ABSTRACT

JHUANG, AN-TING. Spatial Signal Detection Using Continuous Shrinkage Priors. (Under the direction of Montserrat Fuentes and Brian J. Reich.)

The statistical problem of identifying spatial regions affected by an experimental factor has many applications, including epidemiology, neuroscience and materials science. Motivated by the problem of detecting changes in two-dimensional X-ray diffraction data, we propose a Bayesian spatial model for sparse signal detection in Chapter 2. Our model places considerable mass near zero and has heavy tails to reflect the prior belief that the image signal is zero for most pixels and large for an important subset. We show that the spatial prior places mass on nearby locations simultaneously being zero, and also allows for nearby locations to simultaneously be large signals. The form of the prior facilitates efficient computing for large images. We conduct a simulation study to evaluate the properties of the proposed prior and show that it outperforms other spatial models. We apply our method in the analysis of X-ray diffraction data from a two-dimensional area detector to detect changes in the pattern when the material is exposed to an electric field.

In Chapter 3 we extend the spatial horseshoe prior to the spatiotemporal setting. Periodontitis is a chronic inflammatory disease that affects the gum tissue and bone supporting the teeth. The whole-mouth average pocket depth has been commonly used as an indicator of periodontal diagnosis, without considering the spatial correlation across teeth. The objective is to flag local regions of deterioration for individual patients simultaneously. We achieve the objective via the spatial horseshoe prior on spatially-varying linear-trend coefficients. We also consider a low-rank representation to capture the non-stationary covariance structure of the PD data and reduce the computational burden so the method can be applied to thousands of subjects. The empirical results exhibit an improvement in prediction when we use shrinkage priors with nonstationary covariance.

Exposure uncertainty is a common concern in health studies such as the periodontal example of Chapter 3. As an example, in Chapter 4 we consider maternal exposure to ambient air pollution during pregnancy has been linked with adverse birth outcomes in recent studies. However, there are remaining uncertainties and challenges in understanding the effects of particulate matter exposure on neonatal health including uncertainty of exposure measurement errors using environmental monitoring data, the variation in

sources of particulate matter, the identification of a critical window of exposure, and the residential mobility. Our access to geocoded addresses for the pregnant women gives us an opportunity to examine potential effects of air pollution on fetal development. In Chapter 4, we investigate the misclassification effect of fine particulate matter due to residential mobility on evaluating the risk of congenital heart defects. Consideration of maternal residence during pregnancy does not lead to significant difference in the empirical results. However, the simulation study shows that the estimation bias and mean squared errors increase as the moving proportion is higher and the moving distance is longer.

© Copyright 2018 by An-Ting Jhuang

All Rights Reserved

Spatial Signal Detection Using Continuous Shrinkage Priors

by
An-Ting Jhuang

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2018

APPROVED BY:

Ana-Maria Staicu

Luo Xiao

Montserrat Fuentes
Co-chair of Advisory Committee

Brian J. Reich
Co-chair of Advisory Committee

DEDICATION

To my parents and brother.

BIOGRAPHY

An-Ting Jhuang was born in Taichung and grew up in Yuanlin, a small town in Taiwan. She graduated from Taichung Girls' Senior High School in 2006 and attended National Chengchi University in Taipei, Taiwan. At NCCU, An-Ting developed an interest in statistics and graduated summa cum laude with Bachelor's and Master's degree in Statistics in 2010 and 2012, respectively. She joined the Department of Statistics at North Carolina State University as a graduate student in 2012. Under the direction of Drs. Montserrat Fuentes and Brian J. Reich, An-Ting earned her Ph.D. in Statistics from NC State in 2018.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisors Dr. Montserrat Fuentes and Dr. Brian J. Reich for their mentoring, support, and inspiration. I have learned so much over the last few years from them. The staff and faculty in the Department of Statistics, North Carolina State University always create an energetic environment and provide assistance as needed. I appreciate all of my friends in the United States and Taiwan for the distractions during difficult times. Last, I want to thank my parents and brother for their endless support and strong belief. I could not have done it without what I have received in these years. Thank you.

TABLE OF CONTENTS

List of Tables	vii
List of Figures	ix
Chapter 1 Introduction	1
1.1 Spatial sparse signal detection	1
1.2 Spatial exposure assignment	2
Chapter 2 Spatial Signal Detection Using Continuous Shrinkage Priors	4
2.1 Introduction	4
2.2 Description of the 2-D XRD data	6
2.3 Model description	8
2.3.1 The univariate horseshoe prior	8
2.3.2 The spatial horseshoe prior (SHP)	8
2.4 Theoretical properties	9
2.4.1 Properties of the spatial horseshoe prior	9
2.4.2 Properties of the spatial horseshoe posterior	12
2.5 Simulation Study	13
2.5.1 Data generation	13
2.5.2 Models	14
2.5.3 Evaluation metrics	16
2.5.4 Results	16
2.6 Analysis of the 2-D XRD data	21
2.6.1 Model comparisons	21
2.6.2 Summary of fitted models	21
2.7 Conclusion	25
Chapter 3 Spatiotemporal Signal Detection Using Continuous Shrinkage Priors	26
3.1 Introduction	26
3.2 Data description	29
3.3 Model description	31
3.3.1 Spatiotemporal model	31
3.3.2 Low-rank representation	32
3.3.3 Prior specification	33
3.4 Computing details	34
3.5 Data application	35
3.5.1 Model comparisons	35
3.5.2 Estimated parameters	36

3.5.3	Summary of fitted models	42
3.6	Discussion	48
Chapter 4 Misclassification Effect of Fine Particulate Matter due to Residential Mobility on Evaluating Risk of Congenital Heart Defects		49
4.1	Introduction	49
4.2	Methods	51
4.2.1	Study population	51
4.2.2	Exposure assignment	51
4.2.3	Model description	53
4.3	Results	53
4.4	Simulation Study	64
4.4.1	Data generation	64
4.4.2	Measures for evaluation	65
4.4.3	Results	65
4.5	Discussion	72
APPENDIX		86
Appendix A	Supplementary Materials for Chapter 2	87
A.1	Theoretical derivations	87
A.1.1	The joint horseshoe distribution	87
A.1.2	Concentration around zero	89
A.1.3	Tail behavior	92
A.2	Computing details	96
A.2.1	Setup	96
A.2.2	Full conditional distributions	97
A.2.3	The Hamiltonian Monte Carlo algorithm	98
A.3	Simulation plots	100

LIST OF TABLES

Table 2.1	Summary of the simulation study under Signal 1 by error variance σ^2 for the Gaussian, spatial horseshoe with quadratic (SHS_quad), spline 1 (SHS_B-spline 1) and spline 2 (SHS_B-spline 2) shrinkage models and the soft-thresholded Gaussian process (STGP) model. .	17
Table 2.2	Summary of the simulation study under Signal 2 by error variance σ^2 for the Gaussian, spatial horseshoe with quadratic (SHS_quad), spline 1 (SHS_B-spline 1) and spline 2 (SHS_B-spline 2) shrinkage models and the soft-thresholded Gaussian process (STGP) model. .	18
Table 2.3	Comparison of prediction on the 2-D XRD data using the Gaussian and spatial horseshoe (SHS) models with k_a basis functions in the angle and k_r basis functions in the radius (SHS1-SHS6). Methods are compared using $100 \times (\text{mean squared error} - 1)$ as MSE, coverage (%) and computing time per 100 iterations in minutes.	22
Table 3.1	Comparison of prediction accuracy for pocket depth (PD) data using the Gaussian and spatial horseshoe (SHS) models using the low-rank representation of the sample covariance and conditional autoregressive (CAR) covariance with the number of basis functions $L = 11, 53$. Methods are compared using mean squared error (MSE), coverage and computing time for 100 Markov Chain Monte Carlo (MCMC) iterations in minutes.	37
Table 4.1	Maternal information for congenital heart defect cases and controls. We consider maternal information. Variables are summarized using counts (percentage) except for age which is summarized using mean (standard deviation).	56
Table 4.2	Model evaluation	57
Table 4.3	Modeling results of the multinomial logit model with weekly exposures using autoregressive (AR) prior for left-ventricular outflow tract obstruction (LVOTO). The asterisk * indicates significance at the level 0.05.	60
Table 4.4	Modeling results of the multinomial logit model with weekly exposures using autoregressive (AR) prior for right-ventricular outflow tract obstruction (RVOTO). The asterisk * indicates significance at the level 0.05.	61
Table 4.5	Modeling results of the multinomial logit model with weekly exposures using autoregressive (AR) prior for conotruncal defects. The asterisk * indicates significance at the level 0.05.	62

Table 4.6	Modeling results of the multinomial logit model with weekly exposures using autoregressive (AR) prior for septal defects. The asterisk * indicates significance at the level 0.05.	63
Table 4.7	Comparison of stationary and mobile model estimates of weekly effects β_w for data simulated with average moving distance $a = 10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30%	66
Table 4.8	Probability (%) of selecting correct model (mobile PM) using DIC and WAIC under β_w with average moving distance $a=10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30%	67
Table 4.9	100×Correlation of weekly $PM_{2.5}$ between stationary and mobile addresses under all simulation scenarios.	67
Table 4.10	Comparison of stationary and mobile model estimates of overall effect $\beta = 0$ for data simulated with average moving distance $a = 10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30%	68
Table 4.11	Probability (%) of selecting correct model (mobile PM) using DIC and WAIC under $\beta = 0$ with average moving distance $a=10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30%	68
Table 4.12	100×Correlation of overall $PM_{2.5}$ between stationary and mobile addresses under all simulation scenarios.	69
Table 4.13	Comparison of stationary and mobile model estimates of overall effect $\beta = 0.5$ for data simulated with average moving distance $a = 10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30%	69
Table 4.14	Probability (%) of selecting correct model (mobile PM) using DIC and WAIC under $\beta = 0.5$ with average moving distance $a=10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30%	70
Table 4.15	Comparison of stationary and mobile model estimates of overall effect $\beta = 1$ for data simulated with average moving distance $a = 10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30%	70
Table 4.16	Probability (%) of selecting correct model (mobile PM) using DIC and WAIC under $\beta = 1$ with average moving distance $a=10, 30$ and 50 km.	71

LIST OF FIGURES

Figure 2.1	The map of the standardized intensity $Y(\mathbf{s})$ (left) with the maximum electric field 2 kV/mm at time $t = 20$ seconds using the sample mean $\hat{\mu}(\mathbf{s})$ and the sample standard deviation $\hat{\sigma}(\mathbf{s})$ of the 100 baseline images without electric field and the smoothed $Y(\mathbf{s})$ (right) using a Gaussian kernel with the bandwidth 0.75 pixels. . .	7
Figure 2.2	Log prior density of β_1 and β_2 in the first quadrant for (a) independent horseshoe priors ($\log P_0$) and (b)-(d) the spatial horseshoe prior ($\log P_1$) with $\rho_\beta \in \{0, 0.5, 0.9\}$	11
Figure 2.3	Induced prior density plot of the complexity measure $df = \text{tr}(\mathbf{I} - \mathbf{K})$ with independent and common shrinkage with $\rho_\beta \in \{0, 0.5, 0.9\}$. .	13
Figure 2.4	True surface $\beta_0(\mathbf{s})$ (left) and a simulated data set $Y(\mathbf{s})$ with error standard deviation $\sigma = 0.5$ (middle) and $\sigma = 2$ (right).	14
Figure 2.5	Simulation results for Signal 2 with $\sigma = 0.5$ for the four models (rows): the first row shows the true signal $\beta_0(\mathbf{s})$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(\mathbf{s})$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models. 19	
Figure 2.6	Simulation results for Signal 2 with $\sigma = 2$ for the four models (rows): the first row shows the true signal $\beta_0(\mathbf{s})$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(\mathbf{s})$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models. 20	
Figure 2.7	Results of the Gaussian model (top row), and the best prediction performance model SHS6 (bottom row) and columns in different statistics; the first column plots the observations and posterior mean of shrinkage process $\lambda(\mathbf{s})$ from the model SHS6, the second column plots the truncated (at ± 0.5) posterior mean of signal $\beta(\mathbf{s})$ to illustrate contrast between the two models, and the third column plots the difference in posterior probability that $\beta(\mathbf{s})$ is positive, i.e. $P(\beta(\mathbf{s}) > 0 \mid \mathbf{Y})_{\text{SHS6}} - P(\beta(\mathbf{s}) > 0 \mid \mathbf{Y})_{\text{Gaussian}}$, and the signal map for the model SHS6 with the Bayesian spatial false discovery rate controlled at 0.05.	23
Figure 2.8	Density plot (left) and quantile-quantile plot (right) of the posterior mean of signal $\beta(\mathbf{s})$ from the Gaussian and SHS6 models.	24
Figure 2.9	Time series plot of the posterior of $D = \sum_{i=1}^n [\beta(\mathbf{s}_i)]^2$ (boxplots) from the model SHS6 applied separately for a bipolar electric field from -2 kV/mm to 2 kV/mm (line).	24

Figure 3.1	Mean and standard deviation of pocket depth (PD) in millimeters across all subjects and their visits during the first two years. Tooth numbering system is described in numbers and texts for the 28 teeth (2 – 15 and 18 – 31) excluding the 4 wisdom teeth. There are 6 locations measured at each tooth numbered from the buccal surface to lingual surface and the mesial surface to distal surface.	30
Figure 3.2	The first four eigenvectors of the sample covariance.	38
Figure 3.3	The fifth to eighth eigenvectors of the sample covariance.	39
Figure 3.4	The first four eigenvectors of the conditional autoregressive (CAR) covariance.	40
Figure 3.5	The fifth to seventh and the twelfth eigenvectors of the conditional autoregressive (CAR) covariance.	41
Figure 3.6	Posterior mean of β_k and posterior probability $P(\beta_{1k} > 0 \mid \mathbf{Y})$, $k = 1, \dots, 168$ for Subject 1 in the Gaussian and SHS models with $L = 53$ eigenvectors.	43
Figure 3.7	Posterior mean of β_k and posterior probability $P(\beta_{2k} > 0 \mid \mathbf{Y})$, $k = 1, \dots, 168$ for Subject 2 in the Gaussian and SHS models with $L = 53$ eigenvectors.	44
Figure 3.8	Density plot (left) and quantile-quantile plot (right) of the posterior mean of site-specific linear-trend coefficient β_{ik} for subjects $i = 1, \dots, 7, 279$ and $k = 1, \dots, 168$ from the Gaussian and SHS models with the low-rank representation of the sample covariance under 90% explained variation.	45
Figure 3.9	The average rejection rate across subjects by teeth from the Gaussian and SHS models with $L = 53$ eigenvectors.	46
Figure 3.10	Posterior mean of the correlation matrix corresponding to the covariances Σ_a and Σ_z from the SHS model using the sample covariance (left column) and conditional autoregressive (CAR) covariance (right column) in the basis function matrix $\mathbf{Q}$	47
Figure 4.1	Health Regions in Texas.	52
Figure 4.2	Plot of weekly $\text{PM}_{2.5}$ at stationary/mobile addresses.	57
Figure 4.3	95% credible intervals of estimated coefficients for weekly $\text{PM}_{2.5}$ at mobile addresses (solid line) and stationary addresses (dashed line) with AR(1) prior.	58
Figure 4.4	95% credible intervals of estimated coefficients for PCs of weekly $\text{PM}_{2.5}$ at mobile addresses (solid line) and stationary addresses (dashed line).	58
Figure 4.5	95% credible intervals of estimated coefficients for overall average of $\text{PM}_{2.5}$ during weeks 2-8 at mobile addresses (solid line) and stationary addresses (dashed line).	59

- Figure A.1 Simulation results for Signal 1 with $\sigma = 0.5$ for the four models (rows): the first row shows the true signal $\beta_0(\mathbf{s})$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(\mathbf{s})$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models.100
- Figure A.2 Simulation results for Signal 1 with $\sigma = 1$ for the four models (rows): the first row shows the true signal $\beta_0(\mathbf{s})$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(\mathbf{s})$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models.101
- Figure A.3 Simulation results for Signal 1 with $\sigma = 2$ for the four models (rows): the first row shows the true signal $\beta_0(\mathbf{s})$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(\mathbf{s})$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models.102
- Figure A.4 Simulation results for Signal 2 with $\sigma = 1$ for the four models (rows): the first row shows the true signal $\beta_0(\mathbf{s})$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(\mathbf{s})$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models.103

Chapter 1

Introduction

Spatial data is now almost everywhere. There is a wide spectrum of statistical methodology established for different types of spatial data (e.g. point-referenced, areal and point-pattern), with a thorough overview in Cressie (2015) and Gelfand et al. (2010). The basic idea behind spatial statistics is that observations are dependent and decreases with the distance between points. The relaxation of the independence assumption better quantifies uncertainties and leads to more accurate and precise estimation and prediction. The application areas are broad including public health, meteorology, forensic science and so on. In this dissertation, we focus on sparse signal detection and misclassified exposure assignment for spatial data.

1.1 Spatial sparse signal detection

The objective of sparse signal detection is to identify a small number of non-zero signals from noise. It can be viewed as a large-scale multiple testing procedure. Computational burden and false discovery rate (Benjamini and Hochberg, 1995) are the two crucial elements of this problem. The corresponding statistical methods have been established followed by the progress of computational power, in both the frequentist and Bayesian frameworks (Mallick and Yi, 2013). It is usual to deal with high-dimensional data due to the advancement in technology. A common application of spatial sparse signal detection is to find problematic regions using biomedical images such as magnetic resonance imaging, diffusion tensor imaging, electroencephalography data etc (Mori and Zhang, 2006; Gryparis et al., 2009; Derado et al., 2013; Liu et al., 2016; Kang et al., 2018).

Distinct from the usual applications in biomedical imaging data, our approach in Chapter 2 was motivated by a material science analysis of two-dimensional X-ray diffraction (2-D XRD) data. The research goal is to examine if the material has structural changes if an electric field is applied. The statistical objective is to develop a powerful technique to identify the locations of signals in two-dimensional images. In Chapter 2, we propose the spatial horseshoe (SHS) prior, the first continuous shrinkage prior for spatial data so far to solve the scientific problem. The SHS prior achieves sparsity and avoids the problems of poor mixing and slow convergence in classical two-component mixture priors (Carvalho et al., 2010; Johnson and Rossell, 2012). A simulation study demonstrates that the proposed method is effective at identifying spatial signals. In the application of the 2-D XRD data, we find improved cross-validation performance compared to other methods.

Next, we extend the SHS prior in spatiotemporal settings for oral health surveillance. The whole-mouth average pocket depth (PD) has been frequently used as an indicator of periodontal diagnosis. However, we can facilitate the intervention if we can specifically identify places in the mouth where periodontal disease is progressing. In Chapter 3, we introduce a Bayesian spatiotemporal model to detect local regions of deterioration for individual patients via the SHS prior on the site-specific linear-trend coefficients. We consider a spatial shrinkage prior to reflect the prior belief that there are usually only a few unhealthy sites in one’s mouth during a short period of time. In addition, the consideration of a low-rank representation captures the nonstationary spatial covariance structure of the PD data and to relieve the computational burden of analyzing the data of thousands of subjects.

1.2 Spatial exposure assignment

Aside from disease surveillance, environmental health is another important issue in public health. Environmental exposures indicate that being in the presence of some substance in an environment such as air pollutants and hazardous waste, which can lead to health problems (Elliott et al., 2001). Namely, we obtain the value of an exposure based on a specific time and space. The associations between exposures and health outcomes are commonly the main focus in studies of environmental epidemiology. The findings can expand scientific discovery and guide the public health policy. However, the empirical

results may not be valid due to confounding, spatial misalignment error and other factors (Sheppard et al., 2012; Sarnat et al., 2013; Avanasì et al., 2016). Various methods have already been introduced to calibrate the measurement error (Molitor et al., 2007; Gryparis et al., 2009; Berrocal et al., 2012; Alexeeff et al., 2016).

On the other hand, it is challenging to incorporate residential mobility and evaluate its effect on the relationship between exposures and health outcomes because lack of residential information or poor data quality. In exposure assignment, it can be problematic to assume that people have not moved during a period, especially for a long time (Canfield et al., 2006; Chen et al., 2010). In Chapter 4, we investigate the effect of misclassification of fine particulate matter exposure due to ignoring residential mobility during pregnancy on the modeled risk for congenital heart defects (CHDs). The availability of geocoded latitude and longitude for the infants in the National Birth Defects Prevention Study (NBDPS) enables us to examine potential effects of air pollution on CHDs. We find that the strong correlation of exposures at mobile and stationary addresses leads to negligible difference in empirical results. In the simulation study, we obtain that the misclassification effect of exposures grows when the moving proportion and distance increase, presented in larger bias and mean squared errors.

Chapter 2

Spatial Signal Detection Using Continuous Shrinkage Priors

2.1 Introduction

X-ray diffraction (XRD) is a powerful technique to characterize the atomic structure of a material. There are nearly 300,000 structures cataloged in the International Center for Diffraction Data (Thomas, 2012; Editorial, 2014). For example, we analyze the XRD pattern of lead zirconate titanate (PZT). PZT is a useful functional material known as a piezoelectric. As a piezoelectric, PZT changes shape when an electric field is applied, making it useful for actuation, positional control and energy conversion. The material we investigate in the present study is a commercially available PZT polycrystalline ceramic material known by the tradename K350 (Piezo Technologies). This material has been studied previously under varying temperature, electric field, pressure and deformation, making it an established material (e.g., Katrusiak, 2008; Dutta and Singh, 2011; Gorfman et al., 2011; Esteves et al., 2015) that is suitable for the present investigation. The data in the present experiment are acquired using a two-dimensional area detector that measures diffracted intensity; this method can be referred to as 2-D XRD. The data are shown in Figure 2.1 and described in detail in Section 2.2. In a 2-D XRD image, each pixel represents the z-score under an electric field relative to several baseline images. There are few z-scores absolutely greater than two, and these pixels tend to cluster in rings indicative of atomic structure. Therefore, we analyze these data using sparse spatial signal detection methods. Our objective is to develop a more powerful statistical method

for detecting and quantifying structural changes in 2-D XRD patterns under different experimental conditions.

The statistical problem of identifying spatial regions affected by an experimental factor has many applications, including epidemiology, neuroscience and materials science. As a result, there is a rich literature on the topic. The spatial scan statistic (Naus, 1965; Kulldorff, 1997; Costa and Kulldorff, 2009; Liao et al., 2017) searches for regions with different means than the background, but does not estimate the signal and is not well suited for large images with many significant subregions. The two-dimensional fused lasso (Friedman et al., 2007) and the graphical lasso (Friedman et al., 2008) are penalized regression methods that account for spatial structure in the signal using penalties to encourage spatial smoothness. Spatial wavelet shrinkage methods impose a threshold on coefficients in the wavelet domain to recover a sparse signal (Donoho and Johnstone, 1994; Taswell, 2000; Jansen, 2001; Yadav et al., 2014; He and Xiang, 2017). These regularization methods can be applied to high-dimensional data, but require presetting the tuning parameters via cross validation (Mallick and Yi, 2013) and fail to account for all sources of uncertainty.

Our approach builds on Bayesian variable selection methods. This allows us to fully account for uncertainty in the posterior distribution and incorporate known atomic structure in the prior. An intuitive sparse prior is a two-component mixture (spike and slab) with one component concentrated near zero for the unimportant features and the other diffuse for the signals (Mitchell and Beauchamp, 1988; George and McCulloch, 1993; Yuan and Lin, 2005; Geweke, 1996; George, 1997; O’hara and Sillapää, 2009; Rokov and George, 2016). Two-component mixture priors can be extended to the spatial setting to identify subregions of interest (Goldsmith et al., 2014; Boehm Vock et al., 2015; Li et al., 2015; Kang et al., 2018). However, the two-component construction is computationally challenging because posterior sampling requires a search over a considerable space of complex models and is plagued by slow convergence and poor mixing (Carvalho et al., 2010; Johnson and Rossell, 2012; Mallick and Yi, 2013).

The computational difficulties of spike-and-slab priors are abated by continuous shrinkage priors (Carvalho et al., 2010; Griffin and Brown, 2010; Armagan et al., 2013; Bhattacharya et al., 2014; Bhadra et al., 2016b; Piironen and Vehtari, 2016). Rather than a discrete mixture over two components, these priors continuously mix over shrinkage parameters. For example, Carvalho et al. (2010) proposed the horseshoe prior that assumes

a normal prior with mean zero and standard deviation that follows a half-Cauchy prior. The horseshoe prior has high concentration around zero for sparsity, heavy tails to avoid excessive shrinkage of signals and attractive theoretical properties (Datta and Ghosh, 2013; van der Pas et al., 2014; Bhadra et al., 2016a; van der Pas et al., 2017).

To our knowledge, we propose the first continuous shrinkage prior for spatial data. We extend the horseshoe prior to account for spatial dependence in the signal at nearby observations. We prove that the proposed spatial horseshoe prior has the univariate horseshoe prior marginally at each location, and that the induced joint distribution for pairs of nearby sites has higher concentration around zero and heavier tails compared to independent horseshoe priors. The form of the continuous shrinkage prior permits simple expressions for the gradients of the posterior, and thus we use the Hamiltonian Monte Carlo (HMC) algorithm (Neal, 1994) for efficient sampling in high-dimensions. A simulation study demonstrates that the proposed method is effective at identifying spatial signals. When applied to the 2-D XRD data, we find improved cross-validation performance compared to other methods.

Chapter 2 is organized as follows. We describe the data in Section 2.2. In Sections 2.3 and 2.4, we introduce the proposed shrinkage prior and show its theoretical properties. In Section 2.5, we evaluate the proposed prior through the simulation study. In Section 2.6, we apply our method to the 2-D XRD data and compare with other methods. We conclude with remarks and comments in Section 2.7.

2.2 Description of the 2-D XRD data

The 2-D XRD data were acquired using high-energy X-rays generated at a synchrotron source (Advanced Photon Source, Argonne National Laboratory). The beamline 11-ID-C was used with an energy of 105 keV. The sample was located in the X-ray beam and a silicon-based detector was placed in the transmitted direction, approximately 2 meters from the sample position. Electrical connections were made from a high voltage power supply to the top conductive electrode of the PZT sample. The bottom electrode was grounded. 2-D XRD patterns were measured at sequentially increasing and then decreasing electric field amplitudes from -2 kV/mm to 2 kV/mm. The objective is to examine if introducing electric field changes the structure of the PZT sample and to capture the diffraction pattern. The detector records data over time which yields images

that are 2048×2048 pixel matrices. The intensity of the XRD measured in counts per second (cps) indicates the relative abundance at each pixel. In addition, we transform the image of size 2048×2048 into size 1023×1023 by removing the first and last rows consisting of all zeros due to experimental setup and extracting all odd rows and columns. We surmise that this step is reasonable in that the intensity is close to the values nearby, and the diffraction pattern is still apparent.

We compare the diffraction pattern of PZT with electric field to that without electric field and investigate the locations where changes occur. There are 100 baseline images recorded without electric field. We define the response $Y(\mathbf{s})$ as the standardized intensity, $Y(\mathbf{s}) = \frac{X(\mathbf{s}) - \hat{\mu}(\mathbf{s})}{\hat{\sigma}(\mathbf{s})}$, where $X(\mathbf{s})$ is the intensity at pixel $\mathbf{s}$, $\hat{\mu}(\mathbf{s})$ is sample mean of the 100 baseline images at pixel $\mathbf{s}$, and $\hat{\sigma}(\mathbf{s})$ is sample standard deviation of the 100 baseline images at pixel $\mathbf{s}$. The yellow rings in Figure 2.1 indicate the larger change of diffraction pattern at $t = 20$ seconds. We choose the image at time $t = 20$ seconds since this is when the maximum electric field is applied.

Figure 2.1: The map of the standardized intensity $Y(\mathbf{s})$ (left) with the maximum electric field 2 kV/mm at time $t = 20$ seconds using the sample mean $\hat{\mu}(\mathbf{s})$ and the sample standard deviation $\hat{\sigma}(\mathbf{s})$ of the 100 baseline images without electric field and the smoothed $Y(\mathbf{s})$ (right) using a Gaussian kernel with the bandwidth 0.75 pixels.

2.3 Model description

2.3.1 The univariate horseshoe prior

Define Y as the response variable and β as the signal variable. The likelihood for Y with horseshoe prior for β is

$$Y \mid \beta \sim N(\beta, \sigma^2), \beta \mid \lambda \sim N(0, \tau^2 \lambda^2), \lambda \sim C^+(0, 1), \quad (2.1)$$

where σ^2 and τ^2 are the error variances, and λ follows the standard half-Cauchy distribution on the positive reals. If $\sigma^2 = 1$, the posterior mean is

$$E(\beta \mid Y) = \int_0^1 (1 - \kappa) Y P(\kappa \mid Y) d\kappa = [1 - E(\kappa \mid Y)] Y, \quad (2.2)$$

where $\kappa = \frac{1}{1+\lambda^2}$ is the shrinkage coefficient. $E(\kappa \mid Y)$ determines the amount of shrinkage towards zero. The name of the prior comes from the horseshoe-shaped $\text{Beta}(\frac{1}{2}, \frac{1}{2})$ distribution on the shrinkage coefficient κ that is induced by the half-Cauchy distribution on λ . The shape of the shrinkage density implies that the shrinkage coefficient is close to either zero or one with high probability which shrinks null signals towards zero and avoids shrinking the true signals. This property facilitates separating signals from the noise. Although marginally over λ this prior for β does not have a closed form of the probability density function, it enjoys nice theoretical properties including tight bounds, high concentration near zero and heavy tails (Carvalho et al., 2010).

2.3.2 The spatial horseshoe prior (SHP)

To extend the horseshoe prior to the spatial setting, define $Y(\mathbf{s})$ as the real-valued response at a spatial location $\mathbf{s} = (s_1, s_2) \in \mathbb{R}^2$ and $\beta(\mathbf{s})$ as the true signal at $\mathbf{s}$. For data observed at locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, define $\mathbf{Y} = [Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)]^T$ as the $n \times 1$ response vector and $\boldsymbol{\beta} = [\beta(\mathbf{s}_1), \dots, \beta(\mathbf{s}_n)]^T$ as the signal vector. Extending (2.1), let

$$\mathbf{Y} \mid \boldsymbol{\beta} \sim N(\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n), \boldsymbol{\beta} \mid \boldsymbol{\lambda} \sim N(\mathbf{0}, \boldsymbol{\Lambda} \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \boldsymbol{\Lambda}), \quad (2.3)$$

where σ^2 is the error variance, $\mathbf{I}_n$ is the identity matrix of size n , $\boldsymbol{\Lambda}$ is diagonal with the diagonal elements $\boldsymbol{\lambda} = [\lambda(\mathbf{s}_1), \dots, \lambda(\mathbf{s}_n)]^T$, and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ is a spatial covariance matrix.

Independence of the response $\mathbf{Y}$ given the signals $\boldsymbol{\beta}$ is a strong assumption, but justified for the 2-D XRD data where the distribution of photons should theoretically follow a Poisson process with independent counts across pixels.

Spatial shrinkage is induced by the spatial process model for $\lambda(\mathbf{s})$. We propose a Gaussian copula model (Nelsen, 2006) that preserves the marginal half-Cauchy distribution and captures spatial dependence,

$$\lambda(\mathbf{s}) = \lambda_0 l[\theta(\mathbf{s})], \quad (2.4)$$

where λ_0 is a global scale parameter, $\theta(\mathbf{s})$ is a latent spatial Gaussian process with mean zero and variance one for all $\mathbf{s}$, $l(\cdot) = F_{C^+}^{-1}[\Phi(\cdot)]$ is the half-Cauchy link function, $F_{C^+}^{-1}(\cdot)$ is the inverse cumulative density function of the half-Cauchy distribution, and $\Phi(\cdot)$ is the standard normal cumulative distribution function. The half-Cauchy link function provides a marginal half-Cauchy distribution for $l[\theta(\mathbf{s})]$ if $E[\theta(\mathbf{s})] = 0$ and $\text{Var}[\theta(\mathbf{s})] = 1$. We interpret $l[\theta(\mathbf{s})]$ as a local shrinkage parameter and λ_0 as a global shrinkage parameter, respectively.

The latent shrinkage process $\theta(\mathbf{s})$ could follow another Gaussian process prior with mean zero, variance one and spatial correlation. We also consider a low-rank representation of the latent shrinkage process

$$\theta(\mathbf{s}) = \sum_{j=1}^J \frac{x_j(\mathbf{s})}{\sqrt{\sum_{j=1}^J x_j^2(\mathbf{s})}} b_j, \quad (2.5)$$

where $x_j(\mathbf{s})$ is the j^{th} basis function for location $\mathbf{s}$ and $b_j \stackrel{iid}{\sim} N(0, 1)$. The spatial model gives the same marginal distribution as (2.1), but with spatial dependence both in the distribution for $\boldsymbol{\beta} \mid \boldsymbol{\lambda}$ and the distribution for $\boldsymbol{\lambda}$. We will elaborate on the properties of this model in Section 2.4.

2.4 Theoretical properties

2.4.1 Properties of the spatial horseshoe prior

In this section we study the properties of the SHP for a pair of spatial locations as a function of the distance between the locations. We compare the SHP with the simple

model with the independent β and λ , namely, $\Sigma_\beta = \mathbf{I}_2$ and $\lambda_1, \lambda_2 \stackrel{iid}{\sim} C^+(0, 1)$. The joint density of β in this special case is (Carvalho et al., 2010)

$$P_0(\beta_1, \beta_2) = \frac{1}{2\pi^3} e^{\frac{1}{2}(\beta_1^2 + \beta_2^2)} E_1\left(\frac{\beta_1^2}{2}\right) E_1\left(\frac{\beta_2^2}{2}\right), \quad (2.6)$$

where $E_1(\cdot)$ is the exponential integral function. Next, we consider the joint shrinkage parameter $\lambda = \lambda(\mathbf{s}_1) = \lambda(\mathbf{s}_2) \sim C^+(0, 1)$ and the covariance Σ_β with the diagonal elements 1 and the spatial correlation ρ_β which stands in for the distance between the two locations. Then we obtain (Supplementary Materials A.1) the joint density

$$P_1(\beta_1, \beta_2) = 2^{-\frac{1}{2}} \pi^{-\frac{3}{2}} (1 - \rho_\beta^2)^{\frac{1}{2}} (\beta_1^2 - 2\rho_\beta \beta_1 \beta_2 + \beta_2^2)^{-\frac{1}{2}} \\ - (2\pi)^{-1} (1 - \rho_\beta^2)^{\frac{1}{2}} e^{\frac{1}{2}(\beta_1^2 - 2\rho_\beta \beta_1 \beta_2 + \beta_2^2)} \operatorname{erfc}\left(\sqrt{\frac{\beta_1^2 - 2\rho_\beta \beta_1 \beta_2 + \beta_2^2}{2}}\right), \quad (2.7)$$

where $\operatorname{erfc}(\cdot)$ is the complementary error function.

Figure 2.2 plots $\log[P_0(\beta_1, \beta_2)]$ and $\log[P_1(\beta_1, \beta_2)]$ for $\rho_\beta = 0, 0.5, 0.9$. When $\rho_\beta = 0$, the log density with a common shrinkage parameter has higher spike around zero and heavier tail than the log density with independent shrinkage parameters. Increasing the ρ_β gives the higher concentration on the 45-degree line of the log density.

Concentration around zero

The densities $P_0(\beta_1, \beta_2)$ and $P_1(\beta_1, \beta_2)$ both have high concentration towards zero. It can be shown that $\lim_{\beta_1, \beta_2 \rightarrow 0} P_0(\beta_1, \beta_2) = \infty$ and $\lim_{\beta_1, \beta_2 \rightarrow 0} P_1(\beta_1, \beta_2) = \infty$. Moreover, we show that the density with joint shrinkage concentrates towards zero more than the density with independent shrinkage by proving that the ratio of $P_1(\beta_1, \beta_2)$ to $P_0(\beta_1, \beta_2)$ diverges as β_1 and β_2 approach zero. Let R be the ratio,

$$R = \frac{P_1(\beta_1, \beta_2)}{P_0(\beta_1, \beta_2)} \\ = \pi^{\frac{3}{2}} e^{-\frac{1}{2}(\beta_1^2 + \beta_2^2)} \left(\frac{\beta_1^2 + \beta_2^2}{2}\right)^{-\frac{1}{2}} \left[E_1\left(\frac{\beta_1^2}{2}\right) E_1\left(\frac{\beta_2^2}{2}\right)\right]^{-1} - \pi^2 \frac{\operatorname{erfc}\left(\sqrt{\frac{\beta_1^2 + \beta_2^2}{2}}\right)}{E_1\left(\frac{\beta_1^2}{2}\right) E_1\left(\frac{\beta_2^2}{2}\right)}. \quad (2.8)$$

We show that $\lim_{\beta_1, \beta_2 \rightarrow 0} R = \infty$ with the divergence rate $(\sqrt{\beta_1^2 + \beta_2^2} \log \beta_1 \log \beta_2)^{-1}$ (Supplementary Materials A.2). This indicates that joint sparsity is achieved more rapidly

(a) Independence

(b) SHP with $\rho_\beta = 0$

(c) SHP with $\rho_\beta = 0.5$

(d) SHP with $\rho_\beta = 0.9$

Figure 2.2: Log prior density of β_1 and β_2 in the first quadrant for (a) independent horseshoe priors ($\log P_0$) and (b)-(d) the spatial horseshoe prior ($\log P_1$) with $\rho_\beta \in \{0, 0.5, 0.9\}$.

in $P_1(\beta_1, \beta_2)$ than in $P_0(\beta_1, \beta_2)$.

Tail behavior

We also investigate the capability to retain signals of large values through examining the tail densities with joint and independent shrinkage. In Figure 2.2, we observe that the tail density is lower in $P_0(\beta_1, \beta_2)$ than $P_1(\beta_1, \beta_2)$ and becomes larger when the spatial dependence is stronger. The higher tail density implies a better capability to avoid excessive shrinkage. Theoretically, it is shown that $\lim_{\beta_1, \beta_2 \rightarrow \infty} R = \infty$ with the divergence rate $\frac{\beta_1^2 \beta_2^2}{\sqrt{\beta_1^2 + \beta_2^2}}$ (Supplementary Materials A.3). This result suggests the higher tail density with joint shrinkage and its stronger power to incorporate large signals.

2.4.2 Properties of the spatial horseshoe posterior

Let the error variance σ^2 be 1, then the posterior mean can be expressed as $E(\boldsymbol{\beta} \mid \mathbf{Y}, \boldsymbol{\lambda}) = (\mathbf{I}_2 - \mathbf{K})\mathbf{Y}$, where the shrinkage matrix $\mathbf{K} = \mathbf{I}_2 - (\mathbf{I}_2 + \boldsymbol{\Omega}_\beta^{-1})^{-1}$ and $\boldsymbol{\Omega}_\beta = \boldsymbol{\Lambda} \boldsymbol{\Sigma}_\beta \boldsymbol{\Lambda}$. To examine the joint shrinkage induced by the SHP, we consider the one-number complexity summary (Efron, 2004)

$$\text{df} = \text{tr}(\mathbf{I} - \mathbf{K}) = \frac{2(1 - \rho_\beta^2)\lambda_1^2\lambda_2^2 + \lambda_1^2 + \lambda_2^2}{(1 - \rho_\beta^2)\lambda_1^2\lambda_2^2 + \lambda_1^2 + \lambda_2^2 + 1}. \quad (2.9)$$

If $\lambda_1 = \lambda_2 = \lambda$, then

$$\text{df} = \frac{2(1 - \rho_\beta^2)\lambda^4 + 2\lambda^2}{(1 - \rho_\beta^2)\lambda^4 + 2\lambda^2 + 1}. \quad (2.10)$$

In this special case with $\rho_\beta = 0$, $\text{df} = \frac{2\lambda^2}{\lambda^2 + 1}$, and thus its prior distribution is under the SHP resembles the horseshoe $\text{Beta}(\frac{1}{2}, \frac{1}{2})$ density scaled by 2. Figure 2.3 plots the density of df. The SHP with common shrinkage has spikes at zero and two corresponding to joint shrinkage. This property encourages noises to be shrunk to zero and signals to be retained at the original values. The density of df with independent λ has a spike around one which corresponds to shrinking only one of β_1 and β_2 , instead of shrinkage or non-shrinkage for both variables. This characteristic even holds when $\rho_\beta = 0.9$ which is contradictory to the expectation that a high correlation between two variables leads to similar shrinkage behaviors.

Figure 2.3: Induced prior density plot of the complexity measure $df = \text{tr}(\mathbf{I} - \mathbf{K})$ with independent and common shrinkage with $\rho_\beta \in \{0, 0.5, 0.9\}$.

2.5 Simulation Study

We investigate how the SHP performs in detecting sparse signals compared with other methods using a simulation study.

2.5.1 Data generation

We consider the two true surfaces:

1. $\beta_0(\mathbf{s}) = 6I[(s_1 - 10)^2 + (s_2 - 12)^2 \leq 1]$
 $+5I[(s_1 - 10)^2 + (s_2 - 12)^2 > 1 \text{ and } (s_1 - 10)^2 + (s_2 - 12)^2 \leq 4]$
 $+4I[(s_1 - 10)^2 + (s_2 - 12)^2 > 4 \text{ and } (s_1 - 10)^2 + (s_2 - 12)^2 \leq 16]$
 $+3I[(s_1 - 10)^2 + (s_2 - 12)^2 > 16 \text{ and } (s_1 - 10)^2 + (s_2 - 12)^2 \leq 36]$

and

2. $\beta_0(\mathbf{s}) = 6I[(s_1 - 10)^2 + (s_2 - 12)^2 \leq 1]$
 $+5I[(s_1 - 10)^2 + (s_2 - 12)^2 > 1 \text{ and } (s_1 - 10)^2 + (s_2 - 12)^2 \leq 4]$
 $+4I[(s_1 - 10)^2 + (s_2 - 12)^2 > 4 \text{ and } (s_1 - 10)^2 + (s_2 - 12)^2 \leq 16]$
 $+3I[(s_1 - 10)^2 + (s_2 - 12)^2 > 16 \text{ and } (s_1 - 10)^2 + (s_2 - 12)^2 \leq 36]$
 $+4I[(s_1 - 28)^2 + (s_2 - 30)^2 \leq 4]$

$$+3I[(s_1 - 28)^2 + (s_2 - 30)^2 > 4 \text{ and } (s_1 - 28)^2 + (s_2 - 30)^2 \leq 16].$$

The observations are generated with $Y(\mathbf{s}) \mid \beta_0(\mathbf{s}) \stackrel{\text{indep}}{\sim} N(\beta_0(\mathbf{s}), \sigma^2)$ on a 40×40 grid of $n = 1,600$ locations. The proportion of signals with nonzero $\beta_0(\mathbf{s})$ is 7.06% and 10.13% in the two surfaces. In addition, we vary the error variance $\sigma^2 = 0.5^2, 1^2, 2^2$. For each combination of true surface and σ^2 , we generate $N = 100$ data sets. Figure 2.4 shows the two surfaces and a representative simulated data set for each surface.

Figure 2.4: True surface $\beta_0(\mathbf{s})$ (left) and a simulated data set $Y(\mathbf{s})$ with error standard deviation $\sigma = 0.5$ (middle) and $\sigma = 2$ (right).

2.5.2 Models

Although the SHP can accommodate other covariance structures, we use the conditionally autoregressive prior (CAR, Carlin and Banerjee, 2003) for Σ_β . The CAR covariance is $\Sigma_\beta = (\mathbf{M} - \rho_\beta \mathbf{A})^{-1}$, where $\mathbf{M}$ is the diagonal matrix with the elements $m_1, \dots, m_n$ indicating the number of neighbors for locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, ρ_β is the spatial dependence parameter, and $\mathbf{A}$ is the adjacency matrix with $\mathbf{A}_{ij} = 1$ if locations $\mathbf{s}_i$ and $\mathbf{s}_j$ are neighbors

and $\mathbf{A}_{ij} = 0$ otherwise. Note that while $\rho_\beta \in (0, 1)$ determines the strength of spatial dependence, it is not a correlation parameter as in Section 2.4. Computational details are given in Supplementary Materials B.

We fit five models for $\beta(\mathbf{s})$ to each simulated dataset. The first four are versions of the spatial horseshoe differentiated by their flexibility in modeling the shrinkage process $\lambda(\mathbf{s})$. The fifth method is soft-thresholded Gaussian process (STGP) model of Kang et al. (2018) that places prior mass exactly at zero.

Gaussian: $\lambda(\mathbf{s})$ as a constant across space, $\lambda(\mathbf{s}) = \lambda_0$ which leads to $\beta \mid \lambda_0 \sim N(\mathbf{0}, \lambda_0^2 (\mathbf{M} - \rho_\beta \mathbf{A})^{-1})$. The error variance σ^2 and the scale parameter λ_0^2 follow uninformative inverse gamma prior $\text{IG}(0.1, 0.1)$, and the spatial dependence parameter ρ_β follows a beta prior $\text{Beta}(10, 1)$. This is the usual Gaussian CAR model.

SHS_quad: Low-rank representation using the half-Cauchy link function with quadratic basis expansions, $\lambda(\mathbf{s}_i) = \lambda_0 l[\theta(\mathbf{s}_i)]$, where $l[\cdot]$ is the half-Cauchy link function, $\theta(\mathbf{s}) = \sum_{j=1}^J \frac{x_j(\mathbf{s})}{\sqrt{\sum_{j=1}^J x_j^2(\mathbf{s})}} b_j$, priors $b_j \stackrel{iid}{\sim} N(0, 1)$, $x_1(\mathbf{s}_i) = 1$, $x_2(\mathbf{s}_i) = \mathbf{s}_{i1}$, $x_3(\mathbf{s}_i) = \mathbf{s}_{i2}$, $x_4(\mathbf{s}_i) = \mathbf{s}_{i1}^2$, $x_5(\mathbf{s}_i) = \mathbf{s}_{i2}^2$ and $x_6(\mathbf{s}_i) = \mathbf{s}_{i1}\mathbf{s}_{i2}$ for $i = 1, \dots, n$.

SHS_B-spline 1: Low-rank representation using the half-Cauchy link function with the B-spline basis expansions with five degrees of freedom for each coordinate, $\lambda(\mathbf{s}_i) = \lambda_0 l[\theta(\mathbf{s}_i)]$, where $l[\cdot]$ is the half-Cauchy link function, $\theta(\mathbf{s}_i) = \sum_{j=1}^{100} x_j(\mathbf{s}_i) b_j$, priors $b_j \stackrel{iid}{\sim} N(0, 1)$, and $x_j(\mathbf{s}_i)$ is the j^{th} product of B-spline bases at location $\mathbf{s}_i$.

SHS_B-spline 2: Same structure as the model SHS_B-spline 1, but with ten degrees of freedom for each coordinate.

STGP: Soft-thresholded Gaussian process prior for $\beta(\mathbf{s})$. Let $\tilde{\beta}(\mathbf{s})$ follow a multivariate normal distribution with zero mean and CAR covariance with first-order neighborhood. Then $\beta(\mathbf{s}) = g_\kappa[\tilde{\beta}(\mathbf{s})]$, where g_κ is the soft-thresholding function to map $\tilde{\beta}(\mathbf{s})$ near zero to exact zero and thus gives a sparse prior,

$$g_\kappa(x) = \begin{cases} 0 & \text{if } |x| \leq \kappa \\ \text{sign}(x)(|x| - \kappa) & \text{if } |x| > \kappa \end{cases}.$$

The thresholding parameter κ follows a uniform prior $U(0, 10)$ and controls the degree of sparsity.

2.5.3 Evaluation metrics

We compare methods in terms of mean squared error (MSE), coverage probability, Type I error and power. For simulated data set k , let $\hat{\beta}_k(\mathbf{s})$ be the posterior mean and $[l_k(\mathbf{s}), u_k(\mathbf{s})]$ be the 95% credible interval of $\beta(\mathbf{s})$. We list the formulae of the measures as follows:

$$\begin{aligned} \text{MSE} &= \frac{1}{Nn} \sum_{k=1}^N \sum_{i=1}^n [\hat{\beta}_k(\mathbf{s}_i) - \beta_0(\mathbf{s}_i)]^2 \\ \text{Coverage} &= \frac{1}{Nn} \sum_{k=1}^N \sum_{i=1}^n I\{\beta_0(\mathbf{s}_i) \in [l_k(\mathbf{s}_i), u_k(\mathbf{s}_i)]\} \\ \text{Type I error} &= \frac{1}{N} \sum_{k=1}^N \frac{\sum_{i=1}^n I\{P[\hat{\beta}_k(\mathbf{s}_i) > 0] > 0.95\} \cdot I[\beta_0(\mathbf{s}_i) = 0]}{\sum_{i=1}^n I[\beta_0(\mathbf{s}_i) = 0]} \\ \text{Power} &= \frac{1}{N} \sum_{k=1}^N \frac{\sum_{i=1}^n I\{P[\hat{\beta}_k(\mathbf{s}_i) > 0] > 0.95\} \cdot I[\beta_0(\mathbf{s}_i) \neq 0]}{\sum_{i=1}^n I[\beta_0(\mathbf{s}_i) \neq 0]}. \end{aligned}$$

2.5.4 Results

The results are given in Tables 2.1 and 2.2. MSE decreases as flexibility of spatial shrinkage process $\lambda(\mathbf{s})$ increases. SHS_B-spline models have the smallest MSE in the low-noise and mid-noise cases while STGP outperforms other models in MSE under the other scenario. Coverage and type I error are at or near the nominal level for all methods except for the Gaussian model in the low-noise cases. All models have strong power when the error variance is small. However, large error variance distinguishes the models. STGP model has about 66% the power of SHS models. The loss of power in STGP model results from excessive shrinkage from the soft-thresholding formulation, especially in the large-noise scenario.

Figures 2.5 and 2.6 illustrate the simulation results under Signal 2 with $\sigma = 0.5$ and 2, respectively. Increasing the flexibility of the shrinkage process improves signal identification and leads to a smoother signal surface. The Gaussian model tends to have more false positives, especially when the error variance is small. This explains why it has strong power but poor MSE. The simulation plots under other scenarios are shown in Supplementary Materials C. For the computing time, analyzing a simulated data set takes 0.33, 0.87, 1.23, 3.40 and 10.29 minutes for Gaussian, SHS_quad, SHS_B-spline 1,

SHS_B-spline 2 and STGP models, respectively.

Table 2.1: Summary of the simulation study under Signal 1 by error variance σ^2 for the Gaussian, spatial horseshoe with quadratic (SHS_quad), spline 1 (SHS_B-spline 1) and spline 2 (SHS_B-spline 2) shrinkage models and the soft-thresholded Gaussian process (STGP) model.

Statistics	Model	$\sigma^2 = 0.5^2$		$\sigma^2 = 1^2$		$\sigma^2 = 2^2$	
		Estimate	SE	Estimate	SE	Estimate	SE
100×MSE	Gaussian	10.95	0.07	17.95	0.14	27.33	0.27
	SHS_quad	4.99	0.03	10.21	0.09	19.30	0.24
	SHS_B-spline 1	3.88	0.06	8.44	0.14	18.06	0.40
	SHS_B-spline 2	4.55	0.06	9.75	0.09	19.74	0.30
	STGP	5.83	0.10	9.66	0.29	15.92	0.27
Coverage (%)	Gaussian	91.67	0.17	97.46	0.05	98.66	0.05
	SHS_quad	98.09	0.04	98.68	0.03	99.28	0.03
	SHS_B-spline 1	97.91	0.51	97.78	0.68	98.61	0.60
	SHS_B-spline 2	97.61	0.36	97.31	0.08	98.22	0.13
	STGP	98.15	0.07	99.48	0.03	99.62	0.02
Type I error (%)	Gaussian	7.25	0.14	2.60	0.06	1.18	0.05
	SHS_quad	1.79	0.05	1.27	0.04	0.79	0.04
	SHS_B-spline 1	1.73	0.40	1.91	0.55	1.06	0.19
	SHS_B-spline 2	1.96	0.23	2.27	0.09	1.44	0.15
	STGP	0.56	0.03	0.15	0.02	0.01	0.00
Power (%)	Gaussian	100.00	0.00	99.97	0.02	98.00	0.14
	SHS_quad	100.00	0.00	99.86	0.03	95.36	0.25
	SHS_B-spline 1	100.00	0.00	99.86	0.03	95.65	0.25
	SHS_B-spline 2	100.00	0.00	99.94	0.02	96.60	0.27
	STGP	100.00	0.00	98.48	0.15	67.85	0.62

Table 2.2: Summary of the simulation study under Signal 2 by error variance σ^2 for the Gaussian, spatial horseshoe with quadratic (SHS_quad), spline 1 (SHS_B-spline 1) and spline 2 (SHS_B-spline 2) shrinkage models and the soft-thresholded Gaussian process (STGP) model.

Statistics	Model	$\sigma^2 = 0.5^2$		$\sigma^2 = 1^2$		$\sigma^2 = 2^2$	
		Estimate	SE	Estimate	SE	Estimate	SE
100×MSE	Gaussian	13.73	0.09	24.25	0.17	37.25	0.33
	SHS_quad	7.56	0.05	15.75	0.12	28.30	0.29
	SHS_B-spline 1	6.19	0.04	13.57	0.18	27.90	0.50
	SHS_B-spline 2	5.97	0.05	13.45	0.10	27.72	0.38
	STGP	5.83	0.10	18.76	0.39	24.69	0.38
Coverage (%)	Gaussian	88.09	0.24	96.47	0.07	98.59	0.05
	SHS_quad	96.74	0.08	97.74	0.06	98.76	0.05
	SHS_B-spline 1	97.33	0.11	97.75	0.22	98.29	0.40
	SHS_B-spline 2	97.40	0.12	96.87	0.07	98.13	0.08
	STGP	98.15	0.07	98.42	0.07	99.31	0.03
Type I error (%)	Gaussian	9.79	0.17	3.67	0.08	1.39	0.05
	SHS_quad	2.87	0.08	2.07	0.07	1.08	0.05
	SHS_B-spline 1	2.40	0.09	1.85	0.14	1.45	0.35
	SHS_B-spline 2	2.19	0.09	2.53	0.09	1.36	0.08
	STGP	0.56	0.03	0.65	0.05	0.03	0.01
Power (%)	Gaussian	100.00	0.00	99.90	0.02	94.93	0.21
	SHS_quad	100.00	0.00	99.79	0.04	92.69	0.27
	SHS_B-spline 1	100.00	0.00	99.66	0.05	92.43	0.32
	SHS_B-spline 2	100.00	0.00	99.79	0.03	92.65	0.33
	STGP	100.00	0.00	98.91	0.09	64.36	0.56

Figure 2.5: Simulation results for Signal 2 with $\sigma = 0.5$ for the four models (rows): the first row shows the true signal $\beta_0(\mathbf{s})$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(\mathbf{s})$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models.

Figure 2.6: Simulation results for Signal 2 with $\sigma = 2$ for the four models (rows): the first row shows the true signal $\beta_0(\mathbf{s})$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(\mathbf{s})$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models.

2.6 Analysis of the 2-D XRD data

2.6.1 Model comparisons

In this section we apply the proposed model in Section 2.3 to the 2-D XRD data. We tailor the SHP model to the 2-D XRD data to capture the ring-shaped pattern visible in Figure 2.1. Rather than using the basis expansions of square coordinates in the simulation study, we consider a basis expansion of the radius, $r \geq 0$, and the angle, $a \in [0, 2\pi)$, from the central point. We use Fourier and B-spline basis expansions for angle and radius, respectively. Let $A_1(a), \dots, A_{k_a}(a)$ be the Fourier basis functions and $B_1(r), \dots, B_{k_r}(r)$ be the B-spline basis functions, where k_a and k_r are the numbers of basis functions in the angle and the radius, respectively. The Fourier basis functions are $A_1(a) = \sin(a)$, $A_2(a) = \cos(a)$, $A_3(a) = \sin(2a)$, and $A_4(a) = \cos(2a)$, etc. At pixel $\mathbf{s}$, the basis functions consist of $J = k_a \times k_r$ products of Fourier and B-spline basis expansions, i.e. $\mathbf{x}(\mathbf{s}) = [x_1(\mathbf{s}), \dots, x_J(\mathbf{s})]^T = \{A_1[a(\mathbf{s})] \times B_1[r(\mathbf{s})], \dots, A_{k_a}[a(\mathbf{s})] \times B_{k_r}[r(\mathbf{s})]\}^T$. The priors and computing details remain the same as in Section 2.5.

We implement five-fold cross validation to evaluate the prediction in models with varying flexibility of shrinkage across space. We consider $(k_a, k_r) = (4, 10), (4, 25), (4, 50), (4, 100), (8, 5)$ and $(8, 50)$. Table 2.3 presents the results as MSE-1, since for the z-scores with independent standard normal error, one is the lowest achievable prediction MSE. Generally, predicted MSE declines as flexibility of shrinkage grows. However, a larger number of basis functions of angle tends to have a smaller prediction error when the total number of basis expansions is the same, e.g. SHS1 versus SHS5 and SHS3 versus SHS6. According to Wilcoxon signed rank test, MSE of the models SHS1-SHS6 is significantly less than MSE of the Gaussian model. Coverage is close to the nominal level 95% for all models.

2.6.2 Summary of fitted models

In this section we fit the Gaussian and SHS models as above to the full dataset. Because the number of pixels is large, we implement the Bayesian spatial false discovery rate (BSFDR) procedure (Sun et al., 2015) with rate 0.05 to control for multiple testing. We consider the one-sided null and alternative hypotheses $H_0 : \beta(\mathbf{s}) \leq 0$ and $H_1 : \beta(\mathbf{s}) > 0$, respectively. The BSFDR procedure provides a critical probability that the null

Table 2.3: Comparison of prediction on the 2-D XRD data using the Gaussian and spatial horseshoe (SHS) models with k_a basis functions in the angle and k_r basis functions in the radius (SHS1-SHS6). Methods are compared using $100 \times (\text{mean squared error} - 1)$ as MSE, coverage (%) and computing time per 100 iterations in minutes.

Model	k_a	k_r	MSE		Coverage (%)		Computing time (mins)
			Estimate	SE	Estimate	SE	
Gaussian	-	-	7.42	0.17	94.72	0.02	2.78
SHS1	4	10	7.31	0.17	94.78	0.02	6.57
SHS2	4	25	7.27	0.17	94.79	0.02	9.91
SHS3	4	50	7.23	0.17	94.80	0.03	17.45
SHS4	4	100	7.21	0.17	94.80	0.02	41.61
SHS5	8	5	7.24	0.17	94.80	0.02	7.38
SHS6	8	25	7.12	0.18	94.84	0.02	8.62

hypothesis is rejected if the posterior probability is greater than the critical probability, other than the nominal value 0.95. The critical probabilities are 92.93% for all models, except for 91.92% for the SHS4 model. The number of pixels for which H_{0s} is rejected are also similar for the Gaussian and SHS models, from 1.42% to 2.02%.

Figure 2.7 displays the fitted results for the Gaussian and SHS6 (with the lowest MSE) models. The posterior mean of the shrinkage parameter $\lambda(\mathbf{s})$ (Figure 2.7, bottom left) is large for only a few radii and for only a subset of the angles on the rings. These areas also have the largest posterior mean of $\beta(\mathbf{s})$ and posterior probability that $\beta(\mathbf{s})$ is positive. The map of difference in posterior probability (Figure 2.7, top right) exhibits that the model SHS6 encourages shrinkage along the rings. Comparing models, we find more significant pixels and stronger spatial clustering of the signal in the SHS model compared to the Gaussian model. In addition, Figure 2.8 illustrates that the density of posterior mean $\beta(\mathbf{s})$ from the SHP model (SHS6) has higher concentration around zero and heavier tails than the density of $\beta(\mathbf{s})$ from the Gaussian model.

Figure 2.7 is an analysis of one time point ($t = 20$). We investigate the temporal trend of signals by fitting the model SHS6 for the standardized intensity $Y(\mathbf{s})$ at all time points. We use the sum of squared signals, $D = \sum_{i=1}^n [\beta(\mathbf{s}_i)]^2$, as a measure of strength of structural change of the PZT sample. Figure 2.9 displays the boxplots of the posterior of D and the electric field by time. Generally, the structural change becomes larger when the magnitude of electric field increases. The measure D is smaller in the first half of

Figure 2.7: Results of the Gaussian model (top row), and the best prediction performance model SHS6 (bottom row) and columns in different statistics; the first column plots the observations and posterior mean of shrinkage process $\lambda(\mathbf{s})$ from the model SHS6, the second column plots the truncated (at ± 0.5) posterior mean of signal $\beta(\mathbf{s})$ to illustrate contrast between the two models, and the third column plots the difference in posterior probability that $\beta(\mathbf{s})$ is positive, i.e. $P(\beta(\mathbf{s}) > 0 \mid \mathbf{Y})_{\text{SHS6}} - P(\beta(\mathbf{s}) > 0 \mid \mathbf{Y})_{\text{Gaussian}}$, and the signal map for the model SHS6 with the Bayesian spatial false discovery rate controlled at 0.05.

experimental period under a positive charge compared to the second half of period under a negative charge. The reason for the rapid increase in D between 54 and 58 seconds is a material phenomenon known as ferroelectric switching; the electric field amplitude at this time exceeds that required for substantial reorientation of material elements, leading to a significant change in the diffraction pattern. The example of this method on ferroelectrics demonstrates the versatility of the method at efficiently assessing 2-D XRD images for the purposes of materials characterization.

Figure 2.8: Density plot (left) and quantile-quantile plot (right) of the posterior mean of signal $\beta(s)$ from the Gaussian and SHS6 models.

Figure 2.9: Time series plot of the posterior of $D = \sum_{i=1}^n [\beta(s_i)]^2$ (boxplots) from the model SHS6 applied separately for a bipolar electric field from -2 kV/mm to 2 kV/mm (line).

2.7 Conclusion

In this paper, we propose a new method for sparse signal detection with application to the 2-D XRD imaging data. To our knowledge, the SHP is the first continuous shrinkage prior for spatial data. Theoretically, we prove that the SHP has the univariate horseshoe prior marginally at each location, and that the induced joint distribution for pairs of nearby sites has higher concentration around zero and heavier tails compared to independent horseshoe priors. Further, we facilitate the computation via the HMC algorithm (Neal, 1994) for high-dimensional data. The simulation and empirical results both show the improvement of estimation and prediction when we consider spatial dependence.

A potential limitation of our method is that the covariance structure of data is assumed to be independent given the mean vector β . This is sufficient for the 2-D XRD imaging data because the distribution of photons ought to follow a Poisson process with independent counts across pixels, but it may not hold for different data. A second restriction is the assumption of CAR covariance structure with the first order neighbors. The covariance structure can be modified to meet needs in other situations.

In the future, the proposed method can be extended to Bayesian variable selection in spatial linear regression. In spatial regression, the covariates effect can vary by site (Gelfand et al., 2003), and the SHP for the spatially-varying effects would encourage sparsity in these effects. We can also extend the SHP to a spatio-temporal horseshoe model for longitudinal data. We can use the extended model to test for spatio-temporal anomalies in a surveillance study (e.g. Li et al., 2012).

Chapter 3

Spatiotemporal Signal Detection Using Continuous Shrinkage Priors

3.1 Introduction

Periodontitis is a chronic inflammatory disease that affects the gum tissue and bone supporting the teeth. In the United States during 2009 and 2010, 47.2% of adults aged 30 years and older have some form of periodontal disease, which influences approximately 64.7 million people (Eke et al., 2012). Recent studies have shown that periodontitis has an association with oral cancer (Cheng et al., 2017), cardiovascular diseases (Persson and Persson, 2008; Berlin-Broner et al., 2016; Nicolosi et al., 2017), inflammatory bowel diseases (Vavricka et al., 2013; Johannsen et al., 2015; Agossa et al., 2017; Papageorgiou et al., 2017) and other diseases (Long et al., 2017; Ferreira et al., 2017). The disease can lead to adverse dental outcomes and rising insurance premiums if patients do not receive treatments in time.

Sensitive methods to identify periodontal disease progression could minimize its effects over an at-risk patients lifetime. The development of oral health surveillance was introduced in Beltrán-Aguilar et al. (2003). As an indicator of periodontal diagnosis, the whole-mouth average pocket depth (PD) has been commonly used in analyses, without considering spatial correlation across teeth. If we can specifically detect places in the mouth where periodontal disease is progressing, we can intervene with treatments including scaling and root planing, medication and surgery. Our goal is to flag regions where the disease is progressing.

Disease surveillance has been a crucial and popular topic in epidemiology. Defined by the World Health Organization, it is the ongoing systematic collection, collation, analysis and interpretation of data and dissemination of information in order for action to be taken (World Health Organization, 2007). Early detection of anomalies can enhance coordination and control activities, especially useful for infectious diseases (Banos and Lacasa, 2007). Disease surveillance methods stemmed from identifying an unusual temporal pattern during a study period. The incorporation of spatial information strengthens the power of surveillance, and can localize outbreaks of a disease or characterize variations in regional patterns (Unkel et al., 2012).

In general, there are test-based and model-based approaches for space-time disease surveillance. Statistical tests for outbreak detection have a long history. First, the spatiotemporal interaction suggests possibility of etiological mechanism for infectious diseases or presence of additional causative agents for non-infectious diseases (Aldstadt, 2007). The tools to test space-time interaction include Knox test (Knox and Bartlett, 1964), Mantel’s test (Mantel, 1967) and k nearest neighbor test (Jacquez, 1996). The second type is the cumulative sum (CUSUM) methods for detecting a change point in an underlying process. Rogerson (1997) and Rogerson and Yamada (2004) extended the CUSUM methods in a spatial setting, without and with consideration of spatial dependence across sub-regions, respectively. However, CUSUM methods are sensitive to outliers, require heavy computing for the threshold parameter, and lead to difficulty in interpretation in a more complicated setting such as the consideration of spatial dependence. The space-time scan statistic (Kulldorff, 2001; Kulldorff et al., 2005) is another popular method and inspired a Bayesian test to detect unusual temporal patterns in small area data (Li et al., 2012). This type of method identifies disease clusters in space and time, and provides more information than the space-time interaction tests. More statistical tests for outbreak detection can be found in Robertson et al. (2010) and Rogerson and Jacquez (2016).

Compared to test-based methods, model-based approaches estimate disease risk and include other variables of interest, and thus provide better insight into etiology, spread, prediction and control of the disease. A variety of spatiotemporal regression methods have been proposed. Most methods of disease surveillance are based on aggregated data (e.g. count data in specified times and areas). It is usually assumed that the number of cases in an area during a time period follows a Poisson distribution. Various methods

have been established upon distinction in estimating the Poisson mean parameter, such as residual-based approach (Vidal Rodeiro and Lawson, 2006), exponentially weighted moving average approach (Zhou and Lawson, 2008), hidden Markov models (Watkins et al., 2009; Heaton et al., 2012) and others (KnorrHeld, 2000; Lawson et al., 2004; Corberán-Vallet and Lawson, 2011). Some other models were built on a Binomial distribution of the number of cases in a given area and time point (Diggle et al., 2004; Kleinman et al., 2004; Lawson et al., 2004; Abellan et al., 2008). The recent development of disease surveillance methods incorporates different sources of data such as electronic health records and social media data (Hay et al., 2013; Lee et al., 2016; Simonsen et al., 2016).

Distinct from most surveillance methods on aggregated data, there are various methods developed using individual-specific periodontal data to address oral health surveillance problems. The previous methods in spatial settings consider cross-section data while accounting for complex covariance structures (Reich et al., 2007; Jin et al., 2016), informative missingness (Reich and Bandyopadhyay, 2010; Reich et al., 2013) and non-Gaussian responses (Reich et al., 2013; Cai and Bandyopadhyay, 2017). Reich and Hodges (2008) developed a spatiotemporal model to monitor periodontal progression using a nonstationary dependence structure in time and space. However, it is time-consuming if we fit the model to thousands of subjects and does not exploit sparsity or shrinkage towards zero.

We propose a Bayesian spatiotemporal model to detect problematic sites in the mouth via the spatial horseshoe (SHS; Jhuang et al., 2017) prior on the site-specific linear-trend coefficients. We consider a shrinkage prior to reflect the prior belief that there are usually only a few unhealthy sites in one’s mouth during a short period of time. As the first spatial continuous shrinkage prior, the SHS prior incorporates spatial dependence in the signal at nearby observations. Jhuang et al. (2017) proved that the SHS prior has nice theoretical properties such as high concentration around zero for sparsity and heavy tails to avoid excessive shrinkage. Moreover, spatial continuous shrinkage priors abate the problems of poor mixing and slow convergence using the two-component mixture priors in a spatial setting (Goldsmith et al., 2014; Boehm Vock et al., 2015; Li et al., 2015; Kang et al., 2018). Second, we consider a low-rank representation to capture the nonstationary spatial covariance structure of the PD data and to relieve the computational burden for thousands of subjects.

Chapter 3 proceeds as follows. We describe the data in Section 3.2. In Section 3.3 we introduce the spatiotemporal model for sparse signal detection. We delineate the computing details in Section 3.4. In Section 3.5, we apply our method to the PD data. We conclude with remarks and comments in Section 3.6.

3.2 Data description

Our data consists of PD observations for 25,763 subjects in suburban Minneapolis. The data were collected by Health Partners, a non-profit health care provider. The subjects were over 18 years old as of January 1, 2007, with continuous dental insurance coverage during the sampling years, 2007–2014. PD is defined as the distance from the gingival margin to the base of gingival pocket. PD was measured at 168 sites in a subject’s mouth and was recorded as an integer $PD \in \{0, 1, \dots, 9\}$. To obtain more comprehensive coverage of a tooth, there are 6 locations measured for each tooth from the mesial to distal (i.e. anterior to posterior) part on the buccal side which is contiguous to the inside of the cheek, and then the same procedure on the lingual side which is adjacent to the tongue side. Figure 3.1 illustrates the tooth numbering system and measured locations for each tooth excluding wisdom teeth.

The objective is to detect unhealthy sites at an early stage before severe progression. Hence, we use the data during the first two years for each subject. The short period also makes the assumption of a linear change of PD in time reasonable. We exclude subjects with less than four visits in the first two years. There are 7,279 subjects satisfying this condition with the number of visits ranging from 4 to 8. Figure 3.1 displays the average and standard deviation of PD among the 7,279 subjects. The average PD is between 1.1 mm and 3.0 mm, higher in the posterior side than the anterior side. The variation of PD exhibits a similar pattern, with the standard deviation ranging from 0.7 mm to 1.8 mm.

(b) Standard deviation of pocket depth

Figure 3.1: Mean and standard deviation of pocket depth (PD) in millimeters across all subjects and their visits during the first two years. Tooth numbering system is described in numbers and texts for the 28 teeth (2 – 15 and 18 – 31) excluding the 4 wisdom teeth. There are 6 locations measured at each tooth numbered from the buccal surface to lingual surface and the mesial surface to distal surface.

3.3 Model description

3.3.1 Spatiotemporal model

Define y_{ijk} as the measured PD in millimeters for subject $i = 1, \dots, n_p$ at visit $j = 1, \dots, n_{vi}$ and at site $k = 1, \dots, n_s$, where $n_p = 7,279$ is the number of subjects, $n_s = 168$ is the number of sites, and n_{vi} is the number of visits for subject i . PD is recorded as an integer between 0 mm to 9 mm. Let y_{ijk}^* be a latent variable for subject i at site k and the j^{th} visit that is related to the observed PD as $y_{ijk} = \max\{[y_{ijk}^*], 0\}$, where $[x]$ is the nearest integer to x . We use the data at the visits in the first two years and assume a latent linear trend of PD in time. We propose a spatiotemporal model for the latent PD. For subject i ,

$$y_{ijk}^* = \alpha_{ik} + \beta_{ik} t_{ij} + \varepsilon_{ijk}, \quad (3.1)$$

where α_{ik} is the baseline PD at site k , β_{ik} is the slope at site k , t_{ij} is the years since baseline for the j^{th} visit, and ε_{ijk} is the random error.

To capture the prior belief that only a few sites which have deteriorated during the study period, we assume that the slope β_{ik} marginally follows a horseshoe prior (Carvalho et al., 2010). The horseshoe prior can be written hierarchically as

$$\beta_{ik} \mid \lambda_{ik} \sim N(0, \lambda_{ik}^2), \lambda_{ik} \sim C^+(0, 1), \quad (3.2)$$

where λ_{ik} is the prior standard deviation and follows the standard half-Cauchy distribution on the positive reals. Marginally over λ_{ik} , the prior for β_{ik} has a mass concentration near zero and heavy tails. The shape of the density shrinks null signals towards zero and avoids shrinking the true signals. This property facilitates separating signals from the noise.

Define the vector of slopes for subject i as $\boldsymbol{\beta}_i = (\beta_{i1}, \dots, \beta_{in_s})^T$. To incorporate spatial dependence into a multivariate horseshoe prior, we propose to set $\beta_{ik} = \lambda_{ik} \zeta_{ik}$, where λ_{ik} is a shrinkage parameter with half-Cauchy prior and ζ_{ik} is normal. Spatial shrinkage is induced by the spatial process model for $\boldsymbol{\lambda}_i = (\lambda_{i1}, \dots, \lambda_{in_s})^T$. We propose a Gaussian copula model (Nelsen, 2006) that preserves the marginal half-Cauchy distribution as in (3.2) and captures spatial dependence,

$$\lambda_{ik} = f(\delta_{ik}), \quad (3.3)$$

where δ_{ik} is the k^{th} variable of the latent process $\boldsymbol{\delta}_i = (\delta_{i1}, \dots, \delta_{in_s})^T$, $f(\cdot) = F_{C+}^{-1}[\Phi(\cdot)]$ is the half-Cauchy link function, $F_{C+}^{-1}(\cdot)$ is the inverse cumulative density function of the half-Cauchy distribution, and $\Phi(\cdot)$ is the standard normal cumulative distribution function.

The model can be expressed as

$$\mathbf{y}_{ij}^* = \boldsymbol{\alpha}_i + t_{ij} \cdot f(\boldsymbol{\delta}_i) \odot \boldsymbol{\zeta}_i + \boldsymbol{\varepsilon}_{ij}, \quad (3.4)$$

where $\mathbf{y}_{ij}^* = (y_{ij1}^*, \dots, y_{ijn_s}^*)^T$ is the vector of latent variables for subject i at the j^{th} visit, $\boldsymbol{\alpha}_i = (\alpha_{i1}, \dots, \alpha_{in_s})^T$ is the vector of baseline PD for subject i , the operator $\odot$ defines the pointwise vector product, $\boldsymbol{\delta}_i$ is the spatial latent vector, $\boldsymbol{\zeta}_i = (\zeta_{i1}, \dots, \zeta_{in_s})^T$ is the normal vector, and $\boldsymbol{\varepsilon}_{ij} = (\varepsilon_{ij1}, \dots, \varepsilon_{ijn_s})^T$ is the vector of random error.

3.3.2 Low-rank representation

We use a low-rank representation to capture the complex spatial dependence of PD data and to facilitate computing for the vectors $\boldsymbol{\alpha}_i$, $\boldsymbol{\delta}_i$ and $\boldsymbol{\zeta}_i$. Let $\mathbf{Q}$ be an $n_s \times L$ basis function matrix that determines the covariance of PD among sites, then we set $\boldsymbol{\alpha}_i = \mathbf{Q}\mathbf{a}_i$, $\boldsymbol{\delta}_i = \mathbf{Q}\mathbf{d}_i$ and $\boldsymbol{\zeta}_i = \mathbf{Q}\mathbf{z}_i$ so the model becomes

$$\mathbf{y}_{ij}^* = \mathbf{Q}\mathbf{a}_i + t_{ij} \cdot f(\mathbf{Q}\mathbf{d}_i) \odot \mathbf{Q}\mathbf{z}_i + \boldsymbol{\varepsilon}_{ij}, \quad (3.5)$$

where $\mathbf{a}_i = (a_{i1}, \dots, a_{iL})^T$ is the vector related to baseline PD for subject i , $\mathbf{d}_i = (d_{i1}, \dots, d_{iL})^T$ is the vector related to spatial latent vector, and $\mathbf{z}_i = (z_{i1}, \dots, z_{iL})^T$ is the vector related to slope.

We use principal component analysis (PCA; Pearson, 1901) to form the basis function matrix $\mathbf{Q}$. PCA does not merely increase computational efficiency, but provides interpretable decomposition of our dental data. Denote the total visit times as $N_v = \sum_{i=1}^{n_p} n_{vi}$ and $\mathbf{S}$ as the $n_s \times n_s$ sample covariance matrix of the N_v response vectors $\mathbf{y}_{ij} = (y_{ij1}, \dots, y_{ijn_s})^T$ for $i = 1, \dots, n_p$ and $j = 1, \dots, n_{vi}$. The eigen decomposition of $\mathbf{S}$ is $\mathbf{S} = \tilde{\mathbf{Q}}\tilde{\mathbf{D}}\tilde{\mathbf{Q}}^T$, where $\tilde{\mathbf{Q}}$ is the matrix of ordered eigenvectors $\mathbf{q}_{(i)}$ in the i^{th} column, $\tilde{\mathbf{D}}$ is the diagonal matrix with the ordered eigenvalues $d_1 \geq \dots \geq d_{n_s}$. We take the basis matrix $\mathbf{Q}$ to be the first L columns of $\tilde{\mathbf{Q}}$. The choice of L depends on the proportion of explained variation, $\sum_{l=1}^L d_l / \sum_{k=1}^{n_s} d_k$.

In this Chapter, we assume that the vectors $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ are both expanded using the

same basis function matrix $\mathbf{Q}$. However, it is possible to have a different basis for different model components. For example, one option is to perform PCA on the sample covariance of least squares estimates of $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$ and use them as the basis function for $\boldsymbol{\alpha}_i$ and $\boldsymbol{\beta}_i$.

3.3.3 Prior specification

We select normal priors for the vector related to baseline PD $\mathbf{a}_i$, the vector related to slope $\mathbf{z}_i$ and the latent vector $\mathbf{d}_i$, $\mathbf{a}_i \sim N(\boldsymbol{\mu}_a, \sigma_{ai}^2 \boldsymbol{\Sigma}_a)$, $\mathbf{d}_i \sim N(\mathbf{0}, \sigma_{di}^2 \boldsymbol{\Sigma}_d)$ and $\mathbf{z}_i \sim N(\mathbf{0}, \gamma_i^2 \boldsymbol{\Sigma}_z)$. The prior mean $\boldsymbol{\mu}_a$ follows an uninformative prior, $N(\mathbf{0}, 100^2 \mathbf{I}_L)$. The mean of $\mathbf{a}_i$ is non-zero to capture the overall mean spatial trend. The priors for the variance parameters σ_{ai}^2 , σ_{di}^2 and γ_i^2 are the uninformative inverse gamma distribution, $\text{IG}(0.1, 0.1)$. The random error ε_{ijk} follows an independent and identical normal prior with zero mean and variance σ_ε^2 , with an uninformative inverse gamma hyperprior $\text{IG}(0.1, 0.1)$ for σ_ε^2 .

We select inverse Wishart priors for the covariance matrices $\boldsymbol{\Sigma}_a$ and $\boldsymbol{\Sigma}_z$. The covariance of the intercepts and slopes across subjects may not be the same as the sample covariance, and our model allows for this due to the assignment of an inverse Wishart prior for $\boldsymbol{\Sigma}_a$ and $\boldsymbol{\Sigma}_z$. We hope that the basis matrix $\mathbf{Q}$ captures the main features and allow $\boldsymbol{\Sigma}_a$ to specify the best covariance in the span of $\mathbf{Q}$. The slopes $\boldsymbol{\beta}_i$ are the product of the two terms, $f(\boldsymbol{\delta}_i)$ and $\boldsymbol{\zeta}_i$, so it will be difficult to estimate the scale of both $\boldsymbol{\delta}_i$ and $\boldsymbol{\zeta}_i$. We therefore fix $\boldsymbol{\Sigma}_d$ at $\mathbf{D}$, the diagonal matrix with the first L eigenvalues $d_1, \dots, d_L$. Further, to preserve the half-Cauchy marginal distribution for $\boldsymbol{\lambda}_i$, we modify the link function to be $f(\delta_{ik}) = F_{C^+}^{-1}[\Phi(\delta_{ik}/\sqrt{w_k})]$ where w_k is the k^{th} diagonal element $\mathbf{QDQ}^T$. This gives an identifiable model that is still quite flexible, since the slope process $\boldsymbol{\beta}_i$ is nonstationary with $\text{Cov}(\beta_{ik}, \beta_{ik'}) = \lambda_{ik} \lambda_{ik'} \text{Cov}(\zeta_{ik}, \zeta_{ik'})$ depending on the two sites k and k' for $k, k' = 1, \dots, n_s$ and $k \neq k'$.

In summary, we formulate the priors

$$\begin{aligned} \mathbf{a}_i \mid \boldsymbol{\mu}_a, \boldsymbol{\Sigma}_a &\sim N(\boldsymbol{\mu}_a, \sigma_{ai}^2 \boldsymbol{\Sigma}_a), \mathbf{d}_i \mid \boldsymbol{\Sigma}_d \sim N(\mathbf{0}, \mathbf{D}), \\ \mathbf{z}_i \mid \boldsymbol{\Sigma}_z &\sim N(\mathbf{0}, \gamma_i^2 \boldsymbol{\Sigma}_z), \\ \boldsymbol{\mu}_a &\sim N(\mathbf{0}, 100^2 \mathbf{I}_L), \sigma_{ai}^2, \gamma_i^2, \sigma_\varepsilon^2 \sim \text{IG}(0.1, 0.1), \\ \boldsymbol{\Sigma}_a^{-1}, \boldsymbol{\Sigma}_z^{-1} &\sim \text{Wishart}(L, \mathbf{D}^{-1}), \end{aligned} \tag{3.6}$$

where L is the degrees of freedom.

3.4 Computing details

We perform Markov Chain Monte Carlo (MCMC) sampling using R. We implement blocked Metropolis-Hastings (MH) sampling for the vectors $\mathbf{d}_i$ and $\mathbf{z}_i$. The full conditional distributions for these parameters are

$$\begin{aligned}
P(\mathbf{d}_i \mid \cdot) &\propto \left[\prod_{j=1}^{n_{v_i}} P(\mathbf{y}_{ij}^* \mid \mathbf{d}_i) \right] \times P(\mathbf{d}_i) \\
&\propto \exp \left[-\frac{1}{2} \sum_{j=1}^{n_{v_i}} (\mathbf{y}_{ij}^* - \boldsymbol{\mu}_{ij})^T (\sigma_\varepsilon^2 \mathbf{I})^{-1} (\mathbf{y}_{ij}^* - \boldsymbol{\mu}_{ij}) - \frac{1}{2} \mathbf{d}_i^T \mathbf{D}^{-1} \mathbf{d}_i \right], \text{ and} \\
P(\mathbf{z}_i \mid \cdot) &\propto \left[\prod_{j=1}^{n_{v_i}} P(\mathbf{y}_{ij}^* \mid \mathbf{z}_i) \right] \times P(\mathbf{z}_i) \\
&\propto \exp \left[-\frac{1}{2} \sum_{j=1}^{n_{v_i}} (\mathbf{y}_{ij}^* - \boldsymbol{\mu}_{ij})^T (\sigma_\varepsilon^2 \mathbf{I})^{-1} (\mathbf{y}_{ij}^* - \boldsymbol{\mu}_{ij}) - \frac{1}{2\gamma_i^2} \mathbf{z}_i^T \boldsymbol{\Sigma}_z^{-1} \mathbf{z}_i \right],
\end{aligned} \tag{3.7}$$

where the latent mean vector $\boldsymbol{\mu}_{ij} = \mathbf{Q}\mathbf{a}_i + t_{ij} f(\mathbf{Q}\mathbf{d}_i) \odot \mathbf{Q}\mathbf{z}_i$. We use Gaussian candidate distributions $N(\mathbf{0}, \mathbf{D})$ and $N(\mathbf{0}, \gamma_i^2 \boldsymbol{\Sigma}_z)$. We tune the blocked MH algorithm of $\mathbf{d}_i$ and $\mathbf{z}_i$ via $\mathbf{D}$ and $\boldsymbol{\Sigma}_z$ to give acceptance probability near 40%. We monitor convergence using trace plots of several representative parameters.

Gibbs sampling is used for the remaining parameters: the vectors $\mathbf{y}_{ij}^*$, $\mathbf{a}_i$, $\boldsymbol{\mu}_a$, the parameters $\sigma_{ai}^2, \gamma_i^2, \sigma_\varepsilon^2$ and the inverse covariance matrices $\boldsymbol{\Sigma}_a^{-1}, \boldsymbol{\Sigma}_z^{-1}$. Given the priors in (3.6), the full conditional distributions used for the Gibbs updates are given below. Define the latent mean as $\mu_{ijk} = \alpha_{ik} + t_{ij} \beta_{ik}$. The latent PD for subject i at visit j and site k , y_{ijk}^* , follows a truncated normal distribution with mean μ_{ijk} , variance σ_ε^2 , lower bound $\min\{0, y_{ijk} - 0.5\}$ and upper bound $y_{ijk} + 0.5$, i.e.,

$$y_{ijk}^* \mid y_{ijk}, \mu_{ijk}, \sigma_\varepsilon^2 \sim \text{TN}[\mu_{ijk}, \sigma_\varepsilon^2, \min(0, y_{ijk} - 0.5), y_{ijk} + 0.5], \tag{3.8}$$

where $\text{TN}(\mu, \sigma^2, l, u)$ is the truncated normal density with the mean μ , the variance σ^2 , the lower bound l and the upper bound u . The low-rank vector of baseline PD $\mathbf{a}_i$ follows a multivariate normal posterior distribution.

$$\mathbf{a}_i \mid \cdot \sim N \left\{ \mathbf{W}_{ai} \mathbf{Q}^T \left[\frac{1}{\sigma_\varepsilon^2} \sum_{j=1}^{n_{v_i}} (\mathbf{y}_{ij}^* - t_{ij} \boldsymbol{\beta}_i) + \frac{1}{\sigma_{ai}^2} \boldsymbol{\Sigma}_a^{-1} \boldsymbol{\mu}_a \right], \mathbf{W}_{ai} \right\}, \tag{3.9}$$

where $\mathbf{W}_{ai} = (\frac{1}{\sigma_{ai}^2} \mathbf{\Sigma}_a^{-1} + \frac{n_{vi}}{\sigma_\varepsilon^2} \mathbf{Q}^T \mathbf{Q})^{-1}$. In the next layer, the posterior mean $\boldsymbol{\mu}_a$, the scale parameters $\sigma_{ai}^2, \gamma_i^2, \sigma_\varepsilon^2$, and the scale matrices $\mathbf{\Sigma}_a^{-1}, \mathbf{\Sigma}_z^{-1}$ all have a full conditional as below. Denote the total visit times as $N_v = \sum_{i=1}^{n_p} n_{vi}$.

$$\begin{aligned}
\boldsymbol{\mu}_a \mid \cdot &\sim N \left[\left(\sum_{i=1}^{n_p} \frac{1}{\sigma_{ai}^2} \mathbf{\Sigma}_a^{-1} + \frac{1}{100^2} \mathbf{I} \right)^{-1} \left(\sum_{i=1}^{n_p} \frac{1}{\sigma_{ai}^2} \mathbf{\Sigma}_a^{-1} \mathbf{a}_i \right), \left(\sum_{i=1}^{n_p} \frac{1}{\sigma_{ai}^2} \mathbf{\Sigma}_a^{-1} + \frac{1}{100^2} \mathbf{I} \right)^{-1} \right], \\
\sigma_{ai}^2 \mid \cdot &\sim \text{IG} \left[0.1 + \frac{1}{2}, 0.1 + \frac{1}{2} (\mathbf{a}_i - \boldsymbol{\mu}_a)^T \mathbf{\Sigma}_a^{-1} (\mathbf{a}_i - \boldsymbol{\mu}_a) \right], \\
\gamma_i^2 \mid \cdot &\sim \text{IG} \left(0.1 + \frac{1}{2}, 0.1 + \frac{1}{2} \mathbf{z}_i^T \mathbf{\Sigma}_z^{-1} \mathbf{z}_i \right), \\
\sigma_\varepsilon^2 \mid \cdot &\sim \text{IG} \left[0.1 + \frac{N_v n_s}{2}, 0.1 + \frac{1}{2} \sum_{i=1}^{n_p} \sum_{j=1}^{n_{vi}} \sum_{k=1}^{n_s} (y_{ijk}^* - \mu_{ijk})^2 \right], \\
\mathbf{\Sigma}_a^{-1} \mid \cdot &\sim \text{Wishart} \left\{ n_p + L, \left[\sum_{i=1}^{n_p} \frac{1}{\sigma_{ai}^2} (\mathbf{a}_i - \boldsymbol{\mu}_a)(\mathbf{a}_i - \boldsymbol{\mu}_a)^T + \mathbf{D}^{-1} \right] \right\}^{-1}, \\
\mathbf{\Sigma}_z^{-1} \mid \cdot &\sim \text{Wishart} \left[n_p + L, \left(\sum_{i=1}^{n_p} \frac{1}{\gamma_i^2} \mathbf{z}_i \mathbf{z}_i^T + \mathbf{D}^{-1} \right) \right]^{-1}.
\end{aligned} \tag{3.10}$$

We generate 10,000 samples and discard the first 2,000 as burn-in for data analysis in Section 3.5.

3.5 Data application

3.5.1 Model comparisons

In this section we apply the proposed model in Section 3.3 to the PD data described in Section 3.2. We fit the model to the visits during the first two years for all 7,279 subjects simultaneously and evaluate the prediction of PD at the next visit for each subject. We compare models with varying flexibility of shrinkage across space and different covariances. We consider two priors (Gaussian and spatial horseshoe) for the slopes $\boldsymbol{\beta}_i$ and two covariances (the sample covariance and conditionally autoregressive covariance) across space via the basis function matrix $\mathbf{Q}$. The Gaussian $\boldsymbol{\beta}_i$ has a constant shrinkage parameter across space, $\lambda_{ik} = 1$ for all subjects $i = 1, \dots, n_p$ and sites $k = 1, \dots, n_s$. In contrast, the spatial horseshoe (SHS) prior for the slopes allows spatially-varying shrinkage,

$\beta_i = f(\delta_i) \cdot \zeta_i$. Regarding the basis function matrix $\mathbf{Q}$, we consider low-rank representation of the sample covariance or conditionally autoregressive structure (CAR; Carlin and Banerjee, 2003) with the first-order neighbors. That is, a site neighbors the one or two sites on the same side of the jaw, the site directly across on the same jaw, and the site directly above/below on the opposite jaw. Therefore, the four most posterior sites in the buccal side have two neighbors, the other sites in the buccal side have three, and all others have four. The CAR covariance is $(\mathbf{M} - \rho\mathbf{A})^{-1}$, where $\mathbf{M}$ is the diagonal matrix with the elements $m_1, \dots, m_{n_s}$ indicating the number of neighbors for sites $1, \dots, n_s$, ρ is the spatial dependence parameter, and $\mathbf{A}$ is the adjacency matrix with $\mathbf{A}_{ij} = 1$ if sites i and j are neighbors and $\mathbf{A}_{ij} = 0$ otherwise. The spatial dependence parameter ρ is set at 0.99. The number of eigenvectors $L = 11, 53$ in the basis function matrix $\mathbf{Q}$ for the sample and CAR covariance are chosen for 70% and 90% explained variation in the sample covariance, respectively.

Table 3.1 presents the prediction results. The SHS model with the low-rank representation of the sample covariance outperforms the other three models. Predicted mean squared error (MSE) declines as flexibility of shrinkage grows. Regarding the basis function matrix $\mathbf{Q}$, MSE decreases by 11.1% to 11.5% using the low-rank representation of the sample covariance compared with using CAR under the 70% explained sample variation. The proportion reduces more, about 27.1% to 34.8%, under the 90% explained sample variation. Coverage is close to the nominal level 95% for all models. The computing times of the Gaussian models are approximately 17 and 34 minutes for $L = 11$ and 53, respectively, for the 100 MCMC iterations. For the SHS model, it takes 23 and 46 minutes for $L = 11$ and 53, respectively.

3.5.2 Estimated parameters

Figures 3.2 and 3.3 illustrate the first to fourth and the fifth to eighth eigenvectors of the sample covariance, respectively. We interpret the first eigenvector as the overall mean of PD; the second as a weighted average of PD with more emphasis on the posterior teeth; the third puts more weight on the teeth in the mandibular side (i.e. lower jaw); and the fourth puts more weight on the posterior teeth but more anterior part than the second eigenvector. The other four eigenvectors in Figure 3.3 exhibit several local features.

Similarly, Figures 3.4 and 3.5 show the first to fourth and the fifth to seventh and twelfth eigenvectors of the CAR covariance in a tooth map, respectively. The first eleven

Table 3.1: Comparison of prediction accuracy for pocket depth (PD) data using the Gaussian and spatial horseshoe (SHS) models using the low-rank representation of the sample covariance and conditional autoregressive (CAR) covariance with the number of basis functions $L = 11, 53$. Methods are compared using mean squared error (MSE), coverage and computing time for 100 Markov Chain Monte Carlo (MCMC) iterations in minutes.

Statistic	Model	Covariance	$L = 11$		$L = 53$	
			Estimate	SE	Estimate	SE
100×MSE	Gaussian	Sample	89.57	1.23	54.82	0.87
		CAR	101.31	1.37	84.08	0.91
	SHS	Sample	83.85	2.02	43.85	0.99
		CAR	94.36	1.44	60.11	1.02
Coverage (%)	Gaussian	Sample	93.72	0.12	94.79	0.13
		CAR	93.86	0.11	96.15	0.08
	SHS	Sample	93.64	0.12	94.31	0.15
		CAR	94.10	0.11	95.40	0.12
Computing time	Gaussian	Sample	16.89	—	34.72	—
		CAR	17.01	—	34.26	—
	SHS	Sample	22.63	—	45.57	—
		CAR	23.90	—	47.13	—

eigenvectors of the CAR covariance change horizontally, from posterior region to anterior, to posterior region in a mouth which are nearly identical for both jaws. Starting from the twelfth eigenvector, there are difference between the maxillary side (i.e. upper jaw) and the mandibular side, and the buccal side and the lingual side. The first eigenvector serves as the overall mean of PD. The second eigenvector puts more emphasis on the left-posterior region and decreases towards the right-posterior region. The third eigenvector is similar to the second in the sample covariance. The remaining eigenvectors in the CAR covariance depict varying characteristics in sub-regions of a mouth.

(a) The first eigenvector

(b) The second eigenvector

(c) The third eigenvector

(d) The fourth eigenvector

Figure 3.2: The first four eigenvectors of the sample covariance.

(a) The fifth eigenvector

(b) The sixth eigenvector

(c) The seventh eigenvector

(d) The eighth eigenvector

Figure 3.3: The fifth to eighth eigenvectors of the sample covariance.

(a) The first eigenvector

(b) The second eigenvector

(c) The third eigenvector

(d) The fourth eigenvector

Figure 3.4: The first four eigenvectors of the conditional autoregressive (CAR) covariance.

(a) The fifth eigenvector

(b) The sixth eigenvector

(c) The seventh eigenvector

(d) The twelfth eigenvector

Figure 3.5: The fifth to seventh and the twelfth eigenvectors of the conditional autoregressive (CAR) covariance.

3.5.3 Summary of fitted models

In this section we fit the Gaussian and SHS models with $L = 53$ eigenvectors to the PD data for the 7279 subjects in their dental visits within the first two years. To avoid excessive false positives, we implement the Bayesian spatial false discovery rate (BSFDR) procedure (Sun et al., 2015) with rate 0.01 to control for multiple testing. We consider the one-sided null and alternative hypotheses $H_0 : \beta_{ik} \leq 0$ and $H_1 : \beta_{ik} > 0$, for $i = 1, \dots, 7,279$ and $k = 1, \dots, 168$. We reject the null if the posterior probability of the alternative exceeds threshold T . The BSFDR procedure determines T so that the false discovery rate is approximately 0.01. The critical probabilities are $T = 94.95\%$ for the Gaussian models and $T = 96.97\%$ for the SHS models. The proportions of sites for which $H_{0_{ik}}$ is rejected across subjects are 4.81% and 7.29% for the Gaussian and SHS models, and so the SHS model appears to be more powerful.

In Figures 3.6 and 3.7, we exhibit the fitted results with $L = 53$ eigenvectors for the two subjects (hereforth labelled "Subject 1" and "Subject 2") with greatest difference in the posterior mean of β_{ik} between the Gaussian and SHS models with $L = 53$ eigenvectors. The posterior mean for Subject 1 in the SHS model is larger in teeth 5, 6, 14, 15 and 18 than the Gaussian model. The map of posterior probability, $P(\beta_{ik} > 0 \mid \mathbf{Y})$, indicates that the PD in the left side of mouth had increased significantly within the first two years. Compared with Gaussian model, SHS finds more significant deterioration of PD in the buccal side of the lower jaw and few in the middle of the right upper jaw (e.g. teeth 2, 5, 6). For Subject 2, the posterior means are larger in the left maxillary side for the SHS model compared with the Gaussian model. The map of posterior probability shows similar results. Comparing models, we find more significant sites and stronger spatial clustering of the signal in the SHS model compared to the Gaussian model.

To summarize the results for all subjects, Figure 3.8 illustrates that the density of the posterior means of the β_{ik} from the SHS model has higher concentration around zero and heavier tails than from the Gaussian model. Moreover, Figure 3.9 plots the average rejection rate among 7,279 subjects by teeth. Aside from teeth in the posterior region, SHS model also detects progression of teeth in the mandibular side and few in the right-maxillary side of mouth. In addition, Figure 3.10 shows the correlation among the basis functions for the covariances Σ_a and Σ_z using low-rank representation of the sample covariance and CAR covariance. It is not surprising that almost all basis functions are unrelated since we performed eigendecomposition for the two covariances. The only

exception is the first and second basis functions in Σ_a have a weak negative correlation using the sample covariance. Adding a prior in the covariance matrices Σ_a and Σ_z allows more flexibility in model fitting.

Fitting our model to the entire dataset of 7,279 patients is time-consuming. However, this is required only once offline to estimate population parameters. Fitting to the model to one subject as would be done in practice it fast. The computing times are 0.27 and 0.36 minutes with $L = 11$ using the low-rank representation of the sample covariance for the Gaussian and SHS models, respectively. When $L = 53$, it takes 0.63 and 1.17 minutes for the Gaussian and SHS models, respectively. Results (not shown) are very similar to the results from the data analysis of all patients.

(a) Posterior mean, Gaussian

(b) Posterior mean, SHS

(c) $P(\beta_{1k} > 0 \mid \mathbf{Y})$, Gaussian

(d) $P(\beta_{1k} > 0 \mid \mathbf{Y})$, SHS

Figure 3.6: Posterior mean of β_k and posterior probability $P(\beta_{1k} > 0 \mid \mathbf{Y})$, $k = 1, \dots, 168$ for Subject 1 in the Gaussian and SHS models with $L = 53$ eigenvectors.

(a) Posterior mean, Gaussian

(b) Posterior mean, SHS

(c) $P(\beta_{2k} > 0 | \mathbf{Y})$, Gaussian

(d) $P(\beta_{2k} > 0 | \mathbf{Y})$, SHS

Figure 3.7: Posterior mean of β_k and posterior probability $P(\beta_{2k} > 0 | \mathbf{Y})$, $k = 1, \dots, 168$ for Subject 2 in the Gaussian and SHS models with $L = 53$ eigenvectors.

Figure 3.8: Density plot (left) and quantile-quantile plot (right) of the posterior mean of site-specific linear-trend coefficient β_{ik} for subjects $i = 1, \dots, 7, 279$ and $k = 1, \dots, 168$ from the Gaussian and SHS models with the low-rank representation of the sample covariance under 90% explained variation.

(a) Gaussian model

(b) SHS model

Figure 3.9: The average rejection rate across subjects by teeth from the Gaussian and SHS models with $L = 53$ eigenvectors.

(a) Σ_a , sample covariance

(b) Σ_a , CAR covariance

(c) Σ_z , sample covariance

(d) Σ_z , CAR covariance

Figure 3.10: Posterior mean of the correlation matrix corresponding to the covariances Σ_a and Σ_z from the SHS model using the sample covariance (left column) and conditional autoregressive (CAR) covariance (right column) in the basis function matrix $\mathbf{Q}$.

3.6 Discussion

In this paper, we propose a spatiotemporal model for detecting local changes in periodontal disease. We implement the SHS prior on the linear time trend by sites for each subject. We introduce low-rank representation to reduce computational load, and obtain a nonstationary spatial covariance which suits the PD data and provides more flexibility. The empirical results show improved prediction when we consider a shrinkage prior and low-rank representation using the sample covariance.

A potential limitation of our model is that the assumption of linear change of the PD in time. This is probably sufficient for the PD data because we only use the data within the subjects' dental visits in two years. The linear trend in time can be modified to meet needs in other situations such as a longer study period (Corberán-Vallet and Lawson, 2016). A second restriction is the identical and independent error in the latent PD. We think this is sufficient for the PD data in that PDs were measured independently across subjects, visits and sites. However, this may not hold for different data.

The proposed model can be extended to Bayesian variable selection in spatiotemporal linear regression in the future. For data in a longer period, a random temporal effect may be included in the model (Corberán-Vallet and Lawson, 2016). We can also extend to multiple disease surveillance by including more data such as Banks et al. (2012) and Corberán-Vallet (2012).

Chapter 4

Misclassification Effect of Fine Particulate Matter due to Residential Mobility on Evaluating Risk of Congenital Heart Defects

4.1 Introduction

The Clean Air Act identifies six criteria pollutants considered harmful to public health and the environment. These pollutants include particulate matter within 2.5 micrometers in aerodynamic diameter ($PM_{2.5}$), particulate matter within 10 micrometers in aerodynamic diameter (PM_{10}), ground-level ozone, lead, carbon monoxide, nitrogen oxides and sulfur oxides. These criteria pollutants have been related to adverse health effects in both animals and humans including cardiovascular and respiratory diseases (Dominici et al., 2006; Franklin et al., 2007; Shah et al., 2013; Costa et al., 2014; Guarnieri and Balmes, 2014; Lee et al., 2014; Ierodiakonou et al., 2016; Khreis et al., 2017). In particular, $PM_{2.5}$ can travel into humans' airways and lungs through inhalation and may even get into bloodstreams through circulation, which lead to health problems (Cohen et al., 2005; Janssen et al., 2013; Xing et al., 2016). Recent studies have linked maternal exposure to ambient air pollution during pregnancy with adverse birth outcomes including preterm birth, low birth weight, intrauterine growth restriction, birth defects and perinatal mortality (Marshall et al., 2010; Warren et al., 2012; Stingone et al., 2014; Lelieveld et al.,

2015; Girguis et al., 2016; Yorifuji et al., 2016; Malley et al., 2017).

Many studies of environmental exposures and birth defects use the ambient concentration at the mothers' addresses at delivery as a proxy for the exposure, which does not account for residential mobility during pregnancy. The validity of these studies is questionable because birth defects generally occur within the first trimester of pregnancy and the mother's address at delivery may differ from her address early in pregnancy (Canfield et al., 2006; Chen et al., 2010). Bell and Belanger (2012) reviewed the fourteen studies across the world on residential mobility during pregnancy for evaluating prenatal environmental exposures. The study regions include several states in the United States, the United Kingdom, Norway, Netherlands and Australia. Overall mobility rates range from 9% to 32% and is highest in the second trimester. These studies do not have consistent results of the association of mobility and confounders such as maternal race, age, smoking, alcohol use, parity and socioeconomic status. In the studies reporting distances moved, most distances are short with a median distance less than 10 kilometers. There are similar findings in the recent studies investigating the association between ambient air pollution and low birth weight (Warren et al., 2017), PM_{10} and fetal growth (Pereira et al., 2016) using the data from the cohort studies in Connecticut. The negligible benefit from accounting for residential mobility may stem from a low proportion of mothers that move (11%) and short moving distance (median 5 kilometers). In addition, Pereira et al. (2016) showed that the influence of exposure metrics via different calculations is larger than residential mobility.

The objective of our study is to investigate the effect on modeled risk for congenital heart defects (CHDs) of misclassification of $PM_{2.5}$ exposure due to ignoring residential mobility during pregnancy. CHDs are the most common and serious birth defects that have significant impact on morbidity, mortality and health care costs in children and adults. Each year the prevalence is nearly 1% among total births, i.e., approximately 40,000 births in the United States (Hoffman and Kaplan, 2002; Reller et al., 2008). There are remaining uncertainties and challenges in understanding the effects of $PM_{2.5}$ exposure on neonatal health including uncertainty of exposure measurement errors using environmental monitoring data and the variation in $PM_{2.5}$ sources. The availability of geocoded latitude and longitude for the infants in the National Birth Defects Prevention Study (NBDPS), whose mothers provided detailed information about behaviors during pregnancy, gives us an opportunity to examine potential effects of air pollution on CHDs.

Chapter 4 is organized as follows. We describe the data and the model in Section 4.2 and interpret the analysis results in Section 4.3. In Section 4.4, we conduct a simulation study on residential mobility to examine bias from misclassifying exposures using addresses at delivery. We conclude with remarks and comments in Section 4.5.

4.2 Methods

4.2.1 Study population

The NBDPS is a population-based case-control study. Each center contributes approximately 300 cases and 100 controls each year. The data are collected through computer-assisted telephone interviews. Cogswell et al. (2009) concluded that control participants in the NBDPS are generally representative of the base populations. More details about the NBDPS can be found in Yoon et al. (2001) and Reefhuis et al. (2015).

We use birth defects data collected in Health Region 11 (Figure 4.1) of Texas between 2004 and 2011 from the NBDPS. We consider singleton deliveries with estimated date of delivery during the study years. Case infants are ascertained and collected by the Texas Birth Defects Registry. Control infants are randomly sampled from live-born and defect-free infants from birth hospitals in Health Region 11 of Texas. The outcomes of interest are left-ventricular outflow tract obstruction (LVOTO), right-ventricular outflow tract obstruction (RVOTO), conotruncal defects and septal defects. Other pregnancy information includes maternal age, race, smoking and alcohol use during pregnancy, high blood pressure, parity, gender of infant and season of birth. The classification system of birth defects was introduced by Botto et al. (2007). In practice, there is a team of clinicians with expertise in pediatric cardiology assigning one detailed category to each case according to medical review. We select simple and isolated cases to maintain etiological homogeneity and avoid possible confounding. The cases in our study population each have only a single CHD without extracardiac anomalies.

4.2.2 Exposure assignment

We use daily $\text{PM}_{2.5}$ estimates provided by Environmental Protection Agency (<https://www.epa.gov/hesc/rsig-related-downloadable-data-files>) since there are only three monitoring stations for air pollution in the study region. The estimates are derived

Figure 4.1: Health Regions in Texas.

using a spatial downscaler model that combines numerical output from the Models-3/Community Multi-scale Air Quality (CMAQ) model and air quality monitoring data from the National Air Monitoring Stations/State and Local Air Monitoring Stations (NAMS/SLAMS). This model provides calibrated exposures in a broader spatial coverage (Berrocal et al., 2010b; Berrocal et al., 2010a; Berrocal et al., 2012).

The geocoded data contains self-reported residential location and period. Most of residential periods begin on the 1st or 16th day of a month and end on the 15th or last day of a month. There are four levels of matching precision. We only include the highest level that self-reported residence can be matched by United States Postal Service address. "Stationary residence" assumes that mothers had lived at their address at delivery during the entire pregnancy; "mobile residence" uses their real addresses each day based on the geocoded data.

The daily $PM_{2.5}$ at maternal residence during gestational weeks 2-8 is matched to fused $PM_{2.5}$ at the nearest census tract centroid. We consider the susceptible window as gestational weeks 2-8, rather than the standard window of cardiac development (weeks 3-8; Moore et al., 2015), due to the potential lag effect of air pollution (Van den Hooven et al., 2012). The three women lived in the west coast are excluded due to lack of air pollution data.

4.2.3 Model description

We compare two Bayesian multinomial logistic regression models. The first assumes the effect of $\text{PM}_{2.5}$ is constant for weeks 2-8,

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \mathbf{x}_i^T \boldsymbol{\beta}_j + \bar{z}_i \gamma_j, \quad (4.1)$$

where $\pi_{ij} = P(Y_i = j)$ is the probability that infant i has CHD j for $j = 1, 2, 3, 4$, $J = 5$ is control, $\mathbf{x}_i$ is a vector of individual-level confounders, $\boldsymbol{\beta}_j$ is the vector of regression coefficients to CHD j including intercept, $\bar{z}_i = \sum_{d=8}^{56} z(\mathbf{s}_i, d)/49$ is the overall average of daily $\text{PM}_{2.5}$ at location $\mathbf{s}_i$ during gestational weeks 2-8, and $\gamma_{j,w}$ is the pollutant parameter for CHD j during the gestational week w . We use $N(\mathbf{0}, 100\mathbf{I})$ priors for $\boldsymbol{\beta}_j$.

In the second model we allow the $\text{PM}_{2.5}$ effect to vary by week.

$$\log\left(\frac{\pi_{ij}}{\pi_{iJ}}\right) = \mathbf{x}_i^T \boldsymbol{\beta}_j + \sum_{w=2}^8 \bar{z}_{i,w} \gamma_{j,w}, \quad (4.2)$$

where $\bar{z}_{i,w} = \sum_{d \in w} z(\mathbf{s}_i, d)/7$ is the average of daily $\text{PM}_{2.5}$ exposure at residence $\mathbf{s}_i$ during gestational week w for subject i . Let $\boldsymbol{\gamma} = (\gamma_{1,2}, \dots, \gamma_{4,8})^T$, the prior of $\boldsymbol{\gamma}$ is $N(\mathbf{0}, \boldsymbol{\Sigma}_w \otimes \boldsymbol{\Sigma}_{\text{CHD}})$ where $\boldsymbol{\Sigma}_w$ is the autoregressive (AR) covariance matrix among weeks with order 1, $\boldsymbol{\Sigma}_{\text{CHD}}$ is the covariance matrix among CHDs with the diagonal element 1 and off-diagonal element 0.3, and $\otimes$ denotes operation of Kronecker product. The dependent prior covariance of $\boldsymbol{\gamma}$ allows more flexibility compared to the independent covariance. We perform Markov Chain Monte Carlo (MCMC) sampling using R and generate 15,000 samples with the first 5,000 as burn-in for data analysis in Section 4.3.

4.3 Results

Table 4.1 shows demographic characteristics for controls and the four CHDs. There are 995 subjects in total with 642 controls, 48, 25, 71 and 209 cases with LVOTO, RVOTO, conotruncal and septal defects, respectively. In general, the average maternal age ranges from 24 to 29. At least 80% of the mothers are Hispanic. Compared with the control group, the proportions of female infant in the LVOTO and conotruncal groups are lower. The proportions of smoking and drinking during pregnancy are higher among cases with RVOTO. In addition, the proportions of maternal HBP are larger in the RVOTO and sep-

tal cases while the LVOTO and conotruncal cases share similar proportions of maternal HBP with controls.

Weekly $PM_{2.5}$ does not change much between stationary and mobile addresses. 87.12% of exposures at mobile addresses remain the same. The correlation between weekly $PM_{2.5}$ at the two types of addresses is 0.95 (Figure 4.2). This echoes the phenomenon that women tend to stay in the same place or move to a nearby location during pregnancy. Even for mothers that move a short distance, values of exposure may not change if two residences are nearby.

We fit six models. We consider overall average of $PM_{2.5}$ (4.1), weekly average of $PM_{2.5}$ (4.2) with prior covariance following an AR(1) structure across weeks, and full (seven) principal components (PCs) of the weekly exposures with an independent prior covariance. For each of these three models, we compare stationary and mobile exposure estimates. We use deviance information (DIC; Spiegelhalter et al., 2002) and Watanabe-Akaike information criterion (WAIC; Watanabe, 2010) for model evaluation. DIC (4.3) is a Bayesian measure of predictive accuracy which is sum of deviance and effective of parameters. WAIC (4.4) is a Bayesian approach to estimate the out-of-sample expectation of log predictive density and can be viewed as approximation of cross validation. Table 4.2 shows that DIC or WAIC of models using mobile addresses is similar to those using stationary addresses. DIC or WAIC of models with overall $PM_{2.5}$ is less than that with weekly exposures due to penalty of higher complexity of models. The model with $PM_{2.5}$ at mobile addresses detects signal of exposures and its DIC and WAIC is slightly less than the model in the same condition but at stationary residence. In addition, this model is easier for interpretation than using PCs of weekly exposures as predictors.

$$DIC = 2 \left[\log P(y|\hat{\alpha}, \hat{\beta}_w) - \frac{2}{M} \sum_{m=1}^M \log P(y|\alpha^{(m)}, \beta_w^{(m)}) \right] \quad (4.3)$$

$$WAIC = 2 \left\{ \sum_{i=1}^n V[\log P(y_i|\alpha^{(m)}, \beta_w^{(m)})] - \sum_{i=1}^n \log \left[\frac{1}{M} \sum_{m=1}^M P(y_i|\alpha^{(m)}, \beta_w^{(m)}) \right] \right\} \quad (4.4)$$

We choose the model using the weekly exposures with an AR(1) prior covariance as the final model. The results show that $PM_{2.5}$ at gestational week 3 has a significant positive association with having RVOTO. The odds is 1.17 times when weekly average

of $\text{PM}_{2.5}$ increases by $1 \mu\text{g}/\text{m}^3$. The 95% credible interval (CI) of odds ratio ranges from 1.01 to 1.38. Although the model using address at delivery has a similar estimate, it does not indicate a statistical significance of $\text{PM}_{2.5}$ at week 3. Figure 4.3 displays 95% CI of estimated coefficients for weekly exposure at stationary and mobile addresses. Similarly, Figures 4.4 and 4.5 illustrate modeling results using PCs of weekly exposure and overall average of $\text{PM}_{2.5}$.

Regarding the confounders, we find that smoking is significantly and positively associated with RVOTO, maternal HBP and septal defects have positive correlation. Delivering a female infant decreases odds by 0.50 and 0.27 for RVOTO and conotruncal defects, respectively. Odds ratio for septal defects decreases by 0.51 for Hispanic mothers. Having previous babies increases odds for LVOTO. Odds for conotruncal defects decreases when mothers deliver in spring or summer. The numerical results are shown in Tables 4.3-4.6.

Table 4.1: Maternal information for congenital heart defect cases and controls. We consider maternal information. Variables are summarized using counts (percentage) except for age which is summarized using mean (standard deviation).

	Controls <i>n</i> = 642		LVOTO <i>n</i> = 48		RVOTO <i>n</i> = 25		Conotruncal <i>n</i> = 71		Septal <i>n</i> = 209	
Female infant	304	(47.35)	12	(25.00)	9	(36.00)	21	(29.58)	99	(47.37)
Smoker	73	(11.37)	3	(6.25)	8	(32.00)	9	(12.68)	33	(15.79)
Alcohol use	152	(23.68)	10	(20.83)	10	(40.00)	21	(29.58)	58	(27.75)
HBP	80	(12.46)	6	(12.50)	6	(24.00)	10	(14.08)	47	(22.49)
Age	25.70	(6.28)	28.69	(4.93)	27.00	(5.93)	24.90	(6.23)	26.52	(6.50)
Race										
Hispanic	581	(90.50)	41	(85.42)	23	(92.00)	62	(87.32)	169	(80.86)
Others	61	(9.50)	7	(14.58)	2	(8.00)	9	(12.68)	40	(19.14)
Parity										
0	198	(30.84)	4	(8.33)	6	(24.00)	26	(36.62)	59	(28.23)
1	138	(21.50)	13	(27.08)	5	(20.00)	20	(28.17)	57	(27.27)
≥ 2	306	(47.66)	31	(64.58)	14	(56.00)	25	(35.21)	92	(44.02)
Season of birth										
Winter	149	(23.21)	7	(14.58)	6	(24.00)	26	(36.62)	59	(28.23)
Spring	152	(23.68)	16	(33.33)	4	(16.00)	12	(16.90)	41	(10.05)
Summer	181	(28.19)	9	(18.75)	11	(44.00)	16	(22.54)	65	(31.10)
Fall	160	(24.92)	16	(33.33)	4	(16.00)	17	(23.94)	55	(26.32)

Figure 4.2: Plot of weekly PM_{2.5} at stationary/mobile addresses.

Table 4.2: Model evaluation

Exposure	DIC		WAIC	
	Stationary	Mobile	Stationary	Mobile
Overall	1014.16	1014.68	973.65	974.17
Weekly AR(1)	1046.67	1046.05	991.25	990.86
Weekly PCA	1045.60	1046.05	989.30	990.15

Figure 4.3: 95% credible intervals of estimated coefficients for weekly $\text{PM}_{2.5}$ at mobile addresses (solid line) and stationary addresses (dashed line) with AR(1) prior.

Figure 4.4: 95% credible intervals of estimated coefficients for PCs of weekly $\text{PM}_{2.5}$ at mobile addresses (solid line) and stationary addresses (dashed line).

Figure 4.5: 95% credible intervals of estimated coefficients for overall average of $PM_{2.5}$ during weeks 2-8 at mobile addresses (solid line) and stationary addresses (dashed line).

Table 4.3: Modeling results of the multinomial logit model with weekly exposures using autoregressive (AR) prior for left-ventricular outflow tract obstruction (LVOTO). The asterisk * indicates significance at the level 0.05.

	Estimate	SD	95% Credible Interval	
			Lower	Upper
Intercept*	-4.13	1.265	-6.74	-1.82
Female infant*	-1.31	0.370	-2.05	-0.61
Smoker	-0.80	0.686	-2.10	0.58
Alcohol use	-0.22	0.409	-1.02	0.58
HBP	0.06	0.452	-0.83	0.93
Age	0.05	0.026	-0.01	0.10
Hispanic	-0.66	0.461	-1.54	0.26
Parity				
1*	1.21	0.577	0.11	2.37
$\geq 2^*$	1.25	0.556	0.20	2.35
Season of birth				
Spring	0.76	0.474	-0.15	1.71
Summer	-0.24	0.531	-1.30	0.79
Fall	0.74	0.474	-0.19	1.65
PM _{2.5}				
Week 2	-0.14	0.089	-0.31	0.03
Week 3	0.12	0.073	-0.03	0.26
Week 4	-0.11	0.082	-0.28	0.04
Week 5	0.12	0.081	-0.04	0.27
Week 6	-0.05	0.081	-0.21	0.10
Week 7	0.03	0.082	-0.13	0.19
Week 8	0.01	0.083	-0.15	0.16

Table 4.4: Modeling results of the multinomial logit model with weekly exposures using autoregressive (AR) prior for right-ventricular outflow tract obstruction (RVOTO). The asterisk * indicates significance at the level 0.05.

	Estimate	SD	95% Credible Interval	
			Lower	Upper
Intercept*	-4.48	1.432	-7.21	-1.57
Female infant	-0.17	0.394	-0.94	0.60
Smoker*	1.29	0.499	0.29	2.25
Alcohol use	0.01	0.470	-0.93	0.91
HBP	0.56	0.482	-0.40	1.47
Age	0.02	0.033	-0.04	0.09
Hispanic	0.37	0.630	-0.88	1.59
Parity				
1	0.61	0.575	-0.54	1.72
≥ 2	0.35	0.541	-0.72	1.41
Season of birth				
Spring	-0.30	0.562	-1.43	0.79
Summer	0.20	0.491	-0.76	1.17
Fall	-0.69	0.633	-1.92	0.55
PM _{2.5}				
Week 2	-0.06	0.093	-0.24	0.12
Week 3*	0.16	0.084	0.01	0.32
Week 4	-0.10	0.107	-0.31	0.11
Week 5	0.01	0.102	-0.19	0.21
Week 6	-0.06	0.103	-0.27	0.14
Week 7	-0.02	0.108	-0.24	0.18
Week 8	0.04	0.092	-0.14	0.22

Table 4.5: Modeling results of the multinomial logit model with weekly exposures using autoregressive (AR) prior for conotruncal defects. The asterisk * indicates significance at the level 0.05.

	Estimate	SD	95% Credible Interval	
			Lower	Upper
Intercept	-0.75	1.038	-2.73	1.29
Female infant*	-0.70	0.276	-1.25	-0.17
Smoker	-0.28	0.435	-1.12	0.59
Alcohol use	0.24	0.313	-0.39	0.84
HBP	0.12	0.380	-0.60	0.89
Age	-0.03	0.024	-0.07	0.02
Hispanic	-0.53	0.397	-1.28	0.27
Parity				
1	0.07	0.340	-0.60	0.72
≥ 2	-0.19	0.336	-0.85	0.48
Season of birth				
Spring*	-0.92	0.399	-1.69	-0.15
Summer*	-0.66	0.348	-1.37	-0.01
Fall	-0.31	0.339	-0.96	0.38
PM _{2.5}				
Week 2	0.10	0.057	-0.01	0.21
Week 3*	-0.16	0.074	-0.30	-0.01
Week 4	-0.06	0.075	-0.20	0.08
Week 5	0.02	0.075	-0.14	0.16
Week 6	0.11	0.063	-0.01	0.23
Week 7	0.02	0.073	-0.12	0.16
Week 8	0.01	0.070	-0.14	0.13

Table 4.6: Modeling results of the multinomial logit model with weekly exposures using autoregressive (AR) prior for septal defects. The asterisk * indicates significance at the level 0.05.

	Estimate	SD	95% Credible Interval	
			Lower	Upper
Intercept	-0.61	0.689	-1.96	0.81
Female infant	0.05	0.163	-0.26	0.37
Smoker	0.26	0.258	-0.23	0.76
Alcohol use	-0.06	0.206	-0.47	0.34
HBP*	0.69	0.210	0.29	1.11
Age	0.02	0.014	-0.01	0.05
Hispanic*	-0.68	0.232	-1.14	-0.23
Parity				
1	0.28	0.226	-0.17	0.71
≥ 2	-0.12	0.217	-0.55	0.30
Season of birth				
Spring	-0.17	0.250	-0.64	0.32
Summer	0.13	0.227	-0.33	0.56
Fall	0.18	0.233	-0.27	0.64
PM _{2.5}				
Week 2	0.01	0.038	-0.07	0.08
Week 3	-0.01	0.043	-0.08	0.09
Week 4	0.01	0.045	-0.09	0.09
Week 5	0.04	0.046	-0.05	0.13
Week 6	-0.08	0.047	-0.17	0.01
Week 7	-0.03	0.049	-0.13	0.06
Week 8	-0.02	0.043	-0.10	0.06

4.4 Simulation Study

4.4.1 Data generation

The objective of the simulation study is to characterize the $\text{PM}_{2.5}$ measurement error on adverse birth outcomes effect estimation. We use the address at delivery of the 995 mothers in real data and simulate their residential history. We assume that pregnant women had moved at most once during gestational weeks 2-8, according to the finding in real data.

We consider logistic regression model with weekly or overall average of $\text{PM}_{2.5}$. For subject i ,

$$\begin{aligned} y_i \mid p_i &\sim \text{Bernoulli}(p_i) \\ \text{logit}(p_i) &= \alpha + \sum_{w=2}^8 \beta_w Z_{i,w} \quad (\text{Weekly average model}) \quad \text{or} \quad (4.5) \\ \text{logit}(p_i) &= \alpha + \beta \bar{Z}_i, \quad (\text{Overall average model}) \end{aligned}$$

where y_i is a binary response indicating whether subject i is a case with any of four types of CHDs or a control, p_i is the probability that subject i has any CHDs, α is the intercept, β_w is the coefficient in week w . For the overall average model, $Z_{i,w}$ is standardized average of $\text{PM}_{2.5}$ in week w , and $\bar{Z}_i$ is standardized overall average of $\text{PM}_{2.5}$ during gestational weeks 2-8 at mobile address for subject i and β is its effect.

We obtain true $\text{PM}_{2.5}$ from mobile addresses including maternal residences in week 2 and simulated addresses. The simulated residences depend on moving distance and direction. We generate a sample of new residences for women who moved. The possible residences are within 100 kilometers from each original residence. Let d be the moving distance in kilometers, S be the county that includes location $\mathbf{s}$, $D(S)$ be the population density of county S from 2010 census data, and t be the moving week from 3 to 8. The probability of moving to residence $\mathbf{s}$ is proportional to the product of density of moving distance $f(d)$ and population density of county S that includes $\mathbf{s}$. That is, $P(\mathbf{s}) \propto f(d) D(S)$, where $\mathbf{s}_0$ is the residence at gestational week 2 and d follows an Exponential distribution with scale parameter a .

There are three factors in the simulation study: stationary or mobile $\text{PM}_{2.5}$ as predictors, moving proportion q and average moving distance a . We consider $q = 0.1, 0.2, 0.3$ and

$a = 10, 30, 50$. The true coefficients are taken to be $\alpha = -2$, $(\beta_2, \beta_3, \beta_4, \beta_5, \beta_6, \beta_7, \beta_8) = (0.2, 0.3, 0.4, 0.1, 0, 0)$ for the weekly simulations and β is set at 0, 0.5, 1 for the overall simulations. Each coefficient in a logistic regression model is estimated with a normal prior $N(0, 10^2)$ and 10,000 MCMC samples are used to summarize the posterior. We generate 1,000 datasets for each combination of q and a .

4.4.2 Measures for evaluation

We compare models in terms of bias, mean squared error (MSE), Type I error, power, DIC and WAIC. For simulated data set k , let $\hat{\beta}_w^k$ be the posterior mean, $[l_w^k, u_w^k]$ be the 95% credible interval of β_w and $\beta_w^{(m)}$ be the m^{th} posterior sample of β_w . We list the formulae of the measures as follows:

$$\text{Bias} = \frac{1}{7N} \sum_{k=1}^N \sum_{w=2}^8 (\hat{\beta}_w^k - \beta_w) \quad (4.6)$$

$$\text{MSE} = \frac{1}{7N} \sum_{k=1}^N \sum_{w=2}^8 (\hat{\beta}_w^k - \beta_w)^2 \quad (4.7)$$

$$\text{Type I error} = \frac{1}{N} \sum_{k=1}^N \frac{\sum_{w=2}^8 I\{0 \notin [l_w^k, u_w^k]\} \cdot I(\beta_w = 0)}{\sum_{w=2}^8 I(\beta_w = 0)} \quad (4.8)$$

$$\text{Power} = \frac{1}{N} \sum_{k=1}^N \frac{\sum_{w=2}^8 I\{0 \notin [l_w^k, u_w^k]\} \cdot I(\beta_w \neq 0)}{\sum_{w=2}^8 I(\beta_w \neq 0)} \quad (4.9)$$

4.4.3 Results

In general, models with mobile PM_{2.5} outperform models with stationary PM_{2.5}. The estimator of β_w using addresses at delivery tends to underestimate and the bias becomes larger when moving proportion or moving distance increases. In addition, the generated probability p_i close to 0 or 1 may lead to bias even when we consider the right model. Next, MSE of models with mobile addresses is smaller, about 33.33% to 66.67% of that with stationary addresses. MSE with stationary PM_{2.5} grows a little bit as moving distance or proportion increases. Type I errors are at or near the nominal level for all models. Power of models with mobile addresses is larger than that with addresses at delivery, by 9.54% to 19.68%.

We use DIC and WAIC to select models between mobile and stationary PM_{2.5}. Table 4.8 shows that DIC and WAIC can choose the correct model with high accuracy. Among all scenarios, the correlation of weekly averages of PM_{2.5} between stationary and mobile addresses is high, from 0.83 to 0.92. It increases as moving distance or proportion decreases. The simulation scenario with $q = 10\%$ and $a = 10$ is the closest to the empirical data. The simulation results with overall PM_{2.5} (Tables 4.10-4.16) are similar to the results with weekly PM_{2.5}.

Table 4.7: Comparison of stationary and mobile model estimates of weekly effects β_w for data simulated with average moving distance $a = 10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30% .

	q	PM _{2.5}	10 km		30 km		50 km	
			Mean	SE	Mean	SE	Mean	SE
100×Bias	10%	Stationary	-1.66	0.070	-1.84	0.071	-2.08	0.072
		Mobile	0.57	0.071	0.58	0.072	0.58	0.072
	20%	Stationary	-1.80	0.076	-2.14	0.076	-2.65	0.077
		Mobile	0.56	0.076	0.58	0.076	0.48	0.073
	30%	Stationary	-1.89	0.072	-2.45	0.073	-3.19	0.072
		Mobile	0.58	0.073	0.53	0.074	0.50	0.071
100×MSE	10%	Stationary	0.08	0.003	0.08	0.003	0.10	0.004
		Mobile	0.05	0.003	0.05	0.003	0.05	0.003
	20%	Stationary	0.09	0.004	0.10	0.004	0.13	0.005
		Mobile	0.06	0.003	0.06	0.003	0.06	0.003
	30%	Stationary	0.09	0.003	0.11	0.004	0.15	0.005
		Mobile	0.06	0.003	0.06	0.003	0.05	0.003
Type I error (%)	10%	Stationary	6.45	0.553	6.60	0.572	6.65	0.573
		Mobile	6.85	0.580	6.60	0.572	6.70	0.583
	20%	Stationary	5.35	0.509	5.90	0.530	6.10	0.532
		Mobile	6.65	0.569	5.90	0.553	5.75	0.519
	30%	Stationary	6.10	0.537	6.25	0.542	6.25	0.551
		Mobile	6.15	0.543	6.75	0.581	6.05	0.631
Power (%)	10%	Stationary	54.52	0.526	54.50	0.523	53.44	0.543
		Mobile	59.72	0.494	60.64	0.480	60.26	0.503
	20%	Stationary	53.50	0.523	52.96	0.517	51.48	0.510
		Mobile	60.36	0.518	60.02	0.505	60.18	0.508
	30%	Stationary	53.70	0.501	52.36	0.516	50.92	0.509
		Mobile	60.16	0.490	61.18	0.493	60.94	0.476

Table 4.8: Probability (%) of selecting correct model (mobile PM) using DIC and WAIC under β_w with average moving distance $a=10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30% .

Standard	q	10 km		30 km		50 km	
		Mean	SE	Mean	SE	Mean	SE
DIC	10%	99.9	0.100	99.6	0.200	99.6	0.200
	20%	99.8	0.141	99.7	0.173	99.9	0.100
	30%	99.7	0.173	99.9	0.100	99.9	0.100
WAIC	10%	99.7	0.173	99.7	0.173	99.7	0.173
	20%	99.7	0.173	99.7	0.173	99.8	0.141
	30%	99.7	0.173	99.8	0.141	99.9	0.100

Table 4.9: $100 \times$ Correlation of weekly $PM_{2.5}$ between stationary and mobile addresses under all simulation scenarios.

q	10 km		30 km		50 km	
	Mean	SE	Mean	SE	Mean	SE
10%	91.62	0.013	90.58	0.019	88.77	0.029
20%	91.13	0.018	89.08	0.026	85.74	0.035
30%	90.63	0.020	87.61	0.030	83.02	0.041

Table 4.10: Comparison of stationary and mobile model estimates of overall effect $\beta = 0$ for data simulated with average moving distance $a = 10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30% .

	q	PM _{2.5}	10 km		30 km		50 km	
			Mean	SE	Mean	SE	Mean	SE
100×Bias	10%	Stationary	-0.28	0.316	-0.28	0.316	-0.28	0.316
		Mobile	-0.24	0.318	-0.27	0.316	-0.23	0.315
	20%	Stationary	0.32	0.312	0.32	0.312	0.32	0.312
		Mobile	-0.26	0.308	-0.15	0.309	-0.29	0.310
	30%	Stationary	-0.31	0.315	-0.31	0.315	-0.31	0.315
		Mobile	-0.21	0.305	-0.33	0.309	-0.39	0.310
100×MSE	10%	Stationary	1.00	0.048	1.00	0.048	1.00	0.048
		Mobile	1.02	0.048	1.00	0.047	1.00	0.047
	20%	Stationary	0.97	0.045	0.97	0.045	0.97	0.045
		Mobile	0.95	0.043	0.96	0.043	0.97	0.044
	30%	Stationary	0.99	0.043	0.99	0.043	0.99	0.043
		Mobile	0.99	0.046	0.97	0.043	0.96	0.043
Type I error (%)	10%	Stationary	5.10	0.696	5.10	0.696	5.10	0.696
		Mobile	5.90	0.745	5.50	0.721	5.80	0.740
	20%	Stationary	5.20	0.702	5.20	0.702	5.20	0.702
		Mobile	5.30	0.709	4.80	0.676	4.80	0.676
	30%	Stationary	5.00	0.690	5.00	0.690	5.00	0.690
		Mobile	4.70	0.670	4.40	0.649	4.50	0.656

Table 4.11: Probability (%) of selecting correct model (mobile PM) using DIC and WAIC under $\beta = 0$ with average moving distance $a=10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30% .

Standard	q	10 km		30 km		50 km	
		Mean	SE	Mean	SE	Mean	SE
DIC	10%	51.3	1.581	51.5	1.581	51.9	1.581
	20%	50.5	1.582	50.5	1.582	51.6	1.581
	30%	50.0	1.582	49.8	1.582	49.6	1.582
WAIC	10%	47.2	1.579	45.0	1.574	47.1	1.579
	20%	46.1	1.577	45.7	1.576	46.8	1.579
	30%	46.8	1.579	45.8	1.576	46.4	1.578

Table 4.12: 100×Correlation of overall PM_{2.5} between stationary and mobile addresses under all simulation scenarios.

q	10 km		30 km		50 km	
	Mean	SE	Mean	SE	Mean	SE
10%	86.13	0.021	84.80	0.025	82.64	0.047
20%	85.43	0.029	82.80	0.042	78.78	0.058
30%	84.72	0.033	80.78	0.038	75.23	0.069

Table 4.13: Comparison of stationary and mobile model estimates of overall effect $\beta = 0.5$ for data simulated with average moving distance $a = 10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30%.

	q	PM _{2.5}	10 km		30 km		50 km	
			Mean	SE	Mean	SE	Mean	SE
100×Bias	10%	Stationary	-7.16	0.291	-7.71	0.294	-8.90	0.299
		Mobile	0.63	0.283	0.72	0.279	1.09	0.278
	20%	Stationary	-7.09	0.292	-8.54	0.296	-10.79	0.296
		Mobile	0.90	0.285	0.83	0.285	0.66	0.283
	30%	Stationary	-7.70	0.286	-9.71	0.286	-13.00	0.293
		Mobile	0.40	0.274	0.40	0.277	0.29	0.271
100×MSE	10%	Stationary	1.36	0.059	1.46	0.063	1.68	0.070
		Mobile	0.80	0.038	0.78	0.037	0.78	0.036
	20%	Stationary	1.36	0.058	1.61	0.063	2.04	0.076
		Mobile	0.82	0.035	0.82	0.036	0.80	0.035
	30%	Stationary	1.41	0.059	1.76	0.064	2.54	0.083
		Mobile	0.75	0.033	0.77	0.033	0.74	0.034
Power (%)	10%	Stationary	99.60	0.200	99.40	0.244	99.30	0.264
		Mobile	100.00	0.000	100.00	0.000	100.00	0.000
	20%	Stationary	99.40	0.244	99.40	0.244	98.70	0.358
		Mobile	100.00	0.000	100.00	0.000	100.00	0.000
	30%	Stationary	99.70	0.173	99.50	0.223	98.90	0.330
		Mobile	100.00	0.000	100.00	0.000	100.00	0.000

Table 4.14: Probability (%) of selecting correct model (mobile PM) using DIC and WAIC under $\beta = 0.5$ with average moving distance $a=10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30% .

Standard	q	10 km		20 km		30 km	
		Mean	SE	Mean	SE	Mean	SE
DIC	10%	96.0	0.620	96.4	0.589	97.5	0.494
	20%	96.0	0.620	97.2	0.522	97.3	0.513
	30%	96.9	0.548	96.8	0.557	98.8	0.344
WAIC	10%	96.4	0.589	96.5	0.581	97.8	0.464
	20%	95.7	0.642	97.1	0.531	97.2	0.522
	30%	96.4	0.589	96.8	0.557	98.9	0.330

Table 4.15: Comparison of stationary and mobile model estimates of overall effect $\beta = 1$ for data simulated with average moving distance $a = 10, 30$ and 50 km and percent of mothers moving $q = 10\%, 20\%$ and 30% .

	q	PM _{2.5}	10 km		30 km		50 km	
			Mean	SE	Mean	SE	Mean	SE
100×Bias	10%	Stationary	-16.24	0.292	-17.95	0.294	-20.90	0.294
		Mobile	1.52	0.331	1.51	0.334	1.57	0.331
	20%	Stationary	-16.36	0.314	-19.75	0.308	-25.38	0.304
		Mobile	2.05	0.347	2.06	0.340	1.40	0.332
	30%	Stationary	-17.51	0.304	-22.43	0.307	-29.71	0.304
		Mobile	1.57	0.345	1.67	0.346	1.65	0.347
100×MSE	10%	Stationary	3.49	0.099	4.09	0.110	5.23	0.124
		Mobile	1.11	0.053	1.14	0.055	1.12	0.051
	20%	Stationary	3.66	0.108	4.85	0.125	7.36	0.153
		Mobile	1.24	0.063	1.20	0.060	1.12	0.052
	30%	Stationary	3.99	0.110	5.97	0.140	9.75	0.186
		Mobile	1.21	0.060	1.22	0.057	1.23	0.061
Power (%)	10%	Stationary	100.00	0.000	100.00	0.000	100.00	0.000
		Mobile	100.00	0.000	100.00	0.000	100.00	0.000
	20%	Stationary	100.00	0.000	100.00	0.000	100.00	0.000
		Mobile	100.00	0.000	100.00	0.000	100.00	0.000
	30%	Stationary	100.00	0.000	100.00	0.000	100.00	0.000
		Mobile	100.00	0.000	100.00	0.000	100.00	0.000

Table 4.16: Probability (%) of selecting correct model (mobile PM) using DIC and WAIC under $\beta = 1$ with average moving distance $a=10, 30$ and 50 km.

Standard	q	10 km		20 km		30 km	
		Mean	SE	Mean	SE	Mean	SE
DIC	10%	99.9	1.000	100.0	0.000	100.0	0.000
	20%	99.9	1.000	100.0	0.000	100.0	0.000
	30%	100.0	0.000	100.0	0.000	100.0	0.000
WAIC	10%	99.9	1.000	100.0	0.000	100.0	0.000
	20%	99.9	1.000	100.0	0.000	100.0	0.000
	30%	100.0	0.000	100.0	0.000	100.0	0.000

4.5 Discussion

In this paper, we investigate the misclassification effect of fine particulate matter due to ignoring residential mobility on evaluating risk of congenital heart defects through Bayesian logistic regression models using Texas birth defects data during 2004-2011. We find that the weekly exposures help detect signals compared to the overall exposure. The weekly $\text{PM}_{2.5}$ at gestational week 3 has a significant positive association with LVOTO while no significance is found in the model using overall PM exposure. Regarding other covariates, smoking and HBP are risk factors of RVOTO and septal defects, respectively. The strong similarity of exposures at mobile and stationary addresses leads to negligible difference in empirical results. In the simulation study, we obtain that the misclassification effect of exposures grows when the moving proportion and distance increase, presented in larger bias and MSE.

Our analysis results may not hold for other regions since the birth defects data was only sampled from Health Region 11 in Texas. However, Warren et al. (2017) also found a negligible difference in exposures at stationary and mobile addresses in Connecticut. In addition, self-reported information including residence, smoking and drinking may be a source of bias which frequently exists in observational studies.

In the future, we can aggregate data from all other NBDPS centers and perform a comprehensive analysis to investigate the misclassification effect of fine particulate matter exposures regarding residential mobility on risk of several types of birth defects. Another extension can be the investigation of the effect of misclassified speciated $\text{PM}_{2.5}$ on birth defects.

REFERENCES

- Abellan, J. J., Richardson, S., and Best, N. (2008). Use of spacetime models to investigate the stability of patterns of disease. *Environmental Health Perspectives*, 116(8):1111–1119.
- Agossa, K., Dendooven, A., Dubuquoy, L., GowerRousseau, C., DelcourtDebruyne, E., and Capron, M. (2017). Periodontal manifestations of inflammatory bowel disease: emerging epidemiologic and biologic evidence. *Journal of Periodontal Research*, 52(3):313–324.
- Aldstadt, J. (2007). An incremental knox test for the determination of the interval between successive cases of an infectious disease. *Stochastic Environmental Research and Risk Assessment*, 21:487–500.
- Alexeeff, S. E., Carroll, R. J., and Coull, B. (2016). Spatial measurement error and correction by spatial SIMEX in linear regression models when using predicted air pollution exposures. *Biostatistics*, 17(2):377–389.
- Armagan, A., Dunson, D. B., and Lee, J. (2013). Generalized double pareto shrinkage. *Statistica Sinica*, 23(1):119–143.
- Avanasi, R., Shin, H.-M., Vieira, V. M., Savitz, D. A., and Bartell, S. M. (2016). Impact of exposure uncertainty on the association between perfluorooctanoate and preeclampsia in the C8 health project population. *Environmental Health Perspectives*, 124:126–132.
- Banks, D., Datta, G., Karr, A., Lynch, J., Niemi, J., and Vera, F. (2012). Bayesian CAR models for syndromic surveillance on multiple data streams: Theory and practice. *Information Fusion*, 13(2):105 – 116. Special Issue on Information Fusion Applications to Human Health and Safety.
- Banos, A. and Lacasa, J. (2007). Spatio-temporal exploration of SARS epidemic. <http://journals.openedition.org/cybergeogeo/12803>.
- Bell, M. L. and Belanger, K. (2012). Review of research on residential mobility during pregnancy: consequences for assessment of prenatal environmental exposures. *Journal of Exposure Science and Environmental Epidemiology*, 22:429–438.
- Beltrán-Aguilar, E. D., Malvitz, D. M., Lockwood, S. A., Rozier, R. G., and Tomar, S. L. (2003). Oral health surveillance: Past, present, and future challenges. *Journal of Public Health Dentistry*, 63(3):141–149.

- Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of Royal Statistical Society: Series B (Statistical Methodology)*, 57(1):289–300.
- Berlin-Broner, Y., Febbraio, M., and Levin, L. (2016). Association between apical periodontitis and cardiovascular diseases: a systematic review of the literature. *International Endodontic Journal*, 50(9):847–859.
- Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2010a). A bivariate spacetime downscaler under space and time misalignment. *The Annals of Applied Statistics*, 4(4):1942–1975.
- Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2010b). A spatiotemporal downscaler for output from numerical models. *Journal of Agricultural, Biological, and Environmental Statistics*, 15:176–197.
- Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2012). Space-time fusion under error in computer model output: an application to modeling air quality. *Biometrics*, 68:837–848.
- Bhadra, A., Datta, J., Li, Y., Polson, N. G., and Willard, B. (2016a). Prediction risk for the horseshoe regression. *ArXiv e-prints*.
- Bhadra, A., Datta, J., Polson, N. G., and Willard, B. (2016b). Default bayesian analysis with global-local shrinkage priors. *Biometrika*, 103(4):955–969.
- Bhattacharya, A., Pati, D., Pillai, N. S., and Dunson, D. B. (2014). Dirichlet-laplace priors for optimal shrinkage. *ArXiv e-prints*.
- Boehm Vock, L. F., Reich, B. J., Fuentes, M., and Dominici, F. (2015). Spatial variable selection methods for investigating acute health effects of fine particulate matter components. *Biometrics*, 71(1):167–177.
- Botto, L. D., Lin, A. E., Riehle-Colarusso, T., Malik, S., and Correa, A. (2007). Seeking causes: Classifying and evaluating congenital heart defects in etiologic studies. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 79(10):714–727.
- Cai, B. and Bandyopadhyay, D. (2017). Bayesian semiparametric variable selection with applications to periodontal data. *Statistics in Medicine*, 36(14):2251–2264.
- Canfield, M. A., Ramadhani, T. A., Langlois, P. H., and Waller, D. K. (2006). Residential mobility patterns and exposure misclassification in epidemiologic studies of birth defects. *Journal of Exposure Science and Environmental Epidemiology*, 16:538–543.
- Carlin, P. B. and Banerjee, S. (2003). Hierarchical multivariate car models for spatio-temporally correlated survival data. *Bayesian Statistics*, 7:45–63.

- Carvalho, C. M., Polson, N. G., and Scott, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.
- Chang, S.-H., Cosman, P. C., and Milstein, L. B. (2011). Chernoff-type bounds for the gaussian error function. *IEEE Transactions on Communications*, 59(11):2939–2944.
- Chen, L., Bell, E. M., Caton, A. R., Druschel, C. M., and Lin, S. (2010). Residential mobility during pregnancy and the potential for ambient air pollution exposure misclassification. *Environmental Research*, 110(2):162–168.
- Cheng, Y. L., Jordan, L., Chen, H., Kang, D., Oxford, L., Plemons, J., Parks, H., and Rees, T. (2017). Chronic periodontitis can affect the levels of potential oral cancer salivary mRNA biomarkers. *Journal of Periodontal Research*, 52(3):428–437.
- Cogswell, M. E., Bitsko, R. H., Anderka, M., Caton, A. R., Feldkamp, M. L., Sherlock, S. M. H., Meyer, R. E., Ramadhani, T., Robbins, J. M., Shaw, G. M., Mathews, T. J., Royle, M., Reefhuis, J., and the National Birth Defects Prevention Study (2009). Control selection and participation in an ongoing, population-based, case-control study of birth defects: The national birth defects prevention study. *American Journal of Epidemiology*, 170(8):975–985.
- Cohen, A. J., Anderson, H. R., Ostra, B., Pandey, K. D., Krzyzanowski, M., Künzli, N., Gutschmidt, K., Pope, A., Romieu, I., Samet, J. M., and Smith, K. (2005). The global burden of disease due to outdoor air pollution. *Journal of Toxicology and Environmental Health*, A(68):1–7.
- Corberán-Vallet, A. (2012). Prospective surveillance of multivariate spatial disease data. *Statistical methods in medical research*, 21(5):457–477.
- Corberán-Vallet, A. and Lawson, A. B. (2011). Conditional predictive inference for online surveillance of spatial disease incidence. *Statistics in Medicine*, 30(26):3095–3116.
- Corberán-Vallet, A. and Lawson, A. B. (2016). Spatial health surveillance. In Lawson, A. B., Banerjee, S., Haining, R. P., and Ugarte, M. D., editors, *Handbook of Spatial Epidemiology*, chapter 27, pages 501–519. Chapman and Hall/CRC, Boca Raton, FL.
- Costa, M. A. and Kulldorff, M. (2009). Applications of spatial scan statistics: a review. In Glaz, J., P. V. W. S., editor, *Scan Statistics: Methods and Applications*, pages 129–152. Birkhäuser Boston, Boston, MA.
- Costa, S., Ferreira, J., Silveira, C., Costa, C., Lopes, D., Relvas, H., Borrego, C., Roebeling, P., Isabel Miranda, A., and Teixeira, J. P. (2014). Integrating health on air quality assessmentreview report on health risks of two major european outdoor air pollutants:PM and NO₂. *Journal of Toxicology and Environmental Health, Part B*, 17:307–340.

- Cressie, N. A. C. (2015). *Statistics for Spatial Data*. Wiley-Blackwell.
- Datta, J. and Ghosh, J. K. (2013). Asymptotic properties of bayes risk for the horseshoe prior. *Bayesian Analysis*, 8(1):111–132.
- Derado, G., Bowman, F. D., and Zhang, L. (2013). Predicting brain activity using a bayesian spatial model. *Statistical Methods in Medical Research*, 22(4):382–397.
- Diggle, P. J., Knorr-Held, L., Rowlingson, B., Su, T., Hawtin, P., and Bryant, T. (2004). On-line monitoring of public health surveillance health data. In Brookmeyer, R. and Stroup, D. F., editors, *Monitoring the Health of Populations: Statistical Principles & Methods for Public Health Surveillance*, chapter 9, pages 233–266. Oxford University Press, New York, NY.
- Dominici, F., Peng, R. D., Bell, M. L., Pham, L., McDermott, A., Zeger, S. L., and Samet, J. M. (2006). Fine particulate air pollution and hospital admission for cardiovascular and respiratory diseases. *Journal of the American Medical Association*, 295(10):1127–1134.
- Donoho, D. L. and Johnstone, J. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455.
- Duane, S., Kennedy, A., Pendleton, B. J., and Roweth, D. (1987). Hybrid monte carlo. *Physics Letters B*, 195(2):216 – 222.
- Dutta, I. and Singh, R. N. (2011). Dynamic in situ x-ray diffraction study of antiferroelectricferroelectric phase transition in strontium-modified lead zirconate titanate ceramics. *Integrated Ferroelectrics*, 131(1):153–172.
- Editorial (2014). Crystallography matters. *Nature Materials*, 13.
- Efron, B. (2004). The estimation of prediction error: covariance penalties and cross-validation. *Journal of the American Statistical Association*, 99(467):619–632.
- Eke, P. I., Dye, B. A., Wei, L., Thornton-Evans, G. O., and Genco, R. J. (2012). Prevalence of periodontitis in adults in the united states: 2009 and 2010. *Journal of Dental Research*, 91(10):914–920.
- Elliott, P., Wakefield, Jon abd Best, N., and Briggs, D. (2001). *Spatial epidemiology: methods and applications*. Oxford University Press.
- Esteves, G., Fancher, C. M., and Jones, J. L. (2015). In situ characterization of polycrystalline ferroelectrics using x-ray and neutron diffraction. *Journal of Materials Research*, 30(3):340356.

- Ferreira, M. C., DiasPereira, A. C., BrancodeAlmeida, L. S., Martins, C. C., and Paiva, S. M. (2017). Impact of periodontal disease on quality of life: a systematic review. *Journal of Periodontal Research*, 52(4):651–665.
- Franklin, M., Zeka, A., and Schwartz, J. (2007). Association between $PM_{2.5}$ and all-cause and specific-cause mortality in 27 US communities. *Journal of Exposure Science and Environmental Epidemiology*, 17:279–287.
- Friedman, J., Hastie, T., Hfling, H., and Tibshirani, R. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2):302–332.
- Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.
- Gazi, O. (2012). New upper bounds for the computation of complementary error function. In *Innovations on Communication Theory*, Istanbul, Turkey.
- Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010). *Handbook of Spatial Statistics*. Chapman & Hall.
- Gelfand, A. E., Kim, H.-J., Sirmans, C. F., and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98(462):387–396.
- George, E. I. and McCulloch, R. E. (1993). Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.
- George, Edward I., M. R. E. (1997). Approaches for bayesian variable selection. *Statistica Sinica*, 7(2):339–373.
- Geweke, J. (1996). Variable selection and model comparison in regression. In *Bayesian Statistics 5*, pages 609–620. University Press.
- Girguis, M. S., Strickland, M. J., Hu, X., Liu, Y., Bartell, S. M., and Vieira, V. M. (2016). Maternal exposure to traffic-related air pollution and birth defects in Massachusetts. *Environmental Research*, 146:1–9.
- Goldsmith, J., Huang, L., and Crainiceanu, C. M. (2014). Smooth scalar-on-image regression via spatial bayesian variable selection. *Journal of Computational and Graphical Statistics*, 23(1):46–64.
- Gorfman, S., Keeble, D. S., Glazer, A. M., Long, X., Xie, Y., Ye, Z.-G., Collins, S., and Thomas, P. A. (2011). High-resolution x-ray diffraction study of single crystals of lead zirconate titanate. *Physical Review B*, 84(2):020102–1–020102–4.

- Griffin, J. E. and Brown, P. J. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, 5(1):171–188.
- Gryparis, A., Paciorek, C. J., Zeka, A., Schwartz, J., and Coull, B. A. (2009). Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics*, 10(2):258–274.
- Guarnieri, M. and Balmes, J. R. (2014). Outdoor air pollution and asthma. *Lancet*, 383(9928):1581–1592.
- Hay, S. I., George, D. B., Moyes, C. L., and Brownstein, J. S. (2013). Big data opportunities for global infectious disease surveillance. *PLOS Medicine*, 10(4):1–4.
- He, Liangtian, W. Y. and Xiang, Z. (2017). Wavelet frame-based image restoration using sparsity, nonlocal and support prior of frame coefficients. *The Visual Computer*, abs/1510.02866.
- Heaton, M. J., Banks, D. L., Zou, J., Karr, A., Datta, G., Lynch, J., and Vera, F. (2012). A spatio-temporal absorbing state model for disease and syndromic surveillance. *Statistics in medicine*, 31:2123–2136.
- Hoffman, J. I. E. and Kaplan, S. (2002). The incidence of congenital heart disease. *Journal of the American College of Cardiology*, 39(12):1890–1900.
- Ierodiakonou, D., Zanobetti, A., Coull, B. A., Melly, S., Postma, D. S., Boezen, H. M., Vonk, J. M., Williams, P. V., Shapiro, G. G., McKone, E. F., Hallstrand, T. S., Koenig, J. Q., Schildcrout, J. S., Lumley, T., Fuhlbrigge, A. N., Koutrakis, P., Schwartz, J., Weiss, S. T., and Gold, D. R. (2016). Ambient air pollution, lung function and airway responsiveness in children with asthma. *Journal of Allergy and Clinical Immunology*, 137(2):390–399.
- Jacquez, G. M. (1996). A k nearest neighbor test for space-time interaction. *Statistics in Medicine*, 15(18):1935–1949.
- Jansen, M. (2001). *Noise reduction by wavelet thresholding*, volume 161. Springer, New York.
- Janssen, N., Fischer, P., Marra, M., Ameling, C., and Cassee, F. (2013). Short-term effects of pm2.5, pm10 and pm2.510 on daily mortality in the netherlands. *Science of The Total Environment*, 463-464:20 – 26.
- Jhuang, A.-T., Fuentes, M., Jones, J. L., Esteves, G., Fancher, C. M., Furman, M., and Reich, B. J. (2017). Spatial signal detection using continuous shrinkage priors. Unpublished Manuscript.

- Jin, I. H., Yuan, Y., and Bandyopadhyay, D. (2016). A Bayesian hierarchical spatial model for dental careis assessment using non-Gaussian Markov random fields. *The Annals of Applied Statistics*, 10(2):884–905.
- Johannsen, A., Fored, M. C., Hkansson, J., Ekbom, A., and Gustafsson, A. (2015). Consumption of dental treatment in patients with inflammatory bowel disease, a register study. *PLoS One*, 10(8).
- Johnson, V. E. and Rossell, D. (2012). Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association*, 107(498):649–660.
- Kang, J., Reich, B. J., and Staicu, A.-M. (2018). Scalar-on-image regression via the soft-thresholded gaussian process. *Biometrika*, 105(1):165–184.
- Katrusiak, A. (2008). High-pressure crystallography. *Acta Crystallographica Section A*, 64(1):135–148.
- Khreis, H., Kelly, C., Tate, J., Parslow, R., Lucas, K., and Nieuwenhuijsen, M. (2017). Exposure to traffic-related air pollution and risk of development of childhood asthma: A systematic review and meta-analysis. *Environment International*, 100(Supplement C):1 – 31.
- Kleinman, K., Lazarus, R., and Platt, R. (2004). A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. *American Journal of Epidemiology*, 159(3):217–224.
- KnorrHeld, L. (2000). Bayesian modelling of inseparable spacetime variation in disease risk. *Statistics in Medicine*, 19:2555–2567.
- Knox, E. G. and Bartlett, M. S. (1964). The detection of space-time interactions. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 13(1):25–30.
- Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics - Theory and Methods*, 26(6):1481–1496.
- Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(1):61–72.
- Kulldorff, M., Heffernan, R., Hartman, J., Assuno, R., and Mostashari, F. (2005). A spacetime permutation scan statistic for disease outbreak detection. *PLOS Medicine*, 2(3):e59.
- Lawson, A., Clark, A., and Vidal Rodeiro, C. (2004). Developments in general and syndromic surveillance for small area health data. *Journal of Applied Statistics*, 31(8):951–966.

- Lee, B.-J., Kim, B., and Kyohong, L. (2014). Air pollution exposure and cardiovascular disease. *Toxicological Research*, 30(2):71–75.
- Lee, E. C., Asher, J. M., Goldlust, S., Kraemer, J. D., Lawson, A. B., and Bansal, S. (2016). Mind the scales: Harnessing spatial big data for infectious disease surveillance and inference. *The Journal of Infectious Diseases*, 214(4):S409–S413.
- Lelieveld, J., Evans, J. S., Fnais, M., Giannadaki, D., and A., P. (2015). The contribution of outdoor air pollution sources to premature mortality on a global scale. *Nature*, 525:367–371.
- Li, F., Zhang, T., Wang, Q., Gonzalez, M. Z., Maresh, E. L., and Coan, J. A. (2015). Spatial bayesian variable selection and grouping for high-dimensional scalar-on-image regression. *The Annals of Applied Statistics*, 9(2):687–713.
- Li, G., Best, N., Hansell, A. L., Ahmed, I., and Richardson, S. (2012). BaySTDetect: detecting unusual temporal patterns in small area data via bayesian model choice. *Biostatistics*, 13(4):695–710.
- Liao, Y., Li, X., and Wang, J. (2017). Chapter 17 - the study on modified spatial scan statistic. In Yang, W., editor, *Early Warning for Infectious Disease Outbreak*, pages 329 – 342. Academic Press.
- Liu, B., Qiu, X., Zhu, T., Tian, W., Hu, R., Ekholm, S., Schifitto, G., and Zhong, J. (2016). Spatial regression analysis of serial dti for subject-specific longitudinal changes of neurodegenerative disease. *NeuroImage: Clinical*, 11:291 – 301.
- Long, J., Cai, Q., Steinwandel, M., Hargreaves, M. K., Bordenstein, S. R., Blot, W. J., Zheng, W., and Shu, X. O. (2017). Association of oral microbiome with type 2 diabetes risk. *Journal of Periodontal Research*, 52:636–643.
- Malley, C. S., Kuylenstierna, J. C., Vallack, H. W., Henze, D. K., Blencowe, H., and Ashmore, M. R. (2017). Preterm birth associated with maternal fine particulate matter exposure: A global, regional and national assessment. *Environmental International*, 101:173–182.
- Mallick, H. and Yi, N. (2013). Bayesian methods for high dimensional linear models. *Journal of Biometrics and Biostatistics*, 5(1).
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27(2 Part 1):209–220.
- Marshall, E. G., Harris, G., and Wartenberg, D. (2010). Oral cleft defects and maternal exposure to ambient air pollutants in New Jersey. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 88(4):205–215.

- Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032.
- Molitor, J., Jerrett, M., Chang, C.-C., Molitor, N.-T., Gauderman, J., Berhane, K., McConnell, R., Lurmann, F., Wu, J., Winer, A., and Thomas, D. (2007). Assessing uncertainty in spatial exposure models for air pollution health effects assessment. *Environmental Health Perspectives*, 115(8):1147–1153.
- Moore, K. L., Persaud, T. V. N., and Torchia, M. G. (2015). *The developing human*. Elsevier Health Sciences, Philadelphia, PA, 10 edition.
- Mori, S. and Zhang, J. (2006). Principles of diffusion tensor imaging and its applications to basic neuroscience research. *Neuron*, 51(5):527 – 539.
- Naus, J. I. (1965). The distribution of the size of the maximum cluster of points on a line. *Journal of the American Statistical Association*, 60(310):532–538.
- Neal, R. M. (1994). An improved acceptance procedure for the hybrid monte carlo algorithm. *Journal of Computational Physics*, 111(1):194–203.
- Neal, R. M. (2011). Chapter 5 - mcmc using hamiltonian dynamics. In Brooks, S., Gelman, A., Jones, G. L., and Meng, X.-L., editors, *Handbook of Markov Chain Monte Carlo*, pages 113 – 162. Chapman Hall & CRC.
- Nelsen, R. B. (2006). *An introduction to copulas*. Springer, New York.
- Ng, E. W. and Geller, M. (1969). A table of integrals of the error functions. *Journal of Research of the National Institute of Standards and Technology - B*, 73(1):1–20.
- Nicolosi, L. N., Lewin, P. G., Rudzinski, J. J., Pompeo, M., Guanica, F., Rodriguez, P., Gelpi, R. J., and Rubio, M. C. (2017). Relation between periodontal disease and arterial stiffness. *Journal of Periodontal Research*, 52(1):122–126.
- O’hara, R. B. and Sillapää, M. J. (2009). A review of bayesian variable selection methods: what, how and which. *Bayesian Analysis*, 4(1):85–118.
- Papageorgiou, S. N., Hagner, M., Nogueira, A. V. B., Franke, A., Jger, A., and Deschner, J. (2017). Inflammatory bowel disease and oral health: systematic review and a metaanalysis. *Journal of Clinical Periodontology*, 44(4):382–393.
- Pearson, K. F. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11):559–572.

- Pereira, G., Bracken, M. B., and Bell, M. L. (2016). Particulate air pollution, fetal growth and gestational length: The influence of residential mobility in pregnancy. *Environmental Research*, 147:269–274.
- Persson, G. R. and Persson, R. E. (2008). Cardiovascular disease and periodontitis: an update on the associations and risk. *Journal of Clinical Periodontology*, 35(s8):362–379.
- Piironen, J. and Vehtari, A. (2016). On the hyperprior choice for the global shrinkage parameter in the horseshoe prior. *ArXiv e-prints*.
- Reefhuis, J., Gilboa, S. M., Anderka, M., Browne, M. L., Feldkamp, M. L., Hobbs, C. A., Jenkins, M. M., Langlois, P. H., Newsome, K. B., Olshan, A. F., Romitti, P. A., Shapira, S. K., Shaw, G. M., Tinker, S. C., Honein, M. A., and the National Birth Defects Prevention Study (2015). The national birth defects prevention study: A review of the methods. *Birth Defects Research Part A: Clinical and Molecular Teratology*, 103(8):656–669.
- Reich, B. J. and Bandyopadhyay, D. (2010). Modeling longitudinal spatial periodontal data: a spatially-adaptive model with tools for specifying priors and checking fit. *Annals of Applied Statistics*, 4(1):439–459.
- Reich, B. J., Bandyopadhyay, D., and Bondell, H. D. (2013). A nonparametric spatial model for periodontal data with non-random missingness. *Journal of the American Statistical Association*, 108(503):820–831.
- 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(3):790–799.
- Reich, B. J., Hodges, J. S., and Carlin, B. P. (2007). Spatial analyses of periodontal data using conditionally autoregressive priors having two classes of neighbor relations. *Journal of the American Statistical Association*, 102(477):44–55.
- Reller, M. D., Strickland, M. J., Riehle-Colarusso, T., Mahle, W. T., and Correa, A. (2008). Prevalence of congenital heart defects in metropolitan Atlanta, 1998–2005. *The Journal of Pediatrics*, 153(6):807–813.
- Robertson, C., Nelson, T. A., MacNab, Y. C., and Lawson, A. B. (2010). Review of methods for spacetemporal disease surveillance. *Spatial and Spatio-temporal Epidemiology*, 1(2):105 – 116.
- Rogerson, P. A. (1997). Surveillance systems for monitoring the development of spatial patterns. *Statistics in Medicine*, 16(18):2081–2093.

- Rogerson, P. A. and Jacquez, G. (2016). Statistical tests for clustering and surveillance. In Lawson, A. B., Banerjee, S., Haining, R. P., and Ugarte, M. D., editors, *Handbook of Spatial Epidemiology*, chapter 8, pages 161–177. Chapman and Hall/CRC, Boca Raton, FL.
- Rogerson, P. A. and Yamada, I. (2004). Monitoring change in spatial patterns of disease: comparing univariate and multivariate cumulative sum approaches. *Statistics in Medicine*, 23(14):2195–2214.
- Rokov, V. and George, E. I. (In press). The spike-and-slab lasso. *Journal of the American Statistical Association*.
- Sarnat, S. E., Sarnat, J. A., Mulholland, J., Isakov, V., Özkaynak, H., Chang, H. H., Klein, M., and Tolbert, P. E. (2013). Application of alternative spatiotemporal metrics of ambient air pollution exposure in a time-series epidemiological study in atlanta. *Journal of exposure science and environmental epidemiology*, 23:595–605.
- Shah, A. S., Langrish, J. P., Nair, H., McAllister, D. A., Hunter, A. L., Donaldson, K., Newby, D. E., and Mills, N. L. (2013). Global association of air pollution and heart failure: a systematic review and meta-analysis. *Lancet*, 382(9897):1039–1048.
- Sheppard, L., Burnett, R. T., Szpiro, A. A., Kim, S.-Y., Jerrett, M., Pope, C., and Brunekreef, B. (2012). Confounding and exposure measurement error in air pollution. *Air quality, atmosphere, & health*, 5:203–216.
- Simonsen, L., Gog, J. R., Olson, D., and Viboud, C. (2016). Infectious disease surveillance in the big data era: Towards faster and locally relevant systems. *The Journal of Infectious Diseases*, 214(4):S380–S385.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.
- Stingone, J. A., Luben, T. J., Daniels, J. L., Fuentes, M., Richardson, D. B., Aylsworth, A. S., Herring, A. H., Anderka, M., Botto, L., Correa, A., Gilboa, S. M., Langlois, P. H., Mosley, B., Shaw, G. M., Siffel, C., F., O. A., and National Birth Defects Prevention Study (2014). Maternal exposure to criteria air pollutants and congenital heart defects in offspring: results from the National Birth Defects Prevention Study. *Environmental Health Perspectives*, 122:863–872.
- Sun, W., Reich, B. J., Cai, T. T., Guindani, M., and Schwartzman, A. (2015). False discovery control in large-scale spatial multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):59–83.

- Taswell, C. (2000). The what, how, and why of wavelet shrinkage denoising. *Computing in Science Engineering*, 2(3):12–19.
- Thomas, J. M. (2012). The birth of x-ray crystallography. *Nature*, 491:186–187.
- Unkel, S., Farrington, C. P., Garthwaite, P. H., Robertson, C., and Andrews, N. (2012). Statistical methods for the prospective detection of infectious disease outbreaks: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 175(1):49–82.
- Van den Hooven, E. H., Pierik, F. H., de Kluizenaar, Y., Willemsen, S. P., Hofman, A., van Ratingen, S. W., Zandveld, P. Y. J., Mackenbach, J. P., Steegers, E. A. P., Miedema, H. M. E., and Jaddoe, V. W. V. (2012). Air pollution exposure during pregnancy, ultrasound measures of fetal growth, and adverse birth outcomes: A prospective cohort study. *Environmental Health Perspectives*, 120(1):150–156.
- van der Pas, S., Szab, B., and van der Vaart, A. (2017). Uncertainty quantification for the horseshoe. *Bayesian Analysis*, 12(4):1221–1274.
- van der Pas, S. L., Kleijn, B. J. K., and van der Vaart, A. W. (2014). The horseshoe estimator: posterior concentration around nearly black vectors. *Electronic Journal of Statistics*, 8(2):2585–2618.
- Vavricka, S. R., Manser, C. N., Hediger, S., Vögelin, M., Scharl, M. M., Biedermann, L. M., Rogler, S., Seibold, F., Sanderink, R., Attin, T., Schoepfer, A., Fried, M., Rogler, G., and Frei, P. (2013). Periodontitis and gingivitis in inflammatory bowel disease: A casecontrol study. *Inflammatory Bowel Diseases*, 19(13):2768–2777.
- Vidal Rodeiro, C. L. and Lawson, A. B. (2006). Monitoring changes in spatiotemporal maps of disease. *Biometrical Journal*, 48(3):463–480.
- Warren, J., Fuentes, M., Herring, A., and Langlois, P. (2012). Spatial-temporal modeling of the association between air pollution exposure and preterm birth: Identifying critical windows of exposure. *Biometrics*, 68(4):1157–1167.
- Warren, J. L., Son, J.-Y., Pereira, G., Leaderer, B. P., and Bell, M. L. (2017). Investigating the impact of maternal residential mobility on identifying critical windows of susceptibility to ambient air pollution during pregnancy. *American journal of epidemiology*, page kwx335.
- Watanabe, S. (2010). Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, 11:35713591.

- Watkins, R. E., Eagleson, S., Veenendaal, B., Wright, G., and Plant, A. J. (2009). Disease surveillance using a hidden markov model. *BMC Medical Informatics and Decision Making*, 9(1):39.
- World Health Organization (2007). Global early warning system for major animal diseases, including zoonoses (GLEWS). <http://www.who.int/zoonoses/outbreaks/glews/en/>.
- Xing, Y.-F., Xu, Y.-H., Shi, M.-H., and Lian, Y.-X. (2016). The impact of pm2.5 on the human respiratory system. *Journal of Thoracic Disease*, 8(1):E69–E74.
- Yadav, M., Yadav, S., and Sharma, D. (2014). Image denoising using orthonormal wavelet transform with stein unbiased risk estimator. In *Electrical, Electronics and Computer Science (SCEECS), 2014 IEEE Students’ Conference*, pages 1–4.
- Yoon, P. W., Rasmussen, S. A., Lynberg, M. C., Moore, C. A., Anderka, M., Carmichael, S. L., Costa, P., Druschel, C., Hobbs, C. A., Romitti, P. A., Langlois, P. H., and Edmonds, L. D. (2001). The national birth defects prevention study. *Public Health Reports*, 116(1_suppl):32–40.
- Yorifuji, T., Kashima, S., and Doi, H. (2016). Acute exposure to fine and coarse particulate matter and infant mortality in Tokyo, Japan (20022013). *Science of the Total Environment*, 551-552:66–72.
- Yuan, M. and Lin, Y. (2005). Efficient empirical bayes variable selection and estimation in linear models. *Journal of the American Statistical Association*, 100(472):1215–1225.
- Zhou, H. and Lawson, A. B. (2008). EWMA smoothing and bayesian spatial modeling for health surveillance. *Statistics in Medicine*, 27(28):5907–5928.

APPENDIX

Appendix A

Supplementary Materials for Chapter 2

A.1 Theoretical derivations

A.1.1 The joint horseshoe distribution

Consider $\boldsymbol{\beta} \mid \lambda \sim N(\mathbf{0}, \lambda^2 \boldsymbol{\Sigma}_{\boldsymbol{\beta}})$ with the joint shrinkage parameter $\lambda \sim C^+(0, 1)$ and a general covariance matrix with the diagonal elements 1 and the covariance $\rho_{\boldsymbol{\beta}}$,

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = \begin{pmatrix} 1 & \rho_{\boldsymbol{\beta}} \\ \rho_{\boldsymbol{\beta}} & 1 \end{pmatrix}.$$

The prior of β_1 and β_2 is

$$\begin{aligned} P_1(\beta_1, \beta_2) &= \int_0^\infty P_1(\beta_1, \beta_2 \mid \lambda) \cdot P(\lambda) \, d\lambda \\ &= \int_0^\infty \frac{1}{2\pi\lambda^2\sqrt{1-\rho_{\boldsymbol{\beta}}^2}} e^{-\frac{\beta_1^2 - 2\rho_{\boldsymbol{\beta}}\beta_1\beta_2 + \beta_2^2}{2\lambda^2(1-\rho_{\boldsymbol{\beta}}^2)}} \cdot \frac{2}{\pi(1+\lambda^2)} \, d\lambda \\ &= \frac{1}{\pi^2\sqrt{1-\rho_{\boldsymbol{\beta}}^2}} \int_0^\infty \frac{1}{\lambda^2(1+\lambda^2)} e^{-\frac{\beta_1^2 - 2\rho_{\boldsymbol{\beta}}\beta_1\beta_2 + \beta_2^2}{2\lambda^2(1-\rho_{\boldsymbol{\beta}}^2)}} \, d\lambda. \end{aligned}$$

Let $u = \frac{1}{\lambda}$, $d\lambda = -\lambda^2 du = -\frac{1}{u^2} du$, $c_1 = \pi^2 \sqrt{1 - \rho_{\beta}^2}$ and $c_2 = \frac{\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2}{2(1 - \rho_{\beta}^2)}$.

$$\begin{aligned}
P_1(\beta_1, \beta_2) &= c_1 \int_0^{\infty} \frac{u^2}{u^2 + 1} e^{-c_2 u^2} du \\
&= c_1 \int_0^{\infty} \left(1 - \frac{1}{u^2 + 1}\right) e^{-c_2 u^2} du \\
&= c_1 \left(\int_0^{\infty} e^{-c_2 u^2} du - \int_0^{\infty} \frac{1}{u^2 + 1} e^{-c_2 u^2} du \right) \\
&= c_1 \left(\frac{\sqrt{\pi}}{2\sqrt{c_2}} - \int_0^{\infty} \frac{1}{u^2 + 1} e^{-c_2 u^2} du \right).
\end{aligned}$$

According to Formula 7 in Section 3.2 of Ng and Geller (1969),

$$\operatorname{erfc}(az) = \frac{2z}{\pi} e^{-a^2 z^2} \int_0^{\infty} \frac{e^{-a^2 t^2}}{t^2 + z^2} dt,$$

where $\operatorname{erfc}(\cdot)$ is the complementary error function. Take $a = \sqrt{c_2}$ and $z = 1$,

$$\begin{aligned}
\operatorname{erfc}(\sqrt{c_2}) &= \frac{2}{\pi} e^{-c_2} \int_0^{\infty} \frac{e^{-c_2 t^2}}{t^2 + 1} dt \\
\int_0^{\infty} \frac{e^{-c_2 t^2}}{t^2 + 1} dt &= \frac{\pi}{2} e^{c_2} \operatorname{erfc}(\sqrt{c_2}).
\end{aligned}$$

Thus,

$$\begin{aligned}
P_1(\beta_1, \beta_2) &= c_1 \left(\frac{\sqrt{\pi}}{2\sqrt{c_2}} - \int_0^{\infty} \frac{1}{u^2 + 1} e^{-c_2 u^2} du \right) \\
&= c_1 \left[\frac{\sqrt{\pi}}{2\sqrt{c_2}} - \frac{\pi}{2} e^{c_2} \operatorname{erfc}(\sqrt{c_2}) \right] \\
&= 2^{-\frac{1}{2}} \pi^{-\frac{3}{2}} (\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2)^{-\frac{1}{2}} \\
&\quad - (2\pi)^{-1} (1 - \rho_{\beta}^2)^{-\frac{1}{2}} e^{\frac{\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2}{2(1 - \rho_{\beta}^2)}} \operatorname{erfc} \left[\sqrt{\frac{\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2}{2(1 - \rho_{\beta}^2)}} \right].
\end{aligned} \tag{A.1}$$

When $\rho_{\beta} = 0$, the density becomes

$$P_1(\beta_1, \beta_2) = 2^{-\frac{1}{2}} \pi^{-\frac{3}{2}} (\beta_1^2 + \beta_2^2)^{-\frac{1}{2}} - (2\pi)^{-1} e^{\frac{1}{2}(\beta_1^2 + \beta_2^2)} \operatorname{erfc} \left(\sqrt{\frac{\beta_1^2 + \beta_2^2}{2}} \right). \tag{A.2}$$

Regarding the bounds of the joint density,

$$\begin{aligned} (|\beta_1| - |\beta_2|)^2 &\leq \beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2 \leq (|\beta_1| + |\beta_2|)^2, \\ \frac{1}{|\beta_1| + |\beta_2|} &\leq (\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2)^{-\frac{1}{2}} \leq \frac{1}{||\beta_1| - |\beta_2||}. \end{aligned} \quad (\text{A.3})$$

According to Theorem 2 in Chang et al. (2011) and Equation 11 in Gazi (2012),

$$\frac{1}{2}e^{-2x^2} \leq \text{erfc}(x) \leq \frac{1}{2}e^{-2x^2} + \frac{1}{2}e^{-x^2}. \quad (\text{A.4})$$

Then,

$$\begin{aligned} \frac{1}{2}e^{-\frac{\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2}{2(1-\rho_{\beta}^2)}} &\leq e^{\frac{\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2}{2(1-\rho_{\beta}^2)}} \text{erfc}\left[\sqrt{\frac{\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2}{2(1-\rho_{\beta}^2)}}\right], \\ e^{\frac{\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2}{2(1-\rho_{\beta}^2)}} \text{erfc}\left[\sqrt{\frac{\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2}{2(1-\rho_{\beta}^2)}}\right] &\leq \frac{1}{2}\left[e^{-\frac{\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2}{2(1-\rho_{\beta}^2)}} + 1\right]. \end{aligned} \quad (\text{A.5})$$

By (A.3) and (A.5), the lower bound and the upper bound of $P_1(\beta_1, \beta_2)$ are

$$\begin{aligned} 2^{-\frac{1}{2}}\pi^{-\frac{3}{2}}\frac{1}{|\beta_1| + |\beta_2|} - (4\pi)^{-1}(1 - \rho_{\beta}^2)^{-\frac{1}{2}}\left[e^{-\frac{\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2}{2(1-\rho_{\beta}^2)}} + 1\right] &\leq P_1(\beta_1, \beta_2), \\ P_1(\beta_1, \beta_2) &\leq 2^{-\frac{1}{2}}\pi^{-\frac{3}{2}}\frac{1}{||\beta_1| - |\beta_2||} - (4\pi)^{-1}(1 - \rho_{\beta}^2)^{-\frac{1}{2}}e^{-\frac{\beta_1^2 - 2\rho_{\beta}\beta_1\beta_2 + \beta_2^2}{2(1-\rho_{\beta}^2)}}. \end{aligned} \quad (\text{A.6})$$

A.1.2 Concentration around zero

We show that the joint density (A.1) asymptotes at the origin. This indicates that the joint horseshoe density encourages spatial clusters of null coefficients to be simultaneously shrunk towards zero for both variables. However, the independent density, the product of two independent horseshoe densities, also asymptotes at zero. Therefore, we demonstrate the benefits of the joint density by examining the ratio of the joint and independent densities at the origin.

First, the concentration of $P_1(\beta_1, \beta_2)$ towards zero is

$$\begin{aligned} \lim_{\beta_1, \beta_2 \rightarrow 0} P_1(\beta_1, \beta_2) &= \lim_{\beta_1, \beta_2 \rightarrow 0} \left\{ 2^{-\frac{1}{2}} \pi^{-\frac{3}{2}} (\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2)^{-\frac{1}{2}} \right. \\ &\quad \left. - (2\pi)^{-1} (1 - \rho_{\beta}^2)^{-\frac{1}{2}} e^{\frac{\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2}{2(1 - \rho_{\beta}^2)}} \operatorname{erfc} \left[\sqrt{\frac{\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2}{2(1 - \rho_{\beta}^2)}} \right] \right\} \\ &= \infty, \end{aligned} \tag{A.7}$$

since $\lim_{\beta_1, \beta_2 \rightarrow 0} e^{\frac{\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2}{2(1 - \rho_{\beta}^2)}} \operatorname{erfc} \left[\sqrt{\frac{\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2}{2(1 - \rho_{\beta}^2)}} \right] = 1 \cdot 1 = 1$ and $\lim_{\beta_1, \beta_2 \rightarrow 0} (\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2)^{-\frac{1}{2}} = \infty$.

Next, let R be the ratio of the joint and independent densities with independent β and $P_0(\beta_1, \beta_2)$ be the independent density. Then,

$$\begin{aligned} R &= \frac{P_1(\beta_1, \beta_2)}{P_0(\beta_1, \beta_2)} \\ &= \frac{2^{-\frac{1}{2}} \pi^{-\frac{3}{2}} (\beta_1^2 + \beta_2^2)^{-\frac{1}{2}} - (2\pi)^{-1} e^{\frac{1}{2}(\beta_1^2 + \beta_2^2)} \operatorname{erfc} \left(\sqrt{\frac{\beta_1^2 + \beta_2^2}{2}} \right)}{\frac{1}{2\pi^3} e^{\frac{1}{2}(\beta_1^2 + \beta_2^2)} E_1 \left(\frac{\beta_1^2}{2} \right) E_1 \left(\frac{\beta_2^2}{2} \right)} \\ &= 2^{\frac{1}{2}} \pi^{\frac{3}{2}} \frac{1}{e^{\frac{1}{2}(\beta_1^2 + \beta_2^2)} (\beta_1^2 + \beta_2^2)^{\frac{1}{2}} E_1 \left(\frac{\beta_1^2}{2} \right) E_1 \left(\frac{\beta_2^2}{2} \right)} - \pi^2 \frac{\operatorname{erfc} \left(\sqrt{\frac{\beta_1^2 + \beta_2^2}{2}} \right)}{E_1 \left(\frac{\beta_1^2}{2} \right) E_1 \left(\frac{\beta_2^2}{2} \right)}. \end{aligned} \tag{A.8}$$

To show the benefit from the joint density, we prove that $\lim_{\beta_1, \beta_2 \rightarrow 0} R = \infty$ by showing that: (a) the former term goes to infinity, and (b) the latter term goes to zero in (A.8).

To show (a), we find an upper bound of $e^{\frac{1}{2}(\beta_1^2 + \beta_2^2)} (\beta_1^2 + \beta_2^2)^{\frac{1}{2}} E_1 \left(\frac{\beta_1^2}{2} \right) E_1 \left(\frac{\beta_2^2}{2} \right)$,

$$\begin{aligned} &e^{\frac{1}{2}(\beta_1^2 + \beta_2^2)} (\beta_1^2 + \beta_2^2)^{\frac{1}{2}} E_1 \left(\frac{\beta_1^2}{2} \right) E_1 \left(\frac{\beta_2^2}{2} \right) \\ &< e^{\frac{1}{2}(\beta_1^2 + \beta_2^2)} (\beta_1^2 + \beta_2^2)^{\frac{1}{2}} \left[e^{-\frac{\beta_1^2}{2}} \log \left(1 + \frac{2}{\beta_1^2} \right) \right] \left[e^{-\frac{\beta_2^2}{2}} \log \left(1 + \frac{2}{\beta_2^2} \right) \right] \\ &= (\beta_1^2 + \beta_2^2)^{\frac{1}{2}} \log \left(1 + \frac{2}{\beta_1^2} \right) \log \left(1 + \frac{2}{\beta_2^2} \right). \end{aligned}$$

Let $v_1 = \frac{1}{\beta_1}$ and $v_2 = \frac{1}{\beta_2}$, then we can rewrite the upper bound as

$$\left(\frac{1}{v_1^2} + \frac{1}{v_2^2}\right)^{\frac{1}{2}} \log(1 + 2v_1^2) \log(1 + 2v_2^2) = \frac{(v_1^2 + v_2^2)^{\frac{1}{2}} \log(1 + 2v_1^2) \log(1 + 2v_2^2)}{|v_1 v_2|}.$$

The numerator and denominator go to infinity as v_1 and v_2 go to infinity (i.e. β_1 and β_2 go to zero). Apply L'Hospital's rule,

$$\begin{aligned} & \frac{\partial^2}{\partial v_1 \partial v_2} (v_1^2 + v_2^2)^{\frac{1}{2}} \log(1 + 2v_1^2) \log(1 + 2v_2^2) \\ &= \frac{\partial}{\partial v_2} \left[v_1 (v_1^2 + v_2^2)^{-\frac{1}{2}} \log(1 + 2v_1^2) \log(1 + 2v_2^2) + (v_1^2 + v_2^2)^{\frac{1}{2}} \frac{4v_1}{1 + 2v_1^2} \log(1 + 2v_2^2) \right] \\ &= -v_1 v_2 (v_1^2 + v_2^2)^{-\frac{3}{2}} \log(1 + 2v_1^2) \log(1 + 2v_2^2) + v_1 (v_1^2 + v_2^2)^{-\frac{1}{2}} \log(1 + 2v_1^2) \frac{4v_2}{1 + 2v_2^2} \\ & \quad + v_2 (v_1^2 + v_2^2)^{-\frac{1}{2}} \frac{4v_1}{1 + 2v_1^2} \log(1 + 2v_2^2) + (v_1^2 + v_2^2)^{-\frac{1}{2}} \frac{4v_1}{1 + 2v_1^2} \frac{4v_2}{1 + 2v_2^2} \end{aligned} \tag{A.9}$$

and

$$\begin{aligned} \frac{\partial^2}{\partial v_1 \partial v_2} |v_1 v_2| &= 1, \quad \text{if } v_1 v_2 > 0 \\ &= -1, \quad \text{if } v_1 v_2 < 0. \end{aligned}$$

The four additive terms in (A.9) approach zero when v_1 and v_2 go to infinity because polynomial functions converge faster than logarithmic functions. Then,

$$\begin{aligned} & \lim_{\beta_1, \beta_2 \rightarrow 0} (\beta_1^2 + \beta_2^2)^{\frac{1}{2}} \log\left(1 + \frac{2}{\beta_1^2}\right) \log\left(1 + \frac{2}{\beta_2^2}\right) \\ &= \lim_{v_1, v_2 \rightarrow \infty} \left(\frac{1}{v_1^2} + \frac{1}{v_2^2}\right)^{\frac{1}{2}} \log(1 + 2v_1^2) \log(1 + 2v_2^2) \\ &= \lim_{v_1, v_2 \rightarrow \infty} \frac{\frac{\partial^2}{\partial v_1 \partial v_2} \left[(v_1^2 + v_2^2)^{\frac{1}{2}} \log(1 + 2v_1^2) \log(1 + 2v_2^2) \right]}{\frac{\partial^2}{\partial v_1 \partial v_2} |v_1 v_2|} \\ &= 0. \end{aligned} \tag{A.10}$$

Since $e^{\frac{1}{2}(\beta_1^2+\beta_2^2)}(\beta_1^2+\beta_2^2)^{\frac{1}{2}}E_1\left(\frac{\beta_1^2}{2}\right)E_1\left(\frac{\beta_2^2}{2}\right)$ is positive and (A.10), we obtain that

$$\lim_{\beta_1, \beta_2 \rightarrow 0} e^{\frac{1}{2}(\beta_1^2+\beta_2^2)}(\beta_1^2+\beta_2^2)^{\frac{1}{2}}E_1\left(\frac{\beta_1^2}{2}\right)E_1\left(\frac{\beta_2^2}{2}\right) = 0.$$

Namely,

$$\lim_{\beta_1, \beta_2 \rightarrow 0} e^{-\frac{1}{2}(\beta_1^2+\beta_2^2)}(\beta_1^2+\beta_2^2)^{-\frac{1}{2}}\left[E_1\left(\frac{\beta_1^2}{2}\right)E_1\left(\frac{\beta_2^2}{2}\right)\right]^{-1} = \infty. \quad (\text{A.11})$$

To show (b), we find an upper bound for the latter term which goes to zero as β_1 and β_2 go to zero. According to the upper bound in (A.4),

$$\begin{aligned} \frac{\text{erfc}\left(\sqrt{\frac{\beta_1+\beta_2}{2}}\right)}{E_1\left(\frac{\beta_1^2}{2}\right)E_1\left(\frac{\beta_2^2}{2}\right)} &< \frac{2\left[\frac{1}{2}e^{-(\beta_1^2+\beta_2^2)} + \frac{1}{2}e^{-\frac{1}{2}(\beta_1^2+\beta_2^2)}\right]}{\frac{1}{4}e^{-\frac{1}{2}(\beta_1^2+\beta_2^2)}\log\left(1+\frac{4}{\beta_1^2}\right)\log\left(1+\frac{4}{\beta_2^2}\right)} \\ &= \frac{2\left[e^{-\frac{1}{2}(\beta_1^2+\beta_2^2)+1}\right]}{\log\left(1+\frac{4}{\beta_1^2}\right)\log\left(1+\frac{4}{\beta_2^2}\right)} \\ &\rightarrow 0, \end{aligned} \quad (\text{A.12})$$

as β_1 and β_2 go to zero.

Thus,

$$\begin{aligned} \lim_{\beta_1, \beta_2 \rightarrow 0} R &= 2^{\frac{1}{2}}\pi^{\frac{3}{2}} \lim_{\beta_1, \beta_2 \rightarrow 0} e^{-\frac{1}{2}(\beta_1^2+\beta_2^2)}\left(\frac{\beta_1^2+\beta_2^2}{2}\right)^{-\frac{1}{2}}\left[E_1\left(\frac{\beta_1^2}{2}\right)E_1\left(\frac{\beta_2^2}{2}\right)\right]^{-1} \\ &\quad - \pi^2 \lim_{\beta_1, \beta_2 \rightarrow 0} \text{erfc}\left(\sqrt{\frac{\beta_1^2+\beta_2^2}{2}}\right)\left[E_1\left(\frac{\beta_1^2}{2}\right)E_1\left(\frac{\beta_2^2}{2}\right)\right]^{-1} \\ &= 2^{\frac{1}{2}}\pi^{\frac{3}{2}} \lim_{\beta_1, \beta_2 \rightarrow 0} e^{-\frac{1}{2}(\beta_1^2+\beta_2^2)}\left(\frac{\beta_1^2+\beta_2^2}{2}\right)^{-\frac{1}{2}}\left[E_1\left(\frac{\beta_1^2}{2}\right)E_1\left(\frac{\beta_2^2}{2}\right)\right]^{-1} - 0 \\ &= \infty. \end{aligned} \quad (\text{A.13})$$

A.1.3 Tail behavior

In this subsection, we show that the joint density allows for both random variables to be simultaneously large to accommodate spatial clusters of large signals. We also exhibit the advantage of the joint density by the ratio of the joint and independent densities.

Let $\lim_{\beta_1, \beta_2 \rightarrow \infty} P_1(\beta_1, \beta_2)$ represent the tail behavior of the joint density,

$$\begin{aligned} \lim_{\beta_1, \beta_2 \rightarrow \infty} P_1(\beta_1, \beta_2) &= \lim_{\beta_1, \beta_2 \rightarrow \infty} \left\{ 2^{-\frac{1}{2}} \pi^{-\frac{3}{2}} (\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2)^{-\frac{1}{2}} \right. \\ &\quad \left. - (2\pi)^{-1} (1 - \rho_{\beta}^2)^{\frac{1}{2}} e^{\frac{\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2}{2(1 - \rho_{\beta}^2)}} \operatorname{erfc} \left[\sqrt{\frac{\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2}{2(1 - \rho_{\beta}^2)}} \right] \right\}. \end{aligned} \quad (\text{A.14})$$

Apparently, $P_1(\beta_1, \beta_2) = 0$ when $|\rho_{\beta}| = 1$. Otherwise, if $0 \leq |\rho_{\beta}| < 1$,

$$\begin{aligned} (\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2)^{-\frac{1}{2}} &= [(\beta_1 - \rho_{\beta} \beta_2)^2 + (1 - \rho_{\beta}^2) \beta_2^2]^{-\frac{1}{2}} \\ &\leq (1 - \rho_{\beta}^2)^{-\frac{1}{2}} |\beta_2|^{-1} \\ &\rightarrow 0, \end{aligned} \quad (\text{A.15})$$

as β_1 and β_2 go to infinity. Let $f(\beta_1, \beta_2, \rho_{\beta})$ be $e^{\frac{\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2}{2(1 - \rho_{\beta}^2)}} \operatorname{erfc} \left[\sqrt{\frac{\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2}{2(1 - \rho_{\beta}^2)}} \right]$,

$$\begin{aligned} f(\beta_1, \beta_2, \rho_{\beta}) &\leq \frac{1}{2} e^{\frac{(\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2)}{2(1 - \rho_{\beta}^2)}} \left[e^{-\frac{(\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2)}{(1 - \rho_{\beta}^2)}} + e^{-\frac{(\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2)}{2(1 - \rho_{\beta}^2)}} \right] \\ &= \frac{1}{2} \left[e^{-\frac{(\beta_1^2 - 2\rho_{\beta} \beta_1 \beta_2 + \beta_2^2)}{2(1 - \rho_{\beta}^2)}} + 1 \right] \\ &\leq \frac{1}{2} \left[e^{-\frac{\beta_1^2 + \beta_2^2}{2(1 - \rho_{\beta}^2)}} + 1 \right] \\ &\rightarrow 0, \end{aligned} \quad (\text{A.16})$$

as β_1 and β_2 go to infinity. By (A.15) and (A.16), we obtain that

$$\lim_{\beta_1, \beta_2 \rightarrow \infty} P_1(\beta_1, \beta_2) = 0.$$

The independent density when β_1 and β_2 are independent also approaches to zero with large signals. Since

$$\frac{K}{4} \log \left(1 + \frac{4}{\beta_1^2} \right) \log \left(1 + \frac{4}{\beta_2^2} \right) < P_0(\beta_1, \beta_2) < K \log \left(1 + \frac{2}{\beta_1^2} \right) \log \left(1 + \frac{2}{\beta_2^2} \right),$$

where $K = (2\pi^3)^{-1}$, it is trivial that $\lim_{\beta_1, \beta_2 \rightarrow \infty} P_0(\beta_1, \beta_2) = 0$.

Follow the same procedure in the previous section, we compare the joint and independent densities with independent β . We show that: (a) the former term goes to infinity, and (b) the latter term goes to zero when β_1 and β_2 approach infinity. To show (a), we find a lower bound which goes to infinity as β_1 and β_2 approach infinity.

$$\begin{aligned}
\frac{e^{-\frac{1}{2}(\beta_1^2 + \beta_2^2)}}{(\beta_1^2 + \beta_2^2)^{\frac{1}{2}} E_1\left(\frac{\beta_1^2}{2}\right) E_1\left(\frac{\beta_2^2}{2}\right)} &> \frac{e^{-\frac{1}{2}(\beta_1^2 + \beta_2^2)}}{(\beta_1^2 + \beta_2^2)^{\frac{1}{2}} e^{-\frac{1}{2}(\beta_1^2 + \beta_2^2)} \log\left(1 + \frac{2}{\beta_1^2}\right) \log\left(1 + \frac{2}{\beta_2^2}\right)} \\
&> \frac{1}{\left[(\beta_1 + \beta_2)^2\right]^{\frac{1}{2}} \log\left(1 + \frac{2}{\beta_1^2}\right) \log\left(1 + \frac{2}{\beta_2^2}\right)} \text{ for } \beta_1, \beta_2 > 0 \\
&= \frac{\frac{1}{\beta_1 + \beta_2}}{\log\left(1 + \frac{2}{\beta_1^2}\right) \log\left(1 + \frac{2}{\beta_2^2}\right)}.
\end{aligned}$$

We use L'Hospital's rule since the numerator and the denominator both go to zero when β_1 and β_2 approach infinity.

$$\begin{aligned}
\frac{\partial^2}{\partial \beta_1 \partial \beta_2} \frac{1}{\beta_1 + \beta_2} &= -\frac{\partial}{\partial \beta_2} \frac{1}{(\beta_1 + \beta_2)^2} \\
&= \frac{2}{(\beta_1 + \beta_2)^3}, \text{ and}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial^2}{\partial \beta_1 \partial \beta_2} \log\left(1 + \frac{2}{\beta_1^2}\right) \log\left(1 + \frac{2}{\beta_2^2}\right) &= \left[\frac{\partial}{\partial \beta_1} \log\left(1 + \frac{2}{\beta_1^2}\right) \right] \left[\frac{\partial}{\partial \beta_2} \log\left(1 + \frac{2}{\beta_2^2}\right) \right] \\
&= \frac{1}{1 + \frac{2}{\beta_1^2}} (-4\beta_1^{-3}) \cdot \frac{1}{1 + \frac{2}{\beta_2^2}} (-4\beta_2^{-3}) \\
&= 16 \beta_1^{-3} \beta_2^{-3} \frac{\beta_1^2 \beta_2^2}{(\beta_1^2 + 2)(\beta_2^2 + 2)} \\
&= \frac{16}{\beta_1 \beta_2 (\beta_1^2 + 2)(\beta_2^2 + 2)}.
\end{aligned}$$

Then,

$$\begin{aligned}
\lim_{\beta_1, \beta_2 \rightarrow \infty} \frac{\frac{1}{\beta_1 + \beta_2}}{\log\left(1 + \frac{2}{\beta_1^2}\right) \log\left(1 + \frac{2}{\beta_2^2}\right)} &= \lim_{\beta_1, \beta_2 \rightarrow \infty} \frac{\frac{2}{(\beta_1 + \beta_2)^3}}{\frac{16}{\beta_1 \beta_2 (\beta_1^2 + 2)(\beta_2^2 + 2)}} \\
&= \frac{1}{8} \lim_{\beta_1, \beta_2 \rightarrow \infty} \frac{\beta_1 \beta_2 (\beta_1^2 + 2)(\beta_2^2 + 2)}{(\beta_1 + \beta_2)^3} \\
&= \infty.
\end{aligned} \tag{A.17}$$

According to (A.12),

$$\begin{aligned}
\frac{\operatorname{erfc}\left(\sqrt{\frac{\beta_1 + \beta_2}{2}}\right)}{E_1\left(\frac{\beta_1^2}{2}\right) E_1\left(\frac{\beta_2^2}{2}\right)} &< \frac{2 \left[e^{-\frac{1}{2}(\beta_1^2 + \beta_2^2) + 1} \right]}{\log\left(1 + \frac{4}{\beta_1^2}\right) \log\left(1 + \frac{4}{\beta_2^2}\right)} \\
&\rightarrow 0,
\end{aligned}$$

as β_1 and β_2 go to infinity because the exponential function converges to zero faster than the logarithmic function. We can also apply L'Hospital's rule to obtain the limit.

Thus,

$$\begin{aligned}
\lim_{\beta_1, \beta_2 \rightarrow \infty} R &= 2^{\frac{1}{2}} \pi^{\frac{3}{2}} \lim_{\beta_1, \beta_2 \rightarrow 0} e^{-\frac{1}{2}(\beta_1^2 + \beta_2^2)} \left(\frac{\beta_1^2 + \beta_2^2}{2} \right)^{-\frac{1}{2}} \left[E_1\left(\frac{\beta_1^2}{2}\right) E_1\left(\frac{\beta_2^2}{2}\right) \right]^{-1} \\
&\quad - \pi^2 \lim_{\beta_1, \beta_2 \rightarrow \infty} \operatorname{erfc}\left(\sqrt{\frac{\beta_1^2 + \beta_2^2}{2}}\right) \left[E_1\left(\frac{\beta_1^2}{2}\right) E_1\left(\frac{\beta_2^2}{2}\right) \right]^{-1} \\
&= \infty - 0 \\
&= \infty.
\end{aligned} \tag{A.18}$$

A.2 Computing details

We perform Markov Chain Monte Carlo (MCMC) sampling using R with the implementation of the Hamiltonian Monte Carlo (HMC) algorithm, Metropolis-Hastings sampling and Gibbs sampling.

A.2.1 Setup

The hierarchical model is

$$\begin{aligned} \mathbf{Y} \mid \boldsymbol{\beta} &\sim N(\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \\ \boldsymbol{\beta} \mid \boldsymbol{\Lambda} &\sim N(\mathbf{0}, \boldsymbol{\Lambda} \boldsymbol{\Sigma}_{\boldsymbol{\beta}} \boldsymbol{\Lambda}), \end{aligned} \tag{A.19}$$

where σ^2 is the error variance, $\mathbf{I}_n$ is the identity matrix of size n , $\boldsymbol{\Lambda}$ is diagonal with the diagonal elements $\lambda(\mathbf{s}_1), \dots, \lambda(\mathbf{s}_n)$, and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ determines spatial covariance structure. For the spatial covariance, we use conditional autoregressive (CAR; Carlin and Banerjee, 2003) structure with the first order neighbors for the gridded data in Sections 5 and 6. The CAR structure is commonly used because it has a simple form and the computation is fast. The CAR covariance is $\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = (\mathbf{M} - \rho_{\boldsymbol{\beta}} \mathbf{A})^{-1}$, where $\mathbf{M}$ is the diagonal matrix with the elements $m_1, \dots, m_n$ indicating the number of neighbors for locations $\mathbf{s}_1, \dots, \mathbf{s}_n$, $\rho_{\boldsymbol{\beta}}$ is the spatial dependence parameter, and $\mathbf{A}$ is the adjacency matrix with $\mathbf{A}_{ij} = 1$ if locations $\mathbf{s}_i$ and $\mathbf{s}_j$ are neighbors and $\mathbf{A}_{ij} = 0$ otherwise. Note that while $\rho_{\boldsymbol{\beta}}$ determines the degree of spatial dependence, it is not a correlation parameter as in Section 4.

The initial values $\boldsymbol{\beta}^{(0)} = [\beta(\mathbf{s}_1)^{(0)}, \dots, \beta(\mathbf{s}_n)^{(0)}]^T$ and $\boldsymbol{\lambda}^{(0)} = [\lambda(\mathbf{s}_1)^{(0)}, \dots, \lambda(\mathbf{s}_n)^{(0)}]^T$ are taken to be

$$\begin{aligned} \beta(\mathbf{s}_i)^{(0)} &= \frac{Y(\mathbf{s}_i)}{\max\{|Y(\mathbf{s}_i)|, i = 1, \dots, n\}} \\ \lambda_0^{(0)} &= \frac{1}{n} \sum_{i=1}^n |\beta(\mathbf{s}_i)^{(0)}| \\ \lambda(\mathbf{s}_i)^{(0)} &= \frac{|\beta(\mathbf{s}_i)^{(0)}|}{\lambda_0^{(0)}}. \end{aligned} \tag{A.20}$$

Then, we obtain the initial coefficients b_j by regressing $\lambda(\mathbf{s}_i)^{(0)}$ on $x_j(\mathbf{s}_i)$, the j^{th} basis expansion for location $\mathbf{s}$, for $j = 1, \dots, J$. Finally, we generate a random draw from the

prior distribution $\text{Beta}(10, 1)$ as the initial value of the parameter ρ_β .

A.2.2 Full conditional distributions

The unknown parameters are β , λ_0 , b_j and ρ_β . We specify an uninformative prior for the scale parameter, $\lambda_0^2 \sim \text{IG}(a, b)$, and a beta prior for the spatial dependence parameter, $\rho_\beta \sim \text{Beta}(10, 1)$. β and λ_0^2 have the conjugate full conditionals,

$$\begin{aligned} \beta \mid \text{rest} &\sim \text{N} \left\{ \left[\frac{1}{\sigma^2} \mathbf{I}_n + \Lambda^{-1}(\mathbf{M} - \rho_\beta \mathbf{A}) \Lambda^{-1} \right]^{-1} \mathbf{Y}, \left[\frac{1}{\sigma^2} \mathbf{I}_n + \Lambda^{-1}(\mathbf{M} - \rho_\beta \mathbf{A}) \Lambda^{-1} \right]^{-1} \right\} \\ \lambda_0^2 \mid \text{rest} &\sim \text{IG} \left[a + \frac{n}{2}, b + \frac{1}{2} \beta^T \Lambda^{-1}(\mathbf{M} - \rho_\beta \mathbf{A}) \Lambda^{-1} \beta \right]. \end{aligned} \quad (\text{A.21})$$

For small datasets, β could be updated using Gibbs sampling. However, for large datasets we update β using the HMC algorithm because it avoids cumbersome matrix manipulation, as described below. We update λ_0^2 using Gibbs sampling.

The full conditionals for b_j and ρ_β are not conjugate. The posterior distributions are proportional to

$$\begin{aligned} P(b_j \mid \text{rest}) &\propto e^{-\frac{1}{2} [b_j^2 + \beta^T \Lambda^{-1}(\mathbf{M} - \rho_\beta \mathbf{A}) \Lambda^{-1} \beta]} \prod_{i=1}^n \lambda(\mathbf{s}_i)^{-1} \\ P(\rho_\beta \mid \text{rest}) &\propto \rho_\beta^9 e^{-\frac{1}{2} \beta^T \Lambda^{-1}(\mathbf{M} - \rho_\beta \mathbf{A}) \Lambda^{-1} \beta} \prod_{i=1}^n (1 - \rho_\beta d_i), \end{aligned} \quad (\text{A.22})$$

where d_i is the i^{th} eigenvalue of $\mathbf{M}^{-\frac{1}{2}} \mathbf{A} \mathbf{M}^{-\frac{1}{2}}$. We can express the product

$$\beta^T \Lambda^{-1}(\mathbf{M} - \rho_\beta \mathbf{A}) \Lambda^{-1} \beta = \sum_{i=1}^n m_i \left[\frac{\beta(\mathbf{s}_i)}{\lambda(\mathbf{s}_i)} \right]^2 - 2 \rho_\beta \sum_{i \sim j} \frac{\beta(\mathbf{s}_i) \beta(\mathbf{s}_j)}{\lambda(\mathbf{s}_i) \lambda(\mathbf{s}_j)}, \quad (\text{A.23})$$

where $i \sim j$ means locations $\mathbf{s}_i$ and $\mathbf{s}_j$ are neighbors. We draw posterior samples of b_j using blocked Metropolis-Hastings algorithm. To facilitate computational efficiency, we update the spatial dependence parameter ρ_β using Gibbs sampling by grid search. Suppose $\rho_\beta \in \{0.001, 0.002, \dots, 0.999\}$, the prior density at each possible value is calculated before MCMC sampling so that we can avoid density computation during each MCMC iteration. Then, ρ_β is sampled from the set $\{0.001, 0.002, \dots, 0.999\}$ with the probability

proportional to the posterior likelihood in (A.22) given the updated parameters $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ in each iteration.

A.2.3 The Hamiltonian Monte Carlo algorithm

For large datasets we use HMC to update $\boldsymbol{\beta}$. HMC is also known as hybrid Monte Carlo because it incorporates MCMC and deterministic simulation models. This algorithm was first introduced by Duane et al. (1987) in physics and Neal (1994) in statistics. Neal (2011) provides a comprehensive introduction and detailed explanation of HMC algorithm. The HMC algorithm avoids the random walk behavior of Metropolis algorithm and allows the Markov chain to move through the parameter space more rapidly. Denote $\beta(\mathbf{s}_i)$ as β_i , then

$$-\log P(\boldsymbol{\beta} \mid \text{rest}) = c + \frac{1}{2} \left[\sum_{i=1}^n m_i \left(\frac{\beta_i}{\lambda_i} \right)^2 - 2 \rho_{\boldsymbol{\beta}} \sum_{i \sim j} \frac{\beta_i \beta_j}{\lambda_i \lambda_j} \right], \quad (\text{A.24})$$

where c is a constant that does not depend on $\boldsymbol{\beta}$, and the gradient is

$$-\frac{\partial \log P(\boldsymbol{\beta} \mid \text{rest})}{\partial \beta_i} = \frac{m_i \beta_i}{\lambda_i^2} - \frac{\rho_{\boldsymbol{\beta}}}{\lambda_i} \sum_{j \in \mathcal{N}_i} \frac{\beta_j}{\lambda_j}, \quad (\text{A.25})$$

where $\mathcal{N}_i$ is the set of neighbors of location $\mathbf{s}_i$.

Hamiltonian dynamics operates on an n -dimensional position vector $\boldsymbol{\beta}$ and an n -dimensional momentum vector $\mathbf{p}$. The system can be described in the Hamiltonian function,

$$H(\boldsymbol{\beta}, \mathbf{p}) = U(\boldsymbol{\beta}) + K(\mathbf{p}), \quad (\text{A.26})$$

where $U(\boldsymbol{\beta})$ is the potential energy, and $K(\mathbf{p})$ is the kinetic energy. In a statistical interpretation, $\boldsymbol{\beta}$ is a vector of interest, and $\mathbf{p}$ is an auxiliary vector facilitating the system. $U(\boldsymbol{\beta})$ is defined as the negative log posterior density in (A.24), and $K(\mathbf{p})$ is the negative log standard normal density, $N(\mathbf{0}, \mathbf{I}_n)$. Under this setting, the Hamiltonian equations of motion over time t can be expressed as

$$\frac{d\beta_i}{dt} = \frac{\partial H}{\partial p_i} = p_i, \quad \frac{dp_i}{dt} = -\frac{\partial H}{\partial \beta_i} = -\frac{\partial U}{\partial \beta_i}. \quad (\text{A.27})$$

The HMC algorithm proceeds in the following steps:

1. Initialize $\mathbf{p}$ from a random draw of the standard normal distribution, $N(\mathbf{0}, \mathbf{I}_n)$.
2. Obtain the candidate state $(\mathbf{p}^*, \boldsymbol{\beta}^*)$ from the leapfrog process, which requires the two parameters, step length ϵ and number of steps L . Repeat the following procedure L times:
 - (a) Make a half-step of $\mathbf{p}$ using the gradient of the log posterior density:

$$\mathbf{p} \leftarrow \mathbf{p} + \frac{1}{2}\epsilon \nabla \log[P(\boldsymbol{\beta} \mid \mathbf{Y})]. \quad (\text{A.28})$$

- (b) Update $\boldsymbol{\beta}$ using the momentum $\mathbf{p}$:

$$\boldsymbol{\beta} \leftarrow \boldsymbol{\beta} + \epsilon \mathbf{p}. \quad (\text{A.29})$$

- (c) Update the momentum $\mathbf{p}$ by a half-step again:

$$\mathbf{p} \leftarrow \mathbf{p} + \frac{1}{2}\epsilon \nabla \log[P(\boldsymbol{\beta} \mid \mathbf{Y})]. \quad (\text{A.30})$$

3. Let $(\mathbf{p}^{t-1}, \boldsymbol{\beta}^{t-1})$ be the state before the leapfrog process. Compute

$$r = -\log \left[\frac{P(\boldsymbol{\beta}^* \mid \mathbf{Y})P(\mathbf{p}^*)}{P(\boldsymbol{\beta}^{t-1} \mid \mathbf{Y})P(\mathbf{p}^{t-1})} \right]. \quad (\text{A.31})$$

4. Then,

$$\boldsymbol{\beta}^t = \begin{cases} \boldsymbol{\beta}^* & \text{if } \log(U) < r, \text{ where } U \sim \text{Uniform}(0, 1) \\ \boldsymbol{\beta}^{t-1} & \text{otherwise} \end{cases}. \quad (\text{A.32})$$

We tune ϵ and L together such that $\epsilon L = 1$ and stop tuning L until it is greater than 100 to give acceptance probability near 40%. We monitor convergence using trace plots of several representative parameters. We generate 15,000 samples and discard the first 5,000 as burn-in for simulated data in Section 2.5 and data analysis in Section 2.6.

A.3 Simulation plots

Figure A.1: Simulation results for Signal 1 with $\sigma = 0.5$ for the four models (rows): the first row shows the true signal $\beta_0(\mathbf{s})$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(\mathbf{s})$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models.

Figure A.2: Simulation results for Signal 1 with $\sigma = 1$ for the four models (rows): the first row shows the true signal $\beta_0(\mathbf{s})$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(\mathbf{s})$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models.

Figure A.3: Simulation results for Signal 1 with $\sigma = 2$ for the four models (rows): the first row shows the true signal $\beta_0(\mathbf{s})$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(\mathbf{s})$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models.

Figure A.4: Simulation results for Signal 2 with $\sigma = 1$ for the four models (rows): the first row shows the true signal $\beta_0(\mathbf{s})$ and results for the soft-thresholded Gaussian process (STGP) model, the second row shows the simulated data $Y(\mathbf{s})$ and results for the Gaussian model, the third and fourth rows show results for spatial horseshoe with quadratic (SHS_quad) and spline (SHS_B-spline 2) shrinkage models.