Statistical Modeling and Inference for the Analysis of Spatial Categorical Data

By

Chongyang Jin

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

(STATISTICS)

at the

UNIVERSITY OF WISCONSIN – MADISON

2012

Date of final oral examination: July 18, 2012

The dissertation is approved by the following members of the Final Oral Committee:

Dr. J. Zhu, Professor, Statistics

Dr. M. Clayton, Professor, Statistics

Dr. R. Gangnon, Associate Professor, Biostatistics and Medical Informatics

Dr. K. Tsui, Professor, Statistics

Dr. Z. Zhang, Associate Professor, Statistics

Abstract

Spatial categorical data on a lattice are becoming increasingly abundant due to the advances of geographic information systems in environmental science. In this thesis, statistical tools are developed for regression analysis of this type of data. The response variable is modeled by a multinomial distribution in a generalized linear mixed model framework. There are two additive components in the linear predictor: a linear regression on covariates and a spatial random effect such that the spatial dependence in the random effect is induced by a multivariate conditional autoregressive model. Bayesian hierarchical modeling is used for statistical inference and Markov chain Monte Carlo algorithms are devised to obtain posterior samples. The methodology is applied to analyze a northern Wisconsin land cover data set in a study that assesses the relationship between forest landscape structure and past social conditions, expanding the analytical tools available in landscape ecology and environmental history.

Besides Bayesian hierarchical modeling, various other statistical methods have been developed for relating a spatial binary response to covariates in environmental and ecological studies, while properly accounting for spatial dependence. However, these methods tend to be computationally intensive and may be sensitive to model misspecification. Thus, a quasi-likelihood estimating equation is further developed for estimating regression coefficients and drawing statistical inference. In addition, a regularization method is proposed for model selection via penalized quasi-likelihood under an adaptive Lasso. A series of simulations are conducted to evaluate the performance of these new methods. For illustration, these methods are also applied to analyze the northern Wisconsin land

cover data set.

Acknowledgements

I would like to take this opportunity to acknowledge a number of people. First of all, I must thank my advisor Dr. Jun Zhu for her knowledge, enthusiasm, support, and guidance, which made my graduate school experience such a positive one. I am very grateful and fortunate to have such an excellent advisor. I also thank my other committee members, Dr. Murray Clayton, Dr. Ronald Gangnon, Dr. Kam-Wah Tsui and Dr. Zhengjun Zhang for their support.

It is also very important for me to thank the co-authors of my paper, Dr. Michelle Steen-Adams, who provided a very interesting land cover data set and brought us this environmental and ecological problem, Dr. Stephan Sain, who shared his idea and Dr. Ronald Gangnon, who gave lots of useful advice.

Finally, special thanks are reserved for my beautiful wife Yue Zhan, my parents, Yiping Jin and Yu Kong, for their love, help and support for so many years.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Spatial categorical data on a lattice arise often in environmental and ecological studies and are becoming increasingly abundant due to the rapid advances of computer technologies such as the Geographical Information Systems (GIS). However, statistical methods for analyzing such data are limited. The purpose of this thesis is to fill some of this void by bringing together several strands of recent research in spatial statistics, developing new statistical models and inferential methods, and making these methodologies applicable for the analysis of spatial categorical data in practice. Although the specific application here concerns land cover data in relationship to land ownership characteristics in an intersecting area of landscape ecology and environmental history, the statistical methodologies developed can be adapted to analyze spatial categorical data in general.

In the first part of the thesis, we propose a spatial multinomial regression model in the generalized linear mixed model (GLMM) framework, such that the response variable follows a multinomial model and the latent spatial random effect follows a spatial lattice model. Because the categorical data are nominal instead of ordinal, the spatial random effect is multivariate at each site on the lattice, rather than univariate as in most other spatial GLMMs. As a result, additional modeling, methodological, and computational challenges are encountered. Our strategy is to utilize a multivariate spatial lattice model, which can be viewed as an extension from a popular, univariate conditional autoregressive (CAR) model. We further employ dimension-reduction techniques to help alleviate the computational burden. In particular, we consider spatial factor analysis models, such that the latent variables at each site are reduced to a smaller number of latent factors. For statistical inference, we adopt Bayesian hierarchical modeling and devise Markov chain Monte Carlo (MCMC) algorithms to obtain the posterior for statistical inference. This part of the thesis is presented in two separate chapters. In Chapter 2, a new methodology is developed including model specification, Bayesian inference, and connection to alternative models. In Chapter 3, the methodology is applied in a case study for illustration. Specifically, a data set of land cover in northern Wisconsin is analyzed in depth and scientific interpretation is drawn from the perspective of landscape ecology and environmental history.

In the second part of the thesis, we focus on spatial binary data on a lattice, which may be viewed as a special case of spatial categorical data. We consider a new method, alternative to Bayesian inference, for the analysis of such spatial binary data. We propose a quasi-likelihood estimating equation (QEE) approach, where a working correlation matrix is specified but no specific spatial dependence model needs to be assumed. For computation, only a few parameters are used for specifying the working correlation matrix and thus, our method mitigates the singularity problem encountered in previous work. Furthermore, model selection is an important topic to address for regression of a spatial binary response variable on covariates, but very few tools are available in the existing literature. Here, we construct a regularized method using penalized quasi-likelihood for variable selection and develop a LARS-type algorithm for fast computation. A simulation study is conducted to empirically study the asymptotic property and robustness of our method. Finally, the method is again applied to analyze the northern Wisconsin land cover data set for illustration. This second part of the thesis is presented in Chapter 4. Conclusions and a brief discussion of future work are given in the final Chapter 5.

Chapter 2

Spatial Multinomial Regression Models and Bayesian Inference

2.1 Overview

Spatial categorical data on a lattice are becoming increasingly abundant due to the advances of computer technologies in environmental and ecological studies. In a study of northern Wisconsin, an interdisciplinary team of ecologists and environmental historians assess the influence of past social conditions on forest landscape structure and change over a 130-year period relative to pre-EuroAmerican settlement baseline conditions. Using a novel spatial multinomial regression for nominal categorical data, we analyze land cover data from this particular study to integrate historic conditions into landscape ecological analysis. In particular, we build on landscape ecology research that demonstrates relationships between land ownership characteristics and forest landscape composition (Theobald et al., 1996; Crow et al., 1999; Brown, 2003). The response variable is land cover defined in terms of dominance by a land cover class within each quarter-section in the study area. The covariates consist of four land ownership metrics: ownership category, ownership size (or, property area), parcel size, and parcel polygon proportion of quarter-section, a measure of fragmentation. See Figure 1 and Section 3.1 for more details.

In recent years, ecologists have called for integration of historical data derived from land surveys and other sources (Whitney, 1994) to illuminate landscape processes, thereby guiding management (Turner, 2005); numerous studies demonstrate the utility of such an approach (e.g, Foster and Aber (2004), Rhemtulla et al. (2007), Foster et al. (2008), Steen-Adams et al. (2011)). Environmental history data can meaningfully improve models of landscape change and past structure in several ways. One, land-use history records have provided information about effects of soil alterations (e.g., presence vs. absence of past cultivation) on nutrient cycling and ecological community composition (Dupouey et al., 2002; Fraterrigo et al., 2005). This body of work demonstrates persisting legacy effects of land-use history on biogeochemical processes and landscape structure (Turner, 2005). Two, investigators have tapped land ownership history records to illuminate landscape structure (Foster et al., 1992; Medley et al., 2003). Similarly, research on current ownership condition documents a positive relationship between parcel size and forest patch size (Stanfield et al., 2002). Three, historical forest survey records have enabled ecologists to characterize the temporal and spatial scale of past disturbance regimes (e.g., Schulte and Mladenoff (2005)). Four, historic vegetation data have characterized baseline ecological conditions and thereby enabled ecologists to estimate the historic ranges of variability of desired parameters (e.g., patch size, ecological community composition, species' geographic ranges, and climatic conditions); such estimates can inform restoration and management (Swetnam et al., 1999). Furthermore, advances in Geographic Information System (GIS) technology enable investigators to tap the information contained in historic records (Knowles, 2002; Heasley, 2003). In sum, historic records can reveal important, otherwise unavailable information to improve models of long-term landscape ecological structure and process. Yet because these records are often best capitalized upon when processed into categorical datasets, ecologists need advanced statistical techniques to model nominal categorical data.

Regression is suitable for addressing the ecological questions in our study, as it provides a flexible, yet principled, approach to linking human dimensions to ecological patterns. Practical statistical methods for analyzing spatially-referenced data with multiple categories are limited, despite the increasing abundance of such data. One possible approach is auto-multinomial models where spatial dependence is accounted for by autoregression (Cressie, 1993; Li, 2006). However, the existing models tend to be restricted to either a constant mean or only one covariate at a time. Introduction of autoregression also results in an unknown normalizing constant in the likelihood function, which makes computation for statistical inference challenging (Besag, 1974; Cressie, 1993; Zhu et al., 2005; Zheng and Zhu, 2008). An alternative approach is to formulate the problem in the framework of spatial generalized linear mixed models (GLMM), where the response variable is categorical and is linked to a regression on covariates and a spatial random effect as two additive components in a link function. Diggle et al. (1998) proposed such a modeling framework for non-Gaussian response variables, but did not consider multinomial models explicitly. For spatial *ordinal* data, Higgs and Hoeting (2010) extended the traditional probit models to a general class of models that impose a spatial process on the continuous latent variable. Statistical inference was carried out using Bayesian hierarchical modeling and Markov chain Monte Carlo (MCMC) algorithms. Both Diggle et al. (1998) and Higgs and Hoeting (2010) considered geostatistical data as opposed to lattice data. In contrast, we believe that methodologies for the analysis of spatial *nominal* data on a lattice are not well-developed. An exception is Paciorek and McLachlan

(2009) who did model nominal data on a spatial lattice but did not focus on regression.

Here, we develop a set of statistical tools for the regression analysis of land cover data on a lattice by bringing together several ideas in recent development of spatial statistics. In Sections 2.2 and 2.3, we develop a spatial multinomial model in the spatial GLMM framework, such that the response variable follows a multinomial model and the latent spatial random effect follows a lattice model. As we will show, there are several challenging issues to address in both modeling and computation. First, the spatial random effect here is multivariate at each spatial site, rather than univariate as in most other spatial GLMMs. Our strategy is to utilize a multivariate lattice model (Mardia, 1988; Sain and Cressie, 2007), which can be viewed as an extension from the popular, univariate conditional autoregressive (CAR) model. Like the CAR model, precision rather than variance matrices are specified in these multivariate lattice models, providing a distinct computational advantage. However, since these lattice models for the latent spatial random effects are embedded in a multinomial model, computation can be still demanding. Thus, we further utilize dimension-reduction techniques to help alleviate the computational burden. In particular, we consider spatial linear latent variable models, such that the latent variables at each spatial location are reduced to a smaller number of latent factors (Wang and Wall, 2003; Zhu et al., 2005).

In Section 2.4, we reparameterize the models in Sections 2.2 and 2.3 and adopt Bayesian hierarchical modeling for statistical inference. For computation, we devise MCMC algorithms to obtain the posterior samples. Technical details regarding the link function, a comparison with a multivariate intrinsic CAR model, and the MCMC algorithms are also given in Section 2.5–2.8. In next chapter, we return to analyze the northern Wisconsin land cover data using our new methodology in Chapter 2. The result of the data analysis is presented followed by a discussion of the ecological and management implications.

2.2 Response Variable Model

2.2.1 Multinomial Model

At site $i = 1, \ldots, I$ on a spatial lattice of size $I \in \mathbb{N}$, we let n_i denote the total number of trials and $y_{ji} \in \{0, \ldots, n_i\}$ denote the number of trials that result in the jth category, where $j = 0, \ldots, J$ and $J \in \mathbb{N}$. Let π_{ji} denote the probability of the jth outcome at site i. Thus, $\sum_{j=0}^{J} y_{ji} = n_i$ and $\sum_{j=0}^{J} \pi_{ji} = 1$. Let $\boldsymbol{y}_i = (y_{0i}, \ldots, y_{Ji})'$ denote a $(J+1)$-dimensional multinomial response variable with n_i trials and $J+1$ possible categories with probabilities $\boldsymbol{\pi}_i = (\pi_{0i}, \ldots, \pi_{Ji})'$. The corresponding probability density function for the ith site is

$$p(\boldsymbol{y}_i | n_i, \boldsymbol{\pi}_i) = \binom{n_i}{\boldsymbol{y}_i} \pi_{0i}^{y_{0i}} \cdots \pi_{Ji}^{y_{Ji}} \tag{2.1}$$

(McCullagh and Nelder, 1989).

2.2.2 Link Function

Following the convention of multinomial models, we consider a log-ratio link of the probability π_{ji} for the jth category relative to π_{0i} for the baseline category (Aitchison, 1987; Sain et al., 2006),

$$Z_{ji} = \log(\pi_{ji}/\pi_{0i}), \quad j = 1, \ldots, J. \tag{2.2}$$

Thus, with $Z_{0i} \equiv 0$,

$$\pi_{ji} = \exp(Z_{ji}) \Big\{ \sum_{j=0}^{J} \exp(Z_{ji}) \Big\}^{-1}, \tag{2.3}$$

where $i = 1, \ldots, I$, and $j = 0, \ldots, J$. Let $\boldsymbol{Z}_i = (Z_{1i}, \ldots, Z_{Ji})'$ denote a J-dimensional vector of latent variables at the ith site. The dimension of $\boldsymbol{Z}_i$ is for J categories all in reference to the baseline category. We will describe models for $\{\boldsymbol{Z}_i : i = 1, \ldots, I\}$ in Section 2.3.

2.3 Latent Variable Model

2.3.1 Full-dimensional Model

Following Mardia (1988) and Sain and Cressie (2007), we model the latent variables $\{\boldsymbol{Z}_i : i = 1, \ldots, I\}$ by a multivariate Gaussian process such that the spatial dependence is induced by a multivariate CAR (MCAR) model. Specifically, the Gaussian MCAR model is specified via the conditional probability distributions

$$\mathrm{E}(\boldsymbol{Z}_i | \boldsymbol{Z}_{i'} : i' \in \mathcal{N}_i) = \boldsymbol{\mu}_i + \sum_{i' \in \mathcal{N}_i} \boldsymbol{\Lambda}_{ii'} (\boldsymbol{Z}_{i'} - \boldsymbol{\mu}_{i'}), \quad \mathrm{Var}(\boldsymbol{Z}_i | \boldsymbol{Z}_{i'} : i' \in \mathcal{N}_i) = \boldsymbol{\Gamma}_i. \tag{2.4}$$

We let the mean μ_{ji} follow a linear regression $\mu_{ji} = \boldsymbol{x}_i' \boldsymbol{\beta}_j$, where $\boldsymbol{x}_i$ is a K-dimensional vector of covariates at the ith site $(i = 1, \ldots, I)$ and $\boldsymbol{\beta}_j$ is a K-dimensional vector of regression coefficients for the jth category $(j = 1, \ldots, J)$. Thus, $\boldsymbol{\mu}_i = (\boldsymbol{x}_i' \boldsymbol{\beta})'$, where $\boldsymbol{\beta} = [\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_J]$ is a $K \times J$ matrix of all the regression coefficients.

Furthermore, $\mathcal{N}_i = \{i' : i' \text{ is a neighbor of } i\}$ denotes the site indices of the neighbors of the ith site according to a neighborhood structure. The $J \times J$ matrix $\boldsymbol{\Lambda}_{ii'}$ consists of autoregressive coefficients between the ith site and its neighbor at the i'th site both within the same category and among different categories and thus, induces spatial dependence across sites and categories. The $J \times J$ conditional variance matrix $\mathbf{\Gamma}_i$ allows dependence among categories at a given site.

Let $\boldsymbol{Z} = (\boldsymbol{Z}_1', \ldots, \boldsymbol{Z}_I')'$ denote an (IJ)-dimensional vector of all the latent variables and let $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_I]'$ denote an $I \times K$ design matrix. Then, by (2.4), the joint probability distribution of $\boldsymbol{Z}$ is multivariate Gaussian

$$\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{\Sigma}), \tag{2.5}$$

where the mean vector and the variance matrix are

$$\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_I')' = \mathrm{vec}\{(\boldsymbol{X}\boldsymbol{\beta})'\}, \quad \mathbf{\Sigma} = \left\{\mathrm{blk}(-\mathbf{\Gamma}_i^{-1}\mathbf{\Lambda}_{ii'})\right\}^{-1},$$

$\mathrm{blk}(\boldsymbol{A}_{ii'})$ denotes a matrix with blocks $\boldsymbol{A}_{ii'}$ row indexed by i and column indexed by i' for $i, i' = 1, \ldots, I$. The following assumptions ensure that the variance matrix $\mathbf{\Sigma}$ is symmetric and positive definitive (Mardia, 1988; Sain and Cressie, 2007):

(A.1) $\mathbf{\Lambda}_{ii'}\mathbf{\Gamma}_{i'} = \mathbf{\Gamma}_i\mathbf{\Lambda}_{i'i}'$, where $\mathbf{\Lambda}_{ii} = -\boldsymbol{I}_J$ and $\mathbf{\Lambda}_{ii'} = \mathbf{0}$ for $i' \notin \mathcal{N}_i \cup \{i\}$.

(A.2) $\mathrm{blk}(-\mathbf{\Gamma}_i^{-1}\mathbf{\Lambda}_{ii'})$ or $\mathrm{blk}(-\mathbf{\Lambda}_{ii'})$ is positive definite.

We will refer to the model specified by (2.1)–(2.2) and (2.4)–(2.5) as a spatial multinomial model with a full-dimensional latent MCAR model or, simply, the *full spatial multinomial model*.

2.3.2 Reduced-dimensional Model

For more than two categories $J > 1$, we further develop a lower-dimensional model for the latent variables $\{Z_{ji}\}$ as

$$\boldsymbol{Z}_i = \boldsymbol{\mu}_i + \mathbf{\Omega}\boldsymbol{Z}_i^*, \tag{2.6}$$

where $\boldsymbol{\mu}_i = (\boldsymbol{x}_i'\boldsymbol{\beta})'$, $\boldsymbol{\Omega}$ is a $J \times J^*$ matrix, and $\boldsymbol{Z}_i^*$ is a J^*-dimensional vector of zero-mean latent variables with $1 \le J^* \le J-1$. There are at least two reasons to consider a lower-dimensional model. One is to reduce the computational cost, as the dimension is reduced from J to $J^* < J$. Two, the new latent variable $\boldsymbol{Z}_i^*$ can be interpreted as shared spatial effects due to unobserved environmental factors. When $J^* = 1$, the shared spatial effect is also known as a common spatial factor (Wang and Wall, 2003).

We then impose an MCAR model on $\boldsymbol{Z}_i^*$ via conditional probability distributions

$$\mathrm{E}(\boldsymbol{Z}_i^* | \boldsymbol{Z}_{i'}^* : i' \in \mathcal{N}_i) = \sum_{i' \in \mathcal{N}_i} \boldsymbol{\Lambda}_{ii'}^* \boldsymbol{Z}_{i'}^*, \quad \mathrm{Var}(\boldsymbol{Z}_i^* | \boldsymbol{Z}_{i'}^* : i' \in \mathcal{N}_i) = \boldsymbol{\Gamma}_i^*, \tag{2.7}$$

where $\boldsymbol{\Lambda}_{ii'}^*$ and $\boldsymbol{\Gamma}_i^*$ are $J^* \times J^*$ matrices satisfying conditions similar to (A.1) and (A.2) for $\boldsymbol{\Lambda}_{ii'}$ and $\boldsymbol{\Gamma}_i$. We will refer to the model specified by (2.1)–(2.2) and (2.6)–(2.7) as a spatial multinomial model with a reduced-dimensional latent MCAR model or, simply, the *reduced spatial multinomial model*. In the case $J^* = 1$, we will refer to the model as a *common-factor spatial multinomial model*.

2.4 Statistical Inference

In this section, we develop statistical inference for both the full and the reduced spatial multinomial models.

2.4.1 Full Spatial Multinomial Model

For a full spatial multinomial model, we let $\boldsymbol{\Lambda}_{ii'} = \boldsymbol{\Lambda}$ for $i' \in \mathcal{N}_i$ and $\boldsymbol{\Gamma}_i = \boldsymbol{\Gamma}$, where $\boldsymbol{\Lambda}$ is a symmetric matrix and $\boldsymbol{\Gamma}$ is a variance matrix both of dimension $J \times J$. Thus, $\boldsymbol{\Lambda}_{i'i} = \boldsymbol{\Lambda}_{ii'}$ and from (A.1), $\boldsymbol{\Lambda\Gamma} = \boldsymbol{\Gamma\Lambda}$. Further, the precision matrix $\boldsymbol{\Sigma}^{-1}$ has $\boldsymbol{\Gamma}^{-1}$ as the ith diagonal block, $-\boldsymbol{\Gamma}^{-1}\boldsymbol{\Lambda}$ as the the upper- and lower-triangular blocks for the (i, i')th block such that $i' \in \mathcal{N}_i$, and $\mathbf{0}$ otherwise. It can be shown that $\boldsymbol{\Sigma} = \text{diag}\{\boldsymbol{\Gamma}\}(\boldsymbol{I} - \boldsymbol{C} \otimes \boldsymbol{\Lambda})^{-1}$, where $\otimes$ denotes the Kronecker product and $\boldsymbol{C} = [\mathcal{I}(i \sim i')]_{i,i'=1}^{I}$ indicates whether site i' is a neighbor of site i according to a neighbor structure. Here $\mathcal{I}(\cdot)$ is an indicator function and $i \sim i'$ denotes that $i' \in \mathcal{N}_i$. Also, let $\boldsymbol{V} = \boldsymbol{\Gamma}^{-1}$. Thus, the model parameters are $\{\boldsymbol{\beta}_j\}_{j=1}^{J}$, $\boldsymbol{V}$, and $\boldsymbol{\Lambda}$.

We adopt Bayesian hierarchical modeling for statistical inference and devise MCMC algorithms to obtain the posterior. Independent prior distributions for the model parameters are assumed, with probability density functions $p(\boldsymbol{\beta}_j)$, $p(\boldsymbol{V})$, and $p(\boldsymbol{\Lambda})$. In particular, we follow Sain and Cressie (2007) and let

$$\begin{aligned} \boldsymbol{\beta}_j &\sim \text{N}(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I}), \quad \boldsymbol{V} \sim \text{W}(\rho, \psi \boldsymbol{I}), \rho > J, \\ p(\boldsymbol{\Lambda}) &\propto \exp\{-\text{vec}(\boldsymbol{\Lambda})'\text{vec}(\boldsymbol{\Lambda})/\xi^2\}, \end{aligned} \tag{2.8}$$

where W refers to a Wishart distribution, σ_β^2, ρ, $\psi \boldsymbol{I}$, and ξ^2 are the hyperparameters corresponding to variance, degrees of freedom, scale matrix, and scale parameter, respectively. The resulting posterior distribution of $\boldsymbol{Z}, \{\boldsymbol{\beta}_j\}, \boldsymbol{V}$, and $\boldsymbol{\Lambda}$ has, up to a proportionality constant, the following probability density function

$$p(\boldsymbol{Z}, \{\boldsymbol{\beta}_j\}, \boldsymbol{V}, \boldsymbol{\Lambda} | \boldsymbol{y}) \;\propto\; p(\boldsymbol{y}|\boldsymbol{Z}) p(\boldsymbol{Z}|\{\boldsymbol{\beta}_j\}, \boldsymbol{V}, \boldsymbol{\Lambda}) \prod_{j=1}^{J} p(\boldsymbol{\beta}_j) p(\boldsymbol{V}) p(\boldsymbol{\Lambda}). \tag{2.9}$$

Even though the prior specification (2.8) is similar to Sain and Cressie (2007), the posterior distribution (2.9) is different, because the data distribution $p(\boldsymbol{y}|\boldsymbol{Z})$ is multinomial, not Poisson as in Sain and Cressie (2007). Thus, new MCMC algorithms need to be devised.

To sample from (2.9), we use a Gibbs sampler and sample from the full conditional distributions of $\boldsymbol{Z}, \{\boldsymbol{\beta}_j\}, \boldsymbol{V}$, and $\boldsymbol{\Lambda}$. It can be shown that the full conditional distribution of $\boldsymbol{\beta}_j$ is Gaussian with mean and variance

$$\boldsymbol{\mu}_{\boldsymbol{\beta}_j} = \boldsymbol{\Sigma}_{\boldsymbol{\beta}_j}\left\{\boldsymbol{X}'\boldsymbol{S}_{jj}\boldsymbol{Z}_{j\cdot} + \boldsymbol{X}'\sum_{k\neq j}\boldsymbol{S}_{jk}(\boldsymbol{Z}_{k\cdot} - \boldsymbol{X}\boldsymbol{\beta}_k)\right\}, \quad \boldsymbol{\Sigma}_{\boldsymbol{\beta}_j} = (\boldsymbol{X}'\boldsymbol{S}_{jj}\boldsymbol{X} + \sigma_\beta^{-2}\boldsymbol{I})^{-1},$$

where $\{\boldsymbol{S}_{jk}\}$ is defined in Section 2.7.1. Further, the full conditional distribution of $\boldsymbol{V}$ is $\mathrm{W}(I+\rho, \bar{\boldsymbol{\Psi}})$, where $\bar{\boldsymbol{\Psi}}$ is defined in Section 2.7.2. Thus, we may directly sample $\{\boldsymbol{\beta}_j\}$ and $\boldsymbol{V}$ from their full conditional distributions. For the remainder model parameters, the full conditional distributions are not in closed form and thus, we use Metropolis-Hastings (MH) algorithms to simulate from the full conditional distributions. In particular, the proposal distributions for $\boldsymbol{\Lambda}$ and $\boldsymbol{Z}$ are set to uniform and Gaussian, respectively. The parameters of these proposal distributions are tuned to improve convergence and mixing of the Markov chains. For details of the MCMC algorithms, see Section 2.7.

2.4.2 Reduced Spatial Multinomial Model

For a reduced spatial multinomial model, we reparameterize as in the previous subsection but replace $\boldsymbol{\Lambda}$ and $\boldsymbol{\Gamma}$ of dimension $J \times J$ with $\boldsymbol{\Lambda}^*$ and $\boldsymbol{\Gamma}^*$ of a lower dimension $J^* \times J^*$. Analogously define $\boldsymbol{V}^* = \boldsymbol{\Gamma}^{*-1}$. The model parameters are $\{\boldsymbol{\beta}_j\}_{j=1}^J$, $\boldsymbol{V}^*$, $\boldsymbol{\Lambda}^*$, and $\boldsymbol{\Omega}$.

Again, independent prior distributions for the model parameters are assumed, with probability density functions $p(\boldsymbol{\Omega})$, $p(\boldsymbol{\beta}_j)$, $p(\boldsymbol{V}^*)$, and $p(\boldsymbol{\Lambda}^*)$. In particular, let

$$\begin{aligned} \mathrm{vec}(\boldsymbol{\Omega}) \sim \mathrm{N}(\boldsymbol{0}, \sigma_\Omega^2\boldsymbol{I}), \quad \boldsymbol{\beta}_j &\sim \mathrm{N}(\boldsymbol{0}, \sigma_\beta^2\boldsymbol{I}), \\ \boldsymbol{V}^* \sim \mathrm{W}(\rho^*, \psi^*\boldsymbol{I}), \rho^* > J^*, \quad p(\boldsymbol{\Lambda}^*) &\propto \exp\{-\mathrm{vec}(\boldsymbol{\Lambda}^*)'\mathrm{vec}(\boldsymbol{\Lambda}^*)/\xi^{*2}\}, \end{aligned} \tag{2.10}$$

where W refers to a Wishart distribution, σ_Ω^2, σ_β^2, ρ^*, $\psi^*\boldsymbol{I}$, and ξ^{*2} are the hyperparameters corresponding to variances, degrees of freedom, scale matrix, and scale parameter, respectively. The resulting posterior distribution of $\boldsymbol{Z}^*, \boldsymbol{\Omega}, \{\boldsymbol{\beta}_j\}, \boldsymbol{V}^*$ and $\boldsymbol{\Lambda}^*$ has the following probability density function, up to a proportionality constant,

$$p(\boldsymbol{Z}^*, \boldsymbol{\Omega}, \{\boldsymbol{\beta}_j\}, \boldsymbol{V}^*, \boldsymbol{\Lambda}^*|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{Z}^*, \boldsymbol{\Omega}, \{\boldsymbol{\beta}_j\})p(\boldsymbol{Z}^*|\boldsymbol{V}^*, \boldsymbol{\Lambda}^*)p(\boldsymbol{\Omega})\prod_{j=1}^{J} p(\boldsymbol{\beta}_j)p(\boldsymbol{V}^*)p(\boldsymbol{\Lambda}^*) \quad (2.11)$$

Again, we use a Gibbs sampler to sample $\boldsymbol{Z}^*, \boldsymbol{\Omega}, \{\boldsymbol{\beta}_j\}, \boldsymbol{V}^*$ and $\boldsymbol{\Lambda}^*$ from their respective full conditional distributions. Because only $\boldsymbol{V}^*$ has conjugacy, we choose to use a MH algorithm to sample the other parameters from their corresponding full conditional distributions. In the case $J^* = 1$, the matrix $\boldsymbol{V}^*$ is reduced to a scalar v^*. The prior of v^* is set to a gamma distribution to achieve conjugacy for v^*. Thus, we may directly sample v^* from its full conditional distribution. Details of the MCMC algorithms are given in Section 2.8.

2.5 Relation to MICAR

Other alternative models for the latent variables are possible such as multivariate intrinsic CAR models (MICAR) (Jin et al., 2005). We compare MCAR models with the MICAR models and demonstrate that there is a close relationship between the two types of models as follows.

As in Jin et al. (2005), let $\boldsymbol{\mu}_i = \boldsymbol{0}$ and the conditional distributions be

$$\boldsymbol{Z}_i|\{\boldsymbol{Z}_{i'}, i' \in \mathcal{N}_i\} \sim \mathrm{N}(\boldsymbol{R}_i \sum_{i' \in \mathcal{N}_i} \boldsymbol{B}_{ii'}\boldsymbol{Z}_{i'}, \boldsymbol{\Gamma}_i),$$

where $\boldsymbol{\Gamma}_i$, $\boldsymbol{R}_i$ and $\boldsymbol{B}_{ii'}$ are $J \times J$ matrices. Thus, the corresponding joint distribution is

$$\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{0}, \{\boldsymbol{\Gamma}^{-1}(\boldsymbol{I}_{IJ} - \boldsymbol{B}_R)\}^{-1}), \quad (2.12)$$

where $\boldsymbol{B}_R$ is an $IJ \times IJ$ matrix with blocks $(\boldsymbol{B}_R)_{ii'} = \boldsymbol{R}_i\boldsymbol{B}_{ii'}$, $(\boldsymbol{B}_R)_{ii} = \boldsymbol{0}$, and $\boldsymbol{\Gamma}$ is an $IJ \times IJ$ block diagonal matrix with diagonal blocks $\boldsymbol{\Gamma}_i$ (Mardia, 1988). It is further assumed that $\boldsymbol{R}_i = \alpha \boldsymbol{I}_J$ for $i = 1, \ldots, I$, $\boldsymbol{B}_{ii'} = \mathcal{I}(i \sim i')/m_i \boldsymbol{I}_J$, with $m_i > 0$, and $\boldsymbol{\Gamma}^{-1} = \boldsymbol{D} \otimes \boldsymbol{\Lambda}$, where $\boldsymbol{D} = \mathrm{diag}\{m_i\}_{i=1}^{I}$ and $\boldsymbol{\Lambda}$ is a $J \times J$ positive definite and symmetric matrix. Then the joint distribution (2.12) becomes

$$\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{0}, \{(\boldsymbol{D} - \alpha \boldsymbol{C}) \otimes \boldsymbol{\Lambda}\}^{-1}), \tag{2.13}$$

where matrix $\boldsymbol{C}$ is given in Section 2.4.1. When $\alpha = 1$, (2.13) becomes an MICAR model,

$$\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{0}, \{\boldsymbol{D} \otimes \boldsymbol{\Lambda} - \boldsymbol{C} \otimes \boldsymbol{\Lambda}\}^{-1}). \tag{2.14}$$

In our full spatial multinomial model, if we set $\boldsymbol{\mu}_i = \boldsymbol{0}$ correspondingly, then

$$\begin{aligned} \boldsymbol{Z}_i | \{\boldsymbol{Z}_{i'}, i' \in \mathcal{N}_i\} &\sim \mathrm{N}(\sum_{i' \in \mathcal{N}_i} \boldsymbol{\Lambda}_{ii'} \boldsymbol{Z}_i, \boldsymbol{\Gamma}_i) \\ \boldsymbol{Z} &\sim \mathrm{N}(\boldsymbol{0}, \{\mathrm{blk}(-\boldsymbol{\Gamma}_i^{-1} \boldsymbol{\Lambda}_{ii'})\}^{-1}). \end{aligned} \tag{2.15}$$

It is straightforward to verify that (2.12) and (2.15) are equivalent, since $\boldsymbol{\Lambda}_{ii'}$ can be reparameterized to be in the form of $\boldsymbol{R}_i \boldsymbol{B}_{ii'}$. After the simplification of parameterizations in Section 2.4.1, (2.15) becomes

$$\boldsymbol{Z} \sim \mathrm{N}(\boldsymbol{0}, (\mathrm{diag}\{\boldsymbol{V}\} - \boldsymbol{C} \otimes \boldsymbol{V} \boldsymbol{\Lambda})^{-1}). \tag{2.16}$$

Equations (2.14) and (2.16) are similar. The difference is that $\boldsymbol{\Gamma} = \boldsymbol{V}^{-1}$ replaces $\boldsymbol{\Lambda}^{-1}/m_i$ to represent the conditional variance at site i and $\boldsymbol{\Lambda}$ replaces $\boldsymbol{I}_J/m_i$ to represent the correlation between neighbors. Thus, (2.14) and (2.16) are the same only if all m_i's are equal in MICAR model and $\boldsymbol{\Lambda}$ is an identity matrix in our MCAR model. The distinctions among conditional random effects at different sites are not considered in our model, but the correlations among categories between neighbors are modeled via the off-diagonal elements in our matrix $\boldsymbol{\Lambda}$.

2.6 Details of Link Functions

A practical issue to consider is the choice of the baseline category in a multinomial model for nominal categorical data, which we will discuss in detail as follows.

2.6.1 Log-ratio Link Function

Here we investigate the effect of baseline category choice on $\boldsymbol{Z}_i$ in terms of the mean and the dependence structures. Consider model (2.2), with category 0 as the baseline and an alternative model

$$Z_{ji}^{\dagger} = \log(\pi_{ji}/\pi_{j^{\dagger}i}), \quad j = 0, \ldots, j^{\dagger}-1, j^{\dagger}+1, \ldots, J, \tag{2.17}$$

with category $j^{\dagger}$ as an alternative baseline category such that $1 \leq j^{\dagger} \leq J$. Let $\boldsymbol{Z}_i^{\dagger} = (Z_{0i}^{\dagger}, \ldots, Z_{(j^{\dagger}-1)i}^{\dagger}, Z_{(j^{\dagger}+1)i}^{\dagger}, \ldots, Z_{Ji}^{\dagger})'$. It can be shown that $\boldsymbol{Z}_i^{\dagger} = \boldsymbol{A}_{j^{\dagger}}\boldsymbol{Z}_i$, where

$$\boldsymbol{A}_{j^{\dagger}} = \begin{bmatrix} \mathbf{0}' & -1 & \mathbf{0}' \\ \boldsymbol{I}_{j^{\dagger}-1} & \mathbf{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{-1} & \boldsymbol{I}_{J-j^{\dagger}} \end{bmatrix}. \tag{2.18}$$

Suppose $\boldsymbol{Z}_i \sim (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ with mean $\boldsymbol{\mu}_i$ and variance $\boldsymbol{\Sigma}_i$ and $\boldsymbol{Z}_i^{\dagger} \sim (\boldsymbol{\mu}_i^{\dagger}, \boldsymbol{\Sigma}_i^{\dagger})$ with mean $\boldsymbol{\mu}_i^{\dagger}$ and variance $\boldsymbol{\Sigma}_i^{\dagger}$. It follows that

$$\boldsymbol{\mu}_i^{\dagger} = \boldsymbol{A}_{j^{\dagger}}\boldsymbol{\mu}_i \quad \text{and} \quad \boldsymbol{\Sigma}_i^{\dagger} = \boldsymbol{A}_{j^{\dagger}}\boldsymbol{\Sigma}_i\boldsymbol{A}_{j^{\dagger}}'.$$

Conversely, we have

$$\boldsymbol{\mu}_i = \boldsymbol{A}_{j^{\dagger}}^{-1}\boldsymbol{\mu}_i^{\dagger} \quad \text{and} \quad \boldsymbol{\Sigma}_i = \boldsymbol{A}_{j^{\dagger}}^{-1}\boldsymbol{\Sigma}_i^{\dagger}(\boldsymbol{A}_{j^{\dagger}}')^{-1}.$$

Since $|\boldsymbol{A}_{j^{\dagger}}| = (-1)^{j^{\dagger}}$, it is obvious that there is a one-to-one correspondence between the two mean vectors $\boldsymbol{\mu}_i$ and $\boldsymbol{\mu}_i^{\dagger}$, as well as the two variance matrices $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Sigma}_i^{\dagger}$.

A cautionary note, however, is that interpretation of the dependence structure in the variance matrix may not be straightforward. We illustrate this by the northern Wisconsin land-cover data. Let π_{OTH}, π_{APB} and π_{AG} denote the probabilities of land cover types OTH ($j = 0$), APB ($j = 1$) and AG ($j = 2$). Suppose OTH is the baseline category, then $\boldsymbol{Z}_i \sim (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ with mean $\boldsymbol{\mu}_i = (\mu_{1i}, \mu_{2i})'$ and variance

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} \Sigma_i^{11} & \Sigma_i^{12} \\ \Sigma_i^{21} & \Sigma_i^{22} \end{bmatrix}, \tag{2.19}$$

where $Z_{1i} = \log(\pi_{\text{APB},i}/\pi_{\text{OTH},i})$, $Z_{2i} = \log(\pi_{\text{AG},i}/\pi_{\text{OTH},i})$, and $\Sigma_i^{12} = \Sigma_i^{21}$.

Now, suppose AG is the baseline category with $j^\dagger = 2$ and by (2.18), $\boldsymbol{A}_2^\dagger = \begin{bmatrix} 0 & -1 \\ 1 & -1 \end{bmatrix}$. Then, $\boldsymbol{Z}_i^\dagger \sim (\boldsymbol{\mu}_i^\dagger, \boldsymbol{\Sigma}_i^\dagger)$ with mean $\boldsymbol{\mu}_i^\dagger = (-\mu_{2i}, \mu_{1i} - \mu_{2i})'$ and variance

$$\boldsymbol{\Sigma}_i^\dagger = \begin{bmatrix} \Sigma_i^{22} & \Sigma_i^{22} - \Sigma_i^{12} \\ \Sigma_i^{22} - \Sigma_i^{12} & \Sigma_i^{11} - 2\Sigma_i^{12} + \Sigma_i^{22} \end{bmatrix}, \tag{2.20}$$

where $Z_{1i}^\dagger = \log(\pi_{\text{OTH},i}/\pi_{\text{AG},i}) = -Z_{2i}$ and $Z_{2i}^\dagger = \log(\pi_{\text{APB},i}/\pi_{\text{AG},i}) = Z_{1i} - Z_{2i}$.

Consider the case where Z_{ji}'s are uncorrelated across categories and thus, $\boldsymbol{\Sigma}_i$ in (2.19) is a diagonal matrix. After changing the baseline category from 0 to 2, $Z_{ji}^\dagger$'s are no longer uncorrelated across categories, as the covariances are not all 0's in (2.20).

2.6.2 Alternative Link Function

Under the log ratio link function (2.2), it has be shown, in previous section, that there is a one-to-one correspondence between the model under two different baseline category choices in terms of the mean vector and the variance-covariance matrix of the latent variables. However, the interpretation of dependence structure may not be straightforward. An alternative approach is to consider the log of the probability π_{ji} for the ith site and the jth category,

$$\log(\pi_{ji}) = \tilde{Z}_{ji} - \log(C_i), \quad j = 0, \ldots, J, \tag{2.21}$$

where $\tilde{Z}_{ji}$ is a latent variable (different from Z_{ji}) and $C_i = \sum_{j=0}^{J} \exp(\tilde{Z}_{ji})$. That is, $\pi_{ji} = \exp(\tilde{Z}_{ji})/C_i$ and is proportional to $\exp(\tilde{Z}_{ji})$. An advantage of (2.21) is that it overcomes the difficulty in interpretation encountered in (2.2), as the dependence structures of the latent variables are not subject to the choice of baseline category. We now model $\tilde{\boldsymbol{Z}}_i$ in a CAR model. Since the only difference is the dimension, we omit repeating the model formulation, but highlight the major differences in notation.

In particular, we replace (2.4) with

$$\mathrm{E}(\tilde{\boldsymbol{Z}}_i|\tilde{\boldsymbol{Z}}_{i'} : i' \in \mathcal{N}_i) = \tilde{\boldsymbol{\mu}}_i + \sum_{i' \in \mathcal{N}_i} \tilde{\boldsymbol{\Lambda}}_{ii'}(\tilde{\boldsymbol{Z}}_{i'} - \tilde{\boldsymbol{\mu}}_{i'}), \quad \mathrm{Var}(\tilde{\boldsymbol{Z}}_i|\tilde{\boldsymbol{Z}}_{i'} : i' \in \mathcal{N}_i) = \tilde{\boldsymbol{\Gamma}}_i, \tag{2.22}$$

where $\tilde{\mu}_{ji} = \boldsymbol{x}_i'\tilde{\boldsymbol{\beta}}_j$, $\tilde{\boldsymbol{\mu}}_i = (\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}})'$, and $\tilde{\boldsymbol{\beta}} = [\tilde{\boldsymbol{\beta}}_0, \ldots, \tilde{\boldsymbol{\beta}}_J]$ is a $K \times (J+1)$ matrix of regression coefficients. The resulting model will be referred to as the full spatial multinomial model under the alternative link function. Note that $\boldsymbol{\beta}_j = \tilde{\boldsymbol{\beta}}_j - \tilde{\boldsymbol{\beta}}_0$ for $j = 1, \ldots, J$, where $\boldsymbol{\beta}_j$ is the vector of regression coefficients in μ_{ji} given below (2.4).

For a lower-dimensional model of the latent variables $\{\tilde{Z}_{ji}\}$, we replace (2.6) with

$$\tilde{\boldsymbol{Z}}_i = \tilde{\boldsymbol{\mu}}_i + \tilde{\boldsymbol{\Omega}}\boldsymbol{Z}_i^*, \tag{2.23}$$

where $\tilde{\boldsymbol{\Omega}}$ is a $(J+1) \times J^*$ matrix and $\boldsymbol{Z}_i^*$ is the same as in (2.6) except that the range of J^* is now $1 \leq J^* \leq J$. We then impose an MCAR model on $\boldsymbol{Z}_i^*$ as in (2.7). We will refer to the resulting model as the reduced spatial multinomial model under the alternative link function.

Let $\tilde{\boldsymbol{Z}}_i = (\tilde{Z}_{0i}, \ldots, \tilde{Z}_{Ji})'$ denote a $(J+1)$-dimensional multivariate latent variable at the ith site. Suppose $\tilde{\boldsymbol{Z}}_i \sim (\tilde{\boldsymbol{\mu}}_i, \tilde{\boldsymbol{\Sigma}}_i)$ with mean $\tilde{\boldsymbol{\mu}}_i$ and variance $\tilde{\boldsymbol{\Sigma}}_i$. Also, for $\boldsymbol{Z}_i \sim$

$(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ with category $j^* = 0$ as the baseline, it can be shown that

$$\boldsymbol{Z}_i = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{Z}}_i, \quad \text{where} \quad \tilde{\boldsymbol{A}} = \begin{bmatrix} \boldsymbol{I}_{j^*} & -\mathbf{1} & \boldsymbol{O} \\ \boldsymbol{O} & -\mathbf{1} & \boldsymbol{I}_{J-j^*} \end{bmatrix} \quad \text{and} \quad j^* = 0. \tag{2.24}$$

It follows that $\boldsymbol{\mu}_i = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{\mu}}_i$ and $\boldsymbol{\Sigma}_i = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{\Sigma}}_i\tilde{\boldsymbol{A}}'$. Because $\log(\pi_{ji})$ involves an additive constant $\log(C_i)$ from (2.21), it depends on $\exp(\tilde{Z}_{ji})$ for $j = 0, \ldots, J$. Thus, the interpretation of $\tilde{\mu}_{ji}$ in relation to the probability π_{ji} is not as straightforward. However, $\log(\pi_{ji}/\pi_{0i}) = \tilde{Z}_{ji} - \tilde{Z}_{0i}$, which has mean $\tilde{\mu}_{ji} - \tilde{\mu}_{0i}$. For linear regression with means $\mu_{ji} = \boldsymbol{x}_i'\boldsymbol{\beta}_j$ $(j \neq 0)$ and $\tilde{\mu}_{ji} = \boldsymbol{x}_i'\tilde{\boldsymbol{\beta}}_j$ $(j = 0, \ldots, J)$, we have $\boldsymbol{\beta}_j = \tilde{\boldsymbol{\beta}}_j - \tilde{\boldsymbol{\beta}}_0$, where $\boldsymbol{x}_i \in \mathbb{R}^K$ is a K-dimensional vector of covariates at the ith site and $\boldsymbol{\beta}_j, \tilde{\boldsymbol{\beta}}_j \in \mathbb{R}^K$ are K-dimensional vectors of regression coefficients for the jth category.

After projecting $\tilde{\boldsymbol{\mu}}_i$ from the $(J+1)$-dimension down to the J-dimension using (2.24), the resulting $\boldsymbol{\mu}_i$ has the same interpretation as the conventional model under log-ratio link.

2.7 Statistical Inference for Full Spatial Multinomial Models

2.7.1 Posterior Sample of $\{\boldsymbol{\beta}_j\}$

Since $\boldsymbol{Z} \sim \mathrm{N}_{IJ}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ under log ratio link ($\boldsymbol{Z} \sim \mathrm{N}_{I(J+1)}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ under alternative link) and $\boldsymbol{\beta}_j \sim \mathrm{N}_K(\mathbf{0}, \sigma_\beta^2 \boldsymbol{I}_K)$, the full conditional probability density of $\boldsymbol{\beta}_j$ is

$$p(\boldsymbol{\beta}_j|\boldsymbol{Z}, \boldsymbol{V}, \boldsymbol{\Lambda}, \boldsymbol{y}) \propto p(\boldsymbol{Z}|\{\boldsymbol{\beta}_j\}, \boldsymbol{V}, \boldsymbol{\Lambda})p(\boldsymbol{\beta}_j) \propto \exp\{-(\boldsymbol{Z} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})/2\}\exp\{-\boldsymbol{\beta}_j'\boldsymbol{\beta}_j/2\sigma_\beta^2\}.$$

Let $\boldsymbol{U} \equiv \boldsymbol{V}\boldsymbol{\Lambda}$, then $\boldsymbol{\Sigma}^{-1} = \mathrm{blk}\big\{\boldsymbol{V}\mathcal{I}(i=i') - \boldsymbol{U}\mathcal{I}(i \sim i')\big\}$ and

$$(\boldsymbol{Z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{Z}-\boldsymbol{\mu}) = \sum_{i=1}^{I}\big\{(\boldsymbol{Z}_i-\boldsymbol{\mu}_i)'\boldsymbol{V}(\boldsymbol{Z}_i-\boldsymbol{\mu}_i) - \sum_{i'\in\mathcal{N}_i}(\boldsymbol{Z}_i-\boldsymbol{\mu}_i)'\boldsymbol{U}(\boldsymbol{Z}_{i'}-\boldsymbol{\mu}_{i'})\big\}.$$

Focusing on the jth category, we collect the terms involving Z_{ji} and μ_{ji} $(i = 1, \ldots, I)$ and obtain

$$\begin{aligned}\sum_{i=1}^{I}\Big[&(Z_{ji}-\mu_{ji})V_{jj}(Z_{ji}-\mu_{ji}) + 2\sum_{k\neq j}(Z_{ji}-\mu_{ji})V_{jk}(Z_{ki}-\mu_{ki})\\ &- \sum_{i'\in\mathcal{N}_i}\big\{(Z_{ji}-\mu_{ji})U_{jj}(Z_{ji'}-\mu_{ji'}) + 2\sum_{k\neq j}(Z_{ji}-\mu_{ji})U_{jk}(Z_{ki'}-\mu_{ki'})\big\}\Big],\end{aligned} \tag{2.25}$$

where V_{jk} (U_{jk}) denotes the (j,k)th element of $\boldsymbol{V}$ $(\boldsymbol{U})$. Let $\boldsymbol{Z}_{j\cdot} = (Z_{j1}, \ldots, Z_{jI})'$ and $\boldsymbol{\mu}_{j\cdot} = \boldsymbol{X}\boldsymbol{\beta}_j = (\mu_{j1}, \ldots, \mu_{jI})'$. Then (2.25) becomes

$$(\boldsymbol{Z}_{j\cdot}-\boldsymbol{\mu}_{j\cdot})'\boldsymbol{S}_{jj}(\boldsymbol{Z}_{j\cdot}-\boldsymbol{\mu}_{j\cdot}) + 2(\boldsymbol{Z}_{j\cdot}-\boldsymbol{\mu}_{j\cdot})'\sum_{k\neq j}\boldsymbol{S}_{jk}(\boldsymbol{Z}_{k\cdot}-\boldsymbol{\mu}_{k\cdot}),$$

where $\boldsymbol{S}_{jk} = V_{jk}\boldsymbol{I}_I - U_{jk}\boldsymbol{C}$ for $j,k = 1, \ldots, J$ under log ratio link $(j,k = 0, \ldots, J$ under alternative link), and $\boldsymbol{C}$ is given in Section 2.4.1. Thus, the full conditional probability density of $\boldsymbol{\beta}_j$ is proportional to

$$\begin{aligned}&\exp\big\{-(\boldsymbol{Z}_{j\cdot}-\boldsymbol{\mu}_{j\cdot})'\boldsymbol{S}_{jj}(\boldsymbol{Z}_{j\cdot}-\boldsymbol{\mu}_{j\cdot})/2 + (\boldsymbol{Z}_{j\cdot}-\boldsymbol{\mu}_{j\cdot})'\sum_{k\neq j}\boldsymbol{S}_{jk}(\boldsymbol{Z}_{k\cdot}-\boldsymbol{\mu}_{k\cdot})\big\}\exp(-\boldsymbol{\beta}_j{}'\boldsymbol{\beta}_j/2\sigma_\beta^2)\\ =\;&\exp\Big[-\boldsymbol{\beta}_j{}'(\boldsymbol{X}'\boldsymbol{S}_{jj}\boldsymbol{X}+\sigma_\beta^{-2}\boldsymbol{I})\boldsymbol{\beta}_j/2 - \boldsymbol{\beta}_j{}'\big\{\boldsymbol{X}'\boldsymbol{S}_{jj}\boldsymbol{Z}_{j\cdot} + \boldsymbol{X}'\sum_{k\neq j}\boldsymbol{S}_{jk}(\boldsymbol{Z}_{k\cdot}-\boldsymbol{X}\boldsymbol{\beta}_k)\big\}\Big].\end{aligned}$$

That is, the full conditional distribution of $\boldsymbol{\beta}_j$ is Gaussian with variance $\boldsymbol{\Sigma}_{\boldsymbol{\beta}_j} \equiv (\boldsymbol{X}'\boldsymbol{S}_{jj}\boldsymbol{X} + \boldsymbol{I}/\sigma_\beta^2)^{-1}$ and mean $\boldsymbol{\mu}_{\boldsymbol{\beta}_j} \equiv \boldsymbol{\Sigma}_{\boldsymbol{\beta}_j}\big\{\boldsymbol{X}'\boldsymbol{S}_{jj}\boldsymbol{Z}_{j\cdot} + \boldsymbol{X}'\sum_{k\neq j}\boldsymbol{S}_{jk}(\boldsymbol{Z}_{k\cdot}-\boldsymbol{X}\boldsymbol{\beta}_k)\big\}$.

2.7.2 Posterior Sample of $\boldsymbol{V}$

Since $\boldsymbol{Z} \sim \mathrm{N}_{IJ}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ under log ratio link $(\boldsymbol{Z} \sim \mathrm{N}_{I(J+1)}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ under alternative link) and $\boldsymbol{V} \sim \mathrm{W}(\rho, \psi\boldsymbol{I})$, let $\boldsymbol{\varepsilon} = \boldsymbol{Z} - \boldsymbol{\mu}$ and $\boldsymbol{\Psi} = \psi\boldsymbol{I}$, then the full conditional probability density of $\boldsymbol{V}$ is

$$\begin{aligned} p(\boldsymbol{V}|\boldsymbol{Z},\{\boldsymbol{\beta}_j\},\boldsymbol{\Lambda},\boldsymbol{y}) &\propto p(\boldsymbol{\varepsilon}|\boldsymbol{V},\boldsymbol{\Lambda})p(\boldsymbol{V}) \\ &\propto |\boldsymbol{\Sigma}|^{-1/2}\exp(-\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon}/2)|\boldsymbol{V}|^{(\rho-J-1)/2}\exp\big\{-\mathrm{tr}(\boldsymbol{\Psi}^{-1}\boldsymbol{V})/2\big\}. \end{aligned}$$

Further, because

$$|\boldsymbol{\Sigma}| = |\mathrm{diag}\{\boldsymbol{\Gamma}\}(\boldsymbol{I}-\boldsymbol{C}\otimes\boldsymbol{\Lambda})^{-1}| \propto |\boldsymbol{\Gamma}|^{I} = |\boldsymbol{V}|^{-I},$$

and

$$\begin{aligned} \boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\varepsilon} &= \mathrm{tr}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{\Sigma}^{-1}) \\ &= \mathrm{tr}\big[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'(\boldsymbol{I}-\boldsymbol{C}\otimes\boldsymbol{\Lambda})\mathrm{diag}\{\boldsymbol{V}\}\big] \\ &= \mathrm{tr}\big[\mathrm{blk}\{\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_{i'}'\}\mathrm{blk}\{\boldsymbol{I}\mathcal{I}(i=i')-\boldsymbol{\Lambda}\mathcal{I}(i\sim i')\}\mathrm{diag}\{\boldsymbol{V}\}\big] \\ &= \mathrm{tr}\Big(\mathrm{blk}\big[\sum_{l=1}^{I}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_l'\{\boldsymbol{I}\mathcal{I}(l=i')-\boldsymbol{\Lambda}\mathcal{I}(l\sim i')\}\big]\mathrm{diag}\{\boldsymbol{V}\}\Big) \\ &= \sum_{i=1}^{I}\mathrm{tr}\Big(\big[\sum_{l=1}^{I}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_l'\{\boldsymbol{I}\mathcal{I}(l=i)-\boldsymbol{\Lambda}\mathcal{I}(l\sim i)\}\big]\boldsymbol{V}\Big) \\ &= \mathrm{tr}\Big(\sum_{i=1}^{I}\big[\sum_{l=1}^{I}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_l'\{\boldsymbol{I}\mathcal{I}(l=i)-\boldsymbol{\Lambda}\mathcal{I}(l\sim i)\}\big]\boldsymbol{V}\Big) \\ &= \mathrm{tr}(\boldsymbol{\Phi}\boldsymbol{V}), \end{aligned}$$

where $\boldsymbol{\varepsilon}_i = (\boldsymbol{x}_i'\boldsymbol{\beta})'$ and $\boldsymbol{\Phi} = \sum_{i=1}^{I}\big[\sum_{l=1}^{I}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_l'\{\boldsymbol{I}\mathcal{I}(l=i)-\boldsymbol{\Lambda}\mathcal{I}(l\sim i)\}\big]$, the full conditional distribution of $\boldsymbol{V}$ is proportional to

$$|\boldsymbol{V}|^{\frac{1}{2}(I+\rho-J-1)}\exp\Big[-\mathrm{tr}\big\{(\boldsymbol{\Phi}+\boldsymbol{\Psi}^{-1})\boldsymbol{V}\big\}/2\Big].$$

Let $\bar{\boldsymbol{\Psi}} = \big\{(\boldsymbol{\Phi}+\boldsymbol{\Phi}')/2+\boldsymbol{\Psi}^{-1}\big\}^{-1}$, the full conditional distribution of $\boldsymbol{V}$ is $\mathrm{W}(I+\rho,\bar{\boldsymbol{\Psi}})$.

2.7.3 Posterior Sample of $\mathbf{\Lambda}$

To ensure that the variance matrix $\mathbf{\Sigma}$ is symmetric and positive definite, the matrix $\mathbf{\Lambda}$ needs to satisfy certain regularity conditions. Here we have assumed that $\mathbf{\Lambda}$ is symmetric and $\mathbf{\Lambda\Gamma} = \mathbf{\Gamma\Lambda}$, which is equivalent to $\mathbf{\Lambda}\boldsymbol{V} = \boldsymbol{V}\mathbf{\Lambda}$. Let

$$\boldsymbol{H} = \boldsymbol{I} - \boldsymbol{C} \otimes \mathbf{\Lambda}, \tag{2.26}$$

the eigenvalues of $\boldsymbol{H}$ can be written as $\{1 - c_i\lambda_j\}$, where c_i, $i = 1, \dots, I$, are the eigenvalues of $\boldsymbol{C}$ and λ_j $(j = 1, \dots, J)$ under log-ratio link $(j = 0, \dots, J$ under alternative link) are the eigenvalues of $\mathbf{\Lambda}$. Therefore, we constrain $\lambda_j \in (1/c_{(1)}, 1/c_{(I)})$, where $c_{(1)}$ is the smallest negative eigenvalue of $\boldsymbol{C}$ and $c_{(I)}$ is the largest positive eigenvalue of $\boldsymbol{C}$. This ensures that all the eigenvalues of $\boldsymbol{H}$ are positive and thus, $\boldsymbol{H}$ and $\mathbf{\Sigma} = \text{diag}\{\mathbf{\Gamma}\}\boldsymbol{H}^{-1}$ are positive definite.

Since $\boldsymbol{Z} \sim \text{N}_{IJ}(\boldsymbol{\mu}, \mathbf{\Sigma})$ under log ratio link ($\boldsymbol{Z} \sim \text{N}_{I(J+1)}(\boldsymbol{\mu}, \mathbf{\Sigma})$ under alternative link) and

$p(\mathbf{\Lambda}) \propto \exp\{-\text{vec}(\mathbf{\Lambda})'\text{vec}(\mathbf{\Lambda})/\xi^2\}$, the full conditional probability density of $\mathbf{\Lambda}$ is

$$\begin{aligned} p(\mathbf{\Lambda}|\boldsymbol{Z}, \{\boldsymbol{\beta}_j\}, \boldsymbol{V}, \boldsymbol{y}) &\propto p(\boldsymbol{Z}|\{\boldsymbol{\beta}_j\}, \boldsymbol{V}, \mathbf{\Lambda})p(\mathbf{\Lambda}) \\ &\propto |\boldsymbol{H}|^{1/2} \exp\Big\{ - (\boldsymbol{Z} - \boldsymbol{\mu})'\mathbf{\Sigma}^{-1}(\boldsymbol{Z} - \boldsymbol{\mu})/2\Big\} \exp\Big\{ - \text{vec}(\mathbf{\Lambda})'\text{vec}(\mathbf{\Lambda})/\xi^2\Big\}. \end{aligned}$$

We use an MH algorithm with a uniform proposal to generate each component of $\mathbf{\Lambda}$ in arbitrary, but pre-determined order. Suppose $\mathbf{\Lambda} = [\lambda_{kl}]$ and the proposal is $\lambda_{kl}^{(*)} \sim \text{U}(\lambda_{kl}^{(n-1)} - \epsilon, \lambda_{kl}^{(n-1)} + \epsilon)$, which is symmetric and satisfies $\mathbf{\Lambda}^{(*)}\mathbf{\Gamma} = \mathbf{\Gamma}\mathbf{\Lambda}^{(*)}$. Then the Hastings ratio of $\mathbf{\Lambda}^{(*)}$ is

$$\begin{aligned} r = \frac{|\boldsymbol{H}^{(*)}|^{1/2}}{|\boldsymbol{H}^{(n-1)}|^{1/2}} &\exp\Big[- (\boldsymbol{Z} - \boldsymbol{\mu})'\big\{\mathbf{\Sigma}^{(*)^{-1}} - \mathbf{\Sigma}^{(n-1)^{-1}}\big\}(\boldsymbol{Z} - \boldsymbol{\mu})/2\Big] \\ &\exp\Big[- \big\{\text{vec}(\mathbf{\Lambda}^{(*)})'\text{vec}(\mathbf{\Lambda}^{(*)}) - \text{vec}(\mathbf{\Lambda}^{(n-1)})'\text{vec}(\mathbf{\Lambda}^{(n-1)})\big\}/\xi^2\Big], \end{aligned}$$

where $\boldsymbol{H}^{(*)}$ and $\boldsymbol{H}^{(n-1)}$ are given in (2.26) by replacing $\boldsymbol{\Lambda}$ with $\boldsymbol{\Lambda}^{(*)}$ and $\boldsymbol{\Lambda}^{(n-1)}$, ${\boldsymbol{\Sigma}^{(*)}}^{-1} = \mathrm{diag}\{\boldsymbol{V}\}\boldsymbol{H}^{(*)}$, and ${\boldsymbol{\Sigma}^{(n-1)}}^{-1} = \mathrm{diag}\{\boldsymbol{V}\}\boldsymbol{H}^{(n-1)}$.

2.7.4 Posterior Sample of $\boldsymbol{Z}$

For the ith site, the full conditional probability density of $\boldsymbol{Z}$ is proportional to

$$\begin{aligned}
& \prod_{j=0}^{J} \pi_{ji}^{y_{ji}} \exp\Big\{ - (\boldsymbol{Z}_i - \boldsymbol{\mu}_i)'\boldsymbol{V}(\boldsymbol{Z}_i - \boldsymbol{\mu}_i)/2 - \sum_{i' \in \mathcal{N}_i} (\boldsymbol{Z}_i - \boldsymbol{\mu}_i)'\boldsymbol{U}(\boldsymbol{Z}_{i'} - \boldsymbol{\mu}_{i'}) \Big\} \\
\propto\ & \prod_{j=0}^{J} \pi_{ji}^{y_{ji}} \exp\Big[- (\boldsymbol{Z}_i - \boldsymbol{\mu}_i)'\boldsymbol{V}(\boldsymbol{Z}_i - \boldsymbol{\mu}_i)/2 - (\boldsymbol{Z}_i - \boldsymbol{\mu}_i)'\boldsymbol{V}\big\{ \sum_{i' \in \mathcal{N}_i} \boldsymbol{\Lambda}(\boldsymbol{Z}_{i'} - \boldsymbol{\mu}_{i'}) \big\} \Big] \\
\propto\ & \prod_{j=0}^{J} \pi_{ji}^{y_{ji}} \exp\left\{ -(\boldsymbol{Z}_i - \boldsymbol{\mu}_i^*)'\boldsymbol{V}(\boldsymbol{Z}_i - \boldsymbol{\mu}_i^*)/2 \right\},
\end{aligned}$$

where $\boldsymbol{U}$ is defined in Section 2.7.1 and $\boldsymbol{\mu}_i^* \equiv \boldsymbol{\mu}_i + \sum_{i' \in \mathcal{N}_i} \boldsymbol{\Lambda}(\boldsymbol{Z}_{i'} - \boldsymbol{\mu}_{i'})$. We use an MH algorithm with a Gaussian proposal, $\mathrm{N}_J(\boldsymbol{Z}_i^{(n-1)}, {\tau_z}^2 \boldsymbol{I}_J)$ ($\mathrm{N}_{J+1}(\boldsymbol{Z}_i^{(n-1)}, {\tau_z}^2 \boldsymbol{I}_{J+1})$ under alternative link), to generate $\boldsymbol{Z}_i^{(*)}$. From $\boldsymbol{Z}_i^{(*)}$, with $Z_{0i}^{(*)} \equiv 0$ ($Z_{0i}^{(*)}$ is sampled from the proposal under alternative link), we have $\pi_{ji}^{(*)} = \exp(Z_{ji}^{(*)})\big\{ \sum_{j=0}^{J} \exp(Z_{ji}^{(*)}) \big\}^{-1}$, for $j = 0, \ldots, J$. Then the Hastings ratio of $\boldsymbol{Z}_i^{(*)}$ is

$$r = \frac{\prod_{j=0}^{J} {\pi_{ji}^{(*)}}^{y_{ji}}}{\prod_{j=0}^{J} {\pi_{ji}^{(n-1)}}^{y_{ji}}} \exp\left\{ -(\boldsymbol{Z}_i^{(*)} - \boldsymbol{\mu}_i^*)'\boldsymbol{V}(\boldsymbol{Z}_i^{(*)} - \boldsymbol{\mu}_i^*)/2 + (\boldsymbol{Z}_i^{(n-1)} - \boldsymbol{\mu}_i^*)'\boldsymbol{V}(\boldsymbol{Z}_i^{(n-1)} - \boldsymbol{\mu}_i^*)/2 \right\}. \tag{2.27}$$

2.8 Statistical Inference for Reduced Spatial Multinomial Models

2.8.1 Posterior Sample of $\boldsymbol{\Omega}$

The full conditional probability density of $\boldsymbol{\Omega}$ is

$$p(\boldsymbol{\Omega}|\boldsymbol{Z}^*, \{\boldsymbol{\beta}_j\}, \boldsymbol{V}^*, \boldsymbol{\Lambda}^*, \boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{Z}^*, \boldsymbol{\Omega}, \{\boldsymbol{\beta}_j\})p(\boldsymbol{\Omega}).$$

Assuming $\boldsymbol{\Omega} = [\omega_{jj^*}]$, with $j = 1, \ldots, J$ under log ratio link ($j = 0, \ldots, J$ under alternative link), $j^* = 1, \ldots, J^*$, we generate $\omega_{jj^*}^{(*)}$ via a Gaussian proposal $\mathrm{N}(\omega_{jj^*}^{(n-1)}, \tau_\omega^2)$. From $\omega_{jj^*}^{(*)}$, we obtain a new $\boldsymbol{\Omega}^{(*)}$, $\boldsymbol{Z}_i^{(*)} = \boldsymbol{\mu}_i^{(n-1)} + \boldsymbol{\Omega}^{(*)}\boldsymbol{Z}_i^{*(n-1)}$, and $\boldsymbol{\pi}_i^{(*)}$ for $i = 1, \ldots, I$, as in Section 2.7.4. Then the Hastings ratio of $\omega_{jj^*}^{(*)}$ is

$$r = \frac{\prod_{i=1}^I \left[\prod_{j=0}^J \pi_{ji}^{(*)^{y_{ji}}}\right]}{\prod_{i=1}^I \left[\prod_{j=0}^J \pi_{ji}^{(n-1)^{y_{ji}}}\right]} \frac{p(\omega_{jj^*}^{(*)})}{p(\omega_{jj^*}^{(n-1)})}.$$

2.8.2 Posterior Sample of $\{\boldsymbol{\beta}_j\}$

The full conditional probability density of $\boldsymbol{\beta}_j$ is

$$p(\boldsymbol{\beta}_j|\boldsymbol{Z}^*, \boldsymbol{\Omega}, \boldsymbol{V}^*, \boldsymbol{\Lambda}^*, \boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{Z}^*, \boldsymbol{\Omega}, \{\boldsymbol{\beta}_j\})p(\boldsymbol{\beta}_j).$$

Similar to Section 2.7.1, we generate the candidate $\boldsymbol{\beta}_j^{(*)}$ via a Gaussian proposal $\mathrm{N}(\boldsymbol{\beta}_j^{(n-1)}, \tau_\beta^2\boldsymbol{I})$. The Hastings ratio is

$$r = \frac{\prod_{i=1}^I \left[\prod_{j=0}^J \pi_{ji}^{(*)^{y_{ji}}}\right]}{\prod_{i=1}^I \left[\prod_{j=0}^J \pi_{ji}^{(n-1)^{y_{ji}}}\right]} \frac{p(\boldsymbol{\beta}_j^{(*)})}{p(\boldsymbol{\beta}_j^{(n-1)})},$$

where $\boldsymbol{\pi}_i^{(*)}$ is obtained via a procedure similar to that in Section 2.7.4.

2.8.3 Posterior Sample of $\boldsymbol{V}^*$ and $\boldsymbol{\Lambda}^*$

The full conditional probability densities of $\boldsymbol{V}^*$ and $\boldsymbol{\Lambda}^*$ are

$$\begin{aligned} p(\boldsymbol{V}^*|\boldsymbol{Z}^*, \boldsymbol{\Omega}, \{\boldsymbol{\beta}_j\}, \boldsymbol{\Lambda}^*, \boldsymbol{y}) &\propto p(\boldsymbol{Z}^*|\boldsymbol{V}^*, \boldsymbol{\Lambda}^*)p(\boldsymbol{V}^*), \\ p(\boldsymbol{\Lambda}^*|\boldsymbol{Z}^*, \boldsymbol{\Omega}, \{\boldsymbol{\beta}_j\}, \boldsymbol{V}^*, \boldsymbol{y}) &\propto p(\boldsymbol{Z}^*|\boldsymbol{V}^*, \boldsymbol{\Lambda}^*)p(\boldsymbol{\Lambda}^*). \end{aligned}$$

Compared with Section 2.7, the parameterizations of $\boldsymbol{V}^*$ and $\boldsymbol{\Lambda}^*$ are the same. Thus, we use a similar procedure for obtaining the full conditional distribution of $\boldsymbol{V}^*$; as well, the Hastings ratio of $\boldsymbol{\Lambda}^*$ by setting $\boldsymbol{\mu}$ to $\mathbf{0}$.

When $J^* = 1$, we let the prior of v^* be $\mathrm{G}(\tilde{\alpha}_1, \tilde{\gamma}_1)$. The full conditional distribution of v^* becomes

$$\begin{aligned} p(v^*|\boldsymbol{Z}^*, \boldsymbol{\Omega}, \{\boldsymbol{\beta}_j\}, \lambda^*, \boldsymbol{y}) &\propto p(\boldsymbol{Z}^*|v^*, \lambda^*)p(v^*) \\ &\propto |\boldsymbol{\Sigma}|^{-1/2} \exp(-\boldsymbol{Z}'\boldsymbol{\Sigma}^{-1}\boldsymbol{Z}/2)(v^*)^{\tilde{\alpha}_1 - 1} \exp(-\tilde{\gamma}_1 v^*) \\ &\propto (v^*)^{\tilde{\alpha}_1 + I/2 - 1} \exp\big\{ -(\boldsymbol{Z}'\boldsymbol{H}\boldsymbol{Z}/2 + \tilde{\gamma}_1)v^* \big\}. \end{aligned}$$

Thus, we sample $v^{*(n)}$ from $\mathrm{G}(\alpha_1^{(n)}, \gamma_1^{(n)})$, where

$$\alpha_1^{(n)} = \tilde{\alpha}_1 + I/2, \quad \gamma_1^{(n)} = \tilde{\gamma}_1 + \boldsymbol{Z}^{*(n-1)\prime}\boldsymbol{H}^{(n-1)}\boldsymbol{Z}^{*(n-1)}/2.$$

2.8.4 Posterior Sample of $\boldsymbol{Z}^*$

The full conditional probability density of $\boldsymbol{Z}^*$ is

$$p(\boldsymbol{Z}^*|\boldsymbol{\Omega}, \{\boldsymbol{\beta}_j\}, \boldsymbol{V}^*, \boldsymbol{\Lambda}^*, \boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{Z}^*, \boldsymbol{\Omega}, \{\boldsymbol{\beta}_j\})p(\boldsymbol{Z}^*|\boldsymbol{V}^*, \boldsymbol{\Lambda}^*).$$

Similar to Section 2.7.4, we use an MH algorithm with a Gaussian proposal, $\mathrm{N}(\boldsymbol{Z}_i^{*(n-1)}, \tau_{Z^*}^2 \boldsymbol{I}_{\boldsymbol{J}^*})$, to generate a candidate $\boldsymbol{Z}_i^{*(*)}$. An additional operation (2.6) is needed to obtain $\boldsymbol{Z}_i^{(*)}$ first and then $\boldsymbol{\pi}_i^{(*)}$. For computing the Hastings ratio, we set all the $\boldsymbol{\mu}_i$'s to $\mathbf{0}$ in (2.27).

Chapter 3

Case Study

3.1 Northern Wisconsin Land Cover Data

A primary research objective is to assess the relationship between land ownership history characteristics and forest landscape structure in northern Wisconsin and inform future forest landscape management. Data are derived from 1915 plat maps, which indicate ownership conditions, as well as the Wisconsin Land Economic Inventory, which indicates land cover in 1930, at an important turning point in the region's history (for details, see Koch (2006); Steen-Adams et al. (2011)). GIS software packages, ArcView 3.3 and ArcMap 9.2 (ESRI, 2002, 2006), are employed to process, map, and display the data. To control for the variable influence of land form and soil type, analysis is restricted to parcels that lie within a single Land Type Association, here, the Ashland Lake-Modified Till Plain (LTA 212-Ya03), which is hierarchically nested within the Province 212, Laurentian Mixed Forest, as defined by the USDA Forest Service National Hierarchical Framework of Ecological Units (NHFEU) (Cleland et al., 1997). Data are on a square lattice and employ spatial characteristics that conform to the rectilinear township and range grid of the U.S. Public Land Survey System (White, 1983; Hubbard, 2009). The unit of analysis of this study is the quarter-section (1/36 township, 160 acres $\approx$ 65 ha), the minimum acceptable spatial scale for the use of U.S. Public Land Survey records as ecological data (Delcourt and Delcourt, 1996; Schulte and Mladenoff, 2001).

The response variable is dominant land cover in 1930 of each quarter-section. Dominant land cover is defined as the land cover that accounts for the largest proportion of all component classes in the quarter-section unit of analysis. We selected 1930 land cover data because of its significance to the region's environmental history and landscape ecology: the data represent landscape conditions after original forest clearance and at the peak of agricultural land use (White and Mladenoff, 1994). There are three land cover types in the response variable, which correspond with the historical period: aspen-paper birch forest (APB) which can develop after forest clearance, agriculture-grassland (AG), and all others in one category (OTH). See Figure 1. Thus, $J = 2$. In addition, $n_i \equiv 1$, since each quarter-section can take on only one cover type. That is, $y_{ji} = 0$ or 1, such that $y_{ji} = 1$ if the ith site is in the jth category, where $j = 0, \ldots, J$ and $i = 1, \ldots, I = 1429$. The probability density (2.1) simplifies to $p(\boldsymbol{y}_i|\boldsymbol{\pi}_i) = \prod_{j=0}^{J} \pi_{ji}^{y_{ji}}$.

The covariates are various land ownership characteristics, as indicated by the 1915 plat maps and are designed to assess important aspects of ownership. Four covariates are of interest: `reserv`, a binary variable indicating whether a quarter section is on an Indian reservation or not, designed to detect the influence of contrasting social-economic histories; `polypr`, a measure of fragmentation, defined as the average size of parcel units that lie within the quarter-section (measured as a proportion); `totown`, total property area (measured in acres) associated with the owner of the largest parcel unit of the quarter-section and designed to assess whether land cover varies with ownership size; and `avparcel`, average size (measured in acres) of all parcels associated with a quarter section. See Figure 1. The covariates `polypr` and `avparcel` measure different aspects of ownership: calculation of `polypr` is limited to parcel units within the quarter-section unit of analysis and is a measure of fragmentation; calculation of `avparcel` is based on entire parcel units, which may extend beyond the quarter-section boundary, and is a measure of parcel size. Moreover, the covariate `totown` is highly skewed and thus, transformed to the log scale. Finally, the continuous covariates are standardized to have mean 0 and standard deviation 1. Figure 2 plots the proportions of the three land cover types against the four covariates.

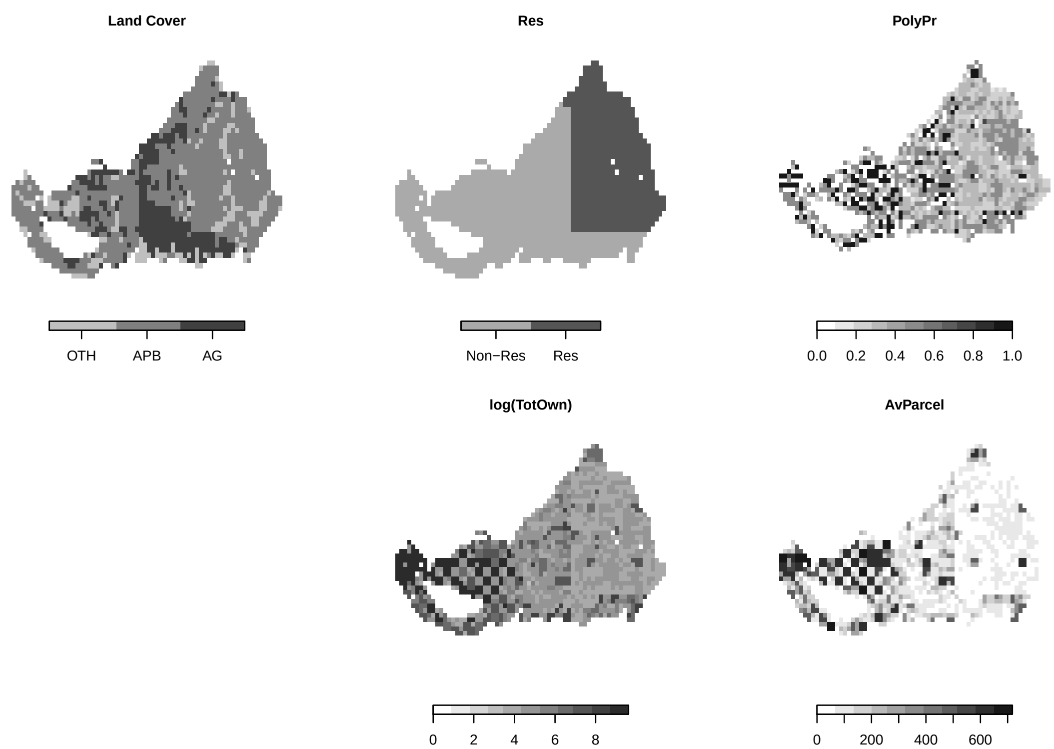

Figure 1: Map of response variable land cover type (top left) and covariates: `reserv` (top middle), `polypr` (top right), `log totown` (bottom middle), and `avparcel` (bottom right).

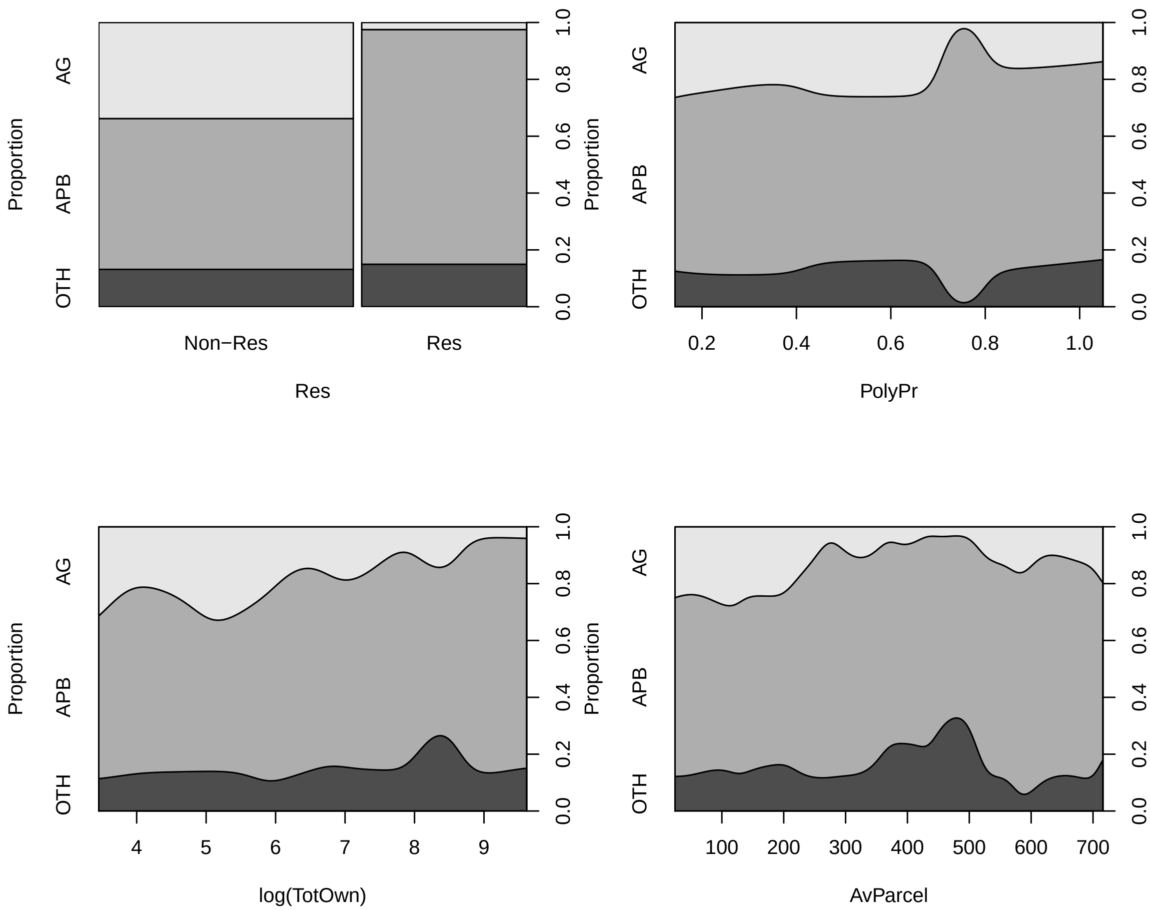

Figure 2: Sample proportions of land cover type versus covariates: `reserv` (top left), `polypr` (top right), `log totown` (bottom left), and `avparcel` (bottom right).

3.2 Model Fitting

We apply the methodology developed in Sections 2.2–2.4 to analyze this data. Both full and reduced spatial multinomial models are considered. Moreover, in the MCAR models for the latent variables, we consider a first-order neighborhood that consists of the four nearest neighbors on the grid in the north, south, west, and east.

When implementing the MCMC algorithms, we select model parameter values for the prior distributions to be diffuse. We also tune the model parameters in the proposal distributions to facilitate the convergence and mixing of the MCMC. Convergence of the MCMC algorithms is judged by examining the trace plots (Figures 3–5). In total, there are 50,000 iterations in each MCMC chain and the first 10,000 iterations are set aside as the burn-in period. The medians and the 95% credible intervals are constructed for the inference of model parameters. The hyperparameter values of the prior distributions are given in Table 7 in Section 3.4 .

We use deviance information criterion (DIC) for model comparison (Spiegelhalter et al., 2002). A smaller DIC value indicates a better model in the sense of model fitting and model parsimony.

3.3 Comparison with Alternative Models

For comparison, we consider the case $\boldsymbol{\Lambda}_{ii'} = \mathbf{0}$ in model (2.1)–(2.2), and (2.4)–(2.5), where the latent variables are independent across sites. We then let $\boldsymbol{Z}_i \sim \mathrm{N}_J(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{\boldsymbol{Z}})$, where $\boldsymbol{\mu}_i = (\boldsymbol{x}_i'\boldsymbol{\beta})'$, $i = 1, \ldots, I$ as before and $\boldsymbol{\Sigma}_{\boldsymbol{Z}} = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_J^2\}$. We will refer to the resulting model as an independent multinomial model. Under the assumption of independent prior distributions $\boldsymbol{\beta}_j \sim \mathrm{N}(0, \sigma_\beta^2 \boldsymbol{I})$ and $\sigma_j^2 \sim \mathrm{IG}(\tilde{\alpha}_2, \tilde{\gamma}_2)$ where IG denotes an inverse gamma distribution, the posterior distribution of $\boldsymbol{Z}$, $\{\boldsymbol{\beta}_j\}$, and $\{\sigma_j^2\}$ is, up to a proportionality constant,

$$p(\boldsymbol{Z}, \{\boldsymbol{\beta}_j\}, \{\sigma_j^2\}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{Z})p(\boldsymbol{Z}|\{\boldsymbol{\beta}_j\}, \{\sigma_j^2\})\left\{\prod_{j=1}^{J} p(\boldsymbol{\beta}_j)p(\sigma_j^2)\right\}. \tag{3.1}$$

We use a Gibbs sampler and sample from the full conditional distributions. We directly sample from the full conditional distributions of $\boldsymbol{\beta}_j$ and σ_j^2, since they have conjugate priors. For $\boldsymbol{Z}$, we use an MH algorithm with a Gaussian proposal to sample from the full conditional distribution. Details for the MCMC algorithms are given as follow.

3.3.1 Posterior Sample of $\boldsymbol{Z}$

We use a Gaussian proposal, $\mathrm{N}_J(\boldsymbol{Z}_i^{(n-1)}, {\tau_z}^2\boldsymbol{I}_J)$ under log ratio link ($\mathrm{N}_{J+1}(\boldsymbol{Z}_i^{(n-1)}, {\tau_z}^2\boldsymbol{I}_{J+1})$ under alternative link), to generate $\boldsymbol{Z}_i^{(*)}$. The Hastings ratio is

$$r = \frac{\prod_{j=0}^{J} {\pi_{ji}^{(*)}}^{y_{ji}} \times p({\boldsymbol{Z}_i}^{(*)}|{\boldsymbol{\mu}_i}^{(n-1)}, \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{(n-1)})}{\prod_{j=0}^{J} {\pi_{ji}^{(n-1)}}^{y_{ji}} \times p({\boldsymbol{Z}_i}^{(n-1)}|{\boldsymbol{\mu}_i}^{(n-1)}, \boldsymbol{\Sigma}_{\boldsymbol{Z}}^{(n-1)})},$$

where $p(\cdot)$ denotes the conditional Gaussian density, $\pi_{ji}^{(*)} = \exp(Z_{ji}^{(*)})\big\{\sum_{j=0}^{J}\exp(Z_{ji}^{(*)})\big\}^{-1}$, $j = 0, \ldots, J$, $\boldsymbol{\mu}_i^{(n-1)} = (\boldsymbol{x}_i'\boldsymbol{\beta}^{(n-1)})'$, and $\boldsymbol{\Sigma}_{\boldsymbol{Z}}^{(n-1)} = \mathrm{diag}\{\sigma_j^{2(n-1)}\}_{j=1}^{J}$.

3.3.2 Posterior Sample of $\{\boldsymbol{\beta}_j\}$

The full conditional probability density of $\boldsymbol{\beta}_j$ is

$$p(\boldsymbol{\beta}_j|\boldsymbol{Z}, \{\sigma_j^2\}, \boldsymbol{y}) \;\; \propto \;\; \prod_{i=1}^{I} p(\boldsymbol{Z}_i|\{\boldsymbol{\beta}_j\}, \{{\sigma_j}^2\})p(\boldsymbol{\beta}_j).$$

Due to the independence across categories, we group the terms involving the jth category and rewrite the conditional probability density as

$$
\begin{aligned}
& p(\boldsymbol{\beta}_j|\boldsymbol{Z},\{\sigma_j^2\},\boldsymbol{y}) \propto \exp\big\{-(\boldsymbol{Z}_{j\cdot}-\boldsymbol{\mu}_{j\cdot})'(\sigma_j^2\boldsymbol{I}_I)^{-1}(\boldsymbol{Z}_{j\cdot}-\boldsymbol{\mu}_{j\cdot})/2-\sigma_\beta^{-2}\boldsymbol{\beta}_j'\boldsymbol{\beta}_j/2\big\} \\
= \; & \exp\Big[-\big\{\boldsymbol{\beta}_j'(\boldsymbol{X}'\boldsymbol{X}/\sigma_j^2+\sigma_\beta^{-2}\boldsymbol{I}_K)\boldsymbol{\beta}_j-\boldsymbol{\beta}_j'(\boldsymbol{X}'\boldsymbol{Z}_{j\cdot}/\sigma_j^2)-(\boldsymbol{Z}_{j\cdot}'\boldsymbol{X}/\sigma_j^2)\boldsymbol{\beta}_j+\boldsymbol{Z}_{j\cdot}'\boldsymbol{Z}_{j\cdot}/\sigma_j^2\big\}/2\Big] \\
= \; & \exp\big\{-(\boldsymbol{\beta}_j'\boldsymbol{A}\boldsymbol{\beta}_j-\boldsymbol{\beta}_j'\boldsymbol{B}-\boldsymbol{B}'\boldsymbol{\beta}_j+\boldsymbol{C})/2\big\},
\end{aligned}
$$

where $\boldsymbol{A}=\boldsymbol{X}'\boldsymbol{X}/\sigma_j^2+\sigma_\beta^{-2}\boldsymbol{I}_K$, $\boldsymbol{B}=\boldsymbol{X}'\boldsymbol{Z}_{j\cdot}/\sigma_j^2$, $\boldsymbol{C}=\boldsymbol{Z}_{j\cdot}'\boldsymbol{Z}_{j\cdot}/\sigma_j^2$. Thus, the full conditional distribution of $\boldsymbol{\beta}_j$ is $\mathrm{N}_K(\boldsymbol{A}^{-1}\boldsymbol{B},\boldsymbol{A}^{-1})$, and we replace $\boldsymbol{Z}_{j\cdot}$ and σ_j^2 with $\boldsymbol{Z}_{j\cdot}^{(n)}$ and $\sigma_j^{2(n-1)}$ in $\boldsymbol{B}$ to obtain the full conditional distributions, $\boldsymbol{\beta}_j^{(n)}\sim\mathrm{N}_K(\boldsymbol{\mu}_{\beta_j}^{(n)},\boldsymbol{\Sigma}_j^{(n)})$, where $\boldsymbol{\Sigma}_j^{(n)}=(\boldsymbol{X}'\boldsymbol{X}/\sigma_j^{2(n-1)}+\sigma_\beta^{-2}\boldsymbol{I}_K)^{-1}$ and $\boldsymbol{\mu}_{\beta_j}^{(n)}=\boldsymbol{\Sigma}_j^{(n)}\boldsymbol{X}'\boldsymbol{Z}_{j\cdot}^{(n)}/\sigma_j^{2(n-1)}$.

3.3.3 Posterior Sample of $\{\sigma_j^2\}$

The full conditional probability density of σ_j^2 is

$$
\begin{aligned}
p(\sigma_j^2|\boldsymbol{Z},\{\boldsymbol{\beta}_j\},\boldsymbol{y}) &\propto p(\boldsymbol{Z}|\{\boldsymbol{\beta}_j\},\{\sigma_j^2\})p(\sigma_j^2) \\
&\propto \sigma_j^{-2(1+\tilde{\alpha}_2)+I}\exp\Big[-\big\{2\tilde{\gamma}_2+(\boldsymbol{Z}_{j\cdot}-\boldsymbol{\mu}_{j\cdot})'(\boldsymbol{Z}_{j\cdot}-\boldsymbol{\mu}_{j\cdot})\big\}/2\sigma_j^2\Big].
\end{aligned}
$$

By conjugacy, we use a Gibbs sampler with $\sigma_j^{2(n)}\sim\mathrm{IG}(\alpha_{2j}^{(n)},\gamma_{2j}^{(n)})$, where $\alpha_{2j}^{(n)}=\tilde{\alpha}_2+I/2$ and $\gamma_{2j}^{(n)}=\tilde{\gamma}_2+(\boldsymbol{Z}_{j\cdot}-\boldsymbol{\mu}_{j\cdot})'(\boldsymbol{Z}_{j\cdot}-\boldsymbol{\mu}_{j\cdot})/2$.

3.4 Data Analysis Results

The DIC values for the three models fitted are 1997, 1326, and 1497, for the independent multinomial model (Figure 3 and Table 1), reduced spatial multinomial model (Figure 4 and Table 2), and full spatial multinomial model (Figure 5 and Table 3), respectively.

That is, the best model is the reduced spatial multinomial model and the second best model is the full spatial multinomial model, whereas the independent model is the worst, based on these DIC values. Tables 1–3 give the medians of the posterior samples of the parameters, along with the 95% credible intervals.

Table 1: Median (the 50th percentile) of the posterior samples for the model parameters along with a 95% credible interval (between the 2.5th and 97.5th percentiles) from an **independent multinomial model**. The land cover types are aspen-paper birch forest, agriculture grassland, and the baseline all others. The covariates are intercept, `reserv, polypr, log totown`, and `avparcel`.

percentile	intercept	`reserv`	`polypr`	`totown`	`avparcel`
aspen-paper birch					
2.5%	1.075	0.464	-0.512	-0.082	-0.001
50%	1.311	0.887	-0.219	0.170	0.336
97.5%	1.581	1.337	0.063	0.430	0.670
agriculture grassland					
2.5%	0.708	-4.410	-0.816	-1.325	-0.302
50%	0.966	-3.752	-0.462	-0.980	0.101
97.5%	1.243	-3.169	-0.125	-0.640	0.550
	σ_1^2	σ_2^2			
2.5%	0.680	0.646			
50%	1.028	0.935			
97.5%	1.627	1.446			

Based on the results of the best model in Table 2, with prediction and probability of each category shown in Figure 6, there appears to be a relationship between land cover and `reserv`, `polypr`, and `totown`. In particular, the regression coefficient for `reserv` is in the positive range for aspen-paper birch and is in the negative range for agriculture grassland with all others as the baseline. In a quarter section that is on an Indian reservation, the log odds of aspen-paper birch increases by 1.363, but the log odds of agriculture grassland decreases by 3.616, relative to all others. The regression coefficients for `polypr` is in the negative ranges for agriculture grassland with all others

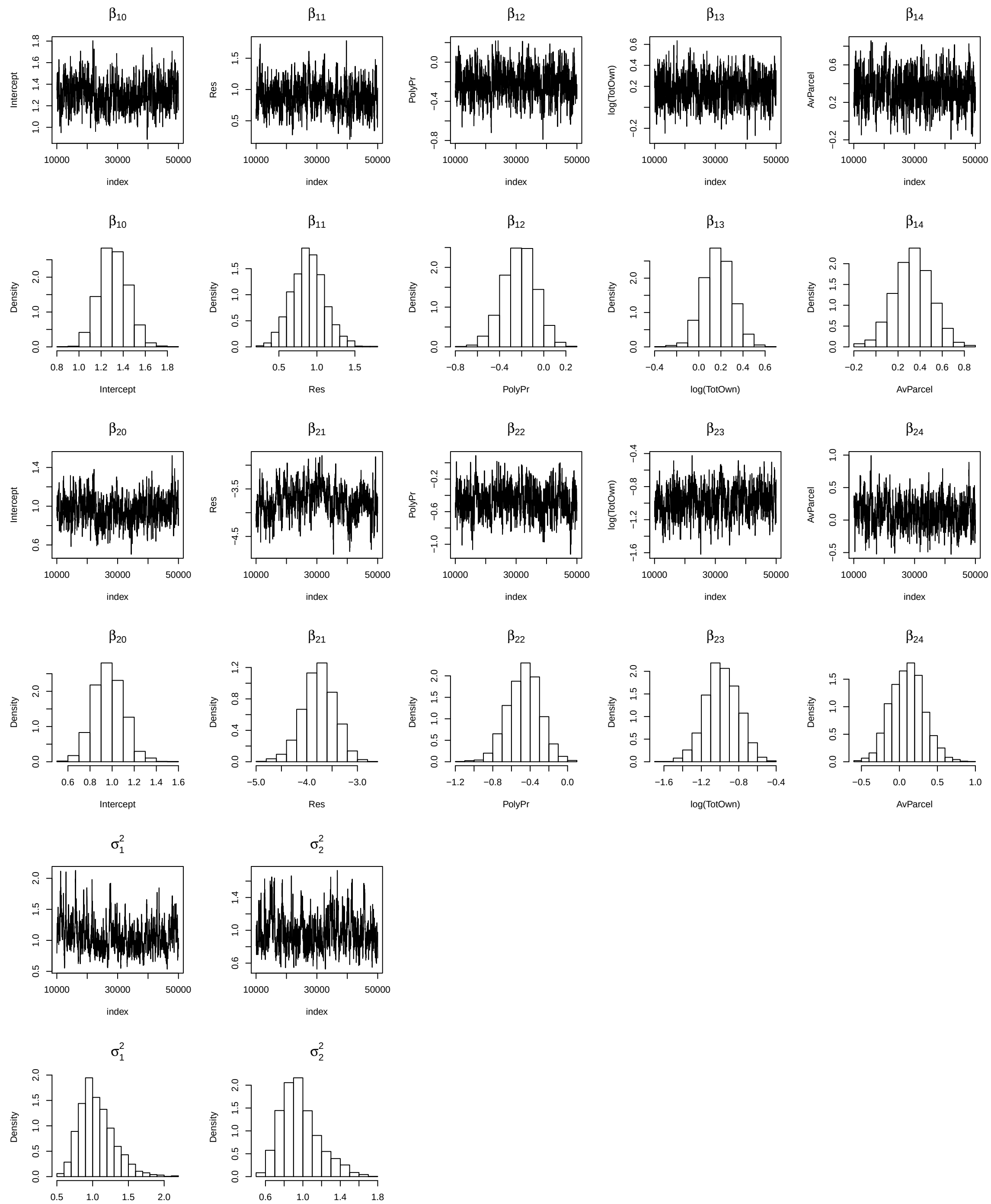

Figure 3: Trace plot and histogram of MCMC sampling of the coefficients (intercept, `reserv`, `polypr`, `log totown`, and `avparcel`) and other parameters ($\{\sigma_j^2\}$) from an **independent multinomial model**.

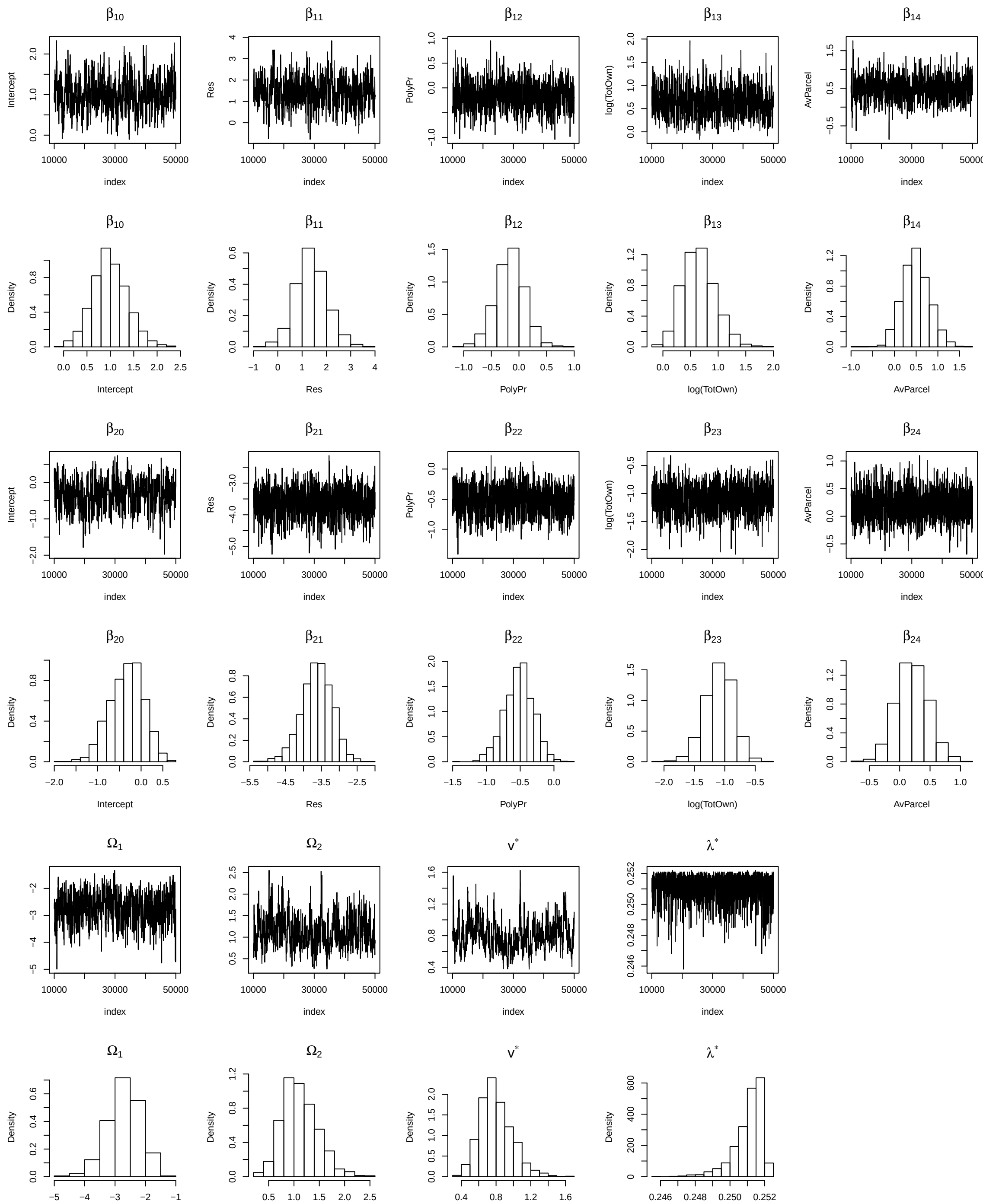

Figure 4: Trace plot and histogram of MCMC sampling of the coefficients (intercept, `reserv`, `polypr`, `log totown`, and `avparcel`) and other parameters ($\boldsymbol{\Omega}$, v^* and λ^*) from a **reduced spatial multinomial model**.

Table 2: Median (the 50th percentile) of the posterior samples for the model parameters along with a 95% credible interval (between the 2.5th and 97.5th percentiles) from a **reduced spatial multinomial model**. The land cover types are aspen-paper birch forest, agriculture grassland, and the baseline all others. The covariates are intercept, `reserv, polypr, log totown`, and `avparcel`.

percentile	intercept	`reserv`	`polypr`	`totown`	`avparcel`
aspen-paper birch					
2.5%	0.277	0.142	-0.695	0.113	-0.090
50%	0.974	1.363	-0.161	0.634	0.494
97.5%	1.750	2.697	0.332	1.281	1.148
agriculture grassland					
2.5%	-1.112	-4.530	-0.961	-1.579	-0.299
50%	-0.311	-3.616	-0.518	-1.083	0.206
97.5%	0.384	-2.818	-0.117	-0.661	0.728
	Ω_1	Ω_2	υ^*	λ^*	
2.5%	-3.817	0.528	0.490	0.249	
50%	-2.672	1.082	0.789	0.251	
97.5%	-1.754	-1.880	1.204	0.252	

Table 2: (cont') Median (the 50th percentile) of the posterior samples for the model parameters along with a 95% credible interval (between the 2.5th and 97.5th percentiles) from a **reduced spatial multinomial model**. The land cover types are aspen-paper birch forest, all others, and the baseline agriculture grassland. The covariates are intercept, `reserv, polypr, log totown`, and `avparcel`.

percentile	intercept	`reserv`	`polypr`	`totown`	`avparcel`
aspen-paper birch					
2.5%	0.238	3.311	-0.333	1.019	-0.542
50%	1.306	5.003	0.356	1.720	0.281
97.5%	2.491	6.749	1.046	2.558	1.150
all others					
2.5%	-0.384	2.818	0.117	0.661	-0.728
50%	0.311	3.616	0.518	1.083	-0.206
97.5%	1.112	4.530	0.961	1.579	0.299

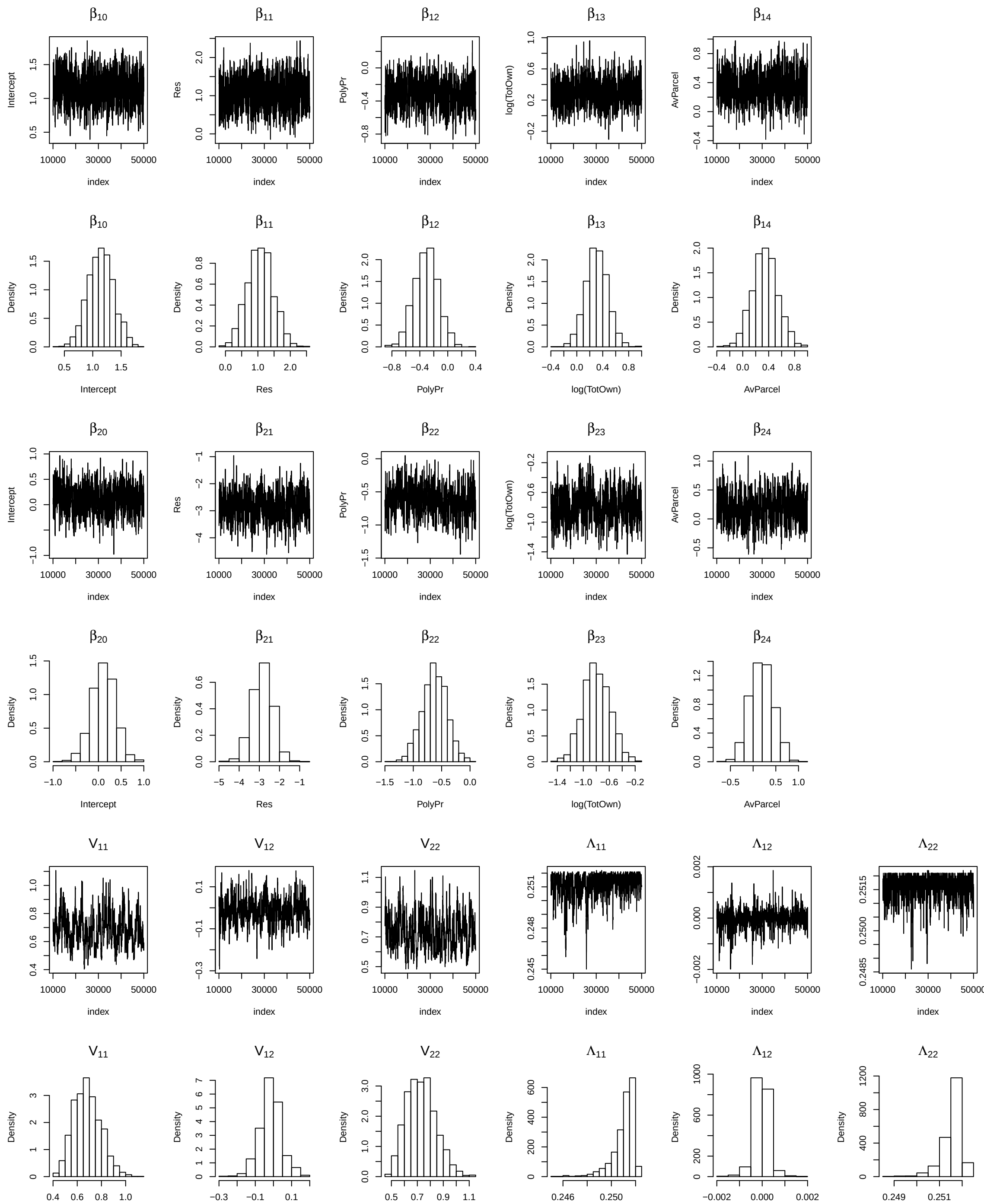

Figure 5: Trace plot and histogram of MCMC sampling of the coefficients (intercept, `reserv`, `polypr`, `log totown`, and `avparcel`) and other parameters ($\boldsymbol{V}$ and $\boldsymbol{\Lambda}$) from a **full spatial multinomial model**.

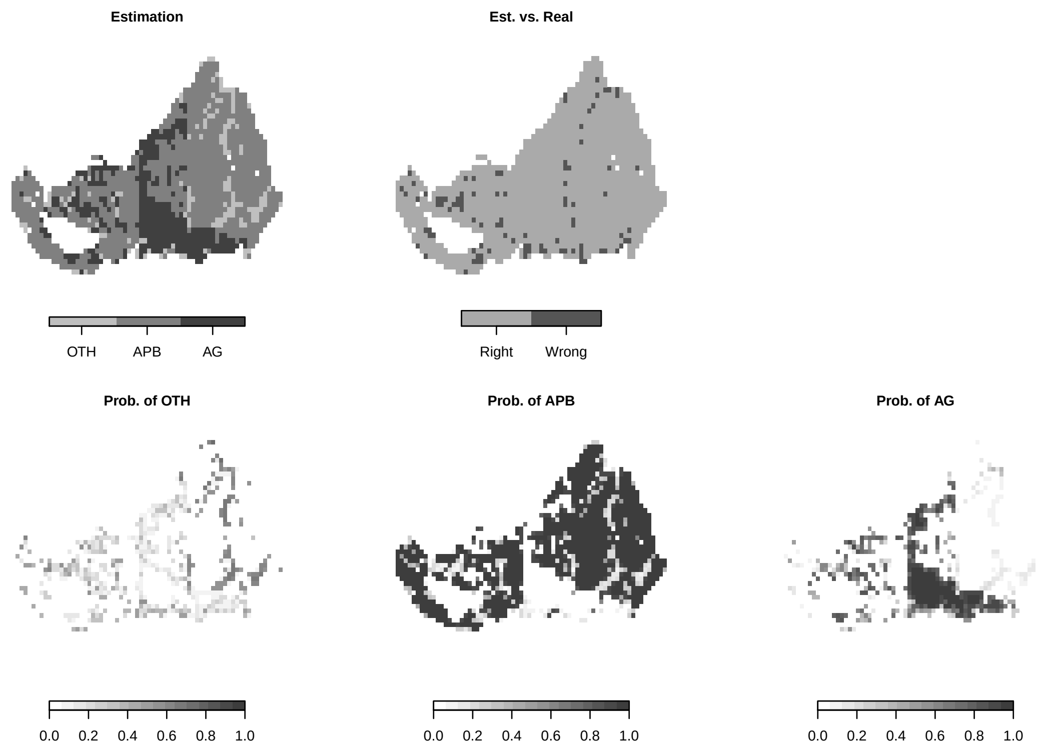

Figure 6: Prediction and probability of each category from a **reduced spatial multinomial model**. Plots in the first row are the predicted/estimated land cover types and the comparison with the observed land cover types. The plots in the second row are the estimated probabilities of sites to be of type aspen-paper birch forest (APB), agriculture-grassland (AG) and all others (OTH).

Table 3: Median (the 50th percentile) of the posterior samples for the model parameters along with a 95% credible interval (between the 2.5th and 97.5th percentiles) from a **full spatial multinomial model**. The land cover types are aspen-paper birch forest, agriculture grassland, and the baseline all others. The covariates are intercept, `reserv, polypr, log totown`, and `avparcel`.

percentile	intercept	`reserv`	`polypr`	`totown`	`avparcel`	
aspen-paper birch						
2.5%	0.695	0.295	-0.655	-0.027	-0.087	
50%	1.157	1.087	-0.314	0.305	0.366	
97.5%	1.609	1.903	0.046	0.661	0.766	
agriculture grassland						
2.5%	-0.370	-3.820	-1.062	-1.260	-0.356	
50%	0.151	-2.787	-0.633	-0.803	0.230	
97.5%	0.629	-1.726	-0.221	-0.333	0.765	
	V_{11}	V_{12}	V_{22}	Λ_{11}	Λ_{12}	Λ_{22}
2.5%	0.422	-0.152	0.493	0.249	-0.001	0.251
50%	0.630	-0.022	0.725	0.251	0.000	0.252
97.5%	0.883	0.087	0.961	0.252	0.000	0.252

as the baseline. That is, as the `polypr` increases by one sample standard deviation 0.292 (or, by one unit), the log odds of agriculture grassland decreases by 0.518 (or, by $0.518/0.292 = 1.774$). Further, the regression coefficient for `totown` (on the log scale) is in the positive range for aspen-paper birch and is in the negative range for agriculture grassland with all others as the baseline. When the log of `totown` increases by one sample standard deviation 1.764 (or, by one unit), the log odds of aspen-paper birch increases by 0.634 (or, by $0.634/1.764 = 0.359$), whereas the log odds of agriculture grassland decreases by 1.083 (or, by $1.083/1.764 = 0.614$) relative to all others. Finally, the 95% credible intervals for the regression coefficients of `avparcel` cover zero for both aspen-paper birch and agriculture grassland, indicating no relationship between land cover types and `avparcel` after accounting for the other covariates. These results are in line with the plots of the data shown in Figure 2.

Based on the results of the other two models in Tables 1 and 3, similar results are observed with the exception that the 95% credible interval for the regression coefficient of `totown` covers zero for aspen-paper birch.

Figures 7–9 and Tables 4–6 show the results of MCMC sampling under the alternative link function. A potential disadvantage of (2.21) is that the dimension of the latent variable increases. Even though the increase is only from dimension J to $J+1$, we have encountered numerical instability when fitting the model to the land cover data, see the trace plot of $\tilde{\boldsymbol{\Omega}}$ in Figure 9. Further investigation will be needed to understand the cause of the computational difficulty and to develop counter measures.

Table 4: Median (the 50th percentile) of the posterior samples for the model parameters along with a 95% credible interval (between the 2.5th and 97.5th percentiles) from an **independent multinomial model** under an alternative link function. The land cover types are aspen-paper birch forest, agriculture grassland, and the baseline all others. The covariates are intercept, `reserv, polypr, log totown`, and `avparcel`.

percentile	intercept	`reserv`	`polypr`	`totown`	`avparcel`
aspen-paper birch					
2.5%	1.383	0.251	-0.638	-0.178	0.018
50%	1.691	0.745	-0.310	0.123	0.395
97.5%	2.044	1.235	0.028	0.433	0.784
agriculture grassland					
2.5%	1.012	-4.655	-0.955	-1.424	-0.282
50%	1.329	-3.974	-0.564	-1.035	0.179
97.5%	1.699	-3.336	-0.193	-0.658	0.671
	$\tilde{\sigma}_0^2$	$\tilde{\sigma}_1^2$	$\tilde{\sigma}_2^2$		
2.5%	0.686	0.674	0.656		
50%	1.075	1.032	0.956		
97.5%	1.804	1.567	1.457		

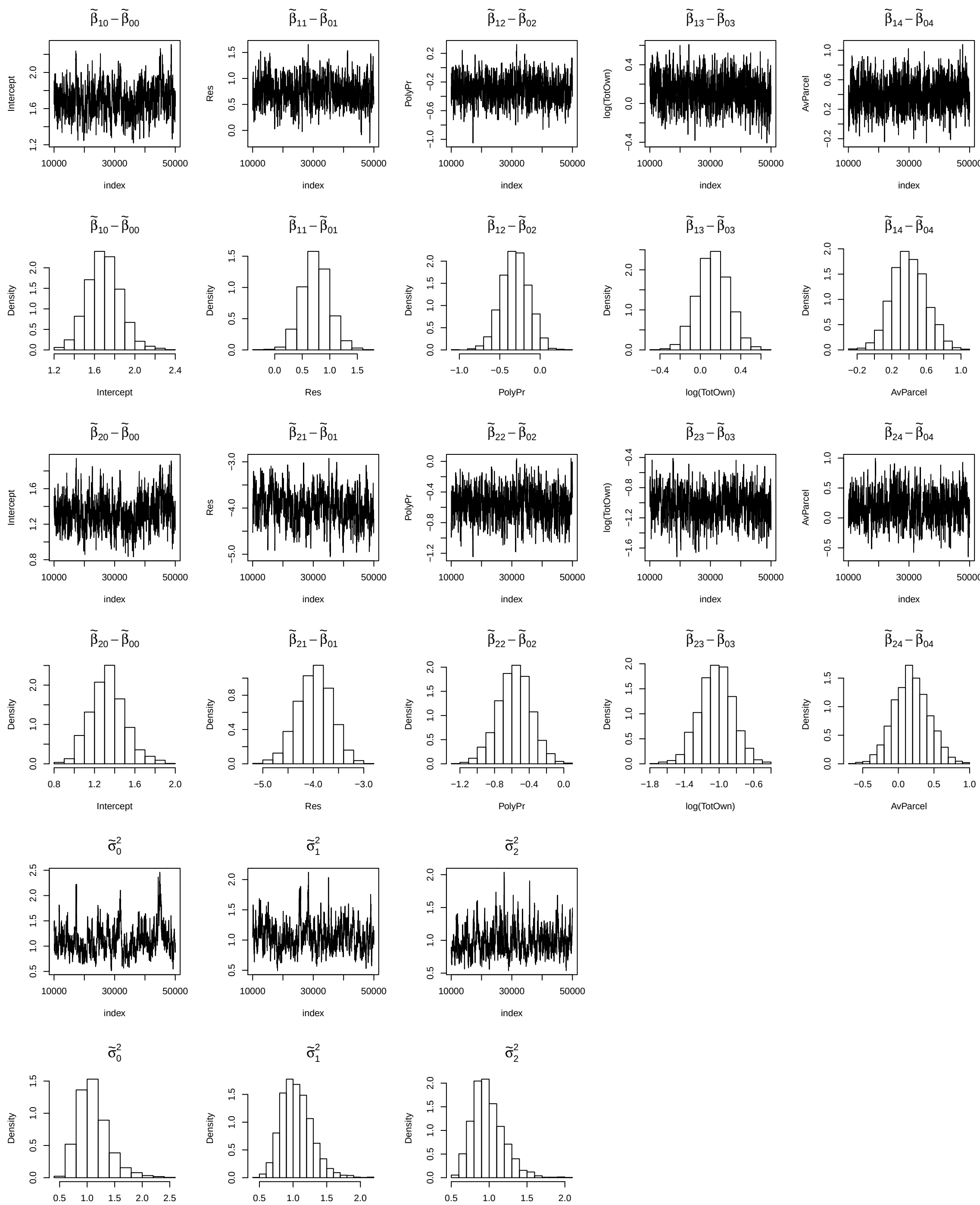

Figure 7: Trace plot and histogram of MCMC sampling of the coefficients (intercept, `reserv`, `polypr`, `log totown`, and `avparcel`) and other parameters ($\{\tilde{\sigma}_j^2\}$) from an **independent multinomial model** under an alternative link function.

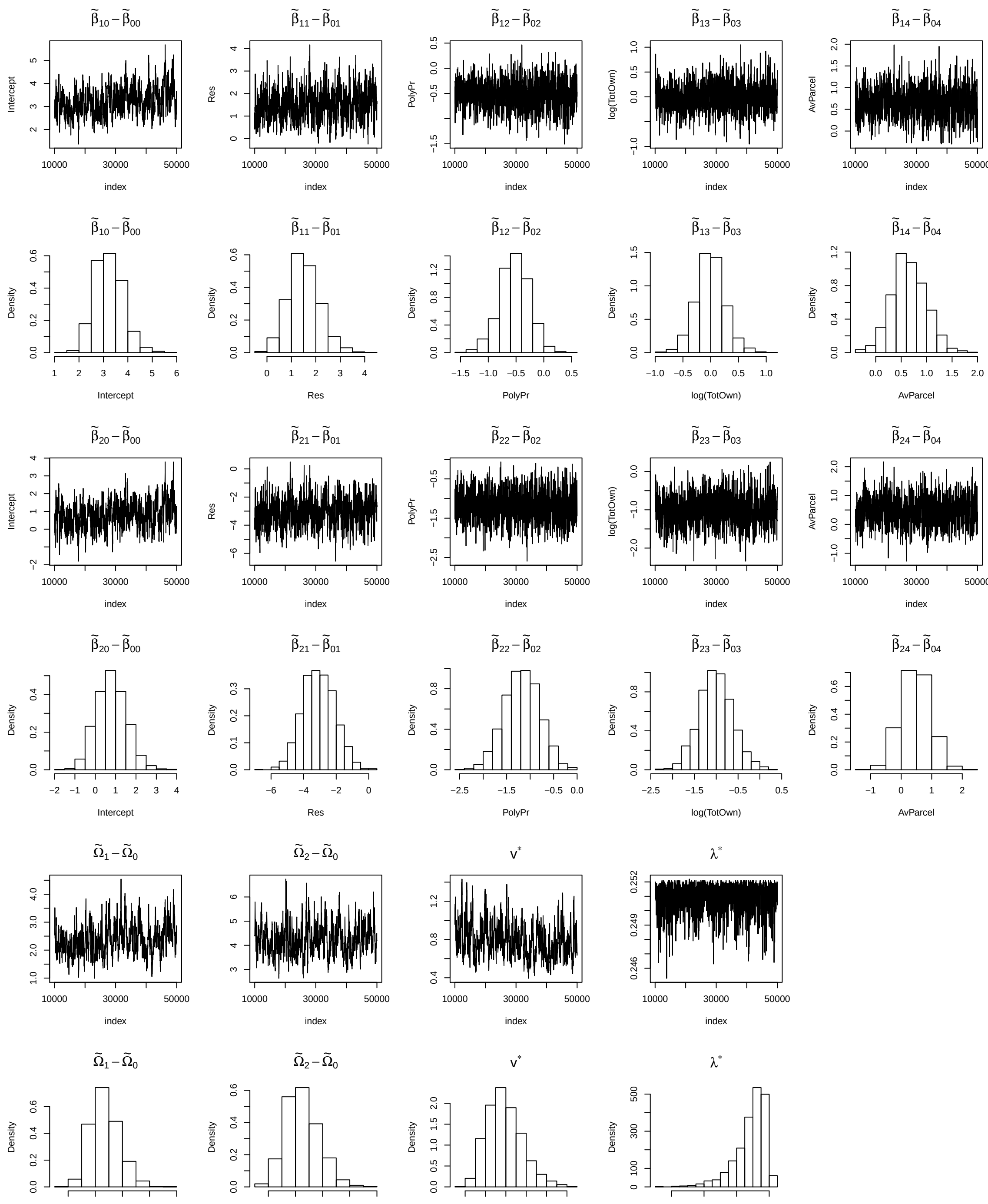

Figure 8: Trace plot and histogram of MCMC sampling of the coefficients (intercept, `reserv, polypr, log totown`, and `avparcel`) and other parameters ($\tilde{\mathbf{\Omega}}$, v^* and λ^*) from a **reduced spatial multinomial model with** $J^* = 1$ under an alternative link function.

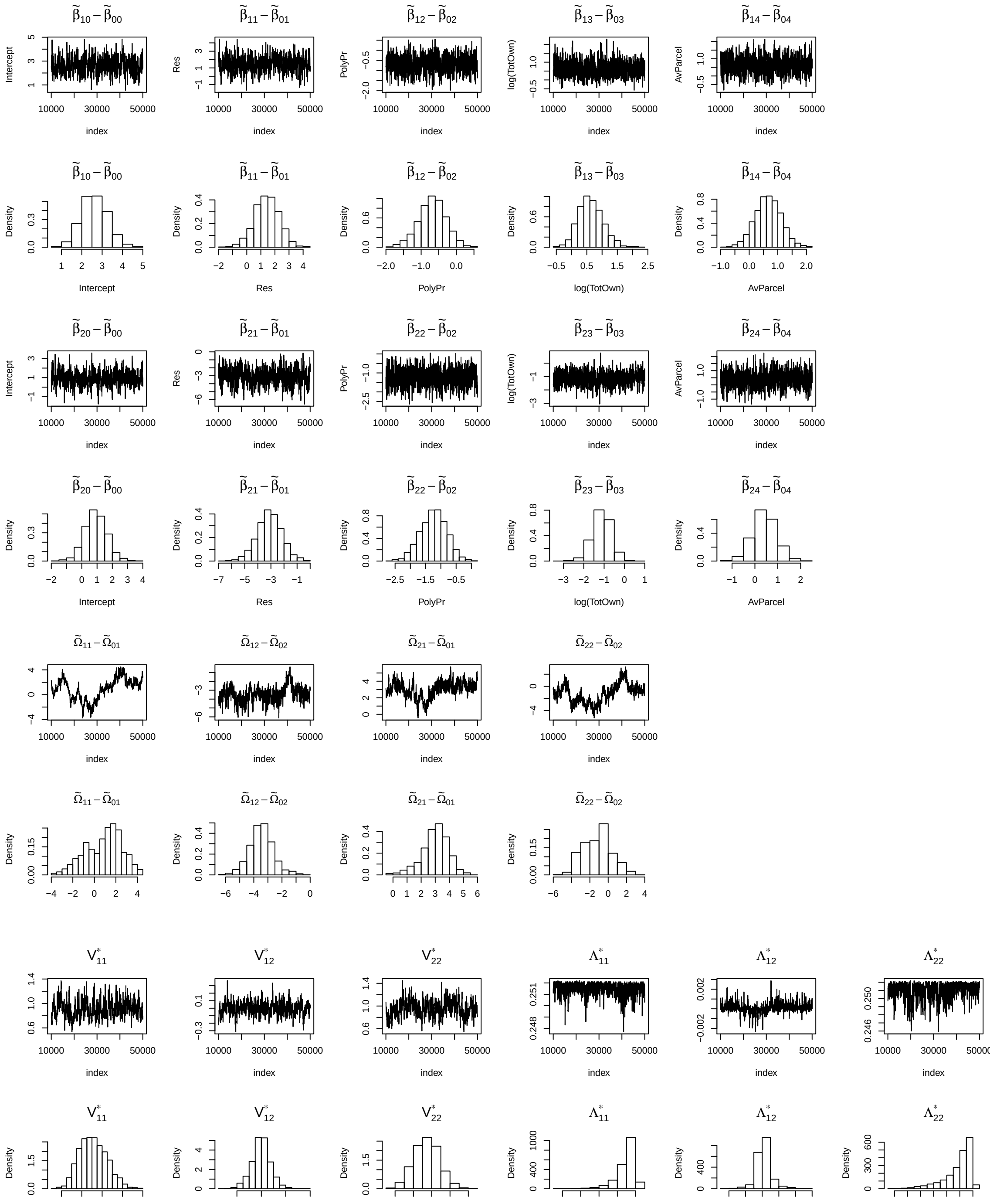

Figure 9: Trace plot and histogram of MCMC sampling of the coefficients (intercept, `reserv`, `polypr`, `log totown`, and `avparcel`) and other parameters ($\tilde{\boldsymbol{\Omega}}$, $\boldsymbol{V}^*$ and $\boldsymbol{\Lambda}^*$) from a **reduced spatial multinomial model with** $J^* = 2$ under an alternative link function.

Table 5: Median (the 50th percentile) of the posterior samples for the model parameters along with a 95% credible interval (between the 2.5th and 97.5th percentiles) from a **reduced spatial multinomial model with** $J^* = 1$ under an alternative link function. The land cover types are aspen-paper birch forest, agriculture grassland, and the baseline all others. The covariates are intercept, `reserv, polypr, log totown`, and `avparcel`.

percentile	intercept	`reserv`	`polypr`	`totown`	`avparcel`
aspen-paper birch					
2.5%	2.216	0.334	-1.101	-0.511	0.005
50%	3.175	1.487	-0.520	-0.015	0.641
97.5%	4.431	2.881	-0.007	0.535	1.343
agriculture grassland					
2.5%	-0.577	-4.973	-1.937	-1.730	-0.452
50%	0.764	-3.082	-1.176	-1.011	0.473
97.5%	2.273	-1.103	-0.452	-0.229	1.380
	$\tilde{\Omega}_1 - \tilde{\Omega}_0$	$\tilde{\Omega}_2 - \tilde{\Omega}_0$	υ^*	λ^*	
2.5%	1.473	3.154	0.502	0.248	
50%	2.306	4.190	0.770	0.251	
97.5%	3.472	5.587	1.170	0.252	

3.5 Ecological and Management Implications

Two covariates have a positive association with aspen-paper birch: `reserv`, a measure of ownership type, and `totown`, a measure of ownership size. This finding illuminates the factors that influence landscape structure in northern Wisconsin, a major topic in landscape ecology. Whereas the influence of biotic and abiotic factors on landscape pattern is well-documented (Turner, 2005), social factors are less well-understood. The relationship between `reserv` and aspen-paper birch is supported by related studies of the region's environmental and landscape history (Steen-Adams et al., 2007, 2011): historical and ecological factors (e.g., U.S. Indian policy and legislation, forest succession dynamics on cutover sites, which were harvested later on-reservation) promoted nearly exclusive management for aspen-paper birch on the Indian reservation, in contrast to off-reservation, during the 1930 study date. The positive relationship between `totown` and

Table 6: Median (the 50th percentile) of the posterior samples for the model parameters along with a 95% credible interval (between the 2.5th and 97.5th percentiles) from a **reduced spatial multinomial model with** $J^* = 2$ under an alternative link function. The land cover types are aspen-paper birch forest, agriculture grassland, and the baseline all others. The covariates are intercept, `reserv, polypr, log totown`, and `avparcel`.

percentile	intercept	`reserv`	`polypr`	`totown`	`avparcel`	
aspen-paper birch						
2.5%	1.450	-0.290	-1.473	-0.092	-0.241	
50%	2.605	1.426	-0.682	0.612	0.653	
97.5%	3.910	3.092	0.051	1.460	1.568	
agriculture grassland						
2.5%	-0.508	-5.024	-2.122	-2.031	-0.613	
50%	0.899	-3.091	-1.230	-1.115	0.414	
97.5%	2.381	-1.253	-0.419	-0.257	1.429	
	V^*_{11}	V^*_{12}	V^*_{22}	Λ^*_{11}	Λ^*_{12}	Λ^*_{22}
2.5%	0.679	-0.171	0.676	0.250	-0.001	0.248
50%	0.903	-0.001	0.944	0.252	0.000	0.251
97.5%	1.199	0.162	1.219	0.252	0.001	0.252
	$\tilde{\Omega}_{11} - \tilde{\Omega}_{01}$	$\tilde{\Omega}_{12} - \tilde{\Omega}_{02}$	$\tilde{\Omega}_{21} - \tilde{\Omega}_{01}$	$\tilde{\Omega}_{22} - \tilde{\Omega}_{02}$		
2.5%	-2.567	-5.143	0.718	-3.887		
50%	1.161	-3.460	3.066	-1.033		
97.5%	3.775	-1.492	4.632	1.961		

Table 7: Hyperparameter values in the prior distributions.

Reduced Spatial Multinomial Models with $J^* = 1$ under the log-ratio link function and an alternative link function	
$\beta_{\cdot\cdot} \sim \mathrm{N}(0, \sigma^2_\beta)$	$\sigma^2_\beta = 1$
$v^* \sim \mathrm{G}(\tilde{\alpha}_1, \tilde{\gamma}_1)$	$\tilde{\alpha}_1 = \tilde{\gamma}_1 = 20$
$\lambda^* \sim \mathrm{N}(0, \xi^2)$	$\xi^2 = 1$
$\Omega_{\cdot\cdot} \sim \mathrm{N}(0, \sigma^2_\Omega)$	$\sigma^2_\Omega = 1$
Full Spatial Multinomial Model under the log-ratio link function Reduced Spatial Multinomial Model with $J^* = 2$ under an alternative link function	
$\beta_{\cdot\cdot} \sim \mathrm{N}(0, \sigma^2_\beta)$	$\sigma^2_\beta = 1$
$\boldsymbol{V} \sim \mathrm{W}(\rho, \psi \boldsymbol{I})$	$\rho = 200$, $\psi = 1/200$
$p(\boldsymbol{\Lambda}) \propto \exp\{-\mathrm{vec}(\boldsymbol{\Lambda})'\mathrm{vec}(\boldsymbol{\Lambda})/\xi^2\}$	$\xi^2 = 1$
Independent Multinomial Models under the log-ratio link function and an alternative link function	
$\beta_{\cdot\cdot} \sim \mathrm{N}(0, \sigma^2_\beta)$	$\sigma^2_\beta = 1$
$\sigma^2_{\cdot} \sim \mathrm{G}(\tilde{\alpha}_2, \tilde{\gamma}_2)$	$\tilde{\alpha}_2 = \tilde{\gamma}_2 = 20$

aspen-paper birch is also plausible. Regional history investigation shows that foresters frequently introduced practices to promote this early successional, pulpwood forest type. For reasons detailed elsewhere (timber contracts, maintenance of stumpage supply to the mills (Steen-Adams et al., 2007, 2011)), timber companies tended to maintain large land holdings, which in northern Wisconsin in the early 20th century meant aspen-paper birch.

Three covariates have a negative association with agriculture-grassland: `reserv`, `totown`, and `polypr`. These results are consistent with related research of the region's environmental and landscape history (Steen-Adams et al., 2007, 2011). Permanent mixed husbandry agriculture (in contrast to shifting cultivation), which 1930 land surveyors would have classified as agriculture, successfully established off the Indian reservation but generally failed on-reservation, corresponding with the variable `reserv`. The negative relationship between land ownership size (`totown`) and agriculture-grassland is also plausible: the conditions of the study area region and historical period promoted small-scale farmsteads, railroad land grantees and real estate companies subdivided land into small parcels affordable to the generally poor, immigrant, farm settlers. Likewise, the negative relationship between agriculture-grassland and quarter-section fragmentation (`polypr`) makes sense for historical land economic reasons: time-consuming, enormous investments of labor and capital were required to develop cutover farmsteads, which promoted agricultural land use on small parcels for practical reasons.

The results of our statistical method to analyze nominal data, specifically historic (1930) land cover pattern, bear several implications for current forest management. One, this approach expanded our understanding of the development of several land covers, and more broadly, the landscape ecology of northern Wisconsin. Specifically, our analysis revealed that several land ownership covariates function as important drivers of landscape pattern (`reserv`, also for agriculture-grassland cover, `polypr`, `totown`); this finding highlights the utility of tailoring policies to specific ownership characteristics to achieve management objectives. Our finding corresponds with those of studies in other regions (e.g., Crow et al. (1999); Stanfield et al. (2002); Spies et al. (2007)) which demonstrate that ownership variables can influence forest landscape pattern, including land cover diversity and patch size, and consequently, the delivery of ecological services to society.

Two, our new statistical methodology, which models the relationship between covariates and categorical response variable while accounting for spatial dependence, contributed capacity to assess historical factors. Our study builds on the expanding literature that demonstrates the utility of historical and long-term data to ecology (Turner, 2005). Investigators must work with the format of existing historical records, however. For example, the source of this study's outcome variable, the Wisconsin Land Economic Inventory, is most appropriately analyzed as nominal data. Our method expands the analytical tools to assess multiple covariates (or, factors). An important management implication is the persistence of social-ecological relationships: the findings of relationships between of 1915 ownership and 1930 land cover (e.g., ownership covariates `totown`, `polypr` and 1930 agriculture-grassland) suggests that social conditions may have a persisting influence on the landscape, indicating the importance of a relatively long-term perspective in developing policies.

Three, our finding of relationships between ownership type (`reserv`) and land cover type suggests the need for greater coordination among ownerships across ecological units (a "multi-ownership perspective"), as other studies have concluded (Spies et al., 2007).

In the absence of a coordinated approach, contrasting motivations, policies, and applicable laws among ownerships can stymie achievement of ecosystem-level management goals.

Four, our statistical method substantially enhanced our capacity to interpret ownership-ecological relationships for specific land covers of interest, which thereby helps managers to discern influential factors on which to focus their interventions. For example, our findings can assist goals to promote agriculture-grassland, a regionally important open land cover. Our result that this land cover type is negatively related to ownership size (i.e., increased likelihood on small ownerships) can signal to managers to develop policies tailored to small owners when the management objective is to promote agriculture grassland.

Chapter 4

Regression for Spatial Binary Data and Penalized QEE

4.1 Overview

Correlated binary data on a spatial lattice are encountered and analyzed extensively in environmental and ecological studies. Various statistical methods have been developed for relating the spatial binary response to covariates in these studies, while properly accounting for spatial dependence.

Besag (1974) used Hammersley-Clifford theorem to construct joint probability of response variables on a lattice by specification of conditional probability models. The resulting models are known as the Markov random field models. In particular, an autologistic model was developed for spatial binary data and accounts for spatial dependence by autoregression. However, the introduction of autoregression results in an unknown normalizing constant in the likelihood function, which makes computation for maximum likehood estimates and statistical inference challenging (see, e.g. Huffer and Wu (1998)).

Alternatively, Diggle et al. (1998) utilized a generalized linear mixed model (GLMM) framework for modeling spatial data. The response variable is modeled by a distribution in the exponential family and is related to covariates and spatial random effects in a link function. Thus, GLMM is suitable for both Gaussian responses and non-Gaussian responses such as binomial and Poisson random variables. Statistical inference can be carried out using Bayesian hierarchical modeling, but computation remains challenging as it involves Markov chain Monte Carlo (MCMC) algorithms.

Other likelihood-based approaches are suitable, but sometimes it is a challenge to attain a full specification of the likelihood function, due to insufficient information and the complicated interactions among the responses. In this case, an estimating equation approach is attractive. Wedderburn (1974) proposed the concept of quasi-likelihood as an extension of generalized least squares method, in which only the mean and variance of the response variables need to be specified. McCullagh and Nelder (1989) generalized the quasi-likelihood to response variables from exponential family of distributions. Wedderburn (1974) and McCullagh and Nelder (1989) assumed that the response variables are independent.

For dependent observations, the correlation matrix for response variables is generally hard to specify. Prentice (1988) developed a generalized estimating equation (GEE) to estimate the correlations, while Lipsitz et al. (1994) used working correlation for longitudinal data that are correlated over time for a given subject but are independent among subjects. In both these methods, however the estimated correlation matrix can be positive semi-definite or even negative definite and thus, the algorithm may not converge, especially when the correlation coefficients are misspecified or the correlation structure is overly complicated.

For spatial binary data, Lin (2008) developed a central limit theorem for a random field under various L_p metrics and derived the consistency and asymptotic normality of quasi-likelihood estimators. Lin (2010) further developed a GEE method for spatial-temporal binary data, but only a single binary covariate was considered and the spatial-temporal dependence is limited to be separable. Lin (2011) introduced a quasi-deviance function for model selection of spatially correlated data, but the proposed method may not be easy to apply for models with overlapping explanatory variables because only strictly non-nested and nested models were considered. This is a limitation that we attempt to overcome when seeking an alternative approach for model selection.

Tibshirani (1996) proposed a least absolute shrinkage and selection operator (Lasso), which performs simultaneous variable selection and parameter estimation. Zou (2006) improved the original Lasso by developing an adaptive Lasso, which utilizes smaller penalties for lager coefficients to remove the estimation bias and hence, enjoys the oracle properties that can only be achieved by original Lasso under certain conditions. Efron et al. (2004) devised the least angle regression algorithm (LARS) which allows computing all Lasso estimates along a path of its tuning parameter with the same computation order as the single ordinary least squares estimate based on the full model. These methods are largely for independent response variables. For dependent data, Dziak and Li (2006) compared several variable selection criteria such as AIC, BIC, C_p, Lasso and SCAD for longitudinal data that are continuous or binary. The SCAD penalty is shown to outperform the other procedures in terms of model selection in a simulation study. Zhu et al. (2010) developed new methodology for selection of covariates and a neighbourhood structure of a spatial linear model via penalized maximum likelihood under an adaptive Lasso.

In this chapter, we develop a new quasi-likelihood method for spatial binary data on a lattice. We propose a quasi-likelihood estimating equation (QEE), where a working correlation matrix is specified but no specific model is assumed about the spatial dependence. Only a few parameters are used to avoid the singularity problem encountered in previous work. Furthermore, we construct a penalized quasi-likelihood and develop a LARS-type algorithm for variable selection. In particular, we generalize the approach taken by Zhu et al. (2010) for Gaussian random field to spatially-correlated binary data. A simulation study and a real data analysis are presented in the end.

4.2 Model and Estimation

4.2.1 Model

Suppose there are n sites on a spatial lattice. The response at each site is binary and is denoted by $Y_i = 1$ or 0, for $i = 1, \ldots, n$. Let $\pi_i = \mathrm{E}(Y_i) = \Pr(Y_i = 1)$ denote the probability of success at site i. Let $\boldsymbol{x}_i = (x_{i0}, \ldots, x_{iK})'$ with $x_{i0} \equiv 1$ denote a $(K+1)$-dimensional vector of explanatory variables at site i and $\boldsymbol{\beta}$ denote the corresponding $(K+1)$-dimensional vector of coefficients of interest. Consider a linear regression in the logit link function of the probability π_i as

$$\mathrm{logit}(\pi_i) = \log\big\{\pi_i/(1-\pi_i)\big\} = \boldsymbol{x}_i'\boldsymbol{\beta}.$$

Then the probability of success at site i relates to the explanatory variables via a logistic function,

$$\pi_i = \pi_i(\boldsymbol{\beta}) = \mathrm{E}(Y_i) = \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\big/\big\{1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\big\}.$$

4.2.2 Quasi-likelihood

Let $\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$ denote an n-dimensional vector of response variables and $\boldsymbol{y} = (y_1, \ldots, y_n)'$ denote the corresponding observations, both at the n sites. Let $\boldsymbol{\pi} = \boldsymbol{\pi}(\boldsymbol{\beta}) = (\pi_1(\boldsymbol{\beta}), \ldots, \pi_n(\boldsymbol{\beta}))'$ denote the n-dimensional mean vector of $\boldsymbol{Y}$. Let $\boldsymbol{\Sigma}$ denote an $n \times n$ covariance matrix of $\boldsymbol{Y}$, which is assumed to be a function of the mean $\boldsymbol{\pi}$ and thus is sometimes denoted by $\boldsymbol{\Sigma}(\boldsymbol{\pi})$. Then the quasi-likelihood function can be written as

$$QL(\boldsymbol{\beta}) = QL(\boldsymbol{\pi}(\boldsymbol{\beta}); \boldsymbol{y}) = \int_{\boldsymbol{t}=\boldsymbol{y}}^{\boldsymbol{t}=\boldsymbol{\pi}} (\boldsymbol{y} - \boldsymbol{t})' \boldsymbol{\Sigma}(\boldsymbol{t})^{-1} d\boldsymbol{t}. \tag{4.1}$$

The integral in (4.1) is on a smooth path in $\mathbb{R}^n$ from $\boldsymbol{t} = \boldsymbol{y}$ to $\boldsymbol{t} = \boldsymbol{\pi}$. By differentiating the QL function, we obtain a quasi-likelihood estimating equation (QEE) for $\boldsymbol{\beta}$,

$$\boldsymbol{D}' \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\pi}) = \boldsymbol{0}, \tag{4.2}$$

where $\boldsymbol{D} = \partial \boldsymbol{\pi}(\boldsymbol{\beta}) \big/ \partial \boldsymbol{\beta}$ is an $n \times (K+1)$ matrix (McCullagh and Nelder (1989)). For $1 \leq i \leq n$ and $0 \leq k \leq K$, the $(i, k+1)$th element of $\boldsymbol{D}$ is

$$D_{ik} = \frac{\partial \pi_i}{\partial \beta_k} = x_{ik} \pi_i (1 - \pi_i).$$

Further, the ith diagonal element of the covariance matrix $\boldsymbol{\Sigma}$ is the variance of Y_i at site i, denoted as $v_i \equiv v_i(\boldsymbol{\beta}) = \mathrm{Var}(Y_i) = \pi_i(1 - \pi_i)$. The (i_1, i_2)th off-diagonal element is the covariance between site i_1 and i_2, $\mathrm{Cov}(Y_{i_1}, Y_{i_2}) = v_{i_1}^{1/2} \rho_{i_1 i_2}(\boldsymbol{\alpha}) v_{i_2}^{1/2}$, where $\boldsymbol{\alpha}$ is a vector of parameters associated with correlation $\rho_{i_1 i_2} = \mathrm{cor}(Y_{i_1}, Y_{i_2})$ between sites i_1 and i_2. Thus, the covariance matrix $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is parameterized by $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.

4.2.3 Estimation of $\boldsymbol{\beta}$

If $\boldsymbol{\alpha}$ were known, the solution of (4.2) can be obtained by a Fisher-scoring algorithm. That is, given a starting value $\hat{\boldsymbol{\beta}}^{(0)}$, we can iteratively compute

$$\hat{\boldsymbol{\beta}}^{(m)} = \hat{\boldsymbol{\beta}}^{(m-1)} + \{\boldsymbol{D}^{(m-1)\prime}\boldsymbol{\Sigma}^{(m-1)^{-1}}\boldsymbol{D}^{(m-1)}\}^{-1}\boldsymbol{D}^{(m-1)\prime}\boldsymbol{\Sigma}^{(m-1)^{-1}}\{\boldsymbol{y} - \boldsymbol{\pi}^{(m-1)}\}, \quad (4.3)$$

where $\boldsymbol{D}^{(m-1)} = \boldsymbol{D}(\hat{\boldsymbol{\beta}}^{(m-1)})$, $\boldsymbol{\Sigma}^{(m-1)} = \boldsymbol{\Sigma}(\boldsymbol{\alpha}, \hat{\boldsymbol{\beta}}^{(m-1)})$, and $\boldsymbol{\pi}^{(m-1)} = \boldsymbol{\pi}(\hat{\boldsymbol{\beta}}^{(m-1)})$, for $m = 1, 2, \ldots$. Let $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}^{(M)}$ denote the quasi-likelihood estimate attained at the Mth iteration, when the convergence criterion is met. The variance of $\hat{\boldsymbol{\beta}}$ is

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \{\boldsymbol{D}^{(M)\prime}\boldsymbol{\Sigma}^{(M)^{-1}}\boldsymbol{D}^{(M)}\}^{-1}.$$

4.2.4 Estimation of $\boldsymbol{\alpha}$

To estimate $\boldsymbol{\alpha}$ and the correlation between sites, we define a neighborhood structure on the lattice. Take Figure 10, a regular grid, as an example. Sites marked as a are the first-order neighbors of site i, b's are the second-order neighbors, c's are the third-order neighbors, and d's are the fourth-order neighbors, etc. We let $i_1 \underset{g}{\sim} i_2$ denote that i_1 and i_2 are neighbors of order g.

There are several different ways to estimate correlation between neighbors. Here we consider two of them.

Method A

Following Prentice (1988), consider a sample correlation of the form

$$Z_{i_1 i_2} = \frac{(y_{i_1} - \pi_{i_1})(y_{i_2} - \pi_{i_2})}{\{\pi_{i_1}(1-\pi_{i_1})\pi_{i_2}(1-\pi_{i_2})\}^{1/2}}.$$

The mean of $Z_{i_1 i_2}$ is $\rho_{i_1 i_2}$ and the variance is

$$w_{i_1 i_2} = 1 + (1-2\pi_{i_1})(1-2\pi_{i2})\{\pi_{i_1}(1-\pi_{i_1})\pi_{i_2}(1-\pi_{i_2})\}^{-1/2}\rho_{i_1 i_2} - \rho_{i_1 i_2}^2,$$

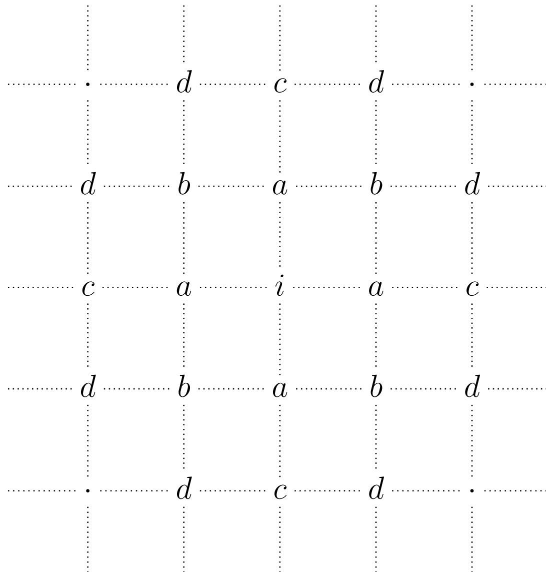

Figure 10: Example sampling sites on a regular grid. Sites a's are the first-order neighbors of site i, b's are the second-order neighbors, c's are the third-order neighbors, and d's are the fourth-order neighbors, etc.

where $i_1 < i_2 \leq n$, $\pi_{i_1} = \pi_{i_1}(\boldsymbol{\beta})$, and $\pi_{i_2} = \pi_{i_2}(\boldsymbol{\beta})$.

Let $\boldsymbol{Z} = (Z_{12}, \ldots, Z_{1n}, Z_{23}, \ldots)'$ and $\boldsymbol{\rho} = (\rho_{12}, \ldots, \rho_{1n}, \rho_{23}, \ldots)'$ denote two $n(n-1)/2$-dimensional vectors. Further, let $\boldsymbol{W} \equiv \boldsymbol{W}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \mathrm{diag}\{w_{12}, \ldots, w_{1n}, w_{23}, \ldots\}$ denote a diagonal matrix of dimension $n(n-1)/2$ and let $\boldsymbol{E} \equiv \boldsymbol{E}(\boldsymbol{\alpha}) = \partial\boldsymbol{\rho}/\partial\boldsymbol{\alpha}$. The vector of correlation parameters $\boldsymbol{\alpha}$ can be estimated by solving a generalized estimating equation

$$\boldsymbol{E}'\boldsymbol{W}^{-1}(\boldsymbol{Z} - \boldsymbol{\rho}) = \boldsymbol{0}. \tag{4.4}$$

An iterative procedure similar to (4.3) for $\boldsymbol{\beta}$ can be applied. That is, given a starting value, $\hat{\boldsymbol{\alpha}}^{(0)}$, iteratively compute

$$\hat{\boldsymbol{\alpha}}^{(m)} = \hat{\boldsymbol{\alpha}}^{(m-1)} + \{\boldsymbol{E}^{(m-1)'}\boldsymbol{W}^{(m-1)^{-1}}\boldsymbol{E}^{(m-1)}\}^{-1}\boldsymbol{E}^{(m-1)'}\boldsymbol{W}^{(m-1)^{-1}}\{\boldsymbol{Z} - \boldsymbol{\rho}^{(m-1)}\}, \tag{4.5}$$

where $\boldsymbol{E}^{(m-1)} = \boldsymbol{E}(\hat{\boldsymbol{\alpha}}^{(m-1)})$, $\boldsymbol{W}^{(m-1)} = \boldsymbol{W}(\hat{\boldsymbol{\alpha}}^{(m-1)}, \boldsymbol{\beta})$, and $\boldsymbol{\rho}^{(m-1)} = \boldsymbol{\rho}(\hat{\boldsymbol{\alpha}}^{(m-1)})$, for $m = 1, 2, \ldots$, until convergence.

Given the starting values, $\hat{\boldsymbol{\alpha}}^{(0)}$ and $\hat{\boldsymbol{\beta}}^{(0)}$, (4.2) and (4.4) can be solved by updating the value of $\hat{\boldsymbol{\alpha}}^{(m-1)}$ and $\hat{\boldsymbol{\beta}}^{(m-1)}$ in (4.3) and (4.5) iteratively until convergence.

Method B

Alternatively, a method of moments approach can be used to estimate $\boldsymbol{\alpha}$, as by Lipsitz et al. (1994). Here, we assume that all the correlations between two gth-order neighbor sites are the same, denoted by ρ_g. Let the vector of correlation parameters be denoted as $\boldsymbol{\alpha} \equiv (\rho_1, \ldots, \rho_G)'$, where G is the highest order of neighbors that we consider. Also, define a working correlation as

$$\hat{\rho}_g \equiv \hat{\rho}_g(\boldsymbol{\beta}) = (N_g - 1)^{-1} \sum_{i_1 \underset{g}{\sim} i_2} \Big[v_{i_1}(\boldsymbol{\beta})^{-1/2} \big\{ y_{i_1} - \pi_{i_1}(\boldsymbol{\beta}) \big\} \big\{ y_{i_2} - \pi_{i_2}(\boldsymbol{\beta}) \big\} v_{i_2}(\boldsymbol{\beta})^{-1/2} \Big],$$

where N_g is the total number of pairs of neighbors with order g. Then, an estimate of the covariance matrix of $\boldsymbol{Y}$ is

$$\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}(\boldsymbol{\beta}) = \boldsymbol{V} + \boldsymbol{V}^{1/2} \bigg(\sum_{g=1}^{G} \hat{\rho}_g \boldsymbol{C}_g \bigg) \boldsymbol{V}^{1/2},$$

where $\boldsymbol{V} \equiv \mathrm{diag}\{v_i\}_{i=1}^n$ and $\boldsymbol{C}_g \equiv \big\{ \mathcal{I}(i_1 \underset{g}{\sim} i_2) \big\}_{i_1, i_2 = 1}^n$ for $g = 1, \ldots, G$ are $n \times n$ matrices. In this way, the matrix $\hat{\boldsymbol{\Sigma}}$ is a function of only $\boldsymbol{\beta}$ and we apply (4.3) to solve (4.2) as before.

4.3 Model Selection

4.3.1 Penalized Quasi-likelihood

Variable selection is an important problem in regression analysis. For regularization methods that involve penalized least squares or penalized likelihoods, the assumption is that some components of $\boldsymbol{\beta}$ are zero and it is of interest to determine and estimate the non-zero coefficients. Here we propose a penalized quasi-likelihood for variable selection in this framework. We refer to $\mathrm{ql}(\boldsymbol{\beta})$ as the quasi-likelihood, which is, more precisely, the log of the quasi-likelihood (4.1). Then we define a penalized quasi-likelihood function as

$$\mathrm{Q}(\boldsymbol{\beta}) = \mathrm{ql}(\boldsymbol{\beta}) - n\sum_{k=0}^{K} p_{\lambda_k}(|\beta_k|),$$

where the second term is a penalty function on the coefficients with regularization (or, tuning) parameters $\{\lambda_k\}$.

Some commonly used penalty functions ar as foolws, smoothly clipped absolute deviation (SCAD) penalty (Fan and Li (2001)) is defined by

$$p_\lambda(|\beta_k|) = \begin{cases} \lambda|\beta_k| & : \quad 0 \leq |\beta_k| < \lambda; \\ \frac{(a^2-1)\lambda^2-(|\beta_k|-a\lambda)^2}{2(a-1)} & : \quad \lambda \leq |\beta_k| < a\lambda; \\ \frac{(a+1)\lambda^2}{2} & : \quad |\beta_k| \leq a\lambda. \end{cases}$$

There are two tuning parameters λ and a and based on a Bayesian argument, it is suggested that $a = 3.7$. For a Lasso penalty, $p_\lambda(|\beta_k|) = \lambda|\beta_k|$ where λ is a tuning parameter (Tibshirani (1996)), whereas for an adaptive Lasso penalty, $p_{\lambda_k}(|\beta_k|) = \lambda_k|\beta_k|$, where λ_k's are tuning parameters (Zou (2006)). One choice of λ_k in an adaptive Lasso is $\lambda_k = \lambda\big/|\hat{\beta}_k|^\gamma$, where $\hat{\beta}_k$ is the result from MLE and $\gamma > 0$. According to the simulation results based on linear models in (Zou (2006)), each penalty function has its own merit and none can universally outperform the other two. However, the adaptive Lasso seems to be able to adaptively combine the strength of the Lassso and the SCAD. That is, with a medium or low level of signal-to-noise ratio (SNR), the adaptive Lasso often outperforms the SCAD, while with a high SNR, the adaptive Lasso is better than Lasso. Thus, we will use the adaptive Lasso penalty, and the resulting penalized quasi-likelihood function is

$$\mathrm{Q}(\boldsymbol{\beta}) = \mathrm{ql}(\boldsymbol{\beta}) - n\sum_{k=0}^{K}\lambda_k(|\beta_k|). \tag{4.6}$$

4.3.2 Maximization of Penalized Quasi-likelihood via LARS

Let $\hat{\boldsymbol{\beta}}^{(0)}$ denote an initial value of $\boldsymbol{\beta}$. At iteration $m = 1, 2, \ldots$, and with $\hat{\boldsymbol{\beta}}^{(m-1)}$ from the previous iteration, we approximate the penalized quasi-likelihood (4.6) by, up to a constant,

$$\mathrm{Q}^*(\boldsymbol{\beta}) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^{(m-1)})'\dot{\mathrm{ql}}(\hat{\boldsymbol{\beta}}^{(m-1)}) + \frac{1}{2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^{(m-1)})'\ddot{\mathrm{ql}}(\hat{\boldsymbol{\beta}}^{(m-1)})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}^{(m-1)}) - n\sum_{k=0}^{K}\lambda_k|\beta_k|$$

where the first-order and second-order derivatives of $\mathrm{ql}(\boldsymbol{\beta})$ with respect to $\boldsymbol{\beta}$ are

$$\begin{aligned}
\dot{\mathrm{ql}}(\hat{\boldsymbol{\beta}}^{(m-1)}) &= \frac{\partial \mathrm{ql}(\hat{\boldsymbol{\beta}}^{(m-1)})}{\partial \boldsymbol{\beta}} = \{\boldsymbol{D}^{(m-1)}\}'\{\boldsymbol{\Sigma}^{(m-1)}\}^{-1}(\boldsymbol{y} - \boldsymbol{\pi}^{(m-1)}),\\
\ddot{\mathrm{ql}}(\hat{\boldsymbol{\beta}}^{(m-1)}) &= \frac{\partial^2 \mathrm{ql}(\hat{\boldsymbol{\beta}}^{(m-1)})}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = -\{\boldsymbol{D}^{(m-1)}\}'\{\boldsymbol{\Sigma}^{(m-1)}\}^{-1}\boldsymbol{D}^{(m-1)},
\end{aligned}$$

and $\boldsymbol{D}^{(m-1)} = \partial\boldsymbol{\pi}(\hat{\boldsymbol{\beta}}^{(m-1)})/\partial\boldsymbol{\beta}$ is an $n \times (K+1)$ matrix.

At iteration $(m-1)$, we propose to update $\hat{\boldsymbol{\beta}}^{(m-1)}$ to

$$\hat{\boldsymbol{\beta}}^{(m)} = \underset{\boldsymbol{\beta}}{\mathrm{argmax}}\{\mathrm{Q}^*(\boldsymbol{\beta})\}, \tag{4.7}$$

and iterate equation (4.7) until convergence.

Let $\boldsymbol{y}^* = (\boldsymbol{A}^{-1})'\big\{\dot{\mathrm{ql}}(\hat{\boldsymbol{\beta}}^{(m-1)}) - \ddot{\mathrm{ql}}(\hat{\boldsymbol{\beta}}^{(m-1)})\hat{\boldsymbol{\beta}}^{(m-1)}\big\}$, $\boldsymbol{X}^* = \boldsymbol{A}\mathrm{diag}\{\lambda_k\}_{k=0}^{K}$, and $\boldsymbol{\beta}^* = \mathrm{diag}\{\lambda_k\}_{k=0}^{K}\boldsymbol{\beta}$, where $\boldsymbol{A}'\boldsymbol{A} = -\ddot{\mathrm{ql}}(\hat{\boldsymbol{\beta}}^{(m-1)})$. It can be shown that the solution of (4.7) can be attained equivalently by

$$\hat{\boldsymbol{\beta}}^{*(m)} = \underset{\boldsymbol{\beta}^*}{\mathrm{argmin}}\Big\{\frac{1}{2}(\boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{\beta}^*)'(\boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{\beta}^*) + n\sum_{k=0}^{K}|\beta_k^*|\Big\}. \tag{4.8}$$

Hence, $\hat{\boldsymbol{\beta}}^{(m)} = \mathrm{diag}\{\lambda_k^{-1}\}_{k=0}^{K}\hat{\boldsymbol{\beta}}^{*(m)}$. Thus, within each iteration, minimization (4.8) can be solved by a LARS algorithm (Efron et al. (2004)).

4.3.3 Regularization Parameter λ

At iteration m, we let

$$\lambda_k = \lambda \log(n)(n|\hat{\beta}_k^{(m-1)}|)^{-1},$$

where $\hat{\beta}_k^{(m-1)}$ is the estimate from last iteraton, for $k = 0, \ldots, K$. Thus, only one tuning parameter λ needs to be estimated. To determine λ, we propose a Bayesian information criterion (BIC) like criterion defined as,

$$\mathrm{BIC}^0(\lambda) = -2\mathrm{ql}^0(\hat{\boldsymbol{\beta}}; \lambda) + \log(n)\mathrm{e}(\lambda),$$

where $\mathrm{e}(\lambda)$ is the number of non-zero coefficients in $\hat{\boldsymbol{\beta}}$. Further, instead of using the integral (4.1), we use $\mathrm{ql}^0(\boldsymbol{\beta}) = \sum_{i=1}^n \Big[y_i \log\big\{\pi_i/(1-\pi_i)\big\} + \log(1-\pi_i)\Big]$ with $\pi_i = \pi_i(\boldsymbol{\beta})$ to approximate the quasi-likelihood $\mathrm{ql}(\boldsymbol{\beta})$ where the observations are assumed to be independent (Dziak (2006)). In each iteration, we select the λ that has the smallest BIC^0 value.

Most regression models contain an intercept term, which is generally not subject to penalty. In that case, we estimate the intercept by its initial value $\hat{\beta}_0^{(0)}$ and only obtain the solution of the other coefficients from (4.8). The same algorithm above may be applied, except that $\boldsymbol{\beta}$ is changed to $\boldsymbol{\beta}_{-0} = (\beta_1, \beta_2, \ldots, \beta_K)'$ and the related vectors and matrices, such as $\dot{\mathrm{ql}}(\cdot)$, $\ddot{\mathrm{ql}}(\cdot)$, $\boldsymbol{y}^*$, $\boldsymbol{X}^*$ and $\boldsymbol{A}$ have a reduced dimension.

4.4 Simulation Study

For simulation, we consider $L \times L$ regular lattices, where $L = 10, 20, 30$, and thus, the sample sizes are $n = 100, 400, 900$. The regression coefficients are set to $\boldsymbol{\beta} =$

$(0.1, 0.5, 0.3, 0, 0)'$. The corresponding covariates at site i are $x_{i0} \equiv 1$, $x_{ik} \sim \mathrm{N}(0,1)$, $k = 1, \ldots, 4$. That is, the last four covariates follow the standard normal distribution and are independent of each other. Furthermore, we use two different methods to generate correlated binary responses Y_i's.

4.4.1 Scenario 1

For each site i, the response Y_i is sumulated from a Bernoulli distribution with probability π_i,

$$Y_i \sim \text{Bernoulli}(\pi_i). \tag{4.9}$$

The success probability π_i is assumed to be

$$\pi_i = \exp(Z_i)\big/\big\{1 + \exp(Z_i)\big\}, \tag{4.10}$$

where the latent variable Z_i is modeled by

$$Z_i = \boldsymbol{x}_i'\boldsymbol{\beta} + \epsilon_i. \tag{4.11}$$

The errors $\{\epsilon_i\}$ are assumed to be a zero-mean Gaussian process such that

$$\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)' \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon), \tag{4.12}$$

where $\boldsymbol{\Sigma}_\epsilon = \mathrm{Var}(\boldsymbol{\epsilon})$ is the covariance matrix of $\boldsymbol{\epsilon}$.

A special case of this process is the conditional auto-regressive of order G (CAR(G)) model with

$$\boldsymbol{\Sigma}_\epsilon = \sigma^2(\boldsymbol{I} - \sum_{g=1}^{G} a_g \boldsymbol{C}_g)^{-1}, \tag{4.13}$$

where $\boldsymbol{C}_g = \big\{\mathcal{I}(i_1 \underset{g}{\sim} i_2)\big\}_{i_1,i_2=1}^{n}$is a neighbor structure matrix as in defined *Method B* in Section 4.2.4 and a_g is an autoregressive coefficient for the gth-order neighbors. The CAR(G) model can be also formulated by the conditional distribution of the error with conditional mean and variance

$$\begin{aligned}
\mathrm{E}(\epsilon_{i_1}|\epsilon_{i_2} : i_2 \underset{g}{\sim} i_1, g = 1,\ldots,G) &= \sum_{g=1}^{G} a_g\Big(\sum_{i_2 \underset{g}{\sim} i_1} \epsilon_{i_2}\Big), \\
\mathrm{Var}(\epsilon_{i_1}|\epsilon_{i_2} : i_2 \underset{g}{\sim} i_1, g = 1,\ldots,G) &= \sigma^2.
\end{aligned}$$

Another special case is the simultaneous auto-regressive of order G (SAR(G)) model with

$$\boldsymbol{\Sigma}_\epsilon = \sigma^2(\boldsymbol{I} - \sum_{g=1}^{G} a_g\boldsymbol{C}_g)^{-1}(\boldsymbol{I} - \sum_{g=1}^{G} a_g\boldsymbol{C}_g')^{-1}, \tag{4.14}$$

which can also be formulated by

$$\boldsymbol{\epsilon} = \boldsymbol{C}\boldsymbol{\epsilon} + \boldsymbol{\nu},$$

where $\boldsymbol{C} = \sum_{g=1}^{G} a_g\boldsymbol{C}_g$ and $\boldsymbol{\nu}$ is an n-dimensional vector of iid errors following $\mathrm{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I}_n)$. We use both CAR and SAR models with a first-order neighbor structure, $a_1 = 0.2$ and $\sigma^2 = 0.25$ to generate Gaussian processes in the simulation study. For each grid size L, a total of 200 data sets are simulated according to (4.9)–(4.12). Figure 2 shows two simulation examples on a 20×20 spatial lattice.

Four different methods are implemented to analyze the simulated data. The first method is MLE, in which an iteratively reweighted least squares algorithm is applied under the generalized linear model (4.9)–(4.12) and the assumption that the responses are independent. The second method is quasi-likelihood estimating equation (QEE) using a working correlation matrix (*Method B*). Four orders of the neighbor structure

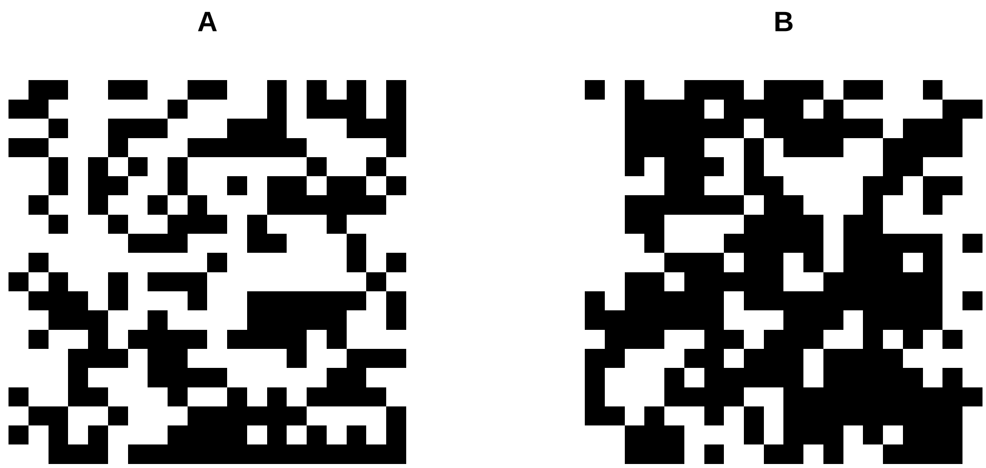

Figure 11: Examples of simulated binary data on a 20×20 lattice with black pixel for $Y_i = 1$ and white pixel for $Y_i = 0$ under a latent Gaussian process: (A) CAR(1) model and (B) SAR(1) model. The parameters of the two spatial autoregressive models are set to $a_1 = 0.2$ and $\sigma^2 = 0.25$.

is considered as in Figure 10 and thus, the vector of correlation parameters is $\boldsymbol{\alpha} = (\rho_1, \rho_2, \rho_3, \rho_4)'$. The third method is penalized QEE (PQEE), in which the adaptive Lasso is applied for variable selection in addition to parameter estimation as in QEE. The last one is QEE with the knowledge that there are only two non-zero coefficients excluding the intercept (or, ORA). We will compare its result with PQEE in order to verify the oracle property of the PQEE. The estimation results of the simulations from CAR and SAR models are shown in the following tables.

Tables 8 and 9 show the means, standard deviations (SD) and mean squared errors (MSE) of the estimates using four estimation methods above applied to the 200 data sets simulated based on CAR(1) and SAR(1) models. The means of the estimates based on simulated data from CAR and SAR models are quite different even for the same estimation method and the same grid size. There is generally more bias with the SAR than the CAR model, possibly because SAR models tend to have stronger correlations as shown in Table 11.

Table 10 shows how many times each coefficient is estimated to be zero among the 200 replicates and the selection rates of different final models via PQEE. The PQEE can choose the zero coefficients correctly with few exceptions, but the non-zero coefficients sometimes are estimated to be zero by mistake, especially for $\beta_2 = 0.3$. It appears that a smaller coefficient has a larger chance to be shrunk to zero, which is plausible. When the sample size is smaller and the spatial correlation is stronger, the non-zero estimates seem to have larger means and more bias under PQEE than the results from other estimation methods in Tables 8 and 9, but this is because we only count those non-zero estimates from PQEE. The means of estimates of β_2 are larger than the true value when grid size is 10×10 and 20×20, and the standard deviations always smaller than other algorithms for the same reason. As the grid size increases, the PQEE algorithm has a better chance to select the true model according to the second part of Table 10.

Table 8: Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). The latent spatial model is CAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 0.1$	0.077	0.079	0.077	0.082	0.255	0.256	0.255	0.25	0.065	0.066	0.065	0.063
	$\beta_1 = 0.5$	0.518	0.52	0.652	0.507	0.241	0.24	0.18	0.233	0.058	0.058	0.055	0.054
	$\beta_2 = 0.3$	0.258	0.254	0.553	0.247	0.259	0.265	0.267	0.251	0.068	0.072	0.133	0.065
	$\beta_3 = 0$	0.004	0.005	−0.143	NA	0.231	0.243	0.495	NA	0.053	0.059	0.241	NA
	$\beta_4 = 0$	−0.001	0.001	−0.231	NA	0.245	0.257	0.538	NA	0.06	0.066	0.319	NA
20x20	$\beta_0 = 0.1$	0.077	0.077	0.077	0.075	0.109	0.108	0.109	0.107	0.012	0.012	0.012	0.012
	$\beta_1 = 0.5$	0.468	0.468	0.466	0.466	0.108	0.11	0.104	0.108	0.013	0.013	0.012	0.013
	$\beta_2 = 0.3$	0.295	0.295	0.34	0.293	0.106	0.108	0.075	0.106	0.011	0.012	0.007	0.011
	$\beta_3 = 0$	0.001	0.002	0.073	NA	0.105	0.109	0.247	NA	0.011	0.012	0.06	NA
	$\beta_4 = 0$	0.003	0.003	0.009	NA	0.098	0.098	0.315	NA	0.01	0.01	0.075	NA
30x30	$\beta_0 = 0.1$	0.101	0.101	0.101	0.101	0.082	0.082	0.082	0.082	0.007	0.007	0.007	0.007
	$\beta_1 = 0.5$	0.473	0.472	0.471	0.472	0.069	0.069	0.069	0.069	0.005	0.006	0.006	0.006
	$\beta_2 = 0.3$	0.287	0.287	0.291	0.287	0.068	0.068	0.062	0.067	0.005	0.005	0.004	0.005
	$\beta_3 = 0$	−0.008	−0.009	−0.043	NA	0.072	0.073	0.177	NA	0.005	0.005	0.029	NA
	$\beta_4 = 0$	0	0	−0.062	NA	0.063	0.063	0.186	NA	0.004	0.004	0.027	NA

Table 9: Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). The latent spatial model is SAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 0.1$	0.085	0.087	0.085	0.086	0.308	0.311	0.308	0.3	0.095	0.096	0.095	0.09
	$\beta_1 = 0.5$	0.484	0.469	0.622	0.462	0.237	0.234	0.143	0.226	0.056	0.055	0.035	0.052
	$\beta_2 = 0.3$	0.26	0.258	0.585	0.244	0.255	0.257	0.161	0.246	0.066	0.067	0.106	0.063
	$\beta_3 = 0$	−0.012	−0.008	−0.155	NA	0.251	0.251	0.541	NA	0.063	0.063	0.292	NA
	$\beta_4 = 0$	−0.039	−0.039	−0.333	NA	0.25	0.253	0.41	NA	0.064	0.065	0.268	NA
20x20	$\beta_0 = 0.1$	0.077	0.077	0.077	0.076	0.145	0.144	0.145	0.143	0.021	0.021	0.021	0.021
	$\beta_1 = 0.5$	0.455	0.454	0.45	0.451	0.106	0.107	0.1	0.105	0.013	0.014	0.012	0.013
	$\beta_2 = 0.3$	0.277	0.278	0.332	0.276	0.108	0.106	0.077	0.104	0.012	0.012	0.007	0.011
	$\beta_3 = 0$	0.007	0.007	0.024	NA	0.098	0.1	0.414	NA	0.01	0.01	0.086	NA
	$\beta_4 = 0$	0.009	0.008	−0.008	NA	0.096	0.096	0.266	NA	0.009	0.009	0.053	NA
30x30	$\beta_0 = 0.1$	0.102	0.102	0.102	0.102	0.112	0.111	0.112	0.111	0.012	0.012	0.012	0.012
	$\beta_1 = 0.5$	0.451	0.451	0.449	0.451	0.074	0.073	0.073	0.073	0.008	0.008	0.008	0.008
	$\beta_2 = 0.3$	0.273	0.273	0.279	0.273	0.068	0.067	0.061	0.067	0.005	0.005	0.004	0.005
	$\beta_3 = 0$	−0.002	−0.001	−0.003	NA	0.075	0.075	0.171	NA	0.006	0.006	0.026	NA
	$\beta_4 = 0$	−0.004	−0.004	−0.059	NA	0.063	0.063	0.178	NA	0.004	0.004	0.03	NA

Table 10: Number of times a coefficient is identified to be zero out of 200 simulations for CAR and SAR models. Also, the rates of different final models selected by penalized quasi-likelihood estimating equation (PQEE). The five digits show whether the five estimated coefficients in selected models are zero or not, with 1 denoting non-zero and 0 denoting zero. In bold are the true model '11100' and the corresponding selection rate.

		CAR	SAR	CAR	SAR
10x10	$\beta_0 = 0.1$	NA	NA	10000 10010 10100 11000 11001 11010 11011	10000 10001 10010 10100 10101 11000 11001
	$\beta_1 = 0.5$	101	106	0.470 0.005 0.030 0.300 0.030 0.025 0.005	0.475 0.020 0.010 0.010 0.015 0.305 0.015
	$\beta_2 = 0.3$	167	170	**11100** 11101 11110 11111	11010 11011 **11100** 11101 11110
	$\beta_3 = 0$	190	188	**0.100** 0.020 0.010 0.005	0.020 0.005 **0.080** 0.020 0.025
	$\beta_4 = 0$	188	185		
20x20	$\beta_0 = 0.1$	NA	NA	10000 11000 **11100** 11101 11110 11111	10000 10100 **11000** 11100 11101 11110
	$\beta_1 = 0.5$	4	4	0.020 0.270 **0.645** 0.015 0.045 0.005	0.015 0.005 **0.350** 0.600 0.020 0.010
	$\beta_2 = 0.3$	58	73		
	$\beta_3 = 0$	190	198		
	$\beta_4 = 0$	196	196		
30x30	$\beta_0 = 0.1$	NA	NA	11000 **11100** 11101 11110	11000 **11100** 11101 11110 11111
	$\beta_1 = 0.5$	0	0	0.030 **0.915** 0.015 0.040	0.065 **0.870** 0.025 0.035 0.005
	$\beta_2 = 0.3$	6	13		
	$\beta_3 = 0$	192	192		
	$\beta_4 = 0$	197	194		

Table 11: Mean of the estimates of correlation parameters, $(\rho_1, \rho_2, \rho_3, \rho_4)'$, by using quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA).

		CAR				SAR		
		QEE	PQEE	ORA		QEE	PQEE	ORA
10x10	ρ_1	0.008	0.006	0.009	ρ_1	0.036	0.034	0.037
	ρ_2	0.007	0.005	0.007	ρ_2	0.023	0.022	0.024
	ρ_3	0.003	0.005	0.005	ρ_3	0.008	0.007	0.009
	ρ_4	−0.009	−0.008	−0.008	ρ_4	0.000	0.000	0.001
20x20	ρ_1	0.009	0.009	0.009	ρ_1	0.052	0.051	0.052
	ρ_2	0.002	0.003	0.003	ρ_2	0.034	0.033	0.034
	ρ_3	0.010	0.009	0.009	ρ_3	0.023	0.023	0.023
	ρ_4	0.000	0.000	0.000	ρ_4	0.017	0.017	0.017
30x30	ρ_1	0.013	0.013	0.013	ρ_1	0.053	0.053	0.053
	ρ_2	0.007	0.007	0.007	ρ_2	0.035	0.035	0.035
	ρ_3	0.005	0.005	0.005	ρ_3	0.025	0.025	0.025
	ρ_4	0.000	0.000	0.000	ρ_4	0.016	0.016	0.016

From Tables 8 and 9, we may find that MLE and QEE tend to give similar results. It is possible that spatial correlations among the responses from simulated data are not strong enough to yield vastly different results under the independent and spatially dependent assumptions. Indeed, the strongest spatial correlation between two nearest neighbors is about 0.05 (Table 11). Even two random variable sets with size 100 from independent standard normal distribution can have a correlation larger than 0.05 with a substantial probability according to a separate simulation. This led us to believe that Scenario 1 is not quite proper to examine the QEE method, which considers spatial correlation between neighbors and is compared against MLE that incorrectly treats the responses as being independent. Another issue is that none of the estimation methods seems to show desirable asymptotic properties of the parameter estimates. For example, the means of the QEE estimates of β_2 are 0.469, 0.454 and 0.451 for SAR models on 10×10, 20×20 and 30×30 latices, which do not tend to the true value 0.5 as the sample size increases. A possible reason for this unexpected result is that the the dependence in the error is not fully captured in the dependence of the response variable after the transformation (4.10) from the latent variable in (4.11) to the response (4.9). In particular, even when the correlation between two error terms at neighboring sites, ρ, is large, it appears that the QEE method can not capture the dependence between the responses at these two sites through $\boldsymbol{\alpha}$. Thus, we consider a different scenario when simulating data sets.

4.4.2 Scenario 2

In this scenario, the binary response Y_i is the indicator variable of the event $\epsilon_i < c_i$ for some fixed cutoff value c_i, such that

$$Y_i = \mathcal{I}(\epsilon_i < c_i). \tag{4.15}$$

We assume $\{\epsilon_i : i = 1, \ldots, n\}$ is a Gaussian process generated from a multivariate normal distribution with zero mean, unit variance, and correlation matrix $\boldsymbol{\Sigma}_\epsilon$,

$$\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)' \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon). \tag{4.16}$$

Since the marginal distribution of ϵ_i is a standard normal distribution, if we set the cutoff c_i to be the π_i quantile of a standard normal distribution, then $\mathrm{P}(Y_i = 1) = \mathrm{E}(Y_i) = \mathrm{P}(\epsilon_i < c_i) = \pi_i$. The probability π_i comes from a logistic function,

$$\pi_i = \mathrm{logit}^{-1}(\boldsymbol{x}_i'\boldsymbol{\beta}) = \exp(\boldsymbol{x}_i'\boldsymbol{\beta}) \big/ \big\{1 + \exp(\boldsymbol{x}_i'\boldsymbol{\beta})\big\}. \tag{4.17}$$

The correlation among binary responses is from the Gaussian process $\{\epsilon_i\}$,

$$\mathrm{cor}(Y_i, Y_{i'}) = \frac{\Phi_2\{c_i, c_{i'}; \mathrm{cor}(\epsilon_i, \epsilon_{i'})\} - \pi_i \pi_{i'}}{\{\pi_i(1-\pi_i)\pi_{i'}(1-\pi_{i'})\}^{1/2}},$$

where Φ_2 denotes the cumulative distribution function of bivariate normal distribution. As in Scenario 1, we use CAR(1) with $\boldsymbol{\Sigma}_\epsilon = \mathrm{diag}\{\sigma_i\}(\boldsymbol{I} - a_1\boldsymbol{C}_1)^{-1}\mathrm{diag}\{\sigma_i\}$ and SAR(1) model with $\boldsymbol{\Sigma}_\epsilon = \mathrm{diag}\{\sigma_i\}(\boldsymbol{I} - a_1\boldsymbol{C}_1)^{-1}(\boldsymbol{I} - a_1\boldsymbol{C}_1')^{-1}\mathrm{diag}\{\sigma_i\}$ to generate the Gaussian processes, where σ_i scales the diagonal elements of $\boldsymbol{\Sigma}_\epsilon$ to be 1, $a_1 = 0.2$, and $\boldsymbol{C}_1$ is a first-order neighbor structure matrix. For each L, a total of 200 data sets are simulated according to (4.15)–(4.17). Figure 12 shows two simulation examples on a 20×20 spatial lattice.

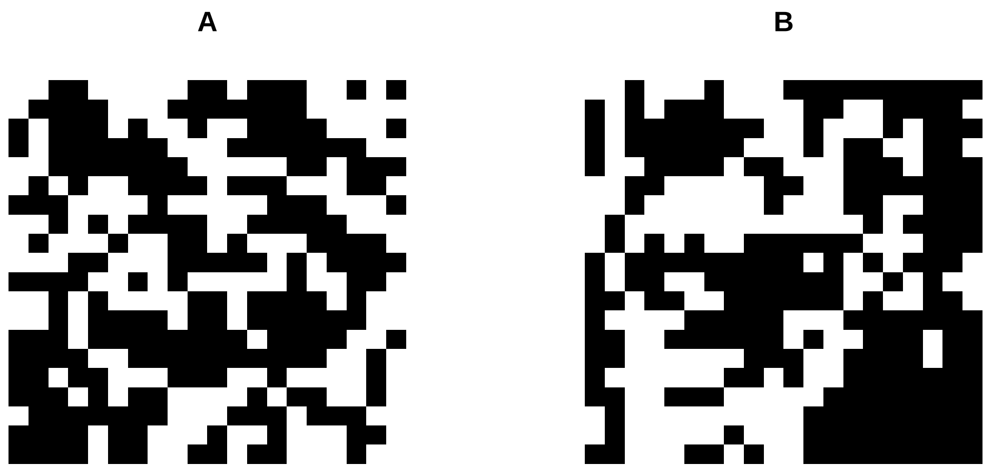

Figure 12: Examples of simulated binary data on a 20×20 lattice with black pixel for $Y_i = 1$ and white pixel for $Y_i = 0$ under a latent Gaussian process: (A) CAR(1) model and (B) SAR(1) model. The parameters of the two spatial autoregressive models are set to $a_1 = 0.2$.

Four different estimation methods are implemented to analyze the simulated data as in Scenario 1. The first method is MLE, that ignores spatial dependence, whereas the second method is QEE using a working correlation matrix. The third is penalized QEE and the last one is QEE under the true model with only two non-zero coefficients. Up to the fourth-order neighbor structure is considered in all three QEE based methods, which means $\boldsymbol{\alpha} \in \mathbb{R}^4$. The estimation results of the simulations from CAR and SAR models are shown in the following tables.

Table 12: Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). The latent spatial model is CAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 0.1$	0.072	0.077	0.072	0.078	0.301	0.297	0.301	0.291	0.091	0.088	0.091	0.085
	$\beta_1 = 0.5$	0.559	0.55	0.689	0.54	0.251	0.248	0.178	0.241	0.066	0.064	0.067	0.059
	$\beta_2 = 0.3$	0.347	0.337	0.613	0.326	0.26	0.271	0.187	0.259	0.07	0.074	0.133	0.067
	$\beta_3 = 0$	−0.01	−0.015	−0.109	NA	0.23	0.235	0.521	NA	0.053	0.055	0.254	NA
	$\beta_4 = 0$	0.006	0.006	0.108	NA	0.235	0.239	0.44	NA	0.055	0.057	0.186	NA
20x20	$\beta_0 = 0.1$	0.079	0.08	0.079	0.08	0.156	0.155	0.156	0.155	0.025	0.024	0.025	0.024
	$\beta_1 = 0.5$	0.51	0.509	0.499	0.506	0.116	0.119	0.112	0.115	0.014	0.014	0.013	0.013
	$\beta_2 = 0.3$	0.315	0.316	0.353	0.315	0.121	0.121	0.093	0.119	0.015	0.015	0.011	0.014
	$\beta_3 = 0$	−0.01	−0.013	−0.038	NA	0.113	0.112	0.25	NA	0.013	0.013	0.051	NA
	$\beta_4 = 0$	−0.015	−0.016	−0.067	NA	0.109	0.107	0.227	NA	0.012	0.012	0.05	NA
30x30	$\beta_0 = 0.1$	0.098	0.098	0.098	0.097	0.111	0.111	0.111	0.111	0.012	0.012	0.012	0.012
	$\beta_1 = 0.5$	0.502	0.499	0.495	0.499	0.075	0.075	0.074	0.075	0.006	0.006	0.006	0.006
	$\beta_2 = 0.3$	0.301	0.3	0.297	0.299	0.066	0.067	0.061	0.066	0.004	0.004	0.004	0.004
	$\beta_3 = 0$	0.003	0.003	−0.032	NA	0.067	0.065	0.225	NA	0.004	0.004	0.026	NA
	$\beta_4 = 0$	−0.001	−0.004	0.166	NA	0.071	0.066	0.053	NA	0.005	0.004	0.03	NA

Table 13: Mean, standard deviation and mean squared error of the estimates of coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). The latent spatial model is SAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 0.1$	0.085	0.086	0.085	0.084	0.517	0.495	0.517	0.487	0.266	0.244	0.266	0.236
	$\beta_1 = 0.5$	0.564	0.554	0.647	0.543	0.241	0.215	0.169	0.211	0.062	0.049	0.05	0.046
	$\beta_2 = 0.3$	0.356	0.347	0.571	0.343	0.251	0.228	0.154	0.224	0.066	0.054	0.097	0.052
	$\beta_3 = 0$	−0.001	−0.01	−0.048	NA	0.236	0.217	0.498	NA	0.055	0.047	0.225	NA
	$\beta_4 = 0$	−0.015	−0.024	−0.03	NA	0.215	0.2	0.453	NA	0.046	0.04	0.177	NA
20x20	$\beta_0 = 0.1$	0.084	0.088	0.084	0.087	0.267	0.256	0.267	0.254	0.071	0.065	0.071	0.064
	$\beta_1 = 0.5$	0.526	0.523	0.506	0.521	0.112	0.108	0.11	0.108	0.013	0.012	0.012	0.012
	$\beta_2 = 0.3$	0.309	0.31	0.34	0.31	0.112	0.102	0.073	0.101	0.013	0.01	0.007	0.01
	$\beta_3 = 0$	−0.006	−0.005	0.052	NA	0.101	0.1	0.284	NA	0.01	0.01	0.056	NA
	$\beta_4 = 0$	−0.004	−0.01	−0.151	NA	0.112	0.092	0.159	NA	0.013	0.008	0.045	NA
30x30	$\beta_0 = 0.1$	0.096	0.094	0.096	0.095	0.182	0.178	0.182	0.178	0.033	0.032	0.033	0.031
	$\beta_1 = 0.5$	0.502	0.501	0.492	0.501	0.08	0.072	0.073	0.071	0.006	0.005	0.005	0.005
	$\beta_2 = 0.3$	0.307	0.303	0.296	0.303	0.068	0.063	0.057	0.063	0.005	0.004	0.003	0.004
	$\beta_3 = 0$	0.002	−0.003	−0.049	NA	0.069	0.061	0.153	NA	0.005	0.004	0.018	NA
	$\beta_4 = 0$	0.005	0.002	0.009	NA	0.064	0.058	0.166	NA	0.004	0.003	0.018	NA

Table 14: Number of times a coefficient is identified to be zero out of 200 simulations for CAR and SAR models. Also, the rates of different final models selected by by penalized quasi-likelihood estimating equation (PQEE). The five digits show whether the five estimated coefficients in selected models are zero or not, with 1 denoting non-zero and 0 denoting zero. In bold are the true model '11100' and the corresponding selection rate.

		CAR	SAR	CAR	SAR
10x10	$\beta_0 = 0.1$	NA	NA	10000 10001 10010 10100 11000 11001 11010	10000 10001 10100 11000 11001 11010
	$\beta_1 = 0.5$	92	89	0.405 0.005 0.005 0.045 0.295 0.015 0.010	0.395 0.005 0.045 0.295 0.005 0.020
	$\beta_2 = 0.3$	148	144	11011 **11100** 11101 11110	**11100** 11101 11110 11111
	$\beta_3 = 0$	191	190	0.005 **0.165** 0.025 0.025	**0.185** 0.020 0.025 0.005
	$\beta_4 = 0$	190	193		
20x20	$\beta_0 = 0.1$	NA	NA	10000 11000 11010 **11100** 11101 11110	11000 11001 **11100** 11101 11110 11111
	$\beta_1 = 0.5$	2	0	0.010 0.245 0.005 **0.685** 0.035 0.010	0.270 0.005 **0.685** 0.025 0.010 0.005
	$\beta_2 = 0.3$	52	55	11111	
	$\beta_3 = 0$	195	197	0.010	
	$\beta_4 = 0$	191	193		
30x30	$\beta_0 = 0.1$	NA	NA	11000 **11100** 11101 11110	11000 **11100** 11101 11110
	$\beta_1 = 0.5$	0	0	0.025 **0.945** 0.020 0.010	0.035 **0.935** 0.015 0.015
	$\beta_2 = 0.3$	5	7		
	$\beta_3 = 0$	198	197		
	$\beta_4 = 0$	196	197		

Table 15: Mean of the estimates of correlation parameters, $(\rho_1, \rho_2, \rho_3, \rho_4)'$, by using quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA).

		CAR				SAR		
		QEE	PQEE	ORA		QEE	PQEE	ORA
10x10	ρ_1	0.137	0.123	0.133	ρ_1	0.268	0.248	0.264
	ρ_2	0.052	0.045	0.051	ρ_2	0.149	0.136	0.146
	ρ_3	0.021	0.018	0.020	ρ_3	0.076	0.067	0.075
	ρ_4	−0.004	−0.004	−0.003	ρ_4	0.031	0.025	0.030
20x20	ρ_1	0.159	0.156	0.158	ρ_1	0.326	0.321	0.325
	ρ_2	0.072	0.071	0.072	ρ_2	0.203	0.198	0.202
	ρ_3	0.045	0.044	0.045	ρ_3	0.133	0.130	0.133
	ρ_4	0.025	0.024	0.025	ρ_4	0.090	0.087	0.089
30x30	ρ_1	0.158	0.157	0.158	ρ_1	0.331	0.328	0.330
	ρ_2	0.073	0.073	0.073	ρ_2	0.213	0.210	0.212
	ρ_3	0.044	0.044	0.044	ρ_3	0.140	0.139	0.140
	ρ_4	0.028	0.028	0.028	ρ_4	0.104	0.103	0.104

Tables 12 and 13 show the means, SDs and MSEs of the estimates from simulated data. The estimates of the coefficients via all methods tend to their true values as the grid size increases, as the means are closer to the truth and the MSEs become smaller. Focusing on one non-zero coefficient, such as $\beta_1 = 0.5$, the means of the estimates based on simulated data from CAR and SAR models tend to be different even with the same method on the same grid size, possibly because CAR and SAR models have different structures and SAR models have stronger correlations according to Table 15. For CAR models, the correlations among the responses from simulated data are not strong enough to make estimates under independent and dependent assumptions markedly different and thus, the MLE and QEE give similar results. However, for SAR models, the correlations are much stronger and the QEE method outperforms the MLE in terms of the means, SDs and MSEs, where the MLE ignores the spatial dependence.

Table 14 shows how many times each coefficient is estimated to be zero in 200 replicates and the rate of different final models via PQEE. The PQEE can choose the zero parameters correctly with few exceptions, but the non-zero parameters sometimes are estimated to be zero by mistake, especially for $\beta_2 = 0.3$. It is reasonable that smaller coefficient has larger probability to be selected out. When the sample size is smaller and the correlation is stronger, the estimates of non-zero coefficients via PQEE have larger means and bias than the results from other methods in Tables 12 and 13, again because only those larger estimates are kept in models and we only use those non-zero estimates. But when the grid size increases to 30×30, the the estimates of non-zero coefficients via different algorithms are very close to each other, and PQEE provide slightly smaller SDs and MSEs. As the grid size increases, the PQEE method have better chances to select the true model according to the second part of Table 14. In addition, the estimates of the non-zero coefficients via PQEE tend to the estimates from the oracle QEE.

The simulation method in this scenario is more reasonable than the previous one, because the correlations among the binary responses are realized via a correlated truncated process. That is, the marginal success probability of each site is a pre-determined constant π_i in (4.17), and it appears to adequately capture the spatial dependence via the estimation of the correlation parameter $\boldsymbol{\alpha}$ under the QEE method.

4.4.3 Other Set-ups

In order to study the robustness of our QEE and penalized QEE methods, we use the simulation methods in Scenario 2 with different set-ups of coefficients. Varying $\boldsymbol{\beta}$ to be $(-0.5, 0.5, 0.3, 0, 0)$ (Table 16), $(0, 0.5, 0.3, 0, 0)$ (Table 17) and $(1, 0.5, 0.3, 0, 0)$ (Table 18) allows to study the effect of different intercepts, whereas varying $\boldsymbol{\beta}$ is be $(0.1, 0.7, 0.3, 0, 0)$ (Table 19), $(0.1, 1, 0.3, 0, 0)$ (Table 20) and $(0.1, 1.5, 0.3, 0, 0)$ (Table 21) allows to study the effect of larger slopes. Both CAR and SAR model are considered and the following tables give the estimation results.

Table 16: Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). $\boldsymbol{\beta} = (-0.5, 0.5, 0.3, 0, 0)'$. The latent spatial model is CAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = -0.5$	-0.52	-0.512	-0.52	-0.499	0.384	0.378	0.384	0.363	0.147	0.143	0.147	0.131
	$\beta_1 = 0.5$	0.536	0.522	0.709	0.514	0.269	0.268	0.212	0.255	0.073	0.072	0.088	0.065
	$\beta_2 = 0.3$	0.3	0.293	0.58	0.294	0.246	0.242	0.112	0.236	0.06	0.058	0.09	0.055
	$\beta_3 = 0$	-0.023	-0.016	-0.163	NA	0.226	0.226	0.562	NA	0.051	0.051	0.303	NA
	$\beta_4 = 0$	-0.009	-0.016	-0.09	NA	0.248	0.242	0.53	NA	0.061	0.059	0.257	NA
20x20	$\beta_0 = -0.5$	-0.505	-0.505	-0.505	-0.503	0.172	0.169	0.172	0.169	0.029	0.029	0.029	0.028
	$\beta_1 = 0.5$	0.514	0.515	0.505	0.512	0.107	0.104	0.1	0.103	0.012	0.011	0.01	0.011
	$\beta_2 = 0.3$	0.303	0.306	0.347	0.304	0.106	0.105	0.076	0.105	0.011	0.011	0.008	0.011
	$\beta_3 = 0$	-0.002	0.002	0.002	NA	0.105	0.101	0.264	NA	0.011	0.01	0.06	NA
	$\beta_4 = 0$	0.016	0.015	0.077	NA	0.116	0.115	0.275	NA	0.014	0.013	0.075	NA
30x30	$\beta_0 = -0.5$	-0.519	-0.516	-0.519	-0.515	0.115	0.114	0.115	0.113	0.014	0.013	0.014	0.013
	$\beta_1 = 0.5$	0.506	0.507	0.501	0.506	0.078	0.077	0.077	0.076	0.006	0.006	0.006	0.006
	$\beta_2 = 0.3$	0.304	0.306	0.308	0.306	0.077	0.075	0.066	0.075	0.006	0.006	0.004	0.006
	$\beta_3 = 0$	-0.002	-0.002	-0.006	NA	0.075	0.073	0.206	NA	0.006	0.005	0.035	NA
	$\beta_4 = 0$	-0.001	0	-0.025	NA	0.08	0.074	0.181	NA	0.006	0.005	0.027	NA

Table 16: (cont') Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). $\boldsymbol{\beta} = (-0.5, 0.5, 0.3, 0, 0)'$. The latent spatial model is SAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
	$\beta_0 = -0.5$	-0.579	-0.553	-0.579	-0.54	0.547	0.517	0.547	0.505	0.304	0.269	0.304	0.256
	$\beta_1 = 0.5$	0.609	0.567	0.694	0.556	0.296	0.273	0.226	0.262	0.099	0.079	0.088	0.071
10x10	$\beta_2 = 0.3$	0.328	0.317	0.564	0.315	0.246	0.217	0.152	0.212	0.061	0.047	0.092	0.045
	$\beta_3 = 0$	-0.028	-0.019	-0.019	NA	0.234	0.205	0.433	NA	0.055	0.042	0.167	NA
	$\beta_4 = 0$	0.001	0.004	0.145	NA	0.244	0.218	0.441	NA	0.059	0.047	0.196	NA
	$\beta_0 = -0.5$	-0.512	-0.506	-0.512	-0.506	0.276	0.269	0.276	0.268	0.076	0.072	0.076	0.072
	$\beta_1 = 0.5$	0.535	0.531	0.513	0.528	0.112	0.102	0.097	0.1	0.014	0.011	0.01	0.011
20x20	$\beta_2 = 0.3$	0.309	0.319	0.342	0.317	0.113	0.107	0.083	0.107	0.013	0.012	0.009	0.012
	$\beta_3 = 0$	0.006	0.013	0.01	NA	0.112	0.097	0.254	NA	0.012	0.01	0.054	NA
	$\beta_4 = 0$	0.007	0.003	0.107	NA	0.111	0.099	0.31	NA	0.012	0.01	0.075	NA
	$\beta_0 = -0.5$	-0.532	-0.529	-0.532	-0.528	0.177	0.171	0.177	0.171	0.032	0.03	0.032	0.03
	$\beta_1 = 0.5$	0.507	0.507	0.497	0.507	0.083	0.076	0.077	0.075	0.007	0.006	0.006	0.006
30x30	$\beta_2 = 0.3$	0.301	0.306	0.299	0.306	0.076	0.07	0.063	0.07	0.006	0.005	0.004	0.005
	$\beta_3 = 0$	-0.002	0.003	-0.192	NA	0.072	0.064	0.033	NA	0.005	0.004	0.038	NA
	$\beta_4 = 0$	-0.004	-0.005	0.044	NA	0.073	0.063	0.149	NA	0.005	0.004	0.02	NA

Table 17: Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). $\boldsymbol{\beta} = (0, 0.5, 0.3, 0, 0)'$. The latent spatial model is CAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 0$	−0.05	−0.04	−0.05	−0.042	0.363	0.358	0.363	0.353	0.134	0.129	0.134	0.126
	$\beta_1 = 0.5$	0.546	0.54	0.683	0.531	0.262	0.248	0.187	0.242	0.07	0.063	0.068	0.059
	$\beta_2 = 0.3$	0.332	0.318	0.583	0.313	0.24	0.246	0.156	0.237	0.058	0.061	0.104	0.056
	$\beta_3 = 0$	−0.032	−0.025	−0.079	NA	0.246	0.247	0.533	NA	0.061	0.061	0.266	NA
	$\beta_4 = 0$	−0.002	0.006	0.275	NA	0.235	0.231	0.548	NA	0.055	0.053	0.342	NA
20x20	$\beta_0 = 0$	−0.003	−0.003	−0.003	−0.002	0.179	0.179	0.179	0.178	0.032	0.032	0.032	0.031
	$\beta_1 = 0.5$	0.512	0.507	0.498	0.506	0.108	0.105	0.108	0.105	0.012	0.011	0.012	0.011
	$\beta_2 = 0.3$	0.305	0.304	0.336	0.305	0.103	0.099	0.076	0.097	0.011	0.01	0.007	0.009
	$\beta_3 = 0$	0.012	0.009	0.169	NA	0.106	0.102	0.204	NA	0.011	0.01	0.063	NA
	$\beta_4 = 0$	0.018	0.018	0.089	NA	0.106	0.107	0.228	NA	0.011	0.012	0.053	NA
30x30	$\beta_0 = 0$	−0.006	−0.006	−0.006	−0.006	0.103	0.102	0.103	0.102	0.011	0.01	0.011	0.01
	$\beta_1 = 0.5$	0.509	0.508	0.502	0.507	0.078	0.075	0.076	0.074	0.006	0.006	0.006	0.006
	$\beta_2 = 0.3$	0.295	0.295	0.294	0.294	0.071	0.069	0.063	0.069	0.005	0.005	0.004	0.005
	$\beta_3 = 0$	0	−0.001	0.149	NA	0.07	0.069	0.016	NA	0.005	0.005	0.022	NA
	$\beta_4 = 0$	0	−0.001	0.096	NA	0.07	0.067	0.156	NA	0.005	0.004	0.027	NA

Table 17: (cont') Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). $\boldsymbol{\beta} = (0, 0.5, 0.3, 0, 0)'$. The latent spatial model is SAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 0$	0.017	0.026	0.017	0.024	0.584	0.552	0.584	0.541	0.34	0.304	0.34	0.292
	$\beta_1 = 0.5$	0.617	0.576	0.684	0.569	0.266	0.237	0.197	0.234	0.084	0.062	0.072	0.059
	$\beta_2 = 0.3$	0.338	0.324	0.598	0.315	0.257	0.234	0.15	0.23	0.067	0.055	0.111	0.053
	$\beta_3 = 0$	−0.021	−0.01	0.117	NA	0.234	0.217	0.505	NA	0.055	0.047	0.237	NA
	$\beta_4 = 0$	0.004	−0.006	0.132	NA	0.245	0.215	0.464	NA	0.06	0.046	0.206	NA
20x20	$\beta_0 = 0$	−0.011	−0.007	−0.011	−0.007	0.292	0.282	0.292	0.282	0.085	0.079	0.085	0.079
	$\beta_1 = 0.5$	0.531	0.527	0.513	0.525	0.113	0.097	0.098	0.097	0.014	0.01	0.01	0.01
	$\beta_2 = 0.3$	0.315	0.317	0.334	0.317	0.104	0.09	0.078	0.089	0.011	0.008	0.007	0.008
	$\beta_3 = 0$	0.01	0.002	0.006	NA	0.107	0.096	0.243	NA	0.012	0.009	0.044	NA
	$\beta_4 = 0$	0.001	0.002	−0.027	NA	0.107	0.098	0.234	NA	0.011	0.01	0.046	NA
30x30	$\beta_0 = 0$	−0.02	−0.023	−0.02	−0.023	0.166	0.162	0.166	0.162	0.028	0.027	0.028	0.027
	$\beta_1 = 0.5$	0.513	0.509	0.501	0.508	0.081	0.073	0.075	0.073	0.007	0.005	0.006	0.005
	$\beta_2 = 0.3$	0.304	0.299	0.29	0.299	0.07	0.062	0.06	0.061	0.005	0.004	0.004	0.004
	$\beta_3 = 0$	0.001	0	0.039	NA	0.073	0.064	0.207	NA	0.005	0.004	0.03	NA
	$\beta_4 = 0$	0.003	−0.002	0.022	NA	0.072	0.064	0.163	NA	0.005	0.004	0.022	NA

Table 18: Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). $\boldsymbol{\beta} = (1, 0.5, 0.3, 0, 0)'$. The latent spatial model is CAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 1$	1.061	1.053	1.061	1.031	0.405	0.397	0.405	0.375	0.167	0.16	0.167	0.141
	$\beta_1 = 0.5$	0.535	0.517	0.701	0.501	0.289	0.283	0.216	0.26	0.084	0.08	0.086	0.067
	$\beta_2 = 0.3$	0.316	0.305	0.62	0.304	0.288	0.29	0.237	0.274	0.083	0.084	0.157	0.075
	$\beta_3 = 0$	0.026	0.025	0.031	NA	0.26	0.263	0.584	NA	0.068	0.069	0.313	NA
	$\beta_4 = 0$	-0.005	0.014	0.279	NA	0.275	0.268	0.489	NA	0.075	0.072	0.297	NA
20x20	$\beta_0 = 1$	1.024	1.024	1.024	1.019	0.19	0.191	0.19	0.189	0.037	0.037	0.037	0.036
	$\beta_1 = 0.5$	0.507	0.51	0.505	0.507	0.123	0.12	0.114	0.119	0.015	0.014	0.013	0.014
	$\beta_2 = 0.3$	0.308	0.311	0.372	0.309	0.129	0.128	0.093	0.127	0.016	0.016	0.014	0.016
	$\beta_3 = 0$	-0.006	-0.004	-0.039	NA	0.117	0.114	0.27	NA	0.014	0.013	0.064	NA
	$\beta_4 = 0$	0.008	0.005	0.135	NA	0.112	0.111	0.316	NA	0.013	0.012	0.093	NA
30x30	$\beta_0 = 1$	1.011	1.013	1.011	1.01	0.126	0.127	0.126	0.126	0.016	0.016	0.016	0.016
	$\beta_1 = 0.5$	0.509	0.51	0.503	0.509	0.089	0.087	0.087	0.087	0.008	0.008	0.008	0.008
	$\beta_2 = 0.3$	0.304	0.304	0.306	0.303	0.071	0.07	0.061	0.071	0.005	0.005	0.004	0.005
	$\beta_3 = 0$	0	0	-0.021	NA	0.077	0.076	0.198	NA	0.006	0.006	0.034	NA
	$\beta_4 = 0$	-0.001	-0.001	-0.058	NA	0.083	0.08	0.186	NA	0.007	0.006	0.026	NA

Table 18: (cont') Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). $\boldsymbol{\beta} = (1, 0.5, 0.3, 0, 0)'$. The latent spatial model is SAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 1$	1.166	1.127	1.166	1.104	0.619	0.597	0.619	0.559	0.409	0.37	0.409	0.321
	$\beta_1 = 0.5$	0.611	0.569	0.729	0.555	0.336	0.3	0.291	0.275	0.125	0.094	0.136	0.078
	$\beta_2 = 0.3$	0.329	0.315	0.603	0.315	0.278	0.254	0.197	0.245	0.078	0.064	0.13	0.06
	$\beta_3 = 0$	0.013	0.014	0.313	NA	0.258	0.233	0.409	NA	0.067	0.054	0.247	NA
	$\beta_4 = 0$	−0.002	−0.002	−0.173	NA	0.262	0.241	0.592	NA	0.068	0.058	0.341	NA
20x20	$\beta_0 = 1$	1.055	1.049	1.055	1.044	0.299	0.295	0.299	0.295	0.092	0.089	0.092	0.088
	$\beta_1 = 0.5$	0.52	0.515	0.507	0.514	0.133	0.122	0.119	0.123	0.018	0.015	0.014	0.015
	$\beta_2 = 0.3$	0.317	0.32	0.365	0.319	0.133	0.121	0.091	0.12	0.018	0.015	0.012	0.015
	$\beta_3 = 0$	−0.018	−0.013	−0.077	NA	0.125	0.116	0.242	NA	0.016	0.014	0.059	NA
	$\beta_4 = 0$	0.013	0.006	0.305	NA	0.119	0.104	0.104	NA	0.014	0.011	0.1	NA
30x30	$\beta_0 = 1$	1.004	1.004	1.004	1.002	0.198	0.195	0.198	0.194	0.039	0.038	0.039	0.038
	$\beta_1 = 0.5$	0.512	0.514	0.503	0.513	0.086	0.078	0.08	0.078	0.007	0.006	0.006	0.006
	$\beta_2 = 0.3$	0.304	0.305	0.299	0.305	0.078	0.068	0.06	0.068	0.006	0.005	0.004	0.005
	$\beta_3 = 0$	−0.004	−0.002	−0.173	NA	0.073	0.063	NA	NA	0.005	0.004	0.03	NA
	$\beta_4 = 0$	−0.001	−0.003	0.164	NA	0.08	0.069	NA	NA	0.006	0.005	0.027	NA

Table 19: Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). $\boldsymbol{\beta} = (0.1, 0.7, 0.3, 0, 0)'$. The latent spatial model is CAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 0.1$	0.061	0.061	0.061	0.061	0.312	0.311	0.312	0.303	0.098	0.098	0.098	0.093
	$\beta_1 = 0.7$	0.764	0.757	0.791	0.744	0.258	0.26	0.22	0.256	0.07	0.07	0.056	0.067
	$\beta_2 = 0.3$	0.332	0.33	0.568	0.322	0.248	0.257	0.149	0.245	0.062	0.067	0.094	0.06
	$\beta_3 = 0$	−0.021	−0.025	−0.056	NA	0.232	0.234	0.511	NA	0.054	0.055	0.241	NA
	$\beta_4 = 0$	0.004	0.002	0.078	NA	0.232	0.235	0.473	NA	0.054	0.055	0.207	NA
20x20	$\beta_0 = 0.1$	0.082	0.084	0.082	0.083	0.16	0.159	0.16	0.159	0.026	0.025	0.026	0.026
	$\beta_1 = 0.7$	0.721	0.718	0.706	0.714	0.126	0.131	0.128	0.127	0.016	0.017	0.016	0.016
	$\beta_2 = 0.3$	0.316	0.314	0.363	0.314	0.123	0.124	0.091	0.123	0.015	0.016	0.012	0.015
	$\beta_3 = 0$	−0.009	−0.014	−0.066	NA	0.118	0.117	0.27	NA	0.014	0.014	0.063	NA
	$\beta_4 = 0$	−0.013	−0.013	0.007	NA	0.113	0.11	0.268	NA	0.013	0.012	0.064	NA
30x30	$\beta_0 = 0.1$	0.098	0.098	0.098	0.097	0.114	0.114	0.114	0.114	0.013	0.013	0.013	0.013
	$\beta_1 = 0.7$	0.7	0.698	0.692	0.697	0.082	0.082	0.082	0.082	0.007	0.007	0.007	0.007
	$\beta_2 = 0.3$	0.3	0.3	0.297	0.3	0.069	0.068	0.062	0.068	0.005	0.005	0.004	0.005
	$\beta_3 = 0$	0.001	0.001	−0.001	NA	0.073	0.071	0.285	NA	0.005	0.005	0.04	NA
	$\beta_4 = 0$	−0.002	−0.005	0.056	NA	0.071	0.067	0.207	NA	0.005	0.004	0.032	NA

Table 19: (cont') Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). $\boldsymbol{\beta} = (0.1, 0.7, 0.3, 0, 0)'$. The latent spatial model is SAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 0.1$	0.074	0.077	0.074	0.075	0.49	0.472	0.49	0.465	0.239	0.222	0.239	0.216
	$\beta_1 = 0.7$	0.82	0.798	0.812	0.783	0.267	0.248	0.218	0.244	0.085	0.071	0.06	0.066
	$\beta_2 = 0.3$	0.372	0.353	0.565	0.347	0.256	0.228	0.149	0.224	0.07	0.054	0.092	0.052
	$\beta_3 = 0$	−0.015	−0.018	−0.125	NA	0.253	0.243	0.526	NA	0.064	0.059	0.273	NA
	$\beta_4 = 0$	−0.015	−0.017	0.02	NA	0.231	0.215	0.56	NA	0.053	0.046	0.275	NA
20x20	$\beta_0 = 0.1$	0.08	0.085	0.08	0.082	0.272	0.266	0.272	0.264	0.074	0.071	0.074	0.07
	$\beta_1 = 0.7$	0.739	0.735	0.716	0.732	0.126	0.128	0.127	0.126	0.017	0.018	0.016	0.017
	$\beta_2 = 0.3$	0.311	0.313	0.344	0.313	0.116	0.11	0.076	0.109	0.013	0.012	0.008	0.012
	$\beta_3 = 0$	−0.003	−0.006	0.031	NA	0.106	0.103	0.273	NA	0.011	0.011	0.057	NA
	$\beta_4 = 0$	−0.003	−0.009	−0.173	NA	0.108	0.095	0.225	NA	0.012	0.009	0.071	NA
30x30	$\beta_0 = 0.1$	0.093	0.092	0.093	0.092	0.177	0.176	0.177	0.176	0.031	0.031	0.031	0.031
	$\beta_1 = 0.7$	0.705	0.7	0.692	0.7	0.085	0.08	0.078	0.079	0.007	0.006	0.006	0.006
	$\beta_2 = 0.3$	0.301	0.3	0.294	0.299	0.074	0.068	0.061	0.068	0.005	0.005	0.004	0.005
	$\beta_3 = 0$	0	−0.003	−0.051	NA	0.074	0.062	0.169	NA	0.005	0.004	0.022	NA
	$\beta_4 = 0$	0	−0.003	0.14	NA	0.069	0.06	NA	NA	0.005	0.004	0.02	NA

Table 20: Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). $\boldsymbol{\beta} = (0.1, 1, 0.3, 0, 0)'$. The latent spatial model is CAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 0.1$	0.084	0.082	0.084	0.082	0.325	0.326	0.325	0.317	0.105	0.106	0.105	0.101
	$\beta_1 = 1$	1.097	1.088	1.045	1.064	0.318	0.323	0.313	0.31	0.11	0.112	0.099	0.099
	$\beta_2 = 0.3$	0.332	0.33	0.596	0.324	0.283	0.281	0.235	0.269	0.081	0.079	0.142	0.073
	$\beta_3 = 0$	−0.004	−0.003	0.041	NA	0.231	0.231	0.562	NA	0.053	0.053	0.273	NA
	$\beta_4 = 0$	0.002	−0.003	0.053	NA	0.247	0.249	0.495	NA	0.061	0.061	0.226	NA
20x20	$\beta_0 = 0.1$	0.076	0.076	0.076	0.075	0.169	0.171	0.169	0.17	0.029	0.03	0.029	0.029
	$\beta_1 = 1$	1.03	1.029	1.015	1.024	0.143	0.149	0.149	0.146	0.021	0.023	0.022	0.022
	$\beta_2 = 0.3$	0.312	0.312	0.369	0.312	0.125	0.122	0.081	0.121	0.016	0.015	0.011	0.015
	$\beta_3 = 0$	−0.005	−0.009	−0.035	NA	0.121	0.119	0.277	NA	0.015	0.014	0.07	NA
	$\beta_4 = 0$	−0.014	−0.014	−0.162	NA	0.115	0.113	0.231	NA	0.013	0.013	0.069	NA
30x30	$\beta_0 = 0.1$	0.101	0.101	0.101	0.1	0.117	0.118	0.117	0.118	0.014	0.014	0.014	0.014
	$\beta_1 = 1$	1.001	0.999	0.993	0.998	0.087	0.087	0.086	0.086	0.008	0.007	0.007	0.007
	$\beta_2 = 0.3$	0.302	0.303	0.303	0.302	0.074	0.073	0.065	0.072	0.005	0.005	0.004	0.005
	$\beta_3 = 0$	−0.001	−0.001	0.252	NA	0.074	0.073	NA	NA	0.005	0.005	0.064	NA
	$\beta_4 = 0$	−0.001	−0.003	0.005	NA	0.074	0.071	0.2	NA	0.005	0.005	0.03	NA

Table 20: (cont') Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). $\boldsymbol{\beta} = (0.1, 1, 0.3, 0, 0)'$. The latent spatial model is SAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 0.1$	0.1	0.103	0.1	0.103	0.506	0.488	0.506	0.482	0.255	0.237	0.255	0.231
	$\beta_1 = 1$	1.183	1.147	1.108	1.126	0.309	0.301	0.296	0.294	0.128	0.112	0.099	0.102
	$\beta_2 = 0.3$	0.366	0.349	0.56	0.345	0.25	0.236	0.153	0.23	0.067	0.058	0.091	0.054
	$\beta_3 = 0$	−0.006	−0.006	−0.006	NA	0.245	0.238	0.474	NA	0.06	0.056	0.202	NA
	$\beta_4 = 0$	0.014	−0.002	−0.069	NA	0.248	0.228	0.49	NA	0.061	0.052	0.223	NA
20x20	$\beta_0 = 0.1$	0.086	0.096	0.086	0.094	0.264	0.259	0.264	0.259	0.069	0.067	0.069	0.067
	$\beta_1 = 1$	1.048	1.048	1.027	1.044	0.143	0.144	0.145	0.142	0.023	0.023	0.022	0.022
	$\beta_2 = 0.3$	0.314	0.311	0.355	0.312	0.125	0.114	0.075	0.113	0.016	0.013	0.009	0.013
	$\beta_3 = 0$	−0.003	−0.009	0.022	NA	0.11	0.105	0.233	NA	0.012	0.011	0.047	NA
	$\beta_4 = 0$	0.007	0.002	−0.052	NA	0.11	0.101	0.231	NA	0.012	0.01	0.049	NA
30x30	$\beta_0 = 0.1$	0.092	0.089	0.092	0.088	0.184	0.185	0.184	0.185	0.034	0.034	0.034	0.034
	$\beta_1 = 1$	1.016	1.014	1.004	1.012	0.097	0.092	0.091	0.091	0.01	0.009	0.008	0.008
	$\beta_2 = 0.3$	0.3	0.298	0.291	0.298	0.074	0.069	0.062	0.069	0.005	0.005	0.004	0.005
	$\beta_3 = 0$	0.004	0	−0.016	NA	0.076	0.065	0.226	NA	0.006	0.004	0.026	NA
	$\beta_4 = 0$	−0.001	−0.005	NA	NA	0.073	0.063	NA	NA	0.005	0.004	NA	NA

Table 21: Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). $\boldsymbol{\beta} = (0.1, 1.5, 0.3, 0, 0)'$. The latent spatial model is CAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 0.1$	0.112	0.114	0.112	0.115	0.343	0.345	0.343	0.337	0.117	0.119	0.117	0.113
	$\beta_1 = 1.5$	1.643	1.613	1.554	1.576	0.399	0.404	0.387	0.384	0.179	0.175	0.152	0.152
	$\beta_2 = 0.3$	0.345	0.342	0.636	0.332	0.271	0.286	0.155	0.274	0.075	0.083	0.137	0.076
	$\beta_3 = 0$	−0.006	−0.015	−0.13	NA	0.256	0.259	0.592	NA	0.065	0.067	0.309	NA
	$\beta_4 = 0$	0.016	0.013	0.188	NA	0.267	0.27	0.578	NA	0.071	0.073	0.344	NA
20x20	$\beta_0 = 0.1$	0.079	0.079	0.079	0.079	0.183	0.181	0.183	0.181	0.034	0.033	0.034	0.033
	$\beta_1 = 1.5$	1.55	1.55	1.531	1.541	0.18	0.182	0.179	0.178	0.035	0.035	0.033	0.033
	$\beta_2 = 0.3$	0.315	0.316	0.386	0.317	0.131	0.127	0.081	0.126	0.017	0.016	0.014	0.016
	$\beta_3 = 0$	−0.008	−0.011	0.05	NA	0.13	0.131	0.326	NA	0.017	0.017	0.096	NA
	$\beta_4 = 0$	−0.011	−0.013	−0.147	NA	0.129	0.125	0.29	NA	0.017	0.016	0.095	NA
30x30	$\beta_0 = 0.1$	0.094	0.096	0.094	0.095	0.121	0.121	0.121	0.12	0.015	0.014	0.015	0.014
	$\beta_1 = 1.5$	1.509	1.507	1.498	1.503	0.117	0.116	0.116	0.115	0.014	0.013	0.013	0.013
	$\beta_2 = 0.3$	0.302	0.302	0.306	0.301	0.078	0.076	0.065	0.075	0.006	0.006	0.004	0.006
	$\beta_3 = 0$	−0.004	−0.003	−0.201	NA	0.081	0.077	0.027	NA	0.007	0.006	0.041	NA
	$\beta_4 = 0$	−0.003	−0.005	0.052	NA	0.08	0.078	0.189	NA	0.006	0.006	0.032	NA

Table 21: (cont') Mean, standard deviation and mean squared error of the estimates of regression coefficients by using maximum likelihood estimation (MLE), quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). $\boldsymbol{\beta} = (0.1, 1.5, 0.3, 0, 0)'$. The latent spatial model is SAR(1).

		mean				SD				MSE			
		GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA	GLM	QEE	PQEE	ORA
10x10	$\beta_0 = 0.1$	0.11	0.12	0.11	0.121	0.526	0.505	0.526	0.495	0.275	0.255	0.275	0.244
	$\beta_1 = 1.5$	1.764	1.711	1.644	1.675	0.411	0.404	0.4	0.397	0.238	0.207	0.18	0.188
	$\beta_2 = 0.3$	0.393	0.373	0.602	0.369	0.265	0.248	0.146	0.243	0.078	0.067	0.112	0.064
	$\beta_3 = 0$	0.026	0.015	0.095	NA	0.264	0.249	0.522	NA	0.07	0.062	0.255	NA
	$\beta_4 = 0$	0.004	−0.005	0.273	NA	0.271	0.254	0.562	NA	0.073	0.064	0.351	NA
20x20	$\beta_0 = 0.1$	0.082	0.086	0.082	0.084	0.284	0.287	0.284	0.287	0.081	0.082	0.081	0.082
	$\beta_1 = 1.5$	1.569	1.564	1.534	1.558	0.19	0.193	0.194	0.19	0.041	0.041	0.039	0.039
	$\beta_2 = 0.3$	0.308	0.308	0.366	0.308	0.129	0.118	0.075	0.117	0.017	0.014	0.01	0.014
	$\beta_3 = 0$	−0.004	−0.007	−0.05	NA	0.123	0.112	0.274	NA	0.015	0.012	0.059	NA
	$\beta_4 = 0$	0.002	0.001	−0.078	NA	0.122	0.11	0.308	NA	0.015	0.012	0.082	NA
30x30	$\beta_0 = 0.1$	0.095	0.094	0.095	0.094	0.191	0.183	0.191	0.183	0.036	0.034	0.036	0.033
	$\beta_1 = 1.5$	1.529	1.528	1.516	1.525	0.125	0.118	0.119	0.117	0.017	0.015	0.014	0.014
	$\beta_2 = 0.3$	0.3	0.3	0.294	0.299	0.079	0.072	0.066	0.071	0.006	0.005	0.004	0.005
	$\beta_3 = 0$	0.001	−0.003	−0.202	NA	0.086	0.074	0.02	NA	0.007	0.005	0.041	NA
	$\beta_4 = 0$	−0.005	−0.008	NA	NA	0.08	0.073	NA	NA	0.006	0.005	NA	NA

Table 22: Number of times a coefficient is identified to be zero out of 200 simulations for CAR and SAR models from different the set-ups.

		(−0.5, 0.5, 0.3)		(0, 0.5, 0, 3)		(1, 0.5, 0.3)		(0.1, 0.7, 0.3)		(0.1, 1, 0.3)		(0.1, 1.5, 0.3)	
		CAR	SAR	CAR	SAR	CAR	SAR	CAR	SAR	CAR	SAR	CAR	SAR
10x10	β_0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	β_1	108	86	97	75	112	103	37	29	1	2	0	0
	β_2	155	151	149	154	154	157	140	131	137	134	143	134
	β_3	192	191	188	192	188	191	189	186	193	190	194	190
	β_4	191	190	191	192	188	191	190	192	189	189	187	192
20x20	β_0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	β_1	2	3	0	2	8	8	0	0	0	0	0	0
	β_2	62	55	51	49	73	67	60	63	66	67	79	84
	β_3	193	194	194	196	193	190	195	196	191	193	192	196
	β_4	189	197	192	194	196	197	190	195	195	192	192	195
30x30	β_0	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
	β_1	0	0	0	0	0	0	0	0	0	0	0	0
	β_2	12	11	8	3	15	15	6	10	9	10	18	16
	β_3	194	198	196	197	193	199	198	197	199	198	195	197
	β_4	195	195	196	195	197	199	197	199	196	200	194	200

Table 22: (cont') The rates of different final models selected by by penalized quasi-likelihood estimating equation (PQEE). The five digits show whether the five estimated coefficients in selected models are zero or not, with 1 denoting non-zero and 0 denoting zero. In bold are the true model '11100' and the corresponding selection rate.

		CAR	SAR
(−0. 5, 0. 5, 0. 3)	10x10	10000 10001 10010 10100 10101 10110 11000 11001 11010 11011 0. 460 0. 005 0. 005 0. 055 0. 010 0. 005 0. 260 0. 020 0. 020 0. 005 **11100** 11101 11110 **0. 145** 0. 005 0. 005	10000 10100 10101 10110 11000 11010 11011 **11100** 11101 0. 370 0. 045 0. 010 0. 005 0. 365 0. 015 0. 005 **0. 135** 0. 030 11110 11111 0. 015 0. 005
	20x20	10000 10100 11000 11001 11010 **11100** 11101 11110 0. 005 0. 005 0. 290 0. 005 0. 010 **0. 610** 0. 050 0. 025	10000 11000 11001 **11100** 11101 11110 0. 015 0. 255 0. 005 **0. 685** 0. 010 0. 030
	30x30	11000 **11100** 11101 11110 0. 060 **0. 885** 0. 025 0. 030	11000 **11100** 11101 11110 0. 055 **0. 910** 0. 025 0. 010
(0, 0. 5, 0. 3)	10x10	10000 10001 10010 10100 11000 11001 11010 **11100** 11101 11110 0. 435 0. 005 0. 005 0. 040 0. 260 0. 015 0. 025 **0. 160** 0. 025 0. 030	10000 10100 10110 11000 11001 11010 11011 **11100** 11101 0. 355 0. 015 0. 005 0. 385 0. 015 0. 010 0. 005 **0. 175** 0. 015 11110 11111 0. 015 0. 005
	20x20	11000 11010 **11100** 11101 11110 0. 250 0. 005 **0. 680** 0. 040 0. 025	10000 10100 11000 11001 **11100** 11101 11110 0. 005 0. 005 0. 235 0. 005 **0. 705** 0. 025 0. 020
	30x30	11000 **11100** 11101 11110 0. 04 **0. 92** 0. 02 0. 02	11000 **11100** 11101 11110 0. 015 **0. 945** 0. 025 0. 015
(1, 0. 5, −0. 3)	10x10	10000 10100 10101 10110 11000 11001 11010 **11100** 11101 11110 0. 500 0. 050 0. 005 0. 005 0. 215 0. 025 0. 030 **0. 120** 0. 025 0. 020 11111 0. 005	10000 10011 10100 10101 10110 11000 11001 11010 **11100** 0. 455 0. 005 0. 035 0. 005 0. 015 0. 295 0. 020 0. 010 **0. 130** 11101 11110 0. 015 0. 015
	20x20	10000 11000 11001 11010 **11100** 11101 11110 0. 040 0. 315 0. 005 0. 005 **0. 590** 0. 015 0. 030	10000 10100 11000 11010 **11100** 11101 11110 0. 035 0. 005 0. 295 0. 005 **0. 600** 0. 015 0. 045
	30x30	11000 **11100** 11101 11110 0. 075 **0. 875** 0. 015 0. 035	11000 **11100** 11101 11110 0. 075 **0. 915** 0. 005 0. 005

Table 22: (cont') The rates of different final models selected by by penalized quasi-likelihood estimating equation (PQEE). The five digits show whether the five estimated coefficients in selected models are zero or not, with 1 denoting non-zero and 0 denoting zero. In bold are the true model '11100' and the corresponding selection rate.

		CAR	SAR
(0.1, 0.7, 0.3)	10x10	10000 10010 10100 11000 11001 11010 11011 **11100** 11101 11110 0.170 0.005 0.010 0.480 0.025 0.015 0.005 **0.240** 0.020 0.030	10000 10100 11000 11001 11010 **11100** 11101 11110 11111 0.130 0.015 0.465 0.025 0.035 **0.285** 0.010 0.030 0.005
	20x20	11000 **11100** 11101 11110 0.300 **0.625** 0.050 0.025	11000 11001 **11100** 11101 11110 0.305 0.010 **0.650** 0.015 0.020
	30x30	11000 **11100** 11101 11110 0.030 **0.945** 0.015 0.010	11000 **11100** 11101 11110 0.050 **0.930** 0.005 0.015
(0.1, 1, 0.3)	10x10	10000 11000 11001 11010 11011 **11100** 11101 11110 0.005 0.640 0.025 0.010 0.005 **0.270** 0.025 0.020	10000 11000 11001 11010 **11100** 11101 11110 0.010 0.615 0.025 0.020 **0.270** 0.030 0.030
	20x20	11000 11001 11010 **11100** 11101 11110 11111 0.320 0.005 0.005 **0.615** 0.015 0.035 0.005	11000 11001 **11100** 11101 11110 11111 0.320 0.015 **0.610** 0.020 0.030 0.005
	30x30	11000 **11100** 11101 11110 0.045 **0.930** 0.020 0.005	11000 **11100** 11110 0.05 **0.94** 0.01
(0.1, 1.5, 0.3)	10x10	11000 11001 11010 11011 **11100** 11101 11110 11111 0.675 0.025 0.010 0.005 **0.240** 0.030 0.010 0.005	11000 11001 11010 **11100** 11101 11110 0.64 0.02 0.01 **0.27** 0.02 0.04
	20x20	11000 11001 **11100** 11101 11110 0.385 0.010 **0.535** 0.030 0.040	11000 11001 11010 **11100** 11101 11110 0.405 0.005 0.010 **0.550** 0.020 0.010
	30x30	11000 11010 **11100** 11101 11110 11111 0.085 0.005 **0.865** 0.025 0.015 0.005	11000 **11100** 11110 0.080 **0.905** 0.015

Table 23: Mean of the estimates of correlation parameters, $(\rho_1, \rho_2, \rho_3, \rho_4)'$, by using quasi-likelihood estimating equation (QEE), penalized QEE (PQEE) and oracle (ORA). (Only those with grid size 30×30 are shown.)

(−0.5, 0.5, 0.3)								(0, 0.5, 0.3)								(1, 0.5, 0.3)							
CAR				SAR				CAR				SAR				CAR				SAR			
	QEE	PQEE	ORA		QEE	PQEE	ORA		QEE	PQEE	ORA		QEE	PQEE	ORA		QEE	PQEE	ORA		QEE	PQEE	ORA
ρ_1	0.156	0.155	0.156	ρ_1	0.329	0.327	0.329	ρ_1	0.159	0.158	0.159	ρ_1	0.334	0.331	0.333	ρ_1	0.143	0.142	0.142	ρ_1	0.315	0.312	0.314
ρ_2	0.073	0.072	0.073	ρ_2	0.213	0.211	0.213	ρ_2	0.073	0.073	0.073	ρ_2	0.214	0.213	0.214	ρ_2	0.065	0.065	0.065	ρ_2	0.197	0.196	0.197
ρ_3	0.043	0.043	0.043	ρ_3	0.138	0.136	0.137	ρ_3	0.046	0.046	0.046	ρ_3	0.144	0.142	0.143	ρ_3	0.041	0.041	0.041	ρ_3	0.131	0.130	0.131
ρ_4	0.027	0.027	0.027	ρ_4	0.102	0.101	0.102	ρ_4	0.025	0.025	0.026	ρ_4	0.103	0.102	0.103	ρ_4	0.023	0.023	0.023	ρ_4	0.093	0.092	0.093

(0.1, 0.7, 0.3)								(0.1, 1, 0.3)								(0.1, 1.5, 0.3)							
CAR				SAR				CAR				SAR				CAR				SAR			
	QEE	PQEE	ORA		QEE	PQEE	ORA		QEE	PQEE	ORA		QEE	PQEE	ORA		QEE	PQEE	ORA		QEE	PQEE	ORA
ρ_1	0.153	0.152	0.153	ρ_1	0.320	0.317	0.319	ρ_1	0.142	0.141	0.141	ρ_1	0.298	0.295	0.298	ρ_1	0.125	0.124	0.125	ρ_1	0.259	0.256	0.259
ρ_2	0.072	0.071	0.072	ρ_2	0.207	0.205	0.206	ρ_2	0.067	0.067	0.067	ρ_2	0.194	0.191	0.194	ρ_2	0.057	0.057	0.058	ρ_2	0.172	0.169	0.172
ρ_3	0.044	0.044	0.044	ρ_3	0.137	0.136	0.137	ρ_3	0.040	0.040	0.040	ρ_3	0.130	0.128	0.130	ρ_3	0.035	0.035	0.035	ρ_3	0.115	0.113	0.115
ρ_4	0.027	0.027	0.027	ρ_4	0.102	0.101	0.102	ρ_4	0.025	0.025	0.025	ρ_4	0.097	0.096	0.097	ρ_4	0.022	0.022	0.022	ρ_4	0.087	0.085	0.087

From the tables above, we observe that our QEE and PQEE methods are reliable over a range of parameter values. For different set-ups, both QEE and PQEE have desirable asymptotic properties, that is, their estimates tend to the true values and the MSEs become smaller as the sample size increases. QEE provides better estimates than MLE when the spatial correlation is stronger, that is, the bias and the SDs are smaller. PQEE also attains the oracle property, that is, the results via PQEE tend to the results from QEE under the true model.

4.5 Data Example

In this section, we consider the land cover data again. As mentioned before, an interdisciplinary team of ecologists and environmental historians are assessing the influence of past land ownership characteristics on land-cover structure in northern Wisconsin, USA. They derived data from historical plat maps, which indicate all the responses and covariates on a lattice. The unit of analysis is the quarter-section (1/36 township, 160 acres $\approx$ 65 ha), and there are 1429 units in the study area. For illustration, we set the response variable to be binary indicating whether the land cover at each site is aspen-paper birch forest (APB) or not (Figure 13). The covariates considered here are: Res, PolyNm, PolyPr, MxPolyPr, TotOwn and AvParcel. Res is whether this site is a reservation or not. PolyNm is the number of parcels in a given quarter-section, which is a measure of parcel density. PolyPr is the the average size of parcel polygon, which is a measure of parcel size. MxPolyPr reports the largest parcel polygon, as a proportion of quarter-section. TotOwn shows the total property area (measured in acres) in this study area associated with owner of the largest parcel polygon in a given quarter-section, and

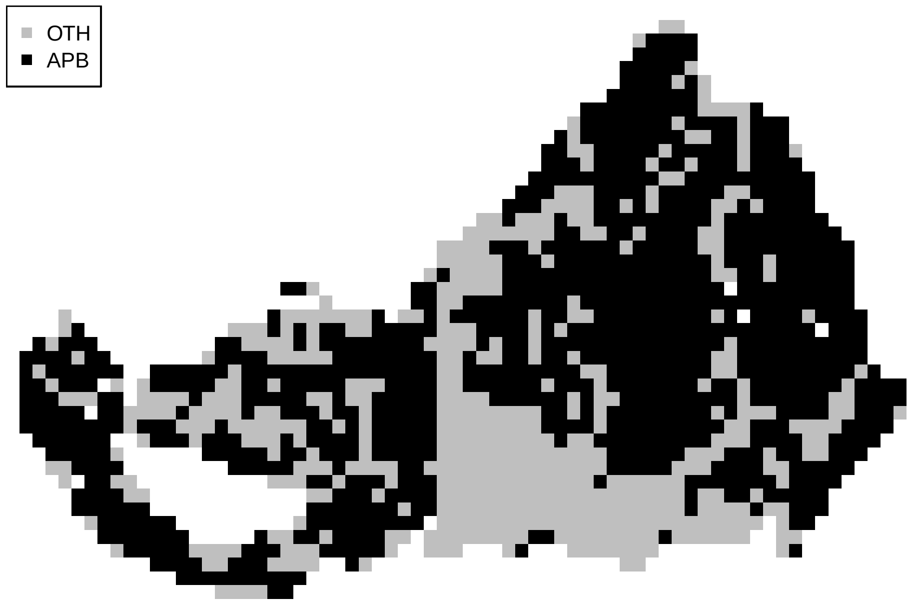

Figure 13: Map of the response variable land cover type: aspen-paper birch (black) or not (grey).

because the values of this covariate are skewed to the right, we use a log transformation. AvParcel is the average size of all parcels that lay in whole or in part within a given quarter-section (measured in acres). All the covariates except Res are standardized to have mean 0 and variance 1 before being used in the regression procedure.

We use three estimation methods, MLE, QEE and PQEE with the results from QEE as initial values, to analyze the land cover data. When implementing QEE and PQEE, five different neighborhood structures are considered: first-order neighbors, first-and second-order neighbors, and so on, up to all first five orders. We use BIC for model comparison, which is calculated under the assumption of independent responses, BIC $= \log(n) \times$ (number of parameters) $- 2 \times$ loglikehood. Since we want to compare the models with different neighbor structures, the number of regression coefficients and the number of correlation parameter estimators are both accounted for in the BIC. The best model is the one with the smallest BIC value. The estimation results are given in Table 24.

Without considering spatial dependent structure, the MLE gives quite good result. After using backward elimination, we obtain a logistic regression model with the smaller BIC as 1636.6. Polynum is selected out of the model, maybe because of its high correlation with PolyPr. When applying QEE and PQEE methods, the estimation assuming a first-order neighborhood structure did not give a converged result and the best models under these two methods both choose a second-order of neighborhood. In PQEE, PolyNm, PolyPr, and AvParcel are selected out and the covariates selected, Res and log(TotOwn), have positive relationship with the probability of APB. The model selected by PQEE is more parsimonious than selected by backward elimination under MLE. It appears that Res and log(TotOwn) have strong positive relationship with the probability of APB, under all three methods. The spatial correlations among quarter-sections are quite strong according to the results from the quasi-likelihood methods (Table 25) and thus, the analysis considering the dependent structure is perhaps more appropriate.

Table 24: Regression coefficient estimates and their standard errors in parentheses via maximum likelihood estimation (MLE) under the full model and the model selected by backward elimination, quasi-likelihood estimating equation (QEE) and penalized QEE (PQEE) under different orders of neighborhood 2–5 in the analysis of the land cover data. BIC values are also given.

	Model	Intercept	Res	PolyNm	PolyPr	MxPolyPr	log(TotOwn)	AvParcel	BIC
MLE	Full	−0.125 (0.077)	2.161 (0.154)	0.027 (0.149)	−0.27 (0.226)	0.373 (0.195)	0.466 (0.105)	0.254 (0.127)	1643.8
	Backward elimination	−0.126 (0.077)	2.162 (0.154)	0	−0.29 (0.197)	0.367 (0.192)	0.465 (0.105)	0.258 (0.125)	1636.6
QEE	Neighborhood	Intercept	Res	PolyNm	PolyPr	MxPolyPr	log(TotOwn)	AvParcel	BIC
	2 orders	0.017 (0.118)	1.639 (0.213)	0.021 (0.119)	−0.137 (0.181)	0.193 (0.155)	0.415 (0.094)	0.127 (0.122)	1673.3
	3 orders	−0.017 (0.127)	1.696 (0.234)	0.041 (0.12)	−0.046 (0.181)	0.089 (0.154)	0.352 (0.09)	0.123 (0.114)	1686.0
	4 orders	−0.008 (0.14)	1.666 (0.248)	0.027 (0.123)	−0.213 (0.184)	0.34 (0.158)	0.366 (0.094)	0.075 (0.114)	1690.3
	5 orders	−0.034 (0.149)	1.725 (0.264)	−0.04 (0.124)	−0.289 (0.187)	0.307 (0.159)	0.389 (0.096)	0.142 (0.118)	1692.6
PQEE	Neighborhood	Intercept	Res	PolyNm	PolyPr	MxPolyPr	log(TotOwn)	AvParcel	BIC
	2 orders	0.017 (0.118)	1.586 (0.211)	0	0	0	0.521 (0.095)	0	1653.5
	3 orders	−0.017 (0.127)	1.694 (0.234)	0	0	0	0.364 (0.09)	0.115 (0.113)	1665.6
	4 orders	−0.008 (0.14)	1.663 (0.249)	0	0	0.146 (0.157)	0.4 (0.094)	0	1671.4
	5 orders	−0.034 (0.149)	1.719 (0.265)	0	0	0.128 (0.158)	0.445 (0.096)	0	1675.2

Table 25: Estimates of the correlation parameters $\rho_1, \ldots, \rho_5$ via quasi-likelihood estimating equation (QEE) and penalized QEE (PQEE) under the orders of neighborhood 2–5.

	QEE					PQEE				
Order of neighborhood	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5	ρ_1	ρ_2	ρ_3	ρ_4	ρ_5
2	0.317	0.214	0	0	0	0.32	0.219	0	0	0
3	0.322	0.217	0.138	0	0	0.323	0.219	0.139	0	0
4	0.318	0.215	0.137	0.134	0	0.32	0.218	0.138	0.135	0
5	0.316	0.213	0.133	0.131	0.089	0.318	0.217	0.136	0.133	0.09

Chapter 5

Summary and Future work

In the first part of the thesis, I have demonstrated via the northern Wisconsin land cover data example that our GLMM methodology is useful for quantifying relationships between a nominal categorical response variable and covariates via regression. The spatial multinomial models have outperformed the independent models in terms of the DIC values and identification of important covariates. The Bayesian hierarchical model has provided a rigorous, yet flexible, framework for statistical inference, in which it is straightforward to obtain the posterior samples of the regression coefficients for any desirable choice of baseline category.

For future research, it would be of interest to work on spatial-temporal models and inference for lattice data that are repeatedly sampled over time. One possible approach is to consider the temporal correlation as another type of spatial correlation, and use a more complex spatial-temporal neighborhood structure. This will result in a larger variance-covariance matrix and correspondingly, more computational challenge when devising and implementing MCMC algorithms. As mentioned in Section 3.4, we encountered numerical instability when using an alternative link function (2.21) to fit the model to the land cover data. Further investigation will be needed to understand the cause of the computational difficulty and to develop counter measures.

In the second part of this thesis, a quasi-likelihood estimating equation (QEE) approach for spatial correlated binary data has performed quite well in both simulations and a real data example. The variable selection via penalized QEE has also given reasonable results. In the future, it would be interesting to extend the method in two directions. One direction is to consider categorical data with more than two categories and the other is to consider spatial-temporal data. A more complicated correlation matrix will need to be specified. In addition, the non-singularity conditions will need to be studied and handled with greater care. Finally for simulation, new approaches will be devised to generate spatially, or spatial-temporally, correlated categorical data.

Bibliography

Aitchison, J. (1987) *The Statistical Analysis of Compositional Data*. London: Chapman and Hall.

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

Brown, D.G. (2003) Land use and forest cover on private parcels in the Upper Midwest, USA, 1970 to 1990. *Landscape Ecology*, **18**, 777–790.

Cleland, D.T., Avers, P.E., McNab, W.H., Jensen, M.E., Bailey, R.G., King, T., and Russell, W.E. (1997) National hierarchical framework of ecological units. In *Ecosystem Management Applications for Sustainable Forest and Wildlife Resources, Eds. Boyce, M. S. and Haney, A.* Yale University Press, New Haven, CT. pp. 181–200.

Cressie, N. (1993) *Statistics for Spatial Data, Rev. Ed.* Wiley, New York.

Crow, T.R., Host, G.E., and Mladenoff, D.J. (1999) Ownership and ecosystem as sources of spatial heterogeneity in a forested landscape, Wisconsin, USA. *Landscape Ecology*, **14**, 449–463.

Delcourt, H.R. and Delcourt, P.A. (1996) Presettlement landscape heterogeneity: Evaluating grain of resolution using General Land Office Survey data. *Landscape Ecology*, **11**, 363–381.

Diggle, P. J., Tawn, J. A., and Moyeed, R. A. (1998) Model-based geostatistics (with discussion). *Applied Statistics*, **47**, 299–350.

Dziak, J. J., (2006) *Penalized quadratic inference functions for variable selection in longitudinal research*. PhD thesis, The Pennsylvania State University.

Dziak, J. J. and Li, R. (2006) An overview on variable selection for longitudinal data. *Quantitative Medical Data Analysis*. World Sciences Publisher, Singapore.

Dupouey, J. L., Dambrine, E., Laffite, J. D., and Moares, D. (2002) Irreversible impact of past land use on forest soils and biodiversity. *Ecology*, **83**, 2978–2984.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004) Least angle regression. *The Annals of Statistics*, **32**, 407–499.

ESRI (2002) *ArcView 3.3*. Redlands, California.

ESRI (2006) *ArcView 9.2*. Redlands, California.

Fan, J. and Li, R. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, **796**, 1348–1360.

Foster, D. R. and Aber, J.D., Editors (2004) *Forests in Time: The Environmental Consequences of 1,000 Years of Change in New England*. Yale University Press, New Haven.

Foster, D. R., Donahue, B., Kittredge, D. B., Motzkin, G., Hall, B., Turner, B. L., and Chilton, E. (2008) New England's forest landscape: Ecological legacies and conservation patterns shaped by agrarian history. In *Agrarian Landscapes in Transition, Eds. Redman, C.L. and Foster, D.R.* Oxford University Press, New York.

Foster, D. R., Zebryk, T., Schoonmaker P., and Lezberg, A. (1992) Post-settlement history of human land-use and vegetation dynamics of a *Tsuga Canadensis* (Hemlock) woodlot in central New England. *Journal of Ecology*, **80**, 773–786.

Fraterrigo J.M., Turner M.G., Pearson S.M., and Dixon P. (2005) Effects of past land use on spatial heterogeneity of soil nutrients in southern Appalachian forests. *Ecological Monographs*, **75**, 215–230.

Heasley, L. (2003) Shifting boundaries on a Wisconsin landscape: Can GIS help historians tell a complicated story? *Human Ecology*, **31**, 183–211.

Higgs, M. D. and Hoeting, J. A. (2010) Clipped latent-variable model for spatially correlated ordered categorical data. *Computational Statistics and Data Analysis*, **54**, 1999–2011.

Hubbard, B.T. (2009) *American Boundaries: The Nation, the States, the Rectangular Survey*. University of Chicago, Chicago.

Huffer, F. W. and Wu, H. L. (1998) Markov chain and Monte Carlo for auto-logistic regression models with application to the distribution of plant species. *Biometrics*, **54**, 509–524.

Knowles, A.K., Editor (2002) *Past Time, Past Place: GIS for History*. ESRI Press, Redlands.

Koch, J. (2006) Touching every forty: John Bordner and the Wisconsin Land Economic Inventory. *Wisconsin Magazine of History*, **89**, 14–25.

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

Li, X. (2006) *Bayesian Analysis of Cross-Classified Spatial Data with Autocorrelation.* PhD thesis, University of Wisconsin, Madison.

Lipsitz, S. R., Kim, K., and Zhao, L. (1994) Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine*, **13**, 1149–1163.

Lin, P. (2008) Estimating equations for spatially correlated data in multi-dimensional space. *Biometrika*, **95**, 847–858.

Lin, P. (2010) Estimating equations for separable spatial-temporal binary data. *Environmental and Ecological Statistics*, **17**, 543–557.

Lin, P. (2011) Quasi-deviance functions for spatially correlated data. *Statistica Sinica*, **21**, 1785–1806.

Mardia, K. V. (1988) Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis*, **24**, 265–284.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models, 2nd Ed.* Boca Raton: Chapman and Hall.

Medley, K.E., Pobocik, C.M., and Okey, B.W. (2003) Historical changes in forest cover and land ownership in a Midwestern U.S. landscape. *Annals of the Association of American Geographers*, **93**, 104–120.

Paciorek, C.J. and McLachlan, J.S. (2009) Mapping ancient forests: Bayesian inference for spatial-temporal trends in forest composition using the fossil pollen proxy record. *Journal of the American Statistical Association*, **104**, 608–621.

Prentice, R. L. (1988) Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**, 1033–1048.

Rhemtulla, J. M., Mladenoff, D. J., and Clayton, M. K. (2007) Regional land cover conversion in the U.S. upper Midwest: magnitude of change and limited recovery (1850-1930-1993). *Landscape Ecology*, **22**, 57–75.

Sain, S. R. and Cressie, N. (2007) A spatial model for multivariate lattice data. *Journal of Econometrics*, **140**, 226–259.

Sain, S. R., Jagtap, S., Mearns, L., and Nychka, D. (2006) A multivariate spatial model for soil water profiles. *Journal of Agricultural, Biological, and Environmental Studies*, **11**, 462–480.

Schulte, L.A. and Mladenoff, D.J. (2001) The original U.S. Public Land Survey records: Their use and limitations in reconstruction pre-European settlement vegetation. *Journal of Forestry*, **99**, 5–10.

Schulte, L. A. and Mladenoff, D. J. (2005) Severe wind and fire regimes in northern forests: Historical variability at the regional scale. *Ecology*, **86**, 431–445.

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

Spies, T.A., Johnson, K.N., Burnett, K.M., Ohmann, J.L., McComb, B.C., Reeves, G.H., Bettinger, P., Kline, J.D., and Garber-Yonts, B. (2007) Cumulative ecological and socioeconomic effects of forest policies in coastal Oregon. *Ecological Applications*, **17**, 5–17.

Stanfield,, B.J., Bliss, J.C., and Spies, T.A. (2002) Land ownership and landscape structure: A spatial analysis of sixty-six Oregon (USA) Coast Range watersheds. *Landscape Ecology*, **17**, 685–697.

Steen-Adams, M.M., Mladenoff, D.J., and Langston, N.E. (2007) White pine in the northern forests: An ecological and management history of white pine on the Bad River Reservation of Wisconsin. *Environmental History*, **12**, 614–648

Steen-Adams, M.M., Mladenoff, D.J., Langston, N.E., Liu, F., and Zhu, J. (2011) Influence of biophysical factors and differences in Ojibwe reservation versus Euro-American social histories on forest landscape change in northern Wisconsin, USA. *Landscape Ecology*, **26**, 1165–1178

Swetnam, T. W., Allen, C. D., and Betancourt, J. L. (1999) Applied historical ecology: Using the past to manage for the future. *Ecological Applications*, **9**, 1189–1206.

Theobald, D.M., Gosnell, H., and Riebsame, W.E. (1996). Land use and landscape change in the Colorado Mountains II: A case study of the East River Valley. *Mountain Research and Development*, **16**, 407–418.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B*, **58**, 267–288.

Turner, M. G. (2005). Landscape ecology: What is the state of the science? *Annual Review of Ecology, Evolution, and Systematics*, **36**, 319–344.

Wedderburn, R. W. M. (1974) Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439–447.

Wang, F. and Wall, M. M. (2003) Generalized common spatial factor model. *Biostatistics*, **4**, 569–582.

White, M.A. and Mladenoff, D.J. (1994) Old-growth forest landscape transitions from pre-European settlement to present. *Landscape Ecology*, **9**, 191–205.

White, C.S. (1983) *A History of the Rectangular Survey System*. U.S. Government Printing Office, Washington DC.

Whitney, G.G. (1994) *From Coastal Forest to Fruited Plain: A History of Environmental Change in Temperate North America*. New York: Cambridge University Press.

Zheng, Y. and Zhu, J. (2008) Markov chain Monte Carlo for spatial-temporal autologistic regression model. *Journal of Computational and Graphical Statistics*, **17**, 123–127.

Zhu, J., Eickhoff, J. C., and Yan, P. (2005) Generalized linear latent variable models for repeated measures of spatially correlated multivariate data. *Biometrics*, **61**, 674–683.

Zhu, J., Huang, H.-C., and Wu, J.-P. (2005) Modeling spatial-temporal binary data using Markov random fields. *Journal of Agricultural, Biological, and Environmental Statistics*, **10**, 212–225.

Zhu, J., Huang, H., and Reyes, P. E. (2010) On selection of spatial linear models for lattice data. *Journal of the Royal Statistical Society: Series B*, **72**, 389–402.

Zou, H. (2006) The adaptive lasso and its oracle properties. *Journal of American Statistical Association*, **101**, 1418–1429.