Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: www.tandfonline.com/journals/uasa20

Bayesian Inference for General Gaussian Graphical Models With Application to Multivariate Lattice Data

Adrian Dobra, Alex Lenkoski & Abel Rodriguez

To cite this article: Adrian Dobra, Alex Lenkoski & Abel Rodriguez (2011) Bayesian Inference for General Gaussian Graphical Models With Application to Multivariate Lattice Data, Journal of the American Statistical Association, 106:496, 1418-1433, DOI: 10.1198/jasa.2011.tm10465

To link to this article: https://doi.org/10.1198/jasa.2011.tm10465

Supplementary materials for this article are available online. Please click the JASA link at http://pubs.amstat.org.

Bayesian Inference for General Gaussian Graphical Models With Application to Multivariate Lattice Data

Adrian DOBRA, Alex LENKOSKI, and Abel RODRIGUEZ

We introduce efficient Markov chain Monte Carlo methods for inference and model determination in multivariate and matrix-variate Gaussian graphical models. Our framework is based on the G-Wishart prior for the precision matrix associated with graphs that can be decomposable or non-decomposable. We extend our sampling algorithms to a novel class of conditionally autoregressive models for sparse estimation in multivariate lattice data, with a special emphasis on the analysis of spatial data. These models embed a great deal of flexibility in estimating both the correlation structure across outcomes and the spatial correlation structure, thereby allowing for adaptive smoothing and spatial autocorrelation parameters. Our methods are illustrated using a simulated example and a real-world application which concerns cancer mortality surveillance. Supplementary materials with computer code and the datasets needed to replicate our numerical results together with additional tables of results are available online.

KEY WORDS: CAR model; G-Wishart distribution; Markov chain Monte Carlo (MCMC) simulation; Spatial statistics.

1. INTRODUCTION

Graphical models (Lauritzen 1996), which encode the conditional independence among variables using a graph, have become a popular tool for sparse estimation in both the statistics and machine learning literatures (Dobra et al. 2004; Meinshausen and Bühlmann 2006; Yuan and Lin 2007; Banerjee, El Ghaoui, and D'Aspremont 2008; Drton and Perlman 2008; Friedman, Hastie, and Tibshirani 2008; Ravikumar, Wainwright, and Lafferty 2010). Implementing model selection approaches in the context of graphical models typically allows a dramatic reduction in the number of parameters under consideration, preventing overfitting and improving predictive capability. In particular, Bayesian approaches to inference in graphical models generate regularized estimators that incorporate model structure uncertainty.

The focus of this article is Bayesian inference in Gaussian graphical models (Dempster 1972) using the G-Wishart prior (Roverato 2002; Atay-Kayis and Massam 2005; Letac and Massam 2007). This class of distributions is extremely attractive since it represents the conjugate family for the precision matrix whose elements associated with edges not in the underlying graph are constrained to be equal to zero. Many recent articles have described various stochastic search methods for Gaussian graphical models (GGMs) with the G-Wishart prior based on marginal likelihoods which, in this case, are given by the ratio of the normalizing constants of the posterior and prior G-Wishart distributions—see the works of Atay-Kayis and Massam (2005), Jones et al. (2005), Carvalho and Scott (2009), Armstrong et al. (2009), Lenkoski and Dobra (2011) and the references therein. Wang and West (2009) proposed a MCMC algorithm for model determination and estimation in matrix-variate GGMs that also involves marginal likelihoods. Although the computation of marginal likelihoods for decomposable graphs is straightforward, similar computations for non-decomposable graphs or matrix-variate GGMs raise significant numerical challenges. This leads to the idea of devising Bayesian model determination methods that avoid the computation of marginal likelihoods.

The contributions of this article are threefold. First, we develop a new Metropolis–Hastings method for sampling from the G-Wishart distribution associated with an arbitrary graph. We discuss our algorithm in the context of the related sampling methods of Wang and Carvalho (2010) and Mitsakakis, Massam, and Escobar (2011), and empirically show that it scales better to graphs with many vertices. Second, we propose novel reversible jump MCMC samplers (Green 1995) for model determination and estimation in multivariate and matrix-variate GGMs. We contrast our approaches with the algorithms of Wong, Carter, and Kohn (2003) and Wang and West (2009) which focus on arbitrary GGMs and matrix-variate GGMs, respectively. Third, we devise a new flexible class of conditionally autoregressive models (CAR) (Besag 1974) for lattice data that rely on our novel sampling algorithms. The link between conditionally autoregressive models and GGMs was originally pointed out by Besag and Kooperberg (1995). However, since typical neighborhood graphs are non-decomposable, fully exploiting this connection within a Bayesian framework requires that we are able to estimate Gaussian graphical models based on general graphs. Our main focus is on applications to multivariate lattice data, where our approach based on matrix-variate GGMs provides a natural approach to create sparse multivariate CAR models.

The organization of the article is as follows. Section 2 formally introduces GGMs and the G-Wishart distribution, along

Adrian Dobra is Assistant Professor, Departments of Statistics, Biobehavioral Nursing, and Health Systems and the Center for Statistics and the Social Sciences, Box 354322, University of Washington, Seattle, WA 98195 (E-mail: *adobra@uw.edu*). Alex Lenkoski is Postdoctoral Research Fellow, Institut für Angewandte Mathematik, Universität Heidelberg, 69115 Heidelberg, Germany (E-mail: *alex.lenkoski@uni-heidelberg.de*). Abel Rodriguez is Assistant Professor, Department of Applied Mathematics and Statistics, University of California, Santa Cruz, CA 95064 (E-mail: *abel@soe.ucsc.edu*). The first author was supported in part by the National Science Foundation (DMS 1120255). The second author gratefully acknowledges support by the joint research project "Spatio/Temporal Graphical Models and Applications in Image Analysis" funded by the German Science Foundation (DFG), grant GRK 1653, as well as the MAThematics Centre Heidelberg (MATCH). The work of the third author was supported in part by the National Science Foundation (DMS 0915272) and the National Institutes of Health (R01GM090201-01). The authors thank the editor, the associate editor, and three reviewers for helpful comments.

Journal of the American Statistical Association
December 2011, Vol. 106, No. 496, Theory and Methods
DOI: 10.1198/jasa.2011.tm10465

with a description of our novel sampling algorithm for the G-Wishart distribution. Section 3 describes our reversible jump MCMC algorithm for GGMs. This algorithm represents a generalization of the work of Giudici and Green (1999) and is applicable beyond decomposable GGMs. Section 4 discusses inference and model determination in matrix-variate GGMs. Unlike the related framework developed by Wang and West (2009) which involves exclusively decomposable graphs, our sampler operates on the joint distribution of the row and column precision matrices, the row and column conditional independence graphs, and the auxiliary variable that needs to be introduced to solve the underlying non-identifiability problem associated with matrix-variate normal distributions. Section 5 reviews conditional autoregressive priors for lattice data and their connection to GGMs. This section discusses both univariate and multivariate models for continuous and discrete data based on generalized linear models. Section 6 presents two illustrations of our methodology: a simulation study and a cancer mortality mapping model. Finally, Section 7 concludes the article by discussing future directions for our work.

2. GAUSSIAN GRAPHICAL MODELS AND THE G–WISHART DISTRIBUTION

We let $\mathbf{X} = \mathbf{X}_{V_p}$, $V_p = \{1, 2, \ldots, p\}$, be a random vector with a p-dimensional multivariate normal distribution $\mathsf{N}_p(\mathbf{0}, \mathbf{K}^{-1})$. We consider a graph $G = (V_p, E)$, where each vertex $i \in V$ corresponds with a random variable X_i and $E \subset V_p \times V_p$ are undirected edges. Here "undirected" means that $(i, j) \in E$ if and only if $(j, i) \in E$. We denote by $\mathcal{G}_p$ the set of all $2^{p(p-1)/2}$ undirected graphs with p vertices. A Gaussian graphical model with conditional independence graph G is constructed by constraining to zero the off-diagonal elements of $\mathbf{K}$ that do not correspond with edges in G (Dempster 1972). If $(i, j) \notin E$, X_i and X_j are conditionally independent given the remaining variables. The precision matrix $\mathbf{K} = (K_{ij})_{1 \le i,j \le p}$ is constrained to the cone P_G of symmetric positive definite matrices with off-diagonal entries $K_{ij} = 0$ for all $(i, j) \notin E$.

We consider the G-Wishart distribution $\mathsf{Wis}_G(\delta, \mathbf{D})$ with density

$$\mathsf{p}(\mathbf{K} \mid G, \delta, \mathbf{D}) = \frac{1}{I_G(\delta, \mathbf{D})} (\det \mathbf{K})^{(\delta-2)/2} \exp\left\{-\frac{1}{2}\langle \mathbf{K}, \mathbf{D}\rangle\right\}, \tag{1}$$

with respect to the Lebesgue measure on P_G (Roverato 2002; Atay-Kayis and Massam 2005; Letac and Massam 2007). Here $\langle \mathbf{A}, \mathbf{B}\rangle = \operatorname{tr}(\mathbf{A}^T\mathbf{B})$ denotes the trace inner product. The normalizing constant $I_G(\delta, \mathbf{D})$ is finite if $\delta > 2$ and $\mathbf{D}$ positive definite (Diaconnis and Ylvisaker 1979). If G is the full graph ($E = V_p \times V_p$), $\mathsf{Wis}_G(\delta, \mathbf{D})$ is the Wishart distribution $\mathsf{Wis}_p(\delta, D)$ (Muirhead 2005).

Since our sampling methods rely on perturbing the Cholesky decompositions of the matrix $\mathbf{K}$, we review some key results. We write $\mathbf{K} \in \mathsf{P}_G$ as $\mathbf{K} = \mathbf{Q}^T(\mathbf{\Psi}^T\mathbf{\Psi})\mathbf{Q}$ where $\mathbf{Q} = (Q_{ij})_{1 \le i \le j \le p}$ and $\mathbf{\Psi} = (\Psi_{ij})_{1 \le i \le j \le p}$ are upper triangular, while $\mathbf{D}^{-1} = \mathbf{Q}^T\mathbf{Q}$ is the Cholesky decomposition of $\mathbf{D}^{-1}$. We see that $\mathbf{K} = \mathbf{\Phi}^T\mathbf{\Phi}$ where $\mathbf{\Phi} = \mathbf{\Psi}\mathbf{Q}$ is the Cholesky decomposition of $\mathbf{K}$. The zero constraints on the off-diagonal elements of $\mathbf{K}$ associated with G induce well-defined sets of free elements $\mathbf{\Phi}^{\nu(G)} = \{\Phi_{ij} : (i, j) \in \nu(G)\}$ and $\mathbf{\Psi}^{\nu(G)} = \{\Psi_{ij} : (i, j) \in \nu(G)\}$ of the matrices $\mathbf{\Phi}$ and $\mathbf{\Psi}$—see proposition 2, page 320, and lemma 2, page 326, of the article by Atay-Kayis and Massam (2005). Here $\nu(G) = \nu_=(G) \cup \nu_<(G)$, $\nu_=(G) = \{(i, i) : i \in V_p\}$ and $\nu_<(G) = \{(i, j) : i < j \text{ and } (i, j) \in E\}$.

We denote by $\mathsf{M}^{\nu(G)}$ the set of incomplete triangular matrices whose elements are indexed by $\nu(G)$ and whose diagonal elements are strictly positive. Both $\mathbf{\Phi}^{\nu(G)}$ and $\mathbf{\Psi}^{\nu(G)}$ must belong to $\mathsf{M}^{\nu(G)}$. The non-free elements of $\mathbf{\Psi}$ are determined through the completion operation (Atay-Kayis and Massam 2005, lemma 2) as a function of the free elements $\mathbf{\Psi}^{\nu(G)}$. Each element Ψ_{ij} with $i < j$ and $(i, j) \notin E$ is a function of the other elements $\Psi_{i'j'}$ that precede it in lexicographical order. Roverato (2002) proved that the Jacobian of the transformation that maps $\mathbf{K} \in \mathsf{P}_G$ to $\mathbf{\Phi}^{\nu(G)} \in \mathsf{M}^{\nu(G)}$ is $\mathsf{J}(\mathbf{K} \to \mathbf{\Phi}^{\nu(G)}) = 2^p \prod_{i=1}^p \Phi_{ii}^{v_i^G+1}$, where $v_i^G = |\{j : j > i \text{ and } (i, j) \in E\}|$. Here $|A|$ denotes the number of elements of the set A. Atay-Kayis and Massam (2005) showed that the Jacobian of the transformation that maps $\mathbf{\Phi}^{\nu(G)}$ to $\mathbf{\Psi}^{\nu(G)}$ is given by $\mathsf{J}(\mathbf{\Phi}^{\nu(G)} \to \mathbf{\Psi}^{\nu(G)}) = \prod_{i=1}^p Q_{ii}^{d_i^G+1}$, where $d_i^G = |\{j : j < i \text{ and } (i, j) \in E\}|$. We have $\det \mathbf{K} = \prod_{i=1}^p \Phi_{ii}^2$ and $\Phi_{ii} = \Psi_{ii}Q_{ii}$. It follows that the density of $\mathbf{\Psi}^{\nu(G)}$ with respect to the Lebesgue measure on $\mathsf{M}^{\nu(G)}$ is

$$\mathsf{p}\big(\mathbf{\Psi}^{\nu(G)} \mid \delta, \mathbf{D}\big) = \frac{2^p}{I_G(\delta, \mathbf{D})} \prod_{i=1}^p Q_{ii}^{v_i^G+d_i^G+\delta} \prod_{i=1}^p \Psi_{ii}^{v_i^G+\delta-1} \times \exp\left(-\frac{1}{2}\sum_{i=1}^p\sum_{j=i}^p \Psi_{ij}^2\right). \tag{2}$$

We note that $v_i^G + d_i^G$ represents the number of neighbors of vertex i in the graph G.

2.1 Existing Methods for Sampling From the G-Wishart Distribution

The problem of sampling from the G-Wishart distribution has received considerable attention in the recent literature. Piccioni (2000) followed by Asci and Piccioni (2007) exploited the theory of regular exponential families with cuts and proposed the block Gibbs sampler algorithm. Their iterative method generates one sample from $\mathsf{Wis}_G(\delta, \mathbf{D})$ by performing sequential adjustments with respect to submatrices associated with each clique of G. The adjustment with respect to a clique C involves the inversion of a $(p - |C|) \times (p - |C|)$ matrix. If G has several cliques involving a small number of vertices but has a large number of vertices p, repeated inversions of high-dimensional matrices must be performed at each iteration. This makes the block Gibbs sampler algorithm impractical for being used in the context of large graphs.

Carvalho, Massam, and West (2007) gave a direct sampling method for $\mathsf{Wis}_G(\delta, \mathbf{D})$ that works well for decomposable graphs. This algorithm was extended by Wang and Carvalho (2010) to non-decomposable graphs by introducing an accept–reject step as follows. Based on theorem 1, page 328, of the article of Atay-Kayis and Massam (2005), Wang and Carvalho (2010) and Mitsakakis, Massam, and Escobar (2011) wrote the density of $\mathbf{\Psi}^{\nu(G)}$ given in (2) as

$$\mathsf{p}\big(\mathbf{\Psi}^{\nu(G)} \mid \delta, \mathbf{D}\big) = \frac{1}{\mathsf{E}_h[f(\mathbf{\Psi}^{\nu(G)})]} f\big(\mathbf{\Psi}^{\nu(G)}\big) h\big(\mathbf{\Psi}^{\nu(G)}\big), \tag{3}$$

where

$$\log f\left(\boldsymbol{\Psi}^{\nu(G)}\right) = -\frac{1}{2} \sum_{(i,j)\notin\nu(G), i<j} \Psi_{ij}^2$$

is a function of the non-free elements of $\boldsymbol{\Psi}$ which, in turn, are uniquely determined from the free elements in $\boldsymbol{\Psi}^{\nu(G)}$. Furthermore, $h(\boldsymbol{\Psi}^{\nu(G)})$ is the product of mutually independent chi-squared and standard normal distributions, that is, $\Psi_{ii}^2 \sim \chi^2_{\delta+\nu_i^G}$ for $i \in V_p$ and $\Psi_{ij} \sim \mathsf{N}(0,1)$ for $(i,j) \in \nu_<(G)$. The expectation $\mathsf{E}_h[f(\boldsymbol{\Psi}^{\nu(G)})]$ is calculated with respect to $h(\cdot)$. Since $f(\boldsymbol{\Psi}^{\nu(G)}) \le 1$ for any $\boldsymbol{\Psi}^{\nu(G)} \in \mathsf{M}^{\nu(G)}$, we have

$$\mathsf{p}\left(\boldsymbol{\Psi}^{\nu(G)} \mid \delta, \mathbf{D}\right) \le M h\left(\boldsymbol{\Psi}^{\nu(G)}\right), \tag{4}$$

where $M = \mathsf{E}_h^{-1}[f(\boldsymbol{\Psi}^{\nu(G)})] \ge 1$. From algorithm A.4 of the book by Robert and Casella (2004), an accept–reject method for sampling from $\mathsf{Wis}_G(\delta, \mathbf{D})$ proceeds by sampling $\boldsymbol{\Psi}^{\nu(G)} \sim h(\cdot)$ and $U \sim \mathsf{Uni}(0,1)$ until $U \le f(\boldsymbol{\Psi}^{\nu(G)})$. The probability of acceptance of a random sample from $h(\cdot)$ as a random sample from $\mathsf{Wis}_G(\delta, \mathbf{D})$ is

$$\mathsf{Pr}\left(U < f\left(\boldsymbol{\Psi}^{\nu(G)}\right)\right) = \mathsf{E}_h\left[f\left(\boldsymbol{\Psi}^{\nu(G)}\right)\right]. \tag{5}$$

Wang and Carvalho (2010) further decomposed G in its maximal prime subgraphs (Tarjan 1985), applied the accept–reject method for marginal or conditional distributions of $\mathsf{Wis}_G(\delta, \mathbf{D})$ corresponding with each maximal prime subgraph, and generated a sample from $\mathsf{Wis}_G(\delta, \mathbf{D})$ by putting together the resulting lower-dimensional sampled matrices.

Mitsakakis, Massam, and Escobar (2011) proposed a Metropolis–Hastings method for sampling from the G-Wishart $\mathsf{Wis}_G(\delta, \mathbf{D})$ distribution. We denote by $\mathbf{K}^{[s]} = \mathbf{Q}^T(\boldsymbol{\Psi}^{[s]})^T\boldsymbol{\Psi}^{[s]}\mathbf{Q}$ the current state of their Markov chain, where $(\boldsymbol{\Psi}^{[s]})^{\nu(G)} \in \mathsf{M}^{\nu(G)}$. Mitsakakis, Massam, and Escobar (2011) generated a candidate $\mathbf{K}' = \mathbf{Q}^T(\boldsymbol{\Psi}')^T\boldsymbol{\Psi}'\mathbf{Q}$ for the next state by sampling $(\boldsymbol{\Psi}')^{\nu(G)} \sim h(\cdot)$ and determining the non-free elements of $\boldsymbol{\Psi}'$ from the free sampled elements $(\boldsymbol{\Psi}')^{\nu(G)}$. The chain moves to $\mathbf{K}'$ with probability

$$\min\left\{f\left((\boldsymbol{\Psi}')^{\nu(G)}\right)/f\left(\left(\boldsymbol{\Psi}^{[s]}\right)^{\nu(G)}\right), 1\right\}. \tag{6}$$

The inequality (4) implies that this independent chain is uniformly ergodic (Mengersen and Tweedie 1996) and its expected acceptance probability is greater than or equal to $\mathsf{E}_h[f(\boldsymbol{\Psi}^{\nu(G)})]$—see lemma 7.9 in the book by Robert and Casella (2004). Hence, the Markov chain of Mitsakakis, Massam, and Escobar (2011) is more efficient than the method of Wang and Carvalho (2010) when the latter algorithm is employed without graph decompositions or when the graph G has only one maximal prime subgraph. On the other hand, the method of Mitsakakis, Massam, and Escobar (2011) involves changing the values of all free elements $\boldsymbol{\Psi}^{\nu(G)}$ in a single step. As a result, if their Markov chain is currently in a region of high probability, it could potentially have to generate many candidates from $h(\cdot)$ before moving to a new state with comparable or larger probability.

2.2 Our Algorithm for Sampling From the G-Wishart Distribution

We introduce a new Metropolis–Hastings algorithm for sampling from the G-Wishart $\mathsf{Wis}_G(\delta, \mathbf{D})$ distribution. In contrast to the work of Mitsakakis, Massam, and Escobar (2011), our approach makes use of a proposal distribution that depends on the current state of the chain and leaves all but one of the free elements in $\boldsymbol{\Psi}^{\nu(G)}$ unchanged. The distance between the current and the proposed state is controlled through a Gaussian kernel with a precision parameter σ_m.

We denote by $\mathbf{K}^{[s]} = \mathbf{Q}^T(\boldsymbol{\Psi}^{[s]})^T\boldsymbol{\Psi}^{[s]}\mathbf{Q}$ the current state of the chain with $(\boldsymbol{\Psi}^{[s]})^{\nu(G)} \in \mathsf{M}^{\nu(G)}$. The next state $\mathbf{K}^{[s+1]} = \mathbf{Q}^T(\boldsymbol{\Psi}^{[s+1]})^T\boldsymbol{\Psi}^{[s+1]}\mathbf{Q}$ is obtained by sequentially perturbing the free elements $(\boldsymbol{\Psi}^{[s]})^{\nu(G)}$. A diagonal element $\Psi^{[s]}_{i_0i_0} > 0$ is updated by sampling a value γ from a $\mathsf{N}(\Psi^{[s]}_{i_0i_0}, \sigma_m^2)$ distribution truncated below at zero. We define the upper triangular matrix $\boldsymbol{\Psi}'$ such that $\Psi'_{ij} = \Psi^{[s]}_{ij}$ for $(i,j) \in \nu(G) \setminus \{(i_0,i_0)\}$ and $\Psi'_{i_0i_0} = \gamma$. The non-free elements of $\boldsymbol{\Psi}'$ are obtained through the completion operation (Atay-Kayis and Massam 2005, lemma 2) from $(\boldsymbol{\Psi}')^{\nu(G)}$. The Markov chain moves to $\mathbf{K}' = \mathbf{Q}^T(\boldsymbol{\Psi}')^T\boldsymbol{\Psi}'\mathbf{Q}$ with probability $\min\{R_m, 1\}$, where

$$\begin{aligned} R_m &= \frac{\mathsf{p}((\boldsymbol{\Psi}')^{\nu(G)} \mid \delta, \mathbf{D})}{\mathsf{p}((\boldsymbol{\Psi}^{[s]})^{\nu(G)} \mid \delta, \mathbf{D})} \frac{\mathsf{p}(\Psi^{[s]}_{i_0i_0} \mid \Psi'_{i_0i_0})}{\mathsf{p}(\Psi'_{i_0i_0} \mid \Psi^{[s]}_{i_0i_0})} \\ &= \frac{\phi(\Psi^{[s]}_{i_0i_0}/\sigma_m)}{\phi(\Psi'_{i_0i_0}/\sigma_m)} \left(\frac{\Psi'_{i_0i_0}}{\Psi^{[s]}_{i_0i_0}}\right)^{\nu^G_{i_0}+\delta-1} R'_m. \end{aligned} \tag{7}$$

Here $\phi(\cdot)$ represents the CDF of the standard normal distribution and

$$R'_m = \exp\left\{-\frac{1}{2}\sum_{i=1}^{p}\sum_{j=i}^{p}\left[(\Psi'_{ij})^2 - \left(\Psi^{[s]}_{ij}\right)^2\right]\right\}. \tag{8}$$

A free off-diagonal element $\Psi^{[s]}_{i_0j_0}$ is updated by sampling a value $\gamma' \sim \mathsf{N}(\Psi^{[s]}_{i_0j_0}, \sigma_m^2)$. We define the upper triangular matrix $\boldsymbol{\Psi}'$ such that $\Psi'_{ij} = \Psi^{[s]}_{ij}$ for $(i,j) \in \nu(G) \setminus \{(i_0,j_0)\}$ and $\Psi'_{i_0j_0} = \gamma'$. The remaining elements of $\boldsymbol{\Psi}'$ are determined by the completion operation from lemma 2 in the article of Atay-Kayis and Massam (2005) from $(\boldsymbol{\Psi}')^{\nu(G)}$. The proposal distribution is symmetric $\mathsf{p}(\Psi'_{i_0j_0} \mid \Psi^{[s]}_{i_0j_0}) = \mathsf{p}(\Psi^{[s]}_{i_0j_0} \mid \Psi'_{i_0j_0})$, thus we accept the transition of the chain from $\mathbf{K}^{[s]}$ to $\mathbf{K}' = \mathbf{Q}^T(\boldsymbol{\Psi}')^T\boldsymbol{\Psi}'\mathbf{Q}$ with probability $\min\{R'_m, 1\}$, where R'_m is given in Equation (7). Since $(\boldsymbol{\Psi}^{[s]})^{\nu(G)} \in \mathsf{M}^{\nu(G)}$, we have $(\boldsymbol{\Psi}')^{\nu(G)} \in \mathsf{M}^{\nu(G)}$ which implies $\mathbf{K}' \in \mathsf{P}_G$. We denote by $\mathbf{K}^{[s+1]}$ the precision matrix obtained after completing all the updates associated with the free elements indexed by $\nu(G)$.

A key computational aspect is related to the dependence of Cholesky decompositions on a particular ordering of the variables involved. Empirically we noticed that the mixing times of Markov chains that make use of our sampling approach can be improved by changing the ordering of the variables in V_p at each iteration. More specifically, a permutation υ is uniformly drawn from the set of all possible permutations Υ_p of V_p. The row and columns of $\mathbf{D}$ are reordered according to υ and a new Cholesky decomposition of $\mathbf{D}^{-1}$ is determined. The set $\nu(G)$

and $\{d_i^G : i \in V\}$ are recalculated given the ordering of the vertices induced by υ. Although the random shuffling of the ordering of the indices worked well for the particular applications we have considered in this article, we do not have any theoretical justification that explains why it improves the computational efficiency of our sampling approaches.

Our later developments from Section 4 involve sampling $\mathbf{K} \sim \mathsf{Wis}_G(\delta, \mathbf{D})$ subject to the constraint $K_{11} = 1$. We have $K_{11}^{[s]} = 1 \Leftrightarrow (\mathbf{\Psi}^{[s]}\mathbf{Q})_{11} = 1 \Leftrightarrow \Psi_{11}^{[s]} = 1/Q_{11}$. We subsequently obtain the next state $\mathbf{K}^{[s+1]}$ of the Markov chain by perturbing the free elements $(\mathbf{\Psi}^{[s]})^{\nu(G)\setminus\{(1,1)\}}$. When defining the triangular matrix $\mathbf{\Psi}'$ we set $\Psi'_{11} = 1/Q_{11}$ which implies that the corresponding candidate matrix $\mathbf{K}'$ has $K'_{11} = 1$. Thus $\mathbf{K}^{[s+1]}$ also obeys the constraint $K_{11}^{[s+1]} = 1$. The random orderings of the variables need to be drawn from the set $\Upsilon_p^{(1,1)}$ of permutations $\upsilon \in \Upsilon_p$ such that $\upsilon(1) = 1$. This way the $(1,1)$ element of $\mathbf{K}$ always occupies the same position.

2.3 The Scalability of Sampling Methods From the G-Wishart Distribution

Practical applications of our novel framework for analyzing multivariate lattice data from Section 5 involve sampling from G-Wishart distributions associated with arbitrary graphs with tens or possibly hundreds of vertices—see Section 6.2 as well as relevant examples from the works of Elliott et al. (2001), Banerjee, Carlin, and Gelfand (2004), Rue and Held (2005), Lawson (2009), Gelfand et al. (2010). We perform a simulation study to empirically compare the scalability of our approach (DLR) for simulating from G-Wishart distributions with respect to the algorithms of Wang and Carvalho (2010) (WC) and Mitsakakis, Massam, and Escobar (2011) (MME). We consider the cycle graph $C_p \in \mathcal{G}_p$ with edges $\{(i, i+1) : 1 \le i \le p-1\} \cup \{(p,1)\}$ and the matrix $\mathbf{A}_p \in \mathsf{P}_{C_p}$ such that $(\mathbf{A}_p)_{ii} = 1$ $(1 \le i \le p)$, $(\mathbf{A}_p)_{i,i-1} = (\mathbf{A}_p)_{i-1,i} = 0.5$ $(2 \le i \le p)$, and $(\mathbf{A}_p)_{1,p} = (\mathbf{A}_p)_{p,1} = 0.4$. We chose C_p because it is the sparsest graph with p vertices and only one maximal prime subgraph, hence no graph decompositions can be performed in the context of the WC algorithm.

We employ the three sampling methods to sample from the G-Wishart $\mathsf{Wis}_{C_p}(103, \mathbf{D}_p)$ distribution, where $\mathbf{D}_p = \mathbf{I}_p + 100\mathbf{A}_p^{-1}$ and $\mathbf{I}_p$ is the p-dimensional identity matrix. This is representative of a G-Wishart posterior distribution corresponding with 100 samples from a $\mathsf{N}_p(0, A_p^{-1})$ and a G-Wishart prior $\mathsf{Wis}_{C_p}(3, I_p)$. The acceptance probability for generating one sample with the DLR algorithm is defined as the average acceptance probabilities of the updates corresponding with diagonal and off-diagonal free elements in $\mathbf{\Psi}^{\nu(G)}$. We calculate Monte Carlo estimates and their standard errors of the acceptance probabilities for our method by running 100 independent chains of length 2500 for each combination $(\sigma_m, p) \in \{0.1, 0.5, 1, 2\} \times \{4, 6, \ldots, 20\}$. The same number of chains of the same length have been run with the MME algorithm for $p = 4, 6, \ldots, 20$. We calculate Monte Carlo estimates of the acceptance probabilities (5) associated with the WC algorithm, that is,

$$\widehat{\mathsf{E}}_h\big[f\big(\mathbf{\Psi}^{\nu(C_p)}\big)\big] = \sum_{s=1}^{2500} f\Big(\big(\mathbf{\Psi}^{[s]}\big)^{\nu(C_p)}\Big), \quad \text{with } \big(\mathbf{\Psi}^{[s]}\big)^{\nu(C_p)} \sim h(\cdot), s = 1, \ldots, 2500.$$

The corresponding standard errors were determined by calculating 100 such estimates for each $p = 4, 6, \ldots, 20$. The results are summarized in Table 1. As we would expect, the acceptance rates for the DLR algorithm decrease as σ_m increases since the proposed jumps in P_{C_p} become larger. More importantly, for each value of σ_m, the acceptance rates for the DLR algorithm decrease very slowly as the number of vertices p grows. This shows that our Metropolis–Hastings method is likely to retain its efficiency for graphs that involve a large number of vertices. On the other hand, the acceptance rates for the WC algorithm are extremely small even for $p = 6$ vertices. This implies that a large number of samples from the instrumental distribution $h(\cdot)$ would need to be generated before one sample from $\mathsf{Wis}_{C_p}(103, \mathbf{D}_p)$ is obtained. The MME algorithm gives slightly better acceptance rates, but they are still small and are indicative of a simulation method that constantly attempts to make large jumps in areas of low probability of $\mathsf{Wis}_{C_p}(103, \mathbf{D}_p)$. Therefore the WC and MME methods might not scale well despite being perfectly valid in theory, hence they cannot be used in the context of the multivariate lattice data models from Section 5. The acceptance rates of the DLR algorithm are a function of the precision parameter σ_m which can be easily adjusted to yield jumps that are not too short but also not too long in the target cone of

Table 1. Monte Carlo estimates and their standard errors (in parentheses) of the acceptance probabilities for our proposed algorithm (DLR) for sampling from the G-Wishart distribution, the Wang and Carvalho (2010) algorithm (WC), and the Mitsakakis, Massam, and Escobar (2011) algorithm (MME)

	DLR					
p	$\sigma_m = 0.1$	$\sigma_m = 0.5$	$\sigma_m = 1$	$\sigma_m = 2$	WC	MME
4	0.953 (2.0e–3)	0.776 (3.0e–3)	0.600 (3.2e–3)	0.389 (3.8e–3)	0.340 (6.4e–3)	0.473 (1.0e–2)
6	0.947 (2.0e–3)	0.751 (2.4e–3)	0.565 (2.9e–3)	0.356 (2.3e–3)	5.08e–2 (2.2e–3)	0.185 (1.0e–2)
8	0.944 (1.8e–3)	0.740 (2.1e–3)	0.551 (2.3e–3)	0.343 (2.6e–3)	7.94e–3 (7.1e–4)	0.078 (0.9e–2)
10	0.943 (1.7e–3)	0.734 (2.0e–3)	0.543 (2.1e–3)	0.336 (2.1e–3)	1.18e–3 (1.7e–4)	0.035 (8.0e–3)
12	0.942 (1.4e–3)	0.729 (1.8e–3)	0.537 (2.1e–3)	0.331 (2.2e–3)	1.56e–4 (5.1e–5)	0.013 (6.0e–3)
14	0.940 (1.5e–3)	0.725 (1.6e–3)	0.532 (1.8e–3)	0.327 (1.9e–3)	2.27e–5 (1.4e–5)	0.005 (4.0e–3)
16	0.940 (1.3e–3)	0.721 (1.6e–3)	0.527 (1.7e–3)	0.324 (1.7e–3)	2.19e–6 (1.8e–6)	0.002 (3.0e–3)
18	0.939 (1.2e–3)	0.717 (1.4e–3)	0.523 (1.5e–3)	0.321 (1.5e–3)	3.87e–7 (8.0e–7)	0.001 (2.0e–3)
20	0.938 (1.0e–3)	0.714 (1.4e–3)	0.520 (1.6e–3)	0.318 (1.5e–3)	2.08e–8 (3.5e–8)	5.4e–4 (9.0e–4)

matrices. This flexibility is key for successfully employing our algorithm for graphs with many vertices.

It may initially appear that failure to handle graphs with a maximal prime subgraph with 20 vertices is not a major shortcoming of an algorithm that relies on graph decompositions. However, the example we consider in Section 6.2 involves a fixed graphical model whose underlying graph has a maximal prime subgraph with 36 vertices (see the Supplementary Materials). Since this graph is one that is constructed from the neighborhood structure of the United States, we can see the importance of the ability to scale when applying the GGM framework to spatial statistical problems.

3. REVERSIBLE JUMP MCMC SAMPLERS FOR GGMS

The previous section was concerned with sampling from the G-Wishart distribution when the underlying conditional independence graph is fixed. In contrast, this section is concerned with performing Bayesian inference in GGMs, which involves determination of the graph G. Giudici and Green (1999) proposed a reversible jump Markov chain algorithm that is restricted to decomposable graphs and employs a hyper inverse Wishart prior (Dawid and Lauritzen 1993) for the covariance matrix $\mathbf{\Sigma} = \mathbf{K}^{-1} = (\Sigma_{ij})_{1\le i,j\le p}$. The efficiency of their approach comes from representing decomposable graphs through their junction trees and from a specification of $\mathbf{\Sigma}$ as an incomplete matrix $\mathbf{\Gamma} = (\Gamma_{ij})_{1\le i,j\le p}$ such that $\Sigma_{ij} = \Gamma_{ij}$ if $i = j$ or if (i, j) is an edge in the current graph (i.e., K_{ij} is not constrained to 0); the remaining elements of $\mathbf{\Gamma}$ are left unspecified but can be uniquely determined using the iterative proportional scaling algorithm (Dempster 1972; Speed and Kiiveri 1986). Brooks, Giudici, and Roberts (2003) presented techniques for calibrating the precision parameter of the normal kernel used by Giudici and Green (1999) to sequentially update the elements of $\mathbf{\Gamma}$ and the edges of the underlying decomposable graph which lead to improved mixing times of the resulting Markov chain. Related work (Scott and Carvalho 2008; Carvalho and Scott 2009; Armstrong et al. 2009) has also focused on decomposable graphs. Graphs of this type have received attention mainly because their special structure is convenient from a computational standpoint. However, decomposability is a serious constraint as it drastically reduces $\mathcal{G}_p$ to a much smaller subset of graphs: the ratio between the number of decomposable graphs and the total number of graphs decreases from 0.95 for $p = 4$ to 0.12 for $p = 8$ (Armstrong 2005).

Roverato (2002), Atay-Kayis and Massam (2005), Jones et al. (2005), Dellaportas, Giudici, and Roberts (2003), Moghaddam et al. (2009), Lenkoski and Dobra (2011) explored various methods for numerically calculating marginal likelihoods for non-decomposable graphs. Besides being computationally expensive, stochastic search algorithms that traverse $\mathcal{G}_p$ based on marginal likelihoods output a set of graphs that have the highest posterior probabilities from all the graphs that have been visited, but do not produce estimates of the precision matrix $\mathbf{K}$ unless a direct sampling algorithm from the posterior distribution of $\mathbf{K}$ given a graph is available. As we have seen in Section 2.3, current direct sampling algorithms (e.g., Wang and Carvalho 2010) might not scale well to graphs with many vertices. Since our key motivation comes from modeling multivariate lattice data (see Section 5), we want to develop a Bayesian method that samples from the joint posterior distribution of precision matrices $\mathbf{K} \in \mathsf{P}_G$ and graphs $G \in \mathcal{G}_p$, thereby performing inference for both $\mathbf{K}$ and G. Wong, Carter, and Kohn (2003) developed such a reversible jump Metropolis–Hastings algorithm by decomposing $\mathbf{K} = \mathbf{T}\mathbf{\Delta}\mathbf{T}$, where $\mathbf{T} = \text{diag}\{K_{11}^{1/2}, \ldots, K_{pp}^{1/2}\}$ and $\mathbf{\Delta} = (\Delta_{ij})_{1\le i,j\le p}$ is a correlation matrix with $\Delta_{ii} = 1$ and $\Delta_{ij} = K_{ij}/(K_{ii}K_{jj})^{1/2}$, for $i < j$. Their prior specification for $\mathbf{K}$ involves independent priors for T_{ii} and a joint prior for the off-diagonal elements of $\mathbf{\Delta}$ whose normalizing constant is associated with all the graphs with the same number of edges.

We propose a new reversible jump Markov chain algorithm that is based on a G-Wishart prior for $\mathbf{K}$. While we do not empirically explore the efficiency of our approach with respect to the Wong, Carter, and Kohn (2003) algorithm, we state that the key advantage of our framework lies in its generalization to matrix-variate data—see Section 4. We are unaware of any similar extension of the Wong, Carter, and Kohn (2003) approach. Another benefit of our method with respect to that of Wong, Carter, and Kohn (2003) is related to its flexibility with respect to prior specifications on $\mathcal{G}_p$—see the discussion below.

We let $\mathcal{D} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\}$ be the observed data of n independent samples from $\mathsf{N}_p(\mathbf{0}, \mathbf{K}^{-1})$. Given a graph $G \in \mathcal{G}_p$, we assume a G-Wishart prior $\mathsf{Wis}_G(\delta_0, \mathbf{D}_0)$ for the precision matrix $\mathbf{K} \in \mathsf{P}_G$. We take $\delta_0 = 3 > 2$ and $D_0 = \mathbf{I}_p$. With this choice the prior for $\mathbf{K}$ is equivalent with one observed sample, while the observed variables are assumed to be a priori independent of each other. Since the G-Wishart prior for $\mathbf{K}$ is conjugate to the likelihood $\mathsf{p}(\mathcal{D} \mid \mathbf{K})$, the posterior of $\mathbf{K}$ given G is $\mathsf{Wis}_G(n + \delta_0, \mathbf{U} + \mathbf{D}_0)$ where $\mathbf{U} = \sum_{j=1}^{n} \mathbf{x}^{(j)}(\mathbf{x}^{(j)})^T$. We also assume a prior $\mathsf{Pr}(G)$ on $\mathcal{G}_p$. We develop a MCMC algorithm for sampling from the joint posterior distribution

$$\mathsf{p}(\mathbf{K}, G \mid \mathcal{D}) \propto \mathsf{p}(\mathcal{D} \mid \mathbf{K})\mathsf{p}(\mathbf{K} \mid G, \delta_0, \mathbf{D}_0)\,\mathsf{Pr}(G),$$

that is well-defined if and only if $\mathbf{K} \in \mathsf{P}_G$. We sequentially update the precision matrix given the current graph and the edges of the graph given the current precision matrix—see Appendix A. The addition or deletion of an edge involves a change in the dimensionality of the parameter space since the corresponding element of $\mathbf{K}$ becomes free or constrained to zero, hence we make use of the reversible jump MCMC methodology of Green (1995). The graph update step also requires the calculation of the normalizing constants of the G-Wishart priors corresponding with the current and the candidate graph. To this end, we make use of the Monte Carlo method of Atay-Kayis and Massam (2005) which converges very fast when computing $I_G(\delta, \mathbf{D})$ [see Equation (1)] even for large graphs when δ is small and $\mathbf{D}$ is set to the identity matrix (Lenkoski and Dobra 2011).

Our framework accommodates any prior probabilities $\mathsf{Pr}(G)$ on the set of graphs $\mathcal{G}_p$ which is a significant advantage with respect to the covariance selection prior from Wong, Carter, and Kohn (2003). Indeed, the prior in the work of Wong, Carter, and Kohn (2003) induces fixed probabilities for each graph and does not allow the possibility of further modifying these probabilities according to prior beliefs. A usual choice is the uniform prior $\mathsf{Pr}(G) = 2^{-m}$ with $m = \binom{p}{2}$, but this prior is biased toward middle-size graphs and gives small probabilities to sparse graphs and to graphs that are almost complete. Here the size of a graph G is defined as the number of edges in G and is denoted

by size$(G) \in \{0, 1, \ldots, m\}$. Dobra et al. (2004) and Jones et al. (2005) assumed that the probability of inclusion of any edge in G is constant and equal to $\psi \in (0, 1)$, which leads to the prior

$$\Pr(G) \propto \psi^{\text{size}(G)}(1-\psi)^{m-\text{size}(G)}. \tag{9}$$

Sparser graphs can be favored with prior (9) by choosing a small value for ψ. Since ψ could be difficult to elicit in some applications, Carvalho and Scott (2009) integrated out ψ from (9) by assuming a conjugate beta distribution $\text{Beta}(a, b)$, which leads to the prior

$$\Pr(G) \propto B(a + \text{size}(G), b + m - \text{size}(G))/B(a, b), \tag{10}$$

where $B(\cdot, \cdot)$ is the beta function. From the work of Scott and Berger (2006) it follows that prior (10) has an automatic multiplicity correction for testing the inclusion of spurious edges. Armstrong et al. (2009) suggested a hierarchical prior on $\mathcal{G}_p$ (the size-based prior) that gives equal probability to the size of a graph and equal probability to graphs of each size, that is,

$$\begin{aligned}\Pr(G) &= \Pr(\text{size}(G))\Pr(G \mid \text{size}(G)) \\ &= \frac{1}{m+1}\binom{m}{\text{size}(G)}^{-1},\end{aligned} \tag{11}$$

which is also obtained by setting $a = b = 1$ in (10). We note that the expected size of a graph under size-based prior is $m/2$, which is also the expected size of a graph under the uniform prior on $\mathcal{G}_p$.

4. REVERSIBLE JUMP MCMC SAMPLER FOR MATRIX–VARIATE GGMS

We extend our framework to the case when the observed data $\mathcal{D} = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}\}$ are associated with a $p_R \times p_C$ random matrix $\mathbf{X} = (X_{ij})$ that follows a matrix-variate normal distribution

$$\text{vec}(\mathbf{X}) \mid \mathbf{K}_R, \mathbf{K}_C \sim \mathsf{N}_{p_R p_C}(\mathbf{0}, [\mathbf{K}_C \otimes \mathbf{K}_R]^{-1}),$$

with probability distribution function (Gupta and Nagar 2000):

$$\begin{aligned}\mathsf{p}(\mathbf{X} \mid \mathbf{K}_R, \mathbf{K}_C) &= (2\pi)^{-p_R p_C/2}(\det \mathbf{K}_R)^{p_C/2}(\det \mathbf{K}_C)^{p_R/2} \\ &\quad \times \exp\{-\tfrac{1}{2}\operatorname{tr}[\mathbf{K}_R \mathbf{X} \mathbf{K}_C \mathbf{X}^T]\}.\end{aligned} \tag{12}$$

Here $\mathbf{K}_R$ is a $p_R \times p_R$ row precision matrix and $\mathbf{K}_C$ is a $p_C \times p_C$ column precision matrix. Furthermore, we assume that $\mathbf{K}_R \in \mathsf{P}_{G_R}$ and $\mathbf{K}_C \in \mathsf{P}_{G_C}$ where $G_R = (V_{p_R}, E_R)$ and $G_C = (V_{p_C}, E_C)$ are two graphs with p_R and p_C vertices, respectively. We consider the rows $\mathbf{X}_{1*}, \ldots, \mathbf{X}_{p_R *}$ and the columns $\mathbf{X}_{*1}, \ldots, \mathbf{X}_{*p_C}$ of the random matrix $\mathbf{X}$. From theorem 2.3.12 of the book by Gupta and Nagar (2000) we have $\mathbf{X}_{i*}^T \sim \mathsf{N}_{p_C}(\mathbf{0}, (\mathbf{K}_R^{-1})_{ii}\mathbf{K}_C^{-1})$ and $\mathbf{X}_{*j} \sim \mathsf{N}_{p_R}(\mathbf{0}, (\mathbf{K}_C^{-1})_{jj}\mathbf{K}_R^{-1})$. The graphs G_R and G_C define graphical models for the rows and columns of $\mathbf{X}$ (Wang and West 2009):

$$\begin{aligned}&\mathbf{X}_{i_1 *} \perp\!\!\!\perp \mathbf{X}_{i_2 *} \mid \mathbf{X}_{(V_{p_R} \setminus \{i_1, i_2\})*} \\ &\quad \Leftrightarrow \quad (\mathbf{K}_R)_{i_1 i_2} = (\mathbf{K}_R)_{i_2 i_1} = 0 \\ &\quad \Leftrightarrow \quad (i_1, i_2) \notin E_R \qquad \text{and} \\ &\mathbf{X}_{*j_1} \perp\!\!\!\perp \mathbf{X}_{*j_2} \mid \mathbf{X}_{(V_{p_C} \setminus \{j_1, j_2\})*} \\ &\quad \Leftrightarrow \quad (\mathbf{K}_C)_{j_1 j_2} = (\mathbf{K}_R)_{j_2 j_1} = 0 \\ &\quad \Leftrightarrow \quad (j_1, j_2) \notin E_C.\end{aligned} \tag{13}$$

Any prior specification for $\mathbf{K}_R$ and $\mathbf{K}_C$ must take into account the fact that the two precision matrices are not uniquely identified from their Kronecker product which means that, for any $z > 0$, $(z^{-1}\mathbf{K}_R) \otimes (z\mathbf{K}_C) = \mathbf{K}_R \otimes \mathbf{K}_C$ represents the same precision matrix for vec$(\mathbf{X})$—see Equation (12). We follow the idea laid out in the article of Wang and West (2009) and impose the constraint $(\mathbf{K}_C)_{11} = 1$. We define a prior for $\mathbf{K}_C$ through parameter expansion by assuming a G-Wishart prior $\mathsf{Wis}_{G_C}(\delta_C, \mathbf{D}_C)$ for the matrix $z\mathbf{K}_C$ with $z > 0$, $\delta_C > 2$, and $D_C \in \mathsf{P}_{G_C}$. It is immediate to see that the Jacobian of the transformation from $z\mathbf{K}_C$ to $(z, \mathbf{K}_C)$ is $J((z\mathbf{K}_C) \to (z, \mathbf{K}_C)) = z^{|\nu(G_C)|-1}$. It follows that our joint prior for $(z, \mathbf{K}_C)$ is given by

$$\begin{aligned}&\mathsf{p}(z, \mathbf{K}_C \mid G_C, \delta_C, \mathbf{D}_C) \\ &\quad = \frac{1}{I_{G_C}(\delta_C, \mathbf{D}_C)}(\det \mathbf{K}_C)^{(\delta_C - 2)/2} \\ &\qquad \times \exp\left\{-\frac{1}{2}\langle \mathbf{K}_C, z\mathbf{D}_C\rangle\right\} z^{p_C(\delta_C-2)/2 + |\nu(G_C)|-1}.\end{aligned}$$

The elements of $\mathbf{K}_R \in \mathsf{P}_{G_R}$ are not subject to any additional constraints, hence we assume a G-Wishart prior $\mathsf{Wis}_{G_R}(\delta_R, \mathbf{D}_R)$ for $\mathbf{K}_R$. We take $\delta_C = \delta_R = 3$, $\mathbf{D}_C = \mathbf{I}_{p_C}$, and $\mathbf{D}_R = \mathbf{I}_{p_R}$. We complete our prior specification by choosing two priors $\Pr(G_C)$ and $\Pr(G_R)$ for the row and column graphs, where $G_R \in \mathcal{G}_{p_R}$ and $G_C \in \mathcal{G}_{p_C}$—see our discussion from Section 3.

We perform Bayesian inference for matrix-variate GGMs by developing a MCMC algorithm for sampling from the joint posterior distribution of the row and column precision matrices, the row and column graphs, and the auxiliary variable z:

$$\begin{aligned}&\mathsf{p}(\mathbf{K}_R, z, \mathbf{K}_C, G_R, G_C \mid \mathcal{D}) \\ &\quad \propto \mathsf{p}(\mathcal{D} \mid \mathbf{K}_R, \mathbf{K}_C)\mathsf{p}(\mathbf{K}_R \mid G_R, \delta_R, \mathbf{D}_R) \\ &\qquad \times \mathsf{p}(z, \mathbf{K}_C \mid G_C, \delta_C, \mathbf{D}_C)\Pr(G_C)\Pr(G_R),\end{aligned} \tag{14}$$

that is defined for $\mathbf{K}_R \in \mathsf{P}_{G_R}$, $\mathbf{K}_C \in \mathsf{P}_{G_C}$ with $(\mathbf{K}_C)_{11} = 1$ and $z > 0$. Equation (14) is written as

$$\begin{aligned}&\mathsf{p}(\mathbf{K}_R, z, \mathbf{K}_C, G_R, G_C \mid \mathcal{D}) \\ &\quad \propto z^{p_C(\delta_C-2)/2 + |\nu(G_C)|-1} \\ &\qquad \times (\det \mathbf{K}_R)^{(np_C + \delta_R - 2)/2}(\det \mathbf{K}_C)^{(np_R + \delta_C - 2)/2} \\ &\qquad \times \exp\left\{-\frac{1}{2}\operatorname{tr}\left[\sum_{j=1}^{n} \mathbf{K}_R \mathbf{x}^{(j)} \mathbf{K}_C \left(\mathbf{x}^{(j)}\right)^T\right.\right. \\ &\qquad \left.\left. + \mathbf{K}_R \mathbf{D}_R + \mathbf{K}_C(z\mathbf{D}_C)\right]\right\} \\ &\qquad \times \Pr(G_C)\Pr(G_R).\end{aligned} \tag{15}$$

The details of our sampling scheme are presented in Appendix B. Updating the row and column precision matrices involves sampling from their corresponding G-Wishart conditional distributions using the Metropolis–Hastings approach from Section 2.2, while updating z involves sampling from its gamma full conditional. The updates of the row and column graphs involve changes in the dimension of the parameter space and require a reversible jump step (Green 1995).

It is relevant to discuss how our MCMC sampler from Appendix B is different from the methodology proposed by Wang and West (2009). First of all, Wang and West (2009) allowed only decomposable row and column graphs which makes their

framework inappropriate for modeling multivariate lattice data where the graph representing the neighborhood structure of the regions involved is typically not decomposable. Second of all, they proposed a Markov chain sampler for the marginal distribution associated with (G_R, G_C) of the joint posterior distribution (15):

$$\Pr(G_R, G_C \mid \mathcal{D}) \propto \mathrm{p}(\mathcal{D} \mid G_R, G_C) \Pr(G_C) \Pr(G_R), \tag{16}$$

which involves the marginal likelihood

$$\mathrm{p}(\mathcal{D} \mid G_R, G_C) = \int \mathrm{p}(\mathcal{D} \mid \mathbf{K}_R, \mathbf{K}_C) \mathrm{p}(\mathbf{K}_R \mid G_R, \delta_R, \mathbf{D}_R) \times \mathrm{p}(z, \mathbf{K}_C \mid G_C, \delta_C, \mathbf{D}_C)\, d\mathbf{K}_R\, d\mathbf{K}_C\, dz. \tag{17}$$

Wang and West (2009) employed the candidate's formula (Besag 1989; Chib 1990) to approximate the marginal likelihood (17). They sampled from (16) by sequentially updating the row and column graphs. The update of the row graph proceeds by sampling a candidate G_R' from an instrumental distribution $\pi(G_R' \mid G_R)$ and accepting it with Metropolis–Hastings probability

$$\min\left\{ \frac{\mathrm{p}(\mathcal{D} \mid G_R', G_C) \Pr(G_R') \pi(G_R \mid G_R')}{\mathrm{p}(\mathcal{D} \mid G_R, G_C) \Pr(G_R) \pi(G_R' \mid G_R)}, 1 \right\}. \tag{18}$$

The update of the column graph is done in a similar manner. Therefore resampling G_R or G_C requires the computation of a new marginal likelihood (17) which entails a considerable computational effort even for graphs with a small number of vertices. We avoid the computation of marginal likelihoods by sampling from the joint distribution (15) in which the row and column precision matrices have not been integrated out. Sampling from the joint marginal distribution (16) seems appealing because it involves the reduced space of row and column graphs. However, the numerical difficulties associated with repeated calculations of marginal likelihoods (17) outweigh the benefits of working in a smaller space. We empirically compare the efficiency of our inference approach for matrix-variate GGMs with the work of Wang and West (2009) in Section 6.1.

5. BAYESIAN HIERARCHICAL MODELS FOR MULTIVARIATE LATTICE DATA

Conditional autoregressive (CAR) models (Besag 1974; Mardia 1988) are routinely used in spatial statistics to model lattice data. In the case of a single observed outcome in each region, the data are associated with a vector $\mathbf{X} = (X_1, \ldots, X_{p_R})^T$ where X_i corresponds to region i. The zero-centered CAR model is implicitly defined by the set of full conditional distributions

$$X_i \mid \{X_j : j \neq i\} \sim \mathsf{N}\left(\sum_{j \neq i} b_{ij} X_j, \lambda_i^2 \right), \qquad i = 1, \ldots, p_R. \tag{19}$$

Therefore, CAR models are just two-dimensional Gaussian Markov random fields. According to Brook's (1964) theorem, this set of full-conditional distributions implies that the joint distribution for $\mathbf{X}$ satisfies

$$\mathrm{p}(\mathbf{X} \mid \boldsymbol{\lambda}) \propto \exp\left\{ -\tfrac{1}{2} \mathbf{X}^T \boldsymbol{\Lambda}^{-1} (\mathbf{I} - \mathbf{B}) \mathbf{X} \right\},$$

where $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1^2, \ldots, \lambda_{p_R}^2\}$ and $\mathbf{B}$ is a $p_R \times p_R$ matrix such that $\mathbf{B} = (b_{ij})$ and $b_{ii} = 0$. In order for $\boldsymbol{\Lambda}^{-1}(\mathbf{I} - \mathbf{B})$ to be a symmetric matrix we require that $b_{ij}\lambda_j^2 = b_{ji}\lambda_i^2$ for $i \neq j$; therefore the matrix $\mathbf{B}$ and vector $\boldsymbol{\lambda}$ must be carefully chosen. A popular approach is to begin by constructing a symmetric proximity matrix $\mathbf{W} = (w_{ij})$, and then set $b_{ij} = w_{ij}/w_{i+}$ and $\lambda_i^2 = \tau^2/w_{i+}$ where $w_{i+} = \sum_j w_{ij}$ and $\tau^2 > 0$. In that case, $\boldsymbol{\Lambda}^{-1}(\mathbf{I} - \mathbf{B}) = \tau^{-2}(\mathbf{E}_{\mathbf{W}} - \mathbf{W})$, where $\mathbf{E}_{\mathbf{W}} = \mathrm{diag}\{w_{1+}, \ldots, w_{p_R+}\}$. The proximity matrix $\mathbf{W}$ is often constructed by first specifying a neighborhood structure for the geographical areas under study; for example, when modeling state or county level data it is often assumed that two geographical units are neighbors if they share a common border. This neighborhood structure can be summarized in a graph $G_R \in \mathcal{G}_{p_R}$ whose vertices correspond to geographical areas, while its edges are associated with areas that are considered neighbors of each other. The proximity matrix $\mathbf{W}$ is subsequently specified as

$$w_{ij} = \begin{cases} 1, & \text{if } j \in \partial_{G_R}(i) \\ 0, & \text{otherwise,} \end{cases} \tag{20}$$

where $\partial_{G_R}(i)$ denotes the set of vertices that are linked by an edge in the graph G_R with the vertex i.

Specifying the joint precision matrix for $\mathbf{X}$ using the proximity matrix derived from the neighborhood structure is very natural; essentially, it implies that observations collected on regions that are not neighbors are conditionally independent from each other given the rest. However, note that the specification in (20) implies that $(\mathbf{E}_{\mathbf{W}} - \mathbf{W})\mathbf{1}_{p_R} = \mathbf{0}$ and therefore the joint distribution on $\mathbf{X}$ is improper. Here $\mathbf{1}_l$ is the column vector of length l with all elements equal to 1. Proper CAR models (Cressie 1973; Sun et al. 2000; Gelfand and Vounatsou 2003) can be obtained by including a spatial autocorrelation parameter ρ, so that

$$X_i \mid \{X_j : j \neq i\} \sim \mathsf{N}\left(\rho \sum_{j \neq i} \frac{w_{ij}}{w_{i+}} X_j, \frac{\tau^2}{w_{i+}} \right). \tag{21}$$

The joint distribution on $\mathbf{X}$ is then multivariate normal $\mathsf{N}_{p_R}(\mathbf{0}, \mathbf{V}_{\mathbf{W}}^{-1}(\tau^{-2}, \rho))$ where $\mathbf{V}_{\mathbf{W}}(\tau^{-2}, \rho) = \tau^{-2}(\mathbf{E}_{\mathbf{W}} - \rho\mathbf{W}) \in \mathsf{P}_{G_R}$. This distribution is proper as long as ρ is between the reciprocals of the minimum and maximum eigenvalues for $\mathbf{W}$. In particular, note that taking $\rho = 0$ leads to independent random effects.

In the spirit of Besag and Kooperberg (1995), an alternative but related approach to the construction of models for lattice data is to let $\mathbf{X} \sim \mathsf{N}_{p_R}(\mathbf{0}, \mathbf{K}^{-1})$ and assign $\mathbf{K}$ a G-Wishart prior $\mathsf{Wis}_{G_R}(\delta_R, (\delta_R - 2)\mathbf{D}_R)$, where $\mathbf{D}_R = \mathbf{V}_{\mathbf{W}}^{-1}(\tau^{-2}, \rho)$. The mode of $\mathsf{Wis}_{G_R}(\delta_R, (\delta_R - 2)\mathbf{D}_R)$ is the unique matrix $\mathbf{K} = (K_{ij}) \in \mathsf{P}_{G_R}$ that satisfies the relations

$$\begin{aligned} (\mathbf{K}^{-1})_{ij} &= (\mathbf{D}_R)_{ij}, \quad \text{if } j \in \partial_{G_R}(i), \qquad \text{and} \\ K_{ij} &= 0, \quad \text{if } j \notin \partial_{G_R}(i). \end{aligned} \tag{22}$$

The matrix $\mathbf{D}_R^{-1} = \mathbf{V}_{\mathbf{W}}(\tau^{-2}, \rho)$ verifies the system (22), hence it is the mode of $\mathsf{Wis}_{G_R}(\delta_R, (\delta_R - 2)\mathbf{D}_R)$. As such, the mode of the prior for $\mathbf{K}_R$ induces the same prior specification for $\mathbf{X}$ as (21). It is easy to see that, conditional on $\mathbf{K}_R \in \mathsf{P}_{G_R}$, we have

$$X_i \mid \{X_j : j \neq i\} \sim \mathsf{N}\left(-\sum_{j \in \partial_{G_R}(i)} \frac{K_{ij}}{K_{ii}} X_j, \frac{1}{K_{ii}} \right). \tag{23}$$

Hence, by modeling $\mathbf{X}$ using a Gaussian graphical model and restricting the precision matrix $\mathbf{K}$ to belong to the cone P_{G_R},

we are inducing a mixture of CAR priors on $\mathbf{X}$ where the priors on

$$b_{ij} = \begin{cases} -K_{ij}/K_{ii}, & \text{if } j \in \partial_{G_R}(i) \\ 0, & \text{otherwise,} \end{cases}$$

and $\lambda_i^2 = 1/K_{ii}$ [see Equation (19)] are induced by the G-Wishart prior $\mathsf{Wis}_{G_R}(\delta_R, (\delta_R - 2)\mathbf{D}_R)$.

The specification of CAR models through G-Wishart priors solves the impropriety problem of intrinsic CAR models and preserves the computational advantages derived from standard CAR specifications while providing greater flexibility. Indeed, the prior is trivially proper because the matrix $\mathbf{K} \in \mathsf{P}_{G_R}$ is invertible by construction. The computational advantages are preserved because the full conditional distributions for each X_i can be easily computed for any matrix $\mathbf{K}$ without the need to perform matrix inversion, and they depend only on a small subset of neighbors $\{X_j : j \in \partial_{G_R}(i)\}$. Additional flexibility is provided because the weights b_{ij} for $j \in \partial_{G_R}(i)$ and smoothing parameters λ_i^2 are being estimated from the data rather than being assumed fixed, allowing for adaptive spatial smoothing. Our approach provides what can be considered as a nonparametric alternative to the parametric estimates of the proximity matrix of Cressie and Chan (1989).

A similar approach can be used to construct proper multivariate conditional autoregressive (MCAR) models (Mardia 1988; Gelfand and Vounatsou 2003). In this case, we are interested in modeling a $p_R \times p_C$ matrix $\mathbf{X} = (X_{ij})$ where X_{ij} denotes the value of the jth outcome in region i. We let $\mathbf{X}$ follow a matrix-variate normal distribution with row precision matrix $\mathbf{K}_R$ capturing the spatial structure in the data (which, as in univariate CAR models, is restricted to the cone P_{G_R} defined by the neighborhood graph G_R), and column precision matrix $\mathbf{K}_C$, which controls the multivariate dependencies across outcomes. It can be easily shown that the row vector $\mathbf{X}_{i*}$ of $\mathbf{X}$ depends only on the row vectors associated with those regions that are neighbors with i—see also Equation (13):

$$\mathbf{X}_{i*}^T \mid \{\mathbf{X}_{j*}^T : j \neq i\} \sim \mathsf{N}_{p_C}\left(-\sum_{j \in \partial_{G_R}(i)} \frac{(\mathbf{K}_R)_{ij}}{(\mathbf{K}_R)_{ii}} \mathbf{X}_{j*}^T, \frac{1}{(\mathbf{K}_R)_{ii}} \mathbf{K}_C^{-1} \right).$$

Given the matrix-variate GGMs framework from Section 4, we can model the conditional independence relationships across outcomes through a column graph $G_C \in \mathcal{G}_{p_C}$ and require $\mathbf{K}_C \in \mathsf{P}_{G_C}$. As opposed to the neighborhood graph G_R which is known and considered fixed, the graph G_C is typically unknown and needs to be inferred from the data.

The matrix-variate GGM formulation for spatial models can also be used as part of more complex hierarchical models. Indeed, CAR and MCAR models are most often used as a prior for the random effects of a generalized linear model (GLM) to account for residual spatial structure not accounted for by covariates. When no covariates are available, the model can be interpreted as a spatial smoother where the spatial covariance matrix controls the level of spatial smoothing in the underlying (latent) surface. Similarly, MCAR models can be used to construct multivariate spatial GLMs. More specifically, consider the $p_R \times p_C$ matrix $\mathbf{Y} = (Y_{ij})$ of discrete or continuous outcomes, and let $Y_{ij} \sim h_j(\cdot \mid \eta_{ij})$ where h_j is a probability mass or probability density function that belongs to the exponential family with location parameter η_{ij}. The spatial GLM is then defined through the linear predictor

$$g^{-1}(\eta_{ij}) = \mu_j + X_{ij} + \mathbf{Z}_{ij}\boldsymbol{\beta}_j, \tag{24}$$

where $g(\cdot)$ is the link function, μ_j is an outcome-specific intercept, X_{ij} is a zero-centered spatial random effect associated with location i, $\mathbf{Z}_{ij}$ is a matrix of observed covariates for outcome j at location i, and $\boldsymbol{\beta}_j$ is the vector of fixed effects associated with outcome j. As an example, by choosing $y_{ij} \sim \mathsf{Poi}(\eta_{ij})$, $g^{-1}(\cdot) = \log(\cdot)$ we obtain a multivariate spatial log-linear model for count data, which is often used for disease mapping in epidemiology (see Section 6.2). We further assign a matrix-variate normal distribution for $\mathbf{X}$ with independent G-Wishart priors for $\mathbf{K}_R$ and $z\mathbf{K}_C$, where $z > 0$ is an auxiliary variable needed to impose the identifiability constraint $(\mathbf{K}_C)_{11} = 1$:

$$\begin{aligned} \mathbf{K}_R \mid \delta_R, \mathbf{D}_R &\sim \mathsf{Wis}_{G_R}(\delta_R, (\delta_R - 2)\mathbf{D}_R), \\ (z\mathbf{K}_C) \mid \delta_C, \mathbf{D}_C &\sim \mathsf{Wis}_{G_C}(\delta_C, \mathbf{D}_C). \end{aligned} \tag{25}$$

Prior elicitation for this class of spatial models is relatively straightforward. Indeed, elicitation of the matrix $\mathbf{D}_R = (\mathbf{E}_\mathbf{W} - \rho\mathbf{W})^{-1}$ requires only the elicitation of the neighborhood matrix $\mathbf{W}$ which also defines the neighborhood graph G_R, along with reasonable values for the spatial autocorrelation parameter ρ. In particular, in the application we discuss in Section 6.2 we assume that, a priori, there is a strong degree of positive spatial association, and choose a prior for ρ that gives higher probabilities to values close to 1 (Gelfand and Vounatsou 2003):

$$\rho \sim \mathsf{Uni}(\{0, 0.05, 0.1, \ldots, 0.8, 0.82, \ldots, 0.90, 0.91, \ldots, 0.99\}). \tag{26}$$

In the case of MCAR models, it is common to assume that the prior value for the conditional covariance between variables is zero, which leads to choosing $\mathbf{D}_C$ from the G-Wishart prior for $z\mathbf{K}_C$ to be a diagonal matrix. We note that the scale parameter τ^2 is no longer identifiable in the context of the joint prior (25), hence it is then sensible to pick $\mathbf{D}_C = \mathbf{I}_{p_C}$.

At this point, a word of caution about the interpretation of the models seems appropriate. The graphs $G_R = (V_{p_R}, E_R)$ and $G_C = (V_{p_C}, E_C)$ induce conditional independence relationships associated with the rows and columns of the random-effects matrix $\mathbf{X}$—see Equation (13). Similar conditional independence relationships hold for the rows and columns of the matrix of location parameters $\boldsymbol{\eta} = (\eta_{ij})$, but these extend to the observed outcomes $\mathbf{Y}$ only if its entries are continuous (for a more thorough discussion, see the Supplementary Materials). Hence, the reader must be careful when interpreting the results of these models when binary or count data are involved; in these cases, any statement about conditional independence needs to be circumscribed to the location parameters $\boldsymbol{\eta}$.

6. ILLUSTRATIONS

6.1 Simulation Study

We empirically compare the reversible jump MCMC sampler for matrix-variate GGMs proposed in Section 4 (RJ-DLR) with the methodology of Wang and West (2009) (WW). We consider the row graph $\bar{G}_R = (V_{p_R}, \bar{E}_R)$ with $p_R = 5$ vertices and with edges $\bar{E}_R = \{(1, i) : 1 < i \leq 5\} \cup \{(2, 3), (3, 4), (4, 5), (2, 5)\}$.

Table 2. Average estimates (left panel) and their standard errors (right panel) of posterior edge inclusion probabilities for the row graph $\bar{G}_R$ obtained using the RJ-DLR algorithm (below diagonal) and the WW algorithm (above diagonal). The boxes identify the edges that are in $\bar{G}_R$

	1	2	3	4	5	1	2	3	4	5
1	–	0.997	0.145	0.739	0.701	–	0.002	0.173	0.365	0.364
2	1	–	0.997	0.676	0.736	0	–	0.007	0.347	0.367
3	1	1	–	0.997	0.147	0	0	–	0.012	0.180
4	1	0.026	1	–	0.998	0	0.024	0	–	0.002
5	1	1	0.040	1	–	0	0	0.048	0	–

We take the column graph $\bar{G}_C$ to be the cycle graph with $p_C = 10$ vertices (see Section 2.3). We note that both the row and column graphs are non-decomposable, with $\bar{G}_R$ being relatively dense and with $\bar{G}_C$ being relatively sparse. We define the row and column precision matrices $\bar{\mathbf{K}}_R \in \mathsf{P}_{\bar{G}_R}$ and $\bar{\mathbf{K}}_C \in \mathsf{P}_{\bar{G}_C}$ to have diagonal elements equal to 1 and nonzero off-diagonal elements equal with 0.4. We generate 100 datasets each comprising $n = 100$ observations sampled from the 5×10 matrix-variate normal distribution $\mathsf{p}(\cdot \mid \bar{\mathbf{K}}_R, \bar{\mathbf{K}}_C)$—see Equation (12).

For each dataset we attempted to recover the edges of $\bar{G}_R$ and $\bar{G}_C$ with the RJ-DLR and the WW algorithms. The two MCMC samplers were run for 10,000 iterations with a burn-in of 1000 iterations. We implemented the RJ-DLR algorithm in R and C++ (code available as Supplemental Material). We used the code developed by Wang and West (2009) for their method. In the RJ-DLR algorithm we set $\sigma_{m,R} = \sigma_{m,C} = \sigma_{g,R} = \sigma_{g,C} = 0.5$, which yields average rejection rates on Metropolis–Hastings updates of about 0.3 for both $\bar{\mathbf{K}}_C$ and $\bar{\mathbf{K}}_R$. We assumed independent uniform priors for the row and column graphs.

Tables 2 and 3 show the average posterior edge inclusion probabilities for $\bar{G}_R$ and $\bar{G}_C$. Our RJ-DLR method recovers the structure of both graphs very well: the edges that belong to each graph receive posterior inclusion probabilities of 1, while the edges that are absent from each graph receive posterior inclusion probabilities below 0.1. The ability of the RJ-DLR algorithm to recover the structures of a dense row graph and of a sparse column graph is quite encouraging since the uniform priors for the row and column graphs favored middle-size graphs.

The WW method shows diminished performance in recovering graphical structures. Edges $(1, 4)$, $(1, 5)$, and $(2, 5)$, which are actually in $\bar{G}_R$, are given average inclusion probabilities of only 0.7. Edge $(1, 3)$, which is in $\bar{G}_R$, receives a very low inclusion probability, while edge $(2, 4)$, which is not in $\bar{G}_R$, is given a high inclusion probability. The WW algorithm includes edges that are in the column graph with relatively high prob-

Table 3. Average estimates (upper panel) and their standard errors (lower panel) of posterior edge inclusion probabilities for the column graph $\bar{G}_C$ obtained using the RJ-DLR algorithm (below diagonal) and the WW algorithm (above diagonal). The boxes identify the edges that are in $\bar{G}_C$

	1	2	3	4	5	6	7	8	9	10
1	–	0.902	0.492	0.298	0.341	0.246	0.362	0.417	0.401	0.948
2	1	–	0.929	0.453	0.362	0.303	0.347	0.339	0.319	0.468
3	0.024	1	–	0.896	0.474	0.338	0.304	0.363	0.370	0.439
4	0.046	0.032	1	–	0.904	0.398	0.398	0.385	0.343	0.336
5	0.022	0.046	0.033	1	–	0.952	0.525	0.374	0.283	0.386
6	0.040	0.038	0.036	0.034	1	–	0.947	0.369	0.259	0.325
7	0.037	0.030	0.052	0.025	0.042	1	—	0.933	0.405	0.383
8	0.047	0.046	0.024	0.053	0.032	0.049	1	–	0.932	0.525
9	0.072	0.082	0.039	0.029	0.036	0.068	0.039	1	–	0.955
10	1	0.057	0.037	0.030	0.047	0.044	0.035	0.066	1	–
1	–	0.228	0.371	0.343	0.393	0.334	0.390	0.418	0.366	0.135
2	0	–	0.169	0.390	0.391	0.373	0.391	0.359	0.382	0.378
3	0.086	0	–	0.240	0.398	0.392	0.372	0.369	0.388	0.399
4	0.129	0.093	0	–	0.225	0.396	0.391	0.403	0.376	0.391
5	0.087	0.125	0.081	0	–	0.127	0.382	0.374	0.356	0.420
6	0.104	0.135	0.131	0.103	0	–	0.134	0.360	0.342	0.385
7	0.109	0.110	0.125	0.085	0.117	0	–	0.151	0.404	0.369
8	0.140	0.139	0.088	0.134	0.095	0.150	0	–	0.166	0.407
9	0.156	0.181	0.111	0.102	0.103	0.173	0.117	0	–	0.140
10	0	0.150	0.103	0.095	0.136	0.132	0.098	0.168	0	–

abilities (typically above 0.9). However, edges that do not belong to $\bar{G}_C$ also appear to be included quite often, with an average inclusion probability between 0.3 and 0.4. These tables also show a large standard deviation of edge inclusion probabilities across the 100 datasets using the WW method. The decreased performance of the WW algorithm versus our own RJ-DLR algorithm is likely attributable to the fact that Wang and West (2009) accounted only for decomposable graphs in their framework, which may be poor at approximating the nondecomposable graphs considered in this example.

6.2 Mapping Cancer Mortality in the United States

Accurate and timely counts of cancer mortality are very useful in the cancer surveillance community for purposes of efficient resource allocation and planning. Estimation of current and future cancer mortality broken down by geographic area (state) and tumor has been discussed in a number of recent articles, including those by Tiwari et al. (2004), Ghosh and Tiwari (2007), Ghosh et al. (2007), and Ghosh, Ghosh, and Tiwari (2008). This section considers a multivariate spatial model on state-level cancer mortality rates in the United States for 2000. These mortality data are based on death certificates that are filed by certifying physicians. They are collected and maintained by the National Center for Health Statistics (*http://www.cdc.gov/nchs*) as part of the National Vital Statistics System. The data are available from the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute (*http://seer.cancer.gov/seerstat*).

The data we analyze consist of mortality counts for 11 types of tumors recorded in the 48 continental states plus the District of Columbia. Since mortality counts below 25 are often deemed unreliable in the cancer surveillance community, we treated them as missing. Along with mortality counts, we obtained population counts in order to model death risk. Figure 1 shows raw mortality rates for four of the most common types of tumors (colon, lung, breast, and prostate). Although the pattern is slightly different for each of these cancers, a number of striking similarities are present; for example, Colorado appears to be a state with low mortality for all of these common cancers, while West Virginia, Pennsylvania, and Florida present relatively high mortality rates.

We consider modeling these data using Poisson multivariate log-linear models with spatial random effects. More concretely, we let Y_{ij} be the number of deaths in state $i = 1, \ldots, p_R = 49$ for tumor type $j = 1, \ldots, p_C = 11$. We set

$$Y_{ij} \mid \eta_{ij} \sim \mathsf{Poi}(\eta_{ij}), \qquad \log(\eta_{ij}) = \log(m_i) + \mu_j + X_{ij}. \quad (27)$$

Here m_i is the population of state i, μ_j is the intercept for tumor type j, and X_{ij} is a zero-mean spatial random effect associated with location i and tumor j. We denote $\widetilde{X}_{ij} = \mu_j + X_{ij}$. We further model $\widetilde{\mathbf{X}} = (\widetilde{X}_{ij})$ with a matrix-variate normal prior:

$$\operatorname{vec}(\widetilde{\mathbf{X}}) \mid \boldsymbol{\mu}, \mathbf{K}_R, \mathbf{K}_C \sim \mathsf{N}_{p_R p_C}\big(\operatorname{vec}\{(\mathbf{1}_{p_R}\boldsymbol{\mu}^T)^T\}, [\mathbf{K}_C \otimes \mathbf{K}_R]^{-1}\big), \quad (28)$$

where $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{p_C})^T$.

We consider four models in total. The first two models we consider (GGM-U and GGM-S) are based on the sparse models for multivariate aerial data described in Section 5, and differ only in terms of the prior assigned on the space of graphs [GGM-U uses a uniform prior, while GGM-S uses the size-based prior from Equation (11)]. Hence, the joint prior for the row and column precision matrices is given in (25) for both models. The column graph G_C is unknown and allowed to vary in $\mathcal{G}_{11}$. The row graph G_R is fixed and derived from the incidence matrix $\mathbf{W}$ corresponding with the neighborhood structure of the U.S. states. More explicitly, each state is associated with a vertex in G_R. Two states are linked by an edge in G_R if they

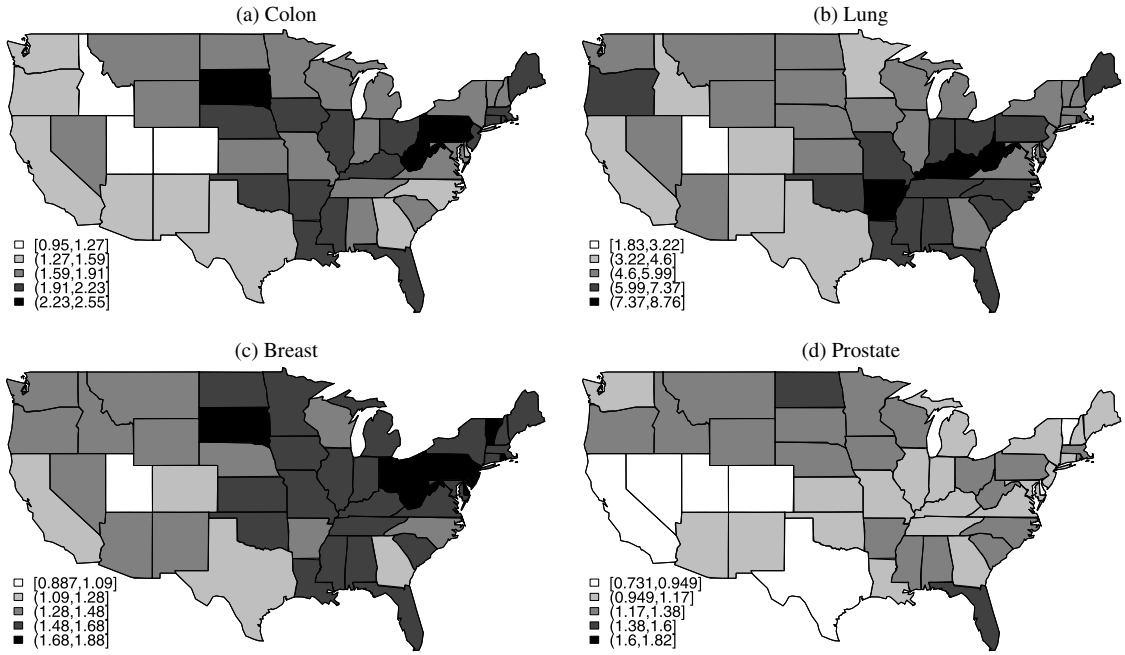

Figure 1. Mortality rates (per 10,000 habitants) in the 48 continental states and the D.C. area corresponding to four common cancers during 2000.

share a common border. The Supplementary Materials describe the neighborhood structure of G_R and provide its decomposition into maximal prime subgraphs. What is interesting about the graph G_R, and relevant in light of the results of Section 2.3, is that its decomposition yields 13 maximal prime subgraphs. Most of these prime components are complete and only contain two or three vertices; however, one maximal prime subgraph has 36 vertices (states), making G_R non-decomposable with a sizable maximal prime subgraph. The degrees of freedom for the G-Wishart priors are set as $\delta_R = \delta_C = 3$, while the centering matrices are chosen as $\mathbf{D}_C = \mathbf{I}_{p_C}$ and $\mathbf{D}_R = (\mathbf{E}_\mathbf{W} - \rho\mathbf{W})^{-1}$.

The third model, which we call model FULL, is obtained by keeping the column graph G_C fixed to the full graph. This is equivalent to replacing the G-Wishart prior for $(z\mathbf{K}_C)$ from Equation (25) with a Wishart prior $\mathsf{Wis}_{p_C}(\delta_C, \mathbf{D}_C)$. Finally, model MCAR is obtained from model FULL by substituting the G-Wishart prior for $\mathbf{K}_R$ from Equation (25) with $\mathbf{K}_R = \mathbf{E}_\mathbf{W} - \rho\mathbf{W}$. The resulting prior distribution for the spatial random effects $\widetilde{\mathbf{X}}$ in model MCAR is precisely the MCAR(ρ, Σ) prior of Gelfand and Vounatsou (2003). In all cases, we complete the model specification by choosing the prior from Equation (26) for the spatial autocorrelation parameter ρ and a multivariate normal prior for the mean rates vector $\boldsymbol{\mu} \sim \mathsf{N}_{p_C}(\boldsymbol{\mu}_0, \boldsymbol{\Omega}^{-1})$ where $\boldsymbol{\mu}_0 = \mu_0 \mathbf{1}_{p_C}$ and $\boldsymbol{\Omega} = \omega^{-2}\mathbf{I}_{p_C}$. We set μ_0 to be the median log incidence rate across all cancers and regions, and ω to be twice the interquartile range in raw log incidence rates.

Posterior inferences for models GGM-U and GGM-S are obtained by extending the sampling algorithm from Appendix B; details are presented in the Supplemental Materials. MCMC samplers for models FULL and MCAR are similarly derived in a straightforward manner. Missing counts were sampled from their corresponding predictive distributions. We monitored the chains to ensure convergence for each model. We set $\sigma_{m,R} = \sigma_{g,R} = 0.2$ and $\sigma_{m,C} = \sigma_{g,C} = 0.5$, which achieved rejection rates of roughly 0.3 when updating elements of both $\mathbf{K}_R$ or $\mathbf{K}_C$. Furthermore, the average acceptance rate of a reversible jump move for the column graph G_C was around 0.22 for both GGM-U and GGM-S.

To assess the out-of-sample predictive performance of the four models, we performed a 10-fold cross-validation exercise. The exercise was run by randomly dividing the non-missing counts (those above 25) in ten bins. For each bin j, we used the samples from the other nine bins as data and treated the samples from bin j as missing. In this comparison, the MCMC sampler for each of the four models was run ten times for 160,000 iterations and the first half of the run was discarded as burn-in. In the sequel, we denote the predicted counts for model $\mathcal{M} \in \{\text{GGM-U}, \text{GGM-S}, \text{FULL}, \text{MCAR}\}$ by $Y_{ij}^{cv}(\mathcal{M})$. We employ the goodness-of-fit (mean squared error of the posterior predictive mean or MSE) and the variability/penalty (mean predictive error or VAR) terms of Gelfand and Ghosh (1998), that is,

$$\text{MSE}(\mathcal{M}) = \frac{1}{|\{(i,j) : Y_{ij} \geq 25\}|} \times \sum_{\{(i,j) : Y_{ij} \geq 25\}} \left(\mathsf{E}_\mathcal{M}[Y_{ij}^{cv}(\mathcal{M})] - Y_{ij}\right)^2,$$

$$\text{VAR}(\mathcal{M}) = \frac{1}{|\{(i,j) : Y_{ij} \geq 25\}|} \sum_{\{(i,j) : Y_{ij} \geq 25\}} \mathsf{Var}_\mathcal{M}[Y_{ij}^{cv}(\mathcal{M})],$$

Table 4. Ten-fold cross-validation predictive scores in the U.S. cancer mortality example. Model GGM-S is best with respect to fit (MSE) and ranked probability score (RPS), while model GGM-U is best with respect to variability (VAR) of the out-of-sample predicted counts

Model	MSE	VAR	RPS
GGM-U	17,379.9	**23,685.2**	62.1
GGM-S	**16,979.8**	24,361.1	**61.8**
FULL	18,959.6	24,530.4	63.2
MCAR	19,211.1	47,568.7	76.7

to compare the predictive mean and variance of the four models. We also calculated the ranked probability score (RPS) (discussed by Czado, Gneiting, and Held 2009 in the context of count data), which measures the accuracy of the entire predictive distribution. The results, which are summarized in Table 4, reveal an interesting progression in the out-of-sample predictive performance of the four methods. Under either the MSE, VAR, or the RPS criteria, model MCAR performs considerably worse than model FULL, which in turn is outperformed by both models GGM-U and GGM-S. Also, the effect of the choice of prior on the column graph space appears negligible as GGM-U and GGM-S have roughly the same predictive performance. It is noteworthy to mention that the improvement in the *RPS* in the sequence of models MCAR, FULL, GGM-U, and GGM-S comes both because of a better prediction of the means as well as because the predictive distributions become sharper—see the article by Gneiting and Raftery (2007) for a discussion of the trade-off between sharpness and calibration in the formation of predictive distributions. The dramatic improvement in predictive performance that results from moving from model MCAR to model FULL is the result of using the G-Wishart distribution to allow greater flexibility in the spatial interactions over the CAR specification suggested by Gelfand and Vounatsou (2003). On the other hand, the additional gain yielded by the use of models GGM-U or GGM-S versus model FULL would seem to come from the increased parsimony of the GGM models.

We also conducted an in-sample comparison of the four models. For each model, we ran ten Markov chains for 160,000 iterations and discarded the first half as burn-in based; Figure 2 is indicative of the performance of the MCMC sampler we employed. These runs used the entire dataset (with the exception of counts below 25, which are still treated as missing), rather than the cross-validation datasets. The Supplemental Materials give detailed tables of fitted values and 95% credible intervals for all counts in the dataset obtained using each of the four models. Table 5 summarizes these results by presenting the empirical coverage rates and the average lengths of these in-sample 95% credible intervals for each of the four models. We see that the methods based on the G-Wishart priors for the spatial component (GGM-U, GGM-S, FULL) have very similar coverage probabilities that are close to the nominal 95%. These three models also tend to have relatively short credible intervals, with GGM-U and GGM-S presenting the smallest (and almost identical) values. In contrast, model MCAR returns credible intervals that are too wide, containing 99% of all observed values. As with the out-of-sample exercise, these results suggest

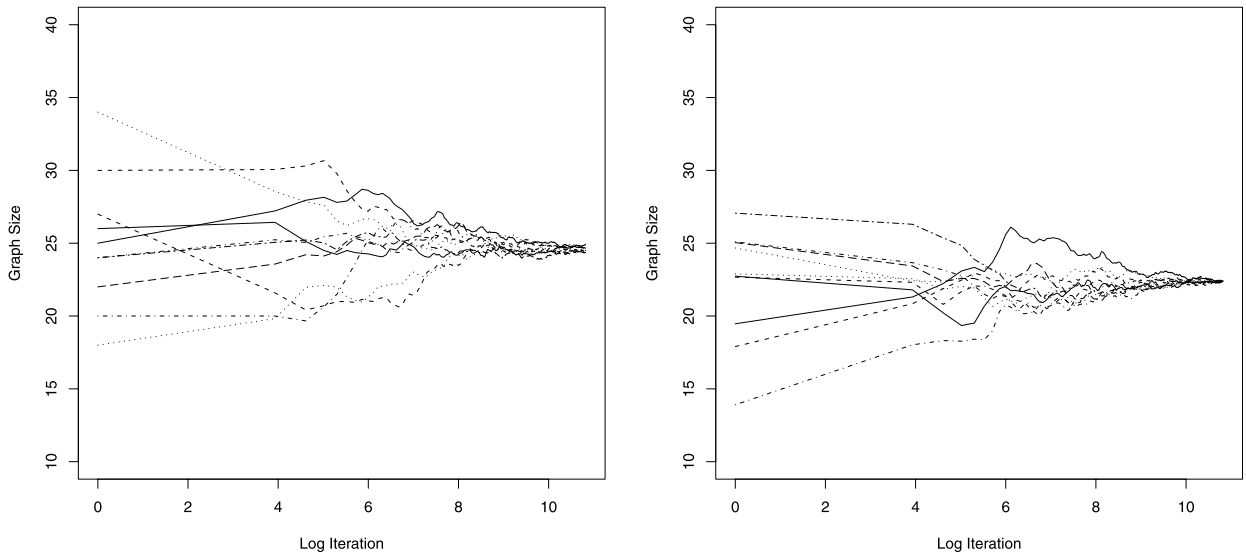

Figure 2. Convergence plot of the average size of the column graph G_C by log iteration for model GGM-U (left panel) and model GGM-S (right panel).

that both the additional flexibility in modeling the spatial dependence and the sharper estimates of the dependence across cancer types provided by the matrix-variate GGMs are key to prevent overfitting.

Finally, we compare the estimates of the column graph G_C under GGM-U and GGM-S to assess the sensitivity of the model to the prior on graph space. Figure 3 shows a heatmap of the posterior edge inclusion probabilities in the column graph G_C, with the lower triangle corresponding to GGM-U and the upper triangle to GGM-S. Overall there is substantial coherence in terms of the effects of the uniform and size-based priors on $\mathcal{G}_{11}$ on the frequency on which edges are added or deleted from the column graph. From the left panel of Figure 2 we see that the average size of the column graph G_C in GGM-U is 24.8 edges while the right panel shows that the average graph size in model GGM-S is 22.5 edges. Both of these numbers are quite close to 27.5—the average graph size under the uniform prior and the size-based prior on $\mathcal{G}_{11}$. This suggests that there is little sensitivity to the prior in terms of graph selection.

7. DISCUSSION

In this article we have developed and illustrated a surprisingly powerful computational approach for multivariate and matrix-variate GGMs that scales well to problems with a moderate number of variables. Convergence seems to be achieved quickly (usually within the first 5000 iterations) and acceptance rates are within reasonable ranges and can be easily controlled. In the context of the sparse multivariate models for aerial data discussed in Section 5 the running times, although longer than for conventional MCAR models, are still short enough to make routine implementation feasible (e.g., in the cancer surveillance example from Section 6.2, about 22 hours on a dual-core 2.8 GHz computer running Linux). However, computations can still be quite challenging when the number of units in the lattice is very high. We plan to explore more efficient implementations that exploit the sparse structure of the Cholesky decompositions of precision matrices induced by GGMs.

Table 5. Nominal coverage rates and mean length of the in-sample 95% credible intervals for the U.S. cancer mortality example obtained using the four Poisson multivariate log-linear models

Model	Coverage rate	Mean length
GGM-U	0.960	**65.464**
GGM-S	0.964	65.474
FULL	0.964	65.885
MCAR	0.990	69.239

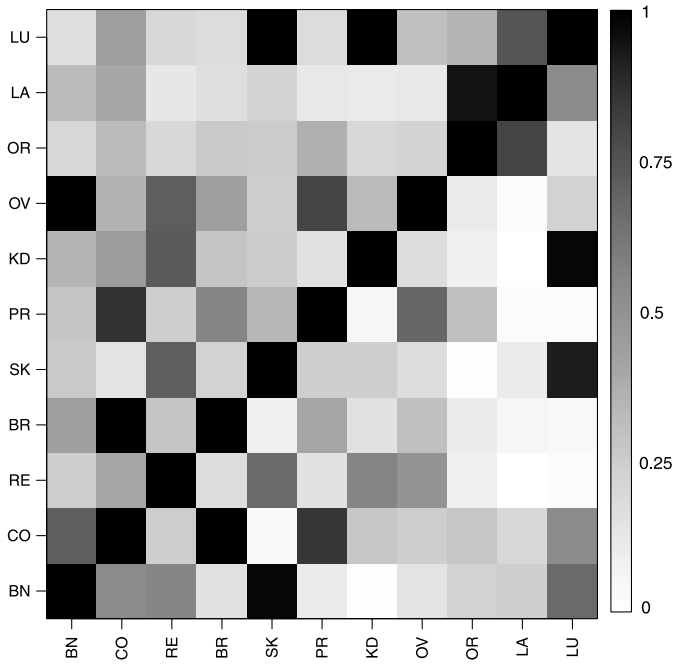

Figure 3. Edge inclusion probabilities for model GGM-U (lower triangle) and model GGM-S (upper triangle) in the U.S. cancer mortality example. The acronyms used are explained in the Supplementary Materials.

We believe that our application to the construction of spatial models for lattice data makes for particularly appealing illustrations. On one hand, the U.S. cancer mortality example we have considered suggests that the additional flexibility in the spatial correlation structure provided by our approach is necessary to accurately model some spatial datasets. Indeed, our approach allows for "nonstationary" CAR models, where the spatial autocorrelation and spatial smoothing parameters vary spatially. On the other hand, to the best of our knowledge, we are unaware of any approach in the literature to construct and estimate potentially sparse MCAR, particularly under a Bayesian approach. In addition to providing insights into the mechanisms underlying the data generation process, the model offers drastically improved predictive performance and sharper estimation of model parameters.

APPENDIX A: DETAILS OF THE REVERSIBLE JUMP MCMC SAMPLER FOR GGMS

The MCMC algorithm from Section 3 sequentially updates the elements of the precision matrix and the edges of the underlying graph as follows. We denote the current state of the chain by $(\mathbf{K}^{[s]}, G^{[s]})$, $\mathbf{K}^{[s]} \in \mathsf{P}_{G^{[s]}}$. Its next state $(\mathbf{K}^{[s+1]}, G^{[s+1]})$, $\mathbf{K}^{[s+1]} \in \mathsf{P}_{G^{[s+1]}}$, is generated by sequentially performing the following two steps. We make use of two strictly positive precision parameters σ_m and σ_g that remain fixed at some suitable small values. We assume that the ordering of the variables has been changed according to a permutation υ selected at random from the uniform distribution on Υ_p. We denote by $(\mathbf{U}+\mathbf{D}_0)^{-1} = (\mathbf{Q}^*)^T\mathbf{Q}^*$ the Cholesky decomposition of $(\mathbf{U}+\mathbf{D}_0)^{-1}$, where the rows and columns of this matrix have been permuted according to υ.

We denote by $\mathsf{nbd}_p^+(G)$ the graphs that can be obtained by adding an edge to a graph $G \in \mathcal{G}_p$ and by $\mathsf{nbd}_p^-(G)$ the graphs that are obtained by deleting an edge from G. We call the one-edge-way set of graphs $\mathsf{nbd}_p(G) = \mathsf{nbd}_p^+(G) \cup \mathsf{nbd}_p^-(G)$ the neighborhood of G in $\mathcal{G}_p$. These neighborhoods connect any two graphs in $\mathcal{G}_p$ through a sequence of graphs such that two consecutive graphs in this sequence are each others' neighbors.

Step 1: Resample the graph. We sample a candidate graph $G' \in \mathsf{nbd}_p(G^{[s]})$ from the proposal

$$q(G' \mid G^{[s]}) = \tfrac{1}{2}\,\mathsf{Uni}\big(\mathsf{nbd}_p^+(G^{[s]})\big) + \tfrac{1}{2}\,\mathsf{Uni}\big(\mathsf{nbd}_p^-(G^{[s]})\big), \quad \text{(A.1)}$$

where $\mathsf{Uni}(A)$ represents the uniform distribution on the discrete set A. The distribution (A.1) gives an equal probability of proposing to delete an edge from the current graph and of proposing to add an edge to the current graph. We favor (A.1) over the more usual proposal distribution $\mathsf{Uni}(\mathsf{nbd}_p(G^{[s]}))$ that is employed, for example, by Madigan and York (1995). If $G^{[s]}$ contains a very large or a very small number of edges, the probability of proposing a move that adds or, respectively, deletes an edge from $G^{[s]}$ is extremely small when sampling from $\mathsf{Uni}(\mathsf{nbd}_p(G^{[s]}))$, which could lead to poor mixing in the resulting Markov chain.

We assume that the candidate G' sampled from (A.1) is obtained by adding the edge (i_0, j_0), $i_0 < j_0$, to $G^{[s]}$. Since $G' \in \mathsf{nbd}_p^+(G^{[s]})$ we have $G^{[s]} \in \mathsf{nbd}_p^-(G')$. We consider the decomposition of the current precision matrix $\mathbf{K}^{[s]} = (\mathbf{Q}^*)^T((\mathbf{\Psi}^{[s]})^T\mathbf{\Psi}^{[s]})\mathbf{Q}^*$ with $(\mathbf{\Psi}^{[s]})^{\nu(G^{[s]})} \in \mathsf{M}^{\nu(G^{[s]})}$. Since the vertex i_0 has one additional neighbor in G', we have $d_{i_0}^{G'} = d_{i_0}^{G^{[s]}}$, $d_{j_0}^{G'} = d_{j_0}^{G^{[s]}} + 1$, $v_{i_0}^{G'} = v_{i_0}^{G^{[s]}} + 1$, $v_{j_0}^{G'} = v_{j_0}^{G^{[s]}}$, and $\nu(G') = \nu(G^{[s]}) \cup \{(i_0, j_0)\}$. We define an upper triangular matrix $\mathbf{\Psi}'$ such that $\Psi'_{ij} = \Psi_{ij}^{[s]}$ for $(i,j) \in \nu(G^{[s]})$. We sample $\gamma \sim \mathsf{N}(\Psi_{i_0j_0}^{[s]}, \sigma_g^2)$ and set $\Psi'_{i_0j_0} = \gamma$. The rest of the elements of $\mathbf{\Psi}'$ are determined from $(\mathbf{\Psi}')^{\nu(G')}$ through the completion operation. The value of the free element $\Psi'_{i_0j_0}$ was set by perturbing the non-free element $\Psi_{i_0j_0}^{[s]}$. The other free elements of $\mathbf{\Psi}'$ and $\mathbf{\Psi}^{[s]}$ coincide.

We take $\mathbf{K}' = (\mathbf{Q}^*)^T((\mathbf{\Psi}')^T\mathbf{\Psi}')\mathbf{Q}^*$. Since $(\mathbf{\Psi}')^{\nu(G')} \in \mathsf{M}^{\nu(G')}$, we have $\mathbf{K}' \in \mathsf{P}_{G'}$. The dimensionality of the parameter space increases by 1 as we move from $(\mathbf{K}^{[s]}, G^{[s]})$ to $(\mathbf{K}', G')$. Since $(\mathbf{\Psi}')^{\nu(G^{[s]})} = (\mathbf{\Psi}^{[s]})^{\nu(G^{[s]})}$, the Jacobian of the transformation from $((\mathbf{\Psi}^{[s]})^{\nu(G^{[s]})}, \gamma)$ to $(\mathbf{\Psi}')^{\nu(G')}$ is equal to 1. Moreover, $\mathbf{\Psi}^{[s]}$ and $\mathbf{\Psi}'$ have the same diagonal elements, hence $\det\mathbf{K}^{[s]} = \prod_{i=1}^p (Q_{ii}^*\Psi_{ii}^{[s]})^2 = \det\mathbf{K}'$. The Markov chain moves to $(\mathbf{K}', G')$ with probability $\min\{R_g^+, 1\}$ where R_g^+ is given by Green (1995)

$$R_g^+ = \sigma_g\sqrt{2\pi}\,Q_{i_0i_0}^*Q_{j_0j_0}^*\Psi_{i_0i_0}^{[s]}\frac{I_{G^{[s]}}(\delta_0,\mathbf{D}_0)}{I_{G'}(\delta_0,\mathbf{D}_0)}\frac{\Pr(G')}{\Pr(G^{[s]})}\frac{|\mathsf{nbd}_p^+(G^{[s]})|}{|\mathsf{nbd}_p^-(G')|}$$
$$\times\exp\left\{-\frac{1}{2}\left[\langle\mathbf{K}'-\mathbf{K}^{[s]},\mathbf{U}+\mathbf{D}_0\rangle - \frac{(\Psi'_{i_0j_0}-\Psi_{i_0j_0}^{[s]})^2}{\sigma_g^2}\right]\right\}. \quad \text{(A.2)}$$

Otherwise the chain stays at $(\mathbf{K}^{[s]}, G^{[s]})$.

Next we assume that the candidate G' is obtained by deleting the edge (i_0, j_0) from $G^{[s]}$. We have $d_{i_0}^{G'} = d_{i_0}^{G^{[s]}}$, $d_{j_0}^{G'} = d_{j_0}^{G^{[s]}} - 1$, $v_{i_0}^{G'} = v_{i_0}^{G^{[s]}} - 1$, $v_{j_0}^{G'} = v_{j_0}^{G^{[s]}}$, and $\nu(G') = \nu(G^{[s]}) \setminus \{(i_0, j_0)\}$. We define an upper triangular matrix $\mathbf{\Psi}'$ such that $\Psi'_{ij} = \Psi_{ij}^{[s]}$ for $(i,j) \in \nu(G')$. The rest of the elements of $\mathbf{\Psi}'$ are determined through completion. The free element $\Psi_{i_0j_0}^{[s]}$ becomes non-free in $\mathbf{\Psi}'$, hence the parameter space decreases by 1 as we move from $(\mathbf{\Psi}^{[s]})^{\nu(G^{[s]})}$ to $(\mathbf{\Psi}')^{\nu(G')} \in \mathsf{M}^{\nu(G')}$. As before, we take $\mathbf{K}' = (\mathbf{Q}^*)^T((\mathbf{\Psi}')^T\mathbf{\Psi}')\mathbf{Q}^*$. The acceptance probability of the transition from $(\mathbf{K}^{[s]}, G^{[s]})$ to $(\mathbf{K}', G')$ is $\min\{R_g^-, 1\}$ where

$$R_g^- = (\sigma_g\sqrt{2\pi}\,Q_{i_0i_0}^*Q_{j_0j_0}^*\Psi_{i_0i_0}^{[s]})^{-1}$$
$$\times\frac{I_{G^{[s]}}(\delta_0,\mathbf{D}_0)}{I_{G'}(\delta_0,\mathbf{D}_0)}\frac{\Pr(G')}{\Pr(G^{[s]})}\frac{|\mathsf{nbd}_p^-(G^{[s]})|}{|\mathsf{nbd}_p^+(G')|}$$
$$\times\exp\left\{-\frac{1}{2}\left[\langle\mathbf{K}'-\mathbf{K}^{[s]},\mathbf{U}+\mathbf{D}_0\rangle + \frac{(\Psi'_{i_0j_0}-\Psi_{i_0j_0}^{[s]})^2}{\sigma_g^2}\right]\right\}. \quad \text{(A.3)}$$

We denote by $(\mathbf{K}^{[s+1/2]}, G^{[s+1]})$, $\mathbf{K}^{[s+1/2]} \in G^{[s+1]}$, the state of the chain at the end of this step.

Step 2: Resample the precision matrix. Given the updated graph $G^{[s+1]}$, we update the precision matrix $\mathbf{K}^{[s+1/2]} = (\mathbf{Q}^*)^T \times (\mathbf{\Psi}^{s+1/2})^T\mathbf{\Psi}^{s+1/2}\mathbf{Q}^*$ by sequentially perturbing the free elements $(\mathbf{\Psi}^{[s+1/2]})^{\nu(G^{[s+1]})}$. For each such element, we perform one iteration of the Metropolis–Hastings algorithm from Section 2.2 with $\delta = n + \delta_0$, $\mathbf{D} = \mathbf{U} + \mathbf{D}_0$, and $\mathbf{Q} = \mathbf{Q}^*$. The standard deviation of the normal proposals is σ_m. We denote by $\mathbf{K}^{[s+1]} \in \mathsf{P}_{G^{[s+1]}}$ the precision matrix obtained after all the updates have been performed.

APPENDIX B: DETAILS OF THE REVERSIBLE JUMP MCMC SAMPLER FOR MATRIX-VARIATE GGMS

Our sampling scheme from the joint posterior distribution (15) is composed of the following five steps that explain the transition of the Markov chain from its current state $(\mathbf{K}_R^{[s]}, \mathbf{K}_C^{[s]}, G_R^{[s]}, z^{[s]})$ to its next state $(\mathbf{K}_R^{[s+1]}, \mathbf{K}_C^{[s+1]}, G_R^{[s+1]}, z^{[s+1]})$. We use four strictly positive precision parameters $\sigma_{m,R}$, $\sigma_{m,C}$, $\sigma_{g,R}$, and $\sigma_{g,C}$.

Step 1: Resample the row graph. We denote $n_R^* = np_C$ and $\mathbf{U}_R^* = \sum_{j=1}^n \mathbf{x}^{(j)}\mathbf{K}_C^{[s]}(\mathbf{x}^{(j)})^T$. We generate a random permutation $\upsilon_R \in \Upsilon_{p_R}$ of

the row indices V_{p_R} and reorder the row and columns of the matrix $\mathbf{U}_R^* + \mathbf{D}_R$ according to υ_R. We determine the Cholesky decomposition $(\mathbf{U}_R^* + \mathbf{D}_R)^{-1} = (\mathbf{Q}_R^*)^T \mathbf{Q}_R^*$. We proceed as described in Step 1 of Appendix A. Given the notations from Appendix A, we take $p = p_R$, $n = n_R^*$, $\mathbf{U} = \mathbf{U}_R^*$, $\delta_0 = \delta_R$, $\mathbf{D}_0 = \mathbf{D}_R$, and $\sigma_g = \sigma_{g,R}$. We denote the updated row precision matrix and graph by $(\mathbf{K}_R^{[s+1/2]}, G_R^{[s+1]})$, $\mathbf{K}_R^{[s+1/2]} \in \mathsf{P}_{G_R^{[s+1]}}$.

Step 2: Resample the row precision matrix. We denote $n_R^* = np_C$ and $\mathbf{U}_R^* = \sum_{j=1}^n \mathbf{x}^{(j)} \mathbf{K}_C^{[s]} (\mathbf{x}^{(j)})^T$. We determine the Cholesky decomposition $(\mathbf{U}_R^* + \mathbf{D}_R)^{-1} = (\mathbf{Q}_R^*)^T \mathbf{Q}_R^*$ after permuting the row and columns of $\mathbf{U}_R^* + \mathbf{D}_R$ according to a random ordering in Υ_{p_R}. The conditional distribution of $\mathbf{K}_R \in \mathsf{P}_{G_R^{[s+1]}}$ is G-Wishart $\mathsf{Wis}_{G_R^{[s+1]}}(n_R^* + \delta_R, \mathbf{U}_R^* + \mathbf{D}_R)$. We make the transition from $\mathbf{K}_R^{[s+1/2]}$ to $\mathbf{K}_R^{[s+1]} \in \mathsf{P}_{G_R^{[s+1]}}$ using Metropolis–Hastings updates described in Section 2.2. Given the notations we used in that section, we take $p = p_R$, $\delta = n_R^* + \delta_R$, $\mathbf{D} = \mathbf{U}_R^* + \mathbf{D}_R$, $\mathbf{Q} = \mathbf{Q}_R^*$, and $\sigma_m = \sigma_{m,R}$.

Step 3: Resample the column graph. We denote $n_C^* = np_R$ and $\mathbf{U}_C^* = \sum_{j=1}^n (\mathbf{x}^{(j)})^T \mathbf{K}_R^{[s+1]} \mathbf{x}^{(j)}$. We sample a candidate column graph $G_C' \in \mathsf{nbd}_{p_C}(G_C^{[s]})$ from the proposal

$$
\begin{aligned}
&q\bigl(G_C' \mid G_C^{[s]}, z^{[s]}\bigr) \\
&\quad = \frac{1}{2} \frac{(z^{[s]})^{|\nu(G_C')|}}{\sum_{G_C'' \in \mathsf{nbd}_{p_C}^+(G_C^{[s]})} (z^{[s]})^{|\nu(G_C'')|}} \mathbf{1}_{\{G_C' \in \mathsf{nbd}_{p_C}^+(G_C^{[s]})\}} \\
&\qquad + \frac{1}{2} \frac{(z^{[s]})^{|\nu(G_C')|}}{\sum_{G_C'' \in \mathsf{nbd}_{p_C}^-(G_C^{[s]})} (z^{[s]})^{|\nu(G_C'')|}} \mathbf{1}_{\{G_C' \in \mathsf{nbd}_{p_C}^-(G_C^{[s]})\}}, \qquad \text{(B.1)}
\end{aligned}
$$

where 1_A is equal to 1 if A is true and is 0 otherwise. The proposal (B.1) gives an equal probability that the candidate graph is obtained by adding or deleting an edge from the current graph.

We assume that G_C' is obtained by adding the edge (i_0, j_0) to $G_C^{[s]}$. We generate a random permutation $\upsilon_C \in \Upsilon_{p_C}^{(1,1)}$ of the row indices V_{p_C} and reorder the row and columns of the matrix $\mathbf{U}_C^* + z^{[s]} \mathbf{D}_C$ according to υ_C. The permutation υ_C is such that $\upsilon_C(1) = 1$, hence the $(1,1)$ element of $\mathbf{K}_C^{[s]}$ remains in the same position. We determine the Cholesky decomposition $(\mathbf{U}_C^* + z^{[s]} \mathbf{D}_C)^{-1} = (\mathbf{Q}_C^*)^T \mathbf{Q}_C^*$ of $(\mathbf{U}_C^* + z^{[s]} \mathbf{D}_C)^{-1}$. We consider the decomposition of the column precision matrix $\mathbf{K}_C^{[s]} = (\mathbf{Q}_C^*)^T (\mathbf{\Psi}_C^{[s]})^T \mathbf{\Psi}_C^{[s]} \mathbf{Q}_C^*$ with $(\mathbf{\Psi}_C^{[s]})^{\nu(G_C^{[s]})} \in \mathsf{M}^{\nu(G_C^{[s]})}$. We define an upper triangular matrix $\mathbf{\Psi}_C'$ such that $(\mathbf{\Psi}_C')_{ij} = (\mathbf{\Psi}_C^{[s]})_{ij}$ for $(i,j) \in \nu(G_C^{[s]})$. We sample $\gamma \sim \mathsf{N}((\mathbf{\Psi}_C^{[s]})_{i_0 j_0}, \sigma_{g,C}^2)$ and set $(\mathbf{\Psi}_C')_{i_0 j_0} = \gamma$. The rest of the elements of $\mathbf{\Psi}_C'$ are determined from $(\mathbf{\Psi}_C')^{\nu(G_C')}$ through the completion operation (Atay-Kayis and Massam 2005, lemma 2). We consider the candidate column precision matrix

$$
\mathbf{K}_C' = (\mathbf{Q}_C^*)^T (\mathbf{\Psi}_C')^T \mathbf{\Psi}_C' \mathbf{Q}_C^*. \qquad \text{(B.2)}
$$

We know that $\mathbf{K}_C^{[s]} \in \mathsf{P}_{G_C^{[s]}}$ must satisfy $(\mathbf{K}_C^{[s]})_{11} = 1$. The last equality implies $(\mathbf{\Psi}_C^{[s]})_{11} = 1/(\mathbf{Q}_C^*)_{11}$, hence $(\mathbf{\Psi}_C')_{11} = 1/(\mathbf{Q}_C^*)_{11}$. Therefore we have $\mathbf{K}_C' \in \mathsf{P}_{G_C'}$ and $(\mathbf{K}_C')_{11} = 1$.

We make the transition from $(\mathbf{K}_C^{[s]}, G_C^{[s]})$ to $(\mathbf{K}_C', G_C')$ with probability $\min\{R_C^+, 1\}$ where

$$
\begin{aligned}
R_C^+ &= \sigma_{g,C} \sqrt{2\pi} (\mathbf{Q}_C^*)_{i_0 i_0} (\mathbf{Q}_C^*)_{j_0 j_0} \bigl(\mathbf{\Psi}_C^{[s]}\bigr)_{i_0 i_0} \\
&\quad \times \frac{I_{G_C^{[s]}}(\delta_C, \mathbf{D}_C)}{I_{G_C'}(\delta_C, \mathbf{D}_C)} \frac{\mathsf{Pr}(G_C')}{\mathsf{Pr}(G_C^{[s]})} \frac{\sum_{G_C'' \in \mathsf{nbd}_{p_C}^+(G_C^{[s]})} (z^{[s]})^{|\nu(G_C'')|}}{\sum_{G_C'' \in \mathsf{nbd}_{p_C}^-(G_C')} (z^{[s]})^{|\nu(G_C'')|}} \\
&\quad \times \exp\left\{ -\frac{1}{2} \left[\langle \mathbf{K}_C' - \mathbf{K}_C^{[s]}, \mathbf{U}_C^* + z^{[s]} \mathbf{D}_C \rangle \right.\right. \\
&\qquad \left.\left. - \frac{((\mathbf{\Psi}_C')_{i_0 j_0} - (\mathbf{\Psi}_C^{[s]})_{i_0 j_0})^2}{\sigma_{g,C}^2} \right] \right\}. \qquad \text{(B.3)}
\end{aligned}
$$

Next we assume that G_C' is obtained by deleting the edge (i_0, j_0) from $G_C^{[s]}$. We define an upper triangular matrix $\mathbf{\Psi}_C'$ such that $(\mathbf{\Psi}_C')_{ij} = (\mathbf{\Psi}_C^{[s]})_{ij}$ for $(i,j) \in \nu(G_C')$. The candidate $\mathbf{K}_C'$ is obtained from $\mathbf{\Psi}_C'$ as in Equation (B.2). We make the transition from $(\mathbf{K}_C^{[s]}, G_C^{[s]})$ to $(\mathbf{K}_C', G_C')$ with probability $\min\{R_C^-, 1\}$ where

$$
\begin{aligned}
R_C^- &= \bigl\{ \sigma_{g,C} \sqrt{2\pi} (\mathbf{Q}_C^*)_{i_0 i_0} (\mathbf{Q}_C^*)_{j_0 j_0} \bigl(\mathbf{\Psi}_C^{[s]}\bigr)_{i_0 i_0} \bigr\}^{-1} \\
&\quad \times \frac{I_{G_C^{[s]}}(\delta_C, \mathbf{D}_C)}{I_{G_C'}(\delta_C, \mathbf{D}_C)} \frac{\mathsf{Pr}(G_C')}{\mathsf{Pr}(G_C^{[s]})} \frac{\sum_{G_C'' \in \mathsf{nbd}_{p_C}^-(G_C^{[s]})} (z^{[s]})^{|\nu(G_C'')|}}{\sum_{G_C'' \in \mathsf{nbd}_{p_C}^+(G_C')} (z^{[s]})^{|\nu(G_C'')|}} \\
&\quad \times \exp\left\{ -\frac{1}{2} \left[\langle \mathbf{K}_C' - \mathbf{K}_C^{[s]}, \mathbf{U}_C^* + z^{[s]} \mathbf{D}_C \rangle \right.\right. \\
&\qquad \left.\left. + \frac{((\mathbf{\Psi}_C')_{i_0 j_0} - (\mathbf{\Psi}_C^{[s]})_{i_0 j_0})^2}{\sigma_{g,C}^2} \right] \right\}. \qquad \text{(B.4)}
\end{aligned}
$$

We denote the updated column precision matrix and graph by $(\mathbf{K}_C^{[s+1/2]}, G_C^{[s+1]})$.

Step 4: Resample the column precision matrix. We denote $n_C^* = np_R$ and $\mathbf{U}_C^* = \sum_{j=1}^n (\mathbf{x}^{(j)})^T \mathbf{K}_R^{[s+1]} \mathbf{x}^{(j)}$. We determine the Cholesky decomposition $(\mathbf{U}_C^* + z^{[s]} \mathbf{D}_C)^{-1} = (\mathbf{Q}_C^*)^T \mathbf{Q}_C^*$ after permuting the row and columns of $\mathbf{U}_C^* + z^{[s]} \mathbf{D}_C$ according to a random ordering in $\Upsilon_{p_C}^{(1,1)}$. The conditional distribution of $\mathbf{K}_C \in \mathsf{P}_{G_C^{[s+1]}}$ with $(\mathbf{K}_C)_{11} = 1$ is G-Wishart $\mathsf{Wis}_{G_C^{[s+1]}}(n_C^* + \delta_C, \mathbf{U}_C^* + z^{[s]} \mathbf{D}_C)$. We make the transition from $\mathbf{K}_C^{[s+1/2]}$ to $\mathbf{K}_C^{[s+1]} \in \mathsf{P}_{G_C^{[s+1]}}$ using the Metropolis–Hastings updates from Section 2.2. Given the notations we used in that section, we take $p = p_C$, $\delta = n_C^* + \delta_C$, $\mathbf{D} = \mathbf{U}_C^* + z^{[s]} \mathbf{D}_C$, $\mathbf{Q} = \mathbf{Q}_C^*$, and $\sigma_m = \sigma_{m,C}$. The constraint $(\mathbf{K}_C)_{11} = 1$ is accommodated as described at the end of Section 2.2.

Step 5: Resample the auxiliary variable. The conditional distribution of $z > 0$ is

$$
\text{Gamma}\left(\frac{p_C(\delta_C - 2)}{2} + \bigl|\nu\bigl(G_C^{[s+1]}\bigr)\bigr|, \frac{1}{2} \operatorname{tr}\bigl(\mathbf{K}_C^{[s+1]} \mathbf{D}_C\bigr) \right). \qquad \text{(B.5)}
$$

Here $\text{Gamma}(\alpha, \beta)$ has density $f(x \mid \alpha, \beta) \propto \beta^\alpha x^{\alpha-1} \exp(-\beta x)$. We sample $z^{[s+1]}$ from (B.5).

SUPPLEMENTARY MATERIALS

Code: Computer code to run all routines discussed in this article, as well as the U.S. cancer mortality data. (DLR-archive.tar.gz.zip, GNU zipped tar file)

Results: Theoretical results related to conditional independence relationships in sparse latent Gaussian processes; the description of the MCMC algorithm used to fit the sparse multivariate spatial Poisson count model from Section 6.2; tables describing the U.S. cancer mortality data, the graphical model constructed from the neighborhood structure of the United States, as well as fitted values and 95% credible intervals for the models considered in the article in Section 6.2. (DLR-supplement.pdf, PDF file)

[Received July 2010. Revised May 2011.]

REFERENCES

Armstrong, H. (2005), "Bayesian Estimation of Decomposable Gaussian Graphical Models," Ph.D. thesis, The University of New South Wales. [1422]

Armstrong, H., Carter, C. K., Wong, K. F., and Kohn, R. (2009), "Bayesian Covariance Matrix Estimation Using a Mixture of Decomposable Graphical Models," *Statistics and Computing*, 19, 303–316. [1418,1422,1423]

Asci, C., and Piccioni, M. (2007), "Functionally Compatible Local Characteristics for the Local Specification of Priors in Graphical Models," *Scandinavian Journal of Statistics*, 34, 829–840. [1419]

Atay-Kayis, A., and Massam, H. (2005), "A Monte Carlo Method for Computing the Marginal Likelihood in Nondecomposable Gaussian Graphical Models," *Biometrika*, 92, 317–335. [1418-1420,1422,1431]

Banerjee, O., El Ghaoui, L., and D'Aspremont, A. (2008), "Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data," *Journal of Machine Learning Research*, 9, 485–516. [1418]

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2004), *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: Chapman & Hall/CRC. [1421]

Besag, J. (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems" (with discussion), *Journal of the Royal Statistical Society, Ser. B*, 36, 192–236. [1418,1424]

——— (1989), "A Candidate's Formula: A Curious Result in Bayesian Prediction," *Biometrika*, 76, 183. [1424]

Besag, J., and Kooperberg, C. (1995), "On Conditional and Intrinsic Autoregressions," *Biometrika*, 82, 733–746. [1418,1424]

Brook, D. (1964), "On the Distinction Between the Conditional Probability and the Joint Probability Approaches in the Specification of the Nearest Neighbour Systems," *Biometrika*, 51, 481–489. [1424]

Brooks, S. P., Giudici, P., and Roberts, G. O. (2003), "Efficient Construction of Reversible Jump Markov Chain Monte Carlo Proposals Distributions," *Journal of the Royal Statistical Society, Ser. B*, 65, 3–55. [1422]

Carvalho, C. M., and Scott, J. G. (2009), "Objective Bayesian Model Selection in Gaussian Graphical Models," *Biometrika*, 96, 1–16. [1418,1422,1423]

Carvalho, C. M., Massam, H., and West, M. (2007), "Simulation of Hyper-Inverse Wishart Distributions in Graphical Models," *Biometrika*, 94, 647–659. [1419]

Chib, S. (1990), "Marginal Likelihood From the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313–1321. [1424]

Cressie, N. A. C. (1973), *Statistics for Spatial Data*, New York: Wiley. [1424]

Cressie, N. A. C., and Chan, N. H. (1989), "Spatial Modeling of Regional Variables," *Journal of American Statistical Association*, 84, 393–401. [1425]

Czado, C., Gneiting, T., and Held, L. (2009), "Predictive Model Assessment for Count Data," *Biometrics*, 65, 1254–1261. [1428]

Dawid, A. P., and Lauritzen, S. L. (1993), "Hyper Markov Laws in the Statistical Analysis of Decomposable Graphical Models," *The Annals of Statistics*, 21, 1272–1317. [1422]

Dellaportas, P., Giudici, P., and Roberts, G. (2003), "Bayesian Inference for Nondecomposable Graphical Gaussian Models," *Sankhyā: The Indian Journal of Statistics*, 65, 43–55. [1422]

Dempster, A. P. (1972), "Covariance Selection," *Biometrics*, 28, 157–175. [1418,1419,1422]

Diaconnis, P., and Ylvisaker, D. (1979), "Conjugate Priors for Exponential Families," *The Annals of Statistics*, 7, 269–281. [1419]

Dobra, A., Hans, C., Jones, B., Nevins, J. R., Yao, G., and West, M. (2004), "Sparse Graphical Models for Exploring Gene Expression Data," *Journal of Multivariate Analysis*, 90, 196–212. [1418,1423]

Drton, M., and Perlman, M. D. (2008), "A SINful Approach to Gaussian Graphical Model Selection," *Journal of Statistical Planning and Inference*, 138, 1179–1200. [1418]

Elliott, P., Wakefield, J., Best, N., and Briggs, D. (2001), *Spatial Epidemiology: Methods and Applications*, New York: Oxford University Press. [1421]

Friedman, J., Hastie, T., and Tibshirani, R. (2008), "Sparse Inverse Covariance Estimation With the Graphical Lasso," *Biostatistics*, 9, 432–441. [1418]

Gelfand, A. E., and Ghosh, S. K. (1998), "Model Choice: A Minimum Posterior Predictive Loss Approach," *Biometrika*, 85, 1–11. [1428]

Gelfand, A. E., and Vounatsou, P. (2003), "Proper Multivariate Conditional Autoregressive Models for Spatial Data Analysis," *Biostatistics*, 4, 11–25. [1424,1425,1428]

Gelfand, A. E., Diggle, P. J., Fuentes, M., and Guttorp, P. (2010), *Handbook of Spatial Statistics*, Boca Raton, FL: Chapman & Hall/CRC. [1421]

Ghosh, K., and Tiwari, R. C. (2007), "Prediction of U.S. Cancer Mortality Counts Using Semiparametric Bayesian Techniques," *Journal of the American Statistical Association*, 102, 7–15. [1427]

Ghosh, K., Ghosh, P., and Tiwari, R. C. (2008), Comment on "The Nested Dirichlet Process," by A. Rodriguez, D. B. Dunson, and A. E. Gelfand, *Journal of the American Statistical Association*, 103 (483), 1147–1149. [1427]

Ghosh, K., Tiwari, R. C., Feuer, E. J., Cronin, K., and Jemal, A. (2007), "Predicting U.S. Cancer Mortality Counts Using State Space Models," in *Computational Methods in Biomedical Research*, eds. R. Khattree and D. N. Naik, Boca Raton, FL: Chapman & Hall/CRC, pp. 131–151. [1427]

Giudici, P., and Green, P. J. (1999), "Decomposable Graphical Gaussian Model Determination," *Biometrika*, 86, 785–801. [1419,1422]

Gneiting, T., and Raftery, A. E. (2007), "Strictly Proper Scoring Rules, Prediction and Estimation," *Journal of the American Statistical Association*, 102, 359–378. [1428]

Green, P. J. (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination," *Biometrika*, 82, 711–732. [1418,1422,1423,1430]

Gupta, A. K., and Nagar, D. K. (2000), *Matrix Variate Distributions. Monographs and Surveys in Pure and Applied Mathematics*, Vol. 104, London: Chapman & Hall/CRC. [1423]

Jones, B., Carvalho, C., Dobra, A., Hans, C., Carter, C., and West, M. (2005), "Experiments in Stochastic Computation for High-Dimensional Graphical Models," *Statistical Science*, 20, 388–400. [1418,1422,1423]

Lauritzen, S. L. (1996), *Graphical Models*, New York: Oxford University Press. [1418]

Lawson, A. B. (2009), *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*, Boca Raton, FL: Chapman & Hall/CRC. [1421]

Lenkoski, A., and Dobra, A. (2011), "Computational Aspects Related to Inference in Gaussian Graphical Models With the G-Wishart Prior," *Journal of Computational and Graphical Statistics*, 20, 140–157. [1418,1422]

Letac, G., and Massam, H. (2007), "Wishart Distributions for Decomposable Graphs," *The Annals of Statistics*, 35, 1278–1323. [1418,1419]

Madigan, D., and York, J. (1995), "Bayesian Graphical Models for Discrete Data," *International Statistical Review*, 63, 215–232. [1430]

Mardia, K. V. (1988), "Multi-Dimensional Multivariate Gaussian Markov Random Fields With Application to Image Processing," *Journal of Multivariate Analysis*, 24, 265–284. [1424,1425]

Meinshausen, N., and Bühlmann, P. (2006), "High Dimensional Graphs and Variable Selection With the Lasso," *The Annals of Statistics*, 34, 1436–1462. [1418]

Mengersen, K. L., and Tweedie, R. L. (1996), "Rates of Convergence of the Hastings and Metropolis Algorithms," *The Annals of Statistics*, 24, 101–121. [1420]

Mitsakakis, N., Massam, H., and Escobar, M. D. (2011), "A Metropolis–Hastings Based Method for Sampling From the G-Wishart Distribution in Gaussian Graphical Models," *Electronic Journal of Statistics*, 5, 18–30. [1418-1421]

Moghaddam, B., Marlin, B., Khan, E., and Murphy, K. (2009), "Accelerating Bayesian Structural Inference for Non-Decomposable Gaussian Graphical Models," in *Advances in Neural Information Processing Systems*, Vol. 22, eds. Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, and A. Culotta, San Mateo, CA: Morgan Kaufmann, pp. 1285–1293. [1422]

Muirhead, R. J. (2005), *Aspects of Multivariate Statistical Theory*, New York: Wiley. [1419]

Piccioni, M. (2000), "Independence Structure of Natural Conjugate Densities to Exponential Families and the Gibbs Sampler," *Scandinavian Journal of Statistics*, 27, 111–127. [1419]

Ravikumar, P., Wainwright, M. J., and Lafferty, J. D. (2010), "High-Dimensional Ising Model Selection Using l_1-Regularized Logistic Regression," *The Annals of Statistics*, 38, 1287–1319. [1418]

Robert, C., and Casella, G. (2004), *Monte Carlo Statistical Methods* (2nd ed.), New York: Springer-Verlag. [1420]

Roverato, A. (2002), "Hyper Inverse Wishart Distribution for Non-Decomposable Graphs and Its Application to Bayesian Inference for Gaussian Graphical Models," *Scandinavian Journal of Statistics*, 29, 391–411. [1418,1419,1422]

Rue, H., and Held, L. (2005), *Gaussian Markov Random Fields: Theory and Applications*, Boca Raton, FL: Chapman & Hall/CRC. [1421]

Scott, J. G., and Berger, J. O. (2006), "An Exploration of Aspects of Bayesian Multiple Testing," *Journal of Statistical Planning and Inference*, 136, 2144–2162. [1423]

Scott, J. G., and Carvalho, C. M. (2008), "Feature-Inclusion Stochastic Search for Gaussian Graphical Models," *Journal of Computational and Graphical Statistics*, 17, 790–808. [1422]

Speed, T. P., and Kiiveri, H. T. (1986), "Gaussian Markov Distributions Over Finite Graphs," *The Annals of Statistics*, 14, 138–150. [1422]

Sun, D., Tsutakawa, R. K., Kim, H., and He, Z. (2000), "Bayesian Analysis of Mortality Rates With Disease Maps," *Statistics and Medicine*, 19, 2015–2035. [1424]

Tarjan, R. E. (1985), "Decomposition by Clique Separators," *Discrete Mathematics*, 55, 221–232. [1420]

Tiwari, R. C., Ghosh, K., Jemal, A., Hachey, M., Ward, E., Thun, M. J., and Feuer, E. J. (2004), "A New Method for Predicting U.S., and State-level Cancer Mortality Counts for the Current Calendar Year," *CA: A Cancer Journal for Clinicians*, 54 (1), 30–40. [1427]

Wang, H., and Carvalho, C. M. (2010), "Simulation of Hyper-Inverse Wishart Distributions for Non-Decomposable Graphs," *Electronic Journal of Statistics*, 4, 1470–1475. [1418-1422]

Wang, H., and West, M. (2009), "Bayesian Analysis of Matrix Normal Graphical Models," *Biometrika*, 96, 821–834. [1418,1419,1423-1427]

Wong, F., Carter, C. K., and Kohn, R. (2003), "Efficient Estimation of Covariance Selection Models," *Biometrika*, 90, 809–830. [1418,1422]

Yuan, M., and Lin, Y. (2007), "Model Selection and Estimation in the Gaussian Graphical Model," *Biometrika*, 94, 19–35. [1418]