Chapter 10

Models for multivariate areal data

In this chapter we explore the extension of univariate CAR methodology (Sections 4.3 and 6.4.3) to the multivariate setting. Such models can be employed to introduce multiple, dependent spatial random effects associated with areal units (as standard CAR models do for a single set of random effects). In this regard, Kim et al. (2001) presented a "twofold CAR" model to model counts for two different types of disease over each areal unit. Similarly, Knorr-Held and Best (2000) have developed a "shared component" model for the above purpose, but their methodology too seems specific to the bivariate situation. Knorr-Held and Rue (2002) illustrate sophisticated MCMC blocking approaches in a model placing three conditionally independent CAR priors on three sets of spatial random effects in a shared component model setting.

Multivariate CAR (MCAR) models can also provide coefficients in a multiple regression setting that are dependent and spatially varying at the areal unit level. For example, Gamerman et al. (2002) investigate a Gaussian Markov random field (GMRF) model (a multivariate generalization of the pairwise difference IAR model) and compare various MCMC blocking schemes for sampling from the posterior that results under a Gaussian multiple linear regression likelihood. They also investigate a "pinned down" version of this model that resolves the impropriety problem by centering the $\boldsymbol{\phi}_i$ vectors around some mean location. These authors also place the spatial structure on the spatial regression coefficients themselves, instead of on extra intercept terms (that is, in (6.27) we would drop the ϕ_i, and replace β_1 by β_{1i}, which would now be assumed to have a CAR structure). Assunção et al. (2002) refer to these models as *space-varying coefficient* models, and illustrate in the case of estimating fertility schedules. Assunção (2003) offers a nice review of the work up to that time in this area. Also working with areal units, Sain and Cressie (2007) offer multivariate GMRF models, proposing a generalization that permits asymmetry in the spatial conditional cross-correlation matrix. They use this approach to jointly model the counts of white and minority persons residing in the census block groups of St. James Parish, LA, a region containing several hazardous waste sites.

We point out that multivariate CAR models are not the only option available for analyzing multivariate areal data. Zhang, Hodges and Banerjee (2009) develop an arguably much simpler alternative approach building upon the techniques of smoothed ANOVA (SANOVA) (see Hodges, Cui, Sargent and Carlin, 2007). Instead of simply shrinking effects without any structure, these authors propose SANOVA to smooth spatial random effects by taking advantage of the spatial structure. The underlying idea is to extend SANOVA to cases in which one factor is a spatial lattice, which is smoothed using a CAR model, and a second factor is, for example, type of disease. Datasets routinely lack enough information to identify the additional structure of MCAR. SANOVA offers a simpler and more intelligible structure than the MCAR while performing as well. Nevertheless, the MCAR and more general CAR-based approaches provide a diverse and rich class of models suitable for capturing complex spatial associations. We focus upon these approaches in the remainder of this section.

10.1 The multivariate CAR (MCAR) distribution

For a vector of univariate variables $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_n)$, zero-centered CAR specifications were detailed in Section 4.3. For the MCAR model we instead let $\boldsymbol{\phi}^T = (\boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \ldots, \boldsymbol{\phi}_n)$ where each $\boldsymbol{\phi}_i = (\phi_{i1}, \phi_{i2}, \ldots, \phi_{ip})^T$ is $p \times 1$. Most multivariate CAR models are members of the family developed by Mardia (1988). Analogous to the univariate case, the joint distribution is derived from the full conditional distributions. Under the MRF assumption, we can specify these conditional distributions as

$$p(\boldsymbol{\phi}_i|\boldsymbol{\phi}_{j\neq i}, \Gamma_i) \;=\; N\left(\sum_{i\sim j} B_{ij}\boldsymbol{\phi}_j \;,\; \Gamma_i\right) \;,\;\; i, j = 1, \ldots, n, \tag{10.1}$$

where Γ_i and B_{ij} are $p \times p$ matrices. Mardia (1988) proved, using a multivariate analogue of Brook's Lemma, that the full conditional distributions in (10.1) yield a joint distribution of the form

$$p(\boldsymbol{\phi} \mid \{\Gamma_i\}) \propto \exp\left\{-\frac{1}{2}\boldsymbol{\phi}^T\Gamma^{-1}(I - \tilde{B})\boldsymbol{\phi}\right\} \;,$$

where Γ is block-diagonal with blocks Γ_i, and $\tilde{B}$ is $np \times np$ with (i, j)-th block B_{ij}.

As in the univariate case, symmetry of $\Gamma^{-1}(I - \tilde{B})$ is required. A convenient special case sets $B_{ij} = b_{ij}I_{p\times p}$, yielding the symmetry condition $b_{ij}\Gamma_j = b_{ji}\Gamma_i$, analogous to (4.14). If as in Subsection 4.3.1 we take $b_{ij} = w_{ij}/w_{i+}$ and $\Sigma_i = w_{i+}^{-1}\Sigma$, then the symmetry condition is satisfied.

Kronecker product notation simplifies the form of $\Gamma^{-1}(I - \tilde{B})$. That is, setting $\tilde{B} = B \otimes I$ with B as in (4.13) and $\Gamma = D^{-1} \otimes \Sigma$ so

$$\Gamma^{-1}(I - \tilde{B}) = (D \otimes \Sigma^{-1})(I - B \otimes I) = (D - W) \otimes \Sigma^{-1} \;.$$

Again, the singularity of $D - W$ implies that $\Gamma^{-1}(I - B)$ is singular. We denote this distribution by $MCAR(1, \Sigma)$.

To consider remedies to the impropriety, Mardia (1988) proposed rewriting (10.1) as

$$p(\boldsymbol{\phi}_i|\boldsymbol{\phi}_{j\neq i}, \Gamma_i) \;=\; N\left(R_i\sum_{i\sim j} B_{ij}\boldsymbol{\phi}_j \;,\; \Gamma_i^{-1}\right) \;,\;\; i, j = 1, \ldots, n,$$

where R_i is $p \times p$. Now $\Gamma^{-1}(I - \tilde{B})$ is revised to $\Gamma^{-1}(I - \tilde{B}_R)$ where $\tilde{B}_R$ has (i, j)th block R_iB_{ij}. In general, then, the symmetry condition becomes

$$(\Gamma_i^{-1}R_iB_{ij})^T = \Gamma_j^{-1}R_jB_{ji} \;\text{ or }\; \Gamma_jB_{ij}^TR_i^T = R_jB_{ji}\Gamma_i \;.$$

See Mardia, (1988), Expression (2.4) in this regard. If, in addition, $\Gamma^{-1}(I - \tilde{B}_R)$ is positive definite, then the conditional distributions uniquely determine the joint distribution

$$\boldsymbol{\phi} \sim N\left(\mathbf{0} \;,\; [\Gamma(I - \tilde{B}_R)]^{-1}\right) \;. \tag{10.2}$$

In particular, if $B_{ij} = b_{ij}I_{p\times p}$ and $b_{ij} = w_{ij}/w_{i+}$, the symmetry condition simplifies to

$$w_{j+}\Gamma_jR_i^T = w_{i+}R_j\Gamma_i \;.$$

Finally, if in addition we take $\Gamma_i = w_{i+}^{-1}\Lambda$, we obtain $\Lambda R_i^T = R_j\Lambda$, which reveals that we must have $R_i = R_j = R$, and thus

$$\Lambda R^T = R\Lambda \;. \tag{10.3}$$

For any arbitrary positive definite Λ, a generic solution to (10.3) is $R = \rho\Lambda^t$. Hence, regardless of t, (10.3) introduces a total of $\binom{p+1}{2} + 1$ parameters. Thus, without loss of generality, we can set $t = 0$, hence $R = \rho I$. Calculations as above yield

$$\Sigma^{-1} = \Gamma^{-1}(I - \tilde{B}_R) = (D - \rho W) \otimes \Lambda^{-1} , \tag{10.4}$$

where Σ is the $np \times np$ variance-covariance matrix of $\boldsymbol{\phi}$. Therefore, Σ has a *separable* structure and is nonsingular under the same restriction to ρ as in the univariate case, and we have

$$\boldsymbol{\phi} \sim N\left(0,\ [(D - \rho W) \otimes \Lambda]^{-1}\right) . \tag{10.5}$$

Following Gelfand and Vounatsou (2003) and Carlin and Banerjee (2003), we denote this model by $MCAR(\rho, \Sigma)$.

The separable MCAR model in (10.5) is obtained under the assumption that $R_i = \rho I_{p\times p}, i = 1, \ldots, n$. While the positive definiteness condition of the covariance matrix follows immediately from its Kronecker product form, which also has computational benefits, the assumption of a common ρ for $j = 1, \ldots, p$ may well be too strong. To elucidate further, suppose $p = 2$ (e.g., two diseases in each county), and define $\boldsymbol{\phi}_1' = (\phi_{11}, \ldots, \phi_{n1})$ and $\boldsymbol{\phi}_2' = (\phi_{12}, \ldots, \phi_{n2})$. Then the MCAR formulation (10.5) can be written as

$$\begin{pmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} (D - \rho W)\Lambda_{11} & (D - \rho W)\Lambda_{12} \\ (D - \rho W)\Lambda_{12} & (D - \rho W)\Lambda_{22} \end{pmatrix}^{-1} \right) , \tag{10.6}$$

where $\Lambda_{ij},\ i = 1, 2,\ j = 1, 2$ are the elements of Λ. More generally, we may need three different ρ_i parameters in (10.6) to explain the correlation between the two types of cancer and across the counties that neighbor each other (Kim et al., 2001). The covariance matrix Σ would then be revised to

$$\Sigma = \begin{pmatrix} (D - \rho_1 W)\Lambda_{11} & (D - \rho_3 W)\Lambda_{12} \\ (D - \rho_3 W)\Lambda_{12} & (D - \rho_2 W)\Lambda_{22} \end{pmatrix}^{-1} , \tag{10.7}$$

where ρ_1 and ρ_2 are the smoothing parameters for the two cancer types, and ρ_3 is the "bridging" or "linking" parameter associating ϕ_{i1} with ϕ_{j2}, $i \neq j$. Unfortunately, with this general covariance matrix, it is difficult to derive the conditions for positive definiteness as they depend upon the unknown Λ matrix.

Carlin and Banerjee (2003) and Gelfand and Vounatsou (2003) generalize the separable MCAR model by allowing two different ρ parameters (say, ρ_1 and ρ_2), and denote this model as $MCAR(\rho_1, \rho_2, \Lambda)$. They write the precision matrix Σ^{-1} as

$$\begin{pmatrix} U_1^T U_1 \Lambda_{11} & U_1^T U_2 \Lambda_{12} \\ U_2^T U_1 \Lambda_{12} & U_2^T U_2 \Lambda_{22} \end{pmatrix} = \begin{pmatrix} U_1^T & 0 \\ 0 & U_2^T \end{pmatrix} (\Lambda \otimes I_{n\times n}) \begin{pmatrix} U_1 & 0 \\ 0 & U_2 \end{pmatrix} , \tag{10.8}$$

where $U_k' U_k = D - \rho_k W,\ k = 1, 2$. Carlin and Banerjee (2003) take U_k to be the Cholesky decomposition of $D - \rho_k W$ so that U_k is an upper-triangular matrix, while Gelfand and Vounatsou (2003) employ a spectral decomposition, i.e., $U_k = \text{Diag}(1 - \rho_k \lambda_i)^{\frac{1}{2}} P' D^{\frac{1}{2}} P$, where the λ_i are the eigenvalues of $D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$ and P is an orthogonal matrix with the corresponding eigenvectors as its columns. Either way, this generalization of the MCAR model permits different smoothing parameters ρ_k for each k (e.g., different strengths of spatial correlation for each type of cancer). As before, Λ controls the nonspatial correlation among cancers at any given location.

The conditions for the covariance matrix to be positive definite are easy to find as long as the Cholesky or spectral decompositions exist and Λ is positive definite. For the $p = 2$ case, these reduce to $|\rho_1| < 1$ and $|\rho_2| < 1$. The spectral approach may be better in terms

of Bayesian computing, since it does not requires calculation of a Cholesky decomposition at each MCMC iteration, a substantial burden particularly for a data set with many spatial regions. Neither of these MCAR structures allows a smoothing parameter ρ on the off-diagonal of the precision matrix as in (10.7); we cannot model the off-diagonal, since it is determined by the diagonal. Finally, since the decomposition of $D - \rho_k W$ is not unique, we can have different MCAR models with the covariance structure (10.8).

Note that the covariance structure (10.8) easily generalizes to an arbitrary number, say, p, of diseases. We write

$$\boldsymbol{\phi} \;\sim\; N_{np}\left(\mathbf{0},\; \left[\mathrm{Diag}(U_1^T,\ldots,U_p^T)(\Lambda \otimes I_{n\times n})\mathrm{Diag}(U_1,\ldots,U_p)\right]^{-1}\right), \tag{10.9}$$

where $U_j^T U_j = D - \rho_j W$, $j = 1,\ldots,p$. We denote the distribution in (10.9) by $MCAR(\rho_1,\ldots,\rho_p,\, \Lambda)$. Note that the off-diagonal block matrices (the U_i's) in the precision matrix in (10.9) are completely determined by the diagonal blocks. Thus, the spatial precision matrices for each disease induce the cross-covariance structure in (10.9).

Kim, Sun, and Tsutakawa (2001) proposed a multivariate CAR model in the bivariate ($p = 2$) case, which they dub the "twofold conditionally autoregressive" model, and which we notate as $2fCAR(\rho_0,\, \rho_1,\, \rho_2,\, \rho_3,\, \tau_1,\, \tau_2)$. They specify the moments of the full conditional distributions as

$$\mathrm{E}(\phi_{ik}|\phi_{il},\phi_{jk},\phi_{jl}) = \frac{1}{2m_i+1}\left(\rho_k \sum_{j\sim i}\phi_{jk} + \rho_3\sqrt{\frac{\tau_l}{\tau_k}}\sum_{j\sim i}\phi_{jl} + \rho_0\sqrt{\frac{\tau_l}{\tau_k}}\phi_{il}\right)$$

and

$$\mathrm{Var}(\phi_{ik}|\phi_{il},\phi_{jk},\phi_{jl}) = \frac{\tau_k^{-1}}{2m_i+1}, \quad i,j = 1,\ldots n,\;\; l,k = 1,2,\;\; l \neq k,$$

where $j \sim i$ again means that region j is a neighbor of region i. Adding the Gaussian MRF structure, they derive the joint distribution arising from these full conditional distributions as

$$\begin{pmatrix}\boldsymbol{\phi}_1\\ \boldsymbol{\phi}_2\end{pmatrix} \sim N\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix},\; \begin{pmatrix} (2D+I-\rho_1 W)\tau_1 & -(\rho_0 I+\rho_3 W)\sqrt{\tau_1\tau_2} \\ -(\rho_0 I+\rho_3 W)\sqrt{\tau_1\tau_2} & (2D+I-\rho_2 W)\tau_2 \end{pmatrix}^{-1}\right), \tag{10.10}$$

where again $\boldsymbol{\phi}_1' = (\phi_{11},\ldots,\phi_{n1})$, $\boldsymbol{\phi}_2' = (\phi_{12},\ldots,\phi_{n2})$, $D = Diag(m_i)$, and W is the adjacency matrix. This model has the same number of parameters in the covariance structure (six) as the general formulation (10.7) in the bivariate case, so they are related to each other. In (10.10), ρ_1 and ρ_2 are the smoothing parameters, while ρ_0 and ρ_3 are the bridging parameters associating ϕ_{i1} with ϕ_{i2} and ϕ_{j2}, $j \neq i$, respectively. Unfortunately, this MCAR model is only designed for the bivariate case ($p = 2$), and seems difficult to generalize to higher dimensions. Also, under this approach it is hard to find conditions that guarantee a positive definite covariance matrix in (10.10). The conditions $|\rho_l| < 1$, $l = 0,1,2,3$ given by Kim et al. (2001) are sufficient but not necessary, and may be overly restrictive for some data sets since they restrict the correlation of ϕ_{i1} with ϕ_{i2} and ϕ_{j2}, $j \neq i$. Finally, this generalization comes at a significant price in terms of computing, since it requires many matrix multiplications, determinant evaluations, and inverses at each MCMC iteration, so can be very time-consuming even when working on a relatively small spatial domain.

10.2 Modeling with a proper, non-separable MCAR distribution

As in Chapter 6, the MCAR specifications of the previous section are employed in models for spatial random effects arising in a hierarchical model. For instance, suppose we have

a linear model with continuous data $\mathbf{Y}_{ik}$, $i = 1, \ldots, n$, $k = 1, \ldots, m_i$, where $\mathbf{Y}_{ik}$ is a $p \times 1$ vector denoting the kth response at the ith areal unit. The mean of the $\mathbf{Y}_{ik}$ is $\boldsymbol{\mu}_{ik}$ where $\mu_{ikj} = (\mathbf{X}_{ik})_j \boldsymbol{\beta}^{(j)} + \phi_{ij}$, $j = 1, \ldots, p$. Here $\mathbf{X}_{ik}$ is a $p \times s$ matrix with covariates associated with $\mathbf{Y}_{ik}$ having jth row $(\mathbf{X}_{ik})_j$, $\boldsymbol{\beta}^{(j)}$ is an $s \times 1$ coefficient vector associated with the jth component of the $\mathbf{Y}_{ik}$'s, and ϕ_{ij} is the jth component of the $p \times 1$ vector $\boldsymbol{\phi}_i$. Given $\left\{\boldsymbol{\beta}^{(j)}\right\}, \{\boldsymbol{\phi}_i\}$ and V, the $\mathbf{Y}_{ik}$ are conditionally independent $N(\boldsymbol{\mu}_{ik}, V)$ variables. Adding a prior for $\left\{\boldsymbol{\beta}^{(j)}\right\}$ and V and one of the MCAR models from Subsection 10.1 for the $\boldsymbol{\phi}_i$ completes the second stage of the specification. Finally, a hyperprior on the MCAR parameters completes the model.

Alternatively, we might change the first stage to a multinomial. Here k disappears and $\mathbf{Y}_i$ is assumed to follow a multinomial distribution with sample size n_i and with $(p+1) \times 1$ probability vector $\boldsymbol{\pi}_i$. Working on the logit scale, using cell $p+1$ as the baseline, we could set $\log\left(\frac{\pi_{ij}}{\pi_{i,p+1}}\right) = \mathbf{X}_i^T \boldsymbol{\beta}^{(j)} + \phi_{ij}$, $j = 1, \ldots, p$, with $\mathbf{X}$'s, $\boldsymbol{\beta}$'s and $\boldsymbol{\phi}$'s interpreted as in the previous paragraph. Many other multivariate first stages could also be used, such as other multivariate exponential family models.

Regardless, model-fitting is most easily implemented using a Gibbs sampler with Metropolis updates where needed. The full conditionals for the $\boldsymbol{\beta}$'s will typically be normal (under a normal first-stage model) or else require Metropolis, slice, or adaptive rejection sampling (Gilks and Wild, 1992). For the $MCAR(1, \Sigma)$ and $MCAR(\rho, \Sigma)$ models, the full conditionals for the $\boldsymbol{\phi}_i$'s will be likelihood-adjusted versions of the conditional distributions that define the MCAR, and are updated as a block. For the $MCAR(\boldsymbol{\rho}, \Sigma)$ model, we can work with either the $\boldsymbol{\phi}$ or the $\boldsymbol{\psi}$ parametrization. With a non-Gaussian first stage, it will be awkward to pull the transformed effects out of the likelihood in order to do the updating. However, with a Gaussian first stage, it may well be more efficient to work on the transformed scale. Under the Gaussian first stage, the full conditional for V will be seen to follow an inverse Wishart, as will Σ. The ρ's do not follow standard distributions; in fact, discretization expedites computation, avoiding Metropolis steps.

We have chosen an illustrative prior for ρ in the ensuing example following three criteria. First, we insist that $\rho < 1$ to ensure propriety but allow $\rho = 0.99$. Second, we do not allow $\rho < 0$ since this would violate the similarity of spatial neighbors that we seek. Third, since even moderate spatial dependence requires values of ρ near 1 (recall the discussion in Subsection 4.3.1) we place prior mass that favors the upper range of ρ. In particular, we put equal mass on the following 31 values: $0, 0.05, 0.1, \ldots, 0.8, 0.82, 0.84, \ldots, 0.90, 0.91, 0.92, \ldots, 0.99$.

Finally, model choice arises here only in selecting among MCAR specifications. That is, we do not alter the mean vector in these investigations; our interest here lies solely in comparing the spatial explanations. Multivariate versions of the Gelfand and Ghosh (1998) criterion (5.14) for multivariate Gaussian data are employed.

Example 10.1 *(Analysis of the child growth data).* Child growth is usually monitored using anthropometric indicators such as height adjusted for age (HAZ), weight adjusted for height (WHZ), and weight adjusted for age (WAZ). Independent analysis of each of these indicators is normally carried out to identify factors influencing growth that may range from genetic and environmental factors (e.g., altitude, seasonality) to differences in nutrition and social deprivation. Substantial variation in growth is common within as well as between populations. Recently, geographical variation in child growth has been thoroughly investigated for the country of Papua New Guinea in Mueller et al. (2001). Independent spatial analyses for each of the anthropometric growth indicators identified complex geographical patterns of child growth finding areas where children are taller but skinnier than average, others where they are heavier but shorter, and areas where they are both short and thin. These geographical patterns could be linked to differences in diet and subsistence agriculture, leading to the analysis presented here; see Gelfand and Vounatsou (2003) for further discussion.

The data for our illustration comes from the 1982–1983 Papua New Guinea National Nutrition Survey (NNS) (Heywood et al., 1988). The survey includes anthropometric measures (age, height, weight) of approximately 28,000 children under five years of age, as well as dietary, socioeconomic, and demographic data about those children and their families. Dietary data include the type of food that respondents had eaten the previous day. Subsequently, the data were coded to 14 important staples and sources of protein. Each child was assigned to a village and each village was assigned to one of 4566 environmental zones (resource mapping units, or RMUs) into which Papua New Guinea has been divided for agriculture planning purposes. A detailed description of the data is given in Mueller et al. (2001).

The nutritional scores, height adjusted for age (HAZ), and weight adjusted for age (WAZ) that describe the nutritional status of a child were obtained using the method of Cole and Green (1992), which yields age-adjusted standard normal deviate Z-scores. The data set was collected at 537 RMUs. To overcome sparseness and to facilitate computation, we collapsed to 250 spatial units. In the absence of digitized boundaries, Delaunay tessellations were used to create the neighboring structure in the spatial units.

Because of the complex, multidimensional nature of human growth, a bivariate model that considers differences in height and weight jointly might be more appropriate for analyzing child growth data in general and to identify geographical patterns of growth in particular. We propose the use of Bayesian hierarchical spatial models and with *multivariate CAR* (MCAR) specifications to analyze the bivariate pairs of indicators, HAZ and WAZ, of child growth. Our modeling reveals bivariate spatial random effects at RMU level, justifying the MCAR specification.

Recalling the discussion of Subsection 10.1, it may be helpful to provide explicit expressions, with obvious notation, for the modeling and the resulting association structure. We have, for the jth child in the ith RMU,

$$\mathbf{Y}_{ij} = \begin{pmatrix} (HAZ)_{ij} \\ (WAZ)_{ij} \end{pmatrix} = \mathbf{X}_{ij}^T \begin{pmatrix} \boldsymbol{\beta}^{(H)} \\ \boldsymbol{\beta}^{(W)} \end{pmatrix} + \begin{pmatrix} \phi_i^{(H)} \\ \phi_i^{(W)} \end{pmatrix} + \begin{pmatrix} \epsilon_{ij}^{(H)} \\ \epsilon_{ij}^{(W)} \end{pmatrix} .$$

In this setting, under say the $MCAR(\rho, \Sigma)$ model,

$$\begin{aligned} &\mathrm{Cov}((HAZ)_{ij}, (HAZ)_{i'j'} \mid \boldsymbol{\beta}^{(H)}, \boldsymbol{\beta}^{(W)}, \rho, \Sigma, V) \\ &\quad = \mathrm{Cov}(\phi_i^{(H)}, \phi_{i'}^{(H)}) + V_{11} I_{i=i', j=j'} , \\ &\mathrm{Cov}((WAZ)_{ij}, (WAZ)_{i'j'} \mid \boldsymbol{\beta}^{(H)}, \boldsymbol{\beta}^{(W)}, \rho, \Sigma, V) \\ &\quad = \mathrm{Cov}(\phi_i^{(W)}, \phi_{i'}^{(W)}) + V_{22} I_{i=i', j=j'} , \\ \text{and} \quad &\mathrm{Cov}((HAZ)_{ij}, (WAZ)_{i'j'} \mid \boldsymbol{\beta}^{(H)}, \boldsymbol{\beta}^{(W)}, \rho, \Sigma, V) \\ &\quad = \mathrm{Cov}(\phi_i^{(H)}, \phi_{i'}^{(W)}) + V_{12} I_{i=i', j=j'} , \end{aligned}$$

where $cov(\phi_i^{(H)}, \phi_{i'}^{(H)}) = (D_W - \rho W)_{ii'} \Sigma_{11}$, $cov(\phi_i^{(W)}, \phi_{i'}^{(W)}) = (D_W - \rho W)_{ii'} \Sigma_{22}$, and $cov(\phi_i^{(H)}, \phi_{i'}^{(W)}) = (D_W - \rho W)_{ii'} \Sigma_{12}$. The interpretation of the components of Σ and V (particularly Σ_{12} and V_{12}) is now clarified.

We adopted noninformative uniform prior specifications on $\boldsymbol{\beta}^{(H)}$ and $\boldsymbol{\beta}^{(W)}$. For Σ and V we use inverse Wishart priors, i.e., $\Sigma^{-1} \sim W(\Omega_1, c_1)$, $V^{-1} \sim W(\Omega_2, c_2)$ where Ω_1, Ω_2 are $p \times p$ matrices and c_1, c_2 are shape parameters. Since we have no prior knowledge regarding the nature or extent of dependence, we choose Ω_1 and Ω_2 diagonal; the data will inform about the dependence *a posteriori*. Since the $\mathbf{Y}_{ij}$'s are centered and scaled on each dimension, setting $\Omega_1 = \Omega_2 = I$ seems appropriate. Finally, we set $c_1 = c_2 = 4$ to provide low precision for these priors. We adopted for ρ_1 and ρ_2 the prior discussed in the previous section. Simulation from the full conditional distributions of the $\boldsymbol{\beta}$'s and the $\boldsymbol{\psi}_i, i = 1, \ldots, n$ is straightforward

Model	G	P	D_∞
$MCAR(1, \Sigma)$	34300.69	33013.10	67313.79
$MCAR(\rho, \Sigma)$	34251.25	33202.86	67454.11
$MCAR(\boldsymbol{\rho}, \Sigma)$	34014.46	33271.97	67286.43

Table 10.1 *Model comparison for child growth data.*

	Height (HAZ)			Weight (WAZ)		
Covariate	2.5%	50%	97.5%	2.5%	50%	97.5%
Global mean	–0.35	–0.16	–0.01	–0.48	–0.25	–0.15
Coconut	0.13	0.20	0.29	0.04	0.14	0.24
Sago	–0.16	–0.07	–0.00	–0.07	0.03	0.12
Sweet potato	–0.11	–0.03	0.05	–0.08	0.01	0.12
Taro	–0.09	0.01	0.10	–0.19	–0.09	0.00
Yams	–0.16	–0.04	0.07	–0.19	–0.05	0.08
Rice	0.30	0.40	0.51	0.26	0.38	0.49
Tinned fish	0.00	0.12	0.24	0.04	0.17	0.29
Fresh fish	0.13	0.23	0.32	0.08	0.18	0.28
Vegetables	–0.08	0.08	0.25	0.02	0.19	0.35
V_{11}, V_{22}	0.85	0.87	0.88	0.85	0.87	0.88
V_{12}	0.60	0.61	0.63			
Σ_{11}, Σ_{22}	0.30	0.37	0.47	0.30	0.39	0.52
Σ_{12}	0.19	0.25	0.35			
ρ_1, ρ_2	0.95	0.97	0.97	0.10	0.80	0.97

Table 10.2 *Posterior summaries of the dietary covariate coefficients, covariance components, and autoregression parameters for the child growth data using the most complex MCAR model.*

as they are standard normal distributions. Similarly, the full conditionals for V^{-1} and Σ^{-1} are Wishart distributions. We implemented the Gibbs sampler with 10 parallel chains.

Table 10.1 offers a comparison of three MCAR models using (5.14), the Gelfand and Ghosh (1998) criterion. The most complex model is preferred, offering sufficient improvement in goodness of fit to offset the increased complexity penalty. Summaries of the posterior quantities under this model are shown in Table 10.2. These were obtained from a posterior sample of size 1,000, obtained after running a 10-chain Gibbs sampler for 30,000 iterations with a burn-in of 5,000 iterations and a thinning interval of 30 iterations. Among the dietary factors, high consumption of sago and taro are correlated with thinner and shorter children, while high consumption of rice, fresh fish, and coconut are associated with both heavier and taller children. Children from villages with high consumption of vegetables or tinned fish are heavier.

The posterior for the correlation associated with Σ, $\Sigma_{12}/\sqrt{\Sigma_{11}\Sigma_{22}}$, has mean 0.67 with 95% credible interval $(0.57, 0.75)$, while the posterior for the correlation associated with V, $V_{12}/\sqrt{V_{11}V_{22}}$, has mean 0.71 with 95% credible interval $(0.70, 0.72)$. In addition, ρ_1 and ρ_2 differ.

10.3 Conditionally specified Generalized MCAR (GMCAR) distributions

Jin, Carlin and Banerjee (2005) expand upon this idea by building the joint distribution for a multivariate Markov random field (MRF) through specifications of simpler conditional

and marginal models. The approach can be regarded as the analogue of the conditioning approach of Royle and Berliner (1999); see Section 9.4.5, for areal models.

This approach is best elucidated in the bivariate setting. We now assume the joint distribution of $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ is

$$\begin{pmatrix}\boldsymbol{\phi}_1\\ \boldsymbol{\phi}_2\end{pmatrix} \sim N\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}^T & \Sigma_{22}\end{pmatrix}\right) ,$$

where the Σ_{kl}, $k, l = 1, 2$ are $n \times n$ covariance matrices. From standard multivariate normal theory, we have $\mathrm{E}(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2) = \Sigma_{12}\Sigma_{22}^{-1}\boldsymbol{\phi}_2$ and $\mathrm{Var}(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2) = \Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T$. Now writing $A = \Sigma_{12}\Sigma_{22}^{-1}$, we can rewrite the joint distribution of $\boldsymbol{\phi}_1$ and $\boldsymbol{\phi}_2$ as

$$\begin{pmatrix}\boldsymbol{\phi}_1\\ \boldsymbol{\phi}_2\end{pmatrix} \sim N\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\Sigma_{11\cdot 2} + A\Sigma_{22}A^T & A\Sigma_{22}\\ (A\Sigma_{22})^T & \Sigma_{22}\end{pmatrix}\right) . \tag{10.11}$$

According to Harville (1997, Corollary 14.8.5), the conditions that ensure the propriety of (10.11) are that Σ_{22} and $\Sigma_{11\cdot 2}$ are positive definite. Since $\boldsymbol{\phi}_1|\boldsymbol{\phi}_2 \sim N(A\boldsymbol{\phi}_2, \Sigma_{11\cdot 2})$ and $\boldsymbol{\phi}_2 \sim N(0, \Sigma_{22})$, we can construct $p(\boldsymbol{\phi}) = p(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2)p(\boldsymbol{\phi}_2)$ where $\boldsymbol{\phi}^T = (\boldsymbol{\phi}_1^T, \boldsymbol{\phi}_2^T)$. For the joint distribution of $\boldsymbol{\phi}$, then, we need to specify the matrices $\Sigma_{11\cdot 2}$, Σ_{22}, and A.

Jin et al. (2005) propose specifying the conditional distribution for $\boldsymbol{\phi}_1 \mid \boldsymbol{\phi}_2$ as $\boldsymbol{\phi}_1|\boldsymbol{\phi}_2 \sim N\left(A\boldsymbol{\phi}_2\,,\, [(D - \rho_1 W)\tau_1]^{-1}\right)$, and the marginal distribution of $\boldsymbol{\phi}_2$ as $\boldsymbol{\phi}_2 \sim N\left(\mathbf{0}\,,\, [(D - \rho_2 W)\tau_2]^{-1}\right)$, where ρ_1 and ρ_2 are the smoothing parameters associated with the conditional distribution of $\boldsymbol{\phi}_1|\boldsymbol{\phi}_2$ and the marginal distribution of $\boldsymbol{\phi}_2$ respectively, and τ_1 and τ_2 scale the precision of $\boldsymbol{\phi}_1|\boldsymbol{\phi}_2$ and $\boldsymbol{\phi}_2$, respectively. The induced joint distribution will always be proper as long as these two CAR distributions are valid, so the positive definiteness of the covariance matrix in (10.11) is easily verified. If $D = Diag(m_i)$ and W be the adjacency matrix, then the positive definiteness conditions require only that $|\rho_1| < 1$ and $|\rho_2| < 1$. Further restricting these parameters between 0 and 1 $0 < \rho_1 < 1$ avoid negative spatial autocorrelation.

Regarding the A matrix, since $\mathrm{E}(\boldsymbol{\phi}_1 \mid \boldsymbol{\phi}_2) = A\boldsymbol{\phi}_2$, we assume its elements are of the form

$$a_{ij} = \begin{cases} \eta_0 & \text{if } j = i\\ \eta_1 & \text{if } j \in N_i \text{ (i.e., if region } j \text{ is a neighbor of region } i)\\ 0 & \text{otherwise}\end{cases} .$$

Thus $A = \eta_0 I + \eta_1 W$ and $E(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2) = (\eta_0 I + \eta_1 W)\boldsymbol{\phi}_2$. Here η_0 and η_1 are the bridging parameters associating ϕ_{i1} with ϕ_{i2} and ϕ_{j2}, $j \neq i$. One could easily augment A with another bridging parameter η_2 associated with the *second-order* neighbors (neighbors of neighbors) in each region, but we do not pursue this generalization here.) Under these assumptions, the covariance matrix in the joint distribution (10.11) can be written as $\Sigma = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12}^T & \Sigma_{22}\end{pmatrix}$, where

$$\begin{aligned}\Sigma_{11} &= [\tau_1(D - \rho_1 W)]^{-1} + (\eta_0 I + \eta_1 W)[\tau_2(D - \rho_2 W)]^{-1}(\eta_0 I + \eta_1 W)\\ \Sigma_{12} &= (\eta_0 I + \eta_1 W)[\tau_2(D - \rho_2 W)]^{-1}\\ \Sigma_{22} &= [\tau_2(D - \rho_2 W)]^{-1} .\end{aligned} \tag{10.12}$$

Jin et al. (2005) denote this new model by $GMCAR(\rho_1, \rho_2, \eta_1, \eta_2, \tau_1, \tau_2)$. This bivariate GMCAR model has the same number of parameters as the twofold CAR model in (10.10), and has one more parameter than the $MCAR(\rho_1, \rho_2, \Lambda)$ model in (10.8).

Setting $\rho_1 = \rho_2 = \rho$ and $\eta_1 = 0$ in (10.12), and using a standard result from matrix theory (Harville, 1997, Corollary 8.5.12), produces the separable precision matrix $\Sigma^{-1} = \Lambda \otimes (D - \rho W)$, where $\tau_1 = \Lambda_{11}$, $\tau_2 = \Lambda_{22} - \frac{\Lambda_{12}^2}{\Lambda_{11}}$, and $\eta_0 = -\frac{\Lambda_{12}}{\Lambda_{11}}$. Further assuming $\rho = 1$ produces an improper MIAR (Multivariate Intrinsic Autoregressive) model. If we assume

$\rho_1 \neq \rho_2$ and $\eta_0 = \eta_1 = 0$, then we ignore dependence between the multivariate components, and the model turns out to be equivalent fitting two separate univariate CAR models. Finally, if we instead assume $\rho_1 = \rho_2 = 0$, $\eta_0 \neq 0$, and $\eta_1 = 0$, the model becomes an i.i.d. bivariate normal model.

The MCAR model in (10.6) has $\mathrm{E}(\boldsymbol{\phi}_1 \mid \boldsymbol{\phi}_2) = -\frac{\Lambda_{12}}{\Lambda_{11}}\boldsymbol{\phi}_2$, which reveals that the conditional mean is merely a scale multiple of $\boldsymbol{\phi}_2$. Since $Var(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2) = [\Lambda_{11}(D - \rho_1 W)]^{-1}$, which is free of $\boldsymbol{\phi}_2$, the distribution of the random variable at a particular site in one field is independent of neighbor variables in another field *given* the value of the related variable at the same area. The extended MCAR model (10.8) has

$$\mathrm{E}(\boldsymbol{\phi}_1 \mid \boldsymbol{\phi}_2) = -\frac{\Lambda_{12}}{\Lambda_{11}}(D - \rho_1 W)^{-\frac{1}{2}}(D - \rho_2 W)^{\frac{1}{2}}\boldsymbol{\phi}_2$$

and $\mathrm{Var}(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2)$ identical to that of model (10.6). Therefore, the distribution of the random variable at a particular site in one field is no longer conditionally independent of neighboring variables in another field. However, this dependence is determined implicitly by ρ_1 and ρ_2 and is difficult to interpret.

By contrast, the GMCAR model has $\mathrm{E}(\boldsymbol{\phi}_1 \mid \boldsymbol{\phi}_2) = (\eta_0 I + \eta_1 W)\boldsymbol{\phi}_2$ and $\mathrm{Var}(\boldsymbol{\phi}_1 \mid \boldsymbol{\phi}_2) = [\tau_1(D - \rho_1 W)]^{-1}$. Thus, while the conditional variance remains free of $\boldsymbol{\phi}_2$, the GMCAR allows spatial information (via the W matrix) to enter the conditional mean in an intuitive way, with a free parameter (η_1) to model the weights. That is, the GMCAR models the conditional mean of $\boldsymbol{\phi}_1$ for a given region as a sensible weighted average of the values of $\boldsymbol{\phi}_2$ for that region *and* a neighborhood of that region.

The GMCAR also allows us to incorporate different weighted adjacency matrices in the $MCAR(\rho, \boldsymbol{\Lambda})$ distribution. Suppose, we wish to extend the precision matrix in model (10.6) to

$$\Sigma^{-1} = \begin{pmatrix} (D_1 - \rho W^{(1)})\Lambda_{11} & (D_3 - \rho W^{(3)})\Lambda_{12} \\ (D_3 - \rho W^{(3)})\Lambda_{12} & (D_2 - \rho W^{(2)})\Lambda_{22} \end{pmatrix} , \tag{10.13}$$

where $D_k = Diag\left(\sum_{j=1}^n W_{1j}^{(k)}, \ldots, \sum_{j=1}^n W_{nj}^{(k)}\right)$ and $W^{(k)}$ is the weighted adjacency matrix with ij-element $W_{ij}^{(k)}$, $k = 1, 2, 3$, and $i, j = 1, \ldots, n$. The conditions for the precision matrix in (10.13) to be positive definite are less obvious. But in our GMCAR case, we obtain

$$\begin{aligned} \boldsymbol{\phi}_1|\boldsymbol{\phi}_2 &\sim N\left((\eta_0 I + \eta_1 W^{(3)})\boldsymbol{\phi}_2\,,\ [\tau_1(D_1 - \rho_1 W^{(1)})]^{-1}\right)\,, \\ \text{and}\ \ \boldsymbol{\phi}_2 &\sim N\left(0,\ [\tau_2(D_2 - \rho_2 W^{(2)})]^{-1}\right)\,. \end{aligned}$$

The conditions for positive definiteness can be easily seen to be $|\alpha_1| < 1$ and $|\alpha_2| < 1$ using the fact that diagonally dominant matrices are always positive definite.

Since we specify the joint distribution for a multivariate MRF directly through specification of simpler conditional and marginal distributions, 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 (i.e., whether to model $p(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2)$ and then $p(\boldsymbol{\phi}_2)$, or $p(\boldsymbol{\phi}_2|\boldsymbol{\phi}_1)$ and then $p(\boldsymbol{\phi}_1)$). This problem is somewhat mitigated 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.

We note that, while the previous theory helps to illuminate structure in specifying multivariate CAR models, from a practical point of view, the reader may simply choose

to implement an analogue of coregionalization to create a multivariate dependence model for areal data. In particular, suppose we assume a common proximity specification for each component of the random effects vector, $\boldsymbol{\phi}$. Then, we could write $\boldsymbol{\phi} = A\boldsymbol{\psi}$ where $\boldsymbol{\psi}_j$, the jth component of $\boldsymbol{\psi}$, is a univariate intrinsic CAR with precision parameter τ_j^2 and each of the component CAR models is independent. As above, we can take A to be lower triangular, with prior specifications discussed earlier. The resulting multivariate CAR model is, of course, improper which is fine as a prior for random effects. Moreover, it is easy to work with since we will only fit the model in the space of the independent CAR models, as we do with the coregionalization model fitting for point-referenced data. We could make the prior proper by making each of the components of $\boldsymbol{\psi}$ proper CAR's, should we wish. Jin, Banerjee and Carlin (2007) develop such coregionalized MCAR distributions, which we discuss in greater detail in Section 10.6. Similar modeling, both in the bivariate and trivariate cases, has been presented in Sang and Gelfand (2009) in the context of modeling temperature extremes (see Section 15.2).

10.4 Modeling using the GMCAR distribution

The $GMCAR(\rho_1, \rho_2, \eta_1, \eta_2, \tau_1, \tau_2)$ models are straightforwardly implemented in a Bayesian framework using MCMC methods. Matters are especially simple with Gaussian likelihoods in the first stage. As a specific example, consider the model

$$Y_{ij} \overset{ind}{\sim} N(Z_{ij}\,,\,\sigma^2)\,,\; i = 1, \ldots, n,\; j = 1, 2\,. \tag{10.14}$$

Assume that $Z_{ij} = \beta_j + \phi_{ij}$, where the ϕ_{ij}'s follow the GMCAR distribution in (10.12). This means that the Z_{ij}'s follow the GMCAR distribution in (10.12) but with $\mathrm{E}[Z_{ij}] = \beta_j$, rather than zero. Then, we easily derive conditional distribution for $\mathbf{Z}_1 \mid \mathbf{Z}_2$

$$\mathbf{Z}_1 \mid \mathbf{Z}_2 \;\sim\; N\left(\beta_1\mathbf{1} + (\eta_0 I + \eta_1 W)(\mathbf{Z}_2 - \beta_2\mathbf{1})\,,\; [\tau_1(D - \rho_1 W)]^{-1}\right)\,,$$

and the marginal distribution $\mathbf{Z}_2 \sim N\left(\beta_2\mathbf{1}\,,\,[\tau_2(D-\rho_2 W)]^{-1}\right)$, where $\mathbf{Z}_1 = (Z_{11}, \ldots, Z_{n1})^T$ and $\mathbf{Z}_2 = (Z_{12}, \ldots, Z_{n2})^T$. Therefore, the joint distribution of $\mathbf{Z}^T = (\mathbf{Z}_1^T, \mathbf{Z}_2^T)$ $p(\mathbf{Z} \mid \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\eta})$ is proportional to

$$\begin{aligned} &\tau_1^{\frac{n}{2}} |D - \rho_1 W|^{\frac{1}{2}} \exp\{-\frac{\tau_1}{2}[\mathbf{Z}_1 - \beta_1\mathbf{1} - (\eta_0 I + \eta_1 W)(\mathbf{Z}_2 - \beta_2\mathbf{1})]' \\ &\quad \times (D - \rho_1 W)[\mathbf{Z}_1 - \beta_1\mathbf{1} - (\eta_0 I + \eta_1 W)(\mathbf{Z}_2 - \beta_2\mathbf{1})]\} \\ &\quad \times \tau_2^{\frac{n}{2}} |D - \rho_2 W|^{\frac{1}{2}} \exp\left[-\frac{\tau_2}{2}(\mathbf{Z}_2 - \beta_2\mathbf{1})'(D - \rho_2 W)(\mathbf{Z}_2 - \beta_2\mathbf{1})\right]\,, \end{aligned} \tag{10.15}$$

where $\boldsymbol{\beta} = (\beta_1, \beta_2)$, $\boldsymbol{\tau} = (\tau_1, \tau_2)$, $\boldsymbol{\eta} = (\eta_0, \eta_1)$, and $\boldsymbol{\alpha} = (\rho_1, \rho_2)$.

The joint posterior distribution $p\left(\boldsymbol{\beta}, \sigma^2, \mathbf{Z}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\eta} \mid \mathbf{Y}_1, \mathbf{Y}_2\right)$ is proportional to

$$L\left(\mathbf{Y}_1, \mathbf{Y}_2 \mid \mathbf{Z}, \sigma^2\right) p\left(\mathbf{Z} \mid \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\eta}\right) p(\boldsymbol{\beta}) p(\boldsymbol{\tau}) p(\boldsymbol{\alpha}) p(\boldsymbol{\eta}) p(\sigma^2)\,, \tag{10.16}$$

where $\mathbf{Y}_1 = (Y_{11}, \ldots, Y_{n1})^T$, $\mathbf{Y}_2 = (Y_{12}, \ldots, Y_{n2})^T$, $L\left(\mathbf{Y}_1, \mathbf{Y}_2 \mid \mathbf{Z}, \sigma^2\right)$ is the likelihood

$$\sigma^{-2n} \exp\left\{-\frac{1}{2\sigma^2}[(\mathbf{Y}_1 - \mathbf{Z}_1)'(\mathbf{Y}_1 - \mathbf{Z}_1) + (\mathbf{Y}_2 - \mathbf{Z}_2)'(\mathbf{Y}_2 - \mathbf{Z}_2)]\right\}\,,$$

$p\left(\mathbf{Z} \mid \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\eta}\right)$ is given by (10.15), and the remaining terms in (10.25) are the prior distributions on $(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\alpha}, \boldsymbol{\eta}, \sigma^2)$.

For the remaining terms, flat priors are chosen for β_1 and β_2, while σ^2 is assigned a vague inverse gamma prior, i.e., a $IG(1, 0.1)$ where we parametrize the $IG(a, b)$ so that $E(\sigma^2) = b/(a-1)$. Next, τ_1 and τ_2 are assigned vague gamma priors, specifically a $G(1, 0.1)$,

which has mean 10 and variance 100. Finally, ρ_1, ρ_2 are given $Unif(0,1)$ priors while η_0 and η_1 are given $N(0,\sigma_1^2)$ and $N(0,\sigma_2^2)$ priors, respectively. Jin et al. (2005) present several simulation studies using these specifications and report robust inference for varying hyperparameter values.

The Gibbs sampler is natural for updating the parameters in this setting because it can take advantage of the conditional specification of the GMCAR model. Each of the full conditional distributions required by the Gibbs sampler must be proportional to (10.25). Furthermore, no matrix inversion is required and only calculations on rather special (e.g., diagonal) n-dimensional matrices are required, regardless of the dimension p ($p = 2$ in our case). To calculate the determinant in (10.15), we have the fact that

$$\begin{aligned} |D - \rho_k W| &= |D^{\frac{1}{2}}(1 - \rho_k D^{-\frac{1}{2}} W D^{-\frac{1}{2}})D^{\frac{1}{2}}| = |D| \prod_{i=1}^{n} (1 - \rho_k \lambda_i) \\ &\propto \prod_{i=1}^{n} (1 - \rho_k \lambda_i)\ ,\ k = 1, 2\ , \end{aligned}$$

where $\lambda_i, i = 1, \ldots n$ are the eigenvalues of the matrix $D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$. The λ_i may be calculated prior to any MCMC iteration. Hence posterior computation for the GMCAR model is simpler and faster than that for existing MCAR models, especially for large areal data sets.

All of the parameters in (10.25) except $\boldsymbol{\eta}$ and $\boldsymbol{\alpha}$ have closed-form full conditionals, and so may be directly updated. For these two remaining parameters, Metropolis-Hastings steps with bivariate Gaussian proposals are convenient (though for $\boldsymbol{\alpha}$, a preliminary logit transformation, having Jacobian $\prod_{k=1}^{2} \rho_k(1 - \rho_k)$, is required). In practice, the ρ_k must be bounded away from 1 (say, by insisting $0 < \rho_k < 0.999$, $k = 1, 2$) to maintain identifiability and hence computational stability.

10.5 Illustration: Fitting conditional GMCAR to Minnesota cancer data

We now use GMCAR distributions as specifications for second-stage random effects in a hierarchical areal data model with a non-Gaussian first stage. Following Jin et al. (2005), we consider modeling the numbers of deaths due to cancers of the lung and esophagus in the years from 1991 to 1998 at the county level in Minnesota. We write the model as

$$Y_{ij} \overset{ind}{\sim} Poisson(E_{ij} e^{Z_{ij}})\ ,\ i = 1, \ldots, 87,\ j = 1, 2\ , \tag{10.17}$$

where Y_{ij} is the observed number of deaths due to cancer j in county i, and E_{ij} is the corresponding expected number of deaths (assumed known). To calculate E_{ij}, we account for each county's age distribution by calculating the expected *age-adjusted* number of deaths due to cancer j in county i as

$$E_{ij} = \sum_{k=1}^{m} \omega_j^k N_i^k\ , i = 1, \ldots, 87,\ j = 1, 2\ ,$$

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. The GMCAR models can be implemented in `BUGS` (see `www.biostat.umn.edu/~brad/software.html` for the code and the data.).

The county-level maps of the raw standardized mortality ratios (i.e., $\text{SMR}_{ij} = Y_{ij}/E_{ij}$) shown in Figure 10.1 exhibit evidence of correlation both across space and between cancers, motivating use of our proposed GMCAR models. Regarding the selection of the proper order

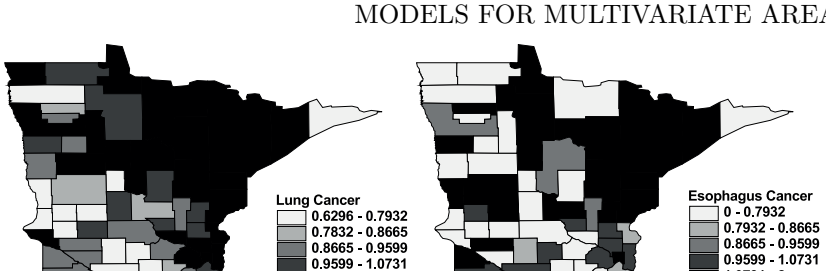

Figure 10.1 *Maps of raw standard mortality ratios (SMR) of lung and esophagus cancer in Minnesota.*

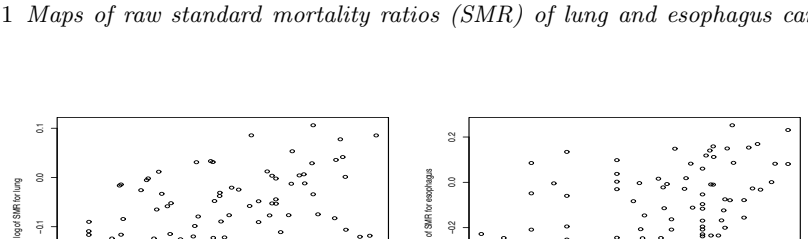

Figure 10.2 *Exploratory plot to help select modeling order: (a) [lung|esophagus], sample correlation 0.394, regression* $t = 3.956$*; (b) [esophagus|lung], sample correlation 0.193, regression* $t = 1.813$.

in which to model the two cancers, Figure 10.2 gives a helpful data-based exploratory plot. We first obtain crude data-based estimates of the spatial random effects as $\hat{\phi}_{i1} = \log(\text{SMR}_{i1})$ and $\hat{\phi}_{i2} = \log(\text{SMR}_{i2})$. Next, recall the linearity of the conditional GMCAR mean for a given ordering (say, lung given esophagus), i.e.,

$$E(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2) = A\boldsymbol{\phi}_2 = A(\eta_0, \eta_1)\boldsymbol{\phi}_2 = (\eta_0 I + \eta_1 W)\boldsymbol{\phi}_2 \, .$$

This motivates obtaining least-squares estimates $\hat{\eta}_0$ and $\hat{\eta}_1$ by minimizing $(\hat{\boldsymbol{\phi}}_1 - A(\eta_0, \eta_1)\hat{\boldsymbol{\phi}}_2)'(\hat{\boldsymbol{\phi}}_1 - A(\eta_0, \eta_1)\hat{\boldsymbol{\phi}}_2)$ as a function of η_0 and η_1. Finally, we plot $A(\hat{\eta}_0, \hat{\eta}_1)\hat{\boldsymbol{\phi}}_2$ versus $\hat{\boldsymbol{\phi}}_1$, and investigate how well the linearity assumption is supported by the data. Repeating this entire process for the reverse order (here, esophagus given lung) produces a second plot, which may be compared in quality to the first. In our case, Figure 10.2(a) (lung given esophagus) indicates more support for linearity, both in its appearance and in its higher sample correlation and regression t statistic.

Using the likelihood in (10.17), we model the random effects Z_{ij} using the $GMCAR(\rho_1, \rho_2, \eta_1, \eta_2, \tau_1, \tau_2)$ with mean $\boldsymbol{\beta}$. In what follows we compare the GMCAR with other existing MCAR models using DIC. In Table 10.3, Models 1–3 are members of our proposed GMCAR class. Specifically, in Model 1, we have the full model with all six parameters, and the conditioning order of the cancers is [lung | esophagus]. Model 2 assumes $\eta_1 = 0$ and uses the same conditioning order as Model 1. In Model 3, we switch the conditioning order to [esophagus | lung] and return to a full model. To compare the GMCAR to existing MCAR models, we take the $MCAR(\rho_1, \rho_2, \Lambda)$ using the Cholesky method for the U_k as Model 4, the same model but using the spectral decomposition for the U_k as Model 5, and the $2fCAR(\rho_0, \rho_1, \rho_2, \rho_3, \tau_1, \tau_2)$ as Model 6. We choose the prior distributions for each parameter as discussed in Section 10.4, and use Metropolis-Hastings and Gibbs sampling to

	model	$\overline{D}$	p_D	DIC
1	GMCAR (full)	483.4	58.2	541.6
2	GMCAR (reduced; $\eta_1 = 0$)	483.0	63.8	546.8
3	GMCAR (full, reverse order)	480.6	63.3	543.9
4	MCAR (Cholesky decomposition)	483.6	61.3	544.9
5	MCAR (spectral decomposition)	483.8	60.6	544.4
6	2fCAR	482.6	65.1	547.7

Table 10.3 *Model comparison using DIC statistics, Minnesota cancer data analysis.*

update all parameters. We use 5,000 iterations as the pre-convergence burn-in period, and then a further 20,000 iterations as our production run for posterior summarization.

Fit measures $\overline{D}$, effective numbers of parameters p_D, and DIC scores for each model are seen in Table 10.3. Model 1 has the smallest p_D and DIC values, so our $GMCAR(\rho_1, \rho_2, \eta_0, \eta_1, \tau_1, \tau_2)$ full model with the conditioning order [lung | esophagus] emerges as best for this data set. The reduced GMCAR Model 2 does less well, suggesting the need to account for bivariate spatial structure in these data. The two MCAR methods perform similarly to each other and to the reduced GMCAR model, while the 2fCAR model does less well, largely because it does not seem to allow sufficient smoothing of the random effects (larger p_D score). Note that effective degrees of freedom may actually be smaller for apparently more complex models that allow more complicated forms of shrinkage, such as Model 1 in this case. We note that our "focus" parameter is the same for each model (both fixed and random effects are in focus), and the Poisson likelihood is also not changing across models. Also, our priors are all noninformative or quite vague (e.g., uniform priors for all ρ parameters). All of this suggests the DIC comparison in Table 10.3 is fair across models. Moreover, the resulting DIC scores were robust to the moderate changes in the prior distributions.

Regarding estimation of the fixed effects, under Model 1 we obtained point and 95% equal-tail interval estimates of 0.602 and (0.0267, 0.979) for ρ_1, and 0.699 and (0.0802, 0.973) for ρ_2. Recall these are spatial association parameters, but while their values are between 0 and 1 they are not "correlations" in the usual sense; the moderate point estimates and wide confidence intervals suggest a relatively modest degree of spatial association in the random effects. It is also important to remember that in this setup, ρ_2 measures spatial association in the esophagus random effects $\boldsymbol{\phi}_2$, while ρ_1 measures spatial association in the lung random effects $\boldsymbol{\phi}_1$ *given* the esophagus random effects $\boldsymbol{\phi}_2$. Thus the interpretation of the ρ_k would be different for Model 3 (due to the different conditioning order), and much different for Model 4 or 5. Note that for the MCAR model, $E(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2)$ and $E(\boldsymbol{\phi}_2|\boldsymbol{\phi}_1)$ both depend on both ρ_1 and ρ_2. But for the GMCAR, $E(\boldsymbol{\phi}_1|\boldsymbol{\phi}_2)$ is free of both ρ_1 and ρ_2, while of course $E(\boldsymbol{\phi}_2) = 0$. Thus for this model, ρ_1 and ρ_2 unambiguously control only their corresponding variance matrices, and can be set without altering the mean structure.

Turning to τ_1 and τ_2, under Model 1 we obtained 32.65, (16.98, 66.71) and 13.73, (4.73, 38.05) as our point and interval estimates, respectively. Since these parameters measure spatial precision for each disease, they suggest slightly more variability in the esophagus random effects, although again comparison is difficult here since τ_2 is a *marginal* precision for $\boldsymbol{\phi}_2$ while τ_1 is a *conditional* precision for $\boldsymbol{\phi}_1$ given $\boldsymbol{\phi}_2$. Along these lines, Figure 10.3 shows estimated posteriors of the conditional variances $\sigma_1^2 = 1/\tau_1$ for several candidate multivariate spatial models. Panel (a) shows the situation for two separate CAR models, a model that ignores any possibility of connection between the cancers. The remaining panels consider the $MCAR(\rho_1, \rho_2, \boldsymbol{\Lambda})$ model, the reduced $GMCAR(\rho_1, \rho_2, \eta_0, \tau_1, \tau_2)$ model, and the full $GMCAR(\rho_1, \rho_2, \eta_0, \eta_1, \tau_1, \tau_2)$ model. The reduction of uncertainty in $\boldsymbol{\phi}_1$ given $\boldsymbol{\phi}_2$ in these more complex models is a measure of the information content between the cancers, and is readily apparent from the histograms and their empirical means.

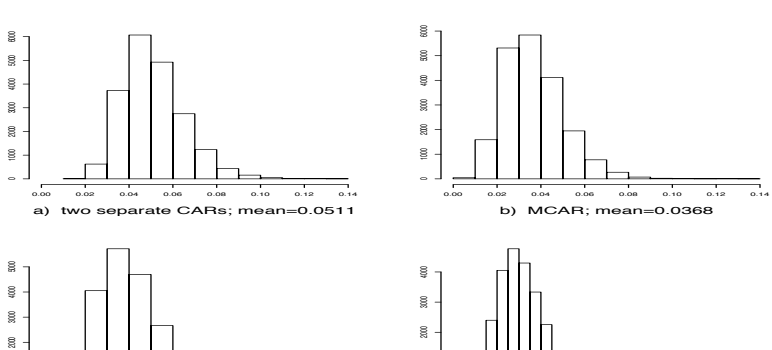

Figure 10.3 *Posterior samples of conditional variances* $\sigma_1^2 = 1/\tau_1$ *for various models: (a) two separate CAR models; (b) MCAR model; (c) reduced GMCAR model; (d) full GMCAR model.*

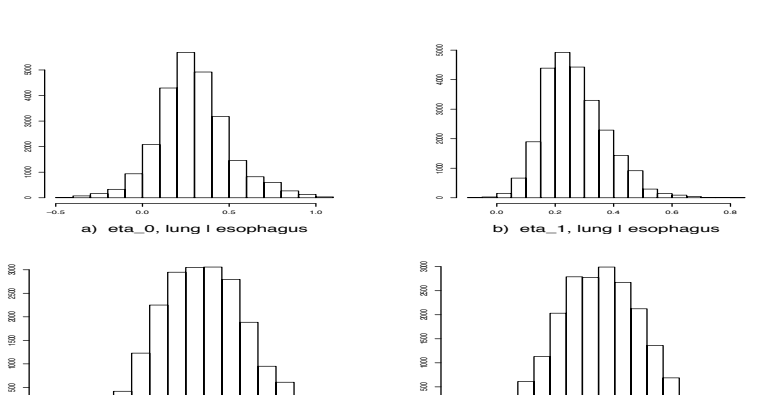

Figure 10.4 *Posterior samples of* η_0 *and* η_1 *using the full GMCAR model with two conditioning orders: (a) estimated posterior for* η_0*, [lung | esophagus]; (b) estimated posterior for* η_1*, [lung | esophagus]; (c) estimated posterior for* η_0*, [esophagus | lung]; (d) estimated posterior for* η_1*, [esophagus | lung].*

DIC's slight preference for Model 1 is consistent with the estimated posteriors of the linking parameters η_0 and η_1 shown in Figure 10.4. The inclusion of 0 within the 95% credible interval for η_1 under the reverse ordering, but not under the natural ordering, is yet further evidence against the former. Note also that the linking parameters η_0 and η_1 have mostly positive support, meaning that the two cancers have positive spatial correlation. This is also evident from the maps of the posterior means of the SMRs for the two cancers under the full model shown in Figure 10.5. Clearly incidence of the two cancers is strongly correlated, with higher fitted ratios extending from the Twin Cities metro area (eastern side, about one third of the way up) to the mining- and tourism-oriented north and northeast, regions where conventional wisdom suggests cigarette smoking may be more common.

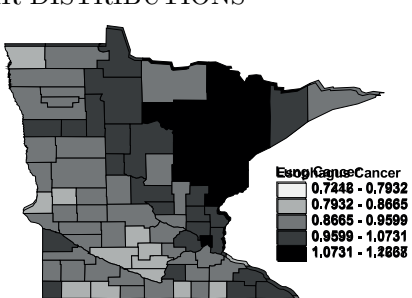

Figure 10.5 *Maps of posterior means of standardized mortality ratios (SMR) of lung and esophagus cancer in Minnesota from the full GMCAR model with conditioning order [lung | esophagus].*

10.6 Coregionalized MCAR distributions

As mentioned in Section 10.3, generalizations of GMCAR to settings with a large number of diseases is encumbered by the dependence of the joint distribution on the sequence of ordering in the hierarachy. To obviate this issue, Jin, Banerjee and Carlin (2007) develop an order-free framework for multivariate areal modeling that allows versatile spatial structures, yet is computationally feasible for many outcomes. This approach is based upon an adaptation of the *linear model of coregionalization* (LMC) to areal data.

The essential idea is to develop richer spatial association models using linear transformations of much simpler spatial distributions. The objective is to 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 U_k's as in (10.6). Unfortunately, except in the separable model with only one smoothing parameter ρ, constructing such dispersion structures is not trivial and leads to identifiability issues on the γ'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. We address this problem using a *linear model of coregionalization* (LMC).

It is worth pointing out that our use of the LMC here is somewhat different from what is 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 10.6.2, independent latent effects produce the $MCAR(\alpha_1, \ldots, \alpha_p; \Lambda)$ in (10.9).

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 $\boldsymbol{\phi} = (\boldsymbol{\phi}_1^T, \ldots, \boldsymbol{\phi}_p^T)^T$ be an $np \times 1$ vector, where each $\boldsymbol{\phi}_j = (\phi_{1j}, \ldots, \phi_{nj})^T$ 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 $\mathbf{u} = (\mathbf{u}_1^T, \ldots, \mathbf{u}_p^T)^T$ 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 $\mathbf{u}_j$.

10.6.1 Case 1: Independent and identical latent CAR variables

First, we will assume that the random spatial processes $\mathbf{u}_j, j = 1, \ldots, p$, are independent and identical. Since each spatial process $\mathbf{u}_j$ is a univariate process over areal units, we might

adopt a CAR structure for each of them, that is

$$\mathbf{u}_j \sim N_n\left(\mathbf{0},\ (D-\alpha W)^{-1}\right),\ j=1,\ldots,p\,. \tag{10.18}$$

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

$$\boldsymbol{\phi} \sim N_{np}\left(\mathbf{0},\ \Sigma\otimes(D-\alpha W)^{-1}\right), \tag{10.19}$$

defining $\Sigma=AA^T$. We denote the distribution in (10.19) by $MCAR(\alpha,\ \Sigma)$. Note that the joint distribution of (10.19) 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 Σ.

Since $\boldsymbol{\phi}=(A\otimes I_{n\times n})\mathbf{u}$, a valid joint distribution of $\boldsymbol{\phi}$ 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 ξ_{min} and ξ_{max} are the minimum and maximum eigenvalues of $D^{-\frac{1}{2}}WD^{-\frac{1}{2}}$. Note if $\alpha=1$ in CAR structure (10.18), which is an ICAR, the joint distribution of $\boldsymbol{\phi}$ in (10.19) becomes the multivariate intrinsic CAR (Gelfand and Vounatsou, 2003).

Currently, `BUGS` offers an implementation of 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 `BUGS` by writing $\boldsymbol{\phi}=(A\otimes I_{n\times n})\mathbf{u}$ and assigning proper CAR priors (via the `car.proper` distribution) for each $\mathbf{u}_j$, $j=1,\ldots,p$ with a common smoothing parameter α. Regarding the prior on A, note that since $AA'=\Sigma$ and A is the Cholesky decomposition of Σ, there is a one-to-one relationship between the elements of Σ and A. In Section 10.7, we argue that assigning a prior to Σ is computationally preferable.

10.6.2 Case 2: Independent but not identical latent CAR variables

Here, we continue to assume that the $\mathbf{u}_j$ are independent, but relax them being identically distributed. Adopting the CAR structure, we assume

$$\mathbf{u}_j \sim N_n\left(\mathbf{0},\ (D-\alpha_j W)^{-1}\right),\ j=1,\ldots,p\,, \tag{10.20}$$

where α_j is the smoothing parameter for the jth 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

$$\boldsymbol{\phi} \sim N_{np}\left(\mathbf{0},\ (A\otimes I_{n\times n})\Gamma^{-1}(A\otimes I_{n\times n})^T\right), \tag{10.21}$$

where $\Sigma=AA^T$ and Γ is an $np\times np$ block diagonal matrix with $n\times n$ diagonal entries $\Gamma_j=D-\alpha_j W$, $j=1,\ldots,p$. We denote the distribution in (10.21) by $MCAR(\alpha_1,\ldots,\alpha_p,\Sigma)$.

It follows from (10.21) 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 Σ, although this might not be the best choice computationally. Through the LMC approach in this case, the distribution in (10.21) is similar to the $MCAR(\alpha_1,\ldots,\alpha_p,\ \Lambda)$ structure (10.9), 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 $MCAR(\alpha_1,\ldots,\alpha_p,\ \Lambda)$ structure, since there is not a unique R_j matrix such that $R_jR_j^T=R_jPP^TR_j^T=D-\alpha_jW$ (P being an arbitrary orthogonal matrix).

Again, a valid joint distribution in (10.21) requires p valid distributions for $\mathbf{u}_j$, i.e., $\frac{1}{\xi_{min}}<\alpha_j<\frac{1}{\xi_{max}}$, $j=1,\ldots,p$. Through the LMC approach, we can also fit the data

with the $MCAR(\alpha_1, \ldots, \alpha_p, \Sigma)$ prior distribution (10.21) 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 α_j for each $\mathbf{u}_j,\ j = 1, \ldots, p$. As mentioned in the preceding section, we assign a prior to $AA^T = \Sigma$ (e.g., an inverse Wishart), and determine A from the one-to-one relationship between the elements of Σ and A; Section 10.7 below provides details.

10.6.3 Case 3: Dependent and not identical latent CAR variables

Finally, in this case we will assume that the random spatial processes $\mathbf{u}_j = (u_{1j}, \ldots, u_{nj})^T,\ j = 1, \ldots, p$ 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 $l, j = 1, \ldots, p$ and $i, k = 1, \ldots, n$ 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 in the univariate case, we specify the ij^{th} conditional distribution as Gaussian with mean

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

and conditional variance $Var(u_{ij} \mid u_{k\neq i,\, j},\ u_{i,\, l\neq j},\ u_{k\neq i,\, l\neq j}) \propto 1/m_i$, where b_{jj} denotes the spatial autocorrelation for the random spatial process $\mathbf{u}_j$ while $b_{jl}\ (l \neq j,\ l, j = 1, \ldots, p)$ 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 $\mathbf{u} = (\mathbf{u}_1^T, \ldots, \mathbf{u}_p^T)^T$ to be

$$\mathbf{u} \ \sim\ N_{np}\left(\mathbf{0},\ (I_{p\times p} \otimes D - B \otimes W)^{-1}\right), \tag{10.22}$$

where I is a $p \times p$ identity matrix and B is a $p \times p$ symmetric matrix with the elements $b_{jl}, j, l = 1, \ldots, p$. As long as the dispersion matrix in (10.22) is positive definite, which boils down to $(I_{p\times p} \otimes D - B \otimes W)$ being positive definite, (10.22) 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 $\xi_i,\ i = 1, \ldots, n$, and the eigenvalues for B as $\zeta_j,\ j = 1, \ldots, p$, 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 $\xi_i \times \zeta_j,\ i = 1, \ldots, n,\ j = 1, \ldots, p$. Hence, the conditions for $I_{p\times p} \otimes D - B \otimes W$ being positive definite become $\xi_i\zeta_j < 1,\ i.e.,\ \frac{1}{\xi_{min}} < \zeta_j < \frac{1}{\xi_{max}},\ i = 1, \ldots, n,\ j = 1, \ldots, p$, where ξ_{min} and ξ_{max} are the minimum and maximum eigenvalues of $D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$. Thus, $\frac{1}{\xi_{min}} < \zeta_j < 1,\ j = 1, \ldots, p$, 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 (10.22). In fact, $\xi_{max} = 1$ and $\xi_{min} < 0$, which makes this formulation easier to work with in practice (e.g., in choosing priors; see Section 10.7) than the alternative parametrization $\frac{1}{\zeta_{min}} < \xi_j < \frac{1}{\zeta_{max}}$.

The model in (10.22) introduces smoothing parameters in the cross-covariance structure through the matrix B, but unlike the $MCAR$ models in Sections 10.6.1 and 10.6.2 does not have the Σ matrix to capture non-spatial variances. To remedy this, we model $\boldsymbol{\phi} = (A \otimes I_{n\times n})\mathbf{u}$ so that the joint distribution for the random effects $\boldsymbol{\phi}$ is

$$\boldsymbol{\phi} \ \sim\ N_{np}\left(\mathbf{0},\ (A \otimes I_{n\times n})\,(I_{p\times p} \otimes D - B \otimes W)^{-1}\,(A \otimes I_{n\times n})^T\right). \tag{10.23}$$

Since $\boldsymbol{\phi} = (A \otimes I_{n\times n})\mathbf{u}$, it is immediately that the validity of (10.22) ensures a valid joint distribution for (10.23). We denote distribution (10.23) by $MCAR(B, \Sigma)$, where $\Sigma = AA^T$. Again, A identifies with the upper-triangular Cholesky square-root of Σ. Note that with $\Sigma = I$ we recover (10.22), which we henceforth denote as $MCAR(B, I)$.

To see the generality of (10.23), we find the joint distribution of $\boldsymbol{\phi}$ reduces to the $MCAR(\alpha_1, \ldots, \alpha_p, \Sigma)$ distribution (10.21) if $b_{jl} = 0$ and $b_{jj} = \alpha_j$, or the $MCAR(\alpha, \Sigma)$ distribution (10.19) if $b_{jl} = 0$ and $b_{jj} = \alpha$, in both cases for $j, l = 1, \ldots, p$. Also note that the distribution in (10.23) 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 $TT^T = APP^TA^T = \Sigma$. Then the covariance matrix in (10.23) can be expressed as $(A \otimes I_{n\times n})\left(I_{p\times p} \otimes D - B \otimes W\right)^{-1}(A \otimes I_{n\times n})^T = (T \otimes I_{n\times n})\left(I_{p\times p} \otimes D - C \otimes W\right)^{-1}(T \otimes I_{n\times n})^T$, where $C = P^TBP$. Without loss of generality, then, we can choose the matrix A as the upper-triangular Cholesky decomposition of Σ.

To understand the features of the $MCAR(B, \Sigma)$ distribution (10.23), we illustrate in the bivariate case ($p = 2$). Define

$$(AA^T)^{-1} = \Sigma^{-1} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{12} & \Lambda_{22} \end{pmatrix}$$

and $B = A^T \begin{pmatrix} \gamma_1\Lambda_{11} & \gamma_{12}\Lambda_{12} \\ \gamma_{12}\Lambda_{12} & \gamma_2\Lambda_{22} \end{pmatrix} A$, where $A = \begin{pmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{pmatrix}$. For convenience, we will denote the entries of B as b_{ij}. Note that the γ's are not identifiable from the matrix Λ and our reparametrization in terms of B must be used to conduct posterior inference on B and Λ (see Section 10.7), from which the cross-covariances may be recovered. The above expression does allow the $MCAR(B, \Sigma)$ distribution (10.23) to be rewritten as

$$\boldsymbol{\phi} \sim N_{2n}\left(\mathbf{0}, \begin{pmatrix} (D-\gamma_1 W)\Lambda_{11} & (D-\gamma_{12}W)\Lambda_{12} \\ (D-\gamma_{12}W)\Lambda_{12} & (D-\gamma_2 W)\Lambda_{22} \end{pmatrix}^{-1}\right), \tag{10.24}$$

which is precisely the general dispersion structure we set out to achieve. Jin et al. (2007) provide explicit expressions for the conditional means and variances, which offer further insight into how the parameters in (10.24) affect smoothing.

10.7 Modeling with coregionalized MCAR's

The $MCAR(B, \Sigma)$ model is straightforwardly implemented in a Bayesian framework using MCMC methods. As in Section 10.6.3, we write $\boldsymbol{\phi} = (A \otimes I_{n\times n})\mathbf{u}$, where $\mathbf{u} = (\mathbf{u}_1^T, \mathbf{u}_2^T)$ and $\mathbf{u}_j = (u_{1j}, \ldots, u_{nj})^T$. The joint posterior distribution is $p\left(\boldsymbol{\beta}, \sigma^2, \mathbf{u}, A, B \mid \mathbf{Y}_1, \mathbf{Y}_2\right)$, which is proportional to

$$L\left(\mathbf{Y}_1, \mathbf{Y}_2 \mid \mathbf{u}, \sigma^2, A\right) p\left(\mathbf{u} \mid B\right) p(B)p(\boldsymbol{\beta})p(A)p(\sigma^2)\,, \tag{10.25}$$

where $\mathbf{Y}_1 = (Y_{11}, \ldots, Y_{n1})^T$ and $\mathbf{Y}_2 = (Y_{12}, \ldots, Y_{n2})^T$, $L\left(\mathbf{Y}_1, \mathbf{Y}_2 \mid \mathbf{u}, \sigma^2, A\right)$ is the data likelihood and $p\left(\mathbf{u} \mid B\right) = N_{np}\left(\mathbf{0},\ (I_{p\times p} \otimes D - B \otimes W)^{-1}\right)$. As mentioned in Section 10.6.3, propriety of this distribution requires the eigenvalues ζ_j of B to satisfy $\frac{1}{\xi_{min}} < \zeta_j < 1$ $(j = 1, \ldots, p)$. 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 $B = P\Delta P^T$, where P is the corresponding orthogonal matrix of eigenvectors and Δ is a diagonal matrix of

ordered eigenvalues, $\zeta_1, \ldots, \zeta_p$. We parameterize the $p \times p$ orthogonal matrix P in terms of the $p(p-1)/2$ Givens angles θ_{ij} for $i = 1, \ldots, p-1$ and $j = i+1, \ldots, p$ (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, $P = G_{12}G_{13}\ldots G_{1p}\ldots G_{(p-1)p}$ where i and j are distinct and G_{ij} is the $p \times p$ identity matrix with the ith and jth 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 θ_{ij} are unique with a domain $(-\pi/2, \pi/2)$ and the eigenvalues ζ_j of B are in the range $(\frac{1}{\xi_{min}}, 1)$, we then put a Uniform$(-\pi/2, \pi/2)$ prior on the θ_{ij} and a Uniform$(\frac{1}{\xi_{min}}, 1)$ prior on the ζ_j. To update θ_{ij}'s or ζ_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 ζ_j must be bounded away from 1 (say, by insisting $\frac{1}{\xi_{min}} < \zeta_j < 0.999$, $j = 1, \ldots, p$) 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 ξ_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 $MCAR(\alpha_1, \ldots, \alpha_p, \Sigma)$ 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 α_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 (10.25), 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 Σ than to put such priors on the elements of A in terms of letting the data drive the inference and obtaining good convergence. Since Σ is a positive definite covariance matrix, the inverse Wishart prior distribution renders itself as a natural choice, that is, $\Sigma^{-1} \sim Wishart\left(\nu, (\nu R)^{-1}\right)$ (see, e.g., Carlin and Louis, 2000, p. 328). Hence, we instead place a prior directly on Σ, and then use the one-to-one relationship between the elements of Σ and the Cholesky factor A. Then, the prior distribution $p(A)$ becomes

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

where $\left|\frac{\partial\Sigma}{a_{ij}}\right|$ is the Jacobian $2^p\prod_{i=1}^{p}a_{ii}^{p-i+1}$. For example, when $p = 2$, the Jacobian is $4a_{22}^2a_{11}$. Rather than updating Σ 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 ν and R in the $Wishart\left(\nu, (\nu R)^{-1}\right)$, since $E(\Sigma^{-1}) = R^{-1}$, if there is no information about the prior mean structure of Σ, 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 $R = Diag(0.1, 0.1)$. Finally, for the remaining terms on the right hand side of (10.25), flat priors are chosen for β_1 and β_2, while σ^2 is assigned a vague inverse gamma prior, i.e., an $IG(1, 0.01)$ parameterized so that $E(\sigma^2) = b/(a-1)$. In this study, $\boldsymbol{\beta}$ and σ^2 have closed-form full conditionals, and so can be directly updated using Gibbs sampling.

10.8 Illustrating coregionalized MCAR models with three cancers from Minnesota

Jin, Banerjee and Carlin (2007) estimate different coregionalized MCAR models methods with a data set consisting 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 159,260 Americans will die in 2014 from lung cancer, accounting for 27% 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`).

Following Jin et al. (2007), we estimate the model

$$Y_{ij} \overset{ind}{\sim} Po(E_{ij}e^{\mu_{ij}}),\ i = 1, \ldots, n,\ j = 1, 2, 3 \tag{10.26}$$

where Y_{ij} is the observed number of cases of cancer type j (one of three types) in region i, $\log \mu_{ij} = \beta_j + \phi_{ij}$ with the β_j's being cancer-specific intercepts and the ϕ_{ij}'s being spatial random effects that are distributed according to some version of the coregionalized MCAR's we discussed earlier. 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 $E_{ij} = \sum_{k=1}^{m} \omega_j^k N_i^k$,$i = 1, \ldots, 87,\ j = 1, 2, 3,\ k = 1, \ldots, 18,$ 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.

The county-level maps of the raw age-adjusted standardized mortality ratios (i.e., $\text{SMR}_{ij} = Y_{ij}/E_{ij}$) shown in Figure 10.6 exhibit evidence of correlation both across space and among the cancers, motivating use of our proposed multivariate lattice model. Using the likelihood in (10.26), we model the random effects ϕ_{ij} using our proposed $MCAR(B, \Sigma)$ model (10.23). In what follows we compare it with other MCAR models, including the $MCAR(\alpha, \Sigma)$ and $MCAR(1, \Sigma)$ from Section 10.6.1, a "three separate CAR's" 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 (10.21) of Section 10.6.2 by choosing the matrix A as the upper-triangular Cholesky decomposition of Σ. Note that we do not consider the order-specific GMCAR model (Section 10.3), since with no natural causal order for these three cancers, it is hard to choose among the six possible conditioning orders.

For priors, we follow the guidelines outlined earlier and use the same specifications as in Jin et al. (2007). Since $p = 3$ in this example, we choose the inverse Wishart distribution with $\nu = 3$ and $R = Diag(0.1, 0.1, 0.1)$ for Σ. For a model comparisons using DIC, 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. To see the relative performance of these models, we use DIC. As in the previous section, the deviance is the same for the models we wish to compare since they differ only in their random effect distributions $p(\boldsymbol{\phi}|B, \Sigma)$.

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 3×3 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 (10.21) with a different smoothing

	model	$\overline{D}$	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 10.4 *Model comparison using DIC statistics, Minnesota cancer data analysis.*

parameter for each cancer. Model 3 assumes a common smoothing parameter α and Model 4 fits the three separate univariate CAR model, while Model 6 is the trivariate i.i.d. model. Fit measures $\overline{D}$, effective numbers of parameters p_D, and DIC scores for each model are seen in Table 10.4. We find that the $MCAR(B, \Sigma)$ model has the smallest $\overline{D}$ 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

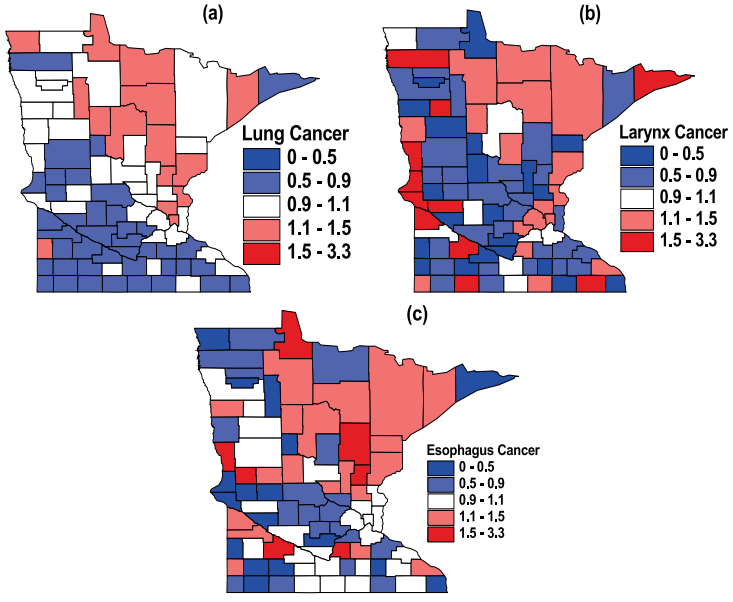

Figure 10.6 *Maps of raw standardized mortality ratios (SMR) of lung, larynx and esophagus cancer in the years from 1990 to 2000 in Minnesota.*

	Lung median (2.5%, 97.5%)	Larynx median (2.5%, 97.5%)	Esophagus mean (2.5%, 97.5%)
β_1, β_2, β_3	–0.093 (–0.179,–0.006)	–0.128 (–0.316, 0.027)	–0.080 (–0.194, 0.025)
Σ_{11}, Σ_{22}, Σ_{33}	0.048 (0.030, 0.073)	0.173 (0.054, 0.395)	0.107 (0.044, 0.212)
ρ_{12}, ρ_{13}		0.277 (–0.112, 0.643)	0.378 (–0.022, 0.716)
ρ_{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 10.5 *Posterior summaries of parameters in $MCAR(B, \Sigma)$ model for Minnesota cancer data.*

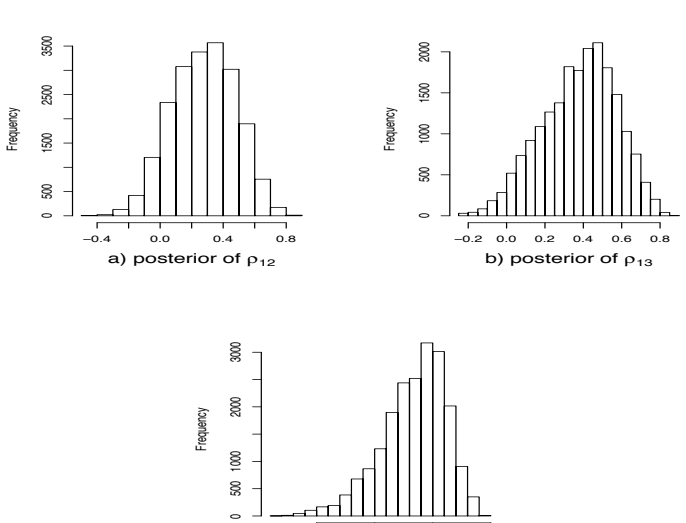

Figure 10.7 *Posterior samples of ρ_{12}, ρ_{13} and ρ_{23} in the Minnesota cancer data analysis using the $MCAR(B, \Sigma)$ model: (a) estimated posterior for correlation ρ_{12} between lung and larynx; (b) estimated posterior for correlation ρ_{13} between lung and esophagus; (c) estimated posterior for correlation ρ_{23} between larynx and esophagus.*

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 ϕ_{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.

We summarize our results from the $MCAR(B, \Sigma)$, which is Model 1 in Table 10.4. Table 10.5 provides posterior means and associated standard deviations for the parameters $\boldsymbol{\beta}$, Σ and b_{ij} in this model, where b_{ij} is the element of the symmetric matrix B. Instead of

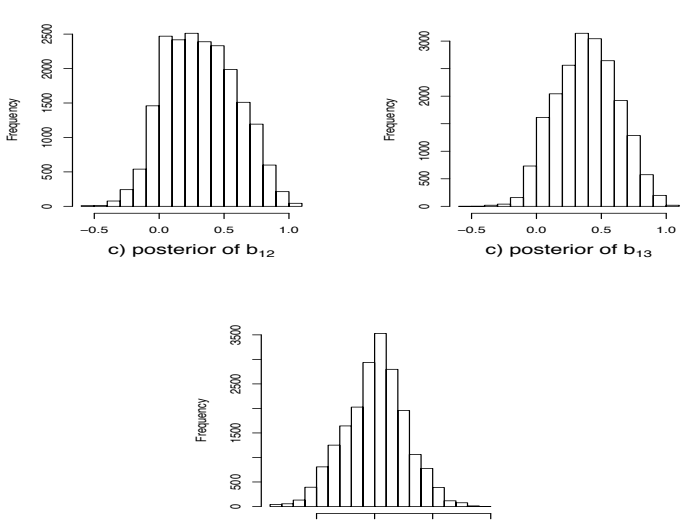

Figure 10.8 *Posterior samples of b_{12}, b_{13} and b_{23} in the Minnesota cancer data analysis using the $MCAR(B, \Sigma)$ model: (a) estimated posterior for b_{12}; (b) estimated posterior for b_{13}; (c) estimated posterior for b_{23}.*

reporting Σ_{12}, Σ_{13} and Σ_{23}, we provide the mean and associated standard deviations for the correlation parameters ρ_{12}, ρ_{13} and ρ_{23}, which are calculated as $\rho_{ij} = \Sigma_{ij}/\sqrt{\Sigma_{ii}\Sigma_{jj}}$. We also plot histograms of the posterior samples ρ_{ij} in Figure 10.7, and histograms of the posterior samples b_{ij} in Figure 10.8.

Table 10.5 and Figure 10.7 reveal correlations between cancers, in particular a strong correlation between lung and esophagus (ρ_{13}). This might explain why the DIC scores for Models 1–4 in Table 10.5 are smaller than that under the separate CAR model. The b_{ij} in Table 10.5 are spatial autocorrelation and cross-spatial correlation parameters for the latent spatial processes $\mathbf{u}_j$, $j = 1, 2, 3$. Figure 10.8 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 10.4, these suggest it is worth fitting our proposed $MCAR(B, \Sigma)$ model to these data.

Turning to geographical summaries, Figure 10.9 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 10.9, 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 10.6, the range of the raw SMRs is seen to be from 0 to 3.3, while in Figure 10.9, the range of the fitted SMRs is from 0.7 to 1.3, due to spatial shrinkage in the random effects.

10.9 Exercises

1. The usual and generalized (but still proper) MCAR models may be constructed using linear transformations of some nonspatially correlated variables. Consider a vector blocked by components, say $\boldsymbol{\phi} = \left(\boldsymbol{\phi}_1^T, \boldsymbol{\phi}_2^T\right)^T$, where each $\boldsymbol{\phi}_i$ is $n \times 1$, n being the number of areal

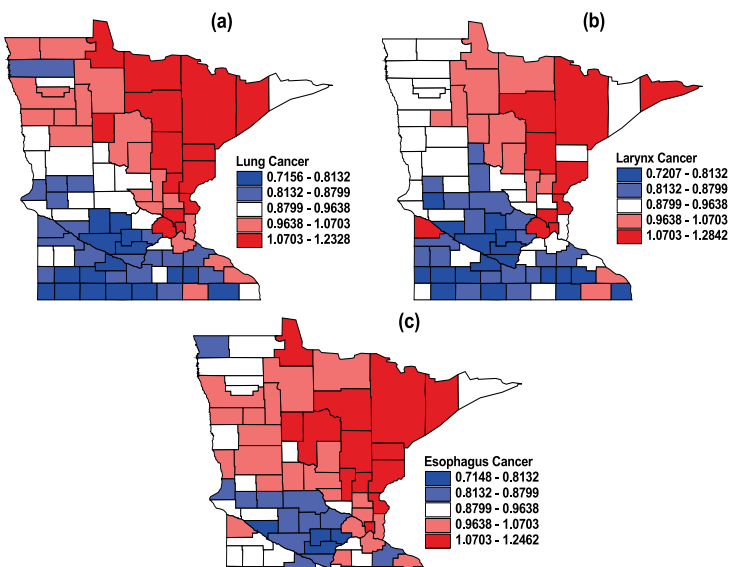

Figure 10.9 *Maps of posterior means of the fitted standard mortality ratios (SMR) of lung, larynx and esophagus cancer in the years from 1990 to 2000 in Minnesota from $MCAR(B, \Sigma)$ model.*

units. Suppose we look upon these vectors as arising from linear transformations

$$\boldsymbol{\phi}_1 = A_1\mathbf{v}_1 \text{ and } \boldsymbol{\phi}_2 = A_2\mathbf{v}_2 \ ,$$

where A_1 and A_2 are any $n \times n$ matrices, $\mathbf{v}_1 = (v_{11}, \ldots, v_{1n})^T$ and $\mathbf{v}_2 = (v_{21}, \ldots, v_{2n})^T$ with covariance structure

$$Cov\,(v_{1i}, v_{1j}) = \lambda_{11} I_{[i=j]},\ Cov\,(v_{1i}, v_{2j}) = \lambda_{12} I_{[i=j]},$$
$$\text{and } Cov\,(v_{2i}, v_{2j}) = \lambda_{22} I_{[i=j]} \ ,$$

where $I_{[i=j]} = 1$ if $i = j$ and 0 otherwise. Thus, although $\mathbf{v}_1$ and $\mathbf{v}_2$ are associated, their nature of association is nonspatial in that covariances remain same for every areal unit, and there is no association between variables in different units.

(a) Show that the dispersion matrix $\Sigma_{(\mathbf{v}_1, \mathbf{v}_2)}$ equals $\Lambda \otimes I$, where $\Lambda = (\lambda_{ij})_{i,j=1,2}$.

(b) Show that setting $A_1 = A_2 = A$ yields a separable covariance structure for $\boldsymbol{\phi}$. What choice of A would render a separable MCAR model, analogous to (10.4)?

(c) Show that appropriate (different) choices of A_1 and A_2 yield the generalized MCAR model with covariance matrix given by (10.7).

2. Derive the covariance matrix in (10.12) for the bivariate $GMCAR(\rho_1, \rho_2, \eta_1, \eta_2, \tau_1, \tau_2)$ model.

3. Show that the covariance matrix in (10.22) can be expressed as:

$$(I_{p\times p} \otimes D)^{\frac{1}{2}} \left(I_{pn\times pn} - B \otimes D^{-\frac{1}{2}} W D^{-\frac{1}{2}}\right) (I_{p\times p} \otimes D)^{\frac{1}{2}} \ .$$

4. Prove that $|I_{p\times p} \otimes D - B \otimes W| \propto \prod_{i=1}^{n} \prod_{j=1}^{p} (1 - \xi_i \zeta_j)$, where ξ_i's are the eigenvalues of $D^{-\frac{1}{2}} W D^{-\frac{1}{2}}$ and ζ_i's are the eigenvalues of B.