8-2019

Bayesian Spatio-Temporal Modeling for Forecasting, Trend Assessment and Spatial Trend Filtering

Stella Coker Watson Self
Clemson University, stellaw@clemson.edu

Bayesian Spatio-Temporal Modeling for Forecasting, Trend Assessment and Spatial Trend Filtering

---

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Statistics

---

by
Stella Coker Watson Self
August 2019

Accepted by:
Dr. Christopher McMahan, Committee Chair
Dr. Andrew Brown
Dr. Robert Lund
Dr. Brook Russell

Abstract

This work develops Bayesian spatio-temporal modeling techniques specifically aimed at studying several aspects of our motivating applications; to include vector-borne disease incidence and air pollution levels. A key attribute of the proposed techniques are that they are scalable to extremely large data sets which consist of spatio-temporally oriented observations. Largely, the scalability of our modeling strategies is accomplished in two primary ways. First, through the introduction of carefully constructed latent random variables we are able to develop Markov chain Monte Carlo (MCMC) sampling algorithms that consist primarily of Gibbs steps. This leads to the fast and easy updating of the model parameters from common distributions. Second, for the spatio-temporal aspects of the models, a novel sampling strategy for Gaussian Markov random fields (GRMFs) that can be easily implemented (in parallel) within MCMC sampling algorithms is used. The performance of the proposed modeling strategies are demonstrated through extensive numerical studies and are further used to analyze vector-borne disease data measured on canines throughout the conterminous United States and $PM_{2.5}$ levels measured at weather stations throughout the Eastern United States.

In particular, we begin by developing a Poisson regression model that can be used to forecast the incidence of vector-borne disease throughout a large geographic area. The proposed model accounts for spatio-temporal dependence through a vector autoregression and is fit through a Metropolis-Hastings based Markov chain Monte Carlo (MCMC) sampling algorithm. The model is used to forecast the prevalnce of Lyme disease (Chapter 2) and Anaplasmosis (Chapter 3) in canines throughout the United States. As a part of these studies we also evaluate the significance of various climatic and socio-economic drivers of disease. We then present (Chapter 4) the development of the *chromatic sampler* for GMRFs. The chromatic sampler is an MCMC sampling technique that exploits the Markov property of GMRFs to sample large groups of parameters in parallel. A greedy algorithm for finding such groups of parameters is presented. The methodology is found

to be superior, in terms of computational effort, to both full block and single-site updating. For assessing spatio-temporal trends, we develop (Chapter 5) a binomial regression model with spatially varying coefficients. This model uses Gaussian predictive processes to estimate spatially varying coefficients and a conditional autoregressive structure embedded in a vector autoregression to account for spatio-temporal dependence in the data. The methodology is capable of estimating both widespread regional and small scale local trends. A data augmentation strategy is used to develop a Gibbs based MCMC sampling routine. The approach is made computationally feasible through adopting the chromatic sampler for GMRFs to sample the spatio-temporal random effects. The model is applied to a dataset consisting of 16 million test results for antibodies to *Borrelia burgdoferi* and used to identify several areas of the United States experiencing increasing Lyme disease risk. For nonparametric functional estimation, we develop (Chapter 6) a Bayesian multidimensional trend filter (BMTF). The BMTF is a flexible nonparameteric estimator that extends traditional one dimensional trend filtering methods to multiple dimensions. The methodology is computationally scalable to a large support space and the expense of fitting the model is nearly independent of the number of observations. The methodology involves discretizing the support space and estimating a multidimensional step function over the discretized support. Two adaptive methods of discretization which allows the data to determine the resolution of the resulting function is presented. The BMTF is then used (Chapter 7) to allow for spatially varying coefficients within a quantile regression model. A data augmentation strategy is introduced which facilitates the development of a Gibbs based MCMC sampling routine. This methodology is developed to study various meteorological drivers of high levels of $PM_{2.5}$, where $PM_{2.5}$ is a particularly hazardous form of air pollution consisting of particles less than 2.5 micrometers in diameter.

Dedication

To my loving husband, for all his support, to my undergraduate professors at Furman University, who first inspired me with the love of mathematics, and to my graduate research mentors at Clemson University, without whom this dissertation would not exist.

Acknowledgments

Many thanks to my dissertation advisor, Dr. Christopher McMahan, without whom none of this work would have been possible. I would also like to thank the other members of my dissertation committee, Drs. Andrew Brown, Robert Lund, and Brook Russell, for their feedback and advice. Much of this work was completed in collaboration with other co-authors. In addition to my committee members, I would like to thank Drs. Jenna Gettings, Yan Liu, Shila Nordone, and Michael Yabsley. I would also like to thank the Companion Animal Parasite Council for supporting this work.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Spatial data arise in a variety of applications, including epidemiology, environmental science, and ecology. Spatial data consist primarily of two types: areal and point process. Areal data occur when observations are pooled over a spatial unit, such as when the total number of cases of a disease are reported at a county or state level. Point process data consists of observation(s) directly associated with a particular point in space, such as the latitude-longitude or the street address. In recent years, a variety of techniques for analyzing both types of these data have been developed. As the size of the typical spatial data set continues to grow over time, much of the research in spatial statistics focuses on developing computationally scalable methodologies. The Bayesian paradigm has proven to be especially useful for the development of these scalable methodologies.

One of the most flexible and widely used spatial modeling tools for point process data is the Gaussian process (GP). As a special case of a stochastic process, a Gaussian process has the property that every finite collection (taken in say space or time) of realizations obeys a multivariate normal distribution. In general, GPs are governed by a mean and covariance function. The covariance between two locations is typically a function of the distance between them, and a variety of different covariance functions have been explored. For more on Gaussian processes, see Cressie [1993]. For an overview of recent developments and extension of Gaussian processes, see Gelfand and Schliep [2016] and Heaton et al. [2017b].

Gaussian process models have been extended in many ways. Palacios and Steel [2006] develop a similar process for t-distributions. Zhang and El-Shaarawi [2009] develop a skew Gaussian process to allow for differing behavior in the left and right tails of the distribution. Lum and

Gelfand [2012] extend Gaussian processes to the quantile regression context via the asymmetric Laplace distribution. Banerjee et al. [2003] and Banerjee and Gelfand [2003] develop methods for quantifying the degree of smoothness of Gaussian process realizations.

One drawback to Gaussian process models is that the dimension of the associated multivariate normal distribution is equal to the number of observations. Furthermore, the covariance matrix of this distribution is dense, making GPs computationally expensive when the number of observations is large. A number of techniques have been developed to address this problem. One of the more commonly used techniques is the Gaussian predictive process (GPP), which reduces the dimension of the problem by defining a parent process at a set of knot locations. Estimation is performed at the knot locations and the process is interpolated to the observation locations via kriging. As the number of knots is typically much smaller than the number of observation locations, this results in significant computational savings. For more on Gaussian predictive processes, see Banerjee et al. [2008]. Another dimension reduction technique is the nearest neighbor Gaussian process model. In this model, the full conditional distribution of the random variable associated with each location is assumed to depend only on a subset of the other locations (the nearest neighbors). This results in a sparse covariance matrix for the joint distribution. For more on nearest neighbor Gaussian processes, see Datta et al. [2016b]. Covariance tapering [Furrer et al., 2006] makes use of the fact the covariance between two locations approaches 0 as the distance between the locations increases. Covariance tapering assumes a compact support which creates a sparse covariance matrix for the joint distribution.

In contrast, Gaussian Markov random fields (GMRFs) are one of the most frequently used techniques for modeling areal data, with conditional autoregressive (CAR) models being the most commonly used. GMRFs are multivariate Gaussian random vectors defined via sparse precision matrices which induce conditional independence properties among the components of the random vector. For more on GMRF's see Rue and Held [2005b]. CAR models are GMRF's whose precision matrix is based on the spatial structure of the data. For example, CAR models typically specify that two locations are conditionally independent if and only if they are not adjacent to each other spatially. For more on CAR models, see Banerjee et al. [2014a]. A number of extensions to GMRFs and CAR models have been developed in recent years. One extension to the original CAR model is the Besag-York-Mollie (BYM) model, which combines a CAR model with a set of independent spatial random effects. See Besag et al. [1991a] for more on Besag-York-Mollie models. Rue and Tjemland

[2002], Song et al. [2008], and Lindgren et al. [2011b] study the relationship between GMRF's and Gaussian processes. Brezger et al. [2007] and Yue and Speckman [2010] extend GMRF's to allow for varying degrees of smoothness. Brown et al. [2017a] study CAR models with neighborhood structures other than those based on physical adjacency.

Spatio-temporal data arises in a variety of applications, including meteorology, the social sciences, and disease forecasting. Modeling the temporal component in these data can be done in a variety of ways, most of which consist of combining a spatial method such as a CAR or BYM model with a technique from time series analysis, such as an autoregressive structure, moving average model, random walk, or spline-based method. For example, Knorr-Held and Besag [1998] combine a CAR model with a random walk in time, while MacNab and Dean [2001] combine a CAR model with a spline-based temporal structure. Sometimes, separate spatial structures are fit for each time point, such as in Waller et al. [1997b] and Knorr-Held and Besag [1998] who fit a BYM model at each time point. Anderson and Ryan [2017] provide a review and comparison of different spatio-temporal modeling techniques.

Spatio-temporal techniques are frequently used for disease mapping and forecasting. Waller et al. [1997b] provide a general Bayesian framework for disease mapping that allows for spatio-temporal interaction. Knorr-Held and Besag [1998] present a Bayesian model for mapping lung cancer risk with an additive spatio-temporal structure. Martinez-Beneito et al. [2008] combine CAR spatial structures with temporal autogregressive structures. Johnson et al. [2017] use heteroskedastic Gaussian processes to forecast dengue fever. Bhatt et al. [2017] use a Gaussian process stacked generalization to develop a disease mapping model based on a GP with a nonlinear mean function. Knorr-Held and Raer [2000] develop a Bayesian method for identifying geographic clusters of increasing disease risk. Anderson et al. [2014] develops a two-stage Bayesian method for identifying clusters of high disease risk, in which clusters are first identified using a hierarchical clustering algorithm, and then separate models are fit to each cluster. Anderson and Ryan [2017] compare several different spatio-temporal models for disease mapping.

Trend filtering, popularized by Hodrick and Prescott [1997], is a commonly used technique for the analysis of time series data. Trend filtering estimates the underlying trend in the data by smoothing out the effects of noise or seasonal cycles. Trend filters penalize the kth order derivatives over the observed time points. There are many variants on trend filtering. The original method, Hodrick-Prescott trend filtering, penalizes the squared backwards finite difference approximation

to the second derivative of the estimated trend function. Seung-Jean et al. [2009] propose another widely used variant of trend filtering, which uses an l_1 penalty to create a piecewise linear estimator. Tibshirani [2014] connects Seung-Jean et al. [2009]'s l_1 trend filter to smoothing splines and locally adaptive regression splines and demonstrates that the l_1 trend filtering estimate will converge to the true trend function if the correct tuning parameter is used. An efficient algorithm for fitting trend filters to data is developed by Ramdas and J. Tibshirani [2014]. Several Bayesian trend filters have recently been developed, including Faulkner and Minin [2018] and Roualdes [2015]. Both these methods are based on shrinkage priors. There have been a few attempts to extend trend filtering to two or more dimensions. For example, Seung-Jean et al. [2009] propose a straightforward two dimensional extension of their one-dimensional ℓ_1 trend filter. However this extension requires data to be collected on a regular grid, which limits its usefulness.

The remainder of this work is organized as follows. First, we develop a Bayesian spatio-temporal model for disease forecasting and apply it to Lyme disease (Chapter 2) and Anaplasmosis (Chapter 3) data collected throughout the conterminous United States. We then develop (Chapter 4) the chromatic sampler for Gaussian Markov random fields, a fast MCMC sampling strategy that relies on parallel updating. Making use of the chromatic sampler, a 'Bayesian model with spatially varying coefficients for assessing trends in disease risk is proposed (Chapter 5). Finally, a novel Bayesian multidimensional trend filtering method is proposed (Chapter 6) and used to develop (Chapter 7) a varying coefficient quantile regression model designed to study the spatially oriented effects of various meteorological drivers on high air pollution levels. We then conclude (Chapter 8) with a summary discussion and a brief overview of potential topics for future work.

Chapter 2

A Bayesian Spatio-Temporal Model for Forecasting the Prevalence of Antibodies to *Borrelia burgdorferi*, Causative Agent of Lyme Disease, in Domestic Dogs within the Contiguous United States

2.1 Introduction

Lyme disease, the most common zoonotic tick-borne disease in the United States and Europe Mead [2015], is caused by bacterial spirochetes from the *Borrelia burgdorferi* sensu lato complex, and is transmitted by ticks in the genus *Ixodes* Stanek et al. [2012], Steere et al. [2004]. *Borrelia burgdorferi*

can infect and cause acute and/or chronic Lyme disease in both humans and dogs. Clinically, there are similarities in disease presentation, and diagnosis and treatment follow similar guidelines. In 2014, the Centers for Disease Control and Prevention (CDC) reported 30,000 confirmed human Lyme disease cases, with an estimated 329,000 additional probable cases based on medical claims information from a large insurance database Nelson et al. [2015], Hinckley et al. [2014]. The Companion Animal Parasite Council (CAPC) reported 250,880 dogs, out of 4 million dogs tested, were positive for antibodies to *B. burgdorferi* in 2015 RNC [2012-current]. While the cost of Lyme disease diagnosis and treatment in dogs is not documented, the cost to the US healthcare system for care of humans with Lyme disease is substantial: treatment of Lyme disease and post-treatment Lyme disease syndrome (PTLDS) cost between $712 million and $1.3 billion annually Adrion et al. [2015]. The incidence of disease has been increasing steadily over the last decade McNabb et al. [2007], and as the number of cases increase, the economic impact of Lyme disease is expected to increase as well.

Clinical Lyme disease manifests similarly in people and dogs, with infection most commonly causing transient fever, anorexia, and arthritis Wormser et al. [2006], Littman et al. [2006]. Early erythema migrans lesions have been observed in up to 75% of human patients Steere et al. [2003a], but are no longer considered pathognomonic for Lyme disease Wormser et al. [2006], Steere et al. [2003a]. Chronic disseminated disease in humans may lead to musculoskeletal, neurologic, dermatologic, and rarely cardiac disease Cohn et al. [2012], Costello et al. [2009], Arvikar and Steere [2015], Thompson et al. [2009], Bachur et al. [2015]. Chronic disease in dogs is more often associated with arthropathy but case reports of renal, neurologic, cardiac, and dermatologic disease exist Littman et al. [2006], Little et al. [2010b], Detmer et al. [2016]. Time to the onset of disease after infection, the incubation period, differs between dogs and humans. Dogs have been reported to have an extended two to five month incubation period before becoming symptomatic Littman et al. [2006], in contrast to three to 30 days in humans Arvikar and Steere [2015]. The first signs of clinical disease in dogs are non-specific, including fever, general malaise, lameness, and swelling of local lymph nodes. These symptoms are likely to be overlooked by dog owners because they are transient, lasting only a few days Krupka and Straubinger [2010]. Detecting the later stages of disease require recognition of pain, however, a standardized protocol for pain assessment in veterinary species is lacking Wiseman-Orr et al. [2004], Mathews et al. [2014], Benito et al. [2013] and mainly relies on dog owners to report disease symptoms. The assessment of pain in dogs can be difficult as they cannot self-report and is often reported by the owner as lethargy, decreased activity, or difficulty getting

up, walking, or navigating stairs. It is often not until the dogs exhibit the characteristic shifting leg lameness several months after infection that owners note any abnormalities. Finally, Lyme disease is most commonly diagnosed by measurement of antibodies specific for *B. burgdorferi* (termed seropositivity). In humans, seropositivity is assessed in a two-step process involving an enzyme-linked immunosorbent assay (ELISA) and Western blot analysis of serum Branda et al. [2010], Lipsett et al. [2015]. Veterinary wellness exams commonly include annual screening for exposure to *B.burdgorferi*, as well as *Ehrlichia* spp., *Anaplasma* spp., and *Dirofilaria immitis* (etiologic agent of Heartworm disease) using a rapid, in-house ELISA platform (SNAP®3Dx®, SNAP®4Dx® and SNAP®4Dx® Plus, IDEXX Laboratories, Inc.). In this point-of-care assay, seropositivity against *B.burgdorferi* is established using the late phase C6 antigen that is indicative of disseminated disease Stillman et al. [2014], Bacon et al. [2003], Embers et al. [2016], Nyman et al. [2006], Wagner et al. [2012a]. Specifically in dogs, seroconversion to *B. burgdorferi* antigen C6 can occur as early as 3-4 weeks post-infection when bacterial burden is high Wagner et al. [2012a]. Multiplex assays suggest antibodies to OspC, OspF, and C6 are characteristic of the intermediate stage of infection (at least 3-4 weeks post-infection), while antibodies to OspF and C6, in the absence of antibodies to OspC, are suggestive of late infection stage Wagner et al. [2012a]. Data from point-of-care tests for C6 seropositivity have been used to develop distribution maps of *B. burgdorferi* seroprevalence in domestic dogs throughout the United States Little et al. [2014], Bowman et al. [2009].

In the United States, *B. burgdorferi* is transmitted by *Ixodes scapularis* on the East coast and Midwest and *Ixodes pacificus* on the West coast, while a variety of wildlife species (e.g., mice, squirrels, shrews) serve as reservoirs Mead [2015]. Although not infected with *B. burgdorferi*, deer and migratory birds play an important role in maintaining and transporting tick vectors Scott [2016], Madhav et al. [2004]. The risk of Lyme disease exposure is therefore related to abundance of a suitable reservoir and exposure to infected ticks. Overall, 96% of human Lyme disease cases reported by the CDC are from 14 states (Connecticut, Delaware, Maine, Massachusetts, Maryland, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, Virginia, and Wisconsin) Adams et al. [2014]. Prevalence of exposure in dogs is similarly high in these states, although the range is expanded beyond the core areas of human cases within these states and into neighboring states such as Ohio, West Virginia and North Dakota RNC [2012-current]. While dogs show no age-specific epidemiological risk profile, children 5-9 years of age are at highest risk of Lyme disease, followed by adults aged 45-59 years RNC [2015]. This bimodal distribution of

human risk is attributed to increased contact with the environment and exposure to infected ticks. Recent assessment of vector distribution identified a 45% increase in *Ixodes* spp. range over the last 18 years Eisen et al. [2016a], with *I. scapularis* firmly established in twice as many counties in the North-Central and Northeastern US as previously believed. Importantly, the Northeastern and North-Central range of *I. scapularis* appears to be merging in the Ohio River Valley, forming a contiguous range of potential vector establishment.

The dynamic change in *Ixodes* spp. ranges suggests ongoing monitoring of these medically important vectors, while important, will be challenging and economically unfeasible on an annual basis. Understanding and forecasting spatial and temporal patterns of risk of exposure to *B. burgdorferi* is thus critical for targeting use of vaccines, preventives measures, and educational campaigns to best protect humans and veterinary species. Building on previously reported risk assessment and surveillance tools for vector-borne disease Bowman et al. [2016b], Liu et al. [2016], we now report on a novel predictive model of canine *B. burgdorferi* exposure using a Bayesian approach to factor assessment and forecasting. As previously described, Bayesian modeling offers a number of advantages over classical approaches Gelman et al. [2014], Harrison and West [1999]. The probabilistic likelihood-based methods are highly flexible and are able to adapt to data availability constraints. These methods are also capable of assessing predictive significance of various covariate factors. The use of data augmentation Markov chain Monte Carlo (MCMC) methods to sample from a posterior distribution provides the opportunity to treat missing data, such as absence of serologic data from certain counties, as latent (missing) variables, even in large populations Besag and Mondal [2013], Hosking et al. [2008]. Finally, a forecast of future seroprevalence, conditional on the past history of data, are easily constructed. The Bayesian methods capably quantify uncertainty both in terms of the potential stochasticity of the disease process and the model parameter estimates.

In what follows, eight factors previously purported to influence canine *B. burgdorferi* seroprevalence will be examined: annual precipitation, annual relative humidity, annual temperature, elevation, percentage forest coverage, percentage surface water coverage, population density, and median household income Stich et al. [2014]. After assessment, a predictive model using the significant factors is developed, and annual *B. burgdorferi* seroprevalence forecasts are constructed. A comparison of actual versus predicted 2015 prevalence is conducted. Intended uses of annual *B. burgdorferi* seroprevalence forecasts for veterinary medicine include: 1) provision of an evidence-based tool used to encourage the year-round use of tick preventive to reduce exposure, and 2) promotion of annual

use of diagnostic testing in areas where the disease is emerging. Finally, based on previous work using canine Lyme disease prevalence to assess human disease risk Little et al. [2010b], Mead et al. [2011a], we highlight the use of canine disease forecast maps to inform risk for human disease in an effort to reduce the burden of human disease on the US healthcare system.

2.2 Materials and Methods

2.2.1 Data Structure

The observed data consist of test results from 11,937,925 *B. burgdorferi* ELISAs performed from 2011 to 2015 in the contiguous United States. The test detects antibodies produced in response to the C6 peptide found on the outer membrane protein, VlsE, of *B. burgdorferi* during infection Levy et al. [2008]. Detection of antibodies does not indicate an active infection, as they will persist over time, even after the infection is resolved. Nor is this a quantitative test, so the time since infection is not known. The seroprevalence reported is a representation of the prevalence of dogs that have been exposed to *B. burgdorferi*, not the prevalence of clinically ill dogs. It is also important to note that the C6 ELISA does not detect antibodies to current vaccines Ting Liang et al. [2000], so vaccinated dogs will not test positive unless the vaccine has failed and the dogs experienced exposure resulting in disseminated infection. The data were provided by a commercial diagnostic laboratory available to veterinary clinics within the study area, IDEXX Laboratories, Inc., and included the county in which the testing clinic resides and the month in which the test was conducted RNI. The submitted samples represent a population of dogs provided veterinary care. No information was available on demographic details of the individuals tested, such as age, sex or breed of dog, nor the travel or testing history of the dog, or the reason why the tests were conducted. Overall, there were 759,103 positive tests, for an empirical prevalence of 0.0635. For the purposes of fitting the model, the data were aggregated by county and by year. Table 2.1 shows the eight considered factors thought to be associated with *B. burgdorferi* prevalence, along with the time period for which data on each factor is available and the level of geographic aggregation (county, state, etc.). For further discussion, including the source of each factor, see Stich et al. [2014], Wang et al. [2014a].

An empirical estimate of the prevalence within each county was constructed in the usual fashion after aggregating the serologic data over the available five year time span; i.e., these estimates were obtained as the proportion of positive tests observed within each reporting county. Figures 2.1

Factors and Notation	Data period	Scale
Annual temperature: $X_{s,1}(t)$	1895 - 2015	Climate Division
Annual precipitation: $X_{s,2}(t)$	1895 - 2015	Climate Division
Annual relative humidity: $X_{s,3}(t)$	2006 - 2015	Climate Division
Elevation: $X_{s,4}(t)$	2012	County
Percentage forest coverage: $X_{s,5}(t)$	2012	County
Percentage surface water coverage: $X_{s,6}(t)$	2010	County
Population density: $X_{s,7}(t)$	2011-2014	County
Median household income: $X_{s,8}(t)$	1997-2014	County

Table 2.1: Factors purported to influence *B. burgdorferi* seroprevalence.

Note, $X_{s,k}(t)$ is used to denote the value of the kth factor in the sth county during the tth year.

and 2.2 display the empirical prevalences and the total number of tests, respectively, within each reporting county. From Figure 2.1, one will note that there appears to be a great deal of positive spatial correlation. Moreover, in studies such as these temporal dependence is expected in the data. Thus, to provide an accurate analysis these spatio-temporal dependencies have to be accounted for in the model. Further, from Figure 2.2, one will note that some counties report relatively small number of tests, thus the interpretation of the aforementioned empirical estimates may be slightly misleading, if one does not consider this effect. For example, a county reporting only 20 test results, with 1 being positive, results in an empirical estimate of 5%, where as the same empirical estimate would be obtained for a county reporting 2000 test results, with 100 being positive. The salient point, more faith should be placed in the latter estimate, when compared to the former, since it is derived from more information.

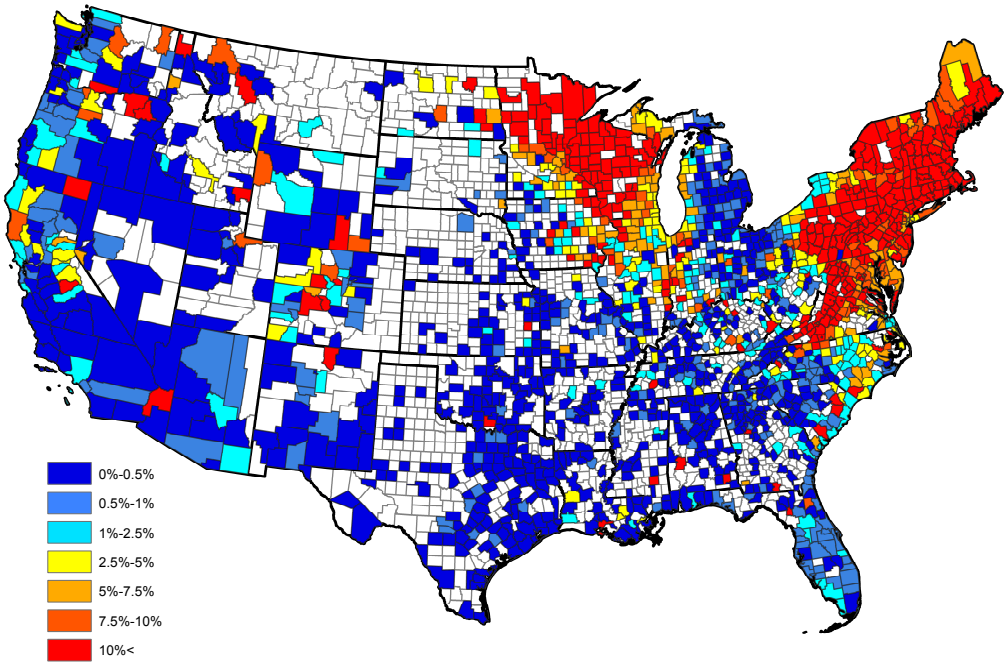


Figure 2.1: County level raw prevalences for *B. burgdorferi* antibodies in domestic dogs aggregated from 2011-2015.

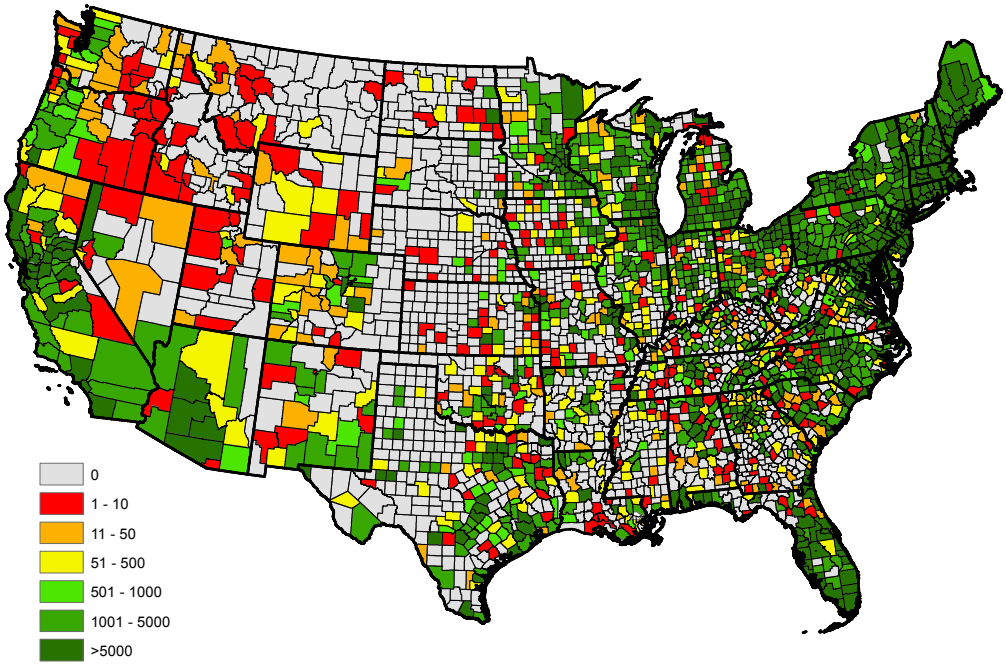


Figure 2.2: Total number of serologic test results for *B. burgdorferi* antibodies in domestic dogs collected within each county during the years of 2011-2015.

In order to construct a baseline *B. burgdorferi* antibody prevalence map, the empirical prevalences displayed in Figure 2.1 were subjected to an ensemble of spatial smoothing techniques. First, weighted head-banging was applied to the empirical prevalences, to acknowledge spatial correlation and diminish the influence of counties which reported only a few tests. For example, there were 57,785 test results reported in South Carolina during the year of 2015, of which 606 were positive, translating to an empirical prevalence estimate of 1%. Further, in Colleton county, South Carolina, only 13 test results were reported, of which 3 were positive, translating to an empirical prevalence estimate of 23%. Obviously, it is inappropriate to believe that the true prevalence is anywhere near as high as the empirical estimate tends to suggest in Colleton county, and that this effect would have been mitigated if more serologic data were available in this region. In the absence of additional serologic data, weighted head-banging provides a robust prevalence estimate at a given spatial location by combining over information that is available in near by areas, thus, for the most part circumventing the small sample size issue. This smoothing procedure used 45 triples and weighted the prevalence of each county proportional to the number of tests from that county. For further

details on weighted head-banging and its uses within the context of disease mapping see Wang et al. [2014a]. Second, in order to render a spatially smooth and complete map Kriging was implemented to interpolate the prevalences of counties not reporting data. Kriging was implemented in ArcGIS using the default parameter settings Environmental Systems Research Institute [2016]. Figure 2.3 provides the baseline *B. burgdorferi* antibody prevalence map resulting from this process, and should be thought of as a baseline for *B. burgdorferi* prevalence in domestic dogs within the contiguous United States. Tincubation conditions and hat is, this figure depicts the general spatial trend of *B. burgdorferi* prevalence in domestic dogs, and will serve as a device which allows one to assess the performance of Bayesian spatio-temporal model developed in the subsequent section, with respect to capturing these trends.

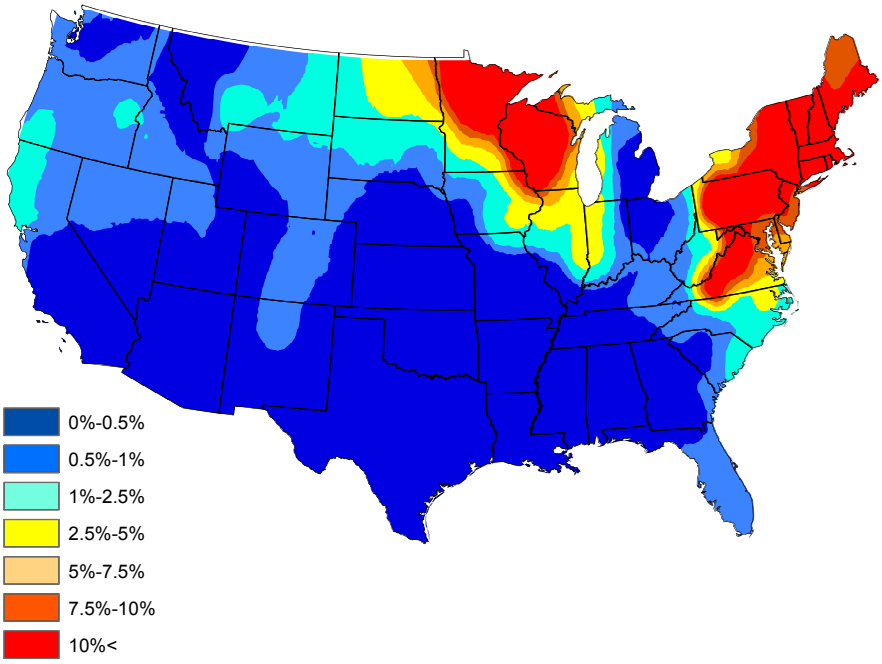


Figure 2.3: Head-banged baseline map showing *B. burgdorferi* antibody prevalences in domestic dogs for an average year during 2011-2015.

2.2.2 Statistical Model

This section describes the statistical model and the techniques used to fit it. The purpose of the model is twofold: to identify environmental and societal risk factors which are significantly associated with the prevalence of *B. burgdorferi* antibodies in dogs and to predict future trends in the prevalence of *B. burgdorferi* antibodies in dogs.

Let $Y_s(t)$ and $n_s(t)$ denote the number of positive and total tests conducted in county s during year t, respectively, for counties $s \in \{1, \ldots, S\}$ and years $t \in \{1, \ldots, T\}$. It is important to note that these values are the raw counts available in the data set; i.e., they are not the post smoothed values resulting from the construction of the baseline map. The available serologic data exhibits both positive spatial and temporal correlation; that is, prevalences in adjacent counties or in successive years tend to have similar values. Thus, to account for these effects this analysis considers using a spatio-temporal model to analyze these data; for additional information about spatial and spatio-temporal models, see Banerjee et al. [2014b], Besag [1974b], Lopez-Quilez and Munoz [2009], Martinez-Beneito et al. [2008], Nobre et al. [2005], Xia and Carlin [1998], Cressie and Wikle [2011b]. In particular, this work uses a Bayesian hierarchical model, which models spatio-temporal dependence through the use of random effects. The details of the distribution of these random effects are described in totality below. In the considered model, the number of positive tests were assumed to follow a Poisson distribution, which is a common choice for modeling count data Lopez-Quilez and Munoz [2009], Martinez-Beneito et al. [2008], Nobre et al. [2005], Xia and Carlin [1998]. Under this specification, the number of tests which are serologically positive for *B. burgdorferi* antibodies in county s during year t (i.e., $Y_s(t)$) is assumed to be distributed as

$$Y_s(t)|n_s(t), p_s(t) \sim \text{Poisson}\left\{n_s(t)p_s(t)\right\}, \tag{2.1}$$

$$\log\left\{p_s(t)\right\} = \beta_0 + \sum_{k=1}^{p} \beta_k X_{s,k}(t) + \xi_s(t), \tag{2.2}$$

where $\log(\cdot)$ denotes the natural logarithm, $'$ denotes the transpose of a matrix (or vector), $\sim$ reads "has the distributional type," and $|$ reads "given." Thus, Eq (2.1) reads "$Y_s(t)$ given $n_s(t)$ and $p_s(t)$ follows a Poisson distribution with mean $n_s(t)p_s(t)$." Furthermore, $\mathbf{X}_s(t) = (X_{s,1}(t), \ldots, X_{s,p}(t))'$ is a vector of covariate information for county s at time t, $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)'$ is the corresponding vector of regression coefficients, and $p_s(t)$ is interpreted as the prevalence of *B. burgdorferi* antibodies in

dogs residing in county s at time t. The random effects, $\xi_s(t)$, are used to account for the spatio-temporal dependence, so that the positive test counts (i.e., $Y_s(t)$'s) are conditionally independent of each other given the total number of tests, the factor information, and the random effects. Note, this does not imply that the $Y_s(t)$'s are independent across varying space s or time t, only that they are conditionally independent once the spatio-temporal correlation is accounted for through the random effects and the other covariate information.

Eq (2.2) specifies the relationship between the prevalence $p_s(t)$ and the covariate information $\mathbf{X}_s(t)$ and the random effect $\xi_s(t)$. This specification is standard for Poisson regression models Lopez-Quilez and Munoz [2009], Martinez-Beneito et al. [2008], Nobre et al. [2005], Xia and Carlin [1998]. In general, different spatio-temporal models specify different structures for the $\{\xi_s(t)\}$. By far, one of the most popular models for areal data is the conditional autoregresive (CAR) model, and it is the one used here Banerjee et al. [2014b]. To further expound on how the CAR model is used in this analysis, it is noted that both spatial and temporal correlation is accounted for through the following multivariate autoregressive model

$$\boldsymbol{\xi}_1 = \boldsymbol{\phi}_1, \tag{2.3}$$

$$\boldsymbol{\xi}_t = \varphi \boldsymbol{\xi}_{t-1} + \boldsymbol{\phi}_t, \text{ for } t = 2, \ldots, T, \tag{2.4}$$

$$\boldsymbol{\phi}_t \sim \text{CAR}(\tau^2; \rho), \text{ for } t = 1, \ldots, T, \tag{2.5}$$

where $\boldsymbol{\xi}_t = (\xi_1(t), \ldots, \xi_S(t))'$ and $\boldsymbol{\phi}_t = (\phi_1(t), \ldots, \phi_S(t))'$ are random vectors. Eq (2.5) specifies that $\boldsymbol{\phi}_t$ are independent and identically distributed random vectors whose distribution follows a CAR model Besag [1974b], Banerjee et al. [2014b]. From Eq (2.4) it is apparent how the model accounts for temporal correlation. That is, in the multivariate autoregressive model depicted above, time-dependence is modeled through a temporal autoregressive model of order one (AR(1)), which is a staple among time series analysis Brockwell and Davis [2002b]. Eq (2.4) relates year t to year $t-1$. The parameter φ is the temporal correlation between consecutive years and lies within $(-1, 1)$. This ensures a causal and stationary solution to the time series model Brockwell and Davis [2002b].

To examine the treatment of spatial correlation, let $\phi = (\phi_1, ..., \phi_S)'$ follow a CAR model, where for the ease of exposition the dependence on t has been suppressed, and it is noted that the subscripts correspond to the S spatial locations. There are several different kinds of CAR models. Usually, CAR models are defined by assigning a univariate distribution for each ϕ_s, whose mean

and variance depends on the spatial relationship between location s and the other locations. This approach uses the CAR model proposed in Besag [1974b], which specifies the conditional distribution for each ϕ_s to be

$$\phi_s \mid \boldsymbol{\phi}_{-s}, \tau^2, \rho, \mathbf{W} \sim \mathrm{N}\left(\rho \frac{\sum_{s' \neq s} w_{s,s'} \phi_{s'}}{\sum_{s' \neq s} w_{s,s'}}, \quad \frac{\tau^2}{\sum_{s' \neq s} w_{s,s'}}\right), \text{ for } s = 1, \ldots, S. \tag{2.6}$$

Here, $\boldsymbol{\phi}_{-s} = (\phi_1, \ldots, \phi_{s-1}, \phi_{s+1}, \ldots, \phi_S)'$ is an $S-1$ dimensional vector that includes every $\phi_{s'}$ except ϕ_s. $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. The $S \times S$ matrix $\mathbf{W}$ is defined by $\mathbf{W} = \{w_{s,s'}\}$, where $w_{s,s'} = 1$ if the s'th and sth counties are adjacent and $w_{s,s'} = 0$ otherwise.

The parameter τ^2 in Eq (3.4) scales the variance structure. Moreover, the conditional variance of ϕ_s, given $\boldsymbol{\phi}_{-s}$, is inversely proportional to the number of counties bordering it. That is, the ϕ_s for counties with more neighbors have a smaller overall variance. This agrees with intuition; if county s has many neighbors, the model has more information to use in estimating ϕ_s, and thus the variance of ϕ_s should be smaller. In Eq (3.4), $\rho \in [0, 1]$ controls the correlation between bordering counties. The conditional mean of ϕ_s, given $\boldsymbol{\phi}_{-s}$, is the average of the $\phi_{s'}$ of the neighboring counties weighted by ρ. Thus, as ρ increases so does the degree of dependence between ϕ_s and the $\phi_{s'}$ of the neighboring counties.

Using Eq (3.4), it can be shown that the joint distribution of $\boldsymbol{\phi}$ is given by the following multivariate normal distribution

$$\boldsymbol{\phi} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Gamma}), \quad \boldsymbol{\Gamma} = \tau^2 \left(\mathbf{D} - \rho \mathbf{W}\right)^{-1},$$

where $\mathbf{W}$ is the adjacency matrix described above and $\mathbf{D}$ is an $S \times S$ diagonal matrix whose sth diagonal element is the number of neighboring counties for county s.

The model is fit using Bayesian Markov Chain Monte Carlo (MCMC) techniques, with posterior samples being used to derive point estimates of the parameters ($\boldsymbol{\beta}, \varphi, \rho$, and τ^2). Thus, to

complete the Bayesian model formulation, the following prior distributions are specified

$$\beta_k \sim \mathrm{N}(0, 1000), \text{ for } k = 0, \ldots, p; \tag{2.7}$$

$$\varphi \sim \mathrm{Uniform}(-1, 1); \tag{2.8}$$

$$\rho \sim \mathrm{Uniform}(0, 1); \tag{2.9}$$

$$\tau^{-2} \sim \mathrm{Gamma}(0.5, 0.05). \tag{2.10}$$

In particular, a diffuse prior distribution is placed on β_k, for $k = 0, ..., p$. This specification permits the prior to exert little, if any, influence on the posterior distribution, thus allowing the data to primarily drive the analysis and subsequently the conclusions. Uninformative (flat) prior distributions are assigned for φ and ρ, for the same reasons. Here "uninformative" means that all possible values of the parameter have equal probability under the prior distribution. The prior for τ^{-2} is chosen as a conjugate prior. Here "conjugate" means that the posterior and prior distributions are from the same distributional family, which simplifies computation. A posterior sampling algorithm was developed, in the usual fashion, to sample all model parameters and random effects from the posterior distributions. This MCMC sampling algorithm uses a combination of Gibbs and Metropolis-Hastings steps. To complete model fitting, $Y_s(t)$ for counties not reporting test results were treated as latent variables, and were sampled along with the model parameters. The posterior sampling algorithm was conducted using code written in **R** and **C++**. For more information about Bayesian models and MCMC methods, see Gelman et al. [2014].

2.3 Results

2.3.1 Model Assessment

This analysis first considered a full model consisting of all 8 climate, geographic, and societal factors. After fitting the full model, credible intervals were computed. Credible intervals are the Bayesian counterpart to frequentist confidence intervals. Table 2.2 provides the point estimates (i.e., the median of the posterior samples) of the regression coefficeints along with 98.75% highest posterior density (HPD) credible intervals for each of the 8 coefficients. These intervals reflect a 90% familywise error rate using a standard Bonferroni correction for multiple comparison. For more information

about Bayesian credible and HPD intervals, see Gelman et al. [2014].

Factor	Estimate	98.75 % HPD Interval
Percentage forest coverage	4.719	[3.535, 5.828]
Percentage surface water coverage	0.518	[0.230, 0.858]
Elevation	0.058	[0.025, 0.089]
Annual relative humidity	0.005	[-0.001, 0.012]
Annual temperature	-0.037	[-0.053, -0.020]
Annual precipitation	0.011	[-0.048, 0.072]
Population density	-3.442e-5	[-5.545e-5, -1.320e-5]
Median household income	0.001	[-0.002, 0.004]

Table 2.2: Parameter estimates for the full model

Table 2.2 indicates that annual relative humidity, annual precipitation, and median household income are not significant predictors of *B. burgdorferi* seroprevalence because their credible intervals contain 0. Removing combinations of these predictors results in 7 potential reduced models. Each of these potential models was fit, and the only model to have all significant predictors was the model which excluded all three of the predictors listed above. This model was selected as the reduced model, and the results for this model are summarized in Table 2.3. The posterior medians of the remaining model parameters are $\varphi = 0.9404$, $\rho = 0.9997$, and $\tau^2 = 0.5958$, and these point estimates validate the claim of strong positive spatial and temporal dependence.

Factor	Estimate	95 % HPD Interval
Percentage forest coverage	4.698	[3.781, 5.629]
Percentage surface water coverage	0.501	[0.244, 0.788]
Elevation	0.052	[0.026, 0.085]
Annual temperature	-0.039	[-0.053, -0.018]
Population density	-3.610e-5	[-5.283e-5, -2.059e-5]

Table 2.3: Parameter estimates for the reduced model.

Most of the predictors in the reduced model behave intuitively. The coefficients for percent forest and water coverage are positive, and the coefficient for population density is negative, as one might expect. Note that the coefficient of elevation is positive and the coefficient of annual temperature is negative, which may appear to be counterintuitive, for further discussion see the Discussion section. In order to assess how well the Bayesian spatio-temporal model explains the data, the correlation between the baseline and model estimated prevalences was computed. In particular, a baseline estimate for each county was extracted from Figure 2.3, and is denoted as $\widetilde{p}_s$, for $s = 1, ..., S$. A model based estimate for each county was then constructed by averaging over

the 5 yearly estimates available from the fitted (reduced) model; i.e., the model estimate for the sth county is given by $\widehat{p}_s = 5^{-1}\sum_{t=1}^{5}\widehat{p}_s(t)$, where $\widehat{p}_s(t)$ is the prevalence estimate resulting from the fitted (reduced) model for the sth county during the tth year. Let $\widetilde{\mathbf{P}}$ and $\widehat{\mathbf{P}}$ denote the collection of baseline and model based estimates, respectively, after removing counties not reporting data. The correlation between these two sets was 0.894 indicating that the Bayesian spatio-temporal model provides a good fit to these data.

2.3.2 Forecasting

Under the Bayesian spatio-temporal model, forecasting *B. burgdorferi* seroprevalence in domestic dogs is tantamount to forecasting the factor levels and the spatio-temporal random effects. In this section, the methods used to forecast these variables are elucidated. First, since the primary goal of this work is to provide for a one year ahead forecast, it is reaonable to assume that certain risk factors are static; i.e., the current years value can be used as the forecasted value since expected changes are negligible. These variables include, forestation, water coverage, and elevation. Thus, the risk factors that need to be forecasted for each county are annual temperature and population density.

To forecast annual temperature, historical temperature records were collected from 1895 to 2015 for each county and modeled as an AR(1) model. The AR(1) model for an annual temperature series $\{F_t\}$ (previously denoted by $\{X_{s,1}(t)\}$ for county s and time t) adheres to the following difference equation

$$F_t = \delta + \gamma F_{t-1} + \omega_t,$$

where $\{\omega_t\}$ is zero mean white noise. Standard statistical software packages (e.g., R and SAS) can be used to easily fit AR(1) models. Let $\widehat{\delta}$ and $\widehat{\gamma}$ denote estimates of δ and γ, respectively, and using these estimates a prediction of the annual temperature at year $t+1$ from temperatures from year 1 to year t is obtained as

$$\widehat{F}_{t+1} = \widehat{\delta} + \widehat{\gamma} F_t.$$

In the proposed forecasting method, $\widehat{F}_{t+1}$ is used as next year's annual temperature value. For more information, see Harrison and West [1999].

Forecasting the county population density for next year requires the county areas and their recent population counts. The US Census provides reliable county population counts for 2010 and

estimated state populations for the years of 1969-2014. A simple linear regression model, with time being the independent variable, was fitted to this state level population data. From this model the county population can be forecasted by first predicting the state population and then partitioning this value into the counties within the state at a proportion that agrees with 2010 Census. The forecasted population density is then obtained by dividing the county population by the county's area.

To forecast the spatial and temporal random effects a year in advance, Eq (2.4) is used. In particular, since the $\boldsymbol{\phi}_t$'s are independent and identically distributed over various years, given values of τ^2 and ρ (available from the posterior samples), $\boldsymbol{\phi}_{t+1}$ is generated randomly from the multivariate normal distribution $\mathrm{N}(\mathbf{0}, \tau^2(\mathbf{D} - \rho\mathbf{W})^{-1})$. Then $\boldsymbol{\xi}_{t+1}$ is set to $\boldsymbol{\xi}_{t+1} = \varphi\boldsymbol{\xi}_t + \boldsymbol{\phi}_{t+1}$. This process is repeated for each pair of ρ and τ^2 available from the posterior sample, thus yielding a sample of the next year's random effects, for further details see Gelman et al. [2014], Harrison and West [1999].

In order to assess the fidelity of the proposed forecasting methods, the 2015 test and factor data were removed and the reduced model was fit to the data collected during the years of 2011-2014. The methods described above were then implemented to forecast the 2015 prevalence of *B. burgdorferi* antibodies in domestic dogs. Figures 2.4 and 2.5 present the observed and forecasted *B. burgdorferi* seroprevalences, respectively, for 2015. Further, Figure 2.6 quantifies the localized predictive capability of the proposed approach by depicting the squared difference between the observed and forecasted *B. burgdorferi* seroprevalences for counties reporting more than 25 test results during 2015. From this figure, one will note that the proposed approach provides an accurate regional forecast throughout the contiguous United States. Note, counties reporting fewer than 25 tests were excluded for the small sample size issues discussed previously. To provide a global assessment of the predictive capacity of the proposed forecasting technique, the weighted correlation (with county weights being set to be the number of tests reported; i.e., $n_s(t)$) between the observed and forecasted prevalence estimates was computed, after removing counties not reporting data, with a value of 0.978 being obtained. Thus, this finding tends to suggest that the proposed approach can be used to accurately forecast future trends in *B. burgdorferi* seroprevalence within the contiguous United States. Here the weighted correlation between two sets, say $\mathbf{A} = \{a_s\}_{s=1}^S$ and $\mathbf{B} = \{b_s\}_{s=1}^S$,

is defined as

$$\mathrm{Corr}(\mathbf{A}, \mathbf{B}) = \frac{\sum_{s=1}^{S} w_s (a_s - \bar{a})(b_s - \bar{b})}{\sqrt{\sum_{s=1}^{S} w_s (a_s - \bar{a})^2 \sum_{s=1}^{S} w_s (b_s - \bar{b})^2}} \tag{2.11}$$

where w_s denotes the weight assigned to the sth observation and

$$\bar{a} = \frac{\sum_{s=1}^{S} w_s a_s}{\sum_{s=1}^{S} w_s} \quad \bar{b} = \frac{\sum_{s=1}^{S} w_s b_s}{\sum_{s=1}^{S} w_s}.$$

Note, the purpose of a weighted correlation is identical to that of the usual correlation, with the exception that it accounts for unequal sample sizes through the weights. Figure 2.7 presents the 2016 forecast of *B. burgdorferi* prevalence within the contiguous United States, after smoothing (Kriging with default parameters were used in the software ArcGIS).

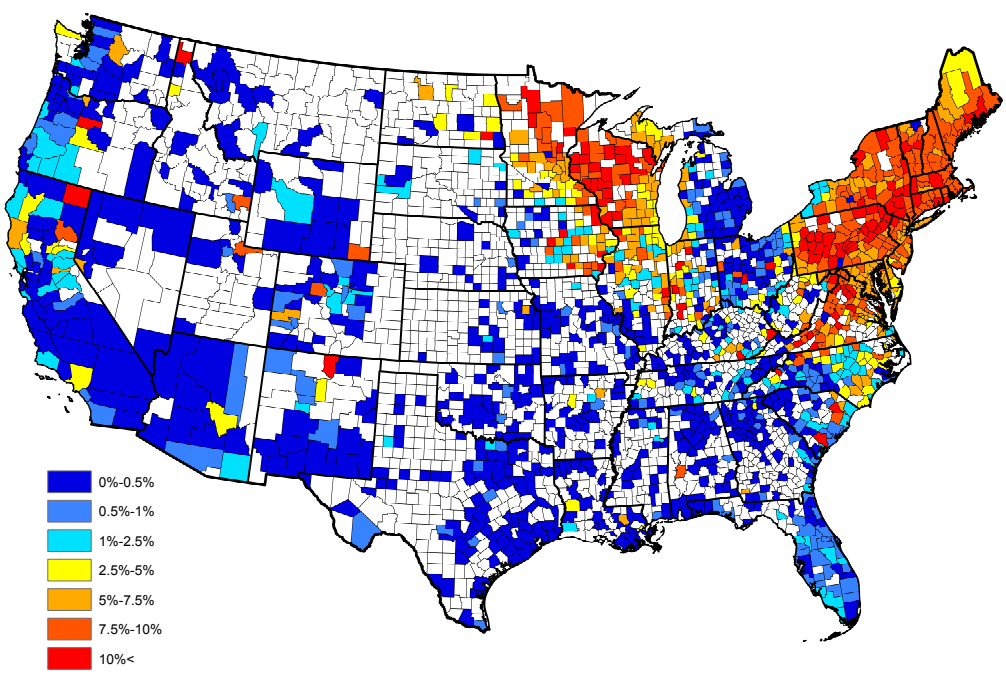


Figure 2.4: Observed *B. burgdorferi* antibody prevalence in domestic dogs for 2015.

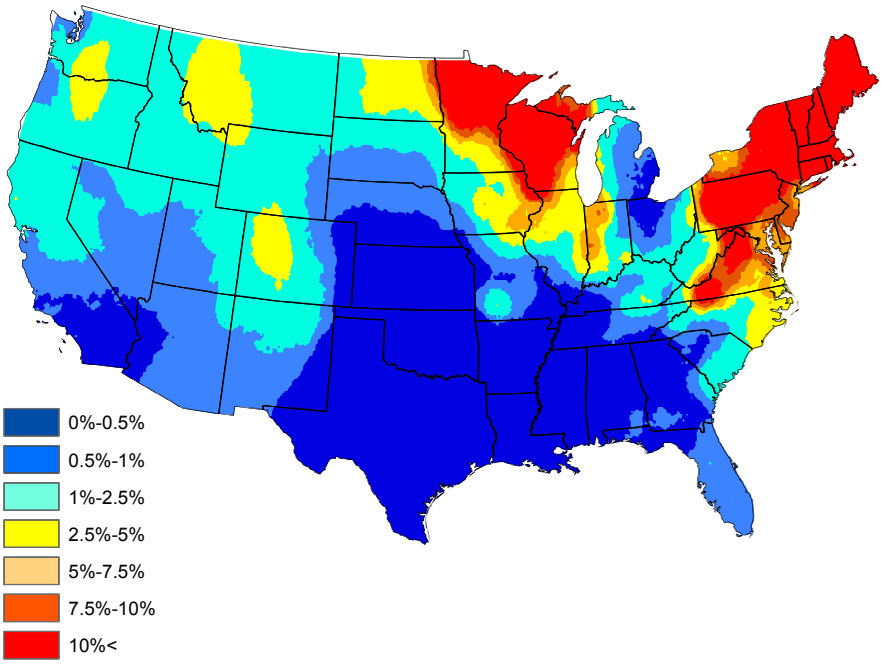


Figure 2.5: Forecasted *B. burgdorferi* antibody prevalence in domestic dogs for 2015.

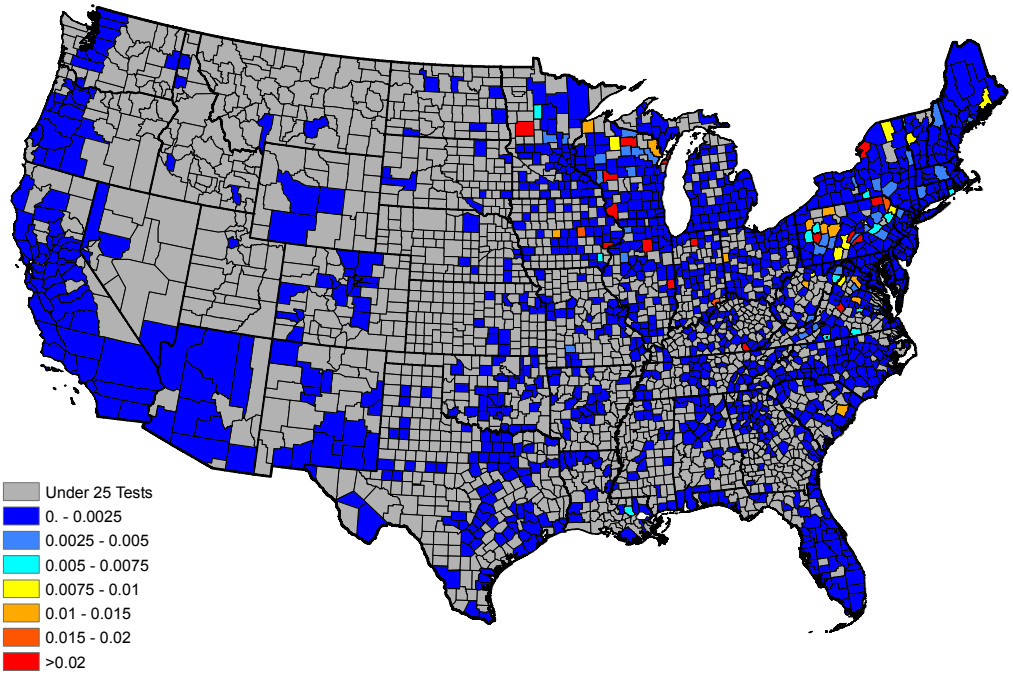


Figure 2.6: Localized predicitve capacity: squared difference between the observed and forecasted *B. burgdorferi* seroprevalences for counties reporting more than 25 test results during 2015.

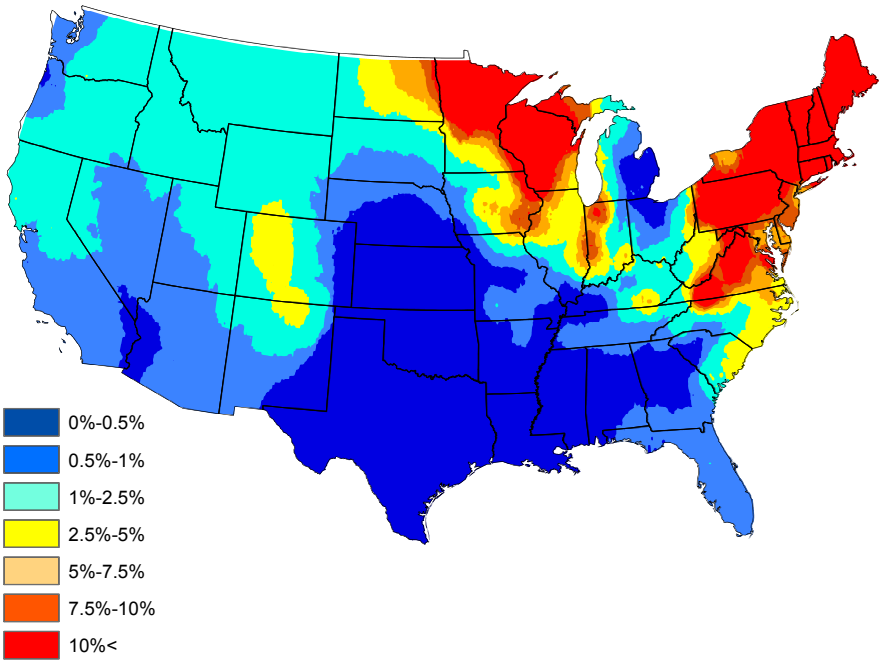


Figure 2.7: Forecasted *B. burgdorferi* antibody prevalence in domestic dogs for 2016.

2.4 Discussion

The objective of this study was to evaluate the space-time patterns and the environmental and socioeconomic drivers of canine *B. burgdorferi* prevalence in the contiguous U.S. using a Bayesian hierarchical modeling approach. Our data were based on serologic testing results for *B. burgdorferi* C6 antigen between 2011-2015. Bayesian hierarchical models have advantages over frequentist approaches for analyzing infectious disease datasets with inherent space-time dependency Hanzlicek et al. [2016], Bowman et al. [2016b], Raghavan et al. [2014, 2016], such as clusters of cases that may be linked due to specific environmental drivers. This is particularly relevant for vector-borne disease as the spatial distribution of most vectors is largely determined by environmental and climatologic conditions and the presence of suitable reservoir hosts Mead [2015], Nieto et al. [2010], Ogden et al. [2013, 2014]. As such, we included relevant ecological covariates, where available, in our model to help explain variability in our aggregated dataset and to strengthen inferences from our Bayesian spatio-temporal model.

The majority of the predictors in our reduced model of covariates are biologically relevant. The coefficients for percent forest and water coverage are positive, and the coefficient for population density is negative, supporting the established role of vector and vertebrate host ecology in disease transmission Ogden et al. [2007]. Tick vectors rely on specific environmental and microhabitat factors for survival while off the host Parham et al. [2015]. The ecology of *B. burgdorferi* is complex and involves numerous vertebrate species that may serve different roles such as reservoirs, dilution hosts, and hosts for the ticks (i.e., white-tailed deer), and all of these are impacted by habitat and anthropogenic changes Levi et al. [2016]. Interestingly, the coefficient of elevation is positive and the coefficient of annual temperature is negative, which is counterintuitive to our understanding of tick ecology and consensus opinion that ticks in the Eastern United States rarely inhabit high elevations (i.e,. Appalachian Mountains) Nieto et al. [2010], Ogden et al. [2013, 2014]. This apparent paradox is likely due to the fact that Lyme disease is more prevalent in the Northeastern United States Adams et al. [2014], which compared with the rest of the country, has a relatively low annual temperature and relatively high elevation. It is important to remember that the proposed model relates the mean of the response to the predictors in a linear fashion and is fitted jointly within the range of the observed data. That is, the relationship between the predictors and the prevalence of Lyme disease is only valid for predictor levels within the range of the observed predictor values, with validity declining at the margins. Further, extrapolations to factor levels not in the range of the considered data set will likely lead to contradictions. Thus the positive coefficient of elevation does not imply that Lyme disease prevalence will continue to increase at extreme elevations, nor does the negative coefficient of annual temperature imply that Lyme disease seroprevalence will continue to increase at extremely low temperatures.

Similar to reported incidence of human Lyme disease Adams et al. [2013], we observe higher 2011-2015 aggregated *B. burgdorferi* seroprevalence in dogs from Connecticut, Delaware, Maine, Massachusetts, Maryland, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont, Virginia, and Wisconsin. However, in contrast, based on data from dogs, we observed an expansion of this endemic range to include Northern California, Southeastern Oregon, Southwestern Idaho, Eastern Colorado and Northern New Mexico (Figure 2.1). Perhaps most striking is the recognized expansion of seropositive dogs on the northern border of the contiguous U.S. along the Canadian border, including North Dakota, and the border of Northern Montana and Idaho. The westward expansion of canine *B. burgdorferi* seroprevalence from Minnesota into North Dakota

mirrors recent reports that Lyme disease is poised to be a significant human public health concern in North Dakota Stone et al. [2015]. Further, this observation supports and extends recent concern over northward expansion of *B. burgdorferi* infected ticks into Canada from the Northeastern and Midwestern United States Gasmi et al. [2016], Nelder et al. [2014].

From five years of historical diagnostic tests, our data show that a Bayesian model can capably quantify *B. burgdorferi* seroprevalence, which ultimately will support qualitative decision-making and surveillance in disease management and response. When comparing actual to forecasted *B. burgdorferi* seroprevalence in 2015, a weighted correlation of 0.978 was achieved, demonstrating significant predictive skill. Finally, using our predictive forecast model we report forecasted 2016 canine *B. burgdorferi* seroprevalence. Of note is the apparent convergence of *B. burgdorferi* infection of dogs from the Northeastern and Mid-Central United States in the Great Lakes region, encompassing Indiana, Ohio, Illinois, Kentucky and Michigan. This observation is supported by recent reports of encroaching *I. scapularis* into this region Eisen et al. [2015], and suggests annual testing of dogs in these states, as well as North Dakota and bordering Canadian provinces is strongly warranted. As an adjunct to annual testing, year-round use of acaracides in dogs can reduce tick infestation, thereby reducing the potential for not only *B. burgdorferi* transmission, but also several other important canine tick-borne pathogens Spencer et al. [2003], Wengenmayer et al. [2014]. Finally, vaccination of dogs against Lyme disease to prevent *B. burgdorferi* infection has been shown to be effective in controlled experimental infection studies LaFleur et al. [2009], Rhodes et al. [2013], Conlon et al. [2000], Zoetis [2016] and in protection against natural infection in dogs living in endemic areas Levy et al. [1993], Levy [2002]. In endemic areas, dogs that are vaccine protocol-compliant are significantly less likely to become infected with *B. burgdorferi* Eschner and Mugnai [2015]. Vaccine studies have concluded that emphasis should be placed on vaccinating dogs at risk for Lyme disease before they are exposed to infected ticks Levy [2002]. As such, we suggest the forecasted areas with an increased likelihood of *B. burgdorferi* transmission outside of the established endemic areas should provide veterinary practitioners with evidence-based options for recommendation of Lyme disease vaccines to clients and protection of dogs against emerging disease.

Limitations to this study include selection bias of the canine population. As mentioned above, samples were submitted for testing by veterinary clinics, and therefore represent dogs under the care of a veterinarian. This suggests that the data are a conservative estimate of the prevalence in domestic dogs because dogs at the highest risk of tick exposure would be those that receive no

veterinary care, those from lower socioeconomic families, or are owned by clients who decline these additional tests during wellness visits. Additionally, a lack of knowledge about the distribution of tick-borne pathogens may limit the testing that veterinarians request; however, these tests are often run during routine heartworm testing so this latter issue is likely a minimal concern. Additional bias could be introduced through variations by region in the use of the diagnostic tests (i.e. whether it is used predominately for wellness visits or for cases in which the disease is suspected prior to testing) or variations by region in the use of reference laboratory services and products. It is also recognized that travel history is not controlled for and will account for some of the cases outside endemic regions. Despite these limitations, these data are acquired on a monthly basis RNC [2012-current], and thus provide a robust and timely source of information about the dynamic change of canine *B. burgdorferi* seroprevalence across the contiguous US, and holds promise for longitudinal studies to better understand the dynamic nature of vector-borne disease over time. There is also evidence that other tick vectors are involved in sustaining the enzootic cycle of *B. burgdorferi ss*, such as *Ixodes affinis*, Maggi et al. [2010] and so it is important when interpreting these results to consider the possibility of other vectors of *B. burgdorferi*. However, *I. affinis* is believed to be uncommonly found on dogs and does not feed on humans so its impact is considered to be minimal.

Though the proposed technique could be used to construct long-term forecasts, caution should be taken. In particular, our approach makes use of forecasted values of the significant factors, with some factors being assumed to be static throughout time (e.g., forestation and surface water coverage). This assumption is reasonable in the short-term, but would obviously be problematic over a much larger time span; e.g., twenty to fifty years. Moreover, in general, when forecasting future trends one should be cautious of long-term forecasts, due to possible violations of assumed model forms not apparent in the available data; e.g., changes in population density that may be spatially variable throughout time. Thus, we promote the use of our approach to provide only short-term forecasts of spatial trends in *B. burgdorferi* seroprevalence.

Similar to domestic dogs, spatial risk models for human Lyme disease are needed to address the rise in human Lyme disease incidence. Ideally, spatial risk models for human Lyme disease would be based on vector abundance inclusive of pathogen burden Eisen and Eisen [2008], Daniels et al. [1998]. In comparison to other vector-borne diseases, Lyme disease risk assessment is facilitated by the observation that *B. burgdorferi* prevalence in *Ixodes* spp. from Lyme disease endemic regions is relatively high Eisen and Eisen [2008], Scott et al. [2016], Hutchinson et al. [2015]. Canine *B.*

burgdorferi seroprevalence has been cited as one potential tool for Lyme disease risk assessment in humans Mead et al. [2011a], Millen et al. [2013a]. As noted by Mead et al. Mead et al. [2011a], canine *B. burgdorferi* seroprevalence greater than 5% at the county level was associated with human risk of Lyme disease Mead et al. [2011a], while less than 1% canine seroprevalence was associated with little to no risk for human cases of Lyme disease Millen et al. [2013a]. We further this suggestion that the use of canine *B. burgdorferi* seropositivity has merit as a risk assessment tool in both endemic and non-endemic regions. In particular, areas where *B. burgdorferi* seroprevalence is greater than 1%, but increases over time, may be an area to focus tick-prevention messages as these areas may be at an increased risk for human Lyme disease. As such, we believe that with further research into the relationship between human and canine Lyme disease, canine *B. burgdorferi* seropositivity has the potential to function as an early warning system for geographic expansion and emerging infection risk in humans Little et al. [2010b], and support targeted vector surveillance efforts to monitor *Ixodes* spp. *B. burgdorferi* infection in a cost-effective manner. As the increasing incidence of Lyme disease continues to put pressure on the United States healthcare system Adrion et al. [2015], the use of canine *B. burgdorferi* seroprevalence to ultimately forecast spatial and temporal patterns of risk of human Lyme disease is a promising tool for targeting public health educational campaigns and resources for vector surveillance to best protect humans and veterinary species from disease.

Chapter 3

A Bayesian Spatio-Temporal Model for Forecasting *Anaplasma* Species Seroprevalence in Domestic Dogs within the Contiguous United States

3.1 Introduction

Canine anaplasmosis is caused by gram-negative intracellular bacteria of the family Anaplasmataceae within the order Rickettsiales Dumler et al. [2001]. *Anaplasma* spp. bacteria are transmitted through the bite of infected ticks, with different tick species transmitting distinct types of *Anaplasma* bacteria in different regions of the country. *A. phagocytophilum* is transmitted by *Ixodes scapularis* and *I. pacificus* and maintained in a vectorreservoir-host system similar to that of *Borrelia burgdorferi* (the causative agent of Lyme disease), with the highest canine *A. phagocytophilum* seroprevalence reported in the Northeast, upper Midwest and along the west coast of California Qurollo et al. [2014]. Although an important canine pathogen, *A. phagocytophilum* is also zoonotic and causes

human disease in the same regions where Lyme disease occurs. In contrast, *A. platys* is presumed to be transmitted by *Rhipicephalus sanguineus*, and has relatively low prevalence across the contiguous United States with a slightly higher prevalence seen in the southern states Qurollo et al. [2014]. Dogs in the southern U.S. (Florida, Georgia, North and South Carolina, Tennessee and Texas) show equivalent seroconversion to both *A phagocytophilum* and *A. platys* Qurollo et al. [2014], suggesting exposure to multiple tick vectors. Veterinary wellness exams commonly include annual screening for exposure to *Anaplasma* spp., as well as *Ehrlichia* spp., *Borrelia burdgorferi* and *Dirofilaria immitis* (heartworm disease agent) using a rapid, in-house enzyme-linked immunosorbent assay (ELISA) platform (SNAP®4Dx® and SNAP®4Dx® Plus Test, IDEXX Laboratories, Inc., Westbrook, ME, USA) Drouin et al. [2004], O'Connor et al. [2006]. These tests detect antibodies to both *A. phagocytophilum* and *A. platys* on a single spot and therefore no in-house speciation is possible. Of four million dogs tested for exposure to *Anaplasma* spp. in 2015, over 100,000 dogs were seropositive. Seroreactivity on these tests are interpreted by veterinary clinicians to indicate tick exposure and a history of transmission of *Anaplasma* spp.

Many, if not most, dogs remain asymptomatic following exposure to *Anaplasma* spp.. For example, in areas such as the northeastern US where disease is endemic, as many as 60% of dogs may have antibodies specific for *Anaplasma* spp. and the majority of these dogs do not have overt evidence of clinical disease Beall et al. [2008], Bowman et al. [2009]. When symptomatic, dogs infected with *A. phagocytophilum* most commonly present with lethargy, fever and anorexia Little [2010]. Thromobocytopenia is a hallmark of symptomatic *A. phagocytophilum* and *A. platys* infection Greig et al. [1996], Kohn et al. [2008], Poitout et al. [2005], presumably because of platelet destruction Lilliehook et al. [1998]. Both splenomegaly and lymphadenopathy are reported and are thought to be associated with reactive lymphoid hyperplasia Lilliehook et al. [1998]; however it is critical to note that this is not specific to anaplasmosis as experimental infection of dogs with *Ehrlicia canis, E. chaffeensis, A. platys* and *A. phagocytophilum* results in similar histopathological lesions in lung, liver and spleen Nair et al. [2016]. Finally, lameness due to neutrophilic polyarthritis, vomiting, diarrhea, neurologic abnormalities and epistaxis have been described Greig et al. [1996], Kohn et al. [2008], Poitout et al. [2005]. Because the majority of seropositive dogs are asymptomatic, current recommendations for veterinary care in *Anaplasma* spp. seroreactive dogs include a complete blood count with platelet count to determine if treatment is necessary Little [2010].

Illnesses caused by tick-borne pathogens in animals and humans have increased over the

last decade Adams et al. [2015], due in part to the geographic expansion of tick populations beyond previously recognized endemic zones Eisen et al. [2016a]. In a previous study, we evaluated potential explanatory factors for *Anaplasma* seroreactivity in dogs and and found seroprevalence increases with increasing precipitation and forestation coverage and decreases with increasing temperature, population density, relative humidity, and elevation McMahan et al. [2016]. Also, socioeconomic status and deer/vehicle collisions were positively and negatively correlated with canine *Anaplasma* spp. seroprevalence, respectively McMahan et al. [2016]. Given that many *Anaplasma* spp. infections are asymptomatic or mild, the relative distribution of disease risk is likely under-appreciated, particularly in non-endemic zones. As such, it would be advantageous to accurately forecast *Anaplasma* spp. seroprevalence on a local scale, providing an *a priori* alert to veterinarians in emerging areas of disease. Annual forecasts of emergent infection can inform veterinary and public health officials to shifting areas of infection, particularly in temperate regions of the US where *Anaplasma* spp. seroprevalence is generally absent, rare, or prevalence is highly influenced by annual variation in biotic or abiotic factors.

In the current study, we utilize the explanatory variables derived from our previous study to develop an *Anaplasma* spp. seroprevalence forecast model that uses Bayesian methods to account for the strong spatial and temporal dependencies that exist in *Anaplasma* spp. seroprevalence between counties and across time. Controlling for this autocorrelation, e.g. the similarity between observations as a function of time, provides us with estimates conducive to forecasting changes in the prevalence of exposure to *Anaplasma* spp. in space and time. The model described herein considers the climate, geographical, and societal factors included in the previous model, with the exception of deer/vehicle collisions, as potential predictors. We report on the fidelity of the forecast model by analyzing the relationship between predicted and actual *Anaplasma* spp. prevalence in 2015, and forecast the prevalence for 2016. Finally, we discuss the potential of canine *Anaplasma* seroprevalence to inform human practitioners, as human anaplasmosis is growing in recognition as a significant problem when present as a co-infection with *Borrelia burgdorferi* and is implicated in complicated, protracted Lyme disease Caulfield and Pritt [2015].

3.2 Materials and Methods

3.2.1 The data and baseline map construction

The data analyzed in this article consists of 11,437,537 *Anaplasma* spp. serology test results for dogs (obtained from IDEXX Laboratories, Inc. RNI). The tests were conducted in the contiguous United States from 2011 to 2015 and 3.21% were positive (i.e., 367,663 positive tests) for antibodies to *Anaplasma* spp. Along with the binary outcomes (positive or negative) the data also provides the county in which the test was conducted, and subsequently the analysis presented herein considers the county level aggregated totals; i.e., the number of positive and negative test results observed within each of the contiguous United States counties during the years of 2011 to 2015. As mentioned above, the tests are reported in the county in which the test was performed. No information is given on the travel or residence of the individual animal. In some cases, the county of testing may not be the county of exposure. There is also no testing history of the dogs or the reason why the tests were conducted, so repeat testing within the year may have occurred.

The explanatory factors considered here are those believed to be related to *Anaplasma* spp. seroprevalence in dogs, for which up to date data are available on a wide geographic scale. Table 3.1 presents the factors considered in our statistical analysis; i.e., climatic variables (annual temperature, precipitation, and relative humidity), geographic variables (county elevation, forestation coverage, and surface water coverage), and socio-economic variables (population density and median household income). For further details about these factors, such as their geographical distributions, please see McMahan et al. [2016]. Fine scale data on tick population levels is desirable for this analysis, but, to our knowledge, does not exist.

Figure 3.1 provides a spatial depiction of the empirical *Anaplasma* spp. seroprevalence, after aggregating over the five years of available data. Here the empirical prevalence for each county is defined to be the number of positive tests divided by the total number of tests conducted. This figure tends to suggest a large degree of spatial correlation; i.e., empirical prevalences from counties close to each other tend to be similar. Moreover, within a given county the empirical prevalences also possess this property across time. These observations lead to the belief that both positive spatial and temporal correlation exists in our data. Thus, to offer a reliable evaluation of the putative factors, as well as to develop a predictive model, these effects must be taken into account. In the next section, a Bayesian regression model which acknowledges and accounts for both the spatial and

Factor	Data period	Scale	Notation	Range
Annual temperature (F)	1895 - 2015	CD	$X_{s,1}(t)$	[34.59, 77.14]
Annual precipitation (in)	1895 - 2015	CD	$X_{s,2}(t)$	[0.30, 10.73]
Annual relative humidity (%)	2006 - 2015	CD	$X_{s,3}(t)$	[17.98, 88.73]
Elevation (ft)	2012	C	$X_{s,4}(t)$	[10, 14495]
Perc. forest coverage (%)	2012	C	$X_{s,5}(t)$	[0.00, 32]
Perc. surface water coverage (%)	2010	C	$X_{s,6}(t)$	[0.00, 91]
Population density (ppsm)	2011-2014	C	$X_{s,7}(t)$	[0.10, 36041.11]
Median household income ($)	1997-2014	C	$X_{s,8}(t)$	[20990, 125635]

Table 3.1: Factors purported to be associated with *Anaplasma* spp. seroprevalence.

For further discussion, including the source of each factor, see Wang et al. [2014a]. Note the following abbreviations are used: persons per square mile (ppsm), climate division (CD), county (C).

temporal correlation is presented. It is worthwhile to note that some counties report relatively small number of test results, thus the interpretation of the empirical estimates depicted in Figure 3.1 may be slightly misleading, if one does not consider this effect. That is, a county reporting only 20 test results, with 1 being positive, results in an empirical estimate of 5%, whereas the same empirical estimate would be obtained for a county reporting 2000 test results, with 100 being positive. The salient point, more faith should be placed in the latter estimate, when compared to the former, since it is derived from a higher sample size.

In order to create a "baseline" map for *Anaplasma* spp. seroprevalence, the empirical estimates depicted in Figure 3.1 were further processed through a series of spatial smoothing techniques. First, a weighted head-banging algorithm was applied to the empirical prevalences. This procedure accomplishes two primary tasks; first, it acknowledges spatial correlation by forming a spatially oriented prevalence estimate through incorporating information from surrounding areas, and second, it down weights the influence of prevalence estimates which are derived from a relatively small number of tests. For example, in Colorado during 2015, 14,908 test results were reported, of which 112 were positive, translating to an observed prevalence of approx 0.75% for the state. Further, in Gunnison county, Colorado, only 3 test results were reported, of which 1 was positive, translating to an observed prevalence of 33%, which is obviously a small sample size issue. Weighted head banging directly acknowledges the available sample size from each county when computing the spatially oriented prevalence estimate. Weighted head banging was implemented using thirteen triples and the county level weights were chosen to be proportional to the number of county level observations

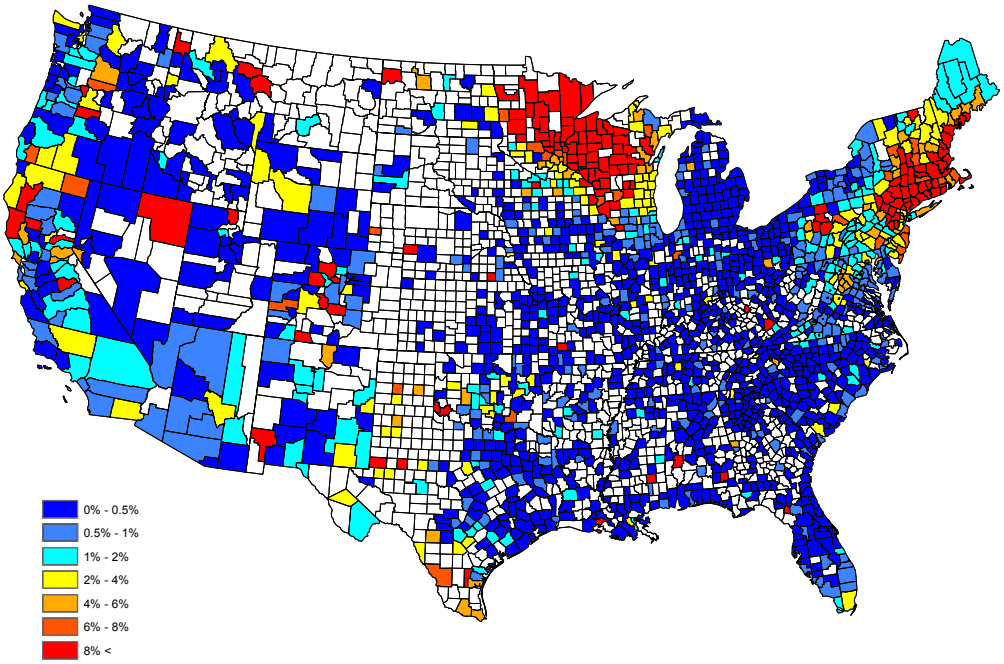


Figure 3.1: Empirical county-by-county *Anaplasma* spp. seroprevalence aggregated over 2011-2015.

which were reported over the five-year period; i.e., counties reporting more data were given more weight. In order to render a spatially complete map, Kriging (a common spatial interpolation technique) was used to interpolate prevalence estimates for counties not reporting data. This technique was implemented in ArcGIS using the default settings. Figure 3.2 provides our "baseline" map for *Anaplasma* spp. seroprevalence in domestic dogs within the contiguous United States. The "baseline" map presented in Figure 3.2 suggests two endemic zones of canine *Anaplasma* seroprevalence in the North-Central and Northeastern US, with lower but apparent seroprevalence in the Mid-Atlantic region, the Dakotas, and Western Texas.

3.2.2 Model

This section outlines the model used to perform the spatio-temporal statistical analysis. As was previously alluded to, the available data exhibits both positive spatial and temporal correlation, and are measured on areal units (i.e., counties). Thus, to analyze these data, we employ a Bayesian hierarchical spatio-temporal regression model, where autocorrelated random effects are utilized to account for the spatial and temporal dependence. The use of Bayesian models for such analyses is

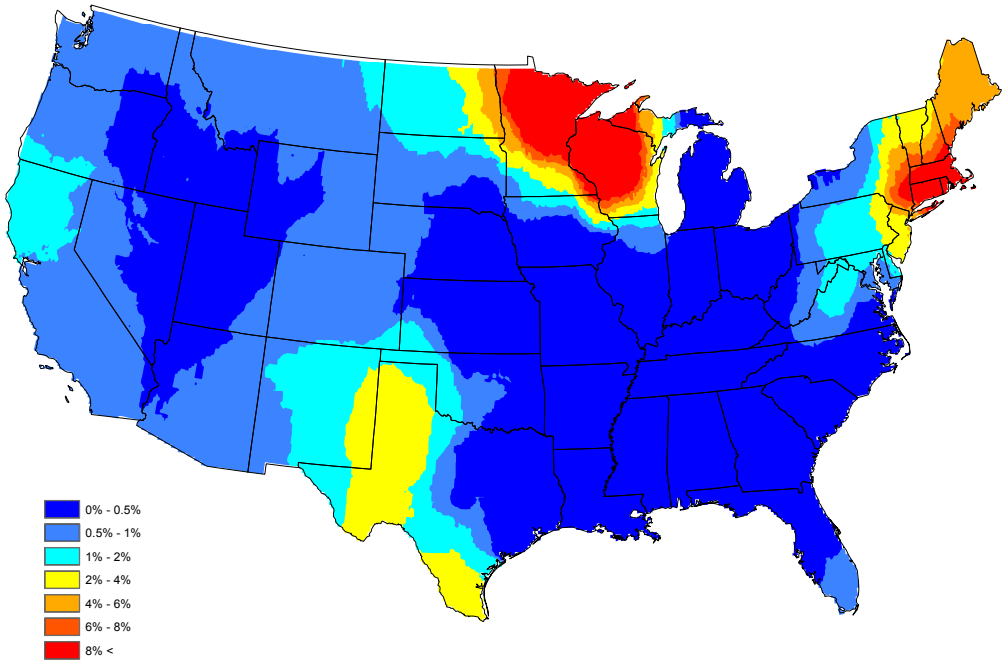


Figure 3.2: Baseline map of *Anaplasma* spp. seroprevalence.

omnipresent in similar application areas; for further discussion and a modern review see Cressie and Wikle [2011b], Lopez-Quilez and Munoz [2009].

For modeling purposes, the number of positive tests and total tests for county s at year t are denoted by $Y_s(t)$ and $n_s(t)$, respectively, where $s \in \{1, \ldots, S\}$ and $t \in \{1, \ldots, T\}$. Following the proposals of Lopez-Quilez and Munoz [2009], Nobre et al. [2005], Xia and Carlin [1998], Martinez-Beneito et al. [2008], we relate the observed testing data to the factor information through the use of a Poisson regression model. Thus, the first level of our Bayesian hierarchical Poisson regression model is given by:

$$Y_s(t)|n_s(t), p_s(t) \sim \text{Poisson}\left\{n_s(t)p_s(t)\right\}, \tag{3.1}$$

$$\log\left\{p_s(t)\right\} = \beta_0 + \sum_{k=1}^{8} \beta_k X_{s,k}(t) + \xi_s(t). \tag{3.2}$$

In Eq (3.1), $p_s(t)$ denotes the *Anaplasma* spp. seroprevalence of county s at time t, $\sim$ means has the distributional type and the symbol $|$ indicates given quantity(ies). Proceeding with this notation, Eq (3.1) indicates that when $n_s(t)$ and $p_s(t)$ are known, the number of positive tests, the $Y_s(t)$, are

conditionally independent across all the counties and they each follow a Poisson distribution with mean $n_s(t)p_s(t)$. In Eq (3.2), $\log(\cdot)$ denotes the natural logarithm, $X_{s,k}(t)$ is the kth factor measured on county s at time t (see Table 3.1), the β_k are regression coefficients, and $\xi_s(t)$ are spatio-temporal random effects used to account for the spatial and temporal dependence.

In order to capture spatial dependence, the proposed approach makes use of a conditional autoregressive (CAR) model. The CAR model was first proposed by Besag [1974b], and has since seen numerous modifications and adaptations; e.g., see Banerjee et al. [2014b]. The version used here, which was adopted from Stern and Cressie [1999], is now described. Let $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_S)'$ denote a random vector which follows a CAR model. Under our specification, one has that

$$\boldsymbol{\phi} \sim \mathrm{N}(\mathbf{0}, \tau^2 \left(\mathbf{D} - \rho \mathbf{W}\right)^{-1}), \tag{3.3}$$

where $\mathbf{W} = \{w_{s,s'}\}$ is a $S \times S$ neighborhood adjacency matrix and $\mathbf{D}$ is a $S \times S$ diagonal matrix whose sth diagonal element denotes the number of counties bordering the sth county. The neighborhood adjacency matrix is constructed such that $w_{s,s'} = 1$ if the sth and s'th counties border one another, and $w_{s,s'} = 0$ otherwise. For notational convenience, we denote the relationship depicted in Eq (3.3) as $\boldsymbol{\phi} \sim \mathrm{CAR}(\tau^2, \rho)$. Under this specification, it can be shown that the conditional distribution of ϕ_s is given by

$$\phi_s \mid \boldsymbol{\phi}_{-s}, \tau^2, \rho, \mathbf{W} \sim \mathrm{N}\left(\rho \frac{\sum_{s' \neq s} w_{s,s'} \phi_{s'}}{\sum_{s' \neq s} w_{s,s'}}, \quad \frac{\tau^2}{\sum_{s' \neq s} w_{s,s'}}\right), \text{ for } s = 1, \ldots, S, \tag{3.4}$$

where $\boldsymbol{\phi}_{-s} = (\phi_1, \ldots, \phi_{s-1}, \phi_{s+1}, \ldots, \phi_S)'$ is a vector of random effects for all counties except the sth one. In Eq (3.4), the conditional expectation of ϕ_s, given its neighbors' random effects, is the scaled (by $\rho \in [0, 1]$) average of the random effects of its neighbors. Therefore, ρ controls the spatial autocorrelation between bordering counties, with $\rho = 0$ indicating spatial independence and ρ close to 1 indicating strong spatial dependence. The parameter τ^2 is a scaling variance parameter, and the conditional variance of ϕ_s is inversely proportional to the number of neighboring counties. Consequently, the conditional variance of ϕ_s is smaller (larger) than the conditional variance of $\phi_{s'}$ if the sth county has more (fewer) neighbors than the s'th county. This is reasonable since more neighbors relates to more information, and hence more certainty (i.e., a smaller variance).

Motivated by Nobre et al. [2005], our proposed model captures the spatial and temporal

dependence through a multivariate autoregressive model of order one which is given by

$$\begin{aligned}
\boldsymbol{\xi}_1 &= \boldsymbol{\phi}_1; \\
\boldsymbol{\xi}_t &= \varphi \boldsymbol{\xi}_{t-1} + \boldsymbol{\phi}_t, \text{ for } t = 2, \ldots, T; \\
\boldsymbol{\phi}_t &\sim \text{CAR}(\tau^2, \rho), \text{ for } t = 1, \ldots, T,
\end{aligned} \tag{3.5}$$

where, $\boldsymbol{\xi}_t = (\xi_1(t), \ldots, \xi_S(t))'$ and $\boldsymbol{\phi}_t = (\phi_1(t), \ldots, \phi_S(t))'$ for $t = 1, \ldots, T$. Note, in our proposed model the spatial random effects $\boldsymbol{\phi}_t$ are independent and identically distributed according to the CAR model described in Eq (3.3). The parameter φ controls the temporal correlation between consecutive years and lies within $(-1, 1)$, this ensures a causal and stationary solution to the time series model (see Brockwell and Davis [2006]). Moreover, in the proposed model $\varphi = 0$ indicates temporal independence, while φ close to 1 indicates strong positive temporal correlation.

To complete the Bayesian model, the following prior distributions were specified:

$$\beta_k \sim \text{N}(0, 1000), \text{ for } k = 0, \ldots, 8; \tag{3.6}$$

$$\varphi \sim \text{Uniform}(-1, 1); \tag{3.7}$$

$$\rho \sim \text{Uniform}(0, 1); \tag{3.8}$$

$$\tau^{-2} \sim \text{Gamma}(0.5, 0.05). \tag{3.9}$$

Note, the prior distributions placed on the β_k are independent and weakly informative so that estimation and inference for these parameters are based mainly on the data. Priors for φ and ρ are uninformative, and were chosen for the same reason. For ease of computation, the prior for τ^{-2} was chosen due to conjugacy (the posterior and prior distributions are from the same distributional family). A Markov chain Monte Carlo (MCMC) posterior sampling algorithm for the model parameters and random effects was constructed, which made use of a combination of Gibbs and Metropolis-Hastings sampling steps Gelman et al. [2014]. In the algorithm, the response variable (i.e., $Y_s(t)$) associated with counties not reporting data were treated as missing data and were subsequently sampled along with the other model parameters. The MCMC sampling algorithm was written and implemented in **C++** and **R**.

3.3 Results

3.3.1 Model Assessment

A primary goal of this work was to evaluate the putative factors (see Table 3.1) with respect to their association with *Anaplasma* spp. seroprevalence within domestic canines in the contiguous United States. To accomplish this task, a full model with all eight factors was fitted to the data. From this fit, point estimates (means of posterior samples) and 95% highest posterior density (HPD) credible intervals were obtained for all regression coefficients; for further discussion and details about HPD intervals see Gelman et al. [2014]. Table 3.2 provides the point estimates and HPD intervals for each of the eight regression coefficients.

Factor	Estimate	95% HPD Interval
Annual temperature (F)	-0.021	[-0.036, -0.008]
Annual precipitation (in)	-0.004	[-0.053, 0.037]
Annual relative humidity (%)	-0.001	[-0.008, 0.004]
Elevation (ft)	0.032	[0.002, 0.061]
Percentage forest coverage (%)	3.039	[1.914, 4.045]
Percentage surface water coverage (%)	0.398	[0.130, 0.692]
Population density (ppsm)	-2.765e-5	[-4.473e-5, -0.976e-5]
Median household income ($)	0.002	[-0.001, 0.005]

Table 3.2: Parameter estimates from the full model.

From Table 3.2, one can ascertain that 3 of the considered factors were insignificant in the full model at the considered significance level; i.e., the 95% HPD intervals for the regression coefficients associated with annual precipitation, annual relative humidity, and median household income contain zero. In order to develop a parsimonious model that only contains significant factors, 7 reduced models were fit to the data, each containing a different combination of the 3 questionable explanatory factors. In each of these additional model fits, these 3 factors were again deemed to be insignificant. Therefore, our selected model makes use of 5 explanatory factors; i.e., annual temperature, percentage forest coverage, percentage surface water coverage, elevation, and population density. Table 3.3 provides the point estimates of the regression coefficients along with 95% HPD intervals for the factors belonging to the selected model. In addition, the point estimates of the other model parameters are $\varphi = 0.935$, $\rho = 0.999$, and $\tau^2 = 0.639$, which reaffirms the assertion that strong positive spatial and temporal correlation are present in the data.

In order to assess the adequacy of our selected model, a figure analogous to Figure 3.2 was

-2.25in0in

Table 3.3: **Parameter estimates from the selected model.**

Factor	Estimate	95% HPD Interval
Annual temperature (F)	-0.019	[-0.031, -0.004]
Percentage forest coverage (%)	2.881	[1.855, 4.065]
Percentage surface water coverage (%)	0.389	[0.112, 0.675]
Elevation (ft)	0.033	[0.005, 0.058]
Population density (ppsm)	-3.090e-5	[-4.635e-5, -1.464e-5]

created using the model based estimates of the county level prevalences. In particular, our fitted model provides a yearly prevalence estimate for each county, regardless of whether the county reports data. Thus, to obtain a single prevalence estimate for each county, we averaged over the 5 yearly estimates available from the model fit; i.e., our aggregated estimate for the sth county is given by $\widehat{p}_s = 5^{-1}\sum_{t=1}^{5}\widehat{p}_s(t)$, where $\widehat{p}_s(t)$ is the prevalence estimate resulting from the selected model for the sth county during the tth year. Figure 3.3 provides a depiction of these results, after Kriging (again default settings were used in ArcGIS). By comparing these two figures one will note that the proposed Bayesian spatio-temporal model appears to provide a good fit to these data. In fact, the correlation between the sets of estimates depicted in Figure 3.2 and Figure 3.3 is 0.902, thus confirming that a strong agreement exists between the empirical and model based estimates. Note, this summary measure did not consider estimates from counties not reporting data.

3.3.2 Forecasting

In this section we provide details on how the model developed in the previous section can be utilized to construct a forecast of future trends of *Anaplasma* spp. seroprevalence across the contiguous United States. Essentially, our strategy relies on forecasting the value of the 5 significant explanatory factors as well as the spatio-temporal random effects for the upcoming year. Since forestation, water coverage, and elevation are relatively stable over time, the forecasted values of these factors are simply taken to be their most recent observations. In order to forecast annual temperature and population density, we make use of the same techniques outlined in Bowman et al. [2016a].

The spatio-temporal random effects for the next year were predicted using the relationship depicted in Eq (3.5). In particular, for each value of τ^2, ρ, φ and $\boldsymbol{\xi}_t$ which are available from the posterior sample, we obtain a predicted value of the spatio-temporal random effect for the next year as $\boldsymbol{\xi}_{t+1} = \varphi\boldsymbol{\xi}_t + \boldsymbol{\phi}_{t+1}$, where $\boldsymbol{\phi}_{t+1}$ was randomly generated from a $\mathrm{N}(\mathbf{0}, \tau^2(\mathbf{D} - \rho\mathbf{W})^{-1})$. This

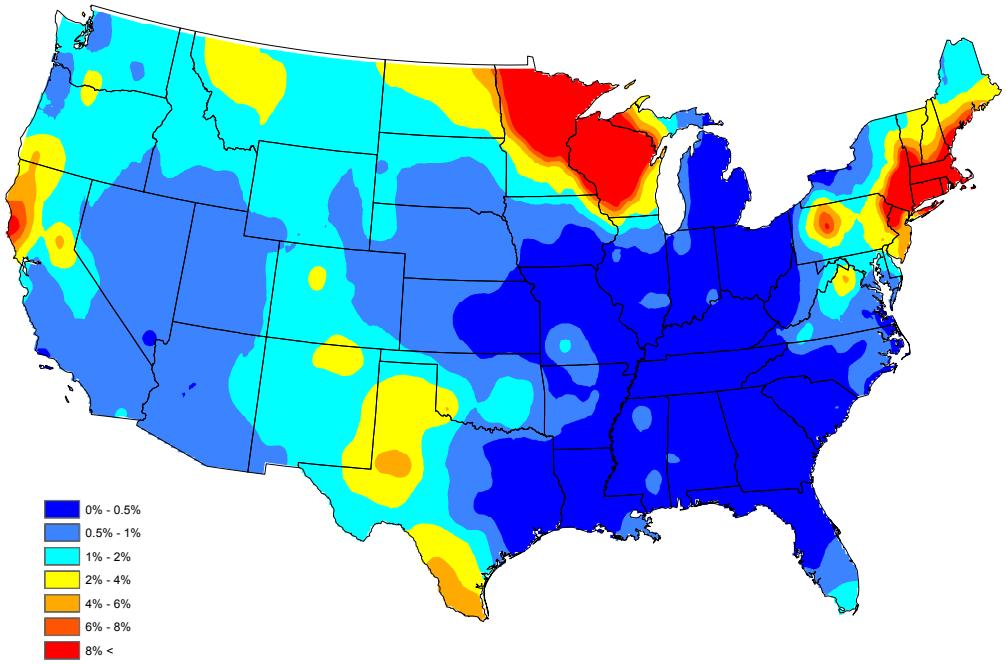


Figure 3.3: Aggregated model-based estimates of *Anaplasma* spp. seroprevalence.

process generates posterior predictive samples (see Gelman et al. [2014] for additional details) of $\boldsymbol{\xi}_{t+1}$, which can then be used to forecast next year's *Anaplasma* spp. seroprevalence.

In order to assess the fidelity of the proposed forecasting procedure, the 2015 test and factor data were removed, and our approach was used to forecast the 2015 county level prevalences using the 2011-2014 test and factor data only. Figure 3.4 and Figure 3.5 present the forecasted and observed prevalences during 2015, respectively. The weighted correlation between the forecasted and observed prevalences (for counties reporting data in 2015), is 0.987, demonstrating that *Anaplasma* spp. seroprevalence can be accurately forecasted through the proposed approach. Here, a weighted correlation, with weights being set to be equal to $n_s(t)$, is used to account for different sample sizes within each county, for further discussion see Bowman et al. [2016a]. Figure 3.6 provides our 2016 forecast of canine *Anaplasma* spp. seroprevalence within the contiguous United States.

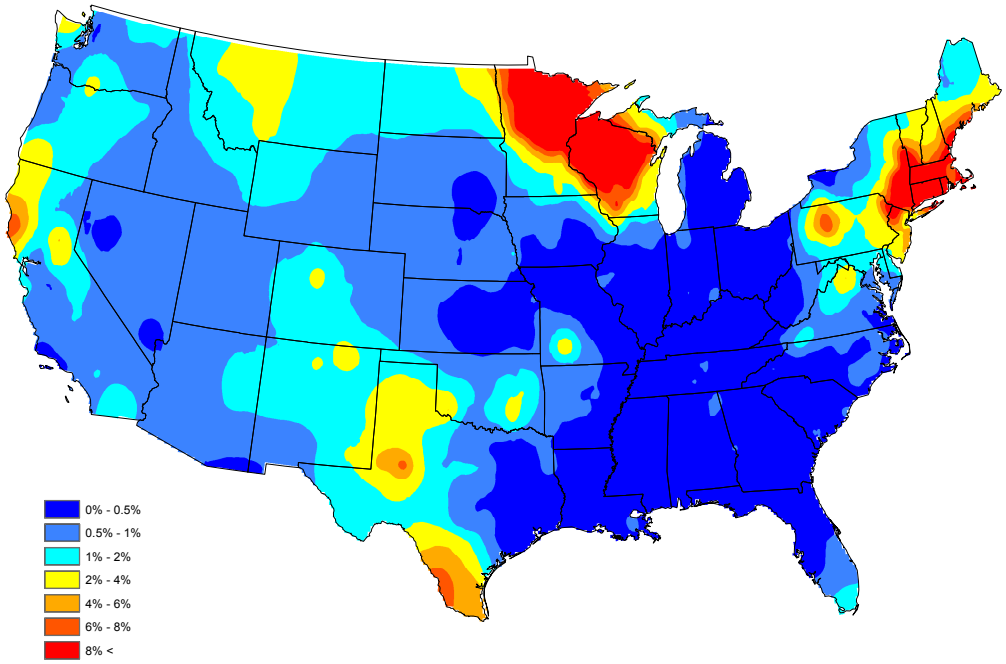


Figure 3.4: 2015 forecasted *Anaplasma* spp. seroprevalence.

3.4 Conclusion and Discussion

In this study, we present for the first time a fully Bayesian approach to forecasting canine *Anaplasma* spp. seroprevalence in the absence of detailed information on the distribution and abundance of the two primary vectors of disease: *Ixodes* spp, and *Rhipicephalus sanguineus* and pathogen prevalence in vectors. Surveillance of these medically important vectors remains of high importance; however, annual analysis of vector distribution and molecular characterization of pathogen burden will remain economically and logistically unfeasible in many areas in the face of changing climate and habitats. As such, we have developed a model for forecasting spatial and temporal patterns of risk of exposure to *Anaplasma* spp. based on canine seroprevalence data. Data in our study are limited by lack of detailed *Anaplasma* species specificity; however, from the recent species-level serologic analysis conducted by Qurollo et al., we can infer that the majority of observed *Anaplasma* spp. seroprevalence in the Northeast, upper Midwest and west coast of California is likely the result of exposure to *A. phagocytophilum*, and seroprevalence in the southern and western US due to exposure to either *A. platys* or *A phagocytophilum* Qurollo et al. [2014]. In California, there is a higher liklihood

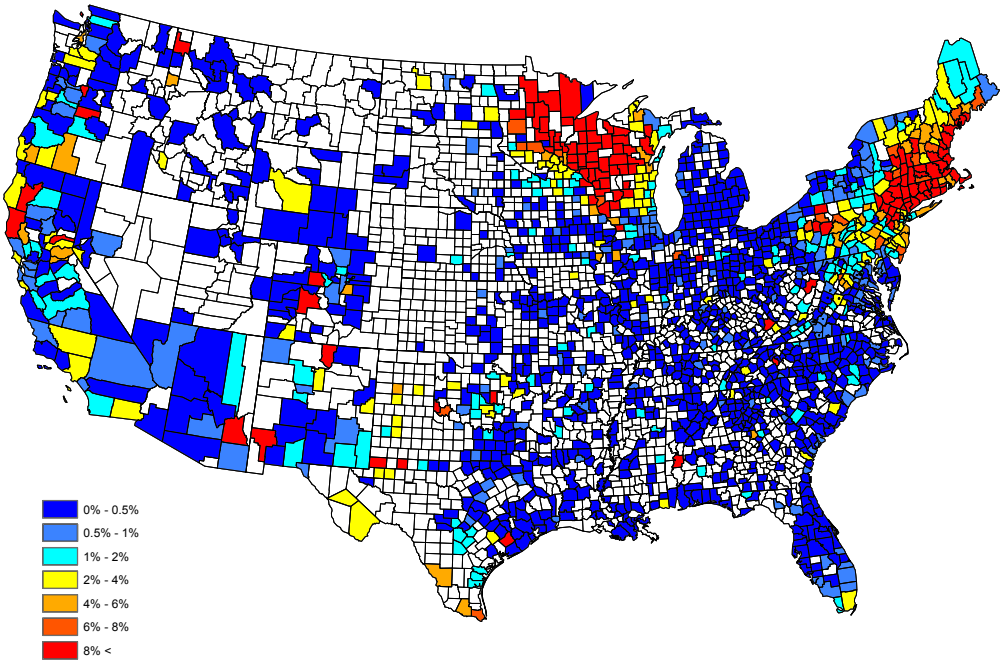


Figure 3.5: 2015 observed *Anaplasma* spp. seroprevalence.

of the antibodies being due to *A phagocytophilum* exposure whereas positives in Texas and other southeastern states is due to *A. platys* exposure.

A major strength of our surveillance effort is that *Anaplasma* spp. seroprevalence data are acquired on a monthly basis RNC [2012-current], providing a robust and timely source of information about the dynamic change of *Anaplasma* spp. exposure across the contiguous US. These data hold strong promise for longitudinal studies to best understand the dynamic nature of *Anaplasma* spp. prevalence over time. For veterinary healthcare practitioners specifically, this information is of critical importance as disease is often mild or unapparent in the dog, requiring both a working knowledge of risk paired with an impetus to pursue further diagnostics to determine when treatment is necessary for the health of the dog. Similarly, a deeper understanding of risk enhances educational opportunities, with the potential to prevent the indiscriminate use of antibiotics Lathers [2001] in asymptomatic dogs without thrombocytopenia or other significant blood count changes. Finally, even if dogs with *Anaplasma* spp. are asympotomatic, this exposure indicates tick infestation which may have or will lead to infection with other tick borne pathogens that may need to be considered.

The forecast model uses 5 years of historical data acquired from IDEXX Laboratories, Inc

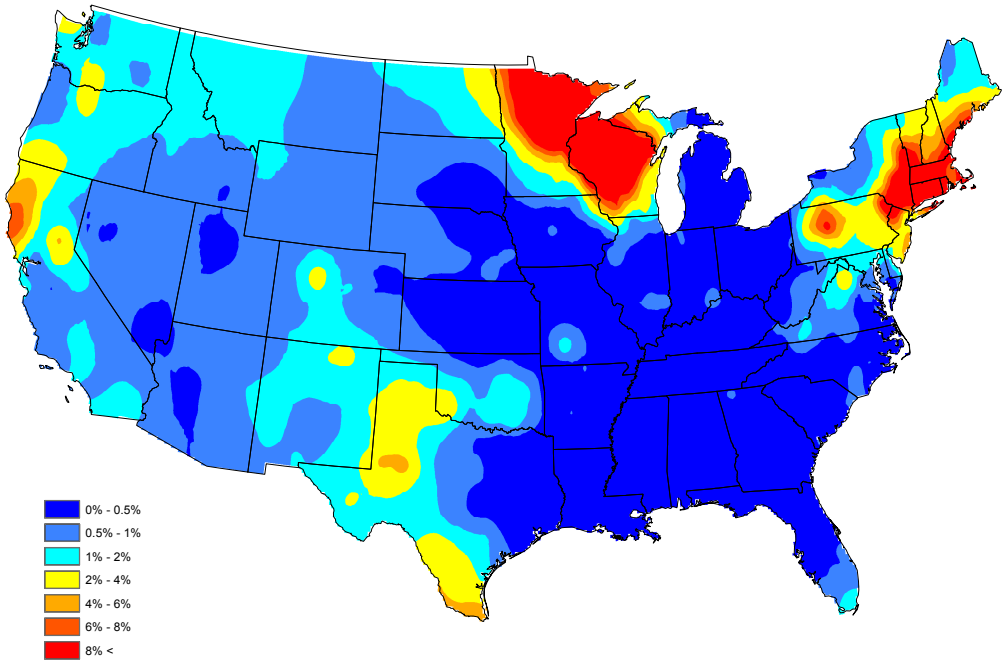


Figure 3.6: 2016 forecasted *Anaplasma* spp. seroprevalence.

and previously published relevant covariate factors McMahan et al. [2016] to help explain variability in our aggregated dataset and to strengthen inferences from our Bayesian spatio-temporal model. As previously described in our explanatory model of *Anaplasma* spp. seroprevalence, the prevalence of seropositivity increases with increasing precipitation and forestation coverage and decreases with increasing temperature, population density, relative humidity, and elevation. Socioeconomic status and deer/vehicle collisions were positively and negatively correlated with canine *Anaplasma* seroprevalence, respectively McMahan et al. [2016].The potential drivers of infection in the current model vary some from the previously published model because of the differences in statistical methods used. For forecasting, a Bayesian model was chosen because of the ability to control for confounding caused by spatial and temporal autocorrelation seen in disease prevalence data. The associations of annual temperature, forest coverage, and population density remain similar between our previous explanatory model and our forecast model. In contrast, percentage of water coverage was significant in a Bayesian model, but not in the explanatory model McMahan et al. [2016]. Elevation was found to have a positive association in a Bayesian forecast model, whereas the association was negative in the explanatory model. This can be explained in part by the distribution of the

two *Anaplasma* spp. of dogs. *Anaplasma phagocytophilum* is the primary cause of seropositivity in dogs in many regions of the United States, and this pathogen is found in high prevalences in the northeastern United States, a region that is higher in elevation and lower temperature, compared to the non-endemic regions of the South and southeastern United States. We acknowledge a limitation in our seroprevalence dataset that it is a "presenceabsence" model presented at a crude spatial scale, and as such covariates such as elevation are perceived as uniform over large geographical areas. By extending our model to include climate variables over a similarly large spatial area it was revealed that the association at the regional level is related to the climatic conditions. Despite the aformentioned limitations, using data from 2011-2014 we evaluated observed versus forecasted *Anaplasma* spp. seroprevalence in 2015, resulting in a weighted correlation between the two maps of 0.987. Given the striking fidelity of our forecast model, we report our 2016 forecast for canine *Anaplasma* seroprevalence in the contiguous United States.

Using our described methodology, canine *Anaplasma* spp. seroprevalence is forecasted to be high in regions of the US endemic for *Ixodes scapularis*: the Mid-central and Northeastern states of Connecticut, Delaware, Maine, Massachusetts, Maryland, Minnesota, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont and Wisconsin. Six of these states: New York, Connecticut, New Jersey, Rhode Island, Minnesota, and Wisconsin, account for 90% of all reported cases of human anaplasmosis due to infection with *A. phagocytophilum* Dahlgren et al. [2015]. Beyond these endemic boundries, we observed moderate to high frequency of canine *Anaplasma* spp. seroprevalence in Northern California, North Dakota and Texas. Our forecast of canine *Anaplasma* spp. seroprevalence in Texas is notable given a recent study by Movilla et al. who documented, for the first time, canine seroprevalence of *Anaplasma* spp. in several states of Mexico Movilla et al. [2016]. The authors reported the highest seroprevalence in northwestern Mexico (16.4%), with the lowest in the north-central states of the country (0.6%). Wildlife species in Texas have been documented to harbor *A. phagocytophilum* as well Yabsley et al. [2006]. It is well established that incursions by humans into natural habitats make the boundary between wildlife, humans and domestic animals more permeable, and thereby make the spillover of vector-borne disease more likely. Collectively, the presence of *Anaplasma* spp. in both domestic dogs in Mexico and wildlife in the south-central US indicate a greater need for annual testing of dogs for *Anaplasma* spp. in southern US border states. Similarly, evidence of canine *Anaplasma* spp. seroprevalence in Northern California and along the Canadian border suggests annual testing for *Anaplasma* spp. in

these regions during veterinary wellness visits is strongly advised.

Human infection with *A. phagocytophilum* has been reported in non-endemic regions of the United States, including southeastern and south-central states Dahlgren et al. [2016, 2015], where high levels of *Ixodes* spp. transmitting *A. phagocytophilum* have yet to be documented. As such, there remains some controversy regarding true establishment of human anaplasmosis outside of endemic states Dahlgren et al. [2016, 2015]. The Centers for Disease Control and Prevention (CDC) suggests some of these human cases may be due to patient travel to states with higher levels of disease, or the misdiagnosis of anaplasmosis in patients actually infected with another tick-borne disease, such as ehrlichiosis or Rocky Mountain Spotted Fever. Our canine seroprevalence data indicate that *Anaplasma* spp. are being transmitted to dogs in Northern Texas and Northern California, as well as along the Canadian border in North Dakota and Montana, although the species is unknown. While canine seroprevalence for *B. burgdorferi* is an established surveillance tool for human Lyme disease Mead et al. [2011b], Millen et al. [2013b], it remains to be determined whether canine *Anaplasma* seroprevalence can provide a similar risk assessment tool for human epidemiologists. Regardless, as *A. phagocytophilum* is a potential zoonotic pathogen, it is important to be aware that these organisms are enzootic in non-endemic regions of the United States and, and in particular, in the regions bordering the south-central US. Finally, it is important to note that while canine Lyme disease caused by *Borrelia burgdorferi* and transmitted by *Ixodes* spp. is now increasing in the Great Lakes region of the US as *Ixodes* ticks converge from the Mid-Central state and Northeastern states Eisen et al. [2016a], a similar pattern of elevated *Anaplasma* spp. seroprevalence was not evident in our current dataset. It is unclear whether *Ixodes scapularis* ticks in this region do not harbor *Anaplasma* spp. or whether canine test data are currently too sparse in this region to detect notable increases in seroprevalence.

While canine anaplasmosis is most often asymptomatic and self-limiting, there is growing recognition in human and veterinary medicine that *Anaplasma* spp. may represent a significant health threat as a co-morbidity. Indeed, co-infection with *Ixodes*-borne pathogens is prevalent and increasingly problematic worldwide Diuk-Wasser et al. [2016]. In humans, most co-infections involve two of the three major human pathogens, *B. burgdorferi* sensu lato, *A. phagocytophilum*, and *Babesia* spp. Such co-infections have been documented to occur in up to 28% of *I. scapularis* ticks in Lyme disease-endemic areas in the US Swanson et al. [2006]. Co-infection with multiple tick-borne pathogens can increase Lyme disease severity and has been attributed to long-term sequela in pa-

tients. Human patients with these concurrent illnesses experienced a greater number of symptoms for a longer duration than patients with Lyme disease alone Krause et al. [2002], Steere et al. [2003b]. Co-exposure of dogs to *B.burgdorferi* and *A. phagocytophilum* has been reported, with co-exposure rates greatest in the Northeast (6.2%) and Mid-Atlantic region (1.8%) of the US Qurollo et al. [2014]. Veterinary patient outcomes in the presence of co-infection are not as thoroughly defined; however, clinically ill dogs that are seropositive for *B.burgdorferi* and *A. phagocytophilum* have been noted to be twice as likely to have lameness, joint pain, and joint effusion than dogs with single infections Beall et al. [2008], Eberts et al. [2011]. In general, veterinary clinicians are increasingly concerned co-infections complicate interpretation of clinical manifestations of disease and therefore potentially confound treatment Qurollo et al. [2014]. As such, in *Ixodes* endemic regions annual screening for multiple vector-borne pathogens is strongly warranted. Beyond annual screening, year-round use of acaracides in dogs can reduce tick infestation, thereby reducing the potential for tick-borne pathogen transmission Spencer et al. [2003], Wengenmayer et al. [2014].

There are a few limitations to this study that need to be addressed. As mentioned above, samples are submitted by veterinary clinics, indicating that this population of dogs are under the care of a veterinarian. This suggests that these data are representative of a subpopulation of dogs that are more likely to receive tick preventatives and thus more protected from exposure to ticks. Therefore, prevalence estimates presented here are likely to be conservative. A second limitation to note is the county of testing. For a small sample of dogs, the county of testing may not reflect the county of residence, and may also misclassify the county of exposure for dogs exposed while traveling. Unfortunately, counties of residence and travel history are not known for these data, so interpretations are made with these limitations in mind.

In summary, we have forecasted the distribution of the canine seroprevalence at the national level with available canine serology data and easily accessible climate, geographical, and societal factors. The high degree of fidelity between actual vs. forecasted seroprevalence in this model and ease of data input provides veterinary and public health officials with an invaluable tool for informing emerging risk of exposure to *Anaplasma* spp. disease potential, and importantly the possibility of risk of co-infection with other tick-borne disease. Information should be shared with dog owners to better facilitate appropriate preventative care of healthy animals and diagnosis and treatment of ill patients. As we continue to learn more about the association between the distributions of anaplasmosis in humans and dogs, this forecasting model may become useful for public health

practitioners as well as a critical source of information on the ecology and changing distribution of *Anaplasma* spp.

Chapter 4

Sampling Strategies for Fast Updating of Gaussian Markov Random Fields

4.1 Introduction

Markov random fields (MRFs) are critical tools in a variety of challenging applications, including disease mapping [Waller et al., 1997a], medical imaging [Higdon, 1998, Brown et al., 2014], and gene microarray [Xiao et al., 2009, Brown et al., 2017a]. Awareness of these models was raised after the seminal work of Besag [1974a]. Since then, they have become popular for modeling temporally- or spatially-dependent areal data due to their interpretability and computational tractability afforded by the conditional independence induced by the Markov property. This property is particularly important for modern Markov chain Monte Carlo [MCMC; Gelfand and Smith, 1990] methods. Indeed, the ease with which Markov random fields can be incorporated into a Gibbs sampling algorithm [Geman and Geman, 1984] has contributed to their popularity in Bayesian statistics.

Gaussian Markov random fields [GMRFs; Rue and Held, 2005a] are simply MRFs in which the conditional distribution of each (scalar) random variable is Gaussian. In the statistics literature they are commonly referred to as condtionally autoregressive [CAR; Banerjee et al., 2015b] models. GMRFs are specified by explicitly defining the precision (inverse covariance) matrix instead

of a covariance function as in Gaussian process modeling [Schabenberger and Gotway, 2005]. This approach may result in an improper distribution such as the intrinsic autogressive model [IAR; Besag and Kooperberg, 1995b]. Further, GMRFs do not usually yield stationary processes due to a so-called "edge effect" in which the marginal variances vary by location. Corrections can be made to yield a stationary process such as a periodic boundary assumption [Fox and Norton, 2016] or algorithmic specification of the precision matrix [Dempster, 1972]. Sometimes the effect can simply be ignored with little effect on inference [Besag and Kooperberg, 1995b]. Efforts have been made to use GMRFs to approximate Gaussian processes with specified covariance functions [e.g., Rue and Tjemland, 2002, Song et al., 2008, Lindgren et al., 2011a], but much work still remains.

Belonging to the Gaussian class of distributions, GMRFs are the most widely studied Markov random fields; see Rue and Held [2005a] for an overview of relevant work. This includes techniques for efficiently sampling from GMRFs. While both single-site and block samplers are straightforward in principle, they can be problematic when working with extremely high-dimensional data. Block sampling involves Cholesky factorizations of large precision matrices and thus carries high computational and memory costs. While a GMRF prior induces sparsity which can be exploited to economize such calculations, conditional posterior precision matrices arising in Bayesian models may depend on parameters that change in each iteration of an MCMC algorithm, which can make repeated calculations extremely time consuming. Single-site samplers work by only considering scalar random variable updates. In addition to being more loop-intensive than block samplers, single-site samplers are known to exhibit slow convergence when the variables are highly correlated [Carlin and Louis, 2009].

There is a trade-off between computational and statistical efficiency when using high-dimensional GMRFs inside an MCMC algorithm. These competing goals have led to recent innovations in alternative sampling approaches for GMRFs. Some of these approaches require considerable expertise in numerical analysis or message passing interface (MPI) protocol, but others are relatively easy to implement and hence can be quite useful for statisticians. Specifically, the recently proposed chromatic Gibbs sampler [Gonzalez et al., 2011a] is easy to implement and is competitive with or even able to dramatically improve upon other existing strategies. It allows a practitioner to parallelize sampling and to take advantage of 'vectorized' calculations in a high-level language such as `R` [R Core Team, 2016] without requiring extensive expertise in numerical analysis.

The chromatic sampler appearing in Gonzalez et al. [2011a] was motivated by and demon-

strated on binary MRFs. However, it is straightforward to carry over the same idea to the Gaussian case. In this chapter, we discuss block updating and single-site updating of GMRFs, and compare them to chromatic sampling. Rather than focusing on theoretical convergence rates or an otherwise overall "best" approach, we view these approaches through the lens of a practitioner looking for easily implemented yet efficient algorithms. To the best of our knowledge, this work is the first time chromatic Gibbs sampling has been directly compared to the standard approaches for sampling of GMRFs.

There exist fast approximation methods for estimating features of a posterior distribution without resorting to Markov chain Monte Carlo. One of the most popular of these is integrated nested Laplace approximation [INLA; Rue et al., 2009], the `R` implementation of which is the `R-INLA` package [Lindgren and Rue, 2015]. Such approximation methods are useful when certain quantities need to be estimated quickly, but they are only approximations and thus are not interchangeable with Markov chain Monte Carlo algorithms that converge to the exact target distribution. Indeed, INLA provides the most accurate approximations around the posterior median and can disagree with MCMC in tail probability approximations [Gerber and Furrer, 2015]. These disagreements are more pronounced in cases where the full conditional distribution of the random field is non-Gaussian (for which INLA uses Laplace approximations) and the GMRF is used as a proposal in a Metropolis-Hastings algorithm. Further, the `R-INLA` package is a "black box" that works well for a set of pre-defined models. For more flexibility to manipulate non-standard models, there is the need to break open the black box to more specifically customize an algorithm. In the context of GMRFs, this requires more direct interaction with the random fields, motivating this work. Efficient strategies such as those considered here are not intended to be substitutes for INLA or other approximation methods. Rather, they are complementary procedures that are useful when one is interested in direct MCMC on challenging posterior distributions.

In Section 4.2, we briefly motivate our sampling problem and review GMRFs. We then compare chromatic sampling to block updating and single-site sampling of GMRFs. In Section 4.3 we compare the performance of single-site sampling, block updating, and the chromatic approach in a study using a simple Bayesian model with spatial random effects on high-dimensional regular lattices as well as a real application involving non-Gaussian polling data. We conclude in Section 4.4 with a discussion.

4.2 MCMC Sampling for Gaussian Markov Random Fields

4.2.1 Motivation

Suppose we have observed data $\boldsymbol{\beta}y = (y_1, \ldots, y_n)^T$ arranged in an areal (lattice) structure, where each y_i is a value summarizing an area i, $i = 1, \ldots, n$. For instance, y_i might be the number of positive disease cases observed in county i in the United States, or the expression level of the ith gene along a chromosome. In the examples we consider in Section 4.3, y_i is either the observed intensity at pixel i in an image or the number of votes cast for a particular candidate in voting precinct i in the state of New York. Suppose further that the y_i's are correlated as a function of "neighborhoods" so that the distribution of y_i is determined by the nearby values.

It is often reasonable to assume that this dependence in $\boldsymbol{\beta}y$ is explained by an unobservable process $\boldsymbol{\beta}x = (x_1, \ldots, x_n)^T$, where x_i is the realization of the process at node i. A typical Bayesian analysis of this problem takes the y_i's to be *conditionally* independent given $\boldsymbol{\beta}x$; $y_i \mid \boldsymbol{\beta}x \overset{indep.}{\sim} F(\cdot \mid \boldsymbol{\beta}x)$, $i = 1, \ldots, n$. In other words, the correlation is assumed to be completely explained by $\boldsymbol{\beta}x$. For more flexibility and to more fully account for sources of uncertainty, one might assume that the distribution of $\boldsymbol{\beta}x$ is determined by a parameter vector $\boldsymbol{\beta}\theta$ (usually of much smaller dimension than $\boldsymbol{\beta}x$) which is itself assigned a hyper-prior. Thus, the Bayesian model is

$$\begin{aligned} \boldsymbol{\beta}y \mid \boldsymbol{\beta}x &\sim f(\cdot \mid \boldsymbol{\beta}x) \\ \boldsymbol{\beta}x \mid \boldsymbol{\beta}\theta &\sim \pi_x(\cdot \mid \boldsymbol{\beta}\theta) \\ \boldsymbol{\beta}\theta &\sim \pi_\theta(\cdot). \end{aligned} \tag{4.1}$$

Inference proceeds by evaluating (or estimating) characteristics of the posterior distribution, determined via Bayes rule as $\pi(\boldsymbol{\beta}x, \boldsymbol{\beta}\theta \mid \boldsymbol{\beta}y) \propto f(\boldsymbol{\beta}y \mid \boldsymbol{\beta}x)\pi_x(\boldsymbol{\beta}x \mid \boldsymbol{\beta}\theta)\pi_\theta(\boldsymbol{\beta}\theta)$. We are concerned in this work with models in which $\boldsymbol{\beta}x \mid \boldsymbol{\beta}\theta$ follows a GMRF.

In modern Bayesian analysis, it is common for the posterior to have no known closed form. Hence, expectations with respect to this distribution cannot be evaluated directly. If one can obtain a sample from this distribution, though, laws of large numbers allow us to approximate quantities of interest via Monte Carlo methods. The most common approach approach to obtaining a sample from a posterior distribution is Markov chain Monte Carlo (MCMC), particularly Gibbs sampling.

One reason for the popularity of Gibbs sampling is the ease with which the general algo-

rithm can be constructed. It proceeds simply by initializing a chain at $(\boldsymbol{\beta}x_0, \boldsymbol{\beta}\theta_0)$ and, at iteration t, sampling $\boldsymbol{\beta}x_t \sim \pi(\boldsymbol{\beta}x \mid \boldsymbol{\beta}\theta_{(t-1)}, \boldsymbol{\beta}y)$ and then $\boldsymbol{\beta}\theta_t \sim \pi(\boldsymbol{\beta}\theta \mid \boldsymbol{\beta}x_t, \boldsymbol{\beta}y)$. Under suitable conditions, ergodic theory [e.g., Robert and Casella, 2004] establishes that the resulting Markov chain $\{(\boldsymbol{\beta}x_t, \boldsymbol{\beta}\theta_t) : t = 0, 1, \cdots\}$ has $\pi(\boldsymbol{\beta}x, \boldsymbol{\beta}\theta \mid \boldsymbol{\beta}y)$ as its limiting distribution. In practice, implementing this algorithm can be challenging due to the need to draw $\boldsymbol{\beta}x \mid \boldsymbol{\beta}\theta, \boldsymbol{\beta}y$ thousands of times, which is computationally expensive when $\boldsymbol{\beta}x$ is high dimensional. This challenge is not entirely circumvented when $\boldsymbol{\beta}x$ follows a GMRF, as we discuss in the sequel.

4.2.2 Gaussian Markov Random Fields

Consider a Gaussian Markov random field (GMRF) $\boldsymbol{\beta}x = (x_1, \ldots, x_n)^T$, where x_i is the realization of the field at node i, $i = 1, \ldots n$. The density of $\boldsymbol{\beta}x$ is given by

$$\pi(\boldsymbol{\beta}x \mid \boldsymbol{\beta}\mu) \propto \exp\left(-\frac{1}{2}\boldsymbol{\beta}x^T\boldsymbol{\beta}Q\boldsymbol{\beta}x + \boldsymbol{\beta}b^T\boldsymbol{\beta}x\right), \tag{4.2}$$

where $\boldsymbol{\beta}\mu \in \mathbb{R}^n$ and $\boldsymbol{\beta}b = \boldsymbol{\beta}Q\boldsymbol{\beta}\mu$. If $\boldsymbol{\beta}Q$ is nonsingular, then this distribution is proper (i.e., $\int \pi(\boldsymbol{\beta}x \mid \boldsymbol{\beta}\mu)d\boldsymbol{\beta}x < \infty$) and the normalizing constant is $(2\pi)^{-n/2}\det(\boldsymbol{\beta}Q)^{1/2}$. Intrinsic GMRFs are such that the $\boldsymbol{\beta}x$ variables have linear constraints so that $\boldsymbol{\beta}Q$ is rank deficient. In this case, we may define the density with proportionality constant $(2\pi)^{-(n-k)/2}\det^*(\boldsymbol{\beta}Q)^{1/2}$, where $n-k$ is the rank of $\boldsymbol{\beta}Q$ and $\det^*(\cdot)$ is the product of the $n-k$ non-zero eigenvalues of $\boldsymbol{\beta}Q$ [Hodges et al., 2003, Rue and Held, 2005a]. Such improper GMRF models are common in Bayesian disease mapping [Waller et al., 1997a] and linear inverse problems [Bardsley, 2012], as they are easily interpretable.

GMRFs may be specified according to an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ indicates nodes (i.e., the vertices) and $\mathcal{E} = \{(i,j) : i \sim j\}$ is the edge set where $i \sim j$ if and only if node i is connected to (i.e., a neighbor of) node j. The precision matrix $\boldsymbol{\beta}Q$ is determined by $(\boldsymbol{\beta}Q)_{ij} \neq 0$ if and only if $(i,j) \in \mathcal{E}$. Specifying the density through the precision matrix $\boldsymbol{\beta}Q$ instead of a covariance matrix induces a Markov property in the random field [Rue and Held, 2005a, Theorem 2.2]. For any node i, $x_i \mid \boldsymbol{\beta}x_{(-i)} \stackrel{d}{=} x_i \mid \boldsymbol{\beta}x_{\mathcal{N}(i)}$, where $\boldsymbol{\beta}x_{(-i)} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)^T$, $\mathcal{N}(i) = \{j : (i,j) \in \mathcal{E}\}$ is the neighborhood of node i, and $\boldsymbol{\beta}x_{\boldsymbol{\beta}A} := (x_i \ : i \in \boldsymbol{\beta}A)^T$ for some index set $\boldsymbol{\beta}A$. That is, x_i is conditionally independent of the rest of the field given its neighbors. Most GMRFs assume that each node has relatively few neighbors, resulting in a sparse precision matrix $\boldsymbol{\beta}Q$. The sparsity of the precision matrix combined with the intuitive and easily-interpretable

Markov property has led to GMRFs being widely used to model dependence in areal data. Indeed, there exists user-friendly software that facilitates incorporating GMRFs into Bayesian spatial models without detailed knowledge of their construction. Examples include the `GeoBUGS` package in `WinBUGS` [Lunn et al., 2000] and the `R` package `CARBayes` [Lee, 2013].

With the need to model extremely large datasets with nontrivial correlation has come the need for efficient sampling techniques whereby posterior distributions of fully Bayesian models can be simulated. When periodic boundary conditions on $\boldsymbol{\beta}x \in \mathbb{R}^n$ can be assumed (i.e., each x_i has the same number of neighbors, including the edge nodes), Fox and Norton [2016] note that the sampling problem can be diagonalized via the Fast Fourier Transform (with complexity $\mathcal{O}(n \log n)$), whence a sample can be drawn by solving a system in $\mathcal{O}(n)$ operations. They propose reducing the total number of draws from the conditional distribution of $\boldsymbol{\beta}x$ by using a "marginal-then-conditional" sampler in which the MCMC algorithm operates by completely collapsing over $\boldsymbol{\beta}x$ and subsequently sampling $\boldsymbol{\beta}x$ using only the approximately independent draws of the hyperparameters obtained from a full MCMC run on their marginal distribution. In many applications, though, the periodic boundary assumption may not be realistic, and sampling from the marginal distribution of hyperparameters can itself be challenging. To avoid the computational difficulties associated with full GMRFs, Cai et al. [2013] propose using a pairwise graphical model as an approximate GMRF for high-dimensional data imputation without specifying the precision matrix directly. The authors admit, however, that this procedure is very hard to implement [Cai, 2014, p. 7]. In cases where we are given $\boldsymbol{\beta}Q$ and $\boldsymbol{\beta}b$ in (4.2) with the goal of estimating $\boldsymbol{\beta}\mu$, Johnson et al. [2013] express the Gibbs sampler as a Gauss-Seidel iterative solution to $\boldsymbol{\beta}Q\mu = \boldsymbol{\beta}b$, facilitating the "Hogwild" parallel algorithm of Niu et al. [2011] in which multiple nodes are updated simultaneously without locking the remaining nodes. In the Gaussian case, Johnson et al. [2013] prove convergence to the correct solution when the precision matrix $\boldsymbol{\beta}Q$ is symmetric diagonally dominant. Motivated by Johnson et al. [2013], Cheng et al. [2015] use results from spectral graph theory to propose a parallel algorithm for approximating a set of sparse factors of $\boldsymbol{\beta}Q^l, \ -1 \leq l \leq 1$, in nearly linear time. They show that it can be used to construct *iid* samples from an approximate distribution. This is opposed to a Gibbs sampler, which produces approximately independent samples from the correct distribution (possibly after thinning). Similar to the Gauss-Seidel splitting considered by Johnson et al. [2013], Liu et al. [2015] propose an iterative approach to approximating a draw from a GMRF in which the corresponding graph is separated into a spanning tree and the missing edges, whence the spanning

tree is randomly perturbed and used as the basis for an iterative linear solve.

The aforementioned algorithms can be difficult to implement and require substantial knowledge of graph theory, numerical analysis, and MPI programming. This makes such approaches inaccessible to many statisticians who nevertheless need to work with large random fields. Further, they are iterative routines for producing a single draw from an approximation to the target distribution. This feature makes them less appealing for users who work in `R` or `MATLAB` (The MathWorks, Natick, MA). It is well-known that loops should be avoided in these languages to avoid repeated data type interpretation and memory overhead issues. In response to these difficulties while still faced with the problem of efficient updating of GMRFs inside a larger MCMC algorithm, additional `R` packages have been made available which are beneficial for manipulating the sparse matrices associated with GMRFs, including `Matrix` [Bates and Maechler, 2016], `SparseM` [Koenker and Ng, 2016], and `spam` [Furrer and Sain, 2010a, Gerber and Furrer, 2015].

4.2.3 Block and Single-Site Gibbs Sampling

In this section, it is helpful to distinguish between sampling $\boldsymbol{\beta}x$ directly from a *prior* GMRF and from the full conditional distribution of $\boldsymbol{\beta}x$ derived from an hierarchical Bayesian model with a GMRF prior on $\boldsymbol{\beta}x$. For an unconditional GMRF, the distribution is of the form $\boldsymbol{\beta}x \sim N(\boldsymbol{\beta}\mu, \boldsymbol{\beta}Q^{-1})$, where $\boldsymbol{\beta}\mu$ and $\boldsymbol{\beta}Q$ are generally unrelated. When drawing from the full conditional distribution as in a Gibbs sampler, the distribution is of the form $\boldsymbol{\beta}x \sim N(\boldsymbol{\beta}Q_p^{-1}\boldsymbol{\beta}b, \boldsymbol{\beta}Q_p^{-1})$, where $\boldsymbol{\beta}Q_p \neq \boldsymbol{\beta}Q$ is an updated precision matrix. For example, in a typical linear model $\boldsymbol{\beta}y \mid \boldsymbol{\beta}x, \boldsymbol{\beta}\Sigma \sim N(\boldsymbol{\beta}Ax, \boldsymbol{\beta}\Sigma)$ with $\boldsymbol{\beta}A$ fixed and $\boldsymbol{\beta}x \sim N(\boldsymbol{\beta}\mu, \boldsymbol{\beta}Q^{-1})$, standard multivariate normal theory yields $\boldsymbol{\beta}x \mid \boldsymbol{\beta}y, \boldsymbol{\beta}\Sigma \sim N(\boldsymbol{\beta}Q_p^{-1}\boldsymbol{\beta}b, \boldsymbol{\beta}Q_p^{-1})$, where $\boldsymbol{\beta}Q_p = \boldsymbol{\beta}A^T\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}A + \boldsymbol{\beta}Q$ and $\boldsymbol{\beta}b = \boldsymbol{\beta}A^T\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}y + \boldsymbol{\beta}Q\mu$.

Two approaches to updating GMRFs inside a Markov chain Monte Carlo algorithm are so-called single site sampling in which individual sites are updated one at a time using the available full conditional distributions, and block Gibbs sampling in which the entire random field is updated all at once via sampling from a known multivariate Gaussian distribution induced by the GMRF. Block sampling improves the convergence of Gibbs samplers in the presence of *a posteriori* correlated variables by allowing the chain to move more quickly through the sample space [Liu et al., 1994]. The drawback is in the manipulation and solution of large covariance matrices necessary for both random variable generation and evaluation of the likelihood in a Metropolis-Hastings algorithm [Metropolis et al., 1953, Hastings, 1970]. Single site updating uses the conditional distributions

of each scalar random variable, thus avoiding large matrix computations. In single-site sampling, though, statistical efficiency may be sacrificed as updating a group of possibly highly correlated parameters one at a time can result in slow exploration of the parameter space, slowing convergence of the Markov chain.

An appealing feature of GMRFs is the ability to specify the distribution of $\boldsymbol{\beta}x$ through a complete set of full conditional distributions, $\{p(x_i \mid \boldsymbol{\beta}x_{(-i)}) : \ i = 1, \ldots, n\}$. For instance, we can assume each $x_i \mid \boldsymbol{\beta}x_{(-i)} \sim N(\eta_i, \sigma_i^2)$, with $\eta_i = \mu_i + \sum_{j \sim i} c_{ij}(x_j - \mu_j)$ and $\sigma_i^2 > 0$, where c_{ij} are specified weights such that $c_{ij} \neq 0$ if and only if $i \sim j$ and $c_{ii} = 0$ for all i. Specification of a Markov random field through these so-called local characteristics was pioneered by Besag [1974a], after which such models came to be known as *conditional autoregressive (CAR) models*. Besag [1974a] used Brook's Lemma and the Hammersley-Clifford Theorem to establish that the set of full conditionals collectively determine a joint density, provided a positivity condition holds among $\boldsymbol{\beta}x$. In this case, we have that $\boldsymbol{\beta}Q_{ij} = (I(i = j) - c_{ij}I(i \neq j))/\sigma_i^2$, where $I(\cdot)$ is the indicator function. The condition $\sigma_j^2 c_{ij} = \sigma_i^2 c_{ji}$, for all i, j, is necessary to ensure symmetry of $\boldsymbol{\beta}Q$. The ease with which these full conditional distributions can be incorporated into a Gibbs sampling algorithm has led to a dramatic increase in the popularity of CAR models over the past twenty years or so [Lee, 2013, Banerjee et al., 2015b].

In single-site Gibbs sampling, we sequentially draw from each univariate distribution with density $p(x_i \mid \boldsymbol{\beta}x_{(-i)}), \ i = 1, \ldots n$. This requires using $\boldsymbol{\beta}x_{(-i)}$ to calculate η_i prior to drawing from each of the n conditional distributions, meaning that single-site updating essentially becomes an $\mathcal{O}(n^2)$ operation. This algorithm has little regard for the ordering of the nodes, making such sampling strategies very easy to implement. Compared to block updating, though, many more Gibbs scans are required to sufficiently explore the support of the distribution. This approach is the most iteration-intensive of any of the approaches considered here. As such, its implementation in `R` can result in a large amount of overhead associated with loops, considerably slowing the entire routine.

Efficient block sampling schemes for GMRFs are discussed in Rue [2001] and Knorr-Held and Rue [2002]. What most of these schemes have in common is the use of a Cholesky factorization of the precision matrix and using the factorization to solve a system of equations. For the case typically encountered in a Gibbs sampler, Rue and Held [2005a] provide an algorithm for simulating from $N(\boldsymbol{\beta}Q_p^{-1}\boldsymbol{\beta}b, \boldsymbol{\beta}Q_p^{-1})$. This algorithm, presented in Algorithm 1 in Appendix A, requires one Cholesky factorization and three linear solves via forward or backward substitution.

In general, for a matrix of dimension $n \times n$, the Cholesky factorization is an $\mathcal{O}(n^3/3)$ operation and each linear solve costs $\mathcal{O}(n^2)$ flops [Golub and Van Loan, 1996]. This can be particularly onerous in a fully Bayesian approach in which hyperpriors are assigned to hyperparameters $\boldsymbol{\beta}\theta$ that appear in the precision matrix $\boldsymbol{\beta}Q_p \equiv \boldsymbol{\beta}Q_p(\boldsymbol{\beta}\theta)$. However, the key to making block updating feasible on high-dimensional data lies in the computational savings that can be achieved when $\boldsymbol{\beta}Q_p$ is sparse. Sparse matrix algebra is itself a non-trivial problem requiring specialized knowledge beyond the expertise of many statisticians. Indeed, concerning this point, Rue and Held [2005a, p. 52] recommend "leaving the issue of constructing and implementing algorithms for factorizing sparse matrices to the numerical and computer science experts." In practice, most statisticians will rely on special functions for sparse matrices such as those found in the `Matrix` [Bates and Maechler, 2016], `SparseM` [Koenker and Ng, 2016], or `spam` [Furrer and Sain, 2010a] packages in `R`. Of these three, `spam` is the most specifically tailored for repeatedly sampling GMRFs in MCMC.

For simulating posterior distributions via block Gibbs sampling, we are interested in drawing from full conditional distributions. In this case, sparsity of the entire precision matrix is contingent upon the sparsity of $\boldsymbol{\beta}A^T\boldsymbol{\beta}\Sigma^{-1}\boldsymbol{\beta}A$. This is often the case in practice. For instance, in disease mapping and related applications, it is common to place a spatially correlated random effect at each location to encourage smoothing of the incidence rate over space [e.g., Waller et al., 1997a, Banerjee et al., 2015b]. In terms of the linear model, this can be expressed as $\boldsymbol{\beta}y - \boldsymbol{\beta}X\beta = \boldsymbol{\beta}Z\gamma + \boldsymbol{\beta}\varepsilon$, where $\boldsymbol{\beta}X\beta$ corresponds to fixed effects and $\boldsymbol{\beta}\gamma$ contains the spatially-varying effects. With site-specific random effects, $\boldsymbol{\beta}Z$ is diagonal or block diagonal. The diagonal case (e.g., $\boldsymbol{\beta}Z = \boldsymbol{\beta}I$) is especially amenable to efficient block Gibbs sampling as well as chromatic sampling, since the underlying graph $\mathcal{G}$ for the full conditional distribution is exactly the same as the prior graph.

For a fully Bayesian model, the full conditional precision matrix associated with the GMRF will in general depend on parameters that are updated in each iteration of an MCMC routine, meaning that the Cholesky factorization has to be recomputed on each iteration. Often, though, the neighborhood structure and thus the sparsity pattern of the Cholesky factor remain fixed. The sparse matrix implementation in the `spam` package exploits this fact to accelerate repeated block GMRF updates. After finding the initial Cholesky factorization using so-called supernodal elimination trees [Ng and Peyton, 1993], `spam` stores the symbolic factorization and only performs numeric factorizations on subsequent iterations. Even with sparse matrix algebra, though, block sampling $\boldsymbol{\beta}x$ from a GMRF in very high dimensions can be problematic due to the computational cost of even

an initial factorization as well as the associated memory overhead [Rue, 2001, p. 331].

4.2.4 Chromatic Gibbs Sampling

Consider the graph representation of the GMRF, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The *local Markov property* says that $x_i \perp \boldsymbol{\beta}x_{-(i,\mathcal{N}(i))} \mid \boldsymbol{\beta}x_{\mathcal{N}(i)}$, where $\boldsymbol{\beta}x_{-(i,\mathcal{N}(i))}$ denotes all x except x_i and the neighborhood of x_i, and $\perp$ denotes (statistical) independence. An extension of the local Markov property is to let $C \subset \mathcal{V}$ denote a separating set, or *cut*, of $\mathcal{G}$ such that nodes in a set $A \subset \mathcal{V}$ are disconnected from nodes in $B \subset \mathcal{V}$ after removing the nodes in C from the graph. Then the *global Markov property* states that $\boldsymbol{\beta}x_A \perp \boldsymbol{\beta}x_B \mid \boldsymbol{\beta}x_C$. Chromatic sampling exploits this property by partitioning the nodes according to a graph coloring whereby each subset can be updated simultaneously in parallel.

A *coloring* $f : \mathcal{V} \to \{1, \ldots, k\},\ k \in \mathbb{N}$, is a collection of labels assigned to nodes on a graph so that no two nodes that share an edge have the same label. A k-coloring induces a partition of the nodes $\{\mathcal{A}_1, \ldots, \mathcal{A}_k\}$, where $\mathcal{A}_j = f^{-1}(\{j\}) \subset \mathcal{V}$. For example, Figure 4.1 displays a 4-coloring that could be used for data that lie on a regular two-dimensional lattice; e.g., imaging data. Given a k-coloring of the MRF graph, we can determine a *cut* C_j corresponding to each color j by assigning all nodes that are not of that color to be in the cut; i.e., $C_j = \mathcal{A}_j^c,\ j = 1, \ldots, k$. Defining cuts in this way for $j = 1, \ldots, k$, we have that $x_i \mid \boldsymbol{\beta}x_{C_j} \overset{\text{indep.}}{\sim} N(\eta_i, \sigma_i^2)$, for all $i \in \mathcal{A}_j$, where each η_i and σ_i^2 depend on $\boldsymbol{\beta}x_{C_j}$. That is, all nodes of the same color are conditionally independent and can be sampled in parallel, given the rest of the field. The use of graph colorings in this way leads Gonzalez et al. [2011a] to term the approach *chromatic Gibbs sampling*.

⊛	⦶	⊛	⦶	$\cdots$	⊛	⦶
◎	⊙	◎	⊙	$\cdots$	◎	⊙
⊛	⦶	⊛	⦶	$\cdots$	⊛	⦶
◎	⊙	◎	⊙	$\cdots$	◎	⊙
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
⊛	⦶	⊛	⦶	$\cdots$	⊛	⦶
◎	⊙	◎	⊙	$\cdots$	◎	⊙

Figure 4.1: An example of a k-coloring ($k = 4$) for nodes on a regular two-dimensional lattice. Each symbol represents a different label.

Algorithm 1 presents a general chromatic Gibbs sampler for GMRFs. An advantage of

Input: Current state of GMRF, $\boldsymbol{\beta}x^{(t)}$, a k-coloring of the MRF graph, $\{\mathcal{A}_j : \ j = 1, \ldots, k\}$.
Output: New draw $\boldsymbol{\beta}x^{(t+1)}$ from the GMRF.

1 **for** $j = 1$ *to* k **do**
2 For $i \in \mathcal{A}_j$, calculate conditional means and standard deviations η_i, σ_i^2 using $\boldsymbol{\beta}x_{\mathcal{A}_1}^{(t+1)}, \ldots, \boldsymbol{\beta}x_{\mathcal{A}_{j-1}}^{(t+1)}, \boldsymbol{\beta}x_{\mathcal{A}_{j+1}}^{(t)}, \ldots, \boldsymbol{\beta}x_{\mathcal{A}_k}^{(t)}$
3 Draw $\boldsymbol{\beta}x_{\mathcal{A}_j} \sim N\left[(\eta_1, \ldots, \eta_{|\mathcal{A}_j^c|})^T, \ \mathrm{diag}(\sigma_i^2, i = 1, \ldots, |\mathcal{A}_j^c|)\right]$
4 **end**
5 **Return** $\boldsymbol{\beta}x^{(t+1)}$

Algorithm 1: Chromatic Gibbs step updating of a GMRF.

using this approach is in step 3 of the algorithm. When updates of the random variables indexed by $\mathcal{A}_j$ is distributed across several processors, the computational effort of updating the entire field can be dramatically reduced, even compared to the approximate linear complexity obtained from sparse matrix factorization. Given p processors and a k-coloring of a Markov random field over n nodes, the chromatic Gibbs sampler generates a new sample in approximately $\mathcal{O}(n/p+k)$ operations [Gonzalez et al., 2011a, p. 326]. Of course, the best computational savings will be achieved by using the *chromatic index* for the coloring (i.e., the minimum k so that a k-coloring of $\mathcal{G}$ exists).

The minimal coloring problem for a graph is NP-Complete and thus very challenging except in simple situations. On regular lattices with commonly assumed neighborhood structures (e.g., Figure 4.1), such colorings can be found by inspection without complicated algorithms. Algorithm 2 is a greedy algorithm for finding a coloring on an arbitrary graph, including irregular lattices that would be encountered in, e.g., disease mapping and climate studies. This algorithm is not guaranteed to find an optimal coloring. It is relatively quick compared to searching for the minimal coloring, though, and any such coloring can be effective, as we demonstrate in Section 4.3. It is important to observe that for fixed sparsity patterns (and hence fixed Markov graphs) that we consider here, graph coloring is a pre-computation. It is only required to run the algorithm once prior running MCMC.

Most computers today have parallel processing capabilities, and any distributed processing over p processors can reduce the computational burden by an approximate factor of $1/p$. Regardless of the number of processors available to the user, though, savings can still be realized when working in a high-level language such as `R` by 'vectorizing' the updating of the conditionally independent sets. Vectorizing still ultimately uses a `for` loop on each set of nodes, but the loops are performed in

Input: MRF graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$.
Output: k-coloring partition $\{\mathcal{A}_1, \mathcal{A}_2, \ldots \mathcal{A}_k\}$, for some k.

Set $j = 1$ and $\mathcal{A}_0 = \emptyset$
while $\mathcal{V} \setminus \cup_{l=0}^{j-1} \mathcal{A}_l \neq \emptyset$ **do**
  $\mathcal{I}_j \leftarrow \mathcal{V} \setminus \cup_{l=0}^{j-1} \mathcal{A}_l$
  $\mathcal{A}_j \leftarrow \emptyset$
  while $|\mathcal{I}_j| > 0$ **do**
    $i \leftarrow \min \mathcal{I}_j$
    $\mathcal{A}_j \leftarrow \mathcal{A}_j \cup \{i\}$
    $\mathcal{I}_j \leftarrow \mathcal{I}_j \setminus (\{i\} \cup \mathcal{N}(i))$
  end
  $j \leftarrow j + 1$
end
$k \leftarrow j - 1$
Return $\{\mathcal{A}_1, ..., \mathcal{A}_k\}$

Algorithm 2: Greedy algorithm for $k-$coloring the nodes of an MRF graph.

a faster language such as `C` or `Fortran`. It also minimizes the overhead associated with interpreting data types; i.e., vectorizing allows `R` to interpret the data type only once for the entire vector instead of repeatedly for each element of the vector.

4.3 Numerical Illustrations

In this section we compare chromatic sampling to block Gibbs updating and single-site sampling with both simulated data on large regular arrays and real, non-Gaussian (binomial) data on an irregular lattice. We emphasize that the computational improvements are realized *without* direct parallel processing. We simply vectorize the simultaneous updating steps, thereby avoiding direct `for` loops in `R`. To implement the block Gibbs sampler, we use the `spam` package [Furrer and Sain, 2010a], since it is specifically tailored for GMRFs inside MCMC routines by storing the sparsity structure for repeated use. To make the `spam` functions as efficient as possible, we follow the authors' suggestion and turn off the symmetry check and safe mode options (`options(spam.cholsymmetrycheck= FALSE, spam.safemodevalidity= FALSE)`).

4.3.1 Simulated Imaging Problem

Image analysis involves attempting to reconstruct a true latent image, where the 'image' may mean a true physical structure as in clinical medical imaging, or an activation pattern or signal as in,

e.g., functional magnetic resonance imaging [fMRI; Lazar, 2008]. The available data consist of pixel values corresponding to color, often on the grayscale taking integer values from 0 to 255. The true values are assumed to have been contaminated with error due to the image acquisition process. This area is one of the original motivating applications for Markov random fields [Besag, 1986, Besag et al., 1991b].

There is growing interest in the statistical analysis of ultra-high dimensional imaging data. For example, structural MR images of the human brain may consist of 20-40 two dimensional slices, each of which has 256×256 resolution or higher. Spatial Bayesian models for even a single slice of such data can involve GMRFs over lattices of dimension $n = 256^2$ [Brown et al., 2017b] and thus are very computationally challenging when drawing inference via Markov chain Monte Carlo. Motivated by such applications, we consider images consisting of $p \times p$ pixels, each of which has an observed value $y_{ij} = x_{ij} + \varepsilon_{ij}$, where x_{ij} is the true value of the $(i,j)^{\text{th}}$ pixel in the latent image and ε_{ij} represents the corresponding contamination. To simulate the data, we take the error terms to be independent, identically distributed $N(0, 0.1^2)$ random variables. The true image in this case is a rescaled bivariate Gaussian density with $x_{ij} = 5\exp\{-\|\boldsymbol{\beta} v_{ij}\|^2/2\}/\pi$, where $\boldsymbol{\beta} v_{ij} = (v_i, v_j) \in [-3,3] \times [-3,3]$ denotes the center of the $(i,j)^{\text{th}}$ pixel, evenly spaced over the grid, and $\|\cdot\|$ denotes the usual Euclidean norm. Figure A.10 depicts the true generated image (in 50×50 resolution) and its corrupted counterpart. To study each of the three sampling algorithms, we consider first an image with dimension $n = p \times p = 50^2$.

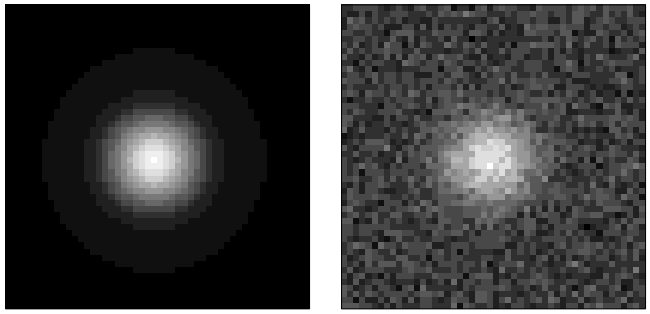


Figure 4.2: True image (left panel) and corrupted image (right panel) for the simulated image reconstruction example. (These particular images have resolution 50×50.)

The assumed model for the observed image is given by $\boldsymbol{\beta} y = \boldsymbol{\beta} 1 \beta_0 + \boldsymbol{\beta}\gamma + \boldsymbol{\beta}\varepsilon$, where $\boldsymbol{\beta} y \in \mathbb{R}^n$

is the vector of the observed pixel values, $\boldsymbol{\beta}1 = (1, \ldots, 1)^T \in \mathbb{R}^n$, $\beta_0 \in \mathbb{R}$ is a constant intercept parameter, $\boldsymbol{\beta}\gamma$ is the vector of spatial effects, and $\boldsymbol{\beta}\varepsilon$ is the vector of errors assumed to follow a $N(\boldsymbol{\beta}0, \sigma^2\boldsymbol{\beta}I)$ distribution. To capture local homogeneity of the images, we assume the spatial random effects obey an intrinsic autoregressive (IAR) model [Besag and Kooperberg, 1995b, Banerjee et al., 2015b]; i.e., $\boldsymbol{\beta}\gamma \sim N\left(\boldsymbol{\beta}0, \tau^2(\boldsymbol{\beta}D - \boldsymbol{\beta}W)^{-1}\right)$, where $\boldsymbol{\beta}W = \{w_{ij} := I(i \sim j)\}_{i,j=1}^n$ is the incidence matrix of the underlying graph and $\boldsymbol{\beta}D = \text{diag}\left(\sum_{j=1}^n w_{ij} : \ i = 1, \ldots, n\right)$. Here we assume each pixel has a first-order neighborhood structure in which each interior pixel has eight neighbors. We ignore edge effects induced by the perimeter pixels of the image. We specify inverse gamma priors for the variance components and a flat prior for the intercept; i.e., $\sigma^2 \sim \text{InvGam}(\alpha, \alpha)$, $\tau^2 \sim \text{InvGam}(\alpha, \alpha)$, and $\pi(\beta_0) \propto 1$. To approximate vague priors for the variance components, we take $\alpha = 0.001$. It has been observed that an inverse gamma prior on τ^2 sometimes can yield undesirable behavior in the posterior [Gelman, 2006]; but our focus is on sampling the random field and thus we use this prior simply for convenience. For posterior sampling, our modeling assumptions lead to a Gibbs sampler having the following full conditional distributions: $\beta_0|\boldsymbol{\beta}y, \boldsymbol{\beta}\gamma, \sigma^2 \sim N\left(\boldsymbol{\beta}1^T(\boldsymbol{\beta}y - \boldsymbol{\beta}\gamma)/n, \sigma^2/n\right)$, $\sigma^2|\boldsymbol{\beta}y, \boldsymbol{\beta}\gamma, \beta_0 \sim \text{InvGam}\left(\alpha + n/2, \alpha + \|\boldsymbol{\beta}y - \boldsymbol{\beta}1\beta_0 - \boldsymbol{\beta}\gamma\|^2/2\right)$,
$\tau^2|\boldsymbol{\beta}\gamma \sim \text{InvGam}\left(\alpha + n/2, \alpha + \boldsymbol{\beta}\gamma^T(\boldsymbol{\beta}D - \boldsymbol{\beta}W)\boldsymbol{\beta}\gamma/2\right)$, and $\boldsymbol{\beta}\gamma|\boldsymbol{\beta}y, \sigma^2, \tau^2 \sim N\left(\boldsymbol{\beta}Q_p^{-1}\boldsymbol{\beta}b, \boldsymbol{\beta}Q_p^{-1}\right)$, where $\boldsymbol{\beta}Q_p = \sigma^{-2}\boldsymbol{\beta}I + \tau^{-2}(\boldsymbol{\beta}D - \boldsymbol{\beta}W)$ and $\boldsymbol{\beta}b = (\boldsymbol{\beta}y - \boldsymbol{\beta}1\beta_0)/\sigma^2$.

To implement the Gibbs sampler, three separate sampling strategies are employed, with the only difference being how we sample the full conditional distribution of $\boldsymbol{\beta}\gamma$. First, since $\boldsymbol{\beta}Q_p$ is sparse, we consider full block Gibbs sampling based on Algorithm 1 in Appendix A to sample $\boldsymbol{\beta}\gamma$ in a single block. The second strategy is single-site Gibbs sampling using the local characteristics, $\gamma_i|\boldsymbol{\beta}\gamma_{(-i)}, \boldsymbol{\beta}y, \sigma^2, \tau^2 \sim N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, n$, where $\mu_i = \sigma_i^2(\sigma^{-2}y_i + \tau^{-2}\sum_{j\in\mathcal{N}(i)} w_{ij}\gamma_j)$ and $\sigma_i^2 = \tau^2\sigma^2\{\sigma^2(\boldsymbol{\beta}D)_{ii} + \tau^2\}^{-1}$. The single-site approach obviates the need to work with large matrices. The final sampling strategy we implement is chromatic Gibbs sampling discussed in Subsection 4.2.4. This approach uses the coloring depicted in Figure 4.1 as a 4-coloring of the pixels in the image. Following the notation in Subsection 4.2.4, we have that $\gamma_i \mid \boldsymbol{\beta}\gamma_{C_j}, \boldsymbol{\beta}y, \sigma^2, \tau^2 \overset{\text{indep.}}{\sim} N(\mu_i, \sigma_i^2)$, $i \in \mathcal{A}_j$, $j = 1, ..., 4$, where μ_i and σ_i^2 are determined from the local characteristics. Observe that all of the necessary conditional means and variances for a given color can be computed simultaneously through matrix multiplication and addition on vectors of dimension $|\mathcal{A}_j^c| < n$, as opposed to the iterative updating of these quantities in the single-site sampler. An important feature of the chromatic sampler is that $\gamma_i \mid \boldsymbol{\beta}\gamma_{C_j}, \boldsymbol{\beta}y, \sigma^2, \tau^2$, $i \in \mathcal{A}_j$, can be drawn simultaneously.

We implement the three sampling strategies so that each procedure draws $10,000$ realizations from the Monte Carlo Markov chain to approximate the posterior distribution of the model parameters. For each approach, three chains are run using dispersed intial values. We assess convergence of the chains via trace plots, Gelman plots [Brooks and Gelman, 1998], scalar and multivariate potential scale reduction factors, and plots of cumulative ergodic averages of scalar hyperparameters. We discard the first 8000 iterations as a burn-in period and assess convergence using the last 2000 realizations of each Markov chain. The simulations, coded entirely in `R`, are carried out on a Dell Precision T3600 desktop running Windows 7 with an Intel Xeon 3.30 GHz CPU and 64 GB of RAM.

Figure 4.3 displays the trace plots and empirical autocorrelation functions for the hyperparameters σ^2 and τ^2 for chromatic, block, and single-site sampling. We see very similar behavior in terms of autocorrelation across all three sampling approaches. From Supplementary Figure A.1, we glean that each sampling approach has approximately converged in the σ^2 and τ^2 chains after 6,000 iterations, although the block sampler has the longest convergence time according to the Gelman plots. While each approach produces estimates of σ^2 and τ^2 that tend to the same value, the single-site approach evidently has larger Monte Carlo standard error than the other two approaches. This partly explains the slight difference in empirical distribution from the single-site sampler versus that of the chromatic and block samplers, as depicted in Figure 4.4. Regardless, the joint and marginal distribution estimates largely agree. This agreement is also evident in Figure 4.5, which displays the posterior mean estimates of the true image $\beta_0\boldsymbol{\beta}1 + \boldsymbol{\beta}\gamma$, the primary quantity of interest. To assess exploration of the posterior distributions, the Figure also depicts point-wise ratios of sample standard deviations for each pair of algorithms. All three samplers produce distributional estimates that are virtually indistinguishable, and the approximate convergence (as measured by the multivariate potential scale reduction factors) is about the same.

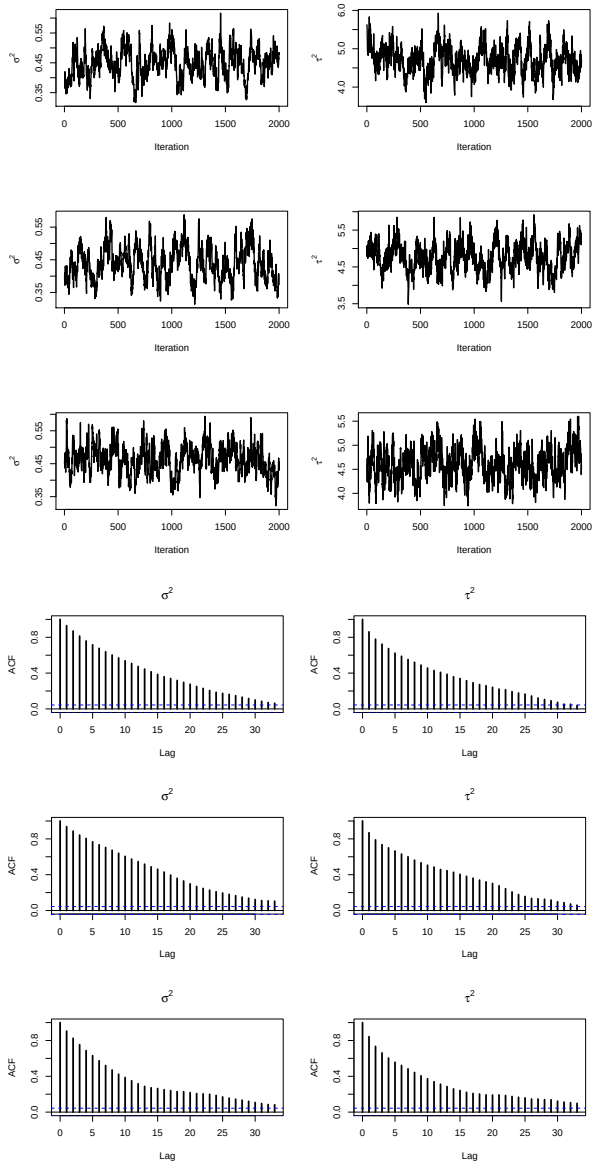


Figure 4.3: MCMC Trace plots (two left columns) and empirical ACF plots (two right columns) of single chains each for σ^2 and τ^2 for the 50×50 regular array example. The top, middle, and bottom rows are from the chromatic, block, and single-site chains, respectively.

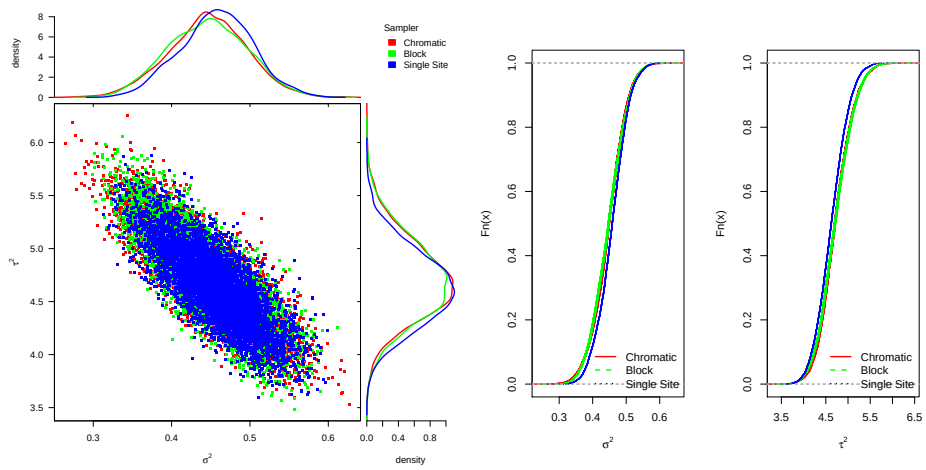


Figure 4.4: Left panel: Scatterplot and estimated marginal posterior densities (left) and empirical CDFs (right) from the three sampling approaches in the 50×50 array example. The left panel was created using code available at `https://github.com/ChrKoenig/R_marginal_plot`.

Sampler	CPU Time (s)	ESS	ACT	CES
Chromatic	26.29	84.64	23.63	0.31
Block	65.86	65.26	30.65	1.00
Single-Site	6268.1	68.97	29.00	90.89

Table 4.1: CPU times to draw 2,000 realizations (including 8,000 burn-in iterations) from one τ^2 Markov chain under each sampling approach in the 50×50 array example. Also reported are the effective sample sizes (ESS), autocorrelation times (ACT), and costs per effective sample (CES).

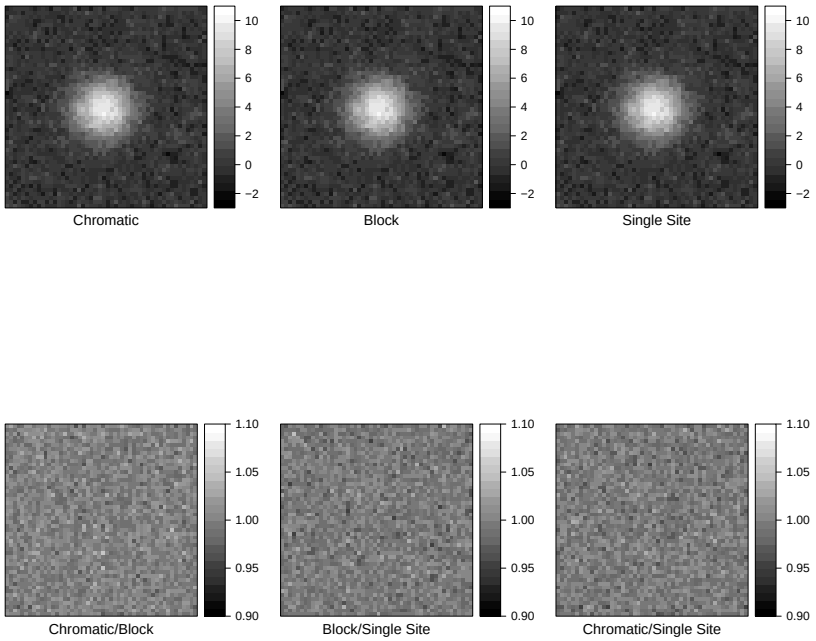


Figure 4.5: Posterior mean estimates of the true image obtained from each sampling approach (top row) along with pairwise standard deviation ratios (bottom row) in the 50×50 array example. The multivariate potential scale reduction factors for the $\boldsymbol{\beta}\varphi := \beta_0\boldsymbol{\beta}1 + \boldsymbol{\beta}\gamma$ chains are 1.22, 1.21, and 1.21 for chromatic, block, and single-site samplers, respectively.

The advantage of the chromatic approach versus the other two is in the computational cost incurred to obtain each sample. The CPU time required to draw 10,000 realizations from one Markov chain each under each sampling approach is displayed in Table 4.1. Of course, the same number of samples from two algorithms is not guaranteed to provide the same quality of posterior

approximation. To accommodate the different convergence characteristics of the three algorithms while still considering total computation time (including the burn-in period), we measure also the *cost per effective sample* [Fox and Norton, 2016], $CES := N^{-1}\kappa T$, where T is the total computation time, N is the size of the retained sample from the Markov chain (after burn-in), and κ is the integrated autocorrelation time [Kass et al., 1998, Carlin and Louis, 2009]. This quantity measures the total computational effort required to generate an effectively *independent* sample from the target distribution. Table 4.1 displays these metrics for the τ^2 chains under each sampling approach. Here we see a 70% improvement in computational effort between independent samples compared to block Gibbs sampling. Single-site sampling is by far the worst performer. It is interesting to note that in this case, the chromatic sampler has the shortest autocorrelation time of the three methods considered.

To further study the performance of block sampling versus chromatic sampling, particularly how they scale with regular arrays of increasing dimension, we repeat the model fitting procedure using data simulated as before, but with images of size $p \times p$, for $p = 80, 128, 256$. To create a more challenging situation, which slows convergence of any MCMC algorithm via poorly identified parameters, we also add considerably more noise to the images by assuming $Var(\boldsymbol{\beta}\varepsilon) = 50^2\boldsymbol{\beta}\boldsymbol{I}$ instead of $0.1^2\boldsymbol{\beta}\boldsymbol{I}$ as before. This makes the underlying spatial field much more weakly identified by the data and thus more strongly determined by the prior. Hence the GMRF parameters ($\beta_0, \boldsymbol{\beta}\boldsymbol{\gamma}$, τ^2) will be more strongly correlated in the posterior. For each p, we run the same model with the same prior specifications. We again run each MCMC algorithm for 10,000 iterations, treating the first 8,000 as burn-in periods.

Supplementary Figures A.2 through A.8 display diagnostics and posterior mean estimates produced by the different sampling procedures under $p = 50, 80, 128$ with noisy data. The single-site sampler was not computationally feasible for images with resolution $p = 80$ or higher and so was not considered. As expected, we see the autocorrelation in the τ^2 chains dramatically increased with the noisy data, regardless of the sampling approach, whereas the data-level variance σ^2 remains well identified. For the 50×50 case, the three approaches still produce parameter estimates that agree with each other, as evident in Supplementary Figure A.4. Considerable differences between chromatic sampling and block updating become apparent at $p = 80$ and $p = 128$. While the σ^2 chains still agree, we see very poor convergence of the β_0 chain under chromatic sampling, in contrast to the much better posterior exploration afforded by the block updates. The ability to

estimate β_0 directly affects the estimate of the hyper-variance τ^2 due to the trade-off between mean and covariance structure that is inherent in any spatial model. Here we see the benefit of block updates on the conditional spatial field. The deterioration of convergence in the β_0 and τ^2 chains is due to the weak identifiability of the underlying field. In many imaging problems, however, the quantity of interest is actually a function of the individual parameters. In this case, we are only interested in $\boldsymbol{\beta}\varphi := \beta_0\boldsymbol{\beta}1 + \boldsymbol{\beta}\gamma$, meaning that we want to explore the so-called *embedded posterior distribution* of $\boldsymbol{\beta}\varphi$. The parameter $\boldsymbol{\beta}\varphi$ is well identified under both sampling approaches, as evident in Supplementary Figure A.7 and Figure 4.6, which display the posterior mean estimates of the underlying images in the 80×80 and 128×128 cases, respectively. Again, the estimates of the latent image obtained from chromatic and block sampling are virtually identical. This phenomenon echoes the observation of Gelfand and Sahu [1999] that even when a Gibbs sampler is run over a vague (or improper) posterior distribution, inferences can still be drawn for certain lower dimensional estimands.

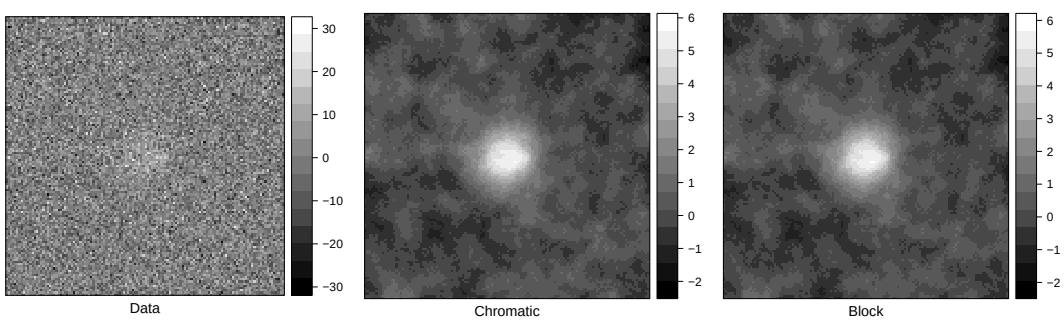


Figure 4.6: Simulated data and posterior mean estimates of the the true image from the chromatic and block sampling approaches in the noisy 128×128 array example. The multivariate potential scale reduction factors for the $\boldsymbol{\beta}\varphi := \beta_0\boldsymbol{\beta}1 + \boldsymbol{\beta}\gamma$ chains are approximately 1.00 for both samplers.

Figure 4.7 displays the total CPU time required to complete 10,000 iterations for $p = 50, 80, 128, 256$. The single site sampler is only feasible in the $p = 50$ case, as it is by far the most inefficient implementation in `R` due to the nested loops that result from sequential updating. Chromatic sampling requires much less computing time, and scales at a lower rate than block updating, due in part to the fact that no Cholesky factorizations are required. Such factorizations with even sparse matrices can be expensive, and repeated multivariate Gaussian draws are still

required even when the symbolic factorization is stored throughout the MCMC routine. There is also considerable memory overhead associated with both sparse Cholseky block updating and chromatic sampling. The total required memory for Cholesky-based block updating depends on the storage scheme used by the sparse matrix implementation. The `spam` implementation in our example uses a variant of the so-called compressed sparse row format [Sherman, 1975]. Chromatic sampling, on the other hand, requires no matrix storage at all, but only lists of identifiers associated with each graph color. The right panel of Figure 4.7 illustrates the consequent savings in memory requirements and how they scale with arrays of increasing dimension. In fact, block updating for 256×256 was not possible due to memory limitations. The Cholesky factorization failed, returning `Cholmod error: 'problem too large'`. In contrast, we see that chromatic sampling scales to extremely large problems, even to a 512×512 array with CPU time 1893.19 seconds (results not shown).

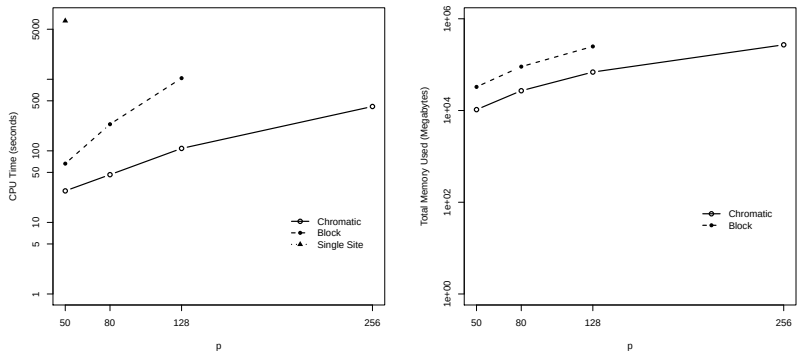


Figure 4.7: CPU time (left) and total memory required (right) for chromatic and block Gibbs sampling to complete 10,000 iterations for simulated noisy $p \times p$ arrays. The CPU time for single-site sampling is also displayed for the $p = 50$ case. In both plots, the y axis is on the log scale.

4.3.2 Binomial Election Data on an Irregular Lattice

Here we examine the performance of the block Gibbs and chromatic sampling strategies on an irregular lattice, since both the structure of $\boldsymbol{\beta}Q$ and the possible colorings of the underlying graph are more complicated. Moreover, we illustrate the performance of these procedures when applied to non-Gaussian data. In particular, we examine geographical trends in voter preference using binomial outcomes. The data were obtained from the Harvard Election Data Archive (`https:`

//dataverse.harvard.edu).

Our data consist of polling results from the 2010 New York Governor's race in which Democratic candidate Andrew Cuomo defeated Republican candidate Carl Paladino and Green Party candidate Howie Hawkins. During this election, the state of New York had $n = 14,926$ precincts, with polling data being available on $14,597$ precincts. The 329 precincts for which data are unavailable is attributable to improper reporting or lack of voter turnout.

Let Y_i be the number of votes cast for the Democratic candidate out of m_i total votes in precinct i, $i = 1, \ldots, n$. Then we assume that $Y_i|p_i, m_i \overset{indep.}{\sim} \text{Bin}(m_i, p_i)$, where $g^{-1}(p_i) = \beta_0 + \gamma_i$, $g(\cdot)$ is the usual logistic link function, and $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_n)^T$ is a vector of random effects inducing spatial homogeneity. We suppose that $\boldsymbol{\beta}\gamma$ follows a *proper* IAR model; i.e., $\boldsymbol{\beta}\gamma \sim N(\boldsymbol{\beta}0, \tau^2(\boldsymbol{\beta}D - \rho\boldsymbol{\beta}W))$, where $\boldsymbol{\beta}D$ and $\boldsymbol{\beta}W$ are as defined in Section 4.3.1. Here, the "propriety parameter" $\rho \in (\lambda_1^{-1}, \lambda_n^{-1})$ ensures that the precision matrix is non-singular, where $\lambda_1 < 0$ and $\lambda_N > 0$ are the smallest and largest eigenvalues of $\boldsymbol{\beta}D^{-1/2}\boldsymbol{\beta}W\boldsymbol{\beta}D^{-1/2}$, respectively [Banerjee et al., 2015b]. Proper IARs are sometimes used as approximations to the standard IAR when a proper prior distribution is desired. For simplicity, we fix $\rho = 0.995$. The model is completed with the prior assumptions that $\beta_0 \sim N(0, 1000)$ and $\tau^2 \sim \text{InvGam}(1, 1)$.

Under the logistic link, we can simplify posterior sampling via data augmentation. This technique exploits the fact that $\exp(\eta)^a(1+\exp(\eta))^{-b} = 2^{-b}\exp(\kappa\eta)\int_0^\infty \exp(-\psi\eta^2/2)p(\psi|b, 0)\, d\psi$, where $\eta \in \mathbb{R}$, $a \in \mathbb{R}$, $b \in \mathbb{R}^+$, $\kappa = a - b/2$, and $p(\cdot|b, 0)$ is the probability density function of a Pólya-Gamma random variable with parameters b and 0 [Polson et al., 2013]. Using this identity, the observed data likelihood can be written as $\pi(\mathbf{Y}|\beta_0, \boldsymbol{\gamma},) \propto \prod_{i=1}^n \exp\{\kappa_i\eta_i\} \times \int_0^\infty \exp(-\psi_i\eta_i^2/2)p(\psi_i|m_i, 0)d\psi_i$, where $\boldsymbol{Y} = (Y_1, ..., Y_n)^T$, $\eta_i = \beta_0 + \gamma_i$ and $\kappa_i = Y_i - m_i/2$. Thus, by introducing ψ_i as latent random variables, we have that

$$\pi(\boldsymbol{Y}, \boldsymbol{\psi}|\beta_0, \boldsymbol{\gamma}) \propto \exp\{-\left((\beta_0\boldsymbol{\beta}1 + \boldsymbol{\beta}\gamma)^T\boldsymbol{D}_{\boldsymbol{\psi}}(\beta_0\boldsymbol{\beta}1 + \boldsymbol{\beta}\gamma) - 2\boldsymbol{\kappa}^T\left(\beta_0\boldsymbol{\beta}1 + \boldsymbol{\beta}\gamma\right)\right)/2\}\prod_{i=1}^n p(\psi_i|m_i, 0),$$

where $\boldsymbol{\psi} = (\psi_1, ..., \psi_n)^T$, $\boldsymbol{\kappa} = (\kappa_1, ..., \kappa_n)$, and $\boldsymbol{D}_{\boldsymbol{\psi}} = \text{diag}(\boldsymbol{\psi})$. By including $\boldsymbol{\beta}\psi$ in the MCMC algorithm, we induce a Gaussian full conditional distribution on $\boldsymbol{\beta}\gamma$, facilitating GMRF updates without having to tune a Metropolis-Hastings algorithm. Additional implementation details are provided in Appendix A.

In order to implement chromatic sampling, a coloring of the underlying Markov graph has

Sampler	CPU Time (s)	ESS	ACT	CES
β_0 Chromatic	222.06	2445.74	2.04	0.0935
β_0 Block	294.16	2732.82	1.83	0.1018
τ^2 Chromatic	222.06	1916.63	2.61	0.1155
τ^2 Block	294.16	2186.021	2.29	0.1332

Table 4.2: CPU times to draw 5,000 realizations (including 5000 burin-in iterations) from one Markov chain under each sampling approach in the New York election example. Also reported are the effective sample sizes (ESS), autocorrelation times (ACT), and costs per effective sample (CES).

to be found. Using the greedy algorithm given in Algorithm 2, we obtain a 7-coloring, so that the chromatic sampler can update the entire $n = 14,926$-dimensional field in seven steps. The coloring is depicted in Supplementary Figure A.11.

We implement Gibbs sampling with both the block Gibbs and chromatic updates for $10,000$ iterations, discarding the first $5,000$ as a burn-in period. The simulations are run on a desktop using Windows 10 with an Intel Core i5-3570 3.40GHz CPU with 16GB of RAM. The trace plots and empirical ACF plots for β_0 and τ^2 are depicted in Supplementary Figure A.12, along with the Gelman plots of these two parameters in Supplementary Figure A.13. We see adequate convergence in the same number iterations under both sampling approaches. Table 4.2 summarizes the results for both samplers in terms of CPU time and cost per effective sample of the intercept and variance terms. We again see a savings in CPU time under chromatic sampling, so much so that it offsets the slightly larger autocorrelation time. Thus we are able to obtain effectively independent samples with less computational effort. The posterior mean maps of the voter Democratic preference (p_i) obtained under each sampling strategy are displayed in Figure 4.8. We see essentially identical results under both strategies.

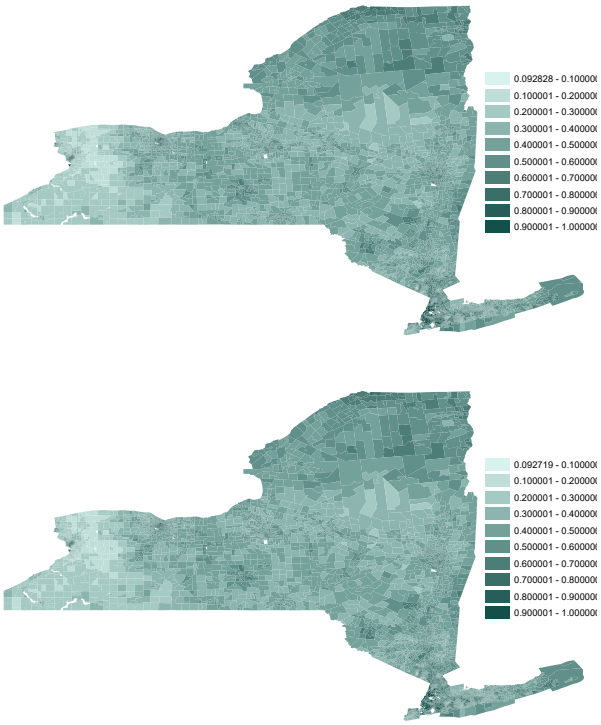


Figure 4.8: Posterior mean maps of voter preference for the Democratic candidate in the binomial election example obtained from the full block Gibbs (top), and chromatic Gibbs (bottom) sampling.

4.3.3 Summary

These numerical experiments illustrate potential improvements that chromatic Gibbs sampling can offer versus the two most common strategies of block sampling via sparse linear algebra and single-site sampling via local characteristics. In simulated image reconstruction, we find that for every considered resolution, the chromatic sampler is computationally much cheaper than the full block Gibbs and single-site samplers. As the dimension increases, we do observe differences between chromatic and block sampling with respect to sampling the GMRF precision parameter and intercept term. This is partly due to the correlation between the random field and the precision parameter that is known to occur as the dimension of the GMRF increases [Rue and Held, 2005a, Agapiou et al., 2014]. Even in this case, however, both approaches are able to estimate the posterior mean of the latent field and obtain equivalent recovery of the quantity of interest. The chromatic sampler is able to do so much more quickly and with much less memory overhead, the latter of which allows the chromatic sampler to scale to images of extremely large dimension beyond the capability of standard

Cholesky factorization routines available in R. The potential advantages extend to irregular arrays and non-Gaussian data, as demonstrated in the election data example.

4.4 Discussion

Over the last twenty years, Gaussian Markov random fields have seen a dramatic increase in popularity in the applied Bayesian community. In this work, we discussed approaches for simulating from Gaussian Markov random fields that are commonly used in practice. We compared the two dominant approaches in the statistics literature, single-site and block updating, to chromatic Gibbs sampling. Each procedure has theoretical guarantees, but our criteria have been pragmatic; i.e., how can statisticians effectively lower the computational cost of sampling from the target distribution without resorting to esoteric knowledge from graph theory, numerical analysis, or parallel programming? Taking this view, we have shown that chromatic sampling is competitive with and often able to improve upon single-site and full block Gibbs.

Motivated by large-scale clinical imaging data, we illustrated potential advantages on a regular array with Gaussian response, finding that chromatic sampling scales to settings where memory limitations prevent direct sparse matrix manipulations. We also considered a real example with binomial election data on an irregular lattice with almost 15,000 areal units, showing that chromatic sampling is useful even without a provably optimal coloring of the MRF graph. Both block sampling and chromatic sampling tend to be far superior to single-site sampling when one is trying to avoid direct iterations in R.

While facilitating parallel or vectorized simultaneous updates, each individual draw under chromatic sampling is still at the level of a single site. Thus, for variables that are highly correlated in the target distribution, convergence can be slow. To handle this, Gonzalez et al. [2011a] propose also a "splash sampler" to combine the blocking principle of updating sets of correlated variables together with the parallelizability afforded by graph colorings. Splash sampling is more involved than simple chromatic sampling. It requires careful construction of undirected acyclic graphs subject to a known tree width determined by individual processor limitations, and hence much more computing effort and more familiarity with graph theory. In the Gaussian case, splash sampling would require repeated Cholesky factorizations, each on matrices of smaller dimension, but without being able to save the sparsity structure. In the presence of highly correlated variables in the target distribution

with GMRF updates, it might be preferable to use ordinary block updates with sparse matrix algebra and the algorithms suggested by Rue [2001]. However, our numerical experiments demonstrate that the gain in computational efficiency from the simple chromatic sampler can still outweigh the loss of statistical efficiency. This leads to an overall improvement in a variety of situations without resorting to more sophisticated approaches that might be inaccessible to most statisticians.

In this chapter, we examined the performance of chromatic sampling versus single-site and block Gibbs on high density data in which the entire study region is sufficiently sampled and in which a first-order Markov neighborhood can capture the salient features of the data. This situation is applicable to many, but not all, analyses of areal data. There remains the issue of how chromatic sampling would perform in the presence of sparse observations from an underlying smooth process, where the autocorrelation of the Markov chain would be expected to be higher than in the high-density case. We leave this question for future work.

Given the current trajectory of modern data analysis, the utility of GMRFs is not likely to diminish anytime soon. However, with their use comes the need for efficient yet accessible sampling strategies to facilitate Bayesian posterior inference along with appropriate measures of uncertainty. This area remains an active area of research among statisticians, computer scientists, and applied mathematicians. Fortunately, the increasingly interdisciplinary environment within which researchers are operating today makes it more likely that significant advancements will be widely disseminated and understood by researchers from a wide variety of backgrounds. This is no doubt a promising trend which will ultimately benefit the broader scientific community as a whole.

Supplementary Material

Appendix A contains additional figures referred to in the text and details concerning the posterior sampling algorithm used in the election example.

Chapter 5

A Large Scale Spatio-Temporal Binomial Regression Model for Estimating Seroprevalence Trends

5.1 Introduction

Lyme disease is a vector-borne disease that impacts both humans and several other mammalian species, with domestic dogs often contracting the infection [Little et al., 2010a]. Disease results via infection by *Borrelia burgdorferi*, a spirochetal bacteria that is transmitted by ticks. In the United States, *B. burgdorferi* is the only known etiologic agent of Lyme disease in dogs; other species can cause Lyme in Europe and Asia. Lyme incidence in humans is considered emerging, with a growing number of high incidence counties being reported [Adams, 2017]. Humans and dogs are infected by the same vectors [Little et al., 2010a]; hence, the risks of exposure for both are related. In fact, dogs are viewed as sentinels for regional Lyme risk in humans [Mead et al., 2011c]. The effects of Lyme disease are usually more severe in humans, often debilitating.

Dogs are regularly tested for exposure to *B. burgdorferi* as part of annual wellness examinations. Commonly, veterinarians use a serologic test that detects antibodies against the C6 peptide that is present in the blood of Lyme-exposed animals. The presence of C6 is indicative of an intermediate or late-term infection and is often detectable 3 to 6 weeks after exposure [Wagner et al.,

2012b]. Among dogs that are infected, only about 5% develop any clinical signs of Lyme disease [Levy and Magnarelli, 1992]. The remaining dogs may either clear the infection without developing the disease, or are subclinically infected and never show symptoms. The routine Lyme testing done on dogs provides an opportunity to measure the proportion of exposed dogs in the relatively healthy canine population that visits veterinary clinics and is tested. This proportion is a prevalence based on serologic (blood) tests and is therefore called a seroprevalence (henceforth shortened to prevalence). In this population, the national prevalence is about 6%, but county Lyme prevalence can exceed 40% in some cases.

Monitoring prevalence is useful for many reasons. Prevalence rates indicate exposure risk within a region, allowing veterinarians to provide effective preventative care and make testing recommendations. Indirectly, prevalence can help identify the range of *Ixodes* spp. tick vectors. This is important because *Ixodes* spp. also transmit other pathogens, including *Anaplasma* spp., *Ehrlichia muris eauclairensis* and *Babesia microti* [Nelder et al., 2016], several of which are also zoonotic. The shared tick vector and extensive Lyme data for dogs makes them a good sentinel for quantifying human Lyme risk. In short, trends in dog prevalence should aid our understanding of Lyme risk changes for humans.

The goal of this chapter is to identify US regions that are experiencing increasing dog Lyme prevalence. Our data are 16,571,562 serologic *B. burgdorferi* tests conducted on domestic dogs in the conterminous United States (US) from January 2012 - December 2016, aggregated by county and month. Figure 5.1 displays the raw prevalence estimates (the proportion of positive tests) after aggregating over all sixty months in the study. Data were reported from 69,876 county-month pairs. To locate where prevalence is increasing, our model needs to have a spatially-varying temporal trend component. To make reliable inferences, the strong positive spatio-temporal dependence of the tests needs to be taken into account. The size of this data set and its large spatio-temporal support (about 3100 distinct counties or county-equivalent regions).

Gaussian processes (GPs) are popular geostatistical modeling tools due to their flexibility and ability to quantify uncertainty in nonparametric regressions [O'Hagan, 1978],[Neal, 1998]. Good GP modeling overviews are Cressie [1993], Rasmussen and Williams [2006], Cressie and Wikle [2011a], and Gelfand and Schliep [2016]. Banerjee et al. [2015a] discuss Bayesian aspects of GPs. Objective prior specification for GP models is studied in Berger et al. [2001]. GPs have become standard tools in a wide variety of applications, including oceanography [Jona-Lasinio et al., 2012],

water quality analysis [Zhang and El-Shaarawi, 2009], image classification [Morales-Álvarez et al., 2017], neuroimaging [Lazar, 2008], and computer experiments [Santner et al., 2003]. GPs have also been previously used to model disease prevalence, including dengue fever [Johnson et al., 2017], malaria [Andrade-Pacheco et al., 2015], and influenza [Senanayake et al., 2016]. Gelfand et al. [2003] allowed GP linear model coefficients to vary smoothly over space, an approach used here to allow for regional prevalence trends.

GP modifications and algorithms for analyzing big spatial data have received significant recent attention, including fixed rank kriging [Cressie and Johannesson, 2008] and LatticeKrig [Nychka et al., 2015]. Both methods employ basis function expansions of spatial random effects to reduce the dimensions of the covariance matrices in the model. Katzfuss [2017] takes a similar approach, applying basis functions to a succession of refined resolutions. Spatial partitioning [e.g., Sang et al., 2011, Heaton et al., 2017a] can be used to split regions into smaller, more manageable sub-regions with computation being accelerated via a conditional independence assumption. Covariance tapering [Furrer et al., 2006] uses a covariance structure with compact support to induce sparsity. Nearest neighbor processes [Datta et al., 2016a] achieve computational efficiency by conditioning on a subset of nearby observations. A similar idea is used by Gramacy and Apley [2015] to find the largest number of neighbors that are computationally feasible for prediction, optimally chosen by minimizing a prediction variance. Heaton et al. [2017b] provide an overview and comparison of these and other procedures. Our approach involves Gaussian predictive processes (GPPs) [GPPs; Banerjee et al., 2008] and is discussed further in Section 5.2.

The most common approach for modeling spatially dependent areal data involves Gaussian Markov random fields [GMRFs; Rue and Held, 2005b], with Gaussian conditional autoregressive (CAR) models [Banerjee et al., 2015a] being particularly popular. As special cases of Markov random fields [Besag, 1974a], GMRFs are collections of jointly distributed Gaussian variables satisfying a Markov dependence structure quantified through a precision matrix. GMRFs are extended to flexible degrees of smoothness in Brezger et al. [2007] and Yue and Speckman [2010]. Brown et al. [2017a] adjust the CAR precision matrix to build a unified model for independent and dependent cases and study neighborhood structures other than those based on physical adjacency. GMRF and GP connections are explored in Rue and Tjemland [2002], Song et al. [2008], and Lindgren et al. [2011b]. CAR models are by now standard in disease mapping problems [e.g., Waller et al., 1997b].

To achieve our goals, a large scale spatio-temporal binomial regression model is developed

that has both GPP and CAR components. The former is used to capture regionally varying trends by treating the trend coefficient as a non-parametric surface over space, while the latter accounts for local heterogeneity. Through data augmentation steps and a novel sampling strategy, a modeling framework is developed that is computationally scalable to large non-Gaussian spatio-temporal data sets. In particular, straightforward Gibbs sampling is facilitated via a data augmentation step involving latent Pólya-Gamma variables. To avoid computationally expensive matrix calculations, a chromatic sampling strategy is used in our Gibbs sampler. Our methodology easily handles missing data. The finite sample properties of our approach are studied via simulation before our Lyme analysis is conducted.

The remainder of this chapter is organized as follows: Section 5.2 describes the model and our GPP and CAR structures. Section 5.3 discusses model fitting procedures, emphasizing computational tractability with large spatio-temporal data. Section 5.4 presents a simulation study supporting our approach and Section 5.5 analyzes the canine serology data described above. Concluding remarks are offered in Section 5.6.

5.2 Modeling Methods

Let Y_{st} denote the number of cases (e.g., positive *B. burgdorferi* tests) observed in n_{st} tests taken in region s at time t, for $s = 1, \ldots, S$ and $t = 1, \ldots, T$. Set $\boldsymbol{Y}_s = (Y_{s1}, \ldots, Y_{sT})'$, $\boldsymbol{Y} = (\boldsymbol{Y}_1', \ldots, \boldsymbol{Y}_S')' \in \mathbb{R}^{ST}$, $\boldsymbol{n}_s = (n_{s1}, \ldots, n_{sT})'$, and $\boldsymbol{n} = (\boldsymbol{n}_1', \ldots, \boldsymbol{n}_S')' \in \mathbb{N}^{ST}$. In addition to the disease tests, the covariates Z_{stq} and X_{stp}, for $q = 1, \ldots, Q$ and $p = 1, \ldots, P$, are assumed available in region s at time t. The Z_{stq} are covariates whose effects are constant over the study area, while X_{stq} are covariates whose associated effects vary by region.

To relate the observed test data to the covariates, a Bayesian generalized linear mixed model [McCullagh and Nelder, 1989, Diggle et al., 1998, Banerjee et al., 2015a] is adopted. Our general model is the binomial regression: $Y_{st}|n_{st}, p_{st} \sim \text{Binomial}(n_{st}, p_{st})$ with

$$\nu_{st} := g^{-1}(p_{st}) = \boldsymbol{Z}_{st}'\boldsymbol{\delta} + \boldsymbol{X}_{st}'\boldsymbol{\beta}(\boldsymbol{\ell}_s) + \xi_{st}; \quad s = 1, \ldots, S; \; t = 1, \ldots, T, \tag{5.1}$$

where $g : \mathbb{R} \to (0, 1)$ is a known link function (e.g., logistic) relating the linear predictor ν_{st} to the prevalence p_{st}, $\boldsymbol{Z}_{st} = (1, Z_{st1}, \ldots, Z_{stQ})' \in \mathbb{R}^{Q+1}$, $\boldsymbol{X}_{st} = (X_{st1}, \ldots, X_{stP})' \in \mathbb{R}^P$, $\boldsymbol{\delta} = (\delta_0, \ldots, \delta_Q)'$

are global regression coefficients, $\boldsymbol{\beta}(\cdot) = (\beta_1(\cdot), \ldots, \beta_P(\cdot))'$ are spatially varying regression coefficients, $\boldsymbol{\ell}_s = (\ell_{s1}, \ell_{s2})'$ is a vector of spatial coordinates (e.g., latitude and longitude) that identifies the centroid of region s, and ξ_{st} is a spatio-temporal random effect. Following Gelfand et al. [2003], the spatially varying regression coefficients are regarded as unknown smooth surfaces over the study region. To model these unknown surfaces while maintaining computational tractability, GPPs are used.

A GP is a stochastic process whose finite dimensional distributions are multivariate normal. A GP $\boldsymbol{\beta}_p(\cdot)$, given a covariance parameter $\boldsymbol{\theta}_p$ and denoted

$$\boldsymbol{\beta}_p \mid \boldsymbol{\theta}_p \sim \mathcal{GP}(\mu_p(\cdot), C(\cdot, \cdot; \boldsymbol{\theta}_p)),$$

is uniquely determined by its mean and covariance, $\mu_p(\boldsymbol{\ell}_s) := E[\beta_p(\boldsymbol{\ell}_s)]$ and $C(\boldsymbol{\ell}_s, \boldsymbol{\ell}_{s'}; \boldsymbol{\theta}_p) := \mathrm{Cov}(\beta_p(\boldsymbol{\ell}_s), \beta_p(\boldsymbol{\ell}_{s'})) = \sigma_p^2 \rho_p(\boldsymbol{\ell}_s, \boldsymbol{\ell}_{s'}; \boldsymbol{\theta}_p)$, where $\rho_p(\cdot, \cdot; \boldsymbol{\theta}_p)$ is a correlation function depending on $\boldsymbol{\theta}_p$. For smoothing and interpolation, a constant mean is often assumed [Bayarri et al., 2007]. Our work *a priori* posits that $\mu_p(\cdot) \equiv 0$ for all p. Thus, $\boldsymbol{\beta}_p = (\beta_p(\boldsymbol{\ell}_1), \ldots, \beta_p(\boldsymbol{\ell}_S))'$, $S \in \mathbb{N}$, follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{C}_p = \sigma_p^2 \boldsymbol{R}_p$, where $(\boldsymbol{R}_p)_{ss'} = \rho_p(\boldsymbol{\ell}_s, \boldsymbol{\ell}_{s'}; \boldsymbol{\theta}_p)$. In general, the covariance matrix inversions and factorizations needed to calculate quantities in our posterior distributions are $\mathcal{O}(S^3)$ in computational time. In Markov chain Monte Carlo algorithms, these operations will need to be repeated thousands of times. Thus, as S grows large, GPs quickly become computationally unwieldy.

To reduce the dimension of the problem, our GPP employs a "parent" process based on a strategically chosen set of knots and interpolates to points of interest via kriging. Let $\{\boldsymbol{\ell}_1^*, \ldots, \boldsymbol{\ell}_{S_p^*}^*\}$ denote the knot set with $S_p^* \ll S$. Define $\boldsymbol{\beta}_p^* = (\beta_p(\boldsymbol{\ell}_1^*), \ldots, \beta_p(\boldsymbol{\ell}_{S_p^*}^*))'$ and note that $\boldsymbol{\beta}_p^* | \sigma_p^2, \boldsymbol{\theta}_p \overset{ind}{\sim} \mathrm{N}(\mathbf{0}, \boldsymbol{C}_p^*)$, for all p, where $\boldsymbol{C}_p^* = \sigma_p^2 \boldsymbol{R}_p^*$ and $(\boldsymbol{R}_p^*)_{ss'} = \rho_p(\boldsymbol{\ell}_s^*, \boldsymbol{\ell}_{s'}^*; \boldsymbol{\theta}_p)$. The GPP simply replaces $\boldsymbol{\beta}_p$ with $\tilde{\boldsymbol{\beta}}_p := E(\boldsymbol{\beta}_p | \boldsymbol{\beta}_p^*; \boldsymbol{\theta}_p) = \tilde{\mathbf{R}}_p^* (\mathbf{R}_p^*)^{-1} \boldsymbol{\beta}_p^*$, where $\tilde{\mathbf{R}}_p^*$ is an $S \times S_p^*$ matrix whose (s, s')th element is $\rho_p(\boldsymbol{\ell}_s, \boldsymbol{\ell}_{s'}^*; \boldsymbol{\theta}_p)$. When S_p^* is not large, $(\mathbf{R}_p^*)^{-1}$ can be quickly computed. For more on GPPs, see Banerjee et al. [2008].

Fully specifying a GPP requires specifying its knot locations. Banerjee et al. [2008] discuss several methods of knot selection, including placing them on a regular grid, selecting them at random from the observation locations, and methods that place more knots in areas with more observations. Finley et al. [2009] suggest choosing knot locations to minimize conditional variances at observation

locations. Guhaniyogi et al. [2011] propose an adaptive knot selection strategy where knot locations are treated as a point process. Following Eidsvik et al. [2012], our knots are chosen via K-means clustering with S_p^* clusters; i.e., using K-means clustering, the S counties are partitioned into S_p^* clusters based on their locations $\boldsymbol{\ell}_s$. The knot locations are taken as the centroids of the S_p^* clusters. For further details on K-means clustering, see Hartigan and Wong [1979].

A variety of ways exist to model spatio-temporal dependence of areal data. Some commonly used methods include intrinsic CAR models, proper CAR models, and the so-called Besag-York-Mollie (BYM) models for disease mapping problems. These methods can be extended to handle spatio-temporal correlation in many ways. For examples, a separate BYM model can be fit at each time point [Waller et al., 1997b], or a CAR model can be combined with a spline-based temporal structure [MacNab and Dean, 2001], or a local autoregressive model in time [Congdon and Southall, 2005]. For a thorough review and comparison of existing spatio-temporal models, see Anderson and Ryan [2017]. A county-by-county exploratory analysis of our Lyme data suggests that a first order autoregressive model is sufficient for handling temporal dependence. Thus, following Rushworth et al. [2014] and Lee and Lawson [2014], a first-order vector autoregression is used with GMRF errors:

$$\boldsymbol{\xi}_t = \zeta \boldsymbol{\xi}_{t-1} + \boldsymbol{\phi}_t, \tag{5.2}$$

where $\boldsymbol{\xi}_t = (\xi_{1t}, ...\xi_{St})'$, $\zeta \in (-1, 1)$ is a parameter controlling temporal correlation, and $\boldsymbol{\xi}_0 = \mathbf{0}$ is taken as a starting condition. We assume the $\boldsymbol{\phi}_t$s to be independent and identically distributed as a proper intrinsically autoregressive model [Besag and Kooperberg, 1995a]; i.e., $\boldsymbol{\phi}_t \sim \mathrm{N}\left(\mathbf{0}, \tau^2(\boldsymbol{D} - \omega \boldsymbol{W})^{-1}\right)$, where $\tau^2 > 0$ and $\omega \in (0, 1)$ is a so-called "propriety parameter" that ensures the precision matrix is non-singular [Banerjee et al., 2015a]. The neighborhood matrix $\boldsymbol{W} \in \mathbb{R}^{S \times S}$ is such that $(\boldsymbol{W})_{ss'}$ is equal to 1 if and only if location s is adjacent to location s', $s \neq s'$, zero otherwise, and

$$\boldsymbol{D} = \mathrm{diag}\left(\sum_{j=1}^{S} (\boldsymbol{W})_{sj}, s = 1, \ldots, S\right).$$

To avoid confounding with the intercept, the standard sum-to-zero constraint

$$\sum_{t=1}^{T} \sum_{s=1}^{S} \xi_{st} = 0,$$

is imposed.

The proposed model is completed by specifying prior distributions on the regression coefficients and the variance and correlation parameters. In the absence of strong prior information, hyperparameters are chosen to induce vague prior distributions. A Gaussian prior is assumed for the global regression coefficients, and inverse Gamma (IG) priors are placed on variance components. A truncated Gaussian prior with support $(-1, 1)$ is specified for ζ. A Beta$(\alpha_\omega, \upsilon_\omega)$ prior is placed on ω to concentrate it close to unity, since previous empirical work has shown that $\omega \approx 1$ is necessary to induce noticeable spatial association [Banerjee et al., 2015a]. These specifications lead to the following hierarchy:

$$
\begin{aligned}
Y_{st}|n_{st}, \nu_{st} &\overset{indep.}{\sim} \text{Binomial}\left(n_{st}, p_{st} = g(\nu_{st})\right), \quad s = 1, \ldots, S;\ t = 1, \ldots, T; \\
\boldsymbol{\beta}^*_p|\sigma^2_p, \boldsymbol{\theta}_p &\overset{indep.}{\sim} \text{N}(\mathbf{0}, \sigma^2_p \boldsymbol{R}^*_p(\boldsymbol{\theta}_p)), \quad p = 1, \ldots, P; \\
\sigma^2_p &\overset{indep.}{\sim} \text{IG}(\alpha_{\sigma^2_p}, \upsilon_{\sigma^2_p}), \quad p = 1, \ldots, P; \\
\boldsymbol{\theta}_p &\overset{i.i.d.}{\sim} \pi(\boldsymbol{\theta}_p), \quad p = 1, \ldots, P; \\
\boldsymbol{\delta} &\sim \text{N}(\mathbf{0}, \sigma^2_\delta \boldsymbol{I}), \quad \sigma^2_\delta > 0; \\
\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1}, \tau^2, \omega, \zeta &\sim \text{N}\left(\zeta\boldsymbol{\xi}_{t-1}, \tau^2(\boldsymbol{D} - \omega\boldsymbol{W})^{-1}\right), \quad t = 1, \ldots, T; \\
\tau^2 &\sim \text{IG}(\alpha_{\tau^2}, \upsilon_{\tau^2}), \quad \alpha_{\tau^2}, \upsilon_{\tau^2} > 0; \\
\omega &\sim \text{Beta}(\alpha_\omega, \upsilon_\omega), \quad \alpha_\omega, \upsilon_\omega > 0; \\
\zeta &\sim \text{Truncated-Normal}(0, \sigma^2_\zeta, -1, 1), \quad \sigma^2_\zeta > 0,
\end{aligned}
\tag{5.3}
$$

where $\nu_{st} = \boldsymbol{Z}'_{st}\boldsymbol{\delta} + \boldsymbol{X}'_{st}\widetilde{\boldsymbol{\beta}}(\boldsymbol{\ell}_s) + \xi_{st}$, $\widetilde{\boldsymbol{\beta}}(\boldsymbol{\ell}_s) = (\widetilde{\beta_1}(\boldsymbol{\ell}_s), ..., \widetilde{\beta}_P(\boldsymbol{\ell}_s))'$, and $\boldsymbol{\xi}_0 = \mathbf{0}$. Each coefficient in $\widetilde{\boldsymbol{\beta}}(\boldsymbol{\ell}_s)$ is obtained from the P predictive processes via $\widetilde{\boldsymbol{\beta}}_p = \widetilde{\boldsymbol{R}}^*_p(\boldsymbol{R}^*_p)^{-1}\boldsymbol{\beta}^*_p$. Appropriate (identical) priors for $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_P$ depend on the correlation function selected in the GPP model.

While the combination of a continuous support GPP and a discrete support GMRF has not been extensively used previously, it is motivated in our application. Since the $\widetilde{\boldsymbol{\beta}}_\ell$ coefficients contain trends that are thought to vary smoothly over space, we appeal to GP models since an explicit covariance function allows for the direct imposition of smoothness assumptions and a meaningful prediction function through kriging — the latter is useful for estimating seroprevalence trends at unobserved locations. The spatio-temporal random effects, on the other hand, are of secondary interest and serve only to smooth the extra-regression variability beyond that explained by the

predictors. Since they are defined over an areal lattice, a CAR model is a natural choice. In Appendix B, we consider replacing the GPP with a CAR model on the regression coefficients and empirically study the results via simulation. We find that the GPP model is able to produce reliable estimates of both "regional" and "local" trends, while the CAR model only estimates local trends. Since both models reliably estimate local trends and the GPP model can also accurately estimate regional trends, the GPP model is used. For further discussion of local and regional trends, including the differences between them, see Section 5.5 and Appendix B.

5.3 Posterior Sampling

5.3.1 Data Augmentation

We assume conditional independence given the covariate effects and spatio-temporal effects and observe that $\boldsymbol{Y}$ depends on the regression coefficients and random effects only through $\boldsymbol{\nu} = (\nu_{11}, \ldots, \nu_{1T}, \nu_{21}, \ldots, \nu_{ST})'$. Hence, the likelihood is

$$f(\boldsymbol{Y}|\boldsymbol{\nu}) \propto \prod_{t=1}^{T} \prod_{s=1}^{S} g(\nu_{st})^{Y_{st}} \{1 - g(\nu_{st})\}^{n_{st} - Y_{st}}. \tag{5.4}$$

To develop a posterior sampling algorithm, let $g(\cdot)$ be the logistic link. Other link functions are possible and can be implemented following Albert and Chib [1993] or Gamerman [1997]. Metropolis-Hastings steps [Metropolis et al., 1953, Hastings, 1970] can be used either component-wise or in blocks, but such samplers can be difficult to tune in high dimensions. To facilitate the derivation of a Gibbs sampler for the regression coefficients and spatio-temporal random effects, a data augmentation scheme is used that leads to sampling these parameters from Gaussian full conditional distributions.

Our data augmentation approach follows Polson et al. [2013] and relies on the fact that $\exp(\nu)^a\{1 + \exp(\nu)\}^{-b} = 2^{-b}\exp(\kappa\nu)\int_0^\infty \exp(-\psi\nu^2/2)\ p(\psi|b, 0)d\psi$, where $a \in \mathbb{R}$, $b \in \mathbb{R}^+$, $\kappa = a - b/2$, and $p(\cdot \mid b, 0)$ is the probability density function of a Pólya-Gamma random variable with

parameters b and 0. From these, under the logistic link, (5.4) can be written as

$$
\begin{aligned}
f(\boldsymbol{Y}|\boldsymbol{\nu}) &\propto \prod_{t=1}^{T}\prod_{s=1}^{S} \exp(\kappa_{st}\nu_{st}) \int_0^\infty \exp(-\psi_{st}\nu_{st}^2/2)p(\psi_{st}|n_{st},0)d\psi_{st} \\
&\propto \prod_{t=1}^{T}\prod_{s=1}^{S} \int_0^\infty f_{Y,\psi}(Y_{st},\psi_{st} \mid \nu_{st})d\psi_{st},
\end{aligned}
$$

where $\kappa_{st} = Y_{st} - n_{st}/2$ and $f_{Y,\psi}$ is the joint density of (Y_{st}, ψ_{st}). By introducing the ψ_{st} as latent random variables to be sampled via MCMC, we obtain

$$
f_{\boldsymbol{Y},\boldsymbol{\psi}}(\boldsymbol{Y},\boldsymbol{\psi} \mid \boldsymbol{\nu}) \propto \exp(-\boldsymbol{\nu}'\boldsymbol{D}_{\boldsymbol{\psi}}\boldsymbol{\nu}/2 + \boldsymbol{\kappa}'\boldsymbol{\nu}) \prod_{t=1}^{T}\prod_{s=1}^{S} p(\psi_{st}|n_{st},0),
$$

where $\boldsymbol{\psi} = (\psi_{11},\ldots,\psi_{1T},\psi_{21},\ldots,\psi_{ST})'$, $\boldsymbol{D}_{\boldsymbol{\psi}} = \text{diag}(\boldsymbol{\psi})$, and $\boldsymbol{\kappa} = (\kappa_{11},\ldots,\kappa_{1T},\kappa_{21},\ldots,\kappa_{ST})'$. Hence, data augmentation yields a Gaussian density in $\boldsymbol{\nu}$ up to a normalizing constant. Consequently, the full conditional distributions for most parameters take a known form and are easy to sample. For specifics, the full conditional distribution of ψ_{st} is Pólya-Gamma, $\boldsymbol{\beta}_p^*$ is multivariate normal, $\boldsymbol{\delta}$ is multivariate normal, σ_p^2 is inverse gamma, τ^2 is inverse gamma, and ζ is truncated normal. Appendix B provides additional conditional distributions.

From the data augmentation, a posterior sampling algorithm involving Gibbs steps for the above parameters can be constructed in the usual manner. Metropolis-Hastings steps are used to sample each $\boldsymbol{\theta}_\ell$ and ω. While the full conditional distribution of $\boldsymbol{\xi}_t$ is multivariate normal, sampling from this high dimensional distribution is computationally expensive. For more efficient repeated updates of $\boldsymbol{\xi}_t$, the Markov structure of the CAR model is exploited to construct a chromatic sampler that updates conditionally independent blocks of $\boldsymbol{\xi}_t$ in parallel. For further discussion, see Gonzalez et al. [2011b] and Brown et al. [2017c].

5.3.2 A Note on Missing Data

In our application, data are not reported at all county-month pairs. To account for this, let $\mathcal{R}$ be the set of all ordered pairs (s,t) for which tests are observed. The augmented likelihood is

$$
f(\boldsymbol{Y}(\mathcal{R}),\boldsymbol{\psi}(\mathcal{R}) \mid \boldsymbol{\nu}(\mathcal{R})) \propto \exp(-\boldsymbol{\nu}(\mathcal{R})'\boldsymbol{D}_{\boldsymbol{\psi}(\mathcal{R})}\boldsymbol{\nu}(\mathcal{R})/2 + \boldsymbol{\kappa}(\mathcal{R})'\boldsymbol{\nu}(\mathcal{R})) \prod_{(s,t)\in\mathcal{R}} p(\psi_{st}|n_{st},0),
$$

where $\boldsymbol{\nu}(\mathcal{R}) = \boldsymbol{Z}(\mathcal{R})\boldsymbol{\delta} + \boldsymbol{X}(\mathcal{R})\widetilde{\boldsymbol{b}} + \boldsymbol{I}(\mathcal{R})\boldsymbol{\xi}$ and the convention that $\boldsymbol{A}(\mathcal{R})$ is the matrix formed by retaining the rows of $\boldsymbol{A}$ whose indices are in $\mathcal{R}$ is used. Here, $\boldsymbol{Z} = (\boldsymbol{Z}_1', \ldots, \boldsymbol{Z}_S')' \in \mathbb{R}^{ST\times(Q+1)}$ with $\boldsymbol{Z}_s = (\boldsymbol{Z}_{s1}, \ldots, \boldsymbol{Z}_{sT})'$. Similarly, $\boldsymbol{X} = \bigoplus_{s=1}^{S} \boldsymbol{X}_s \in \mathbb{R}^{ST\times SP}$ with $\boldsymbol{X}_s = (\boldsymbol{X}_{s1}, \ldots, \boldsymbol{X}_{sT})'$, $\boldsymbol{I}$ is the identity matrix, and $\widetilde{\boldsymbol{b}} = (\widetilde{\boldsymbol{\beta}}'(\boldsymbol{\ell}_1), \ldots, \widetilde{\boldsymbol{\beta}}'(\boldsymbol{\ell}_S))' \in \mathbb{R}^{SP}$. Since $\boldsymbol{\xi} \in \mathbb{R}^{ST}$ is the vector of spatial random effects over *all* locations within the study region for all time points, a well-defined full conditional distribution for $\boldsymbol{\xi}$ is obtained, provided that the prior on $\boldsymbol{\xi}$ is proper. This joint density representation permits the imputation of any missing effects via posterior realizations.

5.4 A Simulation Study

This section studies via simulation how well our methods estimate model coefficients and how GPP knot selection influences results. Data were generated on a regularly spaced $S \times S$ grid over 60 time points, where $S = 13$, and then drawing $Y_{st}|n_{st}, p_{st} \overset{indep.}{\sim}$ Binomial(n_{st}, p_{st}) observations, where

$$g^{-1}(p_{st}) = \delta_0 + \widetilde{\beta_1}(\boldsymbol{\ell}_s)t/60 + \xi_{st}, \quad s = 1, \ldots, S^2;\ t = 1, \ldots, 60,$$

and $g(\cdot)$ is the logistic link. The test counts n_{st} were randomly sampled from a discrete uniform distribution ranging from 100 to 200. The random effects ξ_{st} are generated from the CAR model defined in Section 5.2 with $\zeta = 0.9$, $\tau^2 = 0.005$, $\omega \in \{0.00, 0.55, 0.90\}$, and a neighborhood matrix $\boldsymbol{W}$ set so that two areas are neighbors if and only if they share a common edge or corner. The ω values 0.00, 0.55, and 0.90 correspond to no, weak, and strong spatial dependence, respectively. The true intercept was set to $\delta_0 = -1$ and the surface $\widetilde{\boldsymbol{\beta}}_1(\cdot)$ at each study location is generated from the GPP model in (1). Specifically, a realization of the parent process is first simulated on a 5×5 grid of equally spaced knots. The parent process took $\mu_1(\boldsymbol{\ell}_s^*) \equiv 1$ and $\rho(\boldsymbol{\ell}_s^*, \boldsymbol{\ell}_{s'}^*; \theta_1) = \theta_1^{d_{ss'}^2}$, where $d_{ss'}$ is the Euclidean distance between $\boldsymbol{\ell}_s^*$ and $\boldsymbol{\ell}_{s'}^*$, $\theta_1 = 0.6$, and $\sigma_1^2 = 1.5$. The resulting $\widetilde{\boldsymbol{\beta}}_1(\cdot)$ is depicted in Figure 5.2. Using this surface, 500 independent data sets were generated from the model for each ω.

Our model was fitted to each data set using three separate knot set configurations. The first configuration uses the same knots as those generating the true surface, representing an ideal situation. The other two configurations take 4×4 and 7×7 grids of equally spaced knots. For priors in (1), we take $\alpha_{\sigma_1^2} = \upsilon_{\sigma_1^2} = \alpha_{\tau^2} = \upsilon_{\tau^2} = 2$, $\sigma_\delta^2 = 1000$, $\alpha_\omega = 900$, $\upsilon_\omega = 100$, and $\sigma_\zeta^2 = 10$.

In the GPP, the correlation function was taken as $\rho(\boldsymbol{\ell}_s, \boldsymbol{\ell}_{s'}; \theta_1) = \theta_1^{d^2_{s,s'}}$, the same as the true GPP. A Uniform$(0, 1)$ prior on θ_1 was used. For each data set, 5,000 MCMC iterates are retained after a burn-in of 5,000 samples. Convergence of the chains were assessed via trace plots and judged acceptable.

Figure 5.3 summarizes our results for the temporal trend parameter $\widetilde{\boldsymbol{\beta}}_1(\cdot)$ when $\omega = 0.90$. This includes a spatial depiction of the arithmetic average of the 500 point estimates, as well as empirical biases and mean squared errors. Here, for each data set, a point estimate of $\widetilde{\boldsymbol{\beta}}_1(\cdot)$ was obtained as the mean of the 5,000 retained MCMC iterates. Supplementary Figure B.1 summarizes results for the other ω values. The methods estimate the spatially varying regression coefficient well for every considered ω; i.e., the mean estimates show little bias and have a relatively small mean squared error. Estimator variability increases near the region's edges — this boundary effect is expected and is common in non-parametric regressions. Figure 5.3 shows little practical difference for the estimates obtained under the three different knot configurations, suggesting that the methods can recover the true coefficient surface across the entire study region (assuming the model is correct up to choice of knots).

Two additional simulations were conducted. The first simulation examined the performance of the proposed methodology in the presence of missing data. The second study examined the effects of increasing spatial dimension. Results for these are presented in Appendix B. The results indicate that the proposed methodology performs well in these more challenging situations.

5.5 Lyme Analysis

5.5.1 Background

Our data contain 16,571,562 tests on domestic dogs living throughout the conterminous United States from January 2012 - December 2016. The data were provided by IDEXX Laboratories, Inc. to the Companion Animal Parasite Council (CAPC), who made them available online at `https://www.capcvet.org`. The data are aggregated by month and county; 69,876 county-month pairs report at least one test.

In general, the spatial distribution of a vector-borne disease is strongly influenced by regional environments and the vector's hosts (e.g., deer populations), leading to correlated data [Legendre,

1993]. A strong spatial correlation is seen in these data, as indicated by Figure 5.1 and a Moran's I statistic of 0.378 (p-value ≈ 0). Such data are also positively temporally correlated. Figure 5.4 displays raw county-level prevalence estimates aggregated over all 12 months in the two years of 2012 and 2016. A comparison of these graphics suggest where a significant increase in prevalence is expected, including Western Pennsylvania, Virginia, West Virginia, Minnesota, and Iowa.

5.5.2 Model Building and Seasonality

As an exploratory step, a county-by-county time series analysis of prevalence was conducted for 672 counties reporting a sufficient amount of data; i.e., counties reporting 10 or more positive tests each month. Following Dunsmuir and Scott [2015], a binomial generalized linear model with only a linear time trend (on the logit scale) was fitted to each county's time series. The partial autocorrelations of the Pearson residuals were used to assess autoregressive orders in the usual manner. From this analysis, a first-order autoregressive (AR(1)) model was deemed reasonable for most counties. Next, a generalized linear AR(1) model with the mean structure described above was fitted to each county. Histograms of the probability integral transformations were constructed and used to assess suitability of the AR(1) model. These results support an AR(1) model to account for temporal dependence, providing justification for the form taken in (5.2). For more details on this analysis, see Appendix B.

Given the seasonality of tick activity, seasonality could also be present in Lyme prevalence. Though no strong evidence of seasonality surfaced in our exploratory analysis, a more thorough investigation of seasonality was conducted by fitting the model

$$\nu_{st} = \delta_0 + \widetilde{\beta}_1(\boldsymbol{\ell}_s)I_1(t) + \widetilde{\beta}_2(\boldsymbol{\ell}_s)I_2(t) + \widetilde{\beta}_3(\boldsymbol{\ell}_s)I_3(t) + \widetilde{\beta}_4(\boldsymbol{\ell}_s)t + \xi_{st}, \tag{5.5}$$

where t denotes time (rescaled to the unit interval) and $I_p(t)$ is a seasonal indicator for $p = 1, 2, 3$. Seasons are defined as follows: Winter (December-February), Spring (March-May), Summer (June-August), and Fall (September-November), where winter is regarded as the baseline. This model allows for spatially varying seasonal effects and spatially varying trend effects. While covariates such as county level temperatures and precipitations are available, these are not used in this fit since our goal is to quantify trends, not determine the specific drivers of these trends.

Model (5.5) was fitted with the prior specifications and correlation functions described

in Section 5.4. Two specifications for the GPP model were considered, using 50 and 100 knots, respectively. In both cases, knot placement for all GPP models was done by K-means clustering. For sampling, 30,000 MCMC iterates were generated, with the last 10,000 being retained for inference. Convergence of the MCMC chains was assessed using trace plots. We stress the computational scalability of this approach. This model contains four *a priori* independent coefficient surfaces, each replete with 3,109 spatial locations and 186,540 spatio-temporal random effects.

Two primary findings arise. First, there are no appreciable differences between the estimates using 50 and 100 knots. As both specifications are computationally feasible, all subsequent analyses used 100 knots. Second, there is evidence of seasonality in the location parameters, but these appear constant across space. In particular, Figure 5.5 depicts 95% credible intervals for each county level seasonal effect. These intervals all contain a common non-zero value, indicating that a spatially constant seasonal effect is reasonable. Thus, the simpler model

$$\nu_{st} = \delta_0 + \delta_1 I_1(t) + \delta_2 I_2(t) + \delta_3 I_3(t) + \widetilde{\beta}_1(\boldsymbol{\ell}_s)t + \xi_{st}, \tag{5.6}$$

was considered. Credible intervals at level 95% indicate that the model can be further reduced to

$$\nu_{st} = \delta_0 + \delta_1 I_1^*(t) + \widetilde{\beta}_1(\boldsymbol{\ell}_s)t + \xi_{st}, \tag{5.7}$$

where $I_1^*(t)$ is a seasonal indicator that equals one if t is between March and November, and zero otherwise. Approximate 95% credible intervals for δ_0 and δ_1 are $[-3.95, -3.82]$ and $[-0.20, -0.10]$, respectively.

For further insight, the model in (5.7) was compared to the nonseasonal model

$$\nu_{st} = \delta_0 + \widetilde{\beta}_1(\boldsymbol{\ell}_s)t + \xi_{st}. \tag{5.8}$$

For this model, an approximate 95% credible interval for δ_0 is $[-4.08, -4.03]$. Figure 5.6 displays estimates of $\widetilde{\beta}_1(\cdot)$ from both models. Very similar large-scale patterns in the estimated trends are seen. In short, while seasonality exists in the location parameters, its effect on trends seems negligible.

The temporal trend surface $\widetilde{\boldsymbol{\beta}}$ represents a regional effect estimated at a particular area by incorporating information from a relatively large swath of surrounding areas. While regional

trends are useful for estimating trends in areas with few tests, it may be desirable to separate local effects from regional trends to provide a county-level assessment. Appendix B demonstrates how our modeling framework facilitates such a separation. Specifically, county-specific trends are estimated as follows. Let $\upsilon_s^{(g)}$ be the least-squares estimate of the county s slope obtained at by fitting a simple linear regression to $\{(t, \nu_{st}^{(g)}) : \ t = 1, ..., T\}$, where $\nu_{st}^{(g)}$ is the gth posterior draw of ν_{st} obtained from the MCMC output. Then $\upsilon_s^{(g)}$ can be regarded as a realization of the linear time trend at county s. Using the $\{\upsilon_s^{(g)}\}$ as a random sample from the marginal posterior distribution of υ_s, point estimates and inferences can be obtained for county-level trends.

5.5.3 Results

Figure 5.6 displays the estimated posterior mean of the regional temporal trend surface $\widetilde{\boldsymbol{\beta}}_1$. The regional rate of Lyme prevalence change between January 2012 and December 2016 is positive in all states that are currently recognized as having high human Lyme disease incidence [Centers for Disease Control and Prevention, 2017], including portions of the Northeast and the Upper Midwest. The rate of increase varies by region, with high incidence regions generally exhibiting the greatest changes. These regions include Maine, West Virginia, Virginia, and the northern parts of Minnesota and Wisconsin.

Figure 5.7 displays estimated posterior means of the county-level trends $\upsilon_s, \ s = 1, \ldots, 3109$. Figure 5.8 shows counties where local trends are significantly positive, assessed using approximate 95% equal-tailed credible intervals. Increasing local trends are seen in much of the Northeast, extending southwards through West Virginia and Virginia, and into North Carolina and Tennessee. This region includes localities where Lyme disease is reportedly increasing. Increasing local trends in parts of northwestern Minnesota, northern Wisconsin, and southeastern Iowa are also apparent. In the Great Lakes region, increasing trends are observed in Eastern Ohio, Indiana, and Western Michigan. In much of eastern New England, where human Lyme was first documented, the prevalence appears stable, albeit high.

5.6 Discussion

This chapter developed a computationally feasible binomial regression model for large spatio-temporal data that can identify localized trends. Our novel approach combined several recent advances in

large-scale spatial modeling and MCMC sampling. The end product is a flexible, scalable methodology for modern spatio-temporally referenced count data.

Our proposed approach was used to identify regions of the US experiencing increasing canine Lyme risk. Since human and canine risk are similar, such regions are likely also experiencing increasing human exposure. While human Lyme disease data may not be publicly available and, in many regions, scarce due to lack of testing, our canine prevalence data had over 16 million tests. The size of the spatial domain created computational challenges. While monthly and county-level aggregation reduced the size of the response vector from 16,581,562 tests to 69,876 county-month pairs, a binomial response in an MCMC context typically requires sampling via Metropolis-Hastings steps, which can be difficult to tune in extremely high dimensions (over 180,000, in our case). Under the logistic link, a recently proposed Polyá-Gamma data augmentation was used to facilitate direct Gibbs sampling on full conditional distributions. GPPs were used to model smoothly varying, high-dimensional coefficients through a low-dimensional representation. Local spatio-temporal heterogeneity was captured by random effects following a time-varying Gaussian CAR distribution. Chromatic sampling was used to facilitate efficient updating of the GMRFs in our MCMC algorithm.

This study was motivated by the rise in Lyme disease cases in the United States [Adams, 2017] and, in particular, rising incidence in states not traditionally considered Lyme-endemic. Our results suggest that 1) canine prevalence is rising in tandem with human cases [Kugeler et al., 2015, Hendricks and Mark-Carew, 2017, Centers for Disease Control and Prevention, 2017], 2) Lyme prevalence rates are increasing most in areas where the pathogen has recently encroached, and 3) prevalence in dogs is rising in states traditionally not considered to be of high Lyme risk [Centers for Disease Control and Prevention, 2017], suggesting that human risk is also increasing in these areas. Several recent studies have recognized increasing Lyme risk in traditionally low incidence areas. These areas include Illinois [Herrmann et al., 2014], Iowa [Lingren et al., 2005], North Dakota [Russart et al., 2014], Ohio [Wang et al., 2014b], and Michigan [Lantos et al., 2017]. Significant increases in canine prevalence are also seen in some areas that have not yet reported significant human incidence. Given the proximity of these locations to recognized high-incidence areas, it is reasonable to infer that canine prevalence is more sensitive to changes in Lyme risk and can be used as an early warning system to signal changes in human risk. West Virginia, Western Pennsylvania, and Eastern Ohio are such areas and can be viewed as a leading edge of rising prevalence in Lyme's westward expansion. This is supported by evidence in increased reports of ticks in these regions

[Eisen et al., 2016b].

Examining local, as opposed to regional, trends shows that some adjacent counties have trends in opposite directions. To fully understand this heterogeneity, further ecological analyses are needed. Possible factors to consider include the presence of urban centers, degree of forestation or other habitat factors, tick populations, reservoir presence and densities, vaccination, and preventative medication use. Medication use is likely driven by socioeconomic factors, whereas the other factors are related to climate or changing habitats.

Our approach made several simplifying assumptions. The link function in the model was treated as known, which might be a strong assumption. As poorly specified link functions can induce bias in the estimates of the covariate effects [Neuhaus, 1999], relaxing this assumption could be fruitful. We also assumed that the spatially varying coefficients followed independent and identically distributed Gaussian processes. A more flexible approach would allow these coefficients to be correlated through a multivariate GP [Ver Hoef and Barry, 1998]; however, multivariate GPs are more difficult to use and challenges remain in their development [e.g., Fricker et al., 2013]. The observed prevalence rates suggest that smoothness of the random effects may change by region, suggesting that a heteroskedastic GP might be more appropriate [Binois et al., 2016]. Further, GMRFs are known to oversmooth salient features [Smith and Fahrmeir, 2007]. However, approximating GPs with GMRFs via stochastic partial differential equations to maintain computational feasibility [Lindgren et al., 2011b] could prove promising for our application.

In addition to statistical challenges, future applications of our model include human Lyme disease and heartworm disease, ehrlichiosis, and anaplasmosis in dogs. The ecological, entomological, and environmental implications of the dog Lyme prevalence analysis presented here is the subject of ongoing research.

5.7 Supplementary Material

The supplementary material is contained in Appendix B. Appendix B.1 provides the full conditional distributions required to develop the proposed sampling procedure, Appendix B.2 provides additional simulation results, and Appendix B.3 provides details about the exploratory data analysis conducted in Section 5.5.

Figure 5.1: Observed seroprevalence of *B. burgdorferi*, aggregated over January 2012 to December 2016. White counties are those that did not report any test results.

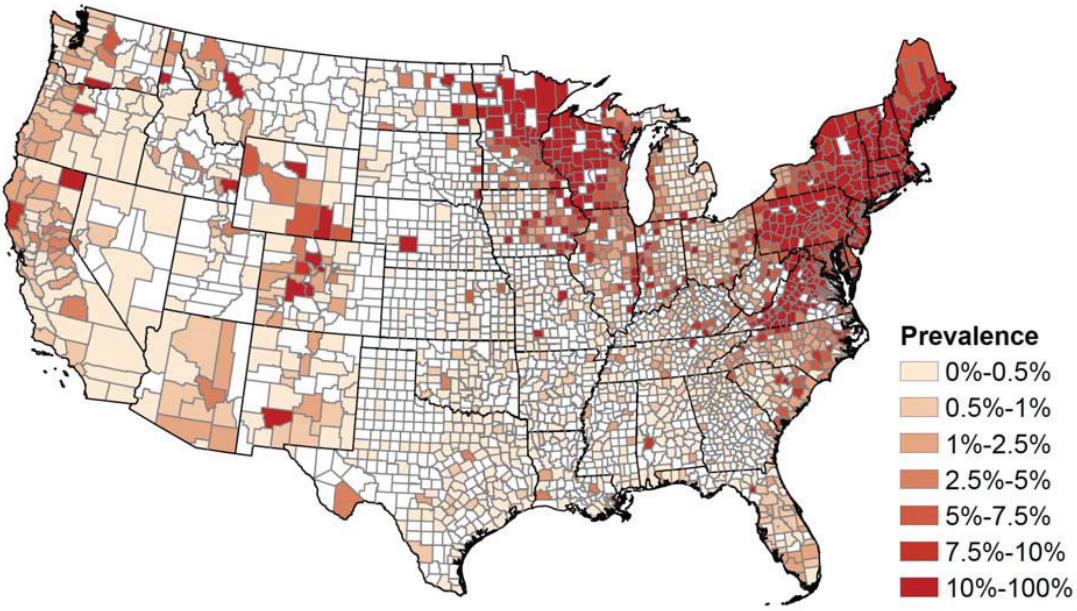


Figure 5.2: The true $\widetilde{\boldsymbol{\beta}}_1$ surface used to generate the independent data sets in the simulation example.

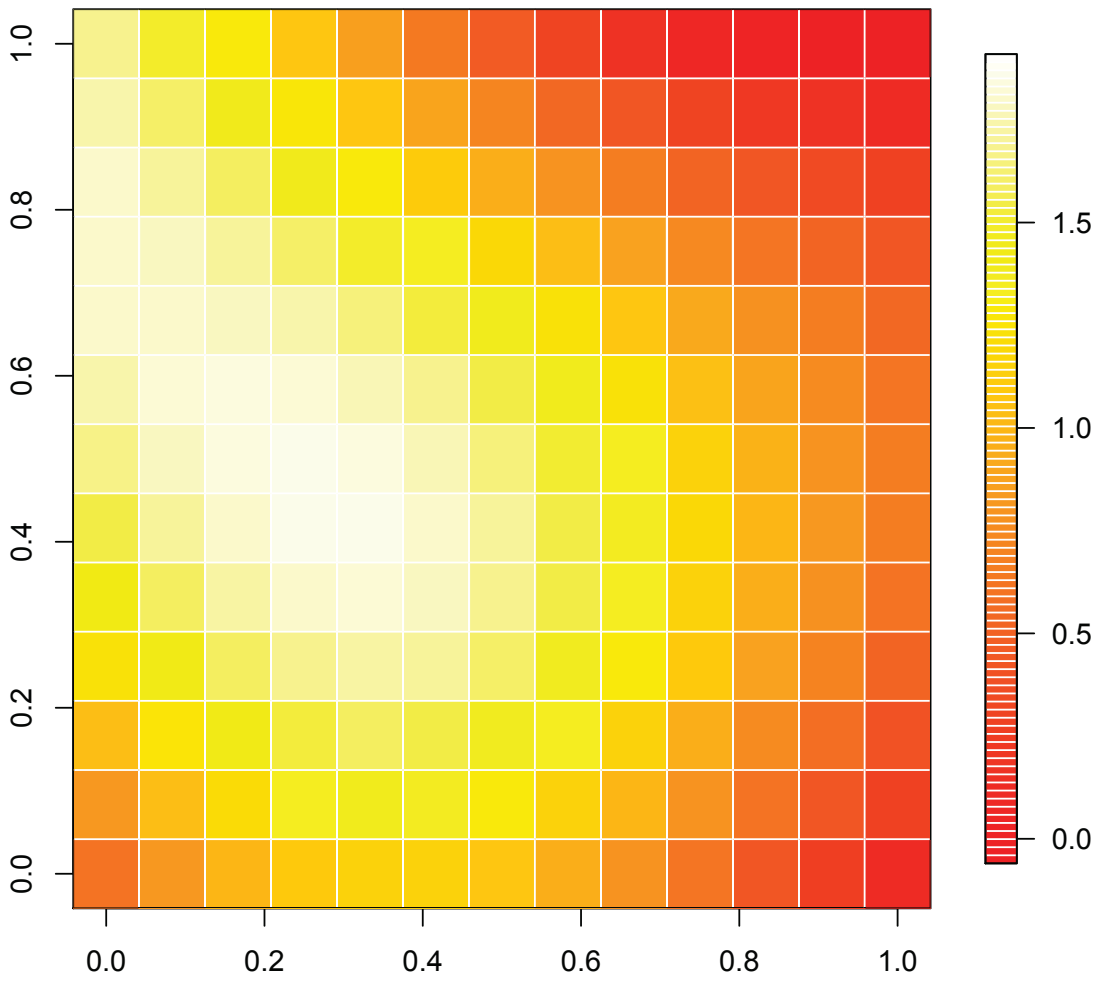


Figure 5.3: Summary of the posterior estimates of $\widetilde{\boldsymbol{\beta}}_1$ obtained in the simulation example when $\omega = 0.90$. Presented results include the sample mean of the posterior estimates (top row), empirical bias (middle row), and empirical mean squared error (bottom row). From left to right the columns correspond to the use of a 4×4, 5×5, and 7×7 grid of knots.

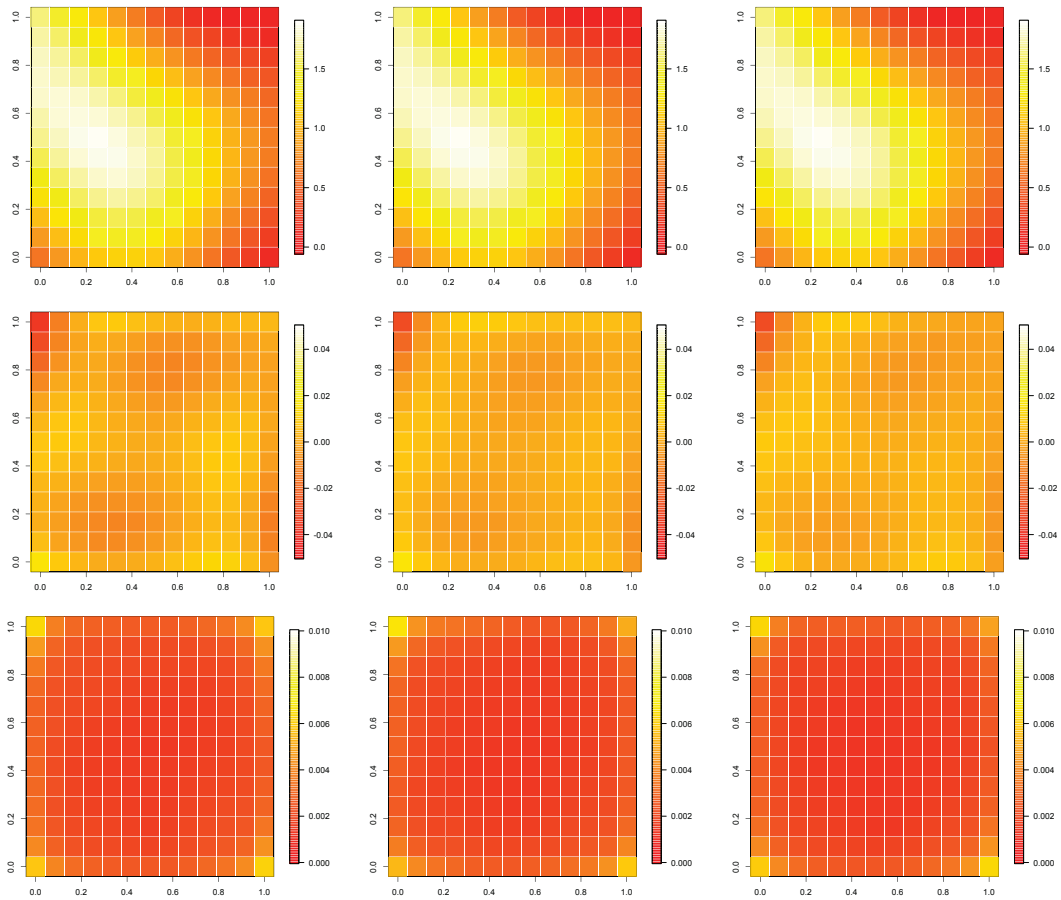
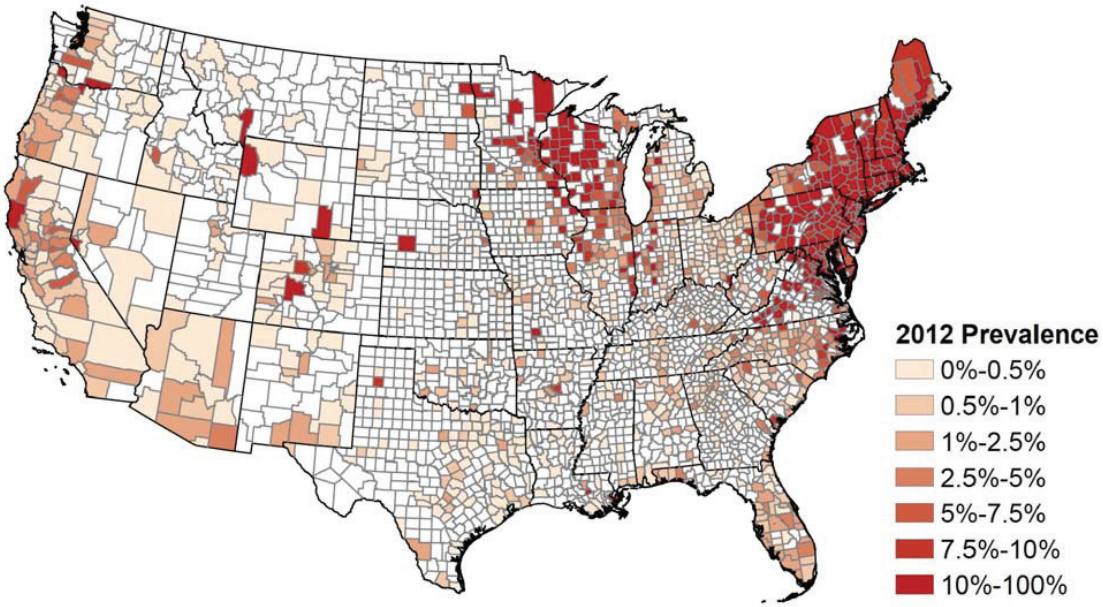


Figure 5.4: Raw reported canine seroprevalences in 2012 (top) and 2016 (bottom). White counties did not report any tests.



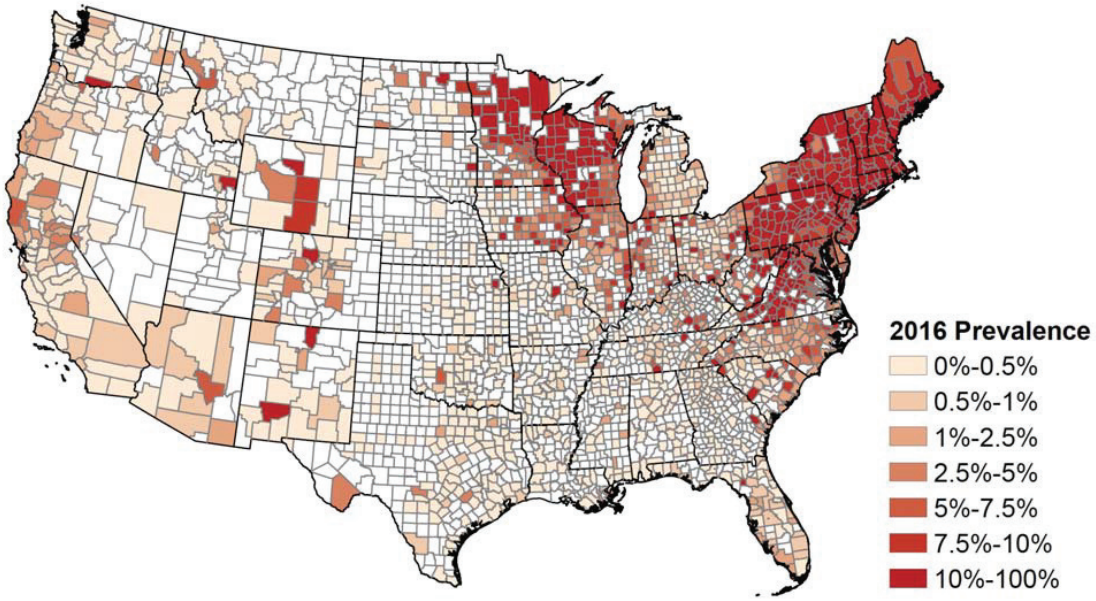


Figure 5.5: A summary of the spatially varying seasonal effects estimate from (5); i.e., estimates (ordered from smallest to largest) of $\widetilde{\beta}_1$ (left panel), $\widetilde{\beta}_2$ (center panel), and $\widetilde{\beta}_3$ (right panel). Included are the posterior mean, and upper and lower endpoints of 95% credible intervals. The red horizontal line is included to demonstrate that all of the credible intervals contain the same constant, thus indicating that a constant seasonal effect might be appropriate.

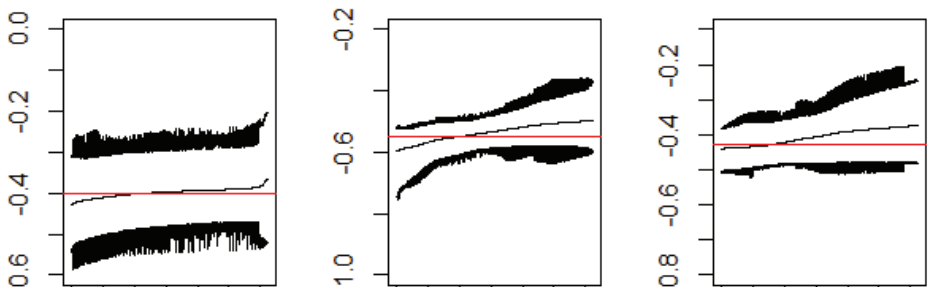


Figure 5.6: Estimate of the regional trend $\widetilde{\boldsymbol{\beta}}_1$ from the seasonal model (5.7) (top) and nonseasonal model (5.8) (bottom) used to analyze the seroprevalence data.

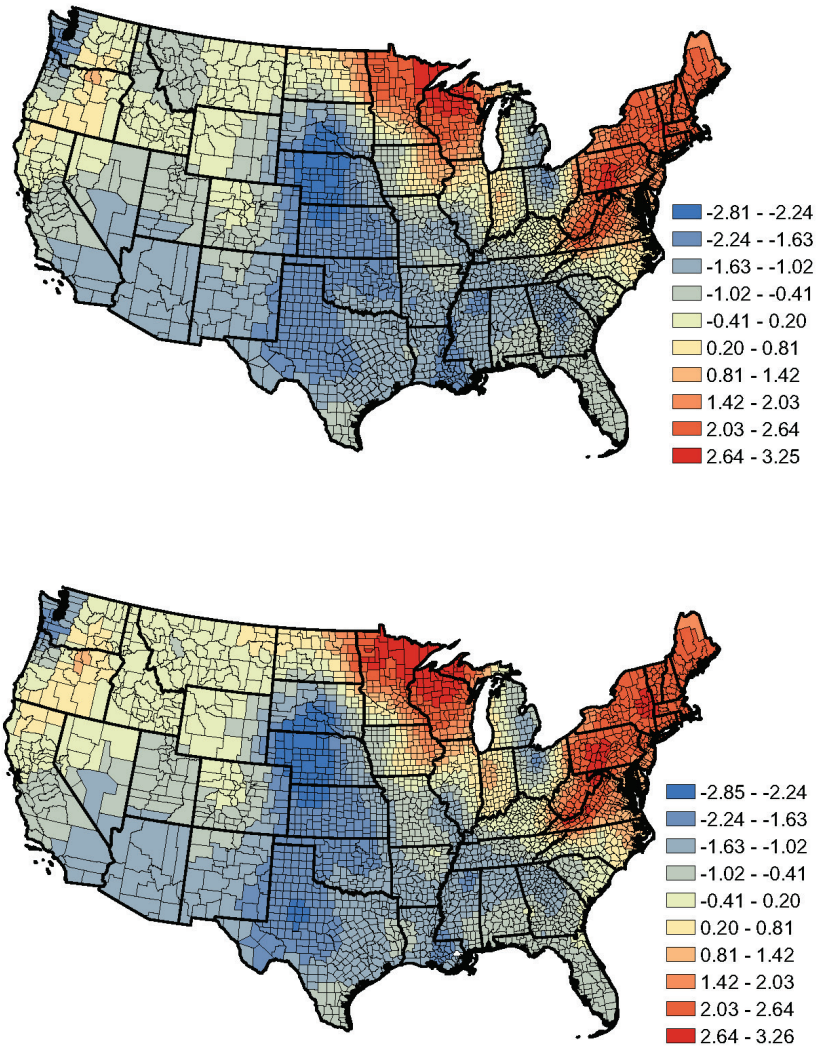


Figure 5.7: County-level trends. The top graphic displays the posterior mean estimate of υ_s from model (7), and the bottom from model (8).

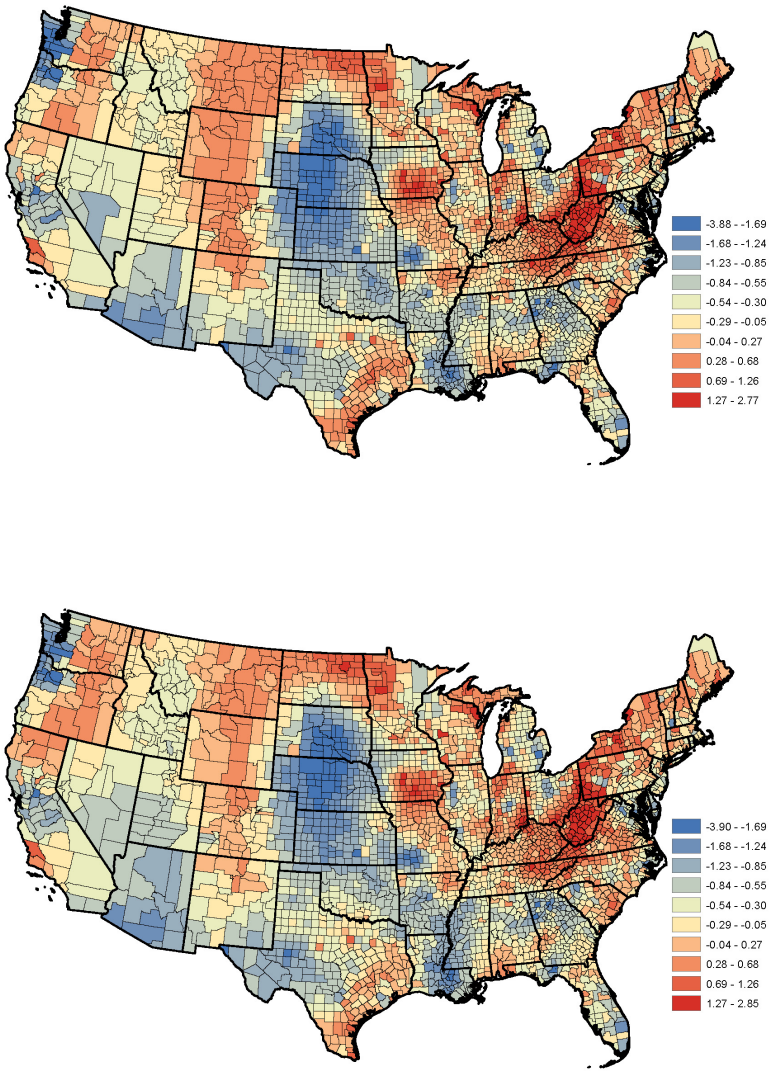
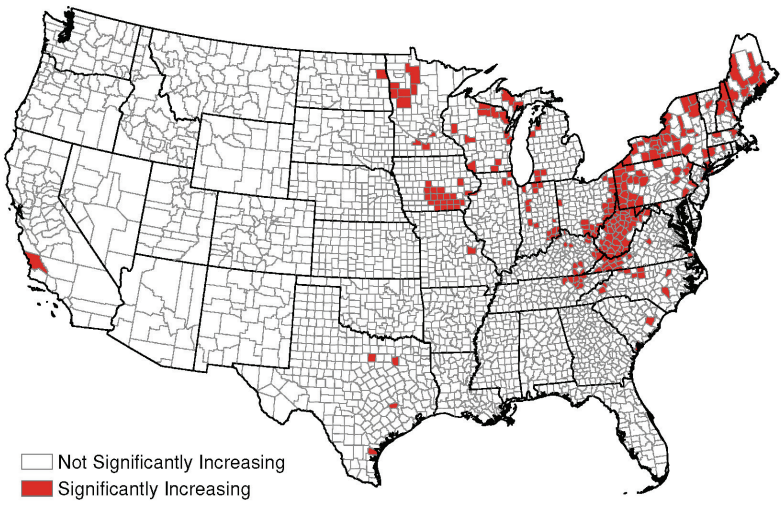
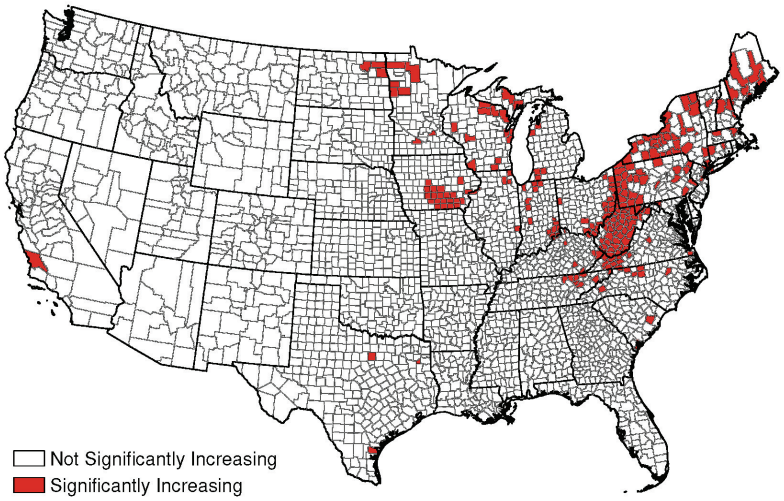


Figure 5.8: Counties where υ_s was significantly positive at the 95 % confidence level. The top graphic corresponds to model (7), and bottom to model (8).





Chapter 6

A Bayesian Multidimensional Trend Filter

6.1 Introduction

Nonparametric estimation is a popular statistical tool because it requires only very general assumptions about the nature of the function being estimated; e.g., smoothness, monotonicity, etc. While this freedom allows for the estimation of a broad class of functions, too much freedom can result in overfitted models with poor predictive performance. To avoid this pitfall, most nonparameteric techniques penalize rapid changes in the estimated function to avoid data chasing. Attempts to extend the penalty structure as the dimension of the covariate space increases are often cumbersome and require extensive tuning, increased computational expense, and apriori assumptions about interactions between covariates. In this work, we propose a nonparametric estimator whose penalty structure extends elegantly with the dimension of the covariate space.

Two of the most widely used frequentist methods for nonparametric estimation are kernel smoothing and spline regression. Kernel smoothers estimate the underlying function with a weighted average of the observed values. Perhaps the most well-known kernel smoother is the Nadarya-Watson estimator, proposed in Nadaraya [1964] and Watson [1964], which weights observations using a kernel function of a pre-specified bandwidth. Other forms of kernel regression include the Priestly-Chao estimator [Priestley and Chao, 1972] and the Gasser-Müller estimator Gasser and Müller [1979].

For a nice comparison of these methods, see Jones et al. [1994]. Spline-based techniques regress the response on spline basis functions. Some important considerations for spline-based methods include the choice of the spline basis, the degree of the splines, and the number and placement of knots. For an introduction to spline-based methods and for further discussion see Marsh et al. [2001]. Smoothing splines are regularized nonparametric estimators which impose a penalty on the second derivative of the estimators. See Wang [2011] for a nice overview of various smoothing spline methods. P-splines are a related regularized spline regression technique which penalize the difference between the coefficients of adjacent splines. Since their initial development in Eilers and Marx [1996], p-splines have received much attention, including extensions to the Bayesian framework (see Lang and Brezger [2004] and Yue et al. [2012]). Unfortunately extending spline-based regression techniques to higher dimensions is plagued by the 'curse of dimensionality'. In particular, the problem of knot placement is difficult in multiple dimensions. The most commonly used method for extending spline models to higher dimensions is to express the multidimensional spline-based function as the tensor product of one dimensional splines, as in Stone [1994], He and Shi [1996], and Lang and Brezger [2004]. Such tensor product based approaches suffer from a number of drawbacks, including the need to tune multiple smoothing parameters.

Gaussian processes are a flexible class of nonparametric functional estimators with a straightforward extension to multiple dimensions and are often used as an alternative to spline regression or kernel smoothing if the covariate space is multidimensional. See Cressie [1993], Rasmussen and Williams [2006], or Cressie and Wikle [2011a] for an overview of GP models. GP models have been particularly widely used in the Bayesian context. See Banerjee et al. [2015a] or Berger et al. [2001] for more on the Bayesian aspects of GP modeling. While GP models are easily extendable to multiple dimensions, they are computationally expensive for large datasets. Many variants of and approximations to GPs have been developed in an attempt to make GP based models scalable to large datasets. Once such approach is covariance tapering [Furrer et al., 2006] which uses a compactly supported covariance function to induce a sparse covariance structure. Other methods such as fixed rank kriging [Cressie and Johannesson, 2008], and [Nychka et al., 2015]'s Lattice Kriging method use basis functions to produce a smaller covariance matrix. One commonly used technique is the Gaussian predictive process (GPP) model of Banerjee et al. [2008], which estimates a parent GP at a small set of knot locations and uses kriging to interpolate the process to the remaining locations. For a overview of the various GP approximation procedures, see Heaton et al. [2017b].

Trend filtering is a common nonparametric technique for time series data. Typically, trend filters are used to estimate the underlying mean function by smoothing out the effects of noise and seasonal cycles. These methods penalize discrete order derivatives over the observed time points. Hodrick and Prescott [1997] popularized trend filtering. Their estimator minimizes the function $g : \mathbb{R}^T \to \mathbb{R}$ defined by:

$$g(\phi_1, ..., \phi_T) = \sum_{t=1}^{T} (y_t - \phi_t)^2 + \lambda \sum_{t=3}^{T} \left\{ (\phi_t - \phi_{t-1}) - (\phi_{t-1} - \phi_{t-2}) \right\}^2 .$$

where y_t is the observation from time t, ϕ_t is the estimate of the trend at time t, and λ is a smoothness parameter. This method was initially explored in Leser [1961]. Note that the penalty term is the squared backwards finite difference approximation to the second derivative of ϕ. Since Hodrick and Prescott's work, trend filtering has been the subject of considerable attention. One of the most notable developments is that of Seung-Jean et al. [2009], who use an l_1 penalty to create a a piecewise linear trend filtering estimator. Their l_1 method minimizes the function $g : \mathbb{R}^T \to \mathbb{R}$

$$g(\phi_1, ..., \phi_T) = \frac{1}{2} \sum_{t=1}^{T} (y_t - \phi_t)^2 + \lambda \sum_{t=2}^{T-1} |(\phi_{t-1} - \phi_t) - (\phi_t - \phi_{t+1})|.$$

Tibshirani [2014] explores the connection between Seung-Jean et al. [2009]'s method and smoothing splines and locally adaptive regression splines and proves that given the correct tuning parameter, the l_1 trend filter will converge to the true trend function. Ramdas and J. Tibshirani [2014] present a fast algorithm for fitting trend filter models based on the alternating direction method of multipliers. Sadhanala and J. Tibshirani [2017] develop an extension to additive models, in which each additive component is estimated with a trend filter. An informed method for selecting the penalty parameter λ is presented in Yam.

Recently, trend filtering has been extended to the Bayesian framework. A Bayesian extension of the Hodrick Prescott method is explored in Trimbur [2006] with inverse gamma and beta priors on the smoothness parameter. Faulkner and Minin [2018] develop a Bayesian trend filtering method called shrinkage prior Markov random fields (SPMRF). The SPMRF is a one dimensional nonparametric Bayesian method which places shrinkage priors on the kth-order differences in function values. Faulkner and Minin [2018] explore different options for shrinkage priors and frame their method in the context of Gaussian Markov random fields. Roualdes [2015] develops another

Bayesian trend filter for time series data based on shrinkage priors. Our method differs from these approaches because it is designed for point process data collected over two or more dimensions, rather than one dimensional time series data.

We provide a straightforward approach to nonparameteric estimation over a multidimensional covariate space. Our method uses a specially constructed Gaussian Markov random field (GMRF) prior distribution to discourage abrupt changes in the estimated function. This prior distribution places an l_2 penalty on an approximation to the derivative of the estimated function, making our method a multidimensional extension of trend filtering methods. This GMRF is defined with respect to a *mesh*, or discretization of the support space. We provide an adaptive method for selecting the mesh which allows the data to determine the overall level of discretization. Our proposed Bayesian multidimensional trend filter avoids the problems of knot placement and tuning parameter selection that plague higher dimensional spline-based estimators. It is also immune to the large sample size problems of GP models, as the computational expense of our method is nearly independent of the number of observations. Furthermore, the method exploits the chromatic sampling strategy of Brown et al. [2017c] to avoid the inversion and/or Cholesky decomposition of large matrices, making it fast and computationally scalable to a large number of observations or a high dimensional covariate space. Additionally our trend filter estimates the underlying trend function across the entire covariate space, not just at the locations at which data was observed. Our method estimates complex interactions between covariates at no additional cost and does not require assumptions about which sort of covariate interactions may be present in the model.

The remainder of the chapter is organized as follows. We develop our methodology in Section 6.2. Section 6.3 describes an efficient, computationally scalable MCMC model fitting procedure, as well as an adaptive method for discretizing the covariate space. In Section 6.4 we present the results of a simulation study comparing the multidimensional trend filter to GP and GPP models. We find that our method is comparable or superior to the other models in terms of bias and MSE, and also faster that the other methods by an order of magnitude. We demonstrate the performance of our method for Gaussian and non-Gaussian data applications in Section 6.5. Section 6.6 provides concluding remarks.

6.2 Methodology

Suppose we have a vector of n observations $\boldsymbol{Y} = (y_1, y_2, ..., , y_n)'$, and we assume a generalized linear model where the mean structure of the ith observation is related to a linear predictor η_i via a link function g. We assume the following form for η_i:

$$\eta_i = \boldsymbol{x}_i\boldsymbol{\beta} + f(\boldsymbol{z}_i),$$

where $\boldsymbol{x}_i = (x_{i1}, ..., x_{1p})$ is a vector of covariates which are assumed to have linear effects, $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)'$ is the associated vector of covariate effects, and $\boldsymbol{z}_i = (z_{i1}, ..., z_{ir})$ is a vector of covariates having an unknown functional relationship to the response. We estimate the unknown function $f(\boldsymbol{z}_i)$ via the Bayesian multidimensional trend filter.

Here we are assuming that $\boldsymbol{z}_1$, $\boldsymbol{z}_2$, ..., $\boldsymbol{z}_n$ are points that may lie anywhere within an r-dimensional hyperrectangle Z, defined using a set of points $a_1 < b_1$, $a_2 < b_2$, ..., $a_r < b_r$, where $Z = \{(w_1, ..., w_r)' : a_j \leq w_j \leq b_j \text{ for } j = 1, ..., r\}$. Recall that a hyperrectangle is the generalization of a rectangle to higher dimensions, and is formally defined as the Cartesian product of intervals. Before applying our method, we split Z into a *mesh* M consisting of disjoint hyperrectangles $c_1, c_2, ..., c_Q$ whose union is Z.

We estimate the unknown function f with a multidimensional step function $\tilde{f}$, defined with reference to the mesh M. That is, for $\boldsymbol{w} \in Z$, $\tilde{f}(\boldsymbol{w}) = \sum_{q=1}^{Q} \phi_q I(\boldsymbol{w} \in c_q)$ where $I(\boldsymbol{w} \in c_q) = 1$ if $\boldsymbol{w} \in c_q$ and 0 otherwise. Our model on η_i then becomes

$$\eta_i = \boldsymbol{x}_i\boldsymbol{\beta} + \boldsymbol{G}_i\boldsymbol{\phi},$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2, ..., \phi_Q)'$ and $\boldsymbol{G}_i = (I(\boldsymbol{w}_i \in c_1), I(\boldsymbol{w}_i \in c_2), ..., I(\boldsymbol{w}_i \in c_Q))$. In vector form we have

$$\boldsymbol{\eta} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{G}\boldsymbol{\phi},$$

where $\boldsymbol{\eta} = (\eta_1, \eta_2 ..., \eta_n)'$, $\boldsymbol{X} = (\boldsymbol{x}_1', \boldsymbol{x}_2', ..., \boldsymbol{x}_n')'$, and $\boldsymbol{G} = (\boldsymbol{G}_1', \boldsymbol{G}_2', ..., \boldsymbol{G}_n')'$.

Trend filtering methods impose a penalty on a derivative of the estimated function to encourage smoothness and prevent overfitting. In the one dimensional setting, the derivative of the trend function can be easily approximated using forward, backwards, or centered difference meth-

ods. In the multidimensional setting, the concept of a derivative encompasses a much broader class of mathematical objects, including partial derivatives, gradients, directional derivatives, mixed partials, Hessian matrices, etc. While one could penalize any of these objects, one simple and easily interpretable quantity to penalize is

$$\sum_{q=1}^{Q} d_q^2 \left(\phi_q - \overline{\phi}_q\right)^2,$$

$N(q)$ denotes the set of hyperrectangles which share an edge or a corner with c_q, $d_q = |N(q)|$ and $\overline{\phi}_q = {d_q}^{-1} \sum_{j \in N(q)} \phi_j$ is the average functional estimate in the neighboring hyperrectangles. Note that $c_q \notin N(q)$. The above is an l_2 penalty on the difference between the step function value in c_q and the average function value at all the surrounding hyperrectangles, weighted by the squared number of neighbors of c_q. This weighting is chosen for two reasons. The first is that it will enforce greater smoothness on the interior of Z and allow for some added flexibility on the boundaries of the region. Secondly, when the hyperrectangle sizes are unequal, larger hyperrectangles will tend to have more neighbors than smaller hyperrectangles. By placing a weaker penalty on smaller hyperrectangles, we allow for additional flexibility in finely discretized areas. We also note that this penalty may be re-expressed as

$$\sum_{q=1}^{Q} \left\{ \sum_{q' \in N(q)} (\phi_q - \phi_{q'}) \right\}^2.$$

Note that for c_q, this quantity is the squared average of the first order difference approximation to the derivative, where the average is taken over all neighboring hyperrectangles.

The trend filtering methodology is introduced via the prior distribution of $\boldsymbol{\phi}$. We assume

$$\pi(\boldsymbol{\phi}|\tau^2) \propto \exp\left\{ -\frac{1}{2\tau^2} \sum_{q=1}^{Q} \left(\sum_{q' \in N(q)} (\phi_q - \phi_{q'}) \right)^2 \right\},$$

which induces the penalty structure considered above. This prior distribution can be expressed as the following intrinsic Gaussian Markov random field

$$\pi(\boldsymbol{\phi}|\tau^2) \sim N(\mathbf{0}, \tau^{-2}\boldsymbol{M}), \quad \tau^2 > 0.$$

where $\boldsymbol{M} = (\boldsymbol{D} - \boldsymbol{W})(\boldsymbol{D} - \boldsymbol{W})$ is the precision matrix, $\boldsymbol{W}$ is the $Q \times Q$ adjacency matrix of the hyperrectangles, and $\boldsymbol{D}$ is diagonal with number of neighbors of each hyperrectangle on the

diagonal. That is, letting $\boldsymbol{A}_{qq'}$ denote the (q, q')th entry of the matrix $\boldsymbol{A}$, we have that $\boldsymbol{W}_{qq'} = 1$ if hyperrectangles q and q' share an edge or a corner, and 0 otherwise, and $\boldsymbol{D}_{qq} = \sum_{q'=1}^{Q} \boldsymbol{W}_{qq'}$. The l_2 penalty on the average of first order approximation to the derivative gives our method its interpretation as a multidimensional trend filter.

We note that $\boldsymbol{M}$ is not invertible, and thus the trend filtering prior is improper. A similar situation arises with intrinsic conditional autogressive (ICAR) models. One solution is introduce a propriety parameter ρ, and take $\boldsymbol{M}^* = (\boldsymbol{D} - \rho\boldsymbol{W})^2$. Letting λ_1 and λ_Q denote the smallest and largest eigenvalues of $\boldsymbol{D}^{-1/2}\boldsymbol{W}\boldsymbol{D}^{-1/2}$, respectively, we have that $(\boldsymbol{D} - \rho\boldsymbol{W})$ is invertible provided $\frac{1}{\lambda_1} \leq \rho \leq \frac{1}{\lambda_Q}$; then $\boldsymbol{M}^*$ is the product of invertible matrices and is thus invertible. See Banerjee et al. [2015a], section 4.3 for more details. Alternatively, one can use the improper prior and proceed with appropriate caution. Intrinsic GMRFs are known to give rise to difficulties concerning the variance parameter τ^2. See Lavine and Hodges [2012] or Lindqvist and Taraldsen [2017] for more details and various techniques for avoiding these complications.

6.2.1 A Note about Higher Order Penalties

We note here that our method can be extended to higher order penalties which induce a smoother estimator. For $k \geq 1$, define $\boldsymbol{M}_k = (\boldsymbol{D} - \boldsymbol{W})^{2k}$. Consider for $k = 2$:

$$\boldsymbol{\phi}'\boldsymbol{M}_2\boldsymbol{\phi} = \{(\boldsymbol{D} - \boldsymbol{W})(\boldsymbol{D} - \boldsymbol{W})\boldsymbol{\phi}\}' \{(\boldsymbol{D} - \boldsymbol{W})(\boldsymbol{D} - \boldsymbol{W})\boldsymbol{\phi}\}.$$

For a vector $\boldsymbol{a}$, let $(\boldsymbol{a})_q$ denote the qth entry of the vector. Note that

$$\begin{aligned}\{(\boldsymbol{D} - \boldsymbol{W})(\boldsymbol{D} - \boldsymbol{W})\boldsymbol{\phi}\}_q &= d_q^2\left(\phi_q - \overline{\phi}_q\right) - \left\{\sum_{q' \in N(q)} d_q\overline{\phi}_q - d_{q'}\overline{\phi}_{q'}\right\} \\ &= \sum_{q' \in N(q)} \left\{d_q(\phi_q - \overline{\phi}_q) - (d_q\overline{\phi}_q - d_{q'}\overline{\phi}_{q'})\right\},\end{aligned}$$

and thus

$$\boldsymbol{\phi}'\boldsymbol{M}_2\boldsymbol{\phi} = \sum_{q=1}^{Q}\left[\sum_{q' \in N(q)}\left\{d_q(\phi_q - \overline{\phi}_q) - d_q(\overline{\phi}_q - \frac{d_{q'}}{d_q}\overline{\phi}_{q'})\right\}\right]^2.$$

This penalty is loosely analogous to a penalty on the second derivative. Recall that $(\phi_q - \overline{\phi}_q)$ is the average of the first order approximation to the derivative taken over the neighborhood of c_q, and

$(\overline{\phi}_q - \overline{\phi}_{q'})$ is the difference between the average function in the neighborhood of c_q and the average function value in the neighborhood of c'_q, and is thus loosely approximates the first derivative of ϕ. Thus in the case where $d_{q'} = d_q$, we have that the expression above is the difference of two approximations to the first derivative, and thus may be thought of as an approximation to the second derivative.

6.3 Model Fitting Procedure

The model is fit using Markov chain Monte Carlo (MCMC) methods to obtain a sample from the posterior distribution. The exact sampling algorithm will depend on the choice of data likelihood. Here we provide a general technique that allows for computationally efficient sampling under different likelihood functions. Our model fitting procedure uses chromatic sampling to update large groups of the ϕ_q terms in parallel [Brown et al., 2017c]. We first provide the details for the Gaussian likelihood, and then explain how to generalize the approach.

6.3.1 Sampling Under the Gaussian Likelihood

Under the Gaussian likelihood, we have

$$f(\boldsymbol{Y}|\boldsymbol{\beta}, \boldsymbol{\phi}, \sigma^2) \sim \mathrm{N}(\boldsymbol{X\beta} + \boldsymbol{G\phi}, \sigma^2 \boldsymbol{I}_n), \quad \sigma^2 > 0,$$

where $\boldsymbol{I}_n$ denotes the $n \times n$ identity matrix. Taking $\pi(\sigma^2) \sim$ Inverse Gamma (IG)$(\alpha_{\sigma^2}, \beta_{\sigma^2})$, and $\pi(\boldsymbol{\beta}) \sim \mathrm{N}(\boldsymbol{0}, \gamma^2 \boldsymbol{I}_p)$, $\gamma^2 > 0$ we obtain the following full conditional distributions:

$$\begin{aligned}
f(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\phi}, \sigma^2, \tau^2) &\sim \mathrm{N}\left(\sigma^{-2}\Sigma_{\boldsymbol{\beta}}\boldsymbol{X}'(\boldsymbol{Y} - \boldsymbol{G\phi}), \Sigma_{\boldsymbol{\beta}}\right) \\
f(\boldsymbol{\phi}|\boldsymbol{Y}, \boldsymbol{\beta}, \sigma^2, \tau^2) &\sim \mathrm{N}\left(\sigma^{-2}\Sigma_{\boldsymbol{\phi}}\boldsymbol{G}'(\boldsymbol{Y} - \boldsymbol{X\beta}), \boldsymbol{\Sigma}_{\boldsymbol{\phi}}\right) \\
f(\sigma^2|\boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\phi}) &\sim \mathrm{IG}\left(n/2 + \alpha_{\sigma^2}, (\boldsymbol{Y} - \boldsymbol{X\beta} - \boldsymbol{G\phi})'(\boldsymbol{Y} - \boldsymbol{X\beta} - \boldsymbol{G\phi})/2 + \beta_{\sigma^2}\right) \\
f(\tau^2|\boldsymbol{\phi}) &\sim \mathrm{IG}\left(Q/2 + \alpha_{\tau^2}, \boldsymbol{\phi M \phi}/2 + \beta_{\tau^2}\right),
\end{aligned}$$

where $\Sigma_{\boldsymbol{\beta}} = (\sigma^{-2}\boldsymbol{X}'\boldsymbol{X} + \gamma^{-2}\boldsymbol{I}_p)^{-1}$ and $\Sigma_{\boldsymbol{\phi}} = (\sigma^{-2}\boldsymbol{G}'\boldsymbol{G} + \tau^{-2}\boldsymbol{M})^{-1}$. In this context, the construction of a Gibbs sampling procedure is straightforward. However in practice, Q may be quite large, rendering both vector block updates of $\boldsymbol{\phi}$ and single site updating via looping prohibitively slow. To

speed up computation, we propose the chromatic sampling approach of Brown et al. [2017c], which relies on parallel univariate updating to avoid looping.

The full conditional distribution of $\boldsymbol{\phi}$ is multivariate Gaussian with a sparse precision matrix. Block Gibbs sampling from this full conditional is feasible when the number of hyperrectangles Q is relatively small [Furrer and Sain, 2010b], but becomes computationally prohibitive as Q increases. Single site updating is feasible for moderate values of Q, but as Q grows this approach suffers from poor mixing. Furthermore, it requires extensive looping, which is known to be slow in R and MATLAB. Rather than block or single site sampling, we employ *chromatic sampling* [Gonzalez et al., 2011b, Brown et al., 2017c]. The chromatic sampler exploits the conditional independence proprieties of the matrix $\boldsymbol{M}$ to parallelize single-site updates, which avoids Cholesky decompositions of large matrices and other time consuming matrix computations. Let $\{\mathcal{A}_1, \ldots, \mathcal{A}_K\}$ be a partition of the set $\{1, 2, ..., Q\}$ in which $\mathcal{A}_k$ is an index set identifying a collection of hyperrectangles such that for all $q, q' \in \mathcal{A}_k$, $\boldsymbol{M}_{qq'} = 0$. Brown et al. [2017c] provide a greedy algorithm for finding such a partition. For a vector $\boldsymbol{a} = (a_1, \ldots, a_Q)'$ and an index set $\mathcal{C}$, define $\boldsymbol{a}(\mathcal{C}) := (a_q : q \in \mathcal{C})'$. The Markov property of the matrix $\boldsymbol{M}$ implies that the elements of $\boldsymbol{\phi}(\mathcal{A}_k)$, given $\boldsymbol{\phi}(\mathcal{A}_k^c)$, are conditionally independent. Therefore, by conditioning on $\boldsymbol{\phi}(\mathcal{A}_k^c)$, the elements of $\boldsymbol{\phi}(\mathcal{A}_k)$ can be sampled from their univariate full conditional distributions in parallel (or through 'vectorized' calculations). This approach is suitable for a very large number of hyperrectangles (e.g., $Q > 100000$) when they are sparsely connected. This is a highly desirable property in our case, as it allows the user to impose a very fine mesh over Z, even if the dimension of Z is large. For further details, see Brown et al. [2017c], who compare block sampling to chromatic sampling for GMRFs. Chromatic sampling requires the following univariate full conditional distribution of ϕ_q:

$$f(\phi_q | \boldsymbol{Y}, \boldsymbol{\beta}, \boldsymbol{\phi}_{-q}, \sigma^2, \tau^2) \sim N(\mu_{\phi_q}, \sigma^2_{\phi_q}),$$

where $\mu_{\phi_q} = \sigma^2_{\phi_q} \left[\tau^{-1}\{\frac{1}{2}(2\boldsymbol{DW} - \boldsymbol{W}^2 + \boldsymbol{D})\boldsymbol{\phi}\}_q + \sigma^{-2}\{\boldsymbol{G}'(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})\}_q\right]$, $\sigma^2_{\phi_q} = \{\sigma^{-2} g_q + \tau^{-2}(d_q^2 + d_q)\}^{-1}$, g_q denotes the number of observations in the qth hyperrectangle, and $\boldsymbol{\phi}_{-q}$ denotes $\boldsymbol{\phi}$ without the qth component.

6.3.2 Sampling Under the Bernoulli and Binomial Likelihood

Under the Bernoulli or binomial likelihood, we assume the probability of success for y_i is given by $p_i = g(\eta_i)$, giving rise to the following likelihood

$$f(\boldsymbol{\beta}, \boldsymbol{\phi}|\boldsymbol{Y}) \propto \prod_{i=1}^{n} \{g(\eta_i)\}^{y_i} \{1 - g(\eta_i)\}^{n_i - y_i},$$

where n_i is equal to the number of trials associated y_i. To facilitate the construction of a Gibbs based sampling procedure, we introduce latent data $\boldsymbol{\psi} = (\psi_1, ..., \psi_n)'$, giving rise to the following augmented likelihood having Gaussian form

$$f(\boldsymbol{\beta}, \boldsymbol{\phi}|\boldsymbol{Y}, \boldsymbol{\psi}) \propto \prod_{i=1}^{n} \exp\left\{-\frac{1}{2}(h_i \eta_i^2 - 2k_i \eta_i)\right\},$$

where the h_is and k_is will depend on the chosen likelihood and link function.

Following Albert and Chib [1993], under the Bernoulli likelihood and the probit link, we take $h_i = 1$ and $k_i = \psi_i$. In this framework, $\psi_i \sim N(\eta_i, 1)$, $\psi_i > 0$ if $y_i = 1$, and $\psi_i \sim N(\eta_i, 1)$, $\psi_i < 0$ if $y_i = 0$. Here the full conditional distribution of the ψ_i is truncated Gaussian with mean η_i and variance 1, with a support set of $(0, \infty)$ if $y_i = 1$ and $(-\infty, 0)$ if $y_1 = 0$. Under the logistic link, we follow the method of Polson et al. [2013] and take the ψ_is to be Pólya-Gamma$(n_i, 0)$ random variables with $h_i = \psi_i$ and $k_i = y_i - n_i/2$. In this framework, the full conditional distribution of ψ_i is Pólya-Gamma(n_i, η_i). In either case, proceeding in this fashion leads to Gaussian full conditional distributions for $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, and an inverse gamma full conditional distribution for τ^2; i.e.,

$$\begin{aligned}
f(\boldsymbol{\beta}|\boldsymbol{Y}, \boldsymbol{\psi}, \boldsymbol{\phi}, \tau^2) &\sim N\left[\Lambda_{\boldsymbol{\beta}}\{\boldsymbol{X}'(\boldsymbol{H}\boldsymbol{G}\boldsymbol{\phi} + \boldsymbol{K})\}, \Lambda_{\boldsymbol{\beta}}\right] \\
f(\boldsymbol{\phi}|\boldsymbol{Y}, \boldsymbol{\psi}, \boldsymbol{\beta}, \tau^2) &\sim N\left[\Lambda_{\boldsymbol{\phi}}\{\boldsymbol{G}'(\boldsymbol{H}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{K})\}, \Lambda_{\boldsymbol{\phi}}\right] \\
f(\tau^2|\boldsymbol{\phi}) &\sim \text{Inverse Gamma}\left(Q/2 + \alpha_{\tau^2}, \boldsymbol{\phi}\boldsymbol{M}\boldsymbol{\phi}/2 + \beta_{\tau^2}\right),
\end{aligned}$$

where $\Lambda_{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{H}\boldsymbol{X} + \gamma^{-1}\boldsymbol{I}_p)^{-1}$, $\Lambda_{\boldsymbol{\phi}} = (\boldsymbol{G}'\boldsymbol{H}\boldsymbol{G} + \tau^{-2}\boldsymbol{M})^{-1}$, $\boldsymbol{H}$ is a diagonal matrix with $\boldsymbol{H}_{ii} = h_i$ and $\boldsymbol{K} = (k_1, ..., k_n)'$. As the full conditional distribuiton of $\boldsymbol{\phi}$ is again multivariate normal with a sparse precision matrix, we can contruct an efficient Gibbs based chromatic sampling approach for the Bernoulli and binomial likelihoods. The univariate full conditional distribution needed for the

chromatic sampling approach is

$$f(\phi_q|\boldsymbol{Y}, \boldsymbol{\phi}_{-q}, \boldsymbol{\beta}, \tau^2) \sim N(\gamma_{\phi_q}, \lambda^2_{\phi_q}),$$

where $\lambda_{\phi_q} = \delta_q^2 \left[\tau^{-1}\{(2\boldsymbol{DW} - \boldsymbol{WW} + \boldsymbol{D})\boldsymbol{\phi}\}_q + \left\{\boldsymbol{G}'(\boldsymbol{HX\beta} + \boldsymbol{K})\right\}_q\right]$ and $\lambda^2_{\phi_q} = \{(\boldsymbol{G}'\boldsymbol{HG})_{qq} + \tau^{-2}(d_q^2 + d_q)\}^{-1}$.

6.3.3 Mesh Selection

Implementing the multidimensional trend filter requires imposing a mesh M on the region Z. A straightforward approach is to impose a regular partition over Z, as described in Section 6.2. However, as demonstrated by a simulation study presented in Section 6.4, using too few or too many hyperrectangles can inflate bias and MSE. To address this issue, we propose a method of adaptively refining the mesh M. Our method is similar to that of Finley et al. [2009], who propose an adaptive method of knot selection for Gaussian predictive process models. Our adaptive refinement method is similar, though for us hyperrectangles play the role of knots. We outline our method as follows:

1. Specify an initial mesh M_0 using a partition P consisting on Q_0 hyperrectangles and define a set of hyperrectangles which are eligible for refinement. This may be all hyperrectangles, all hyperrectangles exceeding a certain size, all hyperrectangles in certain regions, etc. Set $t = 1$.

2. Fit the model using the current mesh M_t, consisting of Q_t hyperrectangles.

3. For each hyperrectangle, calculate

$$e_q = \frac{1}{d_q}\sum_{i=1}^{n}(y_i - \boldsymbol{x}_i\hat{\boldsymbol{\beta}} - \hat{\phi}_q)^2 \cdot I(\boldsymbol{w}_i \in c_q).$$

4. Attempt to minimize $\sum_{q=1}^{Q_t} e_q$, by selecting the hyperrectangle(s) having the largest values of e_q for refinement. We refine a hyperrectangle c_q by splitting it into equally sized pieces. That is we split each interval used to define c_q into two intervals of equal length, and define a partition of c_q using these intervals. This results in 2^r disjoint hyperrectangles of equal size whose union is c_q. We refine the L hyperrectangles having the largest e_q values. Note that refining all hyperrectangles in each iteration is a special case in which we impose finer and finer regular meshes on the support space.

5. Set $t = t + 1$ and return to Step 2 until the stopping criterion has been reached.

We propose two possible stopping criteria. The first method, similar to the approach of Finley et al. [2009], is to specify a maximum number of hyperrectangles Q_M, and terminate the refinement once $Q_t = Q_M$. The second approach is to run the algorithm either for a pre-specified number of iterations, or until no hyperrectangles are eligible for refinement. After the algorithm terminates, a measure of goodness of fit is calculated for each model fit, and the mesh from the model with the best goodness of fit metric is selected as the final mesh. We explore this second approach, the *nonuniform adaptive mesh refinement* via a simulation study in the next section, and compare it the *uniform adaptive mesh refinement* in which we refine all hyperrectangles in each iteration successively impose a finer and finer regular mesh.

6.4 Simulation Study

We use a simulation study to compare our method to some other commonly used nonparametric estimation techniques. Specifically we compare the multidimensional trend filter using a single mesh (the *nonadaptive trend filter*) to a GPP model and the nonuniform and uniform adaptive trend filter to a GP model. For our simulation study, we generated n observations from the following model

$$y_i = f(\boldsymbol{z}_i) + \epsilon_i, \tag{6.1}$$

where $\boldsymbol{z}_i = (z_{1i}, z_{2i})$, $f(\cdot)$ is a function from $\mathbb{R}^2 \to \mathbb{R}$, and $\{\epsilon_i\}_{i=1}^n$ are iid $\mathrm{N}(0, 2)$ random errors. For each $n \in \{500, 1000, 2000, 5000\}$, and each of the three functions f shown in Table 6.1, we generated 500 independent data sets. For each data set, the observation locations $\{\boldsymbol{z}_i\}_{i=1}^n$ where randomly selected from a two-dimensional continuous uniform distribution on $[0, 50] \times [0, 50]$. The distances between locations were re-scaled to improve the numerical stability of the GP model. For these simulations, the data was re-scaled so that $|y_i| < 1$, for $i = 1, ..., n$.

For each dataset, f was estimated using the nonadaptive, nonuniform adaptive, and uniform adaptive trend filters, a GPP with 100 knots placed on a regular grid, and a GP. Due to the computational expense of the GP model, it was fit to only 100 of the 500 datasets for $n = 500, 1000$ and 2000. Fitting the GP model to datasets of size 5000 was computationally prohibitive. The nonadaptive trend filter was fit using a 30×30 square. The adaptive mesh trend filter models were

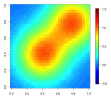
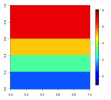
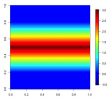
Bimodal Function	$f(\boldsymbol{z}_i) = \exp\left\{-\frac{(z_{i1}-20)^2+(z_{i2}-20)^2}{250}\right\} + \exp\left\{-\frac{(z_{i1}-40)^2+(z_{i2}-40)^2}{250}\right\}$	
Stairstep Function	$\mathrm{f}(\boldsymbol{z}_i) = \begin{cases} 1 & 0 \le z_{i2} \le 10 \\ 3 & 10 < z_{i2} \le 20 \\ 5 & 20 < z_{i2} \le 30 \\ 7 & 30 < z_{i2} \le 50 \end{cases}$	
Spike Function	$\mathrm{f}(\boldsymbol{z}_i) = \begin{cases} 0 & 0 \le z_{i2} \le 10 \\ 3 - \frac{1}{5}\lvert z_{i2} - 25\rvert & 10 < z_{i2} \le 40 \\ 0 & 40 < z_{i2} \le 50 \end{cases}$	

Table 6.1: Caption

initialized using a mesh consisting of a 12×12 square grid. For the uniform adaptive mesh trend filter, the mesh was successively refined 3 times, so that the model was fit using a 12×12, 24×24, 48×48, and 96×96 regular grid. A fivefold cross validation score was computed for each mesh, and the mesh having the best score was selected as the final mesh. For the nonuniform adaptive mesh trend filter, the set of hyperrectangles eligible for refinement in each iteration was the set of hyperrectangles that (1) contained at least one observation, and (2) had been refined less than 4 times (i.e. all hyperrectangles larger than 1/16 their original size). Note that the first criterion is necessary to ensure $e_q > 0$. In each iteration, we refine the 100 hyperrectangles with the largest mean estimated squared error e_q among those hyperrectangles which had been refined less than 4 times. The model was run until no more hyperrectangles were eligible for refinement, which occurs when all hyperrectangles meeting criteria (2) contain no observations. A fivefold cross validation score was computed in each iteration, and the mesh from the iteration with the smallest score was selected as the final mesh. For both the GP and GPP models, we assumed the mean function $\boldsymbol{\mu}(\boldsymbol{z}) = 0$ for all $\boldsymbol{z}$. The covariance function for the GP was specified as $\rho(\boldsymbol{z}_i, \boldsymbol{z}_j : \gamma) = \gamma^{d_{\boldsymbol{z}_i,\boldsymbol{z}_j}}$ where $\gamma \in (0, 1)$ is a model estimated parameter and $d_{\boldsymbol{z}_i,\boldsymbol{z}_j}$ is the Euclidean distance between $\boldsymbol{z}_i$ and $\boldsymbol{z}_j$. The covariance function for the parent process for the GPP's was specified as $\rho(\boldsymbol{z}_i^*, \boldsymbol{z}_j^* : \gamma) = \gamma^{d_{\boldsymbol{z}_i^*,\boldsymbol{z}_j^*}}$, where $d_{\boldsymbol{z}_i^*,\boldsymbol{z}_j^*}$ is the Euclidean distance between knots $\boldsymbol{z}_i^*$ and $\boldsymbol{z}_j^*$.

We report the empirical bias, empirical MSE, empirical standard deviation, and model fitting time in Table 6.2. The bias, MSE, and standard deviation values reported in Table 6.2 represent means over all 500 datasets and over all n observation locations. The time denotes the time to fit a single model, on a computer running Microsoft Windows 10 Home with an Intel(R) Core(TM) i5-3570K CPU and 16 GB of RAM. We see that the bias, MSE, and standard deviation

of all three trend filtering methods are small, and decrease as the sample size increases. The GPP model is an approximation to the GP model which sacrifices some flexibility in exchange for reduced computational expense. Similarly, the nonadaptive trend filter sacrifices the flexibility of a refinable and adaptive mesh for computational savings. We can think of the nonuniform adaptive mesh trend filters and the GP model as more robust modeling techniques, with the non-adpative trend filter and GPP models being low dimensional approximations, and the uniform adaptive trend filter falling somewhere in between. Thus we primarily compare the performance of the GPP model to the non-adaptive trend filter and the GP model to the uniform and nonuniform adaptive trend filter. We see that the performance of the non-adaptive trend filter is comparable to that of the GPP model. Similarly, the performance of the uniform and non-uniform adaptive trend filter is comparable to that of the GP. Furthermore, the trend filtering methods are scalable to values of n which are not feasible for the GP model. Note that the computation time for the non-adaptive and uniform adaptive trend filter is nearly constant as n increases, while the GPP computation time increases like $O(n)$ and the GP increases like $O(n^3)$. The computation time for the nonuniform adaptive mesh trend filter appears to increase substantially with n, but this an artifact of the relative sizes of n and the number of hyperrectangles. In this simulation, the maximum number of hyperrectangles in the adaptive mesh method before no hyperrectangles will meet criterion (2) is $96^2 = 9216$. However, because there are at most $n = 5000$ observations, the algorithm often terminates before reaching the maximum mesh size because all the hyperrectangles meeting criterion (2) contain no observations. This happens much sooner for the smaller sample sizes, resulting in fewer iterations and hence the lower computation time. Once the observations are sufficiently dense, the algorithm will run for the full 31 iterations needed to exhaust crierion (2) before running out of eligible hyperrectangles to refine. Hence the computational time will grow much more slowly with n after this threshold has been reached.

We note that while the GPP and GP model estimate f only at the observation locations, the trend filtering methods estimate f over the covariate space. The estimated functions for the regular mesh trend filter and the adaptive mesh trend filter for $n = 5000$ are shown in Figure 6.1.

n		NATF	GPP	UATF	NUATF	GP
Bimodal Function						
500	Bias	-0.0018	-0.3935	0.0025	0.0020	-0.0005
	MSE	0.0509	0.3065	0.04855	0.0492	0.0513
	SD	0.1739	0.1848	0.1799	0.2073	0.0267
	Time	24.94 sec	3.38 min	16.81 min	59.93 min	27.13 min
1000	Bias	-0.0028	-0.2051	0.0007	-0.0004	-0.0012
	MSE	0.0299	0.2170	0.0276	0.0287	0.0325
	SD	0.1537	0.2779	0.1680	0.1814	0.2557
	Time	24.93 sec	7.70 min	16.73 min	1.89 hrs	3.46 hr
2000	Bias	-0.00004	-0.0564	-0.0027	0.0007	-0.0013
	MSE	0.0168	0.1065	0.0166	0.0178	0.0205
	SD	0.1338	0.2667	0.1460	0.1534	0.22251
	Time	27.56s	19.25 min	17.00 min	2.68 hr	25.13 hr
5000	Bias	0.0003	-0.0130	-0.0002	-0.0005	
	MSE	0.0086	0.0378	0.0087	0.0097	
	SD	0.1065	0.1677	0.1154	0.1192	
	Time	28.10 sec	56.33 min	16.27	4.77 hr	
Stairstep Function						
500	Bias	0.0010	0.0005	-0.0011	0.0067	0.0063
	MSE	0.2561	0.2513	0.2619	0.2748	0.2790
	SD	0.3910	0.3590	0.3968	0.4222	0.4337
	Time	25.22 sec	2.56 min	16.48 min	58.70 min	28.24 min
1000	Bias	0.0003	0.0001	-0.0036	-0.0038	-0.0056
	MSE	0.2236	0.2224	0.2319	0.2357	0.2303
	SD	0.3299	0.2859	0.3524	0.3994	0.4120
	Time	25.97sec	4.50 min	16.46 min	1.99 hr	3.53 hr
2000	Bias	0.0017	0.0018	0.0010	0.0001	-0.0022
	MSE	0.1878	0.1913	0.1801	0.1902	0.2004
	SD	0.2945	0.2323	0.3649	0.4048	0.4049
	Time	24.81 sec	8.46 min	16.28 min	3.01 hr	25.47 hrs
5000	Bias	0.0007	0.0007	-0.0005	-0.0005	
	MSE	0.1370	0.1572	0.1318	0.1334	
	SD	0.2767	0.1721	0.3169	0.3613	
	Time	25.19 sec	21.35min	16.33 min	4.90 hr	
Spike Function						
500	Bias	0.0023	-0.0520	-0.0034	0.0074	0.0054
	MSE	0.1019	0.2899	0.1022	0.1126	0.4922
	SD	0.3513	0.5069	0.3786	0.4076	0.2688
	Time	25.36 sec	2.41 min	15.74 min	59.80 min	27.61 min
1000	Bias	-0.0029	-0.0006	0.0027	0.0014	0.0041
	MSE	0.0650	0.1062	0.0644	0.0745	0.2915
	SD	0.2946	0.3215	0.1348	0.3373	0.2624
	Time	25.46 sec	4.65 min	15.74 min	2.14 hr	4.50 hr
2000	Bias	0.0003	0.0002	-0.0006	0.0018	-0.0032
	MSE	0.0419	0.0411	0.0412	0.0499	0.1500
	SD	0.2448	0.2173	0.2596	0.2803	0.2591
	Time	25.85 sec	8.85 min	16.14 min	3.24 hr	25.94 hr
5000	Bias	- 0.0001	0.0001	-0.0002	0.0003	
	MSE	0.0263	0.0233	0.0229	0.0291	
	SD	0.1896	0.1502	0.1981	0.2167	
	Time	25.68 sec	20.78 min	16.42 min	5.04 hr	

Table 6.2: The empirical bias, empirical MSE, empirical standard deviation, and model fitting time for three functions define in Table 6.1 under the nonadaptive mesh trend filter (NATF), Gaussian predictive process (GPP), uniform adaptive trend filter (UATF), and nonuniform adaptive trend filter (NUATF), and the Gaussian process model (from left to right).

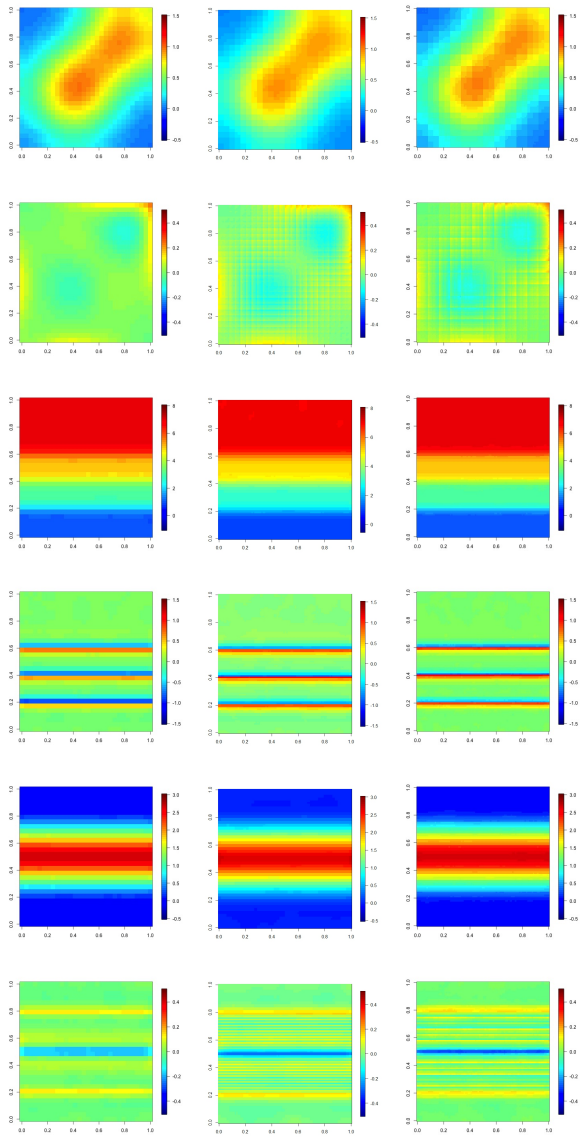


Figure 6.1: Results from the simulation study for the non-adaptive trend filter (column 1), uniform adaptive trend filter (column 2), and nonuniform adaptive trend filter (column 3) for the bimodal function (rows 1-2), the stairstep function (rows 3-4) and the spike function (rows 5-6) having 5000 observations. Depicted are the posterior mean estimate (rows 1,3,5), and empirical bias (rows 2,4,6).

An additional simulation study was used to explore the effect of varying levels of discretization by comparing the nonadaptive trend filter at several different resolutions to the nonuniform adaptive mesh method. Data was generated as described above, with $f(\boldsymbol{z}) = \sin\{\frac{2\pi}{50}(z_1 + z_2)\}$. We

considered 4 different meshes for the nonadaptive trend filter: 10×10, 20×20, 40×40, and 80×80, and 4 sample sizes: $n = 500$, 1000, 2000, and 5000. The adaptive mesh was initialized using a 10 by 10 grid. The 100 hyperrectangles having the largest value of e_q were selected for refinement in each iteration. In each iteration, the hyperrectangles eligible for refinement were those which (1) contained at least 1 observation and (2) had been refined less than 4 times. These configurations result in an 80 by 80 grid as the largest possible mesh before criterion (2) is exhausted, which allows for adequate comparison to the different size regular meshes studied in the simulation. The estimate of $\tilde{f}$ from the iteration having smallest fivefold cross validation score was selected as the final model.

q =	10	20	40	80	NUATF
n = 500					
Bias	0.0019	-0.0013	-0.0062	-0.3672	0.0124
MSE	0.1288	0.0967	0.1035	0.4467	0.1021
SD	0.1257	0.1229	0.0999	0.2060	0.3739
n = 1000					
Bias	-0.0012	0.0043	0.0011	-0.0716	0.0031
MSE	0.0932	0.0631	0.0615	0.1176	0.0624
SD	0.0850	0.0884	0.0742	0.0736	0.3089
n = 2000					
Bias	-0.0022	0.0002	0.0005	0.0109	-0.0013
MSE	0.0492	0.0349	0.0348	0.0567	0.0398
SD	0.2271	0.2433	0.2248	0.2086	0.2522
n = 5000					
Bias	0.0010	-0.0003	-0.00008	-0.0017	- 0.0004
MSE	0.0555	0.0268	0.0202	0.0267	0.0220
SD	0.0272	0.0355	0.0315	0.0286	0.1926

Table 6.3: The empirical bias, empirical MSE, and empirical standard deviation from the nonadaptive mesh with a 10×10, 20×20, 40×40 and 80×80 grid, and the nonuniform adaptive trend filter (from left to right).

The emprical bias, empircal MSE, and emprical standard deviation from the simulation

study comparing the nonadaptive and nonuniform adaptive trend filter are shown in Table 6.3. The values in the table are averages over all 500 datasets and all observation locations. Note that for the nonadaptive trend filter, the bias and MSE tend to decrease and then increase as the number cells increases, indicating using a finer grid increases model performance up to a point, but using too fine a grid decreases performance. We note that the nonuniform adaptive mesh performs comparably to the best performing nonadaptive mesh for each sample size, indicating that the adaptive procedure is resulting in a near optimal mesh.

6.5 Data Application

In this section, we demonstrate the performance of the multidimensional Bayesian trend filter for Gaussian and binomial data, and compare it to the performance of the GPP model.

6.5.1 The Gaussian Case

For the Gaussian case, we use temperature data obtained from the National Oceanic and Atmospheric Administration (NOAA). The data set consists of the minimum daily temperature from January 1, 2017 recorded at 7276 weather stations across the contiguous United States. Figure 6.2 shows the raw data. For this application, we take Z to be the two dimensional latitude-longitude space of the contiguous US, and apply the following model:

$$y_i = f(\boldsymbol{z}_i) + \epsilon_i,$$

where y_i is the observed temperature from the ith weather station, $\boldsymbol{z}_i$ is the location of the ith station and the ϵ_is are assumed to be iid $N(0, \sigma^2)$.

The nonadaptive trend filter was fit using a mesh of 5956 hyperrectangle of width 0.4 degrees. The nonuniform adaptive trend filter was initialized with a mesh of square hyperrectangles of width 2 degrees. The model was iterated 80 times, with the 50 hyperrectangles having the largest mean estimated squared error e_q being refined in each iteration. Fivefold cross validation was used to assess goodness of fit, and the iteration having the smallest score was selected as the final model. The GPP model utilized the same mean and covariance functions described in the simulations. The parent process was defined on a regular grid of 109 knots, and interpolated out to a regular grid of

955 cells. Each iteration was run for 4000 MCMC iterations and the first 2000 were discarded as burn in. Convergence was assessed with trace plots.

Figure 6.2 displays the results from both model fits. The results are from the two trend filter models are very similar, with the adaptive mesh results being somewhat smoother. The GPP model provides a coarser fit.

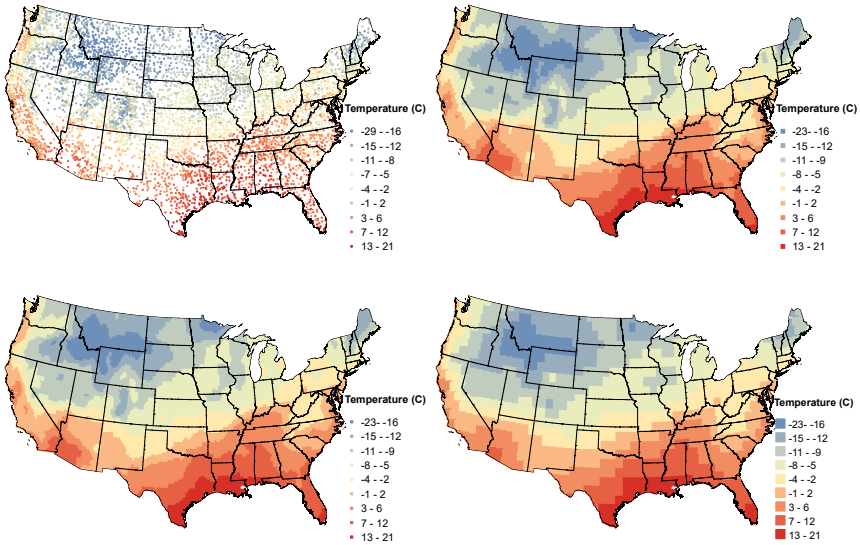


Figure 6.2: The observed minimum daily temperature in degrees Celsius for January 1, 2017 (top left), and estimated minimum daily temperature using the non-adaptive trend filter (top right), nonuniform adaptive trend filter (bottom left), and GPP model (bottom right).

6.5.2 The Binomial Case

For the binomial case, we use a dataset of test results for antibodies to *borrelia Burgdoferi*, the bacteria which causes Lyme disease. These tests were conducted in domestic dogs across the United States from January 1st, 2017 to December 31st, 2017. These tests are aggregated by county, so that the final dataset consists of the number of tests conducted in each county and the number of those tests which were positive. Not all counties report test results. Figure 6.3 displays the raw data. For this application, Z is again the two dimensional latitude-longitude space of the contiguous US. We

assume the following generalized linear model:

$$y_i | n_i, p_i \sim \text{Binomial}(n_i, p_i)$$
$$\eta_i = g^{-1}(p_i) = f(\boldsymbol{z}_i)$$

where y_i is the number of positive tests from county i, n_i is total number of tests from county i, p_i is the unknown disease prevalence in county i, $\boldsymbol{z}_i$ is the location of the centroid of county i, and $g(\cdot)$ is the logistic link function.

We fit the model using the nonadaptive trend filter using grid of hyperrectangles of dimension 1.7×0.7 degrees. The nonuniform adaptive trend filter was initialized using a mesh consisting squares of width 2 degrees, and was iterated 42 times, and the 50 hyperrectangles having the smallest values of e_q were refined in each iteration. The mesh having the smallest fivefold cross validation score was selected as the final model. The GPP model used the same mean and covariance functions used in the simulation study, and the parent process was defined using 100 knot locations randomly selected from the centroids of the counties at which data was observed. The parent process was interpolated out to the remaining county centroids.

Figure 6.3 displays the results from nonadaptive trend filter, the nonuniform adaptive trend filter and the GPP models. The results are fairly similar, with the most notable differences being in the western part of the country. As there is very little data reported in these regions, the higher variability is perhaps not surprising. We note that nonuniform adaptive trend filter model produces a more gradiated estimate in the Midwest and Mid-Atlantic areas where there is more data and prevalence is changing rapidly over space, while producing a rather coarse fit in the Western part of the country where data is scarce. This behavior suggests that the adaptive mesh is performing as it was designed to.

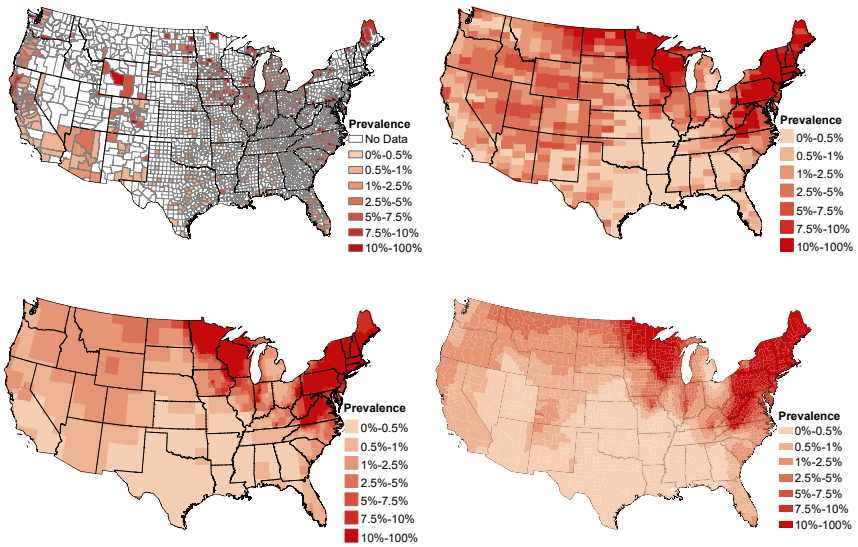


Figure 6.3: The observed 2017 Lyme disease prevalence (top left), and estimated prevalence using the nonadaptive trend filter (top right), nonuniform adaptive trend filter (bottom right) and GPP model (bottom left).

6.6 Conclusion

This chapter presents a computationally efficient and flexible multidimensional nonparametric estimator. The Bayesian multidimensional trend filter has several advantages over existing nonparametric estimators. The computational cost of the trend filter is nearly independent of the number of observations, and it scales well to a large covariate space. The trend filter extends naturally to several dimensions without the need for tensor products or a painstaking tuning process. Furthermore, our method provides an estimate of the trend function across the entire covariate space, not only at observation locations. The method is easily interpretable in the context of trend filtering, with the prior distribution on the trend parameters being a penalty on abrupt changes in the trend function. We also develop an adaptive method of descretization which allows the data to determine the resolution of the trend filtering estimator.

We compared the performance of the multidimensional Bayesian trend filter to Gaussian process and Gaussian predictive process models via a simulation study. We found that the performance of the nonadaptive trend filter is comparable to that of the GPP, and the performance of

the uniform and nonuniform adaptive trend filter is comparable to that of the GP models. The nonadaptive trend filter is by far the fastest method, particularly as the sample size gets large. For example, for 2000 observations, the regular mesh model takes only 28 seconds to run, while the GPP model takes 20 minutes and the GP model takes over 24 hours. These differences become even more pronounced for a sample size of 5000 observations, for which the GP model is computationally infeasible.

While this work focused on two dimensional spatial applications of the Bayesian multidimensional trend filter, the method is easily extendable to a variety of other applications. Other extensions include allowing the prior variance parameter of the trend filter (τ^2) to vary with space to induce stronger penalties in certain areas, and extending the adjacency concept to a larger neighborhood to induce more smoothness. Several aspects of the adaptive mesh trend filter merit further exploration, including the effect of different refinement criteria, the effect of refining more or fewer hyperrectangles in each iteration, and the effect of different measures of goodness of fit for selecting the final mesh. We hope to explore these extensions further in future work.

Chapter 7

A Bayesian Approach to Identifying Meteorological Drivers of High $PM_{2.5}$ Levels through Spatial Quantile Regression and Multidimensional Trend Filtering

7.1 Introduction

Air pollution has long been a public health concern. One form of air pollution considered especially dangerous is $PM_{2.5}$, particulate matter less than 2.5 micrometers in diameter. Because of its small size, $PM_{2.5}$ can become entrapped in the lungs and even enter the bloodstream, making it extremely hazardous to human health. Numerous studies have linked exposure to $PM_{2.5}$ to respiratory problems and cardiovascular disease [Chai et al., 2019, Dabass et al., 2016, Polezer et al., 2018, Wang et al., 2015, Weber et al., 2016]. Each year, an estimated 3.2 million deaths are attributed to $PM_{2.5}$ worldwide [Jerrett, 2015]. Reducing the public health burden of $PM_{2.5}$ requires understanding what factors determine $PM_{2.5}$ levels; understanding which factor or combinations of factors cause

dangerous spikes in $PM_{2.5}$ is of special importance.

Our data consists of daily $PM_{2.5}$ measurements taken at 174 Environmental Protection Agency stations spread across the Eastern United States from 2010 to 2014. Nearly 60% of the US population lives East of the Mississippi River, and the region also has a higher concentration of coal burning power plants than the rest of the country, making the Eastern United States a region of particular interest [Muyskens et al., 2017]. Meteorological variables such as temperature, humidity, and wind speed are known to be associated with $PM_{2.5}$ levels [Porter et al., 2015]. Furthermore, the effects of these covariates have been shown to change over space [Russell et al., 2017]. For this reason it is necessary to allow the effects of covariates to be spatially dependent when modeling $PM_{2.5}$ levels over a large spatial area. We develop a Bayesian quantile regression for this data in which the effects of meteorological covariates are allowed to change with space. As the effects of meteorological drivers on $PM_{2.5}$ levels are believed to change seasonally as well as spatially, we fit our model to the summer and winter $PM_{2.5}$ data separately. The data exhibits spatio-temporal dependence which must be accounted for in the model to allow for reliable inference; toward this end, we include spatio-temporal random effects in our model.

Quantile regression models are commonly used by researchers who wish to draw inference about areas of the response distribution other than the mean, such as the median or the tails. This technique relates covariates to a given quantile of the conditional distribution of the response. Quantile regression was first proposed in Koenker and Bassett [1978]. The rth regression quantile $\boldsymbol{\beta}^*$ is a solution to

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^k}\left\{\sum_{i:y_i\geq \boldsymbol{X}_i\boldsymbol{\beta}} r|y_i-\boldsymbol{X}_i\boldsymbol{\beta}|+\sum_{i:y_i<\boldsymbol{X}_i\boldsymbol{\beta}}(1-r)|y_i-\boldsymbol{X}_i\boldsymbol{\beta}|\right\},$$

where $\boldsymbol{X}_i$ is the k-dimensional vector of covariates associated with observations y_i. Since the work of Koenker and Bassett [1978], quantile regression methods have been developed extensively.

Koenker [2005] provides a nice overview of the body of work on quantile regression up to 2005. We review a few of the more recent advancements here. Wu and Liu [2009] explore variable selection for quantile regression using two methods: SCAD (smoothly clipped absolute deviation) and adaptive LASSO. Single index quantile regression, a method in which multiple covariates are related to the quantile of interest via a single link function, is developed in Wu et al. [2010]. Bondell et al. [2010] propose a constrained method of quantile estimation to prevent individually estimated

quantile curves from crossing each other. A broadly generalizable method of local partitioned quantile regression is explored in Zhang [2017]. Firpo et al. [2009] develops a quantile regression technique that regresses the recented influence function of the response quantile on covariates.

There are several existing Bayesian approaches to quantile regression. Gelfand et al. [2003] develop a Bayesian method for estimating median residual life using Dirichlet processes. Kottas and Krnjajić [2009] develop a Bayesian semiparametric method for quantile regression, which uses a parametric estimator for the quantile function and a nonparametric estimator for the error distribution. Yu and Moyeed [2001] develop a Bayesian approach to quantile regression based on the asymmetric Laplace likelihood. As the asymmetric Laplace likelihood is not particularly tractable for posterior sampling, several different MCMC sampling procedures have been proposed. Choi and Hobert [2013] present a data augmentation MCMC algorithm and sandwich MCMC algorithm for the Laplace likelihood. Kozumi and Kobayashi [2011] present an MCMC algorithm based on a location-scale mixture representation of the asymmetric Laplace distribution. We use this approach for our implementation.

There are a few techniques for applying quantile regression to spatially correlated data. Hallin et al. [2009] develop a local linear quantile estimator for spatial data. Reich et al. [2011] develop a Bayesian method for spatial quantile regression which uses splines whose coefficients are functions of the quantile and spatial location. Gaussian processes are used to ensure the coefficients change smoothly with space. As their method is not feasible for large datasets, they also present an approximation method which can be used for larger datasets. Russell et al. [2017] develop a methodology for simultaneously identifying and estimating significant spatially varying covariate effects for a given regression quantile.

In this chapter, we present a Bayesian approach to spatio-temporal quantile regression. This method incorporates the multidimensional Bayesian trend filter of the previous chapter into the asymmetric Laplace model of Yu and Moyeed [2001]. Covariate effects are allowed to change through space using the multidimensional Bayesian trend filter. The trend filter estimates a covariate effect surface that changes smoothly through space. Furthermore, it produces an estimate of this surface over the entire spatial support, not only at the locations at which data was observed. The spatio-temporal dependence in the data is modeled using random effects embedded in a vector autoregression. Our methodology has several advantages over existing techniques. We account for both spatial and temporal dependence via the random effects. Russell et al. [2017] perform

quantile regression for a similar dataset, but ignore the temporal dependence in the data. Our method is fast and computationally scalable to large datasets and large spatio-temporal supports. The multidimensional trend filter uses Gaussian Markov random fields to estimate smoothly varying surfaces, allowing for fast model fitting via sparse matrix algebra. Finally the Bayesian framework allows for the straightforward assessment of predictor significance without appealing to asymptotics.

The remainder of the chapter is organized as follows. We present our methodology in Section 7.2. We demonstrate the effectiveness of our method with a simulation study in Section 7.3 before applying it the $PM_{2.5}$ data in Section 7.4. Section 7.5 contains a discussion of our findings, and Section 7.6 provides concluding remarks.

7.2 Methodology

In our $PM_{2.5}$ analysis, we have daily observations of $PM_{2.5}$ from air quality monitoring stations taken over several years, though some observations are missing. We wish to study the effects of a set of meteorological covariates on abnormally high $PM_{2.5}$ events. These effects are believed to vary through space, so rather than estimating a single covariate effect, we estimate an effect surface across the entire study area. As the data presents with heavy spatio-temporal dependence, spatio-temporal random effects are included in the model.

Suppose we have responses y_{st} and covariate information x_{stp} collected at locations $\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, ..., \boldsymbol{\ell}_S$ where $s = 1, ..., S$ indexes spatial location, $t = 1, ..., T$, indexes times and $p = 1, ..., P$ indexes the different covariates. We make use of the multidimensional Bayesian trend filter of the previous chapter to estimate the spatially varying covariate effects. The multidimensional trend filter is a nonparameteric estimator consisting of a multidimensional step function and a regularization penalty. In the two dimensional case, the support space is discretized into a set of Q rectangular cells, $(c_1, c_2, ..., c_Q)$ and the step function is estimated at each cell. Specifically, we define $\boldsymbol{\phi}_p = (\phi_{1p}, \phi_{2p}, ..., \phi_{Qp})'$ to be the vector of step function values associated with covariate p estimated at each grid cell in the spatial support of our study area. The regularization penalty is enforced through the prior distribution on $\boldsymbol{\phi}_p$. This prior distribution is given by $\pi(\boldsymbol{\phi}_p) \sim \text{Normal}(\mathbf{0}, \tau^{-1}\boldsymbol{M})$ where $\tau_p^2 > 0$, $\boldsymbol{M} = (\boldsymbol{D} - \boldsymbol{W})(\boldsymbol{D} - \boldsymbol{W})$, and $\tau^{-1}\boldsymbol{M}$ denotes the precision matrix of the distribution (throughout, the notation $\text{Normal}(\mu, \boldsymbol{A})$ will denote a normal distribution with mean μ and *precision* matrix $\boldsymbol{A}$). Here $\boldsymbol{W}$ is the adjacency matrix of the grid of Q cells, i.e. $\boldsymbol{W}_{ij} = 1$ if cells i and j share an edge

or a corner and 0 otherwise, and $\boldsymbol{D}$ is a diagonal matrix with $\boldsymbol{D}_{ii} = \sum_{j=1}^{Q} \boldsymbol{W}_{ij}$. As the rows of $(\boldsymbol{D} - \boldsymbol{W})$ sum to 0, $\boldsymbol{M}$ is not invertible, and thus $\pi(\boldsymbol{\phi})$ is an intrinsic Gaussian Markov random field. This prior distribution imposes a penalty on the average of the first order approximation to the derivative at each grid cell.

Define the function I so that $I(s) = q$ where c_q is the cell containing location $\boldsymbol{\ell}_s$. We assume the following model for the rth quantile of our data:

$$y_{st} = \upsilon + \sum_{p=1}^{P} x_{stp}\phi_{I(s)p} + \psi_{I(s)t} + \epsilon_{st}, \tag{7.1}$$

where υ is a global intercept, and the ϵ_{st} terms are independent errors following the asymmetric Laplace distribution with parameter r. Thus the probability density function of ϵ_{st}, $\pi(\cdot)$, is given by $\pi(\epsilon_{st}) = r(1-r)\exp\{\epsilon_{st}(r - I(\epsilon_{st} < 0)\}$. The ψ_{st} terms are spatio-temporal random effects included to model the spatio-temporal dependence in the data. These random effects will also be modeled using the multidimensional Bayesian trend filter. Thus for $t = 1, .., T$ we define $\boldsymbol{\psi}_t = (\psi_{1t}, \psi_{2t}, ..., \psi_{Qt})'$ to be the vector of spatio-temporal random effects from time point t estimated over the Q grid cells.

Following Kozumi and Kobayashi [2011], we can express the asymmetric Laplace errors as a mixture of exponential and normal random variables, allowing us to rewrite equation 7.1 as

$$y_{st} = \upsilon + \sum_{p=1}^{P} x_{stp}\phi_{I(s)p} + \psi_{I(s)t} + \theta z_{st} + \lambda\sqrt{z_{st}}u_{st}, \tag{7.2}$$

where $\theta = \frac{1-2r}{r(1-r)}$, $\lambda^2 = \frac{2}{r(1-r)}$, $\{z_{st}\} \overset{\text{iid}}{\sim}$ Exponential(1) and $\{u_{st}\} \overset{\text{iid}}{\sim}$ Normal$(0, 1)$.

To completely specify the model, we assign the prior distributions to all the unknown pa-

rameters.

$$\begin{aligned}
\boldsymbol{\phi}_p &\overset{\text{ind.}}{\sim} \text{Normal}(\mathbf{0}, \tau_p^{-2}\boldsymbol{M}) \text{ for } p = 1, ..., P \\
\tau_p^2 &\overset{\text{ind.}}{\sim} \text{Inverse Gamma}(\alpha_{\tau^2}, \beta_{\tau^2}) \text{ for } p = 1, ...P \\
\boldsymbol{\psi}_1 &\sim \text{Normal}(\mathbf{0}, \gamma^{-2}\boldsymbol{M}) \\
\boldsymbol{\psi}_t &\overset{\text{ind.}}{\sim} \text{Normal}(\zeta\boldsymbol{\psi}_{t-1}, \gamma^{-2}\boldsymbol{M}) \text{ for } t = 2, 3, ..., T \\
\gamma^2 &\overset{\text{iid}}{\sim} \text{Inverse Gamma}(\alpha_{\gamma^2}, \beta_{\gamma^2}) \\
\zeta &\overset{\text{iid}}{\sim} \text{Truncated Normal}(0, \sigma_\zeta^2, -1, 1) \\
z_{st} &\overset{\text{iid}}{\sim} \text{Exponential}(1) \text{ for } (s,t) \text{ such that } y_{st} \text{ is observed} \\
u_{st} &\overset{\text{ind.}}{\sim} \text{Normal}(0, 1) \text{ for } (s,t) \text{ such that } y_{st} \text{ is observed.}
\end{aligned}$$

The model is fit using Markov chain Monte Carlo (MCMC) methods. Using equation 7.3 we are able to construct a Gibbs based posterior sampling procedure. In particular, we have normal full conditional distributions for υ, the $\boldsymbol{\phi}_p$s, and the $\boldsymbol{\psi}_t$s, generalized inverse Gaussian full conditional distributions for the z_{st}s, inverse gamma full conditionals for the τ_p^2s and γ_h^2s and truncated normal full conditionals for the ζ_hs.

7.3 Simulation Study

A simulation study was conducted to assess the effect of differing amounts of missing data and differing quantiles of interest on the model. For the simulation study, we took $S = 150$, and generated the $\boldsymbol{\ell}_s$s from a continuous uniform distribution on $[0, 5] \times [0, 5]$. We took $\upsilon = 3$ for the intercept, set $P = 2$ and generated the x_{stp} and u_{st} from a standard normal distribution, and the z_{st} from an exponential(1) distribution. For $r \in \{0.5, 0.95\}$ and $k \in \{0.5, 0.75, 1\}$, data was generated at kST of the location-time pairs using:

$$y_{st} = \upsilon + x_{st1}f_1(\boldsymbol{\ell}_s) + x_{st2}f_2(\boldsymbol{\ell}_s) + \psi_{st} + \theta z_{st} + \lambda\sqrt{z_{st}}u_{st}, \tag{7.3}$$

where $f_1(\boldsymbol{\ell}_s) = \sin(\ell_{s1}) + \cos(\ell_{s2})$ and $f_2(\boldsymbol{\ell}_s) = \exp\left[\frac{1}{4}\left\{(\ell_{s1} - 2.5)^2 + (\ell_{s2} - 2.5)^2\right\}\right]$. Note that if $\boldsymbol{\ell}_s$ and $\boldsymbol{\ell}_{s'}$ are in the same grid cell, then $\phi_{sp} = \phi_{s'p}$, but $f_p(\boldsymbol{\ell}_s) \neq f_p(\boldsymbol{\ell}_{s'})$. The functions f_1 and f_2 used for data generation are shown in Figure 7.1. A 10×10 regular grid was used for the two dimensional

trend filter for a total of $Q = 100$ grid cells. As previously mentioned, we wish to analyze summer and winter $PM_{2.5}$ data separately. While allowing for temporal dependence between observations made in the same season of the same year, we assume temporal independence between years. For our simulation study, we take the number of years to be $R = 4$, and the length of each season to be $M = 10$, for a total of $T = RM = 40$ time points. The spatio-temporal random effects are realizations from the prior distributions specified in Section 7.2, with a slight modification to allow for temporal independence between years. Specifically, we assume $\boldsymbol{\psi}_t \overset{ind.}{\sim} \text{Normal}(\mathbf{0}, \gamma_h^{-2}\boldsymbol{M})$ if t corresponds to the first day of year h, and $\boldsymbol{\psi}_t \overset{ind.}{\sim} \text{Normal}(\zeta_h \boldsymbol{\psi}_{t-1}, \gamma_h^{-2}\boldsymbol{M})$ for all other days in year h. We take $\gamma_h^2 = 0.5$ and $\zeta_h = 0.95$ for all h, and generate the random effects from the prior distributions given above. For each combination of r and k, 500 independent datasets were generated, and the model was fit to each.

Results are summarized in Table 7.1, which contains the empirical bias, empirical MSE, empirical standard deviation and empirical 95% coverage probabilities for each simulation configuration, averaged over all S locations and all 500 datasets. The posterior mean estimate, empirical bias, empirical MSE, empirical standard deviation and empirical 95% coverage probabilities for the covariate surfaces f_1 and f_2 are depicted in Figures 7.2 and 7.3, respectively.

	Bias	MSE	SD	95 % Coverage
$n = ST$, $r = 0.5$				
f_1	0.0010	0.0375	0.2511	97.59
f_2	0.0028	0.0136	0.1447	98.39
υ	0.0005	0.0009	0.0290	100.00
$n = 0.75ST$, $r = 0.5$				
f_1	0.0014	0.0434	0.2738	98.09
f_2	-0.0033	0.0171	0.1567	98.05
υ	-0.0030	0.0012	0.0337	100.00
$n = 0.5ST$, $r = 0.5$				
f_1	0.0013	0.0545	0.3085	98.48
f_2	0.0058	0.0227	0.1741	97.44
υ	0.0012	0.0019	0.0417	100.00
$n = ST$, $r = 0.95$				
f_1	0.0014	0.0945	0.4010	98.74
f_2	0.00046	0.0499	0.2256	94.28
υ	0.0175	0.0061	0.0706	100.00
$n = 0.75ST$, $r = 0.95$				
f_1	-0.0032	0.1202	0.4368	98.49
f_2	0.0088	0.0576	0.2471	94.51
υ	0.0317	0.0084	0.0825	100.00
$n = 0.5ST$, $r = 0.95$				
f_1	-0.0109	0.1552	0.4911	98.49
f_2	0.0093	0.0766	0.2784	94.14
υ	0.0452	0.0127	0.1035	100.00

Table 7.1: The table gives the empirical bias, empirical MSE, empirical standard deviation, and empircal 95% coverage probabilities averaged over all S observed data locations and all data sets.

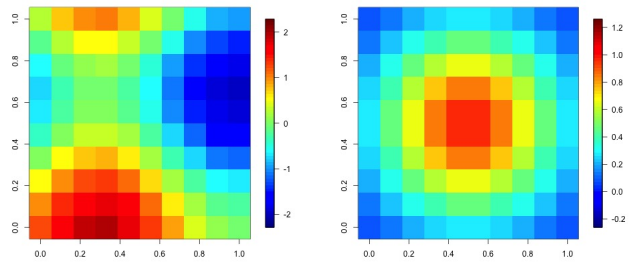


Figure 7.1: The functions f_1 and f_2 used for data generation.

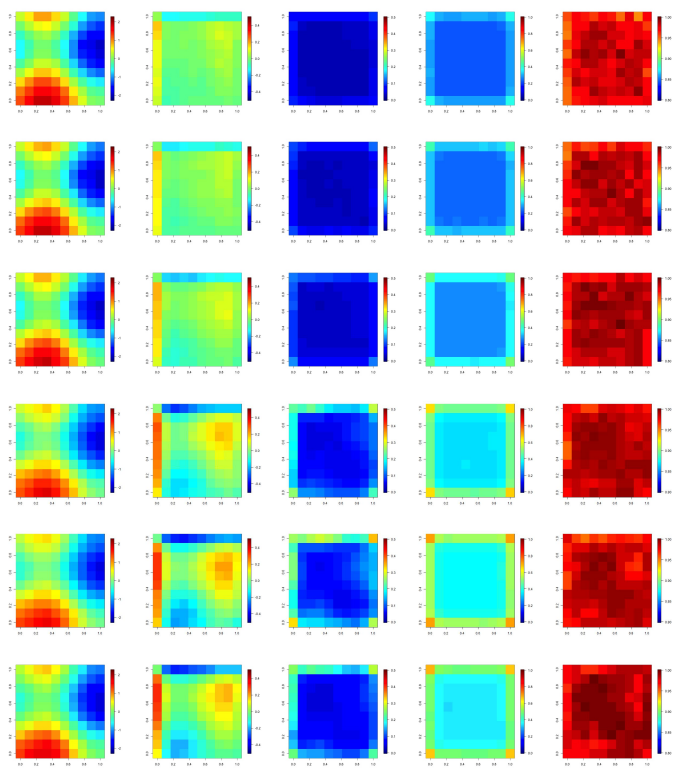


Figure 7.2: Simulation results for f_1. From left the right the columns display the posterior mean estimate, the empirical bias, the empirical MSE, the empirical standard deviation, and 95% empirical coverage probabilities. From top to bottom the rows correspond to 0% missing data with $r = 0.5$, 25% missing data with $r = 0.5$, 50% missing data with $r = 0.5$, 0% missing data with $r = 0.95$, 25% missing data with $r = 0.95$, and 50% missing data with $r = 0.95$.

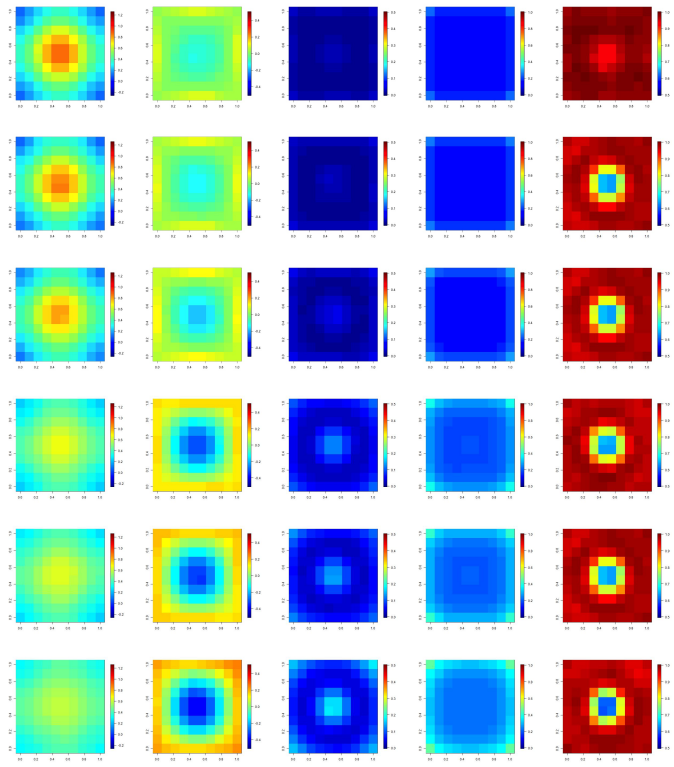


Figure 7.3: Simulation results for f_2. From left the right the columns display the posterior mean estimate, the empirical bias, the empirical MSE, the empirical standard deviation, and 95% empirical coverage probabilities. From top to bottom the rows correspond to 0% missing data with $r = 0.5$, 25% missing data with $r = 0.5$, 50% missing data with $r = 0.5$, 0% missing data with $r = 0.95$, 25% missing data with $r = 0.95$, and 50% missing data with $r = 0.95$.

The simulation study indicates that the methodology performs well overall, with slight decreases in performance when a large amount of data is missing. We also see slight increases in bias, MSE, and standard deviation on the boundaries. This phenomena is widely observed among nonparameteric spatial estimators.

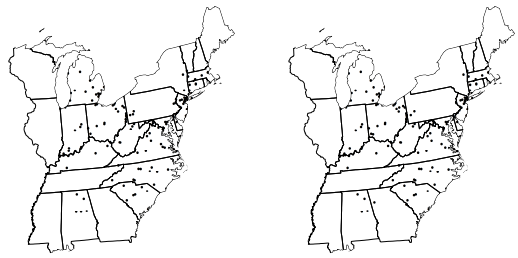


Figure 7.4: The air quality monitoring stations used in the winter analysis (left) and the summer ananlysis (right).

7.4 Data Application

7.4.1 The Data

The $PM_{2.5}$ response was recorded at 174 Environmental Protection Agency stations spread across the Eastern United States from 2010 to 2014. As the meteorological drivers of $PM_{2.5}$ are thought to differ between summer and winter months, we perform separate analysis for the summer months (June- August) and Winter (December - February). Some of the EPA stations had large temporal gaps in the reported data, and so only 150 of the stations were used for the winter analysis and 140 for the summer analysis. The stations involved in each analysis are depicted in Figure 7.4. The study period contained 5 complete summer seasons and 4 complete winter seasons.

The meteorological covariates included in this analysis come from the North American Regional Reanalysis. These covariates include wind speed, daytime and nighttime air temperature, daytime and nighttime height of the planetary boundary layer (HPBL), lower tropospheric stability (LTS), downward shortwave radiative flux (DSRF), percent cloud cover, precipitation, and relative humidity. wind speed is the average of the previous 48 hour period, relative humidity, daytime air temperature and daytime HPBL are daytime averages, nighttime air temperature and nighttime HPBL are nighttime averages, LTS is the average of the previous 24 hour period, DSRF and percent cloud cover are averages of the previous day, and precipitation is a daily presence/absence indicator. For more information on the covariates, see NAR.

7.4.2 Results

We use our proposed methodology to perform four separate analyses; i.e., we analyze the winter $PM_{2.5}$ data at the 0.5 and 0.95 quantiles and the summer $PM_{2.5}$ data at the same two quantiles. In the interests of parsimony, we include only those covariates found to be significantly related to the $PM_{2.5}$ response by Russell et al. [2017]. For the winter analysis, this was wind speed, daytime and nighttime air temperature, daytime and nighttime HPBL, LTS, relative humidity and percent cloud cover. For the summer analysis we include wind speed, daytime and nighttime air temperature, DSWRF, precipitation, relative humidity and cloud cover. We discretized the study area for the trend filter using a regular grid consisting of squares of width 1.5 degrees latitude by 1.2 degrees longitude. We used the prior distributions specified in Section 7.2 with $\alpha_{\tau^2} = \beta_{\tau^2} = \alpha_{\gamma^2} = \beta_{\gamma^2} = 1$, and $\sigma_\zeta^2 = 1000$. The model was run for 4000 MCMC iterations, with first 2000 discarded as burn-in. Convergence was assessed using trace plots. For each of the four models, we report both the estimated covariate effect surfaces and regions of significance for each covariate. Significance was assesed using 95% credible intervals. In particular, if the 95% credible interval for $\boldsymbol{\phi}_{pq}$ was completely positive (negative), then the pth covariate was deemed significantly positively (negatively) related to the response in the region spanned by cell q.

Figure 7.5 shows the estimated covariate effect surfaces for the winter $PM_{2.5}$ data at the 0.5 quantile. Figure 7.6 depicts regions of significance for these covariate surfaces. We find that wind speed and nighttime air temperature are significantly negatively related to median $PM_{2.5}$ levels for most of the study region, and daytime HPBL is significantly negatively related for much of the northern portion of the study area. Daytime air temperature and relative humidity are significantly positively related to median $PM_{2.5}$ levels for large portions of the study region, particularly in the western and southern areas. Nighttime HPBL, LTS, and percent cloud cover were significantly related to median $PM_{2.5}$ levels only for isolated areas.

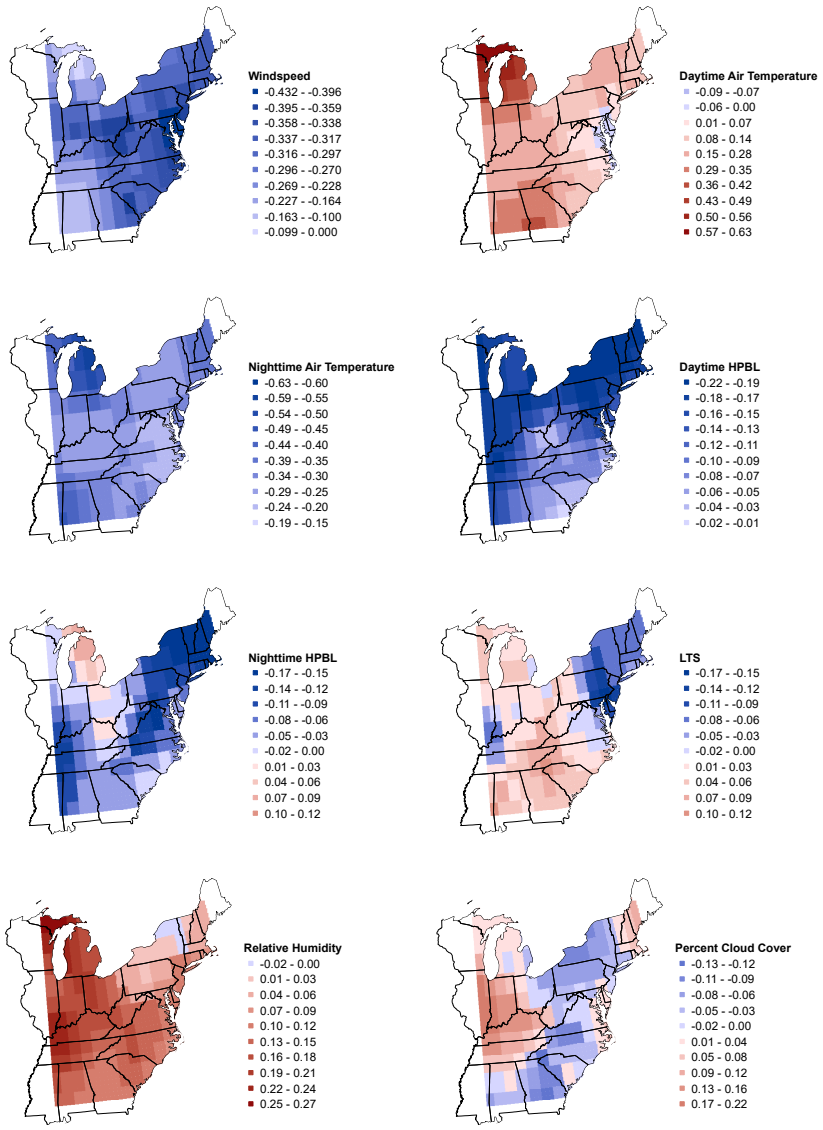


Figure 7.5: The figure depicts the posterior mean estimate of the covariate effect surfaces from the winter data for the 0.5 quantile.

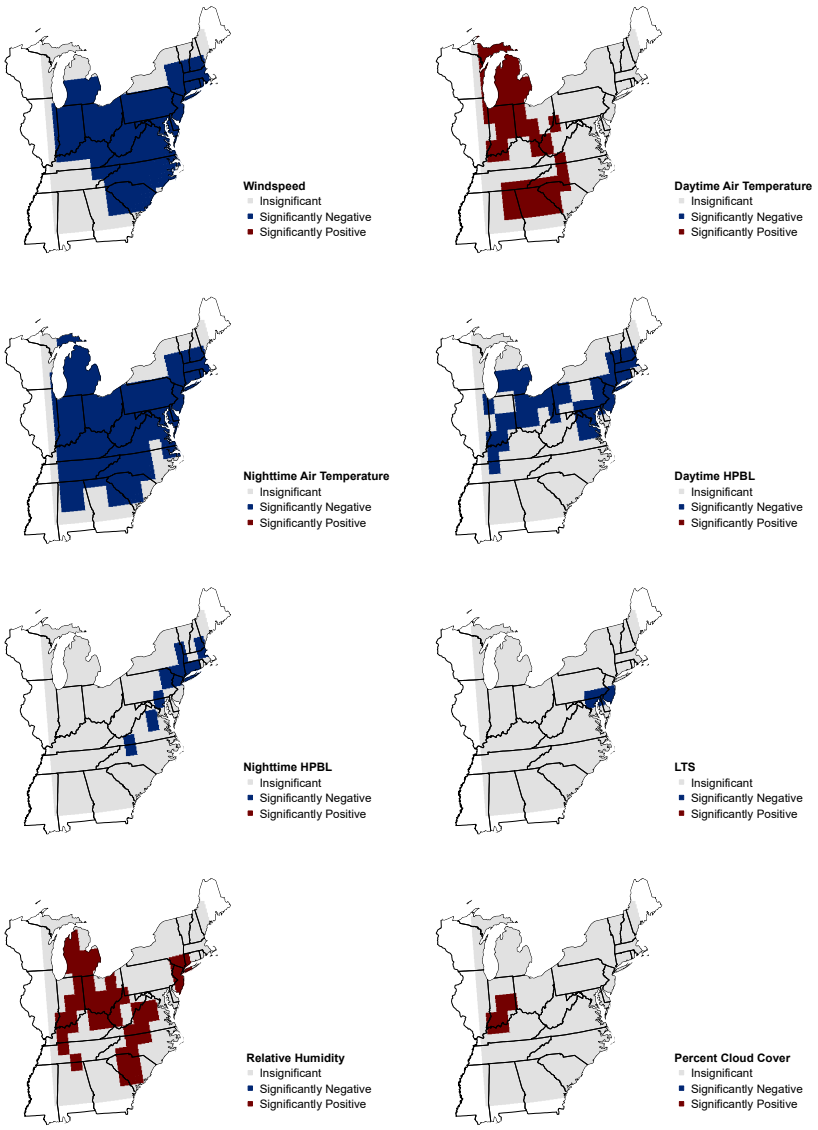


Figure 7.6: The figure depicts the regions of significance of the covariate effect surfaces from the winter data for the 0.5 quantile. Those regions for which a 95% credible interval contained 0 were deemed insignificant.

Figure 7.7 shows the estimated covariate effect surfaces for the winter $PM_{2.5}$ data at the 0.95 quantile. Figure 7.8 depicts regions of significance for these covariate surfaces. Wind speed appears to be the strongest driver of the 0.95 quantile of winter $PM_{2.5}$ levels, having a significantly negative relationship over much of the study area. Nighttime air temperature is significantly negatively

related to the 0.95 quantile of winter $PM_{2.5}$ in the midwest and portions of the Carolinas, while daytime HPBL is significantly negatively related for much of the northeastern portion of the study area. Daytime air temperature and relative humidity are significantly positively related to the 0.95 quantile of winter $PM_{2.5}$ levels in parts of the western and southern study regions. Nighttime HPBL and percent cloud cover are only significant in isolated areas, and LTS is not significant anywhere in the study region.

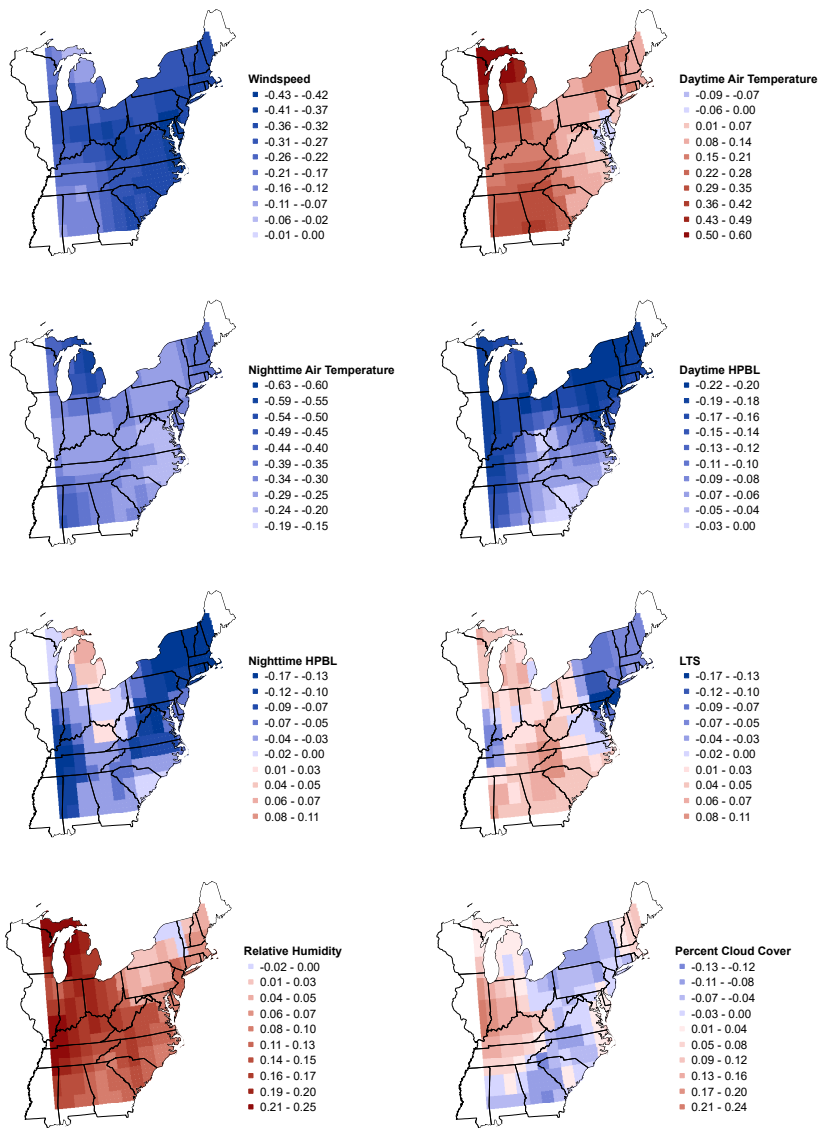


Figure 7.7: The figure depicts the posterior mean estimate of the covariate effect surfaces from the winter data for the 0.95 quantile.

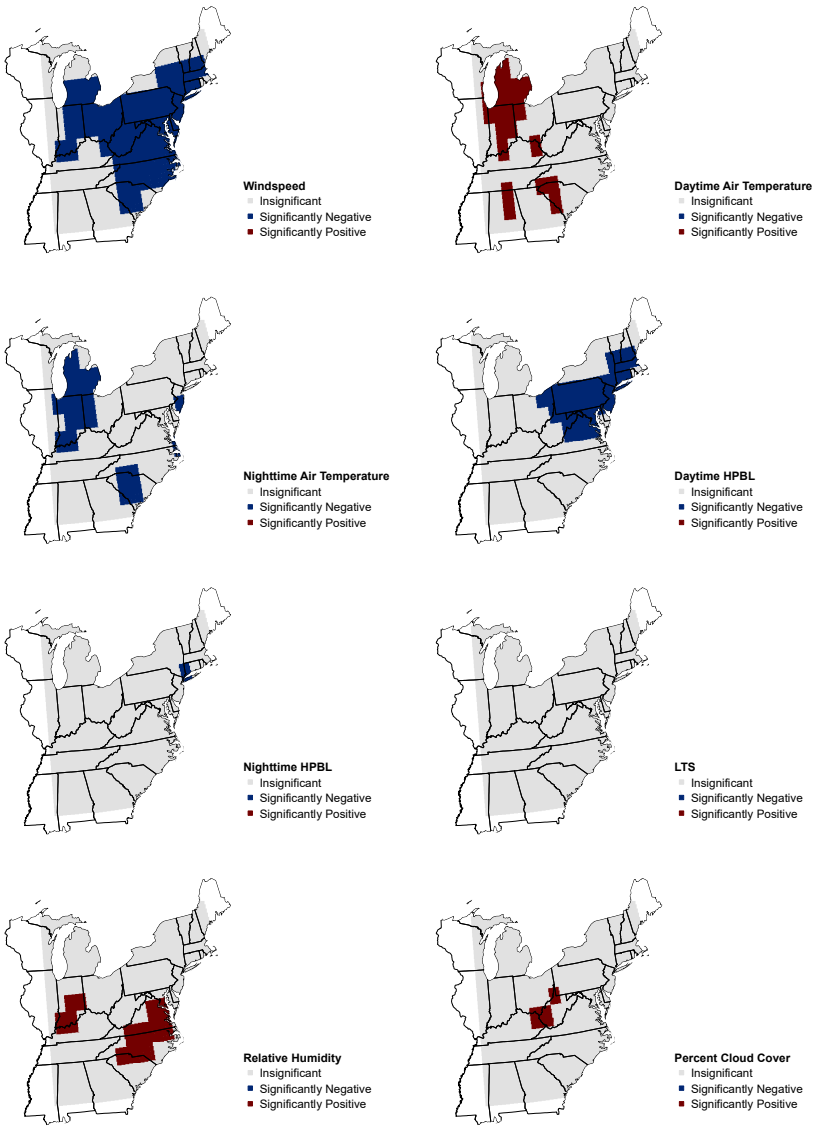


Figure 7.8: The figure depicts the regions of significance of the covariate effect surfaces from the winter data for the 0.95 quantile. Those regions for which a 95% credible interval contained 0 were deemed insignificant.

Figure 7.9 and Figure 7.10 show the estimated covariate surfaces and the regions of significance (respectively) for the summer data at the 0.5 quantile. We see that wind speed is significantly negatively related to median summer $PM_{2.5}$ levels over much of the study area, with nighttime air temperature and precipitation being significantly negatively related to the response in isolated

areas. Daytime air temperature is significantly positively related to median summer $PM_{2.5}$ levels over almost all of the study area. DSWRF is significantly positively associated with the response in mid-Atlantic region, and relative humidity is significantly positively related in the midwest and parts of New England. Percent cloud cover was not significantly related to median summer $PM_{2.5}$ levels anywhere in the study area.

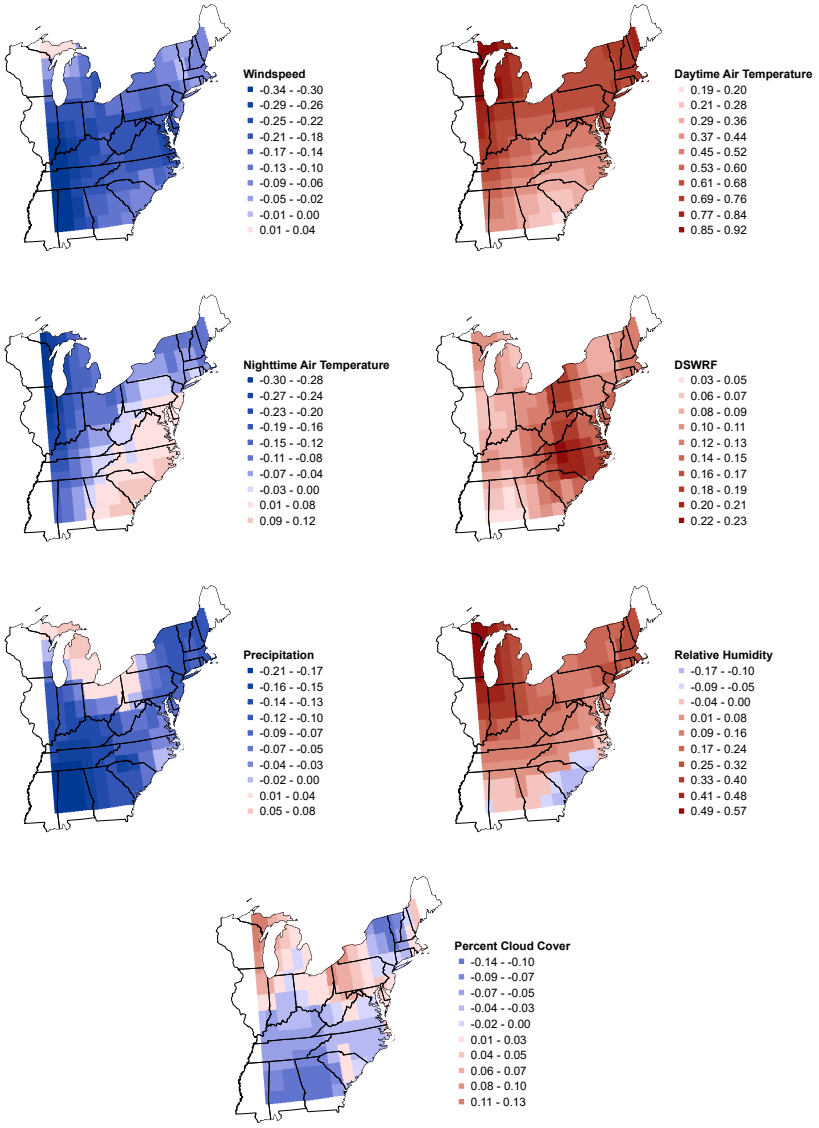


Figure 7.9: The figure depicts the posterior mean estimate of the covariate effect surfaces from the summer data for the 0.5 quantile.

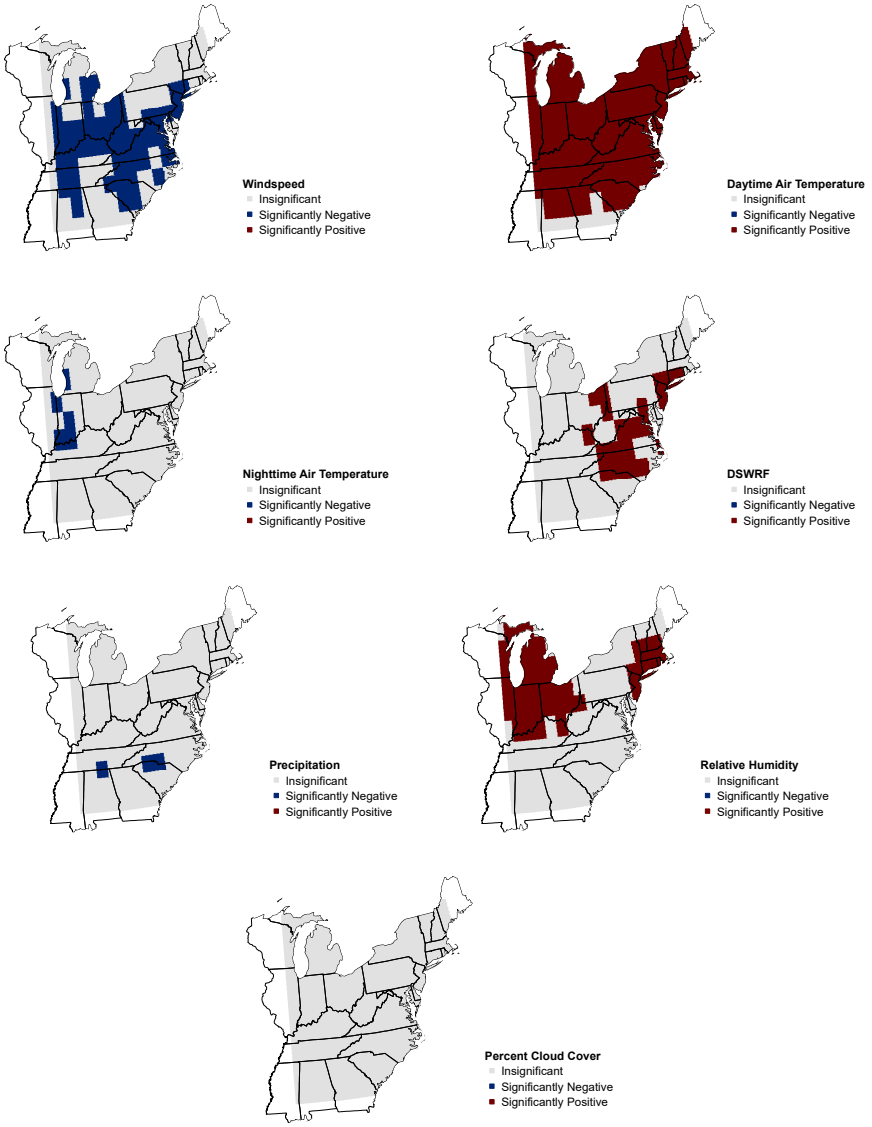


Figure 7.10: The figure depicts the regions of significance of the covariate effect surfaces from the summer data for the 0.5 quantile. Those regions for which a 95% credible interval contained 0 were deemed insignificant.

Figure 7.11 and Figure 7.12 show the estimated covariate surfaces and the regions of significance (respectively) for the summer data at the 0.95 quantile. We see that wind speed is significantly negatively related to high $PM_{2.5}$ levels in the central part of the study region. Daytime air temperature is significantly positively related to the response across most of the study region, and relative

humidity is significantly positively related across much of the northern portion of the study area. The other covariates (nighttime air temperature, DSWRF, precipitation and percent cloud cover) were insignificant across all or almost all of the study area.

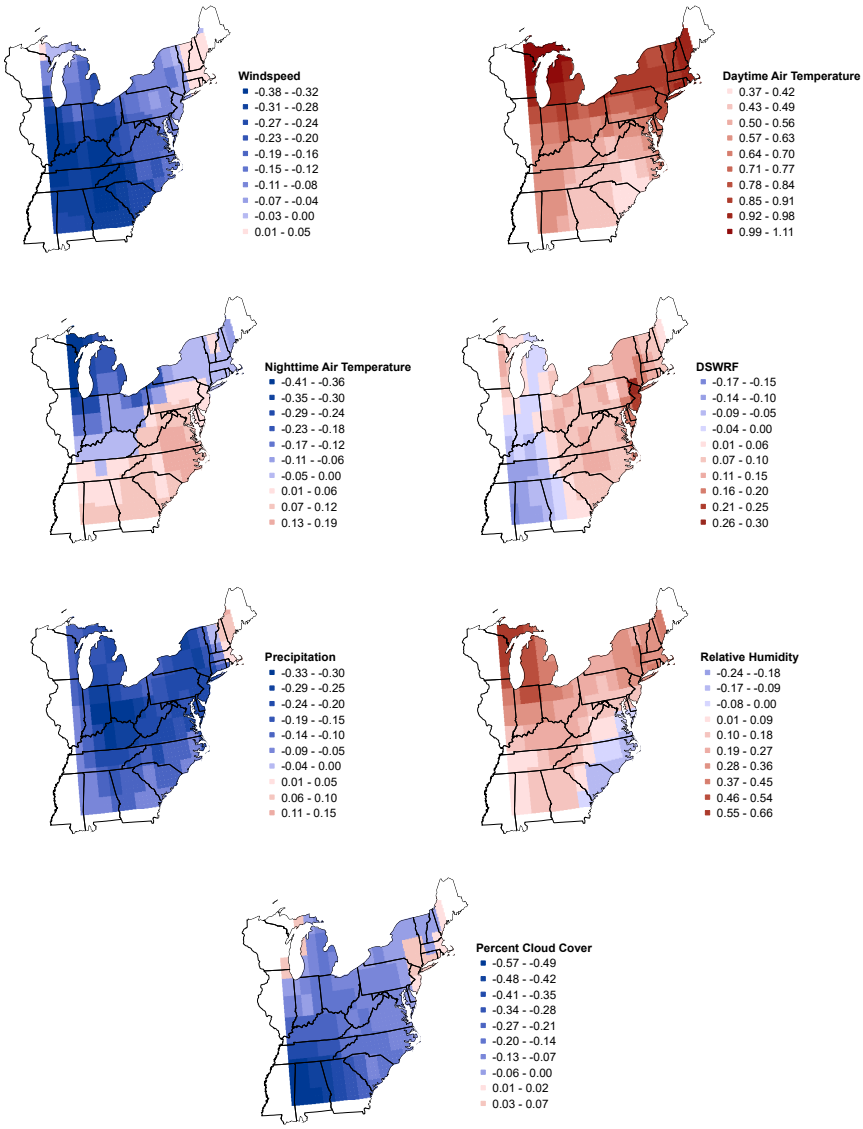


Figure 7.11: The figure depicts the posterior mean estimate of the covariate effect surfaces from the summer data for the 0.95 quantile.

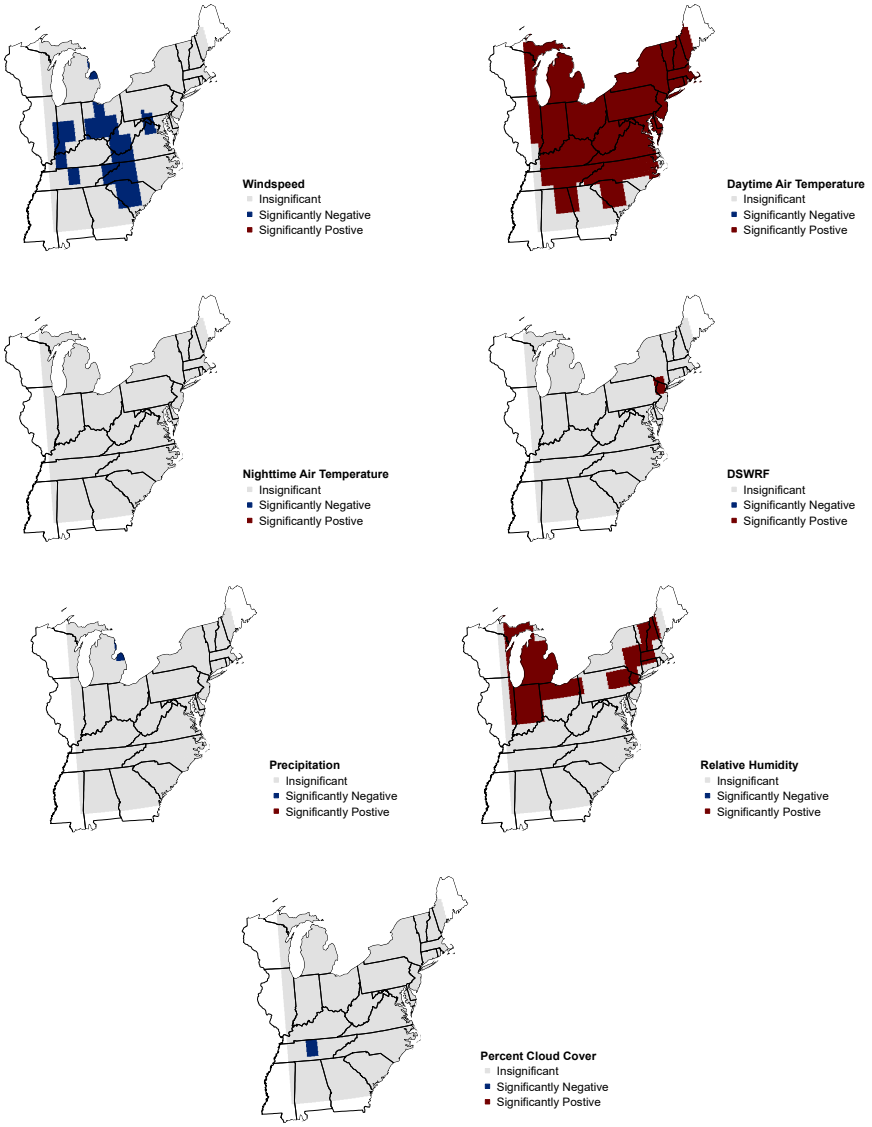


Figure 7.12: The figure depicts the regions of significance of the covariate effect surfaces from the summer data for the 0.95 quantile. Those regions for which a 95% credible interval contained 0 were deemed insignificant.

7.5 Discussion

We note that Russell et al. [2017] also perform a spatial quantile regression for the 0.5 and 0.95 quantiles of a $PM_{2.5}$ response over the eastern United States. However our method has two sub-

stantial advantages over that of Russell et al. [2017]: more reliable inference and reduced estimation uncertainty. Failing to model temporal dependence, when present, often leads to unreliable inference. In particular, neglecting temporal dependence tends to inflate variance estimators and alter the width of confidence intervals. While Russell et al. [2017] ignore the temporal dependence in the data, our method models temporal dependence via a vector autoregression. This approach yields more accurate inference because the random effects are able to capture the temporal autocorrelation in the data. By accounting for an additional source of variation in the data, our method reduces the overall uncertainty in parameter estimation. Our improved inference allows us to detect some relationships missed by Russell et al. [2017].

For the winter $PM_{2.5}$ data, our analysis detected a much larger region of significance for the 0.5 quantile for daytime air temperature than that of Russell et al. [2017]. Our analysis also found that for the 0.5 quantile, LTS was significantly negatively to the response for an isolated region while Russell et al. [2017] found that LTS was significantly positively related to the response for several isolated regions. For the 0.95 quantile, our analysis did not find LTS to be significant anywhere, while Russell et al. [2017] found LTS to be significantly positively related for much of the study area. It is likely that the LTS covariate effect in Russell et al. [2017]'s model was confounding with the unmodeled temporal dependence. Our model, which accounts for the temporal dependence, was able to distinguish between the two. Both analyses found a significant positive relationship between daytime air temperature and the 0.95 quantile of the response in Michigan; however our analysis also found regions of positive significance in the lower Midwest and South, while Russell et al. [2017] found a significantly negative relationship in Eastern Pennsylvannia. Again, we hypothesis this spurious relationship was due to unmodeled temporal dependence. Russell et al. [2017] found a somewhat larger region of positive significance at the 0.95 quantile for daytime HPBL than did our analysis.

Our findings for the summer analysis identify several relationships not found in Russell et al. [2017]. Our analysis found a small region of negative significance in the Midwest for nighttime air temperature at the 0.5 quantile, while Russell et al. [2017] did not find any regions of significance. Both analyses found relative humidity to be significantly positively related to the response at the 0.5 quantile in the Midwest and New England, but Russell et al. [2017] also found relative humidity to be significantly negatively related to the response in a few isolated Southern areas. Our analysis found two isolated regions of negative significance for precipitation at the 0.5 quantile, while Russell

et al. [2017] found no regions of significance for precipitation. For the 0.95 quantile, Russell et al. [2017] found a much larger region of negative significance for percent cloud cover than our analysis did. Both analysis found relative humidity to be significantly positively related to the response at the 0.95 quantile in the Midwest and New England, but Russell et al. [2017] also found relative humidity to be significantly negatively related to the response in isolated Southern areas. It is possible that these areas of negative significance are spurious relationships caused by confounding with unmodeled temporal dependence. Our analysis found precipitation to be largely insignificant at the 0.95 quantile except for a small region in Michigan, while Russell et al. [2017] found precipitation to be significantly negatively related to the response in Kentucky and Tennessee.

7.6 Conclusion

We have developed a Bayesian quantile regression methodology that allows us to estimate the spatially varying effects of meteorological covariates on $PM_{2.5}$. We quantify the relationship between $PM_{2.5}$ and various meteorological covariates for both summer and winter data at the 0.5 and 0.95 quantile. Spatio-temporal random effects are included to account for the spatio-temporal dependence in the data. Our findings include several new relationships not found by Russell et al. [2017], who performed a similar spatial analysis while ignoring the temporal dependence in the data.

Our Bayesian methodology for estimating spatially varying covariate effects for quantile regression uses multidimensional trend filtering to estimate spatially varying covariate effect surfaces. The trend filter is also used to estimate spatio-temporal random effects. The model is fit via MCMC methods. Re-expressing the asymmetric Laplace errors as a combination of normal and exponential random variables allows for the construction of an efficient Gibbs-based MCMC sampling procedure. A simulation study is used to assess the performance of the model and its fitting procedure. We find that our methodology performs well, even when a large amount of data is missing.

This analysis presents several areas for potential future work. One might attempt to perform a single analysis for all the data, rather than analyzing summer and winter data separately. The multidimensional trend filter could be extended to three dimensions to allow the covariate effects to change with time as well as with space. One might also estimate effects for multiple quantiles simultaneously.

Chapter 8

Conclusion

8.1 Discussion

This work summarizes several methodologies for Bayesian spatio-temporal modeling. Chapters 2 and 3 provide disease forecasting methodologies for Lyme disease and Anaplasmosis. These methodologies produce an annual disease forecast for the contiguous United States, and allow for the assessment of significant drivers of disease. Chapter 4 outlines the chromatic sampler for Gaussian Markov random fields, a fast MCMC sampling strategy for fitting models with high dimensional GMRFs. The chromatic sampler is compared to two of the most commonly used GMRF sampling strategies: block updating and single site updating. The chromatic sampler is found to be the fastest approach when the dimension of the underlying GMRF is sufficiently large. Chapter 5 outlines a Bayesian spatio-temporal model for assessing disease trends. This model is then used to identify regions of the country seeing increasing Lyme disease risk. The methodology allows for the simultaneous estimation of both regional and local trends, and employs data augmentation and the chromatic sampler to create a computationally efficient sampling strategy. Chapters 6 and 7 examine a novel nonparametric estimator, the Bayesian multidimensional trend filter. The multidimensional trend filter provides a computationally efficient way to estimate trends or covariate effects for spatio-temporal models with a large support space, though the methodology is not limited to spatio-temporal applications. Chapter 6 develops the methodology, and provides an adaptive, data-driven way for determining the level of resolution of the resulting estimator. Chapter 7 extends the methodology to the quantile regression framework to estimate spatially varying covariate effects for regression quantiles. The

methodology is then used to estimate the spatially varying meteorological drivers of extreme $PM_{2.5}$ events.

Appendices

Appendix A Supplementary Material for Chapter 4

Here the reader may find additional material relating to Chapter 4, including more details about the binomial regression model for the election data, as well as Supplementary Figures discussed in the main text.

A.1 SUPPLEMENTARY ALGORITHMS AND FIGURES

Input: Mean factor $\boldsymbol{\beta} b$, precision matrix $\boldsymbol{\beta} Q_p$.
Output: Draw $\boldsymbol{\beta} x$ from a $N(\boldsymbol{\beta} Q_p^{-1} \boldsymbol{\beta} b, \boldsymbol{\beta} Q_p^{-1})$ distribution.

1. Find the Cholesky factor $\boldsymbol{\beta} Q_p = \boldsymbol{\beta} L L^T$
2. Solve $\boldsymbol{\beta} L w = \boldsymbol{\beta} b$
3. Solve $\boldsymbol{\beta} L^T \boldsymbol{\beta} \mu = \boldsymbol{\beta} w$
4. Sample $\boldsymbol{\beta} z \sim N(\boldsymbol{\beta} 0, \boldsymbol{\beta} I)$
5. Solve $\boldsymbol{\beta} L^T \boldsymbol{\beta} v = \boldsymbol{\beta} z$
6. Compute $\boldsymbol{\beta} x = \boldsymbol{\beta} \mu + \boldsymbol{\beta} v$
7. **Return** $\boldsymbol{\beta} x$

Algorithm 3: Sampling from a typical GMRF-based full conditional encountered in block Gibbs sampling [Rue and Held, 2005a].

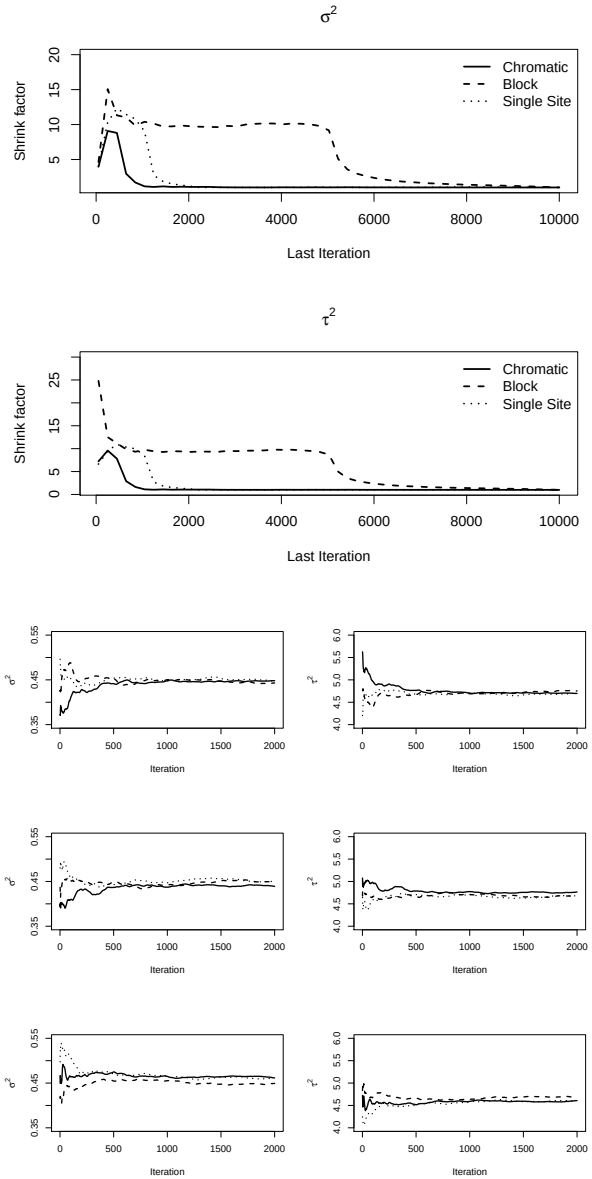


Figure A.1: Left panel: Gelman plots [Brooks and Gelman, 1998] of the potential scale reduction factors versus chain length for the 50×50 regular array example. Right panel: Cumulative averages $\widehat{\sigma^2}^{(k)}$ and $\widehat{\tau^2}^{(k)}$, $k = 1, \ldots, 2,000$ calculated from three independent chains. In the right panel, the top, middle, and bottom rows correspond to chromatic, block, and single-site sampling, respectively.

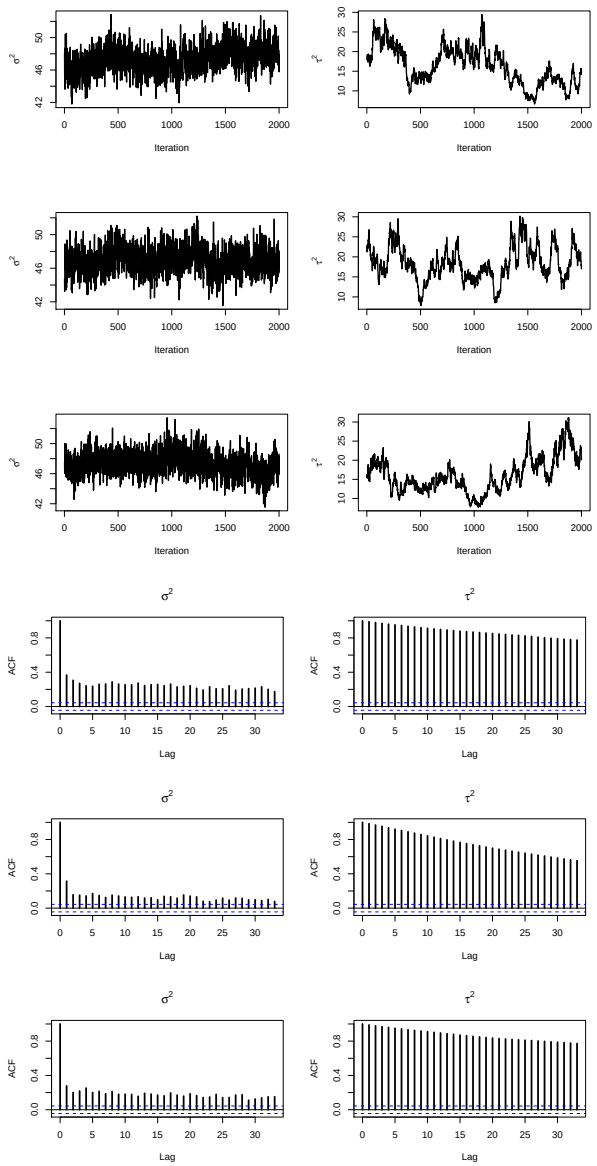


Figure A.2: Left panel: MCMC Trace plots of single chains each for σ^2 and τ^2 for the noisy 50×50 regular array example. Right panel: Empirical ACF plots for these chains. The top, middle, and bottom rows are from the chromatic, block, and single-site chains, respectively.

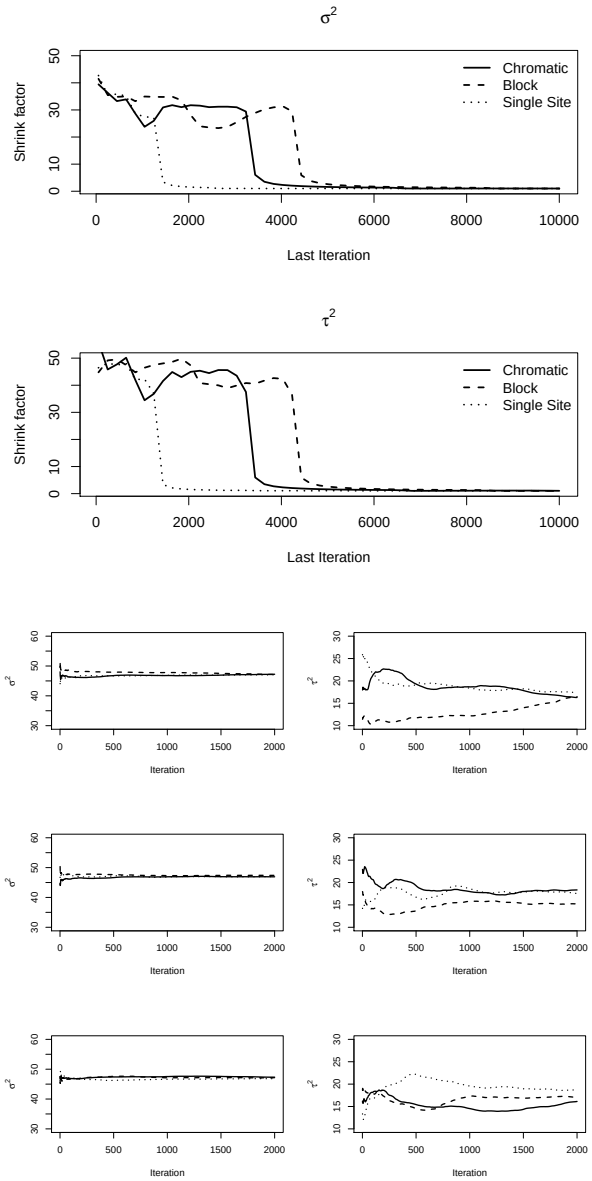


Figure A.3: Left panel: Gelman plots [Brooks and Gelman, 1998] of the potential scale reduction factors versus chain length for the noisy 50×50 regular array example. Right panel: Cumulative averages $\widehat{\sigma^2}^{(k)}$ and $\widehat{\tau^2}^{(k)}$, $k = 1, \ldots, 2000$ calculated from three independent chains. In the right panel, the top, middle, and bottom rows correspond to chromatic, block, and single-site sampling, respectively.

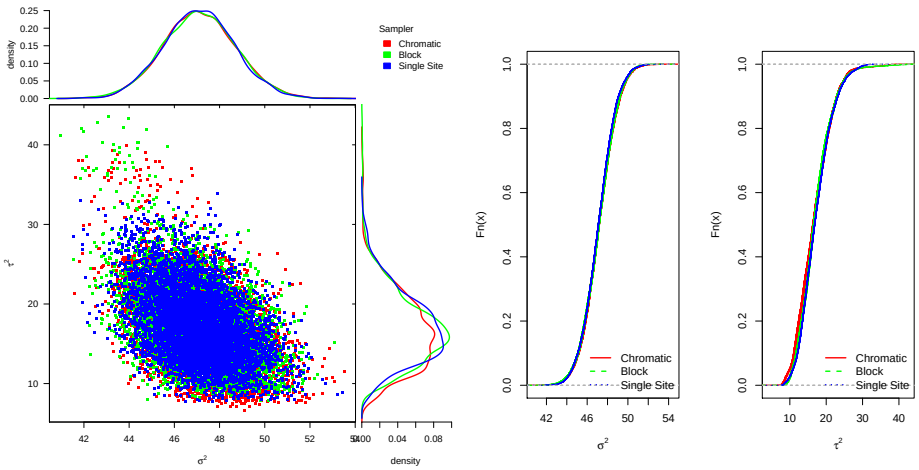


Figure A.4: Left panel: Scatterplot and approximate marginal posterior densities estimated from the three sampling approaches for the noisy 50×50 regular array example. Right panel: Empirical CDFs based on the output. The left panel was created using code available at `https://github.com/ChrKoenig/R_marginal_plot`.

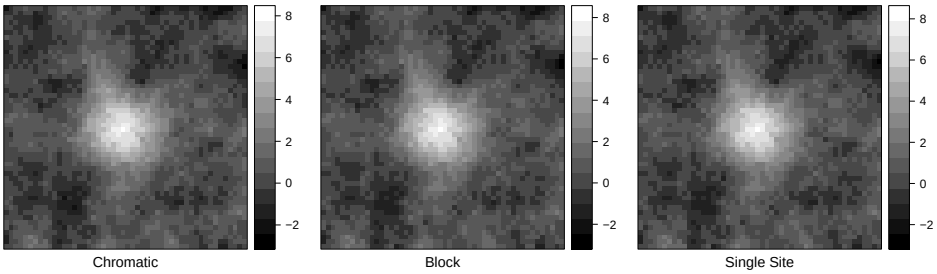


Figure A.5: Posterior mean estimates of the the true underlying image obtained from each sampling approach in the 50×50 regular array example. The multivariate potential scale reduction factors for the $\boldsymbol{\beta}\varphi := \beta_0\boldsymbol{\beta}1 + \boldsymbol{\beta}\gamma$ chains are approximately 1 for each sampler.

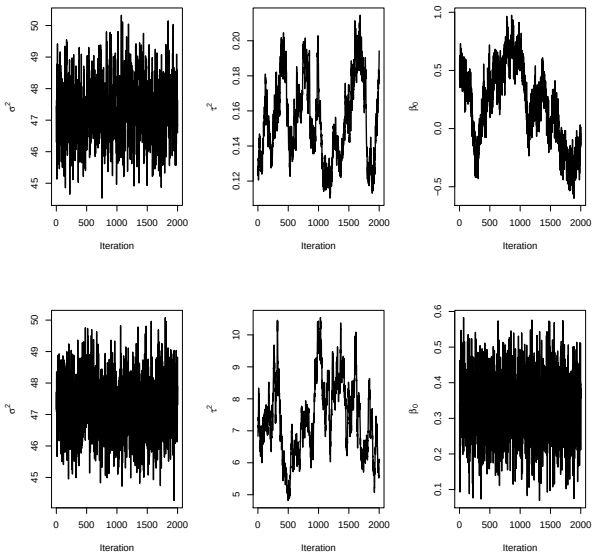


Figure A.6: MCMC Trace plots of single chains each for σ^2, τ^2, and β_0 chains in the noisy 80×80 regular array example. The top and bottom rows are from the chromatic and block chains, respectively.

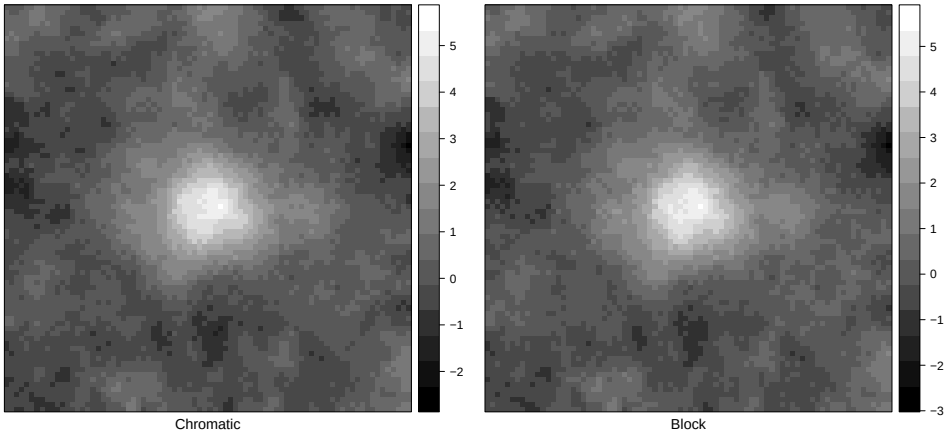


Figure A.7: Posterior mean estimates of the the true underlying image obtained from the chromatic and block sampling approaches in the noisy 80×80 regular array example. The multivariate potential scale reduction factors for the $\boldsymbol{\beta}\varphi := \beta_0 \boldsymbol{\beta} 1 + \boldsymbol{\beta}\gamma$ chains are approximately 1.00 for both samplers.

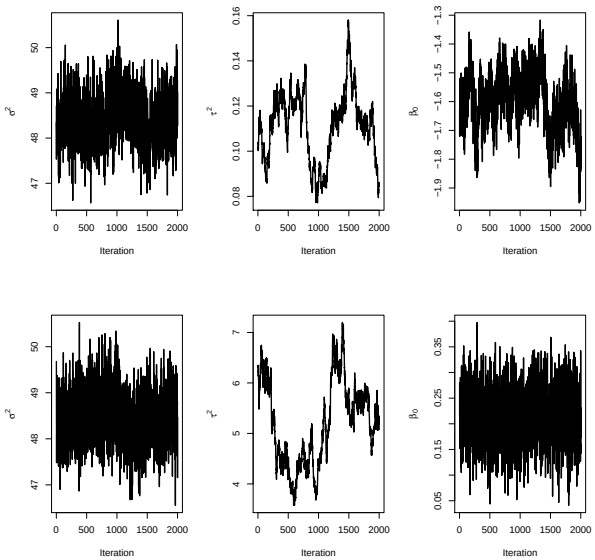


Figure A.8: MCMC Trace plots of single chains each for σ^2, τ^2, and β_0 chains in the noisy 128×128 regular array example. The top and bottom rows are from the chromatic and block chains, respectively.

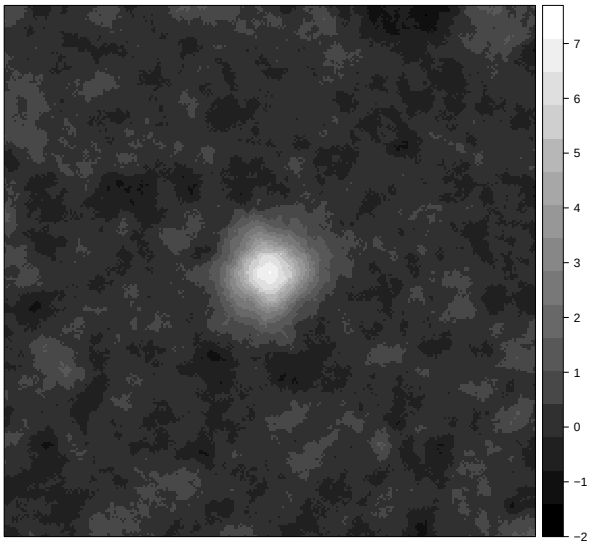


Figure A.9: Posterior mean estimate of the the true underlying image obtained from chromatic sampling in the noisy 256×256 regular array example. The estimate is based on the last 2,000 of 10,000 iterations of the Markov chain.

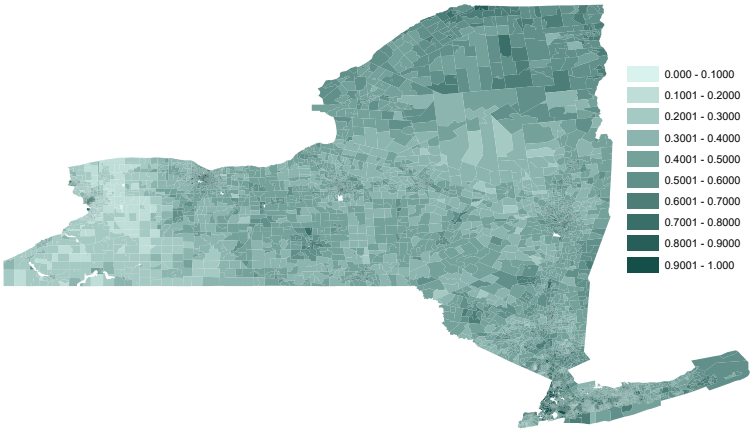


Figure A.10: The raw data from the New York election example. The figure displays the total number of votes for the Democratic candidate in each precinct divided by the total number of votes from that precinct.

Figure A.11: 7-Coloring of the $14,926$ precincts in New York found via Algorithm 2.

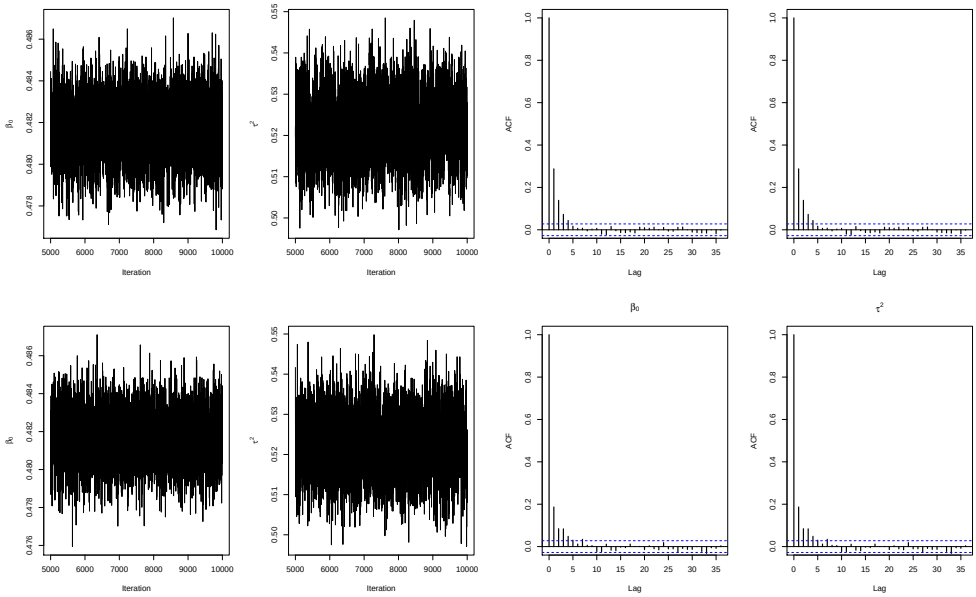


Figure A.12: Left panel: MCMC trace plots of single chains for the β_0 and τ^2 chains from the New York election data example. Right panel: Empirical ACF plots for these chains. The top and bottom rows are from the chromatic and block chains, respectively.

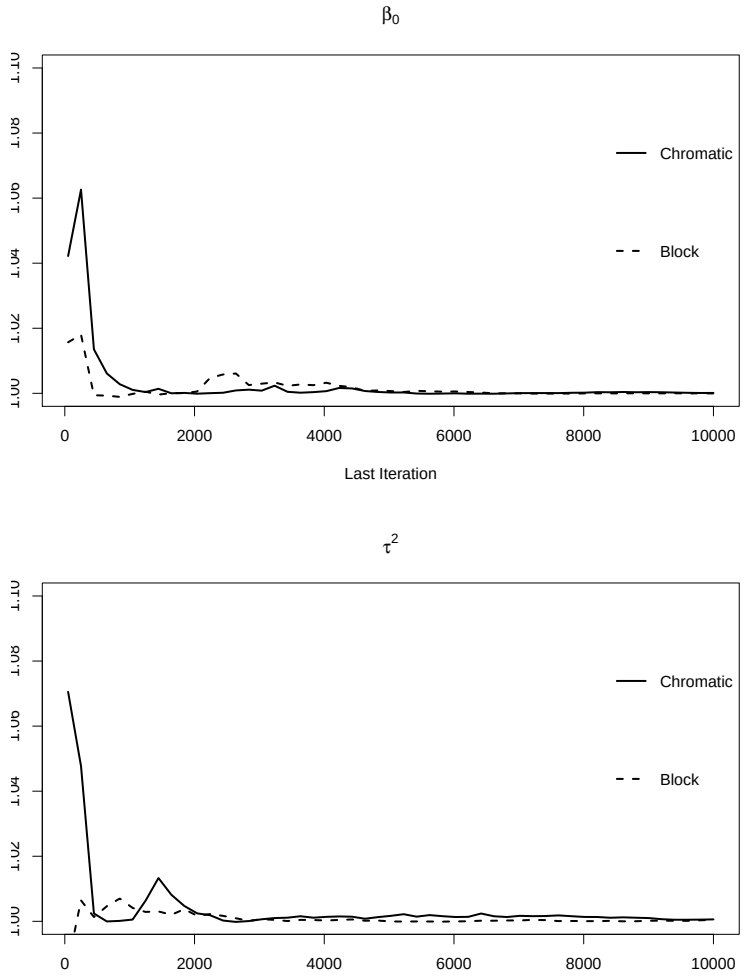


Figure A.13: Gelman plots [Brooks and Gelman, 1998] of the potential scale reduction factors versus chain length for the New York election example.

A.2 DETAILS PERTAINING TO THE SPATIAL BINOMIAL REGRESSION MODEL

It is assumed that $\boldsymbol{\gamma} = (\gamma_1, ..., \gamma_n)^T$ obeys a CAR model, which is given by $N(\mathbf{0}, \tau^2(\boldsymbol{D} - \rho\boldsymbol{W})^{-1})$, with ρ being known. The model is completed by specifying the following priors: $\beta_0 \sim N(0, \sigma_0^2)$ and $\tau^2 \sim IG(\alpha_{\tau^2}, \beta_{\tau^2})$, where the hyperparameters are chosen so these priors are vague. Thus, the unknown parameters that are to be sampled via Markov Chain Monte Carlo (MCMC) are β_0, $\boldsymbol{\psi}$, $\boldsymbol{\gamma}$, and τ^2, with the only difference between the full block Gibbs and chromatic sampler being how the $\boldsymbol{\gamma}$ are sampled; i.e., the former samples all of these elements in a single block while the latter samples independent blocks of these elements from their univariate full conditionals.

To create a posterior sampling algorithm, it is first noted that the full conditional distribution of β_0 is given by $\beta_0|\boldsymbol{\psi}, \boldsymbol{\gamma}, \boldsymbol{Y} \sim N(\mu_{\beta_0}, \sigma^2_{\beta_0})$, where $\mu_{\beta_0} = \sigma^2_{\beta_0}(\mathbf{1}^T\boldsymbol{\kappa} - \mathbf{1}^T\boldsymbol{D}_{\boldsymbol{\psi}}\boldsymbol{\psi})$ and $\sigma^2_{\beta_0} = (\mathbf{1}^T\boldsymbol{D}_{\boldsymbol{\psi}}\mathbf{1} + \sigma_0^{-2})^{-1}$. Next, the full conditional distribution of τ^2 is given by $\tau^2|\boldsymbol{\gamma} \sim IG(\alpha_{\tau^2} + n/2,\ \beta_{\tau^2} + \boldsymbol{\gamma}^T(\boldsymbol{D} - \rho\boldsymbol{W})\boldsymbol{\gamma}/2)$. The full conditional distribution of the latent ψ_i are again Pólya-Gamma; i.e., $\psi_i|\beta_0, \gamma_i \sim PG(m_i, \eta_i)$, for $i = 1, ..., n$. For further details, see Polson et al. [2013]. The full conditional distribution of $\boldsymbol{\gamma}$ is $\boldsymbol{\gamma}|\boldsymbol{Y}, \beta_0, \tau^2, \boldsymbol{\psi} \sim N(\boldsymbol{\mu}_{\boldsymbol{\gamma}}, \boldsymbol{\Sigma}_{\boldsymbol{\gamma}})$, where $\boldsymbol{\mu}_{\boldsymbol{\gamma}} = \boldsymbol{\Sigma}_{\boldsymbol{\gamma}}(\boldsymbol{\kappa} - \boldsymbol{D}_{\boldsymbol{\psi}}\mathbf{1}\beta_0)$ and $\boldsymbol{\Sigma}_{\boldsymbol{\gamma}} = \{\boldsymbol{D}_{\boldsymbol{\psi}} + \tau^{-2}(\boldsymbol{D} - \rho\boldsymbol{W})\}^{-1}$. Through similar arguments, it is easy to show that the univariate full conditional distribution of γ_i is given by $\gamma_i|\boldsymbol{\gamma}_{(-i)}, Y_i, \beta_0, \tau^2, \psi_i \sim N(\mu_{\gamma_i}, \sigma^2_{\gamma_i})$, where $\mu_{\gamma_i} = \sigma^2_{\gamma_i}(\tau^{-2}\rho\sum_{j\in\mathcal{N}(i)}\gamma_j - \psi_i\beta_0 + \kappa_i)$ and $\sigma^2_{\gamma_i} = (\psi_i + \tau^{-2}\boldsymbol{D}_{ii})^{-1}$.

Appendix B Supplementary Material for Chapter 5

B.1 Full Conditional Distributions

This section provides the full conditional distributions of all unknown parameters from the model in chapter 5. To provide for the most general setting possible, these derivations account for possible missing data in time and space. The notation defined in Section 5.3.2 is used. For ease of exposition in the following derivations, the dependencies of these full conditional distributions on remaining parameters are suppressed.

The full conditional distribution of the latent ψ_{st}'s is Pólya-Gamma(n_{st}, ν_{st}). For further details on the structure of the Pólya-Gamma distribution, as well as sampling strategies, see Polson et al. [2013]. Exploiting the normal form in $\boldsymbol{\nu}$, it is easy to establish that the full conditional distribution of $\boldsymbol{\beta}_p^*$ is $\mathrm{N}(\boldsymbol{\mu}_{\boldsymbol{\beta}_p^*}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}_p^*})$, where

$$\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{\beta}_p^*} &= \boldsymbol{\Sigma}_p \left\{ \boldsymbol{T}_p(\mathcal{R})' \boldsymbol{X}_p(\mathcal{R})' \boldsymbol{\kappa} - \boldsymbol{T}_p(\mathcal{R})' \boldsymbol{X}_p(\mathcal{R})' \boldsymbol{D}_{\boldsymbol{\psi}} \boldsymbol{\nu}_{\boldsymbol{\beta}_p^*} \right\}, \\
\boldsymbol{\Sigma}_{\boldsymbol{\beta}_p^*} &= \left\{ \boldsymbol{T}_p(\mathcal{R})' \boldsymbol{X}_p(\mathcal{R})' \boldsymbol{D}_{\boldsymbol{\psi}} \boldsymbol{X}_p(\mathcal{R}) \boldsymbol{T}_p(\mathcal{R}) + \boldsymbol{C}_p^{*-1} \right\}^{-1}.
\end{aligned}$$

In the expression above, we define $\boldsymbol{T}_p$ by stacking the matrix $\widetilde{\boldsymbol{R}}_p^*(\boldsymbol{R}_p^*)^{-1}$ on itself T times, $\odot$ denotes the Hadamard product, and

$$\boldsymbol{\nu}_{\boldsymbol{\beta}_p^*} = \boldsymbol{Z}(\mathcal{R})\boldsymbol{\delta} + \sum_{p' \neq p} \boldsymbol{X}_{p'}(\mathcal{R}) \odot \boldsymbol{T}_{p'}(\mathcal{R}) \boldsymbol{\beta}_{p'}^* + \boldsymbol{I}(\mathcal{R})\boldsymbol{\xi}.$$

Similarly, the full conditional distribution of $\boldsymbol{\delta}$ is $\mathrm{N}(\boldsymbol{\mu}_{\boldsymbol{\delta}}, \boldsymbol{\Sigma}_{\boldsymbol{\delta}})$, where

$$\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{\delta}} &= \boldsymbol{\Sigma}_{\boldsymbol{\delta}} \left\{ \boldsymbol{Z}(\mathcal{R})' \boldsymbol{\kappa} - \boldsymbol{Z}(\mathcal{R})' \boldsymbol{D}_{\boldsymbol{\psi}} \boldsymbol{\nu}_{\boldsymbol{\delta}} \right\}, \\
\boldsymbol{\Sigma}_{\boldsymbol{\delta}} &= \left\{ \boldsymbol{Z}(\mathcal{R})' \boldsymbol{D}_{\boldsymbol{\psi}} \boldsymbol{Z}(\mathcal{R}) + \sigma_{\boldsymbol{\delta}}^{-2} \boldsymbol{I} \right\}^{-1},
\end{aligned}$$

and here

$$\boldsymbol{\nu}_{\boldsymbol{\delta}} = \sum_{p=1}^{P} \boldsymbol{X}_p(\mathcal{R}) \odot \boldsymbol{T}_p(\mathcal{R}) \boldsymbol{\beta}_p^* + \boldsymbol{I}(\mathcal{R})\boldsymbol{\xi}.$$

The full conditional distribution of σ_p^2 is $\mathrm{IG}\Big(S_p^*/2 + \alpha_{\sigma_p^2}, \boldsymbol{\beta}_p^* \boldsymbol{R}_p^{*-1} \boldsymbol{\beta}_p^*/2 + \upsilon_{\sigma_p^2}\Big)$ and the full condi-

tional distribution of τ^2 is $\text{IG}(\alpha^*_{\tau^2}, \beta^*_{\tau^2})$, where

$$\alpha^*_{\tau^2} = TS/2 + \alpha_{\tau^2},$$
$$\beta^*_{\tau^2} = \frac{1}{2}\left(\boldsymbol{\xi}_1^T\{\boldsymbol{D} - \omega\boldsymbol{W}\}\boldsymbol{\xi}_1 + \sum_{t=2}^{T}\{\boldsymbol{\xi}_t - \zeta\boldsymbol{\xi}_{t-1}\}^T\{\boldsymbol{D} - \omega\boldsymbol{W}\}\{\boldsymbol{\xi}_t - \zeta\boldsymbol{\xi}_{t-1}\}\right) + \upsilon_{\tau^2}.$$

The full conditional distribution of ζ is Truncated-Normal$(\mu^*_\zeta, \sigma^*_\zeta, -1, 1)$, where

$$\mu^*_\zeta = \sigma^{2*}_\zeta \sum_{t=2}^{T} \boldsymbol{\xi}'_{t-1}\boldsymbol{A}\boldsymbol{\xi}_t,$$
$$\sigma^{2*}_\zeta = \left(\sum_{t=2}^{T} \boldsymbol{\xi}'_{t-1}\boldsymbol{A}\boldsymbol{\xi}_{t-1} + \sigma_\zeta^{-2}\right)^{-1},$$

and $\boldsymbol{A} = \tau^{-2}(\boldsymbol{D} - \omega\boldsymbol{W})$. We turn our attention to the spatio-temporal random effects. The full conditional distribution of $\boldsymbol{\xi}_t$ is $\text{N}(\boldsymbol{\mu}_{\boldsymbol{\xi}_t}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_t})$, where

$$\boldsymbol{\mu}_{\boldsymbol{\xi}_t} = \boldsymbol{\Sigma}_{\boldsymbol{\xi}_t}\left\{\boldsymbol{I}(\mathcal{R}_t)'\boldsymbol{\kappa}_t - \boldsymbol{I}(\mathcal{R}_t)'\boldsymbol{D}_{\boldsymbol{\psi}_t}\boldsymbol{\nu}_{\boldsymbol{\xi}_t} + \zeta\boldsymbol{A}(\boldsymbol{\xi}_{t-1} + \boldsymbol{\xi}_{t+1})\right\},$$
$$\boldsymbol{\Sigma}_{\boldsymbol{\xi}_t} = \left\{\boldsymbol{I}(\mathcal{R}_t)'\boldsymbol{D}_{\boldsymbol{\psi}_t}\boldsymbol{I}(\mathcal{R}_t) + (1+\zeta^2)\boldsymbol{A}\right\}^{-1} \text{ for } t = 1, ..., T-1,$$
$$\boldsymbol{\Sigma}_{\boldsymbol{\xi}_t} = \left\{\boldsymbol{I}(\mathcal{R}_t)'\boldsymbol{D}_{\boldsymbol{\psi}_t}\boldsymbol{I}(\mathcal{R}_t) + \boldsymbol{A}\right\}^{-1}, \text{for } t = T.$$

Here, $\mathcal{R}_t$ is the collection of ordered pairs (s, t) indexing data that were collected during the tth time period, $\boldsymbol{I}$ is an identity matrix, $\boldsymbol{D}_{\boldsymbol{\psi}_t} = \text{diag}\{\boldsymbol{\psi}(\mathcal{R}_t)\}$, $\boldsymbol{\kappa}_t = \boldsymbol{Y}(\mathcal{R}_t) - \boldsymbol{n}(\mathcal{R}_t)/2$, $\boldsymbol{\xi}_0 = \boldsymbol{\xi}_{T+1} = \boldsymbol{0}$, and

$$\boldsymbol{\nu}_{\boldsymbol{\xi}_t} = \boldsymbol{Z}(\mathcal{R}_t)\boldsymbol{\delta} + \sum_{p=1}^{P} \boldsymbol{X}_p(\mathcal{R}_t) \odot \boldsymbol{T}_p(\mathcal{R}_t)\boldsymbol{\beta}^*_p.$$

The aforementioned full conditional distributions of $\boldsymbol{\xi}_t$ can be used if the number of spatial units is relatively small. When faced with a large number of areal units, it is suggested that the chromatic sampler discussed in Brown et al. [2017c] be used. To accomplish this, one need only have a "coloring" of the areal units and know the full conditional distributions of the ξ_{st}. A coloring of the areal units can be obtained using the algorithm in Brown et al. [2017c]. The (scalar) full conditional

distribution of each ξ_{st} is $\mathrm{N}(\mu_{\xi_{st}}, \sigma^2_{\xi_{st}})$, where

$$\mu_{st} = \frac{\sigma^2_{st}}{\tau^2}\left\{\omega(1+\zeta^2)\boldsymbol{w}_s\boldsymbol{\xi}_t - g_{st}\tau^2(\psi_{st}\nu^*_{st} - \kappa_{st}) + \zeta d_s(\xi_{s,t-1} + \xi_{s,t+1}) - \zeta\omega\boldsymbol{w}_s(\boldsymbol{\xi}_{t-1} + \boldsymbol{\xi}_{t+1})\right\}$$
$$\sigma^2_{st} = \left\{g_{st}\psi_{st} + \frac{d_s(1+\zeta^2)}{\tau^2}\right\}^{-1},$$

for $t = 1, ..., T-1$, and

$$\mu_{st} = \frac{\sigma^2_{st}}{\tau^2}\left\{\omega\boldsymbol{w}_s\boldsymbol{\xi}_t - g_{st}\tau^2(\psi_{st}\nu^*_{st} - \kappa_{st}) + \zeta d_s\xi_{s,t-1} - \zeta\omega\boldsymbol{w}_s\boldsymbol{\xi}_{t-1}\right\}$$
$$\sigma^2_{st} = \left\{g_{st}\psi_{st} + \frac{d_s}{\tau^2}\right\}^{-1}.$$

for $t = T$. Here, $\boldsymbol{w}_s$ denotes the sth row of $\boldsymbol{W}$, d_s denotes the sth diagonal element of $\boldsymbol{D}$, $\nu^*_{st} = \boldsymbol{Z}_{st}\boldsymbol{\delta} + \sum_{p=1}^{P} X_{stp}\tilde{\beta}_p(\boldsymbol{l}_s)$, and $g_{st} = 1$ if $(s,t) \in \mathcal{R}$ and zero otherwise. The full conditional distribution of $\boldsymbol{\theta}_p$ has density

$$\exp\left(-\frac{1}{2}\left\{\widetilde{\boldsymbol{\beta}}'_p \odot \boldsymbol{X}_p(\mathcal{R})'\left[\boldsymbol{D}_{\boldsymbol{\psi}}\boldsymbol{X}_p(\mathcal{R}) \odot \widetilde{\boldsymbol{\beta}}_p + 2\left(\boldsymbol{D}_{\boldsymbol{\psi}}\boldsymbol{\nu}_{\boldsymbol{\beta}^*_p} - \boldsymbol{\kappa}\right)\right] + \boldsymbol{\beta}^{*T}_p\boldsymbol{C}^{*-1}_p\boldsymbol{\beta}^*_p\right\}\right)\det\{\boldsymbol{C}^*_p\}^{\frac{-1}{2}},$$

where the remaining parameters in the conditioning are understood. In general, a Metropolis-Hastings step can be used to sample $\boldsymbol{\theta}_p$, but the appropriate proposal distribution depends on the form of the selected correlation function.

Similarly, the full conditional distribution of ω has density

$$\pi(\omega \mid \cdot) \propto \exp\left\{-\frac{1}{2}\sum_{t=1}^{T}(\boldsymbol{\xi}_t - \zeta\boldsymbol{\xi}_{t-1})'\boldsymbol{A}(\boldsymbol{\xi} - \zeta\boldsymbol{\xi}_{t-1})\right\}\det\{\boldsymbol{A}\}^{\frac{T}{2}}\omega^{\alpha_\omega - 1}(1-\omega)^{v_\omega - 1},$$

where $\boldsymbol{A} = \tau^{-2}(\boldsymbol{D} - \omega\boldsymbol{W})$. A Metropolis-Hastings step is used to sample ω.

B.2 Additional Simulation Results

B.2.1 Investigation of a CAR Model for the Spatially Varying Coefficients

There are many possible models that could be used for the spatially varying coefficients. Gelfand et al. [2003] suggest various ways of defining spatially and spatio-temporally varying coefficients via GP models. Motivated by this work, our proposed methodology uses GPP models to account for spatially varying coefficients. In the case of areal data, a CAR model is a viable alternative for the

spatially varying coefficients. Suppose that, rather than using a GPP for $\widetilde{\boldsymbol{\beta}}_p$ as described in section 5.2, we use a Gaussian CAR model for $\boldsymbol{\beta}_p$. This results in the following hierarchical model:

$$
\begin{aligned}
Y_{st}|n_{st}, \nu_{st} &\overset{indep.}{\sim} \text{Binomial}\left(n_{st}, p_{st} = g(\nu_{st})\right), \quad s = 1, \ldots, S;\ t = 1, \ldots, T; \\
\boldsymbol{\beta}_p|\sigma_p^2, \boldsymbol{\theta}_p &\overset{indep.}{\sim} \text{N}\left(\mathbf{0}, \sigma_p^2(\boldsymbol{D} - \theta_p \boldsymbol{W})^{-1}\right), \quad p = 1, \ldots, P; \\
\sigma_p^2 &\overset{indep.}{\sim} \text{IG}(\alpha_{\sigma_p^2}, \upsilon_{\sigma_p^2}), \quad p = 1, \ldots, P; \\
\boldsymbol{\theta}_p &\overset{i.i.d.}{\sim} \pi(\boldsymbol{\theta}_p), \quad p = 1, \ldots, P; \\
\boldsymbol{\delta} &\sim \text{N}(\mathbf{0}, \sigma_\delta^2 \boldsymbol{I}), \quad \sigma_\delta^2 > 0; \\
\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1}, \tau^2, \omega, \zeta &\sim \text{N}\left(\zeta \boldsymbol{\xi}_{t-1}, \tau^2(\boldsymbol{D} - \omega \boldsymbol{W})^{-1}\right), \quad t = 1, \ldots, T; \\
\tau^2 &\sim \text{IG}(\alpha_{\tau^2}, \upsilon_{\tau^2}), \quad \alpha_{\tau^2}, \upsilon_{\tau^2} > 0; \\
\omega &\sim \text{Beta}(\alpha_\omega, \upsilon_\omega), \quad \alpha_\omega, \upsilon_\omega > 0; \\
\zeta &\sim \text{Truncated-Normal}(0, \sigma_\zeta^2, -1, 1), \quad \sigma_\zeta^2 > 0,
\end{aligned}
\tag{1}
$$

where $\nu_{st} = \boldsymbol{Z}'_{st}\boldsymbol{\delta} + \boldsymbol{X}'_{st}\boldsymbol{\beta}(\boldsymbol{\ell}_s) + \xi_{st}$ and $\boldsymbol{\beta}(\boldsymbol{\ell}_s) = (\beta_1(\boldsymbol{\ell}_s), ..., \beta_P(\boldsymbol{\ell}_s))'$ with $\beta_p(\boldsymbol{\ell}_s)$ being the sth element of $\boldsymbol{\beta}_p$.

We make two conjectures concerning the relative performance of the CAR-based model to the GPP-based model. We empirically evaluate both conjectures:

Conjecture 1: If the spatially varying regression coefficients are smooth, the GPP model yields smaller biases and mean squared errors than the CAR model. To evaluate this conjecture, we generated 500 replicate data sets exactly as in Section 5.4. In addition to fitting the GPP-based model to the data as in Section 5.4, we also fit supplemental model (1) to each data set, using the following priors:

$$
\begin{aligned}
\boldsymbol{\beta}_1|\sigma_1^2, \boldsymbol{\theta}_1 &\sim \text{N}\left(\mathbf{0}, \sigma_1^2(\boldsymbol{D} - \theta_1 \boldsymbol{W})^{-1}\right), \\
\sigma_1^2 &\sim \text{IG}(2, 2), \\
\boldsymbol{\theta}_1 &\sim \text{Beta}(900, 100).
\end{aligned}
$$

Figure B.2 summarizes the parameter estimates obtained under both models. The true coefficient

surface is depicted in Figure 5.2 of the corresponding manuscript. From these results, we see that our proposed GPP approach has smaller empirical bias and empirical mean squared errors than the CAR approach.

Conjecture 2: If the spatially varying regression coefficients at a specific geographic region (e.g., county) is comprised of both a smooth "regional" effect (i.e., a smooth surface over the study area) plus "local" heterogeneity, the proposed GPP approach can distinguish between and estimate these two effects. By contrast, the CAR model can only estimate the local effects. To evaluate this conjecture, we simulated 500 replicate data sets generated as described previously, with the exception that we now add a random draw from a CAR model to the "regional" (low-frequency) coefficient surface to create an underlying smooth surface with local heterogeneity. Both the regional and local effect surfaces are provided in Figure B.3. The proposed GPP approach (i.e., model (3) in the corresponding manuscript) and the CAR model (i.e., supplemental model (1) described above) were again fit to each data set. Model fitting (i.e., prior specifications, number of MCMC iterates sampled, etc.) proceeded in the same manner as described for both the GPP and CAR models. Figure B.4 presents a summary of the estimates of the local effects obtained by both models. Similarly, Figure B.5 presents a summary of the estimates of the regional effects obtained by the GPP model. Only an estimate from the GPP is displayed, since it is not possible to produce a (sensible) estimate of regional effect under the CAR model. In this analysis, local and regional trends based on the GPP model fits are parsed out using the technique outlined in Section 5.5.

From these results, we observe that both procedures provide reliable estimates of the local trends, with the CAR model performing slightly better. This finding is somewhat expected, since the combination of the CAR model on the regression coeffcients and the CAR model on the random effects allow the model to closely adapt to the observed data. However, only the GPP model allows us to estimate both the regional and local trends, at the cost of slightly more bias in the local trend estimates. Again, this is reasonable since most nonparametric smoothers give biased estimates.

This study demonstrates that the GPP model is able to provide an estimate of both the regional and local trends, while other techniques such as the CAR model cannot. This point is buttressed by the fact that, in our application, we are more interested in the larger scale trends than we are the local deviations from these trends. Elaborating, from a scientific perspective, the regional

trends are of interest because they provide more general information on disease trends and indicate broad patterns in disease spread over time. Alternatively, from a public health perspective, local trends are valuable for a variety of reasons, including assessing disease risk, evaluating the effectiveness of local preventative measures, and informing health-care providers about local anomalies in disease prevalence. Hence, it is useful to have a model capable of estimating both types of effects simultaneously. This ultimately motivates our proposed model.

B.2.2 Investigation of the Effects of Missing Data

Here, a simulation study is conducted to assess the performance of our approach in the presence of randomly missing data. The same data generating process described in Section 5.4 is used to create 500 independent data sets, using a 13×13 grid, $\tau^2 = 0.005$, and $\omega = 0.9$. Prior to model fitting, we randomly remove 25%, 50%, and 75% of the generated observations. We fit the models in the exact same fashion described in Section 5.4.

Figure B.6 displays the results. We see consistent reliable recovery of the underlying surface, with only slight increases in bias and mean squared errors, even with more than half of the data missing. This suggests that the proposed approach is robust to randomly missing observations in the sense of still being able to estimate the underlying quantities of interest.

B.2.3 Investigation of the Effects of a Larger Spatial Support

In this study, 500 independent data sets are again generated in the same fashion as described in Section 5.4, with the exception that the data are on a 30×30 regular grid; i.e., $S = 30$. The dimension of this problem is thus $30^2 = 900$. The true surface for the spatially varying coefficients is a single realization from a GPP on this grid with 49 regularly spaced knots. This surface and the results of the simulation study are shown in Figure B.7. In this case, both the bias and the mean squared errors are smaller than when the data are on a 13×13 grid. This is sensible since the larger spatial support provides the model with more information. This study illustrates that our proposed approach is feasible to being fit for larger spatial data sets without sacrificing the quality of the resulting estimates.

B.3 Exploratory Analysis

To reasonably specify our model, we conducted an initial exploratory analysis to suggest potential model structures. This analysis is conducted marginally, using counties that reported tests during every month of the study period and reported at least 10 positive tests. Of the 3,109 counties, 672 counties met this criteria. Our analysis had two stages.

First, for each county we fit, via maximum likelihood, the following ordinary generalized linear model with a binomial likelihood and logistic link:

$$Y_{st}|n_{st}, p_{st} \sim \text{Binomial}(n_{st}, p_{st}), \tag{2}$$

where $p_{st} = \exp(\delta_{0s} + \delta_{1s}t)/\{1 + \exp(\delta_{0s} + \delta_{1s}t)\}$. Following the guidance in Dunsmuir and Scott [2015], we examined the Pearson residuals from these fits to determine if an autoregressive correlation structure can adequately model the temporal correlation. Specifically, Pearson residuals are used to compute the partial autocorrelation functions (PACFs). Generally, a significant PACF at lag 1 with no significant PACF at lags greater than 1 is generally considered as solid evidence that an AR(1) structure is appropriate [Brockwell and Davis, 2002a]. We find that 374 of the 672 models have significant PACFs at one or more lags. Figure B.8 provides a summary of the PACFs across the various model fits.

Based on the results of the first stage, a generalized linear autoregressive model of order 1 was fitted to each county level time series, using the `glarma` package in R [Dunsmuir and Scott, 2015]. In other words, the distributional assumption is the same as in (2), but with $p_{st} = \exp(\delta_{s0} + \delta_{s1}t + Z_{st})/\{1 + \exp(\delta_{s0} + \delta_{s1}t + Z_{st})\}$ and $Z_{st} = \phi_{s1}(Z_{s,t-1} + e_{s,t-1})$. In this implementation, the e_{st} are taken to be the Pearson residuals. As discussed in Dunsmuir and Scott [2015], histograms of the probability integral transformations are constructed from these model fits to assess model suitability. Under a suitable model, these transforms should look like draws from a uniform distribution. Figure B.9 presents histograms of the probability integral transformations from a randomly selected subset of the counties. The fit of the model appears to be adequate for most counties.

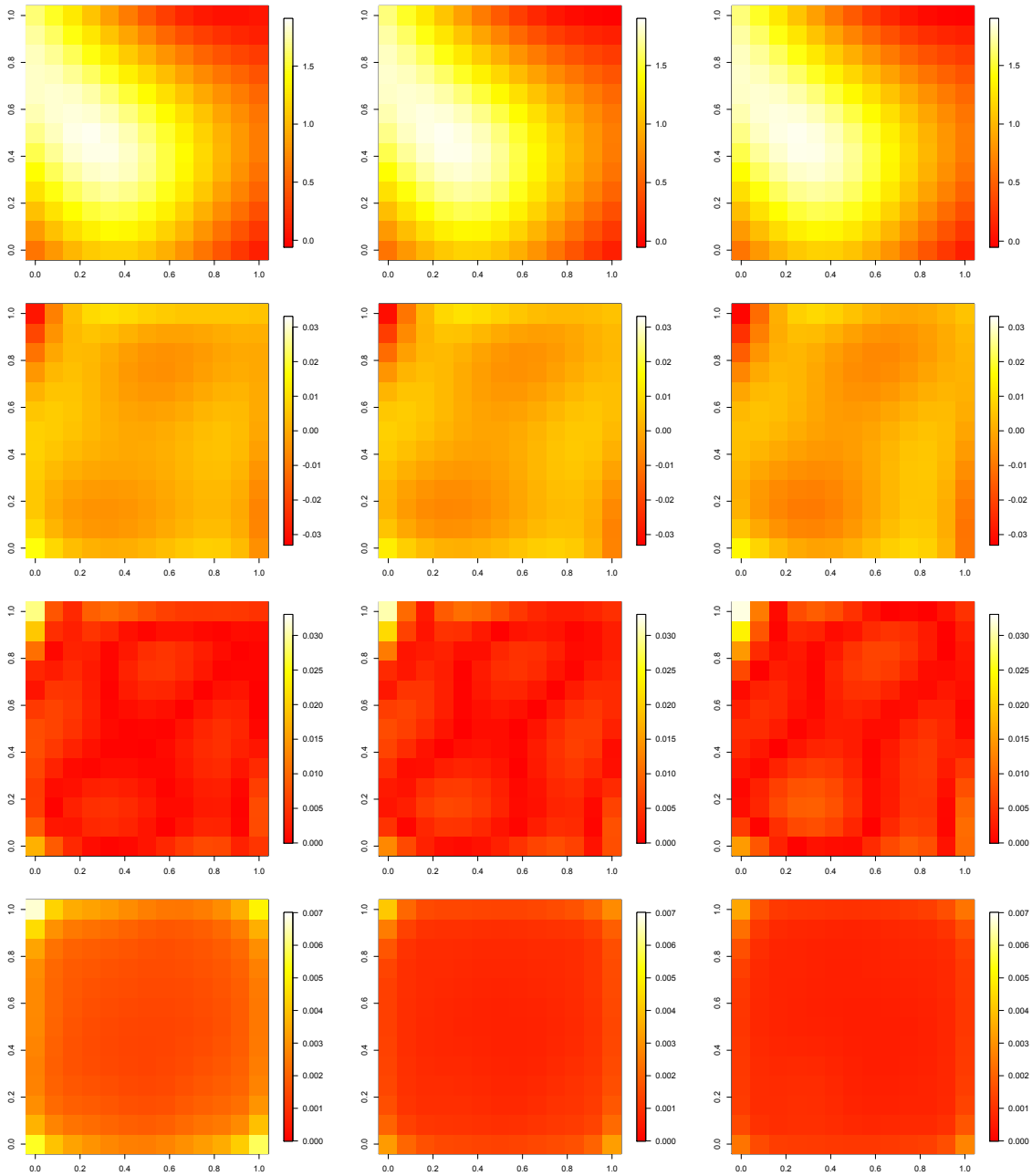


Figure B.1: Summary of the posterior estimates of $\widetilde{\boldsymbol{\beta}}_1$ obtained in the correlation simulation example. Results include the sample mean of the posterior estimates (first row), empirical bias (second row), empirical absolute bias (third row) and empirical mean squared error (fourth row). From left to right, the columns correspond to strongly correlated ($\omega = 0.90$), weakly correlated ($\omega = 0.55$), and uncorrelated data ($\omega = 0.00$). These results use a 5×5 grid of knots.

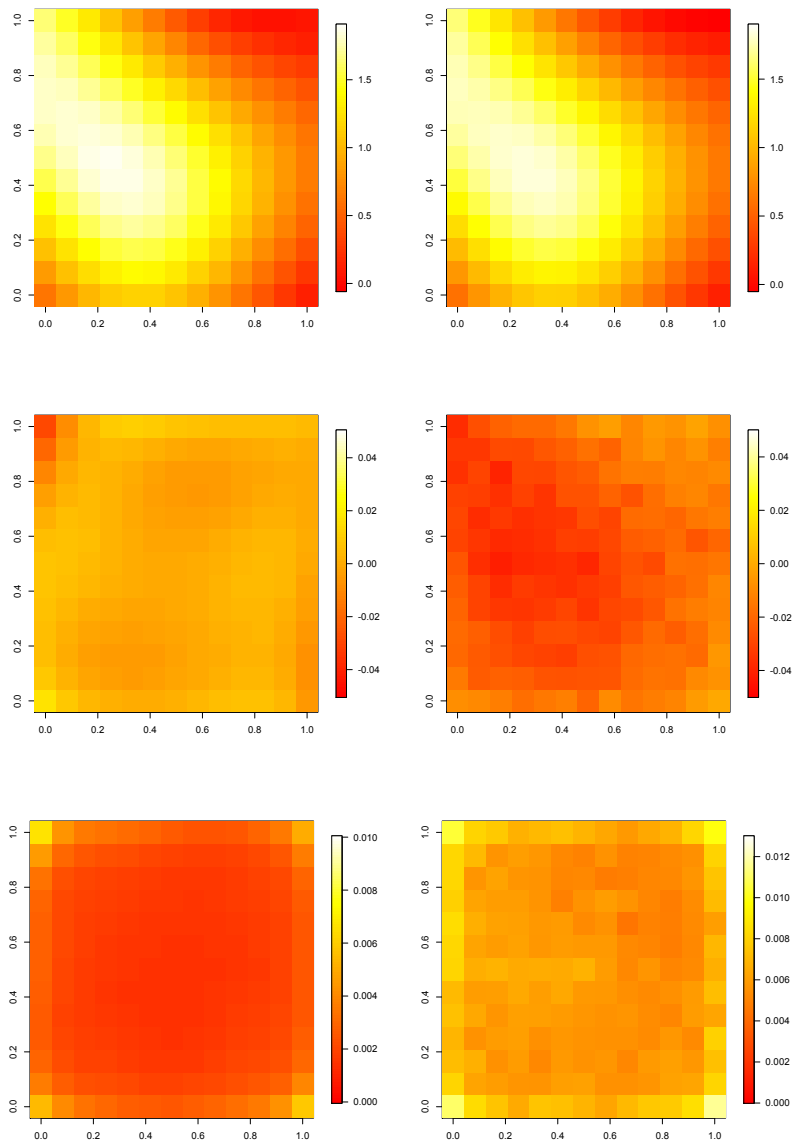


Figure B.2: Summary of the posterior estimates of $\widetilde{\boldsymbol{\beta}}_1$ obtained from the simulation designed to illustrate the differences between the performance of the proposed GPP model and a CAR variant. Results include the sample mean of the posterior estimates (top row), empirical bias (middle row), and empirical mean squared error (bottom row). The left column corresponds to the GPP model, the right column is for the CAR model. The true varying coefficient surface is provided in Figure 5.2 of the manuscript.

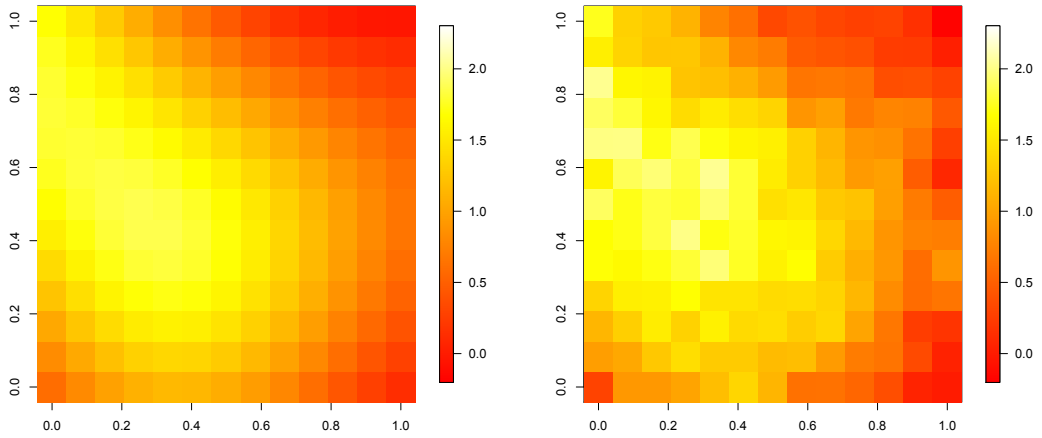


Figure B.3: True varying coefficient surfaces used to generate data in the simulation that illustrates the differences between the performance of the proposed GPP model and a CAR variant. The regional and local trends are provided on the left and right, respectively.

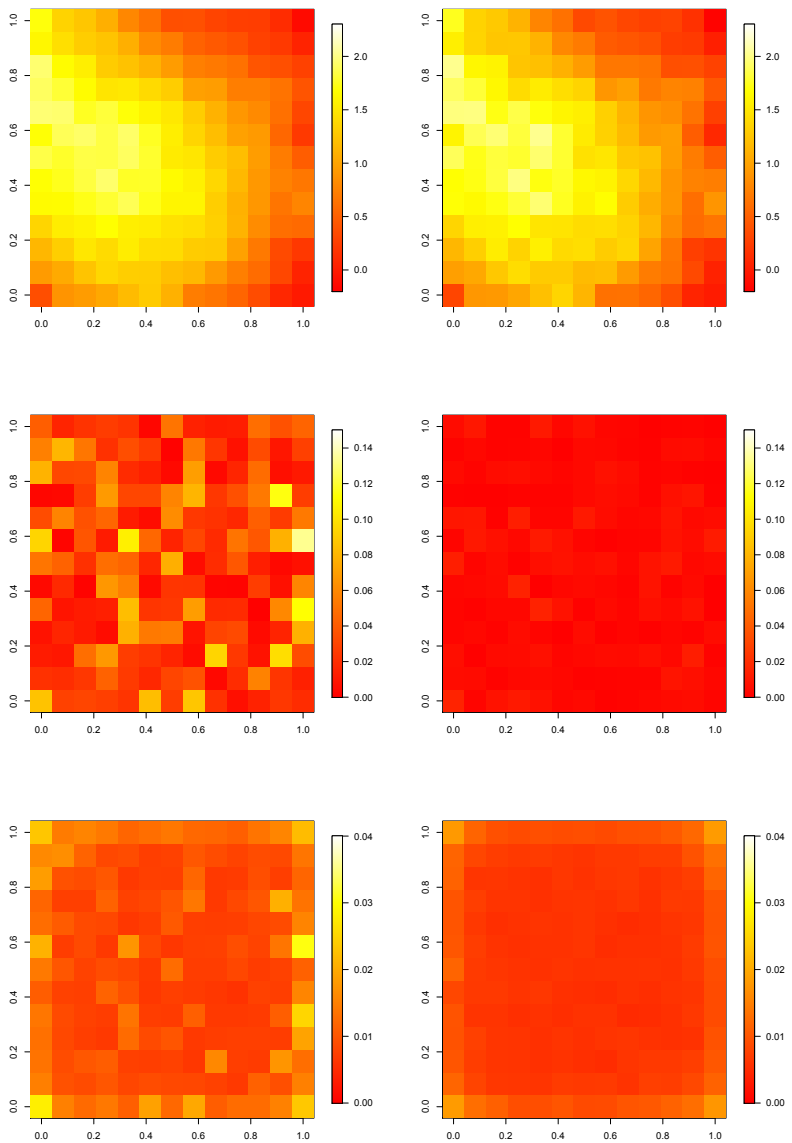


Figure B.4: Summary of the posterior estimates of the local trend (right panel in Figure B.3) obtained from the simulation designed to illustrate the differences between the performance of the proposed GPP model and a CAR variant. Presented are the posterior mean estimate (top row), the empirical absolute bias (middle row), and the empirical MSE (bottom row) associated with estimating the local trend through the use of the proposed GPP model (left column) and the CAR variant (right column).

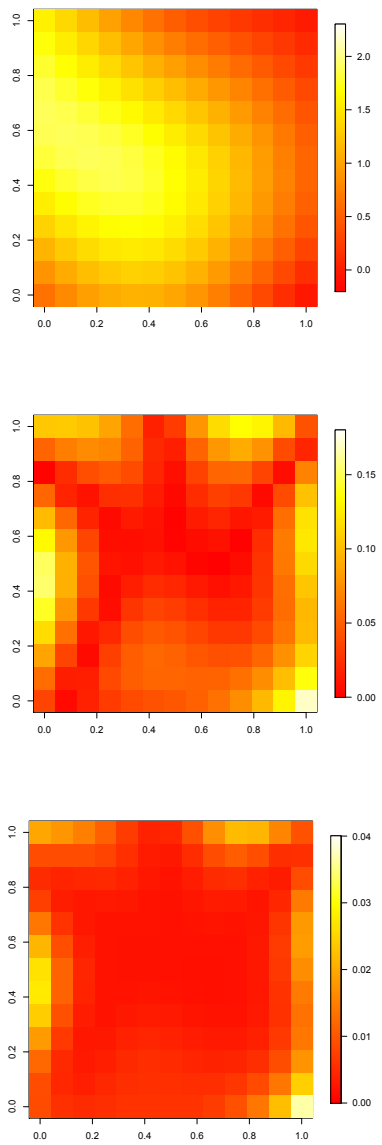


Figure B.5: Summary of the posterior estimates of the local trend (left panel in Figure B.3) obtained from the simulation designed to illustrate the differences between the performance of the proposed GPP model and a CAR variant. Results include the posterior mean estimate (top row), the empirical absolute bias (middle row), and the empirical MSE (bottom row) associated with estimating the regional trend through the use of the proposed GPP model. Note that the CAR variant does not provide an estimate of this quantity.

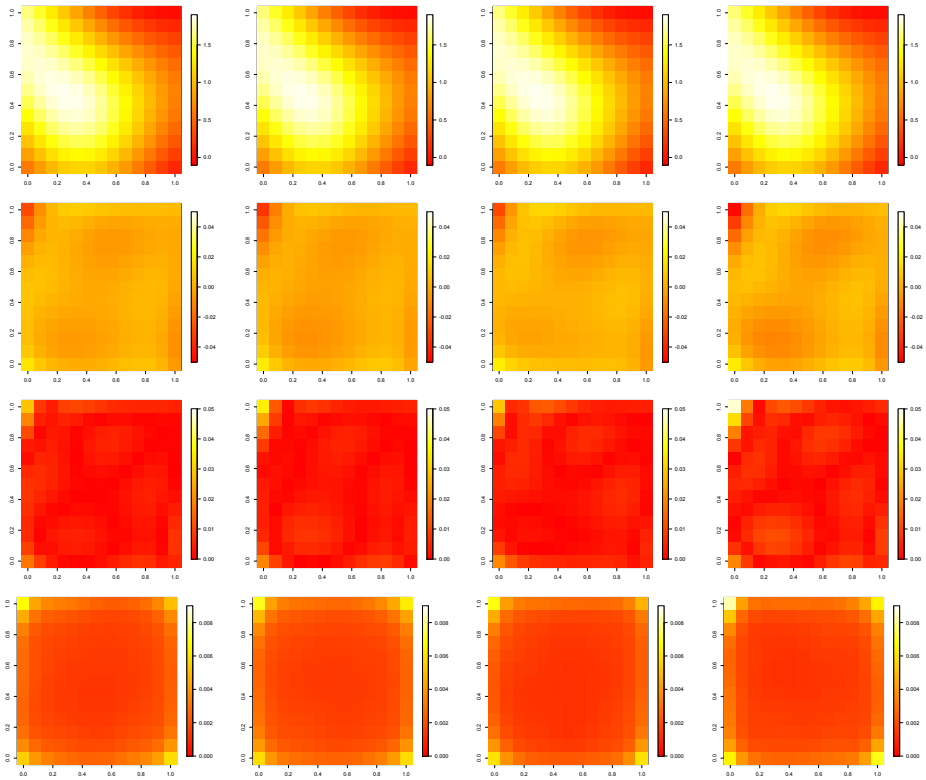


Figure B.6: Summary of the posterior estimates of $\widetilde{\boldsymbol{\beta}}_1$ obtained in the missing data simulation. Results include the sample mean of the posterior estimates (first row), empirical bias (second row), empirical absolute bias (third row), and empirical mean squared error (fourth row). From left to right, the columns correspond to 0%, 25%, 50% and 75% missing data.

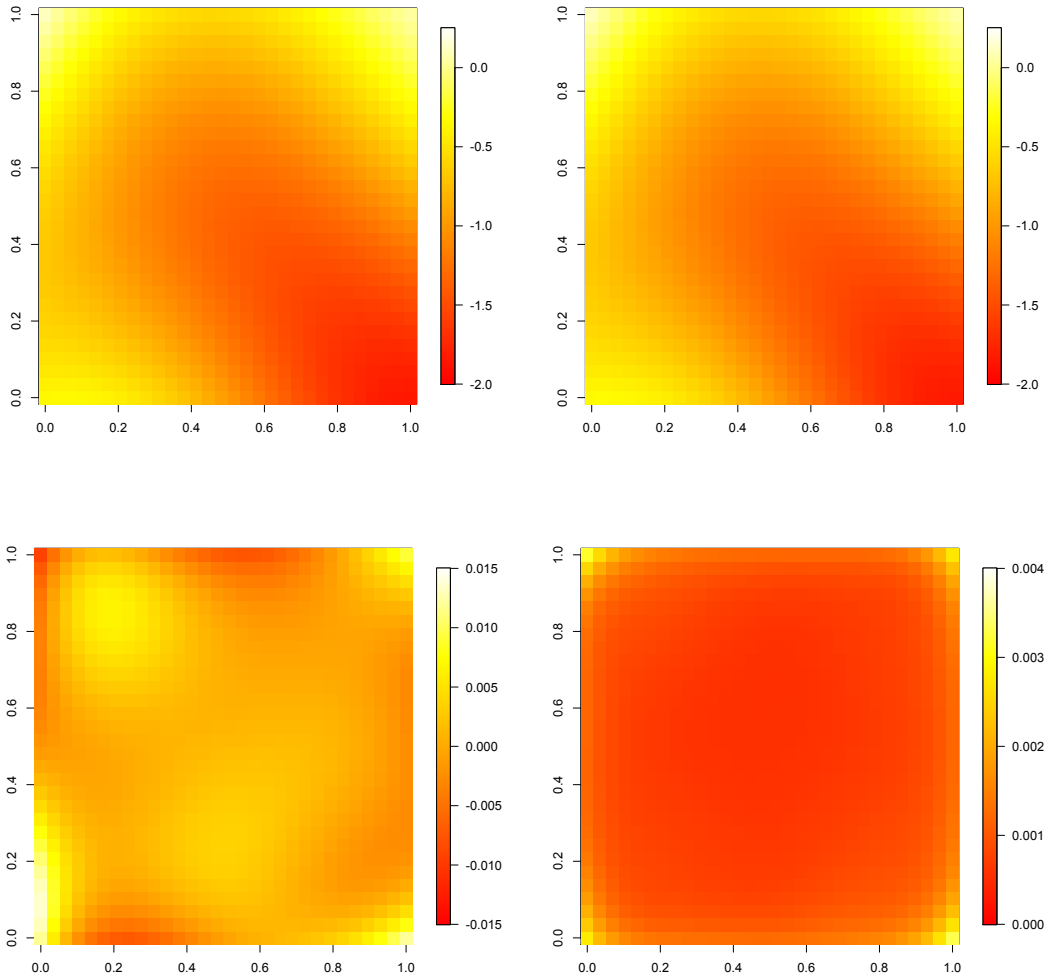


Figure B.7: Summary of the posterior estimates of $\widetilde{\boldsymbol{\beta}}_1$ obtained in the simulation designed to assess a larger spatial support. Results show the true spatially varying coefficient surface used for data generation (top left), the posterior mean estimate of this surface (top right), the empirical bias (bottom left), and the empirical mean square error (bottom right).

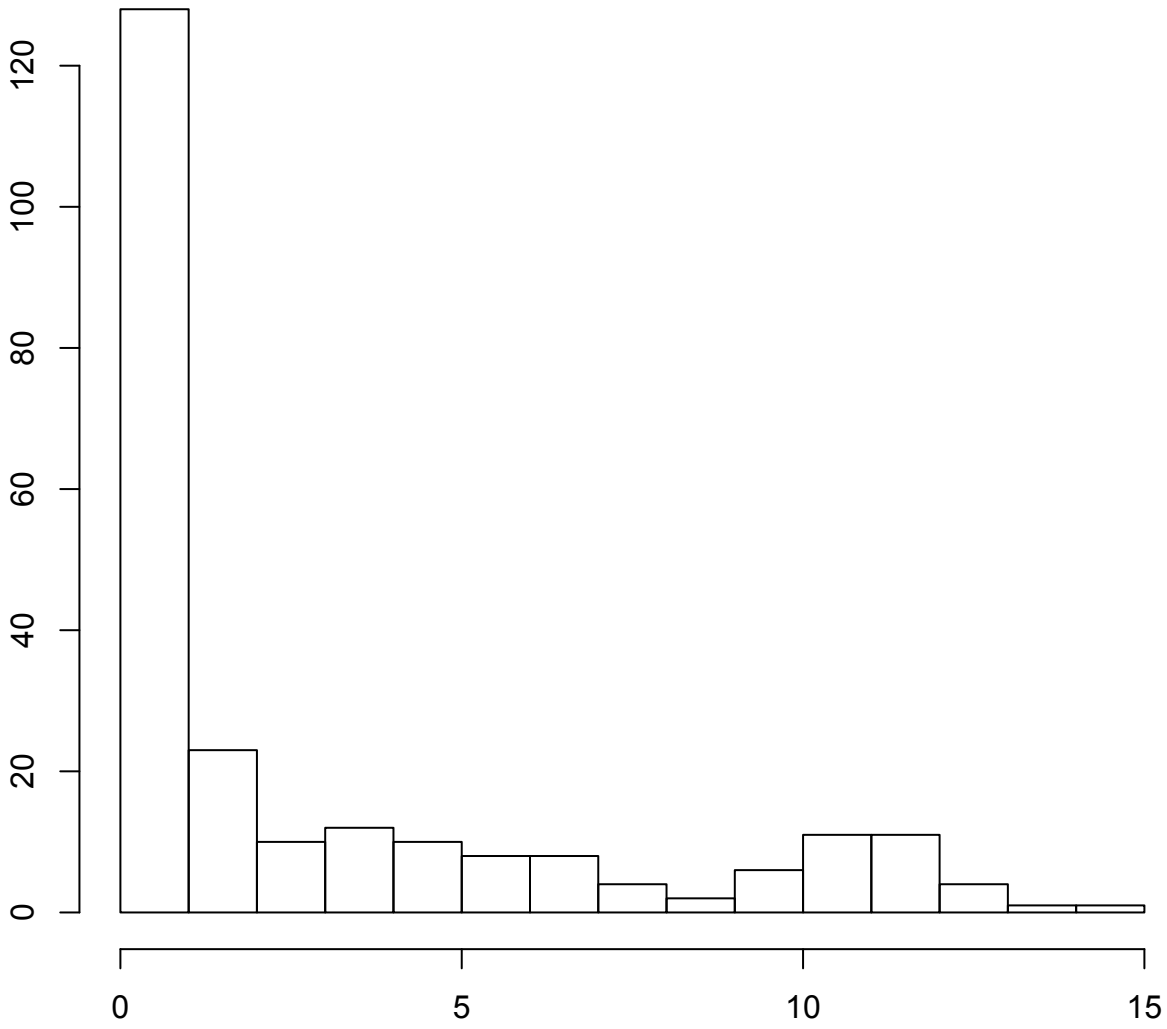


Figure B.8: Summary of the estimated PACFs obtained in stage 1 of the data exploration. A histogram of all lags having significant spikes across all counties is provided.

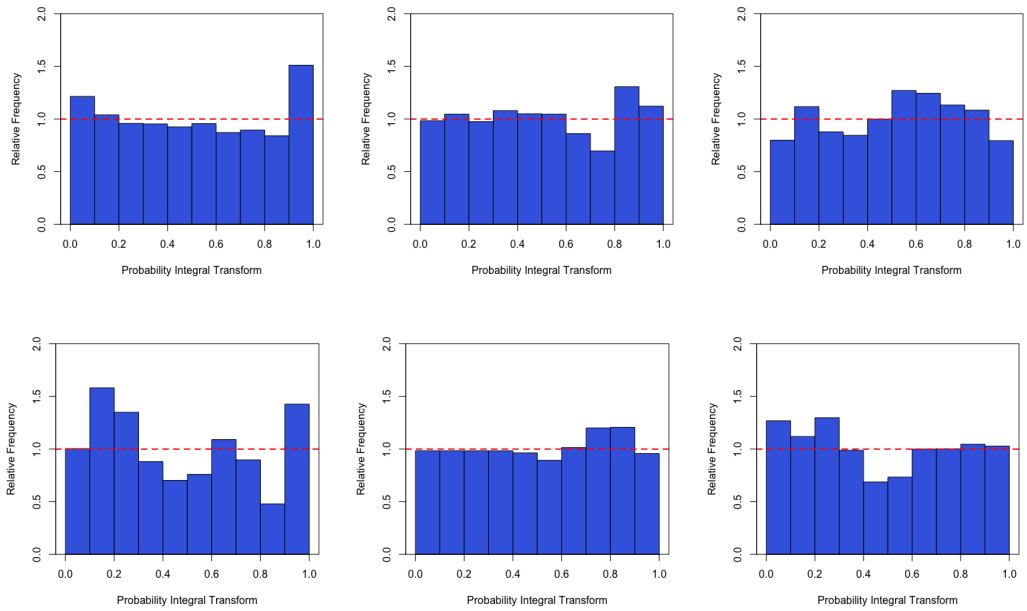


Figure B.9: Histograms of the probability integral transformations from the `glarma` models from 6 randomly selected counties.

Bibliography

North american regional reanalysis (narr). URL https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-regional-reanalysis-narr.

Idexx laboratories, inc. URL http://www.idexx.com/.

A new method for specifying the tuning parameter of 1 trend filtering. *Studies in Nonlinear Dynamics Econometrics*, 224.

Companion animal parasite council, 2012-current. URL www.CAPCvet.org.

Centers for disease control and prevention, 2015. URL http://www.cdc.gov/lyme/stats/index.html.

D. Adams, K. Fullerton, R. Jajosky, P. Sharp, D. Onweh, A. Schley, W. Anderson, A. Faulkner, and K. Kugeler. Summary of notifiable infectious diseases and conditions - united states, 2013. *MMWR Morb Mortal Wkly Rep*, 62(53):1–122, 2015. ISSN 1545-861X (Electronic) 0149-2195 (Linking). doi:10.15585/mmwr.mm6253a1. URL http://www.ncbi.nlm.nih.gov/pubmed/26492038.

D. A. Adams, K. M. Gallagher, R. A. Jajosky, J. Kriseman, P. Sharp, W. J. Anderson, A. E. Aranas, M. Mayes, M. S. Wodajo, D. H. Onweh, J. P. Abellera, Diseases Division of Notifiable, Office of Surveillance Epidemiology Healthcare Information, and C. D. C. Laboratory Services. Summary of notifiable diseases - united states, 2011. *MMWR Morb Mortal Wkly Rep*, 60(53): 1–117, 2013. ISSN 1545-861X (Electronic) 0149-2195 (Linking). URL http://www.ncbi.nlm.nih.gov/pubmed/23820934.

D. A. Adams, R. A. Jajosky, U. Ajani, J. Kriseman, P. Sharp, D. H. Onwen, A. W. Schley, W. J. Anderson, A. Grigoryan, A. E. Aranas, M. S. Wodajo, J. P. Abellera, Control Centers for Disease, and Prevention. Summary of notifiable diseases–united states, 2012. *MMWR Morb Mortal Wkly Rep*, 61(53):1–121, 2014. ISSN 1545-861X (Electronic) 0149-2195 (Linking). URL http://www.ncbi.nlm.nih.gov/pubmed/25233134.

Deborah A Adams. Summary of notifiable infectious diseases and conditions- United States, 2015. *Morbidity and Mortality Weekly Report*, 64, 2017.

E. R. Adrion, J. Aucott, K. W. Lemke, and J. P. Weiner. Health care costs, utilization and patterns of care following lyme disease. *PLoS One*, 10(2):e0116767, 2015. ISSN 1932-6203 (Electronic) 1932-6203 (Linking). doi:10.1371/journal.pone.0116767. URL http://www.ncbi.nlm.nih.gov/pubmed/25650808.

S. Agapiou, J. M. Bardsley, O. Papaspiliopoulos, and A. M. Stuart. Analysis of the Gibbs sampler for hierarchical inverse problems. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):511–544, 2014.

James H. Albert and Siddhartha Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.

C. Anderson and L. M. Ryan. A comparison of spatio-temporal disease mapping approaches including an application to ischaemic heart disease in New South Wales, Australia. *International Journal of Environmental Research and Public Health*, 14(2), 2017. doi:10.1093/biostatistics/kxu005.

C. Anderson, D. Lee, and N. Dean. Identifying clusters in bayesian disease mapping. *Biostatistics*, 15(3):457–469, Nov 2014. doi:10.1093/biostatistics/kxu005.

Ricardo Andrade-Pacheco, Martin Mubangizi, John Quinn, and Neil Lawrence. Monitoring short term changes in infectious disease in Uganda with Gaussian processes. In Ahlame Douzal-Chouakria, José A. Vilar, and Pierre-Franois Marteau, editors, *Advanced Analysis and Learning on Temporal Data*, pages 95–110. Springer, 2015.

S. L. Arvikar and A. C. Steere. Diagnosis and treatment of lyme arthritis. *Infect Dis Clin North Am*, 29(2):269–280, 2015. ISSN 1557-9824 (Electronic) 0891-5520 (Linking). doi:10.1016/j.idc.2015.02.004. URL `http://www.ncbi.nlm.nih.gov/pubmed/25999223`.

R. G. Bachur, C. M. Adams, and M. C. Monuteaux. Evaluating the child with acute hip pain ("irritable hip") in a lyme endemic region. *J Pediatr*, 166(2):407–411 e1, 2015. ISSN 1097-6833 (Electronic) 0022-3476 (Linking). doi:10.1016/j.jpeds.2014.09.040. URL `http://www.ncbi.nlm.nih.gov/pubmed/25444013`.

R. M. Bacon, B. J. Biggerstaff, M. E. Schriefer, R. D. Gilmore, Jr., M. T. Philipp, A. C. Steere, G. P. Wormser, A. R. Marques, and B. J. Johnson. Serodiagnosis of lyme disease by kinetic enzyme-linked immunosorbent assay using recombinant vlse1 or peptide antigens of *Borrelia burgdorferi* compared with 2-tiered testing using whole-cell lysates. *J Infect Dis*, 187(8):1187–1199, 2003. ISSN 0022-1899 (Print) 0022-1899 (Linking). doi:10.1086/374395. URL `http://www.ncbi.nlm.nih.gov/pubmed/12695997`.

S. Banerjee and A.E. Gelfand. On smoothness properties of spatial processes. *Journal of Multivariate Analysis*, 84(1):85 – 100, 2003. ISSN 0047-259X. doi:https://doi.org/10.1016/S0047-259X(02)00016-7. URL `http://www.sciencedirect.com/science/article/pii/S0047259X02000167`.

S. Banerjee, A. E. Gelfand, A. O. Finley, and H. Sang. Gaussian predictive process models for large spatial datasets. *Journal of the Royal Statistical Society: Series B (Methodological)*, 70(4):825–48, 2008.

S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical modeling and analysis for spatial data*. Chapman and Hall/CRC, 2 edition, 2014a.

S. Banerjee, B.P. Carlin, and A. E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, Boca Raton, FL, 2nd. edition, 2014b.

S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman and Hall/CRC, Boca Raton, 2nd edition, 2015a.

S. Banerjee, B. P. Carlin, and A. E. Gelfand. *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC, Boca Raton, 2nd edition, 2015b.

Sudipto Banerjee, Alan E Gelfand, and C. F Sirmans. Directional rates of change under spatial process models. *Journal of the American Statistical Association*, 98(464):946–954, 2003. doi:10.1198/C16214503000000909. URL `https://doi.org/10.1198/C16214503000000909`.

J. M. Bardsley. MCMC-based image reconstruction with uncertainty quantification. *SIAM Journal on Scientific Computing*, 34(3):1316–1332, 2012.

D. Bates and M. Maechler. *Matrix: Sparse and dense matrix classes and methods*, 2016. URL `https://CRAN.R-project.org/package=Matrix`. R package version 1.2-6.

M. J. Bayarri, J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C.-H. Lin, and J. Tu. A framework for validation of computer models. *Technometrics*, 49(2):138–154, 2007.

M. J. Beall, R. Chandrashekar, M. D. Eberts, K. E. Cyr, P. P. Diniz, C. Mainville, B. C. Hegarty, J. M. Crawford, and E. B. Breitschwerdt. Serological and molecular prevalence of *Borrelia burgdorferi*, *Anaplasma phagocytophilum*, and *Ehrlichia* species in dogs from minnesota. *Vector Borne Zoonotic Dis*, 8(4):455–64, 2008. ISSN 1557-7759 (Electronic) 1530-3667 (Linking). doi:10.1089/vbz.2007.0236. URL `http://www.ncbi.nlm.nih.gov/pubmed/18302532`.

J. Benito, V. Depuy, E. Hardie, H. Zamprogno, A. Thomson, W. Simpson, S. Roe, B. Hansen, and B. D. Lascelles. Reliability and discriminatory testing of a client-based metrology instrument, feline musculoskeletal pain index (fmpi) for the evaluation of degenerative joint disease-associated pain in cats. *Vet J*, 196(3):368–373, 2013. ISSN 1532-2971 (Electronic) 1090-0233 (Linking). doi:10.1016/j.tvjl.2012.12.015. URL `http://www.ncbi.nlm.nih.gov/pubmed/23369382`.

James O Berger, Victor de Oliveira, and Bruno Sanso. Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, 96(456):1361–1374, 2001.

J. Besag. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36(2):192–236, 1974a.

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

J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society, Series B*, 48(3):259–302, 1986.

J. Besag and C. Kooperberg. On conditional and intrinsic autoregressions. *Biometrika*, 82:733–746, 1995a.

J. Besag and C. Kooperberg. On conditional and intrinsic autogregressions. *Biometrika*, 82(4): 733–46, 1995b.

J. Besag and D. Mondal. Exact goodness-of-fit tests for markov chains. *Biometrics*, 69(2):488–496, 2013. ISSN 1541-0420 (Electronic) 0006-341X (Linking). doi:10.1111/biom.12009. URL `http://www.ncbi.nlm.nih.gov/pubmed/23432148`.

J. Besag, J. York, and A. Mollie. Bayesian image restoration, with two applications in spatial statistics. *Annuals of the Institute of Statistical Mathematics*, 43:1–20, 1991a.

J. Besag, J. C. York, and A. Mollié. Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43:1–59, 1991b.

Samir Bhatt, Ewan Cameron, Seth R. Flaxman, Daniel J. Weiss, David L. Smith, and Peter W. Gething. Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. *Journal of The Royal Society Interface*, 14(134), 2017.

M. Binois, R. B. Gramacy, and M. Ludkovski. Practical heteroskedastic Gaussian process modeling for large simulation experiments. Arxiv preprint 1611:05902, 2016.

Howard D. Bondell, Brian J. Reich, and Huixia Wang. Noncrossing quantile regression curve estimation. *Biometrika*, 97(4):825–838, 2010. ISSN 00063444. URL `http://www.jstor.org/stable/29777139`.

D. Bowman, S. E. Little, L. Lorentzen, J. Shields, M. P. Sullivan, and E. P. Carlin. Prevalence and geographic distribution of *Dirofilaria immitis, Borrelia burgdorferi, Ehrlichia canis*, and *Anaplasma phagocytophilum* in dogs in the united states: results of a national clinic-based serologic survey. *Vet Parasitol*, 160(1-2):138–48, 2009. ISSN 0304-4017 (Print) 0304-4017 (Linking). doi:10.1016/j.vetpar.2008.10.093. URL `http://www.ncbi.nlm.nih.gov/pubmed/19150176`.

D. D. Bowman, Y. Liu, C. S. McMahan, S. K. Nordone, M. J. Yabsley, and R. B. Lund. Forecasting united states heartworm *Dirofilaria immitis* prevalence in dogs. *Parasit Vectors*, 9(1):540, 2016a. ISSN 1756-3305 (Electronic) 1756-3305 (Linking). doi:10.1186/s13071-016-1804-y. URL `http://www.ncbi.nlm.nih.gov/pubmed/27724981`.

D. D. Bowman, Y. Liu, C. S. McMahan, S. K. Nordone, M. J. Yabsley, and R. B. Lund. Forecasting united states heartworm *Dirofilaria immitis* prevalence in dogs. *Parasit Vectors*, 9(1):540, 2016b. ISSN 1756-3305 (Electronic) 1756-3305 (Linking). doi:10.1186/s13071-016-1804-y. URL `https://www.ncbi.nlm.nih.gov/pubmed/27724981`.

J. A. Branda, M. E. Aguero-Rosenfeld, M. J. Ferraro, B. J. Johnson, G. P. Wormser, and A. C. Steere. 2-tiered antibody testing for early and late lyme disease using only an immunoglobulin g blot with the addition of a vlse band as the second-tier test. *Clin Infect Dis*, 50(1):20–26, 2010. ISSN 1537-6591 (Electronic) 1058-4838 (Linking). doi:10.1086/648674. URL `http://www.ncbi.nlm.nih.gov/pubmed/19947857`.

A. Brezger, L. Fahrmeir, and A. Hennerfeind. Adaptive Gaussian Markov random fields with applications in human brain mapping. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 56(3):327–345, 2007.

P.J. Brockwell and R.A. Davis. *Introduction to Time Series and Forecasting*. Springer-Verlag, New York City, 2002a.

P.J. Brockwell and R.A. Davis. *Introduction to Time Series and Forecasting*. Springer-Verlag, New York City, 2002b.

P.J. Brockwell and R.A. Davis. *Introduction to time series and forecasting*. Springer Science & Business Media, 2nd. edition, 2006.

S. P. Brooks and A. Gelman. General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*, 7(4):434–455, 1998.

D. A. Brown, N. A. Lazar, G. S. Datta, W. Jang, and J. E. McDowell. Incorporating spatial dependence into Bayesian multiple testing of statistical parametric maps in functional neuroimaging. *NeuroImage*, 84:97–112, 2014.

D. A. Brown, G. S. Datta, and N. A. Lazar. A Bayesian generalized CAR model for correlated signal detection. *Statistica Sinica*, 27:1125–1153, 2017a.

D. A. Brown, C. S. McMahan, R. T. Shinohara, and K. L. Linn. Bayesian spatial binary regression for label fusion in structural neuroimaging. ArXiv 1710.10351, 2017b.

D. A. Brown, C. S. McMahan, and S. C. Watson. Sampling strategies for fast updating of Gaussian Markov random fields. Arxiv preprint 1702:05518, 2017c.

Z. Cai. Very large scale Bayesian machine learning. Unpublished doctoral dissertation, Rice University, Department of Computer Science, 2014.

Z. Cai, C. Jermaine, Z. Vagena, D. Logothetis, and L.L. Perez. The pairwise Gaussian random field for high-dimensional data imputation. *IEEE 13th International Conference on Data Mining (ICDM)*, pages 61–70, 2013.

B. P. Carlin and T. A. Louis. *Bayesian Methods for Data Analysis*. Chapman & Hall/CRC, Boca Raton, 3rd edition, 2009.

A. J. Caulfield and B. S. Pritt. Lyme disease coinfections in the united states. *Clin Lab Med*, 35 (4):827–46, 2015. ISSN 1557-9832 (Electronic) 0272-2712 (Linking). doi:10.1016/j.cll.2015.07.006. URL `http://www.ncbi.nlm.nih.gov/pubmed/26593260`.

Centers for Disease Control and Prevention. Data and statistics: Lyme disease, 2017. URL `https://www.cdc.gov/lyme/stats/index.html`.

Guorong Chai, Hua He, Yongzhong Sha, Guangyu Zhai, and Shengliang Zong. Effect of pm2.5 on daily outpatient visits for respiratory diseases in lanzhou, china. *Science of The Total Environment*, 649:1563 – 1572, 2019. ISSN 0048-9697. doi:https://doi.org/10.1016/j.scitotenv.2018.08.384. URL `http://www.sciencedirect.com/science/article/pii/S0048969718333552`.

D. Cheng, Y. Cheng, Y. Liu, R. Peng, and S.-H. Teng. Efficient sampling for Gaussian graphical models via spectral sparsification. In *Journal of Machine Learning Research: Proceedings of the 28th International Conference on Learning Theory*, pages 364–390. 2015.

Hee Min Choi and James P. Hobert. Analysis of mcmc algorithms for bayesian linear regression with laplace errors. *Journal of Multivariate Analysis*, 117:32 – 40, 2013. ISSN 0047-259X.

K. A. Cohn, A. D. Thompson, S. S. Shah, E. M. Hines, T. W. Lyons, E. J. Welsh, and L. E. Nigrovic. Validation of a clinical prediction rule to distinguish lyme meningitis from aseptic meningitis. *Pediatrics*, 129(1):e46–53, 2012. ISSN 1098-4275 (Electronic) 0031-4005 (Linking). doi:10.1542/peds.2011-1215. URL `http://www.ncbi.nlm.nih.gov/pubmed/22184651`.

P. Congdon and H. Southall. Trends in inequality in infant mortality in the north of England, 1921 to 1973, and their association with urban and social structure. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 168(4):679–700, 2005. doi:10.1111/j.1467-985X.2005.00370.x. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-985X.2005.00370.x`.

J. A. Conlon, T. N. Mather, P. Tanner, G. Gallo, and R. H. Jacobson. Efficacy of a nonadjuvanted, outer surface protein a, recombinant vaccine in dogs after challenge by ticks naturally infected with *Borrelia burgdorferi*. *Vet Ther*, 1:96–107, 2000. ISSN 1528-3593 (Print) 1528-3593 (Linking). URL `http://www.ncbi.nlm.nih.gov/pubmed/19757556`.

J. M. Costello, M. E. Alexander, K. M. Greco, A. R. Perez-Atayde, and P. C. Laussen. Lyme carditis in children: presentation, predictive factors, and clinical course. *Pediatrics*, 123(5):e835–841, 2009. ISSN 1098-4275 (Electronic) 0031-4005 (Linking). doi:10.1542/peds.2008-3058. URL `http://www.ncbi.nlm.nih.gov/pubmed/19403477`.

N. Cressie. *Statistics for Spatial Data*. Wiley Press, 1993.

N. Cressie and G. Johannesson. Fixed rank kriging for very large spatial datasets. *Journal of the Royal Statistical Society: Series B (Methodological)*, 70:209–226, 2008.

N. Cressie and C. K. Wikle. *Statistics for Spatio-temporal Data*. Wiley Press, 2011a.

Noel Cressie and Christopher K. Wikle. *Statistics for spatio-temporal data*. Wiley series in probability and statistics. Wiley, Hoboken, N.J., 2011b. ISBN 9780471692744 (cloth).

Arvind Dabass, Evelyn O. Talbott, Arvind Venkat, Judith Rager, Gary M. Marsh, Ravi K. Sharma, and Fernando Holguin. Association of exposure to particulate matter (pm2.5) air pollution and biomarkers of cardiovascular disease risk in adult nhanes participants (20012008). *International Journal of Hygiene and Environmental Health*, 219(3):301 – 310, 2016. ISSN 1438-4639. doi:https://doi.org/10.1016/j.ijheh.2015.12.002. URL `http://www.sciencedirect.com/science/article/pii/S1438463915001595`.

F. S. Dahlgren, K. N. Heitman, N. A. Drexler, R. F. Massung, and C. B. Behravesh. Human granulocytic anaplasmosis in the united states from 2008 to 2012: a summary of national surveillance data. *Am J Trop Med Hyg*, 93(1):66–72, 2015. ISSN 1476-1645 (Electronic) 0002-9637 (Linking). doi:10.4269/ajtmh.15-0122. URL `https://www.ncbi.nlm.nih.gov/pubmed/25870428`.

F. S. Dahlgren, K. N. Heitman, and C. B. Behravesh. Undetermined human ehrlichiosis and anaplasmosis in the united states, 2008-2012: A catch-all for passive surveillance. *Am J Trop Med Hyg*, 94 (2):299–301, 2016. ISSN 1476-1645 (Electronic) 0002-9637 (Linking). doi:10.4269/ajtmh.15-0691. URL `https://www.ncbi.nlm.nih.gov/pubmed/26621564`.

T. J. Daniels, T. M. Boccia, S. Varde, J. Marcus, J. Le, D. J. Bucher, R. C. Falco, and I. Schwartz. Geographic risk for lyme disease and human granulocytic ehrlichiosis in southern new york state. *Appl Environ Microbiol*, 64(12):4663–9, 1998. ISSN 0099-2240 (Print) 0099-2240 (Linking). URL `https://www.ncbi.nlm.nih.gov/pubmed/9835546`.

A. Datta, S. Banerjee, A. O. Finley, and A. E. Gelfand. Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111:800–812, 2016a.

Abhirup Datta, Sudipto Banerjee, Andrew O. Finley, and Alan E. Gelfand. Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111(514):800–812, 2016b.

A. P. Dempster. Covariance selection. *Biometrics*, 28:157–175, 1972.

S. E. Detmer, M. Bouljihad, D. W. Hayden, J. M. Schefers, A. Armien, and A. Wunschmann. Fatal pyogranulomatous myocarditis in 10 boxer puppies. *J Vet Diagn Invest*, 28(2):144–149, 2016. ISSN 1943-4936 (Electronic) 1040-6387 (Linking). doi:10.1177/1040638715626486. URL `http://www.ncbi.nlm.nih.gov/pubmed/26965234`.

P. J. Diggle, J. A. Tawn, and R. A. Moyeed. Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 47(3):299–350, 1998.

M. A. Diuk-Wasser, E. Vannier, and P. J. Krause. Coinfection by ixodes tick-borne pathogens: Ecological, epidemiological, and clinical consequences. *Trends Parasitol*, 32(1):30–42, 2016. ISSN 1471-5007 (Electronic) 1471-4922 (Linking). doi:10.1016/j.pt.2015.09.008. URL `https://www.ncbi.nlm.nih.gov/pubmed/26613664`.

E. E. Drouin, L. J. Glickstein, and A. C. Steere. Molecular characterization of the ospa(161-175) t cell epitope associated with treatment-resistant lyme arthritis: differences among the three pathogenic species of *Borrelia burgdorferi* sensu lato. *J Autoimmun*, 23(3):281–92, 2004. ISSN 0896-8411 (Print) 0896-8411 (Linking). doi:10.1016/j.jaut.2004.06.005. URL `http://www.ncbi.nlm.nih.gov/pubmed/15501399`.

J. S. Dumler, A. F. Barbet, C. P. Bekker, G. A. Dasch, G. H. Palmer, S. C. Ray, Y. Rikihisa, and F. R. Rurangirwa. Reorganization of genera in the families rickettsiaceae and anaplasmataceae in the order rickettsiales: unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six new species combinations and designation of *Ehrlichia equi* and 'hge agent' as subjective synonyms of *Ehrlichia phagocytophila*. *Int J Syst Evol Microbiol*, 51(Pt 6):2145–65, 2001. ISSN 1466-5026 (Print) 1466-5026 (Linking). doi:10.1099/00207713-51-6-2145. URL `http://www.ncbi.nlm.nih.gov/pubmed/11760958`.

William Dunsmuir and David Scott. The glarma package for observation-driven time series regression of counts. *Journal of Statistical Software, Articles*, 67(7):1–36, 2015. ISSN 1548-7660. doi:10.18637/jss.v067.i07. URL `https://www.jstatsoft.org/v067/i07`.

M. D. Eberts, P. P. Vissotto de Paiva Diniz, M. J. Beall, B. A. Stillman, R. Chandrashekar, and E. B. Breitschwerdt. Typical and atypical manifestations of *Anaplasma phagocytophilum* infection in dogs. *J Am Anim Hosp Assoc*, 47, 2011. ISSN 1547-3317 (Electronic) 0587-2871 (Linking). doi:10.5326/JAAHA-MS-5578.

Jo Eidsvik, Andrew O. Finley, Sudipto Banerjee, and Hvard Rue. Approximate Bayesian inference for large spatial datasets using predictive process models. *Computational Statistics and Data Analysis*, 56(6):1362 – 1380, 2012.

Paul H. C. Eilers and Brian D. Marx. Flexible smoothing with b -splines and penalties. *Statist. Sci.*, 11(2):89–121, 05 1996. doi:10.1214/ss/1038425655. URL `https://doi.org/10.1214/ss/1038425655`.

R. J. Eisen and L. Eisen. Spatial modeling of human risk of exposure to vector-borne pathogens based on epidemiological versus arthropod vector data. *J Med Entomol*, 45(2):181–192, 2008. ISSN 0022-2585 (Print) 0022-2585 (Linking). URL `http://www.ncbi.nlm.nih.gov/pubmed/18402133`.

R. J. Eisen, L. Eisen, N. H. Ogden, and C. B. Beard. Linkages of weather and climate with *Ixodes scapularis* and *Ixodes pacificus* (acari: Ixodidae), enzootic transmission of *Borrelia burgdorferi*, and lyme disease in north america. *J Med Entomol*, 2015. ISSN 0022-2585 (Print) 0022-2585 (Linking). doi:10.1093/jme/tjv199. URL `http://www.ncbi.nlm.nih.gov/pubmed/26681789`.

R. J. Eisen, L. Eisen, and C. B. Beard. County-scale distribution of *Ixodes scapularis* and *Ixodes pacificus* (acari: Ixodidae) in the continental united states. *J Med Entomol*, 2016a. ISSN 0022-2585 (Print) 0022-2585 (Linking). doi:10.1093/jme/tjv237. URL `http://www.ncbi.nlm.nih.gov/pubmed/26783367`.

Rebecca J Eisen, Lars Eisen, and Charles B Beard. County-scale distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari : Ixodidae) in the continental United States. *Journal of Medical Entomology*, 53(January):349–386, 2016b. doi:10.1093/jme/tjv237.

M. E. Embers, N. R. Hasenkampf, M. B. Barnes, E. S. Didier, M. T. Philipp, and A. C. Tardo. Five-antigen fluorescent bead-based assay for diagnosis of lyme disease. *Clin Vaccine Immunol*, 23(4):294–303, 2016. ISSN 1556-6811 (Print) 1556-679X (Linking). doi:10.1128/CVI.00685-15. URL `http://www.ncbi.nlm.nih.gov/pubmed/26843487`.

Inc. Environmental Systems Research Institute. Documentation, 2016. URL `http://desktop.arcgis.com/en/documentation/`.

A. K. Eschner and K. Mugnai. Immunization with a recombinant subunit ospa vaccine markedly impacts the rate of newly acquired *Borrelia burgdorferi* infections in client-owned dogs living in a coastal community in maine, usa. *Parasit Vectors*, 8:92, 2015. ISSN 1756-3305 (Electronic) 1756-3305 (Linking). doi:10.1186/s13071-015-0676-x. URL `https://www.ncbi.nlm.nih.gov/pubmed/25890386`.

James R. Faulkner and Vladimir N. Minin. Locally adaptive smoothing with markov random fields and shrinkage priors. *Bayesian Anal.*, 13(1):225–252, 03 2018. doi:10.1214/17-BA1050. URL `https://doi.org/10.1214/17-BA1050`.

Andrew O. Finley, Huiyan Sang, Sudipto Banerjee, and Alan E. Gelfand. Improving the performance of predictive process modeling for large datasets. *Computational Statistics and Data Analysis*, 53 (8):2873 – 2884, 2009.

Sergio Firpo, Nicole M. Fortin, and Thomas Lemieux. Unconditional quantile regressions. *Econometrica*, 77(3):953–973, 2009. ISSN 00129682, 14680262. URL `http://www.jstor.org/stable/40263848`.

C. Fox and R. A. Norton. Fast sampling in a linear-Gaussian inverse problem. *SIAM/ASA Journal on Uncertainty Quantification*, 4:1191–1218, 2016.

T. E. Fricker, J. E. Oakley, and N. M. Urban. Multivariate Gaussian process emulators with non-separable covariance structures. *Technometrics*, 55(1):47–56, 2013.

R. Furrer and S. R. Sain. spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. *Journal of Statistical Software*, 36(10):1–25, 2010a.

R. Furrer, M. G. Genton, and D. Nychka. Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15:502–523, 2006.

Reinhard Furrer and Stephan R. Sain. spam: A sparse matrix R package with emphasis on MCMC methods for Gaussian Markov random fields. *Journal of Statistical Software*, 36(10):1–25, 2010b.

D. Gamerman. Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7:57–68, 1997.

S. Gasmi, N. H. Ogden, P. A. Leighton, L. R. Lindsay, and K. Thivierge. Analysis of the human population bitten by *Ixodes scapularis* ticks in quebec, canada: Increasing risk of lyme disease. *Ticks Tick Borne Dis*, 7(6):1075–1081, 2016. ISSN 1877-9603 (Electronic) 1877-959X (Linking). doi:10.1016/j.ttbdis.2016.09.006. URL `https://www.ncbi.nlm.nih.gov/pubmed/27650641`.

Theo Gasser and Hans-Georg Müller. Kernel estimation of regression functions. In Th. Gasser and M. Rosenblatt, editors, *Smoothing Techniques for Curve Estimation*, pages 23–68, Berlin, Heidelberg, 1979. Springer Berlin Heidelberg. ISBN 978-3-540-38475-5.

A. E. Gelfand and S. K. Sahu. Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, 94(445):247–253, 1999.

A. E. Gelfand and A. F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

A. E. Gelfand, H. J. Kim, C. F. Sirmans, and S. Banerjee. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396, 2003.

Alan E. Gelfand and Erin M. Schliep. Spatial statistics and Gaussian processes: A beautiful marriage. *Spatial Statistics*, 18(Part A):86 – 104, 2016.

A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533, 2006.

A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis*. Chapman & Hall/CRC, Boca Raton, FL, 3rd. edition, 2014.

S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:721–741, 1984.

F. Gerber and R. Furrer. Pitfalls in the implementation of Bayesian hierarchical modeling of areal count data: An illustration using BYM and Leroux models. *Journal of Statistical Software*, 63 (1):1–32, 2015.

G. H. Golub and C. F. Van Loan. *Matrix Computations*. The Johns Hopkins University Press, Baltimore, 3rd edition, 1996.

J. E. Gonzalez, Y. Low, A. Gretton, and C. Guestrin. Parallel Gibbs sampling: From colored fields to thin junction trees. In *Journal of Machine Learning Research: Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 324–332. 2011a.

Joseph Gonzalez, Yucheng Low, Arthur Gretton, and Carlos Guestrin. Parallel Gibbs sampling: From colored fields to thin junction trees. In Geoffrey Gordon, David Dunson, and Miroslav Dudík, editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 324–332, Fort Lauderdale, FL, USA, 11–13 Apr 2011b. PMLR.

R. Gramacy and D. Apley. Local Gaussian process approximation for large computer experiments. *Journal of Computational and Graphical Statistics*, 24:561–578, 2015.

B. Greig, K. M. Asanovich, P. J. Armstrong, and J. S. Dumler. Geographic, clinical, serologic, and molecular evidence of granulocytic ehrlichiosis, a likely zoonotic disease, in minnesota and wisconsin dogs. *J Clin Microbiol*, 34(1):44–8, 1996. ISSN 0095-1137 (Print) 0095-1137 (Linking). URL `http://www.ncbi.nlm.nih.gov/pubmed/8748270`.

Rajarshi Guhaniyogi, Andrew O. Finley, Sudipto Banerjee, and Alan E. Gelfand. Adaptive Gaussian predictive process models for large spatial datasets. *Environmetrics*, 22(8):997–1007, 2011.

Marc Hallin, Zudi Lu, and Keming Yu. Local linear spatial quantile regression. *Bernoulli*, 15(3): 659–686, 08 2009. doi:10.3150/08-BEJ168. URL `https://doi.org/10.3150/08-BEJ168`.

G. A. Hanzlicek, R. K. Raghavan, R. R. Ganta, and G. A. Anderson. Bayesian space-time patterns and climatic determinants of bovine anaplasmosis. *PLoS One*, 11(3):e0151924, 2016. ISSN 1932-6203 (Electronic) 1932-6203 (Linking). doi:10.1371/journal.pone.0151924. URL `https://www.ncbi.nlm.nih.gov/pubmed/27003596`.

J. Harrison and M. West. *Bayesian Forecasting and Dynamic Models*. Springer, New York City, 1999.

J. A. Hartigan and M. A. Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 28(1):100–108, 1979.

W. Hastings. Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, 57:97–109, 1970.

Xuming He and Peide Shi. Bivariate tensor-product b-splines in a partly linear model. *Journal of Multivariate Analysis*, 58(2):162 – 181, 1996. ISSN 0047-259X. doi:https://doi.org/10.1006/jmva.1996.0045. URL `http://www.sciencedirect.com/science/article/pii/S0047259X96900457`.

M. J. Heaton, W. F. Christensen, and M. A. Terres. Nonstationary Gaussian process models using spatial hierarchical clustering from finite differences. *Technometrics*, 59:93–101, 2017a.

M. J. Heaton, A. Datta, A. Finley, R. Furrer, R. Guhaniyogi, R. Gerber, R. Gramacy, et al. Methods for analyzing large spatial data: A review and comparision. Arxiv preprint 1710.05013, 2017b.

Brian Hendricks and Miguella Mark-Carew. Using exploratory data analysis to identify and predict patterns of human Lyme disease case clustering within a multistate region, 2010–2014. *Spatial and Spatio-temporal Epidemiology*, 20:35–43, 2017.

John A Herrmann, Nicole M Dahm, Marilyn O Ruiz, and William M Brown. Temporal and spatial distribution of tick-borne disease cases among humans and canines in Illinois (2000–2009). *Environmental Health Insights*, 8(Supplement 2):15, 2014.

D. M. Higdon. Auxiliary variable methods for Markov chain Monte Carlo with applications. *Journal of the American Statistical Association*, 93(442):585–595, 1998.

A. F. Hinckley, N. P. Connally, J. I. Meek, B. J. Johnson, M. M. Kemperman, K. A. Feldman, J. L. White, and P. S. Mead. Lyme disease testing by large commercial laboratories in the united states. *Clin Infect Dis*, 59(5):676–81, 2014. ISSN 1537-6591 (Electronic) 1058-4838 (Linking). doi:10.1093/cid/ciu397. URL `http://www.ncbi.nlm.nih.gov/pubmed/24879782`.

J. S. Hodges, B. P. Carlin, and Q. Fan. On the precision of the conditionally autoregressive prior in spatial models. *Biometrics*, 59:317–322, 2003.

Robert Hodrick and Edward Prescott. Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29(1):1–16, 1997.

F. J. Hosking, J. A. Sterne, G. D. Smith, and P. J. Green. Inference from genome-wide association studies using a novel markov model. *Genet Epidemiol*, 32(6):497–504, 2008. ISSN 1098-2272 (Electronic) 0741-0395 (Linking). doi:10.1002/gepi.20322. URL `http://www.ncbi.nlm.nih.gov/pubmed/18383184`.

M. L. Hutchinson, M. D. Strohecker, T. W. Simmons, A. D. Kyle, and M. W. Helwig. Prevalence rates of *Borrelia burgdorferi* (spirochaetales: Spirochaetaceae), *Anaplasma phagocytophilum* (rickettsiales: Anaplasmataceae), and *Babesia microti* (piroplasmida: Babesiidae) in host-seeking *Ixodes scapularis* (acari: Ixodidae) from pennsylvania. *J Med Entomol*, 52(4):693–8, 2015. ISSN 0022-2585 (Print) 0022-2585 (Linking). doi:10.1093/jme/tjv037. URL `http://www.ncbi.nlm.nih.gov/pubmed/26335476`.

Michael Jerrett. Atmospheric science: The death toll from air-pollution sources. *Nature*, 525(7569): 330 – 331, 2015. ISSN 1476-4687. URL `http://libproxy.clemson.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mnh&AN=26381981`.

Leah R. Johnson, Robert B. Gramacy, Jeremy Cohen, Erin A. Mordecai, Courtney Murdock, Jason Rohr, Sadie J. Ryan, Anna M. Stewart-Ibarra, and Daniel Weikel. Phenomenological forecasting of disease incidence using heteroskedastic Gaussian processes: A dengue case study. *Annals of Applied Statitics*, In Press, 2017.

M. Johnson, J. Saunderson, and A. Willsky. Analyzing Hogwild parallel Gaussian Gibbs sampling. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 2715–2723. Curran Associates, Inc., 2013. URL `http://papers.nips.cc/paper/5043-analyzing-hogwild-parallel-gaussian-gibbs-sampling.pdf`.

Giovanna Jona-Lasinio, Alan Gelfand, and Mattia Jona-Lasinio. Spatial analysis for wave direction data using wrapped Gassian processes. *Annals of Applied Statistics*, 6(4):1478–1498, 2012.

M. C. Jones, S. J. Davies, and B. U. Park. Versions of kernel-type regression estimators. *Journal of the American Statistical Association*, 89(427):825–832, 1994. ISSN 01621459. URL `http://www.jstor.org/stable/2290908`.

R. E. Kass, B. P. Carlin, A. Gelman, and R. Neal. Markov chain Monte Carlo in practice: A roundtable discussion. *The American Statistician*, 52:93–100, 1998.

M. Katzfuss. A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association*, 112:201–214, 2017.

L. Knorr-Held and H. Rue. On block updating in Markov random field models for disease mapping. *Scandinavian Journal of Statistics*, 29(4):597–614, 2002.

Leonhard Knorr-Held and Julian Besag. Modelling risk from a disease in time and space. *Statistics in Medicine*, 17(18):2045–2060, 1998. doi:10.1002/(SICI)1097-0258(19980930)17:18<2045::AID-SIM943>3.0.CO;2-P.

Leonhard Knorr-Held and Gnter Raer. Bayesian detection of clusters and discontinuities in disease maps. *Biometrics*, 56(1):13–21, 2000.

R. Koenker and P. Ng. *SparseM: Sparse Linear Algebra*, 2016. URL `https://CRAN.R-project.org/package=SparseM`. R package version 1.74.

Roger Koenker. *Quantile Regression*. Econometric Society Monographs. Cambridge University Press, 2005. doi:10.1017/CBO9780511754098.

Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978. URL `http://www.jstor.org/stable/1913643`.

B. Kohn, D. Galke, P. Beelitz, and K. Pfister. Clinical features of canine granulocytic anaplasmosis in 18 naturally infected dogs. *J Vet Intern Med*, 22(6):1289–95, 2008. ISSN 0891-6640 (Print) 0891-6640 (Linking). doi:10.1111/j.1939-1676.2008.0180.x. URL `http://www.ncbi.nlm.nih.gov/pubmed/18783353`.

Athanasios Kottas and Milovan Krnjajić. Bayesian semiparametric modelling in quantile regression. *Scandinavian Journal of Statistics*, 36(2):297–319, 2009. doi:10.1111/j.1467-9469.2008.00626.x. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9469.2008.00626.x`.

Hideo Kozumi and Genya Kobayashi. Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation*, 81(11):1565–1578, 2011.

P. J. Krause, K. McKay, C. A. Thompson, V. K. Sikand, R. Lentz, T. Lepore, L. Closter, D. Christianson, S. R. Telford, D. Persing, J. D. Radolf, A. Spielman, and Group Deer-Associated Infection Study. Disease-specific diagnosis of coinfecting tickborne zoonoses: babesiosis, human granulocytic ehrlichiosis, and lyme disease. *Clin Infect Dis*, 34(9):1184–91, 2002. ISSN 1537-6591 (Electronic) 1058-4838 (Linking). doi:10.1086/339813. URL `https://www.ncbi.nlm.nih.gov/pubmed/11941544`.

I. Krupka and R. K. Straubinger. Lyme borreliosis in dogs and cats: background, diagnosis, treatment and prevention of infections with *Borrelia burgdorferi* sensu stricto. *Vet Clin North Am Small Anim Pract*, 40(6):1103–19, 2010. ISSN 1878-1306 (Electronic) 0195-5616 (Linking). doi:10.1016/j.cvsm.2010.07.011. URL `https://www.ncbi.nlm.nih.gov/pubmed/20933139`.

Kiersten J Kugeler, Grace M Farley, Joseph D Forrester, and Paul S Mead. Geographic distribution and expansion of human Lyme disease, United States. *Emerging Infectious Diseases*, 21(8):1455–1457, 2015.

R. L. LaFleur, J. C. Dant, T. L. Wasmoen, S. M. Callister, D. A. Jobe, S. D. Lovrich, T. F. Warner, O. Abdelmagid, and R. F. Schell. Bacterin that induces anti-ospa and anti-ospc borreliacidal antibodies provides a high level of protection against canine lyme disease. *Clin Vaccine Immunol*, 16(2):253–9, 2009. ISSN 1556-679X (Electronic) 1556-679X (Linking). doi:10.1128/CVI.00373-08. URL `https://www.ncbi.nlm.nih.gov/pubmed/19052162`.

Stefan Lang and Andreas Brezger. Bayesian p-splines. *Journal of Computational and Graphical Statistics*, 13(1):183–212, 2004. doi:10.1198/1061860043010. URL `https://doi.org/10.1198/1061860043010`.

Paul M Lantos, Jean Tsao, Lise E Nigrovic, Paul G Auwaerter, Vance G Fowler, Felicia Ruffin, Erik Foster, and Graham Hickling. Geographic expansion of Lyme disease in Michigan, 2000–2014. In *Open Forum Infectious Diseases*, volume 4. Oxford University Press, 2017.

C. M. Lathers. Role of veterinary medicine in public health: antibiotic use in food animals and humans and the effect on evolution of antibacterial resistance. *J Clin Pharmacol*, 41(6):595–9, 2001. ISSN 0091-2700 (Print) 0091-2700 (Linking). URL `https://www.ncbi.nlm.nih.gov/pubmed/11402627`.

Michael L. Lavine and James S. Hodges. On rigorous specification of icar models. *The American Statistician*, 66(1):42–49, 2012. ISSN 00031305. URL `http://www.jstor.org/stable/23339520`.

N. A. Lazar. *The Statistical Analysis of Functional MRI Data*. Springer Science + Business Media, LLC, New York, 2008.

D. Lee. CARBayes: An R package for Bayesian spatial modeling with conditional autoregressive priors. *Journal of Statistical Software*, 55(13):1–24, 2013.

Duncan Lee and Andrew Lawson. Cluster detection and risk estimation for spatio-temporal health data. *arXiv preprint arXiv:1408.1191*, 2014.

Pierre Legendre. Spatial autocorrelation: Trouble or new paradigm? *Ecology*, 74(6):1659–1673, 1993.

C. E. V. Leser. A simple method of trend construction. *Journal of the Royal Statistical Society. Series B (Methodological)*, 23(1):91–107, 1961. ISSN 00359246. URL `http://www.jstor.org/stable/2983845`.

T. Levi, F. Keesing, R. D. Holt, M. Barfield, and R. S. Ostfeld. Quantifying dilution and amplification in a community of hosts for tick-borne pathogens. *Ecol Appl*, 28(2):484–498, 2016. ISSN 1051-0761 (Print) 1051-0761 (Linking). URL `https://www.ncbi.nlm.nih.gov/pubmed/27209790`.

S. A. Levy. Use of a c6 elisa test to evaluate the efficacy of a whole-cell bacterin for the prevention of naturally transmitted canine *Borrelia burgdorferi* infection. *Vet Ther.*, 3(4):420–424, 2002. ISSN 1528-3593 (Print) 1528-3593 (Linking). URL `https://www.ncbi.nlm.nih.gov/pubmed/12584679`.

S. A. Levy, B. A. Lissman, and C. M. Ficke. Performance of a *Borrelia burgdorferi* bacterin in borreliosis-endemic areas. *J Am Vet Med Assoc*, 202(11):1834–8, 1993. ISSN 0003-1488 (Print) 0003-1488 (Linking). URL `http://www.ncbi.nlm.nih.gov/pubmed/8320151`.

SA Levy and LA Magnarelli. Relationship between development of antibodies to *Borrelia burgdorferi* in dogs and the subsequent development of limb/joint borreliosis. *Journal of the American Veterinary Medical Association*, 200(3):344–347, 1992.

Steven A Levy, Thomas P O'Connor, Jancy L Hanscom, Paulette Shields, Leif Lorentzen, and Anthony A Dimarco. Quantitative measurement of c6 antibody following antibiotic treatment of *Borrelia burgdorferi* antibody-positive nonclinical dogs. *Clinical and vaccine immunology : CVI*, 15(1):115–9, jan 2008. ISSN 1556-679X. doi:10.1128/CVI.00340-07. URL `http://www.ncbi.nlm.nih.gov/pubmed/18003819http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2223868`.

I. Lilliehook, A. Egenvall, and Tvedten H.W. Hematopathology in dogs experimentally infected with a swedish granulocytic *Ehrlichia* species. *Vet Clin Pathol*, 27(4):116–122, 1998. ISSN 1939-165X (Electronic) 0275-6382 (Linking). URL `http://www.ncbi.nlm.nih.gov/pubmed/12075539`.

F. Lindgren and H. Rue. Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63(19):1–25, 2015.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society, Series B*, 73(4):423–498, 2011a.

Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Methodological)*, 73(4):423–498, 2011b.

Bo H. Lindqvist and Gunnar Taraldsen. On the proper treatment of improper distribtuions. Arxiv preprint 1711.02064, 2017.

Mandi Lingren, WA Rowley, Curt Thompson, and Mary Gilchrist. Geographic distribution of ticks (Acari: Ixodidae) in Iowa with emphasis on *Ixodes scapularis* and their infection with *Borrelia burgdorferi*. *Vector-Borne and Zoonotic Diseases*, 5(3):219–226, 2005.

S. C. Lipsett, N. R. Pollock, J. A. Branda, C. D. Gordon, C. R. Gordon, P. M. Lantos, and L. E. Nigrovic. The positive predictive value of lyme elisa for the diagnosis of lyme disease in children. *Pediatr Infect Dis J*, 34(11):1260–1262, 2015. ISSN 1532-0987 (Electronic) 0891-3668 (Linking). doi:10.1097/INF.0000000000000858. URL `http://www.ncbi.nlm.nih.gov/pubmed/26222064`.

S. E. Little. Ehrlichiosis and anaplasmosis in dogs and cats. *Vet Clin North Am Small Anim Pract*, 40(6):1121–40, 2010. ISSN 1878-1306 (Electronic) 0195-5616 (Linking). doi:10.1016/j.cvsm.2010.07.004. URL `http://www.ncbi.nlm.nih.gov/pubmed/20933140`.

S E Little, S R Heise, B L Blagburn, S M Callister, and P S Mead. Lyme borreliosis in dogs and humans in the USA. *Trends in Parasitology*, 26(4):213–218, 2010a.

S. E. Little, S. R. Heise, B. L. Blagburn, S. M. Callister, and P. S. Mead. Lyme borreliosis in dogs and humans in the usa. *Trends Parasitol*, 26(4):213–218, 2010b. ISSN 1471-5007 (Electronic) 1471-4922 (Linking). doi:10.1016/j.pt.2010.01.006. URL `http://www.ncbi.nlm.nih.gov/pubmed/20207198`.

S. E. Little, M. J. Beall, D. D. Bowman, R. Chandrashekar, and J. Stamaris. Canine infection with *Dirofilaria immitis, Borrelia burgdorferi, Anaplasma* spp., and *Ehrlichia* spp. in the united states, 2010-2012. *Parasit Vectors*, 7:257, 2014. ISSN 1756-3305 (Electronic) 1756-3305 (Linking). doi:10.1186/1756-3305-7-257. URL `http://www.ncbi.nlm.nih.gov/pubmed/24886589`.

M. P. Littman, R. E. Goldstein, M. A. Labato, M. R. Lappin, and G. E. Moore. Acvim small animal consensus statement on lyme disease in dogs: diagnosis, treatment, and prevention. *J Vet Intern Med*, 20(2):422–34, 2006. ISSN 0891-6640 (Print) 0891-6640 (Linking). URL `http://www.ncbi.nlm.nih.gov/pubmed/16594606`.

J. S. Liu, W. H. Wong, and A. Kong. Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, 81:27–40, 1994.

Y. Liu, O. Kosut, and A. S. Willsky. Sampling from Gaussian Markov random fields using stationary and non-stationary subgraph pertubations. *IEEE Transactions on Signal Processing*, 63(3):576–589, 2015.

Y. Liu, R. B. Lund, S. K. Nordone, M. J. Yabsley, and C. S. McMahan. A bayesian spatio-temporal model for forecasting the prevalence of antibodies to *Ehrlichia* species in domestic dogs within the contiguous united states. *Parasit Vectors*, In Press, 2016.

A. Lopez-Quilez and F. Munoz. Review of spatio-temporal models for disease mapping. Technical report, Final report for the Euroheis 2 project. Zenodo., 2009. URL `http://www.uv.es/~famarmu/doc/Euroheis2-report.pdf`.

Kristian Lum and Alan E. Gelfand. Spatial quantile multiple regression using the asymmetric Laplace process. *Bayesian Analysis*, 7(2):235–258, 2012.

D. J. Lunn, A. Thomas, N. Best, and D. Spiegelhalter. WinBUGS – A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10:325–337, 2000.

Y. C. MacNab and C. B. Dean. Autoregressive spatial smoothing and temporal spline smoothing for mapping rates. *Biometrics*, 57:949–56, 2001.

Nita K Madhav, John S Brownstein, Jean I Tsao, and Durland Fish. A dispersal model for the range expansion of blacklegged tick (acari: Ixodidae). *Med. Entomol*, 41(5):842–852, 2004.

Ricardo G. Maggi, Sara Reichelt, Marcée Toliver, and Barry Engber. *Borrelia* species in *Ixodes affinis* and *Ixodes scapularis* ticks collected from the coastal plain of north carolina. *Ticks and Tick-borne Diseases*, 1(4):168–171, 2010. ISSN 1877-959X. doi:http://dx.doi.org/10.1016/j.ttbdis.2010.08.003. URL `//www.sciencedirect.com/science/article/pii/S1877959X10000683`.

L.C. Marsh, L.M.D.R. Cormier, and D.R. Cormier. *Spline Regression Models*. Number no. 137 in Online access: Sage Sage Research Methods Online. SAGE Publications, 2001. ISBN 9780761924203. URL `https://books.google.com/books?id=PYbvSt8fejoC`.

M. A. Martinez-Beneito, A. Lopez-Quilez, and P. Botella-Rocamora. An autoregressive approach to spatio-temporal disease mapping. *Stat Med*, 27(15):2874–89, 2008. ISSN 0277-6715 (Print) 0277-6715 (Linking). doi:10.1002/sim.3103. URL `http://www.ncbi.nlm.nih.gov/pubmed/17979141`.

K. Mathews, P. W. Kronen, D. Lascelles, A. Nolan, S. Robertson, P. V. Steagall, B. Wright, and K. Yamashita. Guidelines for recognition, assessment and treatment of pain: Wsava global pain council members and co-authors of this document. *J Small Anim Pract*, 55(6):E10–68, 2014. ISSN 1748-5827 (Electronic) 0022-4510 (Linking). doi:10.1111/jsap.12200. URL `http://www.ncbi.nlm.nih.gov/pubmed/24841489`.

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

C. S. McMahan, D. Wang, M. J. Beall, D. D. Bowman, S. E. Little, P. O. Pithua, J. L. Sharp, R. W. Stich, M. J. Yabsley, and R. B. Lund. Factors associated with *Anaplasma* spp. seroprevalence among dogs in the united states. *Parasit Vectors*, 9:169, 2016. ISSN 1756-3305 (Electronic) 1756-3305 (Linking). doi:10.1186/s13071-016-1431-7. URL `http://www.ncbi.nlm.nih.gov/pubmed/27004557`.

S. J. McNabb, R. A. Jajosky, P. A. Hall-Baker, D. A. Adams, P. Sharp, W. J. Anderson, A. J. Javier, G. J. Jones, D. A. Nitschke, C. A. Worshams, Jr. Richard, R. A., Control Centers for Disease, and Prevention. Summary of notifiable diseases — united states, 2005. *MMWR Morb Mortal Wkly Rep*, 54(53):1–92, 2007. ISSN 1545-861X (Electronic) 0149-2195 (Linking). URL `http://www.ncbi.nlm.nih.gov/pubmed/17392681`.

P. Mead, R. Goel, and K. Kugeler. Canine serology as adjunct to human lyme disease surveillance. *Emerg Infect Dis*, 17(9):1710–2, 2011a. ISSN 1080-6059 (Electronic) 1080-6040 (Linking). doi:10.3201/eid1709.110210. URL `http://www.ncbi.nlm.nih.gov/pubmed/21888800`.

P. Mead, R. Goel, and K. Kugeler. Canine serology as adjunct to human lyme disease surveillance. *Emerg Infect Dis*, 17(9):1710–2, 2011b. ISSN 1080-6059 (Electronic) 1080-6040 (Linking). doi:10.3201/eid1709.110210. URL `http://www.ncbi.nlm.nih.gov/pubmed/21888800`.

P. S. Mead. Epidemiology of lyme disease. *Infect Dis Clin North Am*, 29(2):187–210, 2015. ISSN 1557-9824 (Electronic) 0891-5520 (Linking). doi:10.1016/j.idc.2015.02.010. URL `http://www.ncbi.nlm.nih.gov/pubmed/25999219`.

Paul Mead, Rohan Goel, and Kiersten Kugeler. Canine serology as adjunct to human Lyme disease surveillance. *Emerging Infectious Diseases*, 17(9):1710, 2011c.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1091, 1953.

K. Millen, K. J. Kugeler, A. F. Hinckley, E. W. Lawaczeck, and P. S. Mead. Elevated lyme disease seroprevalence among dogs in a nonendemic county: harbinger or artifact? *Vector Borne Zoonotic Dis*, 13(5):340–341, 2013a. ISSN 1557-7759 (Electronic) 1530-3667 (Linking). doi:10.1089/vbz.2012.1025. URL `http://www.ncbi.nlm.nih.gov/pubmed/23421882`.

K. Millen, K. J. Kugeler, A. F. Hinckley, E. W. Lawaczeck, and P. S. Mead. Elevated lyme disease seroprevalence among dogs in a nonendemic county: harbinger or artifact? *Vector Borne Zoonotic Dis*, 13(5):340–1, 2013b. ISSN 1557-7759 (Electronic) 1530-3667 (Linking). doi:10.1089/vbz.2012.1025. URL `http://www.ncbi.nlm.nih.gov/pubmed/23421882`.

P. Morales-Álvarez, A. Pérez-Suay, R. Molina, and G. Camps-Valls. Remote sensing image classification with large-scale Gaussian processes. *IEEE Transactions on Geoscience and Remote Sensing*, PP(99):1–12, 2017.

R. Movilla, C. Garcia, S. Siebert, and X. Roura. Countrywide serological evaluation of canine prevalence for *Anaplasma* spp., *Borrelia burgdorferi* (sensu lato), *Dirofilaria immitis* and *Ehrlichia canis* in mexico. *Parasit Vectors*, 9(1):421, 2016. ISSN 1756-3305 (Electronic) 1756-3305 (Linking). doi:10.1186/s13071-016-1686-z. URL `https://www.ncbi.nlm.nih.gov/pubmed/27474020`.

John Muyskens, Dan Keating, and Samuel Granados. Mapping how the united states generates its electricity. *The Washington Post*, Mar 2017. URL `https://www.washingtonpost.com/graphics/national/power-plants/?noredirect=on&utm_term=.de2fee192a59`.

E. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142, 1964. doi:10.1137/1109020. URL `https://doi.org/10.1137/1109020`.

A. D. Nair, C. Cheng, C. K. Ganta, M. W. Sanderson, A. R. Alleman, U. G. Munderloh, and R. R. Ganta. Comparative experimental infection study in dogs with *Ehrlichia canis*, *E. chaffeensis*, *Anaplasma platys* and *A. phagocytophilum*. *PLoS One*, 11(2):e0148239, 2016. ISSN 1932-6203 (Electronic) 1932-6203 (Linking). doi:10.1371/journal.pone.0148239. URL `http://www.ncbi.nlm.nih.gov/pubmed/26840398`.

R. M. Neal. Regression and classification using Gaussian process priors. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 6*. Oxford University Press, New York, 1998.

M. P. Nelder, C. Russell, L. R. Lindsay, B. Dhar, S. N. Patel, S. Johnson, S. Moore, E. Kristjanson, Y. Li, and F. Ralevski. Population-based passive tick surveillance and detection of expanding foci of blacklegged ticks *Ixodes scapularis* and the lyme disease agent *Borrelia burgdorferi* in ontario, canada. *PLoS One*, 9(8):e105358, 2014. ISSN 1932-6203 (Electronic) 1932-6203 (Linking). doi:10.1371/journal.pone.0105358. URL `https://www.ncbi.nlm.nih.gov/pubmed/25171252`.

Mark P Nelder, Curtis B Russell, Nina Jain Sheehan, Beate Sander, Stephen Moore, Ye Li, Steven Johnson, Samir N Patel, and Doug Sider. Human pathogens associated with the blacklegged tick *Ixodes scapularis*: A systematic review. *Parasites and Vectors*, 9(1):265, 2016.

C. A. Nelson, S. Saha, K. J. Kugeler, M. J. Delorey, M. B. Shankar, A. F. Hinckley, and P. S. Mead. Incidence of clinician-diagnosed lyme disease, united states, 2005-2010. *Emerg Infect Dis*, 21(9): 1625–1631, 2015. ISSN 1080-6059 (Electronic) 1080-6040 (Linking). doi:10.3201/eid2109.150417. URL `http://www.ncbi.nlm.nih.gov/pubmed/26291194`.

J. M. Neuhaus. Bias and efficiency loss due to misclassified responses in binary regression. *Biometrika*, 86(4):843–855, 1999.

E. g. Ng and B. W. Peyton. Block sparse Cholesky algorithms on advanced uniprocessor computers. *SIAM Journal on Scientific Computing*, 14(5):1034–1056, 1993.

N. C. Nieto, E. A. Holmes, and J. E. Foley. Survival rates of immature *Ixodes pacificus* (acari: Ixodidae) ticks estimated using field-placed enclosures. *J Vector Ecol*, 35(1):43–9, 2010. ISSN 1948-7134 (Electronic) 1081-1710 (Linking). doi:10.1111/j.1948-7134.2010.00026.x. URL `http://www.ncbi.nlm.nih.gov/pubmed/20618646`.

F. Niu, B. Recht, C. Ré, and S. J. Wright. Hogwild! a lock-free approach to parallelizing stochastic gradient descent. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 693–701. Curran Associates, Inc., 2011. URL `http://papers.nips.cc/paper/4390-hogwild-a-lock-free-approach-to-parallelizing-stochastic-gradient-descent.pdf`.

A. A. Nobre, A. M. Schmidt, and H. F. Lopes. Spatio-temporal models for mapping the incidence of malaria in pará. *Environmetrics*, 16(3):291–304, 2005.

D. Nychka, S. Bandyopadhyay, D. Hammerling, F. Lindgren, and S. Sain. A multiresolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, 24:579–599, 2015.

D. Nyman, L. Willen, C. Jansson, S. A. Carlsson, H. Granlund, and P. Wahlberg. Vlse c6 peptide and igg elisa antibody analysis for clinical diagnosis of lyme borreliosis in an endemic area. *Clin Microbiol Infect*, 12(5):496–497, 2006. ISSN 1198-743X (Print) 1198-743X (Linking). doi:10.1111/j.1469-0691.2006.01374.x. URL `http://www.ncbi.nlm.nih.gov/pubmed/16643532`.

T. P. O'Connor, J. L. Hanscom, B. C. Hegarty, R. G. Groat, and E. B. Breitschwerdt. Comparison of an indirect immunofluorescence assay, western blot analysis, and a commercially available elisa for detection of *Ehrlichia canis* antibodies in canine sera. *Am J Vet Res*, 67(2):206–10, 2006. ISSN 0002-9645 (Print) 0002-9645 (Linking). doi:10.2460/ajvr.67.2.206. URL `http://www.ncbi.nlm.nih.gov/pubmed/16454622`.

N. H. Ogden, M. Bigras-Poulin, C. J. O'Callaghan, I. K. Barker, K. Kurtenbach, L. R. Lindsay, and D. F. Charron. Vector seasonality, host infection dynamics and fitness of pathogens transmitted by the tick *Ixodes scapularis*. *Parasitology*, 134(Pt 2):209–227, 2007. ISSN 0031-1820 (Print) 0031-1820 (Linking). doi:10.1017/S0031182006001417. URL `http://www.ncbi.nlm.nih.gov/pubmed/17032476`.

N. H. Ogden, S. Mechai, and G. Margos. Changing geographic ranges of ticks and tick-borne pathogens: drivers, mechanisms and consequences for pathogen diversity. *Front Cell Infect Microbiol*, 3:46, 2013. ISSN 2235-2988 (Electronic) 2235-2988 (Linking). doi:10.3389/fcimb.2013.00046. URL `http://www.ncbi.nlm.nih.gov/pubmed/24010124`.

N. H. Ogden, M. Radojevic, X. Wu, V. R. Duvvuri, P. A. Leighton, and J. Wu. Estimated effects of projected climate change on the basic reproductive number of the lyme disease vector *Ixodes scapularis*. *Environ Health Perspect*, 122(6):631–638, 2014. ISSN 1552-9924 (Electronic) 0091-6765 (Linking). doi:10.1289/ehp.1307799. URL `http://www.ncbi.nlm.nih.gov/pubmed/24627295`.

A. O'Hagan. Curve fitting and optimal design for prediction. *Journal of the Royal Statistical Society: Series B (Methodological)*, 40:1–42, 1978.

M. Blanca Palacios and Mark F. J Steel. Non-gaussian bayesian geostatistical modeling. *Journal of the American Statistical Association*, 101(474):604–618, 2006. doi:10.1198/016214505000001195. URL `https://doi.org/10.1198/016214505000001195`.

P. E. Parham, J. Waldock, G. K. Christophides, D. Hemming, F. Agusto, K. J. Evans, N. Fefferman, H. Gaff, A. Gumel, S. LaDeau, S. Lenhart, R. E. Mickens, E. N. Naumova, R. S. Ostfeld, P. D. Ready, M. B. Thomas, J. Velasco-Hernandez, and E. Michael. Climate, environmental and socio-economic change: weighing up the balance in vector-borne disease transmission. *Philos Trans R Soc Lond B Biol Sci*, 370(1665), 2015. ISSN 1471-2970 (Electronic) 0962-8436 (Linking). doi:10.1098/rstb.2013.0551. URL `http://www.ncbi.nlm.nih.gov/pubmed/25688012`.

F. M. Poitout, J. K. Shinozaki, P. J. Stockwell, C. J. Holland, and S. K. Shukla. Genetic variants of *Anaplasma phagocytophilum* infecting dogs in western washington state. *J Clin Microbiol*, 43(2): 796–801, 2005. ISSN 0095-1137 (Print) 0095-1137 (Linking). doi:10.1128/JCM.43.2.796-801.2005. URL `http://www.ncbi.nlm.nih.gov/pubmed/15695682`.

Gabriela Polezer, Yara S. Tadano, Hugo V. Siqueira, Ana F.L. Godoi, Carlos I. Yamamoto, Paulo A. de Andr, Theotonio Pauliquevis, Maria de Fatima Andrade, Andrea Oliveira, Paulo H.N. Saldiva, Philip E. Taylor, and Ricardo H.M. Godoi. Assessing the impact of pm2.5 on respiratory disease using artificial neural networks. *Environmental Pollution*, 235:394 – 403, 2018. ISSN 0269-7491. doi:https://doi.org/10.1016/j.envpol.2017.12.111. URL `http://www.sciencedirect.com/science/article/pii/S0269749117337107`.

Nicholas G. Polson, James G. Scott, and Jesse Windle. Bayesian inference for logistic models using pólya gamma latent variables. *Journal of the American Statistical Association*, 108(504):1339–1349, 2013.

W. C. Porter, C. L. Heald, D. Cooley, and B. Russell. Investigating the observed sensitivities of air-quality extremes to meteorological drivers via quantile regression. *Atmospheric Chemistry and Physics*, 15, 2015.

M. B. Priestley and M. T. Chao. Non-parametric function fitting. *Journal of the Royal Statistical Society. Series B (Methodological)*, 34(3):385–392, 1972. ISSN 00359246. URL `http://www.jstor.org/stable/2985075`.

B. A. Qurollo, R. Chandrashekar, B. C. Hegarty, M. J. Beall, B. A. Stillman, J. Liu, B. Thatcher, E. Pultorak, B. Cerrito, M. Walsh, and E. B. Breitschwerdt. A serological survey of tick-borne pathogens in dogs in north america and the caribbean as assessed by *Anaplasma phagocytophilum, A. platys, Ehrlichia canis, E. chaffeensis, E. ewingii*, and *Borrelia burgdorferi* species-specific peptides. *Infect Ecol Epidemiol*, 4, 2014. ISSN 2000-8686 (Linking). doi:10.3402/iee.v4.24699. URL `https://www.ncbi.nlm.nih.gov/pubmed/25405006`.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL `https://www.R-project.org/`.

R. K. Raghavan, D. Neises, D. G. Goodin, D. A. Andresen, and R. R. Ganta. Bayesian spatio-temporal analysis and geospatial risk factors of human monocytic ehrlichiosis. *PLoS One*, 9(7):e100850, 2014. ISSN 1932-6203 (Electronic) 1932-6203 (Linking). doi:10.1371/journal.pone.0100850. URL `http://www.ncbi.nlm.nih.gov/pubmed/24992684`.

R. K. Raghavan, D. G. Goodin, D. Neises, G. A. Anderson, and R. R. Ganta. Hierarchical bayesian spatio-temporal analysis of climatic and socio-economic determinants of rocky mountain spotted fever. *PLoS One*, 11(3):e0150180, 2016. ISSN 1932-6203 (Electronic) 1932-6203 (Linking). doi:10.1371/journal.pone.0150180. URL `http://www.ncbi.nlm.nih.gov/pubmed/26942604`.

Aaditya Ramdas and Ryan J. Tibshirani. Fast and flexible admm algorithms for trend filtering. 25, 06 2014.

Carl E. Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.

Brian J. Reich, Montserrat Fuentes, and David B. Dunson. Bayesian spatial quantile regression. *Journal of the American Statistical Association*, 106(493):6–20, 2011. ISSN 01621459. URL `http://www.jstor.org/stable/41415529`.

D. V. Rhodes, C. G. Earnhart, T. N. Mather, P. F. Meeus, and R. T. Marconi. Identification of *Borrelia burgdorferi* ospc genotypes in canine tissue following tick infestation: implications for lyme disease vaccine and diagnostic assay design. *Vet J*, 198(2):412–8, 2013. ISSN 1532-2971 (Electronic) 1090-0233 (Linking). doi:10.1016/j.tvjl.2013.07.019. URL `http://www.ncbi.nlm.nih.gov/pubmed/23962611`.

C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer, New York, 2nd edition, 2004.

Edward A. Roualdes. Bayesian trend filtering. *arXiv:1505.07710*, 2015.

H. Rue. Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society, Series B*, 63(2):325–338, 2001.

H. Rue and L. Held. *Gaussian Markov Random Fields*. Chapman & Hall/CRC, Boca Raton, 2005a.

H. Rue and H. Tjemland. Fitting Gaussian Markov random fields to Gaussian fields. *Scandinavian Journal of Statistics*, 29:31–49, 2002.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society, Series B*, 71(2):319–392, 2009.

Håvard Rue and Leonhard Held. *Gaussian Markov random fields: Theory and applications*. Chapman and Hall/CRC, 2005b.

Alastair Rushworth, Duncan Lee, and Richard Mitchell. A spatio-temporal model for estimating the long-term effects of air pollution on respiratory hospital admissions in greater London. *Spatial and Spatio-temporal Epidemiology*, 10:29–38, 2014.

Nathan M Russart, Michael W Dougherty, and Jefferson A Vaughan. Survey of ticks (Acari: Ixodidae) and tick-borne pathogens in North Dakota. *Journal of Medical Entomology*, 51(5):1087–1090, 2014.

Brook T. Russell, Dewei Wang, and Christopher S. McMahan. Spatially modeling the effects of meteorological drivers of pm2.5 in the eastern united states via a local linear penalized quantile regression estimator. *Environmetrics*, 28(5):e2448, 2017. doi:10.1002/env.2448. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/env.2448`. e2448 env.2448.

Veeranjaneyulu Sadhanala and Ryan J. Tibshirani. Additive models with trend filtering. 02 2017.

H. Sang, M. Jun, and J. Z. Huang. Covariance approximation for large multivariate spatial datasets with an application to multiple climate model errors. *Annals of Applied Statistics*, 5(4):2519–2548, 2011.

T. J. Santner, B. J. Williams, and W. I. Notz. *The Design and Analysis of Computer Experiments*. Springer-Verlag, New York, 2003.

O. Schabenberger and C. A. Gotway. *Statistical Methods for Spatial Data Analysis*. Chapman & Hall/CRC, Boca Raton, 2005.

J. D. Scott, J. E. Foley, K. L. Clark, J. F. Anderson, L. A. Durden, J. M. Manord, and M. L. Smith. Established population of blacklegged ticks with high infection prevalence for the lyme disease bacterium, *Borrelia burgdorferi* sensu lato, on corkscrew island, kenora district, ontario. *Int J Med Sci*, 13(11):881–891, 2016. ISSN 1449-1907 (Electronic) 1449-1907 (Linking). doi:10.7150/ijms.16922. URL `https://www.ncbi.nlm.nih.gov/pubmed/27877080`.

John D Scott. Studies abound on how far north *Ixodes scapularis* ticks are transported by birds. *Ticks and Tick-borne Diseases*, 7:327–328, 2016. doi:10.1016/j.ttbdis.2015.12.001.

Ransalu Senanayake, Simon O'Callaghan, and Fabio Ramos. Predicting spatio-temporal propagation of seasonal influenza using variational Gaussian process regression. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, pages 3901–3907. AAAI Press, 2016.

Kim Seung-Jean, Koh Kwangmoo, and Dimitry Gorinevsky. l1 trend filtering. *SIAM Reivew*, 51(2): 339–360, 2009.

A. H. Sherman. On the efficient solution of sparse systems of linear and nonlinear equations. Unpublished doctoral dissertation, Yale University, 1975.

M. Smith and L. Fahrmeir. Spatial Bayesian variable selection with application to functional magnetic resonance imaging. *Journal of the American Statistical Association*, 102(478):417–431, 2007.

H.-R. Song, M. Fuentes, and S. Ghosh. A comparative study of Gaussian geostatistical models and Gaussian Markov random fields. *Journal of Multivariate Analysis*, 99:1681–1697, 2008.

J. A. Spencer, J. M. Butler, K. C. Stafford, M. B. Pough, S. A. Levy, D. L. Bledsoe, and B. L. Blagburn. Evaluation of permethrin and imidacloprid for prevention of *Borrelia burgdorferi* transmission from blacklegged ticks (*Ixodes scapularis*) to *Borrelia burgdorferi*-free dogs. *Parasitol Res*, 90(3):S106–7, 2003. ISSN 0932-0113 (Print) 0932-0113 (Linking). doi:10.1007/s00436-003-0904-8. URL `http://www.ncbi.nlm.nih.gov/pubmed/12928869`.

G. Stanek, G. P. Wormser, J. Gray, and F. Strle. Lyme borreliosis. *Lancet*, 379(9814):461–473, 2012. ISSN 1474-547X (Electronic) 0140-6736 (Linking). doi:10.1016/S0140-6736(11)60103-7. URL http://www.ncbi.nlm.nih.gov/pubmed/21903253.

A. C. Steere, A. Dhar, J. Hernandez, P. A. Fischer, V. K. Sikand, R. T. Schoen, J. Nowakowski, G. McHugh, and D. H. Persing. Systemic symptoms without erythema migrans as the presenting picture of early lyme disease. *Am J Med*, 114(1):58–62, 2003a. ISSN 0002-9343 (Print) 0002-9343 (Linking). URL http://www.ncbi.nlm.nih.gov/pubmed/12543291.

A. C. Steere, G. McHugh, C. Suarez, J. Hoitt, N. Damle, and V. K. Sikand. Prospective study of coinfection in patients with erythema migrans. *Clin Infect Dis*, 36(8):1078–81, 2003b. ISSN 1537-6591 (Electronic) 1058-4838 (Linking). doi:10.1086/368187. URL https://www.ncbi.nlm.nih.gov/pubmed/12684924.

A. C. Steere, J. Coburn, and L. Glickstein. The emergence of lyme disease. *J Clin Invest*, 113 (8):1093–1101, 2004. ISSN 0021-9738 (Print) 0021-9738 (Linking). doi:10.1172/JCI21681. URL http://www.ncbi.nlm.nih.gov/pubmed/15085185.

H. S. Stern and N. Cressie. *Inference for extremes in disease mapping*, pages 63–84. John Wiley Sons, 1999.

R. W. Stich, B. L. Blagburn, D. D. Bowman, C. Carpenter, M. R. Cortinas, S. A. Ewing, D. Foley, J. E. Foley, H. Gaff, G. J. Hickling, R. R. Lash, S. E. Little, C. Lund, R. Lund, T. N. Mather, G. R. Needham, W. L. Nicholson, J. Sharp, A. Varela-Stokes, and D. Wang. Quantitative factors proposed to influence the prevalence of canine tick-borne disease agents in the united states. *Parasit Vectors*, 7:417, 2014. ISSN 1756-3305 (Electronic) 1756-3305 (Linking). doi:10.1186/1756-3305-7-417. URL http://www.ncbi.nlm.nih.gov/pubmed/25185829.

B. A. Stillman, M. Monn, J. Liu, B. Thatcher, P. Foster, B. Andrews, S. Little, M. Eberts, E. B. Breitschwerdt, M. J. Beall, and R. Chandrashekar. Performance of a commercially available in-clinic elisa for detection of antibodies against *Anaplasma phagocytophilum, Anaplasma platys, Borrelia burgdorferi, Ehrlichia canis*, and *Ehrlichia ewingii* and *Dirofilaria immitis* antigen in dogs. *J Am Vet Med Assoc*, 245(1):80–86, 2014. ISSN 1943-569X (Electronic) 0003-1488 (Linking). doi:10.2460/javma.245.1.80. URL http://www.ncbi.nlm.nih.gov/pubmed/24941391.

B. L. Stone, N. M. Russart, R. A. Gaultney, A. M. Floden, J. A. Vaughan, and C. A. Brissette. The western progression of lyme disease: infectious and nonclonal *Borrelia burgdorferi* sensu lato populations in grand forks county, north dakota. *Appl Environ Microbiol*, 81(1):48–58, 2015. ISSN 1098-5336 (Electronic) 0099-2240 (Linking). doi:10.1128/AEM.02422-14. URL https://www.ncbi.nlm.nih.gov/pubmed/25304515.

Charles J. Stone. The use of polynomial splines and their tensor products in multivariate function estimation. *Ann. Statist.*, 22(1):118–171, 03 1994. doi:10.1214/aos/1176325361. URL https://doi.org/10.1214/aos/1176325361.

S. J. Swanson, D. Neitzel, K. D. Reed, and E. A. Belongia. Coinfections acquired from *ixodes* ticks. *Clin Microbiol Rev*, 19(4):708–27, 2006. ISSN 0893-8512 (Print) 0893-8512 (Linking). doi:10.1128/CMR.00011-06. URL https://www.ncbi.nlm.nih.gov/pubmed/17041141.

A. Thompson, R. Mannix, and R. Bachur. Acute pediatric monoarticular arthritis: distinguishing lyme arthritis from other etiologies. *Pediatrics*, 123(3):959–65, 2009. ISSN 1098-4275 (Electronic) 0031-4005 (Linking). doi:10.1542/peds.2008-1511. URL http://www.ncbi.nlm.nih.gov/pubmed/19255026.

Ryan J Tibshirani. Adaptive piecewise polynomial estimation via trend filtering. *Annals of Statistics*, 42(1):285–323, 2014.

Fang Ting Liang, Richard H Jacobson, Reinhard K Straubinger, Amy Grooters, and Mario T Philipp. Characterization of a borrelia burgdorferi vlse invariable region useful in canine lyme disease serodiagnosis by enzyme-linked immunosorbent assay. *Journal of Clinical Microbiology*, 38(11): 4160–4166, 2000.

Thomas M. Trimbur. Detrending economic time series: a bayesian generalization of the hodrick-prescott filter. *Journal of Forecasting*, 25(4):247–273, 2006.

J. M. Ver Hoef and R. P. Barry. Constructing and fitting models for cokriging and multivariable spatial prediction. *Journal of Statistical Planning and Inference*, 69:275–294, 1998.

B. Wagner, H. Freer, A. Rollins, D. Garcia-Tapia, H. N. Erb, C. Earnhart, R. Marconi, and P. Meeus. Antibodies to *Borrelia burgdorferi* ospa, ospc, ospf, and c6 antigens as markers for early and late infection in dogs. *Clin Vaccine Immunol*, 19(4):527–535, 2012a. ISSN 1556-679X (Electronic) 1556-679X (Linking). doi:10.1128/CVI.05653-11. URL `http://www.ncbi.nlm.nih.gov/pubmed/22336289`.

Bettina Wagner, Heather Freer, Alicia Rollins, David Garcia-Tapia, Hollis N Erb, Christopher Earnhart, Richard Marconi, and Patrick Meeus. Antibodies to *Borrelia burgdorferi* OspA, OspC, OspF, and C6 antigens as markers for early and late infection in dogs. *Clinical and Vaccine Immunology*, 19(4):527–535, 2012b.

L. A. Waller, B. P. Carlin, H. Xia, and A. E. Gelfand. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92(438):607–617, 1997a.

Lance A. Waller, Bradley P. Carlin, Hong Xia, and Alan E. Gelfand. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92(438):607–617, 1997b.

Chenchen Wang, Yifan Tu, Zongliang Yu, and Rongzhu Lu. Pm2.5 and cardiovascular diseases in the elderly: An overview. *International Journal Of Environmental Research And Public Health*, 12 (7):8187 – 8197, 2015. ISSN 1660-4601. URL `http://libproxy.clemson.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=mnh&AN=26193289`.

D. Wang, D. D. Bowman, H. E. Brown, L. C. Harrington, P. E. Kaufman, T. McKay, C. T. Nelson, J. L. Sharp, and R. Lund. Factors influencing u.s. canine heartworm (*Dirofilaria immitis*) prevalence. *Parasit Vectors*, 7:264, 2014a. ISSN 1756-3305 (Electronic) 1756-3305 (Linking). doi:10.1186/1756-3305-7-264. URL `http://www.ncbi.nlm.nih.gov/pubmed/24906567`.

Peng Wang, Meaghan N Glowacki, Armando E Hoet, Glen R Needham, Kathleen A Smith, Richard E Gary, and Xin Li. Emergence of *Ixodes scapularis* and *Borrelia burgdorferi*, the Lyme disease vector and agent, in Ohio. *Frontiers in Cellular and Infection Microbiology*, 4, 2014b.

Yuedong Wang. *Spline Smoothing : Methods and Applications*. Chapman and Hall/CRC, 2011.

Geoffrey S. Watson. Smooth regression analysis. *Sankhy: The Indian Journal of Statistics, Series A (1961-2002)*, 26(4):359–372, 1964. ISSN 0581572X. URL `http://www.jstor.org/stable/25049340`.

Stephanie A. Weber, Tabassum Z. Insaf, Eric S. Hall, Thomas O. Talbot, and Amy K. Huff. Assessing the impact of fine particulate matter (pm2.5) on respiratory-cardiovascular chronic diseases in the new york city metropolitan area using hierarchical bayesian model estimates. *Environmental Research*, 151:399 – 409, 2016. ISSN 0013-9351. doi:https://doi.org/10.1016/j.envres.2016.07.012. URL `http://www.sciencedirect.com/science/article/pii/S0013935116302961`.

C. Wengenmayer, H. Williams, E. Zschiesche, A. Moritz, J. Langenstein, R. K. Roepke, and A. R. Heckeroth. The speed of kill of fluralaner (bravecto) against *Ixodes ricinus* ticks on dogs. *Parasit Vectors*, 7:525, 2014. ISSN 1756-3305 (Electronic) 1756-3305 (Linking). doi:10.1186/s13071-014-0525-3. URL `https://www.ncbi.nlm.nih.gov/pubmed/25406442`.

M. L. Wiseman-Orr, A. M. Nolan, J. Reid, and E. M. Scott. Development of a questionnaire to measure the effects of chronic pain on health-related quality of life in dogs. *Am J Vet Res*, 65 (8):1077–1084, 2004. ISSN 0002-9645 (Print) 0002-9645 (Linking). URL `http://www.ncbi.nlm.nih.gov/pubmed/15334841`.

G. P. Wormser, R. J. Dattwyler, E. D. Shapiro, J. J. Halperin, A. C. Steere, M. S. Klempner, P. J. Krause, J. S. Bakken, F. Strle, G. Stanek, L. Bockenstedt, D. Fish, J. S. Dumler, and R. B. Nadelman. The clinical assessment, treatment, and prevention of lyme disease, human granulocytic anaplasmosis, and babesiosis: clinical practice guidelines by the infectious diseases society of america. *Clin Infect Dis*, 43(9):1089–1134, 2006. ISSN 1537-6591 (Electronic) 1058-4838 (Linking). doi:10.1086/508667. URL `http://www.ncbi.nlm.nih.gov/pubmed/17029130`.

Tracy Z. Wu, Keming Yu, and Yan Yu. Single-index quantile regression. *Journal of Multivariate Analysis*, 101(7):1607 – 1621, 2010. ISSN 0047-259X. doi:https://doi.org/10.1016/j.jmva.2010.02.003. URL `http://www.sciencedirect.com/science/article/pii/S0047259X10000333`.

Yichao Wu and Yufeng Liu. Variable selection in quantile regression. *Statistica Sinica*, 19(2):801–817, 2009. ISSN 10170405, 19968507. URL `http://www.jstor.org/stable/24308857`.

H. Xia and B. P. Carlin. Spatio-temporal models with errors in covariates: mapping ohio lung cancer mortality. *Stat Med*, 17(18):2025–43, 1998. ISSN 0277-6715 (Print) 0277-6715 (Linking). URL `http://www.ncbi.nlm.nih.gov/pubmed/9789912`.

G. Xiao, C. Reilly, and A. B. Khodursky. Improved detection of differentially expressed genes through incorporation of gene location. *Biometrics*, 65:805–814, 2009.

M. J. Yabsley, J. Romines, and V. F. Nettles. Detection of *Babesia* and *Anaplasma* species in rabbits from texas and georgia, usa. *Vector Borne Zoonotic Dis*, 6(1):7–13, 2006. ISSN 1530-3667 (Print) 1530-3667 (Linking). doi:10.1089/vbz.2006.6.7. URL `https://www.ncbi.nlm.nih.gov/pubmed/16584322`.

Keming Yu and Rana A. Moyeed. Bayesian quantile regression. *Statistics Probability Letters*, 54 (4):437 – 447, 2001. ISSN 0167-7152. doi:https://doi.org/10.1016/S0167-7152(01)00124-9. URL `http://www.sciencedirect.com/science/article/pii/S0167715201001249`.

Yu Yue and Paul L. Speckman. Nonstationary spatial Gaussian Markov random fields. *Journal of Computational and Graphical Statistics*, 19(1):96–116, 2010.

Yu Ryan Yue, Paul L. Speckman, and Dongchu Sun. Priors for bayesian adaptive spline smoothing. *Annals of the Institute of Statistical Mathematics*, 64(3):577–613, Jun 2012. doi:10.1007/s10463-010-0321-6. URL `https://doi.org/10.1007/s10463-010-0321-6`.

Hao Zhang and Abdel El-Shaarawi. On spatial skew-Gaussian processes and applications. *Environmetrics*, 21(1):33–47, 2009.

Zhengyu Zhang. Local partitioned quantile regression. *Econometric Theory*, 33(5):10811120, 2017. doi:10.1017/S0266466616000293.

Zoetis. Data on file, study report no. b865r-us-12-018, zoetis inc, 2016. URL `https://www.zoetisus.com/products/dogs/vanguard-crlyme/img/assets/crlyme_efficacy_study.pdf`.