldsymbol{G})|\cdot] \propto [\mathbf{B}_1|\mathbf{B}_0, \boldsymbol{\Sigma}_\eta][\boldsymbol{\Sigma}_\eta]$

$$\mathbf{g}|\cdot \sim N(\mathbf{Ab}, \mathbf{A}),$$

where,

$$\mathbf{A} = (\mathbf{B}_0 \otimes \mathbf{I})'\boldsymbol{\Sigma}_\eta^{-1}(\mathbf{B}_0 \otimes \mathbf{I}) + \boldsymbol{\Sigma}_g^{-1},$$

$$\mathbf{b} = (\mathbf{B}_1 \otimes \mathbf{I})'\boldsymbol{\Sigma}_\eta^{-1}\text{vec}(\mathbf{B}_0) + \boldsymbol{\Sigma}_g^{-1}\tilde{\mathbf{g}}).$$

and

$$\mathbf{B}_1 \equiv (\mathbf{a}_1, \ldots, \mathbf{a}_T)',$$

$$\mathbf{B}_0 \equiv (\mathbf{a}_0, \ldots, \mathbf{a}_{T-1})'.$$

- $[x_t^m|\cdot]$ for $t = 1, \ldots, T.$

M-H step:

1. Generate $x_t^{m*} \sim N(x_t^{m,(j-1)}, \theta)$ at the jth MCMC iteration, and compute ratio:

$$r = \frac{[\mathbf{Z}_t|\boldsymbol{\Phi}(x_t^{m*}, y_t^{m,(j-1)}), \mathbf{a}_t][x_t^{m*}|x_{t-1}^{m,(j)}, \nu_x^{(j-1)}]}{[\mathbf{Z}_t|\boldsymbol{\Phi}(x_t^{m,(j-1)}, y_t^{m,(j-1)}), \mathbf{a}_t][x_t^{m,(j-1)}|x_{t-1}^{m,(j)}, \nu_x^{(j-1)}]}.$$

2. Set $x_t^{m,(j)} = x_t^{m*}$ with probability min(r,1); otherwise set $x_t^{m,(j)} = x_t^{m,(j-1)}$

- $[y_t^m|\cdot]$ for $t = 1, \ldots, T.$

M-H step:

1. Generate $y_t^{m*} \sim N(y_t^{m,(j-1)}, \theta)$ at the jth MCMC iteration, and compute ratio:

$$r = \frac{[\mathbf{Z}_t|\mathbf{\Phi}(x_t^{m,(j)}, y_t^{m*}), \mathbf{a}_t][y_t^{m*}|y_{t-1}^{m,(j)}, \nu_y^{(j-1)}]}{[\mathbf{Z}_t|\mathbf{\Phi}(x_t^{m,(j)}, y_t^{m,(j)}), \mathbf{a}_t][y_t^{m,(j-1)}|y_{t-1}^{m,(j)}, \nu_y^{(j-1)}]}.$$

2. Set $y_t^{m,(j)} = y_t^{m*}$ with probability min(r,1); otherwise set $y_t^{m,(j)} = y_t^{m,(j-1)}$

LITERATURE CITED

Abbott, M. B. and D. R. Basco. 1989. *Computational Fluid Dynamics: An Introduction for Engineers)*. John Wiley & Sons.

Aitchison, J. and C. H. Ho. 1989. The multivariate Poisson-log normal distribution. *Biometrika*, 76:643–653.

Anderson, C. J., C. K. Wikle, Q. Zhou, and J. A. Royle. 2007. Population influences on tornado reports in the united states. *Weather and Forecasting*, 22(3):571–579.

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

Berg, R. K. Fish populations of the wild and scenic Missouri River, Montana: Federal aid to fish and wildlife restoration project FW-3-R, job 1A. Technical report, 1981.

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

Berliner, L. M., C. K. Wikle, and R. F. Milliff. *Bayesian Inference in Wavelet Based Models, a Springer Lecture-Notes volume*, chapter Multiresolution wavelet analyses in hierarchical Bayesian turbulence models, pages 341–359. Springer, 1999.

Berry, C. R. J., M. L. Wildhaber, and D. L. Galat. Fish distribution and abundance. population structure and habitat use of benthic fishes along the Missouri and lower Yellowstone Rivers. Technical report, 2005.

Berry, C. R. J. and B. A. Young. Introduction to the benthic fishes studies. population structure and habitat use of benthic fishes along the Missouri and lower Yellowstone Rivers. Technical report, 2001.

Biller, C. and L. Fahrmeir. 2001. Bayesian varying-coefficient models using adaptive regression splines. *Statistical Modelling*, 1:195–211.

Bramblett, R. G. and R. G. White. Habitat use and movements of pallid and shovelnose sturgeon in the Yellowstone and Missouri Rivers in Montana and South Dakota. In *Transactions of the American Fisheries Society*, 130, pages 1006–1025, 2001.

Brumback, B., D. Ruppert, and M. P. Wand. 1999. Comment on "variable selection and function estimation in additive nonparametric regression using a data-based prior" by shively, kohn, and wood. *Journal of the American Statistical Association*, 94:794–797.

Burnett, D. S. 1987. *Finite Element Analysis*. Addison-Wesley Publishing Company.

Campbell, J. T. The Poisson correlation function. In *Proceedings of the Edinburgh Mathematical Society (series 2)*, 4, pages 18–26, 1938.

Carey, G. F. and J. T. Oden. 1984. *Finite Elements: Computational Aspects, Volume III*. Prentice-Hall.

Carlin, B. P., A. E. Gelfand, and A. F. M. Smith. 1992. Hierarchical bayesian analysis of changepoint problems. *Applied Statistics*, 41(2):389–405.

Casella, G. and E. I. George. 1992. Explaining the gibbs sampler. *The American Statistician*, 46:167–174.

Celeux, G., F. Forbes, C. P. Robert, and D. M. Titterington. 2003. Deviance information criteria for missing data models. *Cahiers du Ceremade*, 0325.

Chib, S. and R. Winkelmann. 2001. Markov chain monte carlo analysis of correlated count data. *Journal of Business and Economic Statistics*, 19(4):428–435.

Clarke, K. R. and R. H. Green. 1988. Statistical design and analysis for a "biological effects" study. *Marine Ecology Progress Series*, 46:213–226.

Congdon, P. 2001. *Bayesian statistical modelling*. Wiley.

Congdon, P. 2005. *Bayesian Models for Categorical Data*. Wiley.

Crainiceanu, C. M., D. Ruppert, and M. P. Wand. 2006. Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statisical Software*, 14(14):159–175.

Cressie, N. and H.-C. Huang. 1999. Classes of nonseparable spatio-temporal stationary covariance functions. *Journal of American Statistical Association*, 94:1330–1340.

Daubechies, I. 1988. Orthonormal bases of compactly supported wavelets. *Comm. Pure Appl. Math.*, 41:909–996.

Dempster, A., N. Laird, and D. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38.

Denison, D. G. T., B. K. Mallick, and A. F. M. Smith. 1998. Bayesian MARS. *Journal of the American Statistical Association*, 8:337–346.

DeWeaver, E. and S. Nigam. 2002. Linearity in enso's atmospheric response. *Journal of Climate*, 15:24462461.

Dixon, M. and G. Wiener. 1993. TITAN: The thunderstorm identification, tracking, analysis and nowcasting–a radar-based methodology. *Journal of Atmospheric and Oceanic Technology*, 10(6):785–797.

Eubank, R. L., C. Huang, Y. M. Maldonado, N. Wang, S. Wang, and R. J. Buchanan. 2004. Smoothing spline estimation in varying-coefficient models. *Journal of the Royal Statistical Society, Series B*, 66(3):653–667.

Fahrmeir, L. and L. O. Echavarria. 2006. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Applied Stochastic Models in Business and Industry*, 34:1–14.

Fernandes, V. V., A. M. Schmidt, and H. S. Migon. Modelling zero-inflated spatio-temporal processes. Technical report, IM-UFRJ, IBGE, COPPE-UFRJ, Brazil, 2006.

Ferreira, M. A. R., M. West, H. K. H. Lee, and D. Higdon. 2006. Multi-scale and hidden resolution time series models. *Bayesian Analysis*, 1(4):947–968.

Gelfand, A. E. and S. K. Ghosh. 1998. Model choice: A minimum posterior predictive loss approach. *Biometrika*, 85(1):1–11.

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

Golding, B. 2000. Quantitative percipitation forecasting in the UK. *Journal of Hydrology*, 239:286–305.

Greene, W. 2003. *Econometric Analysis, 5th Edition*. Prentice Hall.

Guttorp, P., W. Meiring, and P. D. Sampson. 2006. A space-time analysis of ground-level ozone data. *Environmetrics*, 5(3):241–254.

Hall, D. B. 2000. Zero-inflated Poisson and binomial regression with random effects: A case study. *Biometrics*, 56:1030–1039.

Hannachi, A. 2001. Toward a nonlinear identification of the atmospheric response to ENSO. *Journal of Climate*, 14:2138–2149.

Hartley, H. 1958. Maximum likelihood estimation from incomplete data. *Biometrics*, 14: 174–194.

Harvey, A. and S. J. Koopman. 1993. Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association*, 88(424):1228–1236.

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

Holmes, C. C. and B. K. Mallick. 2003. Generalized nonlinear modeling with multivariate free-knot regression splines. *Journal of the American Statistical Association*, 98(462): 352–368.

Johannesson, G., N. Cressie, and H.-C. Huang. 2007. Dynamic multi-resolution spatial models. *Environmental and Ecological Statistics*, 14(1):5–25.

Johnson, J. T., P. L. Mackeen, A. Witt, E. D. Mitchell, G. J. Stumpf, E. M. D., and K. W. Thomas. 1993. The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm. *Weather and Forecasting*, 13(2):263–276.

Johnson, N. L., S. Kotz, and N. Balakrishnan. 1997. *Discrete Multivariate Distributions*. John Wiley & sons, Inc., New York.

Kocherlakota, S. and K. Kocherlakota. 1992. *Bivariate Discrete Distributions*. Marcel Dekker, New York.

Kwon, Y. W. and H. Bang. 2000. *The Finite Element Method Using MATLAB (Second Edition)*. CRC Press.

Lam, K. F., H. Xue, and Y. B. Cheung. 2006. Semiparametric analysis of zero-inflated count data. *Biometrics*, 62(4):996–1003.

Lambert, D. 1992. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, 34:1–14.

Li, C.-S., J.-C. Lu, J. Park, K. Kim, P. A. Brinkley, and J. P. Peterson. 1999. Multivariate zero-inflated Poisson models and their applications. *Technometrics*, 41(1):29–38.

Liebelt, P. B. 1967. *An Introduction to Optimal Estimation.* Addison-Wesley.

Lindley, D. V. and A. F. M. Smith. 1972. Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Series B*, 34:1–41.

Littell, R. C., G. A. Milliken, W. W. Stroup, R. D. Wolfinger, and O. Schabenberber. 2006. *SAS for Mixed Models, 2nd Edition.* SAS Publishing.

Luo, Z. and G. Wahba. 1998. Spatial-temporal analysis of temperature using smoothing spline ANOVA. *Journal of Climate*, 11:18–28.

Lynch, D. R. 2005. *Numerical Partial Differential Equations for Environmental Scientists and Engineers: A First Practical Course.* Springer.

Lynch, D. R. and D. J. McGillicuddy. 2001. Objective Analysis for coastal regimes. *Contin. Shelf Res.*, 21:1299–1315.

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

Martin, T. G., B. A. Wintle, J. R. Rhodes, P. M. Kuhnert, S. A. Field, S. J. Low-Choy, A. J. Tyre, and H. P. Possigham. 2005. Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. *Ecology Letters*, 8:1235–1246.

McCullagh, P. and J. A. Nelder. 1989. *Generalized Linear Models, 2nd edition.* Chapman & Hall/CRC.

Moilanen, A. 2002. Implications of empirical data quality for metapopulation model parameter estimation and application. *Oikos*, 84:516–530.

Nowak, R. D. Multiscale hidden markov models for bayesian image analysis. Technical report, MSU Technical Report MSU-ENGR-004-98, 1998.

Nychka, D., C. Wikle, and J. Royle. 2002. Multiresolution models for nonstationary spatial covariance functions. *Statistical Modelling: An International Journal*, 2:315–331.

O'Connell, M. A. and R. D. Wolfinger. 1997. Spatial regression models, response surfaces, and process optimization. *Journal of Computational and Graphical Statistics*, 6(2):224–241.

Okubo, A. and S. A. Levin. 2002. *Diffusion and Ecological Problems.* Springer.

Pierce, C., C. G. Collier, P. J. Hardaker, and C. M. Haggett. 2000. Gandolf: a systeem for generating automated nowcasts of convective percipitation. *Meteorological Applications*, 8:341–360.

Robert, C. P. and G. Casella. 2005. *Monte Carlo Statistical Methods.* Springer-Verlag.

Royle, J. A. 2006. Site occupancy models with heterogeneous detection probabilities. *Biometrics*, 62(1):97–102.

Ruelle, R. and K. Keenlyne. The suitability of shovelnose sturgeon as a pallid sturgeon surrogate. In *Proceedings of the South Dakota Academy of Science*, 73, pages 67–82, 1994.

Ruppert, D., M. P. Wand, and R. J. Carroll. 2003. *Semiparanetric Regression*. Cambridge University Press.

Seed, A. W. 2003. A dynamic and spatial scaling approach to advection forecasting. *Journal of Applied Meteorology*, 42:381–388.

Shumway, R. H. and D. S. Stoffer. 2005. *Time Series Analysis and Its Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Skalski, G. T. and J. F. Gilliam. 2000. Modeling diffusive spread in a heterogenous popoulation: A movement study with stream fish. *Ecology*, 81(6):1685–1700.

Smith, M. and R. Kohn. 1997. A Bayesian approach to nonparametric bivariate regression. *Journal of the American Statistical Association*, 92(440):1522–1535.

Sparrevohn, C. R., A. Nielsen, and J. G. Støttrup. 2002. Diffusion of fish from a single release point. *Canadian Journal of Fisheries and Aquatic Sciences*, 59(6):844–853.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. V.der Linde. 2002. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B*, 64(4):583–616.

Spiegelhalter, D. J., A. Thomas, N. G. Best, and D. Lunn. WinBUGS version 1.4. Technical report, Imperial College and MRC Biostatistics Unit, London, 2003.

Stein, M. L. 1986. A simple model for spatial-temporal processes. *Water Resources Research*, 22:2107–2110.

Stein, M. L. 2005. Space-time covariance functions. *Journal of American Statistical Association*, 100:320–321.

Stroud, J. R., P. Muller, and B. Sanso. 2001. Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society, Series B*, 63:673–689.

Tong, H. 1993. *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press.

Trenberth, K. E. 1997. The defenition of El Niño. *Bulletin of the American Meteorological Society*, 78:2771–2777.

Tsionas, E. G. 2001. Zero-inflated Poisson regression, with an application to defects in manufacturing. *Communications in Statistics: Theory and Methods*, 30(2):243–255.

van der Linde, A., K.-H. Witzko, and K.-H. Jockel. 1995. Spatial-temporal analysis of mortality using splines. *Biometrics*, 41(4):1352–1360.

Wang, K., K. K. Yau, and A. H. Lee. 2002. A zero-inflated Poisson mixed model to analyze diagnosis related groups with majority of same-day hospital stays. *Computer Methods and Programs in Biomedicine*, 68(3):195–203.

Welsh, A., R. Cunningham, C. Donnelly, and D. Lindenmayer. 1996. Modelling the abundance of rare species- statistical models for counts with extra zeros. *Ecological Modelling*, 88:297–308.

West, M. and J. Harrison. 1989. *Bayesian Forecasting and Dynamic Models*. Springer Verlag, New York.

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

Wikle, C. K. and C. J. Anderson. 2003. Climatological analysis of tornado report counts using a hierarchical bayesian spatio-temporal model. *Journal of Geophysical Research-Atmospheres*, 108(D24):9005. doi:10.1029/2002JD002806.

Wikle, C. K., L. M. Berliner, and N. Cressie. 1998. Hierarchical Bayesian space-time models. *J. Envr. Ecol. Stat.*, 5:117–154.

Wikle, C. K., L. M. Berliner, and R. F. Milliff. 2003. Hierarchical bayesian approach to boundary value problems with stochastic boundary conditions. *Monthly Weather Review*, 131(6):1051–1062.

Wikle, C. K. and N. Cressie. 1999. A dimension-reduced approach to space-time kalman filtering. *Biometrika*, 86(4):815–829.

Wikle, C. K., R. F. Millif, D. Nychka, and L. M. Berliner. 2001. Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds. *Journal of American Statistical Association*, 96:382–397.

Wunsch, C. 1996. *The Ocean Circulation Inverse Problem*. Cambridge University Press.

Xu, K., C. K. Wikle, and N. Fox. 2005. A kernel based spatio temporal dynamical model for nowcasting radar precipitation. *Journal of the American Statistical Association*, 100: 1133–1144.

Zabel, R. 2002. Using "travel time" data to characterize the behavior of migrating animals. *The American Naturalist*, 159(4):372–387.

Zabel, R. W. and J. J. Anderson. 1997. A model of the travel time of migrating juvenile salmon, with an application to Snake River spring chinook. *North American Journal of Fisheries Management*, 17:93–100.

Zienkiewicz, O. C. and R. L. Taylor. 2000. *The Finite Element Method: Volume 3- Fluid Dynamics)*. Butterworth-Hienemann.

VITA

Ali Arab was born in Tehran, Iran, on July 13, 1975. Son of Elaheh and Hossein Arab, he was raised in Tehran. He earned his B.S. from Iran University of Science and Technology (1999) in Applied Mathematics and Operations Research. He moved to the United States on January 2000 and pursued a M.S. degree in Mathematics and Statistics at the Southern Illinois University- Edwardsville under Steven R. Rigdon. Upon completion of his Masters degree in 2002, Ali 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 an Assistant Professor of Mathematics in the Department of Mathematics at Georgetown University in Washington D.C., and will begin in August of 2007.