 DOI: 10.1111/j.1541-0420.2005.00359.x

Generalized Hierarchical Multivariate CAR Models for Areal Data

Xiaoping Jin, Bradley P. Carlin,* and Sudipto Banerjee

Division of Biostatistics, School of Public Health, University of Minnesota,
Mayo Mail Code 303, Minneapolis, Minnesota 55455-0392, U.S.A.
**email:* brad@biostat.umn.edu

Summary. In the fields of medicine and public health, a common application of areal data models is the study of geographical patterns of disease. When we have several measurements recorded at each spatial location (for example, information on $p \geq 2$ diseases from the same population groups or regions), we need to consider *multivariate* areal data models in order to handle the dependence among the multivariate components as well as the spatial dependence between sites. In this article, we propose a flexible new class of generalized multivariate conditionally autoregressive (GMCAR) models for areal data, and show how it enriches the MCAR class. Our approach differs from earlier ones in that it directly specifies the joint distribution for a multivariate Markov random field (MRF) through the specification of simpler conditional and marginal models. This in turn leads to a significant reduction in the computational burden in hierarchical spatial random effect modeling, where posterior summaries are computed using Markov chain Monte Carlo (MCMC). We compare our approach with existing MCAR models in the literature via simulation, using average mean square error (AMSE) and a convenient hierarchical model selection criterion, the deviance information criterion (DIC; Spiegelhalter et al., 2002, *Journal of the Royal Statistical Society, Series B* **64,** 583–639). Finally, we offer a real-data application of our proposed GMCAR approach that models lung and esophagus cancer death rates during 1991–1998 in Minnesota counties.

Key words: Areal data; Conditionally autoregressive (CAR) model; Hierarchical Bayesian model; Markov chain Monte Carlo (MCMC) simulation; Multivariate data; Spatial statistics.

1. Introduction

The analysis of spatially referenced data has been an increasingly active area of both methodological and applied statistical research. Sophisticated computer programs known as geographic information systems (GISs) have allowed health science databases to incorporate geographical information about the units being studied. Such databases have in turn generated interest among statisticians to develop and analyze models that can account for spatial clustering and variation. For data collected over geographic regions (*areal* data) such as counties, census tracts, zip codes, and so on, the most commonly used are conditionally autoregressive (CAR) specifications, pioneered by Besag (1974). CAR distributions are sometimes used as the likelihood for the observations themselves in one-stage models, or as the distribution of the random effects in the mean structure in hierarchical models. In the fields of medicine and public health, a common application of such models is the study of regional patterns of disease. In the United States, publicly available data on precise locations of disease cases are fairly uncommon due to strict confidentiality regulations. Summaries of disease at a regional level, however, are often relatively easy to obtain.

CAR models are most appropriate in the univariate case, as when mapping a single disease. When we have multivariate areal data (say, information on $p \geq 2$ diseases over the same regions), an obvious first choice would be to use p separate univariate CAR models. But because a number of diseases may share the same set of (spatially distributed) risk factors, or the presence of one disease might encourage or inhibit the presence of another over a region, we may need a multivariate areal model to properly analyze this kind of data. This will permit modeling of dependence among the multivariate components while maintaining spatial dependence between sites.

Several multivariate areal models have been proposed to date. Mardia (1988) described the theoretical background for multivariate normal Markov random field (MRF) specifications. Billheimer et al. (1997) developed a hierarchical statistical model for compositional monitoring data utilizing a multivariate MRF in a state-space setting. Kim, Sun, and Tsutakawa (2001) presented a "twofold CAR" model for counts of two different diseases over each areal unit. Sain and Cressie (2002) discussed a multiobjective version of the CAR model that allows for flexible modeling of the spatial dependence structure, the cross-correlations in particular. Most recently, Carlin and Banerjee (2003) and Gelfand and Vounatsou (2003) developed multivariate CAR (MCAR) models for hierarchical modeling based on the family of Mardia (1988).

In this article, we introduce a new flexible class of generalized multivariate CAR (GMCAR) models for areal data, and show how it enriches the existing MCAR class. Reminiscent of the approach of Royle and Berliner (1999) in the case of

geostatistical (point-referenced spatial) data, our method directly specifies the joint distribution for a multivariate MRF through the specification of simpler conditional and marginal models. We then employ these GMCAR distributions as specifications for second-stage random effects in hierarchical areal data models. In particular, we consider modeling the death rates from lung and esophagus cancers in the years from 1991 to 1998 in Minnesota counties, a setting in which association would be expected both within and across the areal units.

The format of our article is as follows. In Section 2, we briefly review the various existing CAR and MCAR models, and point out the advantages and disadvantages of each. Section 3 introduces the GMCAR class, while Section 4 compares it with the existing MCAR models in terms of average mean square error (AMSE) and deviance information criterion (DIC; Spiegelhalter et al., 2002) score via simulation. Section 5 then applies the GMCAR to our illustrative data set. Finally, Section 6 summarizes our findings and suggests avenues for future research in this burgeoning area.

2. Overview of Univariate and Multivariate CAR Modeling

2.1 *Univariate CAR Modeling*

A fundamental result in the understanding of CAR models is due to Besag (1974). Consider a univariate spatially random variable ϕ_i observed at n areal locations, and define $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_n)'$. Under the MRF assumption, we specify the n *full conditional* distributions as

$$p\big(\phi_i \,\big|\, \phi_j, j \neq i, \tau_i^{-1}\big) = N\left(\alpha \sum_{i\sim j} b_{ij}\phi_j, \tau_i^{-1}\right), \quad i, j = 1, \ldots, n, \tag{1}$$

where $i \sim j$ denotes that region j is a *neighbor* (typically defined in terms of spatial adjacency) of region i. Now from the Hammersley–Clifford Theorem and Brook's Lemma (see, e.g., Banerjee, Carlin, and Gelfand, 2004, Section 3.2), the full conditional distributions in (1) uniquely determine the joint distribution,

$$\boldsymbol{\phi} \sim N\big(\mathbf{0}, [D_\tau(I - \alpha B)]^{-1}\big), \tag{2}$$

where B is an $n \times n$ matrix with $b_{ii} = 0$, and $D_\tau = \mathrm{Diag}(\tau_i)$; usually we assume that $D_\tau = \tau D$, where D is an $n \times n$ diagonal matrix. Finally, α is a smoothing parameter, and is often interpreted as measuring spatial association. Notice $\alpha = 0$ corresponds to an independent model, but it is important *not* to view α as a *correlation* parameter. That is, α controls spatial dependence, and its value lies between 0 and 1, but it cannot be interpreted as a correlation coefficient in the usual sense (see, e.g., Wall, 2004 and Section 5).

From the CAR formulation (2), we can choose α, D, and B to obtain various CAR model structures. The most popular CAR implementation (Besag, York, and Mollié, 1991) is the *pairwise difference* formulation, also known as the *intrinsic autoregressive* (IAR) model. In this structure, we set the smoothing parameter $\alpha = 1$. We also typically take $D = \mathrm{Diag}(m_i)$, where m_i is the number of neighbors of region i, and $B = D^{-1}W$, where W denotes the adjacency matrix of the map (i.e., $w_{ii} = 0$, and $w_{ii'} = 1$ if $i \sim i'$, and 0 otherwise). B is called the *scaled adjacency matrix* in this case. Formulation (2) then becomes

$$\boldsymbol{\phi} \sim N(\mathbf{0}, [\tau(D - W)]^{-1}). \tag{3}$$

Model (3) is simple and easy to fit, but has two major drawbacks. First, $\tau(D - W)$ is singular, and thus (3) is improper. Second, the IAR model (3) contains no parameter to control the strength of spatial dependence among regions.

To overcome these difficulties, several authors prefer $\alpha < 1$. For example, Cressie (1993) assumes $D = I_{n\times n}$ and $B = W$ in the CAR formulation (2), and points out that if $\alpha \in (\lambda_{\min}^{-1}, \lambda_{\max}^{-1})$, where $\lambda_{\min}$ and $\lambda_{\max}$ are the minimum and maximum eigenvalue of the adjacency matrix, respectively, a proper joint distribution results. Carlin and Banerjee (2003) avoid the calculation of eigenvalues by using the scaled adjacency matrix B, and show that taking $|\alpha| < 1$ ensures this model's propriety.

2.2 *Multivariate CAR Modeling*

Most multivariate CAR models are members of the family developed by Mardia (1988). Analogous to the univariate case, the joint distribution is derived from the full conditional distributions. Under the MRF assumption, we can specify these conditional distributions as

$$p\big(\boldsymbol{\upsilon}_i \,\big|\, \boldsymbol{\upsilon}_{j\neq i}, \boldsymbol{\Gamma}_i^{-1}\big) = N\left(R_i \sum_{i\sim j} B_{ij}\boldsymbol{\upsilon}_j, \boldsymbol{\Gamma}_i^{-1}\right), \quad i, j = 1, \ldots, n, \tag{4}$$

where $\boldsymbol{\upsilon}_i = (\phi_{i1}, \phi_{i2}, \ldots, \phi_{ip})'$ is a p-dimensional vector, and $\boldsymbol{\Gamma}_i$, R_i, and B_{ij} are $p \times p$ matrices. For example, this model might be appropriate for a data set on p types of cancer over n counties. Mardia (1988) proved that the full conditional distributions in (4) uniquely determine the joint distribution

$$\boldsymbol{\upsilon} \sim N\big(\mathbf{0}, [\boldsymbol{\Gamma}(I - B_R)]^{-1}\big), \tag{5}$$

where $\boldsymbol{\upsilon}' = (\boldsymbol{\upsilon}_1', \boldsymbol{\upsilon}_2', \ldots, \boldsymbol{\upsilon}_n')$, B_R is $np \times np$ with $(B_R)_{ij} = R_i B_{ij}$, $(B_R)_{ii} = 0$, and $\boldsymbol{\Gamma}$ is an $np \times np$ block diagonal matrix with $p \times p$ diagonal entries $\boldsymbol{\Gamma}_i$.

From the MCAR formulation (5), we can choose different $\boldsymbol{\Gamma}$ and B_R matrices to obtain different MCAR model structures. But to obtain a proper joint distribution (5), we need to make sure that $\boldsymbol{\Gamma}(I - B_R)$ is a positive definite and symmetric matrix. Unfortunately, establishing these conditions can be difficult in general cases. To simplify the formulation, we may first assume that $R_i = \alpha I_{p\times p}$ for $i = 1, \ldots, n$ (where α is again called a smoothing parameter), and $\boldsymbol{\Gamma} = D \otimes \Lambda$. Under these assumptions, (5) becomes

$$\boldsymbol{\upsilon} \sim N(\mathbf{0}, [(D(I - \alpha B)) \otimes \Lambda]^{-1}), \tag{6}$$

where Λ is a $p \times p$ positive definite and symmetric matrix, and the matrices D and B are defined as in Section 2.1. The precision matrix in (6) is the Kronecker product of the univariate CAR form and Λ, and thus the covariance matrix in (6) is positive definite as long as Λ is positive definite and the univariate CAR distribution is valid. Now we can apply all of the univariate CAR structures described in Section 2.1 to obtain different MCAR models. For example, we can generalize the IAR model (3) to the multivariate case simply by setting $\alpha = 1$ above. Alternatively, in (6) we can

assume that $D = \text{Diag}(m_i)$, use the scaled adjacency matrix $B = D^{-1}W$, and take $\alpha \in (-1, 1)$. This model is denoted as MCAR(α, Λ) in Carlin and Banerjee (2003) and Gelfand and Vounatsou (2003). The MCAR(α, Λ) formulation is thus

$$\boldsymbol{v} \sim N(\mathbf{0}, [(D - \alpha W) \otimes \Lambda]^{-1}). \tag{7}$$

All of the above MCAR models are generalized from univariate CAR models under the assumption that $R_i = \alpha I_{p\times p}$, $i = 1, \ldots, n$, and can be used for any dimension p. The positive definiteness condition for $\mathbf{\Gamma}(I - B_R)$ in (5) is then easy to verify, and its Kronecker product form simplifies the calculations, especially matrix inversion and determinant evaluation. But the assumption of a common R_i for all $i = 1, \ldots, n$ may well be too strong in some cases.

To explore this idea, suppose $p = 2$ (e.g., two cancers in each county), and define $\boldsymbol{\phi}_1' = (\phi_{11}, \ldots, \phi_{n1})$ and $\boldsymbol{\phi}_2' = (\phi_{12}, \ldots, \phi_{n2})$. Then, the MCAR formulation (7) can be written as

$$\begin{pmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} (D - \alpha W)\Lambda_{11} & (D - \alpha W)\Lambda_{12} \\ (D - \alpha W)\Lambda_{12} & (D - \alpha W)\Lambda_{22} \end{pmatrix}^{-1} \right), \tag{8}$$

where Λ_{ij}, $i = 1, 2$, $j = 1, 2$ are the elements of Λ. More generally, we may need three different α_i parameters in (8) to explain the correlation between the two types of cancer and across the counties that neighbor each other (Kim et al., 2001). The covariance matrix Σ would then be revised to

$$\Sigma = \begin{pmatrix} (D - \alpha_1 W)\Lambda_{11} & (D - \alpha_3 W)\Lambda_{12} \\ (D - \alpha_3 W)\Lambda_{12} & (D - \alpha_2 W)\Lambda_{22} \end{pmatrix}^{-1}, \tag{9}$$

where α_1 and α_2 are the smoothing parameters for the two cancer types, and α_3 is the "bridging" or "linking" parameter associating ϕ_{i1} with ϕ_{j2}, $i \neq j$. Unfortunately, with this general covariance matrix, it is difficult to check the conditions guaranteeing positive definiteness, since they depend on the unknown Λ matrix. This makes model fitting hard to implement via Markov chain Monte Carlo (MCMC).

Carlin and Banerjee (2003) and Gelfand and Vounatsou (2003) generalize the basic MCAR model by allowing two different α parameters (say, α_1 and α_2), and denote this model as MCAR(α_1, α_2, Λ). They write the precision matrix Σ^{-1} as

$$\begin{pmatrix} R_1'R_1\Lambda_{11} & R_1'R_2\Lambda_{12} \\ R_2'R_1\Lambda_{12} & R_2'R_2\Lambda_{22} \end{pmatrix} = \begin{pmatrix} R_1' & 0 \\ 0 & R_2' \end{pmatrix} (\Lambda \otimes I_{n\times n}) \begin{pmatrix} R_1 & 0 \\ 0 & R_2 \end{pmatrix}, \tag{10}$$

where $R_k'R_k = D - \alpha_k W$, $k = 1, 2$. Carlin and Banerjee (2003) take R_k as the Cholesky decomposition of $D - \alpha_k W$ so that R_k is an upper-triangular matrix. Gelfand and Vounatsou (2003) instead recommend a spectral decomposition, that is, $R_k = \text{Diag}(1 - \alpha_k\lambda_i)^{1/2}P'D^{1/2}P$, where the λ_i are the eigenvalues of $D^{-1/2}WD^{-1/2}$ and P is an orthogonal matrix with the corresponding eigenvectors as its columns. Either way, this generalization of the MCAR model permits different smoothing parameters α_k for each k (e.g., different strengths of spatial correlation for each type of cancer). As before, Λ controls the nonspatial correlation among cancers at any given location.

The conditions for the covariance matrix to be positive definite are easy to find as long as the Cholesky or spectral decompositions exist and Λ is positive definite. For the $p = 2$ case, these reduce to $|\alpha_1| < 1$ and $|\alpha_2| < 1$. The spectral approach may be better in terms of Bayesian computing, since it does not require the calculation of a Cholesky decomposition at each MCMC iteration, a substantial burden particularly for a data set with many spatial regions. Neither of these MCAR structures allows a smoothing parameter α on the off-diagonal of the precision matrix as in (9); we cannot model the off-diagonal, since it is determined by the diagonal. Finally, because the decomposition of $D - \alpha_k W$ is not unique, we can have different MCAR models with the covariance structure (10).

Kim et al. (2001) proposed a multivariate CAR model in the bivariate ($p = 2$) case, which they dub the "twofold conditionally autoregressive" model, and which we notate as 2fCAR(α_0, α_1, α_2, α_3, τ_1, τ_2). They specify the moments of the full conditional distributions as

$$E(\phi_{ik} \mid \phi_{il}, \phi_{jk}, \phi_{jl}) = \frac{1}{2m_i + 1}\left(\alpha_k \sum_{j\sim i} \phi_{jk} + \alpha_3 \sqrt{\frac{\tau_l}{\tau_k}} \sum_{j\sim i} \phi_{jl} + \alpha_0 \sqrt{\frac{\tau_l}{\tau_k}} \phi_{il} \right)$$

and

$$\text{Var}(\phi_{ik} \mid \phi_{il}, \phi_{jk}, \phi_{jl}) = \frac{\tau_k^{-1}}{2m_i + 1}, \qquad i, j = 1, \ldots, n, \quad l, k = 1, 2, \quad l \neq k,$$

where $j \sim i$ again means that region j is a neighbor of region i. Adding the Gaussian MRF structure, they derive the joint distribution arising from these full conditional distributions as

$$\begin{pmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} (2D + I - \alpha_1 W)\tau_1 & -(\alpha_0 I + \alpha_3 W)\sqrt{\tau_1\tau_2} \\ -(\alpha_0 I + \alpha_3 W)\sqrt{\tau_1\tau_2} & (2D + I - \alpha_2 W)\tau_2 \end{pmatrix}^{-1} \right), \tag{11}$$

where again $\boldsymbol{\phi}_1' = (\phi_{11}, \ldots, \phi_{n1})$, $\boldsymbol{\phi}_2' = (\phi_{12}, \ldots, \phi_{n2})$, $D = \text{Diag}(m_i)$, and W is the adjacency matrix. This model has the same number of parameters in the covariance structure (six) as the general formulation (9) in the bivariate case, so they are related to each other. In (11), α_1 and α_2 are the smoothing parameters, while α_0 and α_3 are the bridging parameters associating ϕ_{i1} with ϕ_{i2} and ϕ_{j2}, $j \neq i$, respectively. Unfortunately, this MCAR model is only designed for the bivariate case ($p = 2$), and seems difficult to generalize to higher dimensions. Also, under this approach it is hard to find conditions that guarantee a positive definite covariance matrix in (11). The conditions $|\alpha_l| < 1, l = 0, 1, 2, 3$ given by Kim et al. (2001) are sufficient but not necessary, and may be overly restrictive for some data sets since they restrict the correlation of ϕ_{i1} with ϕ_{i2} and ϕ_{j2}, $j \neq i$. Finally, this generalization comes at a significant price in terms of computing, since it requires many matrix multiplications, determinant evaluations, and inverses at each MCMC iteration, so can be very time-consuming even when working on a relatively small spatial domain.

3. A Generalized MCAR Model

As mentioned in the previous section, it is often difficult to specify a valid joint covariance matrix for multivariate areal data models. Most of the MCAR models described in Section 2.2 work with the precision matrix instead of with the covariance matrix directly, which makes for generally faster computing but also obfuscates interpretation of the results. To avoid this difficulty, we introduce a new approach for multivariate areal data in which we directly specify the joint distribution for a multivariate spatial process through the specification of simpler conditional and marginal forms.

To illustrate our approach, we again start with the case of bivariate areal data ($p = 2$). We now assume the joint distribution of $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ is

$$\begin{pmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix} \right),$$

where the Σ_{kl}, $k, l = 1, 2$ are $n \times n$ covariance matrices. From standard multivariate normal theory, we have $E(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2) = \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\phi}_2$ and $\mathrm{Var}(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2) = \Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'$. Now writing $A = \Sigma_{12}\Sigma_{22}^{-1}$, we can rewrite the joint distribution of $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ as

$$\begin{pmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Sigma_{11\cdot 2} + A\Sigma_{22}A' & A\Sigma_{22} \\ (A\Sigma_{22})' & \Sigma_{22} \end{pmatrix} \right). \quad (12)$$

According to Harville (1997, Corollary 14.8.5), the conditions that ensure the propriety of (12) are that Σ_{22} and $\Sigma_{11\cdot 2}$ are positive definite. Because $\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2 \sim N(A\boldsymbol{\phi}_2, \Sigma_{11\cdot 2})$ and $\boldsymbol{\phi}_2 \sim N(0, \Sigma_{22})$, we can construct $p(\boldsymbol{\phi}) = p(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2)p(\boldsymbol{\phi}_2)$, where $\boldsymbol{\phi}' = (\boldsymbol{\phi}_1', \boldsymbol{\phi}_2')$. To write the joint distribution of $\boldsymbol{\phi}$, thus, we need to specify the matrices $\Sigma_{11\cdot 2}$, Σ_{22}, and A.

Following the univariate CAR structure described in Section 2.1, suppose we assume that the conditional distribution for $\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2$ is $\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2 \sim N(A\boldsymbol{\phi}_2, [(D - \alpha_1 W)\ \tau_1]^{-1})$, and the marginal distribution of $\boldsymbol{\phi}_2$ is $\boldsymbol{\phi}_2 \sim N(\mathbf{0}, [(D - \alpha_2 W)\ \tau_2]^{-1})$, where α_1 is the smoothing parameter associated with the conditional distribution of $\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2$, α_2 is similar for the marginal distribution of $\boldsymbol{\phi}_2$, and τ_1 and τ_2 scale the precision of $\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2$ and $\boldsymbol{\phi}_2$, respectively. The induced joint distribution will always be proper as long as these two CAR distributions are valid, so the positive definiteness of the covariance matrix in (12) is easily verified. Let $D = \mathrm{Diag}(m_i)$ and W again be the adjacency matrix. The positive definiteness conditions then require only that $|\alpha_1| < 1$ and $|\alpha_2| < 1$. We typically further restrict to $0 < \alpha_1 < 1$ and $0 < \alpha_2 < 1$ to avoid negative spatial autocorrelation.

Regarding the A matrix, since $E(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2) = A\boldsymbol{\phi}_2$, we assume its elements are of the form:

$$a_{ij} = \begin{cases} \eta_0 & \text{if } j = i, \\ \eta_1 & \text{if } j \in N_i \text{ (i.e., if region } j \text{ is a neighbor of region } i), \\ 0 & \text{otherwise.} \end{cases}$$

Thus, $A = \eta_0 I + \eta_1 W$ and $E(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2) = (\eta_0 I + \eta_1\ W)\ \boldsymbol{\phi}_2$. Here, η_0 and η_1 are the bridging parameters associating ϕ_{i1} with ϕ_{i2} and ϕ_{j2}, $j \neq i$, respectively, similar to α_0 and α_3 in the twofold CAR model (11). (We could easily generalize our model by augmenting A with another bridging parameter η_2 associated with the *second-order* neighbors [neighbors of neighbors] in each region, but we do not pursue this generalization here.) Under these assumptions, the covariance matrix in the joint distribution (12) becomes

$$\begin{pmatrix} [\tau_1(D - \alpha_1 W)]^{-1} + (\eta_0 I + \eta_1 W)[\tau_2(D - \alpha_2 W)]^{-1}(\eta_0 I + \eta_1 W) & (\eta_0 I + \eta_1 W)[\tau_2(D - \alpha_2 W)]^{-1} \\ [\tau_2(D - \alpha_2 W)]^{-1}(\eta_0 I + \eta_1 W) & [\tau_2(D - \alpha_2 W)]^{-1} \end{pmatrix}. \quad (13)$$

We denote this new model by GMCAR(α_1, α_2, η_0, η_1, τ_1, τ_2). This bivariate GMCAR model has the same number of parameters as the twofold CAR model (11), and one more parameter than the MCAR(α_1, α_2, Λ) model (10).

Many MCAR models we have already encountered emerge as special cases of our GMCAR($\alpha_1, \alpha_2, \eta_0, \eta_1, \tau_1, \tau_2$) model by making various assumptions about its six parameters. When we have more than one α parameter, there is no direct relationship between the two models when $\eta_1 = 0$, though they both have the same number of parameters (a result that also holds in $p > 2$ dimensions). However, assuming $\alpha_1 = \alpha_2 = \alpha$ and using a standard result from matrix theory (Harville, 1997, Corollary 8.5.12), it is easily shown that the resulting GMCAR(α, η_0, $\eta_1 = 0$, τ_1, τ_2) model is exactly the same as the MCAR(α, Λ) model (8). In this case, the functional relationships between the parameters are that $\tau_1 = \Lambda_{11}$, $\tau_2 = \Lambda_{22} - \Lambda_{12}^2/\Lambda_{11}$, and $\eta_0 = -\Lambda_{12}/\Lambda_{11}$. Next, assuming $\alpha = 1$, the GMCAR(1, η_0, τ_1, τ_2) is equivalent to the MCAR(1, Λ) model (the multivariate IAR model). If we assume $\alpha_1 \neq \alpha_2$ and $\eta_0 = \eta_1 = 0$, then we ignore dependence between the multivariate components, and the model turns out to be equivalent fitting two separate univariate CAR models. Finally, if we instead assume $\alpha_1 = \alpha_2 = 0$, $\eta_0 \neq 0$, and $\eta_1 = 0$, the model becomes an i.i.d. bivariate normal model.

The usual MCAR model (8) has $E(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2) = -\Lambda_{12}/\Lambda_{11}\,\boldsymbol{\phi}_2$; the conditional mean is merely a scale multiple of $\boldsymbol{\phi}_2$. Because $\mathrm{Var}(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2) = [\Lambda_{11}(D - \alpha_1 W)]^{-1}$, which is free of $\boldsymbol{\phi}_2$, the distribution of the random variable at a particular site in one field is independent of neighbor variables in another field *given* the value of the related variable at the same area. The extended MCAR model (10) has $E(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2) = -\Lambda_{12}/\Lambda_{11}(D - \alpha_1 W)^{-1/2}(D - \alpha_2 W)^{1/2}\boldsymbol{\phi}_2$, and $\mathrm{Var}(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2)$ identical to that of model (8). Hence, the distribution of the random variable at a particular site in one field is no longer conditionally independent of neighbor variables in another field, but one cannot really "model" this dependence because it is determined implicitly by α_1 and α_2. By contrast, our GMCAR model has $E(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2) = (\eta_0 I + \eta_1 W)\ \boldsymbol{\phi}_2$ and $\mathrm{Var}(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2) = [\tau_1(D - \alpha_1 W)]^{-1}$. Thus, while the conditional variance remains free of $\boldsymbol{\phi}_2$, the GMCAR allows spatial information (via the W matrix) to enter the conditional mean in an intuitive way, with a free parameter (η_1) to model the weights. That is, the GMCAR models the conditional mean of $\boldsymbol{\phi}_1$ for a given region as a sensible weighted average of the values of $\boldsymbol{\phi}_2$ for that region *and* a neighborhood of that region.

As a specific example of a practical modeling benefit obtainable with our GMCAR approach, suppose we wished to include different weighted adjacency matrices in the MCAR(α, $\mathbf{\Lambda}$) distribution, for example, extending the precision matrix in model (8) to

$$\Sigma^{-1} = \begin{pmatrix} \big(D_1 - \alpha W^{(1)}\big)\Lambda_{11} & \big(D_3 - \alpha W^{(3)}\big)\Lambda_{12} \\ \big(D_3 - \alpha W^{(3)}\big)\Lambda_{12} & \big(D_2 - \alpha W^{(2)}\big)\Lambda_{22} \end{pmatrix}, \quad (14)$$

where $D_k = \text{Diag}(\sum_{j=1}^n W_{1j}^{(k)}, \ldots, \sum_{j=1}^n W_{nj}^{(k)})$ and $W^{(k)}$ is the weighted adjacency matrix with ij-element $W_{ij}^{(k)}$, $k = 1, 2, 3$, and $i, j = 1, \ldots, n$. The conditions for this new precision matrix being positive definite precision matrix are not at all clear. But in our GMCAR case, we obtain

$$\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2 \sim N\big(\big(\eta_0 I + \eta_1 W^{(3)}\big)\boldsymbol{\phi}_2, \big[\tau_1\big(D_1 - \alpha_1 W^{(1)}\big)\big]^{-1}\big),$$

and

$$\boldsymbol{\phi}_2 \sim N\big(\mathbf{0}, \big[\tau_2\big(D_2 - \alpha_2 W^{(2)}\big)\big]^{-1}\big).$$

The conditions for positive definiteness are easily shown to be $|\alpha_1| < 1$ and $|\alpha_2| < 1$ using a diagonal dominance argument.

Because we specify the joint distribution for a multivariate MRF directly through the specification of simpler conditional and marginal distributions, a practical consideration is the order of our hierarchical modeling (i.e., whether to model $p(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2)$ and then $p(\boldsymbol{\phi}_2)$, or $p(\boldsymbol{\phi}_2 \,|\, \boldsymbol{\phi}_1)$ and then $p(\boldsymbol{\phi}_1)$). In some cases, the conditional modeling order can be determined by a chronology or perhaps causality in events. For example, in an analysis of multivariate pollutant data, Schmidt and Gelfand (2003) model particulate matter as a function of meteorological variables (temperature and humidity), since this is more scientifically plausible than the other way around. Similarly, Gelfand et al. (2004) discuss the analysis of commercial real estate data for which the selling price P for a block of apartments is intuitively thought of as a function of the income I generated by that block. A natural modeling order is thus I, followed by P given I. In other cases, one's ability to understand and interpret model parameters might depend on the order in which they are modeled. An example of this case is the work of Royle and Berliner (1999), where the concentration of ozone at a particular location is scientifically explainable given the maximum temperature at that location, but not the other way around. The modeling by Zhu, Carlin, and Gelfand (2003) of areal average ozone levels followed by pediatric asthma rates given these levels offers another example.

In the absence of any natural ordering, one can always treat the choice of order as a model selection issue, and choose the conditioning order that produces the best fit to the data. This is equivalent to comparing two joint model specifications. Other authors (e.g., Berkhout and Plug, 2004) have used this approach with success; along similar lines, Held et al. (2005) use a latent variable approach to order cancers based on their causal connection to the use of alcohol, tobacco, or both. In our setting, use of the DIC for choosing among complex hierarchical models seems promising; we explore this idea in Sections 4 and 5.

One might think we could symmetrically specify the conditional distributions $p(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2)$ and $p(\boldsymbol{\phi}_2 \,|\, \boldsymbol{\phi}_1)$ and use Brook's Lemma to find the joint distribution, thus avoiding the order issue entirely. But this leads to another problem, namely how to specify these two conditional distributions so that they are compatible and determine a valid joint distribution. Because positive definiteness conditions turn out to be difficult to find from this perspective, we do not consider it further in this article, although some thoughts for future investigation are given in our closing Section 6.

Finally, our approach can be easily generalized to dimensions p greater than 2. For example, in the trivariate case $p = 3$, defining $\boldsymbol{\phi}_1' = (\phi_{11}, \ldots, \phi_{n1})$, $\boldsymbol{\phi}_2' = (\phi_{12}, \ldots, \phi_{n2})$, and $\boldsymbol{\phi}_3' = (\phi_{13}, \ldots, \phi_{n3})$, we can specify valid conditional distributions $p(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2, \boldsymbol{\phi}_3)$ and $p(\boldsymbol{\phi}_2 \,|\, \boldsymbol{\phi}_3)$, and a valid marginal distribution $p(\boldsymbol{\phi}_3)$. The joint distribution of $\boldsymbol{\phi}' = (\boldsymbol{\phi}_1', \boldsymbol{\phi}_2', \boldsymbol{\phi}_3')$ is written as $p(\boldsymbol{\phi}) = p(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2, \boldsymbol{\phi}_3)p(\boldsymbol{\phi}_2 \,|\, \boldsymbol{\phi}_3)p(\boldsymbol{\phi}_3)$, which of course is always proper. Also, the computational burden does not increase much with the dimension p, since it involves only n-dimensional matrix calculations.

4. Simulation Studies

To see the advantage of our GMCAR models, we begin with some simulation studies. They are based on the spatial layout of the 87 counties in the state of Minnesota, a fairly typical layout and the one used by our Section 5 data set. We assume the data Y_{ij} arise from a Gaussian model

$$Y_{ij} \overset{\text{ind}}{\sim} N\big(\beta_j + \phi_{ij}, \sigma^2\big), \quad i = 1, \ldots, n, \quad j = 1, 2, \quad (15)$$

where the β_j's are fixed constants. In Studies 1 and 2, we generate $\boldsymbol{\phi}_1 = (\phi_{11}, \ldots, \phi_{1n})'$ and $\boldsymbol{\phi}_2 = (\phi_{21}, \ldots, \phi_{2n})'$ from our proposed GMCAR(α_1, α_2, η_0, η_1, τ_1, τ_2) model (13); specifically, $\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2 \sim N\,((\eta_0 I + \eta_1 W)\boldsymbol{\phi}_2, [\tau_1(D - \alpha_1 W)]^{-1})$, and $\boldsymbol{\phi}_2 \sim N(\mathbf{0}, [\tau_2(D - \alpha_2 W)]^{-1})$, where $D = \text{Diag}(m_i)$ and the adjacency matrix W are based on the Minnesota county map. The designs of Studies 1 and 2 differ only in that we set the bridging parameter $\eta_1 = 0$ in Study 2.

To compare our proposed GMCAR models with the existing MCAR models, we consider two more simulation designs. In Study 3, we generate $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ from model (10), the MCAR(α_1, α_2, Λ) distribution (using the Cholesky method to obtain the R_k matrices), and in Study 4 we generate $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ from model (11), the 2fCAR(α_0, α_1, α_2, α_3, τ_1, τ_2). The true values of the parameters assumed by each study are shown in Table 1.

4.1 *Bayesian Computation*

Our proposed GMCAR(α_1, α_2, η_0, η_1, τ_1, τ_2) models are straightforwardly implemented in a Bayesian framework using MCMC methods. To improve MCMC convergence, we used hierarchical centering (Gelfand, Sahu, and Carlin, 1995) to reparameterize model (15) to

$$Y_{ij} \overset{\text{ind}}{\sim} N\big(Z_{ij}, \sigma^2\big), \quad i = 1, \ldots, n, \quad j = 1, 2. \quad (16)$$

Because $Z_{ij} = \beta_j + \phi_{ij}$, we still can model Z_{ij} with the GMCAR model (13), but the mean of Z_{ij} becomes β_j rather than 0. In this case, we have the conditional distribution for $\mathbf{Z}_1 \,|\, \mathbf{Z}_2$,

$$\mathbf{Z}_1 \,|\, \mathbf{Z}_2 \sim N\big(\beta_1 \mathbf{1} + (\eta_0 I + \eta_1 W) \times (\mathbf{Z}_2 - \beta_2 \mathbf{1}), [\tau_1(D - \alpha_1 W)]^{-1}\big),$$

Table 1
The true values of parameters in Studies 1–4

Study	True model	β_1	β_2	σ^2	$\tau_1(\Lambda_{11})$	$\tau_2(\Lambda_{22})$	α_1	α_2	$\eta_0(\alpha_0)$	$\eta_1(\alpha_3)$	Λ_{12}
1	GMCAR	−2.0	−5.0	0.01	10	10	0.20	0.90	0.90	0.50	—
2	GMCAR	−2.0	−5.0	0.01	10	10	0.20	0.90	0.90	0	—
3	MCAR	−2.0	−5.0	0.01	10	15	0.20	0.90	—	—	6.1
4	2fCAR	−2.0	−5.0	0.01	10	10	0.20	0.90	0.90	0.50	—

and the marginal distribution $\mathbf{Z}_2 \sim N(\beta_2\mathbf{1}, [\tau_2(D - \alpha_2 W)]^{-1})$, where $\mathbf{Z}_1 = (Z_{11}, \ldots, Z_{n1})'$ and $\mathbf{Z}_2 = (Z_{12}, \ldots, Z_{n2})'$. Thus, the joint distribution of $\mathbf{Z}' = (\mathbf{Z}_1', \mathbf{Z}_2')$ is

$$
\begin{aligned}
p(\mathbf{Z} \,|\, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\eta}) &\propto \tau_1^{n/2} |D - \alpha_1 W|^{1/2} \\
&\times \exp\left\{ -\frac{\tau_1}{2} [\mathbf{Z}_1 - \beta_1 \mathbf{1} - (\eta_0 I + \eta_1 W)(\mathbf{Z}_2 - \beta_2 \mathbf{1})]' \right. \\
&\qquad \left. \times (D - \alpha_1 W)[\mathbf{Z}_1 - \beta_1 \mathbf{1} - (\eta_0 I + \eta_1 W)(\mathbf{Z}_2 - \beta_2 \mathbf{1})] \right\} \\
&\times \tau_2^{n/2} |D - \alpha_2 W|^{1/2} \\
&\times \exp\left[-\frac{\tau_2}{2} (\mathbf{Z}_2 - \beta_2 \mathbf{1})'(D - \alpha_2 W)(\mathbf{Z}_2 - \beta_2 \mathbf{1}) \right], \qquad (17)
\end{aligned}
$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2)$, $\boldsymbol{\tau} = (\tau_1, \tau_2)$, $\boldsymbol{\eta} = (\eta_0, \eta_1)$, and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$. The joint posterior distribution is

$$
\begin{aligned}
&p\big(\boldsymbol{\beta}, \sigma^2, \mathbf{Z}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\eta} \,|\, \mathbf{Y}_1, \mathbf{Y}_2\big) \\
&\quad \propto L\big(\mathbf{Y}_1, \mathbf{Y}_2 \,|\, \mathbf{Z}, \sigma^2\big) \\
&\qquad \times p(\mathbf{Z} \,|\, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\eta}) p(\boldsymbol{\beta}) p(\boldsymbol{\tau}) p(\boldsymbol{\alpha}) p(\boldsymbol{\eta}) p(\sigma^2), \qquad (18)
\end{aligned}
$$

where $\mathbf{Y}_1 = (Y_{11}, \ldots, Y_{n1})'$ and $\mathbf{Y}_2 = (Y_{12}, \ldots, Y_{n2})'$. The first term on the right-hand side of (18) is the likelihood, $L(\mathbf{Y}_1, \mathbf{Y}_2 \,|\, \mathbf{Z}, \sigma^2) \propto \sigma^{-2n} \exp\{-(1/2\sigma^2)[(\mathbf{Y}_1 - \mathbf{Z}_1)'(\mathbf{Y}_1 - \mathbf{Z}_1) + (\mathbf{Y}_2 - \mathbf{Z}_2)'(\mathbf{Y}_2 - \mathbf{Z}_2)]\}$. The second term on the right-hand side of (18) is (17), and the remaining terms are the prior distributions on $(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \sigma^2)$. For the remaining terms, flat priors are chosen for β_1 and β_2, while σ^2 is assigned a vague inverse gamma prior, that is, an IG(1, 0.1) where we parameterize the IG(a, b) so that $E(\sigma^2) = b/(a - 1)$. Next, τ_1 and τ_2 are assigned vague gamma priors, specifically a $G(1, 0.1)$, which has mean 10 and variance 100. Finally, α_1, α_2 are given Unif(0, 1) priors while η_0 and η_1 are given $N(0, \sigma_1^2)$ and $N(0, \sigma_2^2)$ priors, respectively. For convenience we set $\sigma_1 = \sigma_2 = 10$, since in our experience there appears to be little change in our results from using larger values.

The Gibbs sampler (Gelfand and Smith, 1990; Carlin and Louis, 2000, Section 5.4.2) is natural for updating the parameters in this setting, since it can take advantage of the conditional specification of the GMCAR model. Each of the full conditional distributions required by the Gibbs sampler must be proportional to (18). In finding and updating the full conditionals, it is easily shown that no matrix inversion is required, and that calculations on rather special (e.g., diagonal) n-dimensional matrices are all that are required regardless of the dimension p (recall $p = 2$ in our case). To calculate the determinant in (17), we have the fact that

$$
\begin{aligned}
|D - \alpha_k W| &= \left| D^{1/2} \big(1 - \alpha_k D^{-1/2} W D^{-1/2}\big) D^{1/2} \right| \\
&= |D| \prod_{i=1}^{n} (1 - \alpha_k \lambda_i) \propto \prod_{i=1}^{n} (1 - \alpha_k \lambda_i), \quad k = 1, 2,
\end{aligned}
$$

where $\lambda_i, i = 1, \ldots, n$ are the eigenvalues of the matrix $D^{-1/2} W D^{-1/2}$. The λ_i may be calculated prior to any MCMC iteration. Hence, posterior computation for the GMCAR model is simpler and faster than that for existing MCAR models, especially for large areal data sets.

All of the parameters in (18) except $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ have closed-form full conditionals, and so may be directly updated. For these two remaining parameters, Metropolis–Hastings steps with bivariate Gaussian proposals are convenient (though for $\boldsymbol{\alpha}$, a preliminary logit transformation, having Jacobian $\prod_{k=1}^{2} \alpha_k(1 - \alpha_k)$, is required). In practice, the α_k must be bounded away from 1 (say, by insisting $0 < \alpha_k < 0.999$, $k = 1, 2$) to maintain identifiability and hence computational stability.

4.2 *Simulation Results*

To check the performance of our proposed GMCAR models, we simulated $N = 100$ data sets, and fit several different models to each in every study. For each data set and each model, we first ran a few initially overdispersed parallel MCMC chains, and monitored them using measurements of sample autocorrelations within the chains, cross-correlations between parameters, and plots of sample traces. From these, we decided to use 5000 iterations for the preconvergence "burn-in" period, and then a further 15,000 iterations as our "production" run for posterior summarization. While our GMCAR models can be in the `WinBUGS` package (see `www.biostat.umn.edu/~brad/software.html`), for the purpose of our simulation we prefer to rely on our own programs written in `C` and executed in `R` (`www.r-project.org`) using the `.C` function. Random number generation and posterior summarization were also accomplished in `R`.

To choose among competing models, we turned to a simple and intuitively appealing hierarchical modeling extension of AIC called the *deviance information criterion*, or DIC. This criterion is based on the posterior distribution of the *deviance* statistic, $D(\boldsymbol{\theta}) = -2 \log f(\mathbf{y} \,|\, \boldsymbol{\theta}) + 2 \log h(\mathbf{y})$, where $f(\mathbf{y} \,|\, \boldsymbol{\theta})$ is the likelihood function for the observed data vector $\mathbf{y}$ given the parameter vector $\boldsymbol{\theta}$ on which we focus, and $h(\mathbf{y})$ is some standardizing function of the data alone (and which thus has no impact on model selection). The DIC is defined analogously to the AIC as the posterior expected deviance plus

the "effective" number of parameters, i.e., $\text{DIC} = \overline{D} + p_D$. Spiegelhalter et al. (2002) show that p_D is reasonably defined as $E_{\theta|y}[D] - D(E_{\theta|y}[\boldsymbol{\theta}]) = \overline{D} - D(\bar{\boldsymbol{\theta}})$. Because small values of $\overline{D}$ indicate good fit while small values of p_D indicate a parsimonious model, small values of the sum (DIC) indicate preferred models. Note DIC is scale-free (because $\overline{D}$ is), and so no particular score has any intrinsic meaning; only the *ordering* of DIC scores across models is meaningful.

To calculate DIC for our simulated data, we need to calculate $D(\boldsymbol{\theta})$; note this is the same for all the models we wish to compare since they differ only in their random effect distributions, which we do not consider to be part of the likelihood. Setting $2 \log h(\mathbf{y}) = 0$ in $D(\boldsymbol{\theta})$, we have $D(\boldsymbol{\theta}) \equiv D(\mathbf{Z}, \sigma^2) = -2 \log L(\mathbf{Y}_1, \mathbf{Y}_2 \,|\, \mathbf{Z}, \sigma^2)$.

In addition to comparing the models, we also check the AMSE performance of each. Because the true Z_{ij} values are known in the simulation, this is estimated as

$$\widehat{\text{AMSE}} = \frac{1}{Nnp} \sum_{t=1}^{N} \sum_{j=1}^{p} \sum_{i=1}^{n} \left(\widehat{Z}_{ij}^{(t)} - Z_{ij} \right)^2$$

with associated Monte Carlo standard error estimate

$$\widehat{\text{se}}(\widehat{\text{AMSE}}) = \sqrt{\frac{1}{(Nnp)(Nnp-1)} \sum_{t=1}^{N} \sum_{j=1}^{p} \sum_{i=1}^{n} \left[\left(\widehat{Z}_{ij}^{(t)} - Z_{ij} \right)^2 - \widehat{\text{AMSE}} \right]^2},$$

where in our case we have $N = 100$, $p = 2$, and $n = 87$.

In Tables 2 and 3, Models 1–3 are members of the class of proposed GMCAR(α_1, α_2, η_0, η_1, τ_1, τ_2) models under various assumptions described in Section 3. Model 1 is a full model with all six parameters. Model 2 is a reduced model that sets $\eta_1 = 0$ (i.e., no effect of Z_{j2} on $E(Z_{i1} \,|\, \mathbf{Z}_2)$ for counties $j \neq i$). Model 3 returns to the full model, but after reversing the conditioning order to $[\mathbf{Z}_2 \,|\, \mathbf{Z}_1]$. Models 4 and 5 are existing MCAR models, namely an MCAR(α_1, α_2, Λ) (again using the Cholesky method for the R_k) and a 2fCAR(α_0, α_1, α_2, α_3, τ_1, τ_2), respectively.

Table 2 summarizes the difference in DIC score between each model and that study's true model (so negative values correspond to a model "beating" the true model). In each case, we provide 2.5, 50, and 97.5 percentiles for this difference. In addition, Table 3 gives the estimated AMSEs and their associated Monte Carlo standard errors for each model in each simulation study. Here, we also calculate the percentage change in estimated AMSE for each model compared to the true model in each study, that is, $\triangle_\ell = (\widehat{\text{AMSE}}_\ell - \widehat{\text{AMSE}}_{\text{true}})/\widehat{\text{AMSE}}_{\text{true}} \times 100$ for models $\ell = 1, \ldots, 5$; again negative values indicate superiority over the true model.

Tables 2 and 3 are reasonably consistent, in that large DIC differences generally correspond to large AMSEs. In Study 1, the true GMCAR(α_1, α_2, η_0, η_1, τ_1, τ_2) full model easily beats the other models, and DIC and AMSE also reveal the correct conditioning order, since Model 3 finishes well behind Model 1 in this study using either metric. In other studies, however, the conditioning order seems not to matter (i.e., Models 1 and 3 are quite close). This makes sense, since the true linking parameter η_1 is not significantly different from 0 in these studies ($\alpha_3 = 0.5 \neq 0$ in Study 4, but this corresponds to an η_1 value only around 0.1–0.2).

In Studies 3 and 4, GMCAR Models 1 and 2 actually perform as well as the true models in terms of both median DIC difference and AMSE. This suggests the GMCAR is able to pick up small departures of our data from existing MCAR models, and do so efficiently. Note also that the twofold CAR model seems to do poorly throughout, finishing in no better than a statistical "tie" with the three GMCAR models in both tables even when it is the true model (Study 4). The full GMCAR Model 1 is competitive in all four studies, suggesting good performance with little risk of overfitting.

5. Data Example

We now turn from the Gaussian simulation studies of Section 4 to an example that features non-Gaussian data. The data consist of the numbers of deaths due to cancers of the lung and esophagus in the years from 1991 to 1998 at the county level in Minnesota. These diseases are rare enough relative to the population in each county that a Poisson spatial regression model (see, e.g., Banerjee et al., 2004, Section 5.4) is appropriate. We write the model as

$$Y_{ij} \overset{\text{ind}}{\sim} \text{Poisson}\left(E_{ij} e^{Z_{ij}}\right), \quad i = 1, \ldots, 87, \quad j = 1, 2, \quad (19)$$

where Y_{ij} is the observed number of deaths due to cancer j in county i, and E_{ij} is the corresponding expected number of deaths (assumed known). To calculate E_{ij}, we have to take each county's age distribution into account. To do so, we calculate the expected *age-adjusted* number of deaths due to cancer j in county i as

Table 2

Percentiles of estimated DIC difference between the true model and the other models at each simulation study. Model 1 = GMCAR (full); Model 2 = GMCAR (reduced; $\eta_1 = 0$); Model 3 = GMCAR (full, reverse order); Model 4 = MCAR; Model 5 = 2fCAR.

	Study 1			Study 2			Study 3			Study 4		
Model	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%	2.5%	50%	97.5%
1	—	—	—	−6.24	2.74	5.64	−15.3	−0.59	8.57	−12.1	−2.24	2.37
2	11.8	38.7	76.0	—	—	—	−8.68	−1.89	10.5	−15.6	−3.49	1.50
3	−4.80	19.6	56.7	−11.3	−0.06	8.38	−11.9	2.32	13.0	−14.3	−2.15	3.16
4	3.56	34.9	68.8	−3.62	1.48	8.47	—	—	—	−10.0	0.83	4.89
5	3.03	23.7	65.1	0.50	30.4	63.7	2.76	20.9	53.3	—	—	—

The symbol "—" indicates the model is the true model for this study.

Table 3
Average mean squared error (AMSE, $\times 10^{-3}$), associated Monte Carlo standard errors (SE, $\times 10^{-5}$), and percentage change in AMSE ($\triangle$, %) relative to the true model in each simulation study

	Study 1		Study 2		Study 3		Study 4	
Model	AMSE (SE)	$\triangle$	AMSE (SE)	$\triangle$	AMSE (SE)	$\triangle$	AMSE (SE)	$\triangle$
1	7.51 (8.26)	—	7.91 (8.57)	1.54	8.49 (9.20)	−3.08	5.46 (5.92)	−7.92
2	10.2 (11.3)	35.8	7.79 (8.45)	—	8.46 (9.17)	−3.42	5.41 (5.86)	−8.77
3	8.17 (8.91)	8.79	7.70 (8.44)	−1.16	8.81 (9.67)	0.571	5.44 (5.87)	−8.26
4	9.22 (10.1)	22.8	7.82 (8.50)	0.385	8.76 (9.55)	—	5.85 (6.34)	−1.35
5	8.22 (8.80)	9.45	9.86 (10.9)	26.6	11.2 (12.3)	27.8	5.93 (6.44)	—

The symbol "—" indicates the model is the true model for this study.

$$E_{ij} = \sum_{k=1}^{m} \omega_j^k N_i^k, \quad i = 1, \ldots, 87, \quad j = 1, 2,$$

where $\omega_j^k = (\sum_{i=1}^{87} D_{ij}^k)/(\sum_{i=1}^{87} N_i^k)$ is the age-specific death rate due to cancer j for age group k over all Minnesota counties, D_{ij}^k is the number of deaths in age group k of county i due to cancer j, and N_i^k is the total population at risk in county i, age group k.

The county-level maps of the age-adjusted standardized mortality ratios (SMRs) (i.e., $\text{SMR}_{ij} = Y_{ij}/E_{ij}$) shown in Figure 1 exhibit the evidence of correlation both across space and between cancers, motivating use of our proposed GMCAR models. Regarding the selection of the proper order in which to model the two cancers, Figure 2 gives a helpful data-based exploratory plot. We first obtain crude data-based estimates of the spatial random effects as $\hat{\phi}_{i1} = \log(\text{SMR}_{i1})$ and $\hat{\phi}_{i2} = \log(\text{SMR}_{i2})$. Next, recall the linearity of the conditional GMCAR mean for a given ordering (say, lung given esophagus), that is,

$$E(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2) = A\boldsymbol{\phi}_2 = A(\eta_0, \eta_1)\boldsymbol{\phi}_2 = (\eta_0 I + \eta_1 W)\boldsymbol{\phi}_2.$$

This motivates obtaining least-squares estimates $\hat{\eta}_0$ and $\hat{\eta}_1$ by minimizing $(\hat{\boldsymbol{\phi}}_1 - A(\eta_0, \eta_1)\hat{\boldsymbol{\phi}}_2)'(\hat{\boldsymbol{\phi}}_1 - A(\eta_0, \eta_1)\hat{\boldsymbol{\phi}}_2)$ as a function of η_0 and η_1. Finally, we plot $A(\hat{\eta}_0, \hat{\eta}_1)\hat{\boldsymbol{\phi}}_2$ versus $\hat{\boldsymbol{\phi}}_1$, and investigate how well the linearity assumption is supported by the data. Repeating this entire process for the reverse order (here, esophagus given lung) produces a second plot, which may be compared in quality to the first. In our case, Figure 2a (lung given esophagus) indicates more support for linearity, both in its appearance and in its higher sample correlation and regression t-statistic.

Using likelihood (19), we model the random effects Z_{ij} in the same way as in Section 4 using the GMCAR(α_1, α_2, η_0, η_1, τ_1, τ_2) with mean $\boldsymbol{\beta}$. In what follows we compare the GMCAR with other existing MCAR models using DIC. In Table 4, Models 1–3 are members of our proposed GMCAR class paralleling those in Section 4. Specifically, in Model 1, we have the full model with all six parameters, and the conditioning order of the cancers is [lung | esophagus]. Model 2 assumes $\eta_1 = 0$ and uses the same conditioning order as Model 1. In Model 3, we switch the conditioning order to [esophagus | lung] and return to a full model. To compare the GMCAR to existing MCAR models, we take the MCAR(α_1, α_2, Λ) using the Cholesky method for the R_k as Model 4, the same model but using the spectral method for the R_k as Model 5, and the 2fCAR(α_0, α_1, α_2, α_3, τ_1, τ_2) as Model 6. We choose the same prior distributions for each parameter as in Section 4, and use Metropolis–Hastings and Gibbs sampling to update all parameters. We use 5000 iterations as the preconvergence burn-in period, and then a further 20,000 iterations as our production run for posterior summarization.

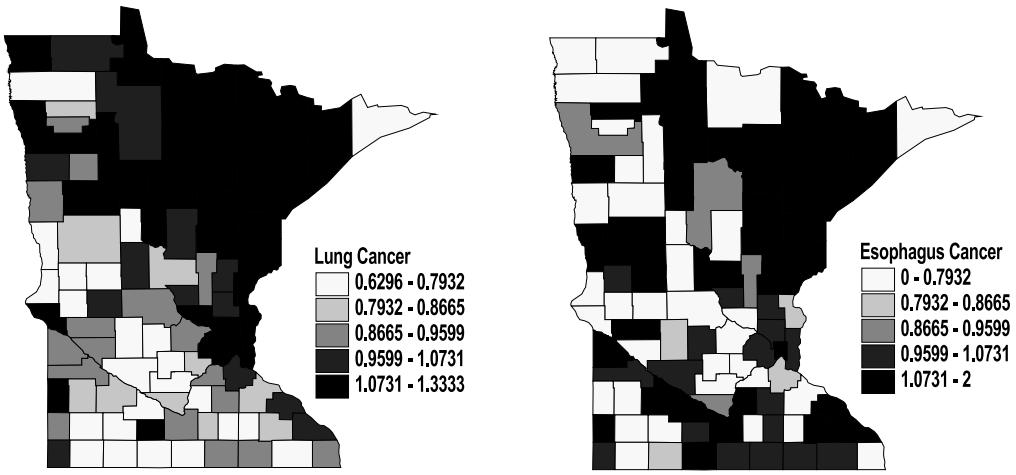

Figure 1. Maps of age-adjusted SMR for lung and esophagus cancer in Minnesota.

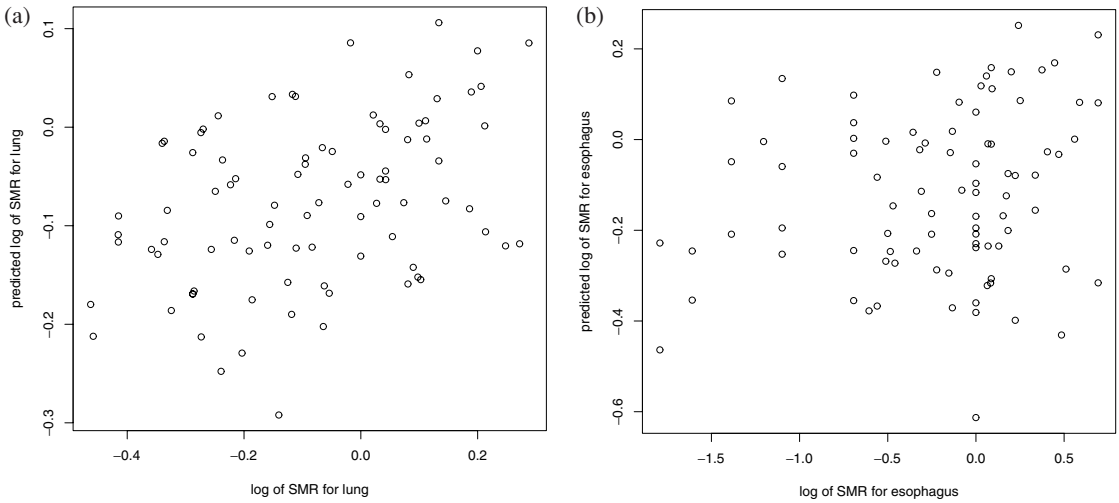

Figure 2. Exploratory plot to help select modeling order: (a) [lung | esophagus], sample correlation 0.394, regression $t = 3.956$; (b) [esophagus | lung], sample correlation 0.193, regression $t = 1.813$.

Fit measures $\overline{D}$, effective numbers of parameters p_D, and DIC scores for each model are seen in Table 4. Model 1 has the smallest p_D and DIC values, so our GMCAR(α_1, α_2, η_0, η_1, τ_1, τ_2) full model with the conditioning order [lung | esophagus] emerges as best for this data set. The reduced GMCAR Model 2 does less well, suggesting the need to account for bivariate spatial structure in these data. The two MCAR methods perform similarly to each other and to the reduced GMCAR model, while the 2fCAR model does less well, largely because it does not seem to allow sufficient smoothing of the random effects (larger p_D score). Note that effective degrees of freedom may actually be smaller for apparently more complex models that allow more complicated forms of shrinkage, such as Model 1 in this case. We note that our "focus" parameter is the same for each model (both fixed and random effects are in focus), and the Poisson likelihood is also not changing across models. Also, our priors are all noninformative or quite vague (e.g., uniform priors for all α parameters). All of this suggests the DIC comparison in Table 4 is fair across models. Moreover, the resulting DIC scores were robust to the moderate changes in the prior distributions.

Regarding estimation of the fixed effects, under Model 1 we obtained point and 95% equal-tail interval estimates of 0.602 and (0.0267, 0.979) for α_1, and 0.699 and (0.0802, 0.973) for α_2. Recall these are spatial association parameters, but while their values are between 0 and 1 they are not "correlations" in the usual sense; the moderate point estimates and wide confidence intervals suggest a relatively modest degree of spatial association in the random effects. It is also important to remember that in this setup, α_2 measures spatial association in the esophagus random effects $\boldsymbol{\phi}_2$, while α_1 measures spatial association in the lung random effects $\boldsymbol{\phi}_1$ *given* the esophagus random effects $\boldsymbol{\phi}_2$. Thus, the interpretation of the α_k would be different for Model 3 (due to the different conditioning order), and much different for Models 4 or 5. Note that for the MCAR model, $E(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2)$ and $E(\boldsymbol{\phi}_2 \,|\, \boldsymbol{\phi}_1)$ both depend on both α_1 and α_2. But for the GMCAR, $E(\boldsymbol{\phi}_1 \,|\, \boldsymbol{\phi}_2)$ is free of both α_1 and α_2, while of course $E(\boldsymbol{\phi}_2) = 0$. Thus, for this model, α_1 and α_2 unambiguously control only their corresponding variance matrices, and can be set without altering the mean structure.

Table 4
Model comparison using DIC statistics, Minnesota cancer data analysis

	Model	$\overline{D}$	p_D	DIC
1	GMCAR (full)	483.4	58.2	541.6
2	GMCAR (reduced; $\eta_1 = 0$)	483.0	63.8	546.8
3	GMCAR (full, reverse order)	480.6	63.3	543.9
4	MCAR (Cholesky decomposition)	483.6	61.3	544.9
5	MCAR (spectral decomposition)	483.8	60.6	544.4
6	2fCAR	482.6	65.1	547.7

Turning to τ_1 and τ_2, under Model 1 we obtained 32.65, (16.98, 66.71) and 13.73, (4.73, 38.05) as our point and interval estimates, respectively. Because these parameters measure spatial precision for each disease, they suggest slightly more variability in the esophagus random effects, although again comparison is difficult here since τ_2 is a *marginal* precision for $\boldsymbol{\phi}_2$ while τ_1 is a *conditional* precision for $\boldsymbol{\phi}_1$ given $\boldsymbol{\phi}_2$. Along these lines, Figure 3 shows estimated posteriors of the conditional variances $\sigma_1^2 = 1/\tau_1$ for several candidate multivariate spatial models. Figure 3a shows the situation for two separate CAR models, a model that ignores any possibility of connection between the cancers. The remaining panels consider the MCAR(α_1, α_2, $\boldsymbol{\Lambda}$) model, the reduced GMCAR(α_1, α_2, η_0, τ_1, τ_2) model, and the full GMCAR(α_1, α_2, η_0, η_1, τ_1, τ_2) model. The reduction of uncertainty in $\boldsymbol{\phi}_1$ given $\boldsymbol{\phi}_2$ in these more complex models is a measure of the information content between the cancers, and is readily apparent from the histograms and their empirical means.

DIC's slight preference for Model 1 is consistent with the estimated posteriors of the linking parameters η_0 and η_1 shown in Figure 4. The inclusion of 0 within the 95% credible interval for η_1 under the reverse ordering, but not under the

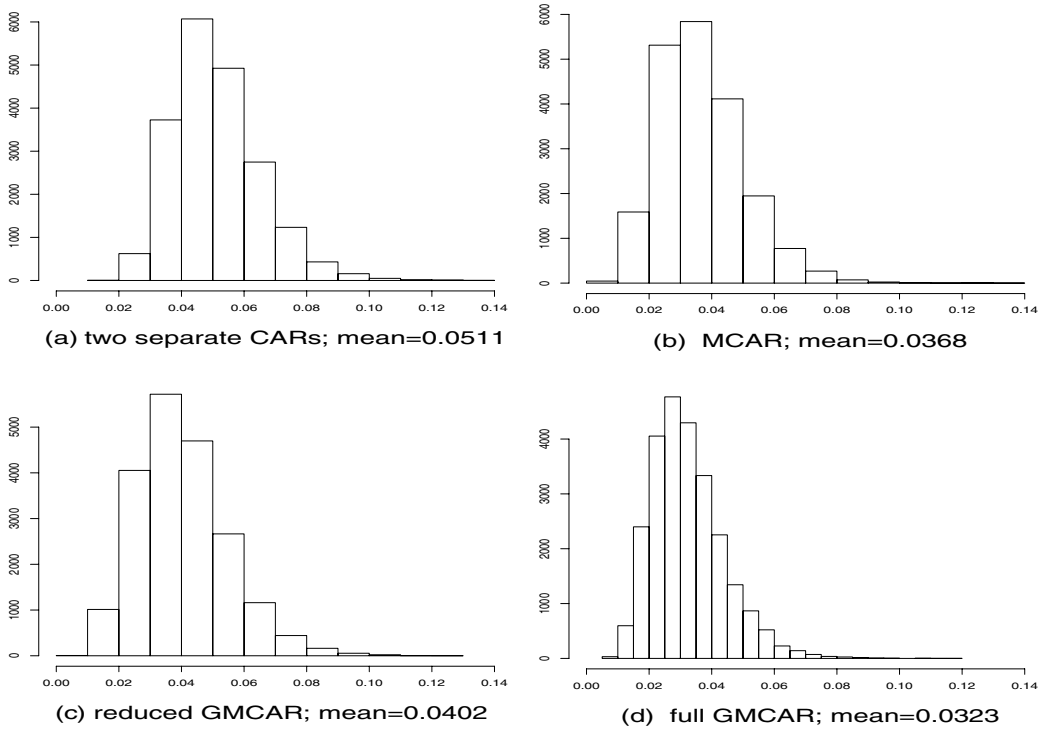

Figure 3. Posterior samples of conditional variances $\sigma_1^2 = 1/\tau_1$ for various models: (a) two separate CAR models; (b) MCAR model; (c) reduced GMCAR model; and (d) full GMCAR model.

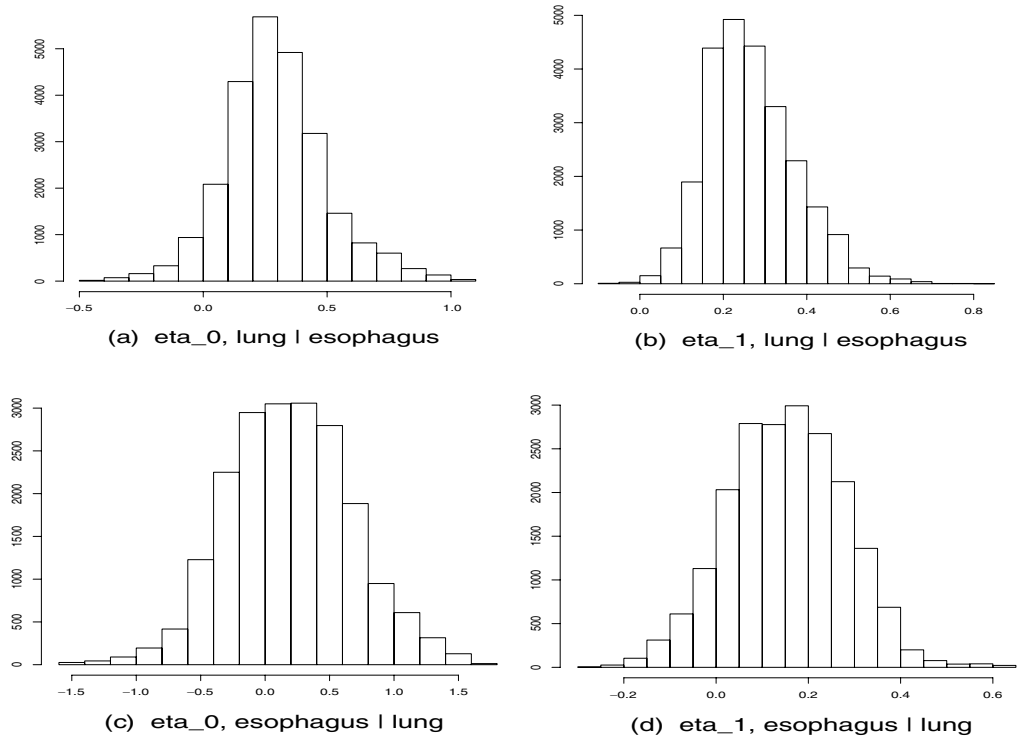

Figure 4. Posterior samples of η_0 and η_1 using the full GMCAR model with two conditioning orders: (a) estimated posterior for η_0, [lung | esophagus]; (b) estimated posterior for η_1, [lung | esophagus]; (c) estimated posterior for η_0, [esophagus | lung]; and (d) estimated posterior for η_1, [esophagus | lung].

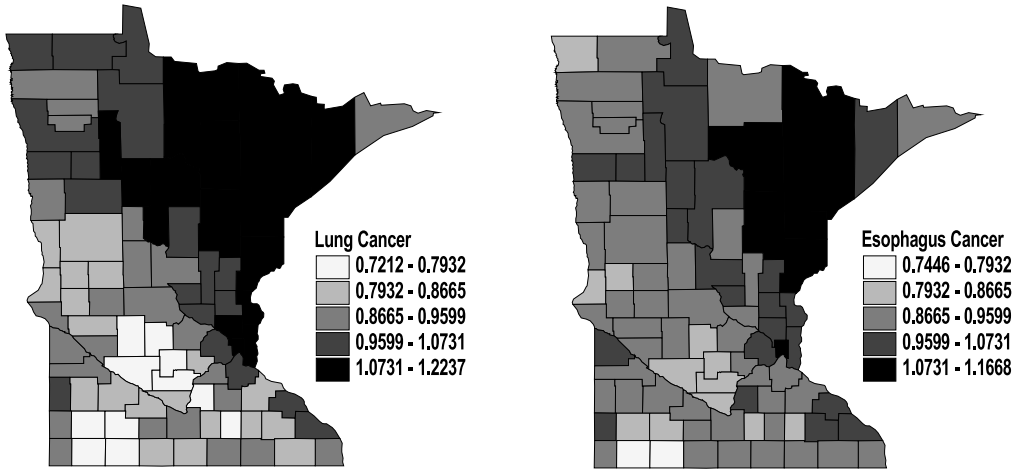

Figure 5. Maps of posterior mean SMR for lung and esophagus cancer in Minnesota from the full GMCAR model with conditioning order [lung | esophagus].

natural ordering, is yet further evidence against the former. Note also that the linking parameters η_0 and η_1 have mostly positive support, meaning that the two cancers have positive spatial correlation. This is also evident from the maps of the posterior means of the SMRs for the two cancers under the full model shown in Figure 5. Clearly, incidence of the two cancers is strongly correlated, with higher fitted ratios extending from the Twin Cities metro area (eastern side, about one third of the way up) to the mining- and tourism-oriented north and northeast, regions where conventional wisdom suggests cigarette smoking may be more common.

6. Summary and Future Research

In this article, we have introduced a flexible class of generalized multivariate CAR (GMCAR) models for complex areal data. We have shown that our generalized framework includes most existing multivariate CAR models as special cases, yet can still be efficiently computed using MCMC algorithms. Our simulations and data examples demonstrate the GMCAR's efficiency as well as its improved performance over the existing alternatives using average MSE and DIC score.

While our Section 5 example considered only disease *rates*, GMCAR models can also be used with time-to-event data to investigate geographical patterns in the *hazard function*. For example, each patient in a study may provide multiple survival times from the onset of each of several cancers along with his or her county of residence. Specifically, suppose the jth patient in the ith region has been diagnosed with a set of cancers C_{ij}, and let t_{ijk} denote the survival time for the (i, j)th individual from diagnosis of the kth type of cancer ($k \in C_{ij}$). Then, an appropriate Cox proportional hazards model might be

$$h(t_{ijk}) = h_0(t_{ijk}) \exp\left(\mathbf{x}_{ijk}^{\mathrm{T}}\boldsymbol{\beta} + u_{(i,j)} + v_k + \phi_{ik}\right),$$

where the $u_{(i,j)}$ and v_k are patient- and cancer-specific effects, while the ϕ_{ik} are spatially correlated frailties for the kth cancer type occurring in the ith county. These frailties are usually weakly identified by the data, so that modeling them in a flexible yet computationally efficient manner is crucial. In particular, the ϕ_{ik} capture space-cancer interactions can be flexibly and efficiently modeled with our GMCAR class, extending work by Carlin and Banerjee (2003) in the case of a single survival time that can be attributable to one of several diseases.

Finally, we note that most of the multivariate CAR modeling proceeds from conditional specifications. As mentioned near the end of Section 3, concerns arise over the ordering of the variables in the hierarchical approach. In particular, usual techniques of using full conditionals to identify joint distributions encounter difficult issues with regard to propriety and this is exacerbated when one deals with several variables $p > 2$. A joint modeling approach, though perhaps computationally more demanding, can lead to more flexible classes of models. One such approach builds rich spatial structures from linear transformations of simpler latent variables. For instance, we can develop alternate GMCAR-type models using $\boldsymbol{\phi} = T\boldsymbol{\psi}$, where T is a suitably specified square matrix. Note that by modeling the joint distribution, the incompatibility of conditional model building (i.e., different joint distributions for different orderings) is avoided. However, the issue of the identifiability of T crops up, and further parameterization is needed.

One such parameterization that leads to order-free specifications can be developed by first considering proper spatial p-variate Gaussian random variables:

$$\boldsymbol{\psi} \sim N\big(\mathbf{0}, (I_{p\times p} \otimes D - B \otimes W)^{-1}\big),$$

where B is a $p \times p$ matrix of parameters that ensures the invertibility of the dispersion matrix. We can further consider the linear transformation $\boldsymbol{\phi} = (A \otimes I_{n\times n})\boldsymbol{\psi}$, where A is also $p \times p$, leading to

$$\boldsymbol{\phi} \sim N\big(\mathbf{0}, (A \otimes I_{n\times n})(I_{p\times p} \otimes D - B \otimes W)^{-1}\big(A^{\mathrm{T}} \otimes I_{n\times n}\big)\big).$$

A crucial fact is that, by restricting A to be a triangular matrix, the elements in A and B can be uniquely identified for dispersion matrices of the form given in (9). Also, the distribution of $\boldsymbol{\phi}$ is invariant over choice of an upper- or lower-triangular A (up to a reparameterization of B). However, the

implementation of these models can be prohibitive, particularly involving the elements of B that ensure propriety of ϕ. This is currently being investigated.

Acknowledgements

The work of the first and second authors was supported in part by NIH grant 5-R01-ES07750-07, while that of the second and third authors was supported in part by NIH grant 1-R01-CA95955-01. The authors are grateful to Professor Melanie Wall for providing and permitting analysis of the Minnesota cancer data set, as well as for helpful discussions that were essential to this project's completion.

References

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

Berkhout, P. and Plug, E. (2004). A bivariate Poisson count data model using conditional probabilities. *Statistica Neerlandica* **58,** 349–364.

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

Besag, J., York, J. C., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* **43,** 1–59.

Billheimer, D., Cardoso, T., Freeman, E., Guttorp, P., Ko, H.-W., 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 survival data (with discussion). In *Bayesian Statistics 7*, J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith, and M. West (eds), 45–63. Oxford: Oxford University Press.

Carlin, B. P. and Louis, T. A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd edition. Boca Raton, Florida: Chapman and Hall/CRC Press.

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

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,** 11–25.

Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995). Efficient parametrizations for normal linear mixed models. *Biometrika* **82,** 479–488.

Gelfand, A. E., Schmidt, A. M., Banerjee, S., and Sirmans, C. F. (2004). Multivariate spatial process models: Conditional and unconditional Bayesian approaches using coregionalization (with discussion). *Test* **51,** 1–57.

Harville, D. D. (1997). *Matrix Algebra from a Statistician's Perspective*. New York: Springer-Verlag.

Held, L., Natário, I., Fenton, S. E., Rue, H., and Becker, N. (2005). Towards joint disease mapping. *Statistical Methods in Medical Research* **14,** 61–82.

Kim, H., Sun, D., and Tsutakawa, R. K. (2001). A bivariate Bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model. *Journal of the American Statistical Association* **96,** 1506–1521.

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

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

Sain, S. R. and Cressie, N. (2002). Multivariate lattice models for spatial environmental data. *Proceedings of the American Statistical Association, Statistics and the Environment Section* [CD-ROM], 2820–2825. Alexandria, Virginia: American Statistical Association.

Schmidt, A. M. and Gelfand, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research* **108**(D24), article 8783. Available at http://www.agu.org/pubs/crossref/2003.../2002JD002905.shtml.

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

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

Zhu, L., Carlin, B. P., and Gelfand, A. E. (2003). Hierarchical regression with misaligned spatial data: Relating ambient ozone and pediatric asthma ER visits in Atlanta. *Environmetrics* **14,** 537–557.

Received February 2004. *Revised October* 2004.
Accepted January 2005.