Spatiotemporal Gradient Modeling with Applications

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Harrison S. Quick

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Advised by Bradley P. Carlin and Sudipto Banerjee

July, 2013

Acknowledgments

I owe the greatest debt of gratitude to my advisors, Brad Carlin and Sudipto Banerjee, for guiding me through a field of study that I greatly enjoy. Not only have they pushed me to succeed as a researcher and as an educator, but they have done so as though I were a colleague, not simply a student. Thanks also must go out to my current and former committee members — Dipankar Bandyopadhyay for giving me a different perspective of spatial data analysis, John Hughes for his inspirational enthusiasm as an educator, and Gurumurthy Ramachandran for helping me think beyond the statistical models and remember the importance of their application. I am also grateful to my past collaborators — Kathy Huppler-Hullsiek, Jason Baker, Traci Toomey, and Darin Erickson — for the opportunity to work on a variety of projects over the past 5 years. I must also thank my high school math teacher, David Vanderwerf, for noticing my interest in math and basketball (and my lack of athletic ability) and encouraging me to take statistics for the basketball team, and my undergraduate statistics professor, Mezbahur Rahman, for putting up with me for 7 courses, introducing me to the field of Biostatistics, and for helping me get into a great graduate program. Finally, I must thank my parents for allowing me to make a mess of their condo for the past 4 years and for their endless encouragement.

Dedication

To my parents, Scott and Mary Quick; may a career in Biostatistics help me support you in the life you've grown (or want to grow) accustomed to.

Abstract

Advances in Geographical Information Systems (GIS) have led to enormous recent growth in spatiotemporal databases and associated statistical modeling with applications in various scientific disciplines, including environmental monitoring, ecological systems, forestry, hydrology, meteorology and public health. After inferring on a spatiotemporal process for a given dataset, inferential interest may turn to estimating rates of change, or *gradients*, over space and time. The primary focus of this thesis is to further develop the methodology required for statistical inference on areally-referenced temporal and spatiotemporal gradient processes.

We begin by first departing from the rather rich literature in space-time modeling by considering the setting where space is discrete but time is continuous. Our major objective here is to carry out inference on gradients of a temporal process in our dataset of monthly county level asthma hospitalization rates in the state of California, while also accounting for spatial similarities of the temporal process across neighboring counties. In addition to using a more flexible stochastic process embedded within a dynamic Markov random field framework that permits inference on the temporal gradient process, we also develop methods for allowing region-specific variance components, leading to variable smoothing in our spatial regions.

We then move to the continuous space, continuous time setting. Here, we develop, within a flexible spatiotemporal process model setting, a framework to estimate arbitrary directional gradients over space at any given timepoint, temporal derivatives at any given spatial location and, finally, mixed spatiotemporal gradients that reflect rapid change in spatial gradients over time and vice-versa. After illustrating the use of our methodology on a dataset comprising daily PM2.5 concentrations in California, we show how the method can be implemented to analyze highly censored data (e.g., data below

detectable limits) and apply these methods to data collected during the cleanup efforts of the *Deepwater Horizon* (BP) oil spill.

Through the use of these methods, we believe researchers can gain significant insight into potentially important spatiotemporally varying risk factors that may as of yet be unknown (or at least not accounted for). Furthermore, the gradient process in and of itself can provide valuable information, for instance by being adapted to alert public health officials of dramatically rising pollution levels in a particular region, potentially leading to a reduction in exposure and, ultimately, a reduction in the incidence of poor health outcomes.

Contents

Acknowledgments **i**

Dedication **ii**

Abstract **iii**

List of Tables **viii**

List of Figures **ix**

1 Background **1**

1.1 Overview of spatiotemporal modeling 1

1.2 Modeling of temporal and spatiotemporal gradients 3

1.2.1 Discrete space, continuous time setting 5

1.2.2 Continuous space, continuous time setting 6

1.3 Motivating data sets 7

1.3.1 California asthma hospitalization data 8

1.3.2 California air quality data 9

1.3.3 *Deepwater Horizon* oil spill data 10

2 Modeling temporal gradients in regionally aggregated data **12**

2.1 Areally referenced temporal processes 12

2.2 Hierarchical modeling 16

2.3 Gradient analysis 17

2.4 Simulation studies 19

2.5 Analysis of the California asthma hospitalization data 21
2.6 Discussion 27

3 Heteroscedastic CAR models for areally referenced temporal processes 31
3.1 Dynamic CAR models 32
3.2 The heteroscedastic CAR (HCAR) model 34
3.2.1 Modeling temporal gradients 37
3.2.2 Diagnostic for Determining Spatial Outliers 38
3.3 Simulation studies 39
3.3.1 Estimating $\mathbf{Z}$ and $\mathbf{Z}'$ 39
3.3.2 Parameter Estimation 42
3.4 Analysis of the California asthma hospitalization data 42
3.5 Discussion 45

4 On Bayesian Inference for Rates of Change in Spatiotemporal Process Settings 48
4.1 Calculus of spatiotemporal gradients 49
4.2 Distribution theory for spatiotemporal gradients 51
4.3 Hierarchical modeling and inference 55
4.4 Simulated data example 58
4.5 Analysis of the California air quality data 60
4.6 Discussion 62

5 Analysis of Censored Data Using the Spatiotemporal Gradient Process 68
5.1 Hierarchical modeling and inference 70
5.2 Simulated data example 72
5.3 Analysis of the *Deepwater Horizon* oil spill data 73
5.4 Discussion 76

6 Conclusion 83
6.1 Summary of major findings 83
6.2 Extensions and future work 84

6.2.1 Computing and software development 84
6.2.2 Investigation of relationship between gradients and missing covariates 85
6.2.3 Temporal predictive process models 85

References **87**

Appendix A. Data Preprocessing Algorithm for Asthma Hospitalization Data **95**
A.1 Prior specification 95
A.2 The algorithm 96

Appendix B. Validity of $\mathbf{C}_{\mathbf{W}}(\boldsymbol{\Delta}, \delta)$ **97**

Appendix C. Derivations for $\boldsymbol{\nabla}\mathbf{K}_0$ and $\mathbf{C}_{\nabla Z}(\mathbf{0}, 0)$ **99**

List of Tables

2.1 Comparisons between our areally referenced Gaussian process model and the three alternatives. p_D is a measure of model complexity, as it represents the effective number of parameters. Smaller values of DIC and Dawid-Sebastiani (D-S) scores indicate a better trade-off between in sample model fit and model complexity.
* Both DIC and D-S shown are standardized relative to our areally referenced Gaussian Process model. 23

2.2 Parameter estimates for asthma hospitalization data, where estimates for $\bar{\tau}_\cdot^2$ represent the median (95% CI) of the $\tau_i^2, i = 1, \ldots, N_s = 58$. 24

3.1 Comparison of the CAR, SACAR, and HCAR models for $\mathbf{Z} = (Z_1, \ldots, Z_{N_s})$. Here, A is a diagonal matrix with $A_{ii} = \frac{1}{\sigma_i} \sum_{j \sim i} \frac{1}{\sigma_j}$ and T is a diagonal matrix with $T_{ii} = \sigma_i$. Note that the CAR model is a special case of both the SACAR and HCAR models where $\sigma_i = \sigma$ for all i. 35

3.2 Posterior medians and 95% credible intervals (CI) for $\boldsymbol{\beta}$, ϕ, and α from our asthma hospitalization rate data. 47

4.1 Parameter estimates from the analysis of the PM2.5 data. ϕ_s is on a scale of 1 unit = 100 km, and ϕ_t is on a scale of 1 unit = 1 week. 61

5.1 Parameter estimates from the analysis of the BP data. ϕ_s is on a scale of 1 unit = 5 km, and ϕ_t is on a scale of 1 unit = 5 minutes. 75

List of Figures

1.1 Raw asthma hospitalization rates per 100,000 people, aggregated over year 9
1.2 Monitoring sites observed in the PM2.5 data. 10
1.3 Locations of the 447 observation sites from the *Deepwater Horizon* data set. Red circles denote VOC levels above the limit of detection (0.1ppm), while black circles denote censored observations. Observations from each day are connected using lines, where each day has its own combination of color and line type. 11
2.1 Spatiotemporal random effects and temporal gradients for a region based on one dataset from the second simulation study. Solid black lines denote true sinusoidal curves based on the model in eq. (2.9), while gray bands represent 95% credible intervals. 21
2.2 Spatial random effects for asthma hospitalization data, by year 25
2.3 Comparison between the spatiotemporal random effects in Los Angeles and San Francisco Counties, and an investigation of temporal gradients in Los Angeles County. Point estimates in black and corresponding 95% CI bands in gray. Figures in the top panel illustrate the differences in the temporal trends of the random effects between the two counties. The bottom-left figure displays the temporal gradients computed between months in Los Angeles County, and the bottom-right figure displays the subset of the gradients which are further described in the text. 26

2.4 Posterior predicted curves (and 95% credible bounds) for the daily asthma hospitalization rates in Los Angeles County between November 15, 1995 to January 15, 1996. This county and interval was selected due the presence of a significantly positive gradient between November and December and a significantly negative gradient between December and January. The true hospitalization are also shown for comparison purposes, though the model was fit using only the monthly aggregates. 27
2.5 Temporal gradients for transition from August to September over time . 28
3.1 Posterior bands for $\mathbf{Z}_i$ in Alameda (left) and Lassen (right) Counties plotted against the true value (black line) and mean of neighbors (red line) in our simulation study. 33
3.2 95% CI (gray bands) of $\mathbf{Z}$ for Alameda (left) and Lassen (right) Counties from the simulation study. Solid black lines denote the true underlying curves and dashed red lines denote the mean of each county's neighboring regions. Compared to the fit in Figure 3.1, Alameda County has vastly better fit, and Lassen County has slightly narrower credible bands. . . . 40
3.3 The first two panels display estimated variance parameters, $\widehat{\sigma}_i$, from the simulation studies. Simulation study #1 is based on the sinusoidal curves constructed in (3.3), and simulation study #2 is constructed from our model, (3.5), using the parameter estimates from the first study. Color cutoffs are (3,4,5,6,7,8), ranging from dark blue to dark red. The third panel displays the estimated $Q_i(\mathbf{Z}_i)$ from the first simulation study, where the color cutoffs used are (50,60,70,80,90,100). The fourth panel is a reference for identifying key counties mentioned in the text. Here, Alameda County is shaded black, Imperial County is shaded blue, Lassen County is shaded red, and Orange County is shaded orange. 41
3.4 The first two panels display the estimated variance parameters from the asthma hospitalization data. Color cutoffs are (1,2,3,4,5,6) for σ_i and τ_i, ranging from dark blue to dark red. The third panel displays our estimated outlier diagnostic, Q_i, with color cutoffs (175,200,225,250,275,300). 43
3.5 Average month-to-month temporal gradients from the California asthma hospitalization data. 45

4.1 Comparison of the true temporal and the s_{i2} mixed gradients for a particular time point and their posterior median values based on our gradient theory. The color scheme here goes from highly negative (blue) to highly positive (red), centered around 0 (light gray), with locations of observed locations plotted as open circles. The third panel highlights significant gradients; i.e., gradients whose 95% CI's contain 0 are assigned value 0, while gradients whose 95% CI's do not contain 0 are assigned their posterior medians. 65

4.2 Predicted levels of PM2.5 for the month of June 2009. The maps are situated so that the first column of days are on Sunday, the second are Mondays, etc. 66

4.3 Estimated spatial and temporal gradient plots for the transition period between June 2 and June 3, as well as a topographical map of California for comparison purposes. Spatial gradients in the southern portion of the state appear to correspond to the Transverse Ranges, which wrap around north and east of Los Angeles. We are also able to identify the Coast Ranges, which appear to cause negative eastern gradients along the Pacific Coast. The temporal gradients display the sharp increase in the Los Angeles area seen in Figure 4.2. 67

5.1 Underlying process for the data generated by (5.7). Values range from -10 (dark blue) to 10 (dark red), where here any shade of blue denotes a negative, and thus censored, value. 78

5.2 Comparison of the true temporal and the s_{i2} mixed gradients for time point 3 (see Figure 5.1) and their posterior median values based on our gradient theory. The color scheme here goes from highly negative (blue) to highly positive (red), centered around 0 (light gray), with locations of observed locations plotted as open circles. The third panel highlights significant gradients; i.e., gradients whose 95% CI's contain 0 are assigned value 0, while gradients whose 95% CI's do not contain 0 are assigned their posterior medians. 79

5.3 Predicted VOC surface on June 26th for sixteen different time points, where blue and white shading denote levels below and near the limit of detection ($\log(0.1) = -2.3$), respectively, and darker shades of red indicate higher levels above detectable limits (up to 2 on the log-scale, or exp(2)=7.39ppm). Here, circles denote observations within 30 minutes of the labeled times, where black circles denote censored observations and red denotes observations above 0.1ppm. Green plus-sign (+) denotes the site of the oil rig. 80

5.4 Predicted VOC surface on June 26 for four different time points, where white shading denotes levels near the limit of detection ($\log(0.1) = -2.3$) and darker shades of red indicate higher levels above detectable limits (up to 2 on the log-scale, or exp(2)=7.39ppm). Also plotted is a black line denoting the path of the ship collecting measurements at intervals of 12 minutes from 9:00AM to 6:30PM. Here, circles denote observations within 30 minutes of the labeled times, where black circles denote censored observations and red denotes observations above 0.1ppm. Green plus-sign (+) denotes the site of the oil rig. For a sense of scale, the distance between the observations in the fourth panel are roughly 4km away from the site of the spill. 81

5.5 Estimated spatial and temporal gradients for 1:00PM on June 26 for the BP oil spill data. Negative predicted gradients are shaded in blue and positive gradients in red, with white shading corresponding to values near zero. Also shown here are the path of the ship carrying workers (black line), the location of uncensored and censored observations occurring between 12:30PM and 1:30PM (denoted by red and black circles, respectively), and the site of the *Deepwater Horizon* oil rig (green plus-sign). 82

Chapter 1

Background

Technological advances in spatially-enabled sensor networks, and geospatial information storage, analysis, and distribution systems have led to a burgeoning of spatial-temporal databases. Accounting for associations across space and time constitutes a routine component in analyzing geographically and temporally referenced datasets. The inference garnered through these analyses often supports decisions with important scientific implications, and it is therefore critical to accurately assess inferential uncertainty. The obstacle for researchers is increasingly not access to the right data, but rather implementing appropriate statistical methods and software.

1.1 Overview of spatiotemporal modeling

There is by now a considerable literature in spatiotemporal modeling; see, for example, the recent books by Banerjee, Carlin, and Gelfand (2004) and Cressie and Wikle (2011) and references therein for overviews. Space-time modeling can broadly be classified as considering one of the following four settings: (a) space is viewed as continuous, but time is taken to be discrete, (b) space and time are both continuous, (c) space and time are both discrete, and (d) space is viewed as discrete, but time is taken to be continuous. Almost exclusively, the existing literature considers the first three settings. Perhaps the most pervasive case is the first. Here, the data are regarded as a time series of spatial process realizations. Early approaches include the STARMA (Pfeifer and Deutsch, 1980, a, b) and STARMAX (Stoffer, 1986) models, which add spatial covariance structure to

standard time series models. Handcock and Wallis (1994) employ stationary Gaussian process models with an AR(1) model for the time series at each location to study global warming. Building upon previous work in the setting of dynamic models by West and Harrison (1997), several authors, including Stroud et al. (2001) and Gelfand et al. (2005), proposed dynamic frameworks to model residual spatial and temporal dependence.

When space and time are both viewed as continuous, the preferred approach is to construct stochastic processes using space-time covariance functions. Gneiting (2002) built upon earlier work by Cressie and Huang (1999) to propose general classes of nonseparable, stationary covariance functions that allow for space-time interaction terms for spatiotemporal random processes. Stein (2005) considered a variety of properties of space-time covariance functions and how these were related to process spatial-temporal interactions.

In settings where both space and time are discrete there has been much spatiotemporal modeling based on a Markov random field (MRF) structure in the form of conditionally autoregressive (CAR) specifications. See, for example, Waller et al. (1997), who developed such models in the service of disease mapping and Gelfand et al. (1998), whose interest was in single family home sales. Pace et al. (2000) work with simultaneous autoregressive (SAR) models extending them to allow temporal neighbors as well as spatial neighbors. Later examples include the space-time interaction CAR model proposed by Schmid and Held (2004), the dynamic CAR model proposed by Martinez-Beneito et al. (2008), the proper Gaussian MRF process models of Vivar and Ferreira (2009), and the latent structure models approach from Lawson et al. (2010).

Finally, there is the setting where space is discrete and time is continuous. This can be envisioned when, for instance, we have a collection of N_s functions of time over N_s regions, but the functions are posited to be spatially associated. That is, functions arising from neighboring regions are believed to resemble each other. The functional data analysis literature (Ramsay and Silverman, 1997, and references therein) deals almost exclusively with kernel smoothers and roughness-penalty type (spline) models; recent discrete-space, continuous time examples using spline-based methods include the works by MacNab and Gustafson (2007) and Ugarte et al. (2010). Baladandayuthapani et al. (2008) consider spatially correlated functional data modeling for point-referenced data by treating space as continuous. A recent review by Delicado et al. (2010) reveals that

spatially associated functional modeling of time has received little attention, especially for regionally aggregated data.

Here we will focus on the two continuous time settings, the benefits of which are twofold. First and foremost, investigators (or, in our case, public health officials) may desire understanding of the local effects of temporal impact at a resolution finer than that at which the data were sampled. For instance, despite collecting data monthly, there may be interest in making inference on a particular week or even a given day of that month. While there is a wealth of literature in this domain, dynamic space-time models that treat time discretely can offer statistically legitimate inference only at the level the data were observed. Second, the modeling also allows us to subsequently carry out inference on temporal (and in the continuous space setting, spatiotemporal) *gradients*; that is, the rate of change of the underlying process over time.

The remainder of this thesis proceeds as follows: In Section 1.2, we provide a background on the concept of temporal and spatiotemporal gradient processes, as well as the importance of investigating the gradient process. Next, we introduce the three data sets to be analyzed by our gradient methods in Section 1.3. Chapter 2 proposes our methodology for statistical inference on the temporal gradient process in regionally aggregated spatial data, and Chapter 3 pertains to an extension of those methods aimed at handling outlying spatial regions. In Chapter 4, we discuss our methodology for investigating the spatiotemporal gradient process in the continuous space, continuous time setting, followed by an illustration of the use of these methods in highly censored data in Chapter 5. Finally, we conclude in Chapter 6, summarizing our findings and presenting suggestions for future work.

1.2 Modeling of temporal and spatiotemporal gradients

In this thesis, our goal is to detect spatiotemporal changes in the residuals that remain after accounting for important covariates; significant changes may correspond to changes in spatiotemporal covariates still missing from our model. While the residuals themselves could be beneficial in detecting missing covariates, spatiotemporal gradients can be more useful in detecting covariates that operate on much finer scales. For example, time points with significantly high residual gradients are likely to point toward

missing covariates whose rapid changes on a finer scale impact the outcome. On the other hand, the residual process estimated from discrete time models is likely to smooth over any patterns arising from such local behavior of covariates. By modeling the gradient process itself, we can also identify *statistically significant* gradients in space and/or time; identifying these significant gradients can help facilitate the search for missing covariates by narrowing the focus to a particular spatial region or time period.

Such rapid changes over space and/or time often arise in many natural phenomena. Examples include geographic features such as mountain ranges causing sharp spatial changes in precipitation levels, the degree of traffic congestion increasing abruptly at rush hour, and sudden changes in land values in urban areas, as discussed by Majumdur et al. (2006). These rapid or abrupt changes are often not manifested as discontinuities in the residuals, but rather as high rates of change. Depending upon the context, outliers over space and time may be defined as $(\mathbf{s}, t)$'s associated with high rates of change, where $\mathbf{s}$ and t denote a spatial location and time point, respectively.

Unfortunately, the factors underlying sudden changes in observed spatiotemporal data are often unknown, and not easily accounted for in spatiotemporal regression models. Therefore, estimating gradients using the smoothness of the process becomes relevant and is often referred to as *wombling*, named after a seminal paper by Womble (1951), and has been explored in a variety of purely spatial contexts by Barbujani et al. (1989), Fortin (1994, 1997), Fortin and Drapeau (1995), Banerjee et al. (2003), Banerjee and Gelfand (2006) and Gabriel et al. (2011); see Banerjee (2010) for an overview and references therein for further details. Estimation of rates of change are also becoming increasingly relevant in computer/physical modeling involving differential equations. Examples include earlier work by O'Hagan (1992) and Morris et al. (1993), who employ derivative information to estimate a purely spatial process, and, more recently, Manolopoulou et al. (2013) explore directional gradient fields in analyzing spatio-dynamic physical models. Our proposed framework will assist in such inference by offering formal assessments, including statistical significance, for space-time gradients.

In addition to being used to identify missing covariates, these methods may also be used to alert public health officials of impending crises in real-time. Motivated by the need for timely detection of bioterror outbreaks in the modern world, Reis et al.

(2003) develop and discuss methods for detecting sudden changes in phenomena such as influenza hospitalizations or sales of over-the-counter medications. They argue that by monitoring coarsely aggregated data (say, counts over a period of 3 days instead of daily counts), one could reduce the effect of noise and improve both the sensitivity and specificity of their surveillance system. Later work by Mandl et al. (2004) discusses the importance of accounting for spatial clustering (among other things) into the analysis as a means of detecting changes in a particular region. The spatiotemporal gradient methodology developed here could be used to detect changes in real time, and thus inform rapid-response public health alerts (say, advising members of vulnerable populations to remain indoors in certain areas about to be impacted by a wave of unhealthful air).

1.2.1 Discrete space, continuous time setting

As the area of spatial gradients has been explored in the aforementioned references, we first extend the existing theory to temporal gradients in areally-referenced temporal data. To this end, we propose a rich class of Bayesian space-time models based upon a dynamic MRF that evolves continuously over time. This accommodates spatial processes that are posited to be spatially indexed over a geographical map with a well-defined system of neighbors. This continuous temporal evolution sets our current article apart from the existing literature. Rather than modeling time using simple parametric forms, as is often done in longitudinal contexts, we employ a stochastic process, enhancing the model's adaptability to the data. The smoothness implications for the underlying process in this context are obvious. In Chapter 2, we deploy a mean square differentiable Gaussian process that provides a tractable gradient process to help us achieve these inferential goals.

How we model dependencies, therefore, is critical for practical implementation of space-time models. By using a simple *separable* model that separates the spatial and temporal associations, one can achieve computational gains from the resulting Kronecker product form of the space-time covariance matrix. This structure, however, limits the model to only one variance parameter for controlling both spatial and temporal variability. Unfortunately, if the true underlying process is less smooth in some regions, or if there are spatial outliers, such a model may simultaneously oversmooth

and undersmooth the space-time effects. Chapter 3 extends our methods to allow for region-specific variance (or, alternatively, *scale*) parameters which allow the spatiotemporal process additional flexibility in outlying regions, providing yet another tool for identifying potential missing covariates.

1.2.2 Continuous space, continuous time setting

After developing the methods for the discrete space, continuous time setting, we will move on to the realm of continuous space, continuous time data. Here we typically encounter stochastic process based models, which are being widely deployed for analyzing spatiotemporal datasets in various scientific disciplines. Such models assume a collection of real-valued random variables $\{Z(\mathbf{s},t) : (\mathbf{s},t) \in \Re^d \times \Re\}$, where $\mathbf{s}$ denotes spatial coordinates residing in $\Re^d$, the d-dimensional Euclidean space (usually $d = 2$ or 3 in spatial statistics but may be larger in other Gaussian process modeling domains such as functional analysis), and t is time with domain $\Re$, the real line. The pair $(\mathbf{s},t)$ is often referred to as a *space-time coordinate*. Depending upon the specific context, the variable $Z(\mathbf{s},t)$ can represent a scientific outcome such as an environmental pollutant or, perhaps, it can be used as a latent variable that generates spatiotemporal random effects to better capture variability in the response. These models treat both space and time as continuous and are usually applied when interpolation over both space and time is desired at arbitrary resolutions.

Spatiotemporal modeling has enjoyed much attention over last two decades; see, for example, the book by Cressie and Wikle (2011) and the references therein. When space and time are both continuous, the preferred approach is to specify a space-time covariance function for the spatiotemporal process $Z(\mathbf{s},t)$, which in turn will ensure a valid joint distribution for the partial realizations of the process over a finite collection of locations and time points. Christakos (1992, 2000), Jones and Zhang (1997), Cressie and Huang (1999), Brown et al. (2000), de Iaco et al. (2002), Gneiting (2002), Ma (2003) and Hartfield and Gunst (2003) all offer classes of stationary space-time covariance functions. Subsequently, Stein (2005) elucidated several properties of space-time covariance functions and how the smoothness of the realizations can impact inference on spatiotemporal interactions.

In Chapter 4, we seek to address fully model-based inference on spatiotemporal gradients under continuous space, continuous time settings. Our contribution is to offer, within a flexible spatiotemporal process model setting, a framework to estimate arbitrary directional gradients over space at any given timepoint, temporal derivatives at any given spatial location and, finally, *mixed* spatiotemporal gradients that reflect rapid change in temporal gradients over space and vice-versa. We seek to achieve such inference without compromising on rich and flexible spatiotemporal process models. In the case of nonseparable stationary spatiotemporal Gaussian processes, we offer distributional details for the aforementioned gradients, and demonstrate how to implement full sampling-based inference within a Bayesian framework. This enables exact inference on gradients and functions thereof (e.g., the maximum spatial gradient over all directions) in posterior predictive fashion by supplying an entire posterior distribution for the gradients at arbitrary spatial locations and timepoints *after* posterior distributions for all the model parameters have been obtained. The Bayesian framework also allows seamless inference on gradients associated with the residual process, as we can recover its full posterior distribution although no true residuals can be observed. To the best of our knowledge, no such treatment of inference on spatiotemporal temporal gradients has been undertaken. Finally, in Chapter 5, we extend these methods to the case where we observed censored data, which is essential for an analysis of one of our data sets below (see Section 1.3.3).

1.3 Motivating data sets

In this chapter, we introduce the data sets which motivated much of this work. First, in Section 1.3.1, we discuss the California asthma hospitalization data, a data set which includes regionally aggregated observations collected over many time points. In Section 1.3.2, we describe the California air quality data, which was collected over time at a number of fixed monitoring sites throughout the state of California. Finally, in Section 1.3.3, we describe data collected during the cleanup efforts resulting from the *Deepwater Horizon* (BP) oil spill; due to the nature of this data, measurements are observed at a variety of unique spatial locations and times.

1.3.1 California asthma hospitalization data

Our first dataset consists of asthma hospitalization rates in the state of California. According to the California Department of Health Services (2003), millions of residents of California suffer from asthma or asthma-like symptoms. As many studies have indicated (e.g. English et al. 1998), asthma rates are related to, among other things, pollution levels and socioeconomic status (SES)—two variables that likely induce a spatiotemporal distribution on such rates. Weather and climate also likely play a role, as cold air can trigger asthma symptoms.

The data were collected daily from 1991 to 2008 from each of the state's 58 counties. We consider all hospital discharges where asthma was the primary diagnosis, which are categorized as extrinsic (allergic), intrinsic (non-allergic), or other. Due to confidentiality, data for days with between one and four hospitalizations of a specific category are missing; this affected 38% of our observations, including more than 50% of those from 21 counties. To remedy this, county-specific values for these days are imputed using a method similar to Besag's iterated conditional modes method (Besag, 1986); see Appendix A for details. For our analysis, the data are aggregated by month, for a total of 216 observations per county over the 18-year period, and then *rates* per 100,000 residents are computed; the conversion from counts to rates for the purpose of fitting Gaussian spatiotemporal models is common in the literature (see, for instance, Short et al., 2002). While the vast majority of rates are less than 20 hospitalizations per month per 100,000 people, the range of the rates extends from 0 to 90. As can be seen in Figure 1.1, hospitalization for asthma demonstrates a statewide decreasing trend early in the study period, and appears to stabilize in later years. Here, we map the raw *annual* (summed over month) hospitalization rates, which have values between 0 and 340 hospitalizations per 100,000.

We attempt to capture the effect of socioeconomic status by including population density in our model, using data from the 2000 U.S. Census and land area measurements from the National Association of Counties. To account for pollution, we use data from the Air Resources Board of the California Environmental Protection Agency which counts the number of days in each month with average ozone levels above 0.07 ppm over 8 consecutive hours, the state standard. Because our ozone data is compiled at the *air basin* level, county-specific values are calculated by taking the maximum value

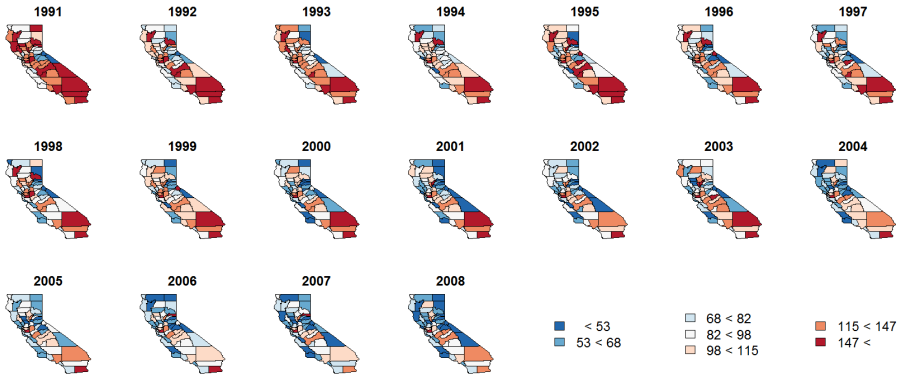


Figure 1.1: Raw asthma hospitalization rates per 100,000 people, aggregated over year

of all air basins that the county belonged to. Generally, ozone levels are highest during the summer months, with the highest values in southern California and the Central Valley region, and show little variation between years. As hospitalization rates are higher among youth and the black population, county-level covariates for percent under 18 and percent black are also included. These demographic covariates both have their highest values in southern California, though counties in the Central Valley region also have larger black populations.

1.3.2 California air quality data

Our second data set consists of concentrations of particulate matter less than 2.5 micrometers in diameter (PM2.5) measured at monitoring sites throughout the state of California, as collected by the Air Resources Board of the California Environmental Protection Agency. These data were typically observed once every three days at each site, with observations staggered such that a subset of sites are observed on, say, Day 1, while another set of sites are observed on Day 2, etc. In addition to the geographic location of each monitoring site (in latitude and longitude) and the date of the observation, the data set also contains each site's elevation. Merely for Markov chain Monte Carlo (MCMC) convergence purposes, we have centered and scaled this elevation covariate to have mean 0 and variance 1. While the full data set contains observations from 1998 to 2009 from a large number of locations, due to computational costs (e.g.,

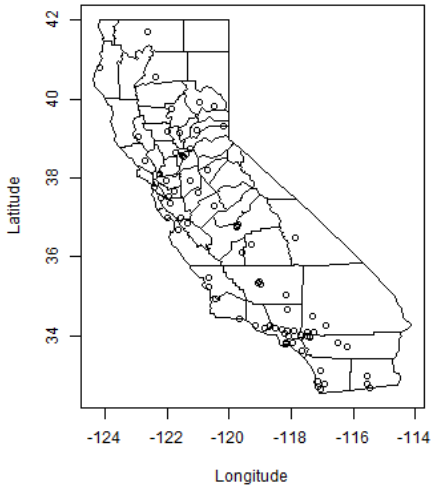


Figure 1.2: Monitoring sites observed in the PM2.5 data.

computing time, memory allocation, etc.), we restrict our present analysis to data from the month of June 2009. As such, our data consist of 78 locations and a total of 773 observations. Figure 1.2 displays the locations from which our data were observed. We use the sinusoidal projection $s_{i1} = R\lambda\cos\theta$ and $s_{i2} = R\theta$, where $R = 6371$km is the radius of the earth, and θ and λ denote latitude and longitude, respectively. Distances were computed by letting one spatial unit equal 100 km, while temporal distances were computed by letting one temporal unit equal one week.

1.3.3 *Deepwater Horizon* oil spill data

Our third and final data set is comprised of levels of volatile organic compounds (VOC) collected during the cleanup efforts resulting from the explosion of the *Deepwater Horizon* oil rig on April 20, 2010 and the subsequent oil spill. Occurring roughly 41 miles off the coast of Louisiana, these 447 observations were collected during the period from June 20 to July 20 by ships carrying workers. Measurements were taken, often at regular intervals (say, once every 12 minutes), while cleanup efforts were conducted over a period of a few hours. As illustrated in Figure 1.3, we may observe multiple boats in different locations of a spatial region over the course of a day — as such, our data is irregular in both space and time, thus requiring a continuous space, continuous time

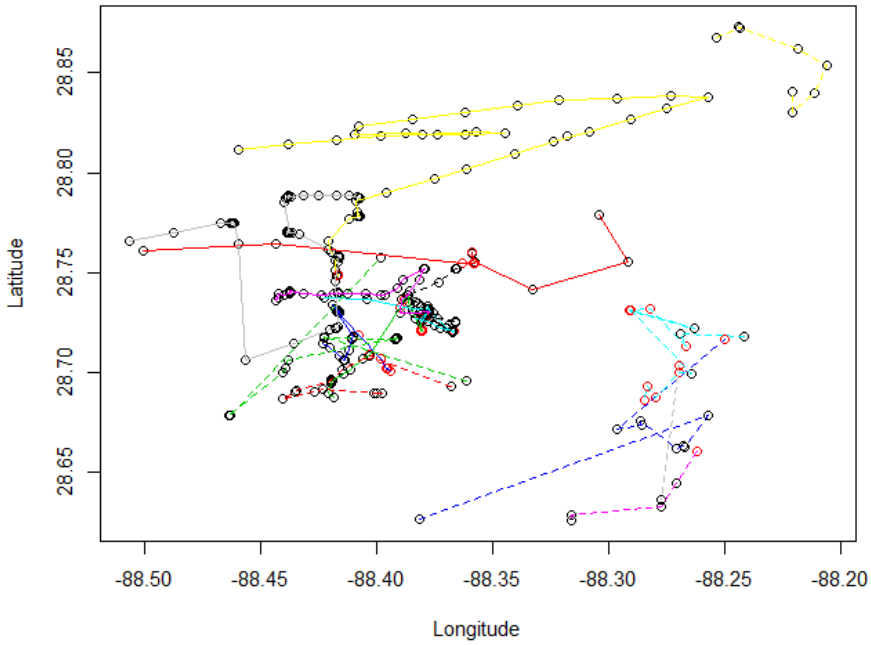


Figure 1.3: Locations of the 447 observation sites from the *Deepwater Horizon* data set. Red circles denote VOC levels above the limit of detection (0.1ppm), while black circles denote censored observations. Observations from each day are connected using lines, where each day has its own combination of color and line type.

framework. Furthermore, over 83% of our VOC observations are below the detectable limit of 0.1ppm (parts per million), thus we require techniques for handling censoring in our analysis.

Chapter 2

Modeling temporal gradients in regionally aggregated data

This chapter begins in Section 2.1 with an outline of a class of dynamic MRFs indexed continuously over time. Section 2.2 then provides details on the Bayesian hierarchical models that emerge from our rich space-time structures, while Section 2.3 derives the posterior predictive inferential procedure for temporal gradient process, verified via simulation in Section 2.4. Section 2.5 describes the detailed analysis of our dataset, while Section 2.6 summarizes and concludes.

2.1 Areally referenced temporal processes

As mentioned in Chapter 1, our methodological contribution in this chapter is a modeling framework for areally referenced outcomes that, it can be reasonably assumed, arise from an underlying stochastic process continuous over time. To be specific, consider a map of a geographical region comprising N_s regions that are delineated by well-defined boundaries, and let $Y_i(t)$ be the outcome arising from region i at time t. For every region i, we believe that $Y_i(t)$ exists, at least conceptually, at every time point. However, the observations are collected not continuously but at discrete time points, say $\mathcal{T} = \{t_1, t_2, \ldots, t_{N_t}\}$. For the time being, we will assume that the data comes from the same set of time points in $\mathcal{T}$ for each region. This is not necessary for the ensuing development, but will facilitate the notation.

A spatial random effects model for our data assumes

$$Y_i(t) = \mu_i(t) + Z_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{ind}{\sim} N(0, \tau_i^2) \text{ for } i = 1, 2, \ldots, N_s \,, \tag{2.1}$$

where $\mu_i(t)$ captures large scale variation or trends, for example using a regression model, and $Z_i(t)$ is an underlying areally-referenced stochastic process over time that captures smaller-scale variations in the time scale while also accommodating spatial associations. Each region also has its own variance component, τ_i^2, which captures residual variation not captured by the other components.

The process $Z_i(t)$ specifies the probability distribution of correlated space-time random effects while treating space as discrete and time as continuous. We seek a specification that will allow temporal processes from neighboring regions to be more alike than from non-neighbors. As regards spatial associations, we will respect the discreteness inherent in the aggregated outcome. Rather than model an underlying response surface continuously over the region of interest, we want to treat the $Z_i(t)$'s as functions of time that are smoothed across neighbors.

The neighborhood structure arises from a discrete topology comprising a list of neighbors for each region. This is described using an $N_s \times N_s$ adjacency matrix $W = \{w_{ij}\}$, where $w_{ij} = 0$ if regions i and j are not neighbors and $w_{ij} = c \neq 0$ when regions i and j are neighbors, denoted by $i \sim j$. By convention, the diagonal elements of W are all zero. To account for spatial association in the $Z_i(t)$'s, a temporally evolving MRF for the areal units at any arbitrary time point t specifies the full conditional distribution for $Z_i(t)$ as depending only upon the neighbors of region i,

$$p(Z_i(t) \,|\, \{Z_{j \neq i}(t)\}) \sim N\left(\sum_{j \sim i} \alpha \frac{w_{ij}}{w_{i+}} Z_j(t), \frac{\sigma^2}{w_{i+}}\right), \tag{2.2}$$

where $w_{i+} = \sum_{j \sim i} w_{ij}$, $\sigma^2 > 0$, and α is a propriety parameter described below. This means that the $N_s \times 1$ vector $\mathbf{Z}(t) = (Z_1(t), Z_2(t), \ldots, Z_{N_s}(t))^T$ follows a multivariate normal distribution with zero mean and a precision matrix $\frac{1}{\sigma^2}(D - \alpha W)$, where D is a diagonal matrix with w_{i+} as its i-th diagonal elements. The precision matrix is invertible as long as $\alpha \in (1/\lambda_{(1)}, 1/\lambda_{(n)})$, where $\lambda_{(1)}$ (which can be shown to be negative) and $\lambda_{(n)}$ (which can be shown to be 1) are the smallest (i.e., most negative) and largest eigenvalues of $D^{-1/2} W D^{-1/2}$, respectively, and this yields a proper distribution for $\mathbf{Z}(t)$ at each timepoint t.

The MRF in (2.2) does not allow temporal dependence; the $\mathbf{Z}(t)$'s are independently and identically distributed as $N\left(\mathbf{0}, \sigma^2(D-\alpha W)^{-1}\right)$. We could allow time-varying parameters σ_t^2 and α_t so that $\mathbf{Z}(t) \overset{ind}{\sim} N\left(\mathbf{0}, \sigma_t^2(D-\alpha_t W)^{-1}\right)$ for every t. If time were treated discretely, then we could envision dynamic autoregressive priors for these time-varying parameters, or some transformations thereof. However, there are two reasons why we do not pursue this further. First, we do not consider time as discrete because that would preclude inference on temporal gradients, which, as we have mentioned, is our primary objective here. Second, time-varying hyperparameters, especially the α_t's, in MRF models are usually weakly identified by the data; they permit very little prior-to-posterior learning and often lead to over-parametrized models that impair predictive performance over time.

Here we prefer to jointly build spatial-temporal associations into the model using a multivariate process specification for $\mathbf{Z}(t)$. A highly flexible and computationally tractable option is to assume that $\mathbf{Z}(t)$ is a zero-centered multivariate Gaussian process, $GP(\mathbf{0}, K_Z(\cdot,\cdot))$, where the matrix-valued covariance function (e.g., "*cross-covariance* matrix function", Cressie, 1993) $K_Z(t,u) = \text{cov}\{\mathbf{Z}(t), \mathbf{Z}(u)\}$ is defined to be the $N_s \times N_s$ matrix with (i,j)-th entry $\text{cov}\{Z_i(t), Z_j(u)\}$ for any $(t,u) \in \Re^+ \times \Re^+$. Thus, for any two positive real numbers t and u, $K_Z(t,u)$ is an $N_s \times N_s$ matrix with (i,j)-th element given by the covariance between $Z_i(t)$ and $Z_j(u)$. These multivariate processes are *stationary* when the covariances are functions of the separation between the time-points, in which case we write $K_Z(t,u) = K_Z(\Delta)$, and *fully symmetric* when $K_Z(t,u) = K_Z(|\Delta|)$, where $\Delta = t-u$. For a detailed exposition on covariance functions, see Chapter 7 of Banerjee et al. (2003); Gelfand and Banerjee (2010) and Gneiting and Guttorp (2010) also provide overviews for continuous settings.

To ensure valid joint distributions for process realizations, we use a constructive approach similar to that used in *linear models of coregionalization* (LMC) and, more generally, belonging to the class of multivariate latent process models (see Section 7.2 of Banerjee et al. 2003). We assume that $\mathbf{Z}(t)$ arises as a (possibly temporally-varying) linear transformation $\mathbf{Z}(t) = A(t)\mathbf{v}(t)$ of a simpler process $\mathbf{v}(t) = (v_1(t), v_2(t), \ldots, v_{N_s}(t))^T$, where the $v_i(t)$'s are univariate temporal processes, independent of each other, and with unit variances. This differs from the conventional LMC approach based on *spatial* processes, which treats space as continuous. The matrix-valued covariance function for

$\mathbf{v}(t)$, say $K_{\mathbf{v}}(t,u)$, thus has a simple diagonal form and $K_Z(t,u) = A(t)K_{\mathbf{v}}(t,u)A(u)^T$. The dispersion matrix for $\mathbf{Z}$ is $\Sigma_Z = \mathcal{A}\Sigma_{\mathbf{v}}\mathcal{A}^T$, where $\mathcal{A}$ is a block-diagonal matrix with $A(t_j)$'s as blocks, and $\Sigma_{\mathbf{v}}$ is the dispersion matrix constructed from $K_{\mathbf{v}}(t,u)$. Constructing simple valid matrix-valued covariance functions for $\mathbf{v}(t)$ automatically ensures valid probability models for $\mathbf{Z}(t)$. Also note that for $t = u$, $K_{\mathbf{v}}(t,t)$ is the identity matrix so that $K_Z(t,t) = A(t)A(t)^T$ and $A(t)$ is a square-root (e.g. obtained from the triangular Cholesky factorization) of the matrix-valued covariance function at time t.

The above framework subsumes several simpler and more intuitive specifications. One particular specification that we pursue here assumes that each $v_i(t)$ follows a stationary Gaussian Process $GP(0, \rho(\cdot,\cdot;\boldsymbol{\phi}))$, where $\rho(\cdot,\cdot;\phi)$ is a positive definite correlation function parametrized by $\boldsymbol{\phi}$ (e.g. Stein, 1999), so that $\mathrm{cov}(v_i(t), v_i(u)) = \rho(t,u;\phi)$ for every $i = 1, 2, \ldots, N_s$ for all non-negative real numbers t and u. Since the $v_i(t)$ are independent across i, we have that $\mathrm{cov}\{v_i(t), v_j(u)\} = 0$ for $i \neq j$.

The matrix-valued covariance function for $\mathbf{Z}(t)$ is $K_Z(t,u) = \rho(t,u;\boldsymbol{\phi})A(t)A(u)^T$. If we further assume that $A(t) = A$ is constant over time, then the process $\mathbf{Z}(t)$ is stationary if and only if $\mathbf{v}(t)$ is stationary. Further, we obtain a *separable* specification, so that $K_Z(t,u) = \rho(t,u;\boldsymbol{\phi})AA^T$. Letting A be some square-root (e.g. Cholesky) of the $N_s \times N_s$ dispersion matrix $\sigma^2(D-\alpha W)^{-1}$ and $R(\boldsymbol{\phi})$ be the $N_t \times N_t$ temporal correlation matrix having (i,j)-th element $\rho(t_i,t_j;\boldsymbol{\phi})$ yields

$$K_Z(t,u) = \sigma^2\rho(t,u;\boldsymbol{\phi})(D-\alpha W)^{-1} \text{ and } \Sigma_Z = R(\boldsymbol{\phi}) \otimes \sigma^2(D-\alpha W)^{-1}\,. \tag{2.3}$$

It is straightforward to show that the marginal distribution from this constructive approach for each $\mathbf{Z}(t_i)$ is $N\left(\mathbf{0}, \sigma^2(D-\alpha W)^{-1}\right)$, the same marginal distribution as the temporally independent MRF specification in (2.2). Therefore, our constructive approach ensures a valid space-time process, where associations in space are modeled discretely using a MRF, and those in time through a continuous Gaussian process.

This separable specification is easily interpretable as it factorizes the dispersion into a spatial association component (areal) and a temporal component. Another significant practical advantage is its computational feasibility. Estimating more general space-time models usually entails matrix factorizations with $O(N_s^3N_t^3)$ computational complexity. The separable specification allows us to reduce this complexity substantially by avoiding factorizations of $N_sN_t \times N_sN_t$ matrices. One could design algorithms to work

with matrices whose dimension is the smaller of N_s and N_t, thereby accruing massive computational gains.

2.2 Hierarchical modeling

In this section, we build a hierarchical modeling framework to analyze the data in Section 1.3.1 using the likelihood from our spatial random effects model in (2.1) and the distributions emerging from the temporal Gaussian process discussed in Section 2.1. The mean $\mu_i(t)$ in (2.1) is often indexed by a parameter vector $\boldsymbol{\beta}$, for example a linear regression with regressors indexed by space and time so that $\mu_i(t;\boldsymbol{\beta}) = \mathbf{x}_i(t)^T\boldsymbol{\beta}$.

The posterior distributions we seek can be expressed as

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{Z} \,|\, \mathbf{Y}) \propto p(\boldsymbol{\phi}) \times IG(\sigma^2 \,|\, a_\sigma, b_\sigma) \times \left(\prod_{i=1}^{M} IG(\tau_i^2 \,|\, a_\tau, b_\tau) \right) \times N(\boldsymbol{\beta} \,|\, \mu_\beta, \Sigma_\beta) \\ \times Beta(\alpha \,|\, a_\alpha, b_\alpha) \times N(\mathbf{Z} \,|\, \mathbf{0}, R(\phi) \otimes \sigma^2(D - \alpha W)^{-1}) \\ \times \prod_{j=1}^{N_t} \prod_{i=1}^{N_s} N(Y_i(t_j) \,|\, \mathbf{x}_i(t_j)^T\boldsymbol{\beta} + Z_i(t_j), \tau_i^2) \,, \end{aligned} \tag{2.4}$$

where $\boldsymbol{\theta} = \{\boldsymbol{\phi}, \alpha, \sigma^2, \boldsymbol{\beta}, \tau_1^2, \tau_2^2, \ldots, \tau_{N_s}^2\}$ and $\mathbf{Y}$ is the vector of observed outcomes defined analogous to $\mathbf{Z}$. The parametrizations for the standard densities are as in Carlin and Louis (2009). We assume all the other hyperparameters in (2.4) are known.

Recall the separable matrix-valued covariance function in (2.3). The correlation function $\rho(\cdot;\boldsymbol{\phi})$ determines process smoothness and we choose it to be a fully symmetric Matérn correlation function given by

$$\rho(t,u;\boldsymbol{\phi}) = \rho(\Delta;\boldsymbol{\phi}) = \frac{1}{\Gamma(\phi_2)2^{\phi_2-1}} \left(2\sqrt{\phi_2}|\Delta|\phi_1\right)^{\phi_2} \mathcal{K}_{\phi_2}\left(2\sqrt{\phi_2}|\Delta|\phi_1\right), \tag{2.5}$$

where $\boldsymbol{\phi} = \{\phi_1, \phi_2\}$, $\Delta = t-u$, $\Gamma(\cdot)$ is the Gamma function, $\mathcal{K}_{\phi_2}(\cdot)$ is the modified Bessel function of the second kind, and ϕ_1 and ϕ_2 are non-negative parameters representing rate of decay in temporal association and smoothness of the underlying process, respectively.

We use MCMC to evaluate the joint posterior in (2.4), using Metropolis steps for updating $\boldsymbol{\phi}$ and Gibbs steps for all other parameters. Sampling-based Bayesian inference seamlessly delivers inference on the residual spatial effects: if t_0 is an arbitrary unobserved timepoint, then, for any region i, we sample from the posterior predictive

distribution $p(Z_i(t_0) \,|\, \mathbf{Y}) = \int p(Z_i(t_0) \,|\, \mathbf{Z}, \boldsymbol{\theta})\ p(\boldsymbol{\theta}, \mathbf{Z} \,|\, \mathbf{Y}) d\boldsymbol{\theta}\ d\mathbf{Z}$. This is achieved using *composition sampling*: for each sampled value of $\{\boldsymbol{\theta}, \mathbf{Z}\}$, we draw $Z_i(t_0)$, one for one, from $p(Z_i(t_0) \,|\, \mathbf{Z}, \boldsymbol{\theta})$, which is Gaussian. Also, our sampler easily adapts to situations where $Y_i(t)$ is missing (or not monitored) for some of the time points in region i. We simply treat such variables as missing values and update them, from their associated full conditional distributions, which of course are $N(\mathbf{x}_i(t)^T \boldsymbol{\beta} + Z_i(t), \tau_i^2)$. We assume that all predictors in $\mathbf{x}_i(t)$ will be available in the space-time data matrix, so this temporal interpolation step for missing outcomes is straightforward and inexpensive.

Model checking is facilitated by simulating *independent* replicates for each observed outcome: for each region i and observed timepoint t_j, we sample from

$$p(Y_{rep,i}(t_j) \,|\, \mathbf{Y}) = \int N(Y_{rep,i}(t_j) \,|\, \mathbf{x}_i(t_j)^T \boldsymbol{\beta} + Z_i(t_j), \tau_i^2)\ p(\boldsymbol{\beta}, Z_i(t_j), \tau_i^2 \,|\, \mathbf{Y}) d\boldsymbol{\beta} dZ_i(t_j) d\tau_i^2,$$

where $p(\boldsymbol{\beta}, Z_i(t_j), \tau_i^2 \,|\, \mathbf{Y})$ is the marginal posterior distribution of the unknowns in the likelihood. Sampling from the posterior predictive distribution is straightforward, again, using composition sampling.

2.3 Gradient analysis

Again, our primary goal is to carry out statistical inference on temporal gradients with data arising from a temporal process indexed discretely over space. We will do so using the notions of smoothness of a Gaussian process and its derivative. Adler (2009), Mardia et al. (1996) and Banerjee and Gelfand (2003) discuss derivatives (more generally, linear functionals) of Gaussian processes, while Banerjee, Gelfand and Sirmans (2003) lay out an inferential framework for directional gradients on a spatial surface. Most of the existing work on derivatives of stochastic processes deal either with purely temporal or purely spatial processes (see, e.g., Banerjee, 2010). Here, we consider gradients for a temporal process indexed discretely over space.

Assume that $\{Z_i(t) : t \in \Re^+\}$ is a stationary random process for each region i.[1] The process is L_2 (or mean square) continuous at t_0 if $\lim_{t \to t_0} E\left|Z_i(t) - Z_i(t_0)\right|^2 = 0$. The notion of a mean square differentiable process can be formalized using the analogous

[1] Stationarity is not required. We only use it to ensure smoothness of realizations and to simplify forms for the induced covariance function.

definition of total differentiability of a function in a non-stochastic setting (see, e.g., Banerjee and Gelfand, 2003): $Z_i(t)$ is mean square differentiable at t_0 if it admits a first order linear expansion for any scalar h,

$$Z_i(t_0 + h) = Z_i(t_0) + hZ_i'(t) + o(h) \tag{2.6}$$

in the L_2 sense as $h \to 0$, where we say that $\frac{d}{dt}Z_i(t) = Z_i'(t_0)$ is the *gradient* or *derivative* process derived from the *parent* process $Z_i(t)$. In other words, we require

$$\lim_{h\to 0} E\left(\frac{Z_i(t_0 + h) - Z_i(t_0)}{h} - Z_i'(t_0)\right)^2 = 0. \tag{2.6'}$$

Equations (2.6) and (2.6′) ensure that mean square differentiable processes are mean square continuous.

For a univariate stationary process, smoothness in the mean square sense is determined by its covariance or correlation function. A stationary multivariate process $\mathbf{Z}(t)$ with matrix-valued covariance function $K_Z(\Delta)$ will admit a well-defined gradient process $\mathbf{Z}'(t) = (Z_1'(t), \ldots, Z_{N_s}'(t))^T$ if and only if $K_Z''(0)$ exists, where $K_Z''(0)$ is the element-wise second-derivative of $K_Z(\Delta)$ evaluated at $\Delta = 0$.

A Gaussian process with a Matérn correlation function has sample paths that are $\lceil \phi_2 - 1 \rceil$ times differentiable. As $\phi_2 \to \infty$, the Matérn correlation function converges to the squared exponential (or the so-called Gaussian) correlation function, which is infinitely differentiable and leads to acute oversmoothing. When $\phi_2 = 0.5$, the Matérn correlation function is identical to the exponential correlation function (see, e.g., Stein, 1999). To ensure that the underlying process is differentiable so that the gradient process exists, we need to restrict $\phi_2 > 1$. However, letting $\phi_2 > 2$ usually leads to oversmoothing as the data can rarely distinguish among values of the smoothness parameter greater than 2. Hence, we restrict $\phi_2 \in (1, 2]$. We could either assign a prior on this support or simply fix ϕ_2 somewhere in this interval. Since it is difficult to elicit informative priors for the smoothness parameter, we would most likely end up with a uniform prior. In our experience, not only does this deliver only modest posterior learning and lead to an increase in computing (both in terms of MCMC convergence and estimating the resulting correlation function and its derivative), but the substantive inference is almost indistinguishable from what is obtained by fixing ϕ_2.

As such, in our subsequent analysis we fix $\phi_2 = 3/2$, which has the side benefit of yielding the closed form expression $\rho(\Delta;\phi_1) = (1+\phi_1|\Delta|)\times\exp(-\phi_1|\Delta|)$. The first and second order derivatives for the matrix-valued covariance function in (2.3) can now be obtained explicitly as

$$K'_Z(\Delta) = -\sigma^2\phi_1^2\Delta\exp(-\phi_1|\Delta|)(D-\alpha W)^{-1} \text{ and } -K''_Z(0) = \sigma^2\phi_1^2(D-\alpha W)^{-1}\ . \tag{2.7}$$

Turning to inference for gradients, we seek the joint posterior predictive distribution,

$$\begin{aligned} p(\mathbf{Z}'(t_0)\,|\,\mathbf{Y}) &= \int p(\mathbf{Z}'(t_0)\,|\,\mathbf{Y},\mathbf{Z},\boldsymbol{\theta})\ p(\mathbf{Z}\,|\,\boldsymbol{\theta},\mathbf{Y})\ p(\boldsymbol{\theta}\,|\,\mathbf{Y})d\boldsymbol{\theta}d\mathbf{Z} \\ &= \int p(\mathbf{Z}'(t_0)\,|\,\mathbf{Z},\boldsymbol{\theta})\ p(\mathbf{Z}\mid\boldsymbol{\theta},\mathbf{Y})\ p(\boldsymbol{\theta}\mid\mathbf{Y})d\boldsymbol{\theta}d\mathbf{Z}\ , \end{aligned} \tag{2.8}$$

where the second equality follows from the fact that the gradient process is derived entirely from the parent process and so $p(\mathbf{Z}'(t_0)\,|\,\mathbf{Y},\mathbf{Z},\boldsymbol{\theta})$ does not depend on $\mathbf{Y}$.

We evaluate (2.8) using composition sampling. First, we obtain $\boldsymbol{\theta}^{(1)},\boldsymbol{\theta}^{(2)},\ldots,\boldsymbol{\theta}^{(M)} \sim p(\boldsymbol{\theta}\mid\mathbf{Y})$ and $\mathbf{Z}^{(j)} \sim p(\mathbf{Z}\mid\boldsymbol{\theta}^{(j)},\mathbf{Y}), j = 1,2,\ldots,M$, where M is the number of (post-burn-in) posterior samples. Next, for each j we draw $\mathbf{Z}^{(j)} \sim p(\mathbf{Z}\mid\boldsymbol{\theta}^{(j)},\mathbf{Y})$, and finally $\mathbf{Z}'(t_0)^{(j)} \sim p(\mathbf{Z}'(t_0)\,|\,\mathbf{Z}^{(j)},\boldsymbol{\theta}^{(j)})$. The conditional distribution for the gradient can be seen to be multivariate normal with mean and variance-covariance matrix given by

$$\begin{aligned} \boldsymbol{\mu}_{Z'\,|\,Z,\theta} &= \mathrm{cov}(\mathbf{Z}'(t_0),\mathbf{Z})\mathrm{var}(\mathbf{Z})^{-1}\mathbf{Z} = -(K'_Z)^T\Sigma_Z^{-1}\mathbf{Z} \\ \text{and } \Sigma_{Z'\,|\,Z,\theta} &= -K''_Z(0) - (K'_Z)^T\Sigma_Z^{-1}(K'_Z)\ , \end{aligned}$$

where $\Sigma_Z^{-1} = \frac{1}{\sigma^2}R(\phi)^{-1}\otimes(D-\alpha W)$ and $(K'_Z)^T$ is an $N_s \times N_sN_t$ block matrix whose j-th block is given by the $N_s \times N_s$ matrix $K'_Z(\Delta_{0j})$, with $\Delta_{0j} = t_j - t_0$. Note that $\Sigma_{Z'\,|\,Z,\theta}$ is an $N_sN_t \times N_sN_t$ matrix, but we can use the properties of the MRF to only invert $N_t \times N_t$ matrices.

2.4 Simulation studies

To validate our model's ability to correctly estimate both our model parameters and the underlying temporal gradients, we have constructed two separate simulation studies using the $N_s = 58$ counties of California as our spatial grid and $N_t = 50$ observations

per county, where $\mathcal{T} = \{1, 2, \ldots, 50\}$. Each simulation study consists of 100 datasets comprised of 2,900 observations generated from (2.1), where $\mu_i(t) = \mathbf{x}_i(t)^T\boldsymbol{\beta}$, using the same parameter values, and our results are based on 5,000 MCMC samples after a burn-in period of 5,000 iterations.

In an effort to obtain simulated outcomes comparable to those from our real data, our first simulation study uses an intercept and the four covariates described in Section 1.3.1, and we set the 5×1 vector, $\boldsymbol{\beta}$, as the least squares estimates from our real data. We also set $\phi = 1$, $\alpha = 0.90$, and $\sigma^2 = 18$, which are then used to generate true values for $\mathbf{Z}$, while our τ_i^2 are drawn from an Inverse Gamma distribution centered at 1 with modest variance. For each of the 100 simulated datasets, we constructed 95% Bayesian credible intervals for each parameter and recorded the number of times they included their true values (i.e., their "frequentist coverage"). We found this coverage to be between $93-97\%$ for the 5 β's, about 87% for the random effect variance σ^2, and around 90% on the average for the 58 τ_i^2's, with the majority of them having 95% coverage. Coverage was poor for $\tau_i^2 < 0.15$; in situations where small variances are to be expected, this issue could be avoided or alleviated by rescaling the data or specifying a prior with a larger mass near 0, respectively. The spatiotemporal random effects, $\mathbf{Z}$, also enjoyed satisfactory coverage; the average coverage over the 2,900 space-time random effects was around 95.5%. By contrast, the coverage for the propriety parameter, α, and the spatial range parameter, ϕ, reveal biases, with coverages less than 50%. This is not entirely unexpected as spatial and temporal range parameters of this type are known to be weakly identified by the data (e.g., Zhang, 2004). Furthermore, the biases for ϕ and α are not substantial, with their posterior medians only 8% above and 5% below their true values, respectively. In an effort to verify the robustness of our model to these biases, we repeated the simulation with both ϕ and α fixed at their true values and were able to reproduce our results.

Having demonstrated the ability of our model to correctly estimate model parameters, the focus of our second simulation study is to validate the theory of our temporal gradient processes. To do this, we assumed

$$Y_i(t_j) \overset{\text{ind}}{\sim} N\left(5 + x_{i1} * \sin\left(\frac{t_j}{2}\right) + x_{i2} * \cos\left(\frac{t_j}{2}\right), \tau_i^2\right), \tag{2.9}$$

where x_{i1} is the ith county's percent black and x_{i2} is the ith county's ozone level from

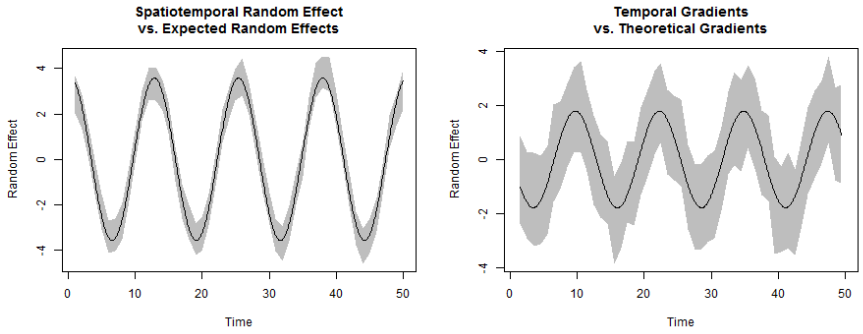


Figure 2.1: Spatiotemporal random effects and temporal gradients for a region based on one dataset from the second simulation study. Solid black lines denote true sinusoidal curves based on the model in eq. (2.9), while gray bands represent 95% credible intervals.

April 1991, as described in Section 1.3.1; this was done in order to induce spatial clustering. As there was no evidence of an association between the coverage of the random effects, $\mathbf{Z}$, and the region-specific variance parameters, values of τ_i^2 were generated from a Uniform(0.5,2.0) distribution in order to avoid the extreme values of the inverse gamma and focus our attention on the random effects themselves. After generating 100 datasets based on these parameters, we then modeled the data using only an intercept, leaving the spatiotemporal random effects to capture the sinusoidal curve, and conducted the gradient analysis at the midpoints of each time interval.

Figure 2.1 displays the true spatiotemporal random effects and temporal gradients for a particular region, along with their 95% CI estimated from one of the 100 datasets. As can be seen, our Gaussian process model accurately estimates both the random effects and the temporal gradients. Across all 100 datasets, 98.3% of the the theoretical gradients derived using elementary calculus were covered by their respective 95% CI, confirming the validity of the gradient theory derived in Section 2.3.

2.5 Analysis of the California asthma hospitalization data

As first mentioned in Section 1.3.1, our dataset is comprised of monthly asthma hospitalization rates in the counties of California over an 18-year period. As such, $N_t = 12 * 18 = 216$, and we will again use $t_j = j = 1, 2, \ldots, N_t$. The covariates in this model

include population density, ozone level, the percent of the county under 18, and percent black. Population-based covariates are calculated for each county using the 2000 U.S. Census, thus they do not vary temporally. However, the covariate for ozone level is aggregated at the air basin level and varies monthly, though show little variation annually. In order to accommodate seasonality in the data, monthly fixed effects are included, using January as a baseline. Thus, after accounting for the monthly fixed effects and the four covariates of interest, $\mathbf{x}_i(t)$ is a 16×1 vector.

To justify the use of the model we've described, we compare it to three alternative models using the DIC criterion (Spiegelhalter et al., 2002) and a predictive model choice criterion using strictly proper scoring rules proposed by Gneiting and Raftery (2007, eq. (27)). Following Czado et al. (2009), we refer to this as the Dawid-Sebastiani (D-S) score (Dawid and Sebastiani, 1999). These models are all still of the form

$$Y_i(t) = \mathbf{x}_i(t)'\boldsymbol{\beta} + Z_i(t) + \epsilon_i(t), \quad \epsilon_i(t) \overset{ind}{\sim} N(0, \tau_i^2) \text{ for } i = 1, 2, \ldots, N_s \,, \tag{2.10}$$

but with different $Z_i(t)$. Our first model is a simple linear regression model which ignores both the spatial and the temporal autocorrelation, i.e., $Z_i(t) = 0 \; \forall \; i, t$. The second model allows for a random intercept and random temporal slope, but ignores the spatial nature of the data, i.e., here $Z_i(t) = \alpha_{0i} + \alpha_{1i}t$, where $\alpha_{ki} \overset{\text{iid}}{\sim} N(0, \sigma_k^2)$, for $k = 0, 1$. In this model, to preserve model identifiability, we must remove the global intercept from our design matrix, $\mathbf{x}_i(t)$. Our third model builds upon the second, but introduces spatial autocorrelation by letting $\boldsymbol{\alpha}_k = (\alpha_{k1}, \ldots, \alpha_{kN_s})' \sim CAR(\sigma_k^2), k = 0, 1$. The results of the model comparison can be seen in Table 2.1, which indicates that our Gaussian process model has the lowest DIC value and D-S score, and is thus the preferred model and the only one we consider henceforth. The surprisingly large p_D for the areally referenced Gaussian process model arises due to the very large size of the dataset (58 counties $\times$ 216 timepoints).

The estimates for our model parameters can be seen in Table 2.2. The coefficients for the monthly covariates indicate decreased hospitalization rates in the summer months, a trend which is consistent with previous findings. The coefficients for population density, percent under 18, and percent black are all significantly positive, also as expected. The coefficient for ozone level is significantly negative, however, which is surprising but consistent with the patterns in the monthly trends for both hospitalization rates and

	p_D	DIC*	D-S*
Simple Linear Regression	79	9,894	16,166
Random Intercept and Slope	165	4,347	10,403
CAR Model	117	7,302	13,436
Areally Referenced Gaussian Process	5,256	0	0

Table 2.1: Comparisons between our areally referenced Gaussian process model and the three alternatives. p_D is a measure of model complexity, as it represents the effective number of parameters. Smaller values of DIC and Dawid-Sebastiani (D-S) scores indicate a better trade-off between in sample model fit and model complexity.
* Both DIC and D-S shown are standardized relative to our areally referenced Gaussian Process model.

ozone levels. This result may also be confounded by the absence of other climate-related factors and the sensitivity of asthma admissions to acute weather effects.

There is a large range of values for the county-specific residual variance parameters, τ_i^2. Perhaps not surprisingly, the magnitude of these terms seems to be negatively correlated with the population of the given counties, demonstrating the effect a (relatively) small denominator can have when computing and modeling rates. The strong spatial story seen in the maps is reflected by the size of σ^2 compared to the majority of the τ_i^2. There is also relatively strong temporal correlation, with $\phi = 0.9$ corresponding to $\rho(t_i, t_j; \phi) \geq 0.4$ for $|t_j - t_i|$ less than 2 months.

Maps of the yearly (averaged across month) spatiotemporal random effects can be seen in Figure 2.2. Since here we are dealing with the *residual* curve after accounting for a number of mostly non-time-varying covariates, it comes as no surprise that the spatiotemporal random effects capture most of the variability in the model, including the striking decrease in yearly hospitalization rates over the study period. It also appears that our model is providing a better fit to the data in the years surrounding 2000, perhaps indicating that we could improve our fit by allowing our demographic covariates to vary temporally. Our model also appears to be performing well in the central counties, where asthma hospitalization rates remained relatively stable for much of the study period.

Parameter	Median (95% CI)	Parameter	Median (95% CI)
β_0 (Intercept)	9.17 (8.93, 9.42)	β_{10} (July)	-3.78 (-4.21, -3.37)
β_1 (Pop Den)	0.60 (0.49, 0.70)	β_{11} (August)	-3.58 (-4.02, -3.13)
β_2 (Ozone)	-0.18 (-0.28, -0.08)	β_{12} (September)	-1.96 (-2.37, -1.54)
β_3 (% Under 18)	1.24 (1.15, 1.34)	β_{13} (October)	-1.36 (-1.73, -1.00)
β_4 (% Black)	1.12 (1.01, 1.24)	β_{14} (November)	-0.71 (-1.02, -0.42)
β_5 (February)	-0.25 (-0.46, -0.04)	β_{15} (December)	0.63 (0.41, 0.86)
β_6 (March)	-0.21 (-0.48, 0.07)	ϕ	0.90 (0.84, 0.97)
β_7 (April)	-1.47 (-1.81, -1.12)	α	0.77 (0.71, 0.80)
β_8 (May)	-1.17 (-1.53, -0.8)	σ^2	21.52 (20.18, 23.06)
β_9 (June)	-2.79 (-3.21, -2.4)	$\bar{\tau}_\cdot^2$	3.32 (0.18, 213.16)

Table 2.2: Parameter estimates for asthma hospitalization data, where estimates for $\bar{\tau}_\cdot^2$ represent the median (95% CI) of the $\tau_i^2, i = 1, \ldots, N_s = 58$.

In the top panel of Figure 2.3, we compare the monthly temporal profiles of the random effects for Los Angeles and San Francisco Counties. For Los Angeles County, the spatiotemporal random effects (top-left panel) decrease at a consistent, moderate rate throughout the length of the study with several large spikes prior to 2000. In contrast, San Francisco County's random effects (top-right) have fewer and less dramatic spikes. In addition, San Francisco County appears to have had a changepoint in its spatiotemporal random effects around 2000, where they transition from a fairly steady decline to a period of lower variability and very little mean change. Further investigation may reveal a corresponding change in social, environmental, or health care reimbursement policy. The bottom-left panel shows the temporal trend of the gradients in Los Angeles County, which reveal the large degree of variability in the random effects. In fact, as more clearly shown in the bottom-right panel of Figure 2.3, the September to October gradient was significantly positive five times between 1995 and 2001, and three times during this period (1995, 1997, and 1999) the November to December gradients were significantly positive, but were immediately followed by significantly *negative* gradients from December to January, a pattern that is seen throughout the region.

A strength of using a continuous-time model for these data is that it seamlessly

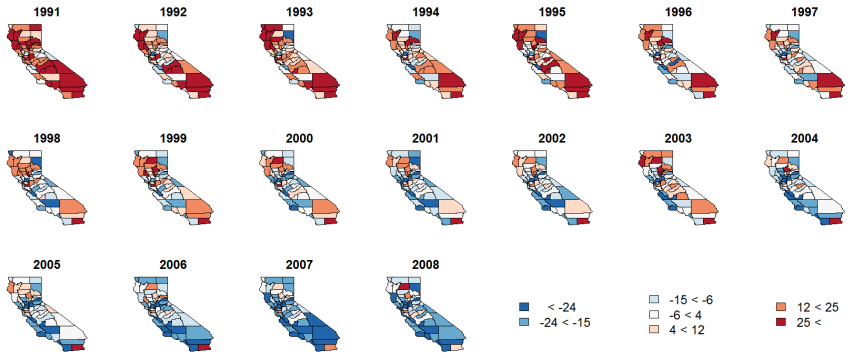


Figure 2.2: Spatial random effects for asthma hospitalization data, by year

permits prediction at a finer resolution than that of the observed data. Upon seeing the significant gradients in Los Angeles County in November and December of 1995, public health officials may ask for a more detailed report than a monthly aggregation can provide. If a discrete-time model were used, researchers would be required to refit the model, pre-specifying at which unobserved time points to conduct inference; however, with this model, we can use the posterior predictive distribution to interpolate values at any time. As a demonstration of this, Figure 2.4 displays the predicted daily values (solid line) and 95% CI bands (dashed lines) every 3 days during the period November 15, 1995 to January 15, 1996, plotted against the true observed rates (open circles). Despite substantial noise in the data and modeling based solely on the aggregate rates for each month (and assigning that value to the temporal midpoint of each month), our predictions and 95% CI bands perform reasonably well.

As our data are aggregated monthly, we felt it was also important to investigate the gradients on a month-to-month basis over the course of the study. For instance, Figure 2.5 reveals the gradients between August and September decrease substantially statewide over the course of the study. Coupling this with the information in Table 2.2, which indicates that hospitalization rates in September are $\beta_{12} - \beta_{11} = 1.62$ per 100,000 higher than those in August, suggests that the difference in asthma hospitalization rates between August and September has decreased nearly 60%, going from roughly 2.31 at the beginning of the period to just 0.97 by the end. An investigation of the raw hospitalization rates shows a similar trend, but this is to be expected since most

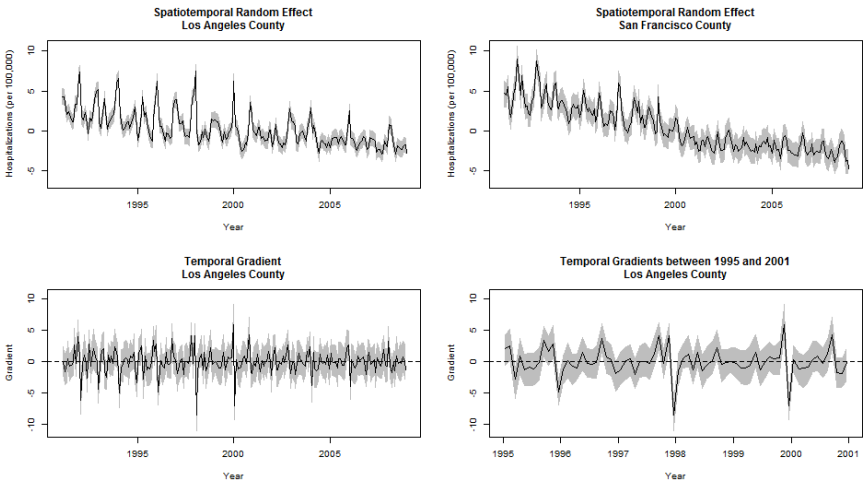


Figure 2.3: Comparison between the spatiotemporal random effects in Los Angeles and San Francisco Counties, and an investigation of temporal gradients in Los Angeles County. Point estimates in black and corresponding 95% CI bands in gray. Figures in the top panel illustrate the differences in the temporal trends of the random effects between the two counties. The bottom-left figure displays the temporal gradients computed between months in Los Angeles County, and the bottom-right figure displays the subset of the gradients which are further described in the text.

of the spatiotemporal variability in the model is accounted for by the random effects. A similar, though not as striking phenomenon occurs between March and April, where the gradients are increasing. As these two pairs of months lie on the transition between the warmer months and the cooler months, this result would seem to suggest that the effect of seasonality has moderated over the length of the study.

One limitation of this analysis is that the data records asthma *hospitalizations*, not overall *prevalence*. This is an important distinction, as factors that trigger symptoms of asthma may not be the same as or have the same impact on asthma hospitalizations. For instance, residents of regions with high risk environments may be better educated about and/or prepared for managing their symptoms, which could lead to a relative decrease in asthma hospitalization rates. Another limitation is that, due to the aggregation of our data, we have an inconvenient interpretation of the daily estimates in Figure 2.4. A more accurate interpretation of these values is that they are the average daily rates for

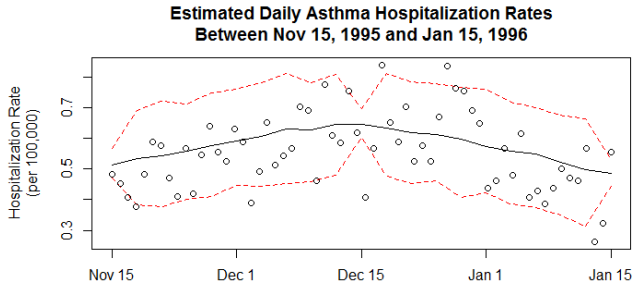


Figure 2.4: Posterior predicted curves (and 95% credible bounds) for the daily asthma hospitalization rates in Los Angeles County between November 15, 1995 to January 15, 1996. This county and interval was selected due the presence of a significantly positive gradient between November and December and a significantly negative gradient between December and January. The true hospitalization are also shown for comparison purposes, though the model was fit using only the monthly aggregates.

the one-month interval centered at a particular day. More generally, the interpretation of predicted values at any time point is determined by the aggregation of the data, but this is certainly not unique to this model.

2.6 Discussion

In this chapter, we have provided an overview of parent and gradient processes, building on previous work in spatiotemporal Gaussian process modeling. We then described our modeling framework and methodology that allows for inference on temporal gradients. An implementation of this work was outlined in Section 2.2, and its theory was verified via simulation. Its use was then illustrated on a real dataset in Section 2.5; where our results showed real insight can be gained from an assessment of temporal gradients in the residual Gaussian process, indicating overall trends as well as motivating a search for temporally interesting covariates still missing from our model (say, one that changes abruptly in San Francisco County around 2000).

We believe there are two primary points of discussion regarding this work, the first of which is the use of modeling time as continuous. If inference is desired at the resolution

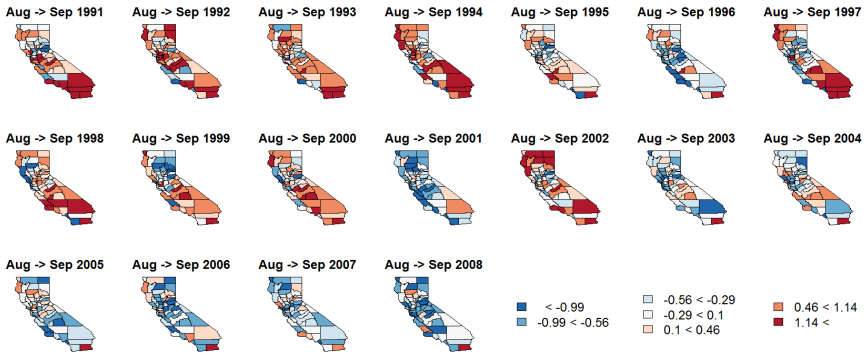


Figure 2.5: Temporal gradients for transition from August to September over time

of the data only, then several of the discrete-time models in the literature would be appropriate. In our example, however, public health officials may be interested in the daily effects of asthma, which can be correlated with effects of daily variation of temperature and a variety of atmospheric pollutants. A practical issue here is that hospitalization data are often more cleanly available as monthly aggregates (say, due to patient confidentiality issues, like those described in Section 1.3.1), and, even when the daily data are available, they tend to be both massive and very likely to have many missing values. Analyzing such data using using discrete-time models would require methods for handling temporal misalignment, while our temporal process-based methods can handle such inference in a posterior predictive fashion.

A second important point of discussion is the importance of significance with respect to these temporal gradients. We believe it depends on the problem being modeled. While we have accounted for monthly differences in our design matrix, the $Z_i(t)$ here may simply be capturing the remaining cyclical trend, and this is why we felt it was more beneficial to focus on a side-by-side comparison of two of California's most populous counties, which motivated a further investigation of Los Angeles County, and the trends of the twelve month-to-month comparisons rather than solely on whether a specific gradient for a particular county was significant. In situations where it's reasonable to assume two time-points are comparable, investigating significant temporal gradients can indicate periods of important changes in the data, which may be caused by rapid changes in missing covariates. We also point out that the methodology for gradients

outlined here can be applied to more general spatial functional data analysis contexts and will be especially useful for estimating gradients from high-resolution samples of the function.

Regarding the specific application of this methodology in this chapter, it bears mentioning that modeling our data as *rates* is not the only option. Often, the counts themselves are modeled directly using a log-linear model, with a Poisson distributional assumption justified as a rare-events approximation to the binomial. In this setting, however, we would no longer be able to rely on the closed form Gibbs sampler for updating our random effects, instead requiring Metropolis updates and a substantial increase in computational burden. Another option is to use a Freeman-Tukey transformation of the rates and a single error variance parameter, τ^2, which is scaled by the county's population, as shown in Freeman and Tukey (1950) and Cressie and Chan (1989), with the goal of justifying the assumption of normality. Given the population sizes we're dealing with, we believe the assumption of normality of our observed rates can be justified as a normal approximation to the binomial. Furthermore, an analysis of the transformed data results in nearly identical substantive findings. However, there is a drawback: by modeling transformed values instead of the rates themselves, we lose the interpretability of the scale for not only our regression parameters, but also the temporal gradients. In our experience, a common question among public health practitioners is, "What does this mean?" As such, we feel that having results which are straightforward to interpret is of the utmost importance, and thus, we chose to model the untransformed rates. Incidentally, we also considered modeling the untransformed rates using a model with a single error variance parameter (scaled by population). Sadly, the simplicity of this model failed to outweigh its loss of flexibility, and, in any case, this model would not be generalizable to non-rate data.

One weakness of our model is if the true underlying process is less smooth in some regions than others, or if there are spatial outliers, our model may simultaneously both oversmooth *and* undersmooth the random effects, $\mathbf{Z}$. In our gradient simulation in Section 2.4, the counties of Alameda (home of Oakland) and Solano have significantly larger percentages of African Americans than any other county in the state. As a result, the true underlying process that we've constructed using (2.9) for these counties takes much more extreme values than their neighbors, resulting in oversmoothing in these

counties and creating the potential for undersmoothing in other counties. While this issue is not unique to our model, this can lead to poor estimation of the temporal gradients, such as biased estimates or wide credible intervals.

Chapter 3

Heteroscedastic CAR models for areally referenced temporal processes

In this chapter, we enrich the continuous-time dynamic CAR models proposed in Chapter 2 with region-specific variance components, in an effort to better control the smoothing. Section 3.1 provides background on dynamic CAR models and the motivation for this work. In Section 3.2, we propose a heteroscedastic CAR (HCAR) model, including a derivation of its temporal gradient process. We then conduct two simulation studies in Section 3.3, in which we investigate the ability of our model to estimate the parent and gradient processes and important model parameters. Our analysis of the California asthma hospitalization rate data using the HCAR follows in Section 3.4 and provides the following tangible benefits for California asthma: an increase in the effect for percent black, sharper temporal gradients, and an improved fit for Imperial County, a region known for its high asthma hospitalization rates. Finally, Section 3.5 summarizes our findings and concludes.

3.1 Dynamic CAR models

Consider again a map comprising N_s regions that are delineated by well-defined boundaries, and let $Y_i(t)$ be the outcome arising from region i at time t. While the region-specific outcome $Y_i(t)$ is conceptualized as a continuous function of time, the observations are again collected at a finite collection of distinct time points $\mathcal{T} = \{t_1, t_2, \ldots, t_{N_t}\}$.

The observation from region i at time t_j is modeled by a space-time regression model

$$Y_i(t_j) = \mu_i(t_j) + Z_i(t_j) + \epsilon_i(t_j), \quad \epsilon_i(t_j) \overset{ind}{\sim} N(0, \tau_i^2), \tag{3.1}$$

for $i = 1, 2, \ldots, N_s$ and $j = 1, 2, \ldots, N_t$, where $\mu_i(t_j) = \mathbf{x}_i(t_j)^T \boldsymbol{\beta}$ captures large scale variation or trends, $\mathbf{x}_i(t_j)$ is a vector of explanatory variables observed at the county level for each timepoint, $\boldsymbol{\beta}$ is the corresponding vector of regression slopes, $Z_i(t_j)$ is the space-time random effect arising from an areally-referenced stochastic process over time, $Z(t)$, that captures smaller-scale variations in the time scale while also accommodating spatial associations, and τ_i^2 captures any residual variation not captured by the other components for region i.

In Chapter 2, we let the process $Z_i(t)$ be looked upon as a random function of time that specifies the probability distribution of correlated space-time random effects. Specifically, we collected the $Z_i(t_j)$'s for all the regions into an $N_s \times 1$ vector function $\mathbf{Z}(t_j)$ for any timepoint t_j and then stacked them into an $N_s N_t \times 1$ column vector $\mathbf{Z} = (\mathbf{Z}(t_1)^T, \mathbf{Z}(t_2)^T, \ldots, \mathbf{Z}(t_{N_t})^T)^T$, where the distribution of $\mathbf{Z}$ is given by

$$\mathbf{Z} \sim N(\mathbf{0}, R(\phi) \otimes \sigma^2 (D - \alpha W)^{-1}) \,, \tag{3.2}$$

where $\sigma^2(D - \alpha W)^{-1}$ is the covariance matrix for the proper CAR model and the matrix $R(\phi)$ is an $N_t \times N_t$ temporal correlation matrix based on the Matérn(3/2). This structure ensured that the $Z_i(t)$'s are mean-square differentiable functions of t and thus legitimizing inference on infinitesmal rates of temporal change.

A limitation of the model in (3.2) is the presence of a single variance parameter σ^2 to capture the scale of temporal and spatial variations. The diagonal elements in $(D - \alpha W)^{-1}$ depend upon the adjacencies in the map and adjust σ^2 accordingly to offer region-specific marginal variances. To illustrate what may go wrong with the specification in (3.2), we conducted a simulation study using the California map with

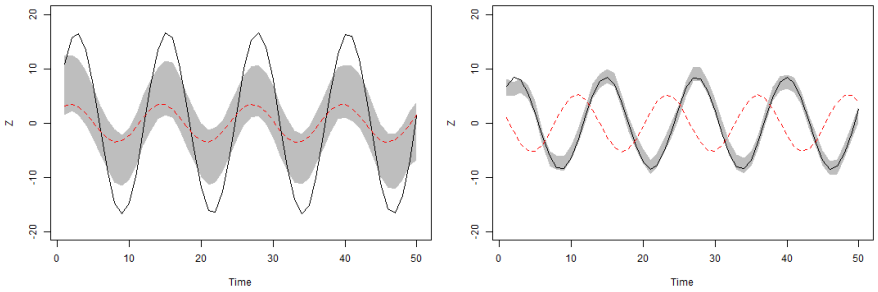


Figure 3.1: Posterior bands for $\mathbf{Z}_i$ in Alameda (left) and Lassen (right) Counties plotted against the true value (black line) and mean of neighbors (red line) in our simulation study.

$N_t = 50$ and observations generated from the model

$$Y_i(t_j) \overset{\text{ind}}{\sim} N\left(x_{i1} * \sin\left(\frac{t_j}{2}\right) + x_{i2} * \cos\left(\frac{t_j}{2}\right), \tau_i^2\right), \tag{3.3}$$

where we induced spatiotemporal clustering by letting x_{i1} and x_{i2} be the ith county's % black and April 1991 ozone level, respectively. In this simulation study, two particular California counties caught our attention — Alameda (home of Oakland) and Solano — due to their significantly larger-than-average African American population densities. This demographic feature generated extreme values for these counties, as displayed for Alameda County in the left panel of Figure 3.1. This simulation study revealed a tendency of the model to oversmooth regions with extreme values in $\mathbf{Z}$, because the estimated value of σ^2 is too small to capture the full shape of the underlying curve. Conversely, this value of σ^2 is likely larger than necessary in some regions, which can lead to unnecessarily wide credible intervals for the temporal and temporal gradient processes. A benefit of the inflated σ^2 is that the fit for regions with temporal patterns far different than their neighbors but with similarly sized extremes does not suffer, as illustrated for Lassen County in the right panel of Figure 3.1. These counties will be the primary focus of our simulation study in Section 3.3.1, at which point we will label their locations on the maps in Figure 3.3.

The issue of differing levels of variability in areal spatial settings has been addressed before. Two primary examples include Lawson and Clark (2002), who split the spatial process into a mixture of a CAR component (L_2) and an L_1 process, and Brewer and

Nolan (2007), who developed an empirical Bayes, pairwise additive approach of the form

$$\pi(Z_i \mid Z_{(i)}) \propto \exp\left[-\frac{1}{2}\sum_{k\sim i}\frac{(Z_i - Z_k)^2}{\sigma_{ik}^2}\right] \tag{3.4}$$

where the pairwise-defined $\sigma_{ik}^2 = \sigma_i^2 + \sigma_k^2$ have replaced σ^2 in the standard CAR model and $Z_{(i)}$ denotes the vector $\mathbf{Z}$ with the ith element removed.

Reich and Hodges (2008) take a similar approach to Brewer and Nolan with their fully Bayesian, spatially adaptive CAR (SACAR) model, in which they assume $\sigma_{ik}^2 = \sigma_i\sigma_k$, where $\sigma_i^2 = \exp(s_0 + s_i)$ and $\mathbf{s} = (s_1, \ldots, s_{N_s})' \sim CAR(\lambda)$ in order to ensure the identifiability of the parameters; that is, the prior distribution of the σ_i^2 *assumes* a spatial structure. In this approach, the s_i are constrained such that $\sum_i s_i = 0$ and s_0 is an intercept term to be estimated, though this model can also permit regression models with predictors for $\log(\sigma_i^2)$. These techniques are based on the improper CAR model (which sets $\alpha = 1$) and designed for purely spatial models, but extending (3.2) to allow for the proper analogs of these structures appears straightforward.

3.2 The heteroscedastic CAR (HCAR) model

To remedy the problem in the left panel of Figure 3.1, we allocate a different variance component σ_i^2 to each region. Rather than assume $\mathbf{Z}(t) \sim CAR(\sigma^2)$ for each time point, we let

$$\pi(Z_i(t) \mid Z_{(i)}(t)) \propto \exp\left[-\frac{1}{2}\sum_{k\sim i}\left(\frac{Z_i(t)}{\sigma_i} - \alpha\frac{Z_k(t)}{\sigma_k}\right)^2\right], \tag{3.5}$$

which is denoted $\mathbf{Z}(t) \sim HCAR(\boldsymbol{\sigma})$. In this structure, the σ_i can be directly viewed as scaling parameters for their respective Z_i, rather than as components of a pairwise structure. A structural comparison between the CAR, the SACAR, and the HCAR is presented in Table 3.1. Extending this structure to our space-time case, we replace $\mathbf{Z}$ in (3.2) with

$$\mathbf{Z} \sim N(\mathbf{0}, R(\phi) \otimes T(D - \alpha W)^{-1}T)\ , \tag{3.6}$$

where T is an $N_s \times N_s$ diagonal matrix with σ_i as its ith diagonal element. Let $\mathbf{Z} \sim HCAR_{ST}(\sigma_1, \ldots, \sigma_{N_t})$ denote the expression in (3.6).

The joint distribution in (3.6) also arises as a nonsingular linear transformation of latent areally-referenced processes following the dynamic CAR distribution in (3.2)

	$\mathbf{Z} \sim CAR(\sigma^2)$	$\mathbf{Z} \sim SACAR(\boldsymbol{\sigma})$	$\mathbf{Z} \sim HCAR(\boldsymbol{\sigma})$
Precision	$(\sigma^2)^{-1}(D - \alpha W)$	$A - \alpha T^{-1} W T^{-1}$	$T^{-1}(D - \alpha W)T^{-1}$
Pairwise	$\sum_i \sum_{j \sim i} \frac{(Z_i - \alpha Z_j)^2}{\sigma^2}$	$\sum_i \sum_{j \sim i} \frac{(Z_i - \alpha Z_j)^2}{\sigma_i \sigma_j}$	$\sum_i \sum_{j \sim i} \left(\frac{Z_i}{\sigma_i} - \alpha \frac{Z_j}{\sigma_j} \right)^2$
Conditional Mean	$\alpha \frac{1}{n_i} \sum_{j \sim i} Z_j$	$\alpha \frac{1}{\sum_{j \sim i} \sigma_j} \sum_{j \sim i} \frac{Z_j}{\sigma_j}$	$\alpha \frac{\sigma_i}{n_i} \sum_{j \sim i} \frac{Z_j}{\sigma_j}$
Conditional Variance	$\frac{\sigma^2}{n_i}$	$\frac{\sigma_i}{\sum_{j \sim i} \frac{1}{\sigma_j}}$	$\frac{\sigma_i^2}{n_i}$

Table 3.1: Comparison of the CAR, SACAR, and HCAR models for $\mathbf{Z} = (Z_1, \ldots, Z_{N_s})$. Here, A is a diagonal matrix with $A_{ii} = \frac{1}{\sigma_i} \sum_{j \sim i} \frac{1}{\sigma_j}$ and T is a diagonal matrix with $T_{ii} = \sigma_i$. Note that the CAR model is a special case of both the SACAR and HCAR models where $\sigma_i = \sigma$ for all i.

with $\sigma^2 = 1$. To be precise, suppose that $U_i(t)$ is a latent process such that the joint distribution of the random vector $\mathbf{U}$, formed analogously to how $\mathbf{Z}$ in Section 3.1, is given by (3.2). If $\mathbf{Z}(t) = T\mathbf{U}(t)$, where T is a nonsingular matrix, then the joint distribution of $\mathbf{Z}$ is of the form in (3.6). This "construction" ensures a valid stochastic process for $\mathbf{Z}(t)$ and a well-defined predictive distributions for any arbitrary time point t_0, as long as T is nonsingular. Because $\mathbf{Z}(t)$ is a well-defined process, we can predict $\mathbf{Z}(t)$ at any time point – a matter of substantial importance for our subsequent inference. The $HCAR_{ST}(\sigma_1, \ldots, \sigma_{N_t})$ emerges as a special case with T assumed to be diagonal with $\sigma_i > 0$ as its diagonal elements. In this spirit, (3.6) can be regarded as a dynamic version of classes explored by Jin et al. (2007).

When it comes to modeling the σ_i, we use an approach similar to that used by Reich and Hodges (2008). We let $\sigma_i = \exp(u_0 + u_i)$ where u_0 represents a baseline for our spatiotemporal variance and the u_i are region-specific adjustments with the constraint that $\sum_i u_i = 0$. Unlike the SACAR model, however, we do not assume a spatial correlation structure on the u_i, as doing so would restrict our model's flexibility

for fitting outlying regions. Instead, we assume $u_i \sim N(0, \gamma^2)$; we then place inverse Gamma priors on both $\sigma_0^2 = \exp(2u_0)$ and γ^2.

While one can interpret the σ_i's as region-specific spatiotemporal variance parameters, further insight is obtained from an alternative parameterization:

$$T(D - \alpha W)^{-1}T = \left(\frac{1}{\sigma_0^2}(D^* - \alpha W^*)\right)^{-1} = \sigma_0^2(D^* - \alpha W^*)^{-1}, \tag{3.7}$$

where

$$\{W^*\}_{ij} = w_{ij}^* = \begin{cases} \frac{\sigma_0}{\sigma_i}\frac{\sigma_0}{\sigma_j} = w_i^* w_j^*, & \text{if } i \sim j \\ 0, & \text{otherwise} \end{cases},$$

and $\{D^*\}_{ii} = n_i^* = \frac{n_i \sigma_0^2}{\sigma_i^2}$. The benefits of this reparameterization are three-fold. First and foremost, a single spatiotemporal variance summary σ_0^2 is perhaps a more intuitive structure due to its similarity to the standard CAR model. Secondly, the $w_i^* = \sigma_0/\sigma_i$ can be viewed as *scaled weight* parameters for the ith region; for instance, regions with $w_i^* \geq 1$ can be viewed as *super* neighbors, as their spatiotemporal processes are constrained to be more similar to their adjacent regions. Finally, $D_{ii}^* = n_i^*$ can now be viewed as the "effective number of neighbors" of region i.

For the remaining parameters, more conventional priors are suitable. Specifically, we place non-informative, conjugate priors on the regression coefficients $\boldsymbol{\beta}$, and assume the τ_i^2 follow inverse gamma priors with mean 1 and infinite variance. Conjugate priors do not exist for our spatial association parameter α or our temporal association parameter ϕ; as such, we have chosen a beta prior with mean 0.9 and an infinite peak at 1 for α, and a uniform prior for ϕ. For the HCAR model presented in (3.5), the joint distribution for the observations and all other unknowns is thus given by

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{Z}, \mid \mathbf{Y}) \propto & N(\boldsymbol{\beta} \mid \mu_\beta, \Sigma_\beta) \times \prod_{i=1}^{N_s} \left[N(u_i \mid 0, \gamma^2) \times IG(\tau_i^2 \mid a_\tau, b_\tau)\right] \\ & \times IG(\sigma_0^2 \mid a_\sigma, b_\sigma) \times IG(\gamma^2 \mid a_\gamma, b_\gamma) \times U(\phi \mid a_\phi, b_\phi) \\ & \times Beta(\alpha \mid a_\alpha, b_\alpha) \times N(\mathbf{Z} \mid \mathbf{0}, R(\phi) \otimes T(D - \alpha W)^{-1}T) \\ & \times \prod_{j=1}^{N_t}\prod_{i=1}^{N_s} N(Y_i(t_j) \mid \mathbf{x}_i(t_j)^T\boldsymbol{\beta} + Z_i(t_j), \tau_i^2), \end{aligned} \tag{3.8}$$

where we are letting $\mu_i(t) = \mathbf{x}_i(t)'\boldsymbol{\beta}$ and $\boldsymbol{\theta} = \{\boldsymbol{\beta}, \boldsymbol{\phi}, \alpha, \sigma_0^2, u_1, \ldots, u_{N_s}, \gamma^2, \tau_1^2, \ldots, \tau_{N_s}^2\}$. We will use MCMC to evaluate (3.8), using Metropolis steps for updating $u_1, \ldots, u_{N_s}$, ϕ, and α, and Gibbs steps for all other parameters.

3.2.1 Modeling temporal gradients

In addition to providing a good fit to the data, our model has been specified in order to permit inference on the temporal gradient process. First developed for two-dimensional, continuous space case in Banerjee et al. (2003), the gradient process can be used to detect sudden changes in the residual surface. Using real estate prices in Baton Rouge, LA, Banerjee et al. show that significant gradients can be used to indicate important predictors, such as proximity to popular shopping centers, still missing from the mean model. Here, we are interested in temporal gradients, which may correspond to features such as sudden temporal changes in weather or public health policy that can affect asthma hospitalization rates.

The development of our temporal gradient process is similar to that in Chapter 2. As the temporal correlation structure of our model remains a Matérn(3/2), inference for the temporal gradient process, $\mathbf{Z}'$, simply requires substituting $T(D - \alpha W)^{-1}T$ for $\sigma^2(D - \alpha W)^{-1}$ in the covariance structure for $\mathbf{Z}$. The result is $\Sigma_Z = R(\phi) \otimes T(D - \alpha W)^{-1}T$ and the expressions for the first and second order derivatives for the matrix-valued covariance function can be derived as

$$K_Z'(\Delta) = -\phi^2 \Delta \exp(-\phi|\Delta|) \left(T(D - \alpha W)^{-1}T\right) \tag{3.9}$$

and $K_Z''(0) = -\phi^2 \left(T(D - \alpha W)^{-1}T\right)$, respectively, where Δ is some temporal distance, say $(t_{j'} - t_j)$. Then the conditional distribution for the gradients, $\mathbf{Z}'$, are found to be multivariate normal with mean and variance-covariance matrix given by

$$\begin{aligned} \boldsymbol{\mu}_{Z' \mid Z,\theta} &= \mathrm{cov}(\mathbf{Z}'(t_0), \mathbf{Z})\mathrm{var}(\mathbf{Z})^{-1}\mathbf{Z} = -(K_Z')^T\Sigma_Z^{-1}\mathbf{Z} \\ \text{and } \Sigma_{Z' \mid Z,\theta} &= -K_Z''(0) - (K_Z')^T\Sigma_Z^{-1}(K_Z') \ , \end{aligned}$$

where $(K_Z')^T$ is an $N_s \times N_sN_t$ block matrix whose j-th block is given by the $N_s \times N_s$ matrix $K_Z'(\Delta_{0j})$, with $\Delta_{0j} = t_j - t_0$. Note that Σ_Z is an $N_sN_t \times N_sN_t$ matrix, but we can use the properties of the MRF to restrict to inverting only $N_t \times N_t$ matrices.

3.2.2 Diagnostic for Determining Spatial Outliers

Another approach for identifying potentially important covariates missing from our model is through spatial outliers, i.e., spatial regions which are markedly different from their neighbors. While many methods focus on the nature of the response surface itself, such as through the use of a Bayesian p-value (Gelman et al., 1996) or via leave-one-out analyses (e.g., Stern and Cressie, 2000), these are typically designed to assess the fit of the entire model. Furthermore, a leave-one-out analysis for our model would depend heavily on the prior specifications for σ_i and τ_i^2, and is thus impractical here. Consequently, our work is more in line with that of Lu and Carlin (2005), who developed areal wombling methods for regional boundary analysis. In our case, geographical features such as mountains are known to affect factors such as air quality, and access to preventive care may also differ between counties.

Here, we have devised a diagnostic which can be used to identify regions that are potential spatial outliers using the model for $\mathbf{Z}$ in (3.5). Extending the expressions listed in Table 3.1 from a single time point to the general space-time case, we find

$$\mathbf{Z}_i \mid \mathbf{Z}_{(i)} \sim N_{N_t}\left(\alpha\frac{\sigma_i}{n_i}\sum_{k\sim i}\frac{\mathbf{Z}_k}{\sigma_k},\ \frac{\sigma_i^2}{n_i}R(\phi)\right). \tag{3.10}$$

Using this, we have constructed the following diagnostic:

$$\begin{aligned} Q_i(\mathbf{Z}_i \mid \mathbf{Z}_{(i)}, \boldsymbol{\theta}, \mathbf{y}) = &\left(\frac{\sqrt{n_i}}{\sigma_i}\mathbf{Z}_i - \alpha\sqrt{n_i}\frac{\sigma_i}{n_i}\sum_{k\sim i}\frac{\mathbf{Z}_k}{\sigma_k}\right)^T \\ &R(\phi)^{-1}\left(\frac{\sqrt{n_i}}{\sigma_i}\mathbf{Z}_i - \alpha\sqrt{n_i}\frac{\sigma_i}{n_i}\sum_{k\sim i}\frac{\mathbf{Z}_k}{\sigma_k}\right). \end{aligned} \tag{3.11}$$

For the sake of simplicity, we suppress the conditional notation and denote (3.11) as simply $Q_i(\mathbf{Z}_i)$. In essence, $Q_i(\mathbf{Z}_i)$ is a measure of posterior learning, as large values correspond to large departures from the conditional mean of $\mathbf{Z}_i$ given by the prior. In this case, however, our prior distribution for $\mathbf{Z}$ assumes a particular degree of spatial smoothing, so large values of $Q_i(\mathbf{Z}_i)$ also indicate potential outlying regions. In practice, we obtain $Q_i(\mathbf{Z}_i^{(\ell)})$ from the ℓth iteration of the sampler and, for repeated ℓ, obtain an estimated posterior distribution for $Q_i(\mathbf{Z}_i)$.

3.3 Simulation studies

To verify the effectiveness of our model, we have conducted two simulation studies. The first is the example presented in (3.3), as this example provided motivation for this work. Here, the goal is to demonstrate the ability of our model to accurately capture the underlying process, $\mathbf{Z}$, and its gradient process, $\mathbf{Z}'$. Then, using the posterior estimates for σ_i, τ_i^2, and ϕ as our true values, we generate data for our second example directly from the model in order to ensure that our parameter estimates are accurate and meaningful. Both simulation studies are illustrated here using the $N_s = 58$ counties of California as our spatial grid, $N_t = 50$ evenly spaced timepoints, and 100 datasets. Each dataset is analyzed using an intercept-only model with a single chain, run for 3,000 iterations.

3.3.1 Estimating Z and Z′

As with the simulation study which motivated this work, we fit our model to 100 datasets generated from (3.3) with τ_i^2 generated from an inverse gamma distribution with shape 3 and rate 2. Figure 3.2 displays the posterior bands for $\mathbf{Z}$ for Alameda and Lassen Counties compared to the true underlying curve (black line) and the mean of the adjacent regions' true underlying curve (red line). Here, we see that the HCAR is able to accurately capture the extreme values in Alameda County, a stark contrast to Figure 3.1, and is also able to overcome the dissimilar temporal pattern seen in the neighbors of Lassen County and provide an equally good fit to the true curve.

Using the county indication key in the right panel of Figure 3.3, we can see that the map of $\widehat{\sigma}_i$ in the left panel reinforces these findings, as both Alameda and Lassen County have larger than average spatiotemporal variance parameters. There also appears to be a spatial association structure to the σ_i, particularly in the San Francisco Bay area near Alameda County. In fact, the σ_i are highly associated with the absolute values of the standardized % black levels that helped generate the data (as well as the $\max|\mathbf{Z}_i|$), suggesting that the σ_i themselves can be used to identify important missing covariates. It's worth noting that, in natural phenomena, $\mathbf{Z}_i$ is unlikely to display the perfect periodicity seen in Figure 3.2, so this strong a relationship may not be typical. These σ_i may also be somewhat inflated, however, as 18% of our τ_i^2 parameters were significantly

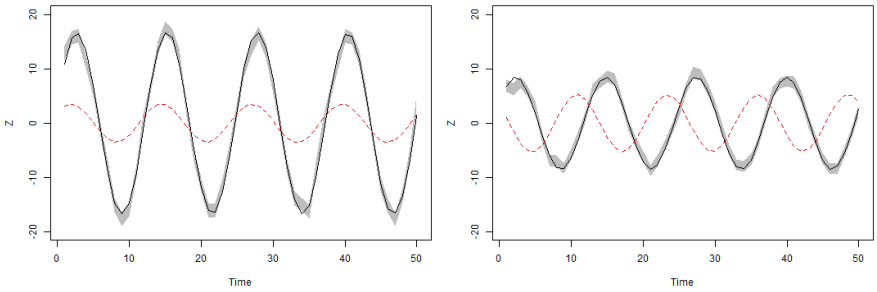
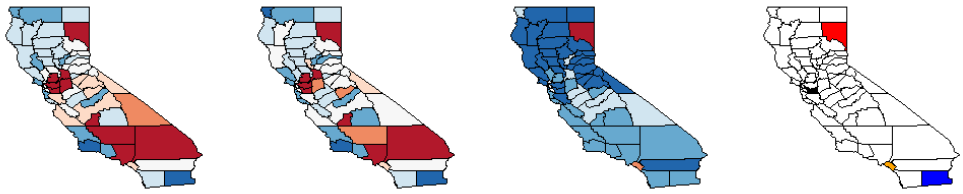


Figure 3.2: 95% CI (gray bands) of $\mathbf{Z}$ for Alameda (left) and Lassen (right) Counties from the simulation study. Solid black lines denote the true underlying curves and dashed red lines denote the mean of each county's neighboring regions. Compared to the fit in Figure 3.1, Alameda County has vastly better fit, and Lassen County has slightly narrower credible bands.

underestimated (true values lying outside their estimated 95% CI), suggesting that some of the variability that should have been attributed to τ_i^2 may have been absorbed by our spatiotemporal process. These biases in τ_i^2 primarily affect regions with larger true τ_i^2, and there does not appear to be any relationship between the coverage of the τ_i^2 and posterior estimates of σ_i. It may be beneficial to re-specify the prior distributions on σ_i and τ_i^2 to see how sensitive our results are to the priors.

Overall, our model performed quite well, with 92.8% of the $Z_i(t_j)$ from the underlying curves being covered by their respective 95% credible intervals. This coverage is comparable to that from the model in equation (3.2), but upon closer inspection, one finds that counties like Alameda now have 85% coverage, a substantial improvement from the previous model, which covered only 32% of the true values. Furthermore, estimating region-specific σ_i provides narrower credible bands for the $\mathbf{Z}_i$, leading to more precise results. Overall, the quality of the estimation is associated with both σ_i and $\max|\mathbf{Z}_i|$, with larger values having poorer estimation, though due to the correlation between σ_i and $\max|\mathbf{Z}_i|$, the cause of this is unclear. One likely reason for this association is that, while the model has been constructed to capture extreme values in $\mathbf{Z}_i$, the magnitude of these extremes is still somewhat underestimated. As seen in previous work, the temporal gradients are well estimated, although a large amount of variability exists,



(a) $\widehat{\sigma}_i$, Sim. #1 (b) $\widehat{\sigma}_i$, Sim. #2 (c) $Q_i(\mathbf{Z}_i)$ (d) Reference Map

Figure 3.3: The first two panels display estimated variance parameters, $\widehat{\sigma}_i$, from the simulation studies. Simulation study #1 is based on the sinusoidal curves constructed in (3.3), and simulation study #2 is constructed from our model, (3.5), using the parameter estimates from the first study. Color cutoffs are (3,4,5,6,7,8), ranging from dark blue to dark red. The third panel displays the estimated $Q_i(\mathbf{Z}_i)$ from the first simulation study, where the color cutoffs used are (50,60,70,80,90,100). The fourth panel is a reference for identifying key counties mentioned in the text. Here, Alameda County is shaded black, Imperial County is shaded blue, Lassen County is shaded red, and Orange County is shaded orange.

leading to extremely high coverage. This provides further evidence that the biases in τ_i may have inflated the estimated variability of our spatiotemporal process, and thus our temporal gradient processes.

As the data in our simulations were generated using the counties' percent black and ozone levels—neither of which are included as predictors in our analysis—evidence suggesting that a county is an outlier should be directly related to one (or both) of these factors. Since the ozone levels were collected at the air basin level, their contribution to the $\mathbf{Z}_i$ is spatially quite smooth, so a county's percent black tends to control the degree of spatial smoothness relative to neighboring counties. Estimation of Q_i, which are mapped in the third panel of Figure 3.3, reveals two significant outliers, Lassen and Orange County, both counties with racial demographics which differ markedly from those of their neighbors. While Lassen County has a high percentage of African Americans, its neighbors do not; similarly, blacks make up only 2.2% of Orange County, 1/3 of the rate in the adjacent counties. On the other hand, counties such as Alameda, which has a

large σ_i but is also surrounded by equally large σ_i, have relatively small Q_i, indicating that they are *not* necessarily spatial outliers. One point worth noting is that the model in (3.2) provides similar coverage for these outlying counties, as the estimated σ^2 was large enough for them to overcome the effect of their neighbors.

3.3.2 Parameter Estimation

Our second simulation study focuses on parameter estimation, that is, verifying that our estimates of σ_i are reasonable. We generated data from

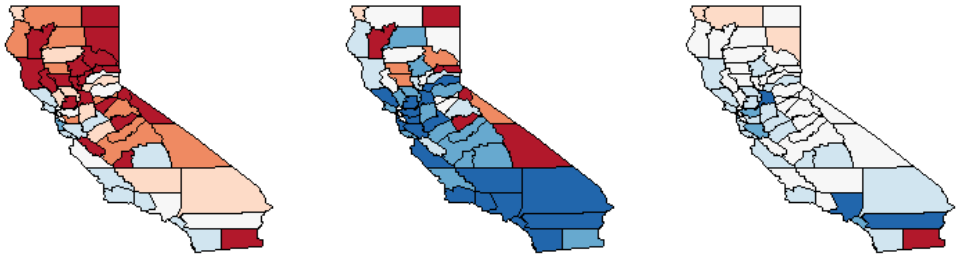
$$Y_i(t_j) = \beta_0 + Z_i(t_j) + \epsilon_i(t_j),\ i = 1, \ldots, N_s,\ j = 1, \ldots, N_t$$

where $\mathbf{Z} \sim N(\mathbf{0}, R(\phi) \otimes T(D - \alpha W)^{-1}T)$ and $\epsilon_i(t_j) \sim N(0, \tau_i^2)$, using the posterior medians of σ_i, τ_i, α, and ϕ from Section 3.3.1. Here, our model performs substantially better, with 94.8% of our $Z_i(t_j)$ and 96% of our τ_i covered by their respective 95% credible intervals. As demonstrated in Chapter 2, the temporal range parameter, ϕ, and spatial association parameter, α, remain difficult to estimate, as, on average, ϕ is overestimated by 14% and α is underestimated by 4%. As shown in the second panel of Figure 3.3, the spatiotemporal variance parameters, σ_i, are well estimated, as they agree with those in the first panel.

Given that these data sets were generated directly from our model, it is not surprising that none of our estimated Q_i indicate spatial outliers in this data. Throughout all of our datasets, the largest estimate of $Q_i(\mathbf{Z}_i) = 65.28$ (47.34, 88.29), which does not differ significantly from the distribution of the remaining counties, which has a median (95% CI) of 52.05 (34.07, 75.60).

3.4 Analysis of the California asthma hospitalization data

As mentioned at the beginning of this chapter, in this section we reanalyze the asthma hospitalization rates using the four covariates (ozone level, population density, percent under the age of 18, and percent black) and fixed effects for each of the 12 months of the year while also allowing each region to have its own variance parameters. Convergence of the MCMC algorithm was rapid, but we ran our sampler for 20,000 iterations to ensure the stability of our estimates. Using DIC, our HCAR model improves upon that



(a) σ_i, spatiotemporal (b) τ_i, random error (c) $Q_i(\mathbf{Z}_i)$

Figure 3.4: The first two panels display the estimated variance parameters from the asthma hospitalization data. Color cutoffs are (1,2,3,4,5,6) for σ_i and τ_i, ranging from dark blue to dark red. The third panel displays our estimated outlier diagnostic, Q_i, with color cutoffs (175,200,225,250,275,300).

of (3.2) by 3000 units despite being more complex, with p_D indicating that our model contains 1111 more effective parameters (6036 compared to 4925).

In Table 3.2, we present the posterior estimates for our model parameters $\boldsymbol{\beta}$ and ϕ. When compared to the results using the CAR model, we find that ϕ has increased 37%, indicating less temporal association in our model, though this may be due to an increase in spatial association, as α has also increased 14%. A number of our regression coefficients have also changed. For instance, the effect for percent black has increased 43%, while the monthly effects for the summer have reduced in magnitude. With regard to the monthly effects, it seems as though our random effects have absorbed much of this temporal change, which will become more apparent when we discuss our temporal gradient analysis.

The left and center panels of Figure 3.4 display maps of our estimated $HCAR_{ST}$ variance parameters, σ_i and τ_i. As shown in our first simulation study, large values of σ_i can indicate regions with high within-county variability (e.g., Alameda County) or high between-county variability (e.g., Lassen County). In this case, the cluster of large σ_i in the northern counties likely indicates the latter explanation. One can reach this conclusion by noticing the lack of spatial smoothness in the raw data (see Figure 1.1),

or by mapping $SD(\mathbf{Z}_i)$ for each region, where $SD(\mathbf{Z}_i)$ denotes the empirical standard deviation of the posterior medians of $\mathbf{Z}_i$. The error variance parameters, τ_i, are strongly negatively correlated with population (i.e., counties with low population have a higher error variance), which coincides with the variance of the rate estimate from a normal approximation to the binomial distribution.

The right panel of Figure 3.4 displays the estimated values of our outlier diagnostic, Q_i. Here, the darker the shade of red, the stronger the evidence that region i is an outlier. This figure highlights Imperial County (extreme lower right), whose $Q_i(\mathbf{Z}_i)$ has median (95% CI) $Q_i = 334$ (291, 382). This is significantly larger than the distribution of the Q_i from the remaining counties, which has a median (95% CI) of 224.09 (162.82, 281.27), indicating that this county is quite different than its neighbors. Giving credence to this result is a recent article in the *LA Times* focused on Imperial County:

> "Imperial County is different because it leads the state for asthmatic children going to the ER and being hospitalized, but experts are unable to pinpoint the cause. Doctors and public health officials said that a combination of whipping winds, pesticide-tinged farmland dust and large numbers of low-income families lacking health insurance contribute to high rates of asthma hospitalizations and ER visits." – *Gorman (2012)*

While this article proposes a number of possible explanations for Imperial County's abnormally high rates of asthma hospitalization, our analysis can help public health officials narrow this list even further by identifying risk factors in Imperial County that are not applicable to its neighbors. For instance, its geographical location is such that the two counties it neighbors — San Diego and Riverside — have their most densely populated areas near the Pacific Ocean, rather than in the desert valley. Also, as mentioned in that article, its close proximity to an industrial portion of northern Mexico may be magnifying these effects.

Turning our attention to temporal gradients, Figure 3.5 shows striking results. Previous research has shown that asthma hospitalization rates are lowest during the summer and highest during the winter months. Here, we map the average month-to-month temporal gradients. These results indicate that hospitalization rates achieve their apex during the late fall and early winter, decline for most of the spring, remain steady through July and August, and then increase heavily during August and September, a cycle that may coincide with agricultural activity in some of the more rural areas of

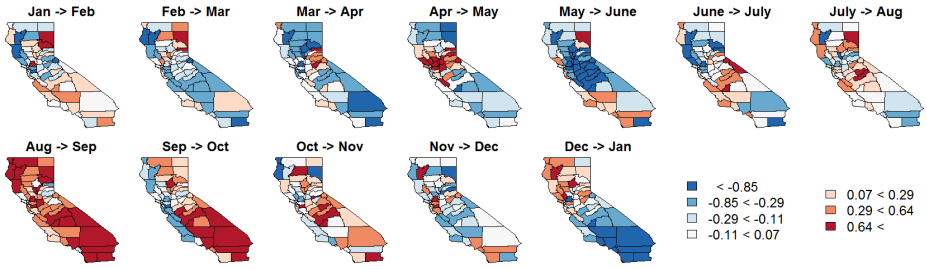


Figure 3.5: Average month-to-month temporal gradients from the California asthma hospitalization data.

the state, particularly those in the Imperial and San Joaquin Valleys, where dry, windy conditions combine with a "bowl effect" caused by the surrounding mountains can lead to prolonged periods of poor air quality. Compared to the results using the CAR model in equation (3.2) proposed in Chapter 2, our results contain more extreme estimated gradients. This is to be expected as: (a) the CAR model led to oversmoothing and thus more gradual rates of change and (b) our spatiotemporal process has absorbed more of the seasonal trend previously contained in the monthly fixed effects.

3.5 Discussion

In this chapter, we have proposed a novel heterogenous CAR method for permitting region-specific variance parameters in a spatiotemporal process, motivated by efforts to better analyze the California asthma hospitalization data. We also examined the validity of our model via simulation, and illustrated its use by reanalyzing the California dataset. Not only did our model produce a better fit to the real data and dramatically alter the estimated effect of race, it was also able to correctly single out a known outlying region. Coupled with improved gradient estimation (by virtue of the improved fit of $\mathbf{Z}$), these results demonstrate the ability of our model to highlight features of spatiotemporal data that can be used by researchers to identify important predictors and risk factors missing from their statistical model.

One point of discussion is the interpretation of the σ_i. In our first simulation study, we discovered that large values of σ_i indicate one of two possibilities: (1) the process $\mathbf{Z}_i$

has a large temporal variance or (2) the process $\mathbf{Z}_i$ is a potential spatial outlier. Because of this, one must be careful not to interpret σ_i^2 as the variance of the temporal process $\mathbf{Z}_i$. A more accurate interpretation of σ_i is as a scale parameter. For instance, we can use the σ_i to construct a spatiotemporal process $\mathbf{U} = (I_{N_t} \otimes T^{-1})\mathbf{Z}$ with unit variance, where $U_i(t_j) = Z_i(t_j)/\sigma_i$, suggesting that our model smooths $\mathbf{Z}_i/\sigma_i$ rather than directly smoothing $\mathbf{Z}_i$. Similarly, these two roles of σ_i suggest that the σ_i themselves are likely to be spatially associated with outlying values corresponding to outlying $\mathbf{Z}_i$. As a result, in Section 3.3 we chose to use posterior estimates from our first simulation to generate data for our second simulation. We also considered using true σ_{i0} generated randomly (say, from their prior distributions), but we decided that this was unnecessary.

One drawback of our outlier diagnostic, $Q_i(\mathbf{Z}_i)$, is the lack of a convenient reference distribution. Here, we were fortunate to have obvious outlying regions in our first simulation study and our data analysis. In general, it may suffice to compare the distribution of the max $Q_i(\mathbf{Z}_i)$ to the distribution of the remaining Q_j. It's also worth noting that the Q_i we propose can be estimated for any CAR-based distribution. Calculating Q_i under the CAR model in (3.2) gives similar results, though the constraint of having a single σ^2 parameter can potentially over- and undersmooth the $\mathbf{Z}_i$, which in this case leads to more extreme values of Q_i on both ends of the spectrum.

As illustrated here, one strength of the modeling framework in proposed in Chapter 2 is its flexibility, which itself is related to the third case of order-free multivariate CAR distributions in Jin et al. (2007). An obvious generalization of our work would be to replace $R(\phi)$ with other temporal correlation structures, as many applications will not require mean-square differentiable temporal processes. In this chapter, we extended their approach to allow for region-specific variance parameters in situations with a large number of temporal observations. Should there be a limited number of temporal observations, however, our methods here could easily be adapted to ensure the stability of our σ_i estimates; for instance, we could induce spatial association in the σ_i as in Reich and Hodges (2008).

Parameter	Median (95% CI) $CAR(\sigma^2)$	$HCAR_{ST}(\boldsymbol{\sigma})$
β_0 (Intercept)	9.17 (8.93, 9.42)	8.98 (8.79, 9.22)
β_1 (Pop Den)	0.60 (0.49, 0.70)	0.62 (0.55, 0.69)
β_2 (Ozone)	-0.18 (-0.28, -0.08)	-0.16 (-0.25, -0.07)
β_3 (% Black)	1.24 (1.15, 1.34)	1.78 (1.68, 1.86)
β_4 (% Under 18)	1.12 (1.01, 1.24)	1.24 (1.16, 1.32)
β_5 (February)	-0.25 (-0.46, -0.04)	-0.27 (-0.49, -0.08)
β_6 (March)	-0.21 (-0.48, 0.07)	-0.05 (-0.28, 0.15)
β_7 (April)	-1.47 (-1.81, -1.12)	-1.18 (-1.46, -0.96)
β_8 (May)	-1.17 (-1.53, -0.80)	-0.92 (-1.20, -0.64)
β_9 (June)	-2.79 (-3.21, -2.40)	-2.05 (-2.33, -1.74)
β_{10} (July)	-3.78 (-4.21, -3.37)	-2.78 (-3.14, -2.47)
β_{11} (August)	-3.58 (-4.02, -3.13)	-2.65 (-3.08, -2.26)
β_{12} (September)	-1.96 (-2.37, -1.54)	-1.47 (-1.93, -1.13)
β_{13} (October)	-1.36 (-1.73, -1.00)	-1.21 (-1.57, -0.88)
β_{14} (November)	-0.71 (-1.02, -0.42)	-0.74 (-1.07, -0.42)
β_{15} (December)	0.63 (0.41, 0.86)	0.57 (0.32, 0.81)
ϕ	0.90 (0.84, 0.97)	1.24 (1.19, 1.29)
α	0.77 (0.71, 0.80)	0.88 (0.86, 0.90)

Table 3.2: Posterior medians and 95% credible intervals (CI) for $\boldsymbol{\beta}$, ϕ, and α from our asthma hospitalization rate data.

Chapter 4

On Bayesian Inference for Rates of Change in Spatiotemporal Process Settings

In this chapter, our methodological innovations entail developments that extend work in purely spatial contexts by Banerjee et al. (2003) to spatiotemporal settings and that extend our work to continuous space, continuous time inferential settings. In Section 4.1 we outline the notion of a spatiotemporal gradient process, followed by the distribution theory for such processes in Section 4.2. Section 4.3 embeds this distribution theory within a Bayesian modeling framework and describes how to carry out posterior predictive inference on spatiotemporal gradients. Section 4.4 presents a simulated data example which aims to verify our proposed theory on gradients. In Section 4.5 we analyze air quality data (PM2.5, particulate matter less than 2.5 micrometers in diameter) from the Air Resources Board of the California Environmental Protection Agency, where we use the spatiotemporal gradient process to identify the impact of mountain ranges on pollution. Finally, Section 4.6 concludes the chapter with a discussion and an eye toward future work.

4.1 Calculus of spatiotemporal gradients

We begin by extending notions of spatial random fields in Adler (2009), Mardia et al. (1996) and Banerjee et al. (2003) to offer a brief overview of the calculus of spatiotemporal gradient processes. Let $Z(\mathbf{s},t)$ be a (weakly) stationary real-valued spatiotemporal process on $\Re^d \times \Re$ with mean 0 and finite second moment and covariance function $\mathrm{Cov}\{Z(\mathbf{s},t), Z(\mathbf{s}',t')\} = K(\boldsymbol{\Delta}_s, \delta_t)$, where $\boldsymbol{\Delta}_s = \mathbf{s}' - \mathbf{s}$ and $\delta_t = t' - t$ for every pair of space-time coordinates $(\mathbf{s},t)$ and $(\mathbf{s}',t')$ in $\Re^d \times R$.

Its restrictions $K(\boldsymbol{\Delta}_s, 0)$ and $K(\mathbf{0}, \delta_t)$ are purely spatial and purely temporal covariance functions, respectively. The function $K(\boldsymbol{\Delta}_s, \delta_t)$ must satisfy the condition of *positive-definiteness* (e.g. Gneiting, 2002) described by the following requirement

$$\mathrm{Var}\left\{\sum_{i=1}^{k} a_i Z(\mathbf{s}_i, t_i)\right\} = \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j K(\mathbf{s}_i - \mathbf{s}_j, t_i - t_j) \geq 0 \tag{4.1}$$

for every finite collection of space-time coordinates $(\mathbf{s}_i, t_i) \in \Re^d \times \Re$ and $a_i \in \Re$, $i = 1, 2, \ldots, k$. Also, to avoid degeneracy, the double sum in (4.1) is constrained to be zero if and only if each $a_i = 0$. This condition is necessary and sufficient for $K(\mathbf{s},t)$ to be a valid stationary covariance function (Bochner, 1955), as it ensures a valid joint probability distribution for every finite collection of space-time coordinates.

The process $Z(\mathbf{s},t)$ is said to be L_2 (or mean-squared) continuous at $(\mathbf{s},t)$ if $E\{Z(\mathbf{s}+\boldsymbol{\Delta}_s, t+\Delta_t) - Z(\mathbf{s},t)\}^2 \to 0$ as $(\boldsymbol{\Delta}_s, \Delta_t) \to 0$. Analogous to the concept of total differentiability for nonstochastic functions, we say that $Z(\mathbf{s},t)$ is mean-square (totally) differentiable at $(\mathbf{s},t)$ if there exists a $d \times 1$ process $\boldsymbol{\nabla}_s Z(\mathbf{s},t)$ and a real valued process $\nabla_t Z(\mathbf{s},t)$ such that for any given vector $\mathbf{u} \in \Re^d$ and any real number v,

$$Z(\mathbf{s}+h\mathbf{u}, t+hv) = Z(\mathbf{s},t) + h\mathbf{u}^T\boldsymbol{\nabla}_s Z(\mathbf{s},t) + hv\nabla_t Z(\mathbf{s},t) + o(h) \;\text{ as }\; h \to 0 \tag{4.2}$$

in the L_2 sense for any scalar h. (Note that we could equivalently write $k = hv$ and let both h and k go to 0 in (4.2) but this is not necessary for our subsequent development.) Therefore, mean square differentiability of $Z(\mathbf{s},t)$ requires that, for any $\mathbf{u} \in \Re^d$ and any real number v,

$$\lim_{h\to 0} E\left(\frac{Z(\mathbf{s}+h\mathbf{u}, t+h) - Z(\mathbf{s},t)}{h} - \mathbf{u}^T\boldsymbol{\nabla}_s Z(\mathbf{s},t) - v\nabla_t Z(\mathbf{s},t)\right)^2 = 0\,. \tag{4.3}$$

We refer to $Z(\mathbf{s},t)$ as the *parent process* from which the spatial and temporal gradient processes $\boldsymbol{\nabla}_s Z(\mathbf{s},t)$ and $\nabla_t Z(\mathbf{s},t)$ are derived. For spatiotemporal gradient processes, we will consider processes associated with spatial directional derivatives, temporal derivatives, and mixed derivatives. The first two emerge from (4.2) and (4.3) as special cases, while the mixed gradient is obtained by taking iterated L_2 limits. We discuss these gradients below.

If we fix $v = 0$ in (4.2) then (4.3) supplies the spatial directional derivative process as the real-valued process obtained from the L_2 limit

$$D_{\mathbf{u}} Z(\mathbf{s},t) = \lim_{h \to 0} \frac{Z(\mathbf{s}+h\mathbf{u},t) - Z(\mathbf{s},t)}{h} = \mathbf{u}^T \boldsymbol{\nabla}_s Z(\mathbf{s},t) \; . \tag{4.4}$$

More generally, if we collect a set of p directions into a $d \times p$ matrix $\mathbf{U} = [\mathbf{u}_1 : \mathbf{u}_2 : \ldots : \mathbf{u}_p]$, we then obtain a $p \times 1$ process $\mathbf{D}_{\mathbf{U}} Z(\mathbf{s},t) = (D_{\mathbf{u}_1} Z(\mathbf{s},t), D_{\mathbf{u}_2} Z(\mathbf{s},t), \ldots, D_{\mathbf{u}_p} Z(\mathbf{s},t))^T$. From (4.4), we see that $D_{\mathbf{U}} Z(\mathbf{s},t) = \mathbf{U}^T \boldsymbol{\nabla}_s Z(\mathbf{s},t)$. Explicit representations of $\boldsymbol{\nabla}_s Z(\mathbf{s},t)$ emerge from specific bases. For example, if we take the standard (or canonical) Euclidean basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_d\}$ as our set of directions, where $\mathbf{e}_i$ has a 1 as its i-th coordinate and 0 everywhere else, then $p = d$ and $\mathbf{U} = \mathbf{I}_d$. Writing $D_{\mathbf{e}_i} Z(\mathbf{s},t) = (\partial/\partial s_i) Z(\mathbf{s},t)$, we obtain

$$\boldsymbol{\nabla}_s Z(\mathbf{s},t) = D_{\mathbf{I}_d} Z(\mathbf{s},t) = \left(\frac{\partial}{\partial s_1} Z(\mathbf{s},t), \frac{\partial}{\partial s_2} Z(\mathbf{s},t), \ldots, \frac{\partial}{\partial s_d} Z(\mathbf{s},t) \right)^T \; , \tag{4.5}$$

where $\mathbf{s} = \sum_{i=1}^d s_i \mathbf{e}_i$, so the s_i's are the coordinates of $\mathbf{s}$ with respect to the canonical basis. In the remainder of this article we restrict to canonical basis representations and refer to $\boldsymbol{\nabla}_s Z(\mathbf{s},t)$ in (4.5) as the *spatial gradient process*. Since $D_{\mathbf{U}} Z(\mathbf{s},t) = \mathbf{U}^T \boldsymbol{\nabla}_s Z(\mathbf{s},t)$, the directional derivative process along a set of arbitrary directions is a linear transformation of the spatial gradient process, and inference for $D_{\mathbf{U}} Z(\mathbf{s},t)$ proceeds immediately from inference for $\boldsymbol{\nabla}_s Z(\mathbf{s},t)$. Finally, when inferring about $D_{\mathbf{u}} Z(\mathbf{s},t)$ it suffices to consider $\mathbf{u}$ to be a unit vector (i.e., $\|\mathbf{u}\| = 1$). Inference on any arbitrary vector $\mathbf{w}$ amounts to simply scaling by the length of $\mathbf{w}$ since the linearity in (4.4) implies that $D_{\mathbf{w}} Z(\mathbf{s},t) = \|\mathbf{w}\| D_{\mathbf{u}} Z(\mathbf{s},t)$.

Purely temporal derivatives are obtained when we fix $\mathbf{u} = \mathbf{0}$ and $v = 1$ in (4.2). Then, (4.3) provides us with the temporal gradient process

$$\nabla_t Z(\mathbf{s},t) = \frac{\partial}{\partial t} Z(\mathbf{s},t) = \lim_{h \to 0} \frac{Z(\mathbf{s},t+h) - Z(\mathbf{s},t)}{h} \; , \tag{4.6}$$

where the limit is taken in the L_2 sense. Analogous to the calculus of nonstochastic functions, we must require $Z(\mathbf{s},t)$ to mean-square total differentiabile as in (4.2), to ensure that the directional gradient processes exist in every direction. Existence of the partial derivative processes in (4.5) and (4.6) do not ensure mean square continuity, and hence differentiability, of $Z(\mathbf{s},t)$ along every direction.

Since the processes $\boldsymbol{\nabla}_s Z(\mathbf{s},t)$ and $\nabla_t Z(\mathbf{s},t)$ are well-defined and have explicit canonical representations with respect to partial derivatives, it is natural to consider gradient processes arising from mixed derivatives. For example, we can construct the the $d \times 1$ process $\nabla_t \boldsymbol{\nabla}_s Z(\mathbf{s},t)$ whose coordinates are given by mixed derivative in the L_2 limit

$$\lim_{h\to 0} \frac{D_{\mathbf{e}_i} Z(\mathbf{s},t+h) - D_{\mathbf{e}_i} Z(\mathbf{s},t)}{h} = \frac{\partial^2}{\partial t \partial s_i} Z(\mathbf{s},t) \;\text{ for }\; i = 1,2,\ldots,d\,.$$

This is the temporal derivative of the spatial gradient, reflecting the continuous rate of change in spatial gradients over time. Alternatively, we could first take the temporal derivative followed by the spatial gradient, which yields the $d \times 1$ process $\boldsymbol{\nabla}_s \nabla_t Z(\mathbf{s},t)$ with coordinates

$$\lim_{h\to 0} \frac{\nabla_t Z(\mathbf{s}+h\mathbf{e}_i,t) - \nabla_t Z(\mathbf{s},t)}{h} = \frac{\partial^2}{\partial s_i \partial t} Z(\mathbf{s},t) \;\text{ for }\; i = 1,2,\ldots,d\,.$$

Assuming that all the mixed second order partial derivatives are mean-square continuous at every space-time coordinate $(\mathbf{s},t)$, we can legitimately change the order of the derivatives so that $\nabla_t \boldsymbol{\nabla}_s Z(\mathbf{s},t) = \boldsymbol{\nabla}_s \nabla_t Z(\mathbf{s},t)$ almost surely for every space-time coordinate. Hence, we can unambiguously define the $d \times 1$ mixed derivative process as

$$\boldsymbol{\nabla}_{s,t} Z(\mathbf{s},t) = \boldsymbol{\nabla}_s \nabla_t Z(\mathbf{s},t) = \nabla_t \boldsymbol{\nabla}_s Z(\mathbf{s},t)\,. \tag{4.7}$$

4.2 Distribution theory for spatiotemporal gradients

Let $Z(\mathbf{s},t)$ be a univariate stationary zero-mean Gaussian random field $GP(0, K(\cdot,\cdot;\boldsymbol{\theta}))$, where $\boldsymbol{\theta}$ is a collection of process parameters. For notational convenience, we suppress the dependence on $\boldsymbol{\theta}$ and simply write $K(\cdot,\cdot)$ for the stationary covariance function defined on $\Re^d \times \Re$. Based upon observations from a finite set of locations, say $\mathcal{S} = \{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{N_s}\}$, and timepoints $\mathcal{T} = \{t_1, t_2, \ldots, t_{N_t}\}$, we wish to predict the gradients outlined in Section 4.1 at an arbitrary space-time coordinate $(\mathbf{s},t)$.

This suggests developing a multivariate process comprising the parent $Z(\mathbf{s},t)$ and its derivative processes. To be precise, let $\mathbf{W}(\mathbf{s},t) = (\mathbf{W}_1(\mathbf{s},t)^T, (\nabla_t \mathbf{W}_1(\mathbf{s},t))^T)^T$ be the $2(d+1)\times 1$ process, where $\mathbf{W}_1(\mathbf{s},t) = (Z(\mathbf{s},t), (\boldsymbol{\nabla}_s Z(\mathbf{s},t))^T)^T$ and $\nabla_t \mathbf{W}_1(\mathbf{s},t) = (\nabla_t Z(\mathbf{s},t), (\nabla_t \boldsymbol{\nabla}_s Z(\mathbf{s},t))^T)^T$ are each $(d+1)\times 1$. Its cross-covariance matrix $\mathbf{C}_{\mathbf{W}}(\boldsymbol{\Delta},\delta)$ is $2(d+1)\times 2(d+1)$ partitioned as

$$\begin{bmatrix} \mathrm{Cov}\{\mathbf{W}_1(\mathbf{s},t), \mathbf{W}_1(\mathbf{s}+\boldsymbol{\Delta}, t+\delta)\} & \mathrm{Cov}\{\mathbf{W}_1(\mathbf{s},t), \nabla_t \mathbf{W}_1(\mathbf{s}+\boldsymbol{\Delta}, t+\delta)\} \\ \mathrm{Cov}\{\nabla_t \mathbf{W}_1(\mathbf{s},t), \mathbf{W}_1(\mathbf{s}+\boldsymbol{\Delta}, t+\delta)\} & \mathrm{Cov}\{\nabla_t \mathbf{W}_1(\mathbf{s},t), \nabla_t \mathbf{W}_1(\mathbf{s}+\boldsymbol{\Delta}, t+\delta)\}\} \end{bmatrix}, \tag{4.8}$$

where each of the above blocks is $(d+1)\times(d+1)$. The matrix in (4.8) need not itself be symmetric or positive definite but must satisfy the following two conditions to be *valid*:

$$\text{(i) } \mathbf{C}_{\mathbf{W}}(\boldsymbol{\Delta},\delta) = \mathbf{C}_{\mathbf{W}}(-\boldsymbol{\Delta},-\delta)^T \text{ and (ii) } \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbf{a}_i^T \mathbf{C}_{\mathbf{W}}(\mathbf{s}_i-\mathbf{s}_j, t_i - t_j)\mathbf{a}_j > \mathbf{0}\ , \tag{4.9}$$

for every $(\boldsymbol{\Delta},\delta) \in \Re^d \times \Re$, $\mathbf{a}_i \in \Re^{2(d+1)} \setminus \{\mathbf{0}\}$, and finite set of space-time coordinates $\{(\mathbf{s}_1,t_1),(\mathbf{s}_2,t_2),\ldots,(\mathbf{s}_n,t_n)\}$. We write (4.8) in terms of the covariance function $K(\boldsymbol{\Delta},\delta)$ and its derivatives by first constructing the corresponding finite-difference process and then passing to limits. Since the finite difference processes arise as linear transformations of the original process, the associated cross-covariance matrices are valid (i.e., they satisfy (4.9)) by construction. This ensures that $\mathbf{C}_{\mathbf{W}}(\boldsymbol{\Delta},\delta)$ in (4.8) is also valid because it arises as limits of the valid finite-difference cross-covariances. Details are provided in Appendix B.

Beginning, then, with the $(1,1)$ block in (4.8), we obtain

$$\mathrm{Cov}\{\mathbf{W}_1(\mathbf{s},t), \mathbf{W}_1(\mathbf{s}+\boldsymbol{\Delta}, t+\delta)\} = \begin{bmatrix} K(\boldsymbol{\Delta},\delta) & (\boldsymbol{\nabla}_s K(\boldsymbol{\Delta},\delta))^T \\ -\boldsymbol{\nabla}_s K(\boldsymbol{\Delta},\delta) & -\mathbf{H}_K(\boldsymbol{\Delta},\delta) \end{bmatrix},$$

where $\boldsymbol{\nabla}_s K(\boldsymbol{\Delta},\delta)$ is the $d\times 1$ spatial gradient of $K(\boldsymbol{\Delta},\delta)$, with elements $(\partial/\partial\Delta_i)K(\boldsymbol{\Delta},\delta)$, and $\mathbf{H}_K(\boldsymbol{\Delta},\delta)$ is the $d\times d$ Hessian with elements $(\partial^2/\partial\Delta_i\partial\Delta_j)K(\boldsymbol{\Delta},\delta)$. Turning to the the $(1,2)$ block in (4.8), similar calculations yield

$$\mathrm{Cov}\{\mathbf{W}_1(\mathbf{s},t), \nabla_t \mathbf{W}_1(\mathbf{s}+\boldsymbol{\Delta}, t+\delta)\} = \begin{bmatrix} \nabla_t K(\boldsymbol{\Delta},\delta) & (\nabla_t \boldsymbol{\nabla}_s K(\boldsymbol{\Delta},\delta))^T \\ -\nabla_t \boldsymbol{\nabla}_s K(\boldsymbol{\Delta},\delta) & -\nabla_t \mathbf{H}_K(\boldsymbol{\Delta},\delta) \end{bmatrix},$$

where $\nabla_t K(\boldsymbol{\Delta}, \delta)$ is the temporal derivative $(\partial/\partial\delta)K(\boldsymbol{\Delta}, \delta)$, $\nabla_t \boldsymbol{\nabla}_s K(\boldsymbol{\Delta}, \delta)$ is the $d \times 1$ vector of mixed derivatives $(\partial^2/\partial\delta\partial\Delta_i)K(\boldsymbol{\Delta}, \delta)$, and $\nabla_t \mathbf{H}_K(\boldsymbol{\Delta}, \delta)$ is the $d \times d$ matrix with elements $(\partial^3/\partial\delta\partial\Delta_i\partial\Delta_j)K(\boldsymbol{\Delta}, \delta)$. The $(2, 1)$ block is obtained from the condition that $\mathbf{C}_{\mathbf{W}}(\boldsymbol{\Delta}, \delta) = \mathbf{C}_{\mathbf{W}}(-\boldsymbol{\Delta}, -\delta)^T$. Hence,

$$\mathrm{Cov}\{\nabla_t \mathbf{W}_1(\mathbf{s}, t), \mathbf{W}_1(\mathbf{s} + \boldsymbol{\Delta}, t + \delta)\} = \begin{bmatrix} \nabla_t K(-\boldsymbol{\Delta}, -\delta) & -(\nabla_t \boldsymbol{\nabla}_s K(-\boldsymbol{\Delta}, -\delta))^T \\ \nabla_t \boldsymbol{\nabla}_s K(-\boldsymbol{\Delta}, -\delta) & -\nabla_t \mathbf{H}_K(-\boldsymbol{\Delta}, -\delta) \end{bmatrix}.$$

Finally, the $(2, 2)$ block is

$$\mathrm{Cov}\{\nabla_t \mathbf{W}_1(\mathbf{s}, t), \nabla_t \mathbf{W}_1(\mathbf{s} + \boldsymbol{\Delta}, t + \delta)\} = \begin{bmatrix} \nabla_t^2 K(\boldsymbol{\Delta}, \delta) & (\nabla_t^2 \boldsymbol{\nabla}_s K(\boldsymbol{\Delta}, \delta))^T \\ -\nabla_t^2 \boldsymbol{\nabla}_s K(\boldsymbol{\Delta}, \delta) & -\nabla_t^2 \mathbf{H}_K(\boldsymbol{\Delta}, \delta) \end{bmatrix},$$

where $\nabla_t^2 K(\boldsymbol{\Delta}, \delta) = (\partial^2/\partial\delta^2)K(\boldsymbol{\Delta}, \delta)$, $\nabla_t^2 \boldsymbol{\nabla}_s K(\boldsymbol{\Delta}, \delta)$ is $d \times 1$ and $\nabla_t^2 \mathbf{H}_K(\boldsymbol{\Delta}, \delta)$ is $d \times d$ with $(\partial^3/\partial\delta^2\partial\Delta_i)K(\boldsymbol{\Delta}, \delta)$ and $(\partial^4/\partial\delta^2\partial\Delta_i\partial\Delta_j)K(\boldsymbol{\Delta}, \delta)$ as their elements respectively.

We now turn to choosing $K(\boldsymbol{\Delta}, \delta)$, which determines the smoothness of the spatiotemporal process realizations, thereby ensuring the existence of the spatiotemporal gradient processes. Earlier work on spatial gradients have, almost invariably, opted for Gaussian processes with the Matérn covariance function (Stein, 1999) as an especially convenient choice for modeling smoothness. When constructing an appropriately smooth spatiotemporal Gaussian process for inferring about gradients, one could conceivably use separable models, which might decompose the spatiotemporal covariance function into the sum or product of a purely spatial and a purely temporal covariance function. One example would be to specify $K(||\boldsymbol{\Delta}||, |\delta|) = K_s(||\boldsymbol{\Delta}||)K_t(|\delta|)$. This appears attractive by not only allowing us to model the smoothness across space and time distinctly from one another, but it also renders computational benefits, as the resulting variance-covariance matrix is a Kronecker product of two lower-dimensional matrices. However, the resulting correlations are sensitive to small perturbations in locations, since they are not smoother away from the origin than they are at the origin (Stein, 2005). Hence, we will restrict our attention to the more complex class of *nonseparable* covariance structures.

Following Gneiting (2002), we construct our covariance function as

$$K(\boldsymbol{\Delta}, \delta) = \frac{\sigma^2}{\psi(|\delta|^2)^{d/2}} \varphi\left(\frac{||\boldsymbol{\Delta}||^2}{\psi(|\delta|^2)}\right),$$

where $\varphi(u),\ u \geq 0$ is a completely monotone function modeling spatial association and $\psi(\delta),\ \delta \geq 0$ is a positive function with a completely monotone derivative modeling temporal association. Letting $\varphi(u)$ be the Matérn correlation function and $\psi(\delta) = (\phi_t^2|\delta|^\alpha + 1)^\beta$ yields

$$K(\boldsymbol{\Delta}, \delta) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)(\phi_t^2|\delta|^{2\alpha}+1)^{\gamma+\beta d/2}} \left(\frac{\phi_s||\boldsymbol{\Delta}||}{(\phi_t^2|\delta|^{2\alpha}+1)^{\beta/2}}\right)^\nu \times K_\nu\left(\frac{\phi_s||\boldsymbol{\Delta}||}{(\phi_s|\delta|^{2\alpha}+1)^{\beta/2}}\right),$$

where $\nu > 0$ and $\alpha \in (0, 1]$ control smoothness with respect to space and time, respectively, ϕ_t and ϕ_s represent correlation decay across time and space, respectively, $\beta \in [0, 1]$ is a space-time interaction term such that $\beta = 0$ corresponds to a separable model, $\gamma \geq 0$ ensures temporal dependence should $\beta = 0$, d is the spatial dimension and K_ν is the modified Bessel function of the second kind. Here, we fix $\beta = 1$ and $\gamma = 0$.

As we require a mean-square differentiable process, we need to choose ν and α in order to achieve the desired level of smoothness. The Matérn correlation function is especially convenient here as it yields sample paths that are $\lceil \nu - 1 \rceil$ times differentiable. However, since $\nu > 2$ leads to oversmoothing, Banerjee et al. (2003) assign a prior on $\nu \in (1, 2]$. This increases the computational burden and, in practice, yields insignificant posterior learning. Hence, we fix $\nu = 3/2$, which also conveniently results in a closed form expression for the Matérn correlation function. Further, letting $\alpha = 1$ renders a mean-square differentiable temporal process yielding a nonseparable and sufficiently smooth covariance function

$$K(\boldsymbol{\Delta}, \delta) = \frac{\sigma^2}{\left(\phi_t^2|\delta|^2+1\right)}\left(1 + \frac{\phi_s||\boldsymbol{\Delta}||}{\left(\phi_t^2|\delta|^2+1\right)^{1/2}}\right)\exp\left[-\frac{\phi_s||\boldsymbol{\Delta}||}{\left(\phi_t^2|\delta|^2+1\right)^{1/2}}\right]. \tag{4.10}$$

Subsequently, in this article, we use (4.10) for our covariance function.

This completes our specifications for $\mathbf{C_W}(\boldsymbol{\Delta}, \delta)$, which, by construction, is a valid cross-covariance as long as the required derivatives of $K(\boldsymbol{\Delta}, \delta)$ exist. Therefore, the collection of random variables $\{\mathbf{W}(\mathbf{s}_i, t_j) : (\mathbf{s}_i, t_j) \in \mathcal{S} \times \mathcal{T}\}$ has a well-defined multivariate normal distribution with a positive-definite $2N_sN_t(d+1) \times 2N_sN_t(d+1)$ variance-covariance matrix, which is an $N_sN_t \times N_sN_t$ block matrix and the block associated with the pair of space-time coordinates $(\mathbf{s}_i, t_k)$ and $(\mathbf{s}_j, t_l)$ is given by the $2(d+1) \times 2(d+1)$ matrix $\mathbf{C_W}(\mathbf{s}_j - \mathbf{s}_i, t_l - t_k)$.

For our subsequent inferential development will be particularly interested in the joint distribution of the set $\{Z(\mathbf{s}_i, t_j) : (\mathbf{s}_i, t_j) \in \mathcal{S} \times \mathcal{T}\}$ and the spatiotemporal gradient

process at an arbitrary space-time coordinate $(\mathbf{s}_0, t_0)$. We collect the $Z(\mathbf{s}_i, t_j)$'s into an $N_s N_t \times 1$ vector by first stacking them over time to form an $N_t \times 1$ column $\mathbf{Z}(\mathbf{s}_i)$ for each $\mathbf{s}_i \in \mathcal{S}$, and then stacking the $\mathbf{Z}(\mathbf{s}_i)$'s to form $\mathbf{Z} = (\mathbf{Z}(\mathbf{s}_1)^T, \mathbf{Z}(\mathbf{s}_2)^T, \ldots, \mathbf{Z}(\mathbf{s}_{N_s})^T)^T$. Note that $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{\Sigma}_Z)$, where $\mathbf{\Sigma}_Z$ is the $N_s N_t \times N_s N_t$ variance-covariance matrix whose elements correspond to each pair of space-time coordinates in $\mathcal{S} \times \mathcal{T}$. The element associated with $(\mathbf{s}_i, t_k)$ and $(\mathbf{s}_j, t_l)$ is $K(\mathbf{\Delta}_{ij}, \delta_{kl})$, where $\mathbf{\Delta}_{ij} = \mathbf{s}_j - \mathbf{s}_i$ and $\delta_{kl} = t_l - t_k$. From (4.10), it is clear that $\mathbf{\Sigma}_Z = \sigma^2 \mathbf{R}_Z(\phi_s, \phi_t)$, where $\mathbf{R}_Z(\phi_s, \phi_t)$ is the corresponding correlation matrix.

Let $\nabla Z(\mathbf{s}, t) = \left((\nabla_s Z(\mathbf{s}, t))^T, \nabla_t Z(\mathbf{s}, t), (\nabla_{s,t} Z(\mathbf{s}, t))^T\right)^T$ be the $(2d+1) \times 1$ spatiotemporal gradient process comprising the spatial, temporal and mixed derivatives of $Z(\mathbf{s}, t)$. For any arbitrary space-time coordinate $(\mathbf{s}_0, t_0)$, we can write $\mathrm{Cov}\{\nabla Z(\mathbf{s}_0, t_0), \mathbf{Z}\}$ as the $(2d+1) \times N_s N_t$ matrix partitioned as $\nabla \mathbf{K}_0 = [\nabla \mathbf{K}_{0,1} : \nabla \mathbf{K}_{0,2} : \ldots : \nabla \mathbf{K}_{0,N_s}]$, where each $\nabla \mathbf{K}_{0,i}$ is $(2d+1) \times N_t$ with j-th column $\nabla K(\mathbf{\Delta}_{i0}, \delta_{j0})$ and with $\nabla K(\mathbf{\Delta}, \delta)$ being defined analogously to $\nabla Z(\mathbf{s}, t)$. Hence, the joint distribution of $\mathbf{Z}$ and $\nabla Z(\mathbf{s}_0, t_0)$ is

$$\begin{pmatrix} \mathbf{Z} \\ \nabla Z(\mathbf{s}_0, t_0) \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{bmatrix} \mathbf{\Sigma}_Z & (\nabla \mathbf{K}_0)^T \\ \nabla \mathbf{K}_0 & \mathbf{C}_{\nabla Z}(\mathbf{0}, 0) \end{bmatrix} \right) , \tag{4.11}$$

where $\mathbf{C}_{\nabla Z}(\mathbf{0}, 0)$ is the cross-covariance matrix of $\nabla Z(\mathbf{s}, t)$ evaluated at $(\mathbf{0}, 0)$. Since $Z(\mathbf{s}, t)$ and $\nabla Z(\mathbf{s}, t)$ are subsets of a valid process $\mathbf{W}(\mathbf{s}, t)$, the positive-definiteness of the above variance-covariance matrix is implicit. The conditional distribution of the spatiotemporal gradient process $\nabla Z(\mathbf{s}_0, t_0)$ given $\mathbf{Z}$ is, therefore, a well-defined normal distribution. Subsequently, we show how to carry out posterior predictive inference on the gradient process.

4.3 Hierarchical modeling and inference

In the following we work with $d = 2$ and consider the spatiotemporal process model

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + Z(\mathbf{s}, t) + \epsilon(\mathbf{s}, t) , \tag{4.12}$$

where $\mu(\mathbf{s}, t)$ captures large scale variation or trends (e.g., a regression model), $Z(\mathbf{s}, t)$ is an underlying spatiotemporal process, and $\epsilon(\mathbf{s}, t)$ is a zero centered white noise process to capture micro-scale variability and other unstructured random deviation in the

data. For (4.12), inferring on gradients associated with $Z(\mathbf{s},t)$ is more general than on gradients associated with $Y(\mathbf{s},t)$. The white-noise process in (4.12) introduces a discontinuity in $Y(\mathbf{s},t)$ precluding the existence of gradients on $Y(\mathbf{s},t)$. However, if the trend function $\mu(\mathbf{s},t)$ is modeled as smooth enough to admit gradients, we can use the theory developed in Sections 4.1 and 4.2 to legitimately infer about $\boldsymbol{\nabla}\mathrm{E}[Y(\mathbf{s},t)\,|\,\mu(\mathbf{s},t),Z(\mathbf{s},t)]$, i.e., the gradients on the conditional expectation surface, which then is simply the sum of the gradients for $\mu(\mathbf{s},t)$ and $Z(\mathbf{s},t)$. Without the white noise process in (4.12) (e.g., in models with no "nugget"), we can legitimately infer on $\boldsymbol{\nabla}Y(\mathbf{s},t)$.

The model in (4.12) is applicable to settings where the $Y(\mathbf{s},t)$ is assumed to be continuous over space and time in that $Y(\mathbf{s},t)$ exists for every space-time coordinate in $\Re^d \times \Re$. While the data will be collected only over a finite subset $\mathcal{S}\times\mathcal{T} \subset \Re^d \times \Re$, inference may be sought at any location and time point in the form of interpolation and forecasting. For notational convenience, we will assume that every location in $\mathcal{S}$ produces observations over the same set of time points in $\mathcal{T}$, but this assumption is not necessary and does not apply to our data analysis in the subsequent sections.

We embed the likelihood from (4.12) in a Bayesian hierarchical framework. Following customary assumptions, we model the trend using regressors indexed by both space and time, i.e., $\mu(\mathbf{s},t) = \mathbf{x}(\mathbf{s},t)^T\boldsymbol{\beta}$, and model $Z(\mathbf{s},t)$ as a spatiotemporal process specified by the covariance kernel in (4.10). Specifications are completed by assigning prior distributions to unknown model parameters, thereby yielding the posterior distribution

$$\begin{aligned} p(\boldsymbol{\theta},\mathbf{Z}\,|\,\mathbf{Y}) \propto\; & U(\phi_s\,|\,a_\phi,b_\phi)\times U(\phi_t\,|\,a_\phi,b_\phi)\times IG(\sigma^2\,|\,a_\sigma,b_\sigma)\times IG(\tau^2\,|\,a_\tau,b_\tau) \\ & \times N(\boldsymbol{\beta}\,|\,\boldsymbol{\mu}_\beta,\boldsymbol{\Sigma}_\beta)\times N\left(\mathbf{Z}\,|\,\mathbf{0},\sigma^2\mathbf{R}_Z(\phi_s,\phi_t)\right) \\ & \times \prod_{i=1}^{N_s}\prod_{j=1}^{N_t} N(Y(\mathbf{s}_i,t_j)\,|\,\mathbf{x}(\mathbf{s}_i,t_j)^T\boldsymbol{\beta}+Z(\mathbf{s}_i,t_j),\tau^2)\;, \end{aligned} \tag{4.13}$$

where $\mathbf{Y}$ is the set of all observed $Y(\mathbf{s}_i,t_j)$'s and $\boldsymbol{\theta}=\{\phi_s,\phi_t,\sigma^2,\tau^2\}$ is the set of process parameters in the spatiotemporal and the white-noise processes. The parametrizations for the standard densities are as in Carlin and Louis (2009). We assume all the other hyperparameters in (4.13) are known.

We use MCMC to draw samples from the posterior distribution in (4.13), using Gibbs steps for all parameters except ϕ_s and ϕ_t, which require Metropolis steps.

Sampling-based Bayesian inference is advantageous here as it seamlessly delivers inference on the residual space-time effects. That is, for an arbitrary space-time coordinate $(\mathbf{s}_0, t_0)$, we can sample from the posterior predictive distribution $p(Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Y}) = \int p(Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Z}, \boldsymbol{\theta}) p(\boldsymbol{\theta}, \mathbf{Z} \,|\, \mathbf{Y}) d\boldsymbol{\theta} d\mathbf{Z}$ using composition sampling. To be precise, for each posterior sample for the parameters in (4.13), say $\{\boldsymbol{\beta}^{(j)}, \boldsymbol{\theta}^{(j)}, \mathbf{Z}^{(j)}\}$, we draw $Z(\mathbf{s}_0, t_0)^{(j)}$ from $p(Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Z}^{(j)}, \boldsymbol{\theta}^{(j)})$ for $j = 1, 2, \ldots, M$, where M is the number of (post-burn-in) posterior samples. Predicting the outcome at $(\mathbf{s}_0, t_0)$ is also straightforward: drawing $Y(\mathbf{s}_0, t_0)^{(j)}$ from its full conditional distribution, $N(\mathbf{x}(\mathbf{s}_0, t_0)^T \boldsymbol{\beta}^{(j)} + Z(\mathbf{s}_0, t_0)^{(j)}, \tau^{2(j)})$, where we assume that $\mathbf{x}(\mathbf{s}_0, t_0)$ is available, yields samples from the posterior predictive distribution $p(Y(\mathbf{s}_0, t_0) \,|\, \mathbf{Y})$. Thus, the sampler adapts easily to situations where $Y(\mathbf{s}_0, t_0)$ is missing (or, in our case, unobserved), as these values can be estimated directly from their posterior predictive distributions.

Model assessment and comparisons are also handled easily using our MCMC output. The posterior predictive distribution may be used to generate samples of our observed $Y(\mathbf{s}_0, t_0)$ to assess how well our model is performing using the diagnostics described in Section 2.5.1 of Carlin and Louis (2009). To compare two or more models, we can use the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002), where a lower score indicates a better, more parsimonious fit.

Turning to inference for gradients, we seek the joint posterior predictive distribution,

$$\begin{aligned} p(\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Y}) &= \int p(\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta}) p(\mathbf{Z} \,|\, \mathbf{Y}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \,|\, \mathbf{Y}) d\boldsymbol{\theta} d\mathbf{Z} \\ &= \int p(\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Z}, \boldsymbol{\theta}) p(\mathbf{Z} \,|\, \mathbf{Y}, \boldsymbol{\theta}) p(\boldsymbol{\theta} \,|\, \mathbf{Y}) d\boldsymbol{\theta} d\mathbf{Z} \end{aligned} \tag{4.14}$$

where the second equality follows from the fact that the gradient process is derived entirely from its parent process, so $p(\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Y}, \mathbf{Z}, \boldsymbol{\theta})$ does not depend upon $\mathbf{Y}$. From (4.11), it follows that $p(\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Z}, \boldsymbol{\theta})$ is multivariate normal with mean $\boldsymbol{\mu}_{\nabla Z \,|\, \cdot} = (\boldsymbol{\nabla} \mathbf{K}_0) \, \boldsymbol{\Sigma}_Z^{-1} \mathbf{Z}$ and variance-covariance matrix $\boldsymbol{\Sigma}_{\nabla Z \,|\, \cdot} = \mathbf{C}_{\nabla Z}(\mathbf{0}, 0) - (\boldsymbol{\nabla} \mathbf{K}_0) \, \boldsymbol{\Sigma}_Z^{-1} (\boldsymbol{\nabla} \mathbf{K}_0)^T$, where the $\nabla Z \,|\, \cdot$ subscript refers to $\nabla Z \,|\, \mathbf{Z}, \boldsymbol{\theta}$. Sampling from (4.14) proceeds, again, using composition: for each $\boldsymbol{\theta}^{(j)}$ and $\mathbf{Z}^{(j)}$, we draw $\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0)^{(j)} \sim p(\nabla Z(\mathbf{s}_0, t_0) \,|\, \mathbf{Z}^{(j)}, \boldsymbol{\theta}^{(j)})$, which results in draws from (4.14). Therefore, Bayesian inference on spatiotemporal gradients proceeds in posterior predictive fashion, and requires only the post-convergence MCMC samples for the spatiotemporal model parameters.

Using derivations detailed in Appendix C and letting $A_{0j} = (\phi_t^2 |\delta_{0j}|^2 + 1)$, the j-th

column of the submatrix $\boldsymbol{\nabla}\mathbf{K}_{0,i}$ in the definition of $\boldsymbol{\nabla}\mathbf{K}_0$ just above (4.11), is available in closed form as

$$\boldsymbol{\nabla}_s K(\boldsymbol{\Delta}_{i0}, \delta_{0j}) = -\phi_s^2 \frac{1}{A_{0j}^2} \exp\left[-\frac{\phi_s||\boldsymbol{\Delta}_{i0}||}{A_{0j}^{1/2}}\right] \boldsymbol{\Delta}_{i0}, \tag{4.15}$$

$$\nabla_t K(\boldsymbol{\Delta}_{i0}, \delta_{0j}) = -2\phi_t^2 \exp\left[-\frac{\phi_s||\boldsymbol{\Delta}_{i0}||}{A_{0j}^{1/2}}\right] \left[\frac{1}{A_{0j}^2} + \frac{\phi_s||\boldsymbol{\Delta}_{i0}||}{A_{0j}^{5/2}} - \frac{\phi_s^2||\boldsymbol{\Delta}_{i0}||^2}{2A_{0j}^3}\right] \delta_{0j}, \tag{4.16}$$

$$\text{and } \boldsymbol{\nabla}_{s,t} K(\boldsymbol{\Delta}_{i0}, \delta_{0j}) = 2\phi_t^2\phi_s^2 \exp\left[-\frac{\phi_s||\boldsymbol{\Delta}_{i0}||}{A_{0j}^{1/2}}\right] \left[\frac{2}{A_{0j}^3} - \frac{\phi_s||\boldsymbol{\Delta}_{i0}||}{A_{0j}^{7/2}}\right] \boldsymbol{\Delta}_{i0}\delta_{0j}\ . \tag{4.17}$$

The cross-covariance matrix $\mathbf{C}_{\nabla Z}(\mathbf{0}, 0) = \text{diag}(\phi_s^2\mathbf{I}_d, 2\phi_t^2, 4\phi_s^2\phi_t^2\mathbf{I}_d)$.

4.4 Simulated data example

To demonstrate the effectiveness of our methods, we present an experiment using data generated from a true underlying gradient process, enabling us to validate our Bayesian estimation of spatiotemporal gradients. Our data is generated from $N_s = 100$ randomly placed locations on the unit square with coordinates (s_{i1}, s_{i2}), for $i = 1, \ldots, N_t$. Each site is observed $N_t = 15$ times at evenly spaced increments, $t \in \{1, 2, \ldots, 15\}$, and all locations are observed at each time point, giving us a total of $N = 100 \times 15 = 1500$ observations. To analyze these data, we use an intercept-only regression model, allowing the Gaussian process to capture all spatiotemporal variability. Our results are based on 10,000 iterations of the MCMC sampler, discarding the first 5,000 as burn-in.

We generate our data from a model where the true gradient is available in closed form:

$$Y(\mathbf{s}_i, t_j) \sim N\left(5\left[\sin(s_{i1}3\pi) + \cos(s_{i2}3\pi)\cos(t_j\pi/7)\right], \tau^2\right). \tag{4.18}$$

The reason we chose the expression in (4.18) is two-fold: not only does it allow us to assess the quality of our gradient estimates, but it also features an interesting interaction between space and time. Specifically, in the s_{i2} direction, our spatial gradients evolve over time, allowing us to determine if our mixed gradients (and thus the spatial and the temporal gradients) are able to accurately estimate the true values, and if our model is capable of detecting *significant* mixed gradients. By contrast, the s_{i1} direction does *not*

evolve over time, providing us with i.i.d. temporal replicates of their gradients to help identify locations in which the gradient estimation performed well and where it did not. Not only that, but we can also check how our mixed gradients perform when there are truly no interesting trends to be found.

Our Bayes-MCMC procedure accurately estimated the error variance parameter as $\tau^2 = 0.94$ with a 95% CI of (0.87, 1.02), covering its true value of 1. We also fared well regarding the spatiotemporal random effects, $\mathbf{Z}$, with 97% of the 1500 CI's for $Z(\mathbf{s}_i, t_j)$ covering their true values. Gradients are computed on a grid comprised of points $\mathbf{s}_{i0} = (s_{i01}, s_{i02})$ where $s_{i0j} \in \{0.05, 0.15, \ldots, 0.95\}$ for $j = 1, 2$ at times $t_0 \in \{1, 2, \ldots, 15\}$. Overall, our gradient estimation methods performed quite well: 99% of our estimated gradients' CI's covered their true values. In Figure 4.1, we display the true temporal and the s_{i2} mixed gradients at $t_0 = 3$ compared to their estimated values, as well as a map displaying significant gradients. Our estimated gradients appear to accurately capture the true underlying gradient surface, and we generally see significant gradients where we would expect. For instance, the temporal gradient at location (0.5, 0.5) in Figure 4.1(a) is roughly equal to 0 and this location appears to be in the middle of a downward slope in the vertical axis. Looking at this location in Figure 4.1(b), we find that this location indeed has a significantly negative gradient in the s_{i2} direction. Also, though not shown, we find zero significant mixed gradients in the s_{i1} direction, as we would anticipate.

This example illustrates how an investigation of the spatiotemporal gradient process could be used to identify important missing covariates. Suppose the geography of an area features a mountain range which runs north-to-south, and further suppose that your outcome can be affected by the presence of large mountains, as is known to be the case with pollution. In this case, you would expect to see significant spatial gradients in the east/west direction at points near these mountains, as the pollution level on one side will differ dramatically from those on the other side, depending on the source of the pollution and the direction of the wind. Furthermore, you would *not* expect significant mixed gradients in the east/west direction here, as the underlying factor (i.e., the mountain range) remains unchanged for all time points. In this example, our s_{i1} direction displays such a phenomenon. In contrast, factors such as weather patterns can give rise to spatial, temporal, and spatiotemporal mixed gradients due to their

movement in both space and time. This phenomenon is captured in the s_{i2} direction in this example.

4.5 Analysis of the California air quality data

In this section, we analyze data collected by the Air Resources Board of the California Environmental Protection Agency. As described in Section 1.3.2, these data consist of concentrations of particulate matter less than 2.5 micrometers in diameter (PM2.5) measured at monitoring sites throughout the state of California, with measurements at each site observed typically once every three days over the month of June 2009. In order to make informed decisions regarding air quality warnings, public health officials rely on statistical models to predict each day's pollution levels. In the analysis of pollution data, weather-related covariates such as wind speed and wind direction are typically included, however the only covariate we account for here is the elevation of the monitoring site. We do this deliberately to illustrate how an investigation of the spatiotemporal gradient process can help identify important covariates (in this case, weather and/or topography) still missing from the statistical model.

To justify the use of the model described in Section 4.3, we compare it to a model where

$$\text{Cov}(Z(\mathbf{s}_i, t_j), Z(\mathbf{s}_{i'}, t_{j'})) = (1 + \phi_s||\mathbf{s}_{i'} - \mathbf{s}_i||) \exp\left[-\phi_s||\mathbf{s}_{i'} - \mathbf{s}_i||\right], \text{if } j = j' \qquad (4.19)$$

and 0 otherwise; that is, we compare our model to a purely spatial one. As this model is also based on the Matérn(3/2), it would permit spatial (but not temporal or mixed) gradients in a manner following Banerjee et al. (2003). Using DIC, we find that our model is superior, with DIC = 1472 compared to DIC = 1521 for the purely spatial model. Our model also has a lower value of p_D (effective number of parameters), coming in at 337 compared to 362 for the purely spatial model; this decrease in p_D is likely due to our model's ability to borrow strength across time points, reducing overfitting.

A summary of our full model's parameter estimates can be found in Table 4.1. As expected, elevation is significantly negative, suggesting that locations with higher elevation tend to have lower levels of pollution. Our variance parameters τ^2 and σ^2 can be used to compute $\sigma^2/(\sigma^2 + \tau^2) = 0.83$, indicating that 83% of the variability in

Parameter	Median (95% CI)	Parameter	Median (95% CI)
β_0 (Intercept)	8.954 (8.841, 9.066)	σ^2 (S-T Var)	12.876 (10.352, 16.35)
β_1 (Elevation)	-0.693 (-0.961, -0.419)	ϕ_s (Spatial)	1.713 (1.422, 2.010)
τ^2 (Error Var)	2.592 (2.214, 3.024)	ϕ_t (Time)	4.175 (3.464, 5.075)

Table 4.1: Parameter estimates from the analysis of the PM2.5 data. ϕ_s is on a scale of 1 unit = 100 km, and ϕ_t is on a scale of 1 unit = 1 week.

our data is being explained by the spatiotemporal process. This high a value is not surprising given the limited number of covariates in our model. Finally, our spatial and temporal range parameters, ϕ_s and ϕ_t, can be interpreted as controlling how quickly the correlation between two observations drops off. For instance, spatial correlation between two observations at the same time point falls to less than 0.4 at a distance of roughly 120 km, and the temporal correlation between two observations at the same location falls to less than 0.4 after roughly 2 days.

Figure 4.2 displays the maps of the fitted values (i.e., $\widehat{Y}(\mathbf{s}_i, t_j) = \mathbf{x}(\mathbf{s}_i, t_j)^T\widehat{\boldsymbol{\beta}} + \widehat{Z}(\mathbf{s}_i, t_j)$) for each of the days in the study period. Given our limited number of covariates, the majority of the variation in this figure is due to the variability in $\mathbf{Z}$; as such, our focus here will be on the underlying spatiotemporal process. Here, most of our higher levels of pollution occur in the southern part of the state, stretching from Los Angeles in the west to Riverside and San Bernardino in the east. Not only that, but June 3rd in particular sticks out as a day of high pollution. For the most part, pollution levels in the northern half of the state are on the lower end of the spectrum.

In Figure 4.3, we display maps of the spatial and temporal gradients for the transition between June 2nd and June 3rd; for comparison purposes, we also display a topographical map of California in Figure 4.3(d). Here, it's easy to see the effect of the topography in California, as the Coast Ranges (along the coast from northern California to the Los Angeles area) and Transverse Ranges (stretching from west to east, creating the northern border of Los Angeles) appear to strongly influence the gradient estimates. While not shown here, all three types of standard gradients have significant values in the Los Angeles area, which is not surprising given the amount of activity in Figure 4.2 between June 2nd and June 3rd. We don't find many significant mixed gradients (not

shown), though this is not entirely unexpected — after all, we would not expect wind to be diverted around a mountain one day and suddenly blow *through* the same mountain the next day.

This analysis illustrates two primary features of our model. First, the spatiotemporal gradient process can be used to help identify important factors missing from our model. In our case, we observe significant spatial gradients and a lack of significant mixed gradients, suggesting our model is failing to account for a spatially-varying factor that is fixed over time. Upon viewing the maps in Figure 4.3, an investigator familiar with California's topography should be able to identify that the phenomenon shown in these maps is likely due to fresh ocean air pushing pollution from more coastal communities against the various mountain ranges around Los Angeles and throughout the state. This information could prove invaluable, as accumulating pollution at the foot of these mountains could result in great public health risk, motivating public health alerts or other precautionary measures. Another strength of this model is that we can estimate gradients at *any* location and *any* point in time. This is of particular importance here because it would be impractical (if not impossible) for investigators to measure pollution levels continuously over mountain ranges, making it impossible to adequately account for the topography of California in their statistical model.

We also believe that the spatiotemporal gradient methodology we have developed here could be used in real-time for the purposes of air quality monitoring. For these data, an automated system could have computed spatiotemporal gradients in real-time and alerted a public health official when a significant gradient had been detected between June 2nd and June 3rd. The official could then look at the maps in Figures 4.3(a), 4.3(b), and 4.3(e), notify child and elder care facilities in the affected areas of the increasingly poor air quality, and suggest that their residents remain indoors if possible.

4.6 Discussion

In this chapter, we have outlined a fully process-based inferential framework for infinitemsmal rates of change in space and time within a Bayesian setting. We expand

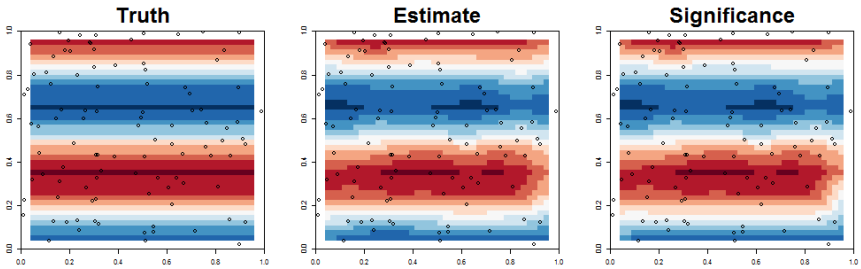
upon earlier work on rates of change in purely spatial or in discrete-space continuous-time settings. Rather than infer on discretized spatiotemporal finite differences as approximations of gradients, which tend to become numerically unstable at sharper resolutions (Banerjee et al., 2003), we have opted to utilize limiting properties of Gaussian processes to develop the required distribution theory for spatiotemporal gradients. This development is non-trivial because space and time can have widely varying scales and, importantly, should be modeled using nonseparable covariance functions. We proposed joint inference on directional rates of change over space, derivatives over time and mixed derivatives over space and time without sacrificing richness and flexibility in modeling.

We obtain the full posterior distribution of spatiotemporal gradients in posterior predictive fashion at arbitrary locations and timepoints. This is attractive from an execution standpoint because we do not need to recompute the posterior distribution of the model parameters when inference is sought on new locations and timepoints. For example, in a real-time surveillance system context, we can store the post burn-in posterior samples of the model parameters and estimate the spatiotemporal gradients using composition sampling, which is fast and exact. Our sampling-based Bayesian framework outlined in Section 4.3 easily adapts to finding gradients from *misaligned* or *unbalanced* spatiotemporal data, where the outcome may not have been observed in the same set of locations across time.

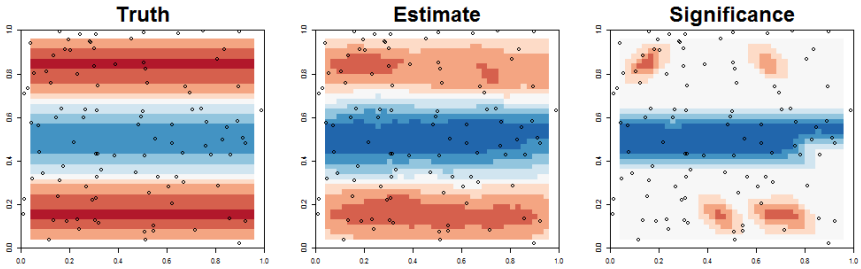
Here, we have focused upon inference on $\boldsymbol{\nabla} Z(\mathbf{s}, t)$. As mentioned in Section 4.3, we can infer on rates of change on the spatial outcome, either through the expected value, $\boldsymbol{\nabla} \mathrm{E}[Y(\mathbf{s}, t) \,|\, \mu(\mathbf{s}, t), Z(\mathbf{s}, t)]$, or through $\boldsymbol{\nabla} Y(\mathbf{s}, t)$ depending upon whether the outcome has a "nugget" (discontinuity) or not. This will require $\boldsymbol{\nabla} \mu(\mathbf{s}, t)$, which may be available in closed form if $\mu(\mathbf{s}, t)$ is a parametric function (e.g. a trend surface) or, more generally, can be calculated from smoothly interpolated space-time covariates. If, on the other hand $\mu(\mathbf{s}, t)$ has systemic discontinuities, such as when $\mu(\mathbf{s}, t)$ includes categorical or discontinuous covariates, then we can legitimately infer only on $\boldsymbol{\nabla} Z(\mathbf{s}, t)$, the smooth part of the mean. An advantage of the Bayesian paradigm is that it seamlessly accommodates each of the above settings including, in particular full inference on the completely unobserved $\boldsymbol{\nabla} Z(\mathbf{s}, t)$ for arbitrary $\mathbf{s}$ and t.

Turning to our analysis of the California air pollution data, supplying the posterior distribution for mixed gradients can help investigators determine the nature of the

important factor(s) not accounted for in our mean model. In the case of our PM2.5 concentration data, the lack of significant mixed gradients suggests that our missing factor varies spatially but remains constant over time. Based on our knowledge of California's topography, the effect of fresh ocean air, and basic meteorology, it is easy to see how the state's mountain ranges could play a role in the concentration of PM2.5. Furthermore, our results from our simulated data example suggest that if California's topography is an important factor for modeling PM2.5, then we should see similar patterns in our spatial gradients over time (as demonstrated in the s_{i1} direction in our example). While not shown here, this is generally the case, despite the lack of obvious trends in the fitted values in Figure 4.2. Similarly, a significant temporal gradient in the absence of a significant mixed gradient in our example could correspond to simple changes in statewide automobile or air traffic patterns. On the other hand, finding significant mixed gradients may suggest something more related to specific weather events, such as a cold front bringing lower temperatures and precipitation. While we did not observe significant activity in the mixed gradients for our pollution data, our simulated data example demonstrated the ability of our model to accurately capture these phenomena.



(a) Temporal Gradient



(b) s_{i2} Mixed Gradient

Figure 4.1: Comparison of the true temporal and the s_{i2} mixed gradients for a particular time point and their posterior median values based on our gradient theory. The color scheme here goes from highly negative (blue) to highly positive (red), centered around 0 (light gray), with locations of observed locations plotted as open circles. The third panel highlights significant gradients; i.e., gradients whose 95% CI's contain 0 are assigned value 0, while gradients whose 95% CI's do not contain 0 are assigned their posterior medians.

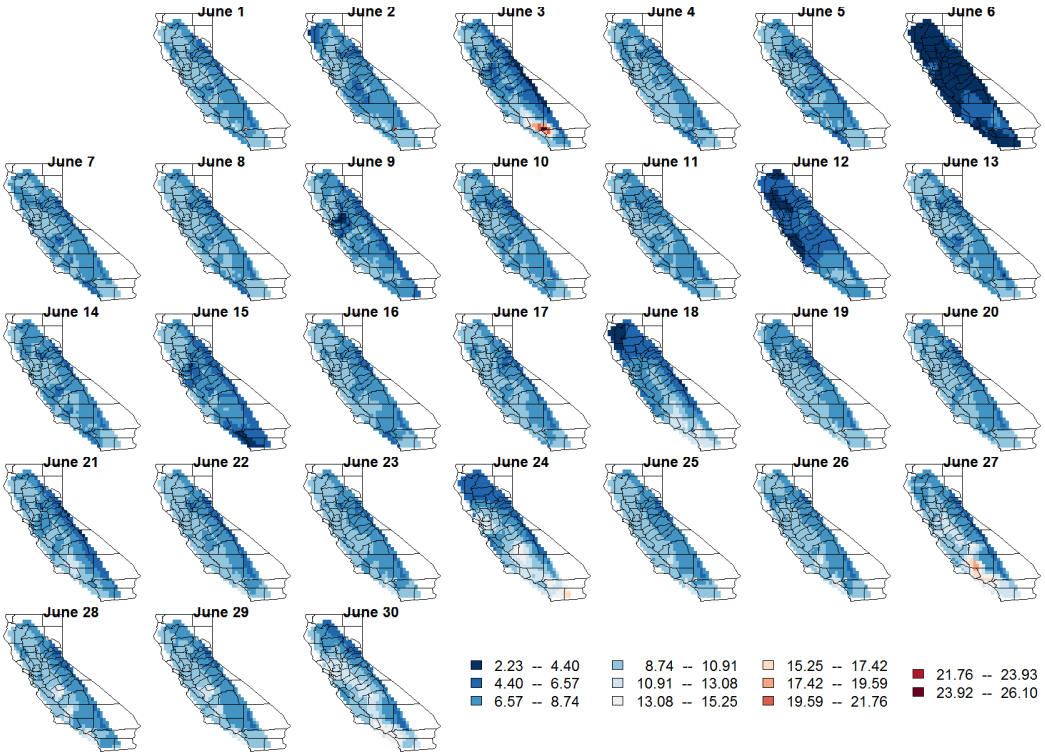
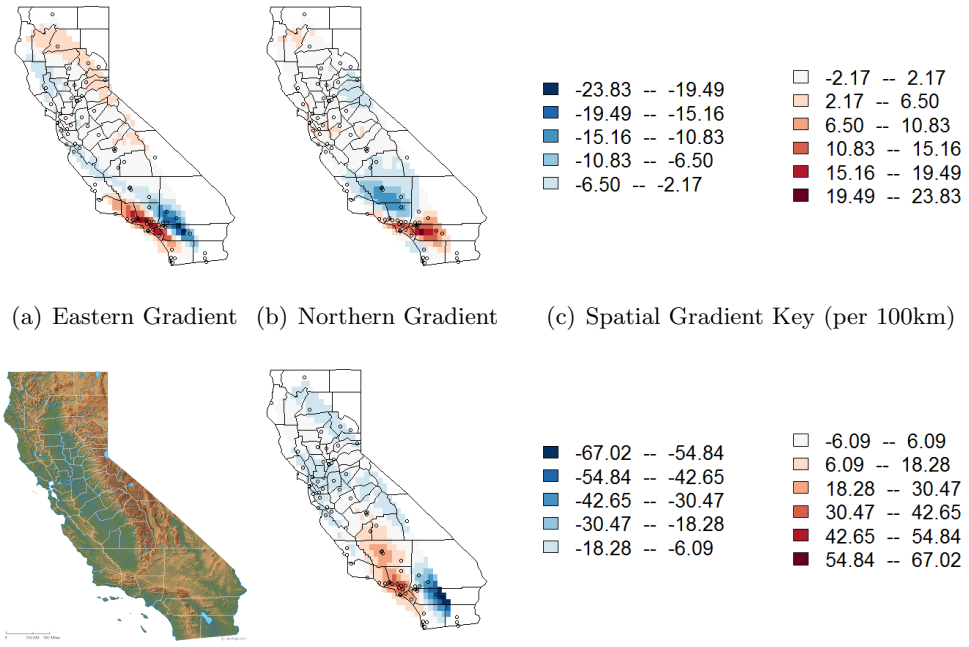


Figure 4.2: Predicted levels of PM2.5 for the month of June 2009. The maps are situated so that the first column of days are on Sunday, the second are Mondays, etc.



(a) Eastern Gradient (b) Northern Gradient (c) Spatial Gradient Key (per 100km)

(d) Topography (e) Temporal Gradient (f) Temporal Gradient Key (per 1 week)

Figure 4.3: Estimated spatial and temporal gradient plots for the transition period between June 2 and June 3, as well as a topographical map of California for comparison purposes. Spatial gradients in the southern portion of the state appear to correspond to the Transverse Ranges, which wrap around north and east of Los Angeles. We are also able to identify the Coast Ranges, which appear to cause negative eastern gradients along the Pacific Coast. The temporal gradients display the sharp increase in the Los Angeles area seen in Figure 4.2.

Chapter 5

Analysis of Censored Data Using the Spatiotemporal Gradient Process

The work in this chapter is motivated by the *Deepwater Horizon* (BP) airborne exposure data described in Section 1.3.3. As mentioned there, these data consist of levels of fine airborne particles to which the workers involved in the cleanup efforts may have been exposed, and our inferential goal is be to analyze how the cloud of pollution disperses over time from the site of the spill. Specifically, we can use the spatiotemporal gradient process to determine if efforts to stem the flow of oil resulted in sharp temporal gradients in pollution levels, whether the cloud of pollution dispersed gradually (the absence of significant spatial gradients), and whether the size/shape of the pollution cloud changed during the course of the cleanup (significant mixed gradients). While at first glance this task seems like a straightforward application of the methods described in Chapter 4—as we again find ourselves with data arising from a continuous space, continuous time setting—here we need to address the high degree of *censoring* in our data before proceeding.

One of the strengths of analyzing data under a Bayesian framework is the seamless nature with which censored or, more generally, *missing* data can be analyzed. Suppose

we have a model of the form

$$Y_i \sim N(\mu, \sigma^2), i = 1, \ldots, N \tag{5.1}$$

where we let $\boldsymbol{\theta} = (\mu, \sigma^2)$. In the case of a single missing observation, Y_{i^*}, one can simply treat this observed value as yet another unknown parameter in our model, and we can proceed as we normally would. Prior to running our MCMC algorithm, we would initialize values for all of our unknown parameters, including the missing Y_{i^*}. Then, in each iteration of our MCMC algorithm, we can sample $\boldsymbol{\theta}^{(\ell)}$ as though we have complete data, and we can then sample a new value, $Y_{i^*}^{(\ell)}$, from its likelihood given the current values of the other model parameters, $\boldsymbol{\theta}^{(\ell)}$. Extending this to the case where we have a collection of missing observations is trivial. The censored data scenario is largely the same, with one change. Instead of the likelihood in (5.1), if we have left-censored data with a single censoring threshold, C, our censored observations will be drawn from

$$Y_{i^*} \sim N(\mu, \tau^2) I(Y_{i^*} < C). \tag{5.2}$$

As described by Gelfand et al. (1992), we have two primary methods for sampling Y_{i^*}: (a) generate values from the likelihood in (5.1) and reject those which are above the threshold or (b) use an inverse CDF method like those described in Section 2.2 of Devroye (1986). In general, option (a) will be less efficient than option (b), particularly if the likelihood of a value being below the censoring threshold is small, thus we choose to use an inverse CDF method for handling our censored values.

Before we proceed, it's important to address one question that may be raised with respect to censored data such as those described in Section 1.3.3: "Why do we care about values below the limit of detection?" Unfortunately, a value below a certain detection limit need not be "negligible" from the standpoint of workers health. Unfortunately, this is not necessarily the case. While it is true that trace amounts of certain compounds may have a negligible impact, carcinogens such as benzene (an organic compound which falls under the umbrella term "VOC") can cause poor health outcomes in very small doses. Furthermore, a common misconception is that an instrument's "limit of detection" is determined by biological guidelines, when in fact they are much more likely to be determined by the technology available, as well as financial limitations. As such, imputing values below even 0.1ppm may very well be substantively important for modeling exposure.

Building on the theory developed in Chapter 4, we begin this chapter by outlining how to incorporate censoring into a hierarchical model in Section 5.1. We then analyze a highly censored simulated data set in Section 5.2, where we demonstrate the ability of our model to accurately estimate the underlying gradient process despite the missingness in the data. Next, in Section 5.3 we turn our attention to the data from the BP oil spill described in Section 1.3.3 and use our methods to investigate rates of change in levels of VOC. Finally, we add some concluding thoughts in Section 5.4.

5.1 Hierarchical modeling and inference

As in Chapter 4, we considered the spatiotemporal process model

$$Y(\mathbf{s},t) = \mu(\mathbf{s},t) + Z(\mathbf{s},t) + \epsilon(\mathbf{s},t)\ , \tag{5.3}$$

where $\mu(\mathbf{s},t)$ captures large scale variation or trends (e.g., a regression model), $Z(\mathbf{s},t)$ is an underlying spatiotemporal process, and $\epsilon(\mathbf{s},t)$ is a zero centered white noise process to capture micro-scale variability and other unstructured random deviation in the data. The model in (5.3) is applicable to settings where the $Y(\mathbf{s},t)$ is assumed to be continuous over space and time, in that $Y(\mathbf{s},t)$ exists for every space-time coordinate in $\Re^d \times \Re$. While the data will be collected only over a finite subset $\mathcal{S}\times\mathcal{T} \subset \Re^d \times \Re$, inference may be sought at any location and time point in the form of interpolation and forecasting.

To construct our censored data model, we first need to define an indicator variable, $D(\mathbf{s},t)$, which equals 1 if $Y(\mathbf{s},t)$ is above the detectable limit, C, and is 0 otherwise. Then, we define the conditional distribution of $Y(\mathbf{s},t)$ given $\boldsymbol{\theta}$ and $D(\mathbf{s},t)$

$$Y(\mathbf{s},t)\,|\,\boldsymbol{\theta}, D(\mathbf{s},t) \sim \begin{cases} N(\mu(\mathbf{s},t) + Z(\mathbf{s},t), \tau^2) & \text{if } D(\mathbf{s},t) = 1 \\ N(\mu(\mathbf{s},t) + Z(\mathbf{s},t), \tau^2)\ I(Y(\mathbf{s},t) < C) & \text{if } D(\mathbf{s},t) = 0 \end{cases}.$$

Embedding this in a Bayesian hierarchical framework, we model the trend using regressors indexed by both space and time, i.e., $\mu(\mathbf{s},t) = \mathbf{x}(\mathbf{s},t)^T\boldsymbol{\beta}$, and model $Z(\mathbf{s},t)$ as a spatiotemporal process specified by the covariance kernel in (4.10). Specifications are completed by assigning prior distributions to unknown model parameters, thereby

yielding the posterior distribution

$$\begin{aligned} p(\boldsymbol{\theta}, \mathbf{Z} \,|\, \mathbf{Y}, \mathbf{D}) \propto\; & U(\phi_s \,|\, a_\phi, b_\phi) \times U(\phi_t \,|\, a_\phi, b_\phi) \times IG(\sigma^2 \,|\, a_\sigma, b_\sigma) \times IG(\tau^2 \,|\, a_\tau, b_\tau) \\ & \times N(\boldsymbol{\beta} \,|\, \boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \times N\left(\mathbf{Z} \,|\, \mathbf{0}, \sigma^2 \mathbf{R}_Z(\phi_s, \phi_t)\right) \\ & \times \prod_{(i,j) \notin \mathcal{D}} N(Y(\mathbf{s}_i, t_j) \,|\, \mathbf{x}(\mathbf{s}_i, t_j)^T \boldsymbol{\beta} + Z(\mathbf{s}_i, t_j), \tau^2) I\{Y(\mathbf{s}_i, t_j) < C\} \\ & \times \prod_{(i,j) \in \mathcal{D}} N(Y(\mathbf{s}_i, t_j) \,|\, \mathbf{x}(\mathbf{s}_i, t_j)^T \boldsymbol{\beta} + Z(\mathbf{s}_i, t_j), \tau^2) \;, \end{aligned} \tag{5.4}$$

where $\mathbf{D}$ is the vector constructed from $D(\mathbf{s}_i, t_j)$, $\mathcal{D} = \{(i,j) \,|\, D(\mathbf{s}_i, t_j) = 1\}$, and $\boldsymbol{\theta} = \{\phi_s, \phi_t, \sigma^2, \tau^2\}$ is the set of process parameters in the spatiotemporal and the white-noise processes. The parametrizations for the standard densities are as in Carlin and Louis (2009). We assume all the other hyperparameters in (5.4) are known.

In the absence of censored data — that is, a "complete" data setting — implementing MCMC to draw samples from the posterior distribution in (5.4) using the combination of Gibbs and Metropolis steps mentioned in Section 4.3 would be straightforward. Thus, our goal here is to describe how one can sample a censored $Y(\mathbf{s}_i, t_j)$ using a random uniform variate and the CDF of a Normal distribution via the inverse CDF method. First, note that the full conditional for each censored $Y(\mathbf{s}_i, t_j)$ is simply

$$\begin{aligned} \pi(Y(\mathbf{s}_i, t_j) \,|\, \boldsymbol{\theta}, \mathbf{Z}, \mathbf{Y}_{(i,j)}, \mathbf{D}) \propto\; & N(Y(\mathbf{s}_i, t_j) \,|\, \mathbf{x}(\mathbf{s}_i, t_j)^T \boldsymbol{\beta} + Z(\mathbf{s}_i, t_j), \tau^2) \\ & \times I\{Y(\mathbf{s}_i, t_j) < C\}, \end{aligned} \tag{5.5}$$

a truncated Normal distribution, where $\mathbf{Y}_{(i,j)}$ denotes the vector $\mathbf{Y}$ with $Y(\mathbf{s}_i, t_j)$ removed. To use the inverse CDF method, we note that

$$\begin{aligned} -\infty < y \leq C \;&\Longrightarrow 0 = F(\infty) < F(y) \leq F(C) \\ &\Longrightarrow 0 < u = \frac{F(y)}{F(C)} \leq \frac{F(C)}{F(C)} = 1, \end{aligned} \tag{5.6}$$

for any cumulative density function, $F(\cdot)$. Then, during the ℓth iteration of our MCMC sampler, to sample $Y^{(\ell)}(\mathbf{s}_i, t_j)$ given $\boldsymbol{\theta}^{(\ell)}$, we simply need to reverse these steps and let $F(\cdot) = F(\cdot \,|\, \boldsymbol{\theta}^{(\ell)})$ be the CDF corresponding to the likelihood from (5.3) evaluated at the current values of $\boldsymbol{\beta}$, Z, and τ^2:

Step 1: Draw $u^{(\ell)} \sim Unif(0, 1)$

Step 2: Let $\psi^{(\ell)} = u^{(\ell)} F(C \mid \boldsymbol{\theta}^{(\ell)})$

Step 3: Let $Y^{(\ell)}(\mathbf{s}_i, t_j) = F^{-1}\left(\psi^{(\ell)} \mid \boldsymbol{\theta}^{(\ell)}\right)$.

Note that this process is equivalent to drawing $Y(\mathbf{s}_i, t_j)$ directly from the truncated Normal in 5.5 but does not depend on the availability of truncated Normal functionality in your programming language. Having demonstrated the ability to impute the censored value (and thus obtain a "complete" data set), our MCMC sampler can proceed as before for our other unknown parameters. Furthermore, we can still take advantage of using sampling-based Bayesian inference by seamlessly obtaining inference on the residual space-time effects and prediction at arbitrary space-time coordinates, $(\mathbf{s}_0, t_0)$, by sampling from their respective posterior predictive distributions; of course, prediction at these coordinates assumes that any covariates $\mathbf{x}(\mathbf{s}_0, t_0)$ are available. Similarly, inference on the spatiotemporal gradient process is conducted in the same way as before.

5.2 Simulated data example

To demonstrate the effectiveness of our methods, we reconduct the experiment from Section 4.4 using data generated from a true underlying gradient process while censoring half of our observations, enabling us to investigate the quality of our spatiotemporal gradient estimation in the presence of censored data. As before, our data is generated from $N_s = 100$ randomly placed locations on the unit square with coordinates (s_{i1}, s_{i2}), for $i = 1, \ldots, N_t$. Each site is observed $N_t = 15$ times at evenly spaced increments, $t \in \{1, 2, \ldots, 15\}$, and all locations are observed at each time point, giving us a total of $N = 100 \times 15 = 1500$ observations. To analyze these data, we use an intercept-only regression model, allowing the Gaussian process to capture all spatiotemporal variability. Our results are based on 10,000 iterations of the MCMC sampler, discarding the first 5,000 as burn-in.

As in Section 4.4, we generate our data from model 4.18 where the true gradient is available in closed form:

$$Y(\mathbf{s}_i, t_j) \sim N\left(5\left[\sin(s_{i1}3\pi) + \cos(s_{i2}3\pi)\cos(t_j\pi/7)\right], \tau^2\right); \tag{5.7}$$

due to the nature of this data and our desire to censor 50% of our observations, we have selected a censoring threshold of $C = 0$. Figure 5.1 displays the surface created by (5.7),

where any shade of blue denotes a negative, and thus censored, value using our threshold of $C = 0$. This figure is particularly helpful, as it helps us identify two features that our spatial gradients will have to encounter. First off, we note that for $s_{i1} \in (0.4, 0.6)$, we are quite likely to encounter censored values, thus we should expect to learn very little information regarding our s_{i2} spatial gradients in this region. Coincidentally, this region corresponds to a relatively flat valley in the s_{i1} direction — that is, we *shouldn't* have any interesting gradients here — so an investigation of these gradients should provide insight into how likely our model is to find significant gradients in a censored region when the underlying gradient process is quite flat. Finally, the temporal gradients are constant across values of s_{i1}, so it will be easy to identify the effect censoring has on these gradients, as well.

Our Bayes-MCMC procedure accurately estimated the error variance parameter as $\tau^2 = 0.95$ with a 95% CI of (0.85, 1.06), covering its true value of 1 and providing nearly identical results to those obtained in Section 4.4. In fact, many of our findings are quite similar to those from the non-censored example in Chapter 4, including the estimation and coverage of both our spatiotemporal random effects and gradients. In Figure 5.2, we recreate the analogous figure from our uncensored example (Figure 4.1), and we find that our results are largely the same. Still, it is easy to notice the effect that the censoring has on the interval $s_{i1} \in (0.4, 0.6)$, as these gradients are much less pronounced (and less likely to be significant).

Based on the results of this example, it appears that even at 50% censoring, we are still able to make important inference on the spatiotemporal gradient process. While regions with a large amount of censoring are less likely to include *significant* gradients, our posterior estimates in these regions remain quite accurate, and there doesn't seem to be much loss of significance in the regions with values above the limit of detection.

5.3 Analysis of the *Deepwater Horizon* oil spill data

As first described in Section 1.3.3, our data set consists of 447 space- and time-specific measurements of VOC collected during the cleanup efforts of the *Deepwater Horizon* oil spill, 375 (83.8%) of which are below the limit of detection (0.1ppm). The data were observed within 30km of the site of the oil rig and were collected over a period of 30 days.

Aside from the spatial location (in latitude and longitude) and time of observation (measured to the nearest second), we have no other information – that is, we have no covariate information such as wind speed, wind direction, water temperature, etc. Spatial distances are computed using a sinusoidal projection: $s_{i1} = R\lambda\cos\theta$ and $s_{i2} = R\theta$, where $R = 6371$km is the radius of the earth, and θ and λ denote latitude and longitude, respectively, and are on a scale where one spatial unit equal 5 km, while temporal distances were computed by letting one temporal unit equal 5 minutes. Following convention, we model our data on the log-scale whence our censoring threshold is $C = \log(0.1) = -2.3$.

A summary of our model's parameter estimates can be found in Table 5.1. Our variance parameters τ^2 and σ^2 can be used to compute $\sigma^2/(\sigma^2 + \tau^2) = 0.90$, indicating that 90% of the variability in our data is being explained by the spatiotemporal process, which is not surprising given that we have no covariate information. Our spatial and temporal range parameters, ϕ_s and ϕ_t, can be interpreted as controlling how quickly the correlation between two observations drops off. For instance, spatial correlation between two observations at the same time point falls to less than 0.4 at a distance of roughly 6 km, and the temporal correlation between two observations at the same location falls to less than 0.4 after roughly 2 hours. One thing to note is that the covariance matrix in this analysis is sensitive to small values of ϕ_t; for values of $\phi_t < 0.04$, the covariance matrix was numerically singular. This is likely due to the extreme temporal range of our data (30 days) compared to the important temporal range (less than 2 hours) and the minimum temporal distance of a few minutes. As a result, we have set our prior for ϕ_t such that it is restricted to values greater than 0.04. This phenomena does not occur for the spatial component, though this could be due to the scope of the spatial range (roughly 40 km) relative to the important temporal range (6 km) or simply difference between the Matérn structure used for spatial correlation compared to the structure used for time. Regardless, future work will be done to determine the precise cause of this and, if necessary, a more appropriate correlation structure for analyzing data with a temporal range that dramatically dwarfs the important temporal scale.

After fitting our model, we interpolated the hourly VOC surface for each day in our study period with a large number of observations. For instance, Figure 5.3 displays 16 hourly surface plots for June 26. Here, "clouds" of red shading denote locations where we expect to observe levels of VOC above the limit of detection (darker shades of red

Parameter	Median (95% CI)	Parameter	Median (95% CI)
β_0 (Intercept)	-5.324 (-5.934, -4.580)	ϕ_s (Spatial)	1.552 (0.484, 2.907)
τ^2 (Error Var)	0.729 (0.482, 1.093)	ϕ_t (Time)	0.043 (0.040, 0.062)
σ^2 (S-T Var)	6.424 (3.840, 9.906)		

Table 5.1: Parameter estimates from the analysis of the BP data. ϕ_s is on a scale of 1 unit = 5 km, and ϕ_t is on a scale of 1 unit = 5 minutes.

indicating larger values) while white denotes borderline values and blue denotes censored values. These correspond to the red and black circles which indicate values above the limit of detection and censored observations, respectively. As we can see, the red cloud of high VOC levels appears to be centered around the site of the *Deepwater Horizon* oil rig, denoted here by a green plus-sign (+). In Figure 5.4, we narrow our focus to the period from 12:00PM to 3:00PM and highlight the path that this particular ship took while collecting measurements on its 10km trip around the affected area. Here, we can more clearly see that the dark shade of red in the 1:00PM and 2:00PM plots is due to the number of uncensored observations, as well as how quickly we began to see elevated levels of VOC.

In Figure 5.5, we again focus on June 26, this time looking at the estimated spatial and temporal gradients at 1:00PM. Immediately, we find that the inflection point of both of our spatial gradients occurs at the site of the oil spill, as evidenced by the change from red to blue shading at this point, though it's worth noting that we may merely be picking up the boundary of our observations, as our observed data on this day is not dispersed evenly on all sides of the *Deepwater Horizon* oil rig. As for changes over time, the area of positive temporal gradients centered around the oil rig correspond not only to the increase in levels of VOC observed in Figure 5.4, but also the largest observed value of VOC in our data set, 8.6ppm, which occurred near the site of the rig at 1:41PM on this day. Unfortunately, the amount of censoring this data (coupled with the irregularity of the locations and times of the observations) appears to hinder our ability to identify *significant* gradients, and our mixed gradients capture little more than noise.

Given the limitations of this data set, we are still able to provide investigators with

a number of interesting results. First and foremost, we are able to provide estimates of the size and shape of the clouds of volatile organic compounds by borrowing strength from other days in the study. Furthermore, it's likely that the compounds that the workers were exposed to are the result of dispersants released to break-up the oil in the water. As such, investigators may be able to use this model in real-time to identify which types of dispersants linger and cause prolonged exposure to workers, information which may then be used to inform future use of dispersants and the conditions under which the workers conduct their cleanup efforts.

5.4 Discussion

In this chapter, not only have we described how one could fit a censored hierarchical model, but we have also demonstrated the ability of our proposed methodology to estimate the underlying gradient process via a simulated data example. Having shown the ability to find substantive findings in such a highly censored setting, we proceeded to an analysis of data collected during the cleanup efforts of the BP oil spill. Despite having less than 17% of our data above the limit of detection, our model was still able to shed valuable insight into the amount of harmful VOCs that the workers engaged in the cleanup were likely exposed to. In practice, these methods could be used to produce warning boundaries beyond which workers must wear protective ventilation masks.

While our BP data set consisted of observations collected at a irregular collection of spatial locations and time points, it's easy to imagine a real-world scenario more in line with our simulated data. For instance, one can envision the monitoring stations in California described in Section 1.3.2 collecting pollution levels with some limit of detection, with data being collected regularly (say, daily) over the course of a month. Inevitably, some proportion of these observations would be censored, but as we have discussed previously, levels below detectable limits can still prove to be quite harmful, thus motivating the use of the methods described in this chapter. Based on the analysis of our simulated data example, there's evidence to suggest that our gradient methodology would perform quite well, even in the presence of censoring.

As for future work, these methods may be used to analyze a much larger set of data collected during the cleanup efforts from the BP oil spill. These data consist of personal

exposure measurements — that is, each worker is assigned a spatiotemporal coordinate using GPS and data regarding exposure to airborne particles is collected via a personal badge each worker wears. These data set will likely encompass a larger, and certainly a more densely populated spatial region (in terms of the number of observations), which should provide more precise results. Furthermore, the use of gradient methods for these data may then be used later to assign workers to exposure groups for the purposes of modeling individual health outcomes.

Sticking with the analysis of data related to the BP oil spill, investigators may be specifically interested in data along the *coastline*, for instance, the amount of oil washing ashore. In this setting, it may be more appropriate to treat space as one-dimensional, as two points on opposite sides of a narrow peninsula would likely be less correlated than two points equally far apart on the *same* side of the peninsula. Provided we can map the two dimensional space to points on a series of line segments, we can then conduct *coastal kriging* — that is, spatial prediction along the coast — and make inference on the gradient process developed here and in Chapter 4.

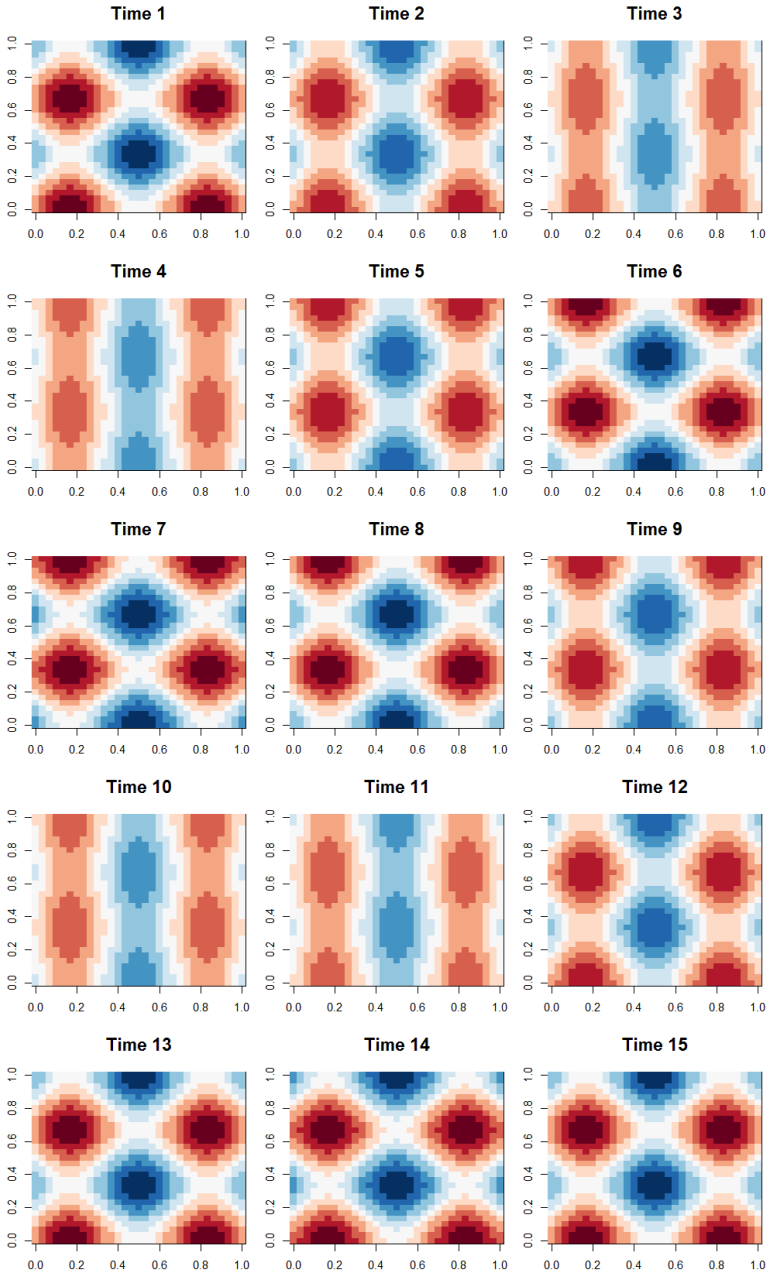
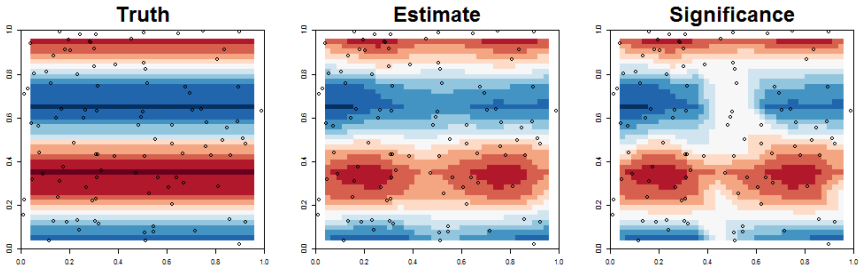
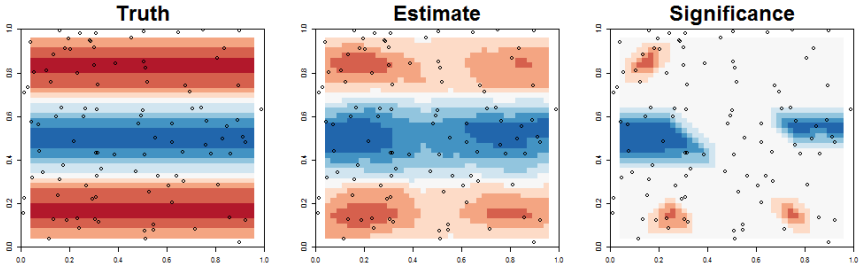


Figure 5.1: Underlying process for the data generated by (5.7). Values range from -10 (dark blue) to 10 (dark red), where here any shade of blue denotes a negative, and thus censored, value.



(a) Temporal Gradient



(b) s_{i2} Mixed Gradient

Figure 5.2: Comparison of the true temporal and the s_{i2} mixed gradients for time point 3 (see Figure 5.1) and their posterior median values based on our gradient theory. The color scheme here goes from highly negative (blue) to highly positive (red), centered around 0 (light gray), with locations of observed locations plotted as open circles. The third panel highlights significant gradients; i.e., gradients whose 95% CI's contain 0 are assigned value 0, while gradients whose 95% CI's do not contain 0 are assigned their posterior medians.

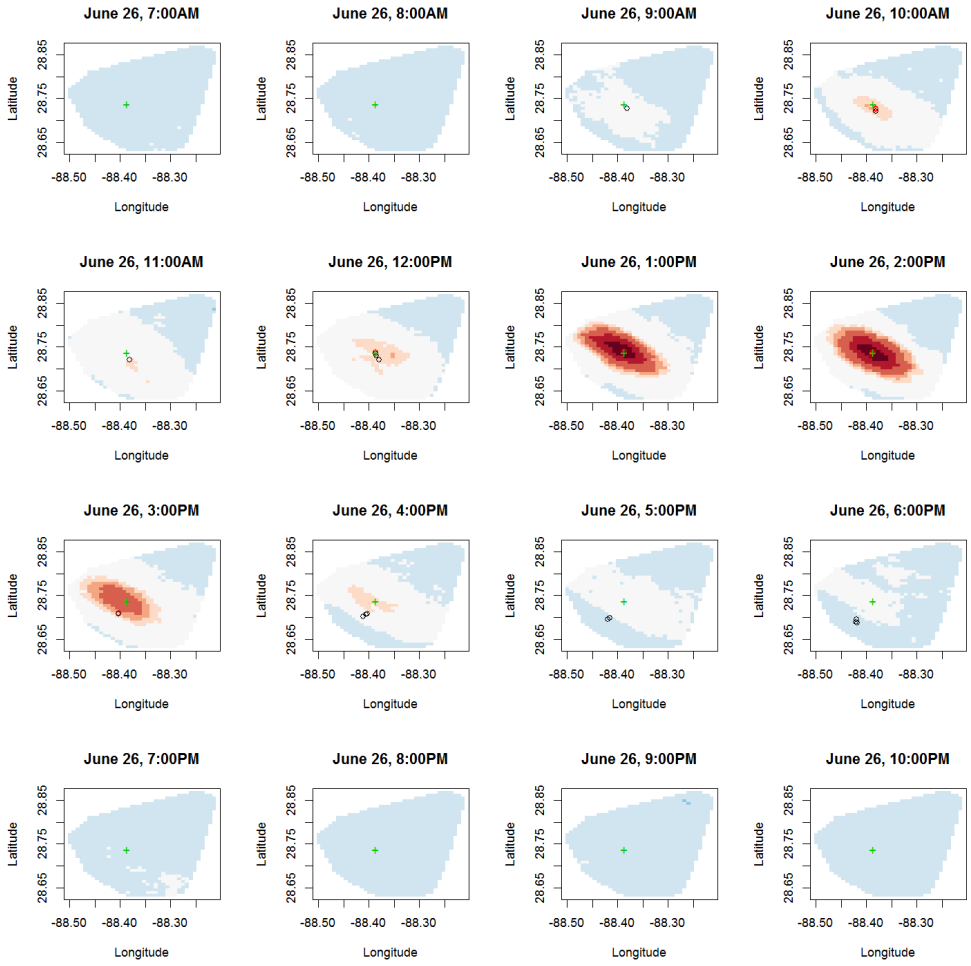


Figure 5.3: Predicted VOC surface on June 26th for sixteen different time points, where blue and white shading denote levels below and near the limit of detection ($\log(0.1) = -2.3$), respectively, and darker shades of red indicate higher levels above detectable limits (up to 2 on the log-scale, or exp(2)=7.39ppm). Here, circles denote observations within 30 minutes of the labeled times, where black circles denote censored observations and red denotes observations above 0.1ppm. Green plus-sign (+) denotes the site of the oil rig.

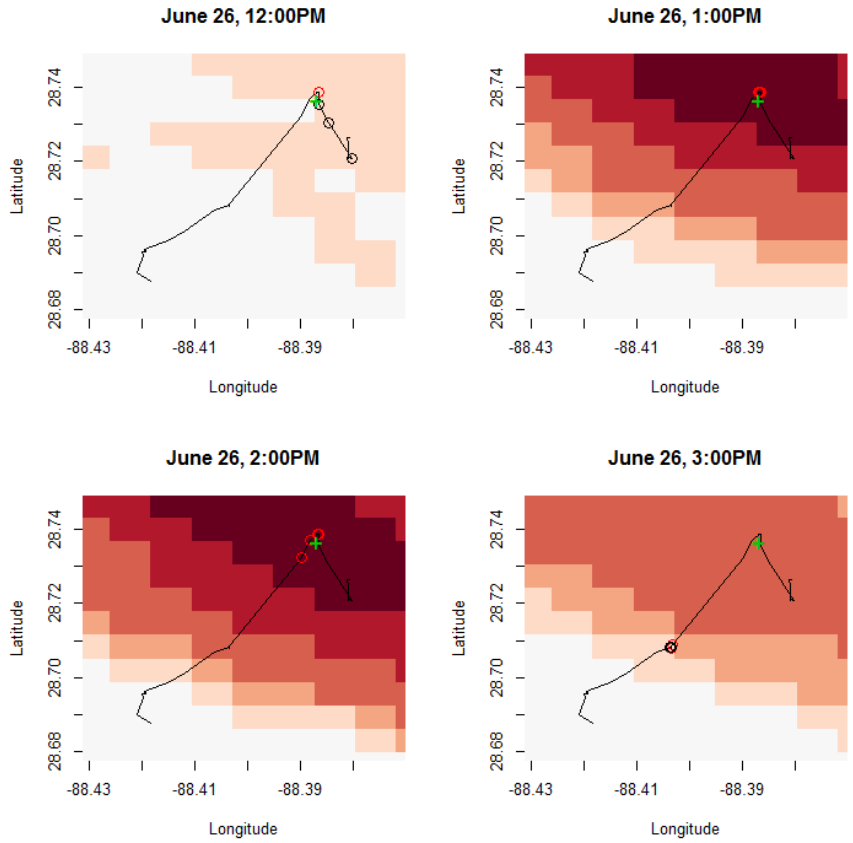
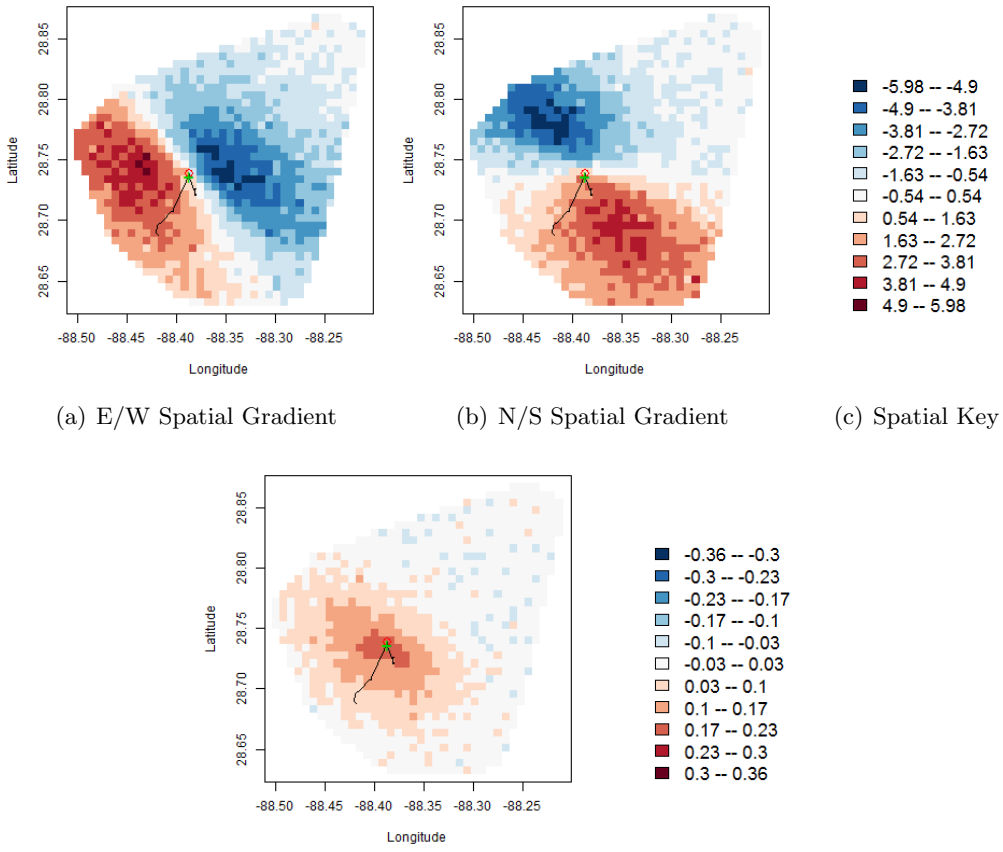


Figure 5.4: Predicted VOC surface on June 26 for four different time points, where white shading denotes levels near the limit of detection ($\log(0.1) = -2.3$) and darker shades of red indicate higher levels above detectable limits (up to 2 on the log-scale, or exp(2)=7.39ppm). Also plotted is a black line denoting the path of the ship collecting measurements at intervals of 12 minutes from 9:00AM to 6:30PM. Here, circles denote observations within 30 minutes of the labeled times, where black circles denote censored observations and red denotes observations above 0.1ppm. Green plus-sign (+) denotes the site of the oil rig. For a sense of scale, the distance between the observations in the fourth panel are roughly 4km away from the site of the spill.



(a) E/W Spatial Gradient (b) N/S Spatial Gradient (c) Spatial Key

(d) Temporal Gradient (e) Temporal Key

Figure 5.5: Estimated spatial and temporal gradients for 1:00PM on June 26 for the BP oil spill data. Negative predicted gradients are shaded in blue and positive gradients in red, with white shading corresponding to values near zero. Also shown here are the path of the ship carrying workers (black line), the location of uncensored and censored observations occurring between 12:30PM and 1:30PM (denoted by red and black circles, respectively), and the site of the *Deepwater Horizon* oil rig (green plus-sign).

Chapter 6

Conclusion

6.1 Summary of major findings

In this thesis, we have extended the concept of spatial gradients (continuous space) to both areally-referenced temporal gradients (discrete space, continuous time) and spatiotemporal gradients (continuous space, continuous time). Through the use of these methods, we believe researchers can gain significant insight into potentially important spatiotemporally varying risk factors that may as of yet be unknown (or at least not accounted for). Furthermore, the gradient process in and of itself can provide valuable information, for instance by being adapted to alert public health officials of dramatically rising pollution levels in a particular region, potentially leading to a reduction in exposure and, ultimately, a reduction in the incidence of poor health outcomes.

In Chapter 2, we focused on the discrete space, continuous time case for the analysis of monthly asthma hospitalization rates in California. Here, we noticed that the overall decreasing trend in hospitalizations was marked by a gradual reduction in the temporal gradients into and out of the summer months, suggesting that the effect of seasonality has been moderated over the 18 year period of the study. We then built upon these results in Chapter 3 by allowing each region to control its own level of smoothing, thereby permitting outlying regions such as Imperial County to avoid being oversmoothed by the effects of their neighbors and be identified by using the proposed outlier diagnostic. In addition to providing improved fit (and subsequently improved gradient estimation), the ability to identify outlying regions can also provide insight into missing covariates

by motivating an investigation into why these regions are different than their neighbors.

Having developed theory for making inference on temporal gradients, in Chapter 4 we then made the natural connection between spatial and temporal gradients by modeling the *spatiotemporal* gradient process, permitting inference on spatiotemporal mixed gradients. Here, we were able to identify the effect of the topography of California on pollution and discussed how inference made from the gradient process could be used for implementing a real-time monitoring system. We then detailed how these methods could be used for modeling censored data in Chapter 5, where we demonstrated its use on the highly censored data collected during the cleanup of the BP oil spill.

6.2 Extensions and future work

In addition to providing immediate important findings, the work in this thesis has also motivated a variety of future projects ranging from practical implementations of the methods presented here to conceptual and theoretical projects.

6.2.1 Computing and software development

In order for these methods to be fully embraced by researchers in non-statistical fields, we need to provide the means with which to implement them conveniently on one's own personal computer. The most obvious way for achieving this would be by creating a freely available `R` package.

Due to the computational benefits of the separable models used in Chapters 2 and 3, these methods can be easily implemented in `R`. For modest sample sizes, one can analyze their data in a matter of hours. For instance, the analyses of the California asthma hospitalization data take only 10 hours to obtain 10,000 MCMC samples for $N = 12,528$, and each data set in the simulation study can be analyzed in 15 minutes (3,000 MCMC samples, $N = 2,900$). Furthermore, these estimates are based on code that has not been optimized to take advantage of advanced computing techniques such as using multiple cores and parallel processing, meaning that additional computational gains can potentially be made.

On the other hand, the methods used in Chapters 4 and 5 can be a bit more computationally burdensome. Despite having substantially fewer observations (e.g., $N = 773$

for the California air quality data), implementing these methods in `R` requires much more effort due to the non-separable model used and the difficulty of inverting "large" matrices. As a result, these methods have been programmed in `C++` and have been designed to utilize the power of today's processors. Unfortunately, this means a considerable amount of effort will be required to make these programs run efficiently on any given computer, and requiring any `R` packages to utilize `C++` calls to retain feasibility.

6.2.2 Investigation of relationship between gradients and missing covariates

When motivating the need for conducting inference on temporal and spatiotemporal gradients in this thesis, we have often claimed that making inference on the gradient process can help researchers identify potential missing covariates, but such a claim has not been formally explored. That is, are there cases where the gradient process can identify areas of sudden change while the residual process gives unclear results?

One means for investigating this question would be to propose a number of missing covariate scenarios and look for telltale signs of each; for instance, an indicator function which discriminates between rush hour and non-rush hour in an analysis of traffic congestion. While we're sure to see an increase and a decrease corresponding to the beginning and end of this period in the residuals, respectively, the gradient process may be much more pronounced. In fact, this may simply be the result of how one *looks* at the residuals; i.e., if you shade your maps using arbitrarily chosen cutoffs (say, quantiles), you may never see sharp changes within a given quantile, particularly in the extremes. On the other hand, one can simply run a complete gradient analysis in a posterior predictive fashion and quickly make a map that highlights significantly positive and negative gradients, immediately providing places to look for more information.

6.2.3 Temporal predictive process models

In Chapters 2 and 3, we discussed methods for analyzing temporal gradients in areally-referenced spatiotemporal data under the assumption that, at each time point, we observed each spatial region. Suppose, however, that this is not the case. More specifically, envision a situation where we have N_s spatial regions, but we only observe a handful of

regions at each time point.

For example, leading up a Presidential election, suppose that statewide polls are conducted once a week, but perhaps the states stagger their polls such that ten state polls are conducted on Monday, ten on Tuesday, etc. One may be tempted to aggregate these polls on a weekly basis, but this ignores the immediate impact on the polls that can occur as a result of each candidate's speeches, advertisements, gaffes, etc., which can occur at any time. Thus analyzing these data on a weekly basis will likely result in a loss of information. That being said, imputing daily values for each state effectively multiplies the size of our data by a factor of 7, which increases the computational burden of this problem by a factor on the order of $7^3 = 343$.

A potential solution to this problem would be to extend the concept of *predictive processes* to the temporal component in our model. First proposed by Banerjee et al. (2008), a predictive process model essentially works by specifying a set of *knots*, $\mathcal{T}^* = \{t_1^*, t_2^*, \ldots, t_{N_t^*}^*\}$, where $N_t^* < N_t$, at which the full covariance model is used to estimate $Z_i(t_j^*)$ — in our election example for instance, one may choose to let each Wednesday be a knot. We then interpolate $Z_i(t_0)$ at any time point, t_0 based on $\mathbf{Z}^* = (\mathbf{Z}_1^{*\prime}, \ldots, \mathbf{Z}_{N_s}^{*\prime})'$, where $\mathbf{Z}_i^* = (Z_i(t_1^*), \ldots, Z_i(t_{N_t^*}))'$. Such a method would effectively reduce the dimension of the problem to a more manageable scale, thus making an otherwise infeasible analysis possible.

References

Adler, R.J. (2009), *The Geometry of Random Fields*, Philadelphia, PA: SIAM—Society for Industrial and Applied Mathematics.

Baladandayuthapani, V., Mallick, B., Hong, M., Lupton, J., Turner, N. and Carroll, R. (2008). Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinoginesis. *Biometrics*, **64**, 64–73.

Banerjee, S. (2010). Spatial gradients and wombling. In *Handbook of Spatial Statistics*, eds. A.E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes. Boca Raton, FL: CRC Press, pp. 559–575.

Banerjee, S., Carlin, B.P. and Gelfand, A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: Chapman and Hall/CRC Press.

Banerjee, S. and Gelfand, A.E. (2003). On smoothness properties of spatial processes. *Journal of Multivariate Analysis*, **84**, 85–100.

Banerjee, S. and Gelfand, A.E. (2006). Bayesian wombling: Curvilinear gradient assessment under spatial process models. *Journal of the American Statistical Association*, **101**, 1787–1501.

Banerjee, S., Gelfand, A.E., Finley, A.O., and Sang, H. (2008). Gaussian predictive process models for large spatial datasets. *Journal of the Royal Statistical Society Series B*, **70**, 825–848.

Banerjee, S., Gelfand, A.E. and Sirmans, C.F. (2003). Directional rates of change under spatial process models. *Journal of the American Statistical Association* **98**, 946–954.

Barbujani, G., Oden, N.L., and Sokal, R.R. (1989). Detecting areas of abrupt change in maps of biological variables. *Systematic Zoology*, **38**, 376–389.

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

Bochner, S. (1955). *Harmonic analysis and the theory of probability*, University of California Press.

Brewer, M. and Nolan, A. (2007). Variable smoothing in Bayesian intrinsic autoregressions. *Environmetrics*, **18**, 841–857.

Brown, P.E., Kåresen, K.F., Roberts, G.O., and Tonellato, S. (2000). Blur-generated non-separable space-time models. *Journal of the Royal Statistical Society, Ser. B*, **62**, 847–860.

California Department of Health Services (2003). California Asthma Facts. http://www.ehib.org/papers/CaliforniaAsthmaFacts010503.pdf

Carlin, B.P. and Louis, T.A. (2009). *Bayesian Methods for Data Analysis*, 3rd ed. Boca Raton, FL: Chapman and Hall/CRC Press.

Christakos, G. (1992). *Random Field Models in Earth Sciences*, San Diego, CA: Academic Press.

Christakos, G. (2000). *Modern Spatiotemporal Geostatistics*, New York: Oxford University Press.

Cressie, N.A.C. (1993). *Statistics for Spatial Data*, 2nd edition. New York: Wiley.

Cressie, N. and Chan, N. (1989). Spatial modeling of regional variables. *Journal of the American Statistical Association*, **84**, 393–401.

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

Cressie, N. and Wikle, C.K. (2011). *Statistics for Spatio-Temporal Data*, 1st ed. Hoboken, NJ: Wiley.

Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics*, **65**, 1254–1261.

Dawid, A.P. and Sebastiani, P. (1999). Coherent dispersion criteria for optimal experimental design. *The Annals of Statistics*, **27**, 65–81.

De Iaco, S., Myers, D.E., and Posa, D. (2002). Space-time variograms and a functional form for total air pollution measurements. *Computational Statistics and Data Analysis*, **41**, 311–328.

Delicado, P., Giraldo, R., Comas, C. and Mateu, J. (2010), Statistics for spatial functional data: some recent contributions. *Environmetrics*, **21**, 224–239.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*, New York: Springer-Verlag.

English, P.B., Von Behren, J., Harnly, M. and Neutra, R.R. (1998). Childhood asthma along the United States/ Mexico border: hospitalizations and air quality in two California counties. *Revista Panamericana de Salud Pública*, **3**, 392–399.

Environmental Protection Agency (2012). Ground-level Ozone: Basic Facts. http://www.epa.gov/air/ozonepollution/basic.html

Fortin, M.J. (1994). Edge detection algorithms for two-dimensional ecological data. *Ecology*, **75**, 956–965.

Fortin, M.J. (1997). Effects of data types on vegetation boundary delineation. *Canadian Journal of Forest Research*, **27**, 1851–1858.

Fortin, M.J. and Drapeau, P. (1995). Delineation of ecological boundaries: Comparisons of approaches and significance tests. *Oikos*, **72**, 323–332.

Freeman, M.F. and Tukey, J.W. (1950). Transformations related to the angular and the square root. *Annals of Mathematical Statistics*, **21**, 607–611.

Gabriel, E., Allard, D., and Barco, J.-N. (2011). Estimating and testing zones of abrupt change for spatial data. *Statistics and Computing*, **21**, 107–120.

Gelfand, A.E. and Banerjee, S. (2010). Multivariate spatial process models. In *Handbook of Spatial Statistics*, eds. A.E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes. Boca Raton, FL: CRC Press, pp. 495–516.

Gelfand, A.E., Banerjee, S. and Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data, *Environmetrics*, **16**, 465–479.

Gelfand, A.E., Smith, A.F.M., and Lee, T.-M. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association*, **87**, 523–532.

Gelman, A., Meng, X.-L., and Stern, H. (1996). Posterior predictive assessment of model fitness via realized discrepancies. *Statistica Sinica*, **6**, 733–807.

Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, **97**, 590–600.

Gneiting, T. and Guttorp, P. (2010). Continuous parameter spatio-temporal processes. In *Handbook of Spatial Statistics*, eds. A.E. Gelfand, P. Diggle, P. Guttorp, and M. Fuentes. Boca Raton, FL: CRC Press, pp. 427–436.

Gneiting, T. and Raftery, A.E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**, 359–378.

Gorman, A. "Imperial County leads state in treatment of children with asthma." *LA Times*, 16 July, 2012. 31 Aug, 2012. <http://articles.latimes.com/2012/jul/16/local/la-me-imperial-county-asthma-20120716>

Handcock, M. and Wallis, J. (1994). An approach to statistical spatiotemporal modelling of meteorological fields. *Journal of the American Statistical Association*, **89**, 368–390.

Hartfield, M. and Gunst, R. (2003). Identification of model components for a class of continuous spatiotemporal models. *Journal of Agricultural, Biological, and Environmental Statistics*, **8**, 105–121.

Jin, X., Banerjee, S., and Carlin, B.P. (2007). Order-free co-regionalized areal data models with application to multiple-disease mapping. *Journal of the Royal Statistical Society, Series B*, **69**, 817–838.

Jones, R.H. and Zhang, Y. (1997). Models for continuous stationary space-time processes. In *Modelling Longitudinal and Spatially Correlated Data*, eds. T.G. Gregoire, D.R. Brillinger, P.J. Diggle, E. Russek-Cohen, W.G. Warren, and R.D. Wolfinger. New York, NY: Springer-Verlag, pp. 289–298.

Lawson, A. and Clark, A. (2002). Spatial mixture relative risk models applied to disease mapping. *Statistics in Medicine*, **21**, 359–370.

Lawson, A.B., Song, H-R., Cai, B., Hossain, M.M., and Huang, K. (2010). Space-time latent component modeling for geo-referenced health data. *Statistics in Medicine*, **29**, 2012–2027.

Liu, J. and Hodges, J.S. (2003). Posterior bimodality in the balanced one-way random effects model. *Journal of the Royal Statistical Society, Series B*, **65**, 247–255.

Lu, H. and Carlin, B.P. (2005). Bayesian areal wombling for geographical boundary analysis. *Geographical Analysis*, **37**, 265–285.

Ma, C. (2003). Families of spatio-temporal stationary covariance models. *Journal of Statistical Planning and Inference*, **116**, 489–501.

MacNab, Y.C. and Gustafson, P. (2007). Regression B-spline smoothing in Bayesian disease mapping: with an application to patient safety surveillance. *Statistics in Medicine*, **26**, 445–4474.

Majumdar, A., Munneke, H.J., Gelfand, A.E., Banerjee, S., and Sirmans, C.F. (2006). Gradients in spatial response surfaces with application to urban land values. *Journal of Business and Economic Statistics*, **24**, 77–90.

Mandl, K.D., Overhage, J.M., Wagner, M.M., Lober, W.B., Sebastiani, P., Mostashari, F., Pavlin, J.A., Gesteland, P.H., Treadwell, T., Koski, E., Hutwagner L., Buckeridge, D.L., Aller, R.D., and Grannis S. (2004). Implementing syndromic surveillance: a practical guide informed by the early experience. *Journal of the American*

Medical Informatics Association. **11**, 141–150.

Manolopoulou, I., Matheu, M.P., Cahalan, M.D., West, M. and Kepler, T.B. (2012). Bayesian spatio-dynamic modeling in cell motility studies: Learning nonlinear toxic fields guiding the immune response. *Journal of the American Statistical Association*, **107**, 855–865.

Mardia, K.V., Kent, J.T., Goodall, C.R. and Little, J.A. (1996). Kriging and splines with derivative information. *Biometrika*, **83**, 207–221.

Martínez-Beneito M.A., López-Quilez A., and Botella-Rocamora P. (2008). An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine*, **27**, 2874-2889.

Morris, M.D., Mitchell, T.J., and Ylvisaker, D. (1993). Bayesian design and analysis of computer experiments: use of derivatives in surface prediction. *Technometrics*, **35**, 243–255.

O'Hagan, A. (1992). Some Bayesian numerical analysis (with discussion). In *Bayesian Statistics 4*, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith. Oxford: Oxford University Press, pp. 345–363.

Pace, R.K., Barry, R., Gilley, O.W., and Sirmans, C.F. (2000). A method for spatiotemporal forecasting with an application to real estate and financial economics. *Journal of Forecasting*, **16**, 229–240.

Pfeifer, P.E. and Deutsch, S.J. (1980a). Independence and sphericity tests for the residuals of space-time ARMA models. *Communications in Statistics – Simulation and Computation*, **9**, 533–549.

Pfeiffer, P.E. and Deutsch, S.J. (1980b). Stationarity and invertibility regions for low order STARMA models. *Communications in Statistics – Simulation and Computation*, **9**, 551–562.

Ramsay, J.O. and Silverman, B.W. (1997). *Functional Data Analysis*, 2nd ed. New York: Springer.

Reich B.J. and Hodges, J.S. (2008). "Modeling longitudinal spatial periodontal data: A spatially adaptive model with tools for specifying priors and checking fit". *Biometrics*, **64**:790–799.

Reis, B.Y., Pagano, M., and Mandl, K.D. (2003). Using temporal context to improve biosurveillance. *Proceedings of the National Academy of Sciences of the United States of America*. **100**, 1961–1965.

Schmid, V. and Held, L. (2004). Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics*, **60**, 1034–1042.

Short, M., Carlin, B.P., and Bushhouse, S. (2002). Using hierarchical spatial models for cancer control planning in Minnesota (United States). *Cancer Causes and Control*, **13**, 903–916.

Spiegelhalter, D.J., Best, N., Carlin, B.P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Ser. B*, **64**, 583–639.

Stein, M.L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer.

Stein, M.L. (2005). Space-time covariance functions. *Journal of the American Statistical Association*, **140**, 310–321.

Stern, H.S. and Cressie, N. (2000). Posterior predictive model checks for disease mapping models. *Statistics in Medicine*, **19**, 2377–2397.

Stoffer, D. (1986) Estimation and identification of space-time ARMAX models in the presence of missing data. *Journal of the American Statistical Association*, **81**, 762-772.

Stroud, J.R., Muller, P. and Sansó, B. (2001). Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society, Ser. B*, **63**, 673–689.

Ugarte, M.D., Goicoa, T., and Militino, A.F. (2010). Spatio-temporal modeling of mortality risks using penalized splines. *Environmetrics*, **21**, 270–289.

Waller, L., Carlin, B.P., Xia, H. and Gelfand, A.E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, **92**, 607–617.

West, M. and Harrison, P. (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer. 2nd edition.

Womble, W.H. (1951). Differential systematics. *Science*, **114**, 315–322.

Zhang, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, **99**, 250–261.

Appendix A

Data Preprocessing Algorithm for Asthma Hospitalization Data

Recall that in the asthma hospitalization data, days with between 1 and 4 asthma hospitalizations are censored for privacy issues. In order to obtain reasonable estimates of the missing, we use an algorithm similar to Besag's Iterated Conditional Modes, but instead calculate iterated conditional *means*. Assuming daily hospitalization counts for the i^{th} county follow a Poisson distribution with rate λ_i, the algorithm depends only on the county's observed counts and a prior distribution for λ_i. This will be done for each county, thus all censored values for a given county will receive the same estimate. We also assume $\lambda_i, i = 1, \ldots, M$ are independent and identically distributed.

A.1 Prior specification

Let $Y_i(t_j) \overset{ind}{\sim} Pois(\lambda_i)$ be the number of hospitalizations on day j in county i and $\mathbf{Y}_i = (Y_i(t_1), Y_i(t_2), \ldots, Y_i(t_D))$ denote the vector of values from the i^{th} county, where D is the number of days in the study, and let $\lambda_i \sim Gam(\alpha, \beta)$. Simple algebra shows

$$\lambda_i \mid \mathbf{Y}_i \sim Gam\left(\sum_j Y_i(t_j) + \alpha, \left(n + \frac{1}{\beta}\right)^{-1}\right), \tag{A.1}$$

which has $E[\lambda_i \mid \mathbf{Y}_i] = \left(\sum_j Y_i(t_j) + \alpha\right)\left(n + \frac{1}{\beta}\right)^{-1}$.

Due to confidentiality, daily counts between 1 and 4 are censored, so in order to continue our analysis, we require an accurate estimated count for these days. Conditioning on this information, it is easily shown that for all censored observations in the i^{th} county,

$$\begin{aligned} E\left[Y_i(t_j) \mid \lambda_i, Y_i(t_j) \in [1,4]\right] &= \frac{\sum_{y=1}^{4} \frac{y e^{-\lambda_i} \lambda_i^y}{y!}}{\sum_{y=1}^{4} \frac{e^{-\lambda_i} \lambda_i^y}{y!}} \\ &= \frac{1 + \lambda_i + \frac{1}{2}\lambda_i^2 + \frac{1}{6}\lambda^3}{1 + \frac{1}{2}\lambda_i + \frac{1}{6}\lambda_i^2 + \frac{1}{24}\lambda_i^3} \end{aligned} \tag{A.2}$$

Naively, one would estimate $E\left[Y_i(t_j) \mid \lambda_i, Y_i(t_j) \in [1,4]\right] = \frac{4+1}{2} = 2.5$, that is, by putting a flat prior over the range. Thus, we wish to choose hyperparameters α and β such that $\lambda^* = E[\lambda] = \alpha\beta$ satisfies $E\left[Y_i(t_j) \mid \lambda^*, Y_i(t_j) \in [1,4]\right] = 2.5$ in (A.2). While this is difficult to do in closed form, with numerical methods one can easily find that $\lambda^* \approx 2.9$. Choosing a prior variance of 100 gives $\beta = \left(\frac{100}{2.9}\right)$ and $\alpha = \frac{2.9^2}{100}$.

A.2 The algorithm

Now that we've found our prior distribution, we can outline how our estimation algorithm proceeds:

Step 1. Set $Y_i(t_j) = 2.5 \; \forall \; Y_i(t_j) \in [1,4]$

Step 2. Let $\hat{\lambda} = E[\lambda_i \mid \mathbf{Y}_i] = \left(\sum_j Y_i(t_j) + \alpha\right)\left(n + \frac{1}{\beta}\right)^{-1}$

Step 3. Let

$$Y_i^*(t_j) = \begin{cases} Y_i(t_j), & \text{if } Y_i(t_j) \notin [1,4] \\ E\left[Y_i(t_j) \mid \hat{\lambda}, Y_i(t_j) \in [1,4]\right] = \frac{1+\hat{\lambda}+\frac{1}{2}\hat{\lambda}^2+\frac{1}{6}\hat{\lambda}^3}{1+\frac{1}{2}\hat{\lambda}+\frac{1}{6}\hat{\lambda}^2+\frac{1}{24}\hat{\lambda}^3}, & \text{if } Y_i(t_j) \in [1,4] \end{cases}$$

Step 4. Calculate $r = \frac{\sum_j Y_i^*(t_j)}{\sum_j Y_i(t_j)}$ and let $Y_i(t_j) = Y_i^*(t_j), j = 1, 2, \ldots, D$. If $r \in [1-\delta, 1+\delta]$ for some prespecified error threshold δ (say, $\delta = 0.01$), then the algorithm is complete. Otherwise, return to Step 2 and repeat.

Convergence follows from Besag (1986) and can be demonstrated via simulation.

Appendix B

Validity of $\mathbf{C_W}(\boldsymbol{\Delta}, \delta)$

The legitimacy of the cross-covariance matrix (4.8) can be established by first constructing the associated finite-difference process and then passing to limits. To be precise, let $\mathbf{U}_h(\mathbf{s},t) = (Z(\mathbf{s},t), Z(\mathbf{s}+h\mathbf{e}_1,t), \ldots, Z(\mathbf{s}+h\mathbf{e}_d,t))^T$ be $(d+1)\times 1$ and let $\mathbf{W}_{1,h}(\mathbf{s},t) = \mathbf{G}_h\mathbf{U}_h(\mathbf{s},t)$, where $\mathbf{G}_h = \begin{bmatrix} 1 & \mathbf{0}' \\ -(1/h)\mathbf{1}_d & (1/h)\mathbf{I}_{d\times d} \end{bmatrix}$. Since $Z(\mathbf{s},t)$ is a Gaussian process, $\mathbf{U}_h(\mathbf{s},t)$ has a nondegenerate Gaussian law for every $h \neq 0$. Therefore, $\mathbf{W}_{1,h}(\mathbf{s},t)$ is a well-defined multivariate Gaussian process because it is a nonsingular linear transformation of $\mathbf{U}_h(\mathbf{s},t)$. If $\mathbf{C}_{1,h}(\boldsymbol{\Delta},\delta)$ is the cross-covariance of $\mathbf{W}_{1,h}(\mathbf{s},t)$, then $\lim_{h\to 0}\mathbf{W}_{1,h}(\mathbf{s},t) = \mathbf{W}_1(\mathbf{s},t)$ is a legitimate multivariate Gaussian process as long as $\lim_{h\to 0}\mathbf{C}_{1,h}(\boldsymbol{\Delta},\delta) = \mathbf{C}_1(\boldsymbol{\Delta},\delta)$ exists, which is, then, the valid cross-covariance matrix of $\mathbf{W}_1(\mathbf{s},t)$. Assuming that our parent spatiotemporal covariance function $K(\boldsymbol{\Delta},\delta)$ is such that $\mathbf{C}_1(\boldsymbol{\Delta},\delta)$ exists, we further construct the $2(d+1)\times 1$ process

$$\mathbf{W}_{h,k}(\mathbf{s},t) = \begin{bmatrix} \mathbf{I}_{d\times d} & \mathbf{O} \\ -(1/k)\mathbf{I}_{d\times d} & (1/k)\mathbf{I}_{d\times d} \end{bmatrix} \begin{bmatrix} \mathbf{W}_{1,h}(\mathbf{s},t) \\ \mathbf{W}_{1,h}(\mathbf{s},t+k) \end{bmatrix}$$

Therefore, for every nonzero h and k, $\mathbf{W}_{h,k}(\mathbf{s},t)$ is a nonsingular linear transformation of a random vector with a nondegenerate Gaussian law and, hence, $\mathbf{W}_{h,k}(\mathbf{s},t)$ is a well-defined process. Let $\mathbf{C}_{h,k}(\boldsymbol{\Delta},\delta)$ be the cross-covariance function for $\mathbf{W}_{h,k}(\mathbf{s},t)$. Then, $\lim_{h,k\to 0}\mathbf{W}_{h,k}(\mathbf{s},t) = \mathbf{W}(\mathbf{s},t)$ is a well-defined multivariate Gaussian process with cross-covariance function $\lim_{h,k\to 0}\mathbf{C}_{h,k}(\boldsymbol{\Delta},\delta) = \mathbf{C}_{\mathbf{W}}(\boldsymbol{\Delta},\delta)$, whenever the latter limit exists. Our choice of $K(\boldsymbol{\Delta},\delta)$, so that its required derivatives exist, ensures that $\mathbf{C}_{\mathbf{W}}(\boldsymbol{\Delta},\delta)$ exists and the spatiotemporal gradient process is well-defined.

The cross-covariance matrix in (4.8) can be constructed by first deriving the cross-covariance matrix of $\mathbf{W}_{h,k}(\mathbf{s},t)$ and then passing to the limit as $h \to 0$ and $k \to 0$. The legitimacy of the finite difference processes ensure that $\mathbf{C_W}(\boldsymbol{\Delta},\delta)$ in (4.8) is valid because it arises as limits of the valid finite-difference cross-covariances. For example, cross-covariance between the mixed spatial gradients $\mathrm{Cov}(\boldsymbol{\nabla}_{st}Z(\mathbf{s},t), \boldsymbol{\nabla}_{st}Z(\mathbf{s}+\boldsymbol{\Delta},t+\delta))$ is obtained as

$$\begin{aligned}
&\lim_{h\to 0}\lim_{k\to 0}\mathrm{Cov}\left[\frac{Z(\mathbf{s}+h\mathbf{u},t+h)-Z(\mathbf{s},t)}{h}, \frac{Z(\mathbf{s}+\boldsymbol{\Delta}+k\mathbf{u},t+\delta+k)-Z(\mathbf{s}+\boldsymbol{\Delta},t+\delta)}{k}\right] \\
&= \lim_{h\to 0}\lim_{k\to 0}\frac{1}{hk}\left[K(\boldsymbol{\Delta}+(k-h)\mathbf{u},\delta+(k-h))\right. \\
&\qquad\qquad \left. -K(\boldsymbol{\Delta}-h\mathbf{u},\delta-h)-K(\boldsymbol{\Delta}+k\mathbf{u},\delta+k)+K(\boldsymbol{\Delta},\delta)\right] \\
&= \lim_{h\to 0}\frac{1}{h}\left[\boldsymbol{\nabla}_{st}K(\boldsymbol{\Delta}-h\mathbf{u},\delta-h)-\boldsymbol{\nabla}_{st}K(\boldsymbol{\Delta},\delta)\right] = -\boldsymbol{\nabla}_{st}\boldsymbol{\nabla}_{st}K(\boldsymbol{\Delta},\delta),
\end{aligned}$$

which has (i,j) element $-(\partial^4/\partial t^2\partial\Delta_i\partial\Delta_j)K(\boldsymbol{\Delta},\delta)$. All the blocks in (4.8) are obtained similarly.

Appendix C

Derivations for $\boldsymbol{\nabla}\mathbf{K}_0$ and $\mathbf{C}_{\nabla Z}(\mathbf{0}, 0)$

To illustrate how to derive $\text{Cov}\{\boldsymbol{\nabla} Z(\mathbf{s}_0, t_0), \mathbf{Z}\} = \boldsymbol{\nabla}\mathbf{K}_0$, we work out the details for deriving $\boldsymbol{\nabla}_s K(\boldsymbol{\Delta}_{i0}, \delta_{0j})$ using the covariance function in (4.10). In order to ease the notation, we again let $A_{0j} = \left(\phi_t^2 |\delta|_{0j}^2 + 1\right)$. Then,

$$\begin{aligned}
\boldsymbol{\nabla}_s K(\boldsymbol{\Delta}_{i0}, \delta_{0j}) &= \text{Cov}(Z(\mathbf{s}_i, t_j), \boldsymbol{\nabla}_s Z(\mathbf{s}_0, t_0)) \\
&= \lim_{h \to 0} \frac{1}{h} \left[\text{Cov}(Z(\mathbf{s}_i, t_j), Z(\mathbf{s}_0 + \mathbf{u}h, t_0) - Z(\mathbf{s}_0, t_0))\right] \\
&= \frac{1}{A_{0j}} \boldsymbol{\nabla}_s \left[\left(1 + \frac{\phi_s ||\boldsymbol{\Delta}_{i0}||}{A_{0j}^{1/2}}\right) \exp\left[-\frac{\phi_s ||\boldsymbol{\Delta}_{i0}||}{A_{0j}^{1/2}}\right]\right] \\
&= -\frac{\phi_s^2}{A_{0j}^2} \exp\left[-\frac{\phi_s ||\boldsymbol{\Delta}_{i0}||}{A_{0j}^{1/2}}\right] \boldsymbol{\Delta}_{i0},
\end{aligned}$$

which is the expression in (4.15). Expressions (4.16) and (4.17) can be derived in a similar fashion, where it is convenient to note that $\partial A_{0j}/\partial\delta = 2\phi_t^2\delta$ and use the chain rule.

The cross-covariance matrix of $\boldsymbol{\nabla} Z(\mathbf{s}, t)$ at $(\mathbf{0}, 0)$, $\mathbf{C}_{\nabla Z}(\mathbf{0}, 0)$, is block diagonal (Section 4.3) and straightforward to derive. Now defining $A = (\phi_t^2|\delta| + 1)$, the first diagonal

block of $\mathbf{C}_{\nabla Z}(\mathbf{0}, 0)$ is

$$\begin{aligned}\mathrm{Cov}(\boldsymbol{\nabla}_s Z(\mathbf{s}, t), \boldsymbol{\nabla}_s Z(\mathbf{s}+\boldsymbol{\Delta}, t+\delta)) &= -\,\boldsymbol{\nabla}_s \left[\boldsymbol{\nabla}_s K(||\boldsymbol{\Delta}||, |\delta|)\right] \\ &= \frac{\phi_s^2}{A^2} \exp\left[-\frac{\phi_s ||\boldsymbol{\Delta}||}{A}^{1/2}\right] \left[I_2 - \frac{\phi_s}{A^{1/2}} \frac{\boldsymbol{\Delta}\boldsymbol{\Delta}^T}{||\boldsymbol{\Delta}||}\right] \\ &\to \phi_s^2 I_2 \text{ as } \boldsymbol{\Delta} \to \mathbf{0},\ \delta \to 0\ .\end{aligned}$$

The remaining diagonal blocks of $\mathbf{C}_{\nabla Z}(\mathbf{0}, 0)$ are obtained similarly.