Biostatistics (2003), **4**, 4, *pp.* 569–582
Printed in Great Britain

Generalized common spatial factor model

FUJUN WANG
Eli Lilly and Company, Indianapolis, IN 46285, USA
MELANIE M. WALL[†]
Division of Biostatistics, University of Minnesota, A460 Mayo MMC 303, 420 Delaware Strt. S.E., Minneapolis, MN 55455, USA
melanie@biostat.umn.edu

SUMMARY

There are often two types of correlations in multivariate spatial data: correlations between variables measured at the same locations, and correlations of each variable across the locations. We hypothesize that these two types of correlations are caused by a common spatially correlated underlying factor. Under this hypothesis, we propose a generalized common spatial factor model. The parameters are estimated using the Bayesian method and a Markov chain Monte Carlo computing technique. Our main goals are to determine which observed variables share a common underlying spatial factor and also to predict the common spatial factor. The model is applied to county-level cancer mortality data in Minnesota to find whether there exists a common spatial factor underlying the cancer mortality throughout the state.

Keywords: Bayesian; Deviance information criterion; Factor analysis; Latent; Markov chain Monte Carlo (MCMC).

1. INTRODUCTION

When several variables are measured at the same locations over a spatial area, they are often correlated with each other. Each of the variables might also be correlated across the locations due to geographic similarities of the different locations. This type of multivariate spatially referenced data is commonly seen in public health or environmental protection research. The literature on multivariate spatial data analysis has been dominated by methods for predicting the value of an outcome variable at an unobserved location by borrowing information from not just the outcome variable at known locations, but also all other variables in the multivariate process: see e.g. Gotway and Hartford (1996) and Le *et al.* (1997) for continuous spatial processes, and Desouza (1992) and Kim *et al.* (2001) for lattice or areal spatial process. Sometimes, our interest is not in any single variable measured, but in what is behind these variables. That is, we treat these variables as indicators of a latent variable of interest that is difficult or impossible to measure directly. We hypothesize that the correlations between variables within locations and the correlations across locations for each variable are caused by the same latent spatial factor. Our purpose here is to find which variables share the latent spatial factor and predict the common spatial factor underlying these variables.

One example of this type of data is the Minnesota mortality data. We find, for example, that if a county has a high death rate for one kind of cancer, it most likely has high rates for some other kinds of cancers.

[†]To whom correspondence should be addressed

We also find that neighboring counties have similar cancer rates. Therefore, the disease rates are correlated both within county and across counties. One reasonable way to explain the two types of correlations in this multivariate spatial data is to assume that all these disease rates share a spatially correlated common factor. This common factor might be interpreted as the public health status of Minnesota counties or simply as the surrogate for unobserved common covariates shared by the specific different cancers.

In this paper we propose the generalized common spatial factor which can be used to explore and model data as in the examples above. Traditionally, the method of 'factor analysis' has been used by researchers in the psychological and behavioral sciences to explore and model common traits among individuals. The generalized common spatial factor model is an extension of the traditional factor analysis model in two ways. First, unlike the traditional common factor model which assumes the factors are independent between observations, we assume the common factors are spatially correlated. Thus, the common factors are used to explain both the correlations within and across locations often seen in different types of multivariate spatial data. Second, the traditional common factor model is applied to normally distributed outcome data, whereas the generalized common spatial factor model extends to handle more types of observed data from an exponential family, in particular Poisson and binomial data.

Recently, other authors have proposed models similar to the generalized common spatial factor model. Christensen and Amemiya (2002) proposed a distribution-free latent variable model to analyze multivariate spatial data. They assume that the observed variables are linear functions of the latent factors and parameters are estimated based on a moment method. The distribution-free assumption makes the model very general, but the applicability of the model is limited by the assumption of linear relationship between the observed and latent variables which may not hold for Poisson and binomial data. Knorr-Held and Best (2001) proposed a shared-component model for detecting joint and selective clustering of two diseases based on Poisson counts. The purpose of their shared-component model is similar to ours, but we assume a continuous underlying spatial common factor instead of a spatial cluster model and, more importantly, the method we propose is straightforward for p-dimensional ($p \geqslant 2$) multivariate spatial data whereas the Knorr-Held and Best model is appealing for only two variables. Furthermore, the generalized common spatial factor model is specified more generally because it can handle different distributions for the observed data.

The rest of this paper is structured as follows. Section 2 describes the form of the generalized common spatial factor model and its specific forms for Poisson and binomial spatial data. Section 3 discusses the estimation of parameters and prediction of the common spatial factor using the Bayesian method with MCMC technique. The model selection criterion DIC is also discussed in Section 3 for use with the generalized common spatial factor model. Section 4 presents an application of the generalized common spatial factor model for Poisson data. Finally, Section 5 gives some discussion and ideas for model extensions.

2. GENERALIZED COMMON SPATIAL FACTOR MODEL

2.1 *Common spatial factor model*

We first introduce the common spatial factor model which extends the traditional factor analysis or common factor model by allowing spatial structure for the underlying factors. It can be written as

$$\mathbf{Z}_i = \boldsymbol{\alpha} + \Lambda f_i + \boldsymbol{\epsilon}_i, \qquad i = 1, \ldots, n, \tag{2.1}$$

where $\mathbf{Z}_i = (Z_{1i}, \ldots, Z_{pi})'$ is the vector of p observed variables at each location s_i in region D, f_i is the underlying common spatial factor at location s_i and $\boldsymbol{\alpha}$ and Λ are $p \times 1$ vectors of intercept and of coefficients called factor loadings. The $p \times 1$ vector $\boldsymbol{\epsilon}_i = (\epsilon_{1i}, \ldots, \epsilon_{pi})'$ is assumed independent of f_i and represents measurement error and any other unmodeled error after modeling each of the observed variables

as a linear combination of one underlying common factor f_i. The ϵ_i is assumed to be i.i.d. multivariate normal with zero mean and diagonal covariance matrix. Further, the common factor $\mathbf{f} = (f_1, \ldots, f_n)'$ is assumed multivariate-normally distributed:

$$\mathbf{f} \sim MVN\left(\mu_f \mathbf{1}_n, \mathbf{C}(\gamma)\right) \tag{2.2}$$

where μ_f is the mean of f_i, $\mathbf{C}(\gamma)$ is the covariance matrix representing the spatial structure, and γ is the vector of parameters in the covariance structure.

There are two common classes of models for the spatial structure that we will consider. One is geostatistical models in which $\mathbf{f}$ is continuously indexed and the spatial correlations depend on the distance between two locations. One possible choice of geostatistical model for $\mathbf{f}$ is the isotropic exponential model, i.e. $\mathbf{C}(\gamma) = (c_{ij})$, and

$$c_{ij} = c_e \mathrm{e}^{-|s_i - s_j|\phi}, \tag{2.3}$$

where $|s_i - s_j|$ is the distance between site s_i and site s_j, c_e is the *sill*, representing the variance in the absence of spatial correlations, ϕ is the *range parameter*, representing the speed of decrease in correlation between two locations as the distance increases, and γ=(c_e, ϕ).

The other class of models for spatial structure we consider is lattice models, in which $\mathbf{f}$ is discretely indexed over a partitioned area, and the spatial correlations depend on the neighborhood structure. The conditional autoregressive (CAR) model (Besag, 1974) is one such lattice model that is commonly used, which can be represented as

$$\mathbf{C}(\gamma) = \tau^2 (I_n - \rho W)^{-1}, \tag{2.4}$$

where ρ is referred to as the 'spatial association' parameter, τ^2 is the conditional variance of $f_i | f_{-i}$, I_n is the $n \times n$ identity matrix, $\gamma = (\rho, \tau^2)$, and $W = (w_{ij})$ is a neighborhood matrix of the lattice, where $w_{ij} = 1$ if subregion i and j share a common boundary, otherwise 0.

2.2 *Generalized common spatial factor model*

We now generalize the model in (2.1) to non-normal (say, Poisson or binomial) data commonly seen in practice.

Let Z_{ij} be the jth ($j = 1, \ldots, p$) random variable observed at location s_i ($i = 1, \ldots, n$). We assume Z_{ij} has a distribution F from an exponential family with mean parameter θ_{ij} and a possibly separate variance parameter σ_j^2, i.e.

$$Z_{ij} | \theta_{ij}, [\sigma_j] \overset{iid}{\sim} F(\theta_{ij}, [\sigma_j]), \qquad i = 1, \ldots, n,\ j = 1, \ldots, p. \tag{2.5}$$

We assume Z_{ij} are independent given the parameter θ_{ij} (and σ_j^2). We now put a spatial common factor model similar to (2.1) on the mean parameter of θ_{ij} with an appropriate link function $g()$, that is

$$g(\theta_{ij}) = O_{ij} + \alpha_j + \lambda_j f_i, \qquad i = 1, \ldots, n, \qquad j = 1, \ldots, p \tag{2.6}$$

where O_{ij} is a known offset we might need for some types of data, the intercept α_j and the slope or 'factor loading' λ_j are fixed unknown parameters that need to be estimated, and f_i is spatially distributed as in (2.2). Note that the link function is actually not linear in the parameters due to the product of unknown parameters λ_j and f_i, thus the generalized common spatial factor model differs from the usual generalized linear model.

Note that there is no measurement error term in (2.6) as in (2.1); actually, the error term is subsumed in (2.5). In one-parameter distributions such as Poisson and binomial, the error term is not separable from the mean, and (2.6) models the mean and variance simultaneously. In two-parameter distributions such as normal and double exponential, (2.6) only models the mean, and the variance parameters must be estimated separately. In the normal case, with identity link function, (2.5) and (2.6) are equivalent to (2.1).

In the following sections, we discuss in detail the cases when Z_{ij} is a Poisson or binomial variable.

2.2.1 *Poisson common spatial factor model.* Poisson data are very common in public health research. Let Z_{ij} be a Poisson random variable with mean θ_{ij}, then we have

$$Z_{ij}|\theta_{ij} \sim POI(\theta_{ij}). \tag{2.7}$$

When using the Poisson distribution to model p different disease counts $j = 1, \ldots, p$ within n regions, $i = 1, \ldots, n$, we can express the Poisson mean θ_{ij} as the product of the relative risk of dying of disease j in region i times the expected number of cases for disease j within a region i based on the age distribution of region i, (e.g. Clayton and Kaldor (1987) and Carlin and Louis (2000, Section 7.8)). For our spatial factor model, on the log scale (i.e. with a log link function), we have

$$\log(\theta_{ij}) = log(E_{ij}) + \lambda_j f_i, \tag{2.8}$$

where E_{ij} (offset) is the expected number of counts at location i for variable j, λ_j and f_i are the same as in (2.6). Equation (2.8) is equivalent to $\lambda_j f_i = log(\theta_{ij}/E_{ij})$. Therefore, $\lambda_j f_i$ can be interpreted as the log Standardized Mortality Ratio (SMR) at region s_i for the variable j, and f_i is the spatially distributed common risk factor. The coefficient λ_j determines how much influence the common risk factor f_i has on the different outcome variable j. Note in (2.8) that the α_j is not included in the model because the expectation E_{ij} is age-adjusted internally.

2.2.2 *Binomial common spatial factor model.* Binomial observations are another type of data commonly seen in public health research and can be handled by the generalized common spatial factor model. Let Z_{ij} be the sum of n_{ij} Bernoulli variable at site i for variable j with parameter θ_{ij}. Then we have

$$Z_{ij} \sim \text{BINOMIAL}(n_{ij}, \theta_{ij}), \tag{2.9}$$

with the logit link function,

$$\text{logit}(\theta_{ij}) = \alpha_j + \lambda_j f_i, \tag{2.10}$$

where $\mathbf{f} = (f_1, \ldots, f_n)'$ is defined as in (2.2). Large (small) values of α_j imply high probability of the corresponding Bernoulli variable being one (zero). The term f_i is the common factor underlying the observed variables at location s_i, representing the common tendency for all Bernoulli variables to be one or zero. Parameter λ_j determines how much influence the common factor will have on making the specific Bernoulli variable j be one or zero.

3. BAYESIAN INFERENCE AND MODEL SELECTION VIA MCMC METHODS

3.1 *Model identifiability and priors*

As in the usual common factor model, we have identifiability problems in the generalized common spatial factor model (2.5)–(2.6). Because f_i is latent, it can be fixed to have any scale. That is, let $\lambda_j^* = c\lambda_j$ and

$f_i^* = \frac{1}{c} f_i$ for $c \neq 0$, then $\alpha_j + \lambda_j^* f_i^* = \alpha_j + \lambda_j f_i$: in other words, the model stays the same. To avoid this arbitrariness, the convention used when the f_i are i.i.d., is to let f_i have a standard normal distribution. To see why this constraint removes the indeterminacy, we suppose f_i has a normal distribution with mean 0 and variance σ^2 instead of 1, let $y_{ij} = \alpha_j + \lambda_j f_i$, and then $E(y_{ij}) = \alpha_j$ and $\text{var}(y_{ij}) = \lambda_j^2 \sigma^2$. If σ^2 increases by a factor c and λ_j^2 decreases by the factor c, $var(y_{ij})$ stays the same. We can see that by letting $\sigma^2 = 1$, the variance of y_{ij} can be uniquely determined by λ_j, thus removing the indeterminacy caused by the product of $\lambda_j^2 \sigma^2$. Thus we will follow this convention to let our spatially correlated factor **f** have a multivariate normal distribution with mean **0** and a spatial covariance structure with a unit variance parameter. That is, we let $c_e = 1$ in (2.3) or $\tau^2 = 1$ in (2.4).

We also notice in (2.6) that if we let $\alpha_j^* = \alpha_j - c\lambda_j$ and $f_i^* = f_i + c$, then $\alpha_j^* + \lambda_j f_i^* = \alpha_j + \lambda_j f_i$ and the model does not change. This indeterminacy can be solved by fixing $\sum_{i=1}^n f_i = 0$. This constraint does not cause problems for the model fitting since it is consistent with the assumption that f_i have expected value zero. Furthermore, by fixing the average of the common spatial factor to be 0, the interpretation for the f_i is straightforward in that positive or negative values of f_i imply the f_i is above or below the average.

To do the Bayesian inference using the MCMC technique, we need to assign prior distributions to the parameters to complete the specification of the Bayesian hierarchical model. The general rule is that we put noninformative priors on the parameters. For the Poisson and binomial common spatial factor models, we assume the following priors:

- λ_j in both models: $\lambda_j \overset{\text{iid}}{\sim} \text{N}(0, \tau_\lambda)$, $j = 1, \ldots, p$. The parameter τ_λ is chosen to be large to make the prior relatively non-informative.
- α_j in binomial model: $\alpha_j \overset{\text{iid}}{\sim} \text{N}(0, \tau_\alpha)$, $j = 1, \ldots, p$. The parameter τ_α is chosen to be large to make the prior relatively non-informative.
- ϕ in the exponential spatial structure: One way is to put uniform prior for ϕ on $(\phi_{\min}, \phi_{\max})$ (Best *et al.*, 2000) where

$$\phi_{\min} = \frac{(-\log 0.5)}{d_{\max}}$$
$$\phi_{\max} = \frac{(-\log 0.01)}{d_{\min}}.$$

  From the definition, we can see that $\phi_{\min}$ corresponds to correlation of 0.5 at the maximum distance ($d_{\max}$) between areas in the study region and $\phi_{\max}$ corresponds to correlation of 0.01 at the minimum distance ($d_{\min}$) between areas in the study region.
- ρ in the CAR spatial structure: To ensure the positive definiteness of the CAR model, ρ has to be $\rho^L = \min(\frac{1}{e_{\min}}, 0) < \rho < \frac{1}{e_{\max}} = \rho^U$, where $e_{\min}$ and $e_{\max}$ are the min and max eigenvalues of neighborhood matrix **W**. Therefore, uniform prior for ρ on (ρ^L, ρ^U) is often used (Best *et al.*, 2000).

3.2 *Estimation and prediction*

In the Bayesian framework, we do not discriminate between estimation and prediction. After putting priors on the parameters, we can derive the posterior distributions analytically or numerically. All the inferences then will be based on these posterior distributions.

In most cases, the analytical forms of the posterior distributions are intractable, and numerical methods have to be used. Standard numerical integration is usually not feasible to get the posterior distribution due

to high dimensionality, so MCMC techniques will be used. One frequently used MCMC technique to get the empirical posterior distributions is the Gibbs sampler, originally developed by Geman and Geman (1984) in image reconstruction and introduced into statistics by Gelfand and Smith (1990). To apply the Gibbs sampler, we need to derive the full conditional likelihood with respect to each parameter.

One advantage of MCMC sampling is that we not only get estimates of the parameters, we also get empirical estimates of the posterior distributions reconstructed from the MCMC samples and we can do more analysis based on the posterior distributions.

Due to the constraints we put on the model, no software package was available. FORTRAN 90 programming language was used to implement the MCMC sampling technique. The numerical issues will be further discussed in the example section.

3.3 *Model selection*

Bayes factors are often used to do Bayesian model selection, but they can be difficult to compute. The commonly used Akaike information criterion (AIC) (Akaike, 1973) and Bayesian information criterion (BIC) (Schwarz, 1978) for model selection are easy to compute but require specification of the number of parameters in each model. This is not a problem for traditional, nonhierarchical models. In our hierarchical generalized common spatial factor model, however, the number of parameters is not clearly defined (Spiegelhalter *et al.*, 2002), the effective number of parameters (or degrees of freedom) could be much smaller than the number of parameters (Hodges and Sargent, 2001). The AIC and BIC cannot be directly applied (Gelfand and Dey, 1994). Spiegelhalter *et al.* (2002) extend the AIC criterion and derive a Deviance Information Criterion (DIC) for comparing complex hierarchical models in which the number of parameters is not clearly defined. DIC is given as

$$DIC = \overline{D(\boldsymbol{\theta})} + p_D \tag{3.1}$$

where $\overline{D(\boldsymbol{\theta})}$ is the average of $D(\boldsymbol{\theta})$ for all MCMC samples of $\boldsymbol{\theta}$, and $D(\boldsymbol{\theta})$, which is called 'Bayesian Deviance', can be defined as $D(\boldsymbol{\theta}) = -2\log(f(y|\boldsymbol{\theta}))$, where $f(y|\boldsymbol{\theta})$ is the likelihood function of the observed data given the parameter $\boldsymbol{\theta}$. The quantity p_D is called 'effective' number of parameters, and is defined as $p_D = \overline{D(\boldsymbol{\theta})} - D(\overline{\boldsymbol{\theta}})$, where $\overline{\boldsymbol{\theta}}$ is the average of MCMC samples of $\boldsymbol{\theta}$.

From the definition of DIC, we see that only the MCMC samples and the likelihood function of the observed data are needed to calculate DIC, therefore it is very convenient to get DIC from MCMC sampling. DIC itself has no well-defined meaning, so when it is used in model selection, the difference of DIC across models is considered. Similar to AIC and BIC, smaller DIC implies better fit.

4. APPLICATION

In this section, we apply the generalized common spatial factor model to Minnesota cancer data. The data set contains the number of deaths due to four types of cancers for Minnesota counties.

4.1 *The data*

We have the complete vital records of Minnesota from 1991 to 1998. In this example, we will focus on numbers of deaths due to cancers of the lung, pancreas, esophagus, and stomach in the years from 1991 to 1998 at the county level. The numbers of deaths from 1991 to 1998 due to the four cancers in county i are denoted as $Z_i = (Z_{i1}, Z_{i2}, Z_{i3}, Z_{i4})$ respectively. Because these diseases are rare relative to the population in each county, disease counts are small enough that the Poisson model is appropriate.

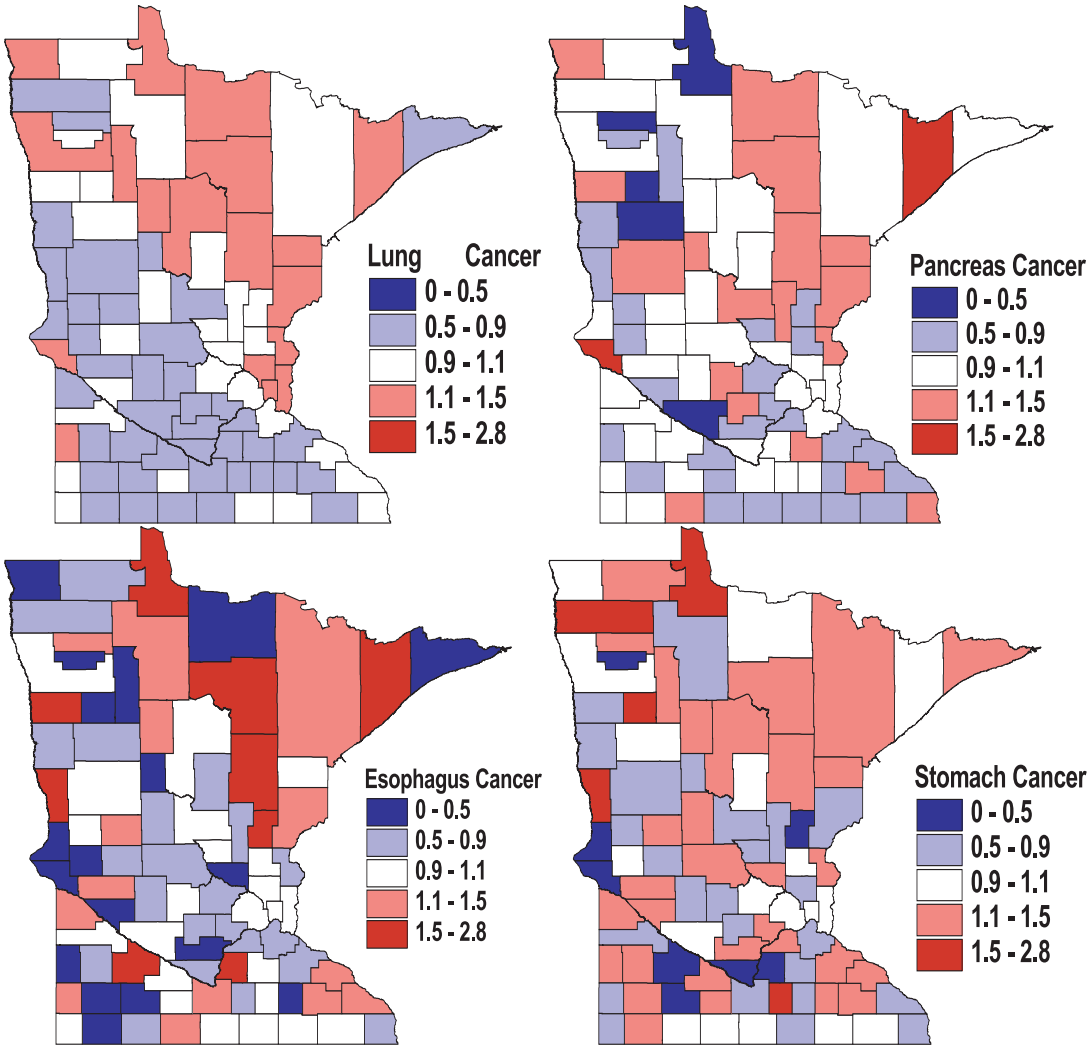

Fig. 1. Minnesota maps of raw data standard mortality ratios of four cancers.

Figure 1 is the maps of raw data standard mortality ratios (SMR), i.e. $\frac{Z_{ij}}{E_{ij}}$, for the 87 Minnesota counties for each of the diseases. The E_{ij} is the age-standardized expected number of deaths due to disease j in county i. Preliminary analyses of these data using a spatial CAR model fit to each variable separately indicate that lung, pancreas, and esophageal cancer deaths exhibit positive spatial autocorrelation (with lung cancer autocorrelation being the strongest) while stomach cancer deaths exhibit slightly negative spatial autocorrelation. In further preliminary analysis, we consider the correlation between each of the four cancers by taking the simple Pearson correlation between $\frac{Z_{ij}-E_{ij}}{\sqrt{E_{ij}}}$ for $j = 1, \ldots, 4$. We find positive correlation between lung, pancreas, and esophageal cancer, i.e. corr(lung, pancreas) $= 0.22$, corr(lung, esophagus) $= 0.23$, corr(pancreas, esophagus) $= 0.24$, but very small or negative correlation between stomach and the others, i.e. corr(stomach, lung) $= 0.05$, corr(stomach, pancreas) $= -0.16$, corr(stomach, esophagus) $= -0.15$.

Table 1. *Parameter Estimates From Poisson Common Spatial Factor Model*

	CAR Model (4 var)		iid Model (4 var)		CAR Model (3 var)	
DIC(p_D)	46171 (30.73)		46193 (44.19)			
Parameter	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
λ_1	0.056 74	0.007 39	0.089 52	0.010 93	0.056 68	0.007 25
λ_2	0.026 13	0.009 46	0.038 14	0.015 95	0.025 74	0.009 40
λ_3	0.046 30	0.015 29	0.075 63	0.024 96	0.046 43	0.015 70
λ_4	0.006 98	0.013 54	−0.001 04	0.023 97	NA	NA
ρ	0.174 54	0.003 08	NA	NA	0.174 37	0.003 45

Because preliminary data analyses indicate that lung, pancreas, and esophageal cancer exhibit both positive spatial autocorrelation and correlation between one another, this suggests that the generalized common spatial factor model might be useful to detect a 'common factor' underlying them. Since stomach cancer does not have strong spatial correlation and is not highly correlated with the other diseases, we suspect the stomach cancer might not share a common spatial factor with the other three diseases. Recall that one of our goals is to use the model to determine which variables 'stick together' to share a common spatial factor. We will keep the 'stomach' variable and apply our Poisson common spatial factor model to see whether the model can detect that stomach cancer does not share a common spatial factor with the other three.

4.2 *Parameter estimation and model selection*

Consider the Poisson common spatial factor model in (2.7) and (2.8) where $i = 1, \ldots, 87$ and $j = 1, \ldots, 4$. It is natural to choose a CAR covariance structure for the spatial common factor $\mathbf{f}$, i.e. $\mathbf{f} \sim MVN(0, \tau^2(I_n - \rho W)^{-1})$.

As discussed in Section 3.1, we set $\tau^2 = 1$ and $\sum_{i=1}^{n} f_i = 0$ for identifiability reasons. To complete the specification of the hierarchical model, we put fairly non-informative priors on other parameters as

$$\lambda_j \overset{\text{iid}}{\sim} N(0, 10^5), \qquad j = 1, \ldots, 4 \tag{4.1}$$

$$\rho \sim \text{Uniform}(-0.322, 0.178). \tag{4.2}$$

There are no closed forms for the posterior distribution of the parameters here so the MCMC technique is used to draw samples from the posterior distributions. The full conditional densities and their first derivatives are derived in the Appendix. It is possible to show that all the full conditionals are log concave (see the Appendix). Therefore, the adaptive rejection sampling (ARS) algorithm (Gilks and Wild, 1992) can be used to do the MCMC sampling.

Two chains with different starting values are run for 4000 iterations and the chains converge very quickly. The last 3000 iterations are used to do the inference based on the MCMC samples of the posterior distributions. The posterior means and standard deviations of the parameters are listed in Table 1. With the model well identified and the relatively non-informative priors, the convergence of these chains is not sensitive to the starting values.

From Table 1, we see that the parameter ρ is significantly different from 0. This indicates the spatial correlation of the common factor is significant. If ρ is not significantly different from 0, it would indicate that the assumption of spatial correlation is not supported by data, and the independence model should be used. Thus, the generalized common spatial factor model not only model the spatial data, but also test the strength of the spatial correlation.

To make sure the spatial model is doing a better job than the independence model, we also run the model assuming the common factor f_i are i.i.d. for comparison. The result is also in Table 1. We use the DIC to compare the two models. Similar to AIC, the value of DIC itself has no well defined meaning but when compared across models, smaller DIC indicates better fit. From this criterion, we can see that the spatial common factor model has smaller DIC, thus fits better than the i.i.d. common factor model.

The λ_j reflect the influence of the common factor on the corresponding outcome variables. Though the actual value of λ_j depends on the scale we choose for the common factor **f**, the ratio between the λ_j should remain the same regardless of the scale chosen for **f**. A ratio of one between two variables means the common factor has the same influence on the two variables. The bigger the ratio is, the bigger influence the common factor has on one variable than on the other. In this example, the different values of λ_j imply the different influence of the common risk factor on different cancer mortality ratios. From the table, we can see that the common factor has the biggest influence on the lung cancer mortality ratio. The influence on lung cancer is twice as big as on pancreas cancer. From Table 1, we also see that λ_4 is not significantly different from 0. This means the common spatial risk factor has no significant influence on the stomach cancer mortality ratio, or in other words, the stomach cancer does not share a common spatial factor with the other three cancers. Therefore, the hypothesis testing of $\lambda_j = 0$ provides a way to detect which diseases 'stick together' and share the common spatial factor.

Since λ_4 is not significant, stomach cancer will not be included in the model to achieve our second goal of predicting the common spatial factor. Thus we re-fit the model without stomach cancer; the result is also in Table 1. From the result, we can see that the λ_j from the three-variable model stay almost the same as those from the four-variable model. This also verifies that the fourth variable is not contributing to identify the common spatial factor and it does not share a common spatial factor with the other variables.

In this example (and this paper), we focus on finding the one common spatial factor that variables share. In general, instead of throwing out variables that do not share the common factor, we could fit a model with multiple factors to fully explore the data. The model with more than one factor contains identifiability issues that have not yet been worked out and will be mentioned further in the discussion section.

4.3 *The predicted spatial common factor*

The final model we use to predict the spatial common factor is the CAR model with three variables. Figure 2 is the map of the predicted spatial common factor and we can see that there is clear spatial pattern. North central Minnesota has higher values of the common factor, which implies those counties tend to have higher rates of deaths due to cancers of lung, pancreas and esophagus. The strong spatial structure is the result of $\rho = 0.174$ which is very close to the upper limit 0.178.

Figure 3 is the plot of the sorted f_i by county with 95% prediction intervals created from the empirical posterior distributions. From the plot, we can see that many predicted f_i are significantly different from each other. County 1 has the highest values of the common factor, while county 65 has the lowest. Counties 27 and 62 have the shortest predictions intervals since they are the two most populated counties in Minnesota encompassing the Twin Cities of Minneapolis and St Paul.

4.4 *Model checking*

As the common spatial factor of the three disease rates, we would expect the predicted **f** to be correlated with the SMRs of these diseases. Thus as a first model check, we compute the correlation coefficient between the mean posterior predicted **f** and $\frac{Z_{ij}-E_{ij}}{\sqrt{E_{ij}}}$ for lung, pancreas and esophagus and get 0.85, 0.31, 0.30, respectively. Thus the common factor is most highly correlated with lung cancer which can also be

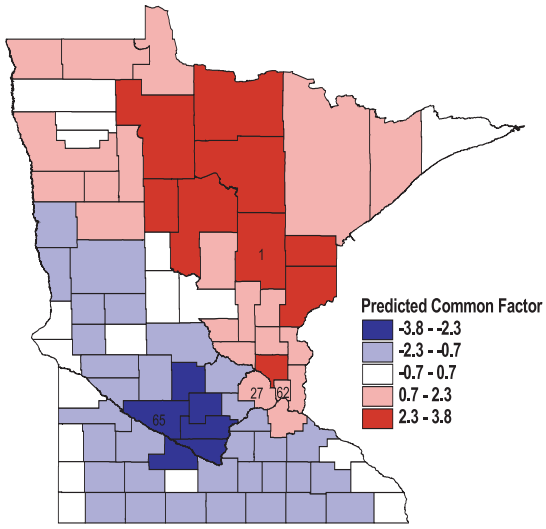

Fig. 2. Predicted Poisson common spatial factor.

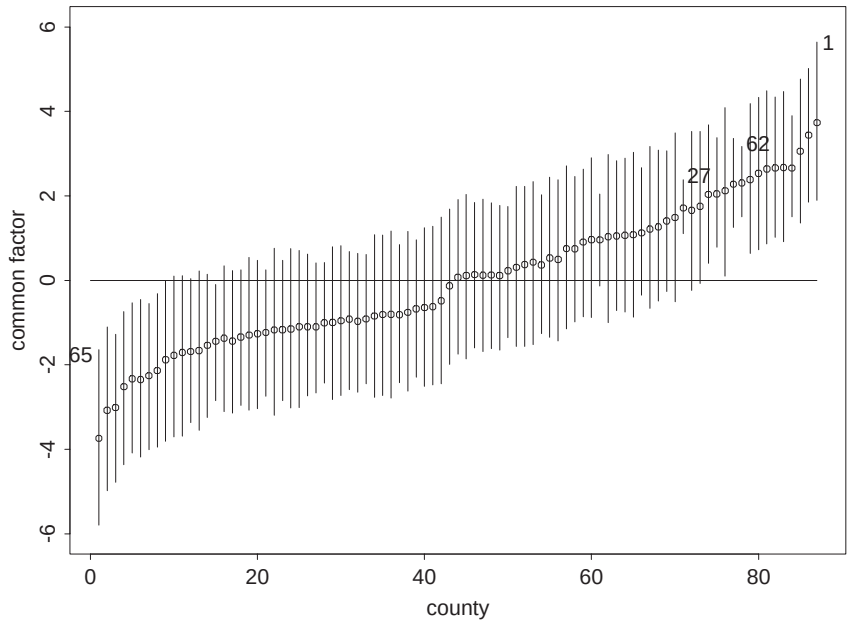

Fig. 3. Predicted Poisson spatial common factor with prediction intervals.

seen from the high similarity between the lung cancer map in Figure 1 and the predicted map in Figure 2. Recall from the preliminary data analysis that corr(lung, pancreas) $= 0.21$, corr(lung, esoph) $= 0.23$ and corr(pancreas, esophagus) $= 0.24$, hence we also see that the common factor is more highly correlated with pancreas and esophagus than any of the individual variables on their own was.

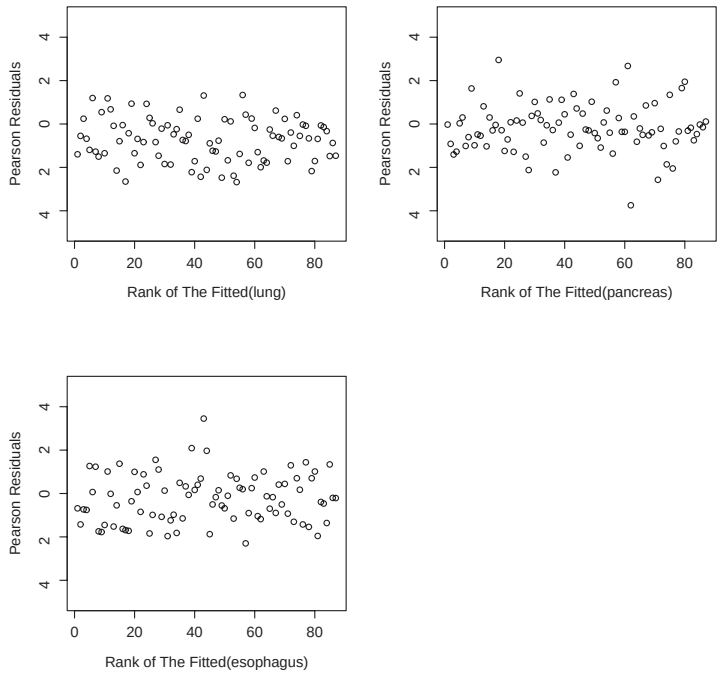

Fig. 4. Pearson residuals from the fit of the Poisson generalized common spatial factor model with CAR spatial structure.

Another model check is performed by examining the Pearson residuals $\frac{(z_{ij}-\hat{\theta}_{ij})}{\sqrt{\hat{\theta}_{ij}}}$ where $\hat{\theta}_{ij}$ is the mean of the posterior for θ_{ij} from the common spatial factor model. Figure 4 is the plot of the residuals against the rank of the fitted counts $\hat{\theta}$ (the rank is used because the range of the actual $\hat{\theta}$ is too wide for an informative plot). The plot shows that the residuals are randomly scattered around zero, not increasing with the fitted counts $\hat{\theta}$ and not indicating any outliers. The variances of these residuals are 0.99, 1.23, and 1.27 respectively, close to 1 as expected. Therefore, the plots indicate the model fits the data well.

5. Discussion

The generalized common spatial factor model not only provides a tool to explain the two types of correlations seen in multivariate spatial data, but also can be used to determine whether a variable included in the model 'sticks together' with other variables to share a common spatial factor by testing whether $\lambda_j = 0$.

Several extensions to model (2.5)–(2.6) may be considered. First, in the example presented in this paper no observed covariates were available for investigation, but, the model can easily be extended to include observed covariate information into (2.6). Second, even though the residuals from the common spatial factor model for the Minnesota cancer counts data did not indicate a problem of overdispersion, it is common to see overdispersion of disease counts data in practice. While the heirarchical structure of (2.7)–(2.8) has the ability to account for overdispersion, the **f** in (2.8) must account for both the heterogeneity across regions and the spatial correlation simultaneously. To add flexibility, we could model the spatial

correlation and heterogeneity separately by adding another random term in (2.8), i.e.

$$\log(\theta_{ij}) = \log(E_{ij}) + \lambda_j f_i + \epsilon_{ij}, \tag{5.1}$$

where ϵ_{ij} is a disease-specific random effect that accounts for the possible heterogeneity in the counts data. The extra flexibility also adds extra identifiability and convergence problems to fitting the model. Eberly and Carlin (2000) investigated the relationship among identifiability, Bayesian learning, and MCMC convergence rates for a common class of univariate spatial models that include separate random effects for spatial structure and heterogeneity. In that case it is shown that only the sum of these random effects is well-identified by the data. Hence, since we are particularly interested in predicting f_i this type of model may be difficult to use.

Another possible extension to the model is that there is more than one 'common factor' underlying the multivariate spatial data. Therefore, one might consider the more general form of (2.6) written as

$$g(\theta_{ij}) = \alpha_j + \sum_{m=1}^{k} \lambda_j^m f_i^m, \qquad i = 1, \ldots, n, \qquad j = 1, \ldots, p \tag{5.2}$$

where k is the number of common factors. One might even imagine treating the number of factors k as an unknown variable, and use Bayesian analysis with reversible jump MCMC method to estimate both the parameters and the number of factors. But, besides the identifiability issues we discussed in Section 3.1, there is an additional nondeterminacy when we have two or more common factors. We know that for an independent normal factor analysis model with two or more factors, any orthogonal transformation (factor rotation) of the λ_{mj} leaves the likelihood unchanged (Johnson and Wichern, 1998). This problem carries over to the generalized spatial factor model. One way to solve this might be to fix certain λ to ensure identifiability. Bock and Gibbons (1996) discuss similar identifiability issues in detail when the latent variables are independent. Additional identifiability issues are introduced by the possibility of more than one factor exhibiting spatial structure as well.

Finally, this model can be naturally extended to model multivariate time series or longitudinal data to find the latent common temporal factors to explain the correlations seen between variables and across time. We only need to replace the spatial covariance structure in (2.2) with a time series covariance structure such as AR(1).

ACKNOWLEDGEMENTS

This work was supported by National Center for Health Statistics grant NCHS-UR6/CCU517477-01. The authors are grateful to the Minnesota Center for Excellence in Health Statistics for providing and permitting analysis of the Minnesota mortality dataset.

APPENDIX

Derivation of full conditionals for Poisson common spatial factor model

First, we write down the joint posterior of the parameters as

$$\begin{aligned} L &= f(\mathbf{Z}|\boldsymbol{\lambda}, \mathbf{f})\, f(\boldsymbol{\lambda})\, f(\mathbf{f}|\rho)\, f(\rho) \\ &\propto \mathrm{e}^{(-\sum_{i=1}^{n}\sum_{j=1}^{m} E_{ij}\mathrm{e}^{\lambda_j f_i})} \prod_{i=1}^{n}\prod_{j=1}^{m} (E_{ij}\mathrm{e}^{\lambda_j f_i})^{Z_{ij}} |C^{-1}|^{1/2} \mathrm{e}^{(-\frac{1}{2}\mathbf{f}'C^{-1}\mathbf{f})} \mathrm{e}^{(-\frac{1}{2}\tau_\lambda \sum_{j=1}^{m}\lambda_j^2)}. \end{aligned}$$

We take the log on both sides for convenience, and plug in $C^{-1} = (I - \rho W)$, we get the log joint posterior of the parameters,

$$\begin{aligned}\log L &\propto \sum_{i=1}^{n}\sum_{j=1}^{m}(Z_{ij}\lambda_j f_i - E_{ij}e^{\lambda_j f_i}) + \tfrac{1}{2}\log|C^{-1}| - \tfrac{1}{2}\mathbf{f}'C^{-1}\mathbf{f} - \tfrac{1}{2}\tau_\lambda\sum_{j=1}^{m}\lambda_j^2 \\ &\propto \sum_{i=1}^{n}\sum_{j=1}^{m}(Z_{ij}\lambda_j f_i - E_{ij}e^{\lambda_j f_i}) + \tfrac{1}{2}\log|I - \rho W| - \tfrac{1}{2}\mathbf{f}'(I - \rho W)\mathbf{f} - \tfrac{1}{2}\tau_\lambda\sum_{j=1}^{m}\lambda_j^2.\end{aligned}$$

It is more efficient to use the full conditionals with respect to (w.r.t.) each parameter instead of using the joint each time. Therefore, the log full conditionals w.r.t. λ_j, f_i and ρ are

$$\begin{aligned}\log L(\lambda_j) &\propto \sum_{i=1}^{n}(Z_{ij}\lambda_j f_i - E_{ij}e^{\lambda_j f_i}) - \tfrac{1}{2}\tau_\lambda\lambda_j^2, \qquad \text{j=1,2,3,4} \\ \log L(f_i) &\propto \sum_{j=1}^{m}(Z_{ij}\lambda_j f_i - E_{ij}e^{\lambda_j f_i}) - \tfrac{1}{2}f_i^2 + \tfrac{\rho}{2}\mathbf{f}'W\mathbf{f}, \qquad i = 1,\ldots,87 \\ \log L(\rho) &\propto \tfrac{1}{2}\log|I - \rho W| + \tfrac{\rho}{2}\mathbf{f}'W\mathbf{f}.\end{aligned}$$

To use the adaptive rejection sampling, we still need the first derivatives of the full conditionals which can be easily derived as

$$\begin{aligned}\frac{\partial \log L(\lambda_j)}{\partial \lambda_j} &= \sum_{i=1}^{n}(Z_{ij} - \theta_{ij})f_i - \tau_\lambda\lambda_j, \qquad j = 1,\ldots,4 \\ \frac{\partial \log L(f_i)}{\partial f_i} &= \sum_{j=1}^{m}(Z_{ij} - \theta_{ij})\lambda_j - f_i + \rho[W\mathbf{f}]_i, \qquad i = 1,\ldots,87 \\ \frac{\partial \log L(\rho)}{\partial \rho} &= -\tfrac{1}{2}\mathrm{tr}\,((I - \rho W)^{-1}W) + \tfrac{1}{2}\mathbf{f}'W\mathbf{f},\end{aligned}$$

where $\theta_{ij} = E_{ij}e^{\lambda_j f_i}$. We now take the second derivatives to show the log concaveness of the full conditionals.

$$\begin{aligned}\frac{\partial^2 \log L(\lambda_j)}{\partial \lambda_j^2} &= -\left(\sum_{i=1}^{n} f_i^2 E_{ij}e^{\lambda_j f_i} + \tau_\lambda\right) < 0, \qquad j = 1,\ldots,4 \\ \frac{\partial^2 \log L(f_i)}{\partial f_i^2} &= -\left(\sum_{j=1}^{m}\lambda_j^2 E_{ij}e^{\lambda_j f_i} + 1 - \rho W_{ii}\right) < 0, \qquad i = 1,\ldots,87 \\ \frac{\partial^2 \log L(\rho)}{\partial \rho^2} &= -\tfrac{1}{2}\mathrm{tr}[((I - \rho W)^{-1}W)^2] < 0.\end{aligned}$$

All the second derivatives are less than zero, indicating the log concaveness of the full conditionals.

REFERENCES

AKAIKE, H. (1973). Petrov, B. N. and Csaki, F. (eds), *Information Theory and an Extension of the Maximum Likelihood Principle*, Second International Symposium on Information Theory. Budapest: Akademiai Kiado, pp. 267–281.

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

BEST, N., MARSHALL, C. AND THOMAS, A. (2000). *Spatial Modeling Using WinBUGS and GeoBUGS*, Short Course: Brisbane. Nov. 30-Dec. 1, 2000

BOCK, R. AND GIBBONS, R. (1996). High-dimensional multivariate probit analysis. *Biometrics* **52**, 1183–1194.

CARLIN, B. AND LOUIS, T. (2000). *Bayes and Empirical Bayes Methods for Data Analysis*, 2nd edn. New York: Chapman & Hall CRC.

CHRISTENSEN, W. F. AND AMEMIYA, Y. (2002). Latent variable analysis of multivariate spatial data. *Journal of the American Statistical Association* **97**, 302–317.

CLAYTON, D. AND KALDOR, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43**, 671–681.

DESOUZA, C. (1992). An approximate bivariate Bayesian method for analyzing small frequencies. *Biometrics* **48**, 1113–1130.

EBERLY, L. E. AND CARLIN, B. P. (2000). Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. *Statistics in Medicine* **19**, 2279–2294.

GELFAND, A. E. AND DEY, D. K. (1994). Bayesian model choice: asymptotics and exact calculations. *Biometrika* **85**, 1–11.

GELFAND, A. E. AND SMITH, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.

GEMAN, S. AND GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of image. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.

GILKS, W. R. AND WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* **41**, 337–348.

GOTWAY, C. AND HARTFORD, A. (1996). Geostatistical methods for incorporating auxiliary information in the prediction of spatial variables. *Journal of Agricultural Biological and Environmental Statistics* **1**, 17–39.

HODGES, J. S. AND SARGENT, D. J. (2001). Counting degrees of freedom in the hierarchical and other richly-parameterized models. *Biometrika* **88**, 367–379.

JOHNSON, R. A. AND WICHERN, D. W. (1998). *Applied Multivariate Statistical Analysis*, 4th edn. Englewood Cliffs, NJ: Prentice-Hall.

KIM, H., SUN, D. AND TSUTAKAWA, R. (2001). A bivariate Bayes method for improving the estimates of mortality rates with twofold conditional autoregressive model. *Journal of the American Statistical Association* **96**, 1506–1521.

KNORR-HELD, L. AND BEST, N. G. (2001). A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society, Series A* **164**, 73–85.

LE, N., SUN, W. AND ZIDEK, J. (1997). Bayesian multivariate spatial interpolater with data missing by design. *Journal of the Royal Statistical Society, Series B* **59**, 501–510.

SCHWARZ, G. (1978). Estimating the dimension of a model. *Annals of Statistics* **6**, 773–784.

SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B.P. AND VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society, Series B* **6**, 583–639.

[*Received July 12, 2002; revised January 30, 2003; accepted for publication February 5, 2003*]