Statistical Theory and Related Fields

ISSN: 2475-4269 (Print) 2475-4277 (Online) Journal homepage: https://www.tandfonline.com/loi/tstf20

Graph-based multivariate conditional autoregressive models

Ye Liang

To cite this article: Ye Liang (2019) Graph-based multivariate conditional autoregressive models, Statistical Theory and Related Fields, 3:2, 158-169, DOI: 10.1080/24754269.2019.1666242

To link to this article: https://doi.org/10.1080/24754269.2019.1666242

Published online: 16 Sep 2019.

STATISTICAL THEORY AND RELATED FIELDS
2019, VOL. 3, NO. 2, 158–169
https://doi.org/10.1080/24754269.2019.1666242

Taylor & Francis
Taylor & Francis Group

Graph-based multivariate conditional autoregressive models

Ye Liang

Department of Statistics, Oklahoma State University, Stillwater, Oklahoma, USA

ABSTRACT

The conditional autoregressive model is a routinely used statistical model for areal data that arise from, for instances, epidemiological, socio-economic or ecological studies. Various multivariate conditional autoregressive models have also been extensively studied in the literature and it has been shown that extending from the univariate case to the multivariate case is not trivial. The difficulties lie in many aspects, including validity, interpretability, flexibility and computational feasibility of the model. In this paper, we approach the multivariate modelling from an element-based perspective instead of the traditional vector-based perspective. We focus on the joint adjacency structure of elements and discuss graphical structures for both the spatial and non-spatial domains. We assume that the graph for the spatial domain is generally known and fixed while the graph for the non-spatial domain can be unknown and random. We propose a very general specification for the multivariate conditional modelling and then focus on three special cases, which are linked to well-known models in the literature. Bayesian inference for parameter learning and graph learning is provided for the focused cases, and finally, an example with public health data is illustrated.

ARTICLE HISTORY
Received 25 November 2018
Revised 6 September 2019
Accepted 7 September 2019

KEYWORDS
Areal data; disease mapping; graphical model; G-wishart distribution; Markov random field; reversible jump

1. Introduction

Areal data, sometimes called lattice data, are usually represented by an undirected graph where each vertex represents an areal unit and each edge represents a neighbouring relationship. A finite set of random variables on an undirected graph, where each vertex is a random variable, is called a Markov random field if it has the Markov property. Hence, the Markov random field models are often used for the areal data. The univariate conditional autoregressive (CAR) model, originated from Besag (1974), is a Gaussian Markov random field model, for which the joint distribution is multivariate Gaussian. Let $\boldsymbol{u} = (u_1, \ldots, u_I)^T$ be a vector of random variables on I areal units (i.e., I vertices). The zero-centred conditional autoregressive model specifies full conditional Gaussian distributions

$$u_i \mid u_{-i} \sim \mathrm{N}\left(\sum_{i' \neq i} b_{ii'} u_{i'}, \tau_i^2\right), \quad i = 1, \ldots, I,$$

where u_{-i} is the collection of $u_{i'}$ for $i' \neq i$. The resulting joint distribution, derived using Brook's lemma, has a density function as follows,

$$f(\boldsymbol{u} \mid \boldsymbol{T}_{\mathrm{CAR}}, \boldsymbol{B}_{\mathrm{CAR}}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{u}^T \boldsymbol{T}_{\mathrm{CAR}}^{-1}(\boldsymbol{I} - \boldsymbol{B}_{\mathrm{CAR}})\boldsymbol{u}\right\},$$

where $\boldsymbol{I}$ is an identity matrix; $\boldsymbol{B}_{\mathrm{CAR}}$ is an $I \times I$ matrix whose off-diagonal entries are $b_{ii'}$ and diagonal entries are zeros, and $\boldsymbol{T}_{\mathrm{CAR}} = \mathrm{diag}\{\tau_1^2, \ldots, \tau_I^2\}$. The joint distribution is multivariate Gaussian if and only if $\boldsymbol{T}_{\mathrm{CAR}}^{-1}(\boldsymbol{I} - \boldsymbol{B}_{\mathrm{CAR}})$ is symmetric and positive definite. A further parameterization on $\boldsymbol{B}_{\mathrm{CAR}}$ and $\boldsymbol{T}_{\mathrm{CAR}}$ is needed to reduce the number of parameters in the model. Consider a so-called adjacency matrix $\boldsymbol{C}_{\mathrm{CAR}}$ for the undirected graph, where the ii'th entry $C_{ii'} = 1$ if unit i and unit i' are neighbours (denoted as $i \sim i'$) and $C_{ii'} = 0$ otherwise. One popular parameterization is to let $b_{ii'} = \rho C_{ii'}/C_{i+}$ and $\tau_i^2 = \sigma^2/C_{i+}$, where C_{i+} is the ith row sum of $\boldsymbol{C}_{\mathrm{CAR}}$, representing the total number of neighbours of unit i. Let $\boldsymbol{D}_{\mathrm{CAR}} = \mathrm{diag}\{C_{1+}, \ldots, C_{I+}\}$. When ρ is strictly between the smallest and largest eigenvalues of $\boldsymbol{D}_{\mathrm{CAR}}^{-1/2}\boldsymbol{C}_{\mathrm{CAR}}\boldsymbol{D}_{\mathrm{CAR}}^{-1/2}$, or sufficiently, when $|\rho| < 1$, and $\sigma^2 > 0$, the joint distribution of $\boldsymbol{u}$ is a zero-mean multivariate Gaussian distribution: $\boldsymbol{u} \sim N\{\boldsymbol{0}, \sigma^2(\boldsymbol{D}_{\mathrm{CAR}} - \rho\boldsymbol{C}_{\mathrm{CAR}})^{-1}\}$. This is called the proper conditional autoregressive model in the literature. When $\rho = 1$, it is called the intrinsic conditional autoregressive model which is an improper distribution due to the singular covariance matrix.

Turning to the multivariate case, consider J responses (e.g., multiple diseases) on I areal units. Let $\boldsymbol{U}$ be an $I \times J$ matrix-variate, where the ijth entry u_{ij} is a random variable for the ith areal unit and jth response. Each column of $\boldsymbol{U}$ is an areal vector for a single response and hence can be modelled by the univariate conditional autoregressive model. However, a multivariate model is desired for the matrix-variate $\boldsymbol{U}$ in

CONTACT Ye Liang ye.liang@okstate.edu Department of Statistics, Oklahoma State University, Stillwater, Oklahoma 74078, USA

order to simultaneously model the dependence across responses. Initially proposed by Mardia (1988), the multivariate conditional autoregressive model specifies full conditional distributions on row vectors of $\boldsymbol{U}$. Let $\boldsymbol{u}_i$ be the ith row vector of $\boldsymbol{U}$. Following Besag (1974), specify

$$\boldsymbol{u}_i \mid \boldsymbol{u}_{-i} \sim \mathrm{N}\left(\sum_{i' \neq i} \boldsymbol{B}_{ii'}\boldsymbol{u}_{i'}, \boldsymbol{\Sigma}_i\right), \quad i = 1, \ldots, I, \quad (1)$$

where $\boldsymbol{B}_{ii'}$ and $\boldsymbol{\Sigma}_i$ are $J \times J$ matrices needing a further parameterization. To make the joint distribution for $\mathrm{vec}(\boldsymbol{U}^T)$ a multivariate Gaussian, $\boldsymbol{B}_{ii'}$ and $\boldsymbol{\Sigma}_i$ must satisfy certain conditions (Mardia, 1988). Gelfand and Vounatsou (2003) showed a convenient parameterization, $\boldsymbol{B}_{ii'} = (\rho C_{ii'}/C_{i+})\boldsymbol{I}_J$ and $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}/C_{i+}$. When $|\rho| < 1$ and $\boldsymbol{\Sigma}$ is positive definite, $\mathrm{vec}(\boldsymbol{U}^T)$ has a zero-mean multivariate Gaussian distribution: $\mathrm{vec}(\boldsymbol{U}^T) \sim N\{\boldsymbol{0}, (\boldsymbol{D}_{\mathrm{CAR}} - \rho\boldsymbol{C}_{\mathrm{CAR}})^{-1} \otimes \boldsymbol{\Sigma}\}$. It is clear that this multivariate specification is a Kronecker-product formula where $(\boldsymbol{D}_{\mathrm{CAR}} - \rho\boldsymbol{C}_{\mathrm{CAR}})^{-1}$ models the covariance structure across rows of $\boldsymbol{U}$ (spatial domain) and $\boldsymbol{\Sigma}$ models the covariance structure across columns of $\boldsymbol{U}$ (response domain). From the modelling perspective, Mardia's specification has a difficulty with parameterization. It is usually difficult to have a meaningful parameterization for $\boldsymbol{B}_{ii'}$ and $\boldsymbol{\Sigma}_i$ unless one pursues a simple formulation. It is arguable that the Mardia's specification presents a conflict, where the between vector variation is specified through an inverse covariance matrix, but the within vector variation is specified through a covariance matrix. It seems more intuitive to either work with the joint covariance or the joint inverse covariance directly. Notice that most multivariate spatial models for point reference data focus on the joint covariance structure. In this paper, we focus on the joint inverse covariance structure of elements in the multivariate areal data. In particular, we consider the joint adjacency structure of the lattice based on graphical structures of both the spatial domain and the response domain. We build a framework for graph-based multivariate conditional autoregressive models and discuss parameterizations under this framework. The advantage is that this framework is very general and we demonstrate it through multiple case examples. Furthermore, we allow graph learning for multiple responses in such models, which is potentially useful for many modern applications.

We shall point out other recent work on multivariate conditional autoregressive models. Kim, Sun, and Tsutakawa (2001) and Jin, Carlin, and Banerjee (2005) proposed conditional autoregressive models for bivariate areal data. Multivariate models were considered by Gelfand and Vounatsou (2003), Jin, Banerjee, and Carlin (2007), MacNab (2011, 2016), Martinez-Beneito (2013), Martinez-Beneito, Botella-Rocamora, and Banerjee (2017) among many others. MacNab (2018) reviewed some recent developments on multivariate Gaussian Markov random field models. We will show that some of the earlier work can be reconstructed in our proposed framework and some can be extended to graphical models. The paper is organised as follows. Section 2 presents the general framework and three special parameterizations. Section 3 presents a real data example using the proposed models. Section 5 contains further discussions and remarks. Technical details are given in the Appendices.

2. Graph-based multivariate conditional autoregressive models

2.1. General framework

Instead of specifying full conditional distributions on vectors like (1), we approach this problem from an element-based perspective. Following Besag (1974), specify full conditional distributions for each element u_{ij} in the matrix-variate $\boldsymbol{U}$ as follows,

$$u_{ij} \mid u_{-\{ij\}} \sim \mathrm{N}\left(\sum_{\{i'j'\} \neq \{ij\}} b_{\{ij\},\{i'j'\}}u_{i'j'}, \tau^2_{ij}\right),$$
$$i = 1, \ldots, I \text{ and } j = 1, \ldots, J,$$

where $\{i'j'\} \neq \{ij\}$ means either $i' \neq i$ or $j' \neq j$. In fact, here we consider a lattice consisting of all elements in $\boldsymbol{U}$. Using Brook's lemma, the resulting joint distribution for $\mathrm{vec}(\boldsymbol{U})$ is

$$f(\mathrm{vec}(\boldsymbol{U}) \mid \boldsymbol{B}, \boldsymbol{T}) \propto \exp$$
$$\times \left\{-\frac{1}{2}\mathrm{vec}(\boldsymbol{U})^T\boldsymbol{T}^{-1}(\boldsymbol{I} - \boldsymbol{B})\,\mathrm{vec}(\boldsymbol{U})\right\},$$

where $\boldsymbol{I}$ is an $IJ \times IJ$ identity matrix, $\boldsymbol{T} = \mathrm{diag}\{\tau^2_{11}, \ldots, \tau^2_{I1}, \ldots, \tau^2_{1J}, \ldots, \tau^2_{IJ}\}$ and $\boldsymbol{B}$ can be expressed block-wisely,

$$\boldsymbol{B} = \begin{pmatrix} \boldsymbol{B}_{11} & \cdots & \boldsymbol{B}_{1J} \\ \vdots & \ddots & \vdots \\ \boldsymbol{B}_{J1} & \cdots & \boldsymbol{B}_{JJ} \end{pmatrix} \quad \text{where}$$

$$\boldsymbol{B}_{jj'} = \begin{pmatrix} b_{\{1j\},\{1j'\}} & \cdots & b_{\{1j\},\{Ij'\}} \\ \vdots & \ddots & \vdots \\ b_{\{Ij\},\{1j'\}} & \cdots & b_{\{Ij\},\{Ij'\}} \end{pmatrix},$$

and the diagonal elements $b_{\{ij\},\{ij\}}$ are zeros. The joint distribution for $\mathrm{vec}(\boldsymbol{U})$ is multivariate Gaussian if and only if $\boldsymbol{T}^{-1}(\boldsymbol{I} - \boldsymbol{B})$ is symmetric and positive definite. It is desired that $\boldsymbol{B}$ and $\boldsymbol{T}$ are further parameterised to reduce the number of parameters in the model. We denote this general model MCAR($\boldsymbol{B}$, $\boldsymbol{T}$) for later use.

Consider the adjacency structure of the undirected graph for the lattice of $\boldsymbol{U}$. In the univariate situation, the adjacency structure is determined by those geographical locations. Two areal units are connected by an edge

if they are neighbours geographically. However, it is not obvious which elements should be neighbours in $\boldsymbol{U}$. Consider that the J responses can be connected through an undirected graph. Let $\boldsymbol{C}^{(s)}$ be the adjacency matrix for all I areal units and $\boldsymbol{C}^{(r)}$ be the adjacency matrix for all J responses. Both the spatial graph and the response graph are then uniquely determined by $\boldsymbol{C}^{(s)}$ and $\boldsymbol{C}^{(r)}$, respectively. Let $\boldsymbol{C}$ be the joint adjacency matrix for the lattice of $\boldsymbol{U}$. A general construction of $\boldsymbol{C}$ can be made through $\boldsymbol{C}^{(s)}$ and $\boldsymbol{C}^{(r)}$,

$$\boldsymbol{C} = \boldsymbol{C}^{(r)} \otimes \boldsymbol{C}^{(s)} + \boldsymbol{C}^{(r)} \otimes \boldsymbol{I}_I + \boldsymbol{I}_J \otimes \boldsymbol{C}^{(s)}. \quad (2)$$

This construction connects u_{ij} with $u_{i'\sim i,j}$, $u_{i,j'\sim j}$ and $u_{i'\sim i,j'\sim j}$, meaning its spatial neighbour, response neighbour and interaction neighbour, respectively. One may add edges for secondary neighbours or drop edges in a specific modelling. For example, some reduced constructions would be (i) $\boldsymbol{C} = \boldsymbol{I}_J \otimes \boldsymbol{C}^{(s)}$ (independent conditional autoregressive models, no dependence between responses); (ii) $\boldsymbol{C} = \boldsymbol{C}^{(r)} \otimes \boldsymbol{I}_I$ (independent multivariate variables, no spatial dependence); and (iii) $\boldsymbol{C} = \boldsymbol{C}^{(r)} \otimes \boldsymbol{I}_I + \boldsymbol{I}_J \otimes \boldsymbol{C}^{(s)}$ (drop edges for interaction neighbours $u_{i'\sim i,j'\sim j}$).

Let $C_{\{ij\},\{i'j'\}}$ denote entries in $\boldsymbol{C}$, analogous to the block-wise notation $b_{\{ij\},\{i'j'\}}$ for $\boldsymbol{B}$. Let $d_j^{(r)}$ be the jth row sum in $\boldsymbol{C}^{(r)}$ and $d_i^{(s)}$ be the ith row sum in $\boldsymbol{C}^{(s)}$. Then the ijth row sum in $\boldsymbol{C}$ is $d_{ij} = d_j^{(r)} d_i^{(s)} + d_j^{(r)} + d_i^{(s)}$. Let $\boldsymbol{D}^{(r)} = \mathrm{diag}\{d_1^{(r)}, \ldots, d_J^{(r)}\}$, $\boldsymbol{D}^{(s)} = \mathrm{diag}\{d_1^{(s)}, \ldots, d_I^{(s)}\}$ and $\boldsymbol{D} = \mathrm{diag}\{d_{11}, \ldots, d_{I1}, \ldots, d_{1J}, \ldots, d_{IJ}\}$. With the adjacency constructions and notations, we then explore further parameterisation on $\boldsymbol{B}$ and $\boldsymbol{T}$ in the following subsections, and specifically, we discuss three specifications made from this general framework, all of which are linked to well-known models in the literature.

2.2. Model 1: nonseparable multifold specification

Kim et al. (2001) developed a twofold conditional autoregressive model for bivariate areal data ($J = 2$), using different linkage parameters for different types of neighbours. Those linkage parameters, in their work, are called smoothing and bridging parameters, representing the strength of information sharing. If we extend their specification to an arbitrary J, we can parameterise $\boldsymbol{B}$ and $\boldsymbol{T}$ in the following way (assuming $i \neq i'$ and $j \neq j'$):

$$b_{\{ij\},\{i'j\}} = \frac{\lambda_j}{d_{ij}} C_{\{ij\},\{i'j\}}, \quad b_{\{ij\},\{ij'\}} = \frac{\psi_{jj'}}{d_{ij}} \sqrt{\frac{\delta_j}{\delta_{j'}}} C_{\{ij\},\{ij'\}},$$

$$b_{\{ij\},\{i'j'\}} = \frac{\phi_{jj'}}{d_{ij}} \sqrt{\frac{\delta_j}{\delta_{j'}}} C_{\{ij\},\{i'j'\}}, \quad \tau_{ij}^2 = \frac{\delta_j}{d_{ij}},$$

where λ_j, $\psi_{jj'}$ and $\phi_{jj'}$ are linkage parameters and δ_j are variance components. Linkage parameters are for three types of neighbour: $u_{i'\sim i,j}$, $u_{i,j'\sim j}$ and $u_{i'\sim i,j'\sim j}$. Having this specification, the conditional mean of u_{ij} essentially is

$$\mathrm{E}(u_{ij} \mid u_{-\{ij\}}) = \frac{1}{d_{ij}} \left(\lambda_j \sum_{i'\sim i} u_{i'j} + \psi_{jj'} \sum_{j'\sim j} \sqrt{\frac{\delta_j}{\delta_{j'}}} u_{ij'} + \phi_{jj'} \sum_{i'\sim i} \sum_{j'\sim j} \sqrt{\frac{\delta_j}{\delta_{j'}}} u_{i'j'} \right),$$

which is a weighted average of all its neighbours in $\boldsymbol{C}$. This specification generalises (Kim et al., 2001)' twofold model and hence could be called a multifold specification. It can be shown that, for this parameterisation, the joint precision matrix is

$$\begin{aligned} \boldsymbol{T}^{-1}(\boldsymbol{I} - \boldsymbol{B}) = (\boldsymbol{\Delta}^{-1/2} \otimes \boldsymbol{I}_I) \Big\{ & \boldsymbol{D} - \boldsymbol{\Lambda} \otimes \boldsymbol{C}^{(s)} \\ & - (\boldsymbol{\Psi} \circ \boldsymbol{C}^{(r)}) \otimes \boldsymbol{I}_I - (\boldsymbol{\Phi} \circ \boldsymbol{C}^{(r)}) \otimes \boldsymbol{C}^{(s)} \Big\} \\ & \times (\boldsymbol{\Delta}^{-1/2} \otimes \boldsymbol{I}_I), \end{aligned} \quad (3)$$

where $\boldsymbol{\Delta} = \mathrm{diag}\{\delta_1, \ldots, \delta_J\}$, $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_J\}$, $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are $J \times J$ symmetric matrices with entries $\psi_{jj'}$ and $\phi_{jj'}$, respectively, and the operator $\circ$ means an element-wise product. A derivation of (3) is given in Appendix 1. Note that only nonzero entries of $\boldsymbol{\Psi}$ and $\boldsymbol{\Phi}$ are parameters in the model, the number of which depends on $\boldsymbol{C}^{(r)}$.

In order to make (3) positive definite, constraints on λ_j, $\psi_{jj'}$ and $\phi_{jj'}$ are needed, assuming that $\delta_j > 0$. In general, it is difficult to find a sufficient and necessary condition for the positive definiteness of (3). Kim et al. (2001)'s solution to this problem was a sufficient condition: $\max\{|\lambda_j|, |\psi_{jj'}|, |\phi_{jj'}|; \forall j, j'\} < 1$, under which the matrix (3) is diagonally dominant and hence is positive definite. Though their proof was under $J = 2$, it is true for any J by the same arguments. The advantage of this condition is that it is simple and implementable. However, this is not a necessary condition meaning that it is impossible to reach all possible positive definite structures for the model under such a condition. In a Bayesian model, priors on parameters λ_j, $\psi_{jj'}$ and $\phi_{jj'}$ can be chosen based on their actual constraints. In our case, a uniform prior $\mathrm{Unif}(-1, 1)$ is adequate for these linkage parameters. Priors on the variance components δ_j can be weakly-informative inverse gamma priors $\mathrm{IG}(a_j, b_j)$. Inference and computation under this model are given in Appendix 2.

2.3. Model 2: separable specification with homogeneous spatial smoothing

Gelfand and Vounatsou (2003)'s Kronecker-product model is a convenient parameterisation of Mardia (1988)'s model. In our framework, this specification can be obtained and extended by having the following parameterisation for $\boldsymbol{B}$ and $\boldsymbol{T}$ (assuming $i \neq i'$ and

$j \neq j'$):

$$b_{\{ij\},\{i'j\}} = \frac{\rho}{d_i^{(s)}} C_{\{ij\},\{i'j\}}, \quad b_{\{ij\},\{ij'\}} = -\frac{\omega_{jj'}}{\omega_{jj}} C_{\{ij\},\{ij'\}},$$
$$b_{\{ij\},\{i'j'\}} = \frac{\rho\omega_{jj'}}{d_i^{(s)}\omega_{jj}} C_{\{ij\},\{i'j'\}}, \quad \tau_{ij}^2 = \frac{1}{d_i^{(s)}\omega_{jj}},$$

where ρ and $\omega_{jj'}$ are linkage parameters, and $1/\omega_{jj}$ are variance components. This parameterisation does not seem straightforward, but is much clearer in the form of conditional mean:

$$\mathrm{E}\left(u_{ij} - \frac{\rho}{d_i^{(s)}} \sum_{i'\sim i} u_{i'j} \,\middle|\, u_{-\{ij\}}\right) = -\sum_{j'\sim j} \frac{\omega_{jj'}}{\omega_{jj}} \left(u_{ij'} - \frac{\rho}{d_i^{(s)}} \sum_{i'\sim i} u_{i'j'}\right). \tag{4}$$

Note that for a single response, the univariate conditional autoregressive model specifies $\mathrm{E}(u_i|u_{-i}) = \rho \sum_{i'\sim i} u_{i'}/d_i^{(s)}$. In the multivariate setting, $\rho \sum_{i'\sim i} u_{i'j}/d_i^{(s)}$ is no longer the conditional mean for $u_{ij} \mid u_{-\{ij\}}$ and their conditional difference is regressed on other differences through $\omega_{jj'}$. This parameterisation yields the joint precision matrix

$$\boldsymbol{T}^{-1}(\boldsymbol{I} - \boldsymbol{B}) = \left\{\boldsymbol{\Omega} \circ (\boldsymbol{I}_J + \boldsymbol{C}^{(r)})\right\} \otimes (\boldsymbol{D}^{(s)} - \rho\boldsymbol{C}^{(s)}) = \boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}} \otimes (\boldsymbol{D}^{(s)} - \rho\boldsymbol{C}^{(s)}), \tag{5}$$

where $\boldsymbol{\Omega}$ is a symmetric $J \times J$ matrix with entries $\omega_{jj'}$. A derivation of (5) is given in Appendix 1. The linkage parameter ρ is interpreted as a spatial smoothing parameter and $\boldsymbol{\Omega}$ controls the dependence across J responses. It is noteworthy that only nonzero entries in $\boldsymbol{\Omega}$ are parameters in the model and we denote $\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}} = \boldsymbol{\Omega} \circ (\boldsymbol{I}_J + \boldsymbol{C}^{(r)})$ for simplicity. The notation $\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}$, commonly used in graphical models, means the precision matrix restricted by graph $\boldsymbol{C}^{(r)}$. The model (5) is a natural extension of Gelfand and Vounatsou (2003)'s model. When $\boldsymbol{C}^{(r)}$ is the complete graph (any two vertices are connected), $\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}$ is free of zero entries. Then let $\boldsymbol{\Sigma} = \boldsymbol{\Omega}^{-1}$ and (5) is equivalent to Gelfand and Vounatsou (2003)'s specification. We call this a completely separable specification because the Kronecker product completely separates the spatial domain and the response domain. This complete separation is often not desirable because it makes the spatial smoothing common for all j. We call this homogeneous spatial smoothing because the linkage ρ is the same for any i and i' which distinguishes Model 2 from Model 3 in the next section.

The joint precision matrix (5) is positive definite if $|\rho| < 1$ and $\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}$ is positive definite. Let $M^+(\boldsymbol{C}^{(r)})$ be the cone of symmetric positive definite matrices restricted by $\boldsymbol{C}^{(r)}$ and then $\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}} \in M^+(\boldsymbol{C}^{(r)})$. In a Bayesian model, a widely used prior on $\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}$ is the G-Wishart distribution (Atay-Kayis & Massam, 2005; Letac & Massam, 2007). The G-Wishart distribution is a conjugate family for the precision matrix of a Gaussian graphical model, whose density function is given by

$$p(\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}} \mid b, \boldsymbol{V}) = I_{\boldsymbol{C}^{(r)}}(b, \boldsymbol{V})^{-1} \left|\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}\right|^{(b-2)/2} \exp \times \left\{-\frac{1}{2}\operatorname{tr}(\boldsymbol{V}\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}})\right\} 1_{\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}} \in M^+(\boldsymbol{C}^{(r)})},$$

where $b > 2$ is the number of degrees of freedom; $\boldsymbol{V}$ is the scale matrix and $I_{\boldsymbol{C}^{(r)}}(\cdot)$ is the normalising constant. It is practically attractive because of its conjugacy. That said, for a prior distribution $\mathrm{GWis}(b, \boldsymbol{V})$ and a given sample covariance matrix $\boldsymbol{S}$ of sample size n, the posterior distribution of $\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}$ is $\mathrm{GWis}(b + n, \boldsymbol{V} + \boldsymbol{S})$. Inference and computation under this model are given in Appendix 2.

2.4. Model 3: separable specification with heterogenous spatial smoothing

Dobra, Lenkoski, and Rodriguez (2011) introduced a multivariate lattice model by giving Kronecker-product G-Wishart priors to the matrix-variate $\boldsymbol{U}$. In our framework, $\boldsymbol{B}$ and $\boldsymbol{T}$ can be parameterised in the following way (assuming $i \neq i'$ and $j \neq j'$):

$$b_{\{ij\},\{i'j\}} = -\frac{\omega_{ii'}^{(s)}}{\omega_{ii}^{(s)}} C_{\{ij\},\{i'j\}}, \quad b_{\{ij\},\{ij'\}} = -\frac{\omega_{jj'}^{(r)}}{\omega_{jj}^{(r)}} C_{\{ij\},\{ij'\}},$$
$$b_{\{ij\},\{i'j'\}} = -\frac{\omega_{ii'}^{(s)}\omega_{jj'}^{(r)}}{\omega_{ii}^{(s)}\omega_{jj}^{(r)}} C_{\{ij\},\{i'j'\}}, \quad \tau_{ij}^2 = \frac{1}{\omega_{ii}^{(s)}\omega_{jj}^{(r)}},$$

which is equivalent to the version of conditional mean

$$\mathrm{E}\left(u_{ij} - \frac{1}{\omega_{ii}^{(s)}} \sum_{i'\sim i} \omega_{ii'}^{(s)} u_{i'j} \,\middle|\, u_{-\{ij\}}\right) = -\sum_{j'\sim j} \frac{\omega_{jj'}}{\omega_{jj}} \left(u_{ij'} - \frac{1}{\omega_{ii}^{(s)}} \sum_{i'\sim i} \omega_{ii'}^{(s)} u_{i'j'}\right). \tag{6}$$

Comparing (6) with (4), instead of a homogeneous spatial smoothing with ρ, it has a heterogeneous specification with $\omega_{ii'}^{(s)}$. This is hence more flexible in the spatial domain. The resulting joint precision matrix is

$$\boldsymbol{T}^{-1}(\boldsymbol{I} - \boldsymbol{B}) = \left\{\boldsymbol{\Omega}^{(r)} \circ (\boldsymbol{I}_J + \boldsymbol{C}^{(r)})\right\} \otimes \left\{\boldsymbol{\Omega}^{(s)} \circ (\boldsymbol{I}_I + \boldsymbol{C}^{(s)})\right\} = \boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}} \otimes \boldsymbol{\Omega}_{\boldsymbol{C}^{(s)}}, \tag{7}$$

where $\boldsymbol{\Omega}^{(r)}$ is a symmetric $J \times J$ matrix with entries $\omega_{jj'}^{(r)}$ and $\boldsymbol{\Omega}^{(s)}$ is a symmetric $I \times I$ matrix with entries $\omega_{ii'}^{(s)}$. A derivation of (7) is given in Appendix 1. We again

use $\mathbf{\Omega}_{\boldsymbol{C}^{(r)}}$ and $\mathbf{\Omega}_{\boldsymbol{C}^{(s)}}$ for simplicity. In model (5), the spatial part is the conventional conditional autoregressive model while in model (7), it is modelled by a more flexible one $\mathbf{\Omega}_{\boldsymbol{C}^{(s)}}$.

The precision matrix (7) is positive definite if both $\mathbf{\Omega}_{\boldsymbol{C}^{(r)}}$ and $\mathbf{\Omega}_{\boldsymbol{C}^{(s)}}$ are positive definite. In a Bayesian model, both can have G-Wishart priors. The specification has an obvious problem of identification: $\mathbf{\Omega}_{\boldsymbol{C}^{(r)}} \otimes \mathbf{\Omega}_{\boldsymbol{C}^{(s)}} = z\mathbf{\Omega}_{\boldsymbol{C}^{(r)}} \otimes (1/z)\mathbf{\Omega}_{\boldsymbol{C}^{(s)}}$, where z is an arbitrary constant scalar. Following Wang and West (2009), one can impose a constraint $\mathbf{\Omega}_{\boldsymbol{C}^{(r)},11} = 1$ and add an auxiliary variable z. Then specify a joint prior on $(z, z\mathbf{\Omega}_{\boldsymbol{C}^{(r)}})$:

$$p(z, z\mathbf{\Omega}_{\boldsymbol{C}^{(r)}} \mid b^{(r)}, \boldsymbol{V}^{(r)}) \propto p_{GWis}(z\mathbf{\Omega}_{\boldsymbol{C}^{(r)}} \mid b^{(r)}, \boldsymbol{V}^{(r)}) \cdot 1, \tag{8}$$

where $p_{GWis}(\cdot)$ is the density of G-Wishart distribution. Transform this joint density to $(z, \mathbf{\Omega}_{\boldsymbol{C}^{(r)}})$ and we obtain the desired joint prior. There is no additional constraint imposed on $\mathbf{\Omega}_{\boldsymbol{C}^{(s)}}$ and let $\mathbf{\Omega}_{\boldsymbol{C}^{(s)}} \sim \text{GWis}(b^{(s)}, \boldsymbol{V}^{(s)})$. Inference and computation under this model are given in Appendix 2.

2.5. Priors for the graph

The two types of graphs used in this modelling framework should be treated differently. On one hand, the spatial graph should be treated known and fixed because the geographical locations and their neighbouring structure is fixed in most scenarios. On the other hand, the response graph should be treated unknown because we often know little about the relationship between multiple responses. In the literature of Gaussian graphical model determination, usually the unknown graph is assumed random and a prior on the graph is assigned. The Markov chain Monte Carlo (MCMC) sampling scheme, such as the reversible jump MCMC (Green, 1995), is often used to sample graphs from the posterior distribution. In this paper, we adopt and slightly modify existing MCMC algorithms for the graph determination (Dobra et al., 2011; Wang & Li, 2012), with computational details given in Appendix 2, for each aforementioned model. For the prior choice of $\boldsymbol{C}^{(r)}$, consider

$$P(\boldsymbol{C}^{(r)}) \propto B(a + \text{size}(\boldsymbol{C}^{(r)}), b + m - \text{size}(\boldsymbol{C}^{(r)}))/B(a, b), \tag{9}$$

where $B(\cdot, \cdot)$ is the beta function, m is the total number of possible edges $\binom{J}{2}$, $\text{size}(\boldsymbol{C}^{(r)}) \in \{0, 1, \ldots, m\}$, and a and b are given hyperparameters. More details about this prior can be found in Scott and Berger (2006) and Scott and Carvalho (2009). The following prior is often used as well (Dobra et al., 2011):

$$P(\boldsymbol{C}^{(r)}) \propto \pi^{\text{size}(\boldsymbol{C}^{(r)})}(1-\pi)^{m-\text{size}(\boldsymbol{C}^{(r)})}, \tag{10}$$

where $\pi \in (0, 1)$ is a given hyperparameter. Sparser graphs can be favoured by choosing a small value for π. The prior (9) can be obtained by integrating π out with a hyperprior $\text{Beta}(a, b)$ on π.

3. An application

We illustrate the proposed models with a real example of disease mapping. It is known that smoking is linked with multiple diseases in the population, of which leading diseases include lung diseases and heart diseases. The dataset under consideration here includes six variables, among which four variables are related to the smoke exposure and the other two are diseases. Obtained from the 2011 Missouri County Level Survey, the four smoke exposure variables are: Current Cigarette Smoking, Current Smokeless Tobacco Use, Current Other Tobacco Use, Exposure to Secondhand Smoke. Data are binary responses to the survey questionnaires (Yes or No), aggregated to each county level. The other two variables, obtained from the Surveillance, Epidemiology and End Results (SEER) program, are the Lung Cancer Mortality and the Heart Diseases Mortality, both of which are counts for each county within a specified time period. To summarise, we have $I = 115$ counties and $J = 6$ response variables. Let $n_{i1}, \ldots, n_{i4}$ be the numbers of respondents in the survey and let E_{i5} and E_{i6} be the age-adjusted expected mortality for the two diseases. Then, the proportions y_{ij}/n_{ij}, $j = 1, \ldots, 4$ are empirical estimates of the prevalences of the survey variables, and the proportions y_{ij}/E_{ij}, $j = 5, 6$ are standardised mortality ratios of the diseases.

Consider a Bayesian hierarchical model for y_{ij}. We use the binomial-logit model and the Poisson-lognormal model (Banerjee, Gelfand, & Carlin, 2004) for $y_{i,1-4}$ and $y_{i,5-6}$, respectively, i.e.,

$$\begin{aligned} &y_{ij} \sim \text{Bin}(n_{ij}, p_{ij}), \quad \text{logit}(p_{ij}) = \beta_j + u_{ij}, \\ &i = 1, \ldots, 115 \text{ and } j = 1, \ldots, 4; \\ &y_{ij} \sim \text{Poi}(E_{ij}\eta_{ij}), \quad \log(\eta_{ij}) = \beta_j + u_{ij}, \\ &i = 1, \ldots, 115 \text{ and } j = 5, 6. \end{aligned}$$

For simplicity, we do not consider other covariates in this example. The primary interest here is to model the random effects u_{ij}, which are expected to be correlated in both the spatial domain and the response domain. To complete the model specification, specify a weakly-informative normal prior for the intercepts β_j and a multivariate conditional autoregressive model MCAR($\boldsymbol{B}, \boldsymbol{T}$) for the random effects $\boldsymbol{U} = \{u_{ij}\}$. We apply the three proposed versions of MCAR($\boldsymbol{B}, \boldsymbol{T}$) here. Hyperparameters for prior distributions are specified as follows. For the graph, noticing that the choice of π in (10) can influence the posterior inference, we consider the prior (10) with both $\pi = 0.2$ in favour of a sparse graph and $\pi = 0.5$ as no preference. All other priors are chosen to be only weakly-informative

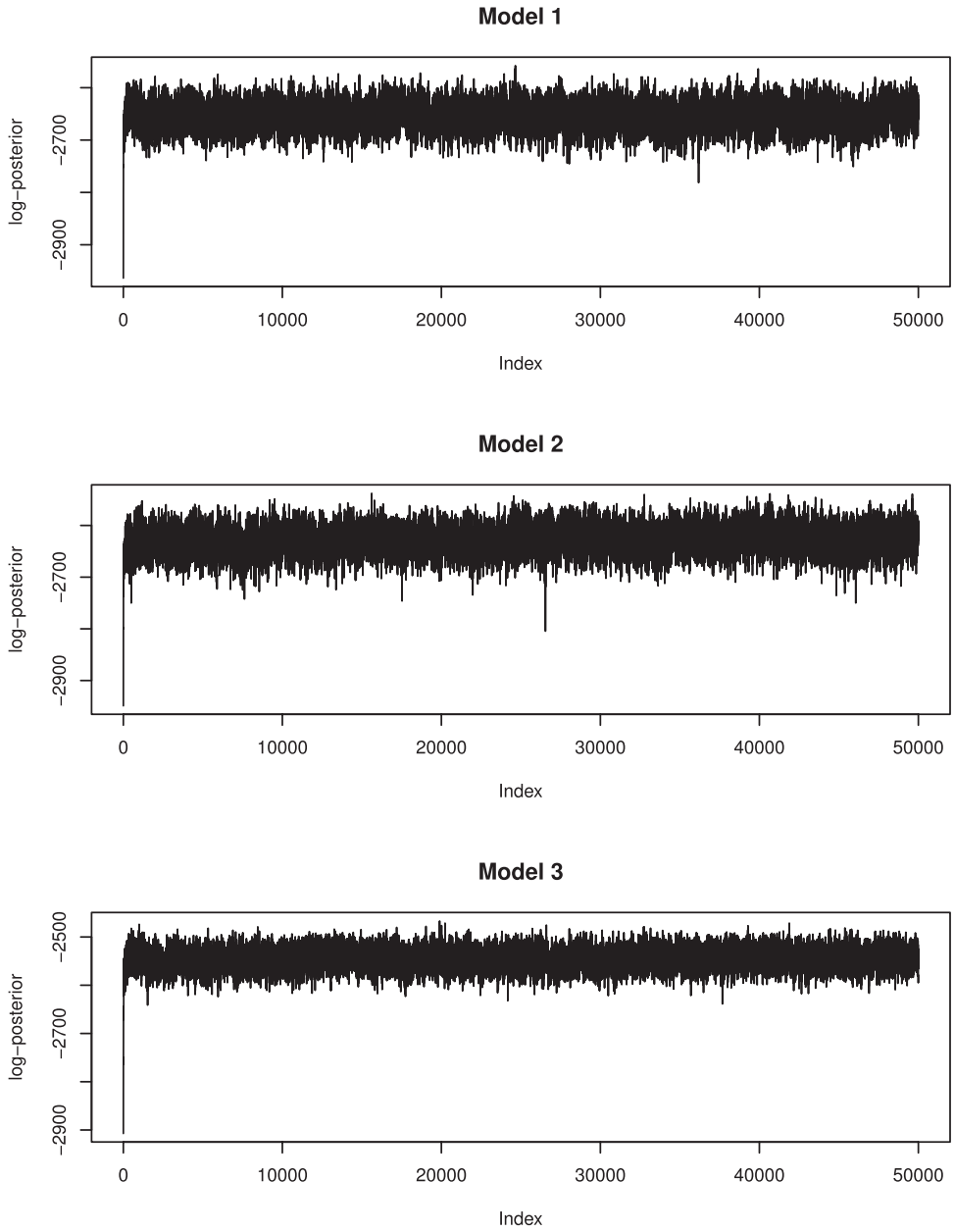

Figure 1. Data analysis: convergence of the log-joint posterior $\log p(\boldsymbol{\theta} \mid \text{data})$ under the three models (first 50,000 iterations), where $\boldsymbol{\theta}$ represents the collection of all parameters.

and have little impact on the posterior inference. In Model 1, we specify hyperparameters in the inverse gamma prior as $a_j = b_j = 0.5$. In Model 2, we specify hyperparameters in the G-Wishart prior as $b = 3$ and $\boldsymbol{V} = \boldsymbol{I}$. In Model 3, we specify hyperparameters in the two G-Wishart priors as $b = 3, \boldsymbol{V} = \boldsymbol{I}, b^{(s)} = 24$ and $\boldsymbol{V}^{(s)} = (b^{(s)} - 2)(\boldsymbol{D} - 0.95\boldsymbol{C})^{-1}$, which implies a prior mode at a proper conditional autoregressive model. For each model, we perform the Markov chain Monte Carlo for 150,000 iterations with a burn-in size of 50,000. Posterior results are based on the remaining samples. Figure 1 shows the convergence of the log-joint-posterior and notice that they all converge quickly.

Table 1 shows the posterior edge inclusion probabilities for the response graph $\boldsymbol{C}^{(r)}$. First, all three models seem to agree on the link between Cigarette Smoking and Secondhand Smoke Exposure, as well as the link between Lung Diseases Mortality and Heart Diseases Mortality. There is a moderate agreement on the links between Secondhand Smoke Exposure and Lung Diseases Mortality, and between Cigarette Smoking and Lung Diseases Mortality. In general, Model 1 tends to be a sparser graph, which is possibly due to the diagonal dominance condition. Model 2 is the simplest model as reflected by its *pD*, the effective number of parameters but has the largest *DIC*. Model 3 is the most flexible model among the three, and the inferred graph tends to be denser than the other two. It is as expected that its *pD* is larger but the overall criterion *DIC* is much smaller than the other two. Second, the edge inclusion probabilities are in general higher when $\pi = 0.5$, as expected,

Table 1. Data analysis: Posterior edge inclusion probabilities for the response graph and deviance information criterion for Models 1–3, respectively.

	Cigarette	Smokeless	Other	Secondhand	Lung	Heart	pD	DIC
				Model 1				
Cigarette		0	0.165	**0.963**	0.276	0	460.8	5766.5
Smokeless	0		0.181	0	0	0.002		
Other	0.258	0		0.032	0.014	0.005		
Secondhand	**0.983**	0	0.062		0.453	0		
Lung	0.301	0	0.040	**0.544**		**0.533**		
Heart	0.005	0	0.039	0.007	**0.772**		460.2	5765.0
				Model 2				
Cigarette		0.213	**0.919**	**1**	**0.756**	0.373	445.1	5813.8
Smokeless	0.340		0.207	0.281	0.249	**0.706**		
Other	**0.872**	0.316		0.390	0.346	**0.898**		
Secondhand	**1**	**0.527**	0.475		**0.732**	0.343		
Lung	**0.584**	0.407	0.412	**0.871**		**1**		
Heart	0.352	**0.942**	**0.838**	0.400	**1**		447.2	5812.6
				Model 3				
Cigarette		0.268	0.328	**0.827**	**0.541**	0.436	543.3	5630.4
Smokeless	0.249		0.417	**0.692**	**0.515**	**0.726**		
Other	0.317	0.403		**0.534**	0.453	**0.510**		
Secondhand	**0.819**	**0.694**	**0.531**		**0.936**	**0.801**		
Lung	**0.521**	**0.500**	0.435	**0.941**		**0.997**		
Heart	0.411	**0.732**	**0.508**	**0.805**	**0.998**		543.1	5629.2

Notes: Inclusion probabilities higher than 0.5 are in bold. The upper-right panel is for $\pi = 0.2$ and the lower-left panel is for $\pi = 0.5$.

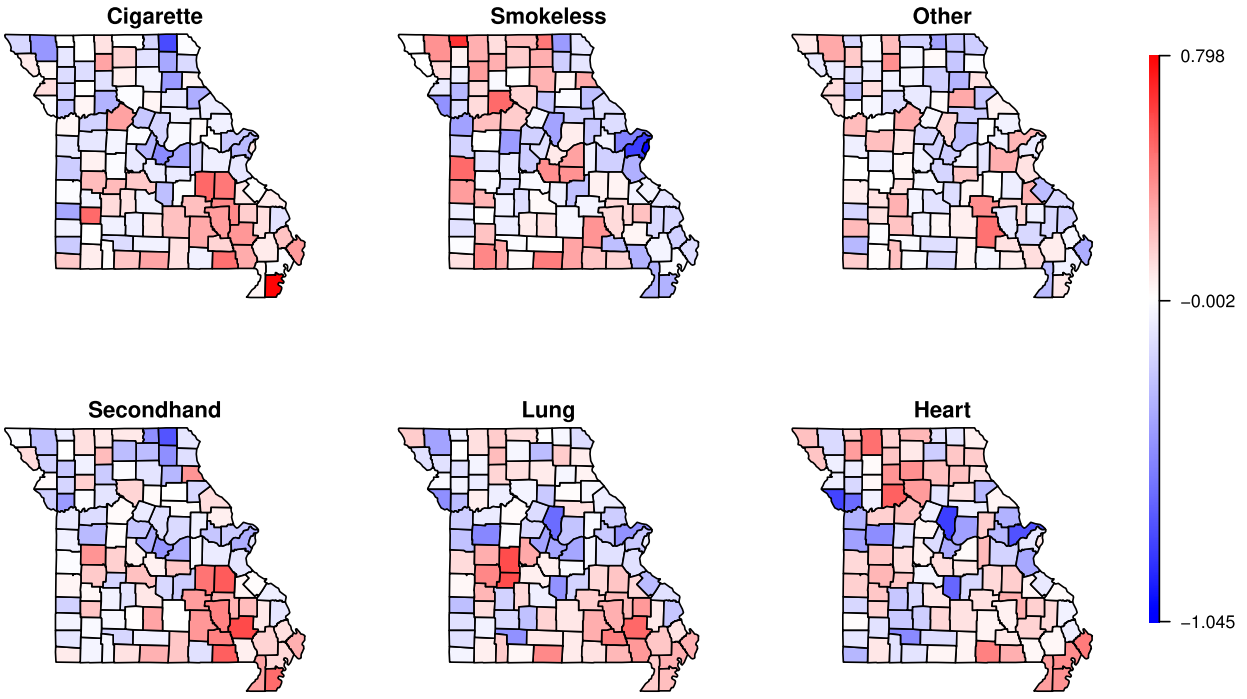

Figure 2. Data analysis: posterior means of spatial random effects u_{ij} under Model 1.

but it has little material impact on the final inferred graph. The *DIC* has little change with different π values. Lastly, Figures 2–4 show the maps of spatial random effects u_{ij} for the three models, respectively, and for a problem of disease mapping, this is often the eventual output for practitioners.

4. Simulation

To validate the proposed algorithms, we perform a simulation study on a 7×7 regular grid ($I = 49$ area units) with $J = 4$ response variables. Consider the true response graph with two edges $C_{13}^{(r)}$ and $C_{24}^{(r)}$. In this simulation study, we do not consider the scenario with misspecified models, and therefore, data are generated under each of the three models and the correct model is then used for inference. The parameter settings are given as follows. For Model 1, $\lambda_j = 0.95$, $\phi_{jj'} = \psi_{jj'} = 0.9$, $\delta_j = 1$ and $\beta_j = 1$. For Model 2, $\rho = 0.9$, $\omega_{jj} = 4$, $\omega_{13} = \omega_{24} = -3.2$ and $\beta_j = 1$. For Model 3, parameters are the same as Model 2 but $\boldsymbol{\Omega}_{\boldsymbol{C}^{(s)}}$ is generated from $\text{GWis}(10, 8(\boldsymbol{D} - 0.9\boldsymbol{C})^{-1})$. We repeat the simulation and inference process for $L = 50$ times and for each time, the MCMC iteration number is 5000.

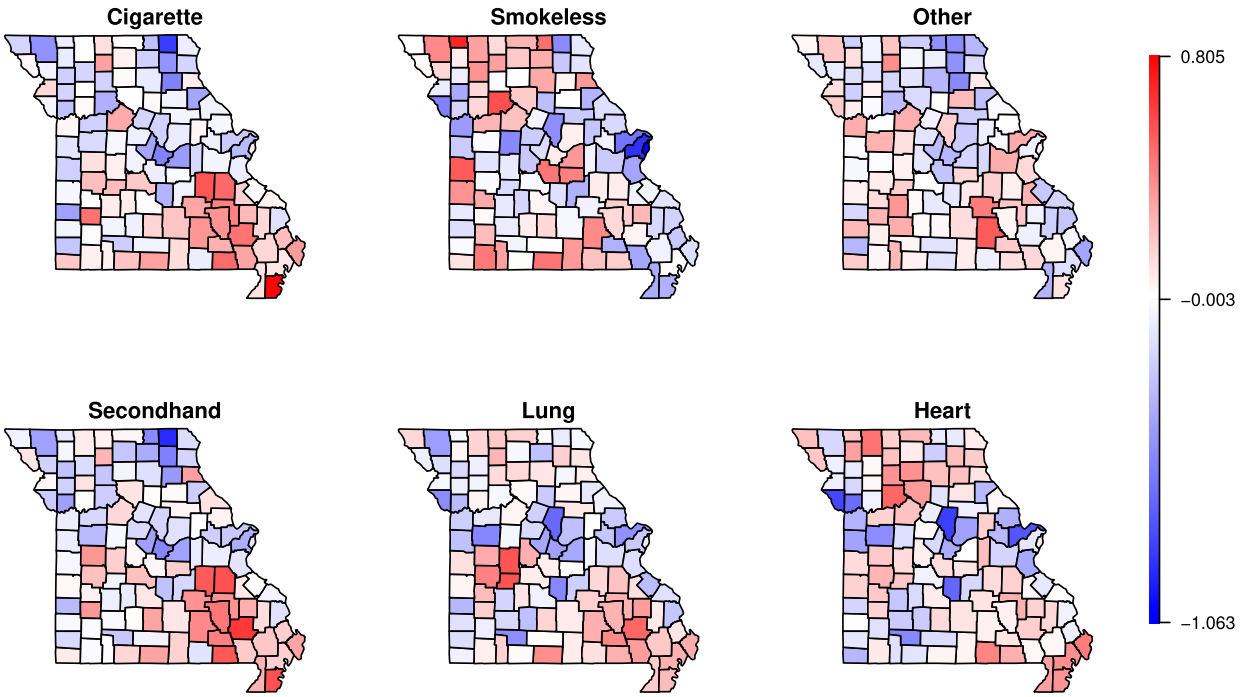

Figure 3. Data analysis: posterior means of spatial random effects u_{ij} under Model 2.

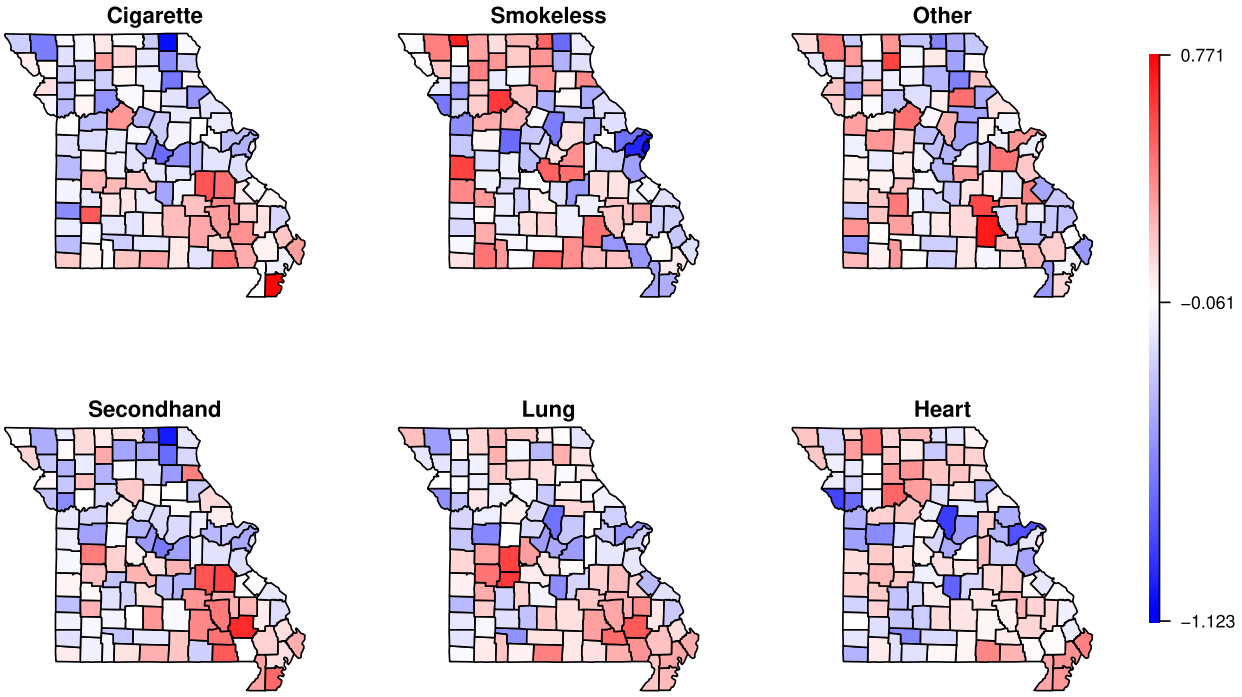

Figure 4. Data analysis: posterior means of spatial random effects u_{ij} under Model 3.

We consider three measures for validating and comparing the three algorithms. The first measure is the mean inclusion probability matrix with standard deviations. We call the second measure the error rate of mis-identified edges. If we use 0.5 as the threshold for identifying an edge in the graph, for each replication, we obtain an inferred graph and then compare with the true graph to record a proportion of wrong edges/non-edges. The error rate is the average proportion of L replications. The third measure is the mean absolute error (MAE) of random effects in the model,

$$\text{MAE} = \frac{1}{L}\frac{1}{J}\frac{1}{I}\sum_{l}\sum_{j}\sum_{i}\left|\frac{\hat{u}_{ijl}-u_{ijl}}{u_{ijl}}\right|$$

where u_{ijl} is the true value and $\hat{u}_{ijl}$ is the posterior mean.

Simulation results are given in Table 2. For all three models, the algorithms can correctly identify the true

Table 2. Simulation: Mean posterior edge inclusion probabilities (standard deviations in parentheses), error rates of mis-identified edges, and the mean absolute errors of random effects.

	Var 2	Var 3	Var 4	Error Rate	MAE
			Model 1		
Var 1	0.128 (0.214)	**0.664** (0.358)	0.088 (0.107)	0.15	3.167
Var 2		0.100 (0.148)	**0.615** (0.352)		
Var 3			0.087 (0.155)		
			Model 2		
Var 1	0.211 (0.108)	**1** (0)	0.200 (0.086)	0.02	2.153
Var 2		0.213 (0.134)	**1** (0)		
Var 3			0.199 (0.107)		
			Model 3		
Var 1	0.246 (0.037)	**0.961** (0.051)	0.292 (0.054)	0.07	2.218
Var 2		0.438 (0.092)	**0.996** (0.008)		
Var 3			0.426 (0.083)		

Note: True edges are in bold.

edges. The algorithm for Model 1 appears to be unstable as the standard deviation is large and tends to underestimate inclusion probabilities, while the algorithm for Model 3 tends to overestimate inclusion probabilities for non-edges. The algorithm for Model 2 presents the smallest error rate and MAE. Note that this simulation study validates the proposed algorithms under correct model specifications and hence the result cannot imply that Model 2 is the best model for a real dataset. In fact, as shown in the data analysis, Model 2 is the simplest specification and is the least preferred model in that case according to *DIC*.

5. Further discussion

In this paper, we proposed a modelling framework for multivariate areal data from a graphical model perspective. We rebuilt three well-known models in our framework and developed Bayesian inference tools for the proposed models. It is our perspective that this framework is very general and can contain other models that are beyond the cases discussed in the paper. For example, Jin et al. (2007) specified a co-regionalised areal data model, in which their Case 3 is a very general specification. We show that this specification can be reproduced and extended in our framework. Consider the Cholesky decomposition $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^T$. Jin et al. (2007)'s Case 3 specification of the joint covariance matrix is $(\boldsymbol{A} \otimes \boldsymbol{I}_I)(\boldsymbol{I}_J \otimes \boldsymbol{D}^{(s)} - \boldsymbol{\Phi} \otimes \boldsymbol{C}^{(s)})^{-1}(\boldsymbol{A} \otimes \boldsymbol{I}_I)^T$ whose inverse is then

$$(\boldsymbol{A}\boldsymbol{A}^T)^{-1} \otimes \boldsymbol{D}^{(s)} - (\boldsymbol{A}^{-1})^T\boldsymbol{\Phi}\boldsymbol{A}^{-1} \otimes \boldsymbol{C}^{(s)} \tag{11}$$

where $\boldsymbol{\Phi}$ is a symmetric $J \times J$ matrix. Let $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} = (\boldsymbol{A}\boldsymbol{A}^T)^{-1}$ and $\boldsymbol{Q} = (\boldsymbol{A}^{-1})^T\boldsymbol{\Phi}\boldsymbol{A}^{-1}$. Obviously it is one-to-one from $(\boldsymbol{A}, \boldsymbol{\Phi})$ to $(\boldsymbol{\Omega}, \boldsymbol{Q})$. Specification (11) is hence equivalent to

$$\boldsymbol{\Omega} \otimes \boldsymbol{D}^{(s)} - \boldsymbol{Q} \otimes \boldsymbol{C}^{(s)}, \tag{12}$$

where $\boldsymbol{Q}$ is a symmetric $J \times J$ matrix with entries $q_{jj'}$. To reproduce this specification in our framework, parameterise $\boldsymbol{B}$ and $\boldsymbol{T}$ as follows (assuming $i \neq i'$ and $j \neq j'$):

$$b_{\{ij\},\{i'j\}} = \frac{q_{jj}}{d_i^{(s)}\omega_{jj}} C_{\{ij\},\{i'j\}}, \quad b_{\{ij\},\{ij'\}} = -\frac{\omega_{jj'}}{\omega_{jj}} C_{\{ij\},\{ij'\}},$$

$$b_{\{ij\},\{i'j'\}} = \frac{q_{jj'}}{d_i^{(s)}\omega_{jj}} C_{\{ij\},\{i'j'\}}, \quad \tau_{ij}^2 = \frac{1}{d_i^{(s)}\omega_{jj}}.$$

This parameterisation leads to the joint precision matrix

$$\begin{aligned} \boldsymbol{T}^{-1}(\boldsymbol{I} - \boldsymbol{B}) &= \left\{\boldsymbol{\Omega} \circ (\boldsymbol{I}_J + \boldsymbol{C}^{(r)})\right\} \otimes \boldsymbol{D}^{(s)} \\ &\quad - \left\{\boldsymbol{Q} \circ (\boldsymbol{I}_J + \boldsymbol{C}^{(r)})\right\} \otimes \boldsymbol{C}^{(s)} \\ &= \boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}} \otimes \boldsymbol{D}^{(s)} - \boldsymbol{Q}_{\boldsymbol{C}^{(r)}} \otimes \boldsymbol{C}^{(s)}. \end{aligned} \tag{13}$$

The expression (13) reduces to (12) which is equivalent to Jin et al. (2007)'s (11) when $\boldsymbol{C}^{(r)}$ is a complete graph. A derivation of (13) is given in Appendix 1. The validity of this model relies on the positive definiteness of (13). Jin et al. (2007) showed that it is positive definite if $\boldsymbol{\Omega}$ is positive definite and eigenvalues of $\boldsymbol{\Phi} = \boldsymbol{A}^T\boldsymbol{Q}\boldsymbol{A}$ are between $1/\xi_{\min}$ and $1/\xi_{\max}$, reciprocals of the smallest and largest eigenvalues of $\boldsymbol{D}^{(s)-1/2}\boldsymbol{C}^{(s)}\boldsymbol{D}^{(s)-1/2}$, which are known constants. The graphical version (13) must also satisfy this condition, that is, $1/\xi_{\min} \leq \lambda(\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}^{-1}\boldsymbol{Q}_{\boldsymbol{C}^{(r)}}) \leq 1/\xi_{\max}$, where $\lambda(\boldsymbol{M})$ is any eigenvalue of $\boldsymbol{M}$. Considering that both $\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}$ and $\boldsymbol{Q}_{\boldsymbol{C}^{(r)}}$ are restricted by the underlying graph, the eigenvalue condition is not easy to implement in computations. This matter is worth investigating in the future.

In general, flexible models are desired for modelling multivariate areal data because overly simplistic models may misspecify the true underlying covariance structure. However, there is almost always a trade-off between the simplicity and the flexibility of a model. It is probably reasonable to allow certain flexibilities for specific purposes, such as in this paper, for learning a graphical relationship between multiple responses. It is usually the practitioner's choice whether a more flexible but complicated model is needed for the problem at

hand, especially when the performance improvement is negligible.

Acknowledgments

The author would like to thank two editors and two anonymous referees for their helpful comments which lead to a much improved paper.

Disclosure statement

No potential conflict of interest was reported by the author.

Notes on contributor

Ye Liang is an associate professor in the Department of Statistics at Oklahoma State University. He received his Ph.D. in statistics from the University of Missouri.

ORCID

Ye Liang http://orcid.org/0000-0002-6513-8962

References

Atay-Kayis, A., & Massam, H. (2005). A Monte Carlo method for computing the marginal likelihood in nondecomposable Gaussian graphical models. *Biometrika*, *92*, 317–335.

Banerjee, S., Gelfand, A., & Carlin, B. (2004). *Hierarchical modeling and analysis for spatial data*. Boca Raton, FL: CRC Press/Chapman & Hall.

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

Dobra, A., Lenkoski, A., & Rodriguez, A. (2011). Bayesian inference for general Gaussian graphical models with applications to multivariate lattice data. *Journal of the American Statistical Association*, *106*, 1418–1433.

Gelfand, A., Sahu, S., & Carlin, B. (1995). Efficient parameterizations for normal linear mixed models. *Biometrika*, *82*, 479–488.

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

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, *82*, 711–732.

Jin, X., Banerjee, S., & Carlin, B. (2007). Order-free co-regionalized areal data models with application to multiple-disease mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *69*, 817–838.

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

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

Letac, G., & Massam, H. (2007). Wishart distributions for decomposable graphs. *The Annals of Statistics*, *35*, 1278–1323.

MacNab, Y. (2011). On Gaussian Markov random fields and Bayesian disease mapping. *Statistical Methods in Medical Research*, *20*, 49–68.

MacNab, Y. C. (2016). Linear models of coregionalization for multivariate lattice data: A general framework for coregionalized multivariate CAR models. *Statistics in Medicine*, *35*, 3827–3850.

MacNab, Y. C. (2018). Some recent work on multivariate Gaussian Markov random fields. *Test*, *27*, 497–541.

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

Martinez-Beneito, M. (2013). A general modeling framework for multivariate disease mapping. *Biometrika*, *100*, 539–553.

Martinez-Beneito, M. A., Botella-Rocamora, P., & Banerjee, S. (2017). Towards a multidimensional approach to Bayesian disease mapping. *Bayesian Analysis*, *12*, 239–259.

Scott, J., & Berger, J. (2006). An exploratory of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference*, *136*, 2144–2162.

Scott, J., & Carvalho, C. (2009). Feature-inclusion stochastic search for Gaussian graphical models. *Journal of Computational and Graphical Statistics*, *17*, 790–808.

Wang, H., & Li, S. (2012). Efficient Gaussian graphical model determination under G-Wishart prior distributions. *Electronic Journal of Statistics*, *6*, 168–198.

Wang, H., & West, M. (2009). Bayesian analysis of matrix normal graphical models. *Biometrika*, *96*, 821–834.

Appendices

Appendix 1. Derivations

A.1 Derivation of Equation (3)

With the parameterisation in Model 1, we have $\boldsymbol{T} = \boldsymbol{D}^{-1}(\boldsymbol{\Delta} \otimes \boldsymbol{I}_I)$ and

$$\boldsymbol{B} = (\boldsymbol{\Delta}^{1/2} \otimes \boldsymbol{I}_I)\boldsymbol{D}^{-1}\left\{\boldsymbol{\Lambda} \otimes \boldsymbol{C}^{(s)} + (\boldsymbol{\Psi} \circ \boldsymbol{C}^{(r)}) \otimes \boldsymbol{I}_I + (\boldsymbol{\Phi} \circ \boldsymbol{C}^{(r)}) \otimes \boldsymbol{C}^{(s)}\right\}(\boldsymbol{\Delta}^{-1/2} \otimes \boldsymbol{I}_I).$$

Then immediately we have expression (3) for $\boldsymbol{T}^{-1}(\boldsymbol{I} - \boldsymbol{B})$.

A.2 Derivation of Equation (5)

With the parameterisation in Model 2, we have $\boldsymbol{T} = (\text{diag}(\boldsymbol{\Omega}) \otimes \boldsymbol{D}^{(s)})^{-1}$ and

$$\boldsymbol{B} = \rho \boldsymbol{I}_J \otimes \boldsymbol{D}^{(s)-1}\boldsymbol{C}^{(s)} - \left(\text{diag}(\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega} \circ \boldsymbol{C}^{(r)}\right) \otimes \boldsymbol{I}_I + \rho\left(\text{diag}(\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega} \circ \boldsymbol{C}^{(r)}\right) \otimes \boldsymbol{D}^{(s)-1}\boldsymbol{C}^{(s)}.$$

Then

$$\begin{aligned}\boldsymbol{T}^{-1}(\boldsymbol{I} - \boldsymbol{B}) &= \text{diag}(\boldsymbol{\Omega}) \otimes \boldsymbol{D}^{(s)} - \rho\,\text{diag}(\boldsymbol{\Omega}) \otimes \boldsymbol{C}^{(s)} \\ &\quad + (\boldsymbol{\Omega} \circ \boldsymbol{C}^{(r)}) \otimes \boldsymbol{D}^{(s)} - \rho(\boldsymbol{\Omega} \circ \boldsymbol{C}^{(r)}) \otimes \boldsymbol{C}^{(s)} \\ &= \left\{\boldsymbol{\Omega} \circ (\boldsymbol{I}_J + \boldsymbol{C}^{(r)})\right\} \otimes \boldsymbol{D}^{(s)} \\ &\quad - \left\{\boldsymbol{\Omega} \circ (\boldsymbol{I}_J + \boldsymbol{C}^{(r)})\right\} \otimes \rho\boldsymbol{C}^{(s)} \\ &= \left\{\boldsymbol{\Omega} \circ (\boldsymbol{I}_J + \boldsymbol{C}^{(r)})\right\} \otimes (\boldsymbol{D}^{(s)} - \rho\boldsymbol{C}^{(s)}),\end{aligned}$$

which is expression (5).

A.3 Derivation of Equation (7)

With the parameterisation in Model 3, we have $\boldsymbol{T} = (\text{diag}(\boldsymbol{\Omega}^{(r)}) \otimes \text{diag}(\boldsymbol{\Omega}^{(s)}))^{-1}$ and

$$\begin{aligned}\boldsymbol{B} = &-\boldsymbol{I}_J \otimes \left(\text{diag}(\boldsymbol{\Omega}^{(s)})^{-1}\boldsymbol{\Omega}^{(s)} \circ \boldsymbol{C}^{(s)}\right) \\ &- \left(\text{diag}(\boldsymbol{\Omega}^{(r)})^{-1}\boldsymbol{\Omega}^{(r)} \circ \boldsymbol{C}^{(r)}\right) \otimes \boldsymbol{I}_I \\ &- \left(\text{diag}(\boldsymbol{\Omega}^{(r)})^{-1}\boldsymbol{\Omega}^{(r)} \circ \boldsymbol{C}^{(r)}\right) \\ &\otimes \left(\text{diag}(\boldsymbol{\Omega}^{(s)})^{-1}\boldsymbol{\Omega}^{(s)} \circ \boldsymbol{C}^{(s)}\right).\end{aligned}$$

Then

$$\begin{aligned}\boldsymbol{T}^{-1}(\boldsymbol{I} - \boldsymbol{B}) &= \text{diag}(\boldsymbol{\Omega}^{(r)}) \otimes \text{diag}(\boldsymbol{\Omega}^{(s)}) + \text{diag}(\boldsymbol{\Omega}^{(r)}) \\ &\quad \otimes (\boldsymbol{\Omega}^{(s)} \circ \boldsymbol{C}^{(s)}) \\ &\quad + (\boldsymbol{\Omega}^{(r)} \circ \boldsymbol{C}^{(r)}) \otimes \text{diag}(\boldsymbol{\Omega}^{(s)}) + (\boldsymbol{\Omega}^{(r)} \circ \boldsymbol{C}^{(r)}) \\ &\quad \otimes (\boldsymbol{\Omega}^{(s)} \circ \boldsymbol{C}^{(s)}) \\ &= \left\{\boldsymbol{\Omega}^{(r)} \circ (\boldsymbol{I}_J + \boldsymbol{C}^{(r)})\right\} \otimes \left\{\boldsymbol{\Omega}^{(s)} \circ (\boldsymbol{I}_I + \boldsymbol{C}^{(s)})\right\},\end{aligned}$$

which is expression (7).

A.4 Derivation of Equation (13)

With the parameterisation in Model 4, we have $\boldsymbol{T} = (\text{diag}(\boldsymbol{\Omega}) \otimes \boldsymbol{D}^{(s)})^{-1}$ and

$$\begin{aligned}\boldsymbol{B} = &\,\text{diag}(\boldsymbol{\Omega})^{-1}\text{diag}(\boldsymbol{Q}) \otimes \boldsymbol{D}^{(s)-1}\boldsymbol{C}^{(s)} \\ &- \text{diag}(\boldsymbol{\Omega})^{-1}\boldsymbol{\Omega} \circ \boldsymbol{C}^{(r)} \otimes \boldsymbol{I}_I \\ &+ \text{diag}(\boldsymbol{\Omega})^{-1}\boldsymbol{Q} \circ \boldsymbol{C}^{(r)} \otimes \boldsymbol{D}^{(s)-1}\boldsymbol{C}^{(s)}.\end{aligned}$$

Then

$$\begin{aligned}\boldsymbol{T}^{-1}(\boldsymbol{I} - \boldsymbol{B}) &= \text{diag}(\boldsymbol{\Omega}) \otimes \boldsymbol{D}^{(s)} - \text{diag}(\boldsymbol{Q}) \otimes \boldsymbol{C}^{(s)} \\ &\quad + \boldsymbol{\Omega} \circ \boldsymbol{C}^{(r)} \otimes \boldsymbol{D}^{(s)} - \boldsymbol{Q} \circ \boldsymbol{C}^{(r)} \otimes \boldsymbol{C}^{(s)} \\ &= \left\{\boldsymbol{\Omega} \circ (\boldsymbol{I}_J + \boldsymbol{C}^{(r)})\right\} \otimes \boldsymbol{D}^{(s)} \\ &\quad - \left\{\boldsymbol{Q} \circ (\boldsymbol{I}_J + \boldsymbol{C}^{(r)})\right\} \otimes \boldsymbol{C}^{(s)},\end{aligned}$$

which is expression (13).

Appendix 2: Bayesian computations

A.5 A hierarchical generalised linear model

For illustration, we now assume a full Bayesian hierarchical model and give computational details for this model. Assume binomial counts y_{ij}/n_{ij} for J responses and I areal units. Specify a Bayesian model as follows, for $i = 1, \ldots, I$ and $j = 1, \ldots, J$,

$$\begin{aligned}&y_{ij} \mid p_{ij} \sim \text{Bin}(n_{ij}, p_{ij}), \quad \text{logit}(p_{ij}) = \beta_j + u_{ij}, \\ &\beta_j \sim \text{N}(0, \tau_0^2), \quad \boldsymbol{U} \mid \boldsymbol{B}, \boldsymbol{T} \sim \text{MCAR}(\boldsymbol{B}, \boldsymbol{T}),\end{aligned}$$

where τ_0^2 is a given constant, $\boldsymbol{U}$ is the matrix-variate of u_{ij}, and priors for $\boldsymbol{B}$ and $\boldsymbol{T}$ depend on the specific parameterisation. This section is organised as follows: we first give details of updating effects parameters β_j and u_{ij}, and then, separately for each model, details of updating parameters of MCAR and updating the random response graph $\boldsymbol{C}^{(r)}$.

A.6 Updating effects parameters

Our experience has shown that the convergence is poor if we directly update β_j and u_{ij}. We apply the hierarchical centring technique (Gelfand, Sahu, & Carlin, 1995) and block sampling. Let $\gamma_{ij} = \beta_j + u_{ij}$ and $\boldsymbol{\gamma} = \text{vec}[(\gamma_{ij})_{I\times J}]$ has a non-centred MCAR prior. We update (γ_{ij}, β_j) instead of (u_{ij}, β_j). The full conditional distribution of γ_{ij} is

$$\begin{aligned}p(\gamma_{ij} \mid \cdot) \propto \frac{e^{y_{ij}\gamma_{ij}}}{(1+e^{\gamma_{ij}})^{n_{ij}}} \exp\Bigg\{-\frac{1}{2\tau_{ij}^2}\Big(\gamma_{ij} - \beta_j \\ - \sum_{\{i'j'\}\neq\{ij\}} b_{\{ij\},\{i'j'\}}(\gamma_{i'j'} - \beta_{j'})\Big)\Bigg\}.\end{aligned}$$

We use Metropolis-Hastings algorithm to sample γ_{ij} from this conditional density. We block sample $\boldsymbol{\beta}$ in the following way. For now denote $\boldsymbol{M} = \boldsymbol{T}^{-1}(\boldsymbol{I} - \boldsymbol{B})$, the joint precision matrix. Let $\boldsymbol{\gamma}^* = \boldsymbol{M}\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}^{**}$ be a $J \times 1$ vector such that γ_1^{**} is the sum of the first I elements in $\boldsymbol{\gamma}^*$, γ_2^{**} is the sum of the second I elements in $\boldsymbol{\gamma}^*$ and so on. Partition $\boldsymbol{M}$ into $J \times J$ blocks and define

$$\boldsymbol{H} = \begin{pmatrix}\mathbf{1}^T\boldsymbol{M}_{11}\mathbf{1} & \cdots & \mathbf{1}^T\boldsymbol{M}_{1J}\mathbf{1} \\ \vdots & \ddots & \vdots \\ \mathbf{1}^T\boldsymbol{M}_{J1}\mathbf{1} & \cdots & \mathbf{1}^T\boldsymbol{M}_{JJ}\mathbf{1},\end{pmatrix}$$

where $\mathbf{1}$ is the all-one vector. Then the full conditional distribution for the vector $\boldsymbol{\beta}$ is

$$(\boldsymbol{\beta} \mid \cdot) \sim \text{N}\left[(\boldsymbol{H} + 1/\tau_0^2\boldsymbol{I})^{-1}\boldsymbol{\gamma}^{**}, (\boldsymbol{H} + 1/\tau_0^2\boldsymbol{I})^{-1}\right].$$

A.7 Model 1: updating δ_j, λ_j, $\psi_{jj'}$, $\phi_{jj'}$ and $C^{(r)}$

Given the current graph $\boldsymbol{C}^{(r)}$, parameters are updated through Gibbs sampling. Recall priors on these parameters: $\delta_j \sim \text{IG}(a_j, b_j)$ and $\lambda_j, \psi_{jj'}, \phi_{jj'} \sim \text{Unif}(-1, 1)$. Let $\boldsymbol{u}_j$ be the jth column vector of $\boldsymbol{U}$, $j = 1, \ldots, J$ and $\boldsymbol{D}_j$ be the jth diagonal block of $\boldsymbol{D}$. The full conditional distribution of δ_j is given by

$$\begin{aligned}p(\delta_j \mid \cdot) \propto \delta_j^{-I/2-a_j-1} \exp\Bigg\{-\frac{1}{2\delta_j}\boldsymbol{u}_j^T(\boldsymbol{D}_j - \lambda_j\boldsymbol{C}^{(s)})\boldsymbol{u}_j \\ + \sum_{j'\sim j}\frac{1}{\sqrt{\delta_j\delta_{j'}}}\boldsymbol{u}_j^T(\psi_{jj'}\boldsymbol{I}_I + \phi_{jj'}\boldsymbol{C}^{(s)})\boldsymbol{u}_{j'} - \frac{b_j}{\delta_j}\Bigg\}.\end{aligned}$$

It can be shown that the transformed one $(\sqrt{1/\delta_j} \mid \cdot)$ is log-concave when $I + 2a_j - 1 > 0$. Thus, we use the adaptive rejection sampling to update δ_j.

Let $\boldsymbol{W}$ be an $I \times J$ matrix, where $\text{vec}(\boldsymbol{W}) = (\boldsymbol{\Delta}^{-1/2} \otimes \boldsymbol{I}_I)\,\text{vec}(\boldsymbol{U})$ and $\boldsymbol{w}_j$ be the jth column vector of $\boldsymbol{W}$. Let $\boldsymbol{M} = \boldsymbol{D} - \boldsymbol{\Lambda} \otimes \boldsymbol{C}^{(s)} - (\boldsymbol{\Psi} \circ \boldsymbol{C}^{(r)}) \otimes \boldsymbol{I}_I - (\boldsymbol{\Phi} \circ \boldsymbol{C}^{(r)}) \otimes \boldsymbol{C}^{(s)}$ as in (3). Then λ_j, $\psi_{jj'}$ and $\phi_{jj'}$ are sequentially updated through following full conditional distributions,

$$\begin{aligned}p(\lambda_j \mid \cdot) &\propto \left|\boldsymbol{M}(\lambda_j)\right|^{1/2} \exp\left(\frac{1}{2}\lambda_j\boldsymbol{w}_j^T\boldsymbol{C}^{(s)}\boldsymbol{w}_j\right), \\ p(\psi_{jj'} \mid \cdot) &\propto \left|\boldsymbol{M}(\psi_{jj'})\right|^{1/2} \exp\left(\psi_{jj'}\boldsymbol{w}_j^T\boldsymbol{w}_{j'}\right), \\ p(\phi_{jj'} \mid \cdot) &\propto \left|\boldsymbol{M}(\phi_{jj'})\right|^{1/2} \exp\left(\phi_{jj'}\boldsymbol{w}_j^T\boldsymbol{C}^{(s)}\boldsymbol{w}_{j'}\right).\end{aligned}$$

We use Metropolis-Hastings algorithm to update these parameters. Note that evaluating the sparse $|\boldsymbol{M}|$ could be computationally intensive. An efficient algorithm, usually based on the Cholesky decomposition, on sparse matrices is helpful.

The graph $\boldsymbol{C}^{(r)}$ is updated through a simple reversible jump MCMC algorithm. Propose a new graph $\boldsymbol{C}^{(r)*}$ by only

adding or deleting one edge from $\boldsymbol{C}^{(r)}$. Without loss of generality, suppose that one edge $\{j0, k0\}$ is added to the new graph. Dimension has been changed by 2 from $(\boldsymbol{C}^{(r)}, \boldsymbol{\Psi}, \boldsymbol{\Phi})$ to $(\boldsymbol{C}^{(r)*}, \boldsymbol{\Psi}^*, \boldsymbol{\Phi}^*)$. Propose $u_1 \sim U(-1, 1)$ and $u_2 \sim U(-1, 1)$, and let $\psi^*_{j0,k0} = u_1$ and $\phi^*_{j0,k0} = u_2$. The Jacobian from $(\boldsymbol{\Psi}, \boldsymbol{\Phi}, u_1, u_2)$ to $(\boldsymbol{\Psi}^*, \boldsymbol{\Phi}^*)$ hence is 1. Choose a Bernoulli jump proposal with odds $q(\boldsymbol{C}^{(r)*}, \boldsymbol{C}^{(r)})/q(\boldsymbol{C}^{(r)}, \boldsymbol{C}^{(r)*}) = p(\boldsymbol{C}^{(r)})/p(\boldsymbol{C}^{(r)*})$ and systematically scan through the graph for updating. Accept the move from $\boldsymbol{C}^{(r)}$ to $\boldsymbol{C}^{(r)*}$ with probability $\min\{1, \alpha\}$ where

$$\alpha = \frac{|\boldsymbol{M}^*|^{1/2}}{|\boldsymbol{M}|^{1/2}} \exp\left\{-\frac{1}{2}\operatorname{vec}(\boldsymbol{W})(\boldsymbol{M}^* - \boldsymbol{M})\operatorname{vec}(\boldsymbol{W})\right\}.$$

A.8 Model 2: updating ρ, $\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}$ and $\boldsymbol{C}^{(r)}$

Given the current graph $\boldsymbol{C}^{(r)}$, parameters are updated through Gibbs sampling. Recall priors on these parameters: $\rho \sim U(-1, 1)$ and $\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}} \sim \text{GWis}(b, \boldsymbol{V})$. Use Metropolis-Hastings algorithm to update ρ. It can be shown that the full conditional distribution for ρ is

$$p(\rho \mid \cdot) \propto \left|\boldsymbol{D}^{(s)} - \rho \boldsymbol{C}^{(s)}\right|^{J/2} \exp \times \left\{\frac{\rho}{2}\operatorname{vec}(\boldsymbol{U})^T(\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}} \otimes \boldsymbol{C}^{(s)})\operatorname{vec}(\boldsymbol{U})\right\}.$$

Let $\boldsymbol{W}_1$ be an $I \times J$ matrix, where $\operatorname{vec}(\boldsymbol{W}_1) = [\boldsymbol{I}_J \otimes (\boldsymbol{D}^{(s)} - \rho \boldsymbol{C}^{(s)})^{1/2}]\operatorname{vec}(\boldsymbol{U})$ and $\boldsymbol{w}_{1,j}$ be the jth column vector of $\boldsymbol{W}_1$. Let $\boldsymbol{S}$ be an $J \times J$ matrix with $s_{jj'} = \boldsymbol{w}^T_{1,j}\boldsymbol{w}_{1,j'}$. Then the full conditional distribution of $\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}$ is

$$p(\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}} \mid \cdot) \propto \left|\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}\right|^{(b+I-2)/2} \exp \times \left[-\frac{1}{2}\text{tr}\left\{\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}(\boldsymbol{V} + \boldsymbol{S})\right\}\right] 1_{\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}} \in M^+(\boldsymbol{C}^{(r)})},$$

which is $\text{GWis}(b + I, \boldsymbol{V} + \boldsymbol{S})$. For sampling from the G-Wishart distribution, we use the block Gibbs sampler, given the set of maximum cliques, introduced by Wang and Li (2012).

The graph $\boldsymbol{C}^{(r)}$ is updated using (Wang & Li, 2012)'s partial analytic structure algorithm (p. 188, Algorithm 2).

A.9 Model 3: updating $\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}$, $\boldsymbol{\Omega}_{\boldsymbol{C}^{(s)}}$ and $\boldsymbol{C}^{(r)}$

Given the current graph $\boldsymbol{C}^{(r)}$, parameters are updated through Gibbs sampling. Recall that we impose a constraint and use a joint prior (8) on $(z, z\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}})$ and have $\boldsymbol{\Omega}_{\boldsymbol{C}^{(s)}} \sim \text{GWis}(b^{(s)}, \boldsymbol{V}^{(s)})$. Let both $\boldsymbol{W}^{(r)}$ and $\boldsymbol{W}^{(s)}$ be $I \times J$ matrices, where $\operatorname{vec}(\boldsymbol{W}^{(r)}) = (\boldsymbol{I}_J \otimes \boldsymbol{\Omega}^{1/2}_{\boldsymbol{C}^{(s)}})\operatorname{vec}(\boldsymbol{U})$ and $\operatorname{vec}(\boldsymbol{W}^{(s)}) = (\boldsymbol{\Omega}^{1/2}_{\boldsymbol{C}^{(r)}} \otimes \boldsymbol{I}_I)\operatorname{vec}(\boldsymbol{U})$. Let $\boldsymbol{w}^{(r)}_j$ be the jth column vector of $\boldsymbol{W}^{(r)}$ and $\boldsymbol{w}^{(s)}_i$ be the ith row vector of $\boldsymbol{W}^{(s)}$. Then let $\boldsymbol{S}^{(r)}$ be $J \times J$ with $s^{(r)}_{jj'} = \boldsymbol{w}^{(r)T}_j \boldsymbol{w}^{(r)}_j$ and $\boldsymbol{S}^{(s)}$ be $I \times I$ with $s^{(s)}_{ii'} = \boldsymbol{w}^{(s)T}_i \boldsymbol{w}^{(s)}_i$. With these notations, we have

$$(z \mid \cdot) \sim Ga(a_z, b_z),$$

where $a_z = J(b-2)/2 + \nu(\boldsymbol{C}^{(r)})$ and $b_z = \text{tr}(\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}}\boldsymbol{V}^{(r)})/2$;

$$(\boldsymbol{\Omega}_{\boldsymbol{C}^{(r)}} \mid \cdot) \sim \text{GWis}(b^{(r)} + I, z\boldsymbol{V}^{(r)} + \boldsymbol{S}^{(r)})$$

and

$$(\boldsymbol{\Omega}_{\boldsymbol{C}^{(s)}} \mid \cdot) \sim \text{GWis}(b^{(s)} + J, \boldsymbol{V}^{(s)} + \boldsymbol{S}^{(s)}).$$

The graph $\boldsymbol{C}^{(r)}$ is updated using (Wang & Li, 2012)'s partial analytic structure algorithm (p. 188, Algorithm 2).