Spatial Models for Large Spatial and Spatiotemporal Data

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

Donald Musgrove

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

Advised by Lynn E. Eberly and John Hughes

April, 2016

Acknowledgments

I am very much indebted to my advisors, Lynn Eberly and John Hughes, for guiding me from start to finish through my dissertation. Special thanks to my committee, Jim Hodges and Galin Jones, for their helpful suggestions. I owe Galin Jones additional appreciation for inviting me to co-author a book chapter featuring work from this dissertation. I also need to thank Kathryn Cullen and Ted Christ for providing crucial financial support when I was *in-between* fellowships. Speaking of fellowships, I am thankful to the Office for Diversity in Graduate Education and the University of Minnesota Informatics Institute for providing fellowships that greatly enhanced my research experience during my time in the division of Biostatistics. Finally, I would like to thank Kevyn Funk, my steadfast partner who left the warm climes of Nevada to join me on my journey to the tundra of the upper Midwest. Kevyn's loving support and kind patience were essential for my successful completion of the PhD.

Dedication

To the everlasting memory of my mother.

Abstract

Regression analysis for areal data is common in numerous fields, including public health, ecology, and econometrics. Often, the goal of such an analysis is to quantify relationships between an outcome or outcomes of interest and covariates. In our present work, we propose several approaches to modeling areal data in areas including neuroimaging, cancer epidemiology, and demography. Much of our work is driven by the need to efficiently model large datasets with spatial dependencies. For instance, we model functional magnetic resonance imaging (fMRI) data using spatial Bayesian variable selection for detecting task-based brain activity. Fitting a full statistical model to large fMRI datasets can be computationally burdensome, so we efficiently implement a full statistical model via a brain parcellation that allows for parallel computations. Brain activity within each parcel is modeled as regressions located on a lattice with latent indicators for regressors, representing zero or non-zero activity levels. A sparse areal mixed model (SAMM) captures spatial dependence among indicator variables for a given stimulus. The SAMM permits more efficient computation than traditional spatial models typically employed. Through simulation we show that our parcellation scheme performs well in various scenarios. We apply our method to data from a task-based fMRI experiment.

Large multivariate and zero-inflated spatial data motivates another area where we develop a computationally efficient model. Our proposed multivariate model employs the SAMM to effectively reduce the dimensionality of the spatial effects and alleviate bias and variance inflation that affects estimation involving fixed effects covariates. We then adapt the multivariate model to handle zero-inflated count data via the truncated Poisson hurdle model. We carry out several simulation studies to evaluate the importance of modeling the correlation between the spatial effects of the different outcomes. We apply our new method on a high-dimensional zero-inflated dataset collected as part

of the 2010 US Census.

Another motivation for our work is the conditional autoregressive (CAR) model and its numerous drawbacks. We develop a covariance selection model that has a marginal interpretation that eases the restriction of interpreting covariate effects conditionally. Additionally, though the proper CAR model's dependence parameter has an intuitive conditional interpretation, the marginal interpretation is complicated and counterintuitive. To overcome these drawbacks, we introduce a copula-based covariance selection model. We achieve unbiased estimation of marginal parameters with an intuitive marginal interpretation. The covariance selection copula's single dependence parameter is the first-order correlation which provides a dependence structure with intuitive marginal interpretations. We develop a computational framework that permits efficient frequentist inference. We evaluate small and large sample performance of our method under simulated conditions and apply our procedure to the widely studied Slovenia stomach cancer dataset.

Contents

List of Tables

List of Figures

Chapter 1

Background

1.1 Overview

Modern computing capabilities have led to a revolution in the collection and storage of large, geographically referenced datasets. Ecologists, epidemiologists, and even neuroimagers routinely collect spatially referenced data where not only does properly modeling the spatial correlation present a challenge, but the size of the dataset represents a difficulty as well. Thus, an obstacle for researchers is implementing statistically principled methods within software that can handle large spatial datasets.

In this dissertation, we are interested in modeling spatial and spatiotemporal dependence in large areal datasets. Areal data are data that have been aggregated over a region of interest, e.g., the number of cancer cases in an administrative area or the neuroactivity in a small region of the brain. As we will see, the most commonly used models for areal data suffer several drawbacks. Chief among these drawbacks is the computational complexity that leads to long (hence expensive) computing times. Thus, a recurring theme of this dissertation is reduction of the computational complexity associated with handling large areal data.

1.2 Areal data modeling: Accounting for spatial correlation

The work presented in this dissertation is largely concerned with the problem of spatial regression in an areal data setting where both the response variable of interest and the fixed effects (predictors of interest) are observed at the level of the areal unit (regions of interest). We desire to make inference on the predictors while appropriately accounting for correlation among the areal units.

Areal data involves spatially aggregated quantities for each areal unit within a partition of a region, e.g., census blocks within a state or municipalities within a country. The spatial nature of areal data arises from the adjacency among areal units, i.e., two areal units that share a border are more likely to be similar than two areal units that do not share a border. The spatial adjacency of areal data is typically described via a graph $G = (V, E)$, where $V = 1, 2, \ldots, n$ are the vertices and E is the set of edges. Each vertex of G corresponds to an areal unit and each edge of G represents the spatial adjacency of two areal units. In this dissertation, G is assumed to be undirected and free of loops and parallel edges.

For spatial regression with areal data, two broad approaches are commonly used, conditional models, e.g., generalized linear mixed models (GLMMs), or marginal models, e.g., generalized estimating equations and copulas. In the case of GLMMs, a typical approach is to add random spatial effects to the regression and place a prior distribution on them. Common prior distributions are in the family of conditional autoregressive (CAR) priors, all of which work with a precision matrix, i.e., the inverse of the covariance matrix among the areal units.

Two common CAR priors are the intrinsic and proper CAR models. Though these priors are still commonly used, they have two drawbacks. First, CAR priors may lead to a problem known as spatial confounding, a complication we will return to throughout this dissertation. Spatial confounding arises when the fixed effects are collinear with the spatial effects resulting in bias and variance inflation. We do not want the introduction of spatial effects to dramatically affect the fixed effects estimates. With spatial confounding, the fixed effects may have reversed signs leading to completely different inference (Hodges and Reich, 2010). In Chapters 2 and 4, we use a reparameterized

CAR prior, the sparse areal mixed model (SAMM), to alleviate bias and variance inflation arising from the confounding. The SAMM uses spatial effects that are forced to be orthogonal to the fixed effects, which greatly alleviates spatial confounding (Hughes and Haran, 2013).

The second problem with CAR priors is the increased computational complexity due in part to the large parameter space. CAR priors introduce a spatial effect for each observation, a potentially prohibitive problem when the number of areal units is large. The SAMM that we employ in Chapters 2 and 4 can effectively lead to a dramatic reduction of the parameter space. Reduction of the parameter space improves the computing speed of the model. These ideas, among others, are central to the contributions of this dissertation.

The marginal model approach to spatial regression is less common but has seen recent growth in popularity, due in part to copulas (Kazianka and Pilz, 2010; Masarotto and Varin, 2012). Copulas offer a simple approach to accounting for dependence with spatially correlated data. In Chapter 3, we propose a Gaussian copula model that results in unbiased inference on the regression coefficients. Estimation with the copula model is carried out in part by using a covariance selection model implemented via iterative proportional scaling (IPS) (Xu et al., 2015). IPS is a covariance estimation scheme based on Gaussian graphical models. Since Gaussian copulas are scale-free, we work with the correlation matrix, rather than the covariance matrix. Thus, we implement a modification of IPS to estimate correlation matrices based on a single unknown parameter, the first-order correlation. The IPS modification allows quantification of a measure of the global level of spatial dependence present in the data. The covariance selection approach is in contrast to CAR priors which result in first-order correlations that vary in a non-intuitive manner (Wall, 2004). Our copula approach scales well computationally and we develop a framework for handling large datasets.

1.3   Areal data modeling: Multivariate response variables

Another area where we seek to offer improvements is the analysis of multivariate spatial data. Here, we take 'multivariate' data to imply data where two or more observations are collected on a single subject, or, in the case of our work, on a single areal unit,

where the correlation between observations does not have a clear structure. In particular, we want to account for two levels of correlation: within-areal unit unstructured correlation and between-areal unit spatial correlation. For example, interest may lie in simultaneously modeling the association between some covariates and five human papillomavirus-related cancer incidence rates in the counties of the United States. In this case, an unstructured correlation structure is likely the best approach since it is not clear whether a simpler correlation structure would be appropriate. On the other hand, county-level unemployment rates collected weekly over several years would clearly have a temporal correlation structure which does not fit into our usage of 'multivariate.'

Multivariate CAR models for areal data have seen use in the literature but these models certainly suffer from spatial confounding and large computational complexity, as discussed previously. Thus, in Chapter 4 we propose approaches based on the SAMM which simultaneously decrease parameter bias and variance inflation due to spatial confounding and dramatically reduce the parameter space leading to a substantially smaller computational complexity.

Lastly, we develop a zero-inflated model for areal count data as a special case of our proposed multivariate SAMM. Our zero-inflated model can be understood as a bivariate model where one model component is composed of counts while the second component is a binary indicator for a non-zero count. As with multivariate data, previous approaches have relied on CAR priors (Neelon et al., 2013). We employ the SAMM which, as expected, substantially reduces the computational complexity of the model leading to faster computing times. Further, the multivariate SAMM offers the same benefits as the univariate SAMM in that bias and variance inflation of the fixed effects is alleviated.

1.4   Dissertation objectives

The rest of the dissertation is laid out as follows. In Chapter 2, we develop a spatial Bayesian variable selection model for high-dimensional neuroimaging data based in part on the SAMM. We employ a fully Bayesian framework for applying the SAMM to model data that is not only spatially dependent, but temporally correlated as well. The SAMM is applied to a latent binary space to carry out spatial Bayesian variable selection. Inference is made by estimating the posterior probabilities that the latent

variables are non-zero. We achieve a computationally efficient algorithm in part by using a partitioning algorithm to split the dataset into independent 'parcels' which can be analyzed simultaneously in parallel. We perform several simulation studies to assess the efficacy of the model. We first examine the effect of the latent spatial field on inference about the posterior probabilities, including accuracy and false positive rate. We then assess the effect of our partitioning algorithm on inference. We apply the model to a single-subject data set from a study on depression in adolescents.

In Chapter 3, we use a covariance selection approach to estimate the covariance matrix in a spatial regression setting for count data. In contrast to Chapter 2, an extra level of difficulty arises in the model due to the variance being a function of the mean, as is typically seen with models for count data. Our proposed approach competes with CAR models and offers appealing theoretical properties. For example, the marginal correlations of the spatial effects implied by the CAR model have non-intuitive and bizarre properties. One such property is that the first-order correlations among areal units are typically non-constant, that is, the correlations vary in a complex way. Our approach uses a covariance selection framework, implemented via a copula, to force the first-order correlations to be constant. We achieve this result by using IPS to estimate the correlation matrix. Our proposed approach is novel because there is a single unknown variable associated with the correlation matrix: the first-order correlation. We carry out several simulation studies to assess the efficacy of the model. In our simulation scenarios, data are simulated on a large grid of $1,600$ areal units as well as a map of the 194 municipalities of Slovenia. We apply our model to the now classic Slovenia stomach cancer data. We achieve results similar to those reported in the literature, but now we have sensible interpretation of the spatial parameter associated with our model.

Chapter 3 may appear disjoint from the rest of the dissertation because it is not built on the SAMM but, similar to our extension of the SAMM in Chapter 2, the work contained in Chapter 3 solves a problem associated with previous modeling approaches. Additionally, this chapter happens to contain the only work that presents a non-Bayesian model. The framework presented can easily be cast in a Bayesian light. If we are uncertain *a priori* about the marginal correlations among areal units, then our covariance selection method can be viewed as a prior distribution that forces similar first-order correlations in the posterior.

In Chapter 4, we extend the basic SAMM presented in the second chapter to a multivariate setting where the data consist of two or more observations collected at each areal unit. We carry out a simulation study to assess the necessity and effectiveness of using a multivariate approach rather than several univariate models. We re-frame a zero-inflated model for count data as a special case of the proposed multivariate model. We propose a zero-inflated model and carry out a simulation study to assess the effect of varying within-areal unit correlation levels on inference of fixed effects of interest. We then apply our model to an interesting dataset from the 2010 US Census. The Census data consist of 2,600 deleted address counts with 75% zeros collected at the block-group level from the state of Iowa.

Finally, we conclude the dissertation with a discussion in Chapter 5. We also give additional computing details as well as a link to the Github repository that includes computing code for each of the models presented in this work.

Chapter 2

Bayesian Spatiotemporal Inference for fMRI Data

2.1 Introduction

This chapter develops a spatial Bayesian variable selection (SBVS) procedure for detecting blood oxygenation level dependent (BOLD) activation in functional magnetic resonance imaging (fMRI) experiments. Functional magnetic resonance imaging is an established method for detecting region-specific activation in the human brain that occurs in response to tasks or stimuli (Bandettini, 2009). During a single-subject fMRI experiment, a participant performs a task in response to a stimulus while three-dimensional images of the subject's brain are collected every 2–3 seconds. The images capture the BOLD signal contrast, allowing us to indirectly observe neuronal activity. The presence (or absence) of neuronal activity in response to stimuli is used to describe brain function networks, to assess brain development, or to assess impairment related to conditions including alcoholism and brain trauma.

Single-subject fMRI data have a complex structure. Each collected image comprises a three-dimensional grid, or lattice, of volume elements (voxels). The BOLD response is observed over time at each voxel, resulting in a large quantity of data (in excess of 20 million observations per person) that exhibit both spatial and temporal dependence. Regression models are used to analyze fMRI data such that the per-voxel regression coefficients quantify the brain activity, relative to a rest condition, in that voxel during

the task; a voxel is considered inactive if the magnitude of the BOLD contrast is not significantly different from zero. The standard non-Bayesian approach to computing estimates of the activation magnitude for a single brain follow, e.g., Friston et al. (1995) and Worsley (2003), which do not incorporate spatial modeling. Several Bayesian approaches have appeared in the neuroimaging literature (Woolrich et al., 2004; Bowman, 2007; Smith and Fahrmeir, 2007). In particular, Smith and Fahrmeir (2007) proposed an SBVS approach in which the regressions are located on a voxel-based lattice, and latent binary indicators represent whether each voxel on the lattice is active or inactive.

The SBVS approach considers voxels as having spatially linked regressions. By borrowing strength across neighboring voxels, this approach results in more reliable inference for voxels where there is little information, i.e., small signal-to-noise ratio (SNR). The primary drawback of the above mentioned SBVS approaches is that temporal correlation within each voxel's time series is ignored for the sake of computational efficiency (Smith et al., 2003; Smith and Fahrmeir, 2007). Our model accounts for both spatial and temporal dependencies via SBVS and an autoregressive (AR) model, respectively. Another innovation of our proposed method is use of the sparse areal mixed model (SAMM) of Hughes and Haran (2013) to model latent indicator variables. The SAMM allows for faster computations than the more commonly used Ising model (Smith and Fahrmeir, 2007).

This work also uses a brain parcellation. A typical adult brain has more than 50,000 voxels, whereas our parcellation method results in parcels potentially composed of far fewer voxels. The parcellation speeds computation in two ways: (1) parcel-level analyses can be done in parallel, and (2) the 'size' of the required spatial model is reduced. Thus, the use of the SAMM along with a parcellation allows for rapid simulation of posterior samples. The computational efficiency of our approach allows for more statistically principled modeling decisions, i.e., modeling need not be driven by computational considerations. The result is a fully Bayesian model than can be applied efficiently.

We show that our method is insensitive to model misspecification and suffers minimal edge effects (edge effects could arise because voxels on the edge of a parcel may be spatially correlated with voxels on the edges of neighboring parcels). Using simulated data, we demonstrate that the parcellation-based method is able to achieve a high

accuracy rate in predicting voxel- and task-specific activations. A full-brain dataset is used to illustrate the applicability of the method. Maps of estimated activation probabilities are produced for these data.

The rest of the chapter is organized as follows. In the next section, our model is outlined and previous methods are described. In Section 2.3, the innovations introduced in our approach are described in detail. In Section 2.4, a simulation study is carried out. In Section 2.5, our model is applied to a single-subject dataset from a study on depression in adolescents (Musgrove et al., 2015). The chapter ends in Section 2.6 with a discussion. Additional supporting material and details including our Markov chain Monte Carlo (MCMC) sampling scheme, description of the parcellation method, and extended application results are given in Appendix A.

2.2 Current approaches for the statistical analysis of fMRI data

For the vth voxel, the voxel time series $\mathbf{y}_v$, $v = 1, \ldots, N$, comprises the BOLD response over T time points. For each such time series we have a linear model:

$$\mathbf{y}_v = \mathbf{Z}\boldsymbol{\alpha}_v + \mathbf{X}\boldsymbol{\beta}_v + \mathbf{W}_v\boldsymbol{\rho}_v + \boldsymbol{\varepsilon}_v. \tag{2.1}$$

The baseline trend matrix $\mathbf{Z}$, $T \times a$, is modeled as a fixed parameter or as a set of polynomial or trigonometric functions (Smith and Fahrmeir, 2007) with unknown effects $\boldsymbol{\alpha}_v$. The activation profile is captured via the p-vector $\boldsymbol{\beta}_v$, which is the magnitude of the BOLD response to the p stimuli. The design matrix $\mathbf{X}$, $T \times p$, is composed of columns that are a function of the task stimuli corresponding to $\boldsymbol{\beta}_v$. Temporal correlation is accounted for by $\boldsymbol{\rho}_v$, an r-vector of autoregressive coefficients, with $\mathbf{W}_v$ the $T \times r$ matrix of lagged prediction errors, while $\boldsymbol{\varepsilon}_v \sim \mathcal{N}(\mathbf{0}, \sigma_v^2\mathbf{I})$. Specification of $\mathbf{W}_v$ follows Penny et al. (2003) such that each row has the form

$$\mathbf{w}_{vt} = \left\{ \left(y_{v,t-1} - \mathbf{z}'_{t-1}\boldsymbol{\alpha}_v - \mathbf{x}'_{t-1}\boldsymbol{\beta}_v \right), \ldots, \left(y_{v,t-r} - \mathbf{z}'_{t-r}\boldsymbol{\alpha}_v - \mathbf{x}'_{t-r}\boldsymbol{\beta}_v \right) \right\}'.$$

If the lag order is greater than the indexed time, i.e., if $l \geq t$ for $l \in \{1, \ldots, r\}$, we follow Penny et al. (2003) and set $w_{vtl} = 0$, where l denotes the lth element of $\mathbf{w}_{vt}$. Spatial dependence between voxels is induced via latent indicator variables $\boldsymbol{\gamma}_v = (\gamma_{v1}, \ldots, \gamma_{vp})'$

such that $\beta_{vj} = 0$ if $\gamma_{vj} = 0$ and $\beta_{vj} \neq 0$ if $\gamma_{vj} = 1$, where j indexes the stimulus effects in $\boldsymbol{\beta}_v$. Estimation of $\boldsymbol{\beta}_v$ and $\boldsymbol{\gamma}_v$ is of primary interest.

SBVS methods were first applied in fMRI by Smith et al. (2003), and more recently in Smith and Fahrmeir (2007). In the latter paper, Smith and Fahrmeir considered a full-brain approach that introduced spatial dependence via an Ising prior for the latent binary indicator variables. The Ising model (Ising, 1925) is a binary Markov random field (MRF) model (Kindermann and Snell, 1980) that was first employed in the study of ferromagnetism. The Ising prior is appealing from a modeling point of view but has an intractable normalizing function. The intractability can be addressed using perfect sampling (Møller, 1999; Møller et al., 2006) or one of various approximations (Gelman and Meng, 1998; Wang and Landau, 2001; Zhang and Ma, 2007). Perfect sampling leads to burdensome computation while approximations do not guarantee that the Markov chain has the desired stationary distribution.

The prior on the regression coefficients in Smith and Fahrmeir (2007) is based on Zellner's g-prior (Smith and Kohn, 1996), which has the form

$$\boldsymbol{\beta}_v \mid \mathbf{y}_v, \boldsymbol{\gamma}_v, \sigma_v^2 \sim \mathcal{N}\left\{\hat{\boldsymbol{\beta}}_v,\, T\sigma_v^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right\}, \tag{2.2}$$

where $\hat{\boldsymbol{\beta}}_v = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}_v$ and the length of the voxel time series T is the 'g' of Zellner's g-prior. Again, we let $\boldsymbol{\gamma}_v = (\gamma_{v1}, \ldots, \gamma_{vp})'$ denote the vector of indicator variables for the vth voxel. For $\boldsymbol{\gamma}_j = (\gamma_{1j}, \ldots, \gamma_{Nj})'$, the vector of binary indicator variables over all voxels for the jth stimulus effect, the Ising prior is given by

$$p\left(\boldsymbol{\gamma}_j \mid \vartheta_j\right) \propto \exp\left[\sum_{v=1}^{N}\left\{\xi_{vj}(\gamma_{vj}) + \sum_{u\sim v}\vartheta_j\omega_{uv}I(\gamma_{vj} = \gamma_{uj})\right\}\right],$$

where $u \sim v$ denotes neighboring voxels, $I(\cdot)$ denotes the indicator function, ξ_{vj} is user-specified and represents prior knowledge of activation locations, ϑ_j is the interaction effect between neighboring voxels, and ω_{uv} is a pre-specified weight. For $\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_N)$, the marginal posterior distribution $p\left(\boldsymbol{\gamma}_j \mid \mathbf{Y}\right)$ is sampled using MCMC, and the probabilities of activation are estimated as the sample means of the indicator variables. Specifically, for K samples from the posterior distribution of $\boldsymbol{\gamma}_j$, the estimated probability of activation is

$$p(\gamma_{vj} = 1 \mid \mathbf{Y}) \approx \frac{1}{K}\sum_{k=1}^{K} p(\gamma_{vj} = 1 \mid \boldsymbol{\gamma}_{\backslash vj}^{[k]}, \mathbf{Y}),$$

where $\boldsymbol{\gamma}_{\backslash vj}$ denotes the full indicator variable $\boldsymbol{\gamma}_j$ with γ_{vj} omitted. The regression coefficients are estimated, using Rao-Blackwellization (Smith and Fahrmeir, 2007), as

$$\mathbb{E}(\boldsymbol{\beta}_v \mid \mathbf{Y}) \approx \frac{1}{K} \sum_{k=1}^{K} \mathbb{E}(\boldsymbol{\beta}_v \mid \boldsymbol{\gamma}_v^{[k]}, \mathbf{y}_v) = \frac{1}{K} \sum_{k=1}^{K} \hat{\boldsymbol{\beta}}_v(\boldsymbol{\gamma}_v^{[k]}).$$

The data-based g-prior of (2.2) was chosen for convenience, as it allows the regression coefficients to be integrated out analytically, with the goal of speeding convergence. Further, magnitudes of the stimulus effects (represented by the regression coefficients) are estimated by least squares. The method is purely spatial and does not account for temporal dependence in the voxel time series.

Noting that the lack of temporal modeling leads to lower quality inference, Lee et al. (2014) developed a spatiotemporal model using the Ising prior as in Smith and Fahrmeir (2007). Lee et al. (2014) captured the temporal dependence via a first-order autoregressive AR(1) model. When the regression coefficients are analytically integrated out of the likelihood, the resulting voxel time series covariance matrix includes an unknown AR(1) coefficient, requiring inversion of a dense $T \times T$ matrix during each iteration of the MCMC sampler. Inversion of a dense $T \times T$ matrix implies a sizable computational burden.

2.3 Extending current approaches

Our approach makes use of a parcellation that divides the brain into non-overlapping parcels. Within each parcel, we apply a spatial Bayesian variable selection method that also accounts for voxel-level temporal correlation. In the context of fMRI, our procedure is novel in that the SAMM is used to model the spatial dependence among the activation indicators. The parcellation and the SAMM together permit efficient sampling even though our model is fully Bayesian.

2.3.1 Partitioning the brain

We partition the brain into chunks to achieve speed gains over previous methods via analysis of parcels generated as follows. First, divide the brain into H_0 cubes, each of which has at most n_0 voxels. Since the brain is not rectangular, many of the parcels will

include fewer than n_0 voxels. With our analysis technique, we found that a minimum of 500 voxels per parcel results in excellent spatial inference. Next, parcels are iteratively resized to enforce a minimum parcel size of 500 voxels while ensuring that the underlying graph is connected. The iterative resizing step moves voxels from small parcels below the 500 voxel threshold to the nearest neighboring parcel. The final dataset for analysis comprises H parcels, with the hth parcel having $n_h \geq 500$ voxels. Computational gains are achieved since the parcels can be analyzed in parallel. The parcellation could potentially be based on an existing anatomical atlas, such as Brodmann areas (Tzourio-Mazoyer et al., 2002), but we found that such an approach leads to an excessive number of parcels with fewer than 500 voxels. Thus, we will apply our divide-and-conquer approach in both our simulation study and application to real data. Details of our parcellation algorithm are given in Appendix A.

2.3.2 Spatial Bayesian variable selection with temporal correlation

We develop a Bayesian variable selection model (George and McCulloch, 1997) for the regression coefficients of the voxel-level regression (2.1). We place a spike-and-slab prior on the regression coefficients $\boldsymbol{\beta}_v$ such that each β_{vj} is drawn from a diffuse normal distribution (the slab) or a point mass at zero (the spike). The spike-and-slab prior reflects the belief that a coefficient is nonzero or zero, respectively. Latent indicator variables $\boldsymbol{\gamma}_v = (\gamma_{v1}, \ldots, \gamma_{vp})'$ are introduced such that the prior for each β_{vj} has the form

$$p(\beta_{vj} \mid \gamma_{vj}) = \gamma_{vj} \mathcal{N}\left(0, \tau_j^2\right) + (1 - \gamma_{vj}) I_0,$$

where τ_j^2 is an unknown stimulus-level variance and I_0 denotes a point mass at zero. We make the natural assumption that the regression coefficients are *a priori* independent conditional on the indicator variables (George and McCulloch, 1997). We model the spatial dependence between voxels by placing a spatial prior on the indicator variables (discussed in Section 2.3.3, below).

The columns of the design matrix $\mathbf{X}$ contain the transformed stimulus function, a hemodynamic response function (HRF) convolved with task onset times and durations. The use of the HRF is motivated by the fact that the BOLD response is delayed in time after the underlying neuronal activation, and occurs continuously, rather than in a binary fashion, at the task onset times. Since the focus of this chapter is on our

spatial prior and our parcellation scheme, we use an 'off-the-shelf' HRF, namely, the gamma density. Other possibilities include the Poisson probability mass function, and the double-gamma canonical HRF (Lindquist et al., 2009).

To account for the serial correlation present in the univariate voxel time series, we employ an AR(2) model that is easy to implement and computationally efficient. The AR(2) model offers a trade-off between the computational complexity of a higher-order AR process and a simpler AR(1) process. Similar to Penny et al. (2003), we include a matrix of lagged prediction errors, denoted $\mathbf{W}_v$, in the regression model. We use a zero mean and vague normal prior on the AR(2) coefficients $\boldsymbol{\rho}_v = (\rho_{v1}, \rho_{v2})'$ such that $\boldsymbol{\rho}_v \sim \mathcal{N}\left(\mathbf{0}, \tau_\rho^2 \mathbf{I}_2\right)$. To complete the voxel-level prior specification, the prior on the error variance is $p\left(\sigma_v^2\right) \propto 1/\sigma_v^2$, and the prior on the variance of the non-zero regression coefficients is $p\left(\tau_j^2\right) \propto 1/\tau_j^2$. Finally, the prior on the baseline trend effects is $\boldsymbol{\alpha}_v \sim \mathcal{N}\left(\mathbf{0}, \tau_\alpha^2 \mathbf{I}_a\right)$. The hyperparameters τ_ρ^2 and τ_α^2 are user-specified. In our simulation studies and application, we set $\tau_\rho^2 = \tau_\alpha^2 = 10^6$.

2.3.3 Sparse areal mixed model prior

Here we describe our spatial prior on the voxel- and task-specific binary indicator variables γ_{vj}. For the hth parcel with $v = 1, \ldots, n_h$ voxels, we assume a sparse areal mixed model (SAMM) (Hughes and Haran, 2013) for each $\boldsymbol{\gamma}_j = (\gamma_{1j}, \ldots, \gamma_{n_h j})'$, $j = 1, \ldots, p$. Specifically, $\boldsymbol{\gamma}_j$ are conditionally independent Bernoulli random variables with a probit link function, i.e.,

$$\gamma_{vj} \mid \eta_{vj} \overset{\text{ind}}{\sim} \mathcal{B}ern\left\{\boldsymbol{\Phi}(\eta_{vj})\right\},$$

$$\eta_{vj} = \psi_{vj} + \mathbf{m}_v' \boldsymbol{\delta}_{sj} + \epsilon_{vj}, \tag{2.3}$$

$$\epsilon_{vj} \sim \mathcal{N}\left(0, 1\right),$$

where $\boldsymbol{\Phi}$ denotes the cumulative distribution function of a standard normal random variable, η_{vj} is an auxiliary variable introduced to facilitate Gibbs sampling (Holmes and Held, 2006), ψ_{vj} is fixed to reflect the prior probability of activation, $\mathbf{m}_v$ is a q-vector of latent spatial predictors, and $\boldsymbol{\delta}_{sj}$ is a q-vector of spatial random effects. The subscript 's' on $\boldsymbol{\delta}_{sj}$ denotes sparse since $\boldsymbol{\delta}_{sj}$ is modeled using the SAMM. Voxels are

located at the vertices of an underlying undirected graph, the structure of which reflects spatial adjacency among voxels. We represent the graph using its parcel-level adjacency matrix $\mathbf{A}$, which is $n_h \times n_h$ with entries given by $\text{diag}(\mathbf{A}) = \mathbf{0}$ and $(\mathbf{A})_{uv} = I(u \sim v)$. In a two-dimensional analysis, we let a voxel neighborhood comprise a voxel and its four adjacent neighbors. With three-dimensional fMRI data, a voxel and its 26 nearest neighbors comprise a neighborhood, i.e., a voxel is the center of a $3 \times 3 \times 3$ cube. The prior for the spatial random effects is based on the SAMM (Hughes and Haran, 2013) and, joint with the hyper-prior κ_j, has the form

$$p(\boldsymbol{\delta}_{sj} \mid \kappa_j)p(\kappa_j) = \mathcal{N}\left\{\boldsymbol{\delta}_{sj} \mid \mathbf{0}, (\kappa_j\mathbf{M}'\mathbf{Q}\mathbf{M})^{-1}\right\} \times \mathcal{G}amma\left(\kappa_j \mid a_\kappa, b_\kappa\right), \qquad (2.4)$$

where κ_j is a smoothing parameter, $\mathbf{M}$ is an $n_h \times q$ matrix, the columns of which are the q principal eigenvectors of $\mathbf{A}$, and $\mathbf{Q} = \text{diag}(\mathbf{A}\mathbf{1}) - \mathbf{A}$ is the graph Laplacian. The vector $\mathbf{m}'_v$ is the vth row of $\mathbf{M}$. The columns of $\mathbf{M}$ are multiresolutional spatial basis vectors that are well suited for spatial smoothing and capture both small-scale and large-scale spatial variation typically exhibited by fMRI data (Woolrich et al., 2004). Eigendecomposition of the adjacency matrix $\mathbf{A}$ represents a special form of the SAMM. As we will see in Chapter 4, the SAMM is typically applied when fixed effects are present in the model. In contrast, here, we are applying the SAMM to a latent binary variable that does not have fixed effects. Following Hughes and Haran (2013), we introduce sparsity by selecting a small subset of the columns of $\mathbf{M}$. Our selected subset of $\mathbf{M}$ permits spatial smoothing while reducing the dimensionality considerably (typically, $q < n_h/2$). Our choice of prior on the smoothing parameter κ_j follows Kelsall and Wakefield (1999) such that the shape $a_\kappa = 1/2$ and the scale $b_\kappa = 2000$, resulting in a large prior mean of the precision and a small prior variance of the spatial random effects.

2.3.4 Posterior computation and inference

Denote the voxel-level parameters as $\boldsymbol{\theta}_v = (\boldsymbol{\alpha}'_v, \boldsymbol{\beta}'_v, \boldsymbol{\rho}'_v, \sigma^2_v)'$, and, for the hth parcel, the parcel-level parameters are denoted as $\boldsymbol{\Theta}_h = (\boldsymbol{\gamma}'_h, \boldsymbol{\eta}'_h, \boldsymbol{\delta}'_{sh}, \boldsymbol{\kappa}'_h, (\boldsymbol{\tau}^2_h)')'$, where $\boldsymbol{\gamma}_h = \left(\boldsymbol{\gamma}'_{h1}, \dots, \boldsymbol{\gamma}'_{hp}\right)'$, $\boldsymbol{\eta}_h = \left(\boldsymbol{\eta}'_{h1}, \dots, \boldsymbol{\eta}'_{hp}\right)'$, each $\boldsymbol{\eta}_{hj} = (\eta_{1j}, \dots, \eta_{n_h j})'$, $\boldsymbol{\delta}_{sh} = \left(\boldsymbol{\delta}'_{sh1}, \dots, \boldsymbol{\delta}'_{shp}\right)'$, $\boldsymbol{\kappa}_h = (\kappa_{h1}, \dots, \kappa_{hp})'$, and $\boldsymbol{\tau}^2_h = \left(\tau^2_{h1}, \dots, \tau^2_{hp}\right)'$. For ease of exposition, we drop the 'h' subscript from the parcel-level parameters. Within the hth

parcel, with n_h voxels, we assume *a priori*

$$p(\boldsymbol{\theta}_v, \boldsymbol{\Theta}_h) = \left\{ \prod_{v=1}^{n_h} p(\boldsymbol{\alpha}_v) p(\boldsymbol{\rho}_v) p(\sigma_v^2) p(\boldsymbol{\beta}_v \mid \boldsymbol{\gamma}_v) \right\} \times$$

$$\left\{ \prod_{j=1}^{p} p(\boldsymbol{\gamma}_j \mid \boldsymbol{\eta}_j) p(\boldsymbol{\eta}_j \mid \boldsymbol{\delta}_{sj}) p(\boldsymbol{\delta}_{sj} \mid \kappa_j) p(\kappa_j) p(\tau_j^2) \right\},$$

where, again, $\boldsymbol{\gamma}_v = (\gamma_{v1}, \ldots, \gamma_{vp})'$. The full conditional distribution is obtained in the usual way by combining priors and the likelihood. To obtain updates for each γ_{vj}, we use a voxel-level likelihood where β_{vj} has been integrated out analytically. For $y_{vt}^* = y_{vt} - \mathbf{z}_t' \boldsymbol{\alpha}_v$, let $e_{vt(j)} = y_{vt}^* - \sum_{j' \neq j}^{p} x_{tj'} \beta_{vj'}$, $e_{vt(j)}^* = e_{vt(j)} - \sum_{l=1}^{2} \rho_{vl} e_{v,t-l(j)}$, and $x_{tj}^* = x_{tj} - \sum_{l=1}^{2} \rho_{vl} x_{t-l,j}$, where $\mathbf{z}_t$ is the tth row of $\mathbf{Z}$. Then, conditional on γ_{vj}, the likelihood can be written as a mixture with two components:

$$L_1 = \tau_j^{-1} \exp \left\{ -\frac{1}{2\sigma_v^2} \sum_{t=1}^{T} \left(e_{vt(j)}^* - x_{tj}^* \beta_{vj} \right)^2 - \frac{1}{2\tau_j^2} \beta_{vj}^2 \right\}$$

and

$$L_0 = \exp \left\{ -\frac{1}{2\sigma_v^2} \sum_{t=1}^{T} e_{vt(j)}^{*2} \right\},$$

where L_1 is the voxel-level likelihood when $\gamma_{vj} = 1$, and L_0 is the likelihood when $\gamma_{vj} = 0$. Integrating β_{vj} out of L_1, we have

$$L_1^* = \tau_j^{-1} \sigma_{vj}^{*-1} \exp \left\{ -\frac{1}{2\sigma_v^2} \sum_{t=1}^{T} e_{vt(j)}^{*2} + \frac{1}{2\sigma_{vj}^{*2}} \left(\frac{1}{\sigma_v^2} \sum_{t=1}^{T} e_{vt(j)}^* x_{tj}^* \right)^2 \right\},$$

where $\sigma_{vj}^{*2} = \sigma_v^{-2} \sum_{t=1}^{T} x_{tj}^{*2} + \tau_j^{-2}$. The conditional posterior probability that $\gamma_{vj} = 1$ is $(1 + \mathcal{P})^{-1}$, where

$$\mathcal{P} = \frac{L_0}{L_1^*} \frac{p\left(\gamma_{vj} = 0 \mid \boldsymbol{\eta}_j \right)}{p\left(\gamma_{vj} = 1 \mid \boldsymbol{\eta}_j \right)}.$$

Conditional on $\gamma_{vj} = 0$, set $\beta_{vj} = 0$. Otherwise, for $\gamma_{vj} = 1$, β_{vj} is updated from its full conditional distribution. We can write $e_{vt(j)}^* = x_{tj}^* \beta_{vj} + \varepsilon_{vt}$, and since the error term is normally distributed, and each β_{vj} has a normal prior distribution, conditional on $\gamma_{vj} = 1$, so the conditional posterior distribution of β_{vj} is given by

$p\left(\beta_{vj} \mid \mathbf{y}_v,\, \boldsymbol{\theta}_v \backslash \beta_{vj},\, \tau_j^2,\, \gamma_{vj}=1\right) \propto \mathcal{N}(\hat{\beta}_{vj},\, \hat{\tau}_{vj}^2),$ where $\hat{\tau}_{vj}^2 = \left(\sum_{t=1}^T x_{tj}^{*2} + \sigma_v^2/\tau_j^2\right)^{-1},$ $\hat{\beta}_{vj} = \hat{\tau}_{vj}^2 \sum_{t=1}^T e_{vt(j)}^* x_{tj}^*,$ and $\boldsymbol{\theta}_v \backslash \beta_{vj}$ denotes $\boldsymbol{\theta}_v$ with β_{vj} omitted.

Posterior sampling of each of $\boldsymbol{\eta}_j$, $\boldsymbol{\delta}_{sj}$, and κ_j uses probit regression with auxiliary variables, conditional on $\boldsymbol{\gamma}_j$ (Holmes and Held, 2006). The full conditional distributions of all remaining variables are derived in Appendix A.

To declare a voxel active or inactive, a threshold on the estimated posterior probability of activation must be chosen. We follow Smith and Fahrmeir (2007) and Lee et al. (2014) in choosing a threshold value of 0.8722. Lee et al. (2014) showed that a threshold of 0.8722 corresponds to a p-value of 0.05. Using a simulation study, we found that this threshold works well in practice (not shown).

2.4   Simulation studies

We carried out two simulation studies to assess the performance of our approach. First, we simulated data from our model under six scenarios to investigate the effects on inference of different values of the spatial smoothing parameter κ and the voxel-level error variance σ_v^2. Next, we investigated the parcellation method's effect on activation detection. The latter simulation study employs the `neuRosim` package (Welvaert et al., 2011) for R (Ihaka and Gentleman, 1996) to simulate two-dimensional datasets with spatiotemporal dependence.

2.4.1   Model performance for varying precisions

We applied our proposed method to data simulated from the prior distributions given in (2.3) and (2.4). For this simulation study, we have a single stimulus. Thus, the subscript 'j' is omitted from the variables in this section. Under investigation is the influence of the spatial smoothing parameter κ and the error variance σ_v^2. We choose $\kappa = 0.5$ or $\kappa = 2$. Simulated probabilities are shown in Figure 2.1 (panels a, e, and i). We chose values of 0.5 and 2 for the error variance.

We fix κ and simulate on the 25×25 square lattice, giving $n = 625$ voxels. On the two-dimensional square lattice, voxels were considered neighbors if they shared a face, thus the maximum number of neighbors per voxel is 4 voxels. The spatial random effects $\boldsymbol{\delta}_s$ are generated according to (2.4), using $q = 298$ eigenvectors of the adjacency matrix

$\kappa = 0.5$

(a) (b) (c) (d)

$\kappa = 1.0$

(e) (f) (g) (h)

$\kappa = 2.0$

(i) (j) (k) (l)
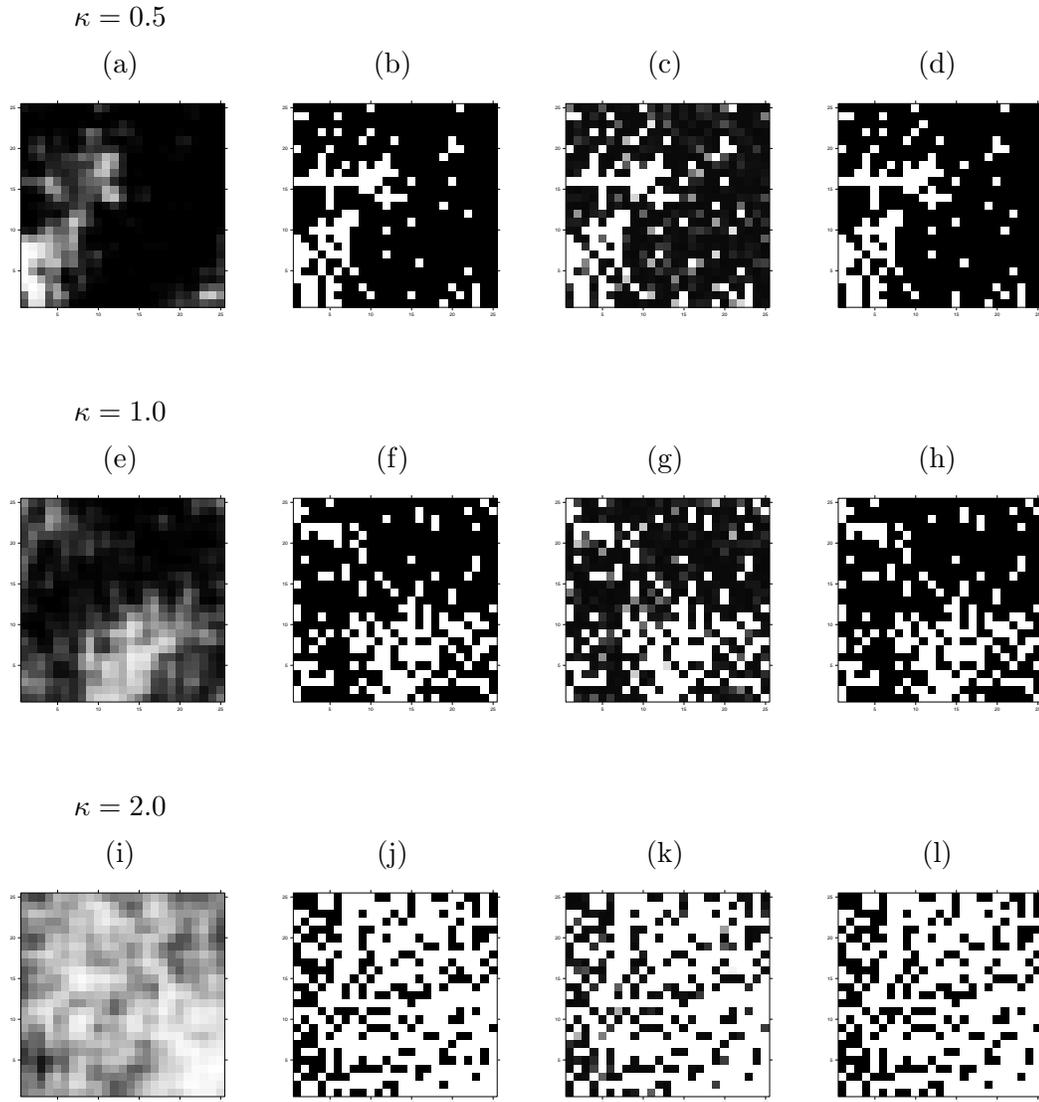
Figure 2.1: Representative simulated and estimated probability and binary maps. Probability maps are generated (first column, a, e, and i) and binary maps are simulated from corresponding Bernoulli distributions (second column, b, f, and j). The SAMM is applied to the simulations and probability maps are estimated (third column, c, g, and k), with binary maps created by thresholding the probability maps at 0.8722 (fourth column, d, h, and l).

A. The selected value of q is specific to the chosen parcel structure. The probability of activation is computed as $\boldsymbol{\Phi}(\eta_v) = \boldsymbol{\Phi}(\mathbf{m}_v'\boldsymbol{\delta}_s)$, where $\mathbf{m}_v'$ is the vth row of the matrix $\mathbf{M}$ of spatial basis vectors. That is, for purposes of the simulation study, the fixed prior probability, ψ_v, was set to zero. The binary indicator for each voxel was simulated as $\gamma_v \sim \mathcal{B}ern\{\boldsymbol{\Phi}(\eta_v)\}$. Simulated activation locations are shown in Figure 2.1 (panels b, f, and j). We used as our design matrix $\mathbf{X}$ a single stimulus time series of length $T = 195$ from our application to a block design study, as shown by the black curve in Figure 2.2. We set the baseline signal to be constant at 1200 for each time point, i.e., $\mathbf{Z}$ is a column vector of ones such that $z_t\alpha_v = 1200$, $t = 1, \ldots, 195$. Each β_v is set to zero or six depending on the simulated value of γ_v. The value of six was chosen to represent a very small percentage increase over baseline. We assume a second-order autoregressive structure with coefficients $\boldsymbol{\rho}_v = (0.4, -0.2)'$, and fix the error variance σ_v^2. The value of $\beta_v = 6$, along with the stimulus time series, corresponds to an average activation amplitude of approximately 6.3. Thus the SNRs for the simulation study are $6.3/\sqrt{0.5} = 8.9$ and $6.3/\sqrt{2} = 4.5$. The latter SNR is similar to that observed in our application to real data.

For each scenario, we simulated and analyzed 100 datasets. We used a fixed-width approach to convergence diagnostics (Flegal et al., 2008). We drew 200,000 samples for each dataset, ensuring that the Monte Carlo standard errors were sufficiently small. We define accuracy as the percentage of voxels correctly classified, i.e., $\frac{1}{625}\sum_{v=1}^{625} I(\hat{\gamma}_v = \gamma_v)$, and false positive rate as the percentage of false positives, i.e., $\frac{1}{625}\sum_{v=1}^{625} I(\hat{\gamma}_v = 1 \mid \gamma_v = 0)$, where $\hat{\gamma}_v$ is the estimate of γ_v. Using a threshold for the estimated posterior means of the γ_v of 0.8722, the accuracy and false positive rate are estimated for each simulated dataset, and the results are presented as medians across all 100 datasets.

The summaries are given in Table 2.1, scenarios 1 through 6. Our approach performs well across all values of the spatial smoothing parameter κ. The median voxel accuracy rates are above 99% for an error variance of 0.5, while the median false positive rate is approximately 2%. As the error variance increases to 2, the accuracy rate remains very high. Figure 2.1 depicts our model's ability to recover the probability (panels c, g, and k) and binary (panels d, h, and l) maps. Regression and autoregressive coefficients, error variances, and spatial parameters were estimated with minimal error (not shown).
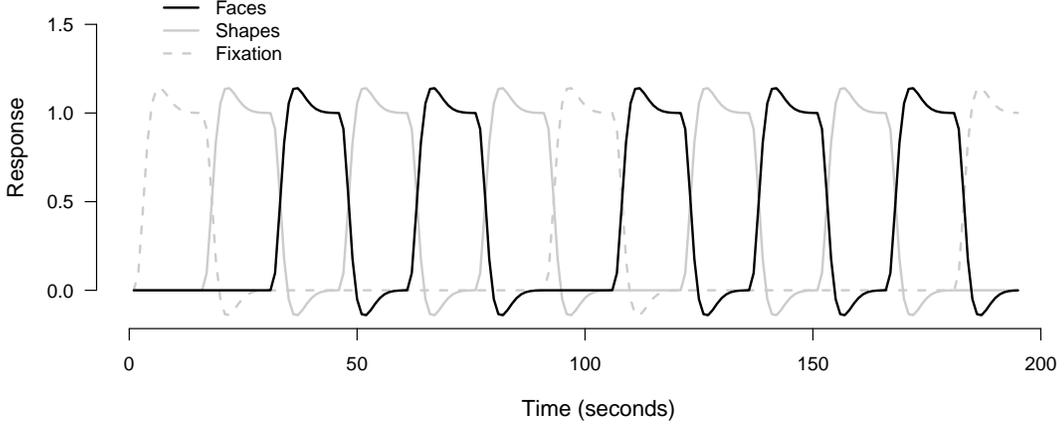
Figure 2.2: Experiment stimuli. Each curve represents a set of stimulus onset times convolved with the gamma density. The gray curves are the shape and fixation stimuli, and the black curve is the faces stimulus. The faces stimulus was used to simulate data for each of the simulation studies.

Table 2.1: Median accuracy and false positive rates for simulated data. Median percentage (minimum and maximum) of correctly classified voxels and false positives over 100 simulated datasets for each scenario are displayed. The Full dataset simulation corresponds to the 50×50 dataset. The Parcellation simulation corresponds to the full 50×50 dataset partitioned into four 25×25 parcels.

Scenario	κ	σ_v^2	Accuracy	False positive
1	0.5	0.5	99.8% (97.6%, 100%)	0.6% (0, 5.4%)
2	1	0.5	99.7% (98.7%, 100%)	0.7% (0, 2.9%)
3	2	0.5	99.7% (98.9%, 100%)	0.6% (0, 2.9%)
4	0.5	2	99.4% (96.6%, 100%)	1.2% (0, 3.6%)
5	1	2	99.5% (97.1%, 100%)	1.0% (0, 4.8%)
6	2	2	99.5% (97.4%, 100%)	1.1% (0, 3.2%)
Full dataset	–	–	98.7% (96.0%, 99.3%)	1.5% (0.7%, 4.4%)
Parcellation	–	–	99.4% (98.9%, 99.7%)	0.66% (0.31%, 1.1%)
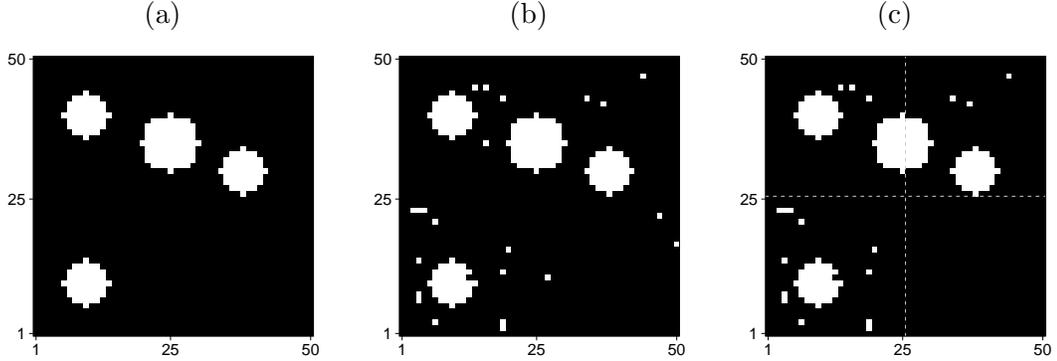
Figure 2.3: Parcellation for a dataset simulated on the 50×50 square lattice. (a) True binary map displaying areas of activation in white. (b) Estimated binary map based on analysis of full dataset. (c) Estimated binary map based on analysis of four 25×25 parcels. Dashed lines represent parcel edges.

2.4.2 Edge effects

For the second simulation study we used the `neuRosim` package (Welvaert et al., 2011) to simulate spatiotemporal fMRI outcomes on the 50×50 square lattice. Four regions of activation were selected, as shown in Figure 2.3a, and the faces stimulus was used, as shown in Figure 2.2. The locations of activation were chosen such that one region of activation overlaps two quadrants, and one quadrant is empty. The 2,500 voxels were simulated from a white noise process or from an autoregressive process of order two with $\rho_1 = 0.4$ and $\rho_2 = -0.2$, dependent on the true value of γ_v shown in Figure 2.3a. The signal-to-noise ratio was chosen to be 4.5, as in our application to real data.

To assess sensitivity to the chosen parcellation, we simulated 100 datasets on the 50×50 lattice, and applied our model to each 50×50 dataset in full. Next, we partitioned each 50×50 dataset into four 25×25 parcels, and carried out estimation on each of the parcels independently. The four parcels were then rejoined before classification accuracy was assessed.

The summaries are given in Table 2.1, scenarios Full Dataset and Partition. We see that, similar to the varying precisions simulation study, the high sensitivity, the high percentage of voxels correctly classified, and the relatively low false positive rate shows that our model is insensitive to parcellation. Figures 2.3b and 2.3c display the

resulting binary maps for a randomly selected dataset. Somewhat surprisingly, the false positive rate under parcellation was slightly lower than for analyses of the full datasets. We investigated the sensitivity and false positive rate among voxels along the partition lines, and the results were similar to those for the full datasets (not shown). Regression and autoregressive coefficients, error variances, and spatial parameters were estimated with minimal error (not shown).

## 2.5 Application to a single subject from an adolescent-depression study

The dataset used to illustrate our method is a single-subject fMRI time series from a visual experiment. The dataset was collected as part of a larger study designed to quantify the interaction between depression and neural response in adolescents with major depressive disorder (Musgrove et al., 2015). Subjects completed a 6.5 minute fMRI task during the scan that had two matching conditions: facial expressions of fear and anger or neutral objects (geometric shapes). The task included 13 counterbalanced blocks (three fixation, five shape, and five faces). Face and shape blocks comprised six consecutive five-second trials. Fixation blocks were presented at the beginning, after five blocks of shapes and faces, and at the conclusion of the task. The task was designed to elicit a negative emotional response (Hariri et al., 2002) with activation expected in regions including the amygdala, anterior cingulate cortex (subgenual and supragenual regions), and prefrontal cortex (ventromedial and dorsolateral regions) (Musgrove et al., 2015).

Imaging was performed using a 3T Siemens Trio MRI Scanner at the Center for Magnetic Resonance Research at the University of Minnesota. The task was projected onto a screen inside the scanner's bore that participants could view using a mirror on the head coil. Using a T2-sensitive echo planar sequence, 195 scans were acquired, 46 axial slices (no gap, interleaved slice acquisition), four mm in thickness (repetition time two seconds, echo time 28 milliseconds, 64×64 array, and a flip angle of $80°$). The voxel size was $3.1 \times 3.1 \times 4.0$ mm^3 and there were 61,000 brain voxels in total. The following preprocessing steps were used (FMRIB software library, FSL; `http://fsl.fmrib.ox.ac.uk`): motion correction using the first volume in the functional series as the reference

volume, slice timing correction, skull stripping, high-pass temporal filtering with a 25 second cutoff, pre-whitening, and registration to the Montreal Neurological Institute (MNI) standard space.

The baseline trend $\mathbf{Z}$ was modeled as six sets of polynomial functions to reduce the effects of head motion and respiration, i.e., $\mathbf{Z}$ is a 195×6 matrix. Onset time indicators for each of the stimuli were convolved with the gamma HRF, as shown in Figure 2.2. The resulting shape and face stimuli time series were used in the design matrix. Our model was fit to a single subject's dataset. The parcellation technique resulted in 90 parcels with a median parcel size of 729 voxels (min: 500, max: 1,000). Use of the SAMM resulted in an average spatial dimension reduction of 72% per parcel. The prior probability of activation was fixed at 2% across all voxels and stimuli by choosing $\psi_{vj} = -2.05$ for $v = 1, \ldots, N$ and $j = 1, \ldots, p$.

Initial values for MCMC were chosen as the maximum likelihood estimate of each parameter. A single chain with no burn-in was used to analyze each parcel. Due to the high dimensionality, it is difficult to use convergence criteria beyond a sufficiently small Monte Carlo standard error (MCSE). Thus, to ensure small MCSEs for all parameter estimates, 200,000 samples were drawn from the posterior distribution for each parcel. The largest MCSE of the estimated posterior means of the activation indicator variables was below 0.005, indicating that a sufficient number of samples was drawn. Figure 2.4 displays axial slices showing voxels with BOLD level changes for the faces stimulus exceeding fixation with a high probability. The voxels highlighted in white have estimated posterior probabilities that exceed 0.8722. Regions highlighted include the amygdala, anterior cingulate cortex, and prefrontal cortex.

The results in Figure 2.4 demonstrate the model's ability to encourage spatial clustering in areas with a high posterior probability of activation. This result reflects the goal of task-based experiments in which an investigator aims to reveal localization of task-related activation effects. Of primary interest for these data is activation of the amygdala, a relatively small functional region located medially within the temporal lobes (shown in Figure 2.4a). Figure 2.4 (panels a and b) show the expected activity in the visual cortex. Figure 2.4 (panels c and d) show activity in the prefrontal cortex, especially the dorsolateral region.

(a) Slice 17 (b) Slice 19 (c) Slice 25 (d) Slice 27
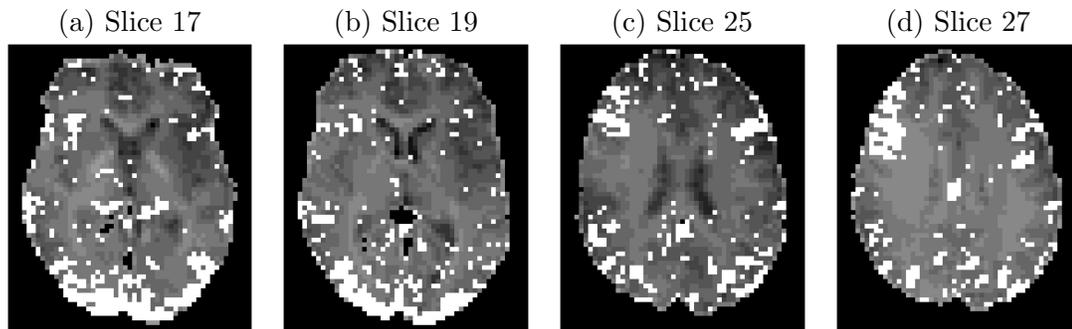
Figure 2.4: The maps portray voxel-specific posterior probabilities, exceeding 0.8722, of increased activity over baseline for the face stimulus. Activated voxels are shown in white and super-imposed over the fMRI slices. Slice numbers denote axial slices (46 slices total). Important contiguous regions of activation include the amygdala (slice 17 (a)), the anterior cingulate cortex (slice 19 (b)), and ventromedial and dorsolateral prefrontal cortex (slices 25 (c) and 27 (d)).

2.6   Discussion

This chapter presented a new approach to spatial Bayesian variable selection for regressions located on a lattice, with an emphasis on applications in single-subject fMRI studies. Our approach is novel in that we avoid the use of Zellner's g-prior, and instead use a spike-and-slab prior, where the spike is a point mass at zero. We also employ a flexible autoregressive model to account for the temporal correlation that is known to exist with voxel time series.

It is important to account for voxel-level temporal correlation, and we found that accounting for spatial dependence is no less important. Our modeling framework efficiently handles temporal autocorrelation in a fully Bayesian way. Regarding spatial dependence, we conducted a small simulation study and compared our method to a non-spatial approach (see Appendix A). We found that our approach maintains the nominal false positive rate while yielding a higher true positive rate, demonstrating that spatial dependence should not be ignored

A potential drawback of our modeling approach is the assumption of space-time separability, as in the works of Lee et al. (2014) and Smith and Fahrmeir (2007). The

high dimensionality of the data makes checking the assumption of space-time separability impossible, thus in the absence of currently non-existent computing technology, we followed previous work and assume space-time separability for our model.

Future research should focus on extending our model to group study settings. For instance, we can estimate group effects by analyzing each of the subject's brains separately then adjusting the posterior probabilities by shrinking towards a group mean. We then have group-level brain maps of posterior probabilities which can be compared between groups.

To ensure convergence of the Markov chains in our MCMC scheme, a large number of draws must be made. For each of the simulation studies and the application to a single subject, we used the fixed-width approach to Markov chain convergence diagnosis (Flegal et al., 2008), in which samples are drawn until all Monte Carlo standard errors are smaller than some pre-selected threshold. We used the `batchmeans` package (Haran and Hughes, 2015) for R to compute Monte Carlo standard errors. Two hundred thousand samples were sufficient to ensure that all Monte Carlo standard errors were sufficiently small.

Computing time is limited by the number of parcels and access to a computing cluster. In our application, computations were carried out in a supercomputing environment using Intel Xeon E5-2680, 2.5 GHz processors. Using this infrastructure, our model required 40 minutes of computing time for the 90 parcels. We have seen that inference appears insensitive to parcellation, and that the joint use of the parcellation and the SAMM greatly reduces the computational burden relative to analysis of a full dataset. The MCMC procedure produces samples from the joint posterior distribution of the model parameters, which obviates adjustment for multiple comparisons in a post-processing step.

Chapter 3

Hierarchical Copula Regression Models for Areal Data

3.1 Introduction

In this chapter we introduce a hierarchical copula-based areal model that alleviates several modeling issues associated with conditional autoregressive (CAR) models. In a first stage, we employ a marginal model while in a second stage, a covariance selection (Dempster, 1972) Gaussian copula is used so that the model's lone spatial dependence parameter has a straightforward marginal interpretation. Covariance selection is carried out using iterative proportional scaling (IPS) to estimate the covariance matrix (Xu et al., 2015). We also develop a computational framework that will allow our method to be applied to large areal datasets.

Areal data are spatially aggregated data that arise in a variety of fields; see, e.g., Smith and Fahrmeir (2007), Reich and Hodges (2008), and Lawson (2014) for applications in neuroimaging, dentistry, and epidemiology, respectively. Often, reliable inference about regression coefficients, rather than prediction at unobserved spatial locations, is the primary goal of areal data analyses. Areal data are typically analyzed using spatial generalized linear mixed models (SGLMM), which employ random effects to account for spatial dependence. The random effects are usually assumed to have arisen from a Gaussian Markov random field (GMRF) (Rue and Held, 2005), i.e., a Gaussian random field with a Markov dependence structure determined by the adjacency among the areal

units (e.g., brain voxels, tooth surfaces, counties). Usually, 'adjacent' means 'shares a boundary.' The most commonly used GMRF model is the CAR model introduced by Besag et al. (1991). Despite its popularity, the CAR model has well known drawbacks. Two important deficiencies are counterintuitive marginal dependence structures (Wall, 2004; Assunção and Krainski, 2009) and spatial confounding, i.e., variance inflation and bias caused by collinearity between the spatial random effects and the fixed-effects predictors (Clayton et al., 1993; Reich et al., 2006; Hodges and Reich, 2010; Hughes and Haran, 2013).

Until recently, spatial copula methods focused on geostatistical data (Bárdossy, 2006; Kazianka and Pilz, 2010; Madsen and Fang, 2011) or applied geostatistical techniques to areal data (Masarotto and Varin, 2012). Hughes (2015) introduced copCAR, which uses the proper CAR model to construct a Gaussian copula and then applies the copula directly to areal outcomes. Although copCAR obviates spatial confounding and has other advantages relative to the traditional application of the CAR model in a mixed model setting, copCAR does face a few challenges.

First, when the outcomes are discrete, the copCAR likelihood is computationally intractable unless the sample size is small (Henn, 2015). This can be addressed by using, e.g., the continuous extension (Madsen, 2009) (a Monte Carlo maximum likelihood approach), the distributional transform approximation (Kazianka and Pilz, 2010), or a composite marginal likelihood approach (Varin, 2008), each of which has drawbacks. First, none of the methods can be expected to work well in all settings (Henn, 2015). For instance, the distributional transform approximation requires fairly smooth cumulative distribution functions or estimation suffers. Continuous extension can be numerically unstable for large sample sizes and it scales poorly when computing parameter variance estimates via a bootstrap. The composite marginal likelihood approach has a large computational complexity which must be overcome via an approximation, potentially resulting in poor inference. Our approach is not computationally intractable and we are able to avoid each of these drawbacks in part by applying the copula to a continuous latent field.

Second, because copCAR applies the copula directly to the outcomes, the GMRF's conditional independence does not carry over to the observations if the observations are discrete (Murray et al., 2013). This is unappealing because the correspondence

between the underlying adjacency structure and the GMRF's conditional independence is a primary reason for applying GMRF models to areal data. Instead, our approach applies the copula to a continuous latent field, allowing the conditional independence to carry over to the outcomes despite their marginal distributions.

Third, applying the copula directly to discrete outcomes has the additional drawback that the marginal distributions are not free of the dependence structure (Genest and Neslehova, 2007). Our hierarchical model eliminates this confounding.

Finally, to highlight a benefit of our new approach, the continuous latent field gives our model more flexibility than the copCAR. For example, our approach can naturally accommodate overdispersed count data by employing a gamma–Poisson formulation (Prates et al., 2015).

The rest of the chapter is organized as follows. In Section 3.2 we introduce our copCS model, where 'CS' denotes covariance selection, which we discuss in Section 3.3. In Section 3.4 we describe our approach to inference for copCS. In Section 3.5 we present the results of a simulation study to assess performance of copCS for realistic sample sizes. In Section 3.6 we apply copCS and competing models to stomach cancer data from Slovenia. In Section 3.7 we give concluding remarks. Details of the IPS algorithm are given in Appendix B.

3.2   The copCS model for areal data

In recent years, copula regression models have become more popular as their utility and ease of use have increased (Kolev and Paiva, 2009). The primary advantage of copula modeling is modularity: the dependence structure and the marginal distributions can be modeled separately and then joined by way of the probability integral transform (Sklar, 1959). See Joe (2014) for an excellent treatment of modeling with copulas.

Our model, copCS, employs a Gaussian copula, which can be constructed as follows. For n areal units, let $\mathbf{\Phi_Q}$ be the cumulative distribution function (CDF) of a multivariate normal random variable with mean zero and $n \times n$ concentration matrix, i.e., inverse correlation matrix, $\mathbf{Q}$. The Gaussian copula can be written

$$\mathcal{C}_{\mathbf{Q}}\left(\mathbf{u}\right) = \mathbf{\Phi_Q}\left\{\Phi^{-1}\left(u_1\right),\,\ldots,\,\Phi^{-1}\left(u_n\right)\right\}, \tag{3.1}$$

where $\mathbf{u} = (u_1, \ldots, u_n)'$, u_i, $i = 1, \ldots, n$, is marginally standard uniform distributed, and Φ^{-1} is the quantile function for a standard normal variable. Thus a sample from $\mathcal{C}_{\mathbf{Q}}$ is a collection of standard uniform random variables having the Gaussian correlation structure defined by $\mathbf{Q}^{-1}$.

We parameterize in terms of $\mathbf{Q}$, rather than $\mathbf{Q}^{-1}$, because it is customary to do so in the context of GMRFs. For copCS, the concentration matrix $\mathbf{Q}$ takes a covariance selection form and can be computed efficiently using a recent extension of the iterative proportional scaling (IPS) algorithm (Xu et al., 2015). We discuss covariance selection and corresponding computation in Section 3.3.

To model areal outcomes, the covariance selection Gaussian copula of (3.1) can be paired with a collection of marginal distributions with CDFs $F_1, \ldots, F_n$. When the F_i are continuous, it is straightforward to apply the copula directly to the outcomes in this way, since the corresponding likelihood is so-called meta-Gaussian. When the marginal distributions are discrete, however, the likelihood is usually intractable because the computational complexity is in $O(2^n)$. Several solutions have been proposed, including the continuous extension (Madsen, 2009), the distributional transform approximation (Kazianka and Pilz, 2010), or a composite marginal likelihood approach (Varin, 2008), as well as treating the marginal distributions as nuisance parameters (Dobra and Lenkoski, 2011). Here we take a different approach that offers several advantages.

We follow Prates et al. (2015) in applying the copula to a continuous latent field rather than directly to the observed areal outcomes. Specifically, for observation y_i, $i = 1, \ldots, n$, we specify a distribution $g_i(y_i \mid \lambda_i)$, where λ_i is a continuous latent variable. Then we choose a distribution $f_i(\lambda_i \mid \boldsymbol{\psi})$ for λ_i, where $\boldsymbol{\psi}$ are hyperparameters associated with f_i. The copula is applied to the λ_i to induce dependence among the λ_i, $i = 1, \ldots, n$, and thus among the outcomes.

In hierarchical form, copCS can be specified generally as

$$
\begin{aligned}
y_i \mid \lambda_i &\sim g_i(y_i \mid \lambda_i), \\
\lambda_i \mid \boldsymbol{\psi} &\sim f_i(\lambda_i \mid \boldsymbol{\psi}), \ i = 1, \ldots, n \\
\mathbf{z} &\sim \mathcal{N}\left\{\mathbf{0}, \mathbf{Q}(\rho_{\mathrm{CS}})^{-1}\right\},
\end{aligned}
$$

where $\mathbf{z} = (z_1, \ldots, z_n)'$, $z_i = \Phi^{-1}\{F_i(\lambda_i \mid \boldsymbol{\psi})\}$, $i = 1, \ldots, n$, F_i is the CDF corresponding to probability distribution function or probability mass function f_i, and $\mathbf{Q}(\rho_{\mathrm{CS}})$

implies that the concentration matrix $\mathbf{Q}$ depends on an unknown first-order correlation parameter ρ_{CS}. The copCS likelihood has the form $\mathcal{L} = \mathcal{L}_{\mathcal{C}} \times \mathcal{L}_{\mathbf{y},\boldsymbol{\lambda}}$, where

$$\mathcal{L}_{\mathcal{C}} = |\mathbf{Q}(\rho_{\text{CS}})|^{1/2} \exp\left\{ -\frac{1}{2}\mathbf{z}' \left(\mathbf{Q}(\rho_{\text{CS}}) - \mathbf{I}\right)\mathbf{z}\right\} \tag{3.2}$$

and

$$\mathcal{L}_{\mathbf{y},\boldsymbol{\lambda}} = \prod_{i=1}^{n} g_i\left(y_i \mid \lambda_i\right) f_i\left(\lambda_i \mid \boldsymbol{\psi}\right). \tag{3.3}$$

The likelihood $\mathcal{L}_{\mathcal{C}}$ is obtained by applying the chain rule to (3.1). For $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)'$, $\boldsymbol{\theta} = (\boldsymbol{\psi}', \boldsymbol{\lambda}', \rho_{\text{CS}})'$ is the full parameter vector. Thus, we optimize the copCS log-likelihood to obtain the maximum likelihood estimate (MLE) $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$. See Section 3.4 for details.

3.2.1 Modeling count data

Here we provide a concrete example of copCS modeling by specifying a model for Poisson distributed outcomes. For $\boldsymbol{\psi} = \left(\boldsymbol{\beta}', \nu\right)'$, we have

$$
\begin{aligned}
y_i \mid \lambda_i &\sim \mathcal{P}ois\left(\lambda_i\right), \\
\lambda_i \mid \boldsymbol{\beta}, \nu &\sim \mathcal{G}amma\left(\nu\mu_i, \nu\right), \ i = 1, \ldots, n \\
\mathbf{z} &\sim \mathcal{N}\left\{\mathbf{0}, \mathbf{Q}\left(\rho_{\text{CS}}\right)^{-1}\right\},
\end{aligned}
$$

where $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients, $\mu_i = \exp\left(\mathbf{x}_i'\boldsymbol{\beta}\right)$, $\mathbf{x}_i$ are the p covariates for the ith areal unit, $\mathbf{z} = (z_1, \ldots, z_n)'$, $z_i = \Phi^{-1}\left\{F_i\left(\lambda_i \mid \boldsymbol{\beta}, \nu\right)\right\}$, $i = 1, \ldots, n$, and F_i is the CDF for the gamma distribution with shape parameter $\nu\mu_i$ and rate parameter ν. Using iterated expectation, we see that

$$\mathbb{E}y_i = \mathbb{E}\left\{\mathbb{E}\left(y_i \mid \lambda_i\right)\right\} = \mathbb{E}\lambda_i = \nu\mu_i/\nu = \mu_i = \exp\left(\mathbf{x}_i'\boldsymbol{\beta}\right),$$

which implies an intuitive marginal interpretation for $\boldsymbol{\beta}$. The variances of λ_i and y_i are

$$\mathbb{V}\lambda_i = \nu\mu_i/\nu^2 = \mu_i/\nu = \exp\left(\mathbf{x}_i'\boldsymbol{\beta}\right)/\nu$$

and

$$\mathbb{V}y_i = \mathbb{V}\left\{\mathbb{E}\left(y_i \mid \lambda_i\right)\right\} + \mathbb{E}\left\{\mathbb{V}\left(y_i \mid \lambda_i\right)\right\} = \mu_i/\nu + \mu_i = \left(1 + 1/\nu\right)\exp\left(\mathbf{x}_i'\boldsymbol{\beta}\right),$$

respectively. Because $\mathbb{E}y_i < \mathbb{V}y_i$ for $\nu < \infty$, the parameterization allows copCS to accommodate overdispersion, which is accounted for by ν. Since the λ_i are continuous, the conditional independence among the z_i carries over to the λ_i, and the conditional independence structure is obtained for the y_i as well.

3.3 Covariance selection

Here we review covariance selection and present our method for estimating the concentration matrix $\mathbf{Q}$ for copCS. For ease of exposition, in this section we use $\mathbf{Q}$ in place of $\mathbf{Q}(\rho_{\mathrm{CS}})$. We begin with the underlying graph $G = (V, E)$, where V and E are finite sets of vertices and edges, respectively. Two vertices are considered adjacent, i.e., share an edge, if and only if (iff) their corresponding areal units share a boundary. Alternative notions of adjacency may be more appropriate in some settings. In any case, the adjacency structure of G leads to conditional independence for the GMRF.

Suppose we have a random vector $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{R})$, where $\mathbf{R}$ is a correlation matrix with inverse $\mathbf{Q}$. In a copula setting, variances are seldom identifiable, so it is typical to specify the inverse of the correlation matrix rather than specifying the inverse covariance matrix. Now, for vertices $i, j \in V$, let $\mathbf{Q}_{ij} \neq 0$ if $(i, j) \in E$, and let $\mathbf{Q}_{ij} = 0$ if $(i, j) \notin E$. If $\mathbf{Q}_{ij} = 0$, w_i and w_j are conditionally independent. That is, iff vertices i and j do not share an edge, the corresponding entry of $\mathbf{Q}$ is 0, and the random variables w_i and w_j are independent conditional on their neighbors (Speed and Kiiveri, 1986).

Some choices for the nonzero entries of $\mathbf{Q}$ lead to a sensible marginal correlation structure, and some do not. We desire a sensible correlation structure so we fix the nonzero entries of $\mathbf{Q}$ such that the first-order correlation takes on some constant value ρ_{CS}. We define the first-order correlation ρ_{CS} as follows. For vertices $i, j \in V$, let $\mathbf{R}_{ij} = \rho_{\mathrm{CS}}$ if $(i, j) \in E$, and let $\mathbf{R}_{ij} \neq \rho_{\mathrm{CS}}$ if $(i, j) \notin E$ (of course $\mathbf{R}_{ij} = 1$ for $i = j$). Thus, ρ_{CS} is the correlation between neighboring areal units. The approach we use to estimate $\mathbf{Q}$ represents a special case of covariance selection, which was investigated first by Dempster (1972) and later by Wermuth and Scheidt (1977) and Besag and Kooperberg (1995). Covariance selection can be accomplished using iterative proportional scaling (Deming and Stephan, 1940). We modified the localized IPS algorithm of Xu et al. (2015) to estimate the concentration matrix $\mathbf{Q}$. Below we offer a sketch of the IPS algorithm for

Q. A complete description of the modified IPS algorithm is given in Appendix B.

The log-likelihood with respect to ρ_{cs}, the first order correlation, can be written

$$\log \mathcal{L}\left(\rho_{\mathrm{cs}} \mid \mathbf{w}\right) = -\frac{1}{2}\left\{\log |\mathbf{Q}| - \mathrm{tr}\left(\mathbf{Q}\mathbf{w}\mathbf{w}'\right)\right\}, \tag{3.4}$$

where $\mathrm{tr}(\cdot)$ denotes the matrix trace. The IPS algorithm takes as input the adjacency matrix $\mathbf{A}$ for the graph G, and the first-order correlation ρ_{cs}. The adjacency matrix has entries $\mathbf{A}_{ij} = I\{(i,j) \in E\}$, where $I\{\cdot\}$ denotes the indicator function. A collection of vertices is considered complete if every set of two vertices in the collection is adjacent. A maximal complete subset of vertices is a clique. That is, a clique is a complete subset that is not contained in a larger complete subset. Let $\mathcal{K}(G)$ denote the set of cliques of graph G.

To estimate $\mathbf{Q}$ using the localized IPS algorithm, we begin by partitioning the clique set $\mathcal{K}(G)$ into M non-overlapping sets, indexed by $m = 1, \ldots, M$. Denote the vertex set for the mth clique set as U_m, and let $\bar{U}_m$ be the set of vertices not in U_m, i.e., $\bar{U}_m$ is the complement of U_m with respect to V. Let $\mathbf{Q}_{U_m}$ denote the submatrix of $\mathbf{Q}$ corresponding to the mth clique set, and let $\mathbf{Q}_{U_m \bar{U}_m}$ be the corresponding block off-diagonal. Let $\mathbf{w}\mathbf{w}' = \mathbf{S}$, $\mathbf{S}_{U_m} = (1 - \rho_{\mathrm{cs}})\mathbf{I}_{\|U_m\|} + \rho_{\mathrm{cs}}\mathbf{J}_{\|U_m\|}$, where $\mathbf{I}$ is the identity matrix, $\mathbf{J}$ is the square matrix of ones, and $\|U_m\|$ is the number of vertices in U_m. Thus, $\mathbf{S}_{U_m}$ is a dense clique set-level correlation matrix. We iteratively update the submatrix of $\mathbf{Q}$ corresponding to U_m as

$$\mathbf{Q}_{U_m} \leftarrow \mathbf{S}_{U_m}^{-1} + \mathbf{Q}_{U_m \bar{U}_m}\mathbf{Q}_{\bar{U}_m}^{-1}\mathbf{Q}'_{\bar{U}_m U_m}, \quad m = 1, \ldots, M,$$

by adjusting $\mathbf{Q}_{U_m}$ for each clique in U_m. Xu et al. (2015) showed that, compared to previous versions of IPS, adjusting $\mathbf{Q}$ locally via a clique partition greatly reduces the computational complexity of the algorithm.

To enumerate the cliques of G, we first reduce the bandwidth of the adjacency matrix by reordering the vertices of G using the Cuthill–McKee algorithm (Cuthill and McKee, 1969). This speeds up the IPS procedure considerably. Next, the cliques are enumerated using the Bron–Kerbosch algorithm (Bron and Kerbosch, 1973). We used the `RBGL` package (Carey et al., 2015) for R (Ihaka and Gentleman, 1996) to implement the clique enumeration and bandwidth reduction algorithms.

IPS is iterative in that the submatrices of $\mathbf{Q}$ corresponding to the U_m, $m = 1, \ldots, M$, are updated until convergence, i.e., when the change in the log-likelihood of (3.4) is less
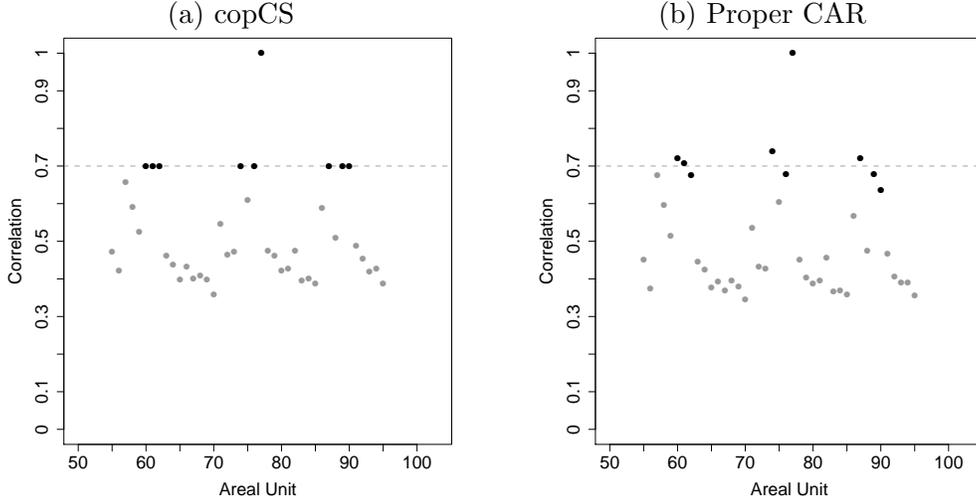
Figure 3.1: Correlations between the 77th municipality and fifty other municipalities of the Slovenia dataset. The black point at correlation equal to 1 corresponds to municipality 77 while the other black points correspond to municipality 77's first-order neighbors. The gray points represent the correlations between municipality 77 and the remaining municipalities with indices between 50 and 100. Panel (a) shows the correlations for the copCS model with $\rho_{CS} = 0.7$. Panel (b) shows the analogous correlations for the proper CAR model with $\rho_{CAR} = 0.985$.

than ε for some small $\varepsilon > 0$. Upon convergence, $\mathbf{Q}$ and the determinant $|\mathbf{Q}|$ are returned. Our choice of $\mathbf{S}_{U_m}$ ensures that the algorithm produces the matrix $\mathbf{Q}$ such that draws from the $\mathcal{N}(\mathbf{0}, \mathbf{Q}^{-1})$ distribution will have the desired properties, namely, constant first-order correlation ρ_{CS} and conditional independence corresponding to G.

Based on the form of each $\mathbf{S}_{U_m}$, $m = 1, \ldots, M$, we must place a restriction on ρ_{CS} to ensure positive definiteness of the concentration matrix $\mathbf{Q}$. For $n_M = \max_m \|U_m\|$, we must have $-1/(n_M - 1) < \rho_{CS} < 1$ to ensure that each $\mathbf{S}_{U_m}$ is invertible and $\mathbf{Q}$ is positive definite. In our work with areal data, it suffices to restrict $0 \leq \rho_{CS} < 1$.

As a brief illustration of how copCS fixes the marginal correlation problem of the proper CAR model, we use a subset of the 194 municipalities of Slovenia (Zadnik and Reich, 2006). Figure 3.1a displays the marginal correlations between municipality 77 and municipalities 50 through 100 of Slovenia for the concentration matrix computed

via IPS with $\rho_{\text{CS}} = 0.7$. Figure 3.1b displays marginal correlations as in panel (a) but for the proper CAR model with $\rho_{\text{CAR}} = 0.985$ (the 'CAR' subscript denotes proper CAR). For the proper CAR correlation matrix, we first compute the precision matrix $\mathbf{\Omega}\left(\rho_{\text{CAR}}\right) = \mathbf{D} - \rho_{\text{CAR}}\mathbf{A}$, where $\mathbf{D}$ is the diagonal matrix with the degrees (i.e., number of neighbors) of the municipalities of Slovenia on its diagonal and $\mathbf{A}$ is the adjacency matrix. Then, the correlation matrix is computed from $\mathbf{\Omega}\left(\rho_{\text{CAR}}\right)^{-1}$. In each plot, the black dots represent municipality 77 (correlation of 1) and its first-order neighbors while the gray dots represent the remaining municipalities. The concentration matrix produced via IPS results in constant marginal correlation for first-order neighbors. In general, the proper CAR model does not lead to constant first-order correlation.

3.4   Frequentist inference for copCS

We use a two-stage estimation procedure similar to that presented by Joe (2005). In the first stage we estimate the parameters of the marginal distributions while in the second stage we estimate the copula parameter ρ_{CS} conditional on the estimates of the marginal parameters.

In the first stage we estimate $\boldsymbol{\psi}$ by integrating the λ_i out of $\mathcal{L}_{\mathbf{y},\boldsymbol{\lambda}}$ (3.3) and optimizing the result,

$$\mathcal{L}_{\mathbf{y}} = \int \prod_{i=1}^{n} g_i\left(y_i \mid \lambda_i\right) f_i\left(\lambda_i \mid \boldsymbol{\psi}\right) d\lambda_i,$$

which can be obtained analytically for proper choices of g_i and f_i. We use a quasi-Newton method (Byrd et al., 1995) to optimize $\mathcal{L}_{\mathbf{y}}$ with respect to $\boldsymbol{\psi}$, arriving at the MLE $\widehat{\boldsymbol{\psi}}$. Then we use $\mathcal{L}_{\mathbf{y},\boldsymbol{\lambda}}$, with $\boldsymbol{\psi}$ fixed at $\widehat{\boldsymbol{\psi}}$, to obtain $\widehat{\boldsymbol{\lambda}}$. For the gamma–Poisson setup described in Section 3.2.1, $\widehat{\boldsymbol{\lambda}}$ is available in closed form (see Section 3.4.2 for details).

In the second stage we estimate the first-order correlation ρ_{CS} by optimizing $\mathcal{L}_{\mathcal{C}}$ (3.2) with $\widehat{\mathbf{z}}$ in place of $\mathbf{z}$, i.e., $\widehat{\mathbf{z}} = (\widehat{z}_1, \ldots, \widehat{z}_n)'$ and $\widehat{z}_i = \Phi^{-1}\{F_i(\widehat{\lambda}_i \mid \widehat{\boldsymbol{\psi}})\}$, $i = 1, \ldots, n$. When n is relatively small (less than 500, say), the matrix $\mathbf{Q}(\rho_{\text{CS}})$ can be estimated quickly, in which case it is straightforward to optimize $\mathcal{L}_{\mathcal{C}}$. For larger n, computing $\mathbf{Q}(\rho_{\text{CS}})$ becomes burdensome. We alleviate the burden by pre-computing $\mathbf{Q}(\rho_{\text{CS}})$ for each value of $\rho_{\text{CS}} \in \{0.01, 0.02, \ldots, 0.99\}$ and then using a B-spline (Hastie and Tibshirani, 1986)
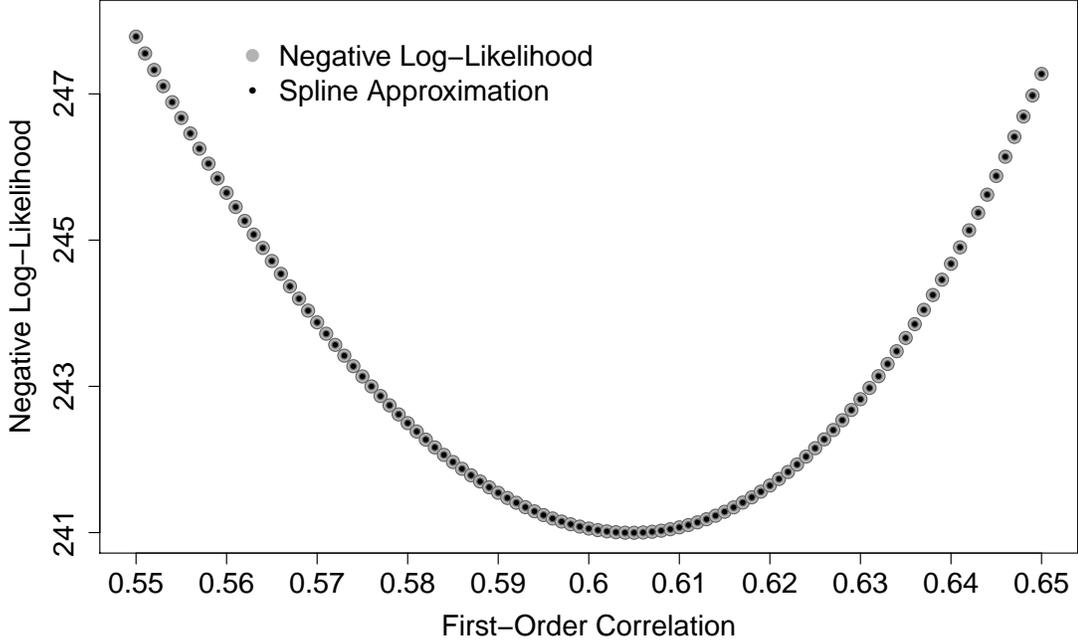
Figure 3.2: B-spline approximation of $\log \mathcal{L}_{\mathcal{C}}$.

approximation of $\log \mathcal{L}_{\mathcal{C}}$ during optimization. To illustrate, Figure 3.2 displays values of $-\log \mathcal{L}_{\mathcal{C}}$ for a single dataset simulated from $\mathcal{N}\left\{\mathbf{0}, \mathbf{Q}\left(\rho_{\mathrm{CS}}\right)^{-1}\right\}$, with $\rho_{\mathrm{CS}} = 0.6$, on the 40×40 lattice ($n = 1,600$, see Section 3.5.2). The gray points are the negative log-likelihoods calculated at $\rho_{\mathrm{CS}} = 0.550, 0.551, \ldots, 0.650$. The black points were calculated by first pre-computing $\mathbf{Q}(\rho_{\mathrm{CS}})$ on a coarse grid of $\rho_{\mathrm{CS}} \in \{0.01, 0.02, \ldots, 0.99\}$. Then, B-spline interpolations of the negative log-likelihood were evaluated at $\rho_{\mathrm{CS}} = 0.550, 0.551, \ldots, 0.650$. As we can see in Figure 3.2, the B-spline interpolation works well at approximating the true negative log-likelihoods.

3.4.1 Computing confidence intervals

Using a preliminary simulation study for gamma–Poisson copCS as a guide, we found that $\widehat{\boldsymbol{\beta}}$ is approximately normally distributed while $\widehat{\nu}$ and $\widehat{\rho}_{\mathrm{CS}}$ are typically right and left skewed, respectively. If we parameterize in terms of $\log \nu$ and $\Phi^{-1}(\rho_{\mathrm{CS}})$ instead, the estimator

$$\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\beta}}', \widehat{\log \nu}, \widehat{\Phi^{-1}(\rho_{\mathrm{CS}})})'$$

is approximately Gaussian distributed. That is, we have

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \Rightarrow \mathcal{N}\left\{\mathbf{0}, \mathcal{I}^{-1}(\boldsymbol{\theta})\mathcal{J}(\boldsymbol{\theta})\mathcal{I}^{-1}(\boldsymbol{\theta})\right\}, \tag{3.5}$$

where $\mathcal{I}$ is the Fisher information matrix and $\mathcal{J}$ is the variance of the score function, i.e.,

$$\mathcal{J}(\boldsymbol{\theta}) = \mathbb{V}\left(\nabla \log \mathcal{L}\right),$$

where ∇ is the gradient.

As an estimate of the Fisher information matrix, we use the observed Fisher information matrix $\widehat{\mathcal{I}}(\boldsymbol{\theta}) = \mathcal{H}(\widehat{\boldsymbol{\theta}})$, where $\mathcal{H}$ is the Hessian matrix $\nabla^2 \log \mathcal{L}$. The Hessian is available as a side effect of optimization. $\mathcal{J}$ is more difficult to estimate, however. We follow Hughes (2015) and use a parametric bootstrap to estimate $\mathcal{J}$ as

$$\widehat{\mathcal{J}}(\boldsymbol{\theta}) = \frac{1}{b}\sum_{k=1}^{b} \nabla\nabla' \log \mathcal{L}^{[k]}, \tag{3.6}$$

where b is the bootstrap sample size, $\nabla\nabla'$ is the outer product of the gradient, and $\mathcal{L}^{[k]}$ is the likelihood for the kth dataset simulated from the model at $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$. This combination of asymptotic inference (3.5) and bootstrap estimation of $\mathcal{J}$ (3.6) is much faster than full bootstrap inference since the former requires only numerical differentiation while the latter requires optimization of the likelihood.

3.4.2   Estimation for gamma–Poisson copCS

For the gamma–Poisson scenario described in Section 3.2.1, ignoring the copula, the likelihood can be written

$$\mathcal{L}_{\mathbf{y},\boldsymbol{\lambda}}\left(\boldsymbol{\beta}, \nu, \boldsymbol{\lambda} \mid \mathbf{y}\right) \propto \prod_{i=1}^{n} \frac{\nu^{\nu\mu_i}}{\Gamma\left(\nu\mu_i\right)} \lambda_i^{y_i + \nu\mu_i - 1} \exp\left\{-\lambda_i\left(1 + \nu\right)\right\}, \tag{3.7}$$

where $\mu_i = \exp\left(\mathbf{x}_i'\boldsymbol{\beta}\right)$. Integrating $\boldsymbol{\lambda}$ out of (3.7) leads to the marginal likelihood

$$\mathcal{L}_{\mathbf{y}}\left(\boldsymbol{\beta}, \nu \mid \mathbf{y}\right) \propto \prod_{i=1}^{n} \frac{\Gamma\left(y_i + \nu\mu_i\right)}{\Gamma\left(\nu\mu_i\right)} \nu^{\nu\mu_i}\left(1 + \nu\right)^{-(y_i + \nu\mu_i)},$$

which we optimize to estimate $\boldsymbol{\beta}$ and ν. Then, the MLEs of the λ_i are

$$\widehat{\lambda}_i = \left(y_i + \widehat{\nu}\widehat{\mu}_i - 1\right)\left(1 + \widehat{\nu}\right)^{-1}, \, i = 1, \ldots, n,$$

where $\widehat{\mu}_i = \exp\left(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}\right)$. Finally, we estimate ρ_{CS} as described in Section 3.4, with

$$\widehat{z}_i = \Phi^{-1}\left\{F_i(\widehat{\lambda}_i \mid \widehat{\nu}, \widehat{\mu}_i)\right\}, \, i = 1, \ldots, n,$$

using $\widehat{z}_i$ as an estimate of z_i, and F_i is the CDF of the gamma distribution with shape parameter $\widehat{\nu}\widehat{\mu}_i$ and rate parameter $\widehat{\nu}$.

3.5 Simulation study

We carried out a simulation study to assess the performance of copCS under realistic conditions. First, we simulated data to mimic the Slovenia cancer data ($n = 194$) that we analyze below in Section 3.6. We then simulated outcomes on the 40×40 square lattice ($n = 1,600$), using the coordinates of the vertices as predictors; the large size of the lattice allowed us to assess large-sample performance.

3.5.1 Simulation scenario: Slovenia map

The plot in Figure 3.3a shows the 194 municipalities of Slovenia, each of which is shaded to reflect the observed standardized incidence rate (SIR) of stomach cancer for that municipality during 1995–2001. A full description of the data is given in Section 3.6. For the simulation study and the subsequent stomach cancer analysis we considered a pair of municipalities to be neighbors iff they share a border. This definition leads to 286 cliques. After reducing the bandwidth of the adjacency matrix, we partitioned the cliques into 11 clique sets, with each set having 26 cliques. The partition permits $\mathbf{Q}(\rho_{\mathrm{CS}})$ to be computed in less than a second.

We used the two-column design matrix from the Slovenia dataset. The design matrix $\mathbf{X}$ is composed of a column of 1's for the intercept and a column that contains the socioeconomic status for each municipality. We also used an offset E_i that is the number of expected deaths from stomach cancer, computed by internal standardization. The regression coefficients were set to $\beta_0 = 0.2$ and $\beta_1 = -0.2$, and we let $\nu = 1$. These values were chosen to match the application results. Finally, we used a first-order correlation value of $\rho_{\mathrm{CS}} = 0.5$ throughout the simulation.

For $\mathbf{z} = (z_1, \ldots, z_{194})'$, we generated data for each of the 1,000 simulations by

sampling $\mathbf{z} \sim \mathcal{N} \left\{ \mathbf{0}, \mathbf{Q}(\rho_{\mathrm{CS}})^{-1} \right\}$. Then, λ_i, $i = 1, \ldots, 194$, was computed via the probability integral transform: $\lambda_i = F_i^{-1} \left\{ \Phi(z_i) \mid \nu, \mu_i \right\}$, where $\mu_i = E_i \exp(\mathbf{x}_i' \boldsymbol{\beta})$, F_i is the CDF of the gamma distribution with shape parameter $\nu \mu_i$ and rate parameter ν, and $\boldsymbol{\beta} = (\beta_0, \, \beta_1)'$. We then sampled the outcomes as $y_i \sim \mathcal{P}ois(\lambda_i)$, $i = 1, \ldots, 194$. We used a bootstrap sample size of $b = 500$ to estimate standard deviations. The delta method was used to obtain the presented standard errors for ν and ρ_{CS}. Simulation results are shown in Table 3.1. As expected, copCS permits unbiased estimation of the regression coefficients. The coverage for β_1 was slightly below the nominal rate. Estimates of the ν and ρ_{CS} parameters tended to be high, but the coverage for those parameters was excellent.

Table 3.1: Simulation results for the Slovenia graph and Poisson marginals. SE denotes standard error estimated from (3.5), SD denotes standard deviation over the 1,000 data sets, and CP denotes coverage probability of the 95% confidence interval.

Parameter	Truth	Mean Est.	Median SE	Simulation SD	95% CP
β_0	0.2	0.198	0.075	0.074	0.941
β_1	-0.2	-0.199	0.042	0.053	0.893
ν	1	1.224	2.451	0.350	0.960
ρ_{CS}	0.5	0.566	0.133	0.082	0.978

3.5.2 Simulation scenario: 40×40 lattice

Next, we carried out a simulation study on a 40×40 square lattice ($n = 1,600$). On the square lattice, we assume two vertices are adjacent if they share a face. The 40×40 lattice has 3,210 cliques. We partitioned the cliques into 16 clique sets, with each set having 195 cliques. We used a two-column design matrix such that row i contains the x and y coordinates of the ith vertex, $i = 1, \ldots, 1600$, where the coordinates were taken from the unit square. Here, we set $E_i = 1$ for all i, and set $\beta_0 = \beta_1 = 0.5$, $\nu = 2$, and $\rho_{\mathrm{CS}} = 0.7$. We then simulated 1,000 datasets using the method described in Section 3.5.1 and applied copCS to each. We used the B-spline approximation method to speed estimation of ρ_{CS}. Figure 3.3b displays one of the simulated datasets.

The simulation results are shown in Table 3.2. The performance was comparable to the Slovenia scenario, except in this 40×40 lattice scenario, the coverage for both
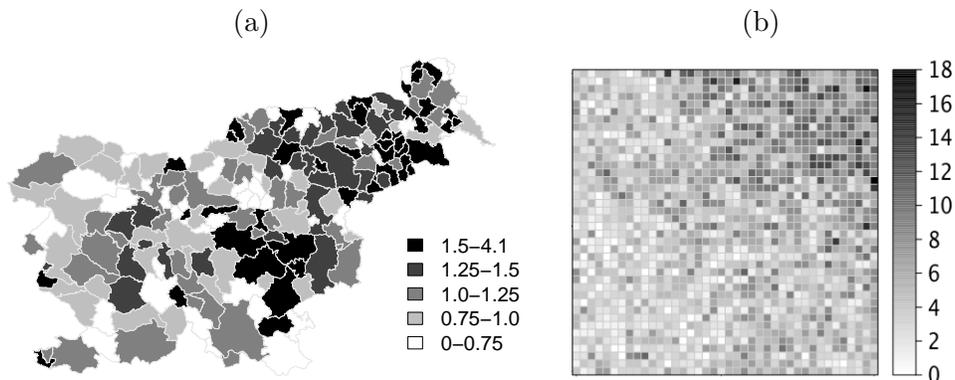
Figure 3.3: Simulation scenarios. Simulated standardized incidence rates for the 194 municipalities of Slovenia (a). Counts simulated on the 40×40 square lattice (b). Darker shades of gray represent larger values.

regression coefficients was very close to the nominal 95%. Similar to the Slovenia scenario, estimates of the ν and ρ_{CS} parameters tended to be high, but the coverage for those parameters was excellent.

Table 3.2: Simulation results for the 40×40 lattice and Poisson marginals. SE denotes standard error estimated from (3.5), SD denotes standard deviation over the 1,000 data sets, and CP denotes coverage probability of the 95% confidence interval.

Parameter	Truth	Mean Est.	Median SE	Simulation SD	95% CP
β_0	0.5	0.502	0.170	0.185	0.927
β_1	0.5	0.491	0.171	0.186	0.926
ν	2	2.601	5.672	0.527	1.000
ρ_{CS}	0.7	0.755	0.266	0.029	1.000

3.6   Application to the Slovenia data

The goal of the initial analysis of the Slovenia stomach cancer data was to investigate the relationship between cancer incidence and socioeconomic status (Zadnik and Reich,

2006), but this dataset has remained popular largely due to its usefulness for investigating spatial confounding (Reich et al., 2006; Hodges and Reich, 2010; Hughes, 2015). The covariate of interest is the municipality-level socioeconomic status (SES).

To these data we applied a nonspatial model, copCS, copCAR, and the areal GLMM with proper CAR random effects. The nonspatial model was the classical Poisson generalized linear model with the canonical log link function. For the copCAR model, we used the distributional transform approach. For both copCS and copCAR we used a bootstrap sample size of $b = 500$. For the areal GLMM, with O_i and E_i denoting the observed and expected number of cancer cases, the model is

$$\log \mathbb{E} O_i = \log E_i + \beta_0 + \beta_1 \mathrm{SES}_i + \delta_i,$$

where β_0 is the intercept, β_1 is the fixed effect for SES_i, and δ_i is the spatial effect for the ith municipality. For the nonspatial model, copCS, and copCAR, the spatial effect δ_i is not included. Since the areal GLMM is typically used in a Bayesian setting, we fit the model using Markov chain Monte Carlo (MCMC) for Bayesian inference. We assumed $\boldsymbol{\beta}$, ρ_{BCAR}, and τ^2 to be independent *a priori*, where the 'B' in ρ_{BCAR} is used to denote Bayesian for the Bayesian areal GLMM. The priors for each of $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ and ρ_{BCAR} were $\mathcal{N}(\mathbf{0}, 10^6 \mathbf{I}_2)$ and $\mathcal{U}(0, 1)$, respectively. The prior for τ^2 was inverse gamma with shape $1/2$ and scale $1/2000$. Posterior samples were drawn using a Metropolis–Hastings random walk. We used the fixed-width approach for convergence diagnosis (Flegal et al., 2008), where samples are drawn until all Monte Carlo standard errors are sufficiently small. We used the `batchmeans` package (Haran and Hughes, 2015) for R to compute Monte Carlo standard errors. Two hundred thousand samples were sufficient to ensure that all Monte Carlo standard errors were less than 0.05. The MCMC estimates stabilized after approximately 80,000 iterations.

The results are shown in Table 3.3. Each of the nonspatial model, copCS, and copCAR found SES to be a significant predictor of SIR while the areal GLMM found that SES is not significantly associated with SIR. Both the copCS and copCAR intervals for the fixed effects estimates are wider than the intervals for the nonspatial model, implying that an adjustment for spatial dependence was necessary. The SES effect estimate of the areal GLMM is biased towards zero, a result known to be due to spatial confounding. Turning our attention to ρ, the copCS estimate of $\rho_{\mathrm{CS}} = 0.513$ implies that

these data exhibit moderate spatial dependence since the first-order correlation for the copula is moderate in strength. Meanwhile, copCAR's estimate of $\rho_{\text{CAR}} = 0.420$ does not have the same simple interpretation. Not shown in the table, the ν parameter of copCS was estimated as 1.129 with 95% confidence interval (0.063, 20.270). The prior variance τ^2 of the areal GLMM with proper CAR random effects was estimated as 0.116 with 95% credible interval (0.052, 0.215).

Table 3.3: Parameter estimates from applying the classical GLM, copCS, copCAR, and the areal GLMM to the Slovenia stomach cancer data. Results are presented as estimate (95% confidence interval).

Model	Intercept	SES	ρ
Nonspatial	0.156 (0.120, 0.192)	-0.137 (-0.175, -0.098)	–
copCS	0.157 (0.048, 0.265)	-0.137 (-0.216, -0.059)	0.513 (0.393, 0.632)
copCAR	0.151 (0.106, 0.195)	-0.122 (-0.174, -0.071)	0.420 (0.095, 0.819)
Areal GLMM	0.124 (0.080, 0.166)	-0.072 (-0.151, 0.017)	0.565 (0.127, 0.927)

3.7 Discussion

In this chapter we presented copCS, a new hierarchical, copula-based regression model for areal data, and developed an efficient computational approach to frequentist inference. By conveniently choosing latent variables that ensure unbiased estimation, we gain an intuitive marginal interpretation of the regression coefficients. In contrast to competing models, the copula dependence parameter has a clear interpretation as the first-order correlation.

An alternative use of covariance selection is as a specification for the dependence structure among random effects in an areal GLMM (Besag and Kooperberg, 1995). For instance, in a Bayesian setting we could easily incorporate our covariance selection method as a prior distribution that forces similar first-order (marginal) correlations in

the posterior. We employed covariance selection in a copula setting because we are primarily interested in the marginal effect of covariates. As mentioned in the introduction, the introduced random effects implies that the areal GLMM has a more complicated conditional interpretation.

A potential drawback of copCS is that it is not a spatial smoother. In particular, visualizations of the mean fit of the model will only involve the regression coefficients. If smoothing is desired, we can apply a method such as spatial filtering (Griffith, 2003), which is an avenue for future research. We can also extend the model to handle alternate distributions, e.g., the beta–binomial distribution for binary or count data. Potential additional future work includes extending the framework to other copula families, e.g., Student's t copula, or Bardossy's v-transform of the Gaussian copula (Bárdossy, 2006).

Our model was implemented in R and C++. The `RcppArmadillo` package (Eddelbuettel and Sanderson, 2014) was instrumental. Numerical optimizations were unconstrained and were carried out using R's `optim` function. The proper CAR areal GLMM was estimated using the `CARBayes` R package (Lee, 2013). Pseudocode for the localized IPS algorithm is presented in Appendix B.

Chapter 4

A Sparse Areal Mixed Model for Multivariate Outcomes, with an Application to Zero-Inflated Census Data

4.1 Introduction

In this chapter we develop a model for multivariate areal data. Existing models for multivariate areal outcomes face two formidable challenges: (1) spatial confounding (Clayton et al., 1993; Reich et al., 2006), which can lead to erroneous regression inference, and (2) an immense computational burden. Our approach addresses both of these challenges by extending the sparse areal mixed model (SAMM) of Hughes and Haran (2013).

The Master Address File (MAF) is a continually updated inventory of addresses for all known living quarters in the United States and Puerto Rico. It is used as a sampling frame for various Census Bureau surveys, including the decennial Census. As discussed in Young et al. (2016), there is interest in building census block-level models to (1) help us understand coverage errors resulting from a particular MAF extract, which is what defines a survey sampling frame, and (2) identify blocks with a large number of expected

housing unit 'deletes,' which could aid targeted address canvassing operations. Housing unit deletes are defined by the Census Bureau as addresses that were deleted from the MAF because they did not correspond to valid housing units (a house, an apartment, a mobile home, a group of rooms, or a single room that is occupied, or intended for occupancy, as a separate living area). The MAF data file comprises approximately 194 million records of housing unit addresses.

In this chapter we analyze a publicly available dataset called the Planning Database (PDB) (`http://goo.gl/LlcwY7`). These data include variables and counts from the 2010 Census and the 2009–13 American Community Survey (ACS). The data are aggregated at the block group level rather than the block level. A census block is the smallest geographic unit used by the Census Bureau. Blocks are typically bounded by streets or creeks/rivers. Within a city, a block corresponds to a city block. In a rural area, blocks may be large and irregularly shaped and bounded by features including roads, streams, or transmission lines. Census blocks are not delineated based on population. A block group, on the other hand, comprises multiple blocks and contains between 600 and 3,000 people. The PDB comprises approximately 220,000 block groups and 300 variables.

Our analysis is concerned with the 2010 Census. We are interested in revealing (potential) causes of address deletes. The block-group-level data are zero-inflated since the aggregated number of deletes has over 90% zeros across the entire US. Some zeros arise from those block groups in which changes in mailing addresses are unlikely. Another source of zeros is areas prone to redevelopment, in which case there is a high 'risk' of future address changes. We include various geographic, demographic, and operational variables from the PDB as predictors. Our work could lead to a more efficient and cost-effective decennial Census. We note that Young et al. (2016) developed zero-inflated models to reveal predictors of housing unit adds and deletes, but their data were from a particular MAF extract and their analysis was non-spatial. Failure to account for consequential spatial dependence can lead to erroneous inference (e.g., confidence intervals for regression coefficients may be too narrow).

The remainder of this chapter is organized as follows. In Section 4.2 we develop our multivariate SAMM and present the results from a simulation study. In Section 4.3 we specialize our model to handle zero-inflated data, and apply the resulting model

to simulated data and to zero-inflated Census data for the 2,600 block groups within the state of Iowa. We conclude the chapter in Section 4.4 with a summary of our work. We provide derivations, computational details, and additional simulation results in Appendix C.

4.2 Our sparse mixed model for multivariate areal data

In this section we develop our multivariate sparse areal mixed model (MSAMM). Our approach is similar to the method of Bradley et al. (2015) in that we, too, employ the orthogonal, multiresolutional spatial basis described by Hughes and Haran (2013) (see also Griffith (2003) and Tiefelsdorf and Griffith (2007)). This basis, known as the Moran basis (Moran, 1950), is appealing from a modeling point of view and also permits efficient computing.

To motivate our development of the MSAMM, we begin by reviewing conditional autoregressive (CAR) models for univariate areal data. The transformed conditional mean vector for these models is given by

$$g(\boldsymbol{\mu}) = g\left\{\mathbb{E}\left(\mathbf{y} \mid \boldsymbol{\beta},\, \boldsymbol{\phi}\right)\right\} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\phi}, \tag{4.1}$$

where g is a link function, $\mathbf{y} = (y_1, \ldots, y_n)'$ are the outcomes, the ith of which is associated with the ith areal unit, $\mathbf{X}$ is an $n \times p$ design matrix, $\boldsymbol{\beta}$ is a p-vector of regression coefficients, and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)'$ are spatially dependent random effects. Note that the outcomes are assumed to be independent conditional on the random effects. Marginally, however, the outcomes are spatially dependent because the random effects are spatially dependent.

The various CAR models are distinguished by the distribution assigned to $\boldsymbol{\phi}$. A common specification for $\boldsymbol{\phi}$ is the intrinsic CAR (Besag and Kooperberg, 1995):

$$p(\boldsymbol{\phi} \mid \tau) \propto \exp\left(-\frac{\tau}{2}\boldsymbol{\phi}'\mathbf{Q}\boldsymbol{\phi}\right),$$

where τ is a smoothing parameter and the precision matrix $\mathbf{Q}$ is equal to $\mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the diagonal matrix with the degrees (i.e., number of neighbors) of the areal units on its diagonal and $\mathbf{A}$ is the adjacency matrix for the underlying graph G, i.e., $\mathbf{A}$ is the binary matrix that encodes the adjacency structure among the areal units. Since

$\mathbf{Q}$ is singular, the intrinsic CAR is improper. A proper alternative has precision matrix $\mathbf{Q}(\rho) = \mathbf{D} - \rho\mathbf{A}$, where ρ is constrained to the interval $[0, 1)$. The parameter ρ can be considered a range parameter, but its effect on the marginal dependence structure is complex and often pathological (Wall, 2004; Assunção and Krainski, 2009). Additional proper CAR specifications exist. For instance, the prior proposed by Leroux et al. (2000) offers an alternate specification of the spatial random effects resulting in a proper prior distribution.

Traditional CAR models present serious challenges. First, there is often collinearity between the spatial random effects and the fixed-effects predictors. This characteristic, which is known as spatial confounding (Clayton et al., 1993), often leads to (1) biased estimation of regression coefficients and (2) substantial variance inflation that may make important covariates appear insignificant. Additionally, CAR random effects permit patterns of spatial repulsion, which we do not expect to observe in the types of data to which these models are usually applied. Spatial repulsion is also known as negative spatial autocorrelation (Boots and Tiefelsdorf, 2000) and occurs when neighboring areal units are mutually dissimilar (this is in contrast to spatial attraction, which occurs when neighboring areal units are mutually similar). Second, computation for CAR models can be extremely burdensome due to (1) the high dimensionality of $\boldsymbol{\phi}$ and (2) the nature of $\boldsymbol{\phi}$'s posterior distribution. It is well known that a univariate Metropolis–Hastings algorithm for sampling from the posterior distribution of $\boldsymbol{\phi}$ leads to a slow mixing Markov chain because the components of $\boldsymbol{\phi}$ exhibit strong *a posteriori* dependence. This has led to a number of approaches for block updating $\boldsymbol{\phi}$. Constructing proposals for these updates is challenging, and the faster mixing comes at the cost of increased running time per iteration—see, for instance, Knorr-Held and Rue (2002), Haran et al. (2003), and Haran and Tierney (2012).

To alleviate spatial confounding, eliminate patterns of spatial repulsion, and greatly reduce computing time and storage requirements, Hughes and Haran (2013) introduced their sparse areal mixed model (SAMM). The SAMM can be developed as follows.

Reich et al. (2006) showed that the traditional CAR models are spatially confounded in the sense that the random effects can 'pollute' the column space $C(\mathbf{X})$, which can lead to a biased and variance-inflated posterior for $\boldsymbol{\beta}$. To see this, first let $\mathbf{P} = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$ be the orthogonal projection onto $C(\mathbf{X})$, so that $\mathbf{I}_n - \mathbf{P}$ is the orthogonal projection

onto $C(\mathbf{X})^{\perp}$, where $C(\mathbf{X})^{\perp}$ is the orthogonal complement of $C(\mathbf{X})$. Now eigendecompose $\mathbf{P}$ and $\mathbf{I}_n - \mathbf{P}$ to obtain orthogonal bases $\mathbf{K}_{n \times p}$ and $\mathbf{L}_{n \times (n-p)}$ for $C(\mathbf{X})$ and $C(\mathbf{X})^{\perp}$, respectively. Then (4.1) can be rewritten as

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{K}\boldsymbol{\gamma} + \mathbf{L}\boldsymbol{\delta},$$

where $\boldsymbol{\gamma}_{p \times 1}$ and $\boldsymbol{\delta}_{(n-p) \times 1}$ are random coefficients. This form shows that $\mathbf{K}$ is the source of the confounding, for $\mathbf{K}$ and $\mathbf{X}$ have the same column space.

Since the columns of $\mathbf{K}$ are merely synthetic predictors (i.e., they have no scientific meaning), Reich et al. (2006) recommend removing them from the model. The resulting model (henceforth the RHZ model) has

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\delta},$$

so that spatial smoothing is restricted to the orthogonal complement of $C(\mathbf{X})$. In a subsequent paper, Hodges and Reich (2010) referred to this technique as restricted spatial regression (RSR).

RSR is not only an effective remedy for confounding but also speeds computing. Because the columns of $\mathbf{L}$ are orthogonal, the RHZ model's random effects are approximately *a posteriori* uncorrelated. This yields a fast mixing Markov chain, and the cost per iteration is reduced because a simple spherical normal proposal is sufficient for updating the random effects. But fitting the RHZ model to large areal datasets is still quite burdensome computationally because the random effects remain high dimensional.

By taking advantage of the underlying graph G, Hughes and Haran (2013) were able to greatly reduce the number of random effects while also improving regression inference. Hughes and Haran (2013) begin by defining the so called Moran operator for $\mathbf{X}$ with respect to G: $(\mathbf{I}_n - \mathbf{P})\mathbf{A}(\mathbf{I}_n - \mathbf{P})$. This operator appears in the numerator of a generalized form of Moran's I, a popular nonparametric measure of spatial dependence for areal data (Moran, 1950):

$$I_{\mathbf{X}}(\mathbf{A}) = \frac{n}{\mathbf{1}'\mathbf{A}\mathbf{1}} \frac{\mathbf{y}'(\mathbf{I}_n - \mathbf{P})\mathbf{A}(\mathbf{I}_n - \mathbf{P})\mathbf{y}}{\mathbf{y}'(\mathbf{I}_n - \mathbf{P})\mathbf{y}}.$$

Boots and Tiefelsdorf (2000) showed that (1) the (standardized) spectrum of a Moran operator comprises the possible values for the corresponding $I_{\mathbf{X}}(\mathbf{A})$, and (2) the eigenvectors comprise all possible mutually distinct patterns of clustering residual to $C(\mathbf{X})$

and accounting for G. The positive (negative) eigenvalues correspond to varying degrees of positive (negative) spatial dependence, and the eigenvectors associated with a given eigenvalue (h_k, say) are the patterns of spatial clustering that data exhibit when the dependence among them is of degree h_k.

In other words, the eigenvectors of the Moran operator form a multiresolutional spatial basis for $C(\mathbf{X})^\perp$ that exhausts all possible patterns that can arise on G. Since we do not expect to observe repulsion in the phenomena to which these models are usually applied, we can use the spectrum of the Moran operator to discard all repulsive patterns, retaining only attractive patterns for our analysis. By retaining only eigenvectors that exhibit positive spatial dependence, we can usually reduce the model dimension by at least half *a priori*. Hughes and Haran (2013) showed that a much greater reduction is usually possible in practice, with 50–100 eigenvectors being sufficient for most datasets.

Let $\mathbf{M}_{n\times q}$ contain the first $q \ll n$ eigenvectors of the Moran operator. Then the SAMM has first stage

$$g(\boldsymbol{\mu}) = \mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\delta}_s,$$

where $\boldsymbol{\delta}_s$ is a q-vector of random coefficients. The subscript 's' on $\boldsymbol{\delta}_s$ denotes sparse since $\boldsymbol{\delta}_s$ is modeled using the SAMM. Specifically, $\boldsymbol{\delta}_s$ is a Gaussian random vector having mean zero and precision matrix $\tau\mathbf{Q}_s$, where $\mathbf{Q}_s = \mathbf{M}'\mathbf{Q}\mathbf{M}$. This implies $p+q+1$ model parameters, compared to $p+n+1$ for the traditional CAR model and $p+(n-p)+1 = n+1$ for the RHZ model. This dramatic reduction in dimension speeds computation considerably, allowing even the largest areal datasets to be analyzed quickly (in minutes or hours rather than days or weeks).

4.2.1 Extending the SAMM to multivariate data

A number of multivariate CAR (MCAR) models have been developed (Carlin and Banerjee, 2003; Gelfand and Vounatsou, 2003; Jin et al., 2005; Martinez-Beneito, 2013). These models have the same drawbacks as their univariate counterparts but of course entail even more burdensome computation. Hence it is desirable to develop a SAMM for multivariate outcomes. We begin by reviewing the MCAR model that is the multivariate analog of the traditional univariate CAR model described above.

Suppose we observe multiple outcomes at each areal unit and that each outcome has its own regression component and collection of spatial effects. Specifically, for

$j \in \{1, \ldots, J\}$ we have outcomes $\mathbf{y}_j = (y_{1j}, \ldots, y_{nj})'$, design matrix $\mathbf{X}_j$, regression co-efficients $\boldsymbol{\beta}_j$, and spatial effects $\boldsymbol{\phi}_j = (\phi_{1j}, \ldots, \phi_{nj})'$. Then the transformed conditional mean vector is given by

$$g_j(\boldsymbol{\mu}_j) = \mathbf{X}_j \boldsymbol{\beta}_j + \boldsymbol{\phi}_j, \ j = 1, \ldots, J.$$

Now collect the $\boldsymbol{\phi}_j$ to form $\boldsymbol{\Phi} = (\boldsymbol{\phi}_1', \ldots, \boldsymbol{\phi}_J')'$, and specify the model

$$p(\boldsymbol{\Phi} \mid \boldsymbol{\Sigma}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{\Phi}'\left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{Q}\right)\boldsymbol{\Phi}\right\},$$

where $\boldsymbol{\Sigma}$ is a $J \times J$ covariance matrix consisting of $J(J+1)/2$ unknown variance and covariance parameters, and $\otimes$ denotes the Kronecker product. The jth diagonal entry of $\boldsymbol{\Sigma}$ is proportional to the variance of the spatial effects corresponding to the jth outcome. The jj' off-diagonal entry is proportional to the covariance between the jth and j'th spatial effects within an areal unit.

Should we require a different precision matrix for each $\boldsymbol{\phi}_j$, the prior on $\boldsymbol{\Phi}$ can be generalized as

$$p(\boldsymbol{\Phi} \mid \boldsymbol{\Sigma}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{\Phi}'\mathrm{bdiag}(\mathbf{R}_1, \ldots, \mathbf{R}_J)'(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n)\mathrm{bdiag}(\mathbf{R}_1, \ldots, \mathbf{R}_J)\boldsymbol{\Phi}\right\},$$

where $\mathrm{bdiag}(\cdot)$ denotes a block diagonal matrix and $\mathbf{R}_j$, $j = 1, \ldots, J$, is such that $\mathbf{R}_j'\mathbf{R}_j = \mathbf{Q}_j$.

Recently, Bradley et al. (2015) introduced the Moran's I (MI) prior, which is a multivariate spatiotemporal model based on the SAMM. We introduce a multivariate model that uses a similar prior but is strictly for spatial data. We call our model the multivariate SAMM (MSAMM). The MSAMM serves as the foundation for the zero-inflated count model that we focus on below in Section 4.3.

Construction of the MSAMM is of course analogous to construction of the SAMM. For $j \in \{1, \ldots, J\}$, let $\mathbf{P}_j = \mathbf{X}_j(\mathbf{X}_j'\mathbf{X}_j)^{-1}\mathbf{X}_j'$, and let $\mathbf{M}_j$ be a matrix the columns of which are the first q eigenvectors of $(\mathbf{I}_n - \mathbf{P}_j)\mathbf{A}(\mathbf{I}_n - \mathbf{P}_j)$. Denote the prior precision matrix as $\mathbf{Q}_{sj} = \mathbf{M}_j'\mathbf{Q}\mathbf{M}_j$, and let $\mathbf{R}_{sj}$ be the upper triangular Cholesky decomposition

of $\mathbf{Q}_{sj}$ so that $\mathbf{R}'_{sj}\mathbf{R}_{sj} = \mathbf{Q}_{sj}$. Then the MSAMM can be specified as

$$g_j(\boldsymbol{\mu}_j) = g_j \left\{ \mathbb{E}\left(\mathbf{y}_j \mid \boldsymbol{\beta}_j,\, \boldsymbol{\delta}_{sj}\right)\right\} = \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{M}_j\boldsymbol{\delta}_{sj},$$

$$p(\boldsymbol{\Delta} \mid \boldsymbol{\Sigma}) \propto \exp\left\{ -\frac{1}{2}\boldsymbol{\Delta}'\mathbf{R}'\left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_q\right)\mathbf{R}\boldsymbol{\Delta}\right\},$$

where $\boldsymbol{\Delta} = (\boldsymbol{\delta}'_{s1}, \ldots, \boldsymbol{\delta}'_{sJ})'$ and $\mathbf{R} = \mathrm{bdiag}\left(\mathbf{R}_{s1},\, \ldots,\, \mathbf{R}_{sJ}\right)$. Once again $\boldsymbol{\Sigma}$ is a $J \times J$ covariance matrix consisting of $J(J+1)/2$ unknown parameters. Thus the MSAMM has $J(p+q+(J+1)/2)$ unknowns, while the MCAR has $J(p+n+(J+1)/2)$ unknowns. This is a considerable reduction so long as $q \ll n$. Moreover, the MSAMM, like the SAMM, alleviates spatial confounding and permits simple, fast updating of the spatial random effects.

In our application and simulation study, we used the same design matrix for all multivariate outcomes, i.e., we used $\mathbf{X}_j = \mathbf{X}$ for all j, which implies a single prior precision matrix $\mathbf{Q}_s$. In this case, the above specification simplifies to

$$g_j(\boldsymbol{\mu}_j) \;=\; g_j \left\{ \mathbb{E}\left(\mathbf{y}_j \mid \boldsymbol{\beta}_j,\, \boldsymbol{\delta}_{sj}\right)\right\} = \mathbf{X}\boldsymbol{\beta}_j + \mathbf{M}\boldsymbol{\delta}_{sj},$$

$$p(\boldsymbol{\Delta} \mid \boldsymbol{\Sigma}) \;\propto\; \exp\left\{ -\tfrac{1}{2}\boldsymbol{\Delta}'\left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{Q}_s\right)\boldsymbol{\Delta}\right\}. \tag{4.2}$$

In either case the precision matrix is invertible, and so the prior distribution is proper.

A key difference between our model and that of Bradley et al. (2015) is our approach to accounting for dependence within an areal unit. The model of Bradley et al. (2015) 'recycles' the spatial component of the linear predictor. That is, the basis expansion $\mathbf{M}\boldsymbol{\delta}_s$ is shared across multivariate outcomes. Our approach is more flexible in that each multivariate outcome has its own spatial component, and we capture within-areal unit dependence by introducing $\boldsymbol{\Sigma}$.

4.2.2 Hyperpriors and computation for the MSAMM

We use independent Gaussian priors for the regression coefficients: $\boldsymbol{\beta}_j \overset{\mathrm{ind}}{\sim} \mathcal{N}(\mathbf{0},\, 10000\,\mathbf{I}_p)$. There are several appealing choices for the prior distribution of $\boldsymbol{\Sigma}$. We follow Huang and Wand (2013) and use the hierarchical half-t prior distribution. The hierarchical half-t prior relies on an inverse-Wishart distribution such that the diagonal elements

of the scale matrix are given inverse-gamma prior distributions. The prior scale of the inverse-Wishart is $2\nu\mathbf{V}$ and the prior degrees of freedom is $\nu + J - 1$. We let $\mathbf{V} = \mathrm{diag}\,(1/v_1, \ldots, 1/v_J)$, where v_j has an inverse-gamma prior with shape $1/2$ and scale $1/\zeta_j^2$. The hyperparameters are given values of $\nu = 2$ and $\zeta_j = 10^5$, $j = 1, \ldots, J$. This approach yields a conjugate prior distribution. Alternatives include the inverse-Wishart prior, the covariance matrix separation strategy (Barnard et al., 2000), and the LKJ prior (Lewandowski et al., 2009).

The full conditional distribution of the spatial effects $\mathbf{\Delta}$ does not have a closed form, so we use a Metropolis update to draw samples from the posterior. When a common design matrix is used across the multivariate outcomes, we can avoid the large number of computations involved with a Kronecker product, as implied by (4.2), by reparameterizing the spatial effects. For $\mathbf{\Psi} = \left(\boldsymbol{\psi}'_{s1}, \ldots, \boldsymbol{\psi}'_{sJ}\right)'$, with each $\boldsymbol{\psi}_{sj}$ $q \times 1$, suppose we have $\mathbf{\Psi} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma} \otimes \mathbf{I}_q)$. Let $\mathbf{R}_s$ be the upper triangular Cholesky decomposition of $\mathbf{Q}_s$ and let $\mathbf{L}_s = \mathbf{R}_s^{-1}$ so that $\mathbf{L}_s\mathbf{L}'_s = \mathbf{Q}_s^{-1}$. We show in Appendix C that $(\mathbf{I}_J \otimes \mathbf{L}_s)\mathbf{\Psi}$ and $\mathbf{\Delta}$ have the same distribution. For estimation with our model, we replace $\mathbf{M}$ with $\mathbf{M}_s = \mathbf{M}\mathbf{L}_s$ and work with $\mathbf{\Psi}$ instead of $\mathbf{\Delta}$. Derivations and details regarding sampling are provided in Appendix C. We also show that a similar reparameterization holds when the design matrices are not similar across multivariate outcomes.

4.2.3   Application of the MSAMM to simulated data

We designed and carried out a simulation study to assess the performance of the MSAMM. We began by simulating count data based on the 512 census block groups from the eight counties of the Des Moines–Ames–West Des Moines metropolitan statistical area (shown in Figure 4.1). For each block group, we simulated three outcomes, each having the same covariates and fixed effects: an intercept term and the proportion of the population within the block group that resided in a rural area (see Section 4.3.3). We set the fixed effects to $\boldsymbol{\beta} = (2, -1)'$ so that rural block groups had a decreased risk relative to urban block groups. We used $\mathbf{\Sigma}$ such that the spatial effects for the first, second, and third outcomes had variances of 6, 8, and 10, respectively, and correlations equal to 0.8.

We constructed $\mathbf{Q}_s$ as detailed in section 4.2.1 and used the same $\mathbf{Q}_s$ for each of the
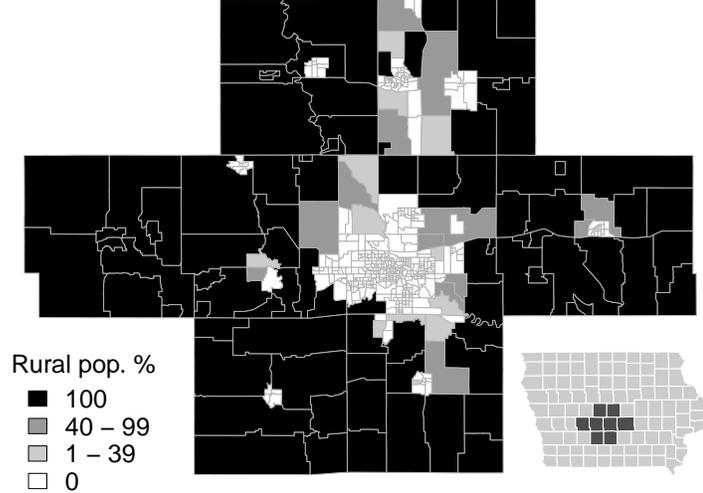
Figure 4.1: Des Moines–Ames–West Des Moines metropolitan statistical area (MSA) percent of population in a rural area at the block-group level. The inset shows the eight counties within the MSA.

three outcomes. Eigendecomposition of the Moran matrix yielded 200 positive eigenvalues, which implies 200 basis vectors exhibiting patterns of positive spatial dependence. We used the basis vectors corresponding to the 50 largest positive eigenvalues to construct $\mathbf{M}$. This choice of basis vectors allowed the outcomes to exhibit both large- and small-scale spatial variation. We then simulated spatial effects from a zero-mean Gaussian distribution with covariance $\boldsymbol{\Sigma} \otimes \mathbf{Q}_s^{-1}$. Finally, for the ith block group and jth outcome, we simulated a count from the Poisson distribution with rate $\exp(\mathbf{x}'_{ij}\boldsymbol{\beta}_j + \mathbf{m}'_{ij}\boldsymbol{\delta}_{sj})$. For our choice of $q = 50$, fitting the MSAMM required estimation of 162 parameters. The analogous proper MCAR has 1,548 parameters.

We applied our method to 1,000 simulated datasets. The results are shown in Table 4.1. In addition to fitting the MSAMM, we also applied three independent SAMMs, i.e., we ignored correlation among the spatial effects within an areal unit. Additional results, including credible interval coverage rates, are included in Appendix C. When an independence model is used, we see that the bias and MSE of the regression coefficient estimates are similar to the MSAMM. It is also clear that our choice of the half-t prior

Table 4.1: Results for simulation study comparing the MSAMM to independent SAMM fits. MSE denotes mean squared error.

Parameter	Truth	MSAMM			Independent SAMMs		
		Mean Est.	Bias	MSE	Mean Est.	Bias	MSE
β_{11}	2	2.000	<0.001	<0.001	2.000	<0.001	<0.001
β_{12}	-1	-0.999	<0.001	0.004	-0.999	<0.001	0.004
β_{21}	2	2.001	<0.001	<0.001	2.000	<0.001	<0.001
β_{22}	-1	-1.001	-0.001	0.004	-1.001	<0.001	0.004
β_{31}	2	2.000	<0.001	<0.001	1.999	<0.001	<0.001
β_{32}	-1	-1.003	-0.003	0.004	-1.004	-0.004	0.004
σ_1^2	6	5.946	-0.054	1.657	6.136	0.136	1.902
σ_2^2	8	7.901	-0.099	2.769	8.194	0.194	3.127
σ_3^2	10	10.025	0.025	4.791	10.436	0.436	5.486
ρ_{12}	0.8	0.806	0.006	0.003	–	–	–
ρ_{13}	0.8	0.804	0.004	0.003	–	–	–
ρ_{23}	0.8	0.808	0.008	0.003	–	–	–

on the covariance matrix results in excellent recovery of the variance and correlation parameters. The left column of Figure 4.2 displays simulated counts for the 3 outcomes for a single simulated dataset. Based on our choice of fixed effects, $\boldsymbol{\beta} = (2, -1)'$, rural areas (larger block groups) have lower counts than more urban areas (smaller block groups). The right column of Figure 4.2 displays the fitted values $\exp(\mathbf{x}'_{ij}\hat{\boldsymbol{\beta}}_j + \mathbf{m}'_{ij}\hat{\boldsymbol{\delta}}_{sj})$.

4.3   An MSAMM for zero-inflated data

In this section we specialize the MSAMM to handle zero-inflated count data. Zero-inflated count models are often applied in non-spatial settings, e.g., in manufacturing (Lambert, 1992), where defective materials are rare and the number of defects is assumed to follow a Poisson distribution, and in the hunger-for-bonus phenomenon that occurs in risk assessment for filed insurance claims (Boucher et al., 2009). Spatial zero-inflated count models have been applied to various types of data, including animal sightings (Agarwal et al., 2002; Ver Hoef and Jansen, 2007; Recta et al., 2012), plant distribution

Observed Counts   Model Fits
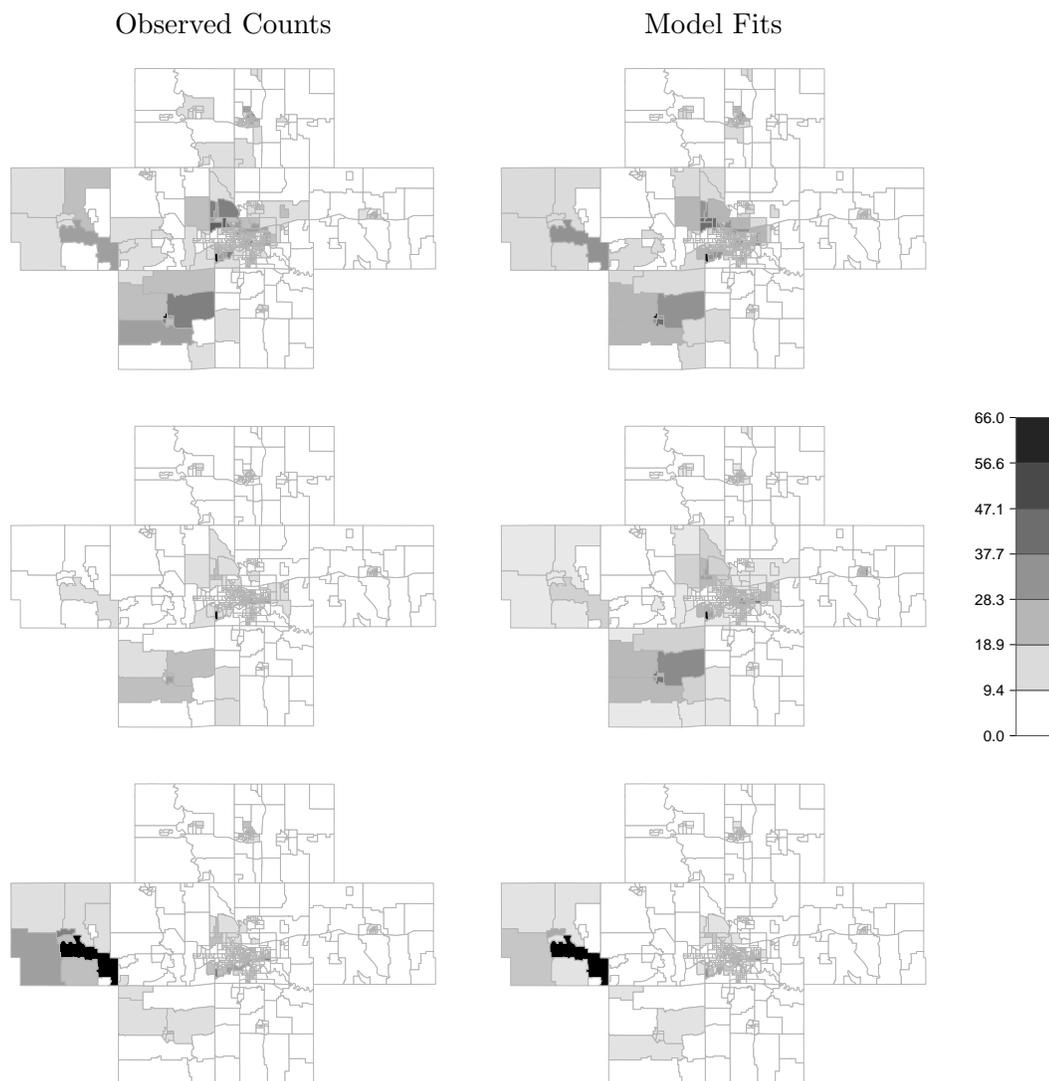
Figure 4.2: Simulated counts and fitted values for a single simulated dataset. Row j shows counts (left column) and corresponding fitted values (right column) for the jth outcome. The within-unit spatial effects have correlation 0.8 for each pair of outcomes.

(Rathbun and Fei, 2006), tornado reports (Wikle and Anderson, 2003), and emergency room visits (Neelon et al., 2013).

4.3.1 Models for zero-inflated areal counts

Two common approaches to modeling zero-inflated counts are the hurdle model and the zero-inflated-Poisson (ZIP) model (Lambert, 1992). For a hurdle model, the outcome is 0 with probability $1-\pi$, and with probability π the outcome arose from a zero-truncated Poisson (ZTP) distribution (Cohen, 1960; Singh, 1978). Formally, the hurdle model is a pseudo-mixture of the form

$$\mathbb{P}(y=0) = 1 - \pi,$$
$$\mathbb{P}(y=k) = \pi \frac{\exp(-\lambda)}{1 - \exp(-\lambda)} \frac{\lambda^k}{k!} \quad (k \in \mathbb{N} : k \geq 1).$$

The ZIP model, on the other hand, is a mixture of the form

$$\mathbb{P}(y=0) = (1 - \pi) + \pi \exp(-\lambda),$$
$$\mathbb{P}(y=k) = \pi \frac{\lambda^k}{k!} \exp(-\lambda) \quad (k \in \mathbb{N}).$$

Let us compare and contrast the hurdle and ZIP models informally. Each model can be viewed as a bivariate model, where one set of observations is binary and the other is a count. For the hurdle model, there is only one source of zeros, namely, the binary process. If the binary outcome is 0, no count is observed. If the binary outcome is 1, a nonzero count is observed. The ZIP model differs in that it posits two sources of zeros. If the binary outcome is 0, no count is observed. If the binary outcome is 1, a (possibly 0) count is observed. This difference has a practical effect: the hurdle model can accommodate both zero-inflation and zero-deflation while the ZIP model can accommodate only zero-inflation (Neelon et al., 2013).

It may also be the case that one or the other of the models makes more sense given the phenomenon of interest. Consider emergency room visits, for example, as in Neelon et al. (2013). A hurdle model is appropriate here; the binary outcome registers whether a given subject visits an emergency room while the count outcome records the number of visits. As a second example, consider animal sightings (Agarwal et al., 2002; Recta et al., 2012). An animal species may not be sighted in a given region even though the

animal is likely present in the region. A ZIP model is appropriate in this scenario. We adopt the hurdle model for the remainder of the chapter.

In the context of our MSAMM, an appropriate hurdle model can be specified as follows. Suppose we have n areal units, in which case our outcomes are $\mathbf{y}_j = (y_{1j}, \ldots, y_{nj})'$, where $j \in \{1, 2\}$ indexes the binary and count components, respectively. For the sake of simplicity, suppose that $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}$. This implies an MSAMM of the form given in (4.2) (as we showed in Section 4.2.1, this model can easily be extended to accommodate different design matrices). Now let the linear predictor for the ith areal unit and the jth outcome be defined as

$$\eta_{ij} = \mathbf{x}_i'\boldsymbol{\beta}_j + \mathbf{m}_i'\boldsymbol{\delta}_{sj}.$$

If we take link functions g_1 and g_2 to be the logit and log functions, respectively, we have

$$\pi_i = \frac{\exp(\eta_{i1})}{1 + \exp(\eta_{i1})},$$
$$\lambda_i = \exp(\eta_{i2}),$$

where π_i is the probability of incidence, i.e., $\mathbb{P}(y_{i1} = 1)$, and λ_i is the ZTP rate for y_{i2}. The within-unit covariance matrix $\boldsymbol{\Sigma}$ is of course 2×2 for this model.

4.3.2 Simulation study

To assess the performance of our areal hurdle model, we carried out a simulation study. We simulated data for the 2,600 census block groups of the US state of Iowa (Figure 4.3). We included an intercept term and, as a covariate, the percentage of housing units occupied by renters (see Section 4.3.3). We used $\boldsymbol{\beta}_1 = (-1, 1)'$ and $\boldsymbol{\beta}_2 = (2, -1)'$. These values for $\boldsymbol{\beta}_1$ indicate that block groups with a high percentage of renters will be more likely to take non-zero values than block groups with a low percentage of renters. The values of $\boldsymbol{\beta}_2$ indicate that as the proportion of renters increases, the ZTP rate decreases. We used 4 and 8 as the diagonal elements of $\boldsymbol{\Sigma}$, and we used five different values for the within-unit correlation: $\rho = 0, 0.2, 0.4, 0.6, 0.8$.

We constructed $\mathbf{Q}_s$ as detailed in Section 4.2.1 and used the same $\mathbf{Q}_s$ for each outcome. Eigendecomposition of the Moran operator yielded 1,000 basis vectors exhibiting

patterns of positive spatial dependence. We used the first $q = 250$ eigenvectors to construct $\mathbf{M}$. This choice of q allowed the responses to exhibit both small- and large-scale spatial variation. We then simulated spatial effects $\boldsymbol{\Delta} = (\boldsymbol{\delta}'_{s1}, \boldsymbol{\delta}'_{s2})'$ from a zero-mean Gaussian distribution with covariance $\boldsymbol{\Sigma} \otimes \mathbf{Q}_s^{-1}$. For the ith block group we simulated y_{i1} from the Bernoulli distribution with success probability $\pi_i = \text{logit}^{-1}(\mathbf{x}'_i\boldsymbol{\beta}_1 + \mathbf{m}'_i\boldsymbol{\delta}_{s1})$. Finally, conditional on $y_{i1} = 1$, we drew y_{i2} from the truncated Poisson distribution with rate $\lambda_i = \exp(\mathbf{x}'_i\boldsymbol{\beta}_2 + \mathbf{m}'_i\boldsymbol{\delta}_{s2})$.

We analyzed 1,000 simulated datasets for each of the five correlations. To assess the importance of modeling the dependence within areal units, we applied both the MSAMM and independent SAMMs to each dataset. Key results are shown in Table 4.2 and Table 4.3. Additional results, including credible interval coverage rates, are included in Appendix C. We see that neglecting within-unit dependence leads to larger biases and, for some parameters, larger mean squared errors. Again, our choice of the half-t prior on the covariance matrix results in excellent recovery of the variance and correlation parameters.

4.3.3 Application to the Iowa data

In this section we apply our areal hurdle model for zero-inflated count data to address deletes from the 2010 US Census within the state of Iowa. Recall that a delete is defined as an address that was deleted from the base count because it did not correspond to a valid housing unit. Figure 4.4 displays a histogram (truncated at 10) showing the zero-inflation present in the data.

In the count component of the model, along with covariates, we introduced an offset calculated via external standardization. Specifically, the offset is the natural log of the total number of addresses in the block group to which a Census form was delivered. We used the same covariates in the binary and count components of the model. Each covariate was a proportion. We used the following covariates and an intercept.

- RURAL_POP: proportion of population living outside of an urban area or urban cluster

- OCCP_HU: proportion of housing units classified as the usual place of residence of the individual or group living there
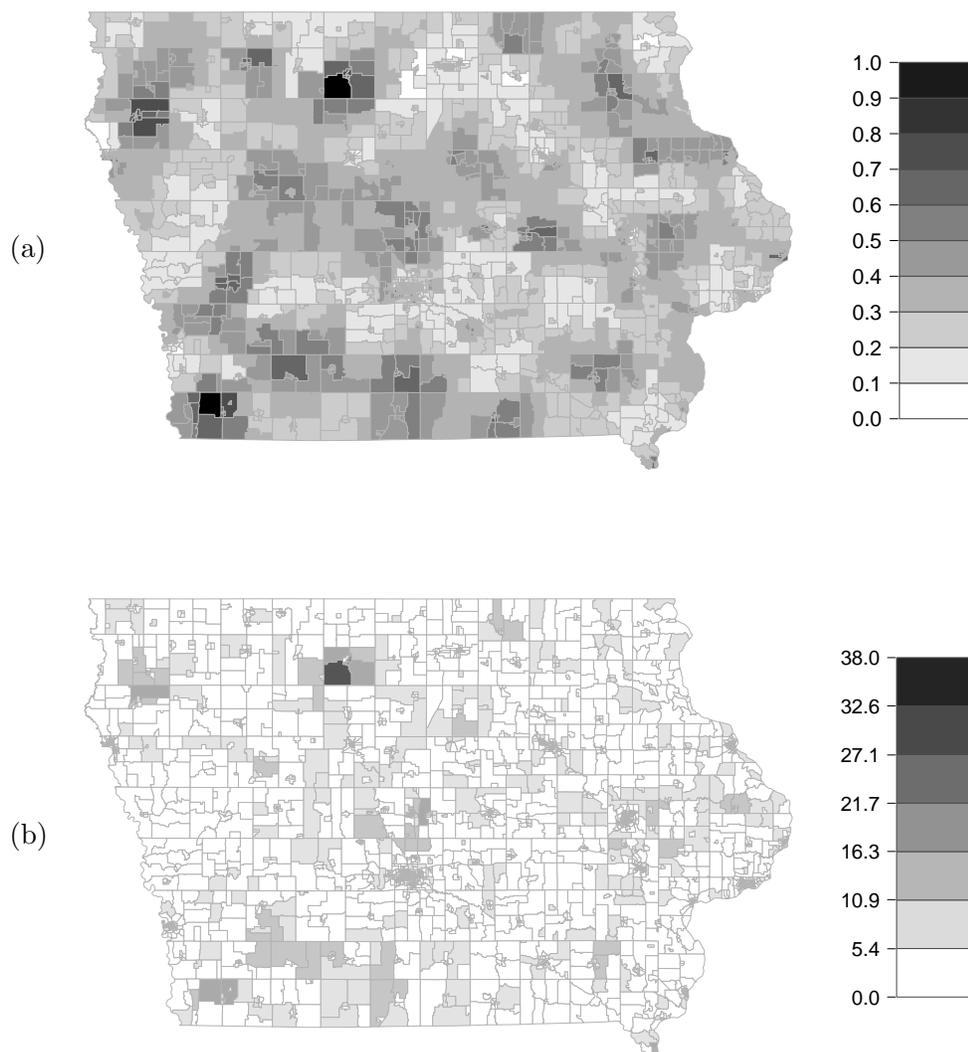
Figure 4.3: A single simulated zero-inflated dataset for the census block groups of Iowa. The proportion of renters in each block group was used as a covariate. Panel (a) displays the probabilities of non-zero counts. Panel (b) displays the ZTP rates, where a given rate is non-zero only if the underlying binary outcome is equal to 1.

Table 4.2: Results for the Iowa simulation study. We analyzed 1,000 simulated zero-inflated datasets for small to moderate correlation values ($\rho = 0$, 0.2, 0.4). MSE denotes mean squared error.

Parameter	Truth	MSAMM			Independent SAMMs		
		Mean Est.	Bias	MSE	Mean Est.	Bias	MSE
β_{11}	-1	-0.998	0.002	0.006	-0.995	0.005	0.006
β_{12}	1	0.994	-0.006	0.058	0.991	-0.009	0.057
β_{21}	2	2.003	0.003	0.001	2.003	0.003	< 0.001
β_{22}	-1	-1.001	-0.001	0.009	-1.001	-0.001	0.009
σ_1^2	4	3.970	-0.030	1.357	3.703	-0.297	1.395
σ_2^2	8	8.274	0.274	0.939	8.153	0.153	0.847
ρ	0	< 0.001	< 0.001	0.014	–	–	–
β_{11}	-1	-0.998	0.002	0.006	-0.995	0.005	0.006
β_{12}	1	0.994	-0.006	0.058	0.991	-0.009	0.057
β_{21}	2	2.004	0.004	< 0.001	2.012	0.012	0.001
β_{22}	-1	-1.000	< 0.001	0.008	-1.003	-0.003	0.008
σ_1^2	4	4.009	0.009	1.336	3.703	-0.297	1.395
σ_2^2	8	8.248	0.248	0.933	8.112	0.112	0.838
ρ	0.2	0.201	0.001	0.013	–	–	–
β_{11}	-1	-0.998	0.002	0.006	-0.995	0.005	0.006
β_{12}	1	0.994	-0.006	0.058	0.991	-0.009	0.057
β_{21}	2	2.004	0.004	< 0.001	2.020	0.020	0.001
β_{22}	-1	-1.000	< 0.001	0.008	-1.004	-0.004	0.008
σ_1^2	4	4.012	0.012	1.345	3.703	-0.297	1.395
σ_2^2	8	8.229	0.229	0.945	8.071	0.071	0.836
ρ	0.4	0.405	0.005	0.012	–	–	–

Table 4.3: Results for the Iowa simulation study. We analyzed 1,000 simulated zero-inflated datasets for moderately strong to strong correlation values ($\rho = 0.6$, 0.8). MSE denotes mean squared error.

| Parameter	Truth	MSAMM Mean Est.	Bias	MSE	Independent SAMMs Mean Est.	Bias	MSE
β_{11}	-1	-0.998	0.002	0.006	-0.995	0.005	0.006
β_{12}	1	0.995	-0.005	0.058	0.991	-0.009	0.057
β_{21}	2	2.005	0.005	<0.001	2.028	0.028	0.002
β_{22}	-1	-1.000	<0.001	0.007	-1.007	-0.007	0.007
σ_1^2	4	4.060	0.060	1.309	3.703	-0.297	1.395
σ_2^2	8	8.238	0.238	0.943	8.018	0.018	0.827
ρ	0.6	0.604	0.004	0.010	–	–	–
β_{11}	-1	-1.000	< 0.001	0.006	-0.995	0.005	0.006
β_{12}	1	0.996	-0.004	0.058	0.991	-0.009	0.057
β_{21}	2	2.007	0.007	< 0.001	2.036	0.036	0.002
β_{22}	-1	-1.000	< 0.001	0.007	-1.008	-0.008	0.007
σ_1^2	4	4.165	0.165	1.263	3.703	-0.297	1.395
σ_2^2	8	8.197	0.197	0.940	7.944	-0.056	0.824
ρ	0.8	0.793	-0.007	0.006	–	–	–
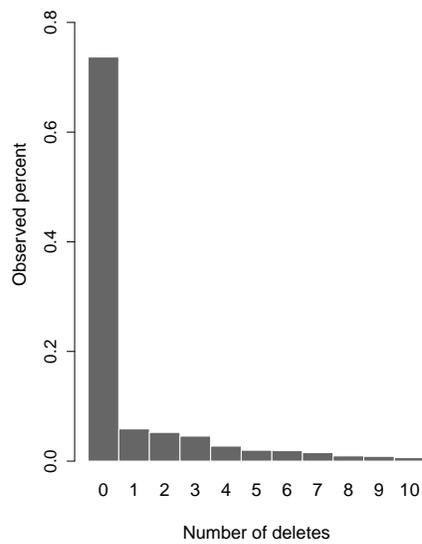
Figure 4.4: Histogram of number of address deletes for the 2,600 block groups of Iowa from the 2010 Census. Approximately 75% of the outcomes are zeros and 4% of the outcomes have greater than 10 counts.

- RENTER_OCCP_HU: proportion of occupied housing units that are not owner occupied, whether they are rented or occupied without payment of rent

- TEA_MAIL: proportion of addresses that received a Census form in the mail and occupants were instructed to complete and return the form

- FIRST_FRM: proportion of addresses where the first form mailed was completed and returned

As in the simulation study, we used the first 250 eigenvectors of the Moran operator, and we applied both the MSAMM and independent SAMMs.

The results are shown in Table 4.4. First, we see that the correlation ρ was estimated as 0.73, with a 95% credible interval of $(0.25, 0.91)$. This suggests that the MSAMM, as opposed to independent SAMMs, is appropriate for these data. This claim is further supported by the fact that the correlation model has a lower deviance information criterion (DIC) (Spiegelhalter et al., 2002) (-7745 versus -7735).

Regarding the regression coefficients for the binary component, we see that all but OCCP_HU have 95% credible intervals that exclude zero. This implies that, in Iowa, the proportion of occupied housing units is not predictive of the occurrence of deletes within a block group. We see that RENTER_OCCP_HU, TEA_MAIL, and FIRST_FRM offer a 'protective effect' against the occurrence of deletes while RURAL_POP is associated with an increased likelihood of deletes. We also see that the variation in the spatial effects for the binary component appears to be small: $\hat{\sigma}_1^2 = 0.67 \, (0.30, 2.04)$.

The regression for the count component tells a somewhat different story. All coefficients save RURAL_POP have 95% credible intervals that exclude zero, implying that the proportion of a block group that is rural is not associated with the number of deletes. Similar to the binary component, we see that RENTER_OCCP_HU, TEA_MAIL, and FIRST_FRM offer a 'protective effect' against the number of deletes, but now OCCP_HU is associated with a greater number of deletes. Evidently, the spatial process is not as smooth for the count component since $\hat{\sigma}_2^2 = 7.74 \, (4.99, 11.47)$.

The results obtained using our areal hurdle model could be valuable to the Census Bureau. When using the most recent covariate values obtained (such as from official government surveys or administrative records), predictions using our model could help to characterize deletes, which is an indication of demographic change or stability in an area.

The spatial component of our model can assist in designing efficient and cost-effective address updating operations. For example, it can inform Census Bureau personnel as to *clusters* of block groups that are candidates for updating in a non-decennial Census setting. Focusing on adjacent regions within a cluster will be advantageous over assessing sets of block groups that might have only 'stable' block groups as neighbors. Such clustering will not always be captured accurately by non-spatial zero-inflated models.

4.4   Discussion

Our proposed methods for handling multivariate and zero-inflated areal data have been demonstrated to effectively reduce the dimensionality of the spatial random effects leading to a reduction in the computational complexity of the spatial models. As with a univariate setting, the SAMM applied in a multivariate setting will alleviate spatial confounding. The simulation studies illustrate the benefit of accounting for correlation between spatial effects for each outcome. The primary benefit is a reduction in bias and mean-squared error of the fixed effects estimators. Additional simulation study results are given in Appendix C and show the 95% coverage probabilities for all model parameters. Table C.1 shows additional results for the multivariate simulation study. Focusing on the fixed effects, we see similar coverage probabilities between the correlation and independent models. Tables C.2 and C.3 display the results of the zero-inflated simulation study, highlighting the importance of accounting for correlation between spatial effects when it is indeed present. In particular, the coverage probabilities of the fixed effects for the independent model decrease as the between spatial effects correlation increases. The application to zero-inflated Census data highlighted the dramatic dimension reduction of the parameter space that is possible while maintaining an efficient estimation scheme. Though similar inference is achieved for the correlation or independent models, the correlation model offers a marginally superior fit as indicated by the DIC values presented in Table 4.4.

For each of the simulation studies and the application to Census data, we used the fixed-width approach to Markov chain convergence diagnosis (Flegal et al., 2008), in which samples are drawn until all Monte Carlo standard errors are smaller than some pre-selected threshold. We used the `batchmeans` package (Haran and Hughes, 2015) in

Table 4.4: Iowa address–delete results for the MSAMM versus independent SAMMs. Results for the binary components of the models are shown in the top portion of the table. Results for the count components are shown in the bottom portion. CI denotes credible interval.

Predictor/Parameter	MSAMM Posterior Mean	95% CI	Independent SAMMs Posterior Mean	95% CI
Intercept	7.23	(4.43, 9.83)	7.48	(4.93, 10.53)
RURAL_POP	2.72	(2.18, 3.30)	2.70	(2.13, 3.29)
OCCP_HU	0.68	(-2.68, 4.60)	0.25	(-3.21, 3.61)
RENTER_OCCP_HU	-3.51	(-5.73, -1.41)	-3.60	(-6.01, -1.16)
TEA_MAIL	-6.95	(-7.54, -6.37)	-6.98	(-7.60, -6.40)
FIRST_FRMS	-7.10	(-9.93, -4.28)	-6.86	(-9.82, -3.86)
σ_1^2	0.67	(0.30, 2.04)	0.94	(0.21, 2.52)
Intercept	3.99	(3.50, 4.50)	3.92	(3.07, 4.69)
RURAL_POP	0.00	(-0.18, 0.15)	0.02	(-0.14, 0.18)
OCCP_HU	2.52	(1.07, 3.59)	2.00	(0.48, 3.17)
RENTER_OCCP_HU	-1.47	(-2.28, -0.68)	-1.16	(-1.99, -0.16)
TEA_MAIL	-2.03	(-2.22, -1.83)	-1.98	(-2.24, -1.73)
FIRST_FRMS	-6.11	(-7.39, -4.65)	-5.50	(-7.00, -2.93)
σ_2^2	7.74	(4.99, 11.47)	8.24	(5.37, 12.01)
ρ	0.73	(0.25, 0.91)	–	
pD	258.85		172.31	
DIC	-7745		-7735	

R to compute Monte Carlo standard errors. Two hundred thousand samples were sufficient to ensure that all Monte Carlo standard errors were less than 0.05. The MCMC estimates stabilized after approximately 100,000 iterations. In our application to zero-inflated count data, computations were carried out using a PC desktop with an i7-3770, 3.4 GHz processor. We found that drawing 200,000 samples from the posterior distributions required approximately 45 minutes of computing time with our MSAMM model. Similarly, the univariate SAMMs required approximately 45 minutes of computing time as well. In comparison, the MCAR models of Section 4.2.1 require in excess of 100 minutes of computing time.

The multivariate Poisson and zero-inflated models were each programmed using R (Ihaka and Gentleman, 1996) and C++ (Eddelbuettel and Francois, 2011). Most of the numerical linear algebra was carried out using the Armadillo C++ library (Sanderson, 2010; Eddelbuettel and Sanderson, 2014). Each of the models presented here will be available in the next version of R package `ngspatial` (Hughes and Cui, 2014). The package will support Gaussian and Bernoulli outcomes as well as Poisson and zero-inflated outcomes.

Chapter 5

Conclusion and Discussion

5.1 Methodological advances

The work presented in this dissertation showcases several methodological advances in modeling areal data. In Chapters 2 and 4, we used a reparameterized conditional autoregressive (CAR) prior, the sparse areal mixed model (SAMM), to reduce the dimensionality of the parameter space and to alleviate bias and variance inflation arising from spatial confounding. The SAMM uses spatial effects that are orthogonal to the fixed effects, which greatly reduces the bias and variance inflation (Hughes and Haran, 2013). In Chapter 3 we presented a new copula-based marginal model that accounts for spatial dependence via a single unknown parameter, the first-order correlation between neighboring areal units. Because we used a marginal model, bias of the fixed effects due to spatial confounding is reduced. Thus, this dissertation work has achieved several improvements over the deficiencies of previous models that are used for the analysis of areal data.

We began our work with Chapter 2 where we developed a novel approach to spatial Bayesian variable selection for temporally correlated regressions located on a lattice, with an emphasis on applications to fMRI studies. Though we employed an AR(2) model to account for within-voxel temporal correlation, our model has the flexibility to handle up to an AR(5) if so desired. Previous methods suffered from extreme computational burdens that severely limited their ability to account for within-voxel temporal correlation in a similar way. We introduced the notion of parcellating the brain and

found that inference appears insensitive to our parcellation technique, and that the joint use of the parcellation and the SAMM greatly reduces the computational burden relative to analysis of a full dataset.

With Chapter 3, we presented copCS, a new hierarchical, copula-based regression model for areal data, and developed an efficient computational framework for inference. The foundation of our approach is covariance selection implemented via iterative proportional scaling (IPS) to quickly compute correlation matrices. By appropriately specifying latent variables, and in contrast to competing CAR-based methods for the analysis of areal data, copCS offers an intuitive marginal interpretation of the regression coefficients as well as a clear understanding of the spatial dependence parameter as the first-order correlation.

In Chapter 4, we proposed methods for handling multivariate and zero-inflated areal data and demonstrated an effective reduction in the dimensionality of the spatial random effects leading to a reduction in the computational complexity of these spatial models. As in a univariate setting, the SAMM applied in a multivariate setting reduces spatial confounding resulting in a reduction in bias and variance inflation of the estimators. As a desirable side effect, the SAMM reduces the dimensionality of the parameter space. This effect is highlighted with the zero-inflated application to Census data where we found that we can dramatically reduce the parameter space while maintaining an efficient estimation scheme.

5.2   Computational advances and software development

In the process of implementing our methodological advances in each of Chapters 2, 3, and 4, it was necessary to overcome several computational challenges. With Chapter 2, the large number of voxels present in a single brain represents a large computational burden. Our use of a parcellation algorithm, as well as the SAMM, allowed us to quickly apply a fully Bayesian model. In Chapter 3, a computational difficulty arises from the need to quickly compute correlation matrices with a conditional dependence structure. Conditional dependence requires that the inverse-correlation matrix has a zero entry whenever the underlying graph vertex is not connected. To accommodate this restriction, we modified the IPS algorithm to handle correlation matrices and carry

out inference on large datasets relatively quickly. Finally, our work in Chapter 4 can be viewed as extensions to previous multivariate models that suffered unfavorable computing drawbacks. By adapting the SAMM to multivariate and zero-inflated settings, we were able to effectively reduce the parameter space resulting in improved computing speeds.

The successful implementation of a model in an applied setting relies on the availability of software. New statistical methods live (or die) based on the accessibility of software that implements the new methodology. Each of the models proposed in this dissertation were implemented using R and C++ and can be accessed via the author's Github repository located at `http://www.github.com/donaldmusgrove/thesis`. As of this writing, work is under way to implement each of the proposed models as R packages available via the Comprehensive R Archive Network (CRAN) (Ihaka and Gentleman, 1996).

5.3   Future directions for research

Similar to the work that preceded our methods development, our work can be advanced in several ways. Our fMRI modeling procedure of Chapter 2 can easily be extended to handle multiple subjects with the end result being a group-level posterior probability map. In Chapter 3, the correlation matrix could be used as part of a prior distribution in a Bayesian setting. For instance, we could use our IPS procedure to produce the correlation matrix as part of the separation strategy of Barnard et al. (2000).

As made evident by our work in Chapter 4, future work with areal data ought to focus on extending models to extremely large datasets. For example, the US Census Bureau has freely available data aggregated at a fine level (block groups) resulting in hundreds of thousands of observations across the US. Using an approach similar to our parcellation technique of Chapter 2, we can easily construct and implement models that effectively account for spatial dependence. To illustrate, the state of New York has 62 counties and approximately 15,000 block groups. Ignoring counties, using the SAMM would require high dimensional matrix multiplications and the eigendecomposition of a $15,000 \times 15,000$ matrix, both of which are computationally burdensome procedures. An alternate approach could partition the block groups by county. The number of block

groups within each county ranges from a minimum of 8 to a maximum of 2049. Thus, we can construct within-county block group adjacency matrices very quickly, and 'link' the counties by the 62×62 between-county adjacency matrix. Thus, we have spatial effects at both the county level and the block group level, effectively reducing the dimension of the problem to a more manageable scale.

References

Agarwal, D. K., Gelfand, A. E., and Citron-Pousty, S. (2002). Zero-inflated models with application to spatial count data. *Environmental and Ecological Statistics* **9,** 341–355.

Assunção, R. and Krainski, E. (2009). Neighborhood dependence in Bayesian spatial models. *Biometrical Journal* **51,** 851–869.

Bandettini, P. A. (2009). Seven topics in functional magnetic resonance imaging. *Journal of Integrative Neuroscience* **8,** 371–403.

Bárdossy, A. (2006). Copula-based geostatistical models for groundwater quality parameters. *Water Resources Research* **42,** W11416.

Barnard, J., McCulloch, R., and Meng, X. L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica* **10,** 1281–1312.

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

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* **43,** 1–20.

Boots, B. and Tiefelsdorf, M. (2000). Global and local spatial autocorrelation in bounded regular tessellations. *Journal of Geographical Systems* **2,** 319–348.

Boucher, J. P., Denuit, M., and Guillen, M. (2009). Number of accidents or number of claims? An approach with zero-inflated Poisson models for panel data. *The Journal of Risk and Insurance* **76,** 821–846.

Bowman, F. D. (2007). Spatiotemporal models for region of interest analyses of functional neuroimaging data. *Journal of the American Statistical Association* **102,** 442–453.

Bradley, J. R., Holan, S. H., and Wikle, C. K. (2015). Multivariate spatio-temporal models for high-dimensional areal data with application to longitudinal employer-household dynamics. *The Annals of Applied Statistics* **9,** 1761–1791.

Bron, C. and Kerbosch, J. (1973). Algorithm 457: Finding all cliques of an undirected graph. *Communications of the Association for Computing Machinery* **16,** 575–577.

Byrd, R., Lu, P., Nocedal, J., and Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *Society for Industrial and Applied Mathematics Journal on Scientific Computing* **16,** 1190–1208.

Carey, V., Long, L., and Gentleman, R. (2015). *RBGL: An interface to the BOOST graph library.* R package version 3.0.

Carlin, B. P. and Banerjee, S. (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion). In Bayarri, M., Berger, J., Bernardo, J., Dawid, A., Heckerman, D., Smith, A., and West, M., editors, *Bayesian Statistics 7.* Oxford University Press.

Clayton, D., Bernardinelli, L., and Montomoli, C. (1993). Spatial correlation in ecological analysis. *International Journal of Epidemiology* **22,** 1193–1202.

Cohen, A. C. (1960). Estimating the parameter in a conditional Poisson distribution. *Biometrics* **16,** 203–211.

Cuthill, E. and McKee, J. (1969). Reducing the bandwidth of sparse symmetric matrices. In *Proceedings of the 1969 24th National Conference*, pages 157–172. Association for Computing Machinery.

Deming, W. E. and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics* **11,** 427–444.

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

Dobra, A. and Lenkoski, A. (2011). Copula Gaussian graphical models and their application to modeling functional disability data. *The Annals of Applied Statistics* **5,** 969–993.

Eddelbuettel, D. and Francois, R. (2011). Rcpp: Seamless R and C++ integration. *Journal of Statistical Software* **40,** 1–18.

Eddelbuettel, D. and Sanderson, C. (2014). RcppArmadillo: Accelerating R with high-performance C++ linear algebra. *Computational Statistics and Data Analysis* **71,** 1054–1063.

Flegal, J. M., Haran, M., and Jones, G. L. (2008). Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science* **23,** 250–260.

Friston, K., Ashburner, J., Frith, C., Poline, J., Heather, J., and Frackowiak, R. (1995). Spatial registration and normalization of images. *Human Brain Mapping* **2,** 165–189.

Gelfand, A. E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* **4,** 11–15.

Gelman, A. and Meng, X. L. (1998). Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statistical Science* **13,** 163–185.

Genest, C. and Neslehova, J. (2007). A primer on copulas for count data. *Astin Bulletin* **37,** 475–515.

George, E. and McCulloch, R. (1997). Approaches for Bayesian variable selection. *Statistica Sinica* **7,** 1126–1141.

Griffith, D. A. (2003). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding Through Theory and Scientific Visualization.* Springer, Berlin.

Haran, M., Hodges, J. S., and Carlin, B. P. (2003). Accelerating computation in Markov random field models for spatial data via structured MCMC. *Journal of Computational and Graphical Statistics* **12,** 249–264.

Haran, M. and Hughes, J. (2015). *batchmeans: Consistent batch means estimation of Monte Carlo standard errors.* Minneapolis, MN. R package version 1.0-2.

Haran, M. and Tierney, L. (2012). On automating Markov chain Monte Carlo for a class of spatial models. *arXiv preprint arXiv:1205.0499* .

Hariri, A., Tessitore, A., Mattay, V., Fera, F., and Weinberger, D. (2002). The amygdala response to emotional stimuli: A comparison of faces and scenes. *NeuroImage* **17,** 317–323.

Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical Science* **1,** 297–310.

Henn, L. (2015). *Challenges in the analysis of data with clustering or correlation.* PhD thesis.

Hodges, J. S. and Reich, B. J. (2010). Adding spatially-correlated errors can mess up the fixed effect you love. *The American Statistician* **64,** 325–334.

Holmes, C. C. and Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis* **1,** 145–168.

Huang, A. and Wand, M. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis* **8,** 439–452.

Hughes, J. (2015). copCAR: A flexible regression model for areal data. *Journal of Computational and Graphical Statistics* **24,** 733–755.

Hughes, J. and Cui, X. (2014). *ngspatial: Fitting the centered autologistic and sparse spatial generalized linear mixed models for areal data.* Minneapolis, MN. R package version 1.0-4.

Hughes, J. and Haran, M. (2013). Dimension reduction and alleviation of confounding for spatial generalized linear mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75,** 139–159.

Ihaka, R. and Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* **5,** 299–314.

Ising, E. (1925). Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik* **31,** 253–258.

Jin, X., Carlin, B. P., and Banerjee, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics* **61,** 950–961.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis* **94,** 401–419.

Joe, H. (2014). *Dependence Modeling with Copulas.* Chapman and Hall/CRC, Boca Raton, USA.

Kazianka, H. and Pilz, J. (2010). Copula-based geostatistical modeling of continuous and discrete data including covariates. *Stochastic Environmental Research and Risk Assessment* **24,** 661–673.

Kelsall, J. and Wakefield, J. (1999). Discussion of "Bayesian models for spatially correlated disease and exposure data", by Best et al. In Berger, J., Bernardo, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics 6*. Oxford University Press.

Kindermann, R. and Snell, J. (1980). *Markov Random Fields and Their Applications.* American Mathematical Society, Providence, RI.

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

Kolev, N. and Paiva, D. (2009). Copula-based regression models: A survey. *Journal of Statistical Planning and Inference* **139,** 3847–3856.

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

Lawson, A. B. (2014). Hierarchical modeling in spatial epidemiology. *Wiley Interdisciplinary Reviews: Computational Statistics* **6,** 405–417.

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

Lee, K. J., Jones, G., Caffo, B., and Bassett, S. S. (2014). Spatial Bayesian variable selection models on functional magnetic resonance imaging time-series data. *Bayesian Analysis* **9,** 699–732.

Leroux, B. G., Lei, X., and Breslow, N. (2000). Estimation of disease rates in small areas: A new mixed model for spatial dependence. *Institute for Mathematics and Its Applications* **116,** 179–191.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* **100,** 1989–2001.

Lindquist, M., Loh, J. M., Atlas, L., and Wager, T. (2009). Modeling the hemodynamic response function in fMRI: Efficiency, bias and mis-modeling. *NeuroImage* **45,** S187–S198.

Madsen, L. (2009). Maximum likelihood estimation of regression parameters with spatially dependent discrete data. *Journal of Agricultural, Biological, and Environmental Statistics* **14,** 375–391.

Madsen, L. and Fang, Y. (2011). Joint regression analysis for discrete longitudinal data. *Biometrics* **67,** 1171–1175.

Martinez-Beneito, M. A. (2013). A general modelling framework for multivariate disease mapping. *Biometrika* **100,** 539–553.

Masarotto, G. and Varin, C. (2012). Gaussian copula marginal regression. *Electronic Journal of Statistics* **6,** 1517–1549.

Møller, J. (1999). Perfect simulation of conditionally specified models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **61,** 251–264.

Møller, J., Pettitt, A., Reeves, R., and Berthelsen, K. (2006). An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants. *Biometrika* **93,** 451–458.

Moran, P. (1950). Notes on continuous stochastic phenomena. *Biometrika* **37,** 17–23.

Murray, J. S., Dunson, D. B., Carin, L., and Lucas, J. E. (2013). Bayesian Gaussian copula factor models for mixed data. *Journal of the American Statistical Association* **108,** 656–665.

Musgrove, D. R., Eberly, L. E., Klimes-Dougan, B., Basgoze, Z., Thomas, K. M., Mueller, B. A., Houri, A., Lim, K. O., and Cullen, K. R. (2015). Impaired bottom-up effective connectivity between amygdala and subgenual anterior cingulate cortex in unmedicated adolescents with major depression: Results from a dynamic causal modeling analysis. *Brain Connectivity* **5,** 608–619.

Neelon, B., Ghosh, P., and Loebs, P. F. (2013). A spatial Poisson hurdle model for exploring geographic variation in emergency department visits. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **176,** 389–413.

Penny, W., Kievel, S., and Friston, K. (2003). Variational Bayesian inference for fMRI time series. *NeuroImage* **19,** 727–741.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2013). *nlme: Linear and nonlinear mixed effects models.* R package version 3.1-113.

Prates, M. O., Dey, D. K., Willig, M. R., and Yan, J. (2015). Transformed Gaussian Markov random fields and spatial modeling of species abundance. *Spatial Statistics* **14,** 382–399.

Rathbun, S. L. and Fei, S. (2006). A spatial zero-inflated Poisson regression model for oak regeneration. *Environmental and Ecological Statistics* **13,** 409–426.

Recta, V., Haran, M., and Rosenberger, J. L. (2012). A two-stage model for incidence and prevalence in point-level spatial count data. *Environmetrics* **23,** 162–174.

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

Reich, B. J., Hodges, J. S., and Zadnik, V. (2006). Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics* **62,** 1197–1206.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Sanderson, C. (2010). Armadillo: An open source C++ linear algebra library for fast prototyping and computationally intensive experiments. Technical report, NICTA.

Singh, J. (1978). A characterization of positive Poisson distribution and its statistical application. *Society for Industrial and Applied Mathematics Journal on Applied Mathematics* **34,** 545–548.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* **8,** 229–231.

Smith, M., B. Pütz, D. A., and Fahrmeir, L. (2003). Assessing brain activity through spatial Bayesian variable selection. *NeuroImage* **20,** 802–815.

Smith, M. and Fahrmeir, L. (2007). Spatial Bayesian variable selection with application to functional magnetic resonance imaging. *Journal of the American Statistical Association* **102,** 417–431.

Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics* **75,** 317–343.

Speed, T. and Kiiveri, H. (1986). Gaussian Markov distributions over finite graphs. *The Annals of Statistics* **14,** 138–150.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64,** 583–639.

Tiefelsdorf, M. and Griffith, D. A. (2007). Semiparametric filtering of spatial autocorrelation: the eigenvector approach. *Environment and Planning A* **39,** 1193–1221.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* **15,** 273–289.

Varin, C. (2008). On composite marginal likelihoods. *AStA Advances in Statistical Analysis* **92,** 1–28.

Ver Hoef, J. M. and Jansen, J. K. (2007). Spacetime zero-inflated count models of Harbor seals. *Environmetrics* **18,** 697–712.

Wall, M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *Journal of Statistical Planning and Inference* **121,** 311–324.

Wang, F. and Landau, D. P. (2001). Efficient, multiple-range random walk algorithm to calculate the density of states. *Physical Review Letters* **86,** 2050–2053.

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., and Rosseel, Y. (2011). neuRosim: An R package for generating fMRI data. *Journal of Statistical Software* **44,** 1–18.

Wermuth, N. and Scheidt, E. (1977). Algorithm AS 105: Fitting a covariance selection model to a matrix. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **26,** 88–92.

Wikle, C. K. and Anderson, C. J. (2003). Climatological analysis of tornado report counts using a hierarchical Bayesian spatiotemporal model. *Journal of Geophysical Research: Atmospheres* **108,** STS 12–1.

Woolrich, M., Jenkinson, M., Brady, J., and Smith, S. (2004). Fully Bayesian spatio-temporal modeling of fMRI data. *IEEE Transactions on Medical Imaging* **23,** 213–231.

Worsley, K. J. (2003). Detecting activation in fMRI data. *Statistical Methods in Medical Research* **12,** 401–418.

Xu, P. F., Guo, J., and Tang, M. L. (2015). A localized implementation of the iterative proportional scaling procedure for Gaussian graphical models. *Journal of Computational and Graphical Statistics* **24,** 205–229.

Young, D. S., Raim, A. M., and Johnson, N. R. (2016). Zero-inflated modelling for

characterizing coverage errors of extracts from the US Census Bureau's Master Address File. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* to appear.

Zadnik, V. and Reich, B. J. (2006). Analysis of the relationship between socioeconomic factors and stomach cancer incidence in Slovenia. *Neoplasma* **53,** 103–110.

Zhang, C. and Ma, J. (2007). Simulation via direct computation of partition functions. *Physical Review E* **76,** 036708.

Appendix A

Appendix for Chapter 2

A.1 Full conditional distributions

Suppose we have v voxels, $v = 1, \ldots, N$, where for each voxel, we observe a time series $\mathbf{y}_v$ of T time points, indexed by t, $t = 1, \ldots, T$, and p stimulus effects each indexed by j, $j = 1, \ldots, p$. For a general autoregressive (AR) model, we will assume a lag of order r. At the voxel level, for the vth voxel, our model can be written

$$\mathbf{y}_v = \mathbf{Z}\boldsymbol{\alpha}_v + \mathbf{X}\boldsymbol{\beta}_v + \mathbf{W}_v\boldsymbol{\rho}_v + \boldsymbol{\varepsilon}_v, \ \ \boldsymbol{\varepsilon}_v \sim \mathcal{N}\left(\mathbf{0}, \sigma_v^2 \mathbf{I}_T\right),$$

$$p\left(\boldsymbol{\alpha}_v\right) = \mathcal{N}\left(\mathbf{0}, \tau_\alpha^2 \mathbf{I}_a\right),$$

$$p\left(\boldsymbol{\rho}_v\right) = \mathcal{N}\left(\mathbf{0}, \tau_\rho^2 \mathbf{I}_r\right),$$

$$p\left(\sigma_v^2\right) \propto 1/\sigma_v^2,$$

$$p\left(\beta_{vj} \mid \gamma_{vj}\right) = \gamma_{vj}\mathcal{N}\left(0, \tau_j^2\right) + \left(1 - \gamma_{vj}\right) I_0,$$

$$p\left(\tau_j^2\right) \propto 1/\tau_j^2, \ \ j = 1, \ldots, p,$$

where $\mathbf{Z}$ is a $T \times a$ matrix and accounts for the baseline trend, $\mathbf{X}$ is the $T \times p$ design matrix, $\mathbf{W}_v$ is the $T \times r$ matrix of lagged prediction errors (see Section 2.2), each of $\boldsymbol{\alpha}_v$, $\boldsymbol{\beta}_v = \left(\beta_{v1}, \ldots, \beta_{vp}\right)'$, and $\boldsymbol{\rho}_v$ are unknown coefficients to be estimated, γ_{vj} is an

unknown binary indicator variable, and the error variance and prior variance σ_v^2 and τ_j^2, respectively, are unknown. The prior variances τ_α^2 and τ_ρ^2 are user-specified.

For the jth stimulus, the spatial component of our model can be written

$$p\left(\gamma_{vj} \mid \eta_{vj}\right) = \mathcal{B}ern\left\{\mathbf{\Phi}\left(\eta_{vj}\right)\right\},$$

$$\eta_{vj} = \psi_{vj} + \mathbf{m}_v'\boldsymbol{\delta}_{sj} + \epsilon_{vj}, \ \ \epsilon_{vj} \sim \mathcal{N}\left(0, \ 1\right), \ \ v = 1, \ldots, N,$$

$$p\left(\boldsymbol{\delta}_{sj} \mid \kappa_j\right) = \mathcal{N}\left\{\mathbf{0}, \ (\kappa_j \mathbf{M}'\mathbf{Q}\mathbf{M})^{-1}\right\},$$

$$p\left(\kappa_j\right) = \mathcal{G}amma\left(1/2, \ 2000\right),$$

where $\mathbf{\Phi}$ denotes the cumulative distribution function of a standard normal random variable, η_{vj} is an auxiliary variable, ψ_{vj} is the prior probability of activation for the vth voxel and jth task, $\mathbf{m}_v$ is the vth row of $\mathbf{M}$ (see Section 2.2), $\boldsymbol{\delta}_{sj}$ is a q-vector of spatial random effects, $\mathbf{Q} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the diagonal matrix with the degrees (i.e., number of neighbors) of the voxels on the diagonal and $\mathbf{A}$ is the adjacency matrix for the underlying graph G (i.e., $\mathbf{A}$ is the binary matrix that encodes the adjacency structure among the voxels), and κ_j is the prior precision of $\boldsymbol{\delta}_{sj}$. The subscript 's' on $\boldsymbol{\delta}_{sj}$ denotes sparse since $\boldsymbol{\delta}_{sj}$ is modeled using the SAMM.

The full conditional distributions of β_{vj} and γ_{vj} are given in Section 2.3.4. The full conditional distribution of $\boldsymbol{\alpha}_v$ is given by $p\left(\boldsymbol{\alpha}_v \mid \mathbf{y}_v, \boldsymbol{\theta}_v \backslash \boldsymbol{\alpha}_v\right) \propto \mathcal{N}\left(\hat{\boldsymbol{\alpha}}_v, \ \hat{\Upsilon}_v\right)$, where $\boldsymbol{\theta}_v = \left(\boldsymbol{\alpha}_v', \boldsymbol{\beta}_v', \boldsymbol{\gamma}_v', \boldsymbol{\rho}_v', \sigma_v^2\right)'$ and $\boldsymbol{\theta}_v \backslash \boldsymbol{\alpha}_v$ denotes that $\boldsymbol{\alpha}_v$ is omitted from $\boldsymbol{\theta}_v$. For $\mathbf{z}_t$ and $\mathbf{x}_t$, the tth rows of $\mathbf{Z}$ and $\mathbf{X}$ respectively, let $\mathbf{z}_t^* = \mathbf{z}_t - \sum_{l=1}^r \mathbf{z}_{t-l}\rho_{vl}$, $\mathbf{x}_t^* = \mathbf{x}_t - \sum_{l=1}^r \mathbf{x}_{t-l}\rho_{vl}$, $y_{vt}^* = y_{vt} - \sum_{l=1}^r y_{v,t-l}\rho_{vl}$, and $e_{vt}^* = y_{vt}^* - \mathbf{x}_t^{*'}\boldsymbol{\beta}_v$. Then we have the regression $e_{vt}^* = \mathbf{z}_t^{*'}\boldsymbol{\alpha}_v + \varepsilon_v$. It easily follows that the posterior mean of $\boldsymbol{\alpha}_v$ is $\hat{\boldsymbol{\alpha}}_v = \sigma_v^2 \left(\sum_{t=1}^T \mathbf{z}_t^*\mathbf{z}_t^{*'} + \sigma_v^2/\tau_\alpha^2 \mathbf{I}_a\right)^{-1} \sum_{t=1}^T \mathbf{z}_t^* e_{vt}^*$ and the posterior covariance of $\boldsymbol{\alpha}_v$ is $\hat{\Upsilon}_v = \sigma_v^2 \left(\sum_{t=1}^T \mathbf{z}_t^*\mathbf{z}_t^{*'} + \sigma_v^2/\tau_\alpha^2 \mathbf{I}_a\right)^{-1}$.

The full conditional distribution of the voxel-level autoregressive coefficients $\boldsymbol{\rho}_v$, $r \times 1$, is given by $p(\boldsymbol{\rho}_v \mid \mathbf{y}_v, \boldsymbol{\theta}_v \backslash \boldsymbol{\rho}_v) \propto \mathcal{N}(\hat{\boldsymbol{\rho}}_v, \hat{\mathbf{R}}_v)$, where $\boldsymbol{\theta}_v \backslash \boldsymbol{\rho}_v$ denotes that $\boldsymbol{\rho}_v$ is omitted from $\boldsymbol{\theta}_v$. Let $\mathbf{w}_v = \mathbf{y}_v - \mathbf{Z}\boldsymbol{\alpha}_v - \mathbf{X}\boldsymbol{\beta}_v$ and $\mathbf{W}_v = (\mathbf{w}_{v1}|\ldots|\mathbf{w}_{vr})$, where $\mathbf{w}_{vk}$ denotes $\mathbf{w}_v$ lagged by k time steps and the kth column of $\mathbf{W}_v$ is $\mathbf{w}_{vk}$. Then, $\hat{\boldsymbol{\rho}}_v = \left(\mathbf{W}_v'\mathbf{W}_v + \sigma_v^2/\tau_\rho^2 \mathbf{I}_r\right)^{-1} \mathbf{W}_v'\mathbf{w}_v$ and $\hat{\mathbf{R}}_v = \sigma_v^2 \left(\mathbf{W}_v'\mathbf{W}_v + \sigma_v^2/\tau_\rho^2 \mathbf{I}_r\right)^{-1}$.

The full conditional distribution of the error variance σ_v^2 is $p(\sigma_v^2 \mid \mathbf{y}_v, \boldsymbol{\theta}_v \backslash \sigma_v^2) \propto \mathcal{IG}(T/2, \hat{b}_{\sigma^2})$, where $\boldsymbol{\theta}_v \backslash \sigma_v^2$ denotes that σ_v^2 is omitted from $\boldsymbol{\theta}_v$, $T/2$ and $\hat{b}_{\sigma^2}$ are the shape and scale of an inverse gamma distribution, respectively, and for $e_{vt} = y_{vt} - \mathbf{z}_t'\boldsymbol{\alpha}_v - \mathbf{x}_t'\boldsymbol{\beta}_v - \mathbf{w}_{vt}'\boldsymbol{\rho}_v$, $\hat{b}_{\sigma^2} = \sum_{t=1}^T e_{vt}^2/2$, where $\mathbf{w}_{vt}$ is the tth row of $\mathbf{W}_v$.

The full conditional distribution of the regression coefficient variance, τ_j^2, depends upon the non-zero $\boldsymbol{\beta}_j = (\beta_{1j}, \ldots, \beta_{nj})'$. Let $q_{1j} = \sum_{v=1}^n \gamma_{vj}$, i.e., q_{1j} is the number of non-zero regressors, then the full conditional distribution of each τ_j^2 is given by $p(\tau_j^2 \mid \boldsymbol{\beta}_j, \boldsymbol{\gamma}_j) \propto \mathcal{IG}(q_{1j}/2, \hat{c}_j)$, where $\hat{c}_j = \sum_{v=1}^n \beta_{vj}^2$.

Now, we move on to the spatial component of the model that relies on probit regression using auxiliary variables. The parameter η_{vj} is an auxiliary variable introduced to facilitate Gibbs sampling (Holmes and Held, 2006). In particular, if $\gamma_{vj} = 1$, η_{vj} is drawn from the positive half of a normal distribution left truncated at zero, while for $\gamma_{vj} = 0$, η_{vj} is drawn from the negative half of a normal distribution right truncated at zero.

The full conditional distribution of η_{vj} is derived by marginalizing $\boldsymbol{\delta}_{sj}$ out of the prior, thus $\eta_{vj} \mid \gamma_{vj}, \kappa_j \sim \mathcal{N}\left(\psi_{vj}, 1 + \kappa_j \mathbf{m}_v' \mathbf{Q}_s^{-1} \mathbf{m}_v\right)$, where $\mathbf{Q}_s = \mathbf{M}'\mathbf{QM}$ is the $q \times q$ prior precision matrix of the spatial random effects. Letting $\nu_{vj}^2 = 1 + \kappa_j \mathbf{m}_v' \mathbf{Q}_s^{-1} \mathbf{m}_v$, the full conditional distribution of each η_{vj} is given by

$$
p\left(\eta_{vj} \mid \gamma_{vj}, \kappa_j\right) \propto
\begin{cases}
\mathcal{N}\left(\psi_{vj}, \nu_{vj}^2\right) I\left(\eta_{vj} > 0\right) & \text{if } \gamma_{vj} = 1 \\
\mathcal{N}\left(\psi_{vj}, \nu_{vj}^2\right) I\left(\eta_{vj} \leq 0\right) & \text{otherwise,}
\end{cases}
$$

where $I(\cdot)$ denotes the indicator function.

The full conditional distribution of $\boldsymbol{\delta}_{sj}$ follows easily since we now have a tractable posterior distribution of the form $p\left(\boldsymbol{\delta}_{sj} \mid \boldsymbol{\eta}_j, \kappa_j\right) \propto \mathcal{N}\left(\hat{\boldsymbol{\delta}}_{sj}, \hat{\mathbf{Q}}_s^{-1}\right)$, where $\hat{\mathbf{Q}}_s = \kappa_j \mathbf{Q}_s + \mathbf{M}'\mathbf{M}$, $\hat{\boldsymbol{\delta}}_{sj} = \hat{\mathbf{Q}}_s^{-1} \mathbf{M}' \left(\boldsymbol{\eta}_j - \boldsymbol{\psi}_j\right)$, $\boldsymbol{\eta}_j = (\eta_{1j}, \ldots, \eta_{nj})'$, and $\boldsymbol{\psi}_j = (\psi_{1j}, \ldots, \psi_{nj})'$.

Finally, the spatial smoothing parameter, κ_j, has a full conditional distribution given by $p(\kappa_j \mid \boldsymbol{\delta}_{sj}) \propto \mathcal{G}amma\left\{(q+1)/2, \hat{b}_\kappa\right\}$, where $\hat{b}_\kappa = \boldsymbol{\delta}_{sj}' \mathbf{Q}_s \boldsymbol{\delta}_{sj}/2 + 1/2000$.

A.2 Algorithm: Brain parcellation

Here we describe the parcellation algorithm used to partition the brain into non–overlapping parcels. Details of the full parcellation algorithm are given in Algorithm 1.

The notation used is unique to this section. Inputs into the algorithm are the set of voxel coordinates A, a base parcel size n_0, and a minimum parcel size n_1. The base parcel size n_0 initializes the algorithm. The output of the algorithm is a list H with each entry containing a matrix of parcel-level voxel coordinates, with the restriction that the graph corresponding to each parcel is connected, i.e., there are no islands in the underlying graph.

Beginning with line 1, the parcellation of A is initialized as H_0. Next, in line 3 a while loop ensures the algorithm runs until the size of each parcel is greater than the minimum parcel size n_1. For lines 4 through 13, the algorithm enforces the minimum parcel size of n_1 voxels. In line 5, if the size of parcel h_i is less than n_1 voxels, the algorithm moves each voxel of h_i to the nearest parcel. In line 7, the parcel h_0 nearest to voxel h_{ij} is found. In line 8, voxel h_{ij} is stored in parcel h_0. After each voxel is moved, parcel h_i is deleted from the partition in line 10. The algorithm continues until all parcels have a minimum size of n_1 voxels.

Algorithm 1 Brain parcellation

Input: Set of coordinates A, base parcel size n_0, and minimum parcel size n_1

Output: List H with each entry containing a matrix of parcel-level voxel coordinates

1: Partition A into $H_0 = \{h_1, \ldots, h_{M_0}\}$, such that $\mathrm{nrow}(h_i) \leq n_0$, $i = 1, \ldots, M_0$, $\bigcup_{i=1}^{M_0} h_i = A$, $h_i \bigcap h_j = \emptyset$ for $i \neq j$, and each voxel in h_i has at least one adjacent voxel

2:   $M \leftarrow M_0$

3: **while** $\min_{i=1,\ldots,M} \mathrm{nrow}(h_i) < n_1$ **do**

4:     **for** $i = 1$ **to** M **do**

5:       **if** $\mathrm{nrow}(h_i) < n_1$ **then**

6:         **for** $j = 1$ **to** $\mathrm{nrow}(h_i)$ **do**

7:           $h_0 \leftarrow \mathrm{which.min.distance}(h_{ij}, H_0 \backslash h_i)$

8:           $h_0.\mathrm{add}(h_{ij})$

9:         **end for**

10:         $H_0 \leftarrow H_0.\mathrm{delete}(h_i)$

11:       **end if**

12:     **end for**

13:     $M \leftarrow H_0.\mathrm{size}$

14: **end while**

15: $H \leftarrow H_0$

16: **return** H

The parcellation algorithm was programmed in R. Using a PC desktop with an i7-3770, 3.4 GHz processor, parcellation of a 61,000 voxel brain into 90 parcels, with an average parcel size of 680 voxels (min: 500, max: 998), requires 20 minutes of computing time.

A.3 Extended results and analysis comparison

In this section, we present comparisons of our model, the SAMM with AR(2) (SAMMAR2), to a non-spatial AR(2) (NSAR2) model. The NSAR2 model is applied in a frequentist setting to each voxel using the gls() function from the `nlme` package (Pinheiro et al., 2013) in R.

Figure A.1 displays axial slices for the SAMMAR2 model in column 1 and the NSAR2 model in column 2. The estimated posterior probabilities from the SAMMAR2 model were thresholded at > 0.8722 while the p-values for the task effect were thresholded at < 0.001 to produce similar results. An important distinction between the two model results is the treatment of the activation probabilities. With the SAMMAR2 model, the binary activation maps represent the probability of activation at the displayed voxels. Here, the probability of activation is at least 0.8722. The same is not true with the activation maps associated with the NSAR2 model.

To further assess differences in activation detection between the SAMMAR2 and NSAR2 models, we conducted a simulation study with four known areas of activation as shown in Figure A.2. To carry out the simulation, we used an approach similar to the simulation study of section 2.4.2, but now on a single 25×25 grid. We generated 100 simulated datasets and applied the SAMMAR2 and NSAR2 models to each dataset. For the NSAR2 model, we calculated p-values for the activation effect and then applied a cluster threshold of $p < 0.001$ and a contiguous area of activation of at least 10 voxels. For the SAMMAR2 model, we applied a posterior probability threshold of $p > 0.8722$ and also used a cluster threshold type approach with a contiguous area of activation of at least 10 voxels. The voxel-level median false positive rates (FPR) were 0 for both models. The maximum FPR for the SAMMAR2 model was 0.05 while the maximum FPR for the NSAR2 model was 0.02. Turning to true positive rate (TPR), the SAMMAR2 model achieves a median TPR of 1.0 while the NSAR2 model suffers with a median TPR of 0.90.
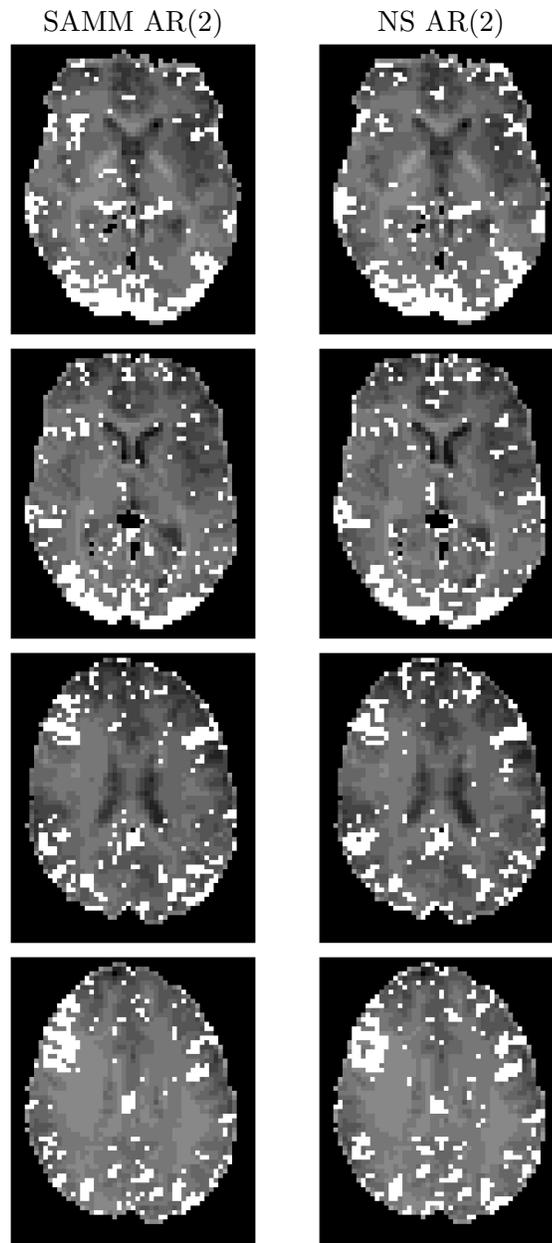
Figure A.1: Model comparisons for increased activity over baseline for the face stimulus. Each row of brain maps corresponds to a slice (top to bottom, slices 17, 19, 25, and 27. The SAMM AR(2) column shows voxel-specific posterior probabilities, exceeding 0.8722, of increased activity over baseline for the face stimulus. The NS AR(2) column shows voxel-specific p-values, below 0.001, of increased activity over baseline for the face stimulus.
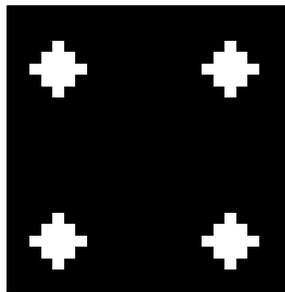
Figure A.2: Simulated activation map. The map is a 25×25 lattice with four symmetric areas of activation. Each area of activation is composed of 13 active voxels.

Appendix B

Appendix for Chapter 3

B.1 Algorithm: Localized implementation of IPS

Here we describe localized iterative proportional scaling (IPS) for computing our concentration matrix $\mathbf{Q}$. On the graph G, denote a partition of the set of cliques $\mathcal{K}(G)$ as $W = \{\mathcal{K}_1, \ldots, \mathcal{K}_M\}$ such that $\bigcup_{m=1}^{M} \mathcal{K}_m = \mathcal{K}(G)$ and $\mathcal{K}_i \bigcap \mathcal{K}_j = \emptyset$ for $i \neq j$. That is, if clique $c \in \mathcal{K}_i$, $c \notin \mathcal{K}_j$. Note that a given vertex may appear in more than one clique, and thus in more than one clique partition. Let $U_m = \bigcup_{c \in \mathcal{K}_m} c$ and $\bar{U}_m = V \setminus U_m$, where V is the set of vertices of G.

Details of the full algorithm are given in Algorithm 2. Inputs into the algorithm are an $n \times n$ adjacency matrix $\mathbf{A}$ (see Section 3.3) and first-order correlation ρ. The output of the algorithm is the concentration matrix $\mathbf{Q}$ and the determinant $|\mathbf{Q}|$. In lines 1 and 2, the Cuthill–McKee and Bron–Kerbosch algorithms, respectively, are applied to the graph represented by the adjacency matrix $\mathbf{A}$. In line 3, for graphs of the type that we work with, it suffices to use an equipartition such that each component of the partition comprises cliques chosen in numerical order based on the edge location as specified in $\mathbf{A}$. The algorithm is initialized in line 4 using $\mathbf{Q} = \mathbf{I}_n$. Lines 6 through 13 carry out localized IPS.

In line 7, each element of the mth clique partition is stored in U_m while $\bar{U}_m$ is the complement of U_m with respect to all vertices. Line 8 constructs the matrix that the clique-level operation O_c is applied to. We call O_c the clique-level operation on $\mathbf{\Lambda}$ and

define the operation as

$$
O_c \mathbf{\Lambda} = \begin{pmatrix} \mathbf{S}_c^{-1} + \mathbf{\Lambda}_{ca}\mathbf{\Lambda}_a^{-1}\mathbf{\Lambda}_{ac} & \mathbf{\Lambda}_{ca} \\ \mathbf{\Lambda}_{ac} & \mathbf{\Lambda}_a \end{pmatrix},
$$

where $a = V \setminus c$ and $\mathbf{S}_c = (1-\rho)\mathbf{I}_{\|c\|} + \rho \mathbf{J}_{\|c\|}$. Lines 9 through 11 loop through each clique of the mth clique partition and update the matrix $\mathbf{\Lambda}$. In line 12, the clique partition-level adjusted matrix is stored. Lines 6 through 13 are repeated until convergence, i.e., the change in the log-likelihood is smaller than ε for small $\varepsilon > 0$.

Algorithm 2 Localized IPS for a concentration matrix

Input: Adjacency matrix $\mathbf{A}$ and first-order correlation ρ

Output: Matrix $\mathbf{Q}$ and determinant $|\mathbf{Q}|$

 1: Reorder $\mathbf{A}$ to minimize its bandwidth
 2: Enumerate all cliques of G in $\mathcal{K}(G)$ associated with $\mathbf{A}$
 3: Find a partition $W = \{\mathcal{K}_1, \ldots, \mathcal{K}_M\}$ such that $\bigcup_{m=1}^{M} \mathcal{K}_m = \mathcal{K}(G)$ and $\mathcal{K}_i \bigcap \mathcal{K}_j = \emptyset$
    for $i \neq j$
 4: $\mathbf{Q} = \mathbf{I}_n$
 5: **while** not convergent **do**
 6:    **for** $m = 1$ **to** M **do**
 7:       $U_m \leftarrow \bigcup_{c \in \mathcal{K}_m} c, \quad \bar{U}_m \leftarrow V \setminus U_m$
 8:       $\mathbf{\Psi} \leftarrow \mathbf{Q}_{U_m \bar{U}_m} \mathbf{Q}_{\bar{U}_m}^{-1} \mathbf{Q}_{\bar{U}_m U_m}, \ \mathbf{\Lambda} \leftarrow \mathbf{Q}_{U_m} - \mathbf{\Psi}$
 9:       **for** each clique $c \in \mathcal{K}_m$ **do**
10:          $\mathbf{\Lambda} \leftarrow O_c \mathbf{\Lambda}$
11:       **end for**
12:       $\mathbf{Q}_{U_m} \leftarrow \mathbf{\Lambda} + \mathbf{\Psi}$
13:    **end for**
14: **end while**
15: **return** $\mathbf{Q}$ and $|\mathbf{Q}|$

The IPS algorithm was programmed in C++ for increased computational speed. Using a PC desktop with an i7-3770, 3.4 GHz processor, computation of $\mathbf{Q}$ requires a few milliseconds for the Slovenia map ($n = 194$) of Section 3.5.1. For the larger 40×40 lattice of Section 3.5.2 ($n = 1,600$), 150 seconds of computing time were required.

Appendix C

Appendix for Chapter 4

C.1 Technical details

C.1.1 Multivariate spatial random effect reparameterization

For the multivariate sparse areal mixed model (MSAMM), when the design matrices are the same across multivariate outcomes, i.e., $\mathbf{X}_1 = \mathbf{X}_2 = \cdots = \mathbf{X}_J$, the likelihood and prior distribution of the spatial effects can be written

$$g_j \left\{ \mathbb{E} \left(\mathbf{y}_j \mid \boldsymbol{\beta}_j, \, \boldsymbol{\delta}_{sj} \right) \right\} \;\; = \;\; \mathbf{X}\boldsymbol{\beta}_j + \mathbf{M}\boldsymbol{\delta}_{sj}, \;\; j = 1, \, \ldots, \, J,$$

$$p \left(\boldsymbol{\Delta} \mid \boldsymbol{\Sigma} \right) \;\; = \;\; \mathcal{N} \left(\mathbf{0}, \, \boldsymbol{\Sigma} \otimes \mathbf{Q}_s^{-1} \right),$$

where $\boldsymbol{\Delta} = \left(\boldsymbol{\delta}'_{s1}, \, \ldots, \, \boldsymbol{\delta}'_{sJ} \right)'$, each $\boldsymbol{\delta}_{sj}$ is $q \times 1$, $\boldsymbol{\Sigma}$ is the $J \times J$ covariance matrix, and $\mathbf{Q}_s$ is the $q \times q$ spatial precision matrix. Computations are eased considerably as follows. Let $\mathbf{R}_s$ be the upper triangular Cholesky decomposition of $\mathbf{Q}_s$ and let $\mathbf{L}_s = \mathbf{R}_s^{-1}$ such that $\mathbf{L}_s \mathbf{L}'_s = \mathbf{Q}_s^{-1}$. Then, for $\boldsymbol{\Psi} = \left(\boldsymbol{\psi}'_{s1}, \, \ldots, \, \boldsymbol{\psi}'_{sJ} \right)'$, each $\boldsymbol{\psi}_{sj}$ is $q \times 1$, and $\boldsymbol{\Psi} \mid \boldsymbol{\Sigma} \sim \mathcal{N} \left(\mathbf{0}, \, \boldsymbol{\Sigma} \otimes \mathbf{I}_q \right)$, we have that $\left(\mathbf{I}_J \otimes \mathbf{L}_s \right) \boldsymbol{\Psi}$ and $\boldsymbol{\Delta}$ have the same distribution. This is easy to see since $\mathbb{E} \left\{ \left(\mathbf{I}_J \otimes \mathbf{L}_s \right) \boldsymbol{\Psi} \right\} = \left(\mathbf{I}_J \otimes \mathbf{L}_s \right) \mathbb{E} \left(\boldsymbol{\Psi} \right) = \mathbf{0}$ and

$$Cov \left\{ \left(\mathbf{I}_J \otimes \mathbf{L}_s \right) \boldsymbol{\Psi} \right\} \;\; = \;\; \left(\mathbf{I}_J \otimes \mathbf{L}_s \right) \left(\boldsymbol{\Sigma} \otimes \mathbf{I}_q \right) \left(\mathbf{I}_J \otimes \mathbf{L}_s \right)'$$

$$= \;\; \boldsymbol{\Sigma} \otimes \mathbf{Q}_s^{-1}.$$

Thus, the likelihood and prior distribution of the spatial effects can now be written

$$g_j\left\{\mathbb{E}\left(\mathbf{y}_j \mid \boldsymbol{\beta}_j,\, \boldsymbol{\psi}_{sj}\right)\right\} \;=\; \mathbf{X}\boldsymbol{\beta}_j + \mathbf{M}\mathbf{L}_s\boldsymbol{\psi}_{sj},\; j = 1,\, \ldots,\, J,$$

$$p\left(\boldsymbol{\Psi} \mid \boldsymbol{\Sigma}\right) \;=\; \mathcal{N}\left(\mathbf{0},\, \boldsymbol{\Sigma} \otimes \mathbf{I}_q\right).$$

Now let's suppose $\mathbf{X}_1 \neq \mathbf{X}_2 \neq \cdots \neq \mathbf{X}_J$. The likelihood and prior distribution of the spatial effects can be written

$$g_j\left\{\mathbb{E}\left(\mathbf{y}_j \mid \boldsymbol{\beta}_j,\, \boldsymbol{\delta}_{sj}\right)\right\} \;=\; \mathbf{X}_j\boldsymbol{\beta}_j + \mathbf{M}_j\boldsymbol{\delta}_{sj},\; j = 1,\, \ldots,\, J,$$

$$p\left(\boldsymbol{\Delta} \mid \boldsymbol{\Sigma}\right) \;=\; \mathcal{N}\left[\mathbf{0},\, \left\{\mathbf{R}'\left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_q\right)\mathbf{R}\right\}^{-1}\right],$$

where $\boldsymbol{\Delta} = \left(\boldsymbol{\delta}'_{s1},\, \ldots,\, \boldsymbol{\delta}'_{sJ}\right)'$, $\mathbf{R} = \mathrm{bdiag}\left(\mathbf{R}_{s1},\, \ldots,\, \mathbf{R}_{sJ}\right)$, and $\mathbf{R}'_{sj}\mathbf{R}_{sj} = \mathbf{Q}_{sj}$, where $\mathbf{R}_{sj}$ is the upper triangular Cholesky decomposition of $\mathbf{Q}_{sj}$. For ease of exposition, let $J = 2$. The following easily extends to the case when $J > 2$. The prior distribution of the spatial effects can be written

$$\begin{pmatrix} \boldsymbol{\delta}_{s1} \\ \boldsymbol{\delta}_{s2} \end{pmatrix} \mid \boldsymbol{\Sigma} \;\sim\; \mathcal{N}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},\, \left\{\begin{pmatrix} \mathbf{R}_{s1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{s2} \end{pmatrix}'\left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_q\right)\begin{pmatrix} \mathbf{R}_{s1} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{s2} \end{pmatrix}\right\}^{-1}\right],$$

$$\sim\; \mathcal{N}\left[\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},\, \begin{pmatrix} \mathbf{L}_{s1} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{s2} \end{pmatrix}\left(\boldsymbol{\Sigma} \otimes \mathbf{I}_q\right)\begin{pmatrix} \mathbf{L}_{s1} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{s2} \end{pmatrix}'\right].$$

where $\mathbf{L}_{sj} = \mathbf{R}_{sj}^{-1}$, $j = 1,\, 2$, and we've used the fact that $\left(\mathbf{R}_{sj}^{-1}\right)' = \left(\mathbf{R}'_{sj}\right)^{-1}$. Now, suppose we have

$$\begin{pmatrix} \boldsymbol{\psi}_{s1} \\ \boldsymbol{\psi}_{s2} \end{pmatrix} \mid \boldsymbol{\Sigma} \sim \mathcal{N}\left\{\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},\, \boldsymbol{\Sigma} \otimes \mathbf{I}_q\right\}.$$

Using basic properties of the multivariate normal distribution, we have that

$$\begin{pmatrix} \mathbf{L}_{s1} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{s2} \end{pmatrix}\begin{pmatrix} \boldsymbol{\psi}_{s1} \\ \boldsymbol{\psi}_{s2} \end{pmatrix} \mid \boldsymbol{\Sigma} \sim \mathcal{N}\left\{\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix},\, \begin{pmatrix} \mathbf{L}_{s1} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{s2} \end{pmatrix}\left(\boldsymbol{\Sigma} \otimes \mathbf{I}_q\right)\begin{pmatrix} \mathbf{L}_{s1} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{s2} \end{pmatrix}'\right\}.$$

Then, since

$$\begin{pmatrix} \mathbf{L}_{s1} & \mathbf{0} \\ \mathbf{0} & \mathbf{L}_{s2} \end{pmatrix}\begin{pmatrix} \boldsymbol{\psi}_{s1} \\ \boldsymbol{\psi}_{s2} \end{pmatrix} = \begin{pmatrix} \mathbf{L}_{s1}\boldsymbol{\psi}_{s1} \\ \mathbf{L}_{s2}\boldsymbol{\psi}_{s2} \end{pmatrix},$$

we can apply a reparameterization similar to the case where design matrices are equivalent across the outcomes. Thus, in general, we can write the likelihood and prior distribution of the spatial effects as

$$g_j \left\{ \mathbb{E} \left(\mathbf{y}_j \mid \boldsymbol{\beta}_j, \, \boldsymbol{\psi}_{sj} \right) \right\} \;=\; \mathbf{X}_j \boldsymbol{\beta}_j + \mathbf{M}_j \mathbf{L}_{sj} \boldsymbol{\psi}_{sj}, \; j = 1, \, \ldots, \, J,$$

$$p \left(\boldsymbol{\Psi} \mid \boldsymbol{\Sigma} \right) \;=\; \mathcal{N} \left(\mathbf{0}, \, \boldsymbol{\Sigma} \otimes \mathbf{I}_q \right),$$

where $\boldsymbol{\Psi} = \left(\boldsymbol{\psi}'_{s1}, \, \ldots, \, \boldsymbol{\psi}'_{sJ} \right)'$.

C.1.2 Full conditional distributions

The full MSAMM, using the reparameterization of Section C.1.1, can be written hierarchically as

$$g_j \left\{ \mathbb{E} \left(\mathbf{y}_j \mid \boldsymbol{\beta}_j, \, \boldsymbol{\psi}_j \right) \right\} \;=\; \mathbf{X}_j \boldsymbol{\beta}_j + \mathbf{M}_{sj} \boldsymbol{\psi}_{sj},$$

$$p \left(\boldsymbol{\Psi} \mid \boldsymbol{\Sigma} \right) \;=\; \mathcal{N} \left(\mathbf{0}, \, \boldsymbol{\Sigma} \otimes \mathbf{I}_q \right),$$

$$p \left(\boldsymbol{\beta}_j \right) \;=\; \mathcal{N} \left(\mathbf{0}, \, b_\beta \mathbf{I}_p \right),$$

$$p \left(v_j \right) \;=\; \mathcal{IG} \left(1/2, \, 1/\zeta_j \right), \; j = 1, \, \ldots, \, J,$$

$$p \left(\boldsymbol{\Sigma} \mid v_1, \ldots, v_J \right) \;=\; \mathcal{IW} \left\{ 2\nu \mathrm{diag}(v_1, \ldots, v_J), \, \nu + J - 1 \right\},$$

where g_j is the link-function, $\mathbf{M}_{sj} = \mathbf{M}_j \mathbf{L}_{sj}$, $\boldsymbol{\Psi} = \left(\boldsymbol{\psi}'_{s1}, \, \ldots, \, \boldsymbol{\psi}'_{sJ} \right)'$, and each of b_β, ν, and ζ_j, $j = 1, \ldots, J$, are hyperparameters to be specified. The full conditional distributions of $\boldsymbol{\Psi}$ and $\boldsymbol{\beta}_j$, $j = 1, \, \ldots, \, J$ are generally not available in closed form, while the full conditional distributions of v_j, $j = 1, \ldots, J$, and $\boldsymbol{\Sigma}$ are inverse-gamma and inverse-Wishart, respectively.

Let $f_j \left\{ \cdot \right\}$ be the likelihood function, $\mathbf{Y} = (\mathbf{y}'_1, \ldots, \mathbf{y}'_J)'$, $\mathbf{B} = \left(\boldsymbol{\beta}'_1, \ldots, \boldsymbol{\beta}'_J \right)'$, and $\boldsymbol{\eta}_j = \mathbf{X}_j \boldsymbol{\beta}_j + \mathbf{M}_{sj} \boldsymbol{\psi}_{sj}$. The full conditional distribution of $\boldsymbol{\Psi}$ is

$$p \left(\boldsymbol{\Psi} \mid \mathbf{Y}, \, \mathbf{B}, \, \boldsymbol{\Sigma} \right) \propto \exp \left\{ -\boldsymbol{\Psi}' \left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_q \right) \boldsymbol{\Psi} / 2 \right\} \prod_{j=1}^{J} f_j \left\{ g_j^{-1} \left(\boldsymbol{\eta}_j \right) \right\}.$$

The full conditional distribution of each $\boldsymbol{\beta}_j$, $j = 1, \ldots, J$, is

$$p\left(\boldsymbol{\beta}_j \mid \mathbf{y}_j, \boldsymbol{\psi}_{sj}\right) \propto \exp\left\{-\boldsymbol{\beta}_j'\boldsymbol{\beta}_j/(2b_{\boldsymbol{\beta}})\right\} \times f_j\left\{g_j^{-1}\left(\boldsymbol{\eta}_j\right)\right\}.$$

Each of $\boldsymbol{\Psi}$ and $\boldsymbol{\beta}_j$, $j = 1, \ldots, J$, do not have a closed form full conditional distribution, and so we use Metropolis updates to draw samples from the posteriors.

The full conditional distribution of each v_j, $j = 1, \ldots, J$, depends only on $\boldsymbol{\Sigma}$. Thus,

$$p\left(v_j \mid \boldsymbol{\Sigma}\right) \propto \mathcal{IG}\left\{\left(\nu + J\right)/2,\ \nu\left(\boldsymbol{\Sigma}^{-1}\right)_{jj} + \zeta_j\right\}.$$

Finally, the full conditional distribution of $\boldsymbol{\Sigma}$ is computed as follows. Reorder $\boldsymbol{\Psi}$ into a $q \times J$ matrix denoted $\boldsymbol{\Psi}_J = \left(\boldsymbol{\psi}_{s1}|\ldots|\boldsymbol{\psi}_{sJ}\right)$, i.e., the jth column of $\boldsymbol{\Psi}_J$ is $\boldsymbol{\psi}_{sj}$. Then, for $\mathbf{V} = \mathrm{diag}\left(v_1,\ \ldots,\ v_J\right)$, the full conditional distribution of $\boldsymbol{\Sigma}$ is

$$p\left(\boldsymbol{\Sigma} \mid \boldsymbol{\Psi}_J, \mathbf{V}\right) \propto \mathcal{IW}\left(\boldsymbol{\Psi}_J'\boldsymbol{\Psi}_J + 2\nu\mathbf{V},\ \nu + q + J - 1\right).$$

We can use Gibbs updates to draw posterior samples for $\boldsymbol{\Sigma}$ and v_j, $j = 1, \ldots, J$.

C.2 Additional simulation results

The following three tables present additional results for the multivariate and zero-inflated simulation studies. For ease of exposition within the chapter, the coverage probabilities of the model parameters were omitted and are presented here for reference.

Table C.1: Additional Iowa multivariate simulation results for 1,000 simulated datasets comparing the MSAMM that accounts for correlation between spatial effects (Correlation Model) to independent SAMM fits (Independent Model). CP denotes coverage probability of the 95% credible interval.

Parameter	Truth	Correlation Model		Independent Model	
		Mean Est.	95% CP	Mean Est.	95% CP
β_{11}	2	2.000	0.952	2.000	0.955
β_{12}	-1	-0.999	0.962	-0.999	0.967
β_{21}	2	2.001	0.944	2.000	0.945
β_{22}	-1	-1.001	0.971	-1.001	0.964
β_{31}	2	2.000	0.947	1.999	0.948
β_{32}	-1	-1.003	0.966	-1.004	0.964
σ_1^2	6	5.946	0.951	6.136	0.950
σ_2^2	8	7.901	0.953	8.194	0.965
σ_3^2	10	10.025	0.940	10.436	0.948
ρ_{12}	0.8	0.806	0.946	–	–
ρ_{13}	0.8	0.804	0.941	–	–
ρ_{23}	0.8	0.808	0.923	–	–

Table C.2: Additional Iowa zero-inflated simulation results for 1,000 simulated datasets comparing the MSAMM that accounts for small to moderate correlation ($\rho = 0$, 0.2, 0.4) between spatial effects (Correlation Model) to independent SAMM fits (Independent Model). CP denotes coverage probability of the 95% credible interval.

Parameter	Truth	Correlation Model		Independent Model	
		Mean Est.	95% CP	Mean Est.	95% CP
β_{11}	-1	-0.998	0.948	-0.995	0.942
β_{12}	1	0.994	0.930	0.991	0.931
β_{21}	2	2.003	0.953	2.003	0.962
β_{22}	-1	-1.001	0.958	-1.001	0.955
σ_1^2	4	3.970	0.945	3.703	0.906
σ_2^2	8	8.274	0.975	8.153	0.977
ρ	0	<0.001	0.952	–	–
β_{11}	-1	-0.998	0.945	-0.995	0.942
β_{12}	1	0.994	0.929	0.991	0.931
β_{21}	2	2.004	0.959	2.012	0.944
β_{22}	-1	-1.000	0.953	-1.003	0.949
σ_1^2	4	4.009	0.950	3.703	0.906
σ_2^2	8	8.248	0.970	8.112	0.974
ρ	0.2	0.201	0.954	–	–
β_{11}	-1	-0.998	0.943	-0.995	0.942
β_{12}	1	0.994	0.929	0.991	0.931
β_{21}	2	2.004	0.962	2.020	0.930
β_{22}	-1	-1.000	0.954	-1.004	0.951
σ_1^2	4	4.012	0.949	3.703	0.906
σ_2^2	8	8.229	0.970	8.071	0.966
ρ	0.4	0.405	0.949	–	–

Table C.3: Additional Iowa zero-inflated simulation results for 1,000 simulated datasets comparing the MSAMM that accounts for moderately strong to strong correlation ($\rho = 0.6, 0.8$) between spatial effects (Correlation Model) to independent SAMM fits (Independent Model). CP denotes coverage probability of the 95% credible interval.

Parameter	Truth	Correlation Model		Independent Model	
		Mean Est.	95% CP	Mean Est.	95% CP
β_{11}	-1	-0.998	0.945	-0.995	0.942
β_{12}	1	0.995	0.932	0.991	0.931
β_{21}	2	2.005	0.966	2.028	0.903
β_{22}	-1	-1.000	0.957	-1.007	0.950
σ_1^2	4	4.060	0.940	3.703	0.906
σ_2^2	8	8.238	0.976	8.018	0.973
ρ	0.6	0.604	0.945	–	–
β_{11}	-1	-1.000	0.945	-0.995	0.942
β_{12}	1	0.996	0.933	0.991	0.931
β_{21}	2	2.007	0.963	2.036	0.853
β_{22}	-1	-1.000	0.959	-1.008	0.946
σ_1^2	4	4.165	0.945	3.703	0.906
σ_2^2	8	8.197	0.979	7.944	0.965
ρ	0.8	0.793	0.953	–	–