GRAPHICAL MODELING OF SPATIAL HEALTH DATA

ADRIAN DOBRA

ABSTRACT. The literature on Gaussian graphical models (GGMs) contains two equally rich and equally significant domains of research efforts and interests. The first research domain relates to the problem of graph determination. That is, the underlying graph is unknown and needs to be inferred from the data. The second research domain dominates the applications in spatial epidemiology. In this context GGMs are typically referred to as Gaussian Markov random fields (GMRFs). Here the underlying graph is assumed to be known: the vertices correspond to geographical areas, while the edges are associated with areas that are considered to be neighbors of each other (e.g., if they share a border). We introduce multi-way Gaussian graphical models that unify the statistical approaches to inference for spatiotemporal epidemiology with the literature on general GGMs. The novelty of the proposed work consists of the addition of the G-Wishart distribution to the substantial collection of statistical tools used to model multivariate areal data. As opposed to fixed graphs that describe geography, there is an inherent uncertainty related to graph determination across the other dimensions of the data. Our new class of methods for spatial epidemiology allow the simultaneous use of GGMs to represent known spatial dependencies and to determine unknown dependencies in the other dimensions of the data.

KEYWORDS: Gaussian graphical models, Gaussian Markov random fields, spatiotemporal multivariate models

CONTENTS

1. Introduction 1
2. GGMs vs. GMRFs 2
3. Multi-way Gaussian Graphical Models 3
4. Inference in Multi-way GGMs 5
5. Multi-way GGMs with Separable Mean Parameters 6
6. An Application to Spatiotemporal Cancer Mortality Surveillance 7
7. Dynamic Multi-way GGMs for Array-variate Time Series 9
8. Discussion 11
References 12

1. Introduction

Graphical models [62, 40] that encode multivariate independence and conditional independence relationships among observed variables \( X = (X_1, \ldots, X_p) \) have a widespread use in major scientific areas (e.g., biomedical and social sciences). In particular, a Gaussian graphical model (GGM) is obtained by setting off-diagonal elements of the precision matrix \( K = \Sigma^{-1} \) to zero of a \( p \)-dimensional multivariate normal model [22]. Employing a GGM instead of a multivariate normal model leads to a significant reduction in the number of parameters that need to be estimated if most elements of \( K \) are constrained to be zero and \( p \) is large. A pattern of zero constraints in \( K \) can be recorded as an undirected graph \( G = (V, E) \) where the set of vertices \( V = \{1, 2, \ldots, p\} \) represent observed variables, while the set of edges \( E \subset V \times V \) link all the pairs of vertices that correspond to off-diagonal elements of \( K \) that have not been set to zero. The absence of an edge between \( X_{v_1} \) and \( X_{v_2} \) corresponds with the conditional independence of these two random variables given the rest and is denoted by \( X_{v_1} \perp X_{v_2} \mid X_{V \setminus \{v_1,v_2\}} \) [60]. This is called the pairwise Markov property relative to \( G \), which in turn implies the local as well as the global Markov properties relative to \( G \) [40]. The local Markov property plays a key role since it gives the regression model induced by \( G \) on each variable \( X_v \). More explicitly, consider the neighbors of \( v \) in \( G \), that is, the set of vertices \( v' \in V \) such that \( (v,v') \in E \). We denote this set by \( \text{bd}_G(v) \). The local Markov property
Inference can be performed based on the best model, i.e. the graph having the highest posterior probability. The local Markov property leads to a further reduction in the set of possible graphs using Markov chain Monte Carlo (MCMC) approaches. As the number of graph increases, MCMC methods are likely to visit only subsets of graphs that have high posterior probabilities. To this end, various papers have proposed stochastic search methods for fast identification of these high posterior probability graphs.

The second research domain on GGMs dominates the applications in spatial epidemiology. In this context GGMs are referred to as Gaussian Markov random fields (GMRFs) [6, 7, 8]. The underlying model based on three-way GGMs, and also present a related theoretical extension of multi-way GGMs to dynamic multi-way GGMs for array-variate time series.

We consider a GGM defined by a graph $G = (V, E)$, where $V$ is the set of vertices and $E$ is the set of edges. The precision matrix $K$ is constrained to belong to the cone $P_G$ of positive definite matrices such that $K_{ij} = 0$ for all $(i, j) \notin E$. The full conditionals are used to describe a full joint distribution of random spatial effects. GMRFs are conditional autoregressions (CAR) models that have a subclass called simultaneous autoregressions (SAR). For a comprehensive account of inference in CAR/SAR/GMRFs see [17, 51, 32].

Key questions relate to conditions in which a joint distribution determined by (1) actually exists and, if it does, whether it is multivariate normal. This leads to particular parametric specifications for the set of conditionals that are more restrictive than the general parametric specification of a GGM.

In this chapter we examine the theoretical differences between GGMs and GMRFs. Developed efficient MCMC methods for inference in univariate and matrix-variate GGMs, and subsequently employed these methods to construct Bayesian hierarchical spatial models for mapping multiple diseases. We extend their results to multi-way GGMs that can capture temporal dependencies in addition to several other relevant dimensions. We exemplify the construction of a Bayesian hierarchical spatiotemporal model based on three-way GGMs, and also present a related theoretical extension of multi-way GGMs to dynamic multi-way GGMs for array-variate time series.

2. GGMs vs. GMRFs

We consider a GGM defined by a graph $G = (V, E)$ for the multivariate normal distribution $N_p(0, K^{-1})$ of a vector $X = (X_1, \ldots, X_p)$. The precision matrix $K$ is constrained to belong to the cone $P_G$ of positive definite matrices such that $K_{ij} = 0$ for all $(i, j) \notin E$. The full conditionals associated with each $X_v$ are expressed as a function of the elements of $K$ as follows:

\begin{equation}
X_v \mid X_{V \setminus \{v\}} = x_{V \setminus \{v\}} \sim N(- \sum_{v' \in E_{bd}(v)} (K_{v'v} / K_{vv'}) x_{v'}, 1 / K_{vv'}). \tag{3}
\end{equation}

Remark that the variables $X_v$ that are not linked by an edge with $X_v$ are dropped from the full conditional (3) because $K_{v'v} = 0$. A GMRF with graph $G$ is parametrized through the full conditionals

\begin{equation}
X_v \mid X_{V \setminus \{v\}} = x_{V \setminus \{v\}} \sim N( \sum_{v' \in E_{bd}(v)} \beta_{v'v} x_{v'}, \sigma_v^2), \text{ for } v \in V. \tag{4}
\end{equation}
The symmetry condition $\beta_{vv'}\sigma_{v'}^2 = \beta_{v'v}\sigma_{v}^2$ for all $v \neq v'$ is necessary for the conditionals \(^4\) to define a proper GGM \(^4\). Additional constraints under which the set of regression parameters $(\beta_{vv'}, \sigma^2_v)$ induce a proper precision matrix $K \in P_G$ with $K_{vv} = \sigma_{v}^{-2}$ and $K_{v'v} = -\beta_{vv'}\sigma_{v}^{-2}$ are given in \(^8\), \(^9\) make use of a symmetric proximity matrix $W$ with $w_{vv'} = 0$, $w_{v'v} > 0$ if $v' \in \mathcal{B}_G(v)$ and $w_{vv'} = 0$ if $v' \notin \mathcal{B}_G(v)$. They define $\beta_{vv'} = \rho w_{vv'}/w_{v'v}$ and $\sigma_{v'}^2 = \sigma^2/v_{v'v}$, where $w_{v'v} = \sum_{v''} w_{v'v''}$. Here $\rho$ is referred to as a spatial autocorrelation parameter. With this choice, for each $\rho \in (-1, 1)$ and $\sigma_v^2 > 0$, the GMRF specified by the full conditionals \(^4\) has a precision $K = \sigma^2(E_W - \rho W)^{-1}$ in $P_G$, where $E_W = \text{diag}(w_{11}, \ldots, w_{pp})$. As such, this widely used parametrization of GMRFs is quite restrictive since not any matrix in the cone $P_G$ can be represented through the two parameters $\rho$ and $\sigma_v^2$ given a particular choice of $W$. This difficulty originates from the parametrization \(^4\) of a GGM. Instead, by imposing proper prior distributions for the precision matrix $K$, we avoid the unnecessary representation of a GGM as the set of full conditionals \(^4\). In particular, we use of the G-Wishart prior $\text{Wis}_G(\delta, D)$ with density

$$p(K \mid G, \delta, D) = \frac{1}{I_G(\delta, D)}(\det K)^{(\delta-2)/2} \exp[-\frac{1}{2}(K, D)],$$

with respect to the Lebesgue measure on $P_G$ \(^5\)\(^6\)\(^4\). Here $(A, B) = \text{tr}(A^T B)$ denotes the trace inner product. The normalizing constant $I_G(\delta, D)$ is finite provided $\delta > 2$ and $D$ positive definite \(^23\). The G-Wishart prior $\text{Wis}_G(\delta, D)$ is conjugate to the normal likelihood. For a thorough account of its numerical properties see \(^4\) and the references therein.

For applications in hierarchical spatial models, \(^26\) set $D = (\delta - 2)\sigma^2(E_W - \rho W)^{-1}$ because, with this choice, the prior mode for $K$ is precisely $\sigma^2(E_W - \rho W)^{-1}$ – the precision matrix of a GMRF. The prior specification for the precision matrix can therefore be completed in a manner similar to the current work from the existing literature on GMRFs. We note that the G-Wishart prior for $K$ induces compatible prior distributions for the regression parameters \(^4\) – see \(^24\). The advantage of this representation of GGMs is a more flexible framework for GMRFs that allows their regression coefficients to be determined from the data rather than being fixed or allowed to vary as a function of only two parameters.

3. Multi-way Gaussian Graphical Models

We develop a framework for analyzing datasets that are associated with a random $L$-dimensional array $X$. Such datasets are quite common in social and biomedical sciences. In particular, spatial epidemiology involves datasets recording SIRs of several diseases observed under different conditions at multiple time points. The notations, definitions and operators related to tensors that appear throughout are introduced in \(^20\)\(^39\). The elements of the observed multi-way array are indexed by $\{i_1, i_2, \ldots, i_L\} : 1 \leq i_l \leq m_l$. The total number of elements of $X$ is $m = \prod_{l=1}^L m_l$. We assume that $X$ follows an array normal distribution

$$\text{vec}(X) \mid K \sim \mathcal{N}_m(0, K^{-1}),$$

whose $m \times m$ precision matrix $K$ is separable across each dimension, i.e.

$$K = K_L \otimes K_{L-1} \otimes \ldots \otimes K_1.$$

The $m_l \times m_l$ precision matrix $K_l$ is associated with dimension $l$, while vec$(X)$ is the vectorized version of $X$. The separability assumption might seem restrictive in the sense that it captures only dependencies across each dimension of the data without directly taking into account the interactions that might exist among two, three or more dimensions. However, this assumption reduces the number of parameters of the distribution of vec$(X)$ from $2^{L-1}m(m+1)$ to $2^{L-1} \prod_{l=1}^L m_l(m_l+1)$ which constitutes a substantial advantage when a sample size is small. The probability density of $X$ as an array is (see \(^36\))

$$p(X \mid K_1, \ldots, K_L) = (2\pi)^{-\frac{L}{2}} \left[ \prod_{l=1}^L (\det K_l)^{\frac{1}{2m_l}} \right]^{\frac{m_l}{2}} \exp \left\{ -\frac{1}{2} (X, X \times \{K_1, \ldots, K_L\}) \right\},$$

where

$$X \times \{K_1, \ldots, K_L\} = X \times_1 K_1 \times_2 \ldots \times_L K_L,$$

is the Tucker product. Here $X \times_l K_l$ is the $l$-mode product of the tensor $X$ and matrix $K_l$. We refer to \(^7\) as the mean-zero $L$-dimensional array normal distribution $\text{AN}_L(0; \{m_1, K_1\}, \ldots, \{m_L, K_L\})$. 

3
Most of the existent literature has focused on two-dimensional (or matrix-variate) arrays – see, for example, [3, 48] and the references therein. [31] studies the separable normal model (6) for \( L = 3 \), while [19] presents theoretical results for matrix-variate distributions that includes (6) with \( L = 2 \) as a particular case. [36] has proposed a Bayesian inference framework for model (6) for an arbitrary number \( L \) of dimensions by assigning independent inverse-Wishart priors for the covariance matrices \( \Sigma_i = K_i^{-1} \) associated with each dimension. Despite its flexibility and generality, the framework of [36] does not allow any further reduction in the number of parameters of model (6). To this end, we propose a framework in which each precision matrix \( K_l \) is constrained to belong to a cone \( P_{G_l} \) associated with a GGM with graph \( G_l \in \mathcal{G}_m \). We denote by \( \mathcal{G}_m \) the set of undirected graphs with \( m_l \) vertices. Sparse graphs associated with each dimension lead to sparse precision matrices, hence the number of parameters that need to be estimated could be significantly smaller than \( 2^{-L} \prod_{l=1}^{L} m_l(m_l+1) \). A similar framework has been proposed in [59] for matrix-variate data (\( L = 2 \)) and for row and column graphs restricted to the class of decomposable graphs. Our framework is applicable for any number of dimensions and allows arbitrary graphs (decomposable and non-decomposable) to be associated with each precision matrix \( K_l \).

The prior specification for \( (K_l)_{l=1}^{L} \) must take into account the fact that two precision matrices are not uniquely identified from their Kronecker product which means that, for any \( z > 0 \) and \( l_1 \neq l_2 \),

\[
K_{l_1} \otimes \cdots \otimes (z^{-1} K_{l_1}) \otimes \cdots \otimes (z K_{l_2}) \otimes \cdots \otimes K_{l_1} = K_{l_2} \otimes \cdots \otimes K_{l_1} \otimes \cdots \otimes K_{l_2} \otimes \cdots \otimes K_{l_1}
\]

represents the same precision matrix for \( \text{vec}(X) \). We follow the basic idea laid out in [59] and impose the constraints

\[
(K_{l})_{11} = 1, \quad \text{for } l = 2, \ldots, L.
\]

Furthermore, we define a prior for \( K_l, l \geq 2 \), through parameter expansion by assuming a G-Wishart prior \( \text{Wis}_{G_l}(\delta_l, D_l) \) for the matrix \( z_i K_l \) with \( \delta_l > 0, D_l > 2 \) and \( D_l \in P_{G_l} \). We denote \( G_l = (V_l, E_l) \), where \( V_l = \{1, 2, \ldots, m_l\} \) are vertices and \( E_l \subset V_l \times V_l \) are edges. We consider the Cholesky decompositions of the precision matrices from [6],

\[
K_l = \phi_l^T \phi_l,
\]

where \( \phi_l \) is an upper triangular matrix with \( (\phi_l)_{ii} > 0, 1 \leq i \leq m_l \). [50] proves that the set \( v(G_l) \) of the free elements of \( \phi_l \) consists of the diagonal elements together with the elements that correspond with the edges of \( G_l \), i.e. \( v(G_l) = \{ (i, i) : i \in V_l \} \cup \{ (i, j) : i < j \text{ and } (i, j) \in E_l \} \).

Once the free elements of \( \phi_l \) are determined, the remaining elements are also known. Specifically, \( (\phi_l)_{ij} = 0 \) if \( j \geq 2 \) and \( (1, j) \notin E_l \). We also have

\[
(\phi_l)_{ij} = -\frac{1}{(\phi_l)_{ii}} \sum_{k=1}^{i-1} (\phi_l)_{ki} (\phi_l)_{kj},
\]

for \( 2 \leq i < j \) and \( (i, j) \notin E_l \). The determination of the elements of \( \phi_l \) that are not free based on the elements of \( \phi_l \) that are free is called the completion of \( \phi_l \) with respect to \( G_l \) [50, 4]. It is useful to remark that the free elements of \( \phi_l \) fully determine the matrix \( K_l \). The development of our framework involves the Jacobian of the transformation that maps \( K_l \in P_{G_l} \) to the free elements of \( \phi_l \) [50]:

\[
J(K_l \rightarrow \phi_l) = 2^{m_l} \prod_{i=1}^{m_l} (\phi_l)_{ii}^{d_{G_l}^i - 1},
\]

where \( d_{G_l}^i \) is the number of elements in \( \text{bd}_{G_l}(i) \cap \{i+1, \ldots, m_l\} \) and \( \text{bd}_{G_l}(i) = \{j : (i, j) \in E_l\} \) is the boundary of vertex \( i \) in \( G_l \).

Our proposed prior specification for the separable normal model (7) is

\[
K_1 | \delta_1, D_1 \sim \text{Wis}_{G_1}(\delta_1, D_1), \quad (zK_l) | \delta_l, D_l \sim \text{Wis}_{G_l}(\delta_l, D_l), \quad \text{for } l = 2, \ldots, L.
\]

The prior for \( K_1 \) is

\[
p(K_1 | G_1) = \frac{1}{I_{G_1}(\delta_1, D_1)} (\det K_1)^{\delta_1^2} \exp \left\{ -\frac{1}{2} K_1, D_1 \right\}.
\]
while the joint prior for \((z_l, K_l)\) is

\[
p(z_l, K_l | G_l) = \frac{1}{I_{G_l}(\delta_l, D_l)} (\det K_l)^\frac{\delta_l-2}{2} \exp \left\{-\frac{1}{2} (K_l, z_l D_l) \right\} z_l^{\frac{m(d_l-2)+|\nu(G_l)|-1}{2}},
\]

for \(l = 2, \ldots, L\).

4. Inference in Multi-way GGMs

We assume that the observed samples \(D = \{x^{(1)}, \ldots, x^{(n)}\}\) are independently generated from the mean-zero array normal distribution \(\mathcal{AN}_n(0; \{m_1, K_1\}, \ldots, \{m_L, K_L\})\). The resulting likelihood is expressed by introducing an additional dimension \(m_{L+1} = n\) with precision matrix \(K_{L+1} = I_n\), where \(I_n\) is the \(n \times n\) identity matrix. We see \(D\) as a \(m_1 \times \ldots \times m_{L+1}\) array that follows an array normal distribution \(\mathcal{AN}_{n+1}(0; \{m_1, K_1\}, \ldots, \{m_{L+1}, K_{L+1}\})\). Furthermore, we define \(\phi_{L+1} = I_n\). The Cholesky decompositions \((9)\) of the precision matrices \(K_l\) give the following form of the likelihood of \(D\):

\[
p(D | K_1, \ldots, K_L) \propto \left( \prod_{l=1}^L (\det K_l) \frac{1}{2} \right) \exp \left\{-\frac{1}{2} ||D \times \{\phi_1, \ldots, \phi_{L+1}\}|| \right\},
\]

where \(||Y|| = \langle Y, Y \rangle\) is the array norm \([39]\). Simple calculations show that the part of the likelihood \((13)\) that depends on the precision matrix \(K_l\) is written as:

\[
p(D | K_1, \ldots, K_L) \propto (\det K_l)^\frac{m}{2} \exp \left\{-\frac{1}{2} (K_l, S_l) \right\},
\]

where \(S_l = D^{(l)}(D^{(l)})^T\) and

\[D^{(l)} = D \times_1 \phi_1 \times_2 \ldots \times_{l-1} \phi_{l-1} \times_l I_l \times_{l+1} \phi_{l+1} \times_{l+2} \ldots \times_{L+1} \phi_{L+1}.
\]

Here \(Y^{(l)}\) is the \(l\)-mode matricization of an array \(Y\) \([39]\). We develop a Markov chain Monte Carlo sampler from the posterior distribution of precision matrices \(K_l \in \mathcal{P}_{G_l}\), graphs \(G_l \in \mathcal{G}_{m_{l+1}}\) and auxiliary variables \(z_l\) for \(1 \leq l \leq L\):

\[
p(K_1, (K_l, G_l, z_l)_{l=2}^L | D) \propto p(D | K_1, \ldots, K_L) p(K_1 | G_1) \prod_{l=2}^L p(z_l, K_l | G_l) \prod_{l=1}^L \pi_{m_l}(G_l)
\]

Here \(\pi_{m_l}(G_l)\) are prior probabilities on the set of graphs \(\mathcal{G}_{m_l}\). The full conditionals of \(K_l\), \(1 \leq l \leq L\) and \(z_l\), \(2 \leq l \leq L\) are \(G\)-Wishart and Gamma, respectively:

\[
p(K_l | \text{rest}) = \text{Wish}_{G_l} \left( \frac{mn}{m_l} + \delta_l, S_l + z_l D_l \right),
\]

\[
p(z_l | \text{rest}) = \text{Gamma} \left( \frac{m(d_l-2)}{2} + |\nu(G_l)|, \frac{1}{2} (K_l, D_l) \right),
\]

where Gamma(\(\alpha, \beta\)) has mean \(\alpha/\beta\). We use the approach for updating \(K_1\) \((1 \leq l \leq L)\) described in \([25]\). Their method sequentially perturbs each free element in the Cholesky decomposition of each precision matrix. The constraint \((8)\) is imposed by not updating the free element \((\phi_l)_{11} = \sqrt{\langle K_l \rangle_{11}} = 1\).

The updates of the graphs \(G_l\) are based on the full joint conditionals of \(K_l\) and \(G_l\), \(1 \leq l \leq L\):

\[
p(K_l, G_l | \text{rest}) \propto \frac{1}{I_{G_l}(\delta_l, D_l)} (\det K_l)^{\frac{m(d_l-2)}{2}} z_l^{\frac{m(d_l-2)+|\nu(G_l)|-1}{2}} \exp \left\{-\frac{1}{2} (K_l, S_l + z_l D_l) \right\},
\]

since, once an edge in \(G_l\) is added or deleted, the corresponding set of free elements of \(K_l\) together with the remaining bound elements must also be updated.

We denote by \(\text{nb}d^{+}_{m_l}(G_l)\) the graphs that can be obtained by adding an edge to a graph \(G_l \in \mathcal{G}_{m_l}\) and by \(\text{nb}d^{-}_{m_l}(G_l)\) the graphs that are obtained by deleting an edge from \(G_l\). We call the one-edge-way set of graphs \(\text{nb}d_{m_l}(G_l) = \text{nb}d^{+}_{m_l}(G_l) \cup \text{nb}d^{-}_{m_l}(G_l)\) the neighborhood of \(G_l\) in \(\mathcal{G}_{m_l}\). These neighborhoods connect
any two graphs in $\mathcal{G}_m$, through a sequence of graphs such that two consecutive graphs in this sequence are each others’ neighbors. We sample a candidate graph $G'_i \in \text{nbhd}_m(G_i)$ from the proposal distribution:

$$q(G'_i \mid G_i, z_i) = \frac{1}{2} \frac{Z_j^{\mid v(G'_i)\mid}}{Z_j^{\mid v(G_i)\mid}} \delta(G'_i \in \text{nbhd}_m(G_i)) + \frac{1}{2} \frac{Z_j^{\mid v(G'_i)\mid}}{Z_j^{\mid v(G_i)\mid}} \delta(G'_i \in \text{nbhd}_m(G_i)),$$

where $\delta_A$ is equal to 1 if $A$ is true and is 0 otherwise. The proposal (16) gives an equal probability that the candidate graph is obtained by adding or deleting an edge from the current graph $G_i$.

We assume that the candidate graph $G'_i$ is obtained by adding an edge $(v_1, v_2)$, $v_1 < v_2$, to $G_i$. We have $v(G'_i) = v(G_i) \cup \{v_1, v_2\}$, $\text{bd}_G(v_1) = \text{bd}_G(v_1) \cup \{v_2\}$ and $d_{v_1}^{G_i} = d_{v_1}^{G'_i} + 1$. We define an upper diagonal matrix $\phi'_i$ such that $(\phi'_i)_{v'_i, v'_2} = (\phi_i)_{v'_i, v'_2}$ for all $(v'_i, v'_2) \in v(G_i)$. The value of $(\phi'_i)_{v_1, v_2}$ is sampled from a $N\left((\phi_i)_{v_1, v_2}, \sigma^2_x\right)$ distribution. The bound elements of $\phi'_i$ are determined through completion with respect to $G'_i$. We form the candidate matrix $K'_i = (\phi'_i)^T \phi'_i \in \mathbb{P}_{G'_i}$. Since the dimensionality of the parameter space increases by one, we must make use of the reversible jump method of [35]. We accept the update of $(K_i, G_i)$ to $(K'_i, G'_i)$ with probability $\min[R_g, 1]$, where

$$R_g = \frac{p(K'_i, G'_i \mid \text{rest}) \, q(G'_i \mid G_i, z_i) \, J\left(K'_i \rightarrow \phi'_i\right) \, \pi_m(G'_i)}{p(K_i, G_i \mid \text{rest}) \, q(G'_i \mid G_i, z_i) \, J(K_i \rightarrow \phi_i) \, \pi_m(G_i)} \times \exp\left(-\frac{(\phi'_i)_{v_1, v_2} - (\phi_i)_{v_1, v_2}}{2\sigma_x^2}\right).$$

Since the free elements of $\phi'_i$ are the free elements of $\phi_i$ and $K'_i$, the Jacobian of the transformation from $\phi_i$ to $\phi'_i$ is equal to 1. Moreover $\det K'_i = \det K_i$ and $J(K'_i \rightarrow \phi'_i) = J(\phi_i \rightarrow \phi'_i)$. It follows that

$$R_g = \sigma_x \sqrt{2\pi(\phi_i)_{v_1, v_2}} \frac{I_{G_i}(\delta_i, D_i)}{I_{\tilde{G}_i}(\delta_i, D_i)} \frac{q(G'_i \mid G_i, z_i) \, \pi_m(G'_i)}{q(G_i \mid G_i, z_i) \, \pi_m(G_i)} \times \exp\left(-\frac{1}{2}(K'_i - K_i, S_i + z_i D_i) + \frac{(\phi'_i)_{v_1, v_2} - (\phi_i)_{v_1, v_2}}{2\sigma_x^2}\right).$$

Next we assume that $G'_i$ is obtained by deleting the edge $(v_1, v_2)$ from $G_i$. We have $v(G'_i) = v(G_i) \setminus \{v_1, v_2\}$, $\text{bd}_G(v_1) = \text{bd}_G(v_1) \setminus \{v_2\}$ and $d_{v_1}^{G_i} = d_{v_1}^{G'_i} - 1$. We define an upper diagonal matrix $\phi'_i$ such that $(\phi'_i)_{v'_i, v'_2} = (\phi_i)_{v'_i, v'_2}$ for all $(v'_i, v'_2) \in v(G_i)$. The bound elements of $\phi'_i$ are obtained by completion with respect to $G'_i$. The candidate precision matrix is $K'_i = (\phi'_i)^T \phi'_i \in \mathbb{P}_{G'_i}$. Since the dimensionality of the parameter space decreases by 1, the acceptance probability of the update of $(K_i, G_i)$ to $(K'_i, G'_i)$ is $\min[R'_g, 1]$, where

$$R'_g = (\sigma_x \sqrt{2\pi(\phi_i)_{v_1, v_2}})^{-1} \frac{I_{G_i}(\delta_i, D_i)}{I_{\tilde{G}_i}(\delta_i, D_i)} \frac{q(G'_i \mid G_i, z_i) \, \pi_m(G'_i)}{q(G_i \mid G_i, z_i) \, \pi_m(G_i)} \times \exp\left(-\frac{1}{2}(K'_i - K_i, S_i + z_i D_i) - \frac{(\phi'_i)_{v_1, v_2} - (\phi_i)_{v_1, v_2}}{2\sigma_x^2}\right).$$

5. Multi-way GGMs with Separable Mean Parameters

So far we have discussed multi-way GGMs associated with array normal distributions with a $m_1 \times \ldots \times m_L$ array mean parameter $\mathbf{M}$ assumed to be zero. In some practical applications this assumption is too restrictive and $\mathbf{M}$ needs to be explicitly accounted for. The observed samples $\mathcal{D} = \{x^{(1)}, \ldots, x^{(m)}\}$ grouped as an $m_1 \times \ldots \times m_L \times n$ array are modeled as

$$\mathcal{D} = \mathbf{M} \circ \mathbf{1}_n + \mathbf{X}, \quad \mathbf{X} \sim \text{AN}_{L+1}(\mathbf{0}; \{m_1, \mathbf{K}_1\}, \ldots, \{m_L, \mathbf{K}_L\}, \{n, \mathbf{K}_{L+1}\}), \quad \mathbf{K}_{L+1} = \mathbf{1}_n.$$
If the sample size $n$ is small or if the observed samples are not independent and their dependence structure is represented by removing the constraint $K_{l+1} = I_{s}$, estimating $m = \prod_{l=1}^{L} m_l$ mean parameters is unrealistic. The matrix-variate normal models of [2] have separable mean parameters defined by row and column means, while [2] extends separable means for general array data. The $L$-dimensional mean array $M$ is written as a sum of distinct mean arrays associated with each dimension:

$$M = \sum_{l=1}^{L} M_l,$$

where $M_l = 1_{m_1} \circ \ldots \circ 1_{m_{l-1}} \circ \mu_l \circ 1_{m_{l+1}} \circ \ldots \circ 1_{m_L}$, $1 \leq l \leq L$.

Here $\mu_l \in \mathbb{R}^{m_l}$ represents the mean vector associated with dimension $l$ of $X$. This particular structure of the mean array $M$ implies the following marginal distribution for each element of the array of random effects $X$:

$$X_{i_1 \ldots i_l i_{l+1}} \sim N\left(\sum_{l=1}^{L} (\mu_l)_{i_l}, \prod_{l=1}^{L} (K_l^{-1})_{i_l i_l'}\right).$$

Thus $\mu_l$ can be interpreted as fixed effects associated with dimension $l$. The dependency between two different elements of $X$ is represented by their covariance

$$\text{Cov}(X_{i_1 \ldots i_l i_{l+1}}, X_{i_1' \ldots i_l' i_{l+1}'}) = \prod_{l=1}^{L} (K_l^{-1})_{i_l i_l'}.$$

We remark that the individual mean arrays $M_l$ are not identified, but their sum $M$ is identified.

Bayesian estimation of $M$ proceeds by specifying independent priors $\mu_l \sim N_{m_l}(\mu_0, \Omega_l^{-1})$. To simplify the notations we take $n = 1$, hence the arrays in equation (17) have only $L$ dimensions. We develop a Gibbs sampler in which each vector $\mu_l$ is updated as follows. Denote $m_{-l} = \prod_{l \neq m_{l}}$ and consider the $l$-mode matricizations $D_{l}, M_l, T_{l}, M$ and $X$. From equation (17) it follows that the $m_l \times m_{-l}$ random matrix

$$\tilde{X}_{l} = X_{l} - \sum_{l \neq m_{l}} (M_{l})_{l},$$

follows a matrix-variate normal distribution with mean $(M_{l})_{l} = \mu_1^{T} 1_{m_{-l}}$, row precision matrix $K_{l}$ and column precision matrix $K_{l} = K_{l} \otimes \ldots \otimes K_{l} \otimes K_{l} \ldots \otimes K_{l}$. It follows that $\mu_l$ is updated by direct sampling from the multivariate normal $N_{m_l}(\mu_l, K_l^{-1})$ where

$$K_{l} = \left(1_{m_1}^{T} 1_{m_{-1}} 1_{m_2} \right) K_l + m_{-l} \Omega_l, \quad m_l = K_{l}^{-1} \left[1_{l} \tilde{X}_{l} 1_{l} m_{-l} \Omega l + m_{-l} \Omega l \right].$$

6. An Application to Spatiotemporal Cancer Mortality Surveillance

In this section we construct a spatial hierarchical model for spatiotemporal cancer mortality surveillance based on the multi-way GGMs just developed. A relevant dataset could comprise counts $y_{i, j}^{*}$ for the number of deaths from cancer $i$ in area $j$ on year $t$, and can be seen as a three-dimensional array of size $m_C \times m_S \times m_T$. Our proposed model accounts for temporal and spatial dependence in mortality counts, as well as dependence across cancer types:

$$y_{iC, jS, iT} | \theta_{iC, jS, iT} \sim \text{Poi}\left(\exp\left(\mu_{iC} + \log(h_{iS, iT}) + \theta_{iC, jS, iT}\right)\right).$$

Here $iC = 1, \ldots, m_C$, $iS = 1, \ldots, m_S$, $iT = 1, \ldots, m_T$, $h_{iS, iT}$ denotes the population in area $iS$ during year $iT$, $\mu_{iC}$ is the mean number of deaths due to cancer $iC$ over the whole period and all locations, and $\theta_{iC, jS, iT}$ is a zero-mean random effect, which is assigned the prior:

$$\Theta = (\theta_{iC, jS, iT}) \sim \text{AN}_{3}(0; \{m_C, K_{C}\}, \{m_S, K_{S}\}, \{m_T, K_{T}\}).$$

The matrix $K_C$ models dependence across cancer types, $K_S$ accounts for spatial dependence across neighboring areas, and $K_T$ accounts for temporal dependence. Since we do not have prior information about dependence across cancer types, the prior for $K_C$ is specified hierarchically by setting

$$K_C | G_C \sim \text{Wish}(\delta_C, I_{m_C}), \quad p(G_C) \propto 1.$$
Thus $G_C \in G_{mc}$ defines the unknown graphical model of cancer types. For the spatial component, we follow the approach of [26] and use a GGM to specify a conditionally autoregressive prior

$$(z_S K_S) \mid G_S \sim \text{Wis}_{G_S} (\delta_S, (\delta_S - 2) D_S),$$

where $D_S = (E_W - \rho W)^{-1}$ and $W$ is the adjacency matrix for the $m_S$ areas, so that $W_{i_S j_S} = 1$ if areas $i_S$ and $j_S$ share a common border, and $W_{i_S j_S} = 0$ otherwise, and $E_W = \text{diag}(I_{m_S} W)$. The graph $G_S$ is fixed and given by the adjacency matrix $W$. Furthermore, we assume that, a priori, there is a strong degree of positive spatial association, and choose a prior for spatial autocorrelation parameter $\rho$ that gives higher probabilities to values close to 1 (see [33]):

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

For the temporal component, the prior for $K_T$ is set to

$$(z_T K_T) \mid G_T \sim \text{Wis}_{G_T} (\delta_T, I_{m_T}).$$

The graph $G_T$ gives the temporal pattern of dependence and could be modeled in a manner similar to the graph $G_C$ for cancer types. Instead of allowing $G_T$ to be any graph with $m_T$ vertices, we can constrain it to belong to a restricted set of graphs, for example, the graphs $G_T^{(1)}, G_T^{(2)}, G_T^{(3)}$ and $G_T^{(4)}$ with vertices $\{1, 2, \ldots, m_T\}$ and edges $E_T^{(1)}, E_T^{(2)}, E_T^{(3)}$ and $E_T^{(4)}$, where

$$E_T^{(1)} = \{(i_T - 1, i_T) : 2 \leq i_T \leq m_T\},$$
$$E_T^{(2)} = E_T^{(1)} \cup \{(i_T - 2, i_T) : 3 \leq i_T \leq m_T\},$$
$$E_T^{(3)} = E_T^{(2)} \cup \{(i_T - 3, i_T) : 4 \leq i_T \leq m_T\},$$
$$E_T^{(4)} = E_T^{(3)} \cup \{(i_T - 4, i_T) : 5 \leq i_T \leq m_T\}.$$

These four graphs define AR(1), AR(2), AR(3) and AR(4) models. We set $\delta_C = \delta_S = \delta_T = 3$. We use a multivariate normal prior for the mean rates vector $\mu = (\mu_1, \ldots, \mu_{mc})^T \sim N_{mc}(\mu_0, \Omega^{-1})$ where $\mu_0$ is fixed and column precision matrix $\Omega = \omega^{-1} I_{mc}$. We set $\mu_0$ to be the median log incidence rate across all cancers, all areas and all time points, and $\omega$ to be twice the interquartile range of the raw log incidence rates.

The MCMC algorithm for this sparse multivariate spatiotemporal model involves iterative updates of the precision matrices $K_C$, $K_S$ and $K_T$ as well as of the graph $G_C$ as described in Section 4. The mean rates $\mu$ are sampled as described in Section 5. Here the three-dimensional mean parameter array $M$ is equal with the mean array associated with the first dimension (cancers), while the mean arrays associated with the other two dimensions (space and time) are set to zero:

$$M = \mu \odot I_{m_S} \odot I_{m_T}.$$
We make use of a Metropolis-Hastings step to sample from to sample from (18). We consider a strictly positive precision parameter \( \sigma \). For each \( i_C = 1, \ldots, m_C \), we update the \( i_C \)-th element of \( \Theta_{iA} \) by sampling \( \gamma \sim N(\tilde{\Theta}_{1:ic}, \tilde{\sigma}^2) \). We define a candidate row vector \( \tilde{\Theta}^{new}_{iA} \) by replacing \( \tilde{\Theta}_{1:ic} \) with \( \gamma \) in \( \tilde{\Theta}_{iA} \). We update the current \( i \)-th row of \( \Theta \) with \( \tilde{\Theta}^{new}_{iA} \) with the Metropolis-Hastings acceptance probability corresponding with (18). Otherwise the \( i \)-th row of \( \Theta \) remains unchanged.

7. Dynamic Multi-way GGMs for Array-variate Time Series

The cancer mortality surveillance application from Section 6 represented the time component as one of the dimensions of the three-dimensional array of observed counts. We give an extension of multi-way GGMs to array-variate time series \( Y_t, t = 1, 2, \ldots, T \), where \( Y_t \in \mathbb{R}^{m_1 \times \ldots \times m_L} \). Our framework generalizes the results from [14] [15] and [59] which assume vector (\( L = 1 \)) or matrix-variate (\( L = 2 \)) time series. We build on the standard specification of Bayesian dynamic linear models [61], and assume that \( Y_t \) is modeled over time by

\[
Y_t = \Theta_t \times_{L+1} F_t^T + \Psi_t, \quad \Psi_t \sim \text{AN}_j(0; \{m_1, v^{-1}_t K_1\}, \{m_2, K_2\}, \ldots, \{m_L, K_L\}),
\]

\[
\Theta_t = \Theta_{t-1} \times_{L+1} H_t + \Gamma_t, \quad \Gamma_t \sim \text{AN}_{L+1}(0; \{m_1, K_1\}, \ldots, \{m_L, K_L\}, \{s, W_t^{-1}\}),
\]

where (a) \( \Theta_t \in \mathbb{R}^{m_1 \times \ldots \times m_L \times s} \) is the state array at time \( t \); (b) \( F_t \in \mathbb{R}^s \) is a vector of known regressors at time \( t \); (c) \( \Psi_t \in \mathbb{R}^{m_1 \times \ldots \times m_L} \) is the array of observational errors at time \( t \); (d) \( H_t \) is a known \( s \times s \) state evolution matrix at time \( t \); (e) \( \Gamma_t \in \mathbb{R}^{m_1 \times \ldots \times m_L \times s} \) is the array of state evolution innovations at time \( t \); (f) \( W_t \) is the \( s \times s \) innovation covariance matrix at time \( t \); and (g) \( v_t > 0 \) is a known scale factor at time \( t \).

Furthermore, the observational errors \( \Psi_t \) and the state evolution errors \( \Gamma_t \), follow zero-mean array normal distributions defined by \( K_1, \ldots, K_L \) and \( W_t \), and are assumed to be both independent over time element-wise and mutually independent as sequences of arrays.

The observation equation (19) and the evolution equation (20) translate into the following dynamic linear model for the univariate time series \( (Y_t)_{i_1, i_2}, t = 1, 2, \ldots, T \):

\[
(Y_t)_{i_1, \ldots, i_L} = F_t^T \Theta_t i_{-L, \ast} + (\Psi_t)_{i_1, \ldots, i_L}, \quad \Psi_t \sim \text{AN}_j(0; \{m_1, v^{-1}_t K_1\}, \{m_2, K_2\}, \ldots, \{m_L, K_L\}),
\]

\[
(\Theta_t)_{i_1, \ldots, i_L, \ast} = H_t (\Theta_{t-1})_{i_1, \ldots, i_L, \ast} + (\Gamma_t)_{i_1, \ldots, i_L, \ast}, \quad \Gamma_t \sim \text{AN}_{L+1}(0; \{m_1, K_1\}, \ldots, \{m_L, K_L\}, \{s, W_t^{-1}\}),
\]

where \( (\Theta_t)_{i_1, \ldots, i_L, \ast} = ((\Theta_t)_{i_1, i_2, \ldots, i_L})_{i_{-L, \ast}} \) and \( (\Gamma_t)_{i_1, \ldots, i_L, \ast} \) are defined in a similar manner. The components \( F_t, H_t, W_t \) are the same for all univariate time series, but the state parameters \( (\Theta_t)_{i_1, \ldots, i_L} \) as well as their scales of measurement defined by \( \prod_{i=1}^{L} (K_i^{-1})_{ii} \) could be different across series. The cross-sectional dependence structure across individual time series at time \( t \) is induced by \( K_1, \ldots, K_L \) and \( W_t \):

\[
\text{Cov}((Y_t)_{i_1, \ldots, i_L}, (Y_{t'})_{i_1', \ldots, i_L'}) = v_t \prod_{i=1}^{L} (K_i^{-1})_{ii'}, \quad \text{Cov}((\Gamma_t)_{i_1, \ldots, i_L, \ast}, (\Gamma_{t'})_{i_1', \ldots, i_L', \ast}) = \prod_{i=1}^{L} (K_i^{-1})_{ii'} W_t.
\]

For example, if \( \prod_{i=1}^{L} (K_i^{-1})_{ii'} \) is large in absolute value, the univariate time series \( (Y_t)_{ii, ii} \) and \( (Y_{t'})_{ii', ii'} \) exhibit significant dependence in the variation of their observational errors and state vectors. Appropriate choices for the matrix sequence \( W_t, t = 1, 2, \ldots, T \), arise from the discount factors discussed in [61] as exemplified, among others, in [59]. The scale factors \( v_t \) can be set to 1, but other suitable values can be employed as needed.

The following result extends Theorem 1 of [59] to array-variate time series.

**Theorem 1.** Let \( \mathcal{D}_0 \) be the prior information and denote by \( \mathcal{D}_t = (Y_t, \mathcal{D}_{t-1}) \) the information available at time \( t = 1, 2, \ldots, T \). We assume to have specified precision matrices \( K_1, \ldots, K_L \), the matrix sequence \( W_t, t = 1, 2, \ldots, T \), as well as an initial prior for the state array at time 0,

\[
(\Theta_0 | \mathcal{D}_0) \sim \text{AN}_{L+1}(M_0; \{m_1, K_1\}, \ldots, \{m_L, K_L\}, \{s, C_0^{-1}\}),
\]
where $M_0 \in \mathbb{R}^{m_1 \times \ldots \times m_L \times s}$ and $C_0$ is an $s \times s$ covariance matrix. For every $t = 1, 2, \ldots, T$, the following distributional results hold: 

(i) posterior at $t - 1$:

$$\left( \Theta_{t-1} \mid D_{t-1} \right) \sim \text{AN}_{L+1}(M_{t-1}; \{m_1, K_1\}, \ldots, \{m_L, K_L\}, (s, C_{t-1}^{-1})),$$

(ii) prior at $t$:

$$\left( \Theta_t \mid D_{t-1} \right) \sim \text{AN}_{L+1}(a_t; \{m_1, K_1\}, \ldots, \{m_L, K_L\}, (s, R_t^{-1})),$$

where $a_t = M_{t-1} \times L+1 \ H_t$ and $R_t = H_t C_{t-1} H_t^T + W_t$.

(iii) one-step forecast at $t - 1$:

$$\left( Y_t \mid D_{t-1} \right) \sim \text{AN}_L(f_t; \{m_1, q_1^{-1} K_1\}, \ldots, \{m_L, K_L\})$$

where $f_t = M_{t-1} \times L+1 \ F_t^T H_t$. We use the filtering equations from Theorem 1 to produce the centered and scaled array data $\bar{Y}_t$.

(iv) posterior at $t$:

$$\left( \Theta_t \mid D_t \right) \sim \text{AN}_{L+1}(M_t; \{m_1, K_1\}, \ldots, \{m_L, K_L\}, (s, C_t^{-1}))$$

where $M_t = a_t + e_t \times L+1 A_{t-1}, C_t = R_t - A_{t-1} A_{t-1}^T q_t$. Here $A_t = q_t^{-1} R_t F_t$ and $e_t = Y_t - f_t$.

The proof of Theorem 1 is straightforward. We write equations (19) and (20) in matrix form:

$$\left( Y_t \right)_{(L+1)} \sim F_t^T \left( \Theta_t \right)_{(L+1)} + \left( \Psi_t \right)_{(L+1)}, \quad \left( \Psi_t \right)_{(L+1)} \sim \text{AN}_m \left( \theta, \nu, K^{-1} \right)$$

$$\left( \Theta_t \right)_{(L+1)} = H_t \left( \Theta_{t-1} \right)_{(L+1)} + \left( \Psi_t \right)_{(L+1)}, \quad \left( \Psi_t \right)_{(L+1)} \sim \text{AN}_m \left( \theta, \nu, K^{-1} \right)$$

where $m = \prod_{i=1}^L m_i$ and $K$ is given in equation (6). The normal theory results laid out in [61] apply directly to the dynamic linear model specified by equations (21) and (22). The predictive distributions relevant for forecasting and retrospective sampling for array-variate time series can be derived from the corresponding predictive distributions for vector data.

We complete the definition and prior specification for the dynamic multi-way GGMs with independent G-Wishart priors from equation (10) for the precision matrices $K_1, \ldots, K_L$ and their corresponding auxiliary variables $z_1, \ldots, z_L$. The graphs $G_1, \ldots, G_L$ associated with the G-Wishart priors receive independent priors $\pi_m(G_l)$ on $G_m$, $l = 1, \ldots, L$. Posterior inference in this framework can be achieved with the following MCMC algorithm that sequentially performs the following steps:

(A) Resampling the precision matrices, graphs and auxiliary variables. By marginalizing over the state arrays $\Theta_1, \ldots, \Theta_L$, we obtain the marginal likelihood [14, 15]:

$$p \left( Y_1, \ldots, Y_T \mid K_1, G_1, (K_l, G_l, z_l)_{l=2}^L \right) = \prod_{i=1}^T p \left( Y_i \mid D_{i-1}, K_1, G_1, (K_l, G_l, z_l)_{l=2}^L \right).$$

The one-step forecast distribution (iii) from Theorem 1 implies that

$$\left( q_t^{-1/2}(Y_t - f_t) \mid D_{t-1} \right) \sim \text{AN}_L(0; \{m_1, K_1\}, \ldots, \{m_L, K_L\}).$$

We use the filtering equations from Theorem 1 to produce the centered and scaled scalar data $\bar{D} = \{q_t^{-1/2}(Y_t - f_t) : t = 1, \ldots, T\}$. Since the elements of $\bar{D}$ are independent and identically distributed, we update each precision matrix $K_l$, graph $G_l$ and auxiliary variable $z_l$ as described in Section 4 based on $\bar{D}$.

(B) Resampling the state arrays. We employ the forward filtering backward algorithm (FFBS) proposed by [12, 30]. Given the current sampled precision matrices, we start by sampling $\Theta_t$ given $D_t$ from the posterior distribution given in (iv) of Theorem 1. Then, for $t = T - 1, T - 2, \ldots, 0$, we sample $\Theta_t$ given $D_t$ and $\Theta_{t+1}$ from the array normal distribution

$$\text{AN}_{L+1} \left( M_t^*; \{m_1, K_1\}, \ldots, \{m_L, K_L\}, (s, (C_{t+1}^{-1})^{-1}) \right),$$

where

$$M_t^* = M_t + (\Theta_{t+1} - a_{t+1}) \times L+1 \left( C_l G_{t+1}^T R_{t+1}^{-1} \right).$$
Recent advances in data collection techniques have allowed the creation of high-dimensional public health datasets that monitor the incidence of many diseases across several areas, time points and additional ecological sociodemographic groupings [28]. Jointly modeling the disease risk associated with each resulting cell count (i.e., a particular disease at a particular time point in a particular region given a particular combination of risk factors) is desirable since it takes into consideration interaction patterns that arise within each dimension or across dimensions. By aggregating data across time, key epidemiological issues related to the evolution of the risk patterns across time might not be given an appropriate answer [1]. The spatial structure of geographical regions must also be properly accounted for [6,9]. Furthermore, since diseases are potentially related and share risk factors, it is critical that individual models should not be developed for each disease [33,56]. Rich, flexible classes of models that capture the joint variation of disease risk in the actual observed data without requiring the aggregation across one or more dimensions will be the fundamental aim of our proposed work related to disease mapping. Multi-way GGMs can be used in Bayesian hierarchical models that produce estimates of disease risk by borrowing strength across time, areas and the other dimensions. Due to the likely presence of small counts in many cells, the degree of smoothing will be controlled through a wide range of parameters that could be constrained to zero according to pre-defined interaction structures (e.g., the neighborhood structure of the areas) or by graphs that received the most support given the data.

We generalize the models from Section [6] and let $Y$ be the $L$-dimensional array of observed disease counts indexed by cells $\{(i_1, \ldots, i_L) : 1 \leq i_l \leq m_l\}$. We assume that the count random variable $Y_{i_1, \ldots, i_L}$ associated with cell $(i_1, \ldots, i_L)$ follows a distribution from an exponential family (e.g., Poisson or binomial) with mean parameter $\theta_{i_1, \ldots, i_L}$, i.e.

$$
Y_{i_1, \ldots, i_L} \mid \theta_{i_1, \ldots, i_L} \overset{iid}{\sim} H(\theta_{i_1, \ldots, i_L}), \text{ for } 1 \leq i_l \leq m_l, 1 \leq l \leq L.
$$

We assume that the cell counts $Y_l$ are conditionally independent given the $L$-dimensional array of parameters $\theta = \{\theta_{i_1, \ldots, i_L} : 1 \leq i_l \leq m_l\}$. Furthermore, given a certain link function $g(\cdot)$ (e.g., log(·)), the parameters $\theta$ follow a joint model

$$
g(\theta_{i_1, \ldots, i_L}) = \nu_{i_1, \ldots, i_L} + X_{i_1, \ldots, i_L},
$$

where $\nu_{i_1, \ldots, i_L}$ is a known offset, while $X = \{X_{i_1, \ldots, i_L} : 1 \leq i_l \leq m_l\}$ is an array of zero-centered random effects. Equation (24) can subsequently include explanatory ecological covariates as needed. The multi-way GGMs are employed in the context of non-Gaussian data as joint distribution for the array of random effects $X$. Thus $X$ is assumed to follow the flexible joint distributions, and Each dimension of the data is represented as a GGM in a particular dimension of the random effects $X$.

This framework accommodates many types of interactions by restricting the set of graphs that are allowed to represent the dependency patterns of the corresponding dimensions. For example, if dimension $l'$ of $Z$ represents time, then the graphs associated with this dimension could be constrained to represent an autoregressive model $AR(q)$, where $q = 1, 2, 3, \ldots$ – see Section 6. Temporal dependence can also be modeled with the dynamic multi-way GGMs from Section 7. If dimension $l''$ represents spatial dependence, one could constrain the space of graphs for dimension $l''$ to consist of only one graph with edges defined by areas that are neighbors of each other in the spirit of [6,16]. As opposed to a modeling framework based on GMRFs, we can allow uncertainty around this neighborhood graph in which case we let the space of graphs for dimension $l''$ to include graphs that are obtained by adding or deleting one, two or more edges from the neighborhood graph. This expansion of the set of spatial graphs is consistent with the hypothesis that interaction occurs not only between areas that are close to each other or share a border, but also between more distant areas. We can also allow all possible graphs to be associated with dimension $l''$ and examine the graphs that receive the highest posterior probabilities. Such graphs can be further compared with the neighborhood graph to see whether the spatial dependency patterns in observed data are actually consistent with the geographical neighborhoods.

To gain further insight on the flexibility of our modeling approach, we examine the case in which a two-dimensional array $Y = \{Y_{ij}\}$ is observed with the first dimension associated with $m_1$ diseases and the second dimension associated with $m_2$ areas. Under the framework of [33,11] the matrix of counts
$Y$ is modeled with a hierarchical Poisson model with random effects distributed as a multivariate CAR (MCAR) model \cite{46}:

\begin{equation}
X' \sim N_{m_1m_2}(0, [K_1 \otimes (E_W - \rho W)]^{-1}).
\end{equation}

This structure of the random effects assumes separability of the association structure among diseases from the spatial structure \cite{55}. The spatial autocorrelation parameter $\rho$ is the only parameter that controls the strength of spatial dependencies, while the precision matrix $K_1$ is not subject to any additional constraints on its elements. In our framework, the random effects $X'$ follow a matrix-variate GGM prior obtained by taking $L = 2$ in equation (5). The same separability of the association structure is assumed, but the precision matrices $K_1$ and $K_2$ follow G-Wishart hyper-priors as in equation (10). The GGMs associated with the diseases are allowed to vary across all possible graphs with $m_1$ vertices, while the GGMs for the spatial structure can be modeled as we described earlier in this section.

Until recently, the application of GGMs with a G-Wishart prior for the precision matrix in large scale Bayesian hierarchical models has been hindered by computational difficulties. For decomposable graphs, the normalizing constant of the G-Wishart distribution is calculated with formulas \cite{50, 4}, and a direct sampler from this distribution existed for several years \cite{13}. But similar results did not exist for non-decomposable graphs. Fortunately, new methodological developments give formulas for the calculation of the G-Wishart distribution for arbitrary graphs \cite{54}, and also a direct sampler for arbitrary graphs \cite{41}. With these key results, the MCMC sampler developed in Section 4 can be significantly improved in its efficiency. The reversible jump algorithm that allows updates in the structure of the graphs associated the dimensions of a multi-way GGM can be subsequently refined to another transdimensional graph updating algorithm which bypasses the calculation of any normalizing constants of the G-Wishart distribution based on the double reversible jump algorithm of \cite{58, 41}. Moreover, the G-Wishart distribution can be replaced altogether in the specification of priors for Bayesian hierarchical spatial models with the graphical lasso prior of \cite{57}. The application of these new theoretical results to spatial health data is a very intense area of research.

References

\cite{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17}
[55] L. Waller and B. P. Carlin. Disease mapping. In A. E. Gelfand, P. J. Diggle, M. Fuentes, and P. Guttorp, editors, *Handbook of Spatial Statistics*, pages 217–243. CRC Press: Taylor & Francis Group, 2010.

[56] F. Wang and M. M. Wall. Generalized common factor spatial model. *Biostatistics*, 4:569–582, 2003.

[57] H. Wang. The Bayesian graphical lasso and efficient posterior computation. *Bayesian Analysis*, 7:771–790, 2012.

[58] H. Wang and S. Z. Li. Efficient gaussian graphical model determination under G-wishart prior distributions. *Electronic Journal of Statistics*, 6:168–198, 2012.

[59] H. Wang and M. West. Bayesian analysis of matrix normal graphical models. *Biometrika*, 96:821–834, 2009.

[60] N. Wermuth. Analogies between multiplicative models in contingency tables and covariance selection. *Biometrics*, 32:95–108, 1976.

[61] M. West and J. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer - Verlag, New York, second edition edition, 1997.

[62] J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. John Wiley & Sons, 1990.

[63] F. Wong, C. K. Carter, and R. Kohn. Efficient estimation of covariance selection models. *Biometrika*, 90:809–830, 2003.

[64] R. Yang and J. O. Berger. Estimation of a covariance matrix using the reference prior. *Ann. Statist.*, 22:1195–211, 1994.

[65] M. Yuan and Y. Lin. Model selection and estimation in the Gaussian graphical model. *Biometrika*, 94:19–35, 2007.

Department of Statistics, Department of Biobehavioral Nursing and Health Systems, Center for Statistics and the Social Sciences and Center for Studies in Demography and Ecology, University of Washington, Box 354322, Seattle, WA 98195

E-mail address: adobra@uw.edu