Scalable Methods for Modeling Dynamic Spatio-Temporal Data

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

By

Yu-Lin Hsu
Bachelor of Mathematics
National Central University, Taiwan, 2019

Director: Dr. Seiyon (Ben) Lee, Professor
Department of Statistics

Summer Semester 2022
George Mason University
Fairfax, VA

Dedication

This is dedicated to my parents, two sisters, my boyfriend, my academic advisor Dr. Seiyon (Ben) Lee, all my committee members and department of Statistics, for all of their support and encouragement.

Acknowledgments

I would like to express my gratitude to my supervisor, Professor Seiyon (Ben) Lee, for his guidance and patience. When I was in the first year of my master's degree in the statistics department, even though I was not a capable student at that moment, Professor Lee guided me through beginning research in spatio-temporal statistics. Without his support, I could not understand the knowledge through books. He helped me break down the problem into small parts, so I could practice implementing Spatio-Temporal statistics in R from easy to hard problems. He is a patient and knowledgeable mentor. He clearly wrote all the formulas on the whiteboard every time I was unclear on the concept. He also encourages me even I made mistakes in experiments. I am so fortunate to work with him.

A special thanks to my boyfriend. He taught me some math, like partial differential equations and Fourier series. He is a PhD in Computer Science in Graphics, so he shared with me what he learn which also can be applied in my field. He stayed with me when I was stressed or depressed. He gave me a lot of encouragement and happiness.

Finally thanks my family for their love and support without whom I would never have enjoyed so many opportunities. Although they were not with me, they often cared about me and made me feel warm over the phone.

Table of Contents

Page

List of Tables . . . vii
List of Figures . . . viii
Abstract . . . ix
1 Introduction . . . 1
2 Statistical Spatial and Spatio-Temporal Models . . . 4
2.1 Spatial linear mixed model (SLMM): . . . 4
2.2 Spatial generalized linear mixed model (SGLMM): . . . 5
2.3 Generalized linear mixed model with Spatio-Temporal random effect: . . . 6
2.3.1 Separable spatio-temporal covariance function . . . 8
2.4 Random Effects with Spatio-Temporal Basis Functions: . . . 9
2.5 Dynamic Spatio-Temporal Models (DSTMs): . . . 9
3 Bayesian statistics and Markov Chain Monte Carlo . . . 12
3.1 Overview of Bayesian Statistics . . . 12
3.2 Markov chain Monte Carlo (MCMC) algorithms . . . 14
3.2.1 Metropolis-Hasting Algorithm . . . 15
3.2.2 Gibbs sampling . . . 16
4 Basis Functions . . . 17
4.1 Polynomial Basis Functions . . . 17
4.2 Wavelets . . . 18
4.3 Radial Basis Functions . . . 18
4.4 Fourier Basis Functions . . . 19
4.5 Eigenvector basis functions . . . 20
5 Dynamic Spatio-temporal Models for Animal Movement . . . 21
5.1 Hierarchical Dynamic Spatio-Temporal Bayesian model: . . . 21
5.2 Our Proposed Method . . . 25
5.2.1 Dimension-Reduction via Basis Representation . . . 25
5.2.2 Spatial and Spatio-temporal Correlation . . . 26
6 Simulation Study . . . 27

6.1 Design . . . 27
6.2 Implementation . . . 27
6.2.1 Proposed Method . . . 28
6.2.2 Separable Spatio-temporal Generalized Linear Mixed Model (Separable SGLMM) . . . 29
6.3 Results . . . 30
6.4 Comparison . . . 33
7 Real World Example . . . 36
7.1 Eurasian collared dove Prevalence from 2001-2010 . . . 36
7.2 Models . . . 36
7.3 Results . . . 38
7.4 Comparison . . . 41
8 Discussion . . . 45
8.1 Summary of Conclusions . . . 45
8.2 Caveats and Directions for Future Research . . . 46
Bibliography . . . 47

List of Tables

Table Page

6.1 Parameter estimation of simulated data by our model. 32

6.2 Parameter estimation of simulated data by separable model. 34

6.3 Results of simulated data by our model. 35

6.4 Results of simulated data by separable model. 35

7.1 Parameter estimation of real data by our model. 40

7.2 Parameter estimation of real data by separable model. 43

7.3 Results of real data by our model. 43

7.4 Results of real data by separable model. 43

List of Figures

Figure Page

4.1 Different types of Basis Functions. 19

4.2 The eigenvector basis functions are extracted by the covariance matrix of the real-world ECD data. 20

6.1 A series of maps of generated data $\boldsymbol{u}_t$ versus estimated $\tilde{\boldsymbol{u}}_t = \boldsymbol{\Phi}\boldsymbol{\gamma}_t$ from our model from time 0 (the initial condition) to 10. 30

6.2 Trace plots and posterior distribution of parameters σ_δ^2, β_0, β_1, σ_η^2, $log(\alpha_0)$, $log(\alpha_1)$ of simulated data. 31

6.3 The estimated diffusion coefficient $\boldsymbol{\delta}$ of simulated data using our model. . . 32

6.4 A series of maps of generated data $\boldsymbol{u}_t$ versus estimated $\tilde{\boldsymbol{v}}_t = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{w}_t$ from separable model from time 1 to 10. 33

6.5 Trace plots and posterior distribution of parameters σ_S^2, a_S, σ_T^2, a_T, β_0, β_1 of simulated data. 34

7.1 A series of maps of $log(\boldsymbol{\lambda}_t)$ and $\boldsymbol{\lambda}_t$, estimated by our model, versus the observed dove counts from 2001 to 2010. 38

7.2 Trace plots and posterior distribution of parameters σ_δ^2, β_0, β_1, σ_η^2, $log(\alpha_0)$, $log(\alpha_1)$ of ECD data. 39

7.3 The estimated diffusion coefficient $\boldsymbol{\delta}$ of ECD data using our model. 40

7.4 A series of maps of $log(\boldsymbol{\lambda}_t)$ and $\boldsymbol{\lambda}_t$, estimated by separable model, versus the observed dove counts from 2001 to 2010. 41

7.5 Trace plots and posterior distribution of parameters σ_S^2, a_S, σ_T^2, a_T, β_0, β_1 of ECD data. 42

Abstract

SCALABLE METHODS FOR MODELING DYNAMIC SPATIO-TEMPORAL DATA

Yu-Lin Hsu

George Mason University, 2022

Thesis Director: Dr. Seiyon (Ben) Lee

Hierarchical spatio-temporal models have been developed to model complex datasets exhibiting spatio-temporal (ST) autocorrelation; however, many of these models are purely descriptive and do not explicitly model the underlying dynamic processes. Animal movement or general movement behaviors are examples of such dynamic processes; that is, animals, or agents, move from one place to another over time, and their migration behavior can change with time and as well as their current (and past) locations. The motivating example for this thesis aims to model the spatio-temporal movement of the Eurasian collared-dove within the continental United States from 2001-2010. Existing studies have modeled animal movement using a reaction-diffusion equation or other systems of differential equation. Recently, dynamic spatio-temporal models (DSTMs) have incorporated these physical processes into a Bayesian hierarchical modeling framework. While DSTMs are extremely flexible, they can be computationally costly to fit and do not scale well to high-dimensional observations.

In this thesis, I propose a computationally-efficient method to fit DSTMs to large space-time count-valued datasets. The proposed scalable DSTM utilizes spatial basis functions to summarize the high-dimensional data as well as a spatial interpolator to assimilate observations at irregularly-spaced locations.

I demonstrate the approach on simulated examples as well as a real-world dataset that tracks the prevalence of the Eurasian collared dove. Through a comparative analysis, the proposed approach is evaluated against a competing method with respect to goodness-of-fit and uncertainty quantification. In addition, I compare the model-fitting walltimes to assess the associated computational costs. The thesis concludes with a summary of the main contributions, discussion of key limitations, and directions for future research.

Chapter 1: Introduction

Spatio-temporal (ST) datasets are characterized by a collection of observations indexed at specific geographic locations and time points. ST data is prevalent across many fields such as epidemiology [1], criminology [2], transportation [3], and ecology [4]. There are several types of ST data; for example, discrete events occurring at distinct locations and times [5], movement bodies' trajectory [6], point-referenced [7], and rasters [8]. Modern spatio-temporal analyses include spatio-temporal clustering [9], detecting anomalous data [10], uncovering spatio-temporal patterns and changes [8,11], creating scalable methods to analyze large datasets [12], and developing complex statistical models for dynamic spatio-temporal data [13].

One salient characteristic of ST data is the spatio-temporal correlation between the observations; that is, observations are correlated with nearby observations in space or/and time [8]. Neglecting this dependence may result in improper inference and inaccurate predictions [14]. Another property of ST data is heterogeneity or non-stationarity [15], where the spatio-temporal dependence structures vary across regions and time.

Movement data among animal populations can be characterized as ST data because the individual agents or animals move from location-to-location across time. There are many motivating factors driving animal movement. Animal species may traverse regions or exhibit movement in order to obtain resources, avoid predators and competitors, and look for mates [16–18]. Building animal movement models can help us understand their ecology, life history and behavior, as well as aid conservation efforts [19]. To account for spatial or temporal animal movement trends, one can fit a species distribution model to the ST data [20]. For example, [21] applied process convolution models to simulate the migration routes of greater white-fronted geese and sandhill cranes. To manage and conserve wildlife habitats, a past

study [22] analyzed the preferred habitat of a whooping crane by an inhomogeneous Poisson point process model. Owing to the small size of pygmy rabbits, [23] used rabbits' burrows to predict their spread via hierarchical spatial models.

In ther literature, there exists a variety of species distribution models such as profile techniques [24], machine learning methods [25], and statistical regression models [26,27]. In this thesis, I employ the general class of spatial regression-based species distribution models, which will be discussed in Chapter 2. More specifically, I will focus on dynamic spatio-temporal models (DSTMs), a special class of statistical spatio-temporal models developed to understand population-level movement for individual species. Due to the complexity of the model, a large number of model parameters, missing data, and the possibility of non-convex objective functions, I have chosen to implement the DSTMs within the Bayesian framework using a sampling-based inference (e.g. Markov Chain Monte Carlo (MCMC)).

To understand the migration behavior of the Eurasian collared doves (ECD), I build upon the work in [28] to develop a computationally-efficient method to fit hierarchical dynamic spatio-temporal Bayesian models. The computational speedup is facilitated by the use of spatial basis functions embedded within a dynamic auto-regressive spatio-temporal model. More specifically, I utilize the class of eigenvector basis extension method. The key benefits of using spatial basis functions are: (1) basis functions can approximate the spatial processes by inferring fewer unknown parameters, as opposed to the 'gold standard' models; (2) basis functions can drastically reduce correlation among the unknown model parameters, which results in faster mixing MCMC algorithm [29]; and (3) bypasses very large matrix operations (e.g. Cholesky decompositions, inverses, and determinants).

The thesis is structured as follows. In Chapter 2, I will introduce common spatial and spatio-temporal geostatistical models. Chapter 3 will provide an overview of Bayesian inference and details of the MCMC algorithm. Next, I will discuss commonly-used spatial and spatio-temporal basis functions in Chapter 4. In Chapter 5, I will discuss the dynamic spatio-temporal models developed in Hooten and Wikle (2008) [28] and introduce our proposed method. In Chapters 6 and 7, experiments with simulated and real-world data are

presented, respectively, as well as comparisons with a competing approach. I conclude the thesis with a summary and directions for future research in Chapter 8.

Chapter 2: Statistical Spatial and Spatio-Temporal Models

Datasets consisting of spatio-temporally correlated observations, such as species distribution, can be modeled with a wide array of geostatistical spatial and spatio-temporal models. In this chapter, I introduce a commonly-used subset of these models, including the spatial generalized linear mixed models (SGLMMs) [29, 30], separable models, basis function models, and dynamic spatio-temporal models. To begin, I will introduce a simple geostatisical model for spatial observations - the spatial linear mixed model.

2.1 Spatial linear mixed model (SLMM):

Let $\boldsymbol{Z} \equiv (Z(s_1), ..., Z(s_n))'$ be a set of n observations at fixed spatial locations, $\boldsymbol{Y} \equiv (Y(s_1), ..., Y(s_n))'$ be the processes. Let s_i be a spatial location observed in some spatial domain $D_s \subseteq R^d$, a subset of d-dimensional Euclidean space. The data $Z(s_i)$ consists of latent space processes $Y(s_i)$ and measurement errors τ^2, and is given by:

$$Z(s_i)|Y(s_i), \tau^2 \sim N(Y(s_i), \tau^2).$$

Let $x(s_i)$ be a p-dimensional vector of covariates with the associated fixed effects $\boldsymbol{\beta}$, and $\boldsymbol{\eta} = (\eta(s_1), ..., \eta(s_n))$ be the unobserved spatial random effect to capture the spatial dependence. Define the linear mixed model with spatial random effect as follows:

$$Y(s_i) = x(s_i)'\boldsymbol{\beta} + \eta(s_i).$$

For the continuous spatial domain, $\boldsymbol{\eta}|\boldsymbol{\theta}$ can be modeled as a Gaussian process with the covariance function $\boldsymbol{\Sigma}_S(\boldsymbol{\theta})$ and covariance function parameters $\boldsymbol{\theta} = (\sigma^2, \rho, \nu)$. Here, $\boldsymbol{\Sigma}_S(\boldsymbol{\theta})$

generates a symmetric positive definite covariance matrix where the point-wise covariance decays over longer distances. One example of covariance function $\boldsymbol{\Sigma}_S(\boldsymbol{\theta})$ is the Matérn class [31], which generates the following isotropic and stationary covariance matrix:

$$\boldsymbol{\Sigma}_S(\sigma^2, \rho, \nu; d) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \big(\sqrt{2\nu}\frac{d}{\rho}\big)^{\nu} K_{\nu}\big(\sqrt{2\nu}\frac{d}{\rho}\big), \tag{2.1}$$

where ρ and ν are parameters of the covariance, K_ν is a Bessel function, and d is the Euclidean distance between locations.

For areal data, conditional autoregressive (CAR) [32] and simultaneous autoregressive (SAR) models [33] can be used to model the spatial random effects $\boldsymbol{\eta}$. Instead of a covariance function 2.1, the CAR model will typically use a precision matrix based upon the neighborhood structure of the areal units. Please see [33] and [34] for additional details on areal models.

2.2 Spatial generalized linear mixed model (SGLMM):

The SGLMM is the extension of the SLMM with non-Gaussian responses. Let $\boldsymbol{Z} \equiv (Z(s_1), ..., Z(s_n))'$ be a set of n observations at fixed spatial locations, $\boldsymbol{Y} \equiv (Y(s_1), ..., Y(s_n))'$ be the processes. Let s_i be a spatial location observed in some spatial domain $D_s \subseteq R^d$, a subset of d-dimensional Euclidean space. The data $Z(s_i)$ consists of latent space processes $Y(s_i)$ and measurement errors τ^2, and is given by:

$$Z(s_i)|Y(s_i), \tau^2 \sim \mathcal{F}(Y(s_i), \tau^2).$$

Let $x(s_i)$ be a p-dimensional vector of covariates with the associated fixed effects $\boldsymbol{\beta}$, and $\boldsymbol{\eta} = (\eta(s_1), ..., \eta(s_n))$ be the unobserved spatial random effect to capture the spatial dependence. Define the generalized linear mixed model with spatial random effect as follows:

$$E(Z(s_i)|Y(s_i)) = g^{-1}(x(s_i)\boldsymbol{\beta} + \eta(s_i)), \tag{2.2}$$

where $g(.)$ is a link function. For example, binary responses can use a probit and logit link function and count responses use a log link function. Similar to the SLMM, the spatial random effects $\boldsymbol{\eta}|\boldsymbol{\theta}$ are modeled as a Gaussian process with the covariance functions $\boldsymbol{\Sigma}_S(\boldsymbol{\theta})$.

SGLMMs can make inference about the regression coefficients and predict at unobserved locations. However, SGLMMs are not suitable for time dependence or space-time interaction data. Besides, the models may have spatial confounding between fixed and random effects, which possible to inflate the variance [35]. Finally, the high-dimensional spatial random effects can be computational demanding to model [31,36].

Spatio-temporal models can capture both spatial and temporal processes simultaneously to account for the spatial and temporal interaction. Spatio-temporal models can be descriptive [37] or dynamic [38,39].

2.3 Generalized linear mixed model with Spatio-Temporal random effect:

Let $\mathbf{Z} \equiv (Z(s_1;t_1), ..., Z(s_n;t_m))'$ be a set of $n \times m$ observations, $\mathbf{Y} \equiv (Y(s_1;t_1), ..., Y(s_n;t_m))'$ be the processes. Let s_i be a spatial location observed in some spatial domain $D_s \subseteq R^d$, a subset of d-dimensional Euclidean space, and t_j be a temporal location observed in some temporal domain D_t, the one-dimensional real line. The data $Z(s_i;t_j)$ consists of latent space-time processes $Y(s_i;t_j)$ and measurement errors τ^2, and is given by:

$$Z(s_i;t_j)|Y(s_i;t_j), \tau^2 \sim \mathcal{F}(Y(s_i;t_j), \tau^2).$$

Let $x(s_i;t_j)$ be a p-dimensional vector of covariates for the fixed effects, and $\boldsymbol{\eta} = (\eta(s_1;t_1), ..., \eta(s_n;t_m))'$ is the random effect to capture the spatio-temporal dependence. Define the generalized linear mixed model with spatio-temporal random effect as follows:

$$E(Z(s_i;t_j)|Y(s_i;t_j)) = g^{-1}(x(s_i;t_j)'\boldsymbol{\beta} + \eta(s_i;t_j)). \tag{2.3}$$

The random effects $\boldsymbol{\eta}|\boldsymbol{\theta}$ are modeled as a Gaussian process with the covariance functions $\boldsymbol{\Sigma}_{ST}(\boldsymbol{\theta})$. The random effects of descriptive spatio-temporal models have several types of valid spatial–temporal covariance functions [40, 41].

There are four classes of spatio-temporal covariance functions $\boldsymbol{\Sigma}_{ST}(\boldsymbol{\theta}) = Cov(Y(s_i;t_j), Y(s_{i'};t_{j'}))$. First, separable in space and time covariance functions [42, 43] can make the computation much easier, but scientists only can use this when no interaction across time and space. I will talk more about the model in the subsection and apply it to simulated and real data experiments, for comparing with our model. The third one is sums-and-products formulation which can easily prove that the covariance function is non-negative-definite [44]. A special case of it is the separable model. Fourth, a sophisticated spatial and temporal interaction could be modeled by a valid non-separable covariance function [45]. Last, stochastic partial differential equation approaches can simulate the random spatio-temporal dynamic processes [46].

The last two covariance functions are flexible and can show spatial and temporal interaction [13]. However, the model would be hard to implement because of the high-dimensional spatio-temporal random effects [47]. In addition, the covariance functions are difficult to scientifically interpret, or even unrealistic [48, 49]. Due to these limitations, alternative spatio-temporal models have been developed such as spatio-temporal basis expansions [13] and dynamic spatio-temporal models [50].

2.3.1 Separable spatio-temporal covariance function

The separable spatio-temporal covariance, mentioned in Wikle's book [39], is a spatio-temporal model, but it doesn't have interaction across space and time. The covariance function can be written as:

$$Cov(Y(s_i;t_j),Y(s_{i'};t_{j'})) = R_S \otimes R_T.$$

R_S and R_T are the spatial and temporal covariance matrices, respectively. These two covariance matrices can be interpreted with the Matérn covariance function with $\nu = 1/2$:

$$R_S = \sigma_S^2 exp(-\frac{1}{a_S}||h||),$$

$$R_T = \sigma_T^2 exp(-\frac{1}{a_T}|\tau|).$$

Here, $||h||$ is the distance between s_i and $s_{i'}$, and $|\tau|$ is the lag between t_j and $t_{j'}$. The parameters a_S and a_T represent spatial- and temporal-dependence, and the parameters σ_S^2 and σ_T^2 are the variance.

To obtain the likelihood function, the inverse of the covariance matrix should be calculated. However, the matrix is really large, if we have S locations and T time points, it is a $ST \times ST$ matrix. Fortunately, instead of computing a large inverse covariance matrix, it is possible to compute $R_S^{-1} \otimes R_T^{-1}$, namely the $S \times S$ and $T \times T$ matrices, respectively. Therefore, the separable spatio-temporal covariance function is still used.

2.4 Random Effects with Spatio-Temporal Basis Functions:

Another popular method is the SGLMM that incorporates spatio-temporal basis functions to model the spatio-temporal random effects:

$$E(Z(s_i;t_j)|Y(s_i;t_j)) = g^{-1}(x(s_i;t_j)'\boldsymbol{\beta} + \eta(s_i;t_j)) \tag{2.4}$$

$$= g^{-1}(x(s_i;t_j)'\beta + \sum_{k=1}^{K} \phi_k(s_i;t_j)\alpha_k + \nu(s_i;t_j)) \tag{2.5}$$

where $\phi_k(s_i;t_j)$ is a spatio-temporal basis function, α_k is a random effect, $\nu(s_i;t_j)$ is a small-scale spatio-temporal random effect not captured by the basis functions. The spatio-temporal basis functions $\boldsymbol{\Phi} \equiv [\phi_1, ..., \phi_K]$ can account for non-separable spatio-temporal dependence.

Basis representation methods are a computationally efficient, yet flexible approach to modeling spatio-temporal data. However, these models are primarily descriptive [51], meaning that the data is modeled based on only observations and omit any of the scientific processes such as diffusion or repulsion.

2.5 Dynamic Spatio-Temporal Models (DSTMs):

Alternatively, dynamic modeling is a conditional probability distribution. The evolution of spatial processes can be discovered in the real world over time [52]. Instead of considering the high-dimensional spatio-temporal covariance functions, the dependence of the dynamic model is incorporated with the evolution of processes from physics, chemistry, biology, and economics [13]. The processes are specifically interpretable because the model considers the Markov chain of the first or more order [52]. The model is more like a time series rather than a spatial process [39]. However, there is a curse of dimensionality in the model [52]. That is, dynamic models will be over-parameterized [52].

For spatio-temporal observations $\boldsymbol{Z}_t = (Z_t(s_1), Z_t(s_2), ..., Z_t(s_{n_t}))'$:

$$\boldsymbol{Z}_t = H_t(Y_t(\mathbf{s}), \boldsymbol{\theta}_t, \epsilon_t(\mathbf{s})), t = 1, ..., T \tag{2.6}$$

where $\boldsymbol{Z}_t$ is the data at time t, $Y_t(\mathbf{s})$ is the latent spatio-temporal process at time t, and H_t can be an independent, linear or nonlinear function to provide the relationship between the data and the latent process, $\boldsymbol{\theta}_t$ are spatially and/or temporally data-model parameters, and ϵ_t are the measurement errors.

The process evolution model:

$$Y_t(\mathbf{s}) = M(Y_{t-1}(\mathbf{s}), \tilde{\boldsymbol{\theta}}_t, \eta_t(\mathbf{s})), \tag{2.7}$$

where M is an linear or nonlinear evolution function, and $\tilde{\boldsymbol{\theta}}_t$ are evolution model parameters, and η_t is a spatial white-noise process.

The are two assumption of independence in DSTMs:

First, $\boldsymbol{Z}_t$ are independent conditional on $Y_t(\mathbf{s})$ and $\boldsymbol{\theta}_t$, so it yields

$$[\{\boldsymbol{Z}_t\}_{t=1}^T | \{Y_t(\mathbf{s})\}_{t=1}^T, \{\boldsymbol{\theta}_t\}_{t=1}^T] = \prod_{t=1}^T [\boldsymbol{Z}_t | Y_t(\mathbf{s}), \boldsymbol{\theta}_t] \tag{2.8}$$

To illustrate the evolution of spatial process in time, I can decompose the process model $Y_t(\mathbf{s})$ to be multiple conditional distributions,

$$[Y_0(\mathbf{s}), Y_1(\mathbf{s}), ..., Y_T(\mathbf{s})] = [Y_T(\mathbf{s}) | Y_{T-1}(\mathbf{s}), Y_{T-2}(\mathbf{s}), ..., Y_0(\mathbf{s})] \times$$

$$[Y_{T-1}(\mathbf{s}) | Y_{T-2}(\mathbf{s}), Y_{T-3}(\mathbf{s}), ..., Y_0(\mathbf{s})] \times ... \times [Y_1(\mathbf{s}) | Y_0(\mathbf{s})] \times [Y_0(\mathbf{s})]$$

The second assumption is Markov property. First-order Markov property means that the current state (t) which conditional on past states (t-1, t-2,..., 1) only depends on the last

time in past (t-1), not all the past state. Hence, I have

$$[Y_0(\mathbf{s}), Y_1(\mathbf{s}), ..., Y_T(\mathbf{s})|\{\tilde{\boldsymbol{\theta}}_t\}_{t=1}^T] = \prod_{t=1}^{T}[Y_t(\mathbf{s})|Y_{t-1}(\mathbf{s}), \tilde{\boldsymbol{\theta}}_t])[Y_0(\mathbf{s})|\tilde{\boldsymbol{\theta}}_0] \tag{2.9}$$

Chapter 3: Bayesian statistics and Markov Chain Monte Carlo

Bayesian statistics [53,54] appears earlier than Frequentist statistics, but is less popular than Frequentist statistics. The reason is that there is no method to deal with an integration problem that has high dimensional parameters in Bayesian statistics. However, the Markov Chain Monte Carlo (MCMC) algorithm is introduced to finally make the Bayesian statistics work.

In order to explain why Bayesian statistics are suitable for ecologists, I will give an overview of Bayesian Statistics. Next, I will talk about the two common algorithms of MCMC—Metropolis-Hastings and Gibbs sampling.

3.1 Overview of Bayesian Statistics

In Bayesian inference, there are three major components - the likelihood function, prior distribution, and the posterior distribution. First, a likelihood function and prior distribution (for the unknown parameters) are specified. Prior distributions may be informed using domain-area expertise. Second, a posterior distribution is computed using the likelihood function and prior distribution. Finally, the posterior distribution is either calculated, or in the majority of cases, approximated using sampling-based methods such as Markov Chain Monte Carlo [55] or variational methods [56].

The goal of Bayesian statistics is to gain information on the posterior distribution of the unknown (and sometimes unobserved) model parameters $\boldsymbol{\theta}$ given the observed data Z. The posterior distribution can be obtained using sampling or variational methods such that we have approximations of the posterior distribution of $\boldsymbol{\theta}$, which is $p(\boldsymbol{\theta}|Z)$. Then, the

posterior median or mean distribution would be the 'best guess' point estimate of our parameters $\boldsymbol{\theta}$. Furthermore, the posterior predictive distribution $p(\tilde{Z}|Z)$ can also be obtained because $p(\tilde{Z}|Z) = \int p(\tilde{Z}|\boldsymbol{\theta})p(\boldsymbol{\theta}|Z)d\boldsymbol{\theta}$. Here, $p(\tilde{Z}|Z)$ represents the posterior distribution of predictions at unobserved locations of covariate values.

Bayesian statistics does not require large dataset [55], but requires prior knowledge of the unknown parameters through a prior distribution $p(\boldsymbol{\theta})$. In other words, the prior distribution contains one's beliefs about the unknown parameters, and these prior can be subjective or objective. Once we specify a prior distribution $p(\boldsymbol{\theta})$ and data distribution (also called likelihood function) $p(Z|\boldsymbol{\theta})$, the posterior distribution can be obtained by using the Bayes' rule:

$$p(\boldsymbol{\theta}|Z) = \frac{p(Z, \boldsymbol{\theta})}{p(Z)} = \frac{p(Z|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(Z)}.$$

The Bayesian modeling framework provides a flexible approach to fit complex statistical models, while also integrating expert knowledge of the unknown model parameters. The Bayesian framework infers the entire distribution of an unknown model parameter, given the observed data; hence, it also provides uncertainty quantification. The Bayesian framework is also amenable to hierarchical models that incorporate scientific knowledge, such as physical or biological concepts; so, the model is closer to the real world.

The parameters $\boldsymbol{\theta}$ are unknown and considered random variables in a Bayesian setting. The posterior distribution of parameters is directly related to a probability statement. To illustrate, the 95% credible interval means that the parameter has a 95% probability of being within the interval, given the observed data.

Due to the absence and presence of animal species, ecologists will not have large data at certain times and in certain places. Moreover, dynamic spatio-temporal models may include high-dimensional spatio-temporally correlated parameters, and the typical Frequentist framework via maximum likelihood may not be feasible. Therefore, Bayesian inference is suitable for ecology analysis [57]. Furthermore, the posterior distribution (also known as the target distribution) can be approximated, regardless of its shape or distributional family.

As long as we have prior knowledge and a data model (likelihood), the posterior distribution can be generated [58].

Prior distributions are only applied in Bayesian inference, not Frequentist inference. Prior distributions can be informative and non-informative. Informative priors mean that some previous experiments provide us with specific knowledge about the parameters. In contrast, non-informative priors indicate knowing nothing about the knowledge, so a uniform distribution is usually applied. Research into non-informative prior distributions is an ongoing and robust research area [59].

3.2 Markov chain Monte Carlo (MCMC) algorithms

Typically, the posterior distribution of the unknown model parameters $p(\boldsymbol{\theta}|Z)$ may not be available in closed form due to the intractable normalizing constant $p(Z)$. Suppose we have observed data Z and parameters $\boldsymbol{\theta}$ of interest, then the posterior distribution is written as,

$$p(\boldsymbol{\theta}|Z) = \frac{p(Z|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(Z)} = \frac{p(Z|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\boldsymbol{\theta}} p(Z|\boldsymbol{\theta})p(\boldsymbol{\theta})\, d\boldsymbol{\theta}},$$

where $p(Z|\boldsymbol{\theta})$ is the likelihood function of $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$ is the prior distribution, and $p(Z)$ is the marginal likelihood function (also called integrated likelihood). When $\boldsymbol{\theta}$ are high-dimensional, the integrated likelihood $\int_{\boldsymbol{\theta}} p(Z|\boldsymbol{\theta})p(\boldsymbol{\theta})\, d\boldsymbol{\theta}$ is generally intractable and not available in analytical form. To bypass these issues, sampling or simulation-based approaches such as MCMC algorithms have been widely used. Here, one can approximate the posterior distribution via sampling; thereby bypassing issues with the intractable integrated likelihood.

In MCMC algorithms, samples from the posterior distribution are drawn, but note that we only need to know the posterior distribution up to the normalizing constant $p(\mathbf{Z})$:

$$p(\boldsymbol{\theta}|Z) \propto p(Z|\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

Since $p(\mathbf{Z})$ is merely a normalizing constant, its general purpose is to ensure that the posterior distribution fully integrates to 1, and it does not affect the overall spread of the posterior distribution. The MCMC algorithm draws samples from the posterior distribution $p(\boldsymbol{\theta}|Z)$. In general, the integral likelihood will be canceled out in the MCMC algorithms, it is actually drawing samples from the likelihood times the prior distribution $p(Z|\boldsymbol{\theta})p(\boldsymbol{\theta})$.

3.2.1 Metropolis-Hasting Algorithm

The Metropolis-Hasting Algorithm [60] is one of the commonly used MCMC algorithms. For Metropolis Hasting, a likelihood function and prior distribution for the parameters of interest are needed. In addition, a proposal density (also called transition kernel) must be specified by the user. One example of a transition kernel is the random-walk kernel where the proposal density $q(x^*|x)$ is a conditional probability of moving from x to x^* via a random walk. If the proposal density is symmetric, then $q(x^*|x) = q(x|x^*)$. To illustrate the algorithm, I give an example of $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ as following:

(i.) Set a start value for $\theta_1^{(0)} = \vartheta_1$, $\theta_2^{(0)} = \vartheta_2$, $\theta_3^{(0)} = \vartheta_3$.

(ii.) Set a proposal density for each parameter.

For t=1...T iteration,

(iii.) first, get a candidate θ_1^* via its proposal density, which can be normal distribution. That is, $\theta_1^* = \theta_1^{(t-1)} + N(0, \sigma_{\theta_1}^2)$, where $\sigma_{\theta_1}^2$ is an selected arbitrary value. Then, calculate

$$\alpha\big(\theta_1^{(t-1)}, \theta_1^*\big) = min\left(1, \frac{p(\theta_1^*|Z)q(\theta_1^{(t-1)}|\theta_1^*)}{p(\theta_1^{(t-1)}|Z)q(\theta_1^*|\theta_1^{(t-1)})}\right) = min\left(1, \frac{p(\theta_1^*|Z)}{p(\theta_1^{(t-1)}|Z)}\right)$$

$$= min\left(1, \frac{p(Z|\theta_1^*)p(\theta_1^*)/p(Z)}{p(Z|\theta_1^{(t-1)})p(\theta_1^{(t-1)})/p(Z)}\right) = min\left(1, \frac{p(Z|\theta_1^*)p(\theta_1^*)}{p(Z|\theta_1^{(t-1)})p(\theta_1^{(t-1)})}\right).$$

Here, if the likelihood function or prior of $\theta_1^{(t-1)}$ and θ_1^* includes other parameters θ_2 or θ_3, the values of $\theta_2^{(t-1)}$ or $\theta_3^{(t-1)}$ should be plug into the function. Accept the candidate $\theta_1^{(t)} = \theta_1^*$ with the probability $\alpha\big(\theta_1^{(t-1)}, \theta_1^*\big)$; otherwise, reject the candidate with the probability

$1 - \alpha\big(\theta_1^{(t-1)}, \theta_1^*\big)$, which means stay in the same value $\theta_1^{(t)} = \theta_1^{(t-1)}$.

Furthermore, update θ_2 and θ_3 one-by-one with the same steps as θ_1 above. Note that if the likelihood function or prior of $\theta_2^{(t-1)}$ and θ_2^* contains parameters θ_1 or θ_3, the values of updated $\theta_1^{(t)}$ or not upgraded $\theta_3^{(t-1)}$ will be plugged into the function. And so on, apply updated $\theta_1^{(t)}$ or $\theta_2^{(t)}$ to the likelihood function or prior of $\theta_3^{(t-1)}$ and θ_3^*.

The algorithm should run until all Markov chains have arrived to the stationary distribution, which, in our case, is the posterior distribution $p(\boldsymbol{\theta}|Z)$. Convergence diagnostics such as the Gelman-Rubin diagnostic [61], batch means standard error [62], or visual inspection of the individual trace plots may help determine convergence to the stationary distribution. Once we arrive at the stationary distribution (i.e. run as many iterations until convergence) an estimate (mean or median) of the parameters from the posterior distribution can be obtained.

3.2.2 Gibbs sampling

The Gibbs sampling algorithm is one of the most well-known methods for implementing Bayesian statistics. Gelman (1992)[61] found that Gibbs sampling is a specialization of the Metropolis-Hasting concept. For Gibbs sampling, we must have a conditional distribution. I will illustrate the algorithm with the same example as above ($\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$) as follows:

(i.) Set a start value for $\theta_1^{(0)} = \vartheta_1$, $\theta_2^{(0)} = \vartheta_2$, $\theta_3^{(0)} = \vartheta_3$.

For t=1...T iteration,

(ii.) Sample $\theta_1^{(t)}$ from $p(\theta_1|\theta_2^{(t-1)}, \theta_3^{(t-1)}, Z)$. Furthermore, draw $\theta_2^{(t)}$ from $p(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, Z)$. Last, take sample $\theta_3^{(t)}$ from $p(\theta_3|\theta_1^{(t)}, \theta_2^{(t)}, Z)$.

Similar to Metropolis Hasting, we run the Gibbs Sampler until the resulting Markov Chain converges to the stationary distribution. However, unlike the Metropolis-Hastings algorithm, the Gibbs sampling has an acceptance probability of 1 because we sample from the full conditional distribution $p(\theta_k|\theta_{-k})$.

Chapter 4: Basis Functions

In this thesis, I address a key limitation for dynamic spatio-temporal models (DSTMs) - high computational costs for model-fitting. As previously discussed, high-dimensional spatio-temporal processes u_t can be difficult for model-fitting due to large matrix operations, slow mixing, and over-parameterization. Over-parameterization means that there are usually more parameters than observations. Model-fitting can be computationally expensive because many parameters should be estimated. Furthermore, mixing in the Markov-chain Monte Carlo (MCMC) algorithm can be slow due to the very large autocorrelation in the spatio-temporal random effects u_t [63]. Slow mixing in MCMC represents that the Monte Carlo would be slow to converge to the stationary distribution. To overcome the high-dimensional problem, spatial or spatio-temporal basis expansions can provide a computationally efficient way to summarize the high-dimensional space-time data.

In our application, I perform a linear basis expansion of spatial/spatio-temporal basis functions $\phi_i(x)$, which are weighted by a basis coefficient (weights) w. A linear basis function model: $Y(x) = \sum_{i=1}^{n} w_i \phi_i(x)$, where $Y(x)$ is any function that depends on x, w_i is a value of weight and $\phi_i(x)$ is a basis function. That is, linear combinations of (a set of) basis functions can form any kind of function. Note that this model is linear in the basis coefficients (weights) w_i, rather than the non-linear basis functions $\phi_i(x)$. In the following subsections, I provide a general overview of some popular basis functions.

4.1 Polynomial Basis Functions

Polynomial basis functions can be constructed from a single set of covariates x. The covariates are transformed by taking various polynomials of the original covariate x. For example, we can construct the following polynomial basis functions: $\{\phi_0(x) = 1, \phi_1(x) = x, \phi_2(x) =$

$x^2, \phi_3(x) = x^3, ...\}$. Note that we can preserve the original covariate by including $\phi_1(x) = x$ into our model. The other polynomial basis functions can be specified by the user.

4.2 Wavelets

Wavelets can be applied to signal or image processing [64], and also be used to solve partial differential equations [65]. Through the dilation and translation of a mother wavelet, we can obtain basis functions from the mother wavelet [66]. Consider a popular set of discrete wavelet transformations - the Haar wavelet. The mother wavelet of Haar transform is written as

$$\Phi(x) = \begin{cases} 1 & \text{if } 0 < x < 1/2 \\ -1 & \text{if } 1/2 < x < 1 \\ 0 & \text{otherwise} \end{cases},$$

and its basis functions are $\{\phi_0(x) = 1,\ \phi_1(x) = \Phi(x),\ \phi_2(x) = \sqrt{2}\Phi(2x),\ \phi_3(x) = \sqrt{2}\Phi(2x-1),\ \ \phi_4(x) = 2\Phi(4x),\ \phi_5(x) = 2\Phi(4x-1),\ \phi_6(x) = 2\Phi(4x-2),\ \phi_7(x) = 2\Phi(4x-3), ...\}$. The Haar wavelet family generates rectangular 'boxes' as the basis functions. There are other kinds of wavelet, such as Meyer wavelet and Daubechies wavelet [67]. Both Meyer wavelets and Daubechies wavelets are orthogonal [68]. Meyer wavelets are symmetric, while Daubechies wavelets can be asymmetric [69].

4.3 Radial Basis Functions

Similar to wavelets, radial basis functions are another example of localized spatial basis functions. A radial basis function approximation can be written as $f(\mathbf{x}) = \sum_k w_k \Phi(||\mathbf{x} - \mathbf{x}_k||)$, where Φ is radial basis function $R^+ \to R$, $\mathbf{x}_k \in R^n$ is the grid point, and $||.||$ is the Euclidean distance [70]. Assume that the tuning a shape parameter c; there are some kinds

of radial basis function, including Gaussian $\Phi(r) = e^{-(cr)^2}$, multiquadric $\Phi(r) = \sqrt{1+(cr)^2}$, inverse quadratic $\Phi(r) = \frac{1}{1+(cr)^2}$, thin plate spline $\Phi(r) = r^2 ln(r)$, and so on.

4.4 Fourier Basis Functions

Fourier 1807 [71] proposed that a function can be expressed by a linear combination of an infinite number of sinusoids. Fourier basis functions are global basis functions, so each basis function covers the entire range of the domain [39].

All Fourier basis functions are sine and cosine,

$$\phi_j(x) = \begin{cases} 1 & \text{if j=0} \\ cos(\pi x j) & \text{if j is odd} \\ sin(\pi x j) & \text{if j is even} \end{cases} .$$

The Fourier basis is the most suitable basis for periodic functions, such as various periodic and seasonal patterns [72]. But it is especially hard for fitting non-stationary functions, such as functions with strong local features [73]. Furthermore, we cannot implement an infinite number of the term of basis functions in practice, so we should balance the accuracy and the time-consuming [73].

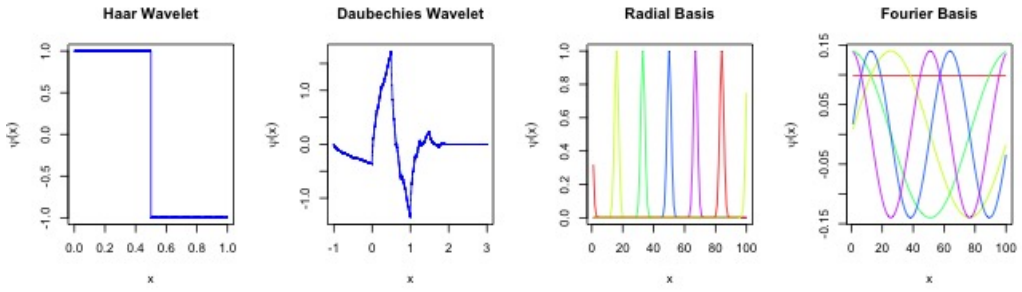

Figure 4.1: Different types of Basis Functions.

4.5 Eigenvector basis functions

There is an eigenvector basis function, which can solve the spatial autocorrelation problem. For the spatial dependence data W, we can construct a Matérn covariance matrix $\boldsymbol{\Sigma}(\theta, d)$. The covariance matrix $\boldsymbol{\Sigma}(\theta, d)$ is a positive definite covariance matrix and d is the distance. The way to do this is to transform the spatial dependence variable into independent variables via adding eigenvector basis functions in the regression, so we can write as $\boldsymbol{\Sigma}(\theta, d) = \boldsymbol{UAU'}$, where $\boldsymbol{U}$ are eigenvectors and $\boldsymbol{A}$ is a diagonal matrix with eigenvalues element. It is appropriate to select only the first m eigenvectors, because the first eigenvector with correspondingly large eigenvalues can explain broad variation [74]. Hence, it can be written as $W \approx \boldsymbol{U_m A_m}^{1/2}\gamma$, where $\gamma \sim N(0, I)$. $\boldsymbol{U_m A_m}^{1/2}$ is the eigenvector basis functions [31]. Eigenvector basis functions (or eigenvector spatial filters) are orthogonal [75] and can be shown in maps to see different patterns. The transformation can help to reduce spatial autocorrelation and speed up the models.

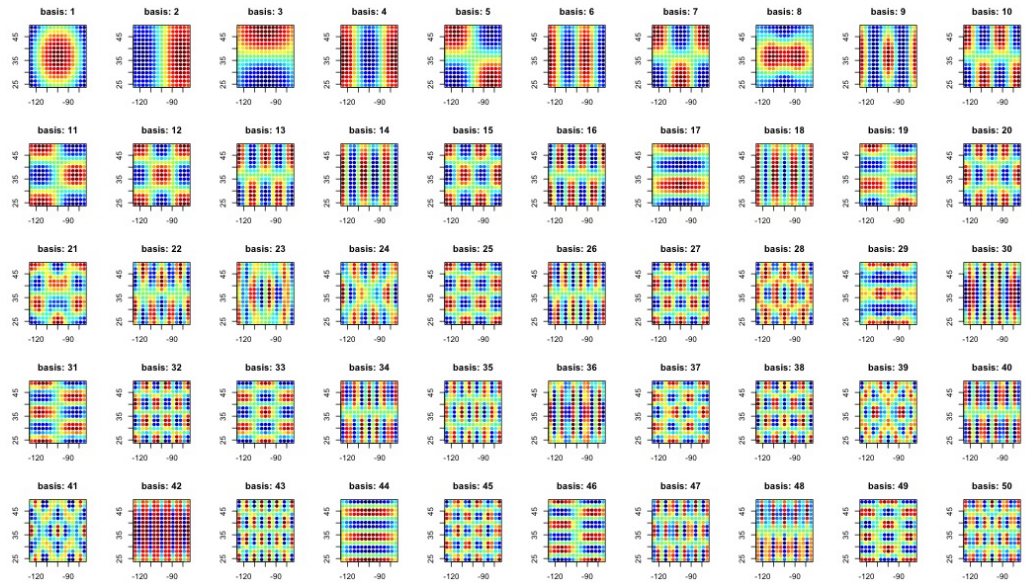

Figure 4.2: The eigenvector basis functions are extracted by the covariance matrix of the real-world ECD data.

Chapter 5: Dynamic Spatio-temporal Models for Animal Movement

In this thesis, I will be extending the model from Hooten and Wikle (2008) [28] to the high-dimensional spatio-temporal setting. The main objective in [28] was to infer the spread of invasive Eurasian collard doves (ECD). This study employed a hierarchical space-time Bayesian model that incorporates the components of a reaction-diffusion PDE. In my thesis, I will expand upon this model by summarizing the high-dimensional spatio-temporal random effects through a basis expansion. Then, the basis approximations will be incorporated into reaction-diffusion PDE of the hierarchical space-time Bayesian model.

In the following section, I will first introduce the model from [28]. Then, I will propose our computationally-efficient basis-expansion approach.

5.1 Hierarchical Dynamic Spatio-Temporal Bayesian model:

The hierarchical dynamic spatio-temporal model from [28] consists of a: (1) data model; (2) process model for the latent space-time processes; (3) process model for the diffusion rate; and (4) prior distributions (models) for the unknown model parameters. The hierarchy represents a top-down structure of conditional distributions where values from the lower stages feed into the models in the upper stages. The hierarchical dynamic spatio-temporal model is outlined as so:

1. **Data Model:**

   - $Z_{s_i,t}|\lambda_{s_i,t} \sim Poisson(\lambda_{s_i,t})$, where $Z_{s_i,t}$ is the count of ECD within different space (i=1,...,N) and time (t=1,...,T), and $\lambda_{s_i,t}$ is the intensity process.

- $log(\boldsymbol{\lambda}_t) = \boldsymbol{K}_t\boldsymbol{u}_t + \boldsymbol{\epsilon}_t$, where $\boldsymbol{K}_t$ is an incidence matrix based on grid connectivity and $\boldsymbol{u}_t = (u_{s_1,t}, ..., u_{s_N,t})'$ is a latent spatio-temporal process at all location at time t, and $\boldsymbol{\epsilon}_t$ is a measurement error.

2. **Process Model for the latent spatio-temporal process:**

- A reaction-diffusion PDE models the latent spatio-temporal process $u_{s_i,t}$.

$$\frac{\partial u_{s_i,t}}{\partial t} = \frac{\partial}{\partial x}\Big(\delta(s_i)\frac{\partial u_{s_i,t}}{\partial x}\Big) + \frac{\partial}{\partial y}\Big(\delta(s_i)\frac{\partial u_{s_i,t}}{\partial y}\Big) + \gamma_0 u_{s_i,t}\Big(1 - \frac{u_{s_i,t}}{\gamma_1}\Big), \tag{5.1}$$

where $s_i \in R^2$, x and y are the longtitude and latitude of location s_i, γ_0 is the ECD population growth rate, γ_1 is the carrying capacity of ECD, and $\delta(s_i)$ is the spatially varying diffusion rate (see process model #3).

- The PDE can be solved by discretization,

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

With assuming $\Delta_t = 1$, the form of $\boldsymbol{u}_t$ can be written as:

$$\boldsymbol{u}_t = \boldsymbol{H}(\boldsymbol{\delta})\boldsymbol{u}_{t-1} + \alpha_0\boldsymbol{u}_{t-1} - \alpha_1\boldsymbol{u}_{t-1}^2 + \boldsymbol{\eta}_t, \tag{5.2}$$

where $\alpha_0 = \gamma_0$ and $\alpha_1 = \gamma_0/\gamma_1$ are the ECD growth parameters, $\boldsymbol{H}$ is a propagator matrix, and $\boldsymbol{\eta}_t$ is the i.i.d error.

- Incorporating (2.10) equation into an MCMC algorithm can be difficult due to the long-range time-dependence (across time points). Instead, a second-order Markov condition can be used to improve computational efficiency.

$$\boldsymbol{u}_t = \boldsymbol{H}(\boldsymbol{\delta})\boldsymbol{u}_{t-1} + \alpha_0\boldsymbol{u}_{t-1} - \alpha_1 diag(\boldsymbol{u}_{t-1})\boldsymbol{u}_{t-2} + \boldsymbol{\eta}_t, \tag{5.3}$$

where the full conditional distribution of $\boldsymbol{u}_t$ can be derived.

3. **Process Model for the diffusion rate:**

- The diffusion rate, which depends on space, can be shown as below.

$$\boldsymbol{\delta} = \boldsymbol{X}\boldsymbol{\beta} + \xi, \tag{5.4}$$

where $\xi|\sigma_\delta^2, \theta \sim N(\boldsymbol{0}, \sigma_\delta^2\boldsymbol{R}(\theta))$, and $\boldsymbol{X}$ includes the intercept and human population density with the associated $\boldsymbol{\beta}$.

- The spatial correlation matrix $\boldsymbol{R}$ depends on the distance ($||d||$: the Euclidian distance between two points) and the spatial range parameter (θ).

$$R(\theta, d) = exp(-\theta||d||), \tag{5.5}$$

which is a Matérn covariance function with $\sigma^2 = 1$, $\nu = 1/2$ and $\rho = 1/\theta$.

4. **Parameter model:**

- Priors for bivariate parameters $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$.

$$log(\boldsymbol{\alpha}) \sim N\Big(\begin{pmatrix}0.001\\0.001\end{pmatrix}, \begin{pmatrix}10 & 2\\2 & 10\end{pmatrix}\Big) \; and \; \boldsymbol{\beta} \sim N\Big(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}10 & 0\\0 & 10\end{pmatrix}\Big).$$

- Priors for univariate parameters σ_ϵ^2, σ_η^2, σ_δ^2, θ.

$$\sigma_\epsilon^2 \sim IG(2.3, 0.4),\ \sigma_\eta^2 \sim IG(2,2),\ \sigma_\delta^2 \sim IG(2,1000)\ and\ \theta \sim Unif(0,10).$$

- At initial time, prior for spatio-temporal process $\boldsymbol{u}_0$.

$$\boldsymbol{u}_0 \sim N(\boldsymbol{0}, 10\boldsymbol{I}),\ and\ \boldsymbol{u}_{-1} = -10 \times \boldsymbol{I}.$$

5. **Other parameters:**

- In data model, $\boldsymbol{\epsilon}_t | \sigma_\epsilon^2 \sim N(\boldsymbol{0}, \sigma_\epsilon^2 \boldsymbol{I}),\ t = 1, ..., T.$
- In process $\boldsymbol{u}_t$ model, $\boldsymbol{\eta}_t | \sigma_\eta^2 \sim N(\boldsymbol{0}, \sigma_\eta^2 \boldsymbol{I})$.
- In process $\boldsymbol{\delta}$ model, $\xi | \sigma_\delta^2, \theta \sim N(\boldsymbol{0}, \sigma_\delta^2 \boldsymbol{R}(\theta))$.

One limitation of this hierarchical DSTM [28] is that it doesn't provide exact inference. This approach artificially imposes a second-order Markov structure in the spatio-temporal random effects, such that temporal dependencies exist for two adjacent timepoints only. Hooten and Wikle (2008) [28] impose the second-order Markov structure because the high-dimensional random effects are hard to update in Metropolis-Hastings algorithm, a Markov chain Monte Carlo (MCMC) method. Another limitation focuses on the negative diffusion rate around New York City. The authors proposed that the human population density in the area around New York is an outlier, so this flexible linear model makes the associated diffusion rate negative. Next, the posterior means of the Poisson intensity processes of ECD in some places will decrease at a certain starting time because there may be some Allee effects [76]. The Allee effect occurs when the population growth rate reduces with low population densities [77]. Finally, this model is capable of forecasting into the future, but there exist so much uncertainty in the forecasts.

In my thesis, I will focus on the computational aspects of the dynamic hierarchical spatio-temporal model. Given the large number of locations and time points, the original

model may not scale to even moderate number of sites and times. The other methodological issues remain an open problem in the literature, and these may be an exciting avenue for future research.

5.2 Our Proposed Method

To make the computation efficient, I apply a basis expansion method to reduce estimated parameters and deal with spatial autocorrelation. Below, I provide a concise overview of how the proposed approach differs from [28]. In Section 6, I provide additional details on the proposed hierarchical model such as the hierarchical structure, data and process models, prior distributions, and formulation of the diffusion rate.

5.2.1 Dimension-Reduction via Basis Representation

The spatial random effect is the product of the basis function and the re-parameterized spatial random effect [51]. In the dynamic spatio-temporal model, the spatio-temporal random effects are modeled as:

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

It may be computationally expensive, when the dimensions of $\boldsymbol{u}_t$ are moderately large. I propose approximating $\boldsymbol{u}_t$ using a basis expansion $\tilde{\boldsymbol{u}}_t = \boldsymbol{\Phi}\boldsymbol{\gamma}_t$, where the dimensions of $\boldsymbol{\gamma}_t$ are much smaller than those of $\boldsymbol{u}_t$.

I reconstruct the process model (reaction-diffusion PDE) using temporally correlated basis coefficients $\boldsymbol{\gamma}_t$,

$$\tilde{\boldsymbol{u}}_t = \boldsymbol{\Phi}\boldsymbol{\gamma}_t = \boldsymbol{H}(\boldsymbol{\delta})\boldsymbol{\Phi}\boldsymbol{\gamma}_{t-1} + \alpha_0 \boldsymbol{\Phi}\boldsymbol{\gamma}_{t-1} - \alpha_1 (\boldsymbol{\Phi}\boldsymbol{\gamma}_{t-1})^2 + \boldsymbol{\eta}_t.$$

Here, the basis coefficients $\boldsymbol{\gamma}_t$ can be modeled using a vector auto-regressive (VAR) process or another temporally-correlated multivariate process. Due to $dim(\boldsymbol{\gamma}_t) << dim(\boldsymbol{u}_t)$, I only

need to estimate fewer parameters than Hooten's paper. $\mathbf{\Phi}$ is the matrix of eigenvector basis functions. Each column ϕ_i represents the multiplication of eigenvector e_i with the square root of its corresponding eigenvalue λ_i.

5.2.2 Spatial and Spatio-temporal Correlation

As previously mentioned, adding basis functions is a way to deal with the spatial autocorrelation problem. Here, $\boldsymbol{\delta}$ exists spatial dependent,

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

When running an MCMC algorithm, the covariance matrix is expensive to compute and parameters are hard to converge. I rebuild the process model with basis expansion,

$$\boldsymbol{\delta} = \boldsymbol{X\beta} + \mathbf{\Phi}\boldsymbol{\gamma_\delta}, \text{ where } \boldsymbol{\gamma_\delta} \sim N(\mathbf{0}, \tilde{\sigma}_\delta^2 I).$$

Elements of $\boldsymbol{\gamma_\delta}$ are the value of weights and their variances are i.i.d $\tilde{\sigma}_\delta^2$. Here, $\mathbf{\Phi}$ is the same eigenvector basis function matrix as constructing $\tilde{\boldsymbol{u}}_t = \mathbf{\Phi}\boldsymbol{\gamma}_t$. I create $\mathbf{\Phi}$ by selecting the first m eigenvectors of the covariance matrix from a distance matrix.

Chapter 6: Simulation Study

6.1 Design

I set up a regularly squared spatial domain with $N = 225$ $(= 15^2)$ grid points for 1 time point; that is, there are $s_1, ..., s_{225}$ for every timing. The spatial domain is restricted to the unit square $\mathcal{D} = [0, 1]^2$, so each s_i is two-dimensional coordinate in $\mathcal{D}$. In our simulated data, I set $T = 10$ time points to demonstrate the temporal dynamics. The diffusion matrix $\boldsymbol{H}$ is based on the formulation from [78]. For the featured model parameters, I chose the median estimated parameters of the posterior distribution from Wikle and Hooten (2008), and then used the median estimates to generate the simulated dataset at $N = 225$ locations and $T = 10$ time points.

I generate the initial spatial random effect $\boldsymbol{u}_0 \sim N(\boldsymbol{0}, \Sigma_0)$, where the covariance matrix $\Sigma_0 = 2 * exp(-||d||/0.2)$, with the distance matrix $||d||$. Then, I have time-varying spatial random effects $\boldsymbol{u}_t \sim N(\boldsymbol{H}(\boldsymbol{\delta})\boldsymbol{u}_{t-1} + \alpha_0\boldsymbol{u}_{t-1} - \alpha_1\boldsymbol{u}_{t-1}^2,\ \sigma_\eta^2\boldsymbol{I})$, where $t = 1, ..., 10$, with $\alpha_0 = 0.005$, $\alpha_1 = 0.001$, and $\sigma_\eta^2 = 1.1$. Here, $\boldsymbol{\delta}$ comes from $N\big(\boldsymbol{X}\boldsymbol{\beta},\ 0.0005 * exp(-||d||/20)\big)$, with $\boldsymbol{\beta} = (\beta_0, \beta_1)' = (0.1, -0.0002)'$. The first column of $\boldsymbol{X}$ is intercept, and the second column is the longitude coordinate of space s_i. Because there are N=225 space grid points, each time-varying spatial random effect $\boldsymbol{u}_t$ is composed by $(u_{s_1,t}, u_{s_2,t}, ..., u_{s_{225},t})'$. Finally, generate each count $n_{s_i,t} \sim Poisson(exp(u_{s_i,t}))$ in space s_i and time t.

6.2 Implementation

I generate the posterior samples via the Metropolis-Hastings algorithm (MCMC). To generate the estimates, I employ two different methods: (1) my proposed method; and (2) a separable spatio-temporal model from Section 2.3.1.

6.2.1 Proposed Method

The first approach is our proposed method—spatial random effects $\tilde{\boldsymbol{u}}_t$ with basis expansions as following:

Data Model:

$$Z_{s_i,t}|\lambda_{s_i,t} \sim Pois(\lambda_{s_i,t}), \text{ where } log(\lambda_{s_i,t}) = \tilde{u}_{s_i,t}$$

Process Model:

- **Latent Spatio-temporal process:**

$$\tilde{\boldsymbol{u}}_t = \boldsymbol{\Phi}\boldsymbol{\gamma}_t = \boldsymbol{H}(\boldsymbol{\delta})(\boldsymbol{\Phi}\boldsymbol{\gamma}_{t-1}) + \alpha_0(\boldsymbol{\Phi}\boldsymbol{\gamma}_{t-1}) - \alpha_1(\boldsymbol{\Phi}\boldsymbol{\gamma}_{t-1})^2 + \boldsymbol{\eta}$$

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

$$\boldsymbol{\gamma}_t \sim N(0, \sigma_\gamma^2\boldsymbol{I}), \text{ where } \sigma_\gamma^2 \text{ is fixed}$$

- **Diffusion rate:**

$$\boldsymbol{\delta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\Phi}\boldsymbol{\gamma}_\delta$$

$$\boldsymbol{\gamma}_\delta \sim N(0, \tilde{\sigma}_\delta^2\boldsymbol{I})$$

Parameter Model:

$$\tilde{\boldsymbol{u}}_0 = \boldsymbol{\Phi}\boldsymbol{\gamma}_0 \sim N(0, 10\boldsymbol{I})$$

$$\tilde{\sigma}_\delta^2,\ \sigma_\eta^2 \sim IG(2,2)$$

$$\boldsymbol{\beta} = \begin{pmatrix}\beta_0\\ \beta_1\end{pmatrix} \sim N\Big(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}10 & 0\\ 0 & 10\end{pmatrix}\Big)$$

$$log(\boldsymbol{\alpha}) = \begin{pmatrix}log(\alpha_0)\\ log(\alpha_1)\end{pmatrix} \sim N\Big(\begin{pmatrix}0.001\\0.001\end{pmatrix}, \begin{pmatrix}10 & 2\\ 2 & 10\end{pmatrix}\Big)$$

Here, $\boldsymbol{u}_t = (u_{s_1,t}, u_{s_2,t}, ..., u_{s_{225},t})'$, and $\boldsymbol{\gamma}_t = (\gamma_{1,t}, \gamma_{2,t}, ..., \gamma_{50,t})'$. To construct the eigenvector basis functions $\boldsymbol{\Phi}$, I extract eigenvectors from the covariance matrix Σ_0, and choose the

first 50 eigenvector basis functions for our model. Hence, $\mathbf{\Phi}$ is an $N \times 50$ eigenbasis matrix and each column is an eigenvector times the squared roots of its corresponding eigenvalue.

6.2.2 Separable Spatio-temporal Generalized Linear Mixed Model (Separable SGLMM)

The second approach models the simulated data using the separable spatio-temporal generalized linear mixed model (separable SGLMM) from Section 2.3.1. The hierarchical spatio-temporal modeling framework is as follows:

Data Model:

$$Z_{s_i,t}|\lambda_{s_i,t} \sim Pois(\lambda_{s_i,t}), \text{where } log(\lambda_{s_i,t}) = \tilde{v}_{s_i,t}$$

Process Model:

- **Latent Spatio-temporal process $\tilde{\boldsymbol{v}}_t = (v_{s_1,t}, v_{s_2,t}, ..., v_{s_{225},t})'$:**

$$\tilde{\boldsymbol{v}}_t = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{w}_t$$

$$\boldsymbol{W} = (vec(\boldsymbol{w}_1, \boldsymbol{w}_2, ..., \boldsymbol{w}_T))' \sim N(\mathbf{0}, R_S \otimes R_T),$$

$$\text{where } R_S = \sigma_S^2 \; exp(-\frac{1}{a_S}||h||), \; R_T = \sigma_T^2 \; exp(-\frac{1}{a_T}|\tau|),$$

Parameter Model:

$$\boldsymbol{\beta} \sim N\Big(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}10 & 0\\0 & 10\end{pmatrix}\Big),$$

$$\sigma_S^2, \sigma_T^2 \sim IG(2,2), \; a_S \sim Unif(0, \sqrt{2}), \; a_T \sim Unif(0, 10).$$

Here, $\otimes$ is the kronecker product, so $R_S \otimes R_T$ is a $N \times t$ matrix. $||h||$ is the spatial distance matrix and $||\tau||$ is the time distance matrix.

6.3 Results

I use the Metropolis-Hasting method to fit our proposed model and separable model to the simulated data. After running MCMC, the posterior distributions of the parameters are obtained. Figure 6.2 shows the parametric trace plots (left), trace plots after burn-in (middle) and posterior distribution plots (right) of the MCMC results. For spatial random effects, maps allow us to visualize the mean estimated parameters of the posterior distribution over time. That is, for i=1,...,225 and t=0,...,10, I take the average of each estimated spatial random effect $\gamma_{s_i,t}$ and perform matrix multiplication $\boldsymbol{\Phi}\boldsymbol{\gamma}_t$ to be the estimated $\tilde{\boldsymbol{u}}_t$. Also, taking the mean of parameters $\boldsymbol{\beta}$ and $\boldsymbol{w}_t$ yields the estimated $\tilde{\boldsymbol{v}}_t$. Figure 6.1 is a map showing the generated data $\boldsymbol{u}_t$ and estimated $\tilde{\boldsymbol{u}}_t$. I list the number of iterations to perform MCMC, mean and 95% credible interval of parameter, model-fitting walltimes, as well as mean squared error.

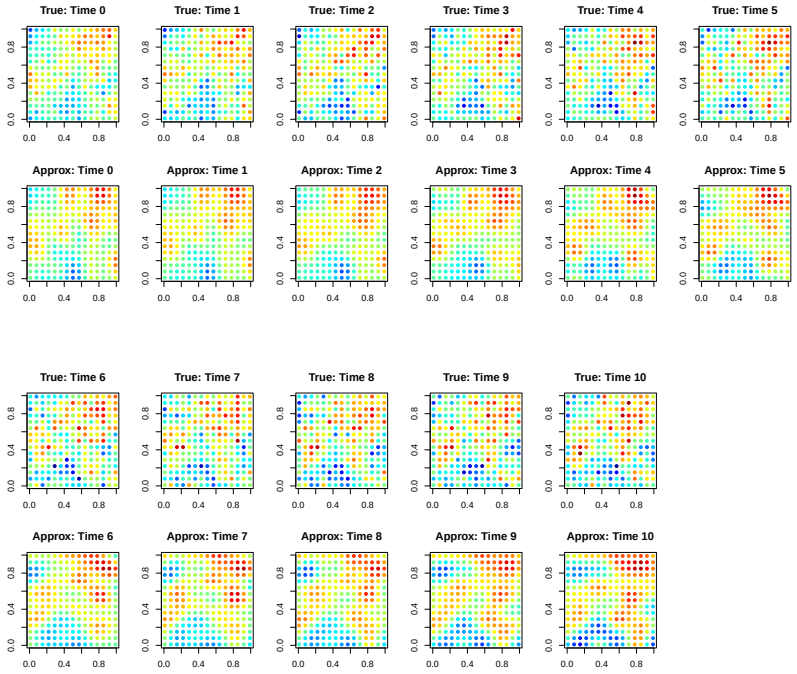

Figure 6.1: A series of maps of generated data $\boldsymbol{u}_t$ versus estimated $\tilde{\boldsymbol{u}}_t = \boldsymbol{\Phi}\boldsymbol{\gamma}_t$ from our model from time 0 (the initial condition) to 10.

For maps, low values are blue, and high are red. The generated data $\boldsymbol{u}_t$ is spatially and temporally correlated, and it is easy to visualize the pattern on a series of maps in Figure 6.1 (first and third rows). Most nearby grid points will have similar colors on a map, which is spatial-dependent. Looking at a grid point of the map for a time point, it will be similar in color to the previous time point and the next time point, which is temporal dependent. In Figure 6.1, the estimated $\tilde{\boldsymbol{u}}_t$ from our model also shows a similar major pattern to the generated data $\boldsymbol{u}_t$. The upper right has higher values (red), and over time, more grid points in the upper right and middle right become red on the estimated and generated maps.

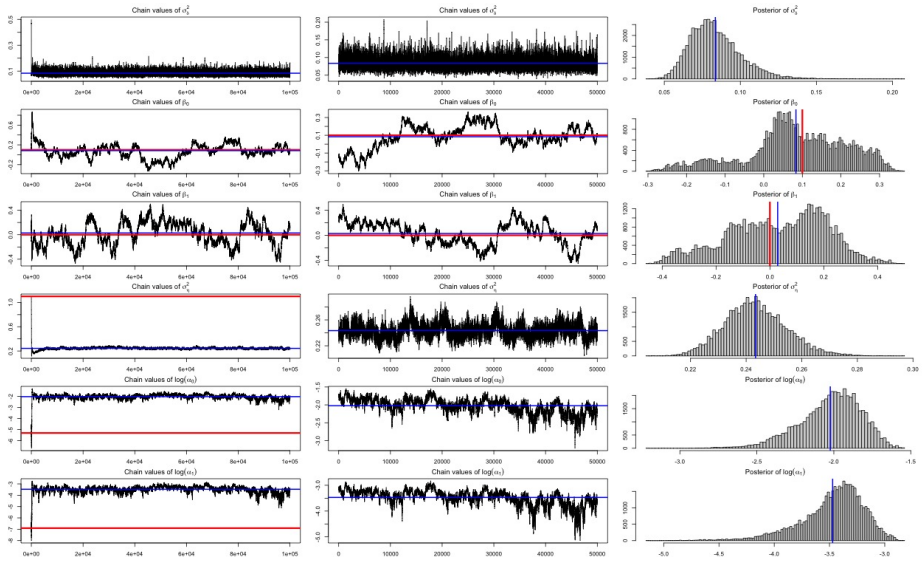

Figure 6.2: Trace plots and posterior distribution of parameters σ_δ^2, β_0, β_1, σ_η^2, $log(\alpha_0)$, $log(\alpha_1)$ of simulated data.

Table 6.1: Parameter estimation of simulated data by our model.

Parameter	Mean (95% CI)
σ_δ^2	0.084 (0.055, 0.118)
β_0	0.083 (-0.182, 0.313)
β_1	0.029 (-0.367,0.308)
σ_η^2	0.243 (0.222,0.266)
$log(\alpha_0)$	-2.023 (-2.423, -1.666)
$log(\alpha_1)$	-3.471 (-4.078, -2.986)

Each parameter estimation from our model is shown in Figure 6.2 and Table 6.1. The plot from the posterior distribution gives us the mean (blue line) of the parameters in Figure 6.2 (right column). The mean of β_0 and β_1 are close to the true value (red line). Yet, both of them are not significant. The mean of $log(\alpha_0)$ and $log(\alpha_1)$ are larger than the generated values (red line). I believe that there is spatial confounding between latitude variables and spatial random effects; therefore, the 95% credible interval does not include the true value. Spatial confounding will cause the variance inflation of the regression coefficients [36]. As we can see the estimated σ_η^2 is smaller than the generated value. It is because the $\boldsymbol{\gamma}_t$ tends to be closer to the mean and each other than observed $\boldsymbol{u}_t$.

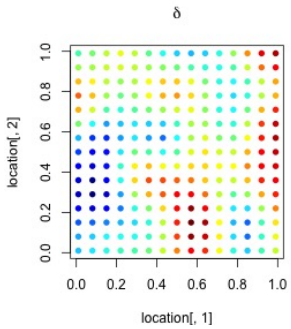

Figure 6.3: The estimated diffusion coefficient $\boldsymbol{\delta}$ of simulated data using our model.

The diffusion coefficient through the map of $\boldsymbol{\delta}$ in Figure 6.3. The spatial dependence can be seen in the diffusion map. Most diffusion coefficients are positive values, with only eight grid points (blue) being negative. Larger positive diffusion coefficients (red) indicate faster diffusion, while smaller positive values (green or cyan) indicate slower diffusion. Negative diffusion coefficients will be explained in the real data experiment.

Each sampler is run for 100,000 iterations with a burn-in of 50,000 iterations. Our model takes about 1959 seconds (32 minutes) to fit the simulated data.

6.4 Comparison

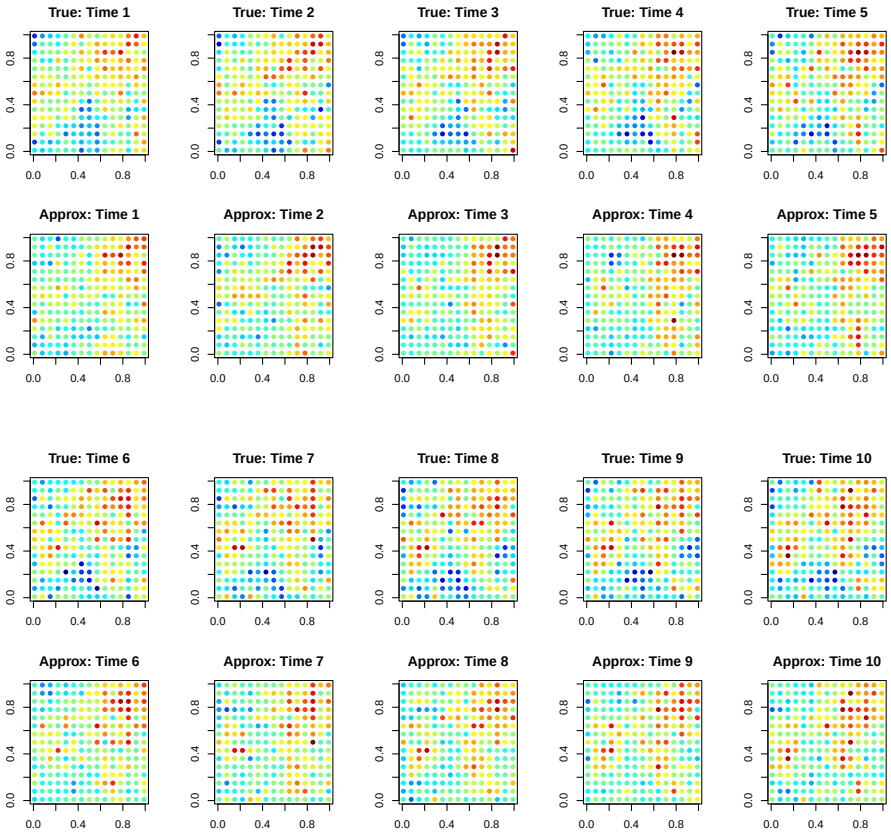

Figure 6.4: A series of maps of generated data $\boldsymbol{u}_t$ versus estimated $\tilde{\boldsymbol{v}}_t = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{w}_t$ from separable model from time 1 to 10.

The separable model is applied to simulated data for comparison with our proposed model. The estimated pattern of $\tilde{\boldsymbol{v}}_t$ closely resembles the generated data $\boldsymbol{u}_t$ on the maps in Figure 6.4. Unlike $\tilde{\boldsymbol{u}}_t$ in Figure 6.1, the $\tilde{\boldsymbol{v}}_t$ is only from $t = 1, ..., 10$. Since the separable model is not a dynamic model, there is no $\tilde{\boldsymbol{v}}_0$. Figure 6.5 and Table 6.2 are the results of MCMC of each parameter estimation from the separable model.

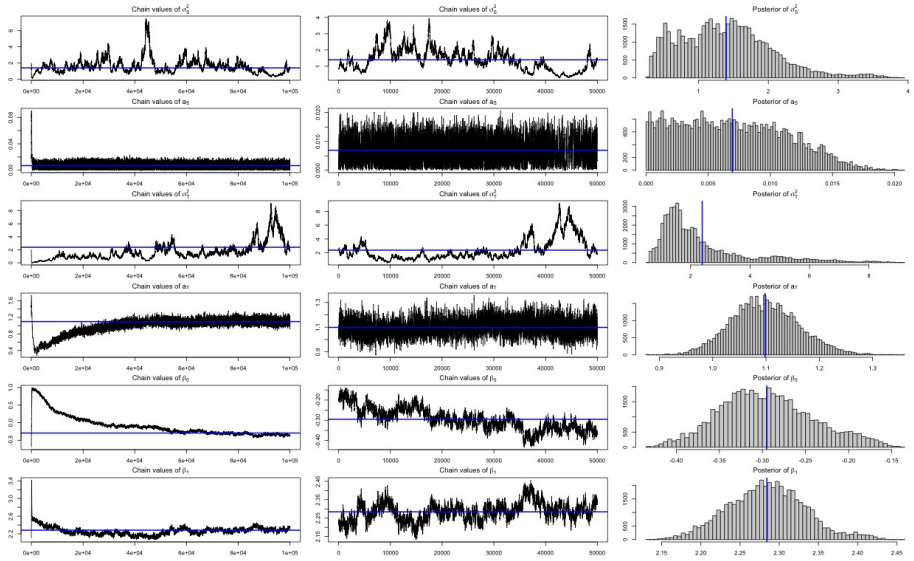

Figure 6.5: Trace plots and posterior distribution of parameters σ_S^2, a_S, σ_T^2, a_T, β_0, β_1 of simulated data.

Table 6.2: Parameter estimation of simulated data by separable model.

Parameter	Mean (95% CI)
σ_S^2	1.394 (0.313, 2.578)
a_S	0.083 (-0.182, 0.313)
σ_T^2	0.007 (0.00002, 0.014)
a_T	1.098 (0.979, 1.227)
β_0	-0.295 (-0.393, -0.181)
β_1	2.284 (2.186, 2.385)

Table 6.3: Results of simulated data by our model.

	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Time 7	Time 8	Time 9	Time 10
SD of $\boldsymbol{\lambda}_t$	0.6900	0.6417	0.5730	0.5452	0.6041	0.5745	0.5923	0.6724	0.9567	1.1057
RMSE of $\tilde{\boldsymbol{u}}_t$	1.249	1.423	1.470	1.545	1.645	1.762	1.866	1.897	1.942	1.995
RMSE of $\boldsymbol{\lambda}_t$	15.573	17.452	25.824	43.907	33.547	30.135	38.297	30.364	29.234	54.556

Table 6.4: Results of simulated data by separable model.

	Time 1	Time 2	Time 3	Time 4	Time 5	Time 6	Time 7	Time 8	Time 9	Time 10
SD of $\boldsymbol{\lambda}_t$	0.8347	0.9092	0.7559	0.8185	0.8571	0.7941	0.9548	0.9242	0.9406	1.0000
RMSE of $\tilde{\boldsymbol{v}}_t$	1.286	1.287	1.210	1.242	1.124	1.167	1.194	1.275	1.225	1.165
RMSE of $\boldsymbol{\lambda}_t$	12.253	17.818	25.369	41.115	26.709	14.109	14.339	13.930	10.196	16.409

For both our proposed model and separable model, I run MCMC for 100,000 iterations and get the posterior distribution after 50,000 burn-ins. Again, I only spend about 1959 seconds (32 minutes) fitting our model to the simulated data. The separable model takes 16708 seconds (4 hrs 40 minutes). Our model is more computationally efficient than the separable model.

I calculate the standard deviation and RMSE for each year to evaluate model fitting in Figures 6.3 and 6.4. The RMSE from separable model is lower (better) than our model. Since there are no basis functions in the separable model, the separable model is not over-smoothed and has high accuracy. But the drawback is that it takes 8.5 times longer than our model to fit the data.

Chapter 7: Real World Example

7.1 Eurasian collared dove Prevalence from 2001-2010

The Eurasian collared dove is an invasive species in the United States with some negative effects on other birds and humans. They drive native North American doves away from food, make obvious noise, and even their excrement will cause indeterminate diseases in human beings [79]. The real-world Eurasian collared dove data and the latitude and longitude coordinates of the routes the doves were spotted on were obtained from the North American Breeding Bird Survey (BBS) website [80]. BBS website collects large numbers of bird data each year during bird breeding months in United State and Canada from 1966 to 2019. A bird expert will count the birds he/she hears or sees on a randomly chosen route. Each route has 50 stops and a bird expert will spend 3 minutes counting doves at each stop. The total count for all 50 stops for each route will be of interest.

Over time, Eurasian collared doves are increasingly found in the US field. The data only records dove counts when the route has doves, so there are no zero or negative values on the route in the dove dataset. I only focus on the United States (except Hawaii and Alaska), so the latitude and longitude coordinates of the doves' routes are in the range [$-124.25°$W,$-75.27°$W] and [24.60°N, 48.85°N], respectively. The observed route k_j^t is time-variant based on the observed dove data. I fit the models on dove data from 2001 to 2010.

7.2 Models

Models that fit real-world Eurasian collared doves are very similar to models that fit simulated data, with few differences. Data model has some changing as following:

Data Model:

$$n_{s_i,t}|\lambda_{s_i,t} \sim ZTP(\lambda_{s_i,t}), \text{ where } log(\boldsymbol{\lambda}_t) = \widetilde{\boldsymbol{K}_t}\tilde{\boldsymbol{u}}_t.$$

We use zero-truncated Poisson (ZTP) distribution because dove data only contains positive integers. As we know the probability mass function (pmf) of Poisson distribution is $p(x;\lambda) = Pr(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}$, where, $\lambda > 0$ and $x = 0, 1, 2, ...$ The pmf of ZTP distribution can be written as:

$$\tilde{p}(x;\lambda) = Pr(X = x|X > 0) = \frac{p(x;\lambda)}{1 - p(0;\lambda)} = \frac{\lambda^x e^{-\lambda}/x!}{1 - e^{-\lambda}} = \frac{\lambda^x}{(e^\lambda - 1)x!}.$$

The mean of ZTP distribution is $E(X) = \frac{\lambda}{1-e^{-\lambda}}$ and the variance is $Var(X) = \frac{\lambda+\lambda^2}{1-e^{-\lambda}} - \frac{\lambda^2}{(1-e^{-\lambda})^2} = E(X)(1 + \lambda - E(X))$.

$\widetilde{\boldsymbol{K}_t}$ is an $M_t \times 225$ inverse-distance weighting matrix for time t. M_t is the total number of the routes over time t. Inverse-distance weighting is a kind of interpolation method. The weights are computed by the distance between grid points s_i and the observed route k_j^t at time t. I set the weighting matrix $(\widetilde{\boldsymbol{K}_t})_{ij} = \frac{w(s_i,k_j^t)}{\sum_{i=1}^{225} w(s_i,k_j^t)}$, where $w(s_i, k_j^t) = 1/d(s_i, k_j^t)^2$. I apply the time-invariant square space domain with the same 15^2 grid points s_i. But the domain is on the US map ([−124.25°W,−75.27°W] and [24.60°N, 48.85°N]).

Another difference from the simulated experiment is the eigenvector basis functions $\boldsymbol{\Phi}$. I extract the first m=50 eigenvectors from the covariance matrix $\Sigma = exp(-3 * \frac{||d||}{||d||_{max} * 0.6})$, where $||d||$ is the distance matrix between the longitude-latitude coordinates of each two routes on the US map.

7.3 Results

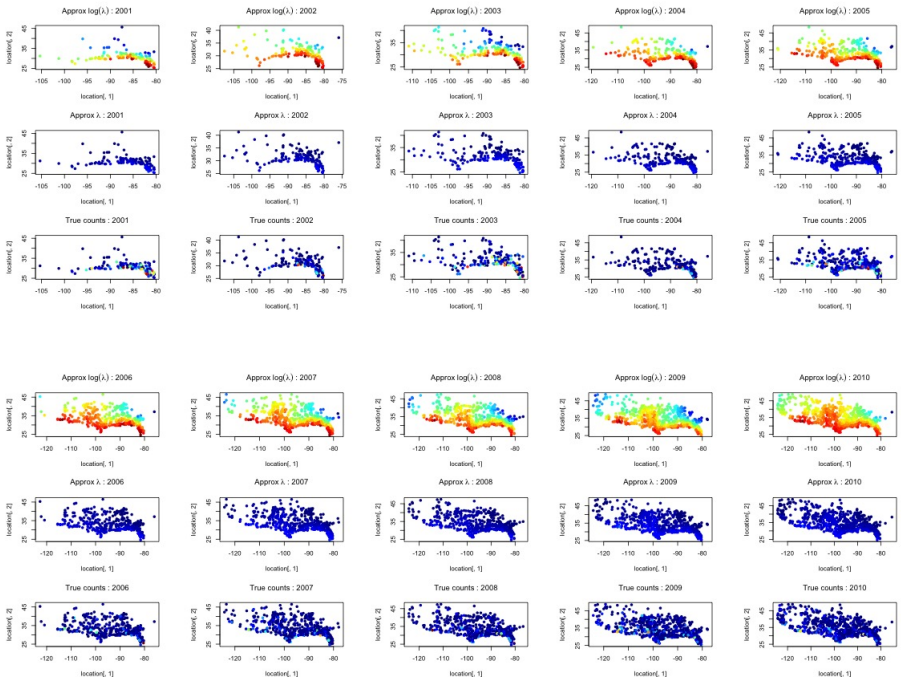

Figure 7.1: A series of maps of $log(\boldsymbol{\lambda}_t)$ and $\boldsymbol{\lambda}_t$, estimated by our model, versus the observed dove counts from 2001 to 2010.

Similar to the simulated example, I fit our proposed model to the real-world data via Metropolis-Hasting technique. Figure 7.2 provides trace plots of our MCMC samples for all 500,000 iterations (left) of the algorithm. Next, I removed the first 100,000 samples for burnin and provide the trace plots (middle) and the density of the posterior distribution (right) of parameter. I also estimate the mean of $\boldsymbol{\gamma}_t$ and obtain a series of maps from 2001 to 2010 in Figure 7.1.

In Figure 7.1, the first and fourth rows are the results of estimated $log(\boldsymbol{\lambda}_t) = \widetilde{\boldsymbol{K}_t}\tilde{\boldsymbol{u}}_t = \widetilde{\boldsymbol{K}_t}\boldsymbol{\Phi}\boldsymbol{\gamma}_t$, which shows spatial random effects with weights on each observed route. The second and fifth rows are the results of estimated $\boldsymbol{\lambda}_t = exp(\widetilde{\boldsymbol{K}_t}\tilde{\boldsymbol{u}}_t) = exp(\widetilde{\boldsymbol{K}_t}\boldsymbol{\Phi}\boldsymbol{\gamma}_t)$, and the third and last rows are the observed bird data. There are more places where doves can be observed in the US, but a series of maps of estimated $\boldsymbol{\lambda}_t$ and the observed dove do not

vary by much. Note that most places are dark blue, which indicates that there are very few counts in those regions. However, the log intensity maps of $log(\boldsymbol{\lambda}_t)$ are able to display the spatial patterns well.

For the maps of $log(\boldsymbol{\lambda}_t)$, spatial random effects become smaller from south to north. It is reasonable because the doves were spread from the southern United States. Over time, the routes with red, orange and yellow became more numerous on the map, as the doves spread out and increased. On the estimated $\boldsymbol{\lambda}_t$ map, there are some lighter blues around the US South (every year) and the US Midwest (2009 and 2010), which means the areas have higher estimated counts. When looking closely at the map, we can also see that the bird data has sporadic different colors in these regions (e.g. red, yellow, green, etc., which means a large number of doves).

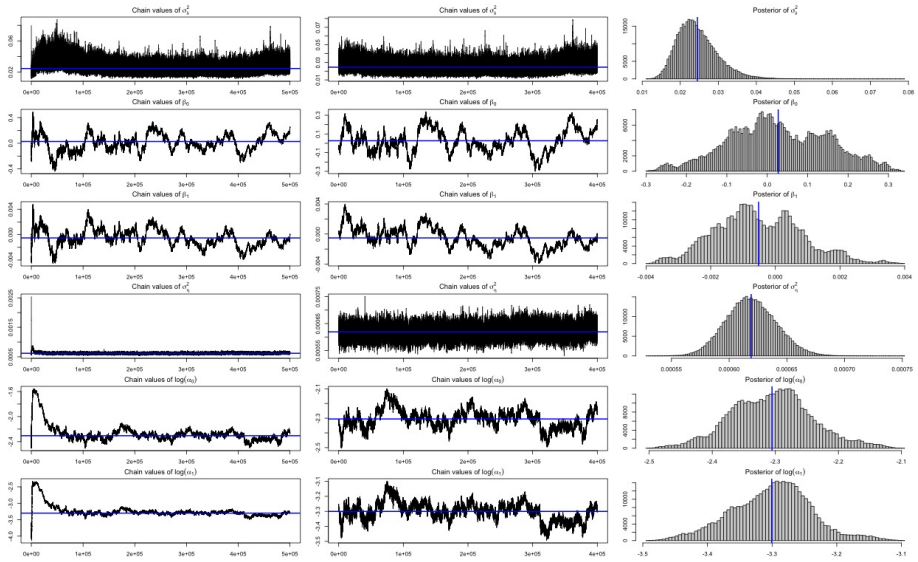

Figure 7.2: Trace plots and posterior distribution of parameters σ_δ^2, β_0, β_1, σ_η^2, $log(\alpha_0)$, $log(\alpha_1)$ of ECD data.

Table 7.1: Parameter estimation of real data by our model.

Parameter	Mean (95% CI)
σ_δ^2	0.025 (0.015, 0.035)
β_0	0.027 (-0.195, 0.271)
β_1	-0.001 (-0.003, 0.002)
σ_η^2	0.0006 (0.0006, 0.0007)
$log(\alpha_0)$	-2.305 (-2.438, -2.177)
$log(\alpha_1)$	-3.301 (-3.440, -3.189)

In Figure 7.2 and Table 7.1, both β_0 and β_1 are insignificant, while $log(\alpha_0)$ and $log(\alpha_1)$ are significant. The estimated growth rate coefficient α_0 can be expressed as $exp(-2.305) = 0.0998$. Since $\alpha_1 = \frac{\alpha_0}{\text{carrying capacity}}$, and we have α_1 is $exp(-3.301) = 0.0368$ and α_0 is $exp(-2.305) = 0.0998$, so we have the carrying capacity is about 2.7.

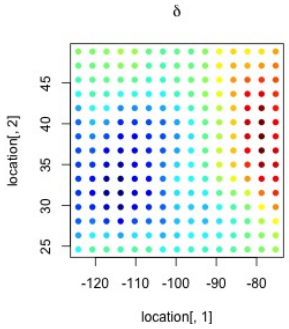

Figure 7.3: The estimated diffusion coefficient $\boldsymbol{\delta}$ of ECD data using our model.

A map of $\boldsymbol{\delta}$ shows the diffusion coefficient for each location in Figure 7.3. Most diffusion coefficients are negative or close to zero. In fact, diffusion coefficient should be positive. Initial conditions lead to the emergence of negative diffusion coefficients. A negative diffusion coefficient (blue) exists in the Midwest. The place didn't have any observed counts in 2001, but over time there were more doves, so that is the reason why there is no positive

diffusion coefficient in this area.

I expect a positive diffusion coefficient in the southern United States, where the birds stayed in 2001, but here the diffusion coefficient is about zero. Because I didn't set the grid points on the sea surface as boundaries, the diffusion couldn't go into the sea, so it offsets the diffusion coefficient. In future work, I will eliminate this problem by taking grid points on the sea surface as boundaries.

7.4 Comparison

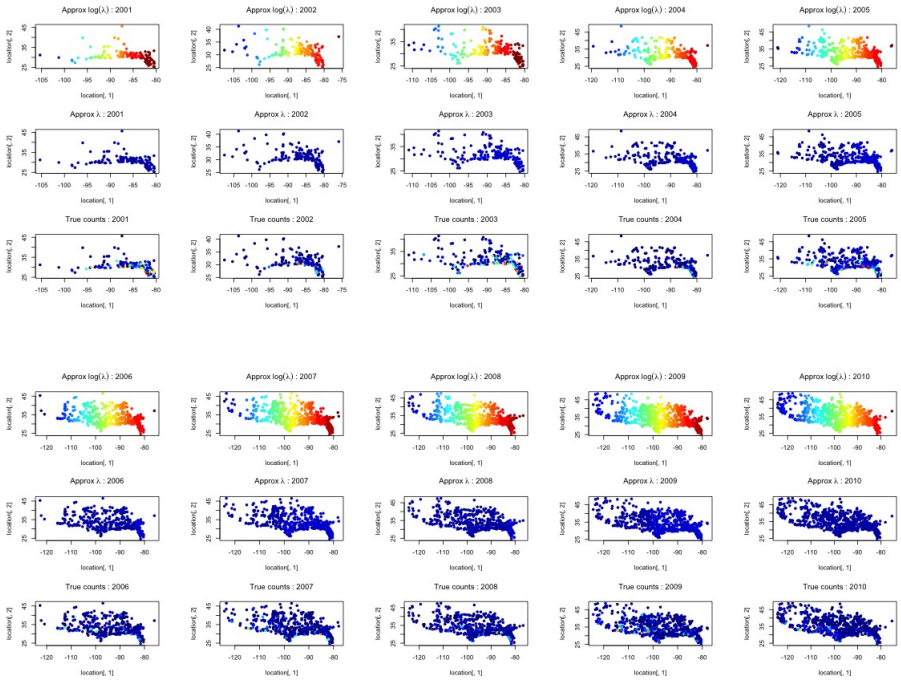

Figure 7.4: A series of maps of $log(\boldsymbol{\lambda}_t)$ and $\boldsymbol{\lambda}_t$, estimated by separable model, versus the observed dove counts from 2001 to 2010.

As I did in the simulation experiment, I also fit the separable model to real data for comparison with our proposed model. As mentioned above, when building our proposed model on real data, only the data model differs from the simulation, and so does the separable model. The changing part of the model is as following:

Data Model:

$$Z_{s_i,t}|\lambda_{s_i,t} \sim ZTP(\lambda_{s_i,t}), \text{ where } log(\boldsymbol{\lambda}_t) = \widetilde{\boldsymbol{K}_t}\tilde{\boldsymbol{v}}_t.$$

In Figure 7.4, the first and fourth rows are the results of $log(\boldsymbol{\lambda}_t) = \widetilde{\boldsymbol{K}_t}\tilde{\boldsymbol{v}}_t$, which shows spatial random effects with weights on each observed route. The second and fifth rows are the results of $\boldsymbol{\lambda}_t = exp(\widetilde{\boldsymbol{K}_t}\tilde{\boldsymbol{v}}_t)$, and the third and last rows are the observed dove. As mentioned above, it's hard to see a pattern of a series of $\boldsymbol{\lambda}_t$ and the observed dove over time (blue dots in most places). The pattern of $\widetilde{\boldsymbol{K}_t}\tilde{\boldsymbol{v}}_t$ is different from $\widetilde{\boldsymbol{K}_t}\tilde{\boldsymbol{u}}_t$, because in the separable model, the $\tilde{\boldsymbol{v}}_t$ only depends on the intercept and latitude covariate.

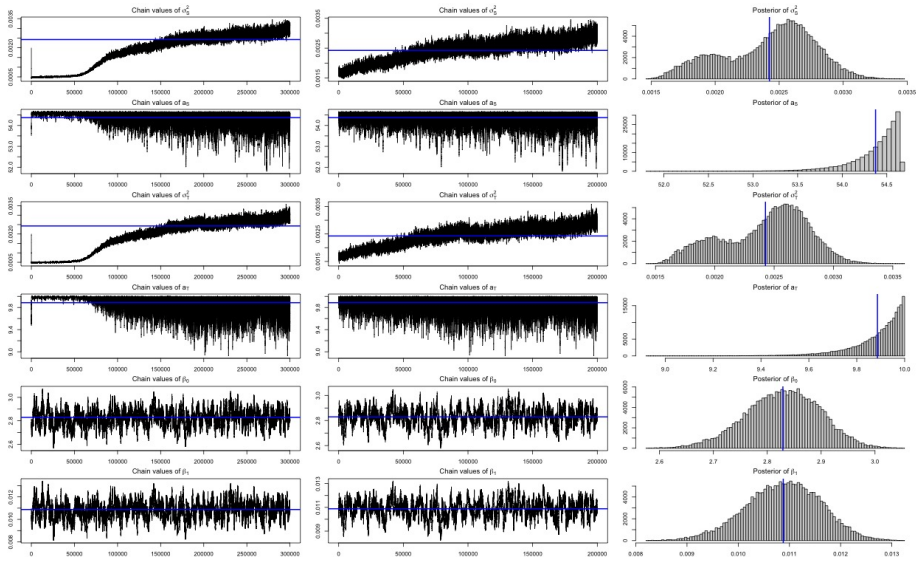

Figure 7.5: Trace plots and posterior distribution of parameters σ_S^2, a_S, σ_T^2, a_T, β_0, β_1 of ECD data.

Table 7.2: Parameter estimation of real data by separable model.

Parameter	Mean (95% CI)
σ_S^2	0.002 (0.002, 0.003)
a_S	54.374 (53.796, 54.657)
σ_T^2	0.002 (0.002, 0.003)
a_T	9.886 (9.656, 10.000)
β_0	2.829 (2.694, 2.963)
β_1	0.011 (0.009, 0.012)

The parameters in the separable model are shown in Figure 7.5 and Table 7.2. The parameters a_S and a_T are significant; the spatial dependence coefficient a_S is 54.374, indicating that the spatial process is highly dependent, and the time dependence coefficient a_T is 9.886, representing that the temporal process is also highly dependent.

Table 7.3: Results of real data by our model.

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
SD of $\boldsymbol{\lambda}_t$	0.3677	0.2949	0.2994	0.2935	0.2683	0.2859	0.2760	0.2858	0.2933	0.3156
RMSE of $log(\boldsymbol{\lambda}_t)$	1.101	1.137	1.143	1.116	1.048	1.100	1.052	1.022	1.086	1.050
RMSE of $\boldsymbol{\lambda}_t$	13.066	14.582	9.533	10.633	8.271	7.350	7.057	8.511	7.383	7.558

Table 7.4: Results of real data by separable model.

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
SD of $\boldsymbol{\lambda}_t$	0.0012	0.0012	0.0011	0.0012	0.0011	0.0012	0.0011	0.0012	0.0012	0.0012
RMSE of $log(\boldsymbol{\lambda}_t)$	1.208	1.259	1.244	1.189	1.164	1.175	1.134	1.113	1.131	1.103
RMSE of $\boldsymbol{\lambda}_t$	14.089	15.446	10.074	11.104	8.676	7.674	7.505	9.090	7.834	8.045

I run my proposed model for 500,000 iterations (burn-in 100,000). It only takes about 7191 seconds (2 hours). On the other hand, the separable model is more expensive to compute, so I only run 300,000 iterations (burn-in 100,000). It takes for about 42614 seconds (11 hours 50 minutes), which is six times slower than our proposed model. The standard deviation and RMSE are calculated in Tables 7.3 and 7.4. The total RMSE of separable model is 9.14, while our model is 8.62. Since the separable model is not dynamic like our model, it may ignore the important underlying dynamic processes.

Chapter 8: Discussion

In this thesis, I propose a computationally efficient method to model dynamic non-Gaussian spatio-temporal datasets with an application to a Eurasian collared dove prevalence study. In this chapter, I summarize the proposed method and provide a brief overview of the simulation study and real-world application. In addition, I include a discussion of some caveats for this study and directions for future research.

8.1 Summary of Conclusions

Dynamic spatio-temporal models (DSTMs) can be used to model complex spatio-temporal datasets by incorporating scientific knowledge of physical processes. DSTMs can capture the evolution of spatial processes across space and time; however, it is difficult to extend these models to high-dimensional (large) datasets. In this thesis, I propose a scalable DSTM model that exploits the spatial basis expansions and temporal correlation in the corresponding basis coefficients.

I demonstrate the method on a simulated example as well as a real-world dataset. In the simulated example, the competing method performs slightly better than my proposed method in terms of the RMSE. However, the proposed method is computationally efficient; the computational walltimes reduce by a factor of 8.5 compared to the competing approach (separable model). In the real world example, the proposed method outperforms the competing method in both goodness-of-fit and computational efficiency. Here, the descriptive spatio-temporal model may not capture the underlying dynamic processes of the real-world dataset.

8.2 Caveats and Directions for Future Research

This thesis proposes a novel approach to model dynamic count-valued spatio-temporal observations using a functional representation of the underlying spatio-temporal processes. However, it is subject to key limitations, which can be addressed in subsequent studies. First, our dynamic spatio-temporal is subject to spatial confounding [35, 36, 81], particularly in the diffusion rate $\boldsymbol{\delta}$. In this study, I set $\boldsymbol{\delta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Phi}\boldsymbol{\gamma}_\delta$ where the design matrix $\mathbf{X}$ includes the corresponding longitude (x-direction) coordinates. Since $\mathbf{X}$ includes the same spatial information as those contained in $\boldsymbol{\Phi}\boldsymbol{\gamma}_\delta$, there may be identifiability issues borne out of confounding between the fixed effects $\boldsymbol{\beta}$ and random effects $\boldsymbol{\Phi}\boldsymbol{\gamma}_\delta$. For a future study, I plan on addressing the confounding issue by extending the restricted spatial models from existing studies [35, 36, 81] to the scalable dynamic spatio-temporal modeling framework. In my thesis, I had only considered an intercept term and observation locations as covariates. Future studies may also include environmental variables (e.g. vegetation, elevation, humidity, or others) to improve accuracy for the real world example.

The choice of spatial basis functions remains an open question in this study. While I used a subset of the leading 50 eigenvectors of a Matérn correlation function, using fewer basis functions may potentially lead to over-smoothing of the underlying spatial surfaces. In future studies, I aim to examine variable selection approaches to choose the appropriate number of spatial basis functions. Moreover, this will lead to a closer examination of the tradeoff between prediction accuracy and computational costs. Finally, from an application point-of-view, we aim to modify our approach to model both Gaussian and non-Gaussian spatio-temporal observations in the climate sciences. These include remotely-sensed observations of cloud mask [82], water vapor [83], and ocean temperatures via ARGO floats [84].

Bibliography

[1] D. Giuliani, M. M. Dickson, G. Espa, and F. Santi, "Modelling and predicting the spatio-temporal spread of covid-19 in italy," *BMC infectious diseases*, vol. 20, no. 1, pp. 1–10, 2020.

[2] T. H. Grubesic and E. A. Mack, "Spatio-temporal interaction of urban crime," *Journal of Quantitative Criminology*, vol. 24, no. 3, pp. 285–306, 2008.

[3] H. Nguyen, W. Liu, and F. Chen, "Discovering congestion propagation patterns in spatio-temporal traffic data," *IEEE Transactions on Big Data*, vol. 3, no. 2, pp. 169–180, 2016.

[4] L. G. Spurgin, K. Bebbington, E. A. Fairfield, M. Hammers, J. Komdeur, T. Burke, H. L. Dugdale, and D. S. Richardson, "Spatio-temporal variation in lifelong telomere dynamics in a long-term ecological study," *Journal of Animal Ecology*, vol. 87, no. 1, pp. 187–198, 2018.

[5] A. C. Robinson, D. J. Peuquet, S. Pezanowski, F. A. Hardisty, and B. Swedberg, "Design and evaluation of a geovisual analytics system for uncovering patterns in spatio-temporal event data," *Cartography and Geographic Information Science*, vol. 44, no. 3, pp. 216–228, 2017.

[6] D. R. Farine, A. Strandburg-Peshkin, T. Berger-Wolf, B. Ziebart, I. Brugere, J. Li, and M. C. Crofoot, "Both nearest neighbours and long-term affiliates predict individual locations during collective movement in wild baboons," *Scientific reports*, vol. 6, no. 1, pp. 1–10, 2016.

[7] A. E. Gelfand, L. Zhu, and B. P. Carlin, "On the change of support problem for spatio-temporal data," *Biostatistics*, vol. 2, no. 1, pp. 31–45, 2001.

[8] G. Atluri, A. Karpatne, and V. Kumar, "Spatio-temporal data mining: A survey of problems and methods," *ACM Computing Surveys (CSUR)*, vol. 51, no. 4, pp. 1–41, 2018.

[9] S. Kisilevich, F. Mansmann, M. Nanni, and S. Rinzivillo, "Spatio-temporal clustering," in *Data mining and knowledge discovery handbook*. Springer, 2009, pp. 855–874.

[10] H. Zhang, Y. Zheng, and Y. Yu, "Detecting urban anomalies using multiple spatio-temporal data sources," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–18, 2018.

[11] L. K. Gehlot, S. M. Jibhakate, P. J. Sharma, P. Patel, and P. Timbadiya, "Spatio-temporal variability of rainfall indices and their teleconnections with el niño-southern

oscillation for tapi basin, india," *Asia-Pacific Journal of Atmospheric Sciences*, vol. 57, no. 1, pp. 99–118, 2021.

[12] R. R. Vatsavai, A. Ganguly, V. Chandola, A. Stefanidis, S. Klasky, and S. Shekhar, "Spatiotemporal data mining in the era of big spatial data: algorithms and applications," in *Proceedings of the 1st ACM SIGSPATIAL international workshop on analytics for big geospatial data*, 2012, pp. 1–10.

[13] N. Cressie and C. K. Wikle, *Statistics for spatio-temporal data*. John Wiley & Sons, 2015.

[14] A. Eklund, T. E. Nichols, and H. Knutsson, "Cluster failure: Why fmri inferences for spatial extent have inflated false-positive rates," *Proceedings of the national academy of sciences*, vol. 113, no. 28, pp. 7900–7905, 2016.

[15] P. Dutilleul, *Spatio-temporal heterogeneity: Concepts and analyses*. Cambridge University Press, 2011.

[16] J. A. Byers, "Simulation of the mate-finding behaviour of pine shoot beetles, tomicus piniperda," *Animal Behaviour*, vol. 41, no. 4, pp. 649–660, 1991.

[17] L. Fahrig, "Non-optimal animal movement in human-altered landscapes," *Functional ecology*, vol. 21, no. 6, pp. 1003–1015, 2007.

[18] T. Adam, C. A. Griffiths, V. Leos-Barajas, E. N. Meese, C. G. Lowe, P. G. Blackwell, D. Righton, and R. Langrock, "Joint modelling of multi-scale animal movement data using hierarchical hidden markov models," *Methods in Ecology and Evolution*, vol. 10, no. 9, pp. 1536–1550, 2019.

[19] D. R. Rubenstein and K. A. Hobson, "From birds to butterflies: animal movement patterns and stable isotopes," *Trends in ecology & evolution*, vol. 19, no. 5, pp. 256–263, 2004.

[20] T. J. Hefley and M. B. Hooten, "Hierarchical species distribution models," *Current Landscape Ecology Reports*, vol. 1, no. 2, pp. 87–97, 2016.

[21] M. B. Hooten, H. R. Scharf, T. J. Hefley, A. T. Pearse, and M. D. Weegman, "Animal movement models for migratory individuals and groups," *Methods in Ecology and Evolution*, vol. 9, no. 7, pp. 1692–1705, 2018.

[22] T. J. Hefley and M. B. Hooten, "On the existence of maximum likelihood estimates for presence-only data," *Methods in Ecology and Evolution*, vol. 6, no. 6, pp. 648–655, 2015.

[23] T. L. Wilson, J. B. Odei, M. B. Hooten, and T. C. Edwards Jr, "Hierarchical spatial models for predicting pygmy rabbit distribution and relative abundance," *Journal of Applied Ecology*, vol. 47, no. 2, pp. 401–409, 2010.

[24] P. A. Hernandez, C. H. Graham, L. L. Master, and D. L. Albert, "The effect of sample size and species characteristics on performance of different species distribution modeling methods," *Ecography*, vol. 29, no. 5, pp. 773–785, 2006.

[25] J. N. Williams, C. Seo, J. Thorne, J. K. Nelson, S. Erwin, J. M. O'Brien, and M. W. Schwartz, "Using species distribution models to predict new occurrences for rare plants," *Diversity and Distributions*, vol. 15, no. 4, pp. 565–576, 2009.

[26] L. B. Buckley, M. C. Urban, M. J. Angilletta, L. G. Crozier, L. J. Rissler, and M. W. Sears, "Can mechanism inform species' distribution models?" *Ecology letters*, vol. 13, no. 8, pp. 1041–1054, 2010.

[27] A. Guisan, T. C. Edwards Jr, and T. Hastie, "Generalized linear and generalized additive models in studies of species distributions: setting the scene," *Ecological modelling*, vol. 157, no. 2-3, pp. 89–100, 2002.

[28] M. B. Hooten and C. K. Wikle, "A hierarchical bayesian non-linear spatio-temporal model for the spread of invasive species with application to the eurasian collared-dove," *Environmental and Ecological Statistics*, vol. 15, no. 1, pp. 59–70, 2008.

[29] M. Haran, "Gaussian random field models for spatial data," *Handbook of Markov Chain Monte Carlo*, pp. 449–478, 2011.

[30] P. J. Diggle, J. A. Tawn, and R. A. Moyeed, "Model-based geostatistics," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 47, no. 3, pp. 299–350, 1998.

[31] B. S. Lee and M. Haran, "Picar: An efficient extendable approach for fitting hierarchical spatial models," *Technometrics*, no. just-accepted, pp. 1–29, 2021.

[32] H. Rue and L. Held, *Gaussian Markov random fields: theory and applications*. CRC press, 2005.

[33] M. M. Wall, "A close look at the spatial structure implied by the car and sar models," *Journal of statistical planning and inference*, vol. 121, no. 2, pp. 311–324, 2004.

[34] S. Banerjee, B. P. Carlin, and A. E. Gelfand, *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2003.

[35] B. J. Reich, J. S. Hodges, and V. Zadnik, "Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models," *Biometrics*, vol. 62, no. 4, pp. 1197–1206, 2006.

[36] J. Hughes and M. Haran, "Dimension reduction and alleviation of confounding for spatial generalized linear mixed models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 75, no. 1, pp. 139–159, 2013.

[37] Q. Su, R. Bergquist, Y. Ke, J. Dai, Z. He, F. Gao, Z. Zhang, and Y. Hu, "A comparison of modelling the spatio-temporal pattern of disease: a case study of schistosomiasis japonica in anhui province, china," *Transactions of The Royal Society of Tropical Medicine and Hygiene*, 2021.

[38] M. Dehghani, B. Saghafian, F. Rivaz, and A. Khodadadi, "Monthly stream flow forecasting via dynamic spatio-temporal models," *Stochastic environmental research and risk assessment*, vol. 29, no. 3, pp. 861–874, 2015.

[39] C. K. Wikle, A. Zammit-Mangion, and N. Cressie, *Spatio-temporal Statistics with R*. Chapman and Hall/CRC, 2019.

[40] N. Cressie and H.-C. Huang, "Classes of nonseparable, spatio-temporal stationary covariance functions," *Journal of the American Statistical Association*, vol. 94, no. 448, pp. 1330–1339, 1999.

[41] T. Gneiting, "Nonseparable, stationary covariance functions for space–time data," *Journal of the American Statistical Association*, vol. 97, no. 458, pp. 590–600, 2002.

[42] M. W. Mitchell, M. G. Genton, and M. L. Gumpertz, "Testing for separability of space–time covariances," *Environmetrics: The official journal of the International Environmetrics Society*, vol. 16, no. 8, pp. 819–831, 2005.

[43] J. Møller and M. Ghorbani, "Aspects of second-order analysis of structured inhomogeneous spatio-temporal point processes," *Statistica Neerlandica*, vol. 66, no. 4, pp. 472–491, 2012.

[44] S. De Iaco and D. Posa, "Strict positive definiteness in geostatistics," *Stochastic environmental research and risk assessment*, vol. 32, no. 3, pp. 577–590, 2018.

[45] J. Tan, Q. Wu, M. Zhang, W. Wei, N. Hatziargyriou, F. Liu, and T. Konstantinou, "Wind power scenario generation with non-separable spatio-temporal covariance function and fluctuation-based clustering," *International Journal of Electrical Power & Energy Systems*, vol. 130, p. 106955, 2021.

[46] C. Salvi and M. Lemercier, "Neural stochastic partial differential equations," *arXiv preprint arXiv:2110.10249*, 2021.

[47] J. L. Harvill, "Spatio-temporal processes," *Wiley interdisciplinary reviews: computational statistics*, vol. 2, no. 3, pp. 375–382, 2010.

[48] M. Kuusela and M. L. Stein, "Locally stationary spatio-temporal interpolation of argo profiling float data," *Proceedings of the Royal Society A*, vol. 474, no. 2220, p. 20180400, 2018.

[49] A. Rodrigues and P. J. Diggle, "A class of convolution-based models for spatio-temporal processes with non-separable covariance structure," *Scandinavian Journal of Statistics*, vol. 37, no. 4, pp. 553–567, 2010.

[50] T. J. Hefley, M. B. Hooten, E. M. Hanks, R. E. Russell, and D. P. Walsh, "Dynamic spatio-temporal models for spatial data," *Spatial statistics*, vol. 20, pp. 206–220, 2017.

[51] C. K. Wikle and M. B. Hooten, "A general science-based framework for dynamical spatio-temporal models," *Test*, vol. 19, no. 3, pp. 417–451, 2010.

[52] C. K. Wikle, "Modern perspectives on statistics for spatio-temporal data," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 7, no. 1, pp. 86–98, 2015.

[53] T. Bayes, "Lii. an essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfr s," *Philosophical transactions of the Royal Society of London*, no. 53, pp. 370–418, 1763.

[54] P. S. Laplace, "Memoir on the probability of the causes of events," *Statistical science*, vol. 1, no. 3, pp. 364–378, 1986.

[55] A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian data analysis*. Chapman and Hall/CRC, 1995.

[56] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[57] R. M. Dorazio, "Bayesian data analysis in population ecology: motivations, methods, and benefits," *Population ecology*, vol. 58, no. 1, pp. 31–44, 2016.

[58] R. Little, "Calibrated bayes, for statistics in general, and missing data in particular," *Statistical Science*, vol. 26, no. 2, pp. 162–174, 2011.

[59] J. Berger, "The case for objective bayesian analysis," *Bayesian analysis*, vol. 1, no. 3, pp. 385–402, 2006.

[60] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of state calculations by fast computing machines," *The journal of chemical physics*, vol. 21, no. 6, pp. 1087–1092, 1953.

[61] A. Gelman and D. B. Rubin, "Inference from iterative simulation using multiple sequences," *Statistical science*, pp. 457–472, 1992.

[62] J. M. Flegal and G. L. Jones, "Batch means and spectral variance estimators in markov chain monte carlo," *The Annals of Statistics*, vol. 38, no. 2, pp. 1034–1070, 2010.

[63] O. F. Christensen, G. O. Roberts, and M. Sköld, "Robust markov chain monte carlo methods for spatial generalized linear mixed models," *Journal of Computational and Graphical Statistics*, vol. 15, no. 1, pp. 1–17, 2006.

[64] W. Yang, D. Yang, Y. Zhao, and J. Gong, "Traffic flow prediction based on wavelet transform and radial basis function network," in *2010 International Conference on Logistics Systems and Intelligent Management (ICLSIM)*, vol. 2. IEEE, 2010, pp. 969–972.

[65] O. V. Vasilyev, D. A. Yuen, S. Paolucci, R. Gruber, and J. Rappaz, "Solving pdes using wavelets," *Computers in Physics*, vol. 11, no. 5, pp. 429–435, 1997.

[66] W. Penny, N. Trujillo-Barreto, and E. Aubert, "Spatio-temporal models for eeg," *Friston et al.(Eds.), Statistical Parametric Mapping: Models for Brain Imaging. Elsevier*, 2006.

[67] A. Cohen, I. Daubechies, and J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Communications on pure and applied mathematics*, vol. 45, no. 5, pp. 485–560, 1992.

[68] L. Chun-Lin, "A tutorial of the wavelet transform," *NTUEE, Taiwan*, vol. 21, p. 22, 2010.

[69] J. H. Mathews, K. D. Fink *et al.*, *Numerical methods using MATLAB*. Pearson prentice hall Upper Saddle River, NJ, 2004, vol. 4.

[70] M. A. Unser and T. Blu, "Wavelets and radial basis functions: A unifying perspective," in *Wavelet Applications in Signal and Image Processing VIII*, vol. 4119. International Society for Optics and Photonics, 2000, pp. 487–493.

[71] J. Fourier, "On the propagation of heat in solid bodies," *Memoir, read before the Class of the Instut de France*, 1807.

[72] L. Yu, Y. Zhao, and L. Tang, "Ensemble forecasting for complex time series using sparse representation and neural networks," *Journal of Forecasting*, vol. 36, no. 2, pp. 122–138, 2017.

[73] N. G. Roland, "Fourier and wavelet representations of functions," *Furman University Electronic Journal of Undergraduate Mathematics*, vol. 6, no. 1, pp. 1–12, 2000.

[74] J. A. F. Diniz-Filho and L. M. Bini, "Modelling geographical patterns in species richness using eigenvector-based spatial filters," *Global Ecology and Biogeography*, vol. 14, no. 2, pp. 177–185, 2005.

[75] D. A. Griffith, "Spatial filtering," in *Spatial Autocorrelation and Spatial Filtering*. Springer, 2003, pp. 91–130.

[76] N. Shigesada, K. Kawasaki *et al.*, "Invasion and the range expansion of species: effects of long-distance dispersal," *Dispersal ecology*, pp. 350–373, 2002.

[77] M. Kuussaari, I. Saccheri, M. Camara, and I. Hanski, "Allee effect and population dynamics in the glanville fritillary butterfly," *Oikos*, pp. 384–392, 1998.

[78] C. K. Wikle and M. B. Hooten, "Hierarchical bayesian spatio-temporal models for population spread," *Applications of computational statistics in the environmental sciences: hierarchical Bayes and MCMC methods*, vol. 145169, 2006.

[79] S. A. Johnson and G. Donaldson-Fortier, "Florida's introduced birds: Eurasian collared-dove (streptopelia decaocto)," *EDIS*, vol. 2009, no. 4, 2009.

[80] C. S. Robbins, D. Bystrak, and P. H. Geissler, "The breeding bird survey: its first fifteen years, 1965-1979," PATUXENT WILDLIFE RESEARCH CENTER LAUREL MD, Tech. Rep., 1986.

[81] J. S. Hodges and B. J. Reich, "Adding spatially-correlated errors can mess up the fixed effect you love," *The American Statistician*, vol. 64, no. 4, pp. 325–334, 2010.

[82] A. Sengupta, N. Cressie, B. H. Kahn, and R. Frey, "Predictive inference for big, spatial, non-Gaussian data: MODIS cloud data and its change-of-support," *Australian & New Zealand Journal of Statistics*, vol. 58, no. 1, pp. 15–45, 2016.

[83] D. Zilber and M. Katzfuss, "Vecchia–Laplace approximations of generalized Gaussian processes for big non-gaussian spatial data," *Computational Statistics & Data Analysis*, vol. 153, p. 107081, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167947320301729

[84] S. R. Jayne, D. Roemmich, N. Zilberman, S. C. Riser, K. S. Johnson, G. C. Johnson, and S. R. Piotrowicz, "The argo program: present and future," *Oceanography*, vol. 30, no. 2, pp. 18–28, 2017.

Curriculum Vitae

Education

- M.S. Statistics, George Mason University (GMU)
- B.S. Mathematics, National Central University (NCU)

Honors and Awards

- Summer Research Fellowship (GMU) — Summer 2022
- Taiwan Study Abroad Scholarship — 2018
- Taiwanese Academic Excellence Award — Fall 2018
- Three times Dean's List Award at NCU — Spring 2015, 2016 and 2017