Statistics in Medicine

Research Article

Received 21 February 2012, Accepted 22 February 2013 Published online 25 March 2013 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.5789

Bayesian semiparametric model with spatially–temporally varying coefficients selection

Bo Cai,[a*†] Andrew B. Lawson,[b] Md. Monir Hossain,[c] Jungsoon Choi,[b] Russell S. Kirby[d] and Jihong Liu[a]

In spatiotemporal analysis, the effect of a covariate on the outcome usually varies across areas and time. The spatial configuration of the areas may potentially depend on not only the structured random intercept but also spatially varying coefficients of covariates. In addition, the normality assumption of the distribution of spatially varying coefficients could lead to potential biases of estimations. In this article, we proposed a Bayesian semiparametric space–time model where the spatially–temporally varying coefficient is decomposed as fixed, spatially varying, and temporally varying coefficients. We nonparametrically modeled the spatially varying coefficients of space–time covariates by using the area-specific Dirichlet process prior with weights transformed via a generalized transformation. We modeled the temporally varying coefficients of covariates through the dynamic model. We also took into account the uncertainty of inclusion of the spatially–temporally varying coefficients by variable selection procedure through determining the probabilities of different effects for each covariate. The proposed semiparametric approach shows its improvement compared with the Bayesian spatial–temporal models with normality assumption on spatial random effects and the Bayesian model with the Dirichlet process prior on the random intercept. We presented a simulation example to evaluate the performance of the proposed approach with the competing models. We used an application to low birth weight data in South Carolina as an illustration.

Keywords: area-specific Dirichlet process; Bayesian space–time models; spatially–temporally varying coefficients; variable selection

1. Introduction

We often encounter spatial–temporal data in various disciplines such as epidemiology, ecology, political sciences, and economics. For example, the average of household income varies across different areas and time. In many applications, spatial–temporal regression models are used to explain the response variable observed over areas and time.

Suppose that the dependent variable y_{it} is observed in the ith spatial unit and the tth time point, for $i = 1, \ldots, n$ and $t = 1, \ldots, T$. We can express a general space–time model as

$$y_{it} \sim f(y_{it}|\cdot), \tag{1}$$

where $f(y_{it}|\cdot)$ denotes a conditional distribution of y_{it} given observed covariates, latent variables, and measurement errors, with mean μ_{it}, $\mu_{it} = E(y_{it})$, which is typically related to a linear predictor η_{it} through a suitable link function $g(\cdot)$, where $\eta_{it} = g(\mu_{it})$. The response variable could be observed as a continuous (e.g., disease rate), categorical (e.g., indication of disease or health status), or count (e.g.,

[a] *Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC, U.S.A.*
[b] *Division of Biostatistics and Epidemiology, Medical University of South Carolina, Charleston, SC, U.S.A.*
[c] *Division of Biostatistics and Epidemiology, Cincinnati Children Hospital Medical Center, Cincinnati, OH, U.S.A.*
[d] *Department of Community and Family Health, University of South Florida, Tampa, FL, U.S.A.*
**Correspondence to: Bo Cai, Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC, U.S.A.*
[†] *E-mail: bcai@sc.edu*

disease or death number) outcome. When the response is an area-referenced count,

$$y_{it} \sim \text{Poisson}(E_{it} \exp(\eta_{it})), \tag{2}$$

where E_{it} is an expected number of events, which is thought of as fixed and sometimes obtained by applying a standard table of sex-specific and age group-specific rates to the population count in district i at time t, n_{it}, subdivided by age and sex [1]. In our case, we set $E_{it} = Rn_{it}$ with $R = \sum_{it} y_{it} / \sum_{it} n_{it}$. The standardization here is often referred to as internal standardization because we have used the same data to compute reference rates R. We can usually express the logarithm of the relative risk, η_{it}, as

$$\eta_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + u_i + v_i + \gamma_t, \tag{3}$$

where $\mathbf{x}_{it} = (1, x_{it2}, \ldots, x_{itp})'$ denotes a $p \times 1$ vector of covariates associated with unit i and time t, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ denotes a $p \times 1$ vector of population parameters, u_i and v_i denote random effects measuring spatial similarity and excess heterogeneity, respectively, and γ_t denotes a structured temporal random component. Conventionally, a multivariate normal prior can model the fixed effects $\boldsymbol{\beta}$. We assume the parameters u_i and v_i to be independent. The parameter v_i captures the heterogeneity among the units, which is chosen to follow an exchangeable normally distributed prior, whereas u_i captures the spatial heterogeneity of data, which is assumed to follow an intrinsic conditional autoregressive (CAR) distribution (a special case of the general class of the Markov random field) [2], $u_i|u_{-i} \sim \text{CAR}(\tau)$, that is, $u_i|u_{-i} \sim \text{N}(\overline{u}_i, (\tau m_i)^{-1})$, where $u_{-i} = (u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n)'$, $\overline{u}_i = m_i^{-1} \sum_{j \in \partial_i} u_j$ with ∂_i denoting the neighbor set of unit i, m_i denoting the number of neighbors of unit i, and τ denoting the precision parameter. We define the constraint $\sum_{i=1}^n u_i = 0$ for the purpose of identifiability of the overall intercept. We assume temporal parameter γ_t to follow an autoregressive prior.

Model (2) is a typical spatiotemporal model for areal data, based on which some hierarchical structures are developed [3–5]. More complex issues occur when the space–time interaction effect (e.g., w_{it}) is included in the predictor (3) [6–8]. Ugarte *et al.* [9] presented the evaluation of the performance of various simple spatiotemporal Bayesian models. Much work, however, assumed that effects of covariates on the response were constant across areas and time. In some applications, this assumption would be inappropriate. For example, the effect of the poverty rate on the low birth weight may vary across different regions and time points. To allow coefficients to vary spatially, among others, Assunção [10] and Gamerman *et al.* [11] proposed spatially varying coefficients models for small-area data. Dreassi *et al.* [12] developed a space model with time-dependent covariates for small-area data. Cai *et al.* [13] proposed a Bayesian regression model with multivariate linear splines for the analysis of space–time data. For point-referenced data, authors have developed some approaches. Gelfand *et al.* [14] proposed a spatial process modeling for univariate and multivariate dynamic spatial data. Paez *et al.* [15] developed spatially varying dynamic coefficient models.

In the aforementioned spatial–temporal models, the spatially varying coefficients are often assumed to follow Gaussian distributions. In practice, the normality assumption is difficult to verify empirically and may be overly restrictive as spatially varying coefficients may follow other distributions and may have clustering issues. Recently, authors have developed some approaches to relax the normality assumption for modeling point-referenced data. Gelfand *et al.* [16] proposed a Bayesian nonparametric spatial modeling with spatial Dirichlet process (DP) mixture models. Duan *et al.* [17] developed a generalized spatial DP. Reich and Fuentes [18] described a multivariate semiparametric Bayesian spatial model for spatial data. In contrast, for areal data, the semiparametric model with spatially–temporally varying coefficients of covariates lacks development. Li *et al.* [19] proposed nonparametric hierarchical models for areal data. They modeled the spatial random intercept by using area-referenced spatial stick-breaking prior with the logit link between the weight and the random variate from the CAR.

In this paper, we focus on developing a Bayesian semiparametric space–time model with spatially–temporally varying coefficients of covariates. We model the spatially varying coefficients by using the area-specific stick-breaking representation for the DP prior with the generalized transformation between the weight in the stick-breaking prior and the spatially specific random variate from the CAR. The generalized transformation includes the linear link, logit link, and probit link, providing more realistic weights associated with the spatial information in the areal data. We modeled temporally varying coefficients by using a dynamic model. Each covariate could have different effects on the response variable, including no effect, the overall-only effect, the spatial-only effect, the temporal-only effect, and the spatial–temporal effect. The proposed model allows for the uncertainty of inclusion of different effects for each covariate. We use the variable selection procedure through determining the probabilities of different effects for each covariate.

The remainder of the article proceeds as follows. Section 2 describes the semiparametric model with spatially–temporally varying coefficients while allowing for uncertainty of inclusion of the coefficients. We describe prior specification and posterior implementation. Section 3 discusses the model evaluation and comparison. Section 4 evaluates the performance of the approach on the basis of a simulated example. Section 5 illustrates the approach via real spatial–temporal data. Finally, Section 6 summarizes and discusses the results.

2. Semiparametric model with uncertainty of spatially–temporally varying coefficients

2.1. The model with selection of spatially–temporally varying coefficients

We consider modeling the logarithm of the relative risk as

$$\eta_{it} = \mathbf{x}_{it}'\boldsymbol{\theta}_{it}, \tag{4}$$

where $\boldsymbol{\theta}_{it} = (\theta_{it1}, \ldots, \theta_{itp})'$ denotes a $p \times 1$ vector of spatially–temporally varying coefficients of covariates. One can decompose each element of coefficients $\boldsymbol{\theta}_{it}$ as $\theta_{itk} = \alpha_k + \beta_{ik} + \gamma_{tk}$, for $k = 1, \ldots, p$, where α_k denotes the global effect of the kth covariate, β_{ik} denotes the spatially structured random effect of the kth covariate, and γ_{tk} denotes the temporally specific effect of the kth covariate. We took the regression coefficients to be independent across covariates. However, this decomposition assumes that each covariate simultaneously has an overall effect and spatially and temporally varying effects on the response. This assumption is too restrictive in general, which might make the fitted model overparameterized because a covariate may have the following: (i) no effect; (ii) only a fixed effect; (iii) only a spatially specific effect given the fixed effect; (iv) only a temporally specific effect given the fixed effect; and (v) all three effects on the response. To account for this uncertainty, we consider defining θ_{itk} as

$$\theta_{itk} = \delta_{1k}\alpha_k + \delta_{1k}(\delta_{2k}\beta_{ik} + \delta_{3k}\gamma_{tk}), \tag{5}$$

where δ_{1k}, δ_{2k}, and δ_{3k} denote the indicator variables for α_k, β_{ik}, and γ_{tk}, respectively. We fix δ_{11}, δ_{21}, and δ_{31} to be 1 for all i and t to reflect some overall spatial–temporal effect. For $k \geqslant 2$, under this construction, a covariate has the following:

1. No effect (i.e., $\theta_{itk} = 0$) for all i and t if $\delta_{1k} = 0$
2. Only a fixed effect (i.e., $\theta_{itk} = \alpha_k$) for all i and t if $\delta_{1k} = 1$ and $\delta_{2k} = \delta_{3k} = 0$
3. Only a spatially specific effect given the fixed effect (i.e., $\theta_{itk} = \alpha_k + \beta_{ik}$) for all t if $\delta_{1k} = \delta_{2k} = 1$ and $\delta_{3k} = 0$
4. Only a temporally specific effect given the fixed effect (i.e., $\theta_{itk} = \alpha_k + \gamma_{tk}$) for all i if $\delta_{1k} = \delta_{3k} = 1$ and $\delta_{2k} = 0$
5. Spatially–temporally varying effects (i.e., $\theta_{itk} = \alpha_k + \beta_{ik} + \gamma_{tk}$) across areas and time if $\delta_{1k} = \delta_{2k} = \delta_{3k} = 1$

We assume that a covariate has no spatially and temporally specific effects if it does not have a global effect. In addition, given a global effect, we assume that a covariate having the spatially specific effect is independent of having the temporally specific effect. Thus, for the priors of the indicators, we have $\pi(\delta_{1k}, \delta_{2k}, \delta_{3k}) = \pi(\delta_{2k}|\delta_{1k})\pi(\delta_{3k}|\delta_{1k})\pi(\delta_{1k})$, where $\pi(\delta_{1k} = 1) = p_{1k}$, $\pi(\delta_{2k} = 0|\delta_{1k} = 0) = \pi(\delta_{3k} = 0|\delta_{1k} = 0) = 1$, $\pi(\delta_{2k} = 1|\delta_{1k} = 1) = p_{2k}$, and $\pi(\delta_{3k} = 1|\delta_{1k} = 1) = p_{3k}$. Then, the prior for δ_{1k}, δ_{2k}, and δ_{3k} is expressed as

$$\pi(\delta_{1k}, \delta_{2k}, \delta_{3k}) = \begin{cases} 1 - p_{1k} & \text{if } \delta_{1k} = \delta_{2k} = \delta_{3k} = 0 \\ p_{1k}(1 - p_{2k})(1 - p_{3k}) & \text{if } \delta_{1k} = 1 \text{ and } \delta_{2k} = \delta_{3k} = 0 \\ p_{1k}p_{2k}(1 - p_{3k}) & \text{if } \delta_{1k} = \delta_{2k} = 1 \text{ and } \delta_{3k} = 0 \\ p_{1k}p_{3k}(1 - p_{2k}) & \text{if } \delta_{1k} = \delta_{3k} = 1 \text{ and } \delta_{2k} = 0 \\ p_{1k}p_{2k}p_{3k} & \text{if } \delta_{1k} = \delta_{2k} = \delta_{3k} = 1 \end{cases}. \tag{6}$$

It is obvious that the sum of the probabilities in (6) equals 1. These priors provide prior probabilities of the five different scenarios. It is shown that the indicator δ_{1k} allows the kth covariate to be included or excluded from the model whereas δ_{2k} and δ_{3k} indicate if the kth covariate has spatially–temporally varying effects given that it is included in the model. The proposed variable selection structure (5) can be thought of as a general case of the variable selection method for spatially–temporally varying effects of covariates. When there are only global effects for the covariates, the proposed model reduces to the model in (3). When the observations are only spatially dependent, the proposed structure reduces to the one by Reich *et al.* [20], where they focused on variable selection in the parametric model with spatially varying coefficients. If the indicators are 1s (i.e., there is no variable selection), the model becomes a spatially–temporally varying coefficient model. Specially, when the observations are only spatially dependent, the proposed model reduces to the spatially varying regression model [10, 11], that is, $\eta_i = \mathbf{x}_i' \boldsymbol{\theta}_i$ with $\boldsymbol{\theta}_i = \boldsymbol{\alpha} + \boldsymbol{\beta}_i$. One might be concerned with the identifiability of the indicators and the coefficients in (5). This concern can be relieved as we are actually interested in $\delta_{1k}\alpha_k$, $\delta_{1k}\delta_{2k}\beta_{ik}$, and $\delta_{1k}\delta_{3k}\gamma_{tk}$, and they are identifiable. On the other hand, Bayesian identifiability concerns the question of whether the data and prior provide information for updating the indicators and the coefficients [21]. For example, when the indicator $\delta_{1k} = 0$, all the coefficients (i.e., α_k, β_{ik}, and γ_{tk}) only rely on the priors. The data will be involved in updating the coefficient(s) when $\delta_{1k} = 1$.

To allow for flexibility of the prior probability, p_{lk}, for $l = 1, 2, 3$, we consider choosing a hyperprior beta distribution for the prior exclusion probability, $p_{lk} \sim \text{Beta}(c_l, d_l)$. Given these prior probabilities, we can easily calculate the full conditional probabilities for different scenarios shown in (6) through the categorical distribution (see details in Appendix 6). For the choice of c_l and d_l $(l = 1, 2, 3)$, following the suggestion by Geisser [22], we choose $c_l = d_l = 1$, which yields the uniform hyperprior. Scott and Berger [23] discussed the choice of priors for the prior probability. They concluded that the objective prior (i.e., the uniform prior) for the prior probability can easily be implemented computationally whereas incorporation of subjective prior information can be beneficial when available. In our case, we have no subjective information about the prior probability of inclusion of the covariates, resulting in choosing a uniform prior. For more details, please refer to Geisser [22], Scott and Berger [23], and Cui and George [24], among others.

2.2. *Nonparametric modeling for spatially varying coefficients*

Typically, we took the prior of the global effect of the kth covariate, α_k, to be $N(0, \tau_{\alpha,k}^{-1})$, where $\tau_{\alpha,k}$ is the precision following a gamma prior $\text{Gamma}(a_{\alpha,k}, b_{\alpha,k})$ with mean $a_{\alpha,k}/b_{\alpha,k}$ and variance $a_{\alpha,k}/b_{\alpha,k}^2$. We could take the conditional specification of the prior for the temporally varying effect of the kth covariate as $\text{N}(\gamma_{t-1,k}, \tau_{\gamma,k}^{-1})$, for $t = 1, \ldots, T$, with $\tau_{\gamma,k}$ being the precision following $\text{Gamma}(a_{\gamma,k}, b_{\gamma,k})$. We chose the initial element γ_{0k} to be 0. For the spatially varying effect of the kth covariate, a conventional choice is the conditional distribution $\beta_{ik}|\beta_{-ik} \sim \text{N}(\overline{\beta}_{ik}, (m_i \tau_{\beta,k})^{-1})$, where $\beta_{-ik} = (\beta_{1k}, \ldots, \beta_{i-1,k}, \beta_{i+1,k}, \ldots, \beta_{nk})'$, $\overline{\beta}_{ik} = m_i^{-1} \sum_{j \in \partial_i} \beta_{jk}$, m_i denotes the number of neighbors of area i, and $\tau_{\beta,k}$ denotes the precision following a gamma prior $\text{Gamma}(a_{\beta,k}, b_{\beta,k})$. For identifiability, the constraint for β_{ik}'s is $\sum_i \beta_{ik} = 0$ for $k = 1, \ldots, p$. However, the normal prior assumption constrains the distributions that the spatially varying random effects may follow. In contrast, the nonparametric prior over distributions provides wider support, typically the support being the space of all distributions (i.e., an infinite dimensional space). As a result, a nonparametric assumption allows for various shapes of the distribution, which may more accurately reflect our prior belief about the true distribution of spatially varying random effects. To allow for uncertainty of distributions that β_{ik} may follow, we consider $\beta_{ik} \sim G_{ik}$, where G_{ik} is an unknown random distribution varying across different areas. We can then choose a prior distribution for G_{ik} with support on the space of all probability measures.

Among the nonparametric processes (e.g., Gaussian process and Pólya tree process), the DP is one of the most prominent random probability measures because of its richness, computational ease, and interpretability. We considered using the DP in our approach for several reasons. First, any distribution over its space can be approximated arbitrarily and accurately in the weak limit by a sequence of draws from the DP [25]. Second, because the distributions drawn from the DP are discrete, the DP has the clustering property that allows for repeated values, implying that multiple β_{ik}'s can take on the same value simultaneously. This feature of the DP is desirable and reflects the attribute of the spatially varying coefficients, which are typically clustered. Third, the stick-breaking representation [26] (which will be described later on) of the DP provides a convenient way of incorporating the area-specific information into the random distribution of β_{ik}'s. Finally, with the nice representation such as stick breaking, we can efficiently implement the DP. For more details on nonparametric Bayesian processes, one may refer to Ghosal and van der Vaart [25].

 Statist. Med. **2013**, 32 3670–3685

We can specify the DP prior as $\text{DP}(MG_0)$, where M is a concentration parameter and G_0 is the base measure of the DP. Under this specification, for any partition $\mathbf{B} = (B_1, \ldots, B_q)'$ of $\mathcal{R}$, we have

$$\{G(B_1), \ldots, G(B_q)\} \sim \text{D}(MG_0(B_1), \ldots, MG_0(B_q)),$$

where $D(\cdot)$ denotes the Dirichlet distribution on the simplex of R^n. This structure centers the distribution at the parametric base distribution, G_0, while allowing the true distribution to deviate from the parametric form. M controls the amount of uncertainty in the parametric assumption. As M tends to 0, most of the samples share the same value sampled from the base measure G_0, whereas when M tends to infinity, the samples are almost independent and identically distributed samples from G_0. One of the popular representations of the DP prior is the Pólya urn representation [27, 28]. Briefly, we can express a Pólya urn prior of β_i as

$$(M+n-1)^{-1}MG_0 + (M+n-1)^{-1}\sum_{s=1}^{k^{(i)}} r_s^{(i)}\delta_{\beta_s^{*(i)}}(\cdot),$$

where $k^{(i)}$ denotes the number of distinct values across all β_j's excluding β_i, $r_s^{(i)}$ denotes the frequency of all β_j's (excluding β_i) being equal to the unique value $\beta_s^{*(i)}$, and $\delta_{\beta^*}(\cdot)$ denotes the degenerate distribution at β^*. Although the Pólya urn Gibbs sampling can be implemented straightforwardly, some limitations remain. When model (1) is not a normal distribution, it is problematic to calculate the probability of generating new samples from the posterior on the basis of the prior and the likelihood because of its nonconjugacy. In addition, from the posterior distribution, we updated the parameter one at a time by using Gibbs sampling. This procedure could lead to a slow-mixing problem. Although an accelerated step [28] can enhance the mixing behavior, slow mixing may still occur because of the inherent property of one-at-a-time updates.

To avoid the limitations with the Pólya urn Gibbs sampling, we considered the blocked Gibbs sampling based on the finite-dimensional Dirichlet priors [29]. With stick-breaking representation [26], we can express the finite-dimensional prior G as $G \overset{d}{=} \sum_{s=1}^{r} \omega_s \delta_{\theta_s}(\cdot)$, where r denotes the number of mixture components, ω_s denotes the weight, and $\delta(\cdot)$ denotes a discrete measure concentrated at θ_s, which is randomly generated from the base measure G_0. For the choice of the truncation of the mixture, Ishwaran and Zarepour [30] suggested using a reasonably large value such as 50 for the sample size.

To allow the unknown distribution of β_{ik} to vary across different areas, we propose to model the spatially varying coefficients by using the area-specific stick-breaking prior. Let $\mathbf{S}_k = (S_{1k}, \ldots, S_{nk})'$ be a configuration determining a classification of $\boldsymbol{\beta}_k = (\beta_{1k}, \ldots, \beta_{nk})'$ into r_k distinct values $\boldsymbol{\beta}_k^* = (\beta_{1k}^*, \ldots, \beta_{r_k k}^*)'$, with $S_{ik} = s$ if β_{ik} in area i belongs to group s for covariate k in terms of the spatially varying effect, that is, $\beta_{ik} = \beta_{sk}^*$, for $s = 1, \ldots, r_k$. Then, we can model β_{ik} as follows

$$S_{ik} \sim \sum_{s=1}^{r_k} \omega_{isk}\delta_s(\cdot),$$

$$\omega_{isk} = V_{isk}^* \prod_{l=1}^{s-1}\left(1 - V_{ilk}^*\right),$$

$$\beta_{sk}^*|\cdot \sim \text{N}\left(0, \tau_{\beta,k}^{-1}\right), \quad \text{for} \quad s = 1, \ldots, r_k,$$

where $V_{isk}^* = u_{isk}V_{sk}$, $V_{sk} \overset{\text{iid}}{\sim} \text{Beta}(1, M_k)$ and $\prod_{l<s}\left(1 - V_{ilk}^*\right) = 1$ for $s = 1$. We define the parameter u_{isk} as a covariate-specific spatial weight that depends on the location-associated random variate. Because $u_{isk} \in (0,1)$, following Ishwaran and James [29], we can show that $\sum_{s=1}^{r_k} \omega_{isk} = 1$ is almost surely in the aforementioned area-specific stick breaking. We used a transformation $g(u_{isk}) = \phi_{isk}$, where ϕ_{isk} is assumed to follow a $\text{CAR}(\tau_k)$ prior. The transformation links the spatial weight to the CAR-distributed variate. Unlike the logit transformation used by Li *et al.* [19], we considered a more general transformation family introduced by Aranda-Ordaz [31],

$$g(u) = \frac{2}{\lambda}\frac{u^\lambda - (1-u)^\lambda}{u^\lambda + (1-u)^\lambda}, \qquad (7)$$

where λ denotes the transformation parameter. The choice of different values of λ results in various link functions. This includes that $\lambda = 0, 0.4, 1$ corresponds to the logit transformation in the limit, the probit link in approximation, and the linear transformation, respectively. We then define the inverse transformation function as

$$u = h(\phi) = \frac{(1+\phi\lambda/2)^{1/\lambda}}{(1+\phi\lambda/2)^{1/\lambda} + (1-\phi\lambda/2)^{1/\lambda}} \text{ for } |\phi\lambda| < 2, \quad (8)$$

$h(\phi) = 0$ for $\phi\lambda \leqslant -2$, and $h(\phi) = 1$ for $\phi\lambda \geqslant 2$. Because the transformation is symmetric, we can focus on $\lambda \geqslant 0$. We choose a uniform prior for λ in the range of $(0, 0.5)$ where the logit and probit links are covered. This setting also allows ϕ to vary in a reasonable range. We chose the prior of the concentration parameter M_k to be Uniform$(0, 10)$ [32].

2.3. *Posterior computation*

We chose priors for the parameters as described in Section 2.1. The posterior computation relies on a blocked Gibbs sampling algorithm in which we iteratively sampled from the full conditional distributions of a block of the parameters. For an update of a single parameter from the nonconjugate distribution, we used adaptive rejection Metropolis sampling [33]. For a block of parameters, the posterior computation relies on the Gibbs sampler and Metropolis–Hastings algorithms. After the values for the parameters are initialized, the proposed MCMC algorithm proceeds in a series of steps outlined in Appendix 6. We generate samples from the joint posterior distribution of the parameters by repeating those steps for a large number of iterations after apparent convergence.

3. Model comparison

The deviance information criterion (DIC) [34] is widely used as a model comparison tool. DIC is shown to be an approximation to a penalized loss function based on the deviance with a penalty derived from a cross-validation argument. However, the implicit approximation is valid only when the effective number of parameters is much smaller than the number of independent observations [35]. Plummer [35] pointed out that in disease mapping, this assumption does not hold, resulting in DIC underpenalizing the complex models. Plummer [35] proposed using penalized loss functions instead of p_D, the effective number of parameter, to assess model adequacy. However, as Plummer [35] noticed, this method requires MCMC runs with each observation left out in turn. Such a calculation is not feasible in general, especially for large data sets. In this article, we considered a comparison method based on the conditional predictive ordinate (CPO) [36–39]. We defined the CPO for the ith observation at time t as the cross-validated marginal posterior predictive density

$$\begin{aligned} \text{CPO}_{it} &= f(y_{it}|\mathbf{y}_{(it)}) \\ &= \int f(y_{it}|\boldsymbol{\theta}) f(\boldsymbol{\theta}|\mathbf{y}_{(it)}, \mathbf{x}_{(it)}) \mathrm{d}\boldsymbol{\theta} \\ &= \left(\int \frac{1}{f(y_{it}|\boldsymbol{\theta}, \mathbf{x}_i)} f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x}) \mathrm{d}\boldsymbol{\theta} \right)^{-1}, \end{aligned}$$

where $\mathbf{y}_{(it)}$ denotes the vector of observations with the ith observation at time t deleted and $\boldsymbol{\theta}$ is the vector of model parameters. We can estimate the cross-validation likelihood by

$$L_{\text{CV}} = \prod_{i=1}^{n} \prod_{t=1}^{T} \text{CPO}_{it}.$$

Because the quantity of the cross-validation likelihood is typically close to 0, we can use the negative cross-validatory predictive log-likelihood [40],

$$NLLK_{\text{CV}} = -\sum_{i=1}^{n} \sum_{t=1}^{T} \log \text{CPO}_{it}.$$

Because a closed form of CPO_{it} is usually unavailable, we could straightforwardly obtain a Monte Carlo estimate of CPO_{it} through MCMC samples $\{\boldsymbol{\theta}^{(s)}\}_{s=1}^{N}$ from the posterior distribution $f(\boldsymbol{\theta}|\mathbf{y}, \mathbf{x})$,

$$\widehat{\text{CPO}}_{it} = \left(\frac{1}{N}\sum_{s=1}^{N}\frac{1}{f(y_{it}|\boldsymbol{\theta}^{(s)}, \mathbf{x}_i)}\right)^{-1},$$

where N is the number of iterations after a burn-in period. Accordingly, we can calculate the estimate of the negative cross-validatory predictive log-likelihood. Because a large CPO indicates agreement between the observation and the model, a model with a smaller $NLLK_{\text{CV}}$ for all observations implies a better fit.

4. A simulation study

We evaluated the performance of the proposed approach, including the accuracy of the estimates, the sensitivity to different choices of hyperparameters, and a comparison of the proposed model with other space–time models. Without loss of generality and for illustration purposes, we created the spatial data using the South Carolina geographical structure containing 46 counties. We generated the data for $n = 46$ counties over $T = 10$ time points on the basis of the model $y_{it} \sim \text{Poisson}(E_{it}\exp(\eta_{it}))$, where the log-relative risk is $\eta_{it} = \mathbf{x}_{it}'\boldsymbol{\theta}_{it}$ with $\mathbf{x}_{it} = (1, x_{it2}, x_{it3}, x_{it4}, x_{it5})'$ and $\boldsymbol{\theta}_{it} = \boldsymbol{\alpha} + \boldsymbol{\beta}_i + \boldsymbol{\gamma}_t$. We chose $\boldsymbol{\alpha} = (1, 1, 1, 1, 0)'$, $\boldsymbol{\beta}_i = (0, \beta_{i2}, \beta_{i3}, 0, 0)'$ with β_{i2} being clustered to follow four different distributions and β_{i3} to follow five different distributions (Figure 1), and $\boldsymbol{\gamma}_t = (0, \gamma_{t2}, 0, \gamma_{t4}, 0)'$ with $\gamma_{t2} \sim \text{N}(\gamma_{t-1,2}, 0.5)$ and $\gamma_{t4} \sim \text{N}(\gamma_{t-1,4}, 1)$, for $t = 1, \ldots, T$. This setting implies that the first covariate (i.e., the intercept) only has an overall effect, the second covariate has the fixed and spatial–temporal effects, the third covariate has the fixed and spatial effects, the fourth covariate has the fixed and temporal effects, and the fifth covariate has no effect. We generated $\mathbf{x}_{itl} \sim \text{Uniform}(0, 1)$ for $l = 2, \ldots, 5$.

We specified the priors for the parameters of the proposed model as follows. We used Gamma(0.05, 0.05) as the prior for $\tau_{\alpha,k}$ and $\tau_{\gamma,k}$. Following Ishwaran and James [29], we chose Gamma(2, 2) as the prior for M_k to encourage both small and large values of M_k. Following Ishwaran and Zarepour [30], we chose $r_k = n = 46$. We chose the prior for the spatially structured random effects β_{ik} as the nonparametric prior described in Section 2.2. We chose prior probabilities in Equation (6) to be 0.5 to express an equal chance for inclusion and exclusion.

We implemented the analysis using the Gibbs sampler described in Section 2.3. We generated 50,000 iterations after a burn-in of 10,000 iterations. We assessed convergence by using a variety of diagnostics described by Cowles and Carlin [41] and implemented it using `CODA` [42] in `R` [43]. The diagnostic tests showed rapid convergence and efficient mixing. We estimated the parameters by thinning the chain by a factor of 5 to obtain a sample size of 10,000.

We compared the proposed model (model 5) with the four competing spatiotemporal models. The following lists the log-relative risks of these models:

Model 1: $\eta_{it} = \mathbf{x}_{it}'\boldsymbol{\alpha} + u_i + v_i + \gamma_t$
Model 2: $\eta_{it} = \mathbf{x}_{it}'\boldsymbol{\alpha} + \mathbf{x}_{it}'\mathbf{u}_i + v_i + \gamma_t$
Model 3: $\eta_{it} = \mathbf{x}_{it}'\boldsymbol{\gamma}_t + u_{it} + v_{it}$
Model 4: $\eta_{it} = \mathbf{x}_{it}'\boldsymbol{\alpha} + u_i + \gamma_t$

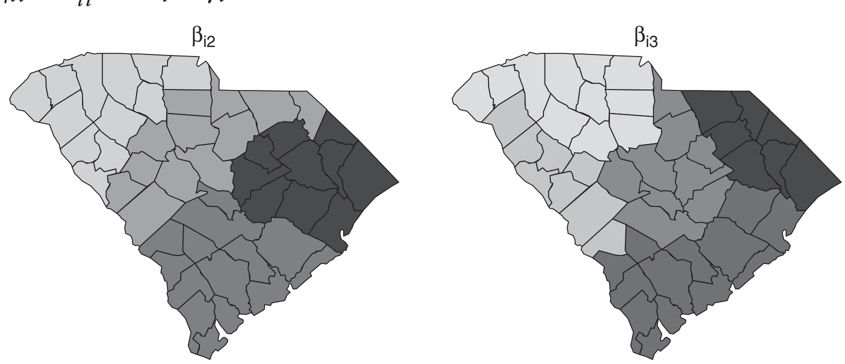

Figure 1. The design of two spatial random effects, β_{i2} and β_{i3}, in the simulation study, where the clusters with different colors in the map show different distributions.

In the four models, we followed conventional settings by specifying the prior of $\boldsymbol{\alpha}$ as $N_p(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$ with $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} \sim \text{IWishart}(p, \boldsymbol{\Sigma}_0^{-1})$, where $\boldsymbol{\Sigma}_0^{-1}$ is a 5×5 precision matrix with a diagonal element of 0.1 and an off-diagonal element of 0.05. We chose the prior of u_i in model 1 as $\text{CAR}(\tau_1^*)$ with $\tau_1^* \sim \text{Gamma}(0.005, 0.005)$. We took the prior of v_i as $N(0, \tau_2^{*-1})$ with $\tau_2^* \sim \text{Gamma}(0.005, 0.005)$. We chose the conditional specification of the prior of γ_t as $N(\gamma_{t-1}, \tau_3^{*-1})$ with $\gamma_0 \sim N(0, \tau_3^{*-1})$ and $\tau_3^* \sim \text{Gamma}(0.005, 0.005)$. In model 2, we chose the prior of $\mathbf{u}_i$ as a multivariate CAR model, $\text{MCAR}(\boldsymbol{\Sigma}_{\mathbf{u}}^{-1})$. In model 3, we chose the prior of $\boldsymbol{\gamma}_t$ as $N(\boldsymbol{\gamma}_{t-1}, \boldsymbol{\tau}_4^{*-1})$. We took the priors of $\mathbf{u}_t$ and $\mathbf{v}_t$ in model 3 as $\text{CAR}(\tau_5^*)$ and $N(\mathbf{0}, \tau_6^{*-1}\mathbf{I})$, respectively, for $t = 1, \ldots, T$. In model 4, we assume the prior of u_i to be a typical DP prior. By relying on the BlackBox component builder, `WinBUGS` [44] allows one to carry out a relatively simple Bayesian statistical modeling by simply specifying a model and the priors for its parameters. For this reason, we implemented models 1–4 using `WinBUGS`. Although it can be conceptually implemented using `WinBUGS`, we write the proposed model in `R` because of slowness and lack of flexibility of `WinBUGS`. From our experience, both `WinBUGS` and `R` programs provide really similar results.

The second column in Table I presents the comparison of the estimated negative cross-validatory predictive log-likelihoods for the four models. It shows that model 5 with the smallest value of $NLLK_{\text{CV,sim}}$ outperforms the other four models. Table II shows the posterior probabilities of inclusion of covariates in the five different cases listed in (6). It is evident that the model selects the designed covariate structure for each covariate with the highest posterior probability.

In the simulation study (and the real data example), we checked sensitivity of the results to the prior specification by repeating the analyses with different hyperparameters. Particularly, we applied the gamma prior, Gamma(0.01, 0.005), for the precision and the uniform prior, Uniform(0, 50), for the standard deviation. Although we did not show the details, there is basically no difference in parameter estimates, inferences, or model ranking for the prior specification. One may also choose other potential priors such as a half-Cauchy prior. According to Gelman [45], the choice of noninformative priors for some scale parameter of the parameter with a common distribution may have a big impact on inferences, especially when the number of clusters is small (say, below five) or the cluster-level variance is close to 0. However, using the traditional gamma prior does not seem to sensitively affect the inference in our cases. The reasons might be that, first, the number of subjects (i.e., clusters) is relatively large ($n = 46$) and, second, because the random effects in our hierarchical model follow a random distribution rather than common distributions, it is not clear which prior for the variance of random effects should be more appropriate. In addition, our sensitivity analysis shows the appropriateness of the prior specification in the proposed model.

Table I. Model comparison based on the negative cross-validatory log-likelihood for the simulated example and application to the low birth weight data in South Carolina.

Model	$NLLK_{\text{CV, sim}}$	$NLLK_{\text{CV,app}}$
Model 1	1189.78	1897.18
Model 2	1166.20	1854.43
Model 3	1175.92	1869.29
Model 4	1191.93	1810.67
Model 5	879.51	1689.83

Table II. Posterior probabilities of inclusion of the four covariates in the simulation example.

	Case				
Predictor	(0,0,0)	(1,0,0)	(1,1,0)	(1,0,1)	(1,1,1)
x_2	0.03	0.06	0.14	0.15	0.62
x_3	0.02	0.04	0.65	0.13	0.16
x_4	0.03	0.09	0.08	0.66	0.14
x_5	0.70	0.16	0.05	0.07	0.02

Case $(\delta_1, \delta_2, \delta_3)$ indicates the five scenarios described in (6).

5. Application to low birth weight data in South Carolina

As an illustration, we applied the approach to the data of county-specific low birth weights (i.e., birth weight is less than 2500 g) across 46 counties in the state of South Carolina during the period 1997–2006. As the observations were made yearly, we included a total of 460 observations in the data. We obtained the number of county-level low birth weights from the South Carolina Department of Health and Environmental Control. We considered the county-level population density (defined as population divided by the total land area in square miles), proportion of African Americans, median household income, and unemployment rate as socioeconomic predictors of low birth weights. The population density, the proportion of African American population, and the household income were acquired from the US census. The unemployment rates were attained from the US Bureau of Labor Statistics. In addition, we also considered aggregate data based on birth certificates for the other known sociodemographic and behavioral risk factors for low birth weights, including the proportion of mothers with less than 12th grade education (i.e., high school), the proportion of mothers smoking during pregnancy, and the proportion of mothers receiving inadequate prenatal care based on the Kotelchuck index. See Kirby *et al.* [46] for details of the choice of the covariates. We calculated the correlation for each pair of the covariates. The range of the correlations is between 0.01 and 0.46, indicating that the covariates have reasonably low correlation. We diagnosed the multicollinearity by calculating the variance inflation factor for each covariate. The range of the variance inflation factor is between 1.15 and 3.68, implying low multicollinearity.

In the data, y_{it} denotes the number of low birth weights in county i during year t, and $\mathbf{x}_{it} = (1, x_{it2}, x_{it3}, x_{it4}, x_{it5}, x_{it6}, x_{it7}, x_{it8})'$ with x_{it2} indicating the county-level population density, x_{it3} the proportion of Black people, x_{it4} the median household income, x_{it5} the unemployment rate, x_{it6} the proportion of mothers with less than 12th grade education, x_{it7} the proportion of mothers smoking during pregnancy, and x_{it8} the proportion of mothers with a Kotelchuck index in county i for year t, for $i = 1, \ldots, 46$ and $t = 1, \ldots, 10$.

We completed the specification of the proposed model by choosing prior Gamma(0.005, 0.005) for $\tau_{\alpha,k}$, $\tau_{\beta,k}$, $\tau_{\gamma,k}$ and τ_k. The prior probability for selection of regression coefficients is chosen to follow Beta(1,1). Because the number of regions is 46, we chose the truncation of the stick-breaking representation as 15 [29]. We also chose larger values that gave similar results. We collected 10,000 samples by thinning 50,000 samples by a factor of 5 after a burn-in of 10,000 iterations.

The third column in Table I shows the estimated negative cross-validatory predictive log-likelihoods for the proposed model along with the four competing models (as outlined in Section 4). We can see that the estimated $NLLK_{\text{CV,app}}$ values for models 1–4 are much higher than that for model 5, evincing that model 5 is the best among all the models. The priors of the parameters and the settings for the hyperparameters used were similar to those in the models of the simulated example.

Table III elucidates the marginal posterior probabilities of inclusion of the seven predictors in terms of fixed effects, spatial random effects, and temporal effects. To detect if the fixed effects and space–time variations are significant, we calculated the Bayes factor on the basis of the marginal posterior probabilities of indicators (i.e., δ_{1k}, δ_{2k}, and δ_{3k}, $k = 2, \ldots, 8$). More precisely, we can calculate the Bayes factor as

$$BF = \frac{\Pr(\delta_{jk} = 1|\mathbf{x}, \mathbf{y}) / \Pr(\delta_{jk} = 1)}{\Pr(\delta_{jk} = 0|\mathbf{x}, \mathbf{y}) / \Pr(\delta_{jk} = 0)}, \quad j = 1, 2, 3,$$

where $\Pr(\delta_{jk} = 1)$ denotes the prior probability of inclusion and $\Pr(\delta_{jk} = 1|\mathbf{x}, \mathbf{y})$ denotes the marginal posterior probability of inclusion. Because we assume the prior probabilities of inclusion and exclusion to be equivalent (i.e., 0.5), the Bayes factor reduces to $\Pr(\delta_{jk} = 1|\mathbf{x}, \mathbf{y}) / \Pr(\delta_{jk} = 0|\mathbf{x}, \mathbf{y})$. We noticed

Table III. Marginal posterior probabilities of inclusion of the seven predictors in the application.

Predictor	Fixed effect	Spatial effect	Temporal effect
Population density	0.91	0.82	0.63
Proportion of Black people	0.97	0.93	0.89
Median household income	0.96	0.91	0.90
Unemployment rate	0.88	0.82	0.66
Proportion of less education	0.85	0.74	0.70
Proportion of smoking	0.92	0.90	0.78
Proportion of Kotelchuck index	0.89	0.84	0.69

Table IV. Estimates and 95% CIs of fixed effects from the posterior samples.

Predictor	Fixed effect	95% CI
Population density	−0.82	(−1.76, 0.16)
Proportion of Black people	1.71	(0.30, 3.06)
Median household income	−1.00	(−1.92, 0.02)
Unemployment rate	0.86	(−0.43, 2.23)
Proportion of less education	0.86	(−0.47, 2.11)
Proportion of smoking	1.13	(−0.21, 2.56)
Proportion of Kotelchuck index	0.38	(−0.87, 1.54)

that when $\Pr(\delta_{jk} = 1|\mathbf{x}, \mathbf{y}) \geqslant 0.94$, the Bayes factor is over 15. From Jeffrey's Bayes factor criteria [46, p. 432], there is a very strong evidence of the effects, with the posterior probability of inclusion being over 0.94. For the fixed effects, the proportion of Black people and the median household income are significantly included in the model with the posterior probability of inclusion over 94%. For the spatial random effects, we concluded that the spatial variation of the effect for the proportion of Black people is marginally very strong (93%), indicating that the effect of the covariate on low birth weights significantly varies across the counties. On the other hand, there is no strong temporal variation of the effect for any covariates, implying little variation of the covariate effects over time.

Table IV provides the estimates and 95% CIs of the fixed effects for the seven predictors based on the posterior samples. It is clear that the African American women significantly have lower-birth-weight babies than the other race group. Besides the proportion of Black people, we noticed that the median household income is marginally negatively associated with the probability of low birth weight (95% CI is (−1.92, 0.02)). This is basically consistent with its posterior probability of inclusion, which is 0.96. For the population density and the proportion of smoking, as their posterior probabilities of inclusion are 0.91 and 0.92, respectively, we anticipated that their 95% CIs cover 0, implying that these two covariates have less impact on explaining the low birth weights. The remaining predictors are not significant in predicting the low birth weight. These results are consistent with previous studies [47, 48].

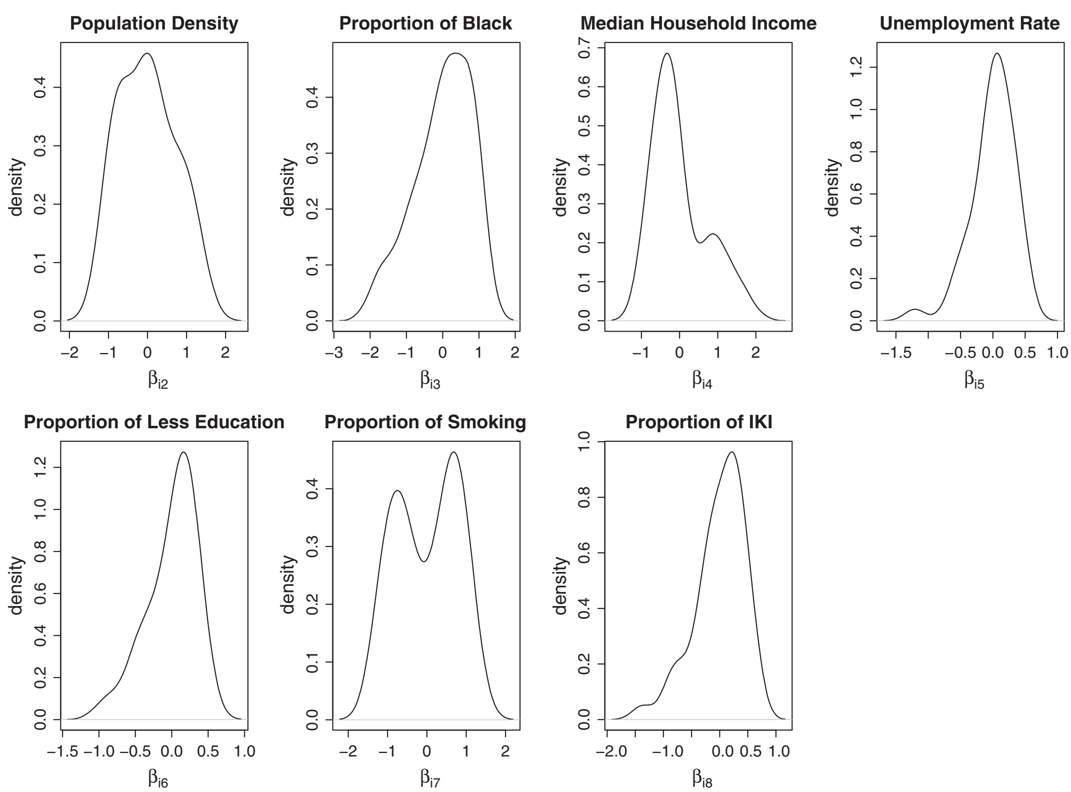

Figure 2. Posterior densities for spatially varying coefficients ($\boldsymbol{\beta}_i$) across areas in the application. IKI, Kotelchuck index.

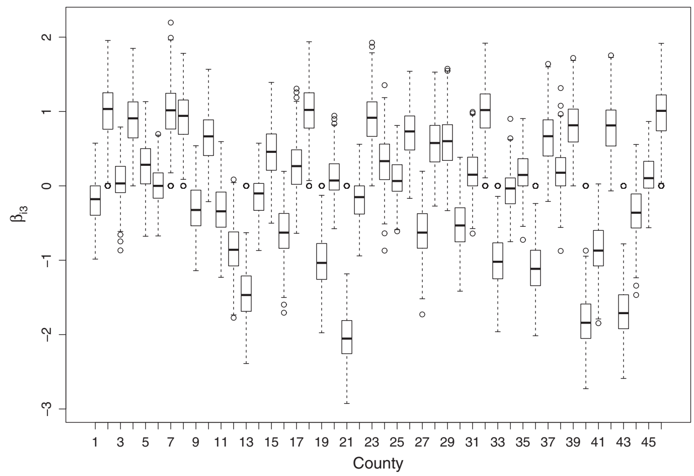

Figure 3. The box plots of the posterior samples of the spatially varying coefficient of the proportion of Black people over county.

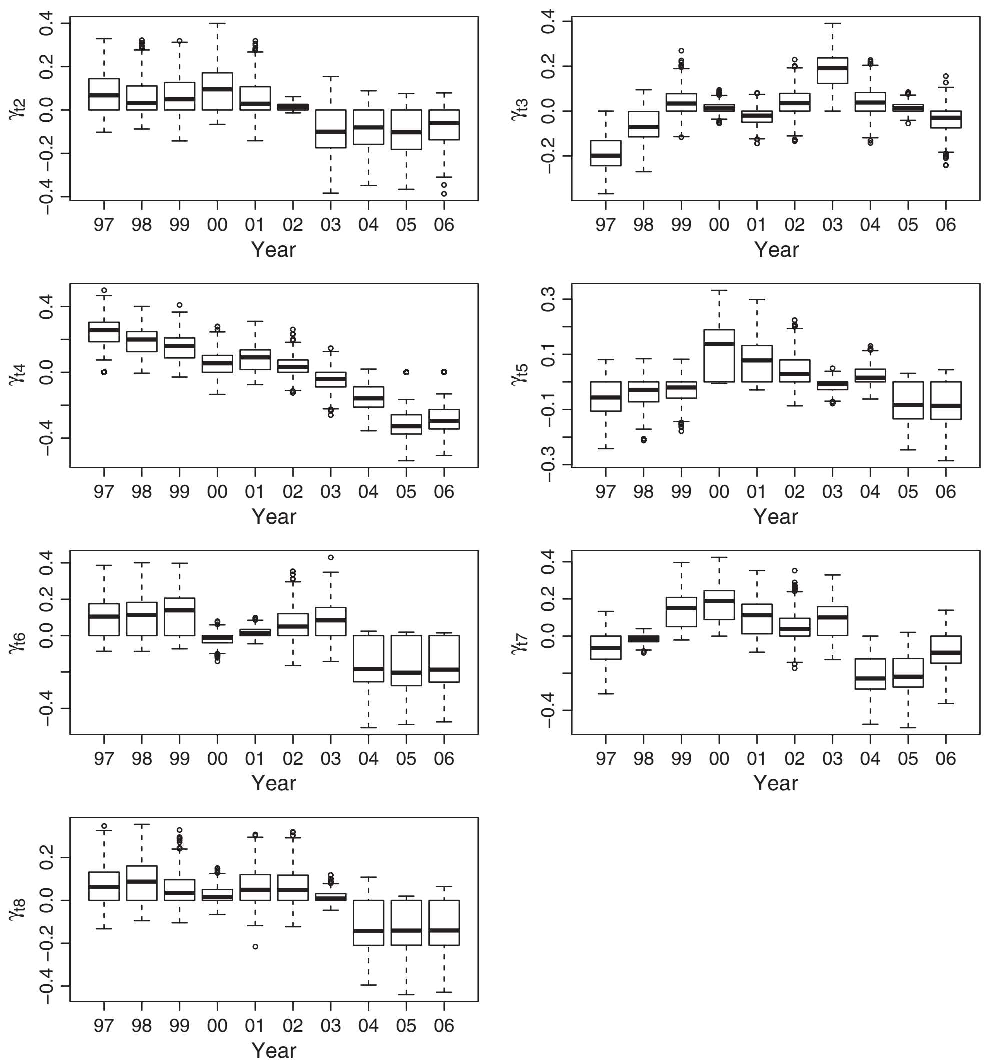

Figure 4. The box plots of the posterior samples of the temporally varying coefficients of the seven covariates over time.

Statistics
in Medicine

Figure 2 depicts the posterior densities for the spatially specific effects of the predictors. It shows that the spatial random effects of different predictors follow different distributions. Figure 3 exemplifies box plots of the posterior samples of the spatially varying coefficient of the proportion of Black people over county, where the points at zero are shown because of variable selection. Figure 4 presents box plots of the posterior samples of the temporally varying coefficients of all covariates over time. We note that although there are some trend variations, the temporally varying coefficients vary in a small scale and almost all of the ranges cover zero, implying an insignificant temporal effect for the covariates. Interestingly, even in the cases where the model coefficients do not present significant variation in time and space, the proposed approach still provides better fitting to the data in this application based on the $NLLK_{\text{CV, app}}$ in Table I. This advantage benefits from the area-specific nonparametric distribution assumption for the spatially varying coefficients along with variable selection, which allows for positive probabilities of excluding coefficients (i.e., zero values).

Figure 5 displays the choropleth maps of comparison of the raw standard mortality ratio of the low birth weight and the estimated standard mortality ratio based on the proposed model in years 1997, 2002, and 2006. We showed that the estimated relative risks of the low birth weight based on the proposed model capture the main geographical pattern and the temporal trend. In general, the relative risk of the low birth weights in many counties of South Carolina was increasing during 1997–2006. More precisely, initially high relative risks were mostly in the center and the east of the region. Then, relative risks were getting worse in the east, whereas those in the center remained in the same interval, and those in the north and the southwest increased remarkably.

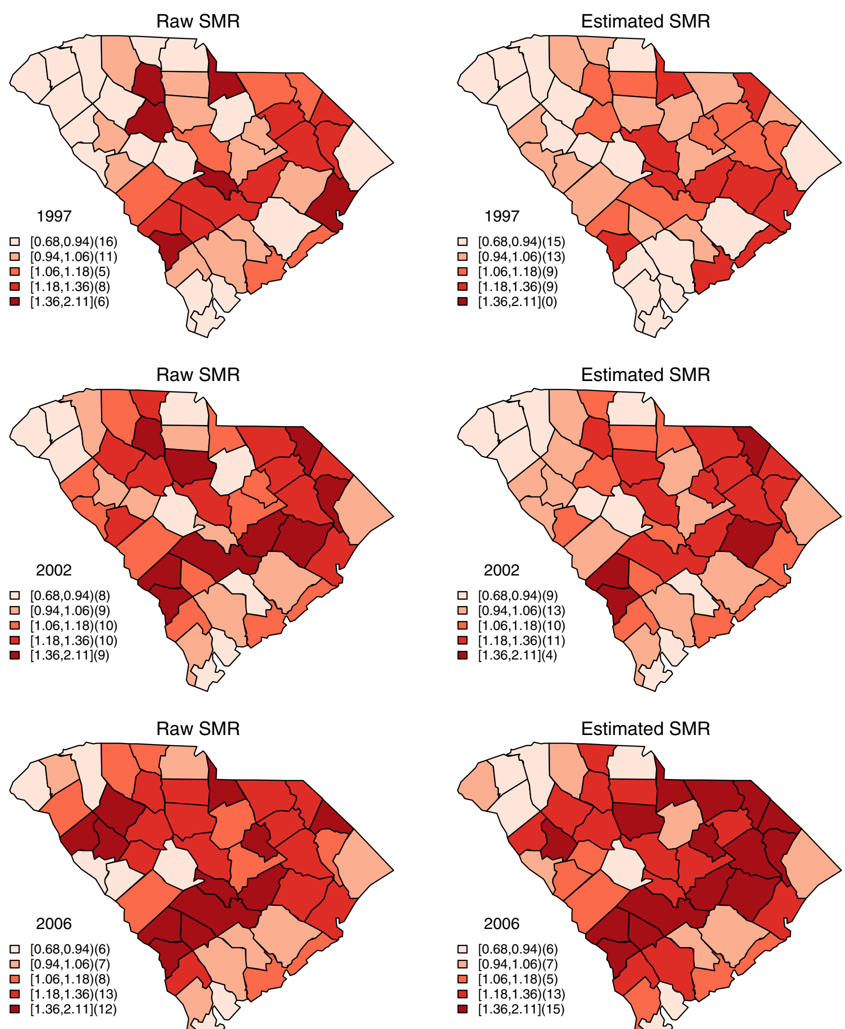

Figure 5. The choropleth maps of comparison of the raw standard mortality ratio (SMR) of low birth weights in South Carolina and the estimated SMR based on the proposed method in years 1997, 2002, and 2006.

6. Discussion

We proposed a Bayesian semiparametric model with variable selection for the analysis of space–time data. The proposed approach relaxes the normality assumption for the spatial random effects of covariates while allowing for uncertainty of the inclusion of the fixed effects, spatial random effects, and temporal effects. The spatial information is incorporated into the nonparametric distributions for the spatial random effects via the generalized transformation, which includes various popular links.

One of the major advantages of the proposed semiparametric model is the ability to flexibly model variation within localized areas of a study region. In the proposed model, we allow a geographically localized definition of the dependence of covariates and provide a flexible method of incorporation of covariates with predefined inclusion probabilities. Although in the examples, the covariate profiles show some impact on the overall county rates, it is evident that the estimated negative cross-validatory predictive log-likelihoods support the proposed model over conventional space–time random-effect models. This suggests that even with the degree of parameterization, there is an overall benefit in the use of such semiparametric models, especially when covariates are to be flexibly accommodated. We noticed computational intensity in the proposed approach, although it is reasonably efficient when it is coded in R. It took about 84 h to run 50,000 iterations for the real data example on a Linux server with Xeon(R) CPU X5355 at 2.66 GHz. Future work will focus on developing more efficient semiparametric space–time models for areal data.

Appendix A. Full conditional distributions in Section 2.3

Step 1: Update $(\delta_{1k}, \delta_{2k}, \delta_{3k})$, for $k = 1, \dots, p$, from its full conditional posterior distribution,

$$\exp\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{itk}y_{it}\theta_{itk} - E_{it}\exp\left(\sum_{k=1}^{p}x_{itk}\theta_{itk}\right)\right)\right]\pi(\delta_{1k}, \delta_{2k}, \delta_{3k}).$$

With this posterior distribution, we can calculate the posterior probability for each scenario in (6). After standardization, we can generate a sample of $(\delta_{1k}, \delta_{2k}, \delta_{3k})$ from a discrete probability measure as

$$(\delta_{1k}, \delta_{2k}, \delta_{3k}) \sim \sum_{l=1}^{5} p^*_{lk}\delta_{\kappa_{lk}}(\cdot),$$

where p^*_{lk} denotes the standardized probability and κ_{lk} denotes different scenarios for $(\delta_{1k}, \delta_{2k}, \delta_{3k})$ shown in (6).

Step 2: Update α_k, for $k = 1, \dots, p$, from its full conditional posterior distribution,

$$\exp\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{itk}y_{it}\delta_{1k}\alpha_k - E_{it}\exp\left(\sum_{k=1}^{p}x_{itk}\theta_{itk}\right)\right) - \frac{\tau_{\alpha,k}}{2}\alpha_k^2\right].$$

Step 3: Update $\tau_{\alpha,k}$, for $k = 1, \dots, p$, from its full conditional posterior distribution,

$$\text{Gamma}\left(a_{\alpha,k} + \frac{1}{2}, b_{\alpha,k} + \frac{\alpha_k^2}{2}\right).$$

Step 4: Update γ_{tk}, for $t = 1, \dots, T$ and $k = 1, \dots, p$, from its full conditional posterior distribution,

$$\exp\left[\sum_{i=1}^{n}\sum_{t=1}^{T}\left(x_{itk}y_{it}\delta_{3k}\gamma_{tk} - E_{it}\exp\left(\sum_{k=1}^{p}x_{itk}\theta_{itk}\right)\right) - \frac{\tau_{\gamma,k}}{2}\sum_{t=1}^{T}(\gamma_{tk} - \gamma_{t-1,k})^2\right].$$

Step 5: Update $\tau_{\gamma,k}$, for $k = 1, \dots, p$, from its full conditional posterior distribution,

$$\text{Gamma}\left(a_{\gamma,k} + \frac{T}{2}, b_{\alpha,k} + \frac{1}{2}\sum_{t=1}^{T}(\gamma_{tk} - \gamma_{t-1,k})^2\right).$$

Step 6: Update β_{ik}, for $i = 1, \ldots, n$ and $k = 1, \ldots, p$. Let $\mathbf{S}_k^* = \left(S_{1k}^*, \ldots, S_{r_k k}^*\right)'$ be the configuration of the current distinct values of $\mathbf{S}_k$. Then, we sample $\beta^*_{S^*_{sk}}$, for $s = 1, \ldots, r_k$, from its full conditional posterior distribution,

$$\exp\left[\sum_{i:S_{ik}=S^*_{sk}} \sum_{t=1}^{T} \left(x_{itk} y_{it} \delta_{2k} \beta^*_{S^*_{sk}} - E_{it} \exp\left(\sum_{k=1}^{p} x_{itk}\theta_{itk}\right)\right) - \frac{\tau_{\beta,k}\beta^{*2}_{S^*_{sk}}}{2}\right].$$

Step 7: Update $\tau_{\beta,k}$, for $k = 1, \ldots, p$, from its full conditional posterior distribution,

$$\text{Gamma}\left(a_{\beta,k} + \frac{n}{2}, b_{\beta,k} + \frac{1}{2}\sum_{i=1}^{n} \beta_{ik}^2\right).$$

Step 8: Update S_{ik}, for $i = 1, \ldots, n$ and $k = 1, \ldots, p$, from its full conditional posterior distribution,

$$\sum_{s=1}^{r_k} \hat{\omega}_{isk} \delta_s(\cdot),$$

where $\hat{\omega}_{isk} \propto \omega_{isk} \exp\left[\sum_{t=1}^{T}\left(x_{itk} y_{it} \delta_{2k} \beta^*_{S^*_{sk}} - E_{it} \exp\left(\sum_{k=1}^{p} x_{itk}\theta_{itk}\right)\right)\right]$, for $s = 1, \ldots, r_k$.

Step 9: Update ω_{isk}, for $s = 1, \ldots, r_k$ and $k = 1, \ldots, p$, as

$$\omega_{isk} = \hat{V}^*_{isk} \prod_{l<s} \left(1 - \hat{V}^*_{ilk}\right),$$

where $\hat{V}^*_{isk} = u_{isk} \hat{V}_{sk}$, $u_{isk} = h(\phi_{isk})$, $\hat{V}_{sk} \sim \text{Beta}(1 + L_{sk}, M_k + \sum_{l=s+1}^{r_k} L_{lk})$, with $L_{sk} = \sharp\{S_{ik} = s,\ \text{for } i = 1, \ldots, n\}$, denoting the number of S_{ik} values that equal s.

Step 10: Update M_k, for $k = 1, \ldots, p$, from its full conditional posterior distribution,

$$\text{Gamma}\left(r_k, -\sum_{l=1}^{r_k-1} \log(1 - \hat{V}_{lk})\right).$$

Step 11: Update ϕ_{isk}, for $k = 1, \ldots, p$, from its full conditional posterior distribution,

$$\sum_{s=1}^{r_k} \hat{V}^*_{isk} \prod_{l<s}\left(1 - \hat{V}^*_{ilk}\right) \delta_{\beta^*_{is}} \exp\left(-\frac{m_i \tau_k}{2}(\phi_{isk} - \overline{\phi}_{isk})^2\right),$$

where $\hat{V}^*_{isk} = u_{isk}\hat{V}_{sk}$ with $u_{isk} = (1 + \phi_{isk}\lambda_k/2)^{1/\lambda_k}/((1 + \phi_{isk}\lambda_k/2)^{1/\lambda_k} + (1 - \phi_{isk}\lambda_k/2)^{1/\lambda_k})$.

Step 12: Update τ_k, for $k = 1, \ldots, p$, from its full conditional posterior distribution,

$$\text{Gamma}\left(a_k + \frac{n r_k}{2}, b_k + \frac{1}{2}\sum_{s=1}^{r_k}\sum_{i\sim j}(\phi_{isk} - \phi_{jsk})^2\right).$$

Step 13: Update λ_k, for $k = 1, \ldots, p$, from its full conditional posterior distribution,

$$\prod_{i=1}^{n}\left(\sum_{s=1}^{r_k} \hat{V}^*_{isk} \prod_{l<s}\left(1 - \hat{V}^*_{ilk}\right)\delta_{\beta^*_{is}}\right).$$

Acknowledgements

The authors would like to thank the editor, the associate editor, and the three referees for valuable comments that greatly improved the presentation of the article. This work was supported by NIH/NHLBI 5R21HL088654-02.

References

1. Bernardinelli L, Montomoli C. Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Statistics in Medicine* 1992; **11**:983–1007.
2. Besag J. Spatial interaction and statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B* 1974; **36**:192–236.
3. Waller LA, Carlin BP, Xia H, Gelfand AE. Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association* 1997; **92**:607–617.
4. Knorr-Held L, Besag J. Modelling risk from a disease in time and space. *Statistics in Medicine* 1998; **17**:2045–2060.
5. Lagazio C, Biggeri A, Dreassi E. Age–period–cohort models and disease mapping. *Environmetrics* 2003; **14**:475–490.
6. Knorr-Held L. Bayesian modelling of inseparable space–time variation in disease risk. *Statistics in Medicine* 2000; **19**:2555–2567.
7. Lagazio C, Dreassi E, Biggeri A. A hierarchical Bayesian model for space time variation of disease risk. *Statistical Modelling* 2001; **1**:17–29.
8. Richardson S, Abellán JJ, Best N. Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire. *Statistical Methods in Medical Research* 2006; **15**:385–407.
9. Ugarte MD, Goicoa T, Ibáñez B, Militino AF. Evaluating the performance of spatio-temporal Bayesian models in disease mapping. *Environmetrics* 2009; **20**:647–665.
10. Assunção RM. Space varying coefficient models for small area data. *Environmetrics* 2003; **14**:453–473.
11. Gamerman D, Moreira ARB, Rue H. Space-varying regression models: specifications and simulation. *Computational Statistics and Data Analysis* 2003; **42**:513–533.
12. Dreassi E, Biggeri A, Catelan D. Space–time models with time dependent covariates for the analysis of the temporal lag between socio-economic factors and mortality. *Statistics in Medicine* 2005; **24**:1919–1932.
13. Cai B, Lawson AB, Hossain Md, Choi J. Bayesian latent structure models with space–time dependent covariates. *Statistical Modeling* 2012; **12**:145–164.
14. Gelfand A, Banerjee S, Gamerman D. Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics* 2005; **16**:465–479.
15. Paez MS, Gamerman D, Landim FMP, Salazar E. Spatially varying dynamic coefficient models. *Journal of Statistical Planning and Inference* 2008; **138**(4):1038–1058.
16. Gelfand AE, Kottas A, MacEachern SN. Bayesian nonparametric spatial modeling with Dirichlet process mixing. *Journal of the American Statistical Association* 2005; **100**:1021–1035.
17. Duan J, Guindani M, Gelfand A. Generalized spatial Dirichlet process models. *Biometrika* 2007; **94**:809–825.
18. Reich BJ, Fuentes M. A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *Annals of Applied Statistics* 2007; **1**:249–264.
19. Li P, Banerjee S, Hanson TA, McBean AM. Nonparametric hierarchical modeling for detecting boundaries in areally referenced spatial datasets. *Research Report*, Division of Biostatistics, University of Minnesota, 2010.
20. Reich BJ, Fuentes M, Herring AH, Evenson KR. Bayesian variable selection for multivariate spatially-varying coefficient regression. *Biometrics* 2010; **66**:772–782.
21. Kuo L, Mallick B. Variable selection for regression models. *Sankhyá B* 1998; **60**:65–81.
22. Geisser S. On prior distribution for binary trials (with discussion). *American Statistician* 1984; **38**(4):244–251.
23. Scott J, Berger J. An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference* 2006; **136**:2144–2162.
24. Cui W, George EI. Empirical Bayes vs. fully Bayes variable selection. *Journal of Statistical Planning and Inference* 2008; **138**:888–900.
25. Ghosal S, van der Vaart AW. *Fundamentals of Nonparametric Bayesian Inference*. Cambridge University Press: Cambridge, (expected in 2013).
26. Sethuraman J. A constructive definition of Dirichlet priors. *Statistica Sinica* 1994; **4**:639–650.
27. Blackwell D, Macqueen JB. Ferguson distributions via Pólya urn schemes. *Annals of Statistics* 1973; **1**:353–355.
28. Bush CA, MacEachern SN. A semiparametric Bayesian model for randomised block designs. *Biometrika* 1996; **83**:275–285.
29. Ishwaran H, James LF. Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association* 2001; **96**:161–173.
30. Ishwaran H, Zarepour M. Markov Chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika* 2000; **87**:371–390.
31. Aranda-Ordaz F. On two families of transformations to additivity for binary response data. *Biometrika* 1981; **68**: 357–363.
32. Ohlssen DI, Sharples LD, Spiegelhalter DJ. Flexible random-effects models using Bayesian semi-parametric models: application to institutional comparisons. *Statistics in Medicine* 2007; **26**:2088–2112.
33. Gilks WR, Best NG, Tan KKC. Adaptive rejection metropolis sampling within Gibbs sampling. *Journal of Applied Statistics* 1995; **44**:455–472.
34. Spiegelhalter DJ, Best NG, Carlin BP, Linde AVD. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 2002; **64**:1–34.
35. Plummer M. Penalized loss functions for Bayesian model comparison. *Biostatistics* 2008; **9**:523–539.
36. Geisser S. *Predictive Inference: An Introduction*. Chapman & Hall: London, 1993.
37. Gelfand AE, Dey D, Chang H. Model determination using predictive distributions with implementation via sampling based methods (with discussion). In *Bayesian Statistics 4*, Bernardo J, *et al.* (eds). Oxford University Press: Oxford, 1992; 147–167.
38. Dey D, Chen MH, Chang H. Bayesian approach for nonlinear random effects models. *Biometrics* 1997; **53**:1239–1252.

39. Sinha D, Dey DK. Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association* 1997; **92**:1195–1212.
40. Draper D, Krnjajić M. Bayesian model specification. *Technical Report*, Department of Applied Mathematics and Statistics, Baskin School of Engineering, University of California, Santa Cruz, 2006.
41. Cowles MK, Carlin BP. Markov Chain Monte Carlo diagnostics: a comparative review. *Journal of the American Statistical Association* 1995; **91**:883–904.
42. Plummer M, Best NG, Cowles K, Vines K. CODA: Convergence diagnosis and output analysis for MCMC. *R News* 2006; **6**:7–11.
43. R Development Core Team. R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2007. ISBN 3-900051-07-0, http://www.R-project.org.
44. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; **10**:325–337.
45. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 2006; **1**:515–534.
46. Jeffreys H. *The Theory of Probability*, 3rd edn. Oxford University Press: Oxford, 1961.
47. Kirby R, Liu J, Lawson A, Choi J, Cai B, Hossain M. Small area low birth weight incidence and socio-economic predictors: a latent spatial structure approach. *Spatial and Spatio-temporal Epidemiology* 2011; **2**(4):265–271.
48. Olson ME, Diekema D, Elliott BA, Renier CM. Impact of income and income inequality on infant health outcomes in the United States. *Pediatrics* 2010; **126**:1165–1173.