

# Order-free Coregionalized Areal Data Models with Application to Multiple Disease Mapping

XIAOPING JIN, SUDIPTO BANERJEE AND BRADLEY P. CARLIN <sup>1</sup>

*Division of Biostatistics, School of Public Health, University of Minnesota,  
Mayo Mail Code 303, Minneapolis, Minnesota 55455-0392, U.S.A.*

Correspondence author: Bradley P. Carlin

telephone: (612) 624-6646

fax: (612) 626-0660

email: brad@biostat.umn.edu

December 20, 2006

---

<sup>1</sup>Xiaoping Jin is Senior Statistician, Early Development Biostatistics, Global Technical Services, Daiichi Medical Research, Inc., 11 Philips Parkway, Montvale, NJ, 07645. Sudipto Banerjee is Assistant Professor of Biostatistics and Bradley P. Carlin is Professor of Biostatistics and Mayo Professor in Public Health at the Division of Biostatistics, School of Public Health, 420 Delaware St. S.E., University of Minnesota, Minneapolis, MN, 55455. The work of all three authors was supported in part by NIH grant 2-R01-ES07750.

# Order-free Coregionalized Areal Data Models with Application to Multiple Disease Mapping

## Summary

With the ready availability of spatial databases and Geographical Information System (GIS) software, statisticians are increasingly encountering multivariate modeling settings featuring associations of more than one type: spatial associations between data locations, and associations between the variables within the locations. While flexible modeling of multivariate point-referenced data have recently been addressed using a *linear model of coregionalization* (LMC; see e.g. Gelfand et al., 2004), existing methods for multivariate areal data (e.g. Kim et al., 2001; Gelfand and Vounatsou, 2003; Jin et al., 2005) typically suffer from unnecessary restrictions on the covariance structure or undesirable dependence on the conditioning order of the variables. In this paper we propose a class of Bayesian hierarchical models for multivariate areal data that avoids these restrictions, permitting flexible and order-free modeling of correlations both between variables and across areal units. Our framework encompasses a rich class of multivariate conditionally autoregressive (MCAR) models that are computationally feasible via modern Markov chain Monte Carlo (MCMC) methods. We illustrate the strengths of our approach over existing models using simulation studies, and also offer a real-data application involving annual lung, larynx, and esophagus cancer death rates in Minnesota counties between 1990 and 2000.

*Key words:* Lattice data; Linear model of coregionalization (LMC); Markov chain Monte Carlo (MCMC); Multivariate conditionally autoregressive (MCAR) model; Spatial statistics.

# 1 Introduction

The last decade has seen an explosion of interest in disease mapping, with recent developments in advanced spatial statistics and increasing availability of computerized Geographic Information System (GIS) technology. For example, the databases from the National Center for Health Statistics (NCHS) or from the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute, publicly available to anyone with a web browser, provide an enormous supply of georeferenced data.

Disease mapping is an epidemiological technique used to describe the geographic variation of disease and to generate etiological hypotheses about the possible causes for apparent differences in risk. Disease maps are used to highlight geographic areas with high and low incidence or mortality rates of a specific disease, and the variability of such rates over a spatial domain. They can also be used to detect spatial clusters which may be due to common environmental, demographical, or cultural effects shared by neighboring regions. However, mapping of crude incidence or mortality rates can be misleading when the population sizes for some of the units are small, resulting in large variability in the estimated rates, and making it difficult to distinguish chance variability from genuine differences. The correct geographic allocation of health care resources would be greatly enhanced by the development of statistical models that allow a more accurate depiction of “true” disease rates and their relation to explanatory variables (e.g. covariates).

For reasons of confidentiality or practicality, disease incidence or mortality data are often reported as counts or rates at a regional level (county, census tract, zip code, etc). Conditionally autoregressive (CAR) models have been widely used for disease mapping with such data. They allow us to borrow strength across regions by using not only the data from a given region, but also the data from neighboring regions. When we have multivariate areal (lattice) data (say, counts of  $p \geq 2$  diseases

over the same regions), an obvious first choice would be to use  $p$  separate univariate CAR models. But correlation across diseases may occur if they share the same set of (spatially distributed) risk factors, or are linked by etiology, a common risk factor, or an affected organ. Moreover, the presence of one disease might encourage or inhibit the presence of another over a region. A multivariate areal model can permit modeling of dependence among those diseases while maintaining spatial dependence between regions. Identifying similar patterns in geographical variation of related diseases in a multivariate way may provide more convincing evidence for any real clustering in the underlying risk than would be available from the analysis of any single disease separately.

Several multivariate areal models have been proposed to date, any of which could be applied to multiple disease mapping. The primary underlying challenge of multivariate areal modeling is to formulate valid probability models that account for association between different variables (e.g., diseases) *within* areal units along with the spatial association *between* areal units. While they provide valuable theoretical insight into joint modeling of areal data, current methods often fall short of offering a template that is at once versatile and practical. Mardia (1988) described the theoretical background for multivariate Gaussian Markov Random Field (MRF) specifications, extending the pioneering work of Besag (1974), but used *separable* models that force identical spatial smoothing for all variables. The “twofold CAR” model of Kim et al. (2001) offers richer spatial covariance structures for counts of two different diseases over each areal unit, but its extension to the case of more than two variables is unclear. Knorr-Held and Best (2001) developed a latent variable “shared component” model for bivariate disease mapping; here extension to more than two diseases is possible (Held et al., 2005), but can be awkward. Sain and Cressie (2002) discussed a multi-objective version of the conditional autoregressive model that allows for flexible modeling of the spatial dependence structure, the cross-correlations in particular, but may become computationally prohibitive. Carlin and Banerjee (2003) and Gelfand and Vounatsou (2003) developed essentially

equivalent multivariate conditionally autoregressive (MCAR) models for hierarchical modeling to include non-separable models, but left room for further generality in the covariance structures.

Adapting the multivariate point-level data approach of Royle and Berliner (1999), Jin et al. (2005) proposed a generalized multivariate conditional autoregressive (GMCAR) model for areal data that formulates the joint distribution for a multivariate MRF by specifying simpler conditional and marginal models. These models are computationally efficient and allow sufficient flexibility in the specifications of the spatial covariance structure. Indeed, many of the above models arise as special cases of the GMCAR. However, an inherent problem with these methods is that their conditional specification imposes a potentially arbitrary order on the variables being modeled, as they lead to different marginal distributions depending upon the conditioning sequence. This problem is somewhat ameliorated in certain (e.g., medical and environmental) contexts where a *natural* order is reasonable, but in many disease mapping contexts this is not the case. Although Jin et al. (2005) suggest using model comparison techniques to decide upon the proper modeling order, since all possible permutations of the variables would need to be considered this seems feasible only with relatively few variables. In any case, the principle of choosing among conditioning sequences using model comparison metrics is perhaps not uncontroversial.

In this paper we develop an order-free framework for multivariate areal modeling that allows versatile spatial structures, yet is computationally feasible for many variables. Our approach is based on a *linear model of coregionalization* (LMC) that has recently been proposed for multivariate point-referenced data (Wackernagel, 2003; Schmidt and Gelfand, 2003; Gelfand et al., 2004). Essentially, the idea is to develop richer spatial association models using linear transformations of much simpler spatial distributions. In this paper, we apply the LMC approach to the analysis of multivariate areal data, with an eye toward developing models for multiple disease mapping. In the process, we arrive at a very versatile framework, which encompasses a rich class of MCAR models (including most of

the existing models) as special cases. In particular, we consider modeling the annual mortality rates from lung, larynx, and esophageal cancer between 1990 and 2000 in Minnesota counties, a setting in which association would be expected both within and across the areal units.

The format of our paper is as follows. In Section 2, we briefly review the spatial modeling of a single disease and multiple diseases. In Section 3, we introduce new LMC-based ways of multivariate spatial modeling. Section 5 discusses the MCMC implementation of our proposed model, and evaluates our approach in terms of average mean square error (AMSE) and the Deviance Information Criteria (DIC) via simulation. Section 6 then illustrates our approach in the aforementioned multiple cancer data mapping setting. Finally, Section 7 summarizes our findings and suggests avenues for future research in this area.

## 2 Spatial modeling for disease mapping

Disease incidence or mortality data are often reported as counts or rates at a regional level (county, census tract, zip code, etc), and are called *areal* (or *lattice*) data. Markov random field (MRF) models for lattice data are based on the Markov property, where the conditional distribution of a site's response given the responses of all the other sites depends only on the observations in the neighborhood of this site. In this paper we define the neighborhood by area adjacency, although other definitions sometimes used (e.g., regions with centroids within a given fixed distance).

### 2.1 Spatial modeling of a single disease

Let  $Y_i$  be the observed number of cases of a certain disease in region  $i$ ,  $i = 1, \dots, n$ , and  $E_i$  be the expected number of cases in this same region. Here the  $Y_i$  are thought of as random variables, while the  $E_i$  are thought of as fixed and known (and are often simply taken as proportional to the number of persons at risk in the region). For rare diseases, a Poisson model approximation to a

binomial sampling distribution for disease counts is often used. Thus, a commonly used likelihood when mapping a single disease is

$$Y_i \stackrel{ind}{\sim} Poisson(E_i e^{\mu_i}), \quad i = 1, \dots, n, \quad (1)$$

where  $\mu_i = \mathbf{x}_i' \boldsymbol{\beta} + \phi_i$ . The  $\mathbf{x}_i$  are explanatory, region-level spatial covariates, having parameter coefficients  $\boldsymbol{\beta}$ . The parameter  $\mu_i$  represents the log-relative risk, estimates of which are often based on the departures of observed from expected counts. We place a form of Gaussian MRF model, commonly referred to as the conditionally autoregressive (CAR) prior, on the random effects  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)'$ , i.e.,

$$\boldsymbol{\phi} \sim N_n(\mathbf{0}, [\tau(D - \alpha W)]^{-1}), \quad (2)$$

where  $N_n$  denotes the  $n$ -dimensional normal distribution,  $D$  is a  $n \times n$  diagonal matrix with diagonal elements  $m_i$  that denote the number of neighbors of area  $i$ , and  $W$  is the adjacency matrix of the map (i.e.,  $W_{ii} = 0$ , and  $W_{i' i} = 1$  if  $i'$  is adjacent to  $i$  and 0 otherwise). In the joint distribution (2),  $\tau^{-1}$  is the spatial dispersion parameter, and  $\alpha$  is the spatial autocorrelation parameter. The CAR prior corresponds to the following conditional distribution of  $\phi_i$ :

$$\phi_i | \phi_j, j \neq i, \sim N\left(\frac{\alpha}{m_i} \sum_{i \sim j} \phi_j, \frac{1}{\tau m_i}\right), \quad i, j = 1, \dots, n, \quad (3)$$

where  $i \sim j$  denotes that region  $j$  is a *neighbor* (typically defined in terms of spatial adjacency) of region  $i$ . The CAR structure (2) reduces to the well-known intrinsic conditionally autoregressive (ICAR) model (Besag et al., 1991) if  $\alpha = 1$ , or an independence model if  $\alpha = 0$ . The ICAR model induces “local” smoothing by borrowing strength from the neighbors, while the independence model assumes independence of spatial rates and induces “global” smoothing. The smoothing parameter  $\alpha$

in the CAR prior (2) controls the strength of spatial dependence among regions, though it has long been appreciated that a fairly large  $\alpha$  may be required to deliver significant spatial correlation; see Wall (2004) for recent discussion and exemplification.

A similar approach proposes a Gaussian convolution prior for the modeling of the random effects  $\phi$ . The random effects  $\phi$  are assumed to be the sum of the two independent components with one having a Gaussian independence prior and the other a Gaussian ICAR prior (Besag et al., 1991). With such a convolution prior, we may capture both the relative contributions of region-wide heterogeneity and local clustering. Although this method has limitations (see e.g. Banerjee, Carlin, and Gelfand, 2004, pp.163–165), since the convolution process priors are among the most widely used we implement this model for our Minnesota cancer data analysis in Section 6.

## 2.2 Spatial modeling of multiple diseases

Now let  $Y_{ij}$  be the observed number of cases of disease  $j$  in region  $i$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, p$ , and let  $E_{ij}$  be the expected number of cases for the same disease in this same region. As in Section 2.1, the  $Y_{ij}$  are thought of as random variables, while the  $E_{ij}$  are thought of as fixed and known. For the first level of the hierarchical model, conditional on the random effects  $\phi_{ij}$ , we assume the  $Y_{ij}$  are independent of each other such that

$$Y_{ij} \stackrel{ind}{\sim} \text{Poisson}(E_{ij}e^{\mathbf{x}'_{ij}\boldsymbol{\beta}_j + \phi_{ij}}), \quad i = 1, \dots, n, \quad j = 1, \dots, p, \quad (4)$$

where the  $\mathbf{x}_{ij}$  are explanatory, region-level spatial covariates for disease  $j$  having (possibly region-specific) parameter coefficients  $\boldsymbol{\beta}_j$ .

Carlin and Banerjee (2003) and Gelfand and Vounatsou (2003) generalized the univariate CAR (2) to a joint model for the random effects  $\phi_{ij}$  under a separability assumption, which permits modeling

of correlation among the  $p$  diseases while maintaining spatial dependence across space. Separability assumes that the association structure separates into a non-spatial and spatial component. More precisely, the joint distribution of  $\boldsymbol{\phi}$  is assumed to be

$$\boldsymbol{\phi} \sim N_{np} \left( \mathbf{0}, [\Lambda \otimes (D - \alpha W)]^{-1} \right), \quad (5)$$

where  $\boldsymbol{\phi} = (\boldsymbol{\phi}'_1, \dots, \boldsymbol{\phi}'_p)'$ ,  $\boldsymbol{\phi}_j = (\phi_{1j}, \dots, \phi_{nj})'$ ,  $\Lambda$  is a  $p \times p$  positive definite matrix that is interpreted as the non-spatial precision matrix (inverse of dispersion matrix) between cancers, and  $\otimes$  denotes the Kronecker product. We denote the distribution in (5) by  $MCAR(\alpha, \Lambda)$ . This distribution can be further generalized by allowing different smoothing parameters for each disease, i.e.,

$$\boldsymbol{\phi} \sim N_{np} \left( \mathbf{0}, [Diag(R_1, \dots, R_p)(\Lambda \otimes I_{n \times n})Diag(R_1, \dots, R_p)']^{-1} \right), \quad (6)$$

where  $R_j R'_j = D - \alpha_j W$ ,  $j = 1, \dots, p$ . We denote the distribution in (6) by  $MCAR(\alpha_1, \dots, \alpha_p, \Lambda)$ . Note that the off-diagonal block matrices (the  $R_i$ 's) in the precision matrix in (6) are completely determined by the diagonal blocks. Thus, the spatial precision matrices for each disease induce the cross-covariance structure in (6).

Recently, Jin et al. (2005) developed a more flexible generalized multivariate CAR (GMCAR) model for the random effects  $\boldsymbol{\phi}$ . For example, in the bivariate case ( $p = 2$ ), they specify the conditional distribution  $\boldsymbol{\phi}_1 | \boldsymbol{\phi}_2$  as  $N \left( (\eta_0 I + \eta_1 W) \boldsymbol{\phi}_2, [\tau_1 (D - \alpha_1 W)]^{-1} \right)$ , and the marginal distribution of  $\boldsymbol{\phi}_2$  as  $N \left( \mathbf{0}, [\tau_2 (D - \alpha_2 W)]^{-1} \right)$ , both of which are univariate CAR as in (2). This formulation yields the models of Kim et al. (2001) as a special case and recognizes explicit smoothing parameters ( $\eta_0$  and  $\eta_1$ ) for the cross-covariances, unlike the MCAR models in (6) where the cross-covariances are not smoothed explicitly.

Kim et al (2001) and Jin et al (2006) demonstrate that explicit smoothing of the cross-covariances

yield better model fits to areally referenced bivariate data. However, to model the random effects  $\phi$  with the GMCAR model, we need to specify the conditioning order, since different conditioning orders will result in different marginal distributions for  $\phi_1$  and  $\phi_2$  and, hence, different joint distributions for  $\phi$ . As mentioned in Section 1, in disease mapping contexts a natural conditioning order is often not evident – a problem that is exacerbated when we have more than two diseases. What we seek, therefore, are models that avoid this dependence on conditional ordering, yet are computationally feasible with sufficiently rich spatial structures.

### 3 Order-free MCAR distributions

Our primary methodological objective is to formulate MCAR distributions that allow explicit smoothing of cross-covariances while at the same time not being hampered by conditional ordering. The most natural model here would parametrize the cross-covariances themselves as  $D - \gamma_{ij}W$ , instead of using the  $R_j$ 's as in (6). Unfortunately, except in the separable model with only one smoothing parameter  $\alpha$ , constructing such dispersion structures is not trivial and leads to identifiability issues on the  $\gamma$ 's (see, e.g., Gelfand and Vonatsou, 2003). Kim et al. (2001) resolve these identifiability issues in the bivariate setting using diagonal dominance, but recognize the difficulty in extending this to the multivariate setting. Our contribution here is to address this problem using a *linear model of coregionalization* (LMC). The LMC is a well-established tool used in multivariate geostatistics (Chilés and Delfiner, 1998; Wackernagel, 2003; Banerjee et al., 2004) to incorporate different spatial ranges for each variable. However, to date this technique has not been employed in areal modeling, which has instead traditionally relied upon conditional specifications.

It is worth pointing out that our use of the LMC here is somewhat broader than usually encountered in geostatistics. In geostatistics we typically transform independent latent effects, which

suffices in meeting the primary goal of introducing a different spatial range for each variable. This is akin to introducing different smoothing parameters for each variable and indeed, as we show below in Section 3.2, independent latent effects lead to the  $MCAR(\alpha_1, \dots, \alpha_p; \Lambda)$  in (6). However, to explicitly smooth the cross-covariances with identifiable parameters, we will relax the independence of latent effects. Still, in our ensuing parametrization, we are able to derive conditions that yield valid joint distributions. To be precise, let  $\boldsymbol{\phi} = (\boldsymbol{\phi}'_1, \dots, \boldsymbol{\phi}'_p)'$  be an  $np \times 1$  vector, where each  $\boldsymbol{\phi}_j = (\phi_{1j}, \dots, \phi_{nj})'$  is  $n \times 1$  representing the spatial effects corresponding to disease  $j$ . We can write  $\boldsymbol{\phi} = (A \otimes I_{n \times n})\mathbf{u}$ , where  $\mathbf{u} = (\mathbf{u}'_1, \dots, \mathbf{u}'_p)'$  is  $np \times 1$  with each  $\mathbf{u}_j$  being an  $n \times 1$  areal process. Indeed, a proper distribution for  $\mathbf{u}$  ensures a proper distribution for  $\boldsymbol{\phi}$  subject only to the non-singularity of  $A$ . The flexibility of this approach is apparent: we obtain different multivariate lattice models with rich spatial covariance structures by making different assumptions about the  $p$  spatial processes  $\mathbf{u}_j$ .

### 3.1 Case 1: Independent and identical latent processes

First, we will assume that the random spatial processes  $\mathbf{u}_j$ ,  $j = 1, \dots, p$ , are independent and identical. Since each spatial process  $\mathbf{u}_j$  is a univariate process over areal units, we might adopt a CAR structure (2) for each of them, that is

$$\mathbf{u}_j \sim N_n(\mathbf{0}, (D - \alpha W)^{-1}), \quad j = 1, \dots, p. \quad (7)$$

Since the  $\mathbf{u}_j$  are independent of each other, the joint distribution of  $\mathbf{u} = (\mathbf{u}'_1, \dots, \mathbf{u}'_p)'$  is  $\mathbf{u} \sim N_{np}(\mathbf{0}, I_{p \times p} \otimes (D - \alpha W)^{-1})$ . Recall that since  $\boldsymbol{\phi} = (A \otimes I_{n \times n})\mathbf{u}$ , the joint distribution of  $\boldsymbol{\phi}$  is

$$\boldsymbol{\phi} \sim N_{np}(\mathbf{0}, \Sigma \otimes (D - \alpha W)^{-1}), \quad (8)$$

defining  $\Sigma = AA'$ . We denote the distribution in (8) by  $MCAR(\alpha, \Sigma)$ . Note that the joint distribution of (8) is identifiable up to  $\Sigma = AA'$ , and is independent of the choice of  $A$ . Thus, without loss of generality, we can specify the matrix  $A$  as the upper-triangular Cholesky decomposition of  $\Sigma$ . The  $MCAR(\alpha, \Sigma)$  distribution given in (8) is exactly the same as the  $MCAR(\alpha, \Lambda)$  structure in (5), as in Carlin and Banerjee (2003) and Gelfand and Vounatsou (2003) with  $\Sigma$  corresponding to  $\Lambda^{-1}$ . Since  $\phi = (A \otimes I_{n \times n})\mathbf{u}$ , a valid joint distribution of  $\phi$  requires valid joint distributions of the  $\mathbf{u}_j$ , which happens if and only if  $\frac{1}{\xi_{min}} < \alpha < \frac{1}{\xi_{max}}$ , where  $\xi_{min}$  and  $\xi_{max}$  are the minimum and maximum eigenvalues of  $D^{-\frac{1}{2}}WD^{-\frac{1}{2}}$ . Note if  $\alpha = 1$  in CAR structure (7), which is an ICAR, the joint distribution of  $\phi$  in (8) becomes the multivariate intrinsic CAR (Gelfand and Vounatsou, 2003).

Currently, the WinBUGS package (<http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>) can fit the  $MCAR(\alpha = 1, \Sigma)$  distribution (using its `mv.car` distribution), but not the  $MCAR(\alpha, \Sigma)$ . However through the LMC approach we still can fit the  $MCAR(\alpha, \Sigma)$  in WinBUGS by writing  $\phi = (A \otimes I_{n \times n})\mathbf{u}$  and assigning proper CAR priors (via the `car.proper` distribution) for each  $\mathbf{u}_j$ ,  $j = 1, \dots, p$  with a common smoothing parameter  $\alpha$ . Regarding the prior on  $A$ , note that since  $AA' = \Sigma$  and  $A$  is the Cholesky decomposition of  $\Sigma$ , there is a one-to-one relationship between the elements of  $\Sigma$  and  $A$ . In Section 4, we argue that assigning a prior to  $\Sigma$  is computationally preferable.

### 3.2 Case 2: Independent but not identical latent processes

In Case 1, we assumed that the random spatial processes  $\mathbf{u}_j$ ,  $j = 1, \dots, p$  were independent and identical. However, it will often be preferable to have  $p$  different spatial processes. In this subsection, we will continue to assume that the  $\mathbf{u}_j$  are independent, but relax their being identically distributed. Adopting the CAR structure (2), the distribution of  $\mathbf{u}_j$  is assumed to be

$$\mathbf{u}_j \sim N_n(\mathbf{0}, (D - \alpha_j W)^{-1}), \quad j = 1, \dots, p, \quad (9)$$

where  $\alpha_j$  is the smoothing parameter for the  $j$ th spatial process. Since the  $\mathbf{u}_j$ 's are independent of each other and  $\boldsymbol{\phi} = (A \otimes I_{n \times n})\mathbf{u}$ , the joint distribution of  $\boldsymbol{\phi}$  is

$$\boldsymbol{\phi} \sim N_{np} \left( \mathbf{0}, (A \otimes I_{n \times n})\Gamma^{-1}(A \otimes I_{n \times n})' \right), \quad (10)$$

where  $\Sigma = AA'$  and  $\Gamma$  is an  $np \times np$  block diagonal matrix with  $n \times n$  diagonal entries  $\Gamma_j = D - \alpha_j W$ ,  $j = 1, \dots, p$ . We denote the distribution in (10) by  $MCAR(\alpha_1, \dots, \alpha_p, \Sigma)$ .

In this case, from the joint distribution in (10) it can be seen that different joint distributions of  $\boldsymbol{\phi}$  having different covariance matrices emerge under different linear transformation matrices  $A$ . To ensure  $A$  is identifiable, we could again specify it to be the upper-triangular Cholesky decomposition of  $\Sigma$ , although this might not be the best choice computationally. Through the LMC approach in this case, the distribution in (10) is similar to the  $MCAR(\alpha_1, \dots, \alpha_p, \Lambda)$  structure (6), developed in Carlin and Banerjee (2003) and Gelfand and Vounatsou (2003). All of these have the same number of parameters, and there is no unique joint distribution for  $\boldsymbol{\phi}$  with the  $MCAR(\alpha_1, \dots, \alpha_p, \Lambda)$  structure, since there is not a unique  $R_j$  matrix such that  $R_j R_j' = R_j P P' R_j' = D - \alpha_j W$  ( $P$  being an arbitrary orthogonal matrix). Carlin and Banerjee (2003) take  $R_j$  as the Cholesky decomposition of  $D - \alpha_j W$ , while Gelfand and Vounatsou (2003) instead recommend a spectral decomposition.

Again, a valid joint distribution in (10) requires  $p$  valid distributions for  $\mathbf{u}_j$ , i.e.  $\frac{1}{\xi_{min}} < \alpha_j < \frac{1}{\xi_{max}}$ ,  $j = 1, \dots, p$ . Through the LMC approach, we can also fit the data with the  $MCAR(\alpha_1, \dots, \alpha_p, \Sigma)$  prior distribution (10) on  $\boldsymbol{\phi}$  in WinBUGS as in the previous subsection by writing  $\boldsymbol{\phi} = (A \otimes I_{n \times n})\mathbf{u}$  and assigning proper CAR priors (via the `car.proper` distribution) with a distinct smoothing parameter  $\alpha_j$  for each  $\mathbf{u}_j$ ,  $j = 1, \dots, p$ . As mentioned in the preceding section, we assign a prior to  $AA' = \Sigma$  (e.g., an inverse Wishart), and determine  $A$  from the one-to-one relationship between the elements of  $\Sigma$  and  $A$ ; Section 4 below provides details.

### 3.3 Case 3: Dependent and not identical latent processes

Finally, in this case we will assume that the random spatial processes  $\mathbf{u}_j = (u_{1j}, \dots, u_{nj})'$ ,  $j = 1, \dots, p$  are neither independent nor identically distributed. We now assume that  $u_{ij}$  and  $u_{i, l \neq j}$  are independent given  $u_{k \neq i, j}$  and  $u_{k \neq i, l \neq j}$ , where  $l, j = 1, \dots, p$  and  $i, k = 1, \dots, n$  implying that latent effects for different diseases in the same region are conditionally independent given those for diseases in the neighboring regions. Based upon the Markov property and similar to the conditional distribution given by (3) in the univariate case, we specify the  $ij^{\text{th}}$  conditional distribution as Gaussian with mean

$$E(u_{ij} | u_{k \neq i, j}, u_{i, l \neq j}, u_{k \neq i, l \neq j}) = b_{jj} \left( \sum_{k \sim i} u_{kj} / m_i \right) + \sum_{l \neq j} \left[ b_{jl} \left( \sum_{k \sim i} u_{kl} / m_i \right) \right],$$

and conditional variance  $Var(u_{ij} | u_{k \neq i, j}, u_{i, l \neq j}, u_{k \neq i, l \neq j}) \propto 1/m_i$ , where  $b_{jj}$  denotes the spatial autocorrelation for the random spatial process  $\mathbf{u}_j$  while  $b_{jl}$  ( $l \neq j$ ,  $l, j = 1, \dots, p$ ) denotes the cross spatial correlation between the random spatial process  $\mathbf{u}_j$  and  $\mathbf{u}_l$ . Putting these conditional distributions together reveals the joint distribution of  $\mathbf{u} = (\mathbf{u}'_1, \dots, \mathbf{u}'_p)'$  to be

$$\mathbf{u} \sim N_{np} \left( \mathbf{0}, (I_{p \times p} \otimes D - B \otimes W)^{-1} \right), \quad (11)$$

where  $I$  is a  $p \times p$  identity matrix and  $B$  is a  $p \times p$  symmetric matrix with the elements  $b_{jl}$ ,  $j, l = 1, \dots, p$ . As long as the dispersion matrix in (11) is positive-definite, which boils down to  $(I_{p \times p} \otimes D - B \otimes W)$  being positive definite, (11) is itself a valid model. To assess non-singularity, note  $I_{p \times p} \otimes D - B \otimes W = (I_{p \times p} \otimes D)^{\frac{1}{2}} \left( I_{pn \times pn} - B \otimes D^{-\frac{1}{2}} W D^{-\frac{1}{2}} \right) (I_{p \times p} \otimes D)^{\frac{1}{2}}$ . Denoting the eigenvalues for  $D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$  as  $\xi_i$ ,  $i = 1, \dots, n$ , and the eigenvalues for  $B$  as  $\zeta_j$ ,  $j = 1, \dots, p$ , one finds (see, e.g., Harville, 1997, Theorem 21.11.1) the eigenvalues for  $B \otimes (D^{-\frac{1}{2}} W D^{-\frac{1}{2}})$  as  $\xi_i \times \zeta_j$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, p$ . Hence,

the conditions for  $I_{p \times p} \otimes D - B \otimes W$  being positive definite become  $\xi_i \zeta_j < 1$ , *i.e.*,  $\frac{1}{\xi_{min}} < \zeta_j < \frac{1}{\xi_{max}}$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, p$ , where  $\xi_{min}$  and  $\xi_{max}$  are the minimum and maximum eigenvalues of  $D^{-\frac{1}{2}}WD^{-\frac{1}{2}}$ . Thus,  $\frac{1}{\xi_{min}} < \zeta_j < 1$ ,  $j = 1, \dots, p$ , ensures the positive definiteness of the matrix  $I_{p \times p} \otimes D - B \otimes W$  and, hence, the validity of the distribution of  $\mathbf{u}$  given in (11). In fact,  $\xi_{max} = 1$  and  $\xi_{min} < 0$  (see, e.g., Banerjee et al., 2004), which makes this formulation easier to work with in practice (e.g. in choosing priors; see Section 4) than the alternative parametrization  $\frac{1}{\zeta_{min}} < \xi_j < \frac{1}{\zeta_{max}}$ .

The model in (11) introduces smoothing parameters in the cross-covariance structure through the matrix  $B$ , but unlike the *MCAR* models in Sections 3.1 and 3.2 does not have the  $\Sigma$  matrix to capture non-spatial variances. To remedy this, we model  $\phi = (A \otimes I_{n \times n})\mathbf{u}$  so that the joint distribution for the random effects  $\phi$  is

$$\phi \sim N_{np} \left( \mathbf{0}, (A \otimes I_{n \times n}) (I_{p \times p} \otimes D - B \otimes W)^{-1} (A \otimes I_{n \times n})' \right). \quad (12)$$

Since  $\phi = (A \otimes I_{n \times n})\mathbf{u}$ , it is immediate that the validity of (11) ensures a valid joint distribution for (12). We denote distribution (12) by *MCAR*( $B, \Sigma$ ), where  $\Sigma = AA'$ . Again,  $A$  identifies with the upper-triangular Cholesky square-root of  $\Sigma$ . Note that with  $\Sigma = I$  we recover (11), which we henceforth denote as *MCAR*( $B, I$ ).

To see the generality of (12), we find the joint distribution of  $\phi$  reduces to the *MCAR*( $\alpha_1, \dots, \alpha_p, \Sigma$ ) distribution (10) if  $b_{jl} = 0$  and  $b_{jj} = \alpha_j$ , or the *MCAR*( $\alpha, \Sigma$ ) distribution (8) if  $b_{jl} = 0$  and  $b_{jj} = \alpha$ , in both cases for  $j, l = 1, \dots, p$ . Also note that the distribution in (12) is invariant to orthogonal transformations (up to a reparametrization of  $B$ ) in the following sense: let  $T = AP$  with  $P$  being a  $p \times p$  orthogonal matrix such that  $TT' = APP'A' = \Sigma$ . Then the covariance matrix in (12) can be expressed as  $(A \otimes I_{n \times n}) (I_{p \times p} \otimes D - B \otimes W)^{-1} (A \otimes I_{n \times n})' = (T \otimes I_{n \times n}) (I_{p \times p} \otimes D - C \otimes W)^{-1} (T \otimes I_{n \times n})'$ , where  $C = P'BP$ . Without loss of generality, then, we can choose the matrix  $A$  as the upper-triangular

Cholesky decomposition of  $\Sigma$ .

To understand the features of the  $MCAR(B, \Sigma)$  distribution (12), we illustrate in the bivariate case ( $p = 2$ ). Define  $(AA')^{-1} = \Sigma^{-1} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12} & \Lambda_{22} \end{pmatrix}$  and  $B = A' \begin{pmatrix} \gamma_1 \Lambda_{11} & \gamma_{12} \Lambda_{12} \\ \gamma_{12} \Lambda_{12} & \gamma_2 \Lambda_{22} \end{pmatrix} A$ , where  $A = \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix}$  and  $B = \begin{pmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \end{pmatrix}$ . Note that the  $\gamma$ 's are not identifiable from the matrix  $\Lambda$  and our reparametrization in terms of  $B$  must be used to conduct posterior inference on  $B$  and  $\Lambda$  (see Section 4), from which the cross-covariances may be recovered. The above expression does allow the  $MCAR(B, \Sigma)$  distribution (12) to be rewritten as

$$\phi \sim N_{2n} \left( \mathbf{0}, \begin{pmatrix} (D - \gamma_1 W) \Lambda_{11} & (D - \gamma_{12} W) \Lambda_{12} \\ (D - \gamma_{12} W) \Lambda_{12} & (D - \gamma_2 W) \Lambda_{22} \end{pmatrix}^{-1} \right), \quad (13)$$

which is precisely the general dispersion structure we set out to achieve.

To see how the parameters in (13) affect smoothing, we obtain the conditional means

$$\begin{aligned} E(\phi_{i1} | \phi_{k \neq i, 1}, \phi_{i2}, \phi_{k \neq i, 2}) &= \left( \frac{-\Lambda_{12}}{\Lambda_{11}} \right) \phi_{i2} + \frac{1}{m_i} \sum_{k \sim i} \left[ \gamma_1 \phi_{k1} - \gamma_{12} \left( \frac{-\Lambda_{12}}{\Lambda_{11}} \right) \phi_{k2} \right] \\ \text{and } E(\phi_{i2} | \phi_{k \neq i, 2}, \phi_{i1}, \phi_{k \neq i, 1}) &= \left( \frac{-\Lambda_{12}}{\Lambda_{22}} \right) \phi_{i1} + \frac{1}{m_i} \sum_{k \sim i} \left[ \gamma_2 \phi_{k2} - \gamma_{12} \left( \frac{-\Lambda_{12}}{\Lambda_{22}} \right) \phi_{k1} \right], \end{aligned}$$

and the conditional variances  $Var(\phi_{i1} | \phi_{k \neq i, 1}, \phi_{i2}, \phi_{k \neq i, 2}) = \frac{\Lambda_{11}^{-1}}{m_i}$  and  $Var(\phi_{i2} | \phi_{k \neq i, 2}, \phi_{i1}, \phi_{k \neq i, 1}) = \frac{\Lambda_{22}^{-1}}{m_i}$ ,  $i, k = 1, \dots, n$ . (The conditional moments for special cases such as the separable model in (5) arise by simply setting  $\gamma_1 = \gamma_2 = \gamma_{12} = \alpha$ .) We note that  $\frac{-\Lambda_{12}}{\Lambda_{11}}$  and  $\frac{-\Lambda_{12}}{\Lambda_{22}}$  appear as regression parameters, regressing one component of  $\phi$  at a given site on the other. The spatial smoothing is incorporated by an autoregressive term for each site in the neighbor set which is corrected at each site for the regression on the other component. The smoothing parameters are different for each

spatial association and cross spatial association. For neighboring regions  $i$  and  $k$ , (13) yields the partial correlations

$$\begin{aligned} \text{corr}(\phi_{i1}, \phi_{k1} | \phi_{1 \setminus (ik)}, \phi_2) &= \frac{\gamma_1}{\sqrt{m_i m_k}}, \quad \text{corr}(\phi_{i2}, \phi_{k2} | \phi_{2 \setminus (ik)}, \phi_1) = \frac{\gamma_2}{\sqrt{m_i m_k}}, \\ \text{corr}(\phi_{i1}, \phi_{i2} | \phi_{1 \setminus i}, \phi_{2 \setminus i}) &= \frac{-\Lambda_{12}}{\sqrt{\Lambda_{11} \Lambda_{22}}}, \\ \text{and } \text{corr}(\phi_{i1}, \phi_{k2} | \phi_{1 \setminus i}, \phi_{2 \setminus k}) &= \text{corr}(\phi_{k1}, \phi_{i2} | \phi_{1 \setminus k}, \phi_{2 \setminus i}) = \frac{\gamma_{12} \Lambda_{12}}{\sqrt{m_i m_k \Lambda_{11} \Lambda_{22}}}, \end{aligned}$$

where  $\phi_l = (\phi_{1l}, \dots, \phi_{nl})'$  and  $\phi_{l \setminus i} = (\phi_{1l}, \dots, \phi_{i-1,l}, \phi_{i+1,l}, \dots, \phi_{nl})'$ , for  $l = 1, 2$  and  $i, k = 1, \dots, n$ . With the  $MCAR(\alpha, \Lambda)$  from (5) or the  $MCAR(\alpha, \Sigma)$  from (8), we have  $\text{corr}(\phi_{i1}, \phi_{k1} | \phi_{1 \setminus (ik)}, \phi_2) = \text{corr}(\phi_{i2}, \phi_{k2} | \phi_{2 \setminus (ik)}, \phi_1) = \frac{\alpha}{\sqrt{m_i m_k}}$ , and  $\text{corr}(\phi_{i1}, \phi_{k2} | \phi_{1 \setminus i}, \phi_{2 \setminus k}) = -\text{corr}(\phi_{i1}, \phi_{k1} | \phi_{1 \setminus (ik)}, \phi_2) \times \text{corr}(\phi_{i2}, \phi_{i1} | \phi_{1 \setminus i}, \phi_{2 \setminus i}) = \frac{\alpha \Lambda_{12}}{\sqrt{m_i m_k \Lambda_{11} \Lambda_{22}}}$ , in both cases for  $i, k = 1, \dots, n$ . Thus Case 3 is the most flexible and general with respect to modeling the conditional correlation structure.

## 4 Bayesian computation

Our proposed  $MCAR(B, \Sigma)$  model is straightforwardly implemented in a Bayesian framework using MCMC methods. As in Section 3.3, we write  $\phi = (A \otimes I_{n \times n})\mathbf{u}$ , where  $\mathbf{u} = (\mathbf{u}_1, \mathbf{u}_2)'$  and  $\mathbf{u}_j = (u_{1j}, \dots, u_{nj})'$ . The joint posterior distribution is

$$p(\boldsymbol{\beta}, \sigma^2, \mathbf{u}, A, B | \mathbf{Y}_1, \mathbf{Y}_2) \propto L(\mathbf{Y}_1, \mathbf{Y}_2 | \mathbf{u}, \sigma^2, A) p(\mathbf{u} | B) p(B) p(\boldsymbol{\beta}) p(A) p(\sigma^2), \quad (14)$$

where  $\mathbf{Y}_1 = (Y_{11}, \dots, Y_{n1})'$  and  $\mathbf{Y}_2 = (Y_{12}, \dots, Y_{n2})'$  and  $L(\mathbf{Y}_1, \mathbf{Y}_2 | \mathbf{u}, \sigma^2, A)$  is the data likelihood.

The second term on the right hand side of (14) is  $p(\mathbf{u} | B) = N_{np}(\mathbf{0}, (I_{p \times p} \otimes D - B \otimes W)^{-1})$ . As mentioned in Section 3.3, propriety of this distribution requires the eigenvalues  $\zeta_j$  of  $B$  to satisfy  $\frac{1}{\xi_{\min}} < \zeta_j < 1$  ( $j = 1, \dots, p$ ). When  $p$  is large, it is hard to determine the intervals over the

elements of  $B$  that result in  $\frac{1}{\xi_{min}} < \zeta_j < 1$ , and thus designing priors for  $B$  that guarantee this condition is awkward. In principle, one might impose the constraint numerically by assigning a flat prior or a normal prior with a large variance for the elements of  $B$ , and then simply check whether the eigenvalues of the corresponding  $B$  matrix are in that range during a random-walk Metropolis-Hastings (MH) update. If the resulting eigenvalues are out of range, the values are thrown out since they correspond to prior probability 0; otherwise we perform the standard MH comparison step. In our experience, however, this does not work well, especially when  $p$  is large.

Instead, here we outline a different strategy to update the matrix  $B$ . Our approach is to represent  $B$  using the spectral decomposition, which we write as  $B = P\Delta P'$ , where  $P$  is the corresponding orthogonal matrix of eigenvectors and  $\Delta$  is a diagonal matrix of ordered eigenvalues,  $\zeta_1, \dots, \zeta_p$ . We parameterize the  $p \times p$  orthogonal matrix  $P$  in terms of the  $p(p-1)/2$  Givens angles  $\theta_{ij}$  for  $i = 1, \dots, p-1$  and  $j = i+1, \dots, p$  (Daniels and Kass, 1999). The matrix  $P$  is written as the product of  $p(p-1)/2$  matrices, each one associated with a Givens angle. Specifically,  $P = G_{12}G_{13} \dots G_{1p} \dots G_{(p-1)p}$  where  $i$  and  $j$  are distinct and  $G_{ij}$  is the  $p \times p$  identity matrix with the  $i$ th and  $j$ th diagonal elements replaced by  $\cos(\theta_{ij})$ , and the  $(i, j)$ -th and  $(j, i)$ -th elements replaced by  $\pm \sin(\theta_{ij})$ , respectively. Since the Givens angles  $\theta_{ij}$  are unique with a domain  $(-\pi/2, \pi/2)$  and the eigenvalues  $\zeta_j$  of  $B$  are in the range  $(\frac{1}{\xi_{min}}, 1)$ , we then put a  $\text{Uniform}(-\pi/2, \pi/2)$  prior on the  $\theta_{ij}$  and a  $\text{Uniform}(\frac{1}{\xi_{min}}, 1)$  prior on the  $\zeta_j$ . To update  $\theta_{ij}$ 's or  $\zeta_j$ 's using random-walk Metropolis-Hastings steps with Gaussian proposals, we need to transform them to have support equal to the whole real line. A straightforward solution here is to use  $g(\theta_{ij}) = \log(\frac{\pi/2 + \theta_{ij}}{\pi/2 - \theta_{ij}})$ , a transformation having Jacobian  $\prod_{i=1}^{p-1} \prod_{j=i+1}^p (\pi/2 + \theta_{ij})(\pi/2 - \theta_{ij})$ . In practice, the  $\zeta_j$  must be bounded away from 1 (say, by insisting  $\frac{1}{\xi_{min}} < \zeta_j < 0.999$ ,  $j = 1, \dots, p$ ) to maintain identifiability and hence computational stability. In fact, with our approach it is also easy to calculate the determinant of the precision matrix, that is,  $|I_{p \times p} \otimes D - B \otimes W| \propto \prod_{i=1}^n \prod_{j=1}^p (1 - \xi_i \zeta_j)$ , where  $\xi_i$  are the eigenvalues of  $D^{-\frac{1}{2}}WD^{-\frac{1}{2}}$ , which can be calculated prior to any MCMC iteration.

For the special case of the  $MCAR(\alpha_1, \dots, \alpha_p, \Sigma)$  models, one could assign each  $\alpha_i \sim U(0, 1)$ , which would be sufficient to ensure a valid model (e.g. Carlin and Banerjee, 2002). However, we also investigated with more informative priors on the  $\alpha_i$ 's such as the  $Beta(2, 18)$  that centers the smoothing parameters closer to 1 and leads to greater smoothing.

With respect to the prior distribution  $p(A)$  on the right hand side of (14), we can put independent priors on the individual elements of  $A$ , such as inverse gamma for the square of the diagonal elements of  $A$  and normal for the off-diagonal elements. In practice, we cannot assign non-informative priors here, since then MCMC convergence is poor. In our experience it is easier to assign a vague (i.e., weakly informative) prior on  $\Sigma$  than to put such priors on the elements of  $A$  in terms of letting the data drive the inference and obtaining good convergence. Since  $\Sigma$  is a positive definite covariance matrix, the inverse Wishart prior distribution renders itself as a natural choice, that is,  $\Sigma^{-1} \sim Wishart(\nu, (\nu R)^{-1})$  (see e.g. Carlin and Louis, 2000, p.328). Hence, we instead place a prior directly on  $\Sigma$ , and then use the one-to-one relationship between the elements of  $\Sigma$  and  $A$ . Then the prior distribution  $p(A)$  becomes

$$p(A) \propto |AA'|^{-\frac{\nu+4}{2}} \exp\left\{-\frac{1}{2}tr[\nu D(AA')^{-1}]\right\} \left|\frac{\partial \Sigma}{\partial a_{ij}}\right|,$$

where  $\left|\frac{\partial \Sigma}{\partial a_{ij}}\right|$  is the Jacobian. For example, when  $p = 2$ , the Jacobian is  $4a_{22}^2 a_{11}$ . Rather than updating  $\Sigma$  as a block using a Wishart proposal, updating the elements  $a_{ij}$  of  $A$  offers better control. These are updated via a random-walk Metropolis, using log-normal proposals for the diagonal elements and normal proposals for the off-diagonal elements. With regard to choosing  $\nu$  and  $R$  in the  $Wishart(\nu, (\nu R)^{-1})$ , since  $E(\Sigma^{-1}) = R^{-1}$ , if there is no information about the prior mean structure of  $\Sigma$ , a diagonal matrix  $R$  can be chosen, with the scale of the diagonal elements being judged using ordinary least squares estimates based on independent models for each response variable. While this

leads to a data-dependent prior, typically the Wishart prior lets the data drive the results, leading to robust posterior inference. In this study we adopt  $\nu = 2$  (i.e., the smallest value for which this Wishart prior is proper) and  $R = \text{Diag}(0.1, 0.1)$ . Finally, for the remaining terms on the right hand side of (14), flat priors are chosen for  $\beta_1$  and  $\beta_2$ , while  $\sigma^2$  is assigned a vague inverse gamma prior, i.e. a  $IG(1, 0.01)$  parameterized so that  $E(\sigma^2) = b/(a - 1)$ . In this study,  $\beta$  and  $\sigma^2$  have closed-form full conditionals, and so can be directly updated using Gibbs sampling.

## 5 Simulation study

To evaluate our new approach for modeling multivariate areal data, we begin with some simulation studies. The studies use the spatial layout of the 87 counties in the state of Minnesota, a fairly typical areal arrangement and the one used by our Section 6 data set. We generated count data from a Poisson distribution, as is typical in disease mapping settings

$$Y_{ij} \stackrel{\text{ind}}{\sim} Po(E_{ij}e^{\mu_{ij}}), \quad i = 1, \dots, n, \quad j = 1, 2 \quad (15)$$

where  $Y_{ij}$  is the observed number of cases of cancer type  $j$  (one of two types) in region  $i$  and  $\log \mu_{ij} = \beta_j + \phi_{ij}$  with the  $\beta_j$ 's being fixed constants set to  $\beta_1 = -0.05$  and  $\beta_2 = -0.01$ . These estimates correspond to fitting (15) to esophagus (relatively low incidence) and lung (high incidence) cancer incidence data from Minnesota (see Section 5) using the actual standardized mortalities  $E_{ij}$ 's in the simulations.

To assess the relative performance of our proposed model, we designed five simulation studies. In Study 1, we generate  $\phi_1 = (\phi_{11}, \dots, \phi_{1n})'$  and  $\phi_2 = (\phi_{21}, \dots, \phi_{2n})'$  from the  $MCAR(B, \Sigma)$  distribution (12), where  $D = \text{Diag}(m_i)$  and the adjacency matrix  $W$  are based on the Minnesota

county map. We specify the distribution in (12) by setting  $A = \begin{pmatrix} 0.3 & 0.1 \\ 0.0 & 0.3 \end{pmatrix}$ , which yields a non-spatial correlation of 0.32, and  $B = \begin{pmatrix} 0.8 & 0.4 \\ 0.4 & 0.1 \end{pmatrix}$ . These choices yielded  $\phi_{ij}$ 's ranging between  $(-1.5, 1.5)$  with correlations between the spatial effects ranging from 0.08 to 0.73. In Study 2, we generate  $\phi$  from the the  $MCAR(B, I)$  model (11) retaining the same  $B$  as in Study 1, while in Study 3, we generate them from the the  $MCAR(\alpha, \Sigma)$  model (8) with the same  $A$  as in Study 1. In Study 4, we generate the  $\phi$  from the  $GMCAR(\alpha_1, \alpha_2, \eta_0, \eta_1, \tau_1, \tau_2)$  models under the conditioning order  $\phi_1 | \phi_2$ . We choose the true parameter values in this GMCAR distribution to be  $\alpha_1 = 0.1$ ,  $\alpha_2 = 0.8$ ,  $\eta_0 = 0.4$ ,  $\eta_1 = 0.3$ ,  $\tau_1 = 10$  and  $\tau_2 = 10$ . Finally, Study 5 examines the effect of model misspecification by using a geostatistical model (instead of an MCAR) as the true random effect distribution. Here  $\phi_1$  and  $\phi_2$  are arise from a Gaussian Process with exponential covariance functions  $\exp(-0.01d_{ij})$  and  $\exp(-0.05d_{ij})$ , respectively, where  $d_{ij}$  is the distance between the centroids of counties  $i$  and  $j$ , and is calculated using the `rdist.earth()` function in R using the `fields` package. In this model we take the same values of  $a_{11}, a_{12}, a_{21}$ , and  $a_{22}$  as in Study 2.

## 5.1 Simulation results: MSE

To evaluate the performance of our proposed model, we simulated  $N = 1,000$  data sets, and fit several multivariate models to each. In each study of Table 1, Model 1 is the  $MCAR(B, \Sigma)$  specified by (12), while Model 2 is the  $MCAR(B, I)$  in (11). Model 3 is the  $MCAR(\alpha, \Sigma)$  from Section 3.2 or, equivalently, the  $MCAR(\alpha, \Lambda)$  in (5). Models 4 and 5 are the order-specific  $GMCAR(\alpha_1, \alpha_2, \eta_0, \eta_1, \tau_1, \tau_2)$  models (Jin et al., 2005) with the conditioning order  $\phi_1 | \phi_2$  and the reverse conditioning order  $\phi_2 | \phi_1$  respectively. Finally, Model 6 is a bivariate I.I.D. model.

For each data set and model in each study, we first ran a few initially overdispersed paral-

labeled MCMC chains, and monitored them using measurements of sample autocorrelations within the chains, cross-correlations between parameters, and plots of sample traces. From these, we decided to use 20,000 iterations for the pre-convergence “burn-in” period, and then a further 20,000 iterations as our “production” run for posterior summarization. Unfortunately, the complexity of model (12) precluded us from using the WinBUGS package, so we instead relied on our own programs written in C and executed in R (<http://www.r-project.org>) using the .C function. Random number generation and posterior summarization were also implemented in R.

To evaluate the relative performance of the models, we compare their average mean squared error ( $AMSE$ ). Since the true  $\phi_{ij}$  values are known in the simulation, the  $AMSE$  for each disease can be estimated as

$$AMSE_j = \frac{1}{Nn} \sum_{r=1}^N \sum_{i=1}^n (\hat{\phi}_{ij}^{(r)} - \phi_{ij})^2$$

with associated Monte Carlo standard error estimate

$$\widehat{se}(AMSE_j) = \sqrt{\frac{1}{(Nn)(Nn-1)} \sum_{r=1}^N \sum_{i=1}^n [(\hat{\phi}_{ij}^{(r)} - \phi_{ij})^2 - AMSE_j]^2},$$

where  $\hat{\phi}_{ij}^{(r)}$  is posterior mean estimate for disease  $j$  at county  $i$  based on the  $r$ -th data set. In our case we have  $N = 1,000$ ,  $n = 87$ , and  $j = 1, 2$ .

Table 1 gives the estimated  $AMSE_j$  values and their associated Monte Carlo standard errors for each disease and each model in each simulation study. The estimated overall  $AMSE$  values are calculated by aggregating over the two diseases. Here we also calculate the percentage change in estimated  $AMSE$  for each model compared to the true model in each study, i.e.,  $\Delta = (AMSE_j^{(k)} - AMSE_j^{(true)}) / AMSE_j^{(true)} \times 100$  for models  $k = 1, \dots, 5$ ; negative values would indicate superiority over the true model.

From Table 1, the performance of the  $MCAR(B, \Sigma)$  appears quite impressive. Not surprisingly,

in Study 1 where it is the true model it excels over the other *MCAR* and *GMCAR* models. In Study 2 its performance is highly comparable with the true model *MCAR*( $B, I$ ). In fact, we find a marginally lower *AMSE* score for  $\phi_1$  ( $-0.13\%$ ) and a virtually identical overall *AMSE* ( $0.80\%$ ). The story is similar for Study 3, where the *MCAR*( $\alpha, \Sigma$ ) is the true model, and now we find a marginally lower overall *AMSE* score ( $-0.40\%$ ). In Study 4, where the *GMCAR* is the true model, it performs a little worse (overall *AMSE* score exceeds that of the true model by  $1.93\%$ ) but is still appreciably better than the other competing models. Finally, in Study 5 where all the models are misspecified, the *MCAR*( $B, \Sigma$ ) remains the best, with considerably lower overall *AMSE* scores and only a hint of a competition from the *GMCAR* (reverse order) model.

The performance of the *MCAR*( $B, I$ ) model is somewhat disappointing. The only study where it performs competitively is in Study 2, where it is in fact the true model. In all of the remaining studies its overall *AMSE* scores beat only those of the bivariate I.I.D. model, and are usually much higher than the *GMCAR* models (especially the one having the better ordering) and the *MCAR*( $\alpha, \Sigma$ ). That this occurs despite the smoothing of the cross-covariances through the matrix  $B$  is probably a reflection of the model's inadequacy in capturing the scaling of the spatial effects offered by  $\Sigma$ . The *MCAR*( $\alpha, \Sigma$ ) model, unlike the *MCAR*( $B, I$ ), incorporates a single smoothing parameter but captures the within-site association between the two diseases through  $\Sigma$ . In fact, the *MCAR*( $\alpha, \Sigma$ )'s overall *AMSE* performance is typically seen to lie between the *MCAR*( $B, \Sigma$ ) and the better *GMCAR* model.

The *GMCAR* model also presents some interesting features. Recall that its parametrization is quite rich and it does allow spatially adaptive smoothing of the variances and cross-covariances. However, its primary drawback lies in its sensitivity to the conditioning order, and this is apparent in Table 1. Except for Study 3, where  $\phi_1|\phi_2$  is the true model,  $\phi_2|\phi_1$  appears to be offering much better estimation. In fact, its performance is seen to be only marginally inferior to the *MCAR*( $B, \Sigma$ )

model and its overall AMSE score never exceeds 3.81% of that of the true model. It typically beats the  $MCAR(\alpha, \Sigma)$  model, except in Study 3 where the latter is the true model and has an overall AMSE score lesser by 3.56%. In Study 5, where all the models are misspecified, the better  $GMCAR$ 's AMSE score exceeds that of the  $MCAR(B, \Sigma)$  by only 3% and is considerably better than all the remaining models.

In summary, while the  $MCAR(B, \Sigma)$  model consistently produces the lowest AMSE scores in Table 1, the absolute differences in AMSE among most of the models are rather small, on the order of  $10^{-2}$  in our scale. The associated Monte Carlo standard errors of these AMSEs are on the order of  $10^{-5}$ , making these differences highly reliable computationally. However, in disease mapping exercises we would not expect to see practical differences among the models. Rather, the main message here is that, despite the rich parametrization of the  $MCAR(B, \Sigma)$  model, it does not appear to be suffering from excessive overfitting (i.e., few negative  $\Delta$ 's with respect to the true model) while at the same time offering very robust estimation compared to its competitors.

## 5.2 Simulation results: DIC

To investigate the predictive performance of our methods, we now turn to the *Deviance Information Criterion*, or DIC (Spiegelhalter et al., 2002) as our model choice criterion. This criterion is based on the posterior distribution of the *deviance* statistic,  $D(\boldsymbol{\theta}) = -2 \log f(\mathbf{y}|\boldsymbol{\theta}) + 2 \log h(\mathbf{y})$ , where  $f(\mathbf{y}|\boldsymbol{\theta})$  is the likelihood function for the observed data vector  $\mathbf{y}$  given the parameter vector  $\boldsymbol{\theta}$  on which we focus, and  $h(\mathbf{y})$  is some standardizing function of the data alone (which thus has no impact on model selection and which we set to 0). The DIC is defined analogously to the AIC as the posterior expected deviance plus the ‘‘effective’’ number of parameters, i.e.,  $DIC = \bar{D} + p_D$ . Spiegelhalter et al. (2002) show that  $p_D$  is reasonably defined as  $E_{\theta|y}[D] - D(E_{\theta|y}[\boldsymbol{\theta}]) = \bar{D} - D(\bar{\boldsymbol{\theta}})$ , i.e., the expected deviance minus the deviance evaluated at the posterior expectations. Since small

values of  $\bar{D}$  indicate good fit while small values of  $p_D$  indicate a parsimonious model, small values of the sum (DIC) indicate preferred models. Note that a model with more parameters can be “more parsimonious” using  $p_D$ , since this criterion measures *effective* model size (i.e., the dimension of the space spanned by the parameters *after* accounting for shrinkage of the random effects toward their grand mean). DIC is scale-free (because  $\bar{D}$  is), and so no particular score has any intrinsic meaning; only the *ordering* of DIC scores across models is meaningful. The use of  $p_D$  and DIC is not without controversy; in particular, Celeux et al. (2006) highlight several missing data settings where *negative*  $p_D$  is a real risk. However, we did not experience this problem with any of our multivariate data models, and so adopt this approach here due to its flexibility and ease of use.

To calculate DIC in our MCMC setting, we only need to calculate the average deviance  $\bar{D}$ , and the deviance of posterior mean,  $D(\bar{\theta})$ . Here  $D(\theta)$  is the same for the models we wish to compare since they differ only in their random effect distributions  $p(\phi|B, \Sigma)$ , which we do not consider to be part of the likelihood. Specifically, setting  $2 \log h(\mathbf{y}) = 0$  in  $D(\theta)$ , we have  $D(\theta) \equiv D(\beta, \phi) = -2 \log L(\mathbf{Y}_1, \mathbf{Y}_2 | \beta, \phi)$ , where  $L$  is the likelihood for the Poisson model (15).

Table 2 gives the simulated results for  $\bar{D}$ ,  $p_D$ , and DIC for the same six models and five studies considered in Table 1. Since the scale of  $D$  (though not  $p_D$ ) is arbitrary, in this table we redefine  $\Delta$  as the absolute (not relative) change when compared to the true model in each case. Thus positive  $\Delta$  values indicate poorer fit, larger effective model size, and poorer overall model selection scores for  $\bar{D}$ ,  $p_D$ , and DIC, respectively.

Overall, the results are similar to those seen for MSE, with generally strong support for the  $MCAR(B, \Sigma)$  model. The large improvements in fit brought about by this model are typically not offset by the relatively modest increases in effective model size, leading to significant improvements in overall DIC score (the “sd” measures in this table are simply sample standard deviations of the MCMC draws themselves, not Monte Carlo standard errors as in Table 1). Not only does it excel

in Study 1, but it actually outperforms the true models in terms of actual DIC in Studies 2 and 3. In fact, in terms of  $\bar{D}$  it outperforms all of the true models in Studies 2–4 but marginally loses out to the true GMCAR model in terms of overall DIC due to a substantial increase in the effective number of parameters. The full  $MCAR(B, \Sigma)$  model also dominates the other models in Study 5 even though it, like the rest of the models, is not designed to capture the true geostatistical nature of the simulated data. Although the differences in the average DIC scores in Table 2 are not very large relative to the standard deviations, the ranking of the models is consistently preserved. For instance, the percentage of replicate datasets for which the  $MCAR(B, \Sigma)$  had the lowest DIC score was roughly 99% in Study 1, 57% in Study 2, 86% in Study 3, 47% in Study 4 (the true GMCAR model had lower DIC in the remaining 53% of the replications), and a complete 100% in Study 5.

The behavior of the remainder of the models is also quite similar with results seen for the AMSE and probably with similar reasonings as given there. For instance, we again note the typically poorer performance of the  $MCAR(B, I)$  relative to the other spatial models; the GMCAR with the better ordering is the closest competitor to the  $MCAR(B, \Sigma)$  while the  $MCAR(\alpha, \Sigma)$  performs somewhere between the two GMCARs in studies 1 and 2, performs better than either of them in Study 3 (true model) and Study 5 (marginally over the better GMCAR) and worse in Study 4.

## 6 Data example: Minnesota cancer data

We now turn from the Gaussian likelihood with  $p = 2$  to a non-Gaussian likelihood with  $p = 3$  in a disease mapping context, and illustrate our methods with a data set extracted from the public-use Surveillance, Epidemiology, and End Results (SEER) mortality database (<http://seer.cancer.gov>). SEER county-level mortality databases (SEER\*Stat Database, 2003) provide the numbers of deaths and corresponding numbers of person-years at risk in quinquennial age brackets for each county in

a particular state and each cancer site. Our data consist of the numbers of deaths due to cancers of the lung, larynx, and esophagus in the years from 1990 to 2000 at the county level in Minnesota. The larynx and esophagus are sites of the upper aerodigestive tract, so they are closely related anatomically. Epidemiological evidence shows a strong and consistent relationship between exposure to alcohol and tobacco and the risk of cancer at these two sites (Baron et al., 1993). Meanwhile, lung cancer is the leading cause of cancer death for both men and women. An estimated 160,440 Americans will die in 2004 from lung cancer, accounting for 28% of all cancer deaths. It has long been established that tobacco, and particularly cigarette smoking, is the major cause of lung cancer. More than 87% of lung cancers are smoking-related (<http://www.lungcancer.org>).

These cancers are rare enough relative to the population in each county that the Poisson spatial regression model (4) in Section 2.2 is appropriate. To calculate the expected counts  $E_{ij}$ , we have to take each county's age distribution (over the 18 age groups) into account. To do so, we calculate the expected *age-adjusted* number of deaths due to cancer  $j$  in county  $i$  as  $E_{ij} = \sum_{k=1}^m \omega_j^k N_i^k$ ,  $i = 1, \dots, 87$ ,  $j = 1, 2, 3$ ,  $k = 1, \dots, 18$ , where  $\omega_j^k = (\sum_{i=1}^{87} D_{ij}^k) / (\sum_{i=1}^{87} N_i^k)$  is the age-specific death rate due to cancer  $j$  for age group  $k$  over all Minnesota counties,  $D_{ij}^k$  is the number of deaths in age group  $k$  of county  $i$  due to cancer  $j$ , and  $N_i^k$  is the total population at risk in county  $i$ , age group  $k$ , which we assume to be the same for each type of cancer.

## 6.1 Model comparison

The county-level maps of the raw age-adjusted standardized mortality ratios (i.e.,  $SMR_{ij} = Y_{ij}/E_{ij}$ ) shown in Figure 1 exhibit evidence of correlation both across space and among the cancers, motivating use of our proposed multivariate lattice model. Using the likelihood in (4), we model the random effects  $\phi_{ij}$  using our proposed  $MCAR(B, \Sigma)$  model (12). In what follows we compare it with other MCAR models, including the  $MCAR(\alpha, \Sigma)$  and  $MCAR(1, \Sigma)$  from Section 3.1, a “three

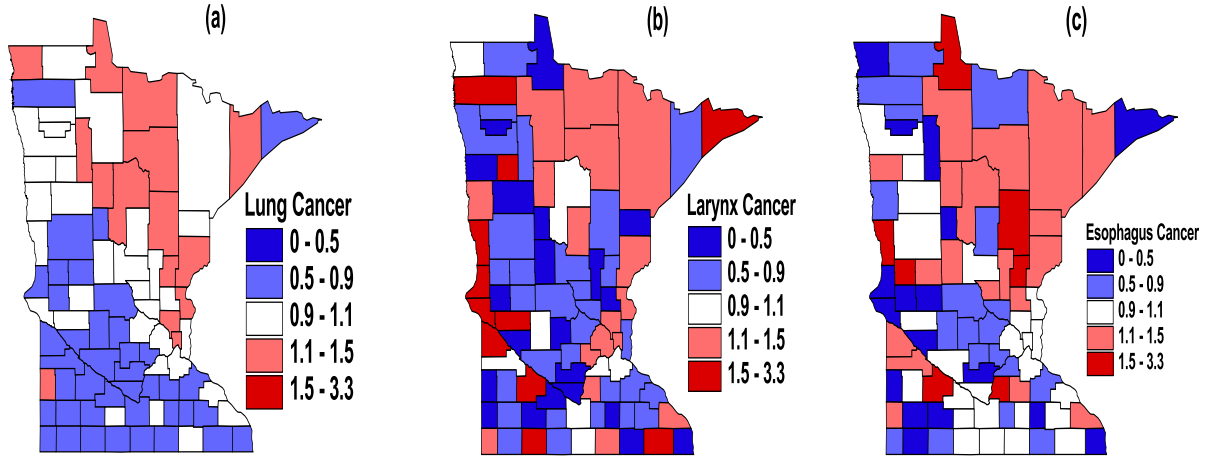


Figure 1: Maps of raw standardized mortality ratios (SMR) of lung, larynx and esophagus cancer in the years from 1990 to 2000 in Minnesota.

separate CARs” model ignoring correlation between cancers, and a trivariate I.I.D. model ignoring correlations of any kind. We also compare one of the  $MCAR(\alpha_1, \alpha_2, \alpha_3, \Sigma)$  models given in (10) of Section 3.2 by choosing the matrix  $A$  as the upper-triangular Cholesky decomposition of  $\Sigma$ . However, we do not consider the order-specific GMCAR model (Jin et al., 2005), since with no natural causal order for these three cancers, it is hard to choose among the six possible conditioning orders.

To see the relative performance of these models, we again use the DIC criterion. As in the previous section, the deviance  $D(\theta)$  is the same for the models we wish to compare since they differ only in their random effect distributions  $p(\phi|B, \Sigma)$ . Specifically, we now have  $D(\theta) \equiv D(\beta, \phi) = -2 \log L(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3 | \beta, \phi)$ , where  $L$  is the likelihood for the Poisson model (4). We choose the same prior distributions for each parameter as in Section 5, and again set the deviance standardizing function  $h(\mathbf{y}) = 0$ . Since  $p = 3$  in this example, we choose the inverse Wishart distribution with  $\nu = 3$  and  $R = \text{Diag}(0.1, 0.1, 0.1)$  for  $\Sigma$ . For a fair DIC comparison, we retain the same “focus” parameters and likelihood across the models. We used 20,000 pre-convergence burn-in iterations followed by a further 20,000 production iterations for posterior summarization.

In what follows, Models 1–6 are multivariate lattice models with different assumptions about

the smoothing parameters. Model 1 is the full model  $MCAR(B, \Sigma)$  (with a  $3 \times 3$  matrix  $B$  whose elements are the six smoothing parameters) while Model 2 is the  $MCAR(B, I)$  model. Model 2 is the  $MCAR(\alpha_1, \alpha_2, \alpha_3; \Sigma)$  model (10) with a different smoothing parameter for each cancer. Model 3 assumes a common smoothing parameter  $\alpha$  and Model 4 fits the three separate univariate CAR model, while Model 6 is the trivariate I.I.D. model. Fit measures  $\bar{D}$ , effective numbers of parameters  $p_D$ , and DIC scores for each model are seen in Table 3. We find that the  $MCAR(B, \Sigma)$  model has the smallest  $\bar{D}$  and DIC values for this data set. The  $MCAR(B, I)$  model again disappoints, excelling over the non-spatial model and the separate CAR models only (very marginally over the latter). The  $MCAR(\alpha, \Sigma)$  and  $MCAR(\alpha_1, \alpha_2, \alpha_3, \Sigma)$  models perform slightly worse than the  $MCAR(B, \Sigma)$  model, suggesting the need for different spatial autocorrelation and cross spatial correlation parameters for this data set. Note that the effective numbers of parameters  $p_D$  in Model 3 is a little larger than in Model 1, even though the latter has three extra parameters. Finally, the MCAR models do better than the separate CAR model or the I.I.D. trivariate model, suggesting that it is worth taking account of the correlations both across counties and among cancers. Model 6 exhibits a large  $p_D$  score, suggesting it does not seem to allow sufficient smoothing of the random effects. This is what we might have expected, since the spatial correlations are missed by this model.

Models 7–11 are the convolution prior models corresponding to Models 1–5 formed by adding I.I.D. effects (following  $N(0, \tau^2)$ ) to the  $\phi_{ij}$ 's. Here the distinctions between the models are somewhat more pronounced due to the added variability in the models caused by the I.I.D. effects. The relative performances of the models remain the same with the  $MCAR(B, \Sigma)$ + I.I.D. model emerging as best. Interestingly, none of the convolution models perform better than their purely spatial counterparts as the improvements in  $\bar{D}$  in the former are insignificant compared to the increase in the effective dimensions brought about. This is indicative of the dominance of the spatial effects over the I.I.D. effects whence the convolution models seem to be rendering overparametrized models.

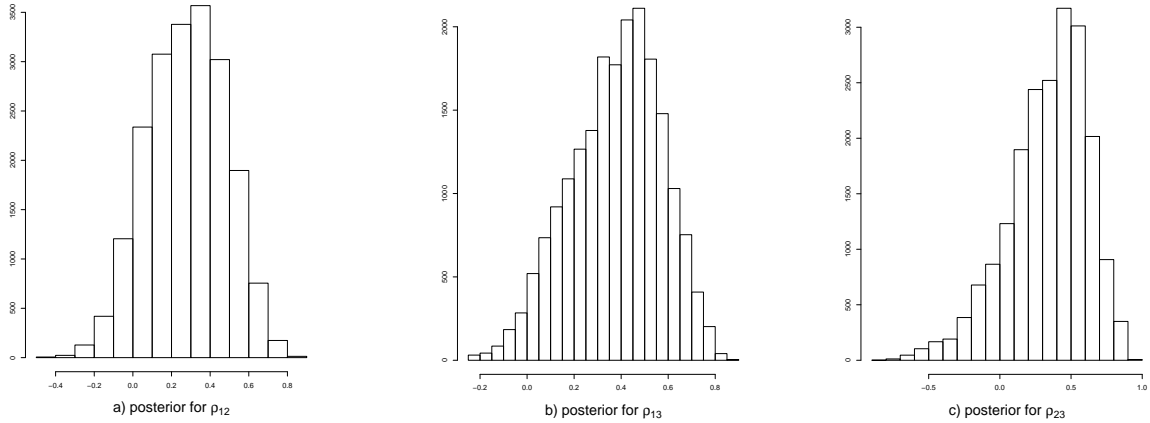


Figure 2: Posterior samples of  $\rho_{12}$ ,  $\rho_{13}$  and  $\rho_{23}$  in the Minnesota cancer data analysis using the  $MCAR(B, \Sigma)$  model: (a) estimated posterior for correlation  $\rho_{12}$  between lung and larynx; (b) estimated posterior for correlation  $\rho_{13}$  between lung and esophagus; (c) estimated posterior for correlation  $\rho_{23}$  between larynx and esophagus.

## 6.2 Results from the selected model

In this subsection, we begin by summarizing our results from the  $MCAR(B, \Sigma)$ , or Model 1 in Table 3. Table 4 provides posterior means and associated standard deviations for the parameters  $\beta$ ,  $\Sigma$  and  $b_{ij}$  in this model, where  $b_{ij}$  is the element of the symmetric matrix  $B$ . Instead of reporting  $\Sigma_{12}$ ,  $\Sigma_{13}$  and  $\Sigma_{23}$ , we provide the mean and associated standard deviations for the correlation parameters  $\rho_{12}$ ,  $\rho_{13}$  and  $\rho_{23}$ , which are calculated as  $\rho_{ij} = \Sigma_{ij} / \sqrt{\Sigma_{ii}\Sigma_{jj}}$ . We also plot histograms of the posterior samples  $\rho_{ij}$  in Figure 2, and histograms of the posterior samples  $b_{ij}$  in Figure 3.

Table 4 and Figure 2 reveal correlations between cancers, in particular a strong correlation between lung and esophagus ( $\rho_{13}$ ). This might explain why the DIC scores for Models 1–4 in Table 4 are smaller than that under the separate CAR model. The  $b_{ij}$  in Table 4 are spatial autocorrelation and cross spatial correlation parameters for the latent spatial processes  $\mathbf{u}_j$ ,  $j = 1, 2, 3$ . Figure 3 shows most of the  $b_{12}$  and  $b_{13}$  posterior samples are positive; the means of these two parameters are 0.323 and 0.389, respectively. Consistent with the DIC results in Table 3, these suggest it is worth fitting our proposed  $MCAR(B, \Sigma)$  model to these data.

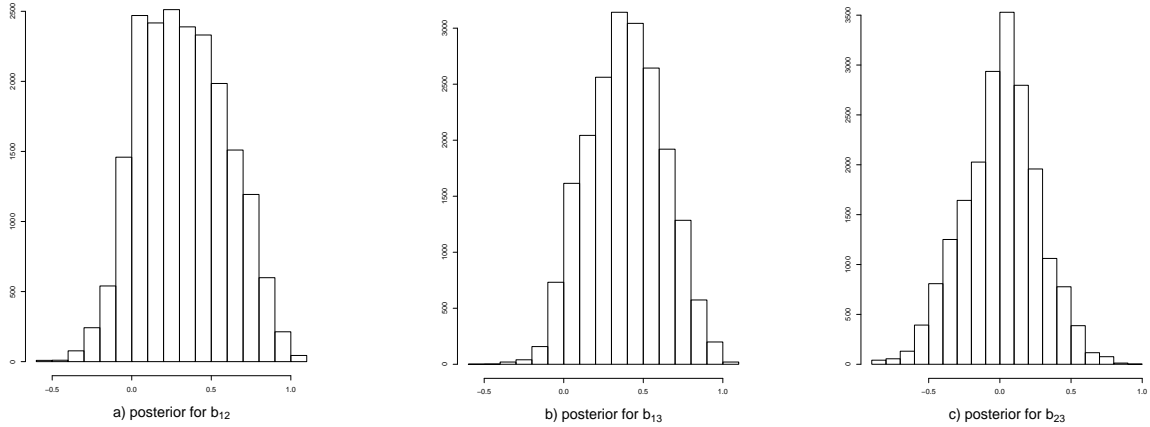


Figure 3: Posterior samples of  $b_{12}$ ,  $b_{13}$  and  $b_{23}$  in the Minnesota cancer data analysis using the  $MCAR(B, \Sigma)$  model: (a) estimated posterior for  $b_{12}$ ; (b) estimated posterior for  $b_{13}$ ; (c) estimated posterior for  $b_{23}$ .

Turning to geographical summaries, Figure 4 maps the posterior means of the fitted standard mortality ratios (SMR) of lung, larynx and esophagus cancer from our  $MCAR(B, \Sigma)$  model. From Figure 4, the correlation among the cancers is apparent, with higher fitted ratios extending from the Twin Cities metro area to the north and northeast (an area where previous studies have suggested smoking may be more common). In Figure 1, the range of the raw SMRs is seen to be from 0 to 3.3, while in Figure 4, the range of the fitted SMRs is from 0.7 to 1.3, due to spatial shrinkage in the random effects.

## 7 Summary and future research

In this paper we have applied the notion of the linear model of coregionalization to the analysis of multivariate areal data, and proposed the  $MCAR(B, \Sigma)$  model for mapping multiple diseases. As with the existing  $MCAR(\alpha, \Lambda)$  and multivariate I.I.D. models in the literature, the  $MCAR(B, \Sigma)$  model is order-free, i.e., independent of the ordering of the variables in the hierarchical model. But the  $MCAR(B, \Sigma)$  model is much more flexible for modeling spatial correlations in multivariate areal

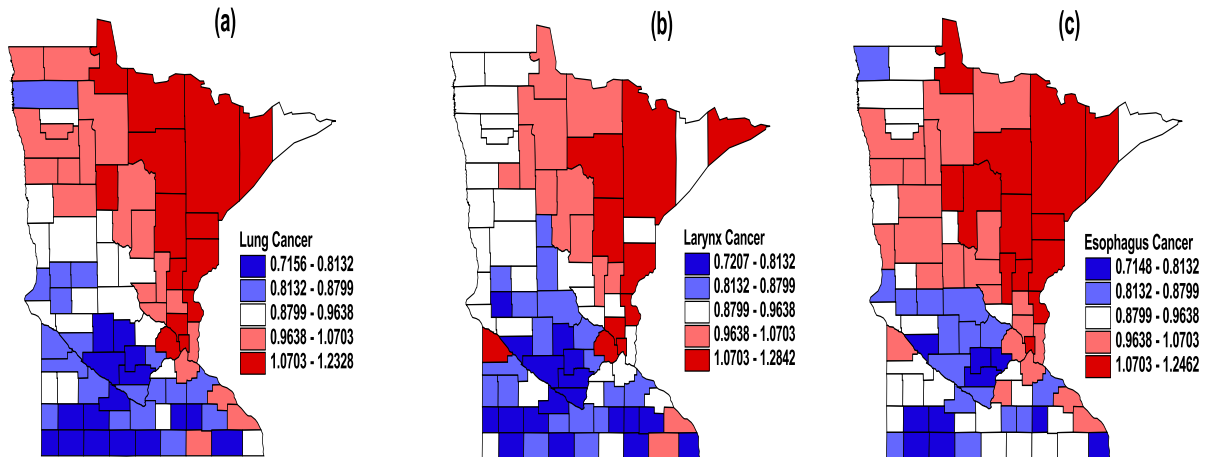


Figure 4: Maps of posterior means of the fitted standard mortality ratios (SMR) of lung, larynx and esophagus cancer in the years from 1990 to 2000 in Minnesota from  $MCAR(B, \Sigma)$  model.

data. Our simulations and data example demonstrate the improved performance of the  $MCAR(B, \Sigma)$  model over the existing alternatives as measured by AMSE or DIC, as well as its efficient implementation using MCMC algorithms. Our approach is also readily extended to higher dimensional ( $p > 2$ ) settings; the computational burden does not increase much with dimension  $p$  since it does not involve large matrix calculations.

Though our emphasis has been on mapping multiple diseases, our proposed MCAR model may also be useful for mapping a single disease. For example, consider a *spatially varying coefficients model* (Assunção, 2003) extending model (1) in Section 2.1 to

$$Y_i \stackrel{ind}{\sim} \text{Poisson}(E_i e^{\mathbf{x}_i' \boldsymbol{\beta} + \varsigma_{i1} z_{i1} + \varsigma_{i2} z_{i2} + \phi_i}), \quad i = 1, \dots, n, \quad (16)$$

where the  $\mathbf{x}_i$  are explanatory, region-level spatial covariates having parameter coefficients  $\boldsymbol{\beta}$ , and  $z_{i1}$  and  $z_{i2}$  are a subset of  $\mathbf{x}_i$  having spatially varying coefficients  $\varsigma_{i1}$  and  $\varsigma_{i2}$ , respectively. We might suspect  $\varsigma_{i1}$ ,  $\varsigma_{i2}$ , or  $\phi_i$  corresponding to counties in geographic proximity to each other might also be similar in magnitude. This in turn means it might be worth fitting a multivariate areal model.

We could take our proposed  $MCAR(B, \Sigma)$  model for  $\Upsilon = (\boldsymbol{\varsigma}'_1, \boldsymbol{\varsigma}'_2, \boldsymbol{\phi}')'$ , where  $\boldsymbol{\varsigma}_1 = (\varsigma_{11}, \dots, \varsigma_{n1})'$ ,  $\boldsymbol{\varsigma}_2 = (\varsigma_{12}, \dots, \varsigma_{n2})'$ , and  $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)'$ .

In our current work, we have only considered mapping the geographic pattern of multiple diseases at a single point in time. However, disease data are often reported over a series of time periods. For example, the SEER database currently provides cancer mortality information for the years from 1969 to 2001. In such cases, we may be interested in temporal effects as well as spatial effects. This motivates an extension of the proposed MCAR model to multivariate *spatiotemporal* data. Knorr-Held (2000) proposed a general framework for spatiotemporal modeling of a single disease using ICAR models, which could perhaps be extended to our multivariate setting. Alternatively, we may work with a multivariate Gaussian AR(1) time series of spatial processes in the setting of dynamic models (Gelfand et al., 2005). This would extend model (4) in Section 2.2 to

$$Y_{ijt} \stackrel{ind}{\sim} \text{Poisson}(E_{ijt} e^{\mathbf{x}'_{ijt} \boldsymbol{\beta}_{jt} + \phi_{ijt}}), \quad i = 1, \dots, n, \quad j = 1, \dots, p, \quad t = 1, \dots, T, \quad (17)$$

where the  $\mathbf{x}_{ijt}$  are explanatory, region-level spatial covariates for disease  $j$  at time period  $t$  having parameter coefficients  $\boldsymbol{\beta}_{jt}$ . Let  $\boldsymbol{\phi}_{j,t+1} = H_j \boldsymbol{\phi}_{j,t} + \boldsymbol{\epsilon}_{jt}$ , where  $H_j$  is a  $n \times n$  matrix,  $\boldsymbol{\phi}_{jt} = (\phi_{1jt}, \dots, \phi_{njt})'$  and  $\boldsymbol{\epsilon}_{jt} = (\epsilon_{1jt}, \dots, \epsilon_{njt})'$ ,  $j = 1, \dots, p$ . Then we can assume that  $H_j = H = \theta_0 I$  or  $H_j = H = \theta_0 I + \theta_1 W$ , where  $W$  is an adjacency matrix and  $I$  is an  $n \times n$  identity matrix. The random effects  $\boldsymbol{\epsilon}_t = (\boldsymbol{\epsilon}'_{1t}, \dots, \boldsymbol{\epsilon}'_{pt})'$  follow a multivariate areal model, such as our proposed  $MCAR(B_t, \Sigma_t)$ , or perhaps the  $MCAR(B, \Sigma)$  if the parameters of the MCAR model do not change over time.

Finally, here we have studied methods for mapping *rare* disease rates, the rarity required to ensure the validity of the Poisson approximation to a binomial sampling distribution. However, it is easy to adapt the statistical methods we have presented here to analyze non-rare diseases by replacing the Poisson likelihood with a binomial distribution for the data (MacNab, 2003). Also while in this paper

we considered only mapping disease rates, we also can apply our  $MCAR(B, \Sigma)$  model methodologies when modeling random effects in a *hazard function* (Carlin and Banerjee, 2003).

## References

- Assunção, R.M. (2003). Space-varying coefficient models for small area data. *Environmetrics*, **14**, 453–473.
- Banerjee, S., Carlin, B.P. and Gelfand, A.E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, Boca Raton, FL: Chapman and Hall/CRC Press.
- Baron, A.E., Franceschi, S., Barra, S., Talamini, R., and La Vecchia, C. (1993). Comparison of the joint effect of alcohol and smoking on the risk of cancer across sites in the upper aerodigestive tract. *Cancer Epidemiology Biomarkers and Prevention*, **2**, 519–523.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. Roy. Statist. Soc., Ser. B*, **36**, 192–236.
- Besag, J., York, J.C., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics (with discussion). *Ann. Inst. Statist. Math.*, **43**, 1–59.
- Carlin, B.P. and Banerjee, S. (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data (with discussion). In *Bayesian Statistics 7*, eds. J.M. Bernardo, M.J. Bayarri, J.O. Berger, A.P. Dawid, D. Heckerman, A.F.M. Smith, and M. West, Oxford: Oxford University Press, pp. 45–63.
- Carlin, B.P. and Louis, T.A. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd ed. Boca Raton, FL: Chapman and Hall/CRC Press.

- Celeux, G., Forbes, F., Robert, C.P., and Titterton, D.M. (2006). Deviance information criteria for missing data models (with discussion). *Bayesian Analysis*, **1**, 651–706.
- Daniels, M.J. and Kass, R.E. (1999). Nonconjugate Bayesian estimation of covariance matrices and its use in hierarchical models. *J. Amer. Statist. Assoc.*, **94**, 1254–1263.
- Gelfand, A.E. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, **4**, 11–25.
- Gelfand, A.E., Banerjee, S. and Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, **16**, 1–15.
- Gelfand, A.E., Schmidt, A.M., Banerjee, S. and Sirmans, C.F. (2004). Multivariate spatial process models: conditional and unconditional Bayesian approaches using coregionalization (with discussion). *Test*, **51**, 1–57.
- Harville, D.D. (1997). *Matrix Algebra from a Statistician's Perspective*. New York: Springer-Verlag.
- Held, L., Natário, I., Fenton, S.E., Rue, H., and Becker, N. (2005). Towards joint disease mapping. *Statistical Methods in Medical Research*, **14**, 61–82.
- Jin, X., Carlin, B.P., and Banerjee, S. (2005). Generalized hierarchical multivariate CAR models for areal data. *Biometrics*, **61**, 950–961.
- Kim, H., Sun, D., and Tsutakawa, R.K. (2001). A bivariate Bayes method for improving the estimates of mortality rates with a twofold conditional autoregressive model. *J. Amer. Statist. Assoc.*, **96**, 1506–1521.
- Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, **19**, 2555–2567.

- Knorr-Held, L. and Best, N.G. (2001). A shared component model for detecting joint and selective clustering of two diseases. *J. Roy. Statist. Soc., Ser. A*, **164**, 73–85.
- MacNab, Y.C. (2003). Hierarchical Bayesian spatial modelling of small-area rates of non-rare disease. *Statist. Med*, **22**, 1761–1773.
- Mardia, K.V. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *Journal of Multivariate Analysis*, **24**, 265–284.
- Royle, J.A. and Berliner, L.M. (1999). A hierarchical approach to multivariate spatial modeling and prediction. *J. Agr. Biol. Env. Statist.*, **4**, 29–56.
- Sain, S.R. and Cressie, N. (2002). Multivariate lattice models for spatial environmental data. In *Proc. of the ASA Section on Statistics and the Environment*, pp.2820–2825.
- Schmidt, A.M. and Gelfand, A.E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *J. Geophysical Research – Atmospheres*, **108**, 8783.
- SEER\*Stat Database (2003). *Mortality – All COD, Public-Use With County, Total U.S. for Expanded Races (1990-2000) <18 Age Groups>*. National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2003.
- Spiegelhalter, D.J., Best, N., Carlin, B.P., and van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion). *J. Roy. Statist. Soc., Ser. B*, **64**, 583–639.
- Wackernagel, H. (2003). *Multivariate Geostatistics: An Introduction with Applications*, 3rd ed. New York: Springer-Verlag.
- Wall, M.M. (2004). A close look at the spatial structure implied by the CAR and SAR models. *J. Statist. Plann. Inf.*, **121**, 311–324.

study	model	$AMSE_1$		$AMSE_2$		overall $AMSE$	
		mean (se)	$\Delta$	mean (se)	$\Delta$	mean (se)	$\Delta$
1	1. $MCAR(B, \Sigma)$	7.51 (8.51)	–	1.34 (0.49)	–	4.43 (8.54)	–
	2. $MCAR(B, I)$	8.36 (8.53)	11.32	2.17 (1.09)	61.94	5.26 (4.30)	18.74
	3. $MCAR(\alpha, \Sigma)$	7.90 (8.48)	5.19	1.54 (1.15)	14.92	4.72 (4.28)	6.54
	4. GMCAR	8.16 (8.55)	8.65	1.71 (1.19)	27.61	4.93 (4.32)	11.29
	5. GMCAR(reverse order)	7.75 (8.51)	3.19	1.39 (0.69)	3.73	4.57 (4.27)	3.16
	6. bivariate I.I.D.	9.92 (8.52)	32.09	2.91 (1.18)	117.16	6.42 (4.30)	44.92
2	1. $MCAR(B, \Sigma)$	7.90 (8.41)	–0.13	2.15 (1.33)	3.86	5.03 (4.26)	0.80
	2. $MCAR(B, I)$	7.91 (8.39)	–	2.07 (1.19)	–	4.99 (4.24)	–
	3. $MCAR(\alpha, \Sigma)$	8.07 (8.41)	2.02	2.58 (1.31)	24.64	5.33 (4.25)	6.81
	4. GMCAR	8.19 (9.51)	3.54	2.82 (1.45)	36.23	5.51 (4.81)	10.42
	5. GMCAR(reverse order)	7.96 (8.51)	0.63	2.39 (1.34)	15.46	5.18 (4.31)	3.81
	6. bivariate I.I.D.	10.53 (10.39)	33.12	3.70 (2.29)	78.74	7.11 (5.32)	42.48
3	1. $MCAR(B, \Sigma)$	8.07 (8.27)	0.25	1.98 (2.25)	–2.94	5.03 (4.18)	–0.40
	2. $MCAR(B, I)$	8.75 (8.29)	8.70	2.61 (1.19)	27.94	5.68 (4.19)	12.48
	3. $MCAR(\alpha, \Sigma)$	8.05 (8.31)	–	2.04 (0.87)	–	5.05 (4.18)	–
	4. GMCAR	8.31 (8.57)	3.23	2.26 (1.24)	10.78	5.29 (4.33)	4.75
	5. GMCAR(reverse order)	8.28 (9.48)	2.86	2.17 (1.30)	6.37	5.23 (4.78)	3.56
	6. bivariate I.I.D.	9.57 (10.34)	18.89	3.38 (2.11)	65.68	6.48 (5.28)	28.32
4	1. $MCAR(B, \Sigma)$	7.80 (8.02)	0.91	1.68 (2.45)	7.01	4.74 (4.19)	1.93
	2. $MCAR(B, I)$	8.23 (8.12)	6.47	2.22 (3.39)	41.40	5.22 (4.40)	12.26
	3. $MCAR(\alpha, \Sigma)$	7.95 (8.20)	2.85	1.94 (2.57)	23.57	4.95 (4.30)	6.45
	4. GMCAR	7.73 (7.98)	–	1.57 (2.42)	–	4.65 (4.17)	–
	5. GMCAR(reverse order)	8.08 (9.38)	4.53	2.01 (3.31)	28.02	5.04 (4.97)	8.38
	6. bivariate I.I.D.	9.68 (9.31)	25.23	3.38 (3.31)	115.29	6.53 (4.94)	40.43
5	1. $MCAR(B, \Sigma)$	9.26 (8.72)	–	2.07 (2.63)	–	5.66 (4.55)	–
	2. $MCAR(B, I)$	10.03 (8.82)	8.32	2.65 (3.03)	28.02	6.34 (4.66)	12.01
	3. $MCAR(\alpha, \Sigma)$	9.98 (8.70)	7.77	2.44 (2.93)	17.87	6.21 (4.59)	9.72
	4. GMCAR	10.28 (10.03)	11.01	3.95 (3.11)	90.82	7.12 (5.25)	25.79
	5. GMCAR(reverse order)	10.19 (9.99)	1.40	3.67 (3.02)	9.66	6.93 (5.22)	3.00
	6. bivariate I.I.D.	14.20 (9.51)	53.35	3.98 (3.52)	92.27	9.09 (5.07)	60.61

Table 1: Average mean squared error ( $AMSE$ ,  $\times 10^{-2}$ ), associated Monte Carlo standard errors (se,  $\times 10^{-5}$ ), and percentage change in  $AMSE$  ( $\Delta\%$ ) relative to the true model in each simulation study with Poisson responses. Except for Study 5, the symbol “–” indicates the model is the true model for this study.

study	model	$\bar{D}$		$p_D$		DIC	
		mean (sd)	$\Delta$	mean (sd)	$\Delta$	mean (sd)	$\Delta$
1	1. $MCAR(B, \Sigma)$	1004.3 (20.6)	–	71.0 (3.6)	–	1075.3 (20.8)	–
	2. $MCAR(B, I)$	1071.9 (21.1)	67.5	59.0 (3.8)	–12.0	1130.8 (21.5)	55.6
	3. $MCAR(\alpha, \Sigma)$	1039.9 (20.4)	35.6	64.1 (3.5)	–6.8	1105.0 (20.6)	28.9
	4. GMCAR	1052.8 (20.1)	48.5	57.9 (3.9)	–13.0	1110.7 (20.7)	35.5
	5. GMCAR (reverse order)	1026.9 (21.6)	32.6	58.2 (3.4)	–12.7	1085.2 (22.0)	19.9
	6. bivariate I.I.D	1112.8 (23.8)	108.4	55.0 (3.8)	–15.9	1167.8 (24.0)	92.5
2	1. $MCAR(B, \Sigma)$	1004.9 (20.6)	–22.6	68.9 (3.7)	13.8	1073.8 (20.9)	–8.8
	2. $MCAR(B, I)$	1027.5 (21.5)	–	55.1 (3.51)	–	1082.6 (21.8)	–
	3. $MCAR(\alpha, \Sigma)$	1043.0 (19.8)	15.5	63.9 (3.6)	8.8	1106.9 (20.2)	24.3
	4. GMCAR	1071.1 (21.4)	43.6	65.0 (3.8)	9.8	1136.1 (21.7)	53.5
	5. GMCAR (reverse order)	1022.3 (21.9)	–5.2	63.0 (3.8)	7.9	1085.3 (22.3)	2.7
	6. bivariate I.I.D	1106.2 (24.7)	78.7	51.0 (3.8)	–4.1	1157.3 (25.0)	74.7
3	1. $MCAR(B, \Sigma)$	1000.5 (20.5)	–39.5	70.0 (3.9)	9.9	1070.5 (20.7)	–29.6
	2. $MCAR(B, I)$	1092.0 (20.6)	51.0	53.0 (3.7)	–7.1	1145.0 (21.1)	44.9
	3. $MCAR(\alpha, \Sigma)$	1040.0 (20.6)	–	60.1 (3.6)	–	1100.1 (20.9)	–
	4. GMCAR	1078.9 (21.1)	38.9	54.0 (3.8)	–6.1	1132.9 (21.4)	32.8
	5. GMCAR (reverse order)	1064.6 (22.9)	24.5	52.1 (3.5)	–8.0	1116.7 (23.3)	16.6
	6. bivariate I.I.D	1134.9 (24.5)	94.9	51.1 (3.9)	–9.0	1186.0 (24.7)	85.9
4	1. $MCAR(B, \Sigma)$	987.2 (20.3)	–7.3	75.0 (3.6)	11.0	1062.2 (20.6)	3.7
	2. $MCAR(B, I)$	1028.6 (21.5)	34.1	73.9 (3.7)	9.9	1102.5 (21.7)	44.1
	3. $MCAR(\alpha, \Sigma)$	1021.9 (20.3)	27.4	77.8 (3.5)	13.8	1099.7 (20.6)	41.2
	4. GMCAR	994.5 (21.6)	–	64.0 (3.9)	–	1058.5 (21.7)	–
	5. GMCAR (reverse order)	1020.5 (22.0)	26.0	66.1 (3.6)	2.1	1086.5 (22.4)	28.0
	6. bivariate I.I.D	1095.7 (24.9)	101.2	60.0 (3.8)	–4.0	1155.8 (25.2)	97.3
5	1. $MCAR(B, \Sigma)$	1152.7 (30.0)	–	84.9 (6.5)	–	1237.6 (30.6)	–
	2. $MCAR(B, I)$	1227.7 (30.5)	75.0	83.9 (6.8)	–1.0	1311.6 (31.3)	74.0
	3. $MCAR(\alpha, \Sigma)$	1191.4 (38.6)	38.7	87.9 (6.5)	3.0	1279.3 (39.0)	41.7
	4. GMCAR	1205.4 (37.6)	52.7	73.9 (7.8)	–11.0	1279.4 (38.7)	41.8
	5. GMCAR (reverse order)	1220.2 (38.4)	67.5	75.9 (7.5)	–9.0	1296.1 (39.0)	58.6
	6. bivariate I.I.D.	1395.3 (44.8)	242.6	59.5 (7.8)	–25.4	1454.8 (45.1)	217.2

Table 2: Simulated  $\bar{D}$ ,  $p_D$ , and DIC scores and associated standard deviations, along with absolute change ( $\Delta$ ) relative to the true model in each simulation study. Except for Study 5, the symbol “–” indicates the model is the true model for this study.

	model	$\bar{D}$	$p_D$	DIC
1	$MCAR(B, \Sigma)$	138.8	82.5	221.3
2	$MCAR(B, I)$	147.6	81.4	229.0
3	$MCAR(\alpha_1, \alpha_2, \alpha_3, \Sigma)$	139.6	86.4	226.0
4	$MCAR(\alpha, \Sigma)$	143.4	81.9	225.3
5	separate CAR	147.6	82.8	230.4
6	trivariate I.I.D.	146.8	91.3	238.1
7	$MCAR(B, \Sigma) +$ trivariate I.I.D	129.6	137.6	267.2
8	$MCAR(B, I) +$ trivariate I.I.D	139.5	155.2	294.7
9	$MCAR(\alpha_1, \alpha_2, \alpha_3, \Sigma) +$ trivariate I.I.D	137.4	155.0	292.4
10	$MCAR(\alpha, \Sigma) +$ trivariate I.I.D	138.2	151.0	289.2
11	separate CAR + trivariate I.I.D.	139.2	162.8	302.0

Table 3: Model comparison using DIC statistics, Minnesota cancer data analysis.

	Lung median (2.5%, 97.5%)	Larynx median (2.5%, 97.5%)	Esophagus mean (2.5%, 97.5%)
$\beta_1, \beta_2, \beta_3$	-0.093 (-0.179, -0.006)	-0.128 (-0.316, 0.027)	-0.080 (-0.194, 0.025)
$\Sigma_{11}, \Sigma_{22}, \Sigma_{33}$	0.048 (0.030, 0.073)	0.173 (0.054, 0.395)	0.107 (0.044, 0.212)
$\rho_{12}, \rho_{13}$		0.277 (-0.112, 0.643)	0.378 (-0.022, 0.716)
$\rho_{23}$			0.337 (-0.311, 0.776)
$b_{11}, b_{22}, b_{33}$	0.442 (-0.302, 0.921)	0.036 (-0.830, 0.857)	0.312 (-0.526, 0.901)
$b_{12}, b_{13}$		0.323 (-0.156, 0.842)	0.389 (-0.028, 0.837)
$b_{23}$			0.006 (-0.519, 0.513)

Table 4: Posterior summaries of parameters in  $MCAR(B, \Sigma)$  model for Minnesota cancer data.