Comparison of Bayesian Spatio-Temporal Models
for Chronic Diseases

Hoon Kim[1] and HeeJeong Lim[2]
[1]*California State Polytechnic University and*
[2]*California State University at Fullerton*

Abstract: This paper discusses a comprehensive statistical approach that will be useful in answering health-related questions concerning mortality and incidence rates of chronic diseases such as cancer and hypertension. The developed spatio-temporal models will be useful to explain the patterns of mortality rates of chronic disease in terms of environmental changes and social-economic conditions. In addition to age and time effects, models include two components of normally distributed residual effects and spatial effects, one to represent average regional effects and another to represent changes of subgroups within region over time. Numerical analysis is based on male lung cancer mortality data from the state of Missouri. Gibbs sampling is used to obtain the posterior quantities. As a result, all models discussed in this article fit well in stabilizing the mortality rates, especially in the less populated areas. Due to the richness of hierarchical settings, easy interpretation of parameters and ease of implementation, any models proposed in this paper can be applied generally to other sets of data.

Key words: Bayesian hierarchical model, CAR, Gibbs sampling, MCMC, mortality rate, SAR.

1. Introduction

The objective of this paper is to develop a comprehensive statistical approach that will be useful in answering health-related questions concerning mortality and incidence rates of chronic diseases such as cancer and hypertension. The methods will employ hierarchical models which include demographic variable (*i.e.,* age) and geographic variable to identify regional differences. The proposed models will also include a longitudinal variable that will describe temporal trends in mortality and incidences for different population groups. The developed spatio-temporal models will be useful to explain the patterns of mortality rates of chronic disease in terms of environmental changes and social-economic conditions.

Within the last two decades, many authors have proposed fully (hierarchical) Bayesian or empirical Bayes estimates of mortality rates with spatial correlation, along with Binomial or Poisson sampling processes. This paper is closely related to the general development in this area, which is typified, among others, by Tsutakawa (1985, 1988), Clayton and Kaldor (1987), Ghosh *et al.* (1998), Best *et al.* (1998), Sun *et al.* (2000), and Kim and Oleson (2008). These studies are based on hierarchical models that use a Poisson model for the first stage and incorporate covariates by various modeling of the Poisson parameters at the second stage. Spatial (Regional) effects are usually treated as random with some distributions whose parameters must be estimated. Hyper-parameters are typically estimated by maximum likelihood methods, and the Poisson parameters and rates are estimated by Bayesian or empirical Bayes method. Tsutakawa (1985) shows that the empirical Bayes method could seriously underestimate the uncertainty in the estimated rates. Ghosh *et al.* (1998) and Best *et al.* (1998) consider special cases of pairwise difference prior (cf. Besag, York and Mollié, 1991) for spatial effects without specifying the degree of spatial correlation. Waller *et al.* (1997) and Knorr-Held (2000) propose spatio-temporal interaction models where the spatial effects are nested within time so that they can examine the evolution of heterogeneity and spatial patterns over time. Sun *et al.* (2000) and Kim and Oleson (2008) assume normal distributions for residual effects which account for the extra-Poisson variations. However, our specification of simultaneously autoregressive (SAR) and conditional autoregressive (CAR) priors includes parameters to represent the degree of spatial correlation among regional areas while we assume both normal and gamma distributions for residual effects based on the choice of the first-stage model.

In this paper, we will modify the spatio-temporal models proposed by Sun *et al.* (2000). Their log-linear mixed model of mortality rates includes two components of normally distributed residual effects and spatial effects, one to represent average regional effects and another to represent changes of subgroups within region over time. Our modifications are made in two directions: first, replace the normally distributed residual effects by those which have log-gamma distributions; second, replace the CAR (Besag, 1974) effects by those which are SAR effects (Ord, 1975; Cliff and Ord, 1981). In order to examine the effect of these modifications we also consider making changes in one direction at a time, giving us a total of four possible models. Those four models are compared using Missouri lung cancer data.

The developed methods will not only be applicable to data obtained from other states or countries, but also be useful in analyzing mortality due to other chronic diseases. Moreover, the methods can be applied more generally to measure problems such as crime rates, accident rates, and high school dropouts rates.

In addition, this study motivates more general research in statistics such as model selection, formulation of prior distributions for hierarchical models, and designs for computer simulation.

The subsequent sections are in the following orders. In Section 2 we introduce multiplicative and log-linear mixed models with residual effects for mortality rates where, for a given period of time, age effects are fixed and spatial effects are random. A CAR model by Besag (1974) and an SAR model by Ord (1975) are chosen to describe the random spatial effects. The necessary prior distributions for the parameters and hyper-parameters are given. In Section 3, we derive the available full conditional distributions for the four proposed models. Our choice of SAR and CAR models allows us to sample all regional effects simultaneously without computing the inverses of matrices each time, which improves the efficiency of the computation greatly. Numerical results on age effects, spatial effects, mortality rates, variance components, spatial correlations and residuals are presented in Section 4. The convergence of Gibbs sampling is investigated using Gelman and Rubin's (1992) diagnostics. Also, model selection using cross-validation predictive density and deviance information criterion is done and followed by conclusions in Section 5.

2. Hierarchical Modeling of Chronic Disease Rates

2.1 Multiplicative and log-linear mixed models

The Poisson distribution arises naturally when data set takes the form of counts; for instance, a major area of application is epidemiology, the study of incidence of diseases. Let y_{ijk} denote the frequencies of deaths due to a specific cause for the kth time period within the jth age group of the ith county. Each cell has its population size n_{ijk}. We assume that y_{ijk} have independent Poisson distributions with means $n_{ijk}p_{ijk}$.

In order to model the effects of county, age and time on p_{ijk}, first we consider the multiplicative model (Tsutakawa, 1988; Kim *et al.*, 2002) having the form,

$$p_{ijk} = Z_i^* \theta_j^* (\mu_j^* W_i^*)^{(t_k - \bar{t})} \epsilon_{ijk}^*, \tag{2.1}$$

where t_k are the midpoints of the kth time period and $\bar{t} = K^{-1} \sum_{k=1}^{K} t_k$, $i = 1, \cdots, I$; $j = 1, \cdots, J$; $k = 1, \cdots, K$. Here Z_i^* are the effects of the ith county, θ_j^* the effects of the jth age group, μ_j^* the overall rates of change over time for the jth age group, W_i^* the rates of change in the ith county over time, and ϵ_{ijk}^* unexplained residual effects for the ijkth cell. We may express this multiplicative model in the log-linear form,

$$\log(p_{ijk}) = Z_i + \theta_j + (\mu_j + W_i)(t_k - \bar{t}) + \epsilon_{ijk}, \tag{2.2}$$

where Z_i, θ_j, μ_j, W_i and ϵ_{ijk} are natural logarithms of Z_i^*, θ_j^*, μ_j^*, W_i^* and ϵ_{ijk}^*, respectively.

In the specification of the model we treat θ_j^* and μ_j^* (or equivalently, θ_j and μ_j) as fixed effects while Z_i^*, W_i^* and ϵ_{ijk}^* (or equivalently, Z_i, W_i and ϵ_{ijk}) are treated as random effects. The parameterization will be made unique by placing restrictions on the distributions of the random effects. Depending on the specification of the distributions for the random effects, several cases can be constructed. We focus on two distributions for residual effects ϵ_{ijk} and two distributions for the spatial effects (Z_i, W_i), resulting in four distinct combinations. These are described below.

2.2 Distributions of residual effects

To estimate the rates of chronic diseases such as lung cancer, the extra variation (residual) arises from socio-economic status, air pollution, smoking prevalence, and many other factors not included in the model. These factors are unknown or hard to quantify in most instances. The role of the residual effects is not only to account for the lack of fit, but also to use it as a tool to detect irregular patterns and changes (Kim *et al.*, 2002). The residual effect will sum up all those county, age group, and period effects which are not spatially diffused (*i.e.*, local). Here we have two distributions for residual effects.

- *Normal*: Suppose the ϵ_{ijk} are independent and identically distributed (*iid*) normal with mean 0 and variance δ_0.

- *Gamma*: Suppose the $\epsilon_{ijk}^* = \exp(\epsilon_{ijk})$ are *iid* gamma (ψ^{-1}, ψ^{-1}) so that ϵ_{ijk}^* have mean 1 and variance ψ.

2.3 Distributions of spatial effects

We represent the spatial correlation among the counties by a CAR model introduced by Besag (1974) and an SAR model by Ord (1975). Although CAR and SAR models have been available for more than 30 years, their relative merits have not been fully investigated.

- *SAR*: The SAR model is defined by the autoregressive relation

$$U_i = \rho \sum_j c_{ij} U_j + \tau_i,$$

  where τ_i are *iid* $\mathcal{N}(0, \delta_u)$. We define the adjacency matrix $\mathbf{C}$ with entries $c_{ij} = 1$ if county i and j are neighbors and $c_{ij} = 0$ otherwise, with the

diagonal entries $c_{ii} = 0$. It can be shown that $(\mathbf{I} - \rho\mathbf{C})$ is nonsingular when ρ is restricted to some interval depending the eigenvalues of $\mathbf{C}$. Then, it follows that $\mathbf{U} = (U_1, U_2, \cdots, U_I)'$ has a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\delta_u(\mathbf{I} - \rho\mathbf{C})^{-2}$.

- *CAR*: The CAR model is defined by the conditional probability density function

$$f(U_i | U_j, j \neq i) = \left(\frac{1}{2\pi\delta_u}\right)^{1/2} \exp\left\{-\frac{1}{2\delta_u}\left(U_i - \rho\sum_{j \neq i} c_{ij}U_j\right)^2\right\},$$

where $\delta_u > 0$ and $|\rho| < 1$. This model is a modified version of that used by Besag (1974) and Ripley (1981). This modified version was also used by Clayton and Kaldor (1987) and Ripley (1988). Sun, Tsutakawa and Speckman (1999) show that if $\lambda_1^{-1} < \rho < \lambda_I^{-1}$ where λ_1 and λ_I are smallest and largest eigenvalues, with $\lambda_1 \leq \cdots < 0 < \cdots \leq \lambda_I$, of the adjacency matrix $\mathbf{C}$, then $\mathbf{U}$ has a multivariate normal distribution with mean $\mathbf{0}$ and nonsingular covariance matrix $\delta_u(\mathbf{I} - \rho\mathbf{C})^{-1}$.

Now, if we let $\mathbf{Z} = (Z_1, \cdots, Z_I)'$, then the spatial effects $\mathbf{Z}$ follow $\mathcal{MVN}_I(\mathbf{0}, \delta_1\mathbf{A}_1^{-\gamma})$ with $\mathbf{A}_1 = \mathbf{I} - \rho_1\mathbf{C}$. Similarly, the spatial effects over time $\mathbf{W} = (W_1, \cdots, W_I)'$ are distributed as $\mathcal{MVN}_I(\mathbf{0}, \delta_2\mathbf{A}_2^{-\gamma})$ with $\mathbf{A}_2 = \mathbf{I} - \rho_2\mathbf{C}$ where (ρ_1, ρ_2) represent the degrees of spatial correlation between regions and (δ_1, δ_2) the dispersion parameters. Note that the choices of γ are 1 for CAR and 2 for SAR. Contrary to CAR model, it is difficult to write the conditional distribution of $(Z_i | Z_j, j \neq i)$ for the SAR model in a higher dimension. Most statisticians prefer to use a CAR model, however, the computational methods for SAR and CAR are almost identical except for the choice of γ. Although it may appear more suitable to work with the gamma distribution under the multiplicative model and the normal distribution under the log-linear model, such pairing is not crucial for Bayesian hierarchical modelings (cf. Kim *et al.*, 2002).

2.4 Choice of priors

One objective of specifying priors is to stabilize the estimates of mortality rates by 'smoothing' the crude map. In hierarchical models, the selection of the hyper-prior distribution of dispersion parameter is crucial, in general, when the sample size is not large enough. The crude map using the standardized mortality rates (Mason and McKay, 1974), for each (i, j, k) estimates (usually the maximum likelihood estimation) of p_{ijk} based only on y_{ijk}, often features large outlying mortality rates in less (sparsely) populated areas, so that the map is visually dominated by the rates having the highest uncertainty. One problem with the

maximum likelihood is that a zero incidence case in any area provides a zero estimate of mortality rate in that area. This approach also fails to account for the possible similarity of spatial effects in adjacent areas. To choose hyper-parameters for the computation, we begin with proper but very flat priors. For both Poisson-Lognormal and Poisson-Gamma Structures, we define $\mathbf{p} = (p_{111}, p_{112}, \cdots, p_{IJK})'$, $\boldsymbol{\theta} = (\theta_1, \cdots, \theta_J)'$, $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_J)'$, $\boldsymbol{\delta} = (\delta_1, \delta_2)'$ and $\boldsymbol{\rho} = (\rho_1, \rho_2)'$ where $\boldsymbol{\delta}$ and $\boldsymbol{\rho}$ are vectors of variance components and spatial correlations, respectively. For the Poisson-Lognormal structure, we assume that for given $(\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{W})$, the first stage prior distributions of p_{ijk} follow equation (2.2), where ϵ_{ijk} are iid $\mathcal{N}(0, \delta_0)$. As mentioned earlier, given $(\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{W})$, $v_{ijk} = \log(p_{ijk})$ have independent normal distributions with means $Z_i + \theta_j + (\mu_j + W_i)(t_k - \bar{t})$ and variance δ_0. The following conditional independence assumptions are also needed.

- Given $(\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{W})$, $\mathbf{p}$ is conditionally independent of $\boldsymbol{\delta}$ and $\boldsymbol{\rho}$;

- Given $(\boldsymbol{\delta}, \boldsymbol{\rho})$, $(\mathbf{Z}, \mathbf{W})$ are independent of $(\boldsymbol{\theta}, \boldsymbol{\mu}, \delta_0)$;

- $\theta_1, \cdots, \theta_J, \mu_1, \cdots, \mu_J, \delta_1, \delta_2, \rho_1, \rho_2$ and δ_0 are mutually independent.

To fully specify the Bayesian hierarchical model, we must specify the second and third stage prior distributions of $\boldsymbol{\theta}$, $\boldsymbol{\mu}$, $\boldsymbol{\delta}$, $\boldsymbol{\rho}$ and δ_0. In particular, we assume that

- The age effects, $\theta_j \sim \mathcal{N}(\xi_{mj}, \delta_{mj})$, $j = 1, \cdots, J$;

- The slope effects of jth age group, $\mu_j \sim \mathcal{N}(\xi_{sj}, \delta_{sj})$, $j = 1, \cdots, J$;

- $\delta_l \sim \mathcal{IG}(a_l, b_l)$, $l = 0, 1, 2$;

- $\rho_1, \rho_2 \sim iid$ $\mathcal{U}(\lambda_1^{-1}, \lambda_I^{-1})$ where λ_1 and λ_I are eigenvalues of the adjacency matrix $\mathbf{C}$.

Here the hyper-parameters $(\xi_{mj}, \delta_{mj}), (\xi_{sj}, \delta_{sj})$ and (a_l, b_l) are fixed constants. On the other hand, for the Poisson-Gamma structure, we assume that for given $(\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{W})$, the first stage prior distributions of p_{ijk} follow equation (2.1), where ϵ_{ijk}^* are iid $\mathcal{G}(\psi^{-1}, \psi^{-1})$. Thus, for given $(\mathbf{Z}, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{W})$, p_{ijk} have independent gamma distributions with shape and scale parameters as ψ^{-1} and $\left\{\psi Z_i^* \theta_j^* (\mu_j^* W_i^*)^{(t_k - \bar{t})}\right\}^{-1}$, respectively. The conditional independence assumptions are same as those for Poisson-Lognormal structure except replacing δ_0 with ψ. The following (conditional) prior distributions are assumed.

- $\theta_j^* \sim \mathcal{IG}(a_{\theta_j}, b_{\theta_j})$, $j = 1, \cdots, J$;

- $\mu_j^* \sim \mathcal{IG}(a_{\mu_j}, b_{\mu_j})$, $j = 1, \cdots, J$;

- $\delta_l \sim \mathcal{IG}(a_l, b_l)$, $l = 1, 2$;

- $\psi \sim \mathcal{IG}(a_\psi, b_\psi)$;

- $\rho_1, \rho_2 \sim \ iid \ \mathcal{U}(\lambda_1^{-1}, \lambda_I^{-1})$.

Again, the hyper-parameters $(a_{\theta_j}, b_{\theta_j})$, (a_{μ_j}, b_{μ_j}), (a_l, b_l) and (a_ψ, b_ψ) are fixed constants.

In order to compare Poisson-Lognormal and Poisson-Gamma structures numerically, we decide to use the values of above hyperparameters which give approximately equal first and second moments in their distributions under the two structures. Detailed values will be given in the following Section.

3. Estimation via MCMC

For the data set considered in this article, we use the Missouri male lung cancer mortality data. This data set contains 115 counties, 4 age groups (45-54, 55-64, 65-74, 75+), and 4 time periods (1973-1977, 1978-1982, 1983-1987, 1988-1992). There are a few large urban areas and many rural communities in Missouri. The state is geologically diverse and has experienced considerable changes in mortality patterns in the recent decades. The total number of parameters $(\mathbf{p}, \boldsymbol{\theta}, \boldsymbol{\mu}, \mathbf{Z}, \mathbf{W}, \boldsymbol{\delta}, \boldsymbol{\rho})$ and δ_0 or ψ is equal to $2,083$. We use a Gibbs sampling (Gelfand and Smith, 1990) to obtain the posterior moments and densities. All of the full conditional distributions for a Gibbs sampling and generating random samples are listed in the Appendices A through C. Some of them are standard distributions such as normal (univariate or multivariate) and inverse gamma, while others require sampling from log-concave densities.

3.1 Specification of prior distributions

When we choose hyper-parameters for the computation, we decide to use the values of hyper-parameters which give approximately equal first and second moments in their distributions under the two different structures. This will enable us to compare the posterior distributions numerically, somewhat free of the impact of prior selections (cf. Kim *et al.*, 2002). In both structures, the values of hyper-parameters for the variance components are $(a_l, b_l) = (a_\psi, b_\psi) = (3.0, 0.1)$. This is equivalent to assuming that the prior of the reciprocal of the variance components has mean 30 and variance 300. Under Poisson-Lognormal structure, the values of hyper-parameters for age and overall slope of age effects are set to $(\xi_{mj}, \delta_{mj}) = (\xi_{sj}, \delta_{sj}) = (-0.5, 0.75)$. These values are comparable to those set $(a_{\theta_j}, b_{\theta_j}) = (a_{\mu_j}, b_{\mu_j}) = (3.0, 2.0)$ for Poisson-Gamma structure. It appears that the posterior estimates for each of the four different models are quite robust in terms of the choices of hyper-parameters.

4. Numerical Results

We define PLS for Poisson-Lognormal-SAR, PLC for Poisson-Lognormal-CAR, PGS for Poisson-Gamma-SAR, and PGC for Poisson-Gamma-CAR models.

4.1 Posterior Quantities

A summary of the posterior quantities of the parameters $(\boldsymbol{\theta}, \boldsymbol{\mu}, \boldsymbol{\delta}, \boldsymbol{\rho})$, δ_0 and ψ for four different models is presented in Table 1. As seen in Table 1, all posterior means except estimates for the spatial correlations $\boldsymbol{\rho}$, from four different models based on $50,000$ Gibbs cycles, are about the same while the posterior standard deviations for $(\boldsymbol{\theta}, \boldsymbol{\rho})$ disagree. It is interesting to note that posterior standard deviations, except for ρ_1, from Poisson-Gamma structure are smaller than or equal to those from the Poisson-Lognormal structure. Such results are likely due to the different tails of the gamma distribution with respect to the lognormal distribution.

Table 1: Posterior estimates based on $50,000$ Gibbs

para-	PGC		PGS		PLC		PLS	
meters	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
θ_1	-5.678	0.0248	-5.676	0.0251	-5.691	0.0358	-5.683	0.0398
θ_2	-4.510	0.0192	-4.510	0.0194	-4.511	0.0320	-4.508	0.0365
θ_3	-3.870	0.0178	-3.870	0.0180	-3.870	0.0304	-3.867	0.0357
θ_4	-3.802	0.0195	-3.802	0.0189	-3.801	0.0320	-3.798	0.0364
μ_1	-0.0086	0.0076	-0.0091	0.0077	-0.0086	0.0082	-0.0071	0.0077
μ_2	0.0121	0.0057	0.0125	0.0059	0.0121	0.0064	0.0123	0.0063
μ_3	0.0204	0.0055	0.0208	0.0053	0.0204	0.0056	0.0203	0.0057
μ_4	0.0344	0.0055	0.0344	0.0057	0.0339	0.0061	0.0332	0.0060
δ_1	0.0178	0.0031	0.0181	0.0032	0.0181	0.0032	0.0173	0.0032
δ_2	0.0030	0.0004	0.0030	0.0004	0.0031	0.0004	0.0031	0.0004
δ_0	-	-	-	-	0.0132	0.0021	0.0131	0.0021
ψ	0.0131	0.0020	0.0131	0.0021	-	-	-	-
ρ_1	0.10940	0.0375	0.06959	0.0271	0.14469	0.0217	0.11239	0.0227
ρ_2	0.00592	0.0599	0.00665	0.0374	0.00820	0.0611	0.00950	0.0395

- *Age effects $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$*: The posterior means of θ_j increase rapidly with age, except for the two upper groups which have similar values. The posterior means of μ_j show a gradual increase with age. The negative value for the youngest group indicates an overall decrease in mortality over the 20 year period.

Age 45-54, Year 73-77   Age 45-54, Year 78-82   Age 45-54, Year 83-87   Age 45-54, Year 88-92

Age 55-64, Year 73-77   Age 55-64, Year 78-82   Age 55-64, Year 83-87   Age 55-64, Year 88-92

Age 65-74, Year 73-77   Age 65-74, Year 78-82   Age 65-74, Year 83-87   Age 65-74, Year 88-92

Age 75+, Year 73-77   Age 75+, Year 78-92   Age 75+, Year 83-87   Age 75+, Year 88-92

| | 0-50 | | 150-250 | | 350-450 | | 550-650 |
| | 50-150 | | 250-350 | | 450-550 | | > 650 |

Figure 1: Frequency estimates of annual mortality rates per 10^5 population by age group, county and time period.

- *Spatial effects* **Z** *and* **W**: It is important to note that the spatial effects for the same Poisson-Gamma or Poisson-Lognormal structure are almost identical regardless of the choice of a CAR or an SAR model. On the other hand, the spatial effects for the different structures under the same CAR or SAR model are different. That is, error structures cause the differences in the spatial effects. Examination of disease maps suggests that such spatial effects in the Poisson-Gamma structure have more shrinkage toward the 'average' value (of 1), compared to those from Poisson-Lognormal structure.
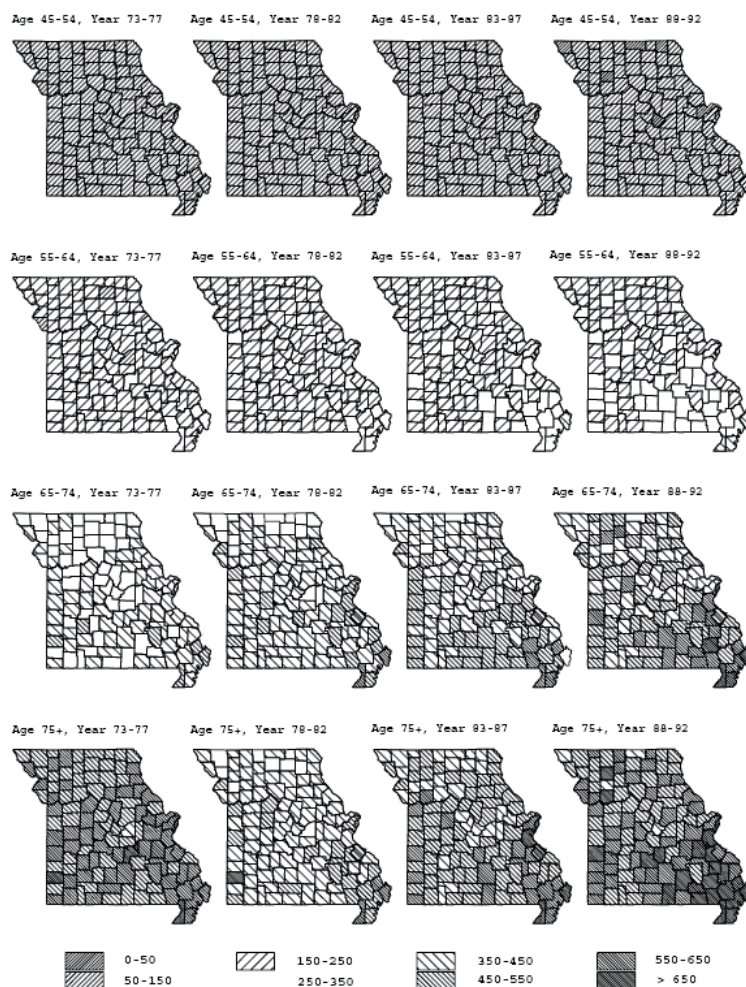
Figure 2: Bayesian posterior estimates of annual mortality rates per 10^5 population by age group, county and time period.

Through the MCMC simulations we have compared the marginal correlations of Z_i between adjacent counties and also between non-adjacent counties. We conclude that marginal correlations based on adjacent counties are, on the average, larger than those based on non-adjacent counties. For a given specific county, we have observed that correlations with adjacent counties are consistently larger than correlations with non-adjacent counties. We also approximate the spatial covariance matrix of $\mathbf{Z}$, $\delta_1 \mathbf{A}_1^{-\gamma}$, using the posterior estimates of (δ_1, ρ_1). We notice that almost all of diagonals and absolute values of off-diagonals of spatial covariance matrix based on Poisson-Gamma structure are less than those based
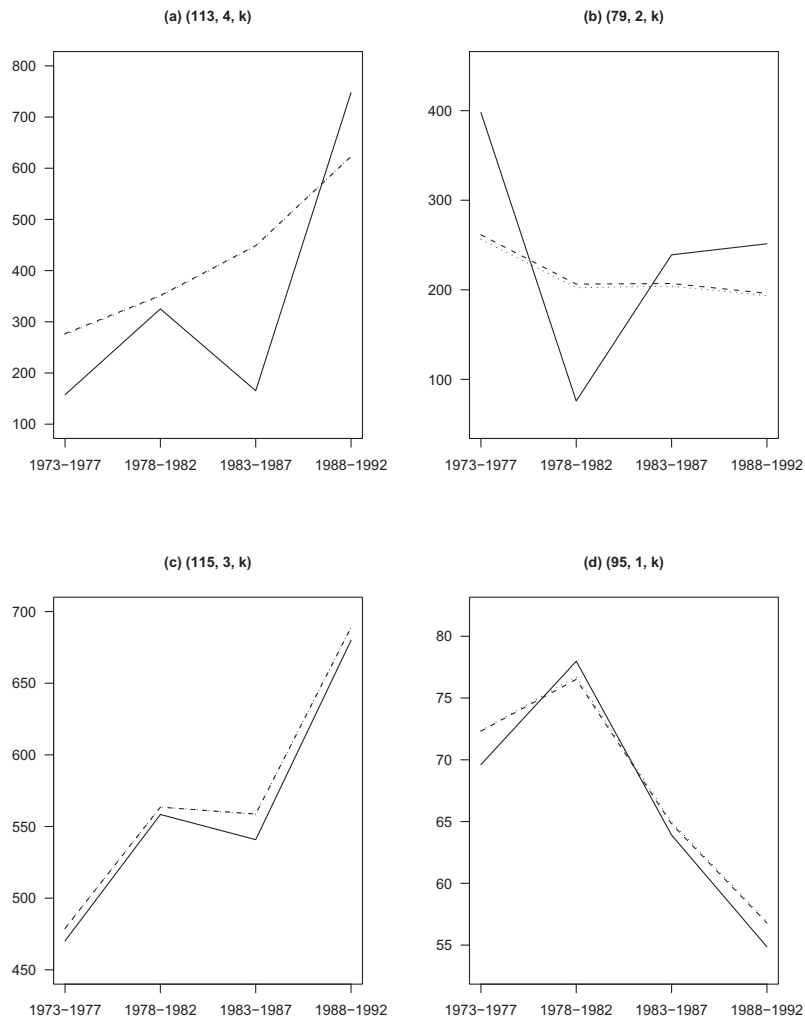
Figure 3: Plots of annual mortality rates for four selected areas over time period given age group. Solid line is for raw rates; dotted line is for PLC; and dashed line is for PLS.

on Poisson-Lognormal structure. The generalized variances, which are the determinants of the matrices, of spatial covariance matrix of $\mathbf{Z}$ indicate that the SAR model gives less variability in Z_i since the generalized variances under SAR models are smaller than those under CAR models.

- *Mortality rates* $\mathbf{p}$: Disease maps show the patterns of possible changes over time and the concentration of counties of high and low rates. Figures 1 and

2 map the frequency and Bayesian posterior estimates (based on the PGC model for the purpose of illustration since other models show the similar patterns) of the rate p_{ijk}, which have been rescaled to represent the annual rates per 10^5 population. In Figure 2, we note that most of the extreme rates in Figure 1 have disappeared. In particular, the isolated cases of high rates among the two lower age groups have disappeared as well as the isolated lower rates among the two upper age groups. It can clearly be seen how the rates have increased over time for the two upper age groups and how they tend to decrease over time for the youngest group. In spite of the general trends, it is worth to note that in each age group there exist several counties where the trend is reversed or in the opposite direction of the general trend.

Figure 3 shows plots of the annual mortality rates for four selected areas. Based on population sizes, we choose the smallest (Worth, county=113), medium (Perry, county=79), largest (St. Louis, county=95) and a special county (St. Louis city, county=115) for the purpose of demonstration. Mortality rates between the Poisson-Gamma and Poisson-Lognormal structures are close enough to each other that a difference is negligible. On the other hand, frequency estimates vary drastically for less populated counties. The Bayesian posterior estimates for densely populated counties are close to the frequency rates with small standard errors, while less populated counties have posterior means around the overall mean rates with larger standard errors. We observe that mortality rates from the Poisson-Gamma structure for less populated counties are relatively larger (toward the overall mean rates) than those from the Poisson-Lognormal structure because of more shrinkage effects. Table 2 gives the summary of the relative errors. The relative errors are defined as $|\hat{p}_{ijk} - \hat{p}_{ijk}^*|/\hat{p}_{ijk}$, where $\hat{p}_{ijk}^*$ is from one model and $\hat{p}_{ijk}$ from another, based on posterior means or standard deviations of the mortality rates. The relative errors support that the same structure generates the similar values of posterior estimates.

Table 2: Relative errors (RE) between models based on posterior means and standard deviations of mortality rates.

model		RE of mean		RE of s.d.	
$\hat{p}_{ijk}$	$\hat{p}_{ijk}^*$	max. value	%(< 0.01)	max. value	%(< 0.01)
PGC	PGS	0.04578	90.76	0.08232	66.30
PGC	PLC	0.12049	44.62	0.31836	40.60
PLC	PLS	0.04201	86.85	0.09188	61.85
PGS	PLS	0.11602	40.87	0.31656	37.23

- *Spatial Correlations ρ*: Table 1 and density plots (not shown) of the posterior distributions of the ρ indicate the spatial correlation ρ_1 for $\mathbf{Z}$ is clearly positive but ρ_2 for $\mathbf{W}$ are centered, roughly about zero. We construct 95% Bayesian credible intervals for each of (ρ_1, ρ_2) and find that the interval for ρ_1 excludes zero while that for ρ_2 does not. The positive values of ρ_1 indicate the similarities between neighboring counties. Based on ρ_1, Poisson-Lognormal structure gives larger posterior means with smaller posterior standard deviations than the Poisson-Gamma structure does. Also, the CAR model gives larger posterior means than does the SAR.

- *Variance Components $\boldsymbol{\delta}$*: The variability in $\mathbf{Z}, \mathbf{W}$ and $\mathbf{e}$ may be seen in the posterior estimates of their variance δ_1, δ_2 and δ_0 or ψ in Table 1. The concentration of the posterior distributions away from zero indicates the importance of $\mathbf{Z}, \mathbf{W}$ and $\mathbf{e}$ in the model, though the estimates are small.

- *Residual Effects $\mathbf{e}ps$*: We have computed the ordinary sums of squares due to errors of the fitted models. About 67.00% (SAR) and 67.04% (CAR) of the total variation in $\log(\hat{p}_{ijk})$ are explained by the linear fit under the Poisson-Lognormal structure, while 65.59% (SAR) and 65.70% (CAR) of the total variation in $\hat{p}_{ijk}$ under the Poisson-Gamma structure. In other words, about $32.96\% \sim 34.41\%$ of the total variation remains in the residuals, which is due to other factors that are not included in the model. Therefore, the residual effects are needed. The detailed discussion on including residual effects in hierarchical models is in Kim *et al.* (2002).

- *Convergence Diagnostic*: In monitoring the convergence of Gibbs sampling, we have run several parallel MCMC chains with different starting values and used graphical monitoring of the chains for a representative subset of the parameters. It is noted that the estimates are quite stable after about $10,000$ iterations. Among $50,000$ cycles of Gibbs sampling, we have chosen to use the last $40,000$ cycles for an estimation purpose following a burn-in of first $10,000$ cycles.

4.2 Model selection

In order to compare two competing models M_1 and M_2 that have equal prior probabilities of being selected, the Bayes factor BF_{12}, the relative weight of evidence for model M_1 compared with model M_2, is defined as

$$BF_{12} = \frac{f(\mathbf{y}|M_1)}{f(\mathbf{y}|M_2)},$$

where $f(\mathbf{y}|M_i)$ denotes the marginal density of the data $\mathbf{y}$ under M_i, $i=1,2$ (Dey, Chen and Chang, 1997). Generally, the Bayes factor is the ratio of the posterior odds for M_1 versus M_2 to the prior odds for M_1 versus M_2. Hence, model selection requires no more than the Bayes factor (Jeffreys, 1961; Kass and Raftery, 1995). Unfortunately in complex hierarchical models, in which parameters generally outnumber observations, the Bayes factor cannot be directly applied (Gelfand and Dey, 1994). Instead we use the alternative methods such as the cross-validation predictive density (Dey *et al.*, 1997) and the deviance information criterion (Spiegelhalter *et al.*, 2002).

A Monte Carlo integration of the predictive density yields the estimated conditional predictive ordinate (CPO). Between competing models M_1 and M_2, each observation favors a specific model of which the CPO is larger. In Table 3, we present the summaries of model selections based on the estimated CPOs. The values in the third and fourth columns indicate that the number of observations (out of $1,840$ total observations) which favor the models M_1 and M_2, respectively. The last column shows that the percentage of supporting the model M_1. The value of 50% implies the data set supports two competing models equally. Poisson-Lognormal structure seems to be the ones as being selected. However, the percentages of supporting (or favoring) M_1 over M_2 are not very far away from 50% mark, so no model is strictly superior to others.

Table 3: Summary of model selections based on the estimated CPOs.

model comparison		favor		$\frac{M_1}{M_1+M_2} \times 100$
M_1	M_2	M_1	M_2	
P G C	P G S	966	874	52.5
P G C	P L C	824	1016	44.8
P G C	P L S	817	1023	44.4
P G S	P L C	808	1032	43.9
P G S	P L S	810	1030	44.0
P L C	P L S	846	994	46.0
P G C	PGC with $\rho_2 = 0$	929	911	50.5
P G S	PGS with $\rho_2 = 0$	924	916	50.2
P L C	PLC with $\rho_2 = 0$	925	915	50.3
P L S	PLS with $\rho_2 = 0$	933	907	50.7

The second method we use is based on the Deviance Information Criterion (DIC). Spiegelhalter *et al.* (2002) compare two or more competing models based on the posterior distributions of the deviance where they identify 'fit' as the posterior mean of the deviance, and 'complexity' (*i.e.* the effective number of parameters, p_D) as the difference between the posterior mean of the deviance and the deviance based on the posterior means of the parameters. The 'fit'

and 'complexity' are then added to form a DIC which may be used for model comparison. These quantities can be obtained easily from an MCMC chain. Detailed discussion can be found in Spiegelhalter *et al.* (2002) and Best *et al.* (1998). For our Poisson model, the deviance $D(\mathbf{p})$ is defined (McCullagh and Nelder, 1989) as

$$D(\mathbf{p}) \;=\; 2 \sum_i \sum_j \sum_k \left[y_{ijk} \log \frac{y_{ijk}}{n_{ijk}p_{ijk}} - (y_{ijk} - n_{ijk}p_{ijk}) \right].$$

First, we compute the 'fit' of a model as $\bar{D} = \mathbf{E}_{\mathbf{p}|\mathbf{Y}}[D(\mathbf{p})]$ and the 'complexity' as $p_D = \bar{D} - D(\bar{\mathbf{p}})$. Then, we compute a DIC as $DIC = \bar{D} + p_D = D(\bar{\mathbf{p}}) + 2p_D$. In Table 4, we present the summaries of deviance statistic for competing models. The values of p_D show that Poisson-Gamma structure has fewer effective parameters (total number of second and third stage parameters is 243), so it may be regarded as more parsimonious. As a whole, all models are virtually indistinguishable in terms of fittings since the differences in DICs, $\bar{D}$s, and p_Ds are not large across competing models.

Table 4: Summary of Deviance Statistic based on 8 different models

model	$\bar{D}$	$D(\bar{\theta})$	p_D	DIC
P G C	1880.621	1636.588	244.033	2124.654
P G S	1880.484	1636.406	244.078	2124.562
P L C	1866.436	1613.662	252.774	2119.210
P L S	1865.907	1615.908	249.999	2115.906
PGC with $\rho_2 = 0$	1880.151	1636.683	243.468	2123.619
PGS with $\rho_2 = 0$	1879.557	1635.434	244.123	2123.680
PLC with $\rho_2 = 0$	1867.611	1613.662	253.949	2121.560
PLS with $\rho_2 = 0$	1867.699	1619.331	248.368	2116.067

5. Conclusions

Four spatio-temporal models discussed fit well in stabilizing the mortality rates though differences exist such that the Poisson-Gamma has more emphasis on providing shrinkage. All four procedures lead to provide more smooth and accurate maps of rates, especially in less populated areas. As discussed, the spatial effects for the same Poisson-Gamma or Poisson-Lognormal structure are very similar regardless of the choice of a CAR or an SAR model. On the other hand, those for the different structures under the same CAR or SAR model are different. That is, error structures cause the differences in the spatial effects. Poisson-Gamma structure has more tendency to shrink the posterior estimates toward the overall

mean rates than does the Poisson-Lognormal structure. As a whole, there exist more differences due to error structures (Lognormal vs. Gamma) than due to spatial correlations (SAR vs. CAR). During MCMC chains, the computational methods for SAR and CAR are almost identical except the values of γ in the covariance matrix of spatial effects. Our choice of CAR model improves the efficiency of the computation greatly over other forms of CAR models appearing in the literature. All posterior means except estimates for spatial correlations $\boldsymbol{\rho}$, based on four different models, are about the same. It is worthwhile to note that posterior standard deviations from Poisson-Gamma structure are smaller than those from Poisson-Lognormal structure.

In conclusion, due to the richness of hierarchical settings, easy interpretation of parameters and ease of implementation, any models proposed in this paper can be applied generally to other sets of data. We may also use one model to validate the performance based on other model.

Appendix A. Available full conditional distributions for Poisson-lognormal structure

We now give the full conditional distributions for our hierarchical structures. Note that the Poisson-Lognormal structure shares the same full conditional distributions except for the choice of γ in the covariance structures of spatial effects described in earlier section. This is also true for the Poisson-Gamma structure.

(i) For any (i, j, k), the conditional posterior distribution of $v_{ijk} = \log(p_{ijk})$ given others (*i.e.*, other parameters) is proportional to

$$\exp\left\{y_{ijk}v_{ijk} - n_{ijk}e^{v_{ijk}} - \frac{1}{2\delta_0}(v_{ijk} - a_{ijk})^2\right\}, \tag{5.1}$$

where

$$a_{ijk} = Z_i + \theta_j + (\mu_j + W_i)(t_k - \bar{t}). \tag{5.2}$$

(ii) For any $j = 1, \cdots, J$,

$$(\theta_j|\text{ others}) \text{ is } \mathcal{N}\left(\frac{\frac{\xi_{mj}}{\delta_{mj}} + \frac{1}{\delta_0}\sum_i\sum_k(v_{ijk} - Z_i)}{(IK/\delta_0) + (1/\delta_{mj})}, \frac{1}{(IK/\delta_0) + (1/\delta_{mj})}\right). \tag{5.3}$$

(iii) Define $c_i = \delta_0^{-1}\sum_{j,k}(v_{ijk} - \theta_j)$ and $\mathbf{c} = (c_1, \cdots, c_I)'$. Then

$$(\mathbf{Z}|\text{ others}) \text{ is } \mathcal{MVN}_I\left(\left(\frac{JK}{\delta_0}\mathbf{I} + \frac{1}{\delta_1}\mathbf{A}_1^\gamma\right)^{-1}\mathbf{c}, \left(\frac{JK}{\delta_0}\mathbf{I} + \frac{1}{\delta_1}\mathbf{A}_1^\gamma\right)^{-1}\right). \tag{5.4}$$

(iv) The conditional distribution of $(\delta_1|$ others) is $\mathcal{IG}(a_1 + \frac{1}{2}I, \; b_1 + \frac{1}{2}\mathbf{Z}'\mathbf{A}_1^\gamma\mathbf{Z})$.

(v) The conditional density of $(\rho_1|$ others) is proportional to

$$|\mathbf{I} - \rho_1\mathbf{C}|^{\gamma/2}\exp\Big\{-\frac{1}{2\delta_1}\mathbf{Z}'(\mathbf{I} - \rho_1\mathbf{C})^\gamma\mathbf{Z}\Big\}. \tag{5.5}$$

(vi) Let $s_t^2 = \sum_{k=1}^K (t_k - \bar{t})^2$. Then

$$(\mu_j|\text{others}) \text{ is } \mathcal{N}\Big(\frac{\frac{\xi_{sj}}{\delta_{sj}} + \frac{1}{\delta_0}[\sum_i \sum_k v_{ijk}(t_k - \bar{t}) - s_t^2\sum_i W_i]}{(s_t^2/\delta_0) + (1/\delta_{sj})}, \; \frac{1}{(s_t^2/\delta_0) + (1/\delta_{sj})}\Big) \tag{5.6}$$

(vii) Define $d_i = \delta_0^{-1}[\sum_j \sum_k v_{ijk}(t_k - \bar{t}) - \mu_j s_t^2]$ and $\mathbf{d} = (d_1, \cdots, d_I)'$. Then

$$(\mathbf{W}|\text{ others}) \text{ is } \mathcal{MVN}_I\Big(\Big(\frac{Js_t^2}{\delta_0}\mathbf{I} + \frac{1}{\delta_2}\mathbf{A}_2^\gamma\Big)^{-1}\mathbf{d}, \; \Big(\frac{Js_t^2}{\delta_0}\mathbf{I} + \frac{1}{\delta_2}\mathbf{A}_2^\gamma\Big)^{-1}\Big). \tag{5.7}$$

(viii) The conditional distribution of $(\delta_2|$ others) is $\mathcal{IG}(a_2 + \frac{1}{2}I, \; b_2 + \frac{1}{2}\mathbf{W}'\mathbf{A}_2^\gamma\mathbf{W})$.

(ix) The conditional density of $(\rho_2|$ others) is proportional to

$$|\mathbf{I} - \rho_2\mathbf{C}|^{\gamma/2}\exp\Big\{-\frac{1}{2\delta_2}\mathbf{W}'(\mathbf{I} - \rho_2\mathbf{C})^\gamma\mathbf{W}\Big\}. \tag{5.8}$$

(x) The conditional distribution of $(\delta_0|$ others) is $\mathcal{IG}(a_0 + \frac{1}{2}IJK, \; b_0 + \frac{1}{2}\sum_i\sum_j\sum_k$ $(v_{ijk} - a_{ijk})^2)$, where a_{ijk} is given by (5.2).

Appendix B. Available full conditional distributions for Poisson-gamma structure

(i) For any (i, j, k), the conditional distribution of $(p_{ijk}|$ others) is $\mathcal{G}(\psi^{-1} + y_{ijk}, c_{ijk} + n_{ijk})$ where $c_{ijk} = \Big\{\psi Z_i^*\theta_j^*(\mu_j^*W_i^*)^{(t_k-\bar{t})}\Big\}^{-1}$.

(ii) For any j, the conditional distribution of $(\theta_j^*|$ others) is $\mathcal{IG}(a_{\theta_j} + \frac{IK}{\psi}, \; b_{\theta_j} + \frac{1}{\psi}\sum_i\sum_k \frac{p_{ijk}}{Z_i^*\theta_j^*(\mu_j^*W_i^*)^{(t_k-\bar{t})}})$. Since $\theta_j = \log(\theta_j^*)$, the Jacobian transformation gives the desired conditional posterior distribution of θ_j.

(iii) For any j, the conditional density of $(\mu_j|$ others) is proportional to

$$\exp\Big[-\frac{1}{\psi}\sum_{i=1}^I\sum_{k=1}^K\frac{p_{ijk}}{Z_i^*\theta_j^*W_i^{*(t_k-\bar{t})}}e^{-\mu_j(t_k-\bar{t})} - a_{\mu_j}\mu_j - b_{\mu_j}e^{-\mu_j}\Big]. \tag{5.9}$$

(iv) We can rewrite $\mathbf{A}_1$ as $\mathbf{A}_1^\gamma = \mathbf{\Gamma}\mathrm{diag}\left[(1-\rho_1\lambda_1)^\gamma, \cdots, (1-\rho_1\lambda_I)^\gamma\right]\mathbf{\Gamma}' \equiv [S_{il}]_{I\mathrm{X}I}$. Then, for any i, the conditional density of $(Z_i|$ others) is proportional to

$$\exp\left[-\frac{JKZ_i}{\psi} - \frac{e^{-Z_i}}{\psi}\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{p_{ijk}}{\theta_j^*(\mu_j^*W_i^*)^{(t_k-\bar{t})}} - \frac{1}{2\delta_1}S_{ii}Z_i^2 + \frac{Z_i}{\delta_1}\sum_{\substack{l=1\\l\neq i}}^{I}Z_lS_{li}\right] \quad (5.10)$$

(v) The conditional distribution of $(\delta_1|$ others) is $\mathcal{IG}(a_1 + \frac{I}{2}, b_1 + \frac{1}{2}\mathbf{Z}'\mathbf{A}_1^\gamma\mathbf{Z})$.

(vi) Given the eigenvectors of adjacency matrix $\mathbf{C}$, write $|\mathbf{A}_1^\gamma| = \prod_{i=1}^{I}(1-\rho_1\lambda_i)^{\gamma/2}$ and $\mathbf{\Gamma}'\mathbf{Z} = (f_1, \cdots, f_I)'$. The conditional density of $(\rho_1|$ others) is proportional to

$$\exp\left[\frac{\gamma}{2}\sum_{i=1}^{I}\log(1-\rho_1\lambda_i) - \frac{1}{2\delta_1}\sum_{i=1}^{I}(1-\rho_1\lambda_i)^\gamma f_i^2\right]. \quad (5.11)$$

(vii) Write $\mathbf{A}_2^\gamma = \mathbf{\Gamma}\mathrm{diag}\left[(1-\rho_2\lambda_1)^\gamma, \cdots, (1-\rho_2\lambda_I)^\gamma\right]\mathbf{\Gamma}' \equiv [T_{il}]_{I\mathrm{X}I}$. Then, for any i, the conditional density of $(W_i|$ others) is proportional to

$$\exp\left[-\frac{1}{\psi}\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{p_{ijk}}{Z_i^*\theta_j^*\mu_j^{*(t_k-\bar{t})}}e^{-W_i(t_k-\bar{t})} - \frac{1}{2\delta_2}T_{ii}W_i^2 + \frac{W_i}{\delta_2}\sum_{\substack{l=1\\l\neq i}}^{I}W_lT_{li}\right] \quad (5.12)$$

(viii) The conditional distribution of $(\delta_2|$ others) is $\mathcal{IG}(a_2 + \frac{I}{2}, b_2 + \frac{1}{2}\mathbf{W}'\mathbf{A}_2^\gamma\mathbf{W})$.

(ix) Write $|\mathbf{A}_2^\gamma| = \prod_{i=1}^{I}(1-\rho_2\lambda_i)^{\gamma/2}$ and $\mathbf{\Gamma}'\mathbf{W} = (g_1, \cdots, g_I)'$. The conditional density of $(\rho_2|$ others) is proportional to

$$\exp\left[\frac{\gamma}{2}\sum_{i=1}^{I}\log(1-\rho_2\lambda_i) - \frac{1}{2\delta_2}\sum_{i=1}^{I}(1-\rho_2\lambda_i)^\gamma g_i^2\right]. \quad (5.13)$$

(x) Define

$$S_\psi = \prod_{i=1}^{I}\prod_{j=1}^{J}\prod_{k=1}^{K}\frac{p_{ijk}}{Z_i^*\theta_j^*(\mu_j^*W_i^*)^{(t_k-\bar{t})}}, \quad T_\psi = b_\psi + \sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{K}\frac{p_{ijk}}{Z_i^*\theta_j^*(\mu_j^*W_i^*)^{(t_k-\bar{t})}} \quad (5.14)$$

Then, the conditional density of $(\psi|$ others) is proportional to

$$\frac{1}{\Gamma(\frac{1}{\psi})^{IJK}}\left(\frac{1}{\psi}\right)^{\frac{IJK}{\psi}+a_\psi+1}S_\psi^{\frac{1}{\psi}}e^{-\frac{T_\psi}{\psi}}. \quad (5.15)$$

Appendix C. Generating random samples

It is easy to generate random samples from normal (univariate or multivariate) as well as gamma and inverse gamma distributions. If we use the parameter p_{ijk} in the computation of Poisson-Lognormal structure, we can show that the conditional density of p_{ijk} given other parameters is not of a closed form, not even log-concave. Thus, sampling from the conditional density of p_{ijk} given other parameters is not trivial (cf. rejection/acceptance sampling, Metropolis algorithm). However, as shown below, in Poisson-Lognormal structure, $v_{ijk} = \log(p_{ijk})$ has a log-concave density since v_{ijk} is a monotone (one-to-one) transformation of p_{ijk}. We will also see that the conditional densities of ρ_1 and ρ_2 in all models are log-concave. Additionally, μ_j, Z_i, W_i and $\eta = 1/\psi$ in Poisson-Gamma structure are log-concave. The adaptive rejection sampling method proposed by Gilks and Wild (1992) was used to generate random samples from these log-concave densities.

Fact 1 The conditional density of $v_{ijk} = \log(p_{ijk})$ given in Poisson-Lognormal structure is log-concave.

Proof. Let $h(v_{ijk})$ be the exponent in (5.1). The second derivative of $h(v_{ijk})$ with respect to v_{ijk} is $-n_{ijk} \exp(v_{ijk}) - \delta_0^{-1}$, which is negative. □

Let $\mathbf{\Gamma}$ be an $I \times I$ orthogonal matrix such that the adjacency matrix $\mathbf{C}$ can be decomposed as $\mathbf{\Gamma} \mathrm{diag}(\lambda_1, \cdots, \lambda_I) \mathbf{\Gamma}'$. Let $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \cdots, \lambda_I)$ and $\mathbf{S} = \mathbf{\Gamma}'\mathbf{Z} = (s_1, \cdots, s_I)'$. Then we have the following equalities,

$$\mathbf{Z}'\mathbf{A}_1^{\gamma}\mathbf{Z} = \mathbf{Z}'(\mathbf{I} - \rho_1\mathbf{C})^{\gamma}\mathbf{Z} = \mathbf{Z}'\mathbf{\Gamma}(\mathbf{I} - \rho_1\mathbf{\Lambda})^{\gamma}\mathbf{\Gamma}'\mathbf{Z} = \sum_{i=1}^{I}(1 - \rho_1\lambda_i)^{\gamma}s_i^2. \quad (5.16)$$

Fact 2 (a) The conditional density of ρ_1, given in (5.5), is equivalent to

$$h(\rho_1) \propto \exp\left\{\frac{\gamma}{2}\sum_{i=1}^{I}\log(1 - \rho_1\lambda_i) - \frac{1}{2\delta_1}\sum_{i=1}^{I}(1 - \rho_1\lambda_i)^{\gamma}s_i^2\right\}, \quad \rho_1 \in (\lambda_1^{-1}, \lambda_I^{-1}) (5.17)$$

(b) The conditional density of ρ_2 given in (5.8) is log-concave.
(c) The conditional densities of ρ_1 and ρ_2 given in (5.11) and (5.13) are also log-concave.

Proof. Part (a) follows from (5.16). For Part (b), the first and second derivatives

of $h(\rho_1)$ are

$$\frac{\partial}{\partial \rho_1} \log[h(\rho_1)] = -\frac{\gamma}{2} \sum_{i=1}^{I} \lambda_i \left\{ \frac{1}{1-\rho_1\lambda_i} - \frac{1}{\delta_1}(1-\rho_1\lambda_i)^{\gamma-1} s_i^2 \right\}, \tag{5.18}$$

$$\frac{\partial^2}{\partial \rho_1^2} \log[h(\rho_1)] = -\frac{\gamma}{2} \sum_{i=1}^{I} \lambda_i^2 \left\{ \frac{1}{(1-\rho_1\lambda_i)^2} + \frac{\gamma-1}{\delta_1}(1-\rho_1\lambda_i)^{\gamma-2} s_i^2 \right\}. \tag{5.19}$$

Here, when $\gamma=1$ or 2, the second derivative of $h(\rho_1)$ is negative. In general, the conditional density of ρ_1 is log-concave if $\gamma \geq 1$. The proof of part (c) follows directly from that of part (b). This proves the results. $\qquad\square$
fact

  For Poisson-Gamma structure,
(a) The conditional density of μ_j is log-concave.
(b) The conditional density of Z_i is log-concave.
(c) The conditional density of W_i is log-concave.
(d) The conditional density of $\eta=1/\psi$ is log-concave.

Proof. For part (a),

$$\frac{\partial}{\partial \mu_j} \log[h(\mu_j)] = \frac{1}{\psi} \sum_{i=1}^{I} \sum_{k=1}^{K} \frac{p_{ijk}(t_k - \bar{t})}{Z_i^* \theta_j^* W_i^{*(t_k-\bar{t})}} e^{-\mu_j(t_k-\bar{t})} - a_{\mu_j} + b_{\mu_j} e^{-\mu_j},$$

$$\frac{\partial^2}{\partial \mu_j^2} \log[h(\mu_j)] = -\frac{1}{\psi} \sum_{i=1}^{I} \sum_{k=1}^{K} \frac{p_{ijk}(t_k - \bar{t})^2}{Z_i^* \theta_j^* W_i^{*(t_k-\bar{t})}} e^{-\mu_j(t_k-\bar{t})} - b_{\mu_j} e^{-\mu_j} < 0.$$

For part (b),

$$\frac{\partial}{\partial Z_i} \log[h(Z_i)] = -JK\psi + e^{-Z_i} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{p_{ijk}}{\theta_j^*(\mu_j^* W_i^*)^{(t_k-\bar{t})}} - \frac{S_{ii}}{\delta_1} Z_i - \frac{1}{\delta_1} \sum_{\substack{l=1 \\ l \neq i}}^{I} Z_l S_{li},$$

$$\frac{\partial^2}{\partial Z_i^2} \log[h(Z_i)] = -e^{-Z_i} \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{p_{ijk}}{\theta_j^*(\mu_j^* W_i^*)^{(t_k-\bar{t})}} - \frac{S_{ii}}{\delta_1} < 0.$$

For part (c),

$$\frac{\partial}{\partial W_i} \log[h(W_i)] = \sum_{j=1}^{J} \sum_{k=1}^{K} \frac{p_{ijk}(t_k - \bar{t})}{Z_i^* \theta_j^* \mu_j^{*(t_k-\bar{t})}} e^{-W_i(t_k-\bar{t})} - \frac{1}{\delta_2} T_{ii} W_i - \frac{1}{\delta_2} \sum_{\substack{l=1 \\ l \neq i}}^{I} W_l T_{li},$$

$$\frac{\partial^2}{\partial W_i^2} \log[h(W_i)] = -\sum_{j=1}^{J} \sum_{k=1}^{K} \frac{p_{ijk}(t_k - \bar{t})^2}{Z_i^* \theta_j^* \mu_j^{*(t_k-\bar{t})}} e^{-W_i(t_k-\bar{t})} - \frac{1}{\delta_2} T_{ii} < 0.$$

For part (d), note that

$$
\begin{aligned}
\log[h(\eta)] &\propto (IJK\eta + a_\psi - 1)\log(\eta) + \eta\log(S_\psi) - \eta T_\psi - IJK\log[\Gamma(\eta)], \\
\frac{\partial}{\partial\eta}\log[h(\eta)] &= IJK\log(\eta) + \frac{IJK\eta + a_\psi - 1}{\eta} + \log(S_\psi) - T_\psi - IJK\frac{\partial}{\partial\eta}\log[\Gamma(\eta)], \\
\frac{\partial^2}{\partial\eta^2}\log[h(\eta)] &= \frac{IJK}{\eta} + \frac{1 - a_\psi}{\eta^2} - IJK\frac{\partial^2}{\partial\eta^2}\log[\Gamma(\eta)],
\end{aligned}
$$

where S_ψ and T_ψ are defined in equation (5.14). From Bowman and Shenton (1988), we have

$$
\frac{\partial^2}{\partial x^2}\log[\Gamma(x)] = \frac{1}{x} + \frac{1}{2x^2} + \frac{2\pi}{x}\int_0^\infty \frac{\sqrt{t}\,e^{2\pi\sqrt{t}}}{(x^2+t)(e^{2\pi\sqrt{t}}-1)^2}dt.
$$

Hence,

$$
\frac{\partial^2}{\partial\eta^2}\log[h(\eta)] = -\frac{IJK - 2 + 2a_\psi}{2\eta^2} - 2\pi\frac{IJK}{\eta}\int_0^\infty \frac{\sqrt{t}\,e^{2\pi\sqrt{t}}}{(\eta^2+t)(e^{2\pi\sqrt{t}}-1)^2}dt < 0.
$$

For given positive integers (I, J, K) and moderate size of a_ψ, the second derivative is negative. This proves the results.

Acknowledgment

References

Besag, J. (1974). Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion). *Journal of the Royal Statistical Society Series B-Statistical Methodology* **36**, 192-236.

Besag, J., York, J. and Mollié, A. (1991). Bayesian Image Restoration, with Two Applications in Spatial Statistics (with discussion). *Annals of the Institute of Statistical Mathematics* **43**, 1-59.

Best, N.G., Waller, L.A., Thomas, A., Conlon, E.M. and Arnold, R.A. (1998). Bayesian Models for Spatially Correlated Disease and Exposure Data. *Sixth Valencia International Meetings on Bayesian Statistics, Valencia, Spain*, 65-82.

Bowman, K.O. and Shenton, L.R. (1988). *Properties of Estimators for the Gamma Distribution.* Marcel Dekker.

Clayton, D. and Kaldor, J. (1987). Empirical Bayes Estimates of Age-standardized Relative Risks for Use in Disease Mapping. *Biometrics* **43**, 671-681.

Cliff, D. and Ord, J.K. (1981). *Spatial Processes: Models and Applications*. Pion.

Dey, D.K., Chen, M., and Chang, H. (1997). Bayesian Approach for Nonlinear Random Effects Models. *Biometrics* **53**, 1239-1252.

Gelfand, A.E. and Dey, D.K. (1994). Bayesian Model Choice: Asymptotics and Exact Calculations. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **56**, 501-514.

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.

Gelman, A. and Rubin, D.B. (1992). Inference from Iterative Simulation Using Multiple Sequences (with discussion). *Statistical Science* **7**, 457-511.

Ghosh, M., Natarajan, K., Stroud, T.W.F. and Carlin, B.P. (1998). Generalized Linear Models for Small Area Estimation. *Journal of the American Statistical Association* **93**, 273-282.

Gilks, W.R. and Wild, P. (1992). Adaptive Rejection Sampling for Gibbs Sampling. *Applied Statistics* **41**, 337-348.

Jeffreys, H. (1961). *Theory of Probability*. Oxford University Press.

Kass, R.E. and Raftery, A.E. (1995). Bayes Factors. *Journal of the American Statistical Association* **90**, 773-795.

Kim, H., Sun, D., and Tsutakawa, R.K. (2002). Lognormal vs. Gamma: Extra Variations. *Biometrical Journal* **44**, 305-323.

Kim, H. and Oleson, J.J. (2008). A Bayesian Dynamic Spatio-Temporal Interaction Model: An Application to Prostate Cancer Incidence. *Geographical Analysis* **40**, 77-96.

Knorr-Held, L. (2000). Bayesian Modeling of Inseparable Space-Time Variation in Disease Risk. *Statistics in Medicine* **19**, 2555-2567.

Mason, T.J. and McKay, F.W. (1974). *U.S. Cancer Mortality by County, 1950-1960.* DHEW publication Number (NIH) 74-615.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models*. Chapman and Hall, London.

Ord, K. (1975). Estimation Methods for Models with Spatial Interaction. *Journal of the American Statistical Association* **70**, 120-126.

Ripley, B.D. (1981). *Spatial Statistics*. Wiley.

Ripley, B.D. (1988). *Statistical Inference for Spatial Processes.* Cambridge University Press.

Spiegelhalter, D.J., Best, N.G., Carlin, B.R. and van der Linde, A (2002). Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **64**, 583-616.

Sun, D., Tsutakawa, R.K. and Speckman, P.L. (1999). Bayesian Inference for CAR (1) Models with Noninformative Priors. *Biometrika* **86**, 341-350.

Sun, D., Tsutakawa, R.K., Kim, H. and He, Z. (2000). Spatio-temporal Interaction with Disease Mapping. *Statistics in Medicine* **19**, 2015-2035.

Tsutakawa, R.K. (1985). Estimation of Cancer Mortality Rates: A Bayesian Analysis of Small Frequencies. *Biometrics* **41**, 69-79.

Tsutakawa, R.K. (1988). Mixed Model for Analyzing Geographic Variability in Mortality Rates. *Journal of the American Statistical Association* **83**, 37-42.

Waller, L.A., Carlin, B.P., Xia, H. and Gelfand, A.E. (1997). Hierarchical Spatio-temporal Mapping of Disease Rates. *Journal of the American Statistical Association* **92**, 607-617.

Hoon Kim
Department of Mathematics and Statistics
California State Polytechnic University
3801 West Temple Ave, Pomona, CA 91768, USA
hoonkim@csupomona.edu

HeeJeong Lim
Department of Mathematics
California State University at Fullerton
800 N. State College Blvd, Fullerton, CA 92831, USA
heejeonglim@fullerton.edu