A Spatial Model for Multivariate Lattice Data

Stephan R. Sain[1] and Noel Cressie[2]

November 4, 2003

ABSTRACT: In this article, we develop Markov random field models for multivariate lattice data. Specific attention is given to building models that incorporate general forms of the spatial correlations and cross-correlations between variables at different sites. The methodology is applied to the problem of environmental equity. Using a Bayesian hierarchical model that is multivariate in form, we examine the racial distribution of residents of southern Louisiana in relation to the location of sites listed with the U.S. Environmental Protection Agency's Toxic Release Inventory.

KEY WORDS: Markov random field, conditional autoregressive model, Bayesian hierarchical model, environmental data analysis.

[1]Geophysical Statistics Project, National Center of Atmospheric Research, Boulder, CO and Department of Mathematics, University of Colorado at Denver, P.O. Box 173364, Denver, CO 80217-3364 USA, V: (303) 556-8463, F: (303) 556-8550, ssain@math.cudenver.edu.

[2](Corresponding author) Department of Statistics, The Ohio State University, Columbus OH 43210-1247 USA, V: (614) 292-5194, F: (614) 292-2096, ncressie@stat.ohio-state.edu

1 Introduction

Many spatial problems, particularly those concerning environmental investigations, are inherently multivariate, in that more than one variable is typically measured at each spatial location. Multivariate spatial databases are becoming much more prevalent with the advent of geographic information systems (GIS) that allow users to display many different spatial data layers.

Formally, consider a spatial location $\mathbf{s}$ and a p-dimensional random variable $\mathbf{Y}(\mathbf{s})$ associated with each location. Letting $\mathbf{s}$ vary over an index set $D \subset \Re^d$ generates a multivariate random field $\{\mathbf{Y}(\mathbf{s}) : \mathbf{s} \in D\}$. For geostatistical data, D is a given subset of $\Re^d$ and $\mathbf{s}$ is assumed to vary continuously throughout D. For lattice data, D is assumed to be a given finite or countable collection of points. Lattices may be either regular, as on a grid, or irregular, such as zip codes, census divisions (counties, tracts, block groups, or blocks), police precincts, game-management units, etc.

Models for multivariate geostatistical data (i.e., data with continuous spatial index) have been extensively explored (e.g., Wackernagel, 1998; Royle and Berliner, 1999; Ver Hoef, Cressie, and Barry, 2003), while models for multivariate lattice data (i.e., data with a countable index set) have received less attention. Mardia (1988) introduced a multivariate Markov random field (MRF) model for image processing, while more recently Billheimer et al. (1997), Pettitt, Weir, and Hart (2002), Carlin and Banerjee (2003), and Gelfand and Vounatsou (2003), have explored these multivariate MRF models and their role in Bayesian hierarchical modeling.

An integral feature of MRF models involves the specification of neighborhoods. For each site or lattice point $\mathbf{s}_i$, a neighborhood is a collection of sites that are spatially close. Neighboring sites can be defined, for example, as two sites separated by a fixed distance or as two sites that share a common boundary. Formally, define $\{N_i\}$ as the collection of neighborhoods, where each N_i is a set of indices representing the neighbors of $\mathbf{s}_i$. Through this specification of neighborhoods, a MRF model of the spatial-dependence structure in lattice data can be constructed.

We consider a class of hierarchical models for lattices, either regular or irregular, with n locations and (potentially) $p > 1$ measurements at each location. Letting Y_{ij} denote the jth observation made at the ith lattice point, the data model is generally written as

$$Y_{ij}|\theta_{ij} \sim f(y|\theta_{ij}), \qquad i = 1, \ldots, n; \;\; j = 1, \ldots, p.$$

Let $\boldsymbol{\theta}$ denote the $n \times p$ matrix of the process parameters and $\boldsymbol{\theta}^v \equiv \text{vec}(\boldsymbol{\theta}')$, which is a $np \times 1$ vector obtained by stacking the columns of $\boldsymbol{\theta}$. Then, the process model for $\boldsymbol{\theta}$ (or perhaps some suitable transformation of $\boldsymbol{\theta}$) is given by,

$$\boldsymbol{\theta}^v \sim \text{N}_{np}(\boldsymbol{\mu}^v, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\mu}^v \equiv \text{vec}(\boldsymbol{\mu}')$ and $\boldsymbol{\mu}$ is the $n \times p$ matrix whose (i, j)th element is $\mu_{ij} = E[\theta_{ij}]$. The large-scale dependence in $\boldsymbol{\theta}$ is captured in the mean $\boldsymbol{\mu}$, while the small-scale, spatial dependence is captured in the covariance matrix $\boldsymbol{\Sigma}$.

The covariance matrix $\boldsymbol{\Sigma}$ is determined by the neighborhood structure of the process on the lattice. Previous efforts at defining the nature of this covariance matrix have used simplified forms that, while computationally efficient, can unduly constrain the covariances; see references at the end of Section 3.2. In this article, we propose a more general and flexible form that allows for varying degrees of spatial dependence for different variables, as well as asymmetric covariances between different variables at different locations.

In the next section, an overview of univariate MRF models is presented, followed by a description of a multivariate model that we call the CAMCAR model. Section 3 presents more details on the parameterization of $\boldsymbol{\Sigma}$, including an interpretation of the spatial parameters. In Section 4, a hierarchical model that incorporates the CAMCAR model is developed, and an analysis using Bayesian inference and Markov chain Monte Carlo (MCMC) is demonstrated on the problem of assessing environmental equity. Section 5 presents an extended CAMCAR model, while Section 6 contains discussion and conclusions.

2 Multivariate MRF Models

In the introduction, we discussed two types of spatial models, geostatistical and MRF. Geostatistical models focus on building covariance functions between the variables as a function of their displacement, which are then used for kriging or cokriging. On the other hand, MRF models for lattice data are based on a type of Markov property in space where the conditional distribution of $\mathbf{Y}(\mathbf{s}_i)$, given all the other values of $\{\mathbf{Y}(\mathbf{s}_j) : j \neq i\}$, depends only on the observations in the neighborhood N_i of $\mathbf{s}_i$. In all that is to follow, we assume the lattice is finite with locations $\{\mathbf{s}_i : i = 1, \ldots, n\}$.

2.1 Univariate Lattice Models

Assume for the moment that $Y(\mathbf{s})$ is univariate. Besag (1974) shows that, under various consistency conditions, the conditional distributions

$$\Pr(Y(\mathbf{s}_i)|\{Y(\mathbf{s}_j) : j \in N_i\}), \qquad i = 1, \ldots, n, \tag{1}$$

can be used to determine the joint distribution

$$\Pr(Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n)),$$

which is called a Markov random field (MRF). Gaussian models that are MRFs are also commonly referred to as conditional autoregressive (CAR) models.

Assuming the conditional distributions in (1) are Gaussian, the ith distribution ($i = 1, \ldots, n$) is specified through

$$\begin{aligned} E[Y(\mathbf{s}_i)|\{Y(\mathbf{s}_j) : j \in N_i\}] &= \mu_i + \sum_{j \in N_i} g_{ij}(Y(\mathbf{s}_j) - \mu_j) \\ \mathrm{var}[Y(\mathbf{s}_i)|\{Y(\mathbf{s}_j) : j \in N_i\}] &= \tau_i^2 \,. \end{aligned}$$

Putting these conditional distributions together yields the joint distribution,

$$\mathbf{Y} \sim \mathrm{N}_n(\boldsymbol{\mu}, (\mathbf{I} - \mathbf{G})^{-1}\mathbf{T}) \,, \tag{2}$$

where $\mathbf{Y} \equiv (Y(\mathbf{s}_1), \ldots, Y(\mathbf{s}_n))'$, $\boldsymbol{\mu} \equiv (\mu_1, \ldots, \mu_n)$, $\mathbf{T} \equiv \text{diag}(\tau_1^2, \ldots, \tau_n^2)$, and $\mathbf{G} \equiv (g_{ij})$. Of course, for the joint distribution in (2) to be well defined, the elements of $\mathbf{G}$ must be chosen so that $(\mathbf{I} - \mathbf{G})^{-1}\mathbf{T}$ is a symmetric, positive-definite matrix.

When covariates are present, it is natural to parameterize $\boldsymbol{\mu}$ by $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ is an $n \times q$ matrix of known covariates and $\boldsymbol{\beta}$ is a $q \times 1$ vector of regression parameters. The small-scale spatial variation contained in $\mathbf{G}$ could be written as $\mathbf{G} = \phi\mathbf{C}$, where $\mathbf{C}$ is a known matrix defining the neighborhood structure. Then ϕ controls the strength of the spatial dependence. The spatial parameter ϕ^2 can sometimes be interpreted as the square of a partial or conditional correlation between neighboring $Y(\mathbf{s}_i)$ and $Y(\mathbf{s}_j)$, depending on the choices of $\mathbf{C}$ and $\mathbf{T}$ (Cressie, 1993, p. 557). Alternatively, it is sometimes appropriate to assume conditional homoscedasticity, where the diagonal matrix $\mathbf{T}$ has constant elements $\tau^2 > 0$.

2.2 Multivariate Lattice Models

Following Mardia (1988), let $\mathbf{Y} \equiv (\mathbf{Y}(\mathbf{s}_1)', \ldots, \mathbf{Y}(\mathbf{s}_n)')'$ and suppose that the $\{\mathbf{Y}(\mathbf{s}_i)\}$ are p-dimensional random vectors whose conditional distribution is a p-variate Gaussian with

$$E(\mathbf{Y}(\mathbf{s}_i)|R_i) = \boldsymbol{\mu}_i + \sum_{j \in N_i} \boldsymbol{\Lambda}_{ij}(\mathbf{Y}(\mathbf{s}_j) - \boldsymbol{\mu}_j), \qquad i = 1, \ldots, n,$$

and

$$\text{var}(\mathbf{Y}(\mathbf{s}_i)|R_i) = \boldsymbol{\Gamma}_i, \qquad i = 1, \ldots, n,$$

where $R_i \equiv \{\mathbf{Y}(\mathbf{s}_j) : j \in N_i\}$ denotes the "rest" of $\mathbf{Y}$ at all locations j in the neighborhood N_i. To ensure the existence of a joint distribution, we make two assumptions. The first guarantees symmetry of var($\mathbf{Y}$):

$$\boldsymbol{\Lambda}_{ij}\boldsymbol{\Gamma}_j = \boldsymbol{\Gamma}_i\boldsymbol{\Lambda}_{ji}', \qquad i, j = 1, \ldots, n,$$

where $\boldsymbol{\Lambda}_{ii} = -\mathbf{I}$, $i = 1, \ldots, n$, and $\boldsymbol{\Lambda}_{ij} \equiv \mathbf{0}$ for $j \notin N_i \cup \{i\}$.

The second assumption guarantees positive-definiteness of var($\mathbf{Y}$):

$$\text{Block}(-\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Lambda}_{ij}) \quad \text{or} \quad \text{Block}(-\boldsymbol{\Lambda}_{ij}) \quad \text{is positive-definite,}$$

where Block($\mathbf{A}_{ij}$) denotes a block matrix with the (i,j)th block given by $\mathbf{A}_{ij}$. Then $\mathbf{Y}$ is $N_{np}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} \equiv (\boldsymbol{\mu}_1', \ldots, \boldsymbol{\mu}_n')' \qquad \text{and} \qquad \boldsymbol{\Sigma} \equiv \{\text{Block}(-\boldsymbol{\Gamma}_i^{-1}\boldsymbol{\Lambda}_{ij})\}^{-1}. \tag{3}$$

Motivated by our modeling of rates in Section 4, we choose the following parameterization:

$$\boldsymbol{\Lambda}_{ij} = \mathbf{m}_i^{-1/2}\boldsymbol{\Lambda}\mathbf{m}_j^{1/2}, \qquad i < j, \;\; j \in N_i,$$

and

$$\boldsymbol{\Gamma}_i = \mathbf{m}_i^{-1/2}\boldsymbol{\Gamma}\mathbf{m}_i^{-1/2}, \qquad i = 1, \ldots, n, \tag{4}$$

where $\mathbf{m}_i$ is a $p \times p$ diagonal matrix of precision measures (see Section 3.3) and $\boldsymbol{\Gamma}$ is a covariance matrix (i.e., a symmetric positive-definite matrix). The symmetry condition implies that

$$\boldsymbol{\Lambda}_{ji} = \boldsymbol{\Gamma}_j\boldsymbol{\Lambda}_{ij}'\boldsymbol{\Gamma}_i^{-1}, \qquad i < j, \;\; j \in N_i,$$

because site i is a neighbor of site j if and only if site j is a neighbor of site i. Hence,

$$\boldsymbol{\Lambda}_{ji} = \mathbf{m}_j^{-1/2}\boldsymbol{\Gamma}\boldsymbol{\Lambda}'\boldsymbol{\Gamma}^{-1}\mathbf{m}_i^{1/2}, \qquad i < j, \;\; j \in N_i.$$

Define $\mathbf{B} \equiv \boldsymbol{\Gamma}^{-1/2}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^{1/2}$. Then, for $i < j$ and $j \in N_i$,

$$\begin{aligned} \boldsymbol{\Lambda}_{ij} &= \mathbf{m}_i^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{B}\boldsymbol{\Gamma}^{-1/2}\mathbf{m}_j^{1/2} \\ \boldsymbol{\Lambda}_{ji} &= \mathbf{m}_j^{-1/2}\boldsymbol{\Gamma}^{1/2}\mathbf{B}'\boldsymbol{\Gamma}^{-1/2}\mathbf{m}_i^{1/2}\,. \end{aligned}$$

The ith diagonal block in the covariance matrix $\boldsymbol{\Sigma}$ can be written as $\mathbf{m}_i^{1/2}\boldsymbol{\Gamma}^{-1}\mathbf{m}_i^{1/2}$. For $i < j$ and $j \in N_i$, the blocks in the upper-triangular part of $\boldsymbol{\Sigma}$ can be written as $-\mathbf{m}_i^{1/2}\boldsymbol{\Gamma}^{-1/2}\mathbf{B}\boldsymbol{\Gamma}^{-1/2}\mathbf{m}_j^{1/2}$ and blocks in the lower-triangular part of $\boldsymbol{\Sigma}$ can be written as $-\mathbf{m}_j^{1/2}\boldsymbol{\Gamma}^{-1/2}\mathbf{B}'\boldsymbol{\Gamma}^{-1/2}\mathbf{m}_i^{1/2}$. Simplifying yields

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma}^*\mathbf{H}^{-1}\boldsymbol{\Gamma}^{*\prime}, \tag{5}$$

where

$$\mathbf{H} \equiv \begin{bmatrix} \mathbf{I} & -\mathbf{B}I(2 \in N_1) & \cdots & -\mathbf{B}I(n \in N_1) \\ -\mathbf{B}'I(1 \in N_2) & \mathbf{I} & \cdots & -\mathbf{B}I(n \in N_2) \\ \vdots & & \ddots & \vdots \\ -\mathbf{B}'I(1 \in N_n) & -\mathbf{B}'I(2 \in N_n) & \cdots & \mathbf{I} \end{bmatrix}, \tag{6}$$

and

$$\mathbf{\Gamma}^* \equiv \begin{bmatrix} \mathbf{m}_1^{-1/2}\mathbf{\Gamma}^{1/2} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{m}_n^{-1/2}\mathbf{\Gamma}^{1/2} \end{bmatrix}.$$

Since $\mathbf{\Gamma}^*$ is block diagonal and all of the diagonal blocks are positive-definite ($\mathbf{\Gamma}$ is a covariance matrix), $\mathbf{\Gamma}^*$ is positive-definite. So, to guarantee that $\mathbf{\Sigma}$ is positive-definite, it is enough to specify the parameter space for $\mathbf{H}$ to be the set of all matrices given by (6) that are positive-definite. We call the model defined by (5) and (6) a (simple) canonical multivariate conditional autoregressive (CAMCAR) model and derive its properties below.

3 Properties of the CAMCAR Model

Recall from (5) that the spatial variability in the CAMCAR model can be expressed in terms of $\mathbf{H}$ and $\mathbf{\Gamma}^*$, which in turn can be expressed in terms of $\mathbf{B}$ and $\{\mathbf{m}_i\}$, respectively. In this section, the behavior of the matrix $\mathbf{H}$ is discussed, particularly conditions on the spatial-correlation parameters $\mathbf{B}$ that give positive-definite $\mathbf{H}$. An interpretation of $\mathbf{B}$ is also presented as well as a comparison of various parameterizations of $\mathbf{B}$. Finally, the role of the $\{\mathbf{m}_i\}$ is discussed.

3.1 The Matrix H

From (6), $\mathbf{H}$ is symmetric and typically sparse (i.e., many zero elements). Furthermore, it is the particular choice of $\mathbf{B}$ that makes $\mathbf{H}$ positive-definite. For example, if $\mathbf{B}$ is constrained to be a symmetric matrix ($\mathbf{B} = \mathbf{B}'$), then $\mathbf{H}$ can be written as $\mathbf{H} = \mathbf{I} - \mathbf{B} \otimes \mathbf{C}$ where $\otimes$ denotes the Kronecker product of two matrices and $\mathbf{C}$ is a known $n \times n$ incidence matrix whose (i, j)th element equals one if sites i and j are neighbors and zero otherwise. Then,

the eigenvalues of $\mathbf{H}$ can be written as $1 - b_i c_j$, where $b_i,\ i = 1, \ldots, p$, are the eigenvalues of $\mathbf{B}$ and $c_j,\ j = 1, \ldots, n$, are the eigenvalues of $\mathbf{C}$. Hence, the eigenvalues of $\mathbf{B}$ can be constrained to ensure that all of the eigenvalues of $\mathbf{H}$ are positive (i.e., $\mathbf{H}$ is positive-definite). In particular, each $b_i,\ \ i = 1, \ldots, p$, must be chosen such that $1/c_1 < b_i < 1/c_n$, where $c_1 < 0$ is the smallest eigenvalue of $\mathbf{C}$ and $c_n > 0$ is the largest.

When $\mathbf{B}$ has a more general and possibly asymmetric specification, $\mathbf{H}$ is not so straightforward to simplify. However, by restricting $\mathbf{H}$ to be strictly diagonally dominant, conditions on the range of the elements of $\mathbf{B}$ can be found that will ensure $\mathbf{H}$ is positive-definite.

For a $n \times n$ matrix $\mathbf{A} = (a_{ij})$, strict diagonal dominance imples that $\mathbf{A}$ has the property that

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|, \qquad i = 1, \ldots, n.$$

In other words, the sum of the absolute value of the off-diagonal elements in each row must be less than the absolute value of its respective diagonal element. If we assume that as well as being diagonally dominant, $\mathbf{A}$ has positive diagonal elements and is symmetric, then it is straightforward to show that it is also positive-definite through the use of Gershgorin's disc theorem (Horn and Johnson, 1990). Gershgorin's disc theorem states that, for a square complex matrix, every eigenvalue of the matrix lies in a disc in the complex plane centered at one of the diagonal elements a_{ii} and with radius $\sum_{j \neq i} |a_{ij}|$. Since $\mathbf{A}$ is symmetric, all eigenvalues of $\mathbf{A}$ are real. Hence, the discs suggested by Gershgorin's theorem reduce to intervals on the real line of the form,

$$\left(a_{ii} - \sum_{j \neq i} |a_{ij}|, a_{ii} + \sum_{j \neq i} |a_{ij}|\right), \quad i = 1, \ldots, n\,.$$

Further, since the diagonal elements of $\mathbf{A}$ are positive and from the definition of diagonal dominance, these intervals contain no negative elements. Hence, all eigenvalues of $\mathbf{A}$ are positive and $\mathbf{A}$ is positive-definite.

For the symmetric matrix $\mathbf{H}$, diagonal dominance yields a sequence of conditions on

the elements of $\mathbf{B}$ of the form,

$$|N_i||b_{kk}| + n_{iL}\sum_{\ell\neq k}^{p}|b_{\ell k}| + n_{iU}\sum_{\ell\neq k}^{p}|b_{k\ell}| < 1, \qquad i = 1,\ldots,n, \quad k = 1,\ldots,p,$$

where $|N_i|$ is the number of elements in N_i, n_{iL} is the number of elements j in N_i such that $j < i$, and n_{iU} is the number of elements j in N_i such that $j > i$.

For regular lattices, these conditions are the same for all values of i except near the boundaries. For an irregular lattice, the conditions can be different for each i, since the number of neighbors might differ. In either event, for each k, the most restrictive of the sequence of conditions is used to ensure that the matrix is diagonally dominant for all i. For example, when $p = 2$, the elements of $\mathbf{B}$ are constrained such that

$$\max_i\{|N_i||b_{11}| + |b_{21}|n_{iL} + |b_{12}|n_{iU}\} < 1$$

and

$$\max_i\{|N_i||b_{22}| + |b_{12}|n_{iL} + |b_{21}|n_{iU}\} < 1. \tag{7}$$

3.2 The Role of B

Examining the conditional covariance of $\mathbf{Y}(\mathbf{s}_i)$ and $\mathbf{Y}(\mathbf{s}_j)$, given $\{\mathbf{Y}(\mathbf{s}_k) : k \neq i, j\}$, can yield insight into the interpretation of the spatial parameters contained in $\mathbf{B}$. To establish the conditional covariance, recall two important results from multivariate statistics (see, e.g., Mardia, Kent, and Bibby, 1979). First, partition a random vector $\mathbf{X}$ as

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{bmatrix},$$

and let $\mathbf{X}$ be distributed as $N(\boldsymbol{\eta}, \boldsymbol{\Omega})$ with

$$\boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\eta}_1 \\ \boldsymbol{\eta}_2 \end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{bmatrix}.$$

Then, the conditional distribution of $\mathbf{X}_1$ given $\mathbf{X}_2$ is multivariate normal with

$$\boldsymbol{\eta}_{1|2} = \boldsymbol{\eta}_1 + \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}(\mathbf{X}_2 - \boldsymbol{\eta}_2) \qquad \text{and} \qquad \boldsymbol{\Omega}_{1|2} = \boldsymbol{\Omega}_{11} - \boldsymbol{\Omega}_{12}\boldsymbol{\Omega}_{22}^{-1}\boldsymbol{\Omega}_{21}. \tag{8}$$

The second result concerns the inverse of partitioned matrices. Let $\mathbf{\Omega}$ be given by

$$\mathbf{\Omega} = \begin{bmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{21} & \mathbf{\Omega}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{\Omega}^{11} & \mathbf{\Omega}^{12} \\ \mathbf{\Omega}^{21} & \mathbf{\Omega}^{22} \end{bmatrix}^{-1}.$$

Then,

$$\mathbf{\Omega}_{11} = (\mathbf{\Omega}^{11} - \mathbf{\Omega}^{12}(\mathbf{\Omega}^{22})^{-1}\mathbf{\Omega}^{21})^{-1} \quad \mathbf{\Omega}_{22} = (\mathbf{\Omega}^{22} - \mathbf{\Omega}^{21}(\mathbf{\Omega}^{11})^{-1}\mathbf{\Omega}^{12})^{-1}$$

and

$$\begin{aligned} \mathbf{\Omega}_{12} &= -(\mathbf{\Omega}^{11})^{-1}\mathbf{\Omega}^{12}\mathbf{\Omega}_{22} = -\mathbf{\Omega}_{11}\mathbf{\Omega}^{12}(\mathbf{\Omega}^{22})^{-1} \\ \mathbf{\Omega}_{21} &= -(\mathbf{\Omega}^{22})^{-1}\mathbf{\Omega}^{21}\mathbf{\Omega}_{11} = -\mathbf{\Omega}_{22}\mathbf{\Omega}^{21}(\mathbf{\Omega}^{11})^{-1}. \end{aligned}$$

Combining these results yields,

$$\begin{aligned} \mathbf{\Omega}_{1|2} &= \mathbf{\Omega}_{11} - (\mathbf{\Omega}^{11})^{-1}\mathbf{\Omega}^{12}\mathbf{\Omega}_{22}\mathbf{\Omega}_{22}^{-1}(\mathbf{\Omega}^{22})^{-1}\mathbf{\Omega}^{21}\mathbf{\Omega}_{11} \\ &= (\mathbf{\Omega}^{11})^{-1}(\mathbf{\Omega}^{11} - \mathbf{\Omega}^{12}(\mathbf{\Omega}^{22})^{-1}\mathbf{\Omega}^{21})\mathbf{\Omega}_{11} \\ &= (\mathbf{\Omega}^{11})^{-1}. \end{aligned}$$

Given the covariance matrix $\mathbf{\Sigma} = \mathbf{\Gamma}^*\mathbf{H}^{-1}\mathbf{\Gamma}^{*\prime}$ defined in (5), then the previous result shows that the conditional covariance of $\mathbf{Y}(\mathbf{s}_i)$ given $\{\mathbf{Y}(\mathbf{s}_j) : j \neq i\}$ is,

$$\mathbf{\Sigma}_{i|-i} = \mathbf{m}_i^{-1/2}\mathbf{\Gamma}\mathbf{m}_i^{-1/2},$$

which is consistent with the results of Section 2.2 and with Mardia (1988). The conditional correlation matrix is then given by

$$\mathbf{R}_{i|-i} = \mathbf{D}^{-1/2}\mathbf{\Gamma}\mathbf{D}^{-1/2}, \tag{9}$$

where $\mathbf{D} \equiv \text{diag}(\gamma_{11}, \ldots, \gamma_{pp})$ and γ_{ii} is the ith diagonal element of $\mathbf{\Gamma}$. Notice that the precision measures $\{\mathbf{m}_i\}$ do not appear in the conditional correlation matrix, which is an attractive property of the CAMCAR model.

For two sites $i < j$ that are neighbors, their conditional covariance given $\{\mathbf{Y}(\mathbf{s}_k) : k \neq i, j\}$ is

$$\begin{aligned} \mathbf{\Sigma}_{ij|-ij} &= \mathbf{\Gamma}_{ij}^* \begin{bmatrix} \mathbf{I} & -\mathbf{B} \\ -\mathbf{B}' & \mathbf{I} \end{bmatrix}^{-1} \mathbf{\Gamma}_{ij}^{*\prime} \\ &= \mathbf{\Gamma}_{ij}^* \begin{bmatrix} (\mathbf{I} - \mathbf{B}\mathbf{B}')^{-1} & (\mathbf{I} - \mathbf{B}\mathbf{B}')^{-1}\mathbf{B} \\ (\mathbf{I} - \mathbf{B}'\mathbf{B})^{-1}\mathbf{B}' & (\mathbf{I} - \mathbf{B}'\mathbf{B})^{-1} \end{bmatrix} \mathbf{\Gamma}_{ij}^{*\prime}, \end{aligned}$$

where

$$\mathbf{\Gamma}^*_{ij} = \begin{bmatrix} \mathbf{m}_i^{-1/2}\mathbf{\Gamma}^{1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{m}_j^{-1/2}\mathbf{\Gamma}^{1/2} \end{bmatrix}.$$

To establish the conditional correlation matrix, note that $\mathbf{\Sigma}_{ij|-ij}$ can be written as

$$\mathbf{\Sigma}_{ij|-ij} = \begin{bmatrix} \mathbf{m}_i^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{m}_j^{-1/2} \end{bmatrix} \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{bmatrix} \begin{bmatrix} \mathbf{m}_i^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{m}_j^{-1/2} \end{bmatrix},$$

where

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{\Gamma}^{1/2}(\mathbf{I} - \mathbf{B}\mathbf{B}')^{-1}\mathbf{\Gamma}^{1/2} & \mathbf{\Gamma}^{1/2}(\mathbf{I} - \mathbf{B}\mathbf{B}')^{-1}\mathbf{B}\mathbf{\Gamma}^{1/2} \\ \mathbf{\Gamma}^{1/2}(\mathbf{I} - \mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{\Gamma}^{1/2} & \mathbf{\Gamma}^{1/2}(\mathbf{I} - \mathbf{B}'\mathbf{B})^{-1}\mathbf{\Gamma}^{1/2} \end{bmatrix}.$$

Observe that the first p diagonal elements of $\mathbf{\Sigma}_{ij|-ij}$ are given by $w_{11,kk}/m_{i,kk}$, where $w_{11,kk}$ is the kth diagonal element of $\mathbf{W}_{11}$ and $m_{i,kk}$ is the kth diagonal element of $\mathbf{m}_i$. Similarly, the last p diagonal elements are given by $w_{22,kk}/m_{j,kk}$. Let $\mathbf{D}_{ij}$ denote the diagonal matrix with these $2p$ diagonal elements. Then the conditional correlation matrix is given by,

$$\begin{aligned} \mathbf{R}_{ij|-ij} &= \mathbf{D}_{ij}^{-1/2}\mathbf{\Sigma}_{ij|-ij}\mathbf{D}_{ij}^{-1/2} \\ &\equiv \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & \mathbf{R}_{22} \end{bmatrix}, \end{aligned}$$

where the elements of $\mathbf{R}_{ab}$ $(a, b = 1, 2)$ are given by

$$R_{ab,kl} = \frac{w_{ij,kl}}{\sqrt{w_{ii,kk}w_{jj,ll}}}, \qquad k, l = 1, \ldots, p.$$

Again, the conditional correlation matrix does not depend on the precision measures $\{\mathbf{m}_i\}$, and the elements $R_{ab,kl}$ are strictly functions of $\mathbf{W}$ and hence of $\mathbf{\Gamma}$ and $\mathbf{B}$.

In the univariate case ($p = 1$), $\mathbf{B} = \phi$, $\mathbf{\Gamma} = \tau^2$, and the conditional covariance matrix simplifies to

$$\begin{aligned} \mathbf{\Sigma}_{ij|-ij} &= \begin{bmatrix} m_i^{-1/2}\tau & 0 \\ 0 & m_j^{-1/2}\tau \end{bmatrix} \begin{bmatrix} (1-\phi^2)^{-1} & (1-\phi^2)^{-1}\phi \\ (1-\phi^2)^{-1}\phi & (1-\phi^2)^{-1} \end{bmatrix} \begin{bmatrix} \tau m_i^{-1/2} & 0 \\ 0 & \tau m_j^{-1/2} \end{bmatrix} \\ &= \begin{bmatrix} \dfrac{\tau^2/m_i}{1-\phi^2} & \dfrac{\tau^2\phi/(m_i m_j)^{1/2}}{1-\phi^2} \\ \dfrac{\tau^2\phi/(m_i m_j)^{1/2}}{1-\phi^2} & \dfrac{\tau^2/m_j}{1-\phi^2} \end{bmatrix}. \end{aligned}$$

The conditional correlation between measurements at sites i and j is then given by ϕ, which is consistent with Cressie and Chan (1989) and Stern and Cressie (1999).

A simplification can be achieved by restricting the form of $\mathbf{B}$. Setting $\mathbf{B} = \phi\mathbf{I}$ yields a conditional covariance $\mathbf{\Sigma}_{ij|-ij}$ of the form

$$\mathbf{\Sigma}_{ij|-ij} = \begin{bmatrix} \frac{1}{1-\phi^2}\mathbf{m}_i^{-1/2}\mathbf{\Gamma}\mathbf{m}_i^{-1/2} & \frac{\phi}{1-\phi^2}\mathbf{m}_i^{-1/2}\mathbf{\Gamma}\mathbf{m}_j^{-1/2} \\ \frac{\phi}{1-\phi^2}\mathbf{m}_j^{-1/2}\mathbf{\Gamma}\mathbf{m}_i^{-1/2} & \frac{1}{1-\phi^2}\mathbf{m}_j^{-1/2}\mathbf{\Gamma}\mathbf{m}_j^{-1/2} \end{bmatrix}.$$

The conditional correlation matrix is then given by

$$\mathbf{R}_{ij|-ij} = \begin{bmatrix} \mathbf{R} & \phi\mathbf{R} \\ \phi\mathbf{R} & \mathbf{R} \end{bmatrix} = \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix} \otimes \mathbf{R}, \tag{10}$$

where $\mathbf{R} \equiv \mathbf{R}_{i|-i}$ in (9). For a fixed sited located at $\mathbf{s}_i$, we see that the relationship between neighboring sites can be characterized as simply a multiple of the conditional correlation matrix (9) for the i-th site. Furthermore, since $\mathbf{R}$ is symmetric, the conditional cross-correlations between different variables at neighboring sites are equal. As noted earlier, the correlations are not functions of the precision measures $\{\mathbf{m}_i\}$, although to preserve positive-definiteness, the range of possible correlations would be determined by the neighborhood structures (e.g., see (7)).

Using a parameterization $\mathbf{B} = \text{diag}(\phi_1, \ldots, \phi_p)$, yields a slightly more general specification. However, the conditional cross-correlation between variable k at site i and variable ℓ at site j is equal to that between variable ℓ at site i and variable k at site j, a consequence that is true for any symmetric choice for $\mathbf{B}$. This restrictive spatial cross-dependence is present in the models of Billheimer et al. (1997), Pettit et al. (2002), Carlin and Banerjee (2003), and Gelfand and Vounatsou (2003). Only the most general specification of $\mathbf{B}$ yields conditional cross-correlations between variables at neighboring sites that have the possibility of being different.

3.3 The Precision Measures $\{\mathbf{m}_i\}$

The role of the precision measures captured by the $p \times p$ diagonal matrices $\{\mathbf{m}_i\}$, is to allow for varying degrees of variability in the data. This is especially true when the ob-

servations at each lattice point represent *rates* computed over regions such as counties, zip codes, census divisions, etc., where the variation in the rates from location to location is determined by the relevant-population number in each region. This suggests, for example, that the diagonal matrix $\mathbf{m}_i$ could be taken to have diagonal elements equal to the relevant-population number for each rate observed at location $\mathbf{s}_i$.

The impact of how the precision of the measurements enters the model becomes clear when comparing CAMCAR with other formulations of the multivariate CAR suggested by Billheimer et al. (1997), Carlin and Banerjee (2003), and Gelfand and Vounatsou (2003), which link the precision of the data at location $\mathbf{s}_i$ to the number of neighbors at $\mathbf{s}_i$ (denoted by $|N_i|$). For example, set

$$\mathbf{\Gamma}_i = \frac{1}{|N_i|}\mathbf{\Gamma}.$$

In addition, set $\mathbf{\Lambda}_{ii} = -\mathbf{I}$ and

$$\mathbf{\Lambda}_{ij} = \mathbf{\Lambda}_i = \frac{\alpha}{|N_i|}\mathbf{I}\,,$$

if j is a neighbor of i and $\mathbf{\Lambda}_{ij} = \mathbf{0}$ otherwise. From (3), this yields a covariance matrix of the form,

$$\mathbf{\Sigma} = (\mathrm{diag}(|N_1|, \ldots, |N_n|) - \alpha\mathbf{C})^{-1} \otimes \mathbf{\Gamma}\,,$$

where $\mathbf{C}$ is an $n \times n$ adjacency matrix with diagonal entries equal to zero and (i, j)th entry equal to one if i and j are neighbors and zero otherwise.

The conditional covariance for two neighboring sites i and j is given by

$$\begin{aligned}
\mathbf{\Sigma}_{ij|-ij} &= \begin{bmatrix} |N_i| & -\alpha \\ -\alpha & |N_j| \end{bmatrix}^{-1} \otimes \mathbf{\Gamma} \\
&= \mathbf{\Gamma}^*_{ij} \begin{bmatrix} \mathbf{I} & -\dfrac{\alpha}{\sqrt{|N_i|\,|N_j|}}\mathbf{I} \\ -\dfrac{\alpha}{\sqrt{|N_i|\,|N_j|}}\mathbf{I} & \mathbf{I} \end{bmatrix}^{-1} \mathbf{\Gamma}^*_{ij}
\end{aligned}$$

$$= \boldsymbol{\Gamma}^*_{ij} \begin{bmatrix} \frac{1}{1-\alpha/\sqrt{|N_i|\ |N_j|}}\mathbf{I} & \frac{\alpha/\sqrt{|N_i|\ |N_j|}}{1-\alpha/\sqrt{|N_i|\ |N_j|}}\mathbf{I} \\ \frac{\alpha/\sqrt{|N_i|\ |N_j|}}{1-\alpha/\sqrt{|N_i|\ |N_j|}}\mathbf{I} & \frac{1}{1-\alpha/\sqrt{|N_i|\ |N_j|}}\mathbf{I} \end{bmatrix} \boldsymbol{\Gamma}^*_{ij},$$

where

$$\boldsymbol{\Gamma}^*_{ij} = \begin{bmatrix} \boldsymbol{\Gamma}^{1/2}/\sqrt{|N_i|} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Gamma}^{1/2}/\sqrt{|N_j|} \end{bmatrix}.$$

The conditional correlation matrix is then

$$\mathbf{R}_{ij|-ij} = \begin{bmatrix} 1 & \alpha/\sqrt{|N_i|\ |N_j|} \\ \alpha/\sqrt{|N_i|\ |N_j|} & 1 \end{bmatrix} \otimes \mathbf{R}, \tag{11}$$

where recall that $\mathbf{R} \equiv \mathbf{R}_{i|-i}$ in (9).

We compare this specification of the MRF with that of the CAMCAR model in (10), where an analogous condition concerning the spatial correlation parameters ($\mathbf{B} = \phi\mathbf{I}$) was used. Therefore, the only difference is the direct use of the precision matrices $\{\mathbf{m}_i\}$ for the CAMCAR model. Unlike for the correlations in (10), the correlations in (11) are functions of the number of neighbors and can vary with each pair of neighbors for an irregular lattice, making interpretation of the spatial-dependence parameter α difficult. Hence, incorporating the precision matrices $\{\mathbf{m}_i\}$ not only into the $\{\boldsymbol{\Gamma}_i\}$, but also into the $\{\boldsymbol{\Lambda}_{ij}\}$ (Section 2.2), yields spatial correlations that are not directly functions of the number of neighbors. (But recall that conditions like (6) imply that the range of the correlations is still dependent on the number of neighbors.)

4 Assessments of Environmental Equity in Louisiana

In 1996, the Shintech Corporation, a Houston-based division of the Japanese conglomerate Shin-Etsu, announced plans to build one of the world's largest polyvinyl chloride plants in the small town of Convent, located in St. James Parish, Louisiana. St. James Parish is located in the heart of Louisiana's Industrial Corridor, the area of the state surrounding

the Mississippi River between Baton Rouge and New Orleans. This area has a heavy concentration of manufacturing and other industrial facilities, including more than half of the facilities in the state listed with the U.S. Environmental Protection Agency's Toxic Release Inventory (TRI).

The proposed plant-siting prompted lawyers with the Environmental Law Clinic at Tulane University to file complaints citing Title VI of the Civil Rights Act and the February 11, 1994 Executive Order 12898, Federal Actions to Address Environmental Justice in Minority Populations and Low-Income Populations. While Shintech ultimately decided not to put the facility in Convent, many agencies and groups, including the U.S. Environmental Protection Agency (EPA), analyzed data from Convent, St. James Parish, and Louisiana as a whole. Many of these studies focused on the socio-demographics and the distribution and type of toxic facility, including those listed with the TRI, to determine whether the proposed siting of the Shintech plant would significantly increase the hazards already affecting the population of Convent and St. James Parish.

The notion of environmental equity is based in the idea that no subpopulation, including minority and/or low income groups, should bear a disproportionate burden from environmental hazards. A wide spectrum of approaches have been used to assess environmental equity. Early work, including the often-cited reports from the General Accounting Office (GAO, 1983) and the United Church of Christ (UCC, 1987), focused on using simple descriptive statistics comparing the racial distribution between areas deemed to be affected by toxic facilities and areas deemed to be unaffected. Recent methodological research in environmental equity has used far more sophisticated techniques that even attempt to link distance from toxic sites to disease occurrence (e.g., Waller et al., 1999, and Carlin and Xia, 1999). For an overview of some statistical issues on environmental equity, see also Waller and Conlon (2000).

In the analysis that follows, we use the CAMCAR model as a component of a hierarchical statistical model that relates population demographics with the location of certain toxic facilities. Specifically, we examine the relationship between the populations of whites

and minorities and the impact of the locations of TRI facilities in a section of Louisiana surrounding St. James Parish. In this case, environmental equity would be manifest by any relationship being the same for whites and minorities.

4.1 Specifying the Hierarchical Statistical Model

Let Y_{i1} and Y_{i2} denote the observed counts of white and minority populations, respectively, for locations $i = 1, \ldots, n$. These locations can be, for example, zip codes or census divisions such as tracts or block groups. In what follows, we build the hierarchical statistical model for the general case of p variables, although in our application $p = 2$. The first level of this model, namely the data model, is then

$$Y_{ik}|\theta_{ik} \sim \text{Poisson}(E_{ik}e^{\theta_{ik}}), \qquad i = 1, \ldots, n; \;\; k = 1, \ldots, p,$$

where E_{ik} are known expected counts for the ith location and the kth variable; see Section 4.2. The parameters θ_{ik} represent departures of the observed counts from the expected counts.

Recall that $\boldsymbol{\theta}^v \equiv \text{vec}(\boldsymbol{\theta}')$, where the vec operator stacks the columns of a matrix. Also, let $\mathbf{X}$ denote a $n \times q$ matrix of covariates and $\boldsymbol{\beta}$ a $q \times p$ matrix of regression coefficients. Then the second level of the hierarchical statistical model, namely the process model, includes regression effects and stochastic spatial-dependence effects expressed through the CAMCAR model,

$$\boldsymbol{\theta}^v|\boldsymbol{\beta}, \mathbf{V}, \mathbf{B} \sim \text{N}_{np}(\boldsymbol{\mu}^v, \boldsymbol{\Sigma})$$

where $\boldsymbol{\mu}^v = \text{vec}(\boldsymbol{\mu}')$, $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, $\mathbf{V} = \boldsymbol{\Gamma}^{-1}$, $\mathbf{m}_i \equiv \text{diag}(E_{i1}, \ldots, E_{ip})$, $i = 1, \ldots, n$, and $\boldsymbol{\Sigma}$ is defined as in (5).

To complete the specification, the third level, namely the prior model for $\boldsymbol{\beta}$, $\mathbf{V}$, and $\mathbf{B}$ is specified. For the regression parameters, assume that each column of $\boldsymbol{\beta}$, denoted by $\boldsymbol{\beta}_k$, $k = 1, \ldots, p$, has the prior distribution

$$\boldsymbol{\beta}_k \sim \text{N}_p(\mathbf{0}, \sigma^{\mathbf{2}}\mathbf{I}).$$

The prior on the inverse covariance matrix $\mathbf{V}$ is given by the Wishart distribution (Mardia, Kent, and Bibby, 1979),

$$\mathbf{V} \sim \text{Wishart}(\rho, (\rho\mathbf{A})^{-1}),$$

where $\rho > p$ and $\mathbf{A}$ is a pre-specified symmetric positive-definite matrix (e.g., $\mathbf{A} = \mathbf{I}$). Finally, the prior on $\mathbf{B}$ is taken to be proportional to,

$$\exp\left\{-(\mathbf{B}^v)'\mathbf{B}^v/\xi^2\right\},$$

where $\mathbf{B}^v = \text{vec}(\mathbf{B}')$. Of course, the prior is truncated for values of $\mathbf{B}$ outside the limits constrained by the diagonal dominance of the matrix $\mathbf{H}$. For small values of ξ, the prior for $\mathbf{B}$ is quite peaked at zero. Hence, posterior distributions with mean away from zero would indicate strong evidence of significant (non-zero) values in the data. Also, note that the prior for $\mathbf{B}$ does not favor symmetric or asymmetric $\mathbf{B}$.

Bayes' theorem yields the posterior distribution, which here is

$$P(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{V}, \mathbf{B}|\mathbf{Y}) \propto P(\mathbf{Y}|\boldsymbol{\theta})P(\boldsymbol{\theta}|\boldsymbol{\beta}, \mathbf{V}, \mathbf{B})P(\boldsymbol{\beta})P(\mathbf{V})P(\mathbf{B}).$$

Substituting in the model specifications at the three levels yields

$$\begin{aligned}
P(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\Gamma}, \mathbf{B}|\mathbf{Y}) \;\propto\; & \prod_{i=1}^{n}\prod_{k=1}^{p} \exp(-E_{ik}e^{\theta_{ik}})(E_{ik}e_{ik}^{\theta})^{Y_{ik}} \\
\times\; & |\mathbf{V}|^{n/2}|\mathbf{H}|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)\right\} \\
\times\; & \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\beta}_1'\boldsymbol{\beta}_1\right\} \times \cdots \times \exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\beta}_p'\boldsymbol{\beta}_p\right\} \\
\times\; & |\mathbf{V}|^{(\rho-p-1)/2} \exp\left\{-\frac{\rho}{2}\text{tr}(\mathbf{A}^{-1}\mathbf{V})\right\} \\
\times\; & \exp\left\{-(\mathbf{B}^v)'\mathbf{B}^v/\xi^2\right\}.
\end{aligned}$$

Sampling from the posterior requires the use of Markov chain Monte Carlo (MCMC) such as the Gibbs sampler (Geman and Geman, 1984; Gelfand and Smith, 1990; Gelfand et al., 1990; see also Gilks, Richardson, and Spiegelhalter, 1996). One iteration of the algorithm requires sampling from

1. $P(\boldsymbol{\beta}_k|\boldsymbol{\beta}_{-k}, \mathbf{V}, \mathbf{B}, \boldsymbol{\theta})$, $k = 1, \ldots, p$

2. $P(\mathbf{V}|\boldsymbol{\beta}, \mathbf{B}, \boldsymbol{\theta})$

3. $P(B_{k\ell}|\boldsymbol{\beta}, \mathbf{V}, \boldsymbol{\theta})$, $k, \ell = 1, \ldots, p$

4. $P(\boldsymbol{\theta}_i|\boldsymbol{\theta}_{-i}, \boldsymbol{\beta}, \mathbf{V}, \boldsymbol{\Gamma}, \mathbf{Y})$, $i = 1, \ldots, n$,

where $\boldsymbol{\beta}_{-k}$ denotes the all columns of $\boldsymbol{\beta}$ except for the kth and similarly for $\boldsymbol{\theta}_{-i}$. Also, $B_{k\ell}$ indicates the (k, ℓ)th element of $\mathbf{B}$. We now show how each component of this iteration can be implemented.

4.1.1 Conditional Distribution of $\boldsymbol{\beta}_k$

For $\boldsymbol{\beta}_k$, the conditional distribution is proportional to

$$\exp\left\{-\frac{1}{2}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)\right\}\exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\beta}_k'\boldsymbol{\beta}_k\right\}. \tag{12}$$

To simplify (12), consider re-ordering the elements of the quadratic form. Note that $\boldsymbol{\theta}^v$ can be written as

$$\boldsymbol{\theta}^v = \begin{bmatrix} \theta_{11} \\ \theta_{12} \\ \vdots \\ \theta_{1p} \\ \vdots \\ \theta_{np} \end{bmatrix}.$$

However, consider writing it as

$$\boldsymbol{\theta}^* = \begin{bmatrix} \theta_{11} \\ \theta_{21} \\ \vdots \\ \theta_{n1} \\ \vdots \\ \theta_{np} \end{bmatrix}.$$

This re-ordering results in $\boldsymbol{\theta}^*$ having a covariance matrix that is an $np \times np$ block matrix of the form Block($\mathbf{S}_{k\ell}$), where $\mathbf{S}_{k\ell}$ is an $n \times n$ matrix derived below.

Then, the conditional distribution in (12) is proportional to

$$\exp\left\{-\frac{1}{2}[(\boldsymbol{\theta}_k^* - \boldsymbol{\mu}_k^*)'\mathbf{S}_{kk}(\boldsymbol{\theta}_k^* - \boldsymbol{\mu}_k^*) + 2(\boldsymbol{\theta}_k^* - \boldsymbol{\mu}_k^*)'\sum_{\ell\neq k}\mathbf{S}_{k\ell}(\boldsymbol{\theta}_\ell^* - \boldsymbol{\mu}_\ell^*)]\right\}\exp\left\{-\frac{1}{2\sigma^2}\boldsymbol{\beta}_k'\boldsymbol{\beta}_k\right\}, \quad (13)$$

where $\boldsymbol{\theta}^*$ is written as $((\boldsymbol{\theta}_1^*)', \ldots, (\boldsymbol{\theta}_p^*)')'$. Noting that $\boldsymbol{\mu}_k^* = \mathbf{X}\boldsymbol{\beta}_k$, (13) simplifies to

$$\exp\left\{-\frac{1}{2}\left[\boldsymbol{\beta}_k'(\mathbf{X}'\mathbf{S}_{kk}\mathbf{X} + \frac{1}{\sigma^2}\mathbf{I})\boldsymbol{\beta}_k - 2\boldsymbol{\beta}_k'\left(\mathbf{X}'\mathbf{S}_{kk}\boldsymbol{\theta}_k^* + \mathbf{X}'\sum_{\ell\neq k}\mathbf{S}_{k\ell}(\boldsymbol{\theta}_\ell^* - \mathbf{X}\boldsymbol{\beta}_\ell)\right)\right]\right\}. \quad (14)$$

But, (14) is proportional to a multivariate normal distribution with covariance matrix given by

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}_k} \equiv (\mathbf{X}'\mathbf{S}_{kk}\mathbf{X} + \frac{1}{\sigma^2}\mathbf{I})^{-1}$$

and mean given by

$$\boldsymbol{\mu}_{\boldsymbol{\beta}_k} \equiv \boldsymbol{\Sigma}_{\boldsymbol{\beta}_k}[\mathbf{X}'\mathbf{S}_{kk}\boldsymbol{\theta}_k^* + \mathbf{X}'\sum_{\ell\neq k}\mathbf{S}_{k\ell}(\boldsymbol{\theta}_\ell^* - \mathbf{X}\boldsymbol{\beta}_\ell)].$$

Hence, sampling from this conditional distribution can be done directly.

It remains to derive the matrices $\mathbf{S}_{kl}$, $k, l = 1, \ldots, p$. If $k = l$, then for $i = 1, \ldots, n$, the ith diagonal element of $\mathbf{S}_{kk}$ is given by

$$m_{ik}^{1/2}V_{kk}m_{ik}^{1/2},$$

where m_{ik} denotes the kth diagonal element of $\mathbf{m}_i$ and V_{kl} denotes the (k, l)th element of $\mathbf{V}$. For the off-diagonal elements, let $\mathbf{U} \equiv (U_{k\ell}) = \mathbf{V}^{1/2}\mathbf{B}\mathbf{V}^{1/2}$. Then the off-diagonal elements are given by

$$m_{ik}^{1/2}U_{kk}m_{jk}^{1/2}$$

if $j \in N_i$, and zero otherwise. The diagonal elements of $\mathbf{S}_{kl}$, where $k \neq l$, are given by

$$m_{ik}^{1/2}V_{kl}m_{il}^{1/2}.$$

The off-diagonal elements of $\mathbf{S}_{kl}$, for $i < j$, are given by

$$m_{ik}^{1/2}U_{kl}m_{jl}^{1/2},$$

and, for $i > j$, are given by

$$m_{ik}^{1/2}U_{lk}m_{jl}^{1/2}.$$

4.1.2 Conditional Distribution of V

The conditional distribution of $\mathbf{V} \equiv \mathbf{\Gamma}^{-1}$ is proportional to

$$|\mathbf{V}|^{n/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)\right\} |\mathbf{V}|^{(\rho-p-1)/2} \exp\left\{-\frac{\rho}{2}\text{tr}(\mathbf{A}^{-1}\mathbf{V})\right\}. \tag{15}$$

In the case of spatial independence ($\mathbf{B} = \mathbf{0}$), (15) would simplify to another Wishart distribution. However, such is not the case here and the Metropolis-Hastings (M-H) algorithm with a Wishart proposal is used to generate realizations from (15); for details of the M-H algorithm, see Metropolis et al. (1953), Hastings (1970), and the overview in Gilks, Richardson and Spiegelhalter (1996). The precision parameter for the Wishart proposal density is chosen to ensure a sufficiently high acceptance rate and reasonable mixing for $\mathbf{V}$.

4.1.3 Conditional Distribution of B

The conditional distribution of $\mathbf{B}$ is proportional to,

$$|\mathbf{H}|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)\right\} \exp\left\{-(\mathbf{B}^v)'\mathbf{B}^v/\xi^2\right\}. \tag{16}$$

In the Gibbs step for $\mathbf{B}$, the M-H algorithm with a uniform proposal is used to generate realizations from (16) for each component of $B_{k\ell}$ conditional on the values of the other components. The order of components generated during this step is randomly selected. The upper and lower bounds of the uniform proposal density are again chosen to ensure a sufficiently high acceptance rate and reasonable mixing for the elements of $\mathbf{B}$.

4.1.4 Conditional Distribution of $\boldsymbol{\theta}_i$

The final conditional distribution involves the process parameters $\boldsymbol{\theta}$. This distribution is proportional to

$$\prod_{i=1}^{n}\prod_{k=1}^{p} \exp(-E_{ik}e^{\theta_{ik}})(E_{ik}e^{\theta_{ik}})^{Y_{ik}} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)'\mathbf{\Sigma}^{-1}(\boldsymbol{\theta}^v - \boldsymbol{\mu}^v)\right\}.$$

However, it is more beneficial to consider the conditional distributions of each $\boldsymbol{\theta}_i$, the collection of process parameters associated with each spatial location. From the definitions

given in Section 2.2, these distributions are proportional to

$$\prod_{k=1}^{p} \exp(-e^{\theta_{ik}} E_{ik})(e^{\theta_{ik}} E_{ik})^{Y_{ik}} \exp\left\{-\frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\mu}_i^*)' \mathbf{m}_i^{1/2} \mathbf{V} \mathbf{m}_i^{1/2} (\boldsymbol{\theta}_i - \boldsymbol{\mu}_i^*)\right\},$$

where

$$\begin{aligned} \boldsymbol{\mu}_i^* \equiv \boldsymbol{\mu}_i \quad &+ \quad \sum_{i<j} \mathbf{m}_i^{-1/2} \mathbf{V}^{-1/2} \mathbf{B} \mathbf{V}^{1/2} \mathbf{m}_j^{1/2} (\boldsymbol{\theta}_j - \boldsymbol{\mu}_j) I(j \in N_i) \\ &+ \quad \sum_{i>j} \mathbf{m}_i^{-1/2} \mathbf{V}^{-1/2} \mathbf{B}' \mathbf{V}^{1/2} \mathbf{m}_j^{1/2} (\boldsymbol{\theta}_j - \boldsymbol{\mu}_j) I(j \in N_i). \end{aligned}$$

Sampling from these distributions is accomplished via the M-H algorithm with a multivariate normal proposal. The covariance matrix on the proposal distribution is set to $\epsilon^2 \mathbf{I}_p$ with ϵ^2 chosen to ensure a sufficiently high acceptance rate and reasonable mixing.

4.2 Results

To illustrate statistical inference for environmental equity, based on Section 4.1, we extracted all of the census block groups within a 50 mile radius of St. James Parish, Louisiana (see Figure 1). Several of the block groups were removed, including those that were entirely covered by water, had no population, or were indicated to be populated entirely by persons on one or more civilian or military ships. After this screening, a total of $n = 1,982$ block groups were analyzed. Actual counts of whites and minorities were extracted and the expected counts $\{E_{ik}\}$ were computed using internal standardization (Mantel and Stark, 1968) based on age and poverty status. Neighbors were defined as any two block groups that shared a boundary.

To examine the impact of toxic facilities, data on the locations of facilities listed in the TRI were examined, and a covariate was constructed by placing a four-mile buffer around each TRI site and counting the number of facilities whose buffers intersect the block group. That is, let

$$X_i = \sum_t I(B(TRI_t, 4) \cap BG_i \neq \emptyset), \qquad i = 1, \ldots, 1982,$$

where $I(\cdot)$ is the indicator function, TRI_t denotes the location of the tth TRI site, BG_i denotes the spatial region corresponding to the ith block group, and $B(x, r)$ denotes a disk

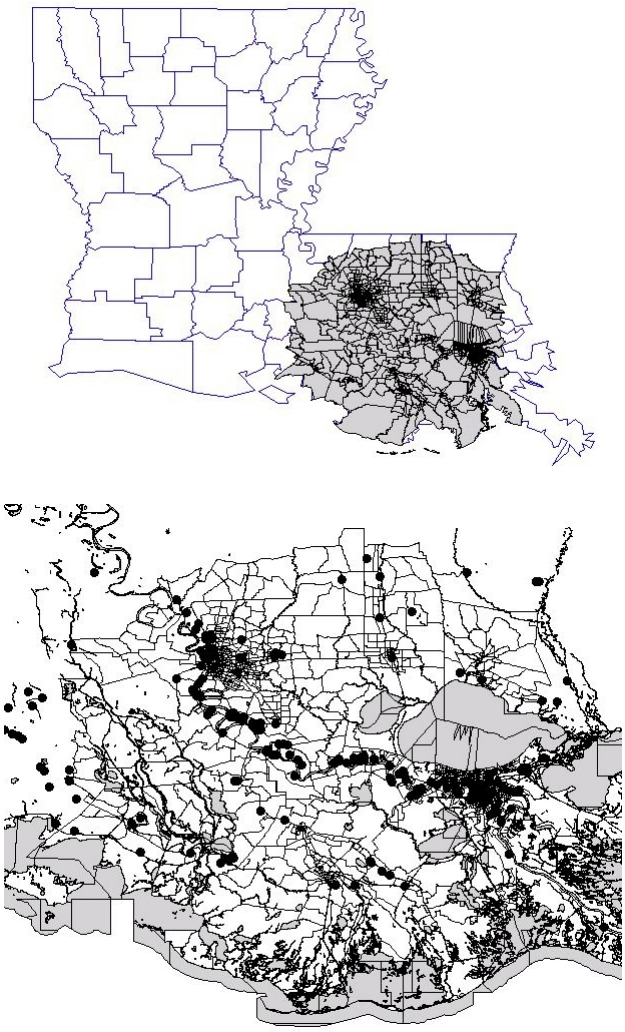

Figure 1: Top frame shows the state of Louisiana with highlighting of the block groups included in the example. Bottom frame shows detail of the highlighted block groups, including locations of the TRI facilities (solid dots) and water bodies (shaded).

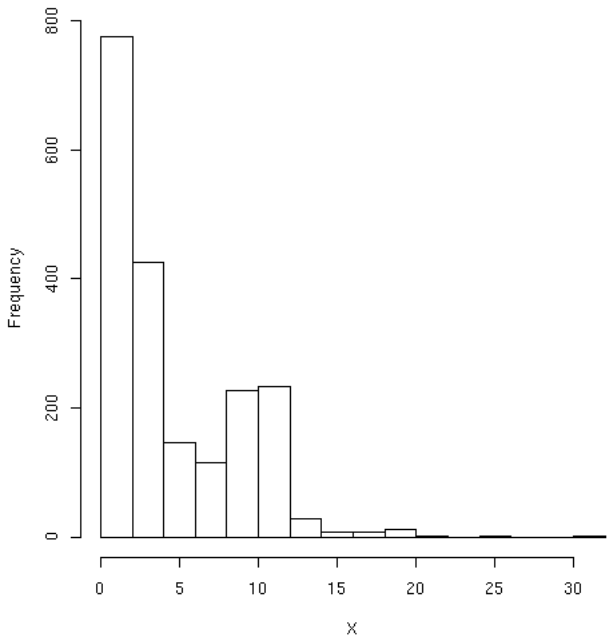

Figure 2: Histogram of the covariate X

centered at x with a radius of r miles. A plot of the locations of the TRI sites is given in Figure 1, while the (non-spatial and spatial) distribution of X is displayed in Figures 2 and 3.

The parameters for the priors on $\boldsymbol{\beta}$, $\mathbf{V}$, and $\mathbf{B}$ are set as follows. For the regression parameters $\boldsymbol{\beta}$, $\sigma^2 = 10$. For $\mathbf{V}$, $\mathbf{A} = \mathbf{I}$ and $\rho = 1.0 \times 10^6$. Finally, the prior parameter ξ for $\mathbf{B}$ is set as $\xi = 2.0 \times 10^{-4}$.

Starting values for the Gibbs sampler were obtained by setting $\theta_{ik} = \log((Y_{ik}+1)/E_{ik})$, for $i = 1, \ldots, n$ and $k = 1, \ldots, p$, where Y_{ik} and E_{ik} are the observed and expected counts. Using these simple estimates for $\boldsymbol{\theta}$, standard least-squares was used to obtain a starting value for the regression coefficients $\boldsymbol{\beta}$, and then a starting value for $\mathbf{V} = \boldsymbol{\Gamma}^{-1}$ was computed from the residuals. A starting value for $\mathbf{B}$ was computed from a coarse grid search to maximize (16) conditional on the other values. The MCMC was run for 10,000 iterations with visual cues suggesting convergence after about 5,000 iterations. Subsequently, four additional chains were run with starting values randomly chosen to be overdispersed relative to the posteriors estimated from the last 5,000 iterations of the initial run.

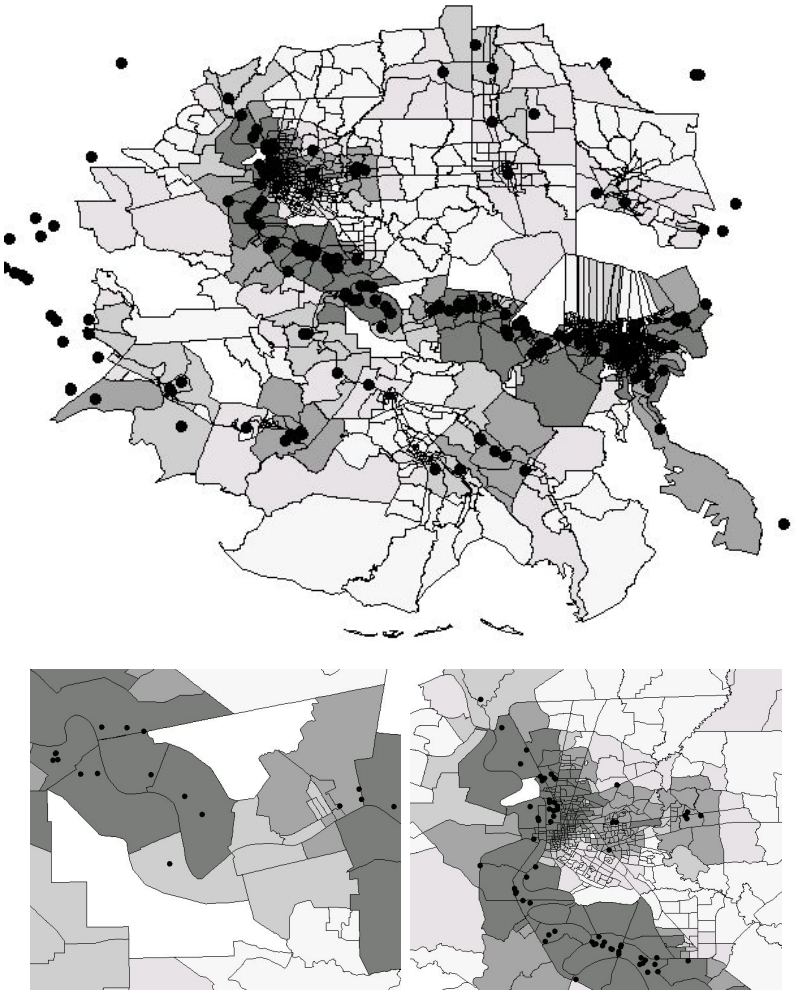

Figure 3: Maps showing the block-group regions shaded by the covariate, X, as defined in the text, with the TRI sites' locations superimposed. Detailed maps include St. James Parish and an area near Baton Rouge. The shading represents a grey scale from light to dark for $X = 0$, 1, 2-3, 4-6, and 7 or greater.

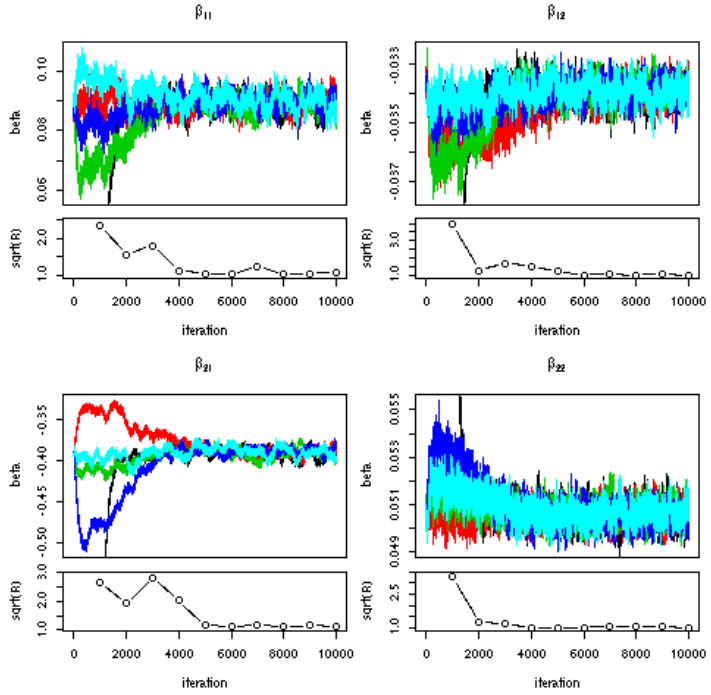

Figure 4: Convergence diagnostics for the regression parameters $\boldsymbol{\beta}$; the five different chains represent different starting values.

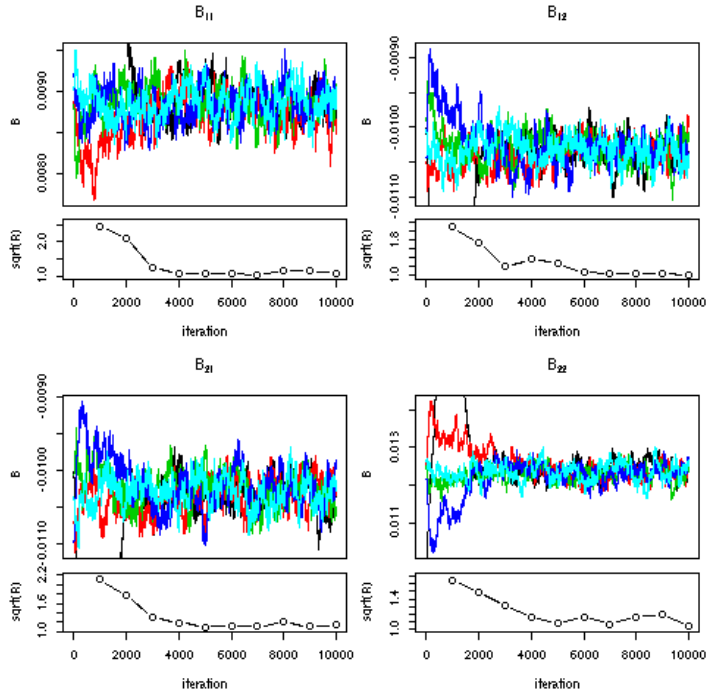

Figure 5: Convergence diagnostics for the spatial parameters $\mathbf{B}$; the five different chains represent different starting values.

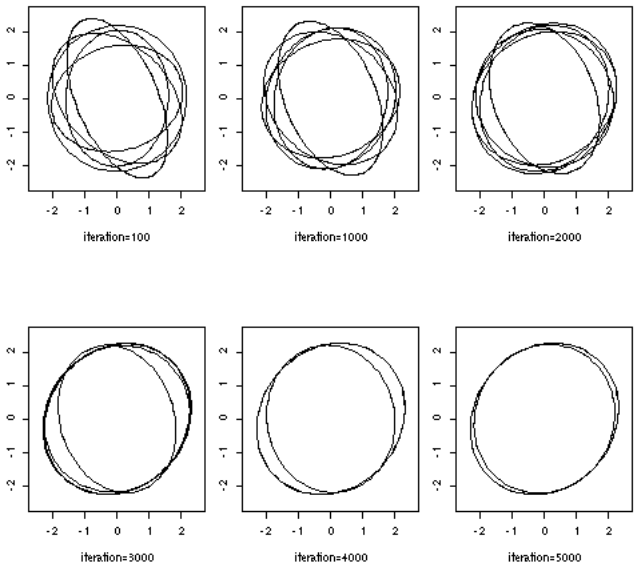

Figure 6: Ellipses representing $\mathbf{V} = \mathbf{\Gamma}^{-1}$ for various iterations of the five runs.

Convergence of the parameters of the model was monitored visually through trace plots as well as through numerical summaries, such as the $\sqrt{R}$-statistic of Gelman (1996). Examples of these plots are shown in Figures 4 and 5. After about 5,000 iterations, the values of the $\sqrt{R}$-statistic tend to one and are below the 1.2 value suggested by Gelman (1996). The convergence of $\mathbf{V}$ is demonstrated in Figure 6, where ellipses are shown representing the values of $\mathbf{V}$ for each run at various iterations. Again, convergence is clear after about 5,000 iterations.

Nearly all of the 3192 ($= 1982 \times 2$) θ_{ik}s also showed signs of convergence in trace plots and with $\sqrt{R}$-statistics near one after about 5,000 iterations. A small fraction of the θ_{ik}s had higher values of $\sqrt{R}$, although these values were associated with block groups having very small values of expected and observed counts. While the expected counts have a median of 432 and are as large as 4,000, there are a few block groups with expected counts representing just a few individuals (< 10). However, these few block groups do not appear to affect the overall statistical analysis.

One manner of addressing questions of environmental equity is to examine the resulting

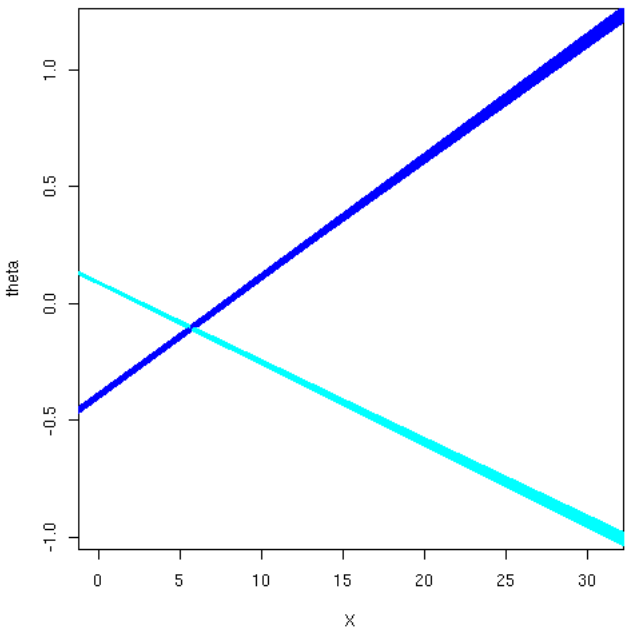

Figure 7: One hundred estimated regression lines of θ on X obtained from the posterior distribution of $\boldsymbol{\beta}$. The darker (lighter) lines with positive (negative) slope refers to the minority (white) population.

regression lines, linking the impact of the toxic facilities (variable X) and the demographic profile summarized by $\boldsymbol{\theta}$. Figure 7 shows estimated regression lines computed from 100 samples randomly selected from the combination of the last 5,000 iterations of the five chains. The clear difference between whites and minorities is striking. In the special case where $X = 0$, so that none of the four-mile buffers surrounding the TRI facilities intersect a particular block group, whites have greater than expected population counts while minorities have less than expected. Importantly, the regression coefficient for X (number of TRI facilities within 4 miles of a given block group) is positive for minorities, indicating that there are increasingly more than expected minorities as X increases. The opposite is true for whites, where the regression coefficient for X is negative.

Examining the impact of the spatial parameter $\mathbf{B}$, the conditional correlations from Section 3.2 are shown in Figure 8. The top frame shows a kernel estimate of the posterior distribution of the within-location correlation between the θs for whites and minorities computed from 5,000 random draws from a set of 25,000 realizations made up of the last

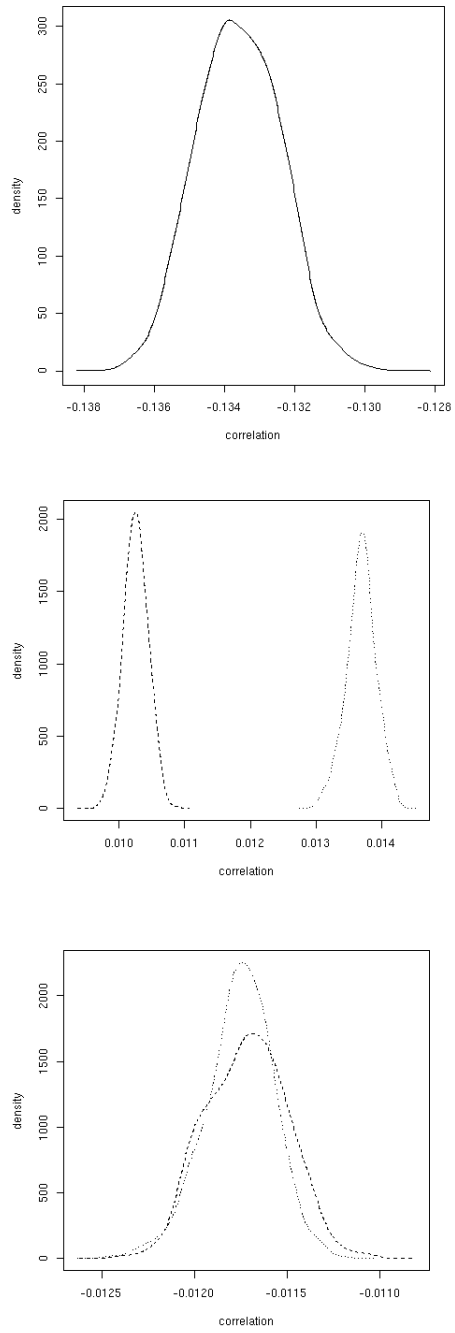

Figure 8: Top frame shows the within-location conditional correlation between the θs for whites and minorities. Middle frame shows the between-location conditional correlations for whites (dashed lines) and minorities (dotted lines). Bottom frame shows the between-location conditional cross-correlations.

5,000 iterations of the five chains. The middle frame shows kernel estimates of the posterior distribution of the between-location correlation between the θs for whites (dashed lines) and the θs for minorities (dotted lines) for two neighboring block groups. The spatial dependence appears to be stronger for the minorities than for the whites.

The bottom frame of Figure 8 shows the cross-correlations between the θs for whites and the θs for minorities at two neighboring block groups. While the model allows for asymmetric cross-correlations, the distributions of the two cross-correlations show a lot of overlap, although there is some evidence that conditional cross-correlation between a minority θ_{i2} and white θ_{j1} at a neighboring site is slightly stronger (more negative) than the conditional cross-correlation between θ_{i1} and θ_{j2}.

Maps of $\boldsymbol{\theta}$ for whites and minorities are shown in Figures 9 and 10, respectively. As expected from an examination of the regression lines, θs for whites are lower than expected near areas with increased numbers of TRI facilities, and vice versa for minorities. One feature of Figure 10 that seems to dominate is the preponderance of block groups near the Mississippi river (the so-called Industrial Corridor) with large θs for minorities, suggesting greater than expected numbers of minorities near this sensitive area known for its large number of toxic facilities.

5 An Extended CAMCAR Model

The covariance structure presented above allows for the possibility of asymmetric spatial cross-correlations, an important step beyond what has been previously considered. In this section, a more general structure for the covariance matrix is outlined that is based on rethinking the nature of the relationship between the data and the lattice. The typical approach to multivariate spatial modeling considers a two-dimensional spatial lattice with multiple measurements at each lattice point. However, a higher-dimensional lattice can be considered with only a single measurement at each lattice point, which allows us to develop a more flexible class of models for $\mathbf{H}$ and, subsequently, $\boldsymbol{\Sigma}$.

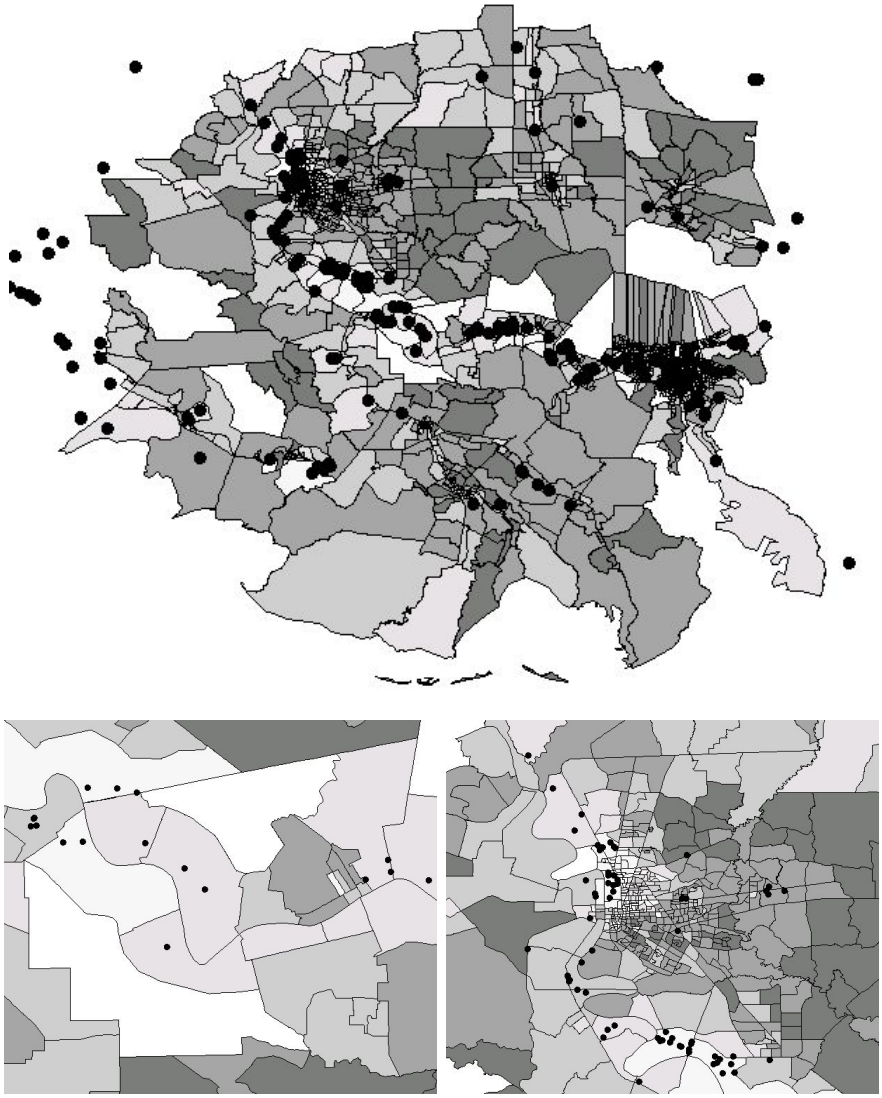

Figure 9: Maps of θs for whites. Detailed maps include St. James Parish and an area near Baton Rouge. Grey-scale shading is from light to dark based on five equal percentile breaks over all θs. Darker (lighter) colors indicate greater (less) than expected population counts.

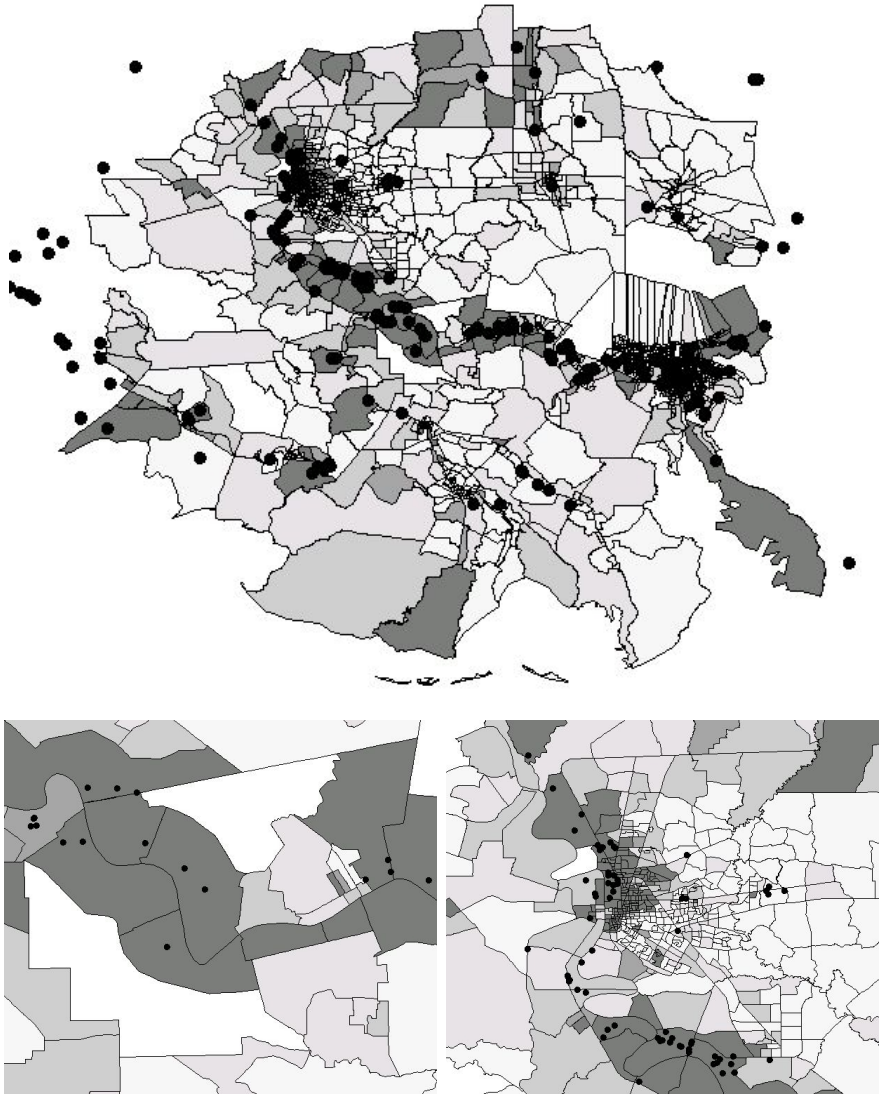

Figure 10: Maps of θs for minorities. Detailed maps include St. James Parish and an area near Baton Rouge. Grey-scale shading from light to dark based on five equal percentile breaks over all θs. Darker (lighter) colors indicate greater (less) than expected population counts.

For example, the joint covariance matrix from (5) can be written as

$$\boldsymbol{\Sigma} = \mathbf{M}\boldsymbol{\Pi}\mathbf{H}^{-1}\boldsymbol{\Pi}\mathbf{M}, \tag{17}$$

where

$$\mathbf{M} = \begin{bmatrix} \mathbf{m}_1^{-1/2} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{m}_n^{-1/2} \end{bmatrix}$$

and

$$\boldsymbol{\Pi} = \begin{bmatrix} \boldsymbol{\pi}^{1/2} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{\pi}^{1/2} \end{bmatrix},$$

with $\boldsymbol{\pi} = \text{diag}(\pi_1^2, \ldots, \pi_p^2)$. The variances $\{\pi_i^2\}$ are analogous to the diagonal elements of $\boldsymbol{\Gamma}$ in (4). Now the matrix $\mathbf{H}$ contains all the correlation information, including the within-location structure, which we write as

$$\mathbf{H} \;=\; \mathbf{I} - \sum_{k<l}\sum \boldsymbol{\rho}_{k\ell}\mathbf{G}_{k\ell} - \sum_{k=1}^{p}\sum_{\ell=1}^{p} \boldsymbol{\phi}_{k\ell}\mathbf{H}_{k\ell}.$$

The matrices $\{\mathbf{G}_{k\ell}\}$ control the relationships within a spatial location, while the matrices $\{\mathbf{H}_{k\ell}\}$ control the relationships across spatial locations.

For example, consider the simple lattice shown on the left frame of Figure 11. There are nine spatial locations and two measurements at each spatial location. This structure could equivalently be represented by the expanded three-dimensional lattice on the right frame of Figure 11, where each horizontal layer in the expanded lattice represents a separate measurement. In this case, the matrix $\mathbf{H}$ becomes,

$$\mathbf{H} \;=\; \mathbf{I} - \boldsymbol{\rho}_{12}\mathbf{G}_{12} - \sum_{k=1}^{2}\sum_{\ell=1}^{2} \boldsymbol{\phi}_{k\ell}\mathbf{H}_{k\ell}.$$

To account for the within-location relationship, indicated by the vertical dashed lines connecting the lattice layers in the right frame of Figure 11, $\mathbf{G}_{12}$ might have be modeled by,

$$\mathbf{G}_{12} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \otimes \mathbf{I}_n.$$

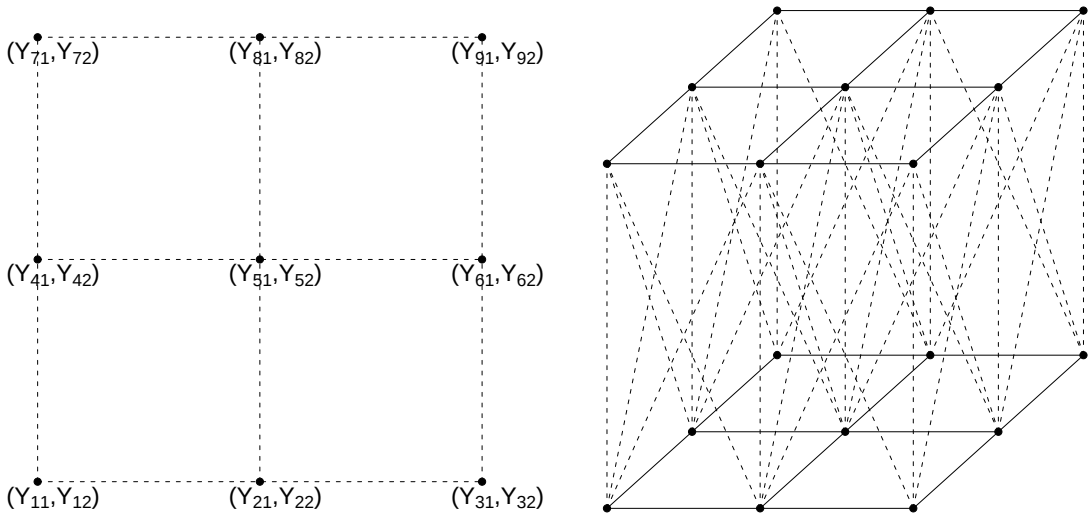

Figure 11: Example of a simple two-dimensional lattice with a bivariate measurement at each lattice point (left) and an expanded three-dimensional lattice with each layer of the lattice representing a different measurement (right).

The spatial correlations are modelled through the matrices $\{\mathbf{H}_{ij}\}$. To account for the within-variable relationships, indicated by the solid lines on each layer of the right frame of Figure 11, the matrices $\mathbf{H}_{11}$ and $\mathbf{H}_{22}$ might be modeled by,

$$\begin{aligned} \mathbf{H}_{11} &= \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \otimes \mathbf{C} \\ \mathbf{H}_{22} &= \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \otimes \mathbf{C}, \end{aligned}$$

where $\mathbf{C}$ is the incidence matrix derived from the neighborhood structure. To account for the cross-correlations, indicated by the diagonal dashed lines connecting the lattice layers in the right frame of Figure 11, the matrices $\mathbf{H}_{12}$ and $\mathbf{H}_{21}$ might be modeled by,

$$\begin{aligned} \mathbf{H}_{12} &= \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \otimes \mathbf{C}_U + \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \otimes \mathbf{C}'_U \\ \mathbf{H}_{21} &= \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} \otimes \mathbf{C}_U + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \otimes \mathbf{C}'_U, \end{aligned}$$

where $\mathbf{C}_U$ is the matrix containing the upper triangular elements of $\mathbf{C}$ with the lower triangular elements all set to zero.

As established previously, restrictions on $\mathbf{H}$ to yield a positive-definite covariance matrix can be determined through a diagonal-dominance argument. Replacing (5) with (17) in the hierarchical model from the previous section could be accomplished straightforwardly, with slight modifications to the priors. For example, the priors on the diagonal elements of $\boldsymbol{\pi}$ could easily be taken as gamma densities. A prior on the correlation parameters, the $\{\rho_{k\ell}\}$ and the $\{\phi_{k\ell}\}$, could be specified in a fashion similar to prior specification of $\mathbf{B}$ in the previous section, namely proportional to $\exp(-\rho_{k\ell}^2/\xi^2)$ and $\exp(-\phi_{k\ell}^2/\xi^2)$, for some small value of ξ.

6 Concluding Remarks

We have introduced a flexible spatial model for multivariate lattice data that can incorporate covariates as well as a very general form of spatial dependence between the variables measured at different spatial locations. This model, which we call the CAMCAR model, incorporates the precision of the data, especially for rates, in such a way as to allow for ease in interpretation of the spatial-dependence parameters. Based on the framework of a MRF for multivariate processes and incorporated within hierarchical statistical models, this work opens the door for the analysis of complex problems, including those often seen in the environmental arena.

In particular, a hierarchical model incorporating the CAMCAR model allows us to use Bayesian inference based on MCMC to assess environmental equity in a region of southern Louisiana. In studying the relationship between white and minority populations and the location of TRI facilities, the analysis suggests some disparity between whites and minorities, but it cannot ascertain causation, only association. Further, the maps in Figures 9 and 10 represent an important tool for use in siting decisions for future facilities.

The model CAMCAR presented in Sections 2.2 and 3 is general in the specification of covariance parameters and the types of dependence that are possible to model. Bayesian model selection critera such as those presented in Spiegelhalter et al. (2002), would be useful to investigate for assessing the fit of the model. The extended CAMCAR model

presented in Section 5 represents an even broader class of models and could be generalized further to allow inference based on different layers of data measured on different types of lattices (e.g., zip codes and census tracts).

Acknowledgment

The research of the first author was supported in part by the National Science Foundation under grant DMS 9815344 and a Faculty Grant Award from the University of Colorado at Denver. The research of the second author was supported by the US Environmental Protection Agency under Assistance Agreement R827257-01-0 and by the Office of Naval Research under grant N00014-02-1-0052.

References

Besag, J.E. (1974), "Spatial interaction and the statistical analysis of lattice systems (with discussion)," *Journal of the Royal Statistical Society, Series B*, **35**, 192-236.

Billheimer, D., Cardoso, T., Freeman, E., Guttorp, P., Ko, H., and Silkey, M. (1997), "Natural variability of benthic species composition in the Delaware Bay," *Environmental and Ecological Statistics*, **4**, 95-115.

Carlin, B.P. and Banerjee, S. (2003), "Hierarchical multivariate CAR models for spatio-temporally correlated data," In *Bayesian Statistics 7* (eds. J.M. Bernardo et al.), 45-63, Oxford: Oxford University Press.

Carlin, B.P. and Xia, H. (1999), "Assessing environmental justice using Bayesian hierarchical models: two case studies," *Journal of Exposure Analysis and Environmental Epidemiology*, **9**, 66-78.

Cressie, N. (1993), *Statistics for Spatial Data, rev. edn.*, New York: John Wiley.

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

Gelfand, A.E., Hills, S.E., Racine-Poon, A., and Smith, A.F.M. (1990), "Illustration of Bayesian inference in normal data models using Gibbs sampling," *Journal of the American Statistical Association*, **85**, 972-985.

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

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

Gelman, A. (1996), "Inference and monitoring convergence," In *Markov Chain Monte Carlo in Practice* (eds. W.R. Gilks, S. Richardson, and D.J. Spiegelhalter), 131-144, London: Chapman and Hall.

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

General Accounting Office (1983), "Siting of hazardous waste landfills and their correlation with racial and economic status of surrounding communities," Washington, D.C.: General Accounting Office.

Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996), "Introducing Markov chain Monte Carlo," In *Markov Chain Monte Carlo in Practice* (eds. W.R. Gilks, S. Richardson, and D.J. Spiegelhalter), 1-19, London: Chapman and Hall.

Hastings, W.K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, **57**, 97-109.

Horn, R.A. and Johnson, C.R. (1990), *Matrix Analysis*, Cambridge, U.K.: Cambridge University Press.

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

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979), *Multivariate Analysis*, New York: Academic Press.

Mantel, N. and Stark, C. (1968), "Computation of indirect adjusted rates in the presence of confounding," *Biometrics*, **24**, 997-1005.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., and Teller, E. (1953), "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, **21**, 1087-1091.

Pettitt, A.N., Weir, I.S., and Hart, A.G. (2002), "A conditional autoregressive Gaussian process for irregularly spaced multivariate data with application to modeling large sets of binary data," *Statistics and Computing*, **12**, 353-367.

Royle, A.M. and Berliner, L.M. (1999), "A hierarchical approach to multivariate spatial modeling and prediction," *Journal of Agricultural, Biological, and Environmental Statistics*, **4**, 29-56.

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

Stern, H.S. and Cressie, N. (1999), "Inference for extremes in disease mapping," In *Disease Mapping and Risk Assessment for Public Health* (eds. A. Lawson et al.), 63-84, Chichester: Wiley.

United Church of Christ Commission for Racial Justice (1987), *Toxic Wastes and Race in the United States: A National Report on the Racial and Socioeconomic Characteristics of Communities with Hazardous Waste Sites*, New York: United Church of Christ.

Ver Hoef, J.M., Cressie, N., and Barry, R.M. (2003), "Flexible spatial models for kriging and cokriging using moving averages and the Fast Fourier Transform (FFT)," *Journal of Computational and Graphical Statistics*, to appear.

Wackernagel, H. (1998), *Multivariate Geostatistics, 2nd edn.*, Berlin: Springer.

Waller, L.A. and Conlon, E.M. (2000), "LULU? NIMBY! Statistical issues in assessments of environmental justice," *STATS*, **28**, 3-13.

Waller, L.A., Louis, T.A., and Carlin, B.P. (1999), "Environmental justice and statistical summaries of differences in exposure distributions," *Journal of Exposure Analysis and Environmental Epidemiology*, **9**, 56-65.