Statistica Sinica **22** (2012), 231-248
doi:http://dx.doi.org/10.5705/ss.2010.106

DEFAULT BAYESIAN ANALYSIS FOR MULTIVARIATE GENERALIZED CAR MODELS

Sarat C. Dass, Chae Young Lim and Tapabrata Maiti

Michigan State University

Abstract: In recent years, multivariate spatial models have been proven to be an effective tool for analyzing spatially related multidimensional data arising from a common underlying spatial process. The Bayesian analysis of these models is popular; however, the selection of an appropriate prior plays an important role in the inference. The two main contributions of this article are the development of shrinkage-type default priors for covariance matrices in these spatial models, and an innovative Gibbs sampling implementation that removes positive definiteness constraints when updating entries of the covariance matrix. The default prior elicitation is non-informative, but results in a proper posterior on the related parameter spaces. This elicitation not only provides robust inference (with respect to prior choice), but also provides improved estimation. In the computational step, the avoidance of sampling from restricted domains provides more stability and efficiency in the Gibbs implementation. Both simulations and data examples are provided to validate and illustrate the proposed methodology.

Key words and phrases: Conditional autoregressive models, default Bayesian analysis, generalized linear mixed models, Gibbs sampling, spectral decomposition.

1. Introduction

The issue of "health disparity" is central to the distribution of federal and state aid based on socio-economic indicators. Health disparity studies analyze how the health status of individuals vary across various socio-economic groups and spatial locations, in particular in relation to a specific disease. Multiple response variables are available as indicators of health status and, as a result, models for multivariate spatial lattice data are an indispensable tool for analyzing health disparity data. Recently, Greco, and Trivisano (2009), Zhang, Hodges, and Banerjee (2009), Jin, Banerjee, and Carlin (2007), Sain and Cressie (2007), Jin, Carlin, and Banerjee (2005), Gelfand and Vounatsou (2003), Carlin and Banerjee (2003), and Kim, Sun, and Tsutakawa (2001) explored multivariate spatial models for lattice data, adopting the Bayesian framework as the natural inferential approach.

Although the Bayesian inferential framework is a natural choice for spatial lattice data, one obvious obstacle is the choice of prior distribution for the model

parameters. The works mentioned are based on standard subjective and, at best, vague priors to account for the lack of subjective knowledge. Subjective specification of priors was the obvious drawback of introducing bias in the estimation procedure, the extent of which may not be easy to gauge in applications. A *default*, or non-informative prior, is therefore preferable for the Bayesian approach. In this paper, we develop a new *default* prior for parameters of generalized linear mixed models (GLMMs) in the multivariate spatial context, where the underlying spatial component has a conditionally autoregressive (CAR) model. This prior is an extension of the reference prior considered in Natarajan and Kass (2000) in the univariate non-spatial context. The simulation analysis indicates that, while subjective prior specifications are sensitive, the proposed prior is not. Since the default prior is not conjugate to the likelihood, our Bayesian computational scheme is different from previous approaches. We adopt a new parametrization of the multivariate CAR model based on the Cholesky and spectral decompositions of matrices; the advantage gained is the removal of positive definiteness constraints on the inverse covariance matrix.

The rest of the article is organized as follows: Section 2 introduces the multivariate CAR model in the spatial context. Section 3 discusses the motivation for developing the proposed default prior by taking spatial information into account. The resulting posterior distribution is shown to be proper for both complete and missing data cases. The Gibbs steps are outlined in Section 4. The Bayesian inferential procedure is illustrated through an application to cancer mortality data from the Surveillance, Epidemiology, and End Results (SEER) Program (`www.seer.cancer.gov`). Section 5 gives the numerical findings for both simulated and real data. This is followed by a brief discussion summarizing our findings in Section 6, and the Appendix.

2. Multivariate Generalized Linear Mixed Models

Assume there are n distinct sites on a spatial domain where observations on p variables are recorded. The multivariate data consist of the p-dimensional random vector $\mathbf{y}_j \equiv (y_{1j}, y_{2j}, \ldots, y_{pj})'$ for the jth site, for $j = 1, \ldots, n$. Corresponding to the response y_{ij}, denote by $x_{ij} = (x_{ij1}, x_{ij2}, x_{ijq_i})'$ the $q_i \times 1$ vector of explanatory variables. The following two-stage hierarchical model is considered for the distribution of the $np \times 1$ vector of all observables $\mathbf{y} \equiv (\mathbf{y}_1', \mathbf{y}_2', \ldots, \mathbf{y}_n')'$: The y_{ij} are independent with a density belonging to the exponential family $f_{ij}(y \,|\, \eta_{ij}) = C(y)\, \exp\{\, \eta_{ij} y - h_i(\eta_{ij})\}$, with canonical parameter η_{ij}, and normalizing constant $\exp\{\, h_i(\eta_{ij})\,\} = \int C(y)\, \exp\{\, \eta_{ij} y\,\}\, dy$; η_{ij} is related to covariate x_{ij} via $\eta_{ij} = x_{ij}'\beta_i + \epsilon_{ij}$, where β_i is a $q_i \times 1$ vector of regression coefficients and ϵ_{ij} are error random variables. The hierarchical specification is completed by $\boldsymbol{\epsilon} \sim N_{np}(0, \boldsymbol{D})$, where $\boldsymbol{\epsilon}_j \equiv (\epsilon_{1j}, \epsilon_{2j}, \ldots, \epsilon_{pj})'$ is the $p \times 1$ error vector at the jth spatial site, $\boldsymbol{\epsilon} \equiv (\boldsymbol{\epsilon}_1', \boldsymbol{\epsilon}_2', \ldots, \boldsymbol{\epsilon}_n')'$ is the $np \times 1$ vector of all the error variables, and

$\boldsymbol{D}$ is the covariance matrix (of dimension $np \times np$) of $\boldsymbol{\epsilon}$. Such models are called generalized linear mixed models (GLMMs; see McCulloch and Searle (2001)).

2.1. Multivariate Gaussian CAR

In the spatial context, the distribution of $\boldsymbol{\epsilon}$, and hence $\boldsymbol{D}$, can be given more structure based on the spatial neighborhood structure $\mathcal{N} = \{ N_j : j = 1, \ldots, n \}$, where N_j is the collection of spatial sites that are neighbors of site j. Such development is known in the literature and we only provide a brief discussion to introduce the relevant notation for the subsequent sections. For more details, see Banerjee, Gelfand, and Carlin (2004) and Sain and Cressie (2007). For the multivariate Gaussian CAR model, $\boldsymbol{\epsilon}$ follows $N_{np}(\mathbf{0}, \boldsymbol{D})$ with $\boldsymbol{D} = \{\text{Block}(-\boldsymbol{\Gamma}_j^{-1}\boldsymbol{\Lambda}_{jk})\}^{-1}$ for some $p \times p$ matrices $\boldsymbol{\Gamma}_j$ and $\boldsymbol{\Lambda}_{jk}$, where $j, k = 1, \ldots, n$. Symmetry and positive definiteness of D is guaranteed by requiring

$$\boldsymbol{\Lambda}_{jk}\,\boldsymbol{\Gamma}_k = \boldsymbol{\Gamma}_j\,\boldsymbol{\Lambda}'_{kj} \tag{2.1}$$

for all pairs (j, k), and $\text{Block}(-\boldsymbol{\Gamma}_j^{-1}\boldsymbol{\Lambda}_{jk})$ is positive definite with $\boldsymbol{\Lambda}_{jj} = -\mathbf{I}$.

The spatial neighborhood weights are defined as $w_{jk} = 1$ if j and k are neighbors, and let 0 otherwise, with $w_{jj} \equiv 0$ for $j = 1, \ldots, n$. Also, let $\boldsymbol{W} = ((w_{jk}))$ (of dimension $n \times n$), and $\boldsymbol{M}$ be the $n \times n$ diagonal matrix $\boldsymbol{M} = \text{diag}(w_{1+}, w_{2+}, \ldots, w_{n+})$ where $w_{j+} \equiv \sum_{k \in N_j} w_{jk}$.

A further parametrization, for computational as well as for practical analysis, is given by $\boldsymbol{\Gamma}_j = \boldsymbol{\Gamma}/w_{j+}$ and $\boldsymbol{\Lambda}_{jk} = (w_{jk}/w_{j+}) \cdot \boldsymbol{H}$, which entails a common covariance matrix $\boldsymbol{\Gamma}$ for all sites j and a common set of dependencies parameters $\boldsymbol{H}$, both rescaled by w_{j+}^{-1}. Such parametrization allows for sites with a larger number of neighbors to have lower variability, which is reasonable to expect. The symmetry requirement of (2.1) is equivalent to $\boldsymbol{H}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}\boldsymbol{H}'$ or, in other words, $\boldsymbol{F} \equiv \boldsymbol{\Gamma}^{-1/2}\boldsymbol{H}\boldsymbol{\Gamma}^{1/2}$ is symmetric, where $\boldsymbol{\Gamma}^{1/2}$ is the (unique) square root matrix of $\boldsymbol{\Gamma}$ and $\boldsymbol{\Gamma}^{-1/2}$ is its inverse. With the above re-parametrization, the inverse of $\boldsymbol{D}$ is

$$\boldsymbol{D}^{-1} = (\boldsymbol{I}_n \otimes \boldsymbol{\Gamma}^{-1/2})(\boldsymbol{M} \otimes \boldsymbol{I}_p - \boldsymbol{W} \otimes \boldsymbol{F})(\boldsymbol{I}_n \otimes \boldsymbol{\Gamma}^{-1/2}), \tag{2.2}$$

where $A \otimes B$ is the Kronecker product of matrices A and B. Positive definiteness of $\boldsymbol{D}^{-1}$ is ensured by restricting eigenvalues of $\boldsymbol{F} = \boldsymbol{\Gamma}^{-1/2}\boldsymbol{H}\boldsymbol{\Gamma}^{1/2}$ to be in $(-1, 1)$; a proof is presented in the Appendix. In the case when the eigenvalues of $\boldsymbol{F}$ are all 1, that is, $\boldsymbol{F} = \boldsymbol{I}_p$, $\boldsymbol{D}^{-1}$ is positive semi-definite. This is because $\boldsymbol{M} \otimes \boldsymbol{I}_p - \boldsymbol{W} \otimes \boldsymbol{I}_p = (\boldsymbol{M} - \boldsymbol{W}) \otimes \boldsymbol{I}_p$ and $(\boldsymbol{M} - \boldsymbol{W})\,\mathbf{1} = 0$, where $\mathbf{1}$ is the $n \times 1$ vector of ones. In this special case, the CAR model is equivalent to the intrinsically autoregressive (IAR) model.

The concatenated vector of all η_{ij}'s at the jth spatial location is denoted by $\boldsymbol{\eta}_j \equiv (\eta_{1j}, \eta_{2j}, \ldots, \eta_{pj})'$. Further, define $\boldsymbol{\eta} \equiv (\boldsymbol{\eta}'_1, \boldsymbol{\eta}'_2, \ldots, \boldsymbol{\eta}'_n)'$ to be the $np \times 1$ vector of all η_{ij} variables. It follows that $\boldsymbol{\eta} \sim N_{np}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{D})$, where

$\boldsymbol{\beta} = (\beta_1', \beta_2', \ldots, \beta_p')'$ is the $q \times 1$ concatenated vector of all regression coefficients with $q = \sum_{i=1}^{p} q_i$, and $\boldsymbol{X} = (\boldsymbol{X}_1' \, \boldsymbol{X}_2' \, \ldots \, \boldsymbol{X}_n')'$ is the $p\,n \times q$ design matrix with $\boldsymbol{X}_j = \text{Block Diagonal}(x_{1j}', x_{2j}', \ldots, x_{pj}')$ denoting the $p \times q$ design matrix at the jth spatial location for $j = 1, \ldots, n$.

2.2. Handling partial observations

Let $\mathcal{P}$ and $\mathcal{C}$ be all pairs (i, j) where y_{ij} is either partially observed or completely observed, respectively. Given $\boldsymbol{\eta}$, the conditional likelihood contribution corresponding to the partially observed data, $\mathcal{D}_{obs}$ say, can be written as

$$\ell(\mathcal{D}_{obs} \,|\, \boldsymbol{\eta}) = \prod_{(i,j)\in\mathcal{C}} f_{ij}(y_{ij} \,|\, \eta_{ij}) \prod_{(i,j)\in\mathcal{P}} F_{ij}(P_{ij} \,|\, \eta_{ij}), \tag{2.3}$$

with $F_{ij}(P_{ij} \,|\, \eta_{ij}) = \int_{y_{ij}\in P_{ij}} f_{ij}(y_{ij} \,|\, \eta_{ij})\, dy_{ij}$ denoting the contribution arising from the partial information that y_{ij} belongs to the set P_{ij}. No contribution is made to the likelihood if y_{ij} is missing, but we may incorporate covariates x_{ij} at the missing site in $\boldsymbol{\eta}$. The (unconditional) likelihood is

$$\ell(\mathcal{D}_{obs} \,|\, \boldsymbol{\beta}, \boldsymbol{F}, \boldsymbol{\Gamma}) = \int_{\boldsymbol{\eta}} \ell(\mathcal{D}_{obs} \,|\, \boldsymbol{\eta})\, f_0(\boldsymbol{\eta} \,|\, \boldsymbol{\beta},\, \boldsymbol{F}, \boldsymbol{\Gamma})\, d\boldsymbol{\eta}, \tag{2.4}$$

where f_0 denotes the density of $N_{np}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{D})$. Examples of partially observed data are common in rare diseases. For example, when mapping cancer incidences, certain counties do not report the exact number of incidences if the total number is less than a known threshold. In this case the partial information is $P_{ij} = \{\, y_{ij} < \tau \,\}$ where τ is the threshold.

3. Default Prior Elicitation

This section discusses appropriate default priors on the model parameters $\boldsymbol{\beta}$, $\boldsymbol{H}$, and $\boldsymbol{\Gamma}$. Since each β_i represents the regression effects to the mean of the observations y_{ij}, it is natural to elicit a standard "flat" non-informative prior on each β_i for $i = 1, \ldots, p$, namely, $\pi_N(\boldsymbol{\beta}) \propto 1$. It is also natural to consider a Jeffrey's type non-informative prior on $\boldsymbol{\Gamma}$ of the form $\pi_J(\boldsymbol{\Gamma}) \propto (\det(\boldsymbol{\Gamma}))^{-(p+1)/2}$. However, a Jeffrey's type prior leads to an improper posterior distribution.

Theorem 1. *Let $\pi_0(\boldsymbol{H})$ be a proper prior on $\boldsymbol{H}$. The default prior specification*

$$\pi_0(\boldsymbol{\beta}, \boldsymbol{H}, \boldsymbol{\Gamma}) = \pi_N(\boldsymbol{\beta}) \times \pi_J(\boldsymbol{\Gamma}) \times \pi_0(\boldsymbol{H}) \tag{3.1}$$

gives a posterior distribution on the parameters that is improper.

Proof: We refer the reader to a proof in the Appendix. The consequence of Theorem 1 is that new motivation is required for the development of a default

prior on $\mathbf{\Gamma}$ that would make the posterior proper. We discuss the development of such a prior in the subsequent paragraphs.

Our justification for the default prior comes from looking at the conditional update of each η_{ij} given the data y_{ij} and the rest of the $\boldsymbol{\eta}$ elements. In the normal-normal case, one can explicitly derive an expression for the weights that represent the contribution of the data, y_{ij}, and the rest of the $\boldsymbol{\eta}$ to the conditional mean of η_{ij}. However, in the case of non-conjugate GLMMs, it is not possible to obtain a closed form expression for the weights. An approximate approach can be considered based on a quadratic expansion of the exponential family pdf $f_{ij}(y \,|\, \eta_{ij}) = C(y)\, \exp\{\, \eta_{ij} y - h_i(\eta_{ij})\}$, to yield a similar analysis as in the normal-normal model.

We consider the prediction of the vector $\boldsymbol{\eta}_j$ given $\boldsymbol{y}_j$ and $\boldsymbol{\eta}_{-j}$ (the rest of the η_{ij}'s excluding the ones at site j). From Taylor's expansion, we have

$$h_i(\eta_{ij}) \approx h_i(\eta^*_{ij}) + (\eta_{ij} - \eta^*_{ij}) h_i^{(1)}(\eta^*_{ij}) + \frac{1}{2}(\eta_{ij} - \eta^*_{ij})^2 h_i^{(2)}(\eta^*_{ij}), \tag{3.2}$$

where $\boldsymbol{\eta}^*_j \equiv (\eta^*_{1j}, \ldots, \eta^*_{pj})$ is a fixed value. Substituting (3.2) in $\prod_{i=1}^p \exp\{\, \eta_{ij} y_{ij} - h_i(\eta_{ij})\,\}$, the latter simplifies, after completing squares, to

$$\exp\left\{ -\frac{1}{2} (\boldsymbol{\eta}_j - \boldsymbol{\eta}^*_j(\boldsymbol{y}_j))' \boldsymbol{H}_j^{(2)} (\boldsymbol{\eta}_j - \boldsymbol{\eta}^*_j(\boldsymbol{y}_j)) \right\} \tag{3.3}$$

with $\boldsymbol{\eta}^*_j(\boldsymbol{y}_j) \equiv \big(\boldsymbol{H}_j^{(2)}\big)^{-1} (\boldsymbol{y}_j - \boldsymbol{h}_j^{(1)} + \boldsymbol{H}_j^{(2)} \boldsymbol{\eta}^*_j)$, where $\boldsymbol{h}_j^{(1)} = \big(h_1^{(1)}(\eta^*_{1j}), \ldots, h_p^{(1)}(\eta^*_{pj})\big)'$ is the $p \times 1$ vector of first derivatives and $\boldsymbol{H}_j^{(2)} = \operatorname{diag}\big(h_1^{(2)}(\eta^*_{1j}), h_2^{(2)}(\eta^*_{2j}), \ldots, h_p^{(2)}(\eta^*_{pj})\big)$ is the diagonal matrix of all second derivatives of h_i, $i = 1, \ldots, p$, evaluated at $\boldsymbol{\eta}^*_j$. Now $\boldsymbol{\eta} \sim N_{np}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{D})$, which implies that the conditional distribution of $\boldsymbol{\eta}_j$ given $\boldsymbol{\eta}_{-j}$ is

$$\exp\left\{ -\frac{w_{j+}}{2} (\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^{CAR})' (\mathbf{\Gamma}^{-1}) (\boldsymbol{\eta}_j - \boldsymbol{\eta}_j^{CAR}) \right\} \tag{3.4}$$

for some $\boldsymbol{\eta}_j^{CAR}$. Combining (3.3) and (3.4), the conditional mean of $\boldsymbol{\eta}_j$ (again by completion of squares) is $(\boldsymbol{H}_j^{(2)} + \mathbf{\Gamma}^{-1} w_{j+})^{-1} \boldsymbol{H}_j^{(2)} \boldsymbol{\eta}^*_j(\boldsymbol{y}_j) + (\boldsymbol{H}_j^{(2)} + \mathbf{\Gamma}^{-1} w_{j+})^{-1} \mathbf{\Gamma}^{-1} w_{j+} \boldsymbol{\eta}_j^{CAR}$ with (matrix) weights $\boldsymbol{W}_{1j} = (\boldsymbol{H}_j^{(2)} + \mathbf{\Gamma}^{-1} w_{j+})^{-1} \boldsymbol{H}_j^{(2)}$ and $\boldsymbol{W}_{2j} = \boldsymbol{I} - \boldsymbol{W}_{1j}$ corresponding to the direct estimate $\boldsymbol{y}_j$ and the population mean, respectively. Since $\boldsymbol{W}_{1j}$ varies with j, we first replace it by its average across all the sites. Thus, we set $w_0 = \frac{1}{np} \sum_j \operatorname{trace}(\boldsymbol{H}_{\boldsymbol{j}}^{(\mathbf{2})} / w_{j+})$. Substituting $\boldsymbol{H}_j^{(2)} / w_{j+}$ by its average $w_0 \, \boldsymbol{I}_p$ in the expression of W_{1j} above, we get the matrix

$$\boldsymbol{U} \equiv \big(w_0 \boldsymbol{I}_p + \mathbf{\Gamma}^{-1} \big)^{-1} w_0 \boldsymbol{I}_p = (\, w_0 \mathbf{\Gamma} + \boldsymbol{I}_p)^{-1} w_0 \mathbf{\Gamma}.$$

Note that $0 \le \boldsymbol{U} \le \boldsymbol{I}_p$ in terms of positive definiteness, with $\det(\boldsymbol{U})$ representing the volume of the weight matrix $\boldsymbol{U}$. Our proposal is to induce a prior on $\mathbf{\Gamma}$ such

that the prior on $\det(\boldsymbol{U})$ is uniform. A similar technique was used by Daniels (1999) and Natarajan and Kass (2000) in the univariate non-spatial context.

To get this uniform distribution on $\boldsymbol{U}$, note that the multivariate Beta family of distributions, given by $f(\boldsymbol{U} \mid a, b) = C(\det(\boldsymbol{U}))^{a-(p+1)/2}(\det(\boldsymbol{I}_p - \boldsymbol{U}))^{b-(p+1)/2}$ with $a > (p-1)/2$ and $b > (p-1)/2$, forms a class of priors on $\boldsymbol{U}$. The uniform prior is obtained by setting $a = b = (p+1)/2$. The resulting prior on $\boldsymbol{\Gamma}$ is

$$\pi_{UV}(\boldsymbol{\Gamma}) = \det\left(\boldsymbol{I}_p + w_0\boldsymbol{\Gamma}\right)^{-(p+1)}. \tag{3.5}$$

This is also the prior developed in Natarajan and Kass (2000) leading to shrinkage-type estimators in the non-spatial context. The uniform volume prior is proper from Theorem 2 of Natarajan and Kass (2000).

The prior on $\boldsymbol{H}$ is induced via $\boldsymbol{F}$. The prior on $\boldsymbol{F}$ is taken to be independent of $\boldsymbol{\Gamma}$ and is constructed as follows: Writing the spectral decomposition of $\boldsymbol{F}$ as

$$\boldsymbol{F} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}', \tag{3.6}$$

we put a uniform prior on $\boldsymbol{Q}$; to have positive definiteness, we put a uniform prior $U(-1, +1)$ on the eigenvalues in $\boldsymbol{\Lambda}$. The default prior on $(\boldsymbol{\beta}, \boldsymbol{F}, \boldsymbol{\Gamma})$ is thus

$$\pi_0(\boldsymbol{\beta}, \boldsymbol{F}, \boldsymbol{\Gamma}) = \pi_N(\boldsymbol{\beta}) \times \pi_{UV}(\boldsymbol{\Gamma}) \times \frac{1}{2^p}. \tag{3.7}$$

For each $i = 1, \ldots, p$, the design matrix corresponding to the ith response variable is the $n \times q_i$ matrix $\tilde{\boldsymbol{X}}_i = (x_{i1}, x_{i2}, \ldots, x_{in})'$. The submatrix $\tilde{\boldsymbol{X}}_{\mathcal{C}_i}$ is formed by taking all rows j of $\tilde{\boldsymbol{X}}_i$ for which $(i, j) \in \mathcal{C}$.

Theorem 2. *Assume that f_{ij}, and F_{ij} in* (2.3) *are bounded above by a constant independent of η_{ij} for each pair (i, j). Under* (3.7)*, the posterior is proper if there exists q_i linearly independent row vectors in $\tilde{\boldsymbol{X}}_{\mathcal{C}_i}$, for each $i = 1, \ldots, p$, such that*

$$\int_{\boldsymbol{\Gamma}}\int_{\boldsymbol{F}}\int_{\boldsymbol{\beta}}\int_{\boldsymbol{\eta}}\Big(\prod_{i=1}^{p}\prod_{j=1}^{q_i} f_{ij}(y_{ij} \mid \eta_{ij})\Big) f_0(\boldsymbol{\eta} \mid \boldsymbol{\beta}, \boldsymbol{F}, \boldsymbol{\Gamma}) d\boldsymbol{\eta}\pi_0(\boldsymbol{\beta}, \boldsymbol{F}, \boldsymbol{\Gamma}) d\boldsymbol{\beta} d\boldsymbol{F} d\boldsymbol{\Gamma} < \infty, \tag{3.8}$$

where f_0 is as in (2.4) *and π_0 is the prior at* (3.7)*.*

Remark 1. Under the assumptions of Theorem 2, $0 \le f_{ij} \le A$ and $0 \le F_{ij} \le B$, say. In our applications, f_{ij} is taken to be either a Poisson pmf or a normal pdf. For the Poisson (or generally for a discrete distribution), it easily follows that $A = B = 1$, independent of i and j. When f_{ij} is normal with mean η_{ij} and fixed standard deviation σ_0, $A = 1/\sqrt{2\pi}\sigma_0$ and $B = 1$, and independent of i and j. Generally for densities, the bound A needs to be established on a case-by-case basis.

Remark 2. Theorem 2 shows that propriety can be achieved if there are at least q_i sites on the lattice for which the y_{ij}s are completely observed. The only further check that we have to perform is to see if the design matrix corresponding to those sites $\tilde{\boldsymbol{X}}_{\mathcal{C}_i}$ is non-singular. This is a requirement for each i, so the conditions of Theorem 2 can be checked separately for each $i = 1, \ldots, p$.

4. Bayesian Inference

A slightly different (yet equivalent) parametrization of the spatial multivariate GLMM is considered in developing the Gibbs sampler. Let $\boldsymbol{L}$ denote the lower triangular matrix obtained from the Cholesky decomposition of $\boldsymbol{\Gamma}^{-1} = \boldsymbol{L}\boldsymbol{L}'$. Define the matrix $\boldsymbol{B} \equiv \boldsymbol{L}\boldsymbol{Q}$ with entries $\boldsymbol{B} = ((b_{uv}))_{u,v=1}^{p}$, where $\boldsymbol{Q}$ as in (3.6). Then we have $\boldsymbol{D}^{-1} = (\boldsymbol{I}_n \otimes \boldsymbol{L})(\boldsymbol{M} \otimes \boldsymbol{I}_p - \boldsymbol{W} \otimes \boldsymbol{F})(\boldsymbol{I}_n \otimes \boldsymbol{L}') = (\boldsymbol{I}_n \otimes \boldsymbol{L})(\boldsymbol{M} \otimes \boldsymbol{I}_p - \boldsymbol{W} \otimes \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}')(\boldsymbol{I}_n \otimes \boldsymbol{L}') = (\boldsymbol{I}_n \otimes \boldsymbol{L}\,\boldsymbol{Q})(\boldsymbol{M} \otimes \boldsymbol{I}_p - \boldsymbol{W} \otimes \boldsymbol{\Lambda})(\boldsymbol{I}_n \otimes \boldsymbol{Q}'\,\boldsymbol{L}') = (\boldsymbol{I}_n \otimes \boldsymbol{B})(\boldsymbol{M} \otimes \boldsymbol{I}_p - \boldsymbol{W} \otimes \boldsymbol{\Lambda})(\boldsymbol{I}_n \otimes \boldsymbol{B}')$.

The advantage of the re-parametrization of $\boldsymbol{D}^{-1}$ in terms of $\boldsymbol{B}$ is that the entries of $\boldsymbol{B}$ are unconstrained. Note that it is possible to obtain $\boldsymbol{Q}$ and $\boldsymbol{L}$ uniquely from $\boldsymbol{B}$ using the QR decomposition of $\boldsymbol{B}' = \mathcal{Q}\mathcal{R}$, where $\mathcal{Q}$ is orthogonal and $\mathcal{R}$ is upper triangular. It follows that $\boldsymbol{L} = \mathcal{R}'$ and $\boldsymbol{Q} = \mathcal{Q}'$. The four main steps of the Gibbs sampler are the following.

• **Update η_{ij},** based on the (conditional) posterior density of η_{ij} given the rest of the parameters:

$$\pi(\eta_{ij} \mid \ldots) \propto \exp\{\, \eta_{ij} y_{ij} - h_i(\eta_{ij}) - \frac{A_{ij}}{2}(\eta_{ij} - \eta_{ij}^*)^2\}, \tag{4.1}$$

where $\eta_{ij}^* \equiv x_{ij}'\beta_i + \epsilon_{ij}^*$, and ϵ_{ij}^* has the expression $\epsilon_{ij}^* = (1/\sum_{v=1}^{p} b_{iv}^2)\epsilon_0$,

$$\epsilon_0 = \underbrace{\sum_{v=1}^{p} b_{iv}^2 \lambda_v \sum_{k \in N_j} \frac{w_{jk}}{w_{j+}} \epsilon_{ik}}_{(A)} - \sum_{v=1}^{p} \sum_{u=1, u \neq i}^{p} b_{iv} b_{uv} \epsilon_{uj} + \sum_{v=1}^{p} \sum_{u=1, u \neq i}^{p} b_{iv} b_{uv} \lambda_v \sum_{k \in N_j} \frac{w_{jk}}{w_{j+}} \epsilon_{uk}, \tag{4.2}$$

with $\epsilon_{uv} = \eta_{uv} - x_{uv}'\beta_{uv}$ for all $(u, v) = 1, \ldots, p$ except for $(u, v) = (i, j)$, and $A_{ij} = w_{j+} \sum_{v=1}^{p} b_{iv}^2$. The update of η_{ij} is done by a griddy grid sampler based on (4.1) for each fixed pair (i, j), and cycling through all the combinations of $(i, j) \in \{(1, 1), \ldots, (1, n), \ldots, (p, n)\}$. Note that when y_{ij} is missing, term (A) in equation (4.1) is absent. When y_{ij} is partially observed, it is first treated as missing and then updated based on the truncated distribution $f_{ij}(\cdot \mid \eta_{ij})$ given that $y_{ij} \in P_{ij}$.

• **Update $\boldsymbol{\beta}$:** The update of $\boldsymbol{\beta}$ requires a re-ordering of the variables involved. For each $i = 1, \ldots, p$, let $\boldsymbol{\eta}_i^r = (\eta_{i1}, \eta_{i2}, \ldots \eta_{in})'$ and $\boldsymbol{\epsilon}_i^r = (\epsilon_{i1}, \epsilon_{i2}, \ldots \epsilon_{in})'$, and let

$\boldsymbol{\eta}^r \equiv ((\boldsymbol{\eta}_1^r)', (\boldsymbol{\eta}_2^r)', \ldots, (\boldsymbol{\eta}_p^r)')$ and $\boldsymbol{\epsilon}^r \equiv ((\boldsymbol{\epsilon}_1^r)', (\boldsymbol{\epsilon}_2^r)', \ldots, (\boldsymbol{\epsilon}_p^r)')$ denote the $np \times 1$ vector of re-ordered entries from $\boldsymbol{\eta}$ and $\boldsymbol{\epsilon}$, respectively. The covariance matrix of $\boldsymbol{\epsilon}^r$ is subsequently a re-ordered version of $\boldsymbol{D}$ given by $(\boldsymbol{D}^r)^{-1} = (\boldsymbol{B} \otimes \boldsymbol{I}_n)(\boldsymbol{I}_p \otimes \boldsymbol{M} - \boldsymbol{\Lambda} \otimes \boldsymbol{W})(\boldsymbol{B}' \otimes \boldsymbol{I}_n)$. If $\tilde{\boldsymbol{X}} = ((\text{Block Diagonal}\{\tilde{\boldsymbol{X}}_i\}))$ denotes the block diagonal matrix consisting of the design matrices for the ith response variable for $i = 1, \ldots, p$, the conditional posterior distribution of $\boldsymbol{\beta}$ is multivariate normal with mean $\boldsymbol{\mu}_{\boldsymbol{\beta}}$, and covariance matrix $\boldsymbol{S}_{\boldsymbol{\beta}}$, where

$$\boldsymbol{\mu}_{\boldsymbol{\beta}} = (\tilde{\boldsymbol{X}}'(\boldsymbol{D}^r)^{-1}\tilde{\boldsymbol{X}})^{-1}(\tilde{\boldsymbol{X}}'(\boldsymbol{D}^r)^{-1}\boldsymbol{\eta}^r) \quad \text{and} \quad \boldsymbol{S}_{\boldsymbol{\beta}} = (\tilde{\boldsymbol{X}}'(\boldsymbol{D}^r)^{-1}\tilde{\boldsymbol{X}})^{-1}. \tag{4.3}$$

• **Update $\boldsymbol{\Lambda}$:** The re-parametrization in terms of $\boldsymbol{B}$ and $\boldsymbol{\Lambda}$ means the diagonal entries of $\boldsymbol{\Lambda}$ can be updated independently of each other. Consider the $p \times n$ matrix, $\boldsymbol{\Upsilon}$, constructed by putting ϵ_{ij} in its ith row and jth column. Define a new matrix $p \times n$ matrix $\boldsymbol{E}$ as $\boldsymbol{E} = \boldsymbol{B}'\boldsymbol{\Upsilon}$, and let e_i' be the ith row of $\boldsymbol{E}$ for $i = 1, \ldots, p$. The conditional posterior density of λ_k is given by

$$\pi(\lambda \mid \ldots) \propto \exp\left\{-\frac{1}{2}\, e_k'(\boldsymbol{M} - \lambda \boldsymbol{W})\, e_k\right\} (\det(\boldsymbol{M} - \lambda \boldsymbol{W}))^{1/2} \tag{4.4}$$

on $-1 \leq \lambda_k \leq 1$, independently for each $k = 1, \ldots, p$. The update of λ_k is based on a griddy grid sampler using (4.4) for each fixed k.

• **Update $\boldsymbol{B}$:** The conditional posterior density of $\boldsymbol{B}$ has the expression $\pi(\boldsymbol{B} \mid \ldots)$

$$\propto \exp\left\{-\frac{1}{2}\sum_{k=1}^{p} e_k'(\boldsymbol{M} - \lambda_k \boldsymbol{W})e_k\right\} \times \det\left(w_0 \boldsymbol{I} + \boldsymbol{B}\boldsymbol{B}'\right)^{-(p+1)} \times \det(\boldsymbol{B}\boldsymbol{B}')^{(n+1)/2}, \tag{4.5}$$

where $e_k' \equiv e_k'(\boldsymbol{B})$ is as defined in (4.4) but now viewed as a function of $\boldsymbol{B}$. The update of $\boldsymbol{B}$ is carried out by updating each entry b_{uv} one at a time. Note that for each (u, v), the conditional posterior density of b_{uv} has the same expression as (4.5) up to a constant of proportionality. Thus, b_{uv} is updated using a griddy grid sampler based on (4.5). The derivation of (4.4) and (4.5) is given in the Appendix.

5. Experimental Results

5.1. Simulation study

We have conducted extensive simulation studies to check the performance of our methodology. The experimental settings closely mimic county-level data for southern lower Michigan as obtained from the SEER database. We took $p = 2$ with y_{1j} and y_{2j} representing binomial responses on $n = 40$ spatial sites, with neighborhood structure determined by the adjacency information

among 40 southern lower Michigan counties. Given $\boldsymbol{\eta}$, the observed data were $y_{ij} \sim$ Binomial $(T_{ij}, [e^{\eta_{ij}}]/[1+e^{\eta_{ij}}])$, for $i = 1, 2$ and $j = 1, \ldots, 40$, independent of each other; T_{ij}, the total number of trials was selected independently from a Uniform$(22, 48)$ distribution. The distribution of $\boldsymbol{\eta}$ was multivariate Gaussian CAR with the following true parameter specifications: $\boldsymbol{\beta}_1 = (-1, 0.3)'$, $\boldsymbol{\beta}_2 = (-0.5, -0.2)'$, $\boldsymbol{\Gamma} = \begin{pmatrix} 0.5 & 0.1 \\ 0.1 & 0.125 \end{pmatrix}$ and $\boldsymbol{F} = \begin{pmatrix} 0.7 & -0.1 \\ -0.1 & 0.2 \end{pmatrix}$. The first components of $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ correspond to the intercept terms. Here $q_1 = q_2 = 2$ and additional covariate information was gathered from independent normal distributions: $X_1 \sim N(0, \sigma^2 = 0.3)$ and $X_2 \sim N(1, \sigma^2 = 0.5)$.

Two types of priors were considered for $\boldsymbol{\Gamma}$: (a) the default prior $\pi_{UV}(\boldsymbol{\Gamma}) \propto \det(\boldsymbol{I_p} + \boldsymbol{\omega_0}\boldsymbol{\Gamma})^{-(p+1)}$; (b) the proper inverse Wishart given by $\pi_{IW}(\boldsymbol{\Gamma}) = IW(\rho, \rho\boldsymbol{A})$, where $\rho \geq p$. The inverse Wishart distribution is a generalization of the inverse gamma for the variance parameter in a multivariate setting. If $\boldsymbol{\Gamma} \sim IW(m, \boldsymbol{\Psi})$, the expectation and variance of entries of $\boldsymbol{\Gamma}$ are given by $E(\boldsymbol{\Gamma}_{kl}) = \Psi_{kl}/(m-p-1)$ and $var(\boldsymbol{\Gamma}_{kl}) = [(m-p+1)\boldsymbol{\Psi}_{kl}^2 + (m-p-1)\boldsymbol{\Psi}_{kk}\boldsymbol{\Psi}_{ll}]/[(m-p)(m-p-1)^2(m-p-3)]$, where $\boldsymbol{\Gamma}_{kl}$ and $\boldsymbol{\Psi}_{kl}$ are the (k,l)th entries of $p \times p$ matrices $\boldsymbol{\Gamma}$ and $\boldsymbol{\Psi}$, respectively. When ρ is large, $E(\boldsymbol{\Gamma}) \approx \boldsymbol{A}$, and $var(\boldsymbol{\Gamma}_{ij}) \approx 1/\rho$, leading to a high concentration of probability around the initial guess of $\boldsymbol{A}$. Thus, this prior does not represent non-informative prior knowledge. Prior (b) was proposed by Sain and Cressie (2007) as the choice of vague prior for $\boldsymbol{\Gamma}$ when ρ is large, which is not the case (actually, Sain and Cressie (2007) put a prior on $\boldsymbol{\Gamma}^{-1}$ as Wishart$(\rho, (\rho\boldsymbol{A})^{-1})$, but this is equivalent to choice (b) since $\boldsymbol{\Gamma} \sim IW(m, \boldsymbol{\Psi})$ iff $\boldsymbol{\Gamma^{-1}} \sim W(m, \boldsymbol{\Psi^{-1}})$).

The priors (a) and (b) for $\boldsymbol{\Gamma}$ above in turn induce priors on $\boldsymbol{B}$. This is based on the transformations $\boldsymbol{\Gamma} \to \boldsymbol{\Gamma^{-1}}$, $\boldsymbol{\Gamma^{-1}} \to \boldsymbol{LL'}$, and $(\boldsymbol{L}, \boldsymbol{Q}) \to \boldsymbol{B}$. The derivation of the Jacobian for the composition transformation from $(\boldsymbol{\Gamma}, \boldsymbol{Q}) \to \boldsymbol{B}$ is given in the Appendix. The priors on $\boldsymbol{B}$ turn out to be $\pi_{UV}(\boldsymbol{B}) = \det(\boldsymbol{\omega_0 I_p} + \boldsymbol{BB'})^{-(p+1)} \det(\boldsymbol{BB'})^{1/2}$ and $\pi_{IW}(\boldsymbol{B}) = \exp\{-(\rho/2)tr(\boldsymbol{ABB'})\} \times \det(\boldsymbol{BB'})^{(\rho-p)/2}$, respectively, for the priors (a) and (b) for $\boldsymbol{\Gamma}$. Prior choices for $\boldsymbol{\beta}$ are (i) the default non-informative constant prior $\pi_N(\boldsymbol{\beta}) \propto 1$, and (ii) the proper subjective prior $\pi_G(\boldsymbol{\beta}_k) \sim N(0, \sigma_k^2 \boldsymbol{I}_{q_k})$, independently for each $k = 1, \ldots, p$. Using (ii), it is easy to see that the posterior for $\boldsymbol{\beta}$ is $N(\boldsymbol{\mu}, \boldsymbol{S})$, where $\boldsymbol{\mu} = \boldsymbol{SX'}(\boldsymbol{D^r})^{-1}\boldsymbol{\eta^r}$, $\boldsymbol{S^{-1}} = \boldsymbol{X'}(\boldsymbol{D^r})^{-1}X + \boldsymbol{\Sigma^{-1}}$ with $\boldsymbol{\Sigma} =$ Block Diagonal$(\sigma_k^2 \boldsymbol{I}_{q_k})$. Thus we investigate three prior choices: (I) $\pi_{UV}(\boldsymbol{B})$ and $\pi_N(\boldsymbol{\beta})$ and (II) $\pi_{IW}(\boldsymbol{B})$ and $\pi_G(\boldsymbol{\beta}_k)$ with $\rho = 5$, $A = I_p$ and $\sigma_k^2 = 100$, and (III) $\pi_{IW}(\boldsymbol{B})$ and $\pi_G(\boldsymbol{\beta}_k)$ with $\rho = 100{,}000$, $A = I_p$ and $\sigma_k^2 = 100$. Here (I) is our proposed default prior, (II) is a proper prior with small ρ, and (III) is a prior with large ρ. Note that (II) is weakly informative whereas (III) is highly informative.

The Gibbs sampler was run for $10{,}000$ iterations and checked for convergence using traceplots and the R-statistic of Gelman and Rubin (1992). We established

Table 1. Deviations measures for averages over $\boldsymbol{\beta}$, $\boldsymbol{\Gamma}$, $\boldsymbol{F}$ and $\boldsymbol{\eta}$ for the complete data case.

θ	$\boldsymbol{\beta}$			$\boldsymbol{\Gamma}$			$\boldsymbol{F}$			$\boldsymbol{\eta}$		
Prior	(I)	(II)	(III)	(I)	(II)	(III)	(I)	(II)	(III)	(I)	(II)	(III)
RMAD	0.253	0.250	0.258	0.799	1.660	2.994	0.956	0.982	0.990	0.557	0.533	0.558
MSE	0.017	0.017	0.017	0.045	0.113	0.336	0.166	0.167	0.169	0.066	0.071	0.085
CP	0.861	0.894	0.925	0.896	0.618	0.000	0.881	0.859	0.859	0.874	0.933	0.932
W	0.383	0.430	0.478	0.546	0.642	0.027	1.236	1.161	1.107	0.787	0.943	1.040

convergence for all experiments with $5,000$ iterations; all diagnostic plots were satisfactory. Outputs from the Gibbs chains were used to compute different statistics to validate and compare the proposed Bayesian methodology. The three prior choices were compared in terms of their ability to derive consistent estimation and prediction results. The deviation measures of comparisons were (1) relative mean absolute deviation (RMAD), (2) mean square error (MSE), (3) empirical 90% highest posterior density (HPD) coverage probabilities (CP), and (4) width of the 90% HPD set (W). Formulas for a generic quantity θ are given by $RMAD_\theta = (E(\theta) - \theta_0)/\theta_0$, $MSE_\theta = E(\theta - \theta_0)^2$, $CP_\theta = P\{\theta_0 \in HPD(\theta)\}$, and W the width of $HPD(\theta)$; in the RHS of each expression, θ represents a sample (or samples) from the posterior, $HPD(\theta)$ is the 90% HPD set calculated based on θ samples, and θ_0 is the true value set in the simulation experiments. We used 500 replications in each experiment and report the averages of the deviation measures. The computational time for all 500 replications in each experimental setup was approximately $20 - 25$ hours. All computations were carried out using an HP ProLiant DL160 machine (a cluster of nodes) with 8 Intel Xeon cores and 24GB of memory at the High Performance Computing Center (HPCC) in Michigan State University.

Table 1 reports a summary of all deviation measures. For convenience of understanding, we report averages over specific components of the unknown parameters; for example, the $\boldsymbol{\beta}$ column reports the average over all β components, $\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}$, and similarly for the other parameters $\boldsymbol{\Gamma}$ and $\boldsymbol{F}$. The last column reports averages over all $40 \times 2 = 80$ predicted values for $\boldsymbol{\eta}$; the deviation measures are calculated based on $\boldsymbol{\eta}_0$, the true values generated in each replication. Entries in Table 1 clearly show the sensitivity of standard prior distributions (i.e., priors (II) and (III)) used in the literature. For example, for a wrong choice of ρ and $\boldsymbol{A}$ in π_{IW}, the coverage can be even 0 along with high MSE and RMAD. This effect can be reduced with a more sensible prior choice, for example, choice (II). On the other hand, π_{UV} always provides sensible results. One might notice that the sensitivity of priors (II) and (III) is highest for $\boldsymbol{\Gamma}$ compared to the other columns in Table 1. This is due to the fact that the prior for $\boldsymbol{\Gamma}$ changes significantly for the three choices (I-III), whereas we always use the default uniform

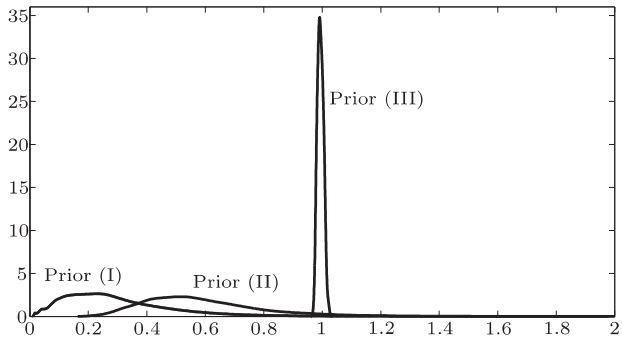


Figure 1. Posterior densities for Γ_{22} corresponding to the three prior choices.

prior for $\boldsymbol{F}$. The regression parameters $\boldsymbol{\beta}$ are less affected by the prior choice compared to $\boldsymbol{\Gamma}$ due to the fact that $\boldsymbol{\beta}$ is related to the mean parameter with large prior variance while $\boldsymbol{\Gamma}$ is related to dispersion. Nevertheless, the standard choice of Gaussian prior on $\boldsymbol{\beta}$ also appears to be somewhat sensitive, but not to the extent of $\boldsymbol{\Gamma}$. Although the $\boldsymbol{\eta}$ components are not fixed model parameters (i.e., they vary from county to county), their inference can also be sensitive to the different prior choices. To explain the discrepancies in the $\boldsymbol{\Gamma}$ entries, Figure 1 plots the posterior densities of Γ_{22}, the $(2,2)$th entry of $\boldsymbol{\Gamma}$, corresponding to the three different prior choices for a arbitrarily chosen replicate. Note that under (III), the prior mean is I_2 whereas the prior variance is 10^{-5}, making it highly concentrated on a value different from the true Γ_{22}; while we understand that the small prior variance is unreasonable, this choice is not uncommon (see, for example, Sain and Cressie (2007)). The situation improves under prior (II) where the prior mean is the same but the prior variance is 0.2, which gives a comparatively higher prior mass around the true Γ_{22}. Overall, the proposed default prior π_{UV} performed well in all respects. This prior is thus a robust choice. We also explored the choice $\boldsymbol{\Gamma} = \begin{pmatrix} 10 & 6 \\ 6 & 5 \end{pmatrix}$. The results were similar to those discussed here. The componentwise univariate spatial analysis was carried out and, as expected, the multivariate analysis had superior performance.

We also performed similar experiments with 10% missing observations. The results are reported in Table 2. Comparative trends similar to the complete data case with priors (I-III) were also observed here.

5.1. Data examples

The Bayesian inferential framework was applied to study bivariate dependence of a number of health-socio-economic indicators in the state of Michigan.

Table 2. Deviation measures for averages over $\boldsymbol{\beta}$, $\boldsymbol{\Gamma}$, $\boldsymbol{F}$, and $\boldsymbol{\eta}$ for the missing data case.

θ	$\boldsymbol{\beta}$			$\boldsymbol{\Gamma}$			$\boldsymbol{F}$			$\boldsymbol{\eta}$		
Prior	(I)	(II)	(III)	(I)	(II)	(III)	(I)	(II)	(III)	(I)	(II)	(III)
RMAD	0.256	0.254	0.263	1.037	1.782	2.992	0.916	0.928	0.929	0.539	0.559	1.132
MSE	0.018	0.018	0.018	0.074	0.143	0.336	0.155	0.157	0.158	0.071	0.075	0.090
CP	0.911	0.936	0.951	0.923	0.634	0.000	0.931	0.890	0.882	0.889	0.936	0.934
W	0.445	0.495	0.534	0.710	0.746	0.026	1.294	1.210	1.157	0.840	0.977	1.059

Two studies were conducted with different pairs of response variables: (1) lung cancer mortality incidence and poverty, and (2) lung cancer mortality and air quality index (AQI) measurements. Study (1) and (2) illustrate the complete and missing data applications, respectively.

We considered lung cancer mortality counts over a period 2001-2005 in Michigan counties obtained from SEER program (SEER*Stat Database: Mortality - All COD, Aggregated With County,Total U.S. (1990-2005) <Katrina/Rita Population Adjustment>). A poverty measure and other socio-economic county attributes were also obtained from the SEER program that is based on US Census 2000 (SEER*Stat Database: County Attributes - Total U.S., 1969-2005 Counties, `www.seer.cancer.gov/seerstat/variables/countyattribs`). The AQI measurements were obtained from the US Environmental Protection Agency (EPA) AQI report site (`http://www.epa.gov/air/data/monaqi.html?st~MI~Michigan`). We considered an average of median AQIs over a period 2001-2005. The AQI is a daily index value calculated for each air pollutant measured; the highest of these index values is the AQI value, and the pollutant responsible for the highest index value is the "Main Pollutant". The criteria pollutants used to calculate AQI are: CO (Carbon monoxide), NO_2 (Nitrogen dioxide), O_3 (Ozone), SO_2 (Sulfur dioxide), $PM_{2.5}$ (Particulate matter smaller than 2.5 micrometers), and PM_{10} (Particulate matter smaller than 10 micrometers). The highest possible AQI value is 500. A day having an AQI value 0 through 50 is considered 'Good' and a day having an AQI value 0 through 50 is considered 'Moderate'. The 32 counties in Michigan had AQI measurements over a period 2001-2005 in the 'Good' or 'Moderate' categories. Emission variables, in particular $PM_{2.5}$, was obtained from US EPA's National Emission Inventory (NEI) database (`http://www.epa.gov/air/data/emisdist.html?st~MI~Michigan`). Since emission data in this website is available only up to 2002, we considered the average of 2001-2002 emission data. The unit is tons per county. In each application, the Gibbs sampler was run for $10,000$ iterations and checked for convergence as in the simulated data. Posterior samples were obtained from the Gibbs chains

for computing the mean, standard deviation, and 90% HPD intervals for all the parameters.

5.1.1. Complete data example: Study 1

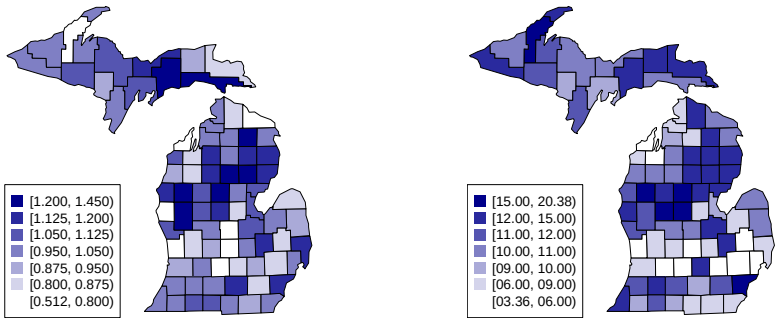
The mortality of lung cancer, the first component of the response variable in Study 1, is rare enough relative to the population in the 68 counties of lower Michigan that a Poisson distribution is appropriate. We write the model (conditional on η_{1j}) as

$$y_{1j} \overset{\text{ind}}{\sim} \text{Poisson}(E_j e^{\eta_{1j}}),\ i = 1, 2 \text{ and } j = 1, \ldots, 68,$$

where y_{1j} measures the observed number of deaths in county j, and E_j is the estimated population at risk; we assume the E_j are known and give a way for calculating them later. The poverty count, y_{2j}, is taken to be the second component of the response variable in Study 1. We model y_{2j} as a Binomial with the number of trials being the total county population and success probability $e^{\eta_{2j}}/(1 + e^{\eta_{2j}})$. The associated covariates for y_{1j} are the intercept, $\text{PM}_{2.5}$, the extents of urbanization, non-white population, and non-industry (these are measured as proportions). Covariates for poverty are the intercept, the extents of urbanization, and non-industry. Thus, $q_1 = 5$ and $q_2 = 3$.

To calculate E_j, we take each county's age distribution into account, as available from Census 2000. The expected age-adjusted number of deaths due in county j is $E_j = \sum_{k=1}^{m} \omega^k N_j^k$, for $j = 1, \ldots, 68$, where $\omega^k = \sum_{j=1}^{68} D_j^k / \sum_{j=1}^{68} N_j^k$ is the age-specific death rate due to lung cancer for age group k, and N_j^k and D_j^k are the total population at risk and the number of deaths in county j for age group k, respectively. The county level maps of the age-adjusted standardized mortality ratios (SMRs), Y_{1j}/E_j for lung cancer, shown in Figure 2, exhibit evidence of correlation over space. Figure 2 also gives the spatial distribution of poverty levels that can be seen to be highly spatially correlated with lung cancer (simple correlation between lung cancer and poverty is around 0.4).

We present summary conclusions of our analysis. The standard errors of the parameter estimates for prior (I) are smaller compared to priors (II) and (III). For example, the average standard error of regression coefficients for lung cancer is 0.61 under prior (I), and 0.65 and 1.28 for (II) and (III), respectively. For the variance component parameters, the standard errors and widths of HPD sets are comparable under priors (I) and (II). Note that the inference for $\mathbf{\Gamma}$ is highly misleading for prior (III), since most of the posterior probability is concentrated around the prior mean. Another difference is that the covariate $\text{PM}_{2.55}$ (related to the particle matter in the air) for lung cancer incidence is positive and significant under (I), whereas it is insignificant under priors (II) and (III). For prior (III),



(a) Obs SMR (b) Poverty(%)

Figure 2. Observed SMR of Lung Cancer and Poverty in Michigan.

the posterior mean takes a negative value that is not very realistic. For brevity other statistics, along with the smooth map of $\boldsymbol{\eta}$, are suppressed.

5.1.2. Missing data example: Study 2

In Study 2, AQI is taken as the second component of the response variable in place of poverty. Air pollutant monitoring stations are sparsely located in 32 of 68 lower Michigan counties and thus constitute missing information. The covariates for AQI are the intercept and the non-industrialization status of the county ($q_2 = 2$). We take y_{2j} to be normally distributed with mean η_{2j} and fixed standard deviation σ_0, estimated using our data and set at 0.1.

Results for the standard errors and width of HPD sets for the parameters are similar to the complete data case. There are two striking features in this application. First, the extent of urbanization for lung cancer incidence is negative and significant under (I), whereas it is positive under the other two priors (which may not be reasonable). Second, the regression coefficient for racial segregation (non-white) is significant under (I) and (II) but not under (III). This shows the sensitivity of the subjective elicitation under the missing data setup as well.

6. Summary and Conclusion

In the absence of reliable subjective information, the use of Jeffreys type non-informative priors or diffuse conjugate priors is popular. However, in the context of the hierarchical spatial multivariate CAR models, we have shown that none of these priors will work; the Jeffreys prior yields a posterior that is improper whereas the diffuse conjugate prior is highly sensitive. This led us to elicit priors on the model parameters that are close to Jeffreys, but still yield a proper posterior for inference. The development of prior elicitation can be thought of

as an extension of Natarajan and Kass (2000) in the spatial context. Besides the prior development, we propose some innovative computational techniques for the Gibbs implementation. Suitable transformations are made on the parameters to avoid sampling from restricted domains, thus providing more stability and efficiency in the Gibbs steps. The methodology has been extended to the case of missing responses in the multi-dimensional setup.

Acknowledgement

The authors thank the editor, associate editor, and referees for their valuable suggestions in improving this paper. The authors also wish to acknowledge the support of NSF DMS grants no. 0706385, 0961649 and 1106450 while conducting this research.

Appendix

Proof of the positive definiteness of $\boldsymbol{D}$. A necessary and sufficient condition for $\boldsymbol{D}$ to be positive definite (pd) is that $\boldsymbol{D}^{-1}$ be positive definite. Since $\boldsymbol{\Gamma}^{-1/2}$ is pd and hence non-singular, it follows from (2.2) that $(\boldsymbol{M}\otimes\boldsymbol{I}_p-\boldsymbol{W}\otimes\boldsymbol{F})$ is positive definite. The eigenvalues of $(\boldsymbol{M}\otimes\boldsymbol{I}_p-\boldsymbol{W}\otimes\boldsymbol{F})$ are those of $(\boldsymbol{M}\otimes\boldsymbol{I}_p-\boldsymbol{W}\otimes\boldsymbol{\Lambda})$, which are all eigenvalues of $\boldsymbol{M}-\lambda_k\boldsymbol{W}$, for $k=1,\ldots,p$. Now, requiring that $\boldsymbol{M}-\lambda_k\boldsymbol{W}$ be diagonally dominant (which implies positive definiteness), it follows that

$$|\lambda_k|\sum_{l\in N_j} w_{jl} < w_{j+} \quad \Rightarrow \quad |\lambda_k|\, w_{j+} < w_{j+} \quad \Rightarrow \quad |\lambda_k| < 1 \tag{A.1}$$

for all $k=1,\ldots,p$.

Proof of Theorem 1. In order to show that the posterior is improper, it is enough to show that the marginal of $\mathbf{y}\equiv(\mathbf{y}_1',\mathbf{y}_2',\ldots,\mathbf{y}_n')'$ does not exist; that is, the integration with respect to the variable $\boldsymbol{\eta}$ and the parameters $\boldsymbol{H}$, $\boldsymbol{\Gamma}$, and $\boldsymbol{\beta}$ is infinity. The marginal of $\mathbf{y}$ is

$$\begin{aligned}\mathbf{m}(\mathbf{y}) = \int_{\boldsymbol{\eta}}\int_{\boldsymbol{H}}\int_{\boldsymbol{\Gamma}}\int_{\boldsymbol{\beta}} &\left(\prod_{i=1}^{p}\prod_{j=1}^{n} f_i(y_{ij}\,|\,\eta_{ij})\right)\frac{1}{(2\pi)^{np/2}}(\det(\mathbf{D}))^{-1/2}\\ &\times\exp\left\{-\frac{1}{2}(\boldsymbol{\eta}-\mathbf{X}\boldsymbol{\beta})'\mathbf{D}^{-1}(\boldsymbol{\eta}-\mathbf{X}\boldsymbol{\beta})\right\}d\boldsymbol{\eta}\,\pi(\boldsymbol{H})d\boldsymbol{H}\frac{1}{(\det(\boldsymbol{\Gamma}))^{(p+1)/2}}d\boldsymbol{\Gamma}\,d\boldsymbol{\beta}.\end{aligned}$$

Let $\boldsymbol{\epsilon}=\boldsymbol{\eta}-\boldsymbol{X}\boldsymbol{\beta}$. Next, write $\det(\boldsymbol{D})=\det(\boldsymbol{\Gamma})^n\times g_0(\boldsymbol{H})$ for some function g_0 of $\boldsymbol{H}$, and note that the expression within the exponent can be simplified to $-(1/2)\mathrm{tr}(\boldsymbol{\Gamma}^{-1}\boldsymbol{S})$, where $\boldsymbol{S}=\sum_{j,l=1}^{n}\boldsymbol{H}_{jl}\boldsymbol{\epsilon}_j\boldsymbol{\epsilon}_l'$ with $\boldsymbol{H}_{jl}\equiv-w_{jl}\,\boldsymbol{H}$ if $j\neq l$, and

$\boldsymbol{H}_{jj} = w_{j+}\,\boldsymbol{I}_p$. Now, integrating with respect to $\boldsymbol{\Gamma}$, the marginal reduces to

$$\boldsymbol{m}(\boldsymbol{y}) \propto \int_{\boldsymbol{H}} \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\epsilon}} \left(\prod_{i=1}^{p} \prod_{j=1}^{n} f_i(y_{ij} \,|\, \epsilon_{ij} + x'_{ij}\beta_i) \right) \frac{1}{(\det(\boldsymbol{S}))^{n/2}} \, d\boldsymbol{\epsilon}\, d\boldsymbol{\beta}\, (g_0(\boldsymbol{H}))^{-1}\, \pi(\boldsymbol{H})\, d\boldsymbol{H},$$

ignoring proportionality constants. Let $\epsilon_{11} = u_{11}$ and $\epsilon_{ij} = u_{ij}u_{11}$ for $(i,j) \neq (1,1)$, with an associated Jacobian of u_{11}^{np-1}. With this substitution, $\boldsymbol{S} = u_{11}^2\, \boldsymbol{S}^*$, where $\boldsymbol{S}^* = \sum_{j,l=1}^{n} \boldsymbol{H}_{jl}\boldsymbol{U}_{jl}$ with $\boldsymbol{U}_{jl} = V_j\, V_l'$, $V_1 = (1, u_{21}, u_{31}, \ldots, u_{p1})'$, and $V_j = (u_{1j}, u_{2j}, \ldots, u_{pj})$ for $j \geq 2$. It follows that $\det(\boldsymbol{S}) = u_{11}^{2p} \det(\boldsymbol{S}^*)$. Then

$$\boldsymbol{m}(\boldsymbol{y}) = \int_{\boldsymbol{H}} \int_{\boldsymbol{\beta}} \int_{\boldsymbol{u}} \left(\prod_{i=1}^{p} \prod_{j=1}^{n} f_i(y_{ij} \,|\, u^*_{ij}u_{11} + x'_{ij}\beta_i) \right) \frac{1}{u_{11}} \frac{1}{(\det(\boldsymbol{S}^*)^{n/2})} \, d\boldsymbol{u}\, d\boldsymbol{\beta}\, (g_0(\boldsymbol{H}))^{-1}\, \pi(\boldsymbol{H})\, d\boldsymbol{H},$$

where $u^*_{ij} = u_{ij}$ for $(i,j) \neq (1,1)$, $u^*_{11} = 1$ and $\boldsymbol{S}^* = \sum_{j,l=1}^{n} \boldsymbol{H}_{jl}\boldsymbol{U}_{jl}$). It follows that the integral with respect to u_{11} diverges around $u_{11} = 0$, proving that $\boldsymbol{m}(\boldsymbol{y}) = \infty$.

Proof of Theorem 2. Without loss of generality, we take the first q_i rows of $\tilde{\boldsymbol{X}}_{\mathcal{C}_i}$ to be the linearly independent rows. It follows that the marginal of $\boldsymbol{y}$,

$$\boldsymbol{m}(\boldsymbol{y}) \leq C_0 \int_{\boldsymbol{F}} \int_{\boldsymbol{\Gamma}} \int_{\boldsymbol{\beta}} \int_{\boldsymbol{\eta}} \left(\prod_{i=1}^{p} \prod_{j=1}^{q_i} f_{ij}(y_{ij} \,|\, \epsilon_{ij} + x'_{ij}\beta_i) \right) f_0(\boldsymbol{\eta} \,|\, \boldsymbol{\beta}\, \boldsymbol{F}, \boldsymbol{\Gamma}) \, d\boldsymbol{\beta}\, d\boldsymbol{\eta}\, \pi(\boldsymbol{F})\, d\boldsymbol{F}\, \pi_{UV}(\boldsymbol{\Gamma})\, d\boldsymbol{\Gamma},$$

where C_0 is a constant depending on A and B and the submatrix $X_i^* = (x_{i1}^{'}, \ldots, x_{iq_i}^{'})'$, is of dimension $q_i \times q_i$ with full rank q_i. With $r_i = (\epsilon_{i1}, \ldots, \epsilon_{iq_i})' + X_i^*\beta_i$ for $i = 1, \ldots, p$, condition (3.8) implies that

$$\begin{aligned} \boldsymbol{m}(\boldsymbol{y}) &\leq \prod_{i=1}^{p} \det(X_i^*) \int_{\boldsymbol{\epsilon}} \int_{\boldsymbol{F}} \int_{\boldsymbol{\Gamma}} f_0(\boldsymbol{\epsilon} \,|\, \boldsymbol{F}, \boldsymbol{\Gamma})\, d\boldsymbol{\epsilon}\, \pi(\boldsymbol{F})\, d\boldsymbol{F}\, \pi_{UV}(\boldsymbol{\Gamma})\, d\boldsymbol{\Gamma} \\ &\leq \prod_{i=1}^{p} \det(X_i^*) < \infty, \end{aligned}$$

since the integrands are all integrable: f_0 integrates to 1 with respect to $\boldsymbol{\epsilon}$, and $\pi(\boldsymbol{F})$ and $\pi_{UV}(\boldsymbol{\Gamma})$, respectively, integrate to 1 with respect to $\boldsymbol{F}$ and $\boldsymbol{\Gamma}$ since they are proper priors.

Derivation of the Jacobian from $(\boldsymbol{\Gamma}, \boldsymbol{Q}) \to \boldsymbol{B}$. We consider three transformation steps: $(\boldsymbol{\Gamma}, \boldsymbol{Q}) \overset{(a)}{\to} (\boldsymbol{\Gamma^{-1}}, \boldsymbol{Q}) \overset{(b)}{\to} (\boldsymbol{L}, \boldsymbol{Q}) \overset{(c)}{\to} \boldsymbol{B}$, where (a) is an inverse

transformation, (b) is Cholesky decomposition, and (c) is QR decomposition. The Jacobian of each transformation are as follows (Muirhead (1982)): (a) $d\boldsymbol{\Gamma} = \det(\boldsymbol{\Gamma^{-1}})^{-(p+1)}\, d\boldsymbol{\Gamma}^{-1}$, (b) $d\boldsymbol{\Gamma^{-1}} = 2^p \prod_{i=1}^p L_{ii}^{p+1-i}\, d\boldsymbol{L'}$, and (c) $d\boldsymbol{B} = \prod_{i=1}^p L_{ii}^{p-i}\, d\boldsymbol{L'}(\boldsymbol{Q}\, d^*\boldsymbol{Q'})$, where $(\boldsymbol{Q}\, d^*\boldsymbol{Q'})$ defines the Haar measure on the set of $p \times p$ orthogonal matrices. Thus, defining $d\boldsymbol{Q} \propto (\boldsymbol{Q} d^*\boldsymbol{Q'})$, we have $d\boldsymbol{\Gamma} d\boldsymbol{Q} = \det(\boldsymbol{\Gamma^{-1}})^{-(p+1)}\, 2^p \prod_{i=1}^p L_{ii}^{p+1-i}\, d\boldsymbol{L'} d\boldsymbol{Q} \propto \det(\boldsymbol{\Gamma^{-1}})^{-(p+1)} \prod_{i=1}^p L_{ii} d\boldsymbol{B} \propto \det(\boldsymbol{\Gamma^{-1}})^{-p-1/2} d\boldsymbol{B} \propto \det(\boldsymbol{BB'})^{-(p+1/2)} d\boldsymbol{B}$.

Derivation of (4.4) and (4.5): The posterior density of $(\boldsymbol{\Lambda}, \boldsymbol{B})$ given the other parameters is

$$\pi(\boldsymbol{\Lambda}, \boldsymbol{B} \mid \cdots) \propto |(\boldsymbol{D}^r)^{-1}|^{1/2} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\eta}^r - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right)' (\boldsymbol{D}^r)^{-1} \left(\boldsymbol{\eta}^r - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right)\right\} \times \det(I_p + w_0(\boldsymbol{BB'})^{-1})^{-(p+1)} \det(\boldsymbol{BB'})^{-(p+1/2)}. \tag{A.2}$$

In the above, $|(\boldsymbol{D}^r)^{-1}|^{1/2} = (\det(\boldsymbol{BB'}))^{n/2} \cdot \prod_{k=1}^p (\det(\boldsymbol{M} - \lambda_k \boldsymbol{W}))^{1/2}$ and

$$\left(\boldsymbol{\eta}^r - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right)' (\boldsymbol{D}^r)^{-1} \left(\boldsymbol{\eta}^r - \tilde{\boldsymbol{X}}\boldsymbol{\beta}\right) = \boldsymbol{\epsilon}^{r'}(\boldsymbol{B} \otimes \boldsymbol{I}_n)(\boldsymbol{I}_p \otimes \boldsymbol{M} - \boldsymbol{\Lambda} \otimes \boldsymbol{W})(\boldsymbol{B'} \otimes \boldsymbol{I}_n)\boldsymbol{\epsilon}^r = \prod_{k=1}^p e_k'(\boldsymbol{M} - \lambda_k \boldsymbol{W})\, e_k,$$

since $(\boldsymbol{B'} \otimes \boldsymbol{I}_n)\, \boldsymbol{\epsilon^r} = (e_1', \ldots, e_p')'$. Also, note that $\det(I_p + w_0(\boldsymbol{BB'})^{-1})^{-(p+1)} = \det(\boldsymbol{BB'} + w_0 \boldsymbol{I}_p)^{-(p+1)} \cdot (\det(\boldsymbol{BB'}))^{(p+1)}$. Substituting these three expressions into (A.2), we get

$$\pi(\boldsymbol{\Lambda}, \boldsymbol{B} \mid \cdots) \propto \prod_{k=1}^p (\det(\boldsymbol{M} - \lambda_k \boldsymbol{W}))^{1/2} \times \exp\Big\{-\frac{1}{2}\sum_{k=1}^p e_k'(\boldsymbol{M} - \lambda_k \boldsymbol{W})\, e_k\Big\} \times \det(\boldsymbol{BB'} + w_0 \boldsymbol{I}_p)^{-(p+1)} \cdot (\det(\boldsymbol{BB'}))^{(n+1)/2}. \tag{A.3}$$

The posteriors in (4.4) and (4.5) are obtained from (A.3) by viewing the other parameter in the pair $(\boldsymbol{\Lambda}, \boldsymbol{B})$ as fixed.

References

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

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

Daniels, M. J. (1999). A Prior for the variance in Hierarchical models. *Canad. J. Statist.* **27**, 569-580.

Gelfand, A. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* **4**, 11-25.

Gelman, A., and Rubin, D. (1992). Inference from iterative simulation using multiple sequences (with discussion). *Statist. Sci.* **7**, 457-511.

Greco, F. P., and Trivisano, C. (2009). A multivariate CAR model for improving the estimation of relative risks. *Statist. Med.* **28**, 1707-1724.

Jin, X., Banerjee, S., and Carlin, B. P. (2007). Order-free coregionalized areal data models with applications to multiple disease mapping. *J. Roy. Statist. Soc. Ser. B* **69**, 817-838.

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

Kim, H.D. Sun, D. and Tsutakawa, R. (2001). A bivariate Bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model. *J. Amer. Statist. Assoc.* **96**, 1506-1521.

McCulloch, C. E. and Searle, S. R. (2001). *Generalized, Linear, and Mixed Models*. Wiley, New York.

Muirhead (1982). *Aspects of Multivariate Statistical Theory*. John Wiley, Canada.

Natarajan, R. and Kass, R. (2000). Reference Bayesian methods for generalized linear mixed models. *J. Multivariate Anal.* **95**, 227-237.

Sain, S. and Cressie, N. (2007). A spatial model for multivariate lattice data. *J. Economet.* **140**, 226-259.

Zhang, Y., Hodges, J. S., and Banerjee, S. (2009). Smoothed ANOVA with spatial effects as a competitor to MCAR in multivariate spatial smoothing. *Ann. Appl. Statist.* **3**, 1805-1830.

Department of Statistics & Probability, Michigan State University, East Lansing, MI 48824, USA.

E-mail: sdass@stt.msu.edu

Department of Statistics & Probability, Michigan State University, East Lansing, MI 48824, USA.

E-mail: lim@stt.msu.edu

Department of Statistics & Probability, Michigan State University, East Lansing, MI 48824, USA.

E-mail: maiti@stt.msu.edu

(Received May 2010; accepted November 2010)