Statistical Modelling 2012; **12**(**2**): 145–164

Bayesian latent structure models with space-time dependent covariates

Bo Cai[1], Andrew B Lawson[2], Md Monir Hossain[3] and Jungsoon Choi[2]
[1]Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC, USA
[2]Division of Biostatistics and Epidemiology, Medical University of South Carolina, Charleston, SC, USA
[3]Center for Clinical and Translational Sciences, The University of Texas, Houston, TX, USA

Abstract: Spatial-temporal data requires flexible regression models which can model the dependence of responses on space- and time-dependent covariates. In this paper, we describe a semiparametric space-time model from a Bayesian perspective. Nonlinear time dependence of covariates and the interactions among the covariates are constructed by local linear and piecewise linear models, allowing for more flexible orientation and position of the covariate plane by using time-varying basis functions. Space-varying covariate linkage coefficients are also incorporated to allow for the variation of space structures across the geographical location. The formulation accommodates uncertainty in the number and locations of the piecewise basis functions to characterize the global effects, spatially structured and unstructured random effects in relation to covariates. The proposed approach relies on variable selection-type mixture priors for uncertainty in the number and locations of basis functions and in the space-varying linkage coefficients. A simulation example is presented to evaluate the performance of the proposed approach with the competing models. A real data example is used for illustration.

Key words: Bayesian regression; latent structure model; piecewise linear splines; space-time models; variable selection

Received February 2010; **revised** June 2010; **accepted** January 2011

1 Introduction

There is an increasing attention in the analysis of spatially and temporally referenced data in both methodological and applied research. Such data are of substantial interest in a variety of disciplines such as epidemiology, ecology, political sciences and economics. For example, one might be interested in geographical patterns and trends of a certain disease in a particular region over time. We start with describing a general space-time model.

Address for correspondence: Bo Cai, Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC, USA. E-mail: bcai@sc.edu

 10.1177/1471082X1001200202

Suppose that the dependent variable y_{it} is observed in the ith spatial unit (e.g., region or individual) and the tth time point with $i = 1, \ldots, n$ and $t = 1, \ldots, T$. A general space-time model can be expressed as

$$y_{it} \sim f(y_{it}|\cdot), \tag{1.1}$$

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., indicates of disease or health status) and count (e.g., disease or death number) outcome. The predictor η_{it} is usually expressed as

$$\eta_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + u_i + v_i + \delta_t, \tag{1.2}$$

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 δ_t denotes a structured temporal random component. Conventionally, the fixed effects $\boldsymbol{\beta}$ can be modelled to follow a multivariate normal prior. The parameters u_i and v_i are assumed to be independent. The parameter v_i captures the heterogeneity among the units which is chosen to follow an exchangeable normally distributed prior, while u_i captures the clustering property of spatial data which is assumed to follow a conditional autoregressive (CAR) distribution (a special case of the general class of Markov random field) (Besag, 1974), $u_i|u_{-i} \sim \mathrm{N}(\overline{u}_i, (\tau m_i)^{-1})$, 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 neighbour set of unit i, m_i denotes the number of neighbours of unit i and τ denotes the precision parameter. The constraint $\sum_{i=1}^{n} u_i = 0$ is defined for the purpose of identifiability of the overall intercept. The temporal parameter δ_t is assumed to follow AR(1) prior.

When responses are count data, model (1) becomes a typical spatio-temporal model based on which some hierarchical models were developed (Waller *et al.*, 1997; Knorr-Held and Besag, 1998; Lagazio *et al.*, 2003; among others). More complex issues occur when the space-time interaction effect, θ_{it}, is included in the predictor (2). Knorr-Held (2000) incorporated a space-time interaction for inseparable space-time variation in disease risk where four types of space-time interaction were described. Richardson *et al.* (2006) proposed a joint spatio-temporal modelling of two diseases with shared space-time interaction. Lagazio *et al.* (2001) focussed on the birth cohort model to assess latent effects associated with temporal trends. Ugarte *et al.* (2009) evaluated the performance of various spatio-temporal Bayesian models. Hossain and Lawson (2010) also evaluated spatial-temporal (ST) small area models but with an emphasis on cluster recovery/detection. When the effect of covariates on the response is the main focus, some space-time models with space-dependent coefficients (e.g., Assunção, 2003; Gamerman *et al.*, 2003) or time-dependent coefficients (Dreassi *et al.*, 2005) were developed. In practice, the space-time dependent effect of a specific space-time covariate on responses is also of substantial interest (e.g., effects of poverty rates on low birth weight may vary across different regions and time points). Although some work has been done (Gelfand *et al.*, 2005; Paez *et al.*, 2008), there is still lack of development for such models. Typically, one may consider a space-time model based on linear combinations of covariates such as $\eta_{it} = \mathbf{x}'_{it}\boldsymbol{\beta}_{it} + u_i + v_i + \delta_t$. This expression can be treated as a general case including the model with a space-time interaction term, the model with space-dependent covariates (Assunção, 2003) and the model with time-dependent covariates (Dreassi *et al.*, 2005). However, linear models prove too rigid when large quantities of data are considered and there exists nonlinearity (Hastie and Tibshirani, 1990).

Outside the context of the space-time data analysis, there exists a fairly rich literature of nonlinear regression modelling in both the frequentist and Bayesian framework. For example, Friedman (1991) proposed multivariate adaptive regression splines by using flexible tensor product splines. Holmes and Mallick (2001, 2003) described a Bayesian approach of piecewise linear model with covariate surface constructed by basis functions. Bigelow and Dunson (2007) extended the method to allow the spline coefficients to be subject specific. Pintore *et al.* (2006) derived a spatially adaptive smoothing splines based on a reproducing kernel Hilbert space representation. Numerous references can be found in Denison *et al.* (2002) and Ruppert *et al.* (2003), and therein.

Within the context of spatial and temporal modelling, Schmid and Held (2004) investigated space-time trends by incorporating intercept terms of covariates of interest with additional spatial component into the model. Banerjee and Johnson (2006) proposed to model single and multi resolution spatially varying growth curves as Gaussian processes that capture associations at single and multiple resolutions. Kneib and Fahrmeir (2006) described a general class of structured additive regression models for categorical responses, allowing for a semiparametric predictor. Zhao *et al.* (2006) developed general design generalized linear mixed models in which random effects with spatial correlation structure are included. Among the methods developed, however, none of them simultaneously considers space- and time-specific effects of space-time covariates on responses.

In this paper, we focus on developing a general space-time model with main interest in the effect of space- and time-dependent covariates on the response. We extend the generalized multivariate regression splines (Holmes and Mallick, 2001) to flexibly accommodate the space- and time-specific covariates, allowing for flexible orientation and position of the covariate plane by using time-varying basis functions. Space-varying covariate linkage coefficients are incorporated for variation of space structures across geographical locations. Such multivariate regression models allow for effects of covariates on responses not only across space but also over time (i.e., interactions) in a flexible manner. We develop an approach which relies on variable selection-type mixture priors for uncertainty in the number and locations of the piecewise linear basis functions and in the space-varying linkage coefficients.

The remainder of the paper is organized as follows. Section 2 describes the space-time latent structure model with multivariate linear splines for a covariate linkage.

Prior specification and posterior implementation are described. Section 3 discusses the model evaluation and comparison. Section 4 evaluates the performance of the approach based on a simulated example. Section 5 illustrates the approach via a real spatial-temporal data. Finally, Section 6 summarizes and discusses the results.

2 Space-time models with latent structure

2.1 The model and prior specification

We consider to model the unknown linear predictor as

$$\eta_{it} = \sum_{k=1}^{K} \beta_{ik}(\mathbf{x}'_{it}\mathbf{u}_{tk})_+, \tag{2.1}$$

where $\boldsymbol{\beta}_i = (\beta_{i1}, \ldots, \beta_{iK})'$ denotes a $K \times 1$ vector of candidate space-specific linkage parameters for the underlying latent effects, $\mathbf{u}_{tk}$ denotes a $p \times 1$ vector of time-varying basis parameters and $(\mathbf{x}_{it}\mathbf{u}_{tk})_+$ denotes a basis function which is an inner product of $\mathbf{x}_{it}$ and $\mathbf{u}_{tk}$ truncated below by 0. To allow each model to include an intercept term, we define $(\mathbf{x}_{it}\mathbf{u}_{t1})_+$ to be one for all i and t. When $\boldsymbol{\beta}_i$ are typically subject-specific coefficients, this formula is a generalization of the formulae by Holmes and Mallick (2001) and Bigelow and Dunson (2007). In model (3), the linkage coefficients $\boldsymbol{\beta}_i$ have spatially structured effects on covariates. To avoid identifiability problem, we constrain $\sum_{i=1}^{n} \beta_{ik} = C_k$, where C_k is some constant (Assunção *et al.*, 2002). One may typically decompose β_{ik} as $\alpha_k + \xi_{ik}$, where α_k denotes the global effect and ξ_{ik} denotes the spatially structured effect with $\sum_{i=1}^{n} \xi_{ik} = 0$ to identify the overall effect due to the location invariance of the CAR prior. An attractive property of the structure of equation (2.1) is that the basis functions are time dependent which provide flexible orientation and position of the covariate plane and important trends in the impact of covariates over time. It is clear that each of the $(K-1)$ non-intercept basis functions contains linear effect for at least one covariate. When a basis contains multiple covariate effects, the proposed model allows for the effects of interactions (i.e., dependence) of space-time covariates on the response. The proposed model can be thought of as a more general spatio-temporal model. For example, if the observations are only spatially dependent, then basis functions are time irrelevant. In this case, the model reduces to the space-varying regression model (Assunção, 2003; Gamerman *et al.*, 2003) with time-dependent covariates, i.e., $\eta_{it} = \mathbf{x}'_{it}\boldsymbol{\theta}_i$.

The proposed approach allows for a flexible number of unknown basis functions and the linkage coefficients. Since the number of basis functions related to covariates is unknown *a priori*, one may consider the reversible jump MCMC (Markov chain Monte Carlo) (Green, 1995) for such models (e.g., Holmes and Mallick, 2001; Bigelow and Dunson, 2007). However, it involves complicated marginal likelihood calculation or approximation. To avoid this complexity, we adopt variable selection-type mixture priors for uncertainty of the number and locations of piecewise basis functions. To allow the kth basis to be effectively excluded from the model, we choose a mixture prior including a point mass at zero and a CAR distribution for the linkage coefficient β_{ik} given the indicator γ_k:

$$\beta_{ik}|\gamma_k \sim \gamma_k \delta_0(\beta_{ik}) + (1-\gamma_k)\mathrm{N}(\overline{\beta}_{\partial_i,k}, (\tau_k m_i)^{-1}), \tag{2.2}$$

where γ_k is an indicator variable which is 1 for exclusion or 0 for inclusion of the kth basis function, $\delta_0(\cdot)$ denotes a point mass at zero, $\overline{\beta}_{\partial_i,k} = m_i^{-1}\sum_{j\in\partial_i}\beta_{jk}$ with ∂_i denoting the neighbour set of unit i, m_i denotes the number of neighbours of unit i and τ_k denotes the precision parameter. We refer to prior (4) as a zero-inflated CAR prior, ZI-CAR(γ_k, τ_k). The prior probability of the kth basis out of the K candidate bases related to covariates being excluded is $p_{1,k0} = \Pr(H_{1,k0} : \beta_{ik} = 0)$. The prior for γ_k is then chosen as a Bernoulli distribution, Bern($p_{1,k0}$). The prior for τ_k is chosen as Gamma(a_τ, b_τ), where Gamma(a, b) is a gamma distribution with mean a/b and variance a/b^2.

To reflect time-dependent measurements in each region, we use multivariate dynamic normal priors for $\mathbf{u}_{tk,-1}$ which can be written as

$$\mathbf{u}_{tk,-1} \sim \mathrm{N}_{p-1}(\rho_k \mathbf{u}_{t-1,k,-1}, \boldsymbol{\nu}), \qquad t = 1, \ldots, T,$$

where ρ_k denotes the variation of the temporal autocorrelation in the risk, $\mathbf{u}_{0k,-1}$ denotes the starting vector of $\mathbf{u}_{tk,-1}$ and $\boldsymbol{\nu}$ denotes a diagonal covariance matrix, $diag(\nu_2, \ldots, \nu_p)$. Due to lack of unique solutions of β_{ik} and $\mathbf{u}_{tk}$ to the same model for each k, following Holmes and Mallick (2001), we normalize each component of $\mathbf{u}_{tk,-1} = \mathbf{u}_{tk}/u_{tk,1} = (u_{tk,2}, \ldots, u_{tk,p})'$, i.e., $||\mathbf{u}_{tk,-1}|| = 1$ for $t = 1, \ldots, T$ and $k = 1, \ldots, K$, so that $\mathbf{u}_{tk}/u_{tk,1}$ can be used for orientation of the plane in $(p-1)$-dimension covariate space and $u_{tk,1}$ for the position of the plane. To flexibly select the components from $p-1$ covariates at each time point, we first choose $\mathbf{u}_{0k,-1}$ to be zero. We then choose a variable selection-type mixture prior with a point mass at zero for variance ν_l, for $l = 2, \ldots, p$,

$$\nu_l \sim \kappa_l \delta_0(\nu_l) + (1-\kappa_l)\mathrm{IG}(\nu_l; a_\nu, b_\nu), \tag{2.3}$$

where κ_l is an indicator variable which is 1 for exclusion or 0 for inclusion of the lth covariate and IG($\cdot$) denotes an inverse gamma distribution. We refer to prior (5) as ZI-IG(κ_l, a_ν, b_ν). The prior for κ_l is chosen as Bern($p_{2,l0}$), where $p_{2,l0}$ denotes the probability of the lth covariate being excluded. The first element of $\mathbf{u}_{tk}$ can be defined as $u_{tk1} = -\mathbf{x}'_{i't,-1}\mathbf{u}_{tk,-1}$, where $\mathbf{x}_{i't,-1}$ is randomly selected with $i' \in \{1, \ldots, n\}$. With probability of $p_{2,l0}$, all u_{tkl} are zeroes, for $t = 1, \ldots, T$ and $k = 1, \ldots, K$, indicating that the lth covariate is excluded from the model. The mixture prior allows for the locations of the splines to vary over time by effectively excluding the elements from each basis function. The overall prior probability of excluding all covariates (except intercept) from the model at time t is $\prod_{l=2}^{p} p_{2,l0}$.

To allow for flexibility of the prior probability, $p_{1,k0}$, we consider choosing a hyper-prior beta distribution for the prior exclusion probability, $p_{1,k0} \sim \text{Beta}(c_1, d_1)$. Given the assumption that all prior probabilities are equal (say, p_0), the full conditional for $p_{1,k0}$ can be easily calculated (see details in Appendix). Similarly, the prior of $p_{2,l0}$ is chosen as a beta distribution, $\text{Beta}(c_2, d_2)$, allowing for more flexibility in adapting the desired model. For the choice of c_i and d_i $(i = 1, 2)$, following the suggestion by Geisser (1984), we choose $c_i = d_i = 1$ which yields the uniform hyperprior. Scott and Berger (2006) discuss the choice of priors for the prior probability. They conclude that the objective prior (i.e., the uniform prior) for the prior probability can easily be implemented computationally while 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 (1984), Scott and Berger (2006, 2008) and Cui and George (2008), among others.

2.2 Posterior computation

The joint posterior distribution for the parameters is

$$\pi(\boldsymbol{\beta}, \mathbf{u}, \boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\gamma}, \boldsymbol{\kappa}|\mathbf{y}, \mathbf{x}) \propto \prod_{i=1}^{n}\prod_{t=1}^{T} f\left\{\sum_{k=1}^{K} \beta_{ik}(\mathbf{x}_{it}'\mathbf{u}_{tk})_+\right\}\pi(\boldsymbol{\beta}|\boldsymbol{\gamma})\pi(\boldsymbol{\gamma})\pi(\mathbf{u}|\boldsymbol{\nu}, \boldsymbol{\kappa})\pi(\boldsymbol{\nu}|\boldsymbol{\kappa})\pi(\boldsymbol{\tau}),$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_n)'$, $\mathbf{u} = (\mathbf{u}_1, \ldots, \mathbf{u}_T)'$ and $f(\cdot)$ can be normal linear, Poisson and logistic regression models for the continuous, count and binary outcomes, respectively, described as

$$\begin{aligned} f(y_{it}|\eta_{it}) &= \sqrt{\frac{\tau}{2\pi}} \exp\left\{-\frac{\tau}{2}\left(y_{it} - \eta_{it}\right)^2\right\} \qquad \text{(normal linear with } \eta_{it} = \mu_{it}) \\ &= \frac{1}{y_{it}!} \exp\left\{y_{it}\eta_{it} - \exp\left(\eta_{it}\right)\right\} \qquad \text{(Poisson with } \eta_{it} = \log \mu_{it}) \\ &= \frac{\exp(y_{it}\eta_{it})}{1 + \exp(\eta_{it})} \qquad \left(\text{logistic with } \eta_{it} = \log \frac{\mu_{it}}{1 - \mu_{it}}\right). \end{aligned}$$

We choose priors for the parameters as described in Section 2.1. The posterior computation relies on a stochastic search variable selection Gibbs sampling algorithm (George and McCulloch, 1993), in which we iteratively sample from the full conditional distributions of each of the parameters. For each element of $\boldsymbol{\beta}_i$ and $\boldsymbol{\nu}$, the posterior has a mixture structure with a point mass at zero and a conjugate (for normal linear) or non-conjugate (for Poisson and logistic) distribution. To sample from the non-conjugate distribution, we use adaptive rejection Metropolis sampling (Gilks *et al.*, 1995). Under the linear normal case, reparameterization allows the model to have conditionally linear structure for each parameter which facilitates the use of conjugate priors. For the purpose of generality, we instead provide a general full conditional posterior distribution for sampling.

The posterior computation relies on the Gibbs sampler and Metropolis-Hastings algorithms. After initializing values for the parameters, the proposed MCMC algorithm proceeds which is detailed in Appendix. Samples from the joint posterior distribution of the parameters are generated by repeating these steps for a large number of iterations after apparent convergence. Obviously, for identity link, the parameters can be sampled from the conjugate full conditional distributions.

As to guidance of how to specify the initial number for truncated planes for a particular analysis, from the simulation experiments that we conducted, we found that a large initial number of truncated planes, K, may provide sufficient space for change of dimension. However, after a minimum necessary number reaches, any further increase only marginally affects the fit while the computation time increases dramatically. Too low values of K, however, result in an inflexible modelling of the unknown linear predictor. Thus, we recommend to start with at least 10 number of truncated planes for small sample sizes and for large sample sizes of $n > 20$, referring to the heuristic rule of thumb given by Ruppert (2002) in the context of penalized splines of $K = \min(40, n/4)$.

3 Model comparison

The deviance information criterion (DIC)(Spiegelhalter *et al.*, 2002) 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 (Plummer, 2008). Plummer (2008) pointed out that in disease mapping, this assumption does not hold, resulting in that DIC under-penalizes the complex models. Plummer (2008) proposed penalized loss functions instead of p_D, the effective number of parameter, to assess model adequacy. However, as Plummer (2008) noticed, this method requires MCMC runs with each observation left out in turn. Such calculation is not feasible in general, especially for large datasets. In this paper, we consider the comparison method based on the conditional predictive ordinate (CPO) (Gelfand *et al.*, 1992; Geisser, 1993; Dey *et al.*, 1997; Sinha and Dey, 1997). The CPO for the ith observation at time t is defined as the cross-validated marginal posterior predictive density

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

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. The cross-validation likelihood can be estimated by

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

Since the quantity of the cross-validation likelihood is typically close to zero, the negative cross-validatory predictive log-likelihood (Spiegelhalter *et al.*, 1996; Draper and Krnjajić, 2006) can be used

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

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

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

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

4 A simulated example

The motivation of this simulation was to evaluate the performance of the proposed approach, including the accuracy of the estimates, the sensitivity to different choices of hyperparameters and comparison of the proposed model with other space-time models. Without loss of generality and for illustration purpose, we created the data based on Belgium map available from GeoBUGS in WinBUGS (Lunn *et al.*, 2000) containing 43 districts. We considered the case of count responses. The data were generated for each of $n = 43$ districts over an observation period of $T = 10$ based on the model $y_{it} \sim \mathrm{Poisson}(E_{it} \exp(\eta_{it}))$, where the log-relative risk $\eta_{it} = \mathbf{x}'_{it}\boldsymbol{\alpha} + \tilde{\mathbf{x}}'_{it}\boldsymbol{\xi}_i + v_i + \delta_t$, where $\mathbf{x}_{it} = \tilde{\mathbf{x}}_{it} = (1, x_{it2}, x_{it3}, x_{it4}, x_{it5})'$. E_{it} is an expected number of events obtained by Rn_{it}, where n_{it} is the population count in district i at time t and $R = \sum_{it} y_{it} / \sum_{it} n_{it}$. This model is similar to the one by Assunção (2003), where space-dependent covariates are included. The fixed effect $\boldsymbol{\alpha}$ was chosen as $(1, 1, 1, 0, 0)'$, implying that the last two covariates are irrelevant. We generated $\boldsymbol{\xi}_i$ from a multivariate CAR, MVCAR($\boldsymbol{\tau}$), where $\boldsymbol{\tau}^{-1}$ is a 3×3 covariance matrix with components along the row as $\{0.5, 0.2, 0.2, 0.2, 0.4, 0.2, 0.2, 0.2, 0.8\}$, $v_i \sim \mathrm{N}(0, 1)$ and $\delta_t \sim \mathrm{N}(\delta_{t-1}, 2)$ for $t = 2, \ldots, T$ with $\delta_1 \sim \mathrm{N}(0, 2)$. We generated $\mathbf{x}_{itl} \sim \mathrm{U}(0, 1)$ for $l = 2, \ldots, 5$. For practical reason, the expected count E_{it} was sampled from $\mathrm{U}(1, 5)$.

We specified the priors for the parameters of the proposed model as follows. We used Gamma(0.05, 0.05) as the prior for τ_k. The prior for the spatially structured random effects β_{ik} was chosen as the prior in (4) with $\mathrm{N}(\overline{\beta}_{\partial_{ik}}, (\tau_k m_{ik})^{-1})$. For the time-varying bases $\mathbf{u}_{tk,-1}$, the prior was chosen as $\mathrm{N}(\mathbf{u}_{t-1,k,-1}, \boldsymbol{\nu})$, where $\nu_l \sim$ZI-IG(κ_l, 0.05, 0.05) for $l = 2, \ldots, 5$. The starting vector of $\mathbf{u}_{tk,-1}$, $\mathbf{u}_{0k,-1}$, was chosen as $\mathbf{0}$ and ρ_k as 1. For a flexible hyperprior beta distribution of $p_{1,k0}$ and $p_{2,l0}$, we chose $c = d = 1$ which yields the uniform hyperprior. Following Holmes and Mallick (2001), we chose the initial number of truncated planes as $K = 30$. We also tried several larger initial numbers of truncated planes which yielded essentially identical results.

We implemented the analysis using the Gibbs sampling algorithm described in Section 2. We generated 50 000 iterations after a burn-in of 10 000 iterations. Convergence was assessed by using a variety of diagnostics described by Cowles and Carlin (1995) and implemented using CODA (Plummer *et al.*, 2006) in R. The diagnostic tests showed rapid convergence and efficient mixing. The parameters were estimated by thinning the chain by factor of 5 to obtain a sample of size 10 000. Sensitivity test of the results to the prior specification was assessed by repeating the analysis with different hyperparameters, which showed very similar results.

We compared the proposed model (Model 5) with the four competing spatio-temporal models. The log-relative risks of these models are listed as follows:

Model 1: $\eta_{it} = \mathbf{x}'_{it}\boldsymbol{\alpha} + \xi_i + \delta_t$,
Model 2: $\eta_{it} = \mathbf{x}'_{it}\boldsymbol{\alpha} + \xi_i + v_i + \delta_t$,
Model 3: $\eta_{it} = \mathbf{x}'_{it}\boldsymbol{\alpha} + \xi_i + v_i + \delta_t + b_{it}$ and
Model 4: $\eta_{it} = \mathbf{x}'_{it}\boldsymbol{\alpha} + \mathbf{x}'_{it}\boldsymbol{\xi}_i + v_i + \delta_t$

In the first three models, we followed conventional settings by specifying the prior of $\boldsymbol{\alpha}$ as $\mathrm{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}})$ with $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} \sim \mathrm{IWishart}(p, \boldsymbol{\Sigma}_0^{-1})$ and $\boldsymbol{\Sigma}_0 = \{0.1, 0.005, 0.005, 0.005, 0.1, 0.005, 0.005, 0.005, 0.1\}$. The prior of ξ_i was chosen as $\mathrm{N}(\overline{\xi}_i, (\tau_1 m_i)^{-1})$ with $\overline{\xi}_i = m_i^{-1}\sum_{j\in\partial_i}\xi_j$ and $\tau_1 \sim$ Gamma(0.005, 0.005). The prior of v_i was taken as $\mathrm{N}(0, \tau_2^{-1})$ with $\tau_2 \sim$ Gamma(0.005, 0.005). The prior of δ_t was chosen as $\mathrm{N}(\delta_{t-1}, \tau_3^{-1})$ with $\delta_0 \sim \mathrm{N}(0, \tau_3^{-1})$ and $\tau_3 \sim$ Gamma(0.005, 0.005). We chose the prior of b_{it} to be $\mathrm{N}(0, \tau_4^{-1})$ with $\tau_4 \sim$ Gamma(0.005, 0.005). For Model 4, the prior of $\boldsymbol{\xi}_i$ was chosen as $\mathrm{N}(\overline{\boldsymbol{\xi}}_{\partial_i}, (\boldsymbol{\Sigma}_{\boldsymbol{\xi}} m_i)^{-1})$ with $\overline{\boldsymbol{\xi}}_{\partial_i} = m_i^{-1}\sum_{j\in\partial_i}\boldsymbol{\xi}_j$ and $\boldsymbol{\Sigma}_{\boldsymbol{\xi}} \sim \mathrm{Wishart}(p, \boldsymbol{\Sigma}_0)$. We implemented Models 1–4 using WinBUGS (Lunn *et al.*, 2000). Although Model 5 can also be implemented by WinBUGS, it is computationally intensive. A C program was instead written to carry out the proposed algorithm.

Table 1 Model comparison based on the negative cross-validatory log-likelihood for the simulated example and the application to the low birth weight data in South Carolina

Model	$NLLK_{CV,\,sim}$	$NLLK_{CV,app}$
Model 1	1160.967	1820.849
Model 2	1161.409	1802.395
Model 3	767.532	1746.393
Model 4	637.534	1763.222
Model 5	634.297	1738.383

The second column in Table 1 presents the comparison of the estimated negative cross-validatory predictive log-likelihoods for the five models. We can see that Model 1 and Model 2 are basically the same. This is due to the fact that the unstructured random effects vary moderately across regions which is consistent with the setting. It is evident that Models 3–5 appear much better than the first two models with Model 5 being the best. Since Model 4 is the model where the data are generated, its performance is very close to Model 5.

In Figure 1, the upper plot represents true and pointwise estimated relative risks with 95% credible intervals from the proposed approach (Model 5) across all districts at time 5. The lower plot shows true and estimated relative risks for district 20 over time based on the five models along with 95% credible intervals from Model 5. It is clear that the proposed model provides closer estimates than the others.

Figure 2 shows posterior densities of variance parameters ν_4 and ν_5 and boxplots of the posterior means for the time-varying basis components u_4 and u_5 at each time point in the simulated example. We can see that the two variances ν_4 and ν_5 corresponding to the time-varying coefficients, u_{tk4} and u_{tk5}, are close to zero, implying that x_{it4} and x_{it5} are not involved. This is consistent with the simulation design.

Sensitivity of the results to the prior specification was assessed by repeating the analysis with different hyperparameters. Figure 3 shows the histograms of the posterior number of truncated planes and the probabilities of inclusion of covariates in the basis functions. We noticed that the average number of components varies insubstantially with various choices of the hyperparameters. It is evident that the last two covariates are basically excluded from the model, which is consistent with the design of the covariates.

5 Application

As an illustration, we applied the approach to the data of county-specific low birth weights (i.e., birth weight is less than 2500 gram) across 46 counties in the state of South Carolina during the period 1997–2007. A number of county-level low birth

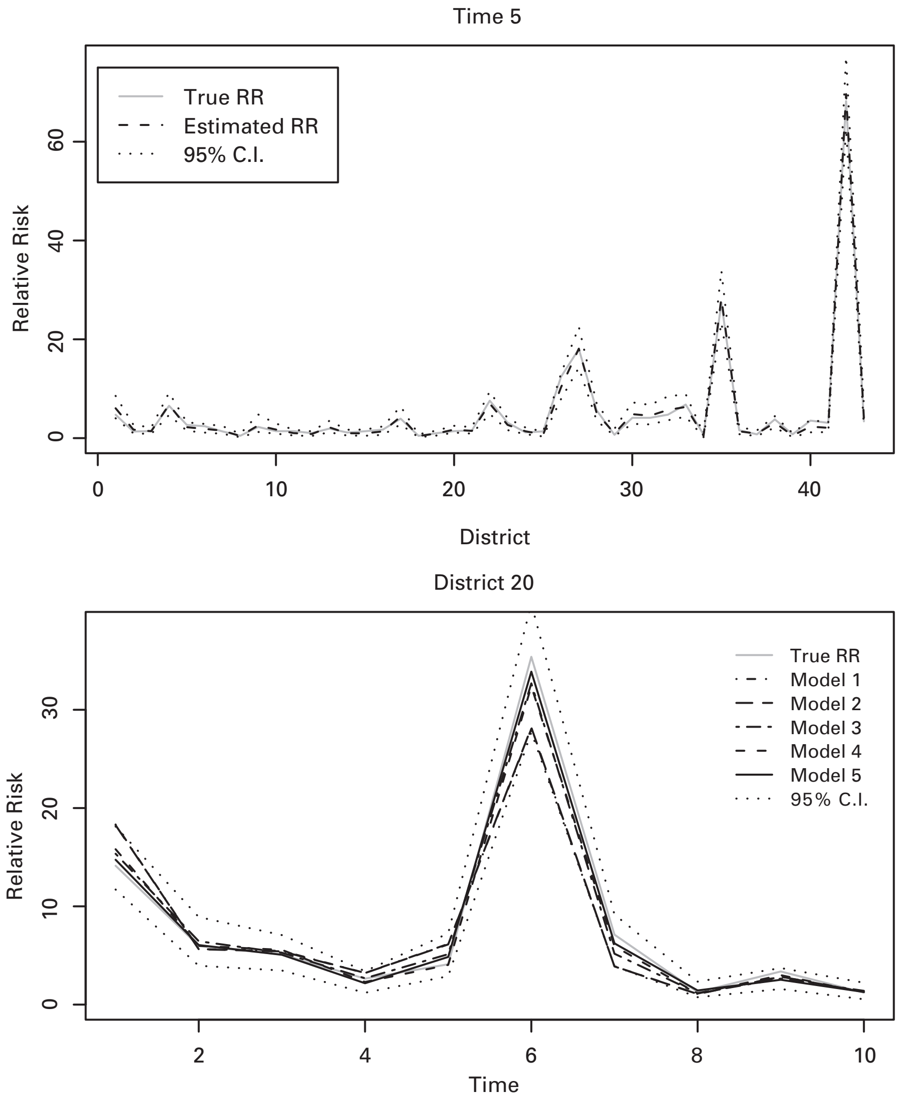

Figure 1 Results from the simulated example. Top panel: true and estimated relative risks with 95% pointwise credible intervals from the proposed approach (Model 5) across all districts at time 5. Bottom panel: true and estimated relative risks with 95% pointwise credible intervals from the proposed approach across time points for district 20, along with posterior estimates from Models 1–4

weights were obtained from South Carolina Department of Health and Environmental Control. The population density, the proportion of African American population, the household income and the poverty rate were acquired from the US census. The unemployment rates were attained from the US Bureau of Labor Statistics.

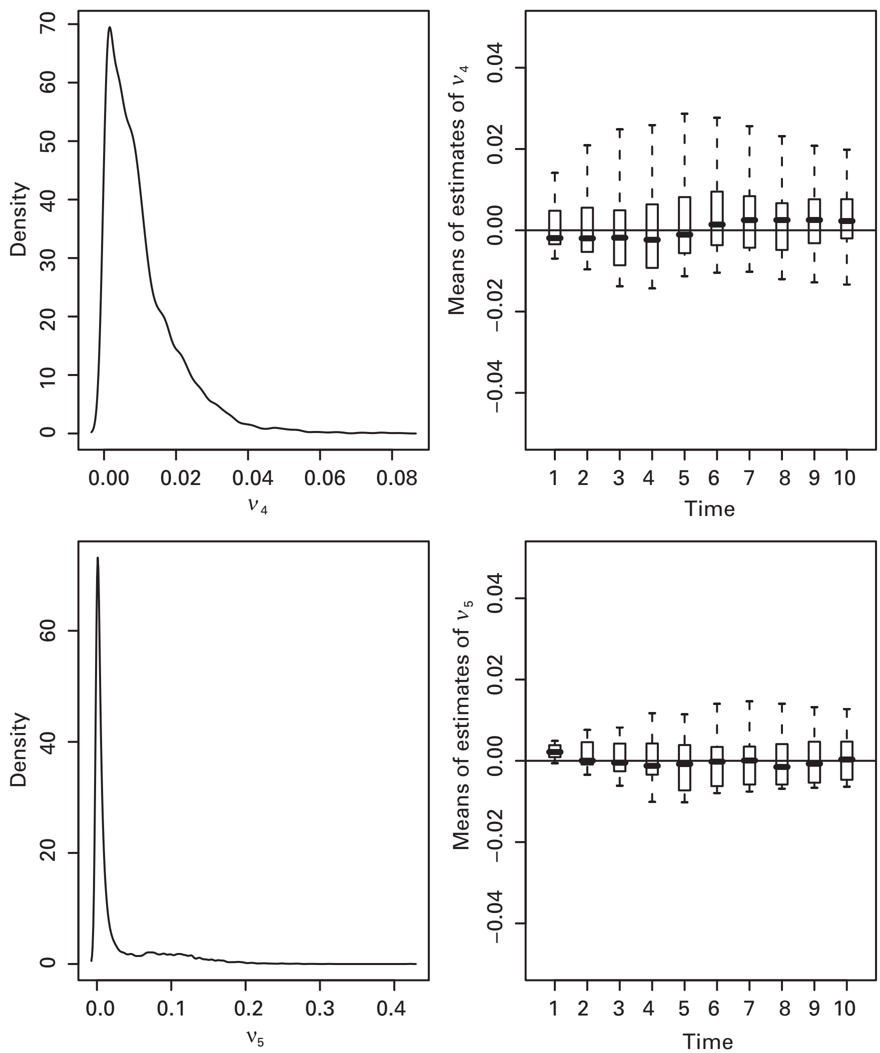

Figure 2 Left panel: posterior densities of the parameters ν_4 and ν_5 in the simulated example. Right panel: boxplots of the posterior means for the time-varying basis components υ_4 and υ_5 at each time point in the simulation. The horizontal line denotes the true value

In the data, y_{it} denotes the number of low birth weights in county i during year t and $\mathbf{x}_{it} = (1, x_{it1}, x_{it2}, x_{it3}, x_{it4}, x_{it5})'$ with x_{it1} indicating the county-level population density, x_{it2} the proportion of black people, x_{it3} the median household income, x_{it4} the poverty rate and x_{it5} the unemployment rate in county i for year t, for $i = 1, \ldots, 46$ and $t = 1, \ldots, 11$. The population density is defined as population divided by the total land area in square miles. The expected low birth weight counts for county i in

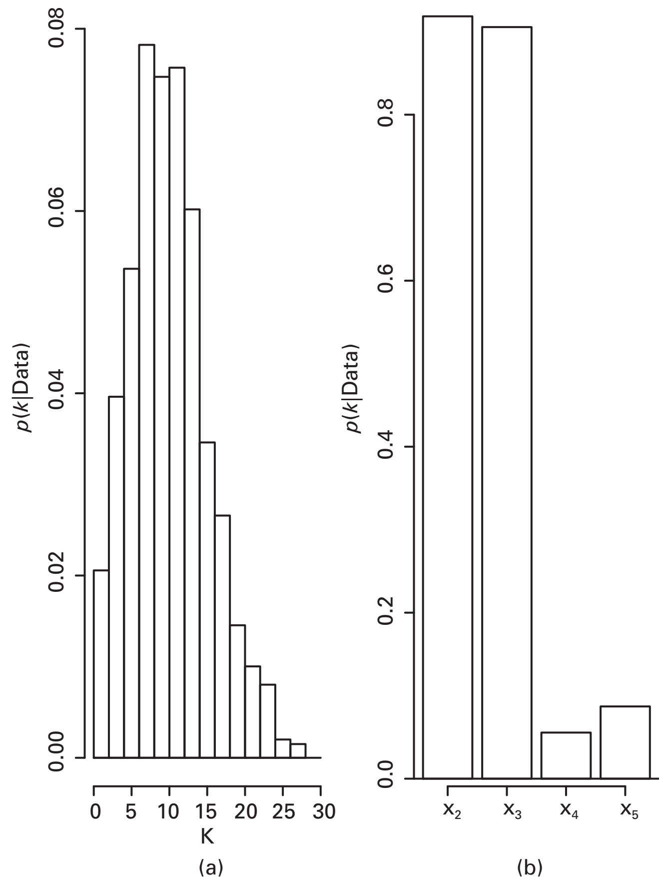

Figure 3 (a) Histogram of the posterior number of truncated planes and (b) the probabilities of inclusion of covariates in the basis functions for the simulated data

year t, E_{it}, is calculated by $n_{it}R$, where n_{it} is the total number of births for county i in year t and R is the overall statewide low birth weight rate which can be calculated by the total low birth weight counts divided by the total number of births over the entire counties and time periods.

We completed the specification of the proposed model by choosing prior Gamma$(0.05, 0.05)$ for τ_k and ZI-IG$(0.05, 0.05)$ for ν_l. The prior probability of a point mass at zero for the variance components of $\boldsymbol{\beta}_i$ and $\mathbf{u}_{tk,-1}$ is chosen to follow Beta$(1, 1)$. Since five covariates were initially included in the model, the initial number of truncated planes was chosen as 30. We collected 10 000 samples by thinning 50 000 samples by factor of 5 after a burn-in of 10 000 iterations.

Figure 4 displays spatial maps of the posterior means and the standard deviations of relative risk in years 1997, 2002 and 2007. Figure 5 shows the posterior means

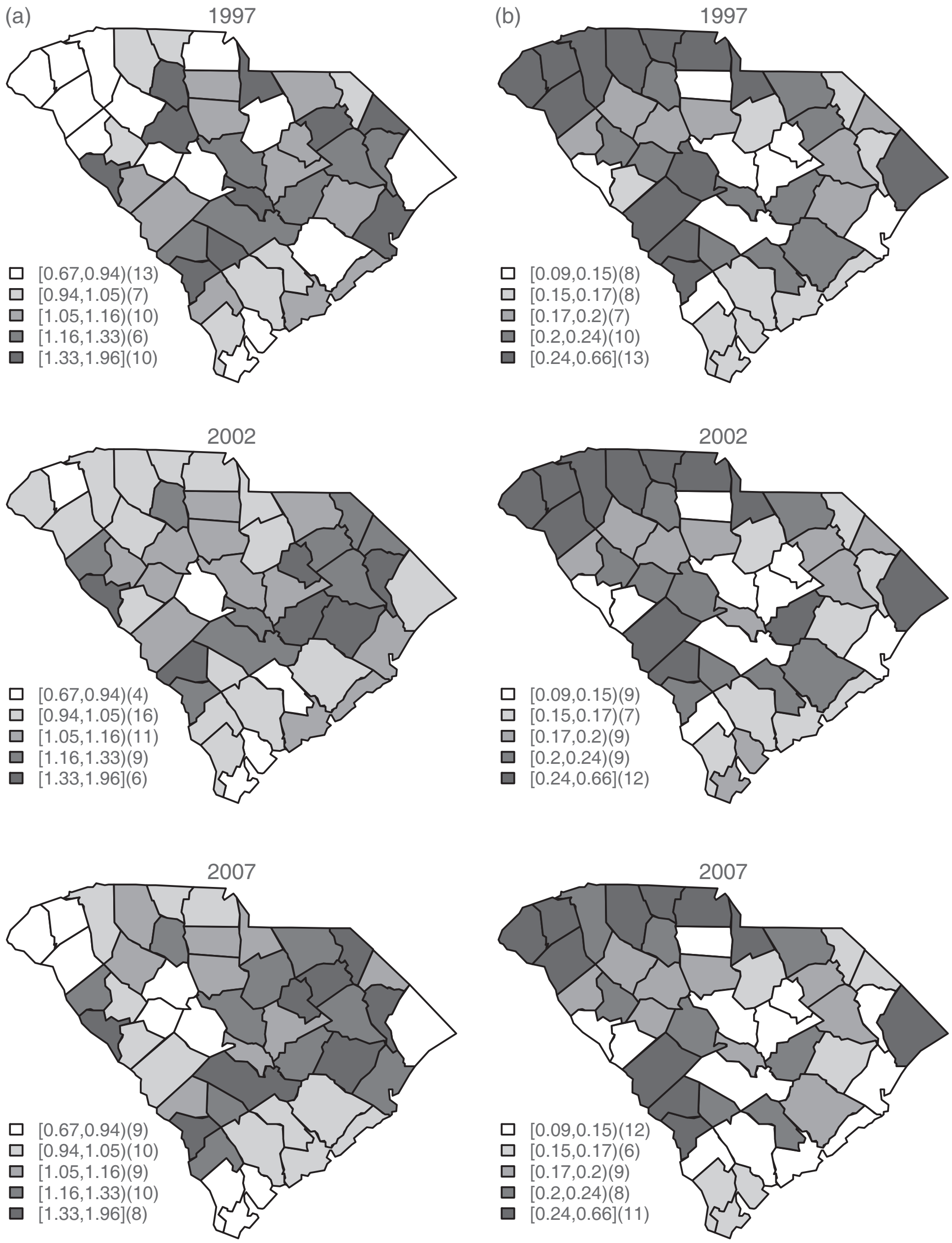

Figure 4 Spatial maps of (a) posterior mean and (b) posterior standard deviation (STD) of relative risks in years 1997, 2002 and 2007 for the low birth weight data in South Carolina. Left panel: posterior means of relative risk. Right panel: posterior STD of relative risk

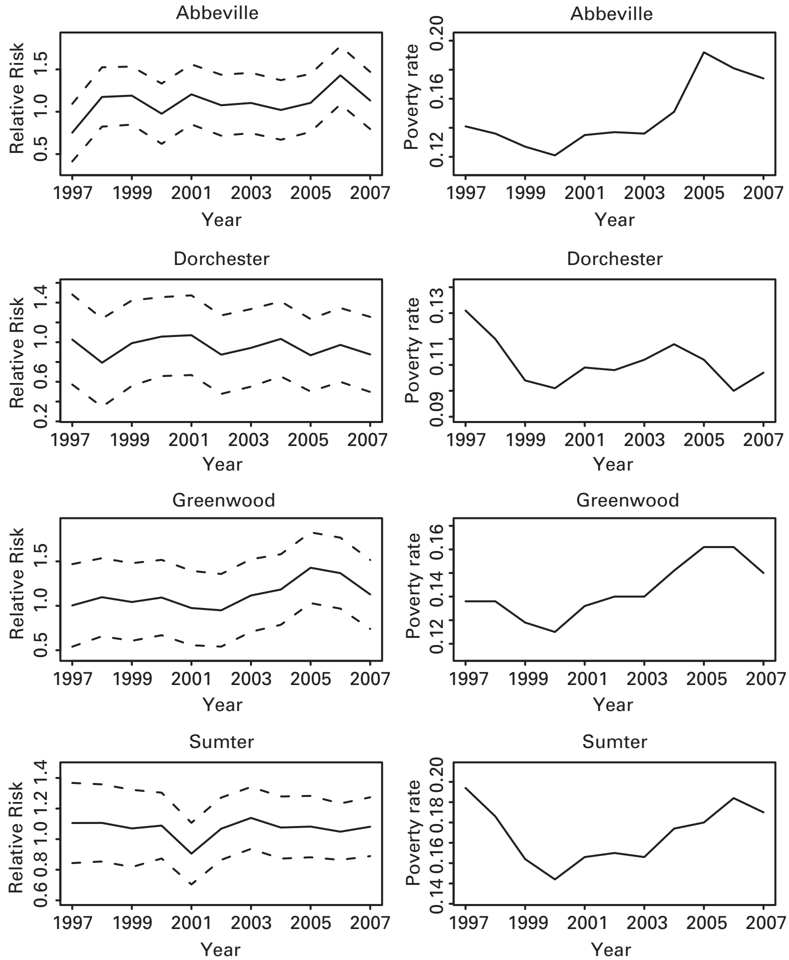

Figure 5 Posterior means and 95% pointwise credible intervals of relative risks of low birth weight for the four counties along with the corresponding poverty rates in SC through 11-year time period. Left panel: solid lines denote posterior means of relative risks and dashed lines denote 95% pointwise credible intervals. Right panel: poverty rate over time

and the 95% pointwise credible intervals for relative risk of low birth weight in four representative counties (randomly selected) along with their corresponding poverty rates over years 1997–2007. We can see that the estimated relative risk of low birth weight in County Dorchester with decreasing poverty rates slightly decreases over the 11-year time period. Counties Abbeville and Greenwood with increasing poverty rates basically have ascending trends over time. County Sumter roughly has a 'V' shape of poverty rate over time and the estimated relative risks, and interestingly the estimated relative risk has a similar curve. We also investigated the sensitivity of the number of truncated planes for $\boldsymbol{\beta}_i$ to various choices of the hyperparameters, which varies slightly. The posterior mean of the number of truncated planes is 12.4 with 95% credible interval (9.6, 16.5). We note that proportion of black people, the median household income, the poverty rate and the unemployment rate are included in the model with posterior probability of inclusion over 97% for each covariate while the population density has over 98% posterior probability of exclusion, implying that the population density can be excluded from the model at 5% significant level.

The third column in Table 1 shows the estimated negative cross-validatory predictive log-likelihoods for the proposed model along with the four competing models. We can see that the estimated $NLLK_{CV,app}$ values for Model 1 and 2 are close while the other three models have much lower $NLLK_{CV,app}$ values. The proposed model (Model 5) has the smallest $NLLK_{CV,app}$ value, evincing that it 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.

6 Discussion

We proposed a Bayesian regression model with multivariate linear splines for the analysis of space-time data. The proposed approach extends generalized multivariate regression splines (Holmes and Mallick, 2001) to flexibly accommodate the space- and time-specific covariates, allowing for flexible orientation and position of the covariate plane by incorporating time-varying basis functions.

One of the major advantages of a semiparametric modelling specification is the ability to flexibly model variation within a localized areas of a study region. In the proposed model, we allow geographically localized definition of the dependence of covariates and provide a flexible method of incorporation of variates via zero-inflation mixture priors. 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 supports 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. Computational intensity is noticed in the proposed approach, though it is reasonably efficient when it is coded in C language. Future work will focus on developing space-time models with nonparametric modelling and clustering on spatial effects coefficients, and on developing a more efficient sampling method.

Acknowledgements

The authors would also like to thank the editor, the associate editor and the referees for valuable comments which greatly improved the presentation of the paper. This work was supported by NIH/NHLBI 1R21HL088654-01A2.

Appendix

Full conditional distributions in Section 2.2.

Step 1: Update β_{ik}, for $k = 1, \ldots, K$, from its full conditional posterior distribution,

$$\prod_{i=1}^{n}\prod_{t=1}^{T} f\left\{\sum_{k=1}^{K}\beta_{ik}(\mathbf{x}_{it}'\mathbf{u}_{tk})_{+}\right\}\exp\left\{-\frac{\tau_k}{2}\sum_{i=1}^{n}\sum_{i\sim j}(\beta_{ik}-\beta_{jk})^2\right\},$$

with the conditional posterior probability

$$1-\hat{p}_{1,k} = Pr(\gamma_k = 0|\boldsymbol{\beta},\boldsymbol{\gamma}_{-k}) = \frac{1}{1+C_k},$$

where $C_k = p_{1,k0}/(1-p_{1,k0}) \times L(\boldsymbol{\beta}_k = 0, \boldsymbol{\beta}_{-k}, \mathbf{u}, \boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\gamma})/L(\boldsymbol{\beta}_k = \boldsymbol{\beta}_k, \boldsymbol{\beta}_{-k}, \mathbf{u}, \boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\gamma})$ with $L(\boldsymbol{\beta}_k, \boldsymbol{\beta}_{-k}, \mathbf{u}, \boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\gamma}) = \prod_{i=1}^{n}\prod_{t=1}^{T} f\{\sum_{k=1}^{K}\beta_{ik}(\mathbf{x}_{it}'\mathbf{u}_{tk})_{+}\}$ and $\boldsymbol{\beta}_k = (\boldsymbol{\beta}_{1k}, \ldots, \boldsymbol{\beta}_{nk})'$. Otherwise, β_{ik} is assigned to be zero. Simultaneously, γ_k ($k = 1, \ldots, K$) can be sampled from its full conditional posterior distribution, Bern($\hat{p}_{1,k}$).

Step 2: Update $p_{1,k0}$ from its full conditional distribution,

$$p_{1,k0}|\boldsymbol{\gamma} \sim \mathrm{Beta}(c_1 + n_{\boldsymbol{\gamma}}, d_1 + K - n_{\boldsymbol{\gamma}}),$$

where $\boldsymbol{\gamma}$ corresponds to a model from the model space $\mathcal{M}$ containing 2^K models, and $n_{\boldsymbol{\gamma}}$ denotes the number of excluded predictors in the model, i.e., $\sum_{k=1}^{K}\gamma_k$.

Step 3: Update τ_k, for $k = 1, \ldots, K$, from its full conditional posterior distribution,

$$\mathrm{Gamma}\left(a_\tau + \frac{n}{2}, b_\tau + \frac{1}{2}\sum_{i=1}^{n}\sum_{i\sim j}(\beta_{ik}-\beta_{jk})^2\right).$$

Step 4: Update ν_l, for $l = 2, \ldots, p$, from its full conditional distribution with a point mass at zero

$$\kappa_l\delta_0(\nu_l) + (1-\kappa_l)\mathrm{IG}\left\{a_\nu + \frac{KT}{2}, b_\nu + \frac{1}{2}\sum_{t=1}^{T}\sum_{k=1}^{K}(u_{tkl} - \rho_k u_{t-1,kl})^2\right\},$$

where $\kappa_l \sim$ Bern($\hat{p}_{2,l}$) with $\hat{p}_{2,l} = (p_{2,l0}L(\boldsymbol{\beta}, \mathbf{u}_l = 0, \boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\gamma}))/(p_{2,l0}L(\boldsymbol{\beta}, \mathbf{u}_l = 0, \boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\gamma}) + (1-p_{2,l0})L(\boldsymbol{\beta}, \mathbf{u}_l = \mathbf{u}_l, \mathbf{u}_{-l}, \boldsymbol{\tau}, \boldsymbol{\nu}, \boldsymbol{\gamma}))$ and $\mathbf{u}_l = \{u_{tkl}\}_{t,k}$.

Step 5: Update $\mathbf{u}_{tk}$ from its full conditional distribution, for $t = 1, \ldots, T$ and $k = 1, \ldots, K$,

$$\prod_{i=1}^{n}\prod_{t=1}^{T} f\left\{\sum_{k=1}^{K}\beta_{ik}(\mathbf{x}_{it}'\mathbf{u}_{tk})_{+}\right\}\exp\left\{-\frac{1}{2}(\mathbf{u}_{tk}-\rho_k\mathbf{u}_{t-1,k})'\boldsymbol{\nu}^{-1}(\mathbf{u}_{tk}-\rho_k\mathbf{u}_{t-1,k})\right\},$$

For each t and k, we standardize the components of $\mathbf{u}_{tk,-1}$ and $u_{tk1} = \mathbf{x}_{it,-1} \mathbf{u}_{tk,-1}$.

Step 6: Update $p_{2,l0}$ from its full conditional distribution,

$$p_{2,l0}|\boldsymbol{\kappa} \sim \text{Beta}(c_2 + n_{\boldsymbol{\kappa}}, d_2 + p - 1 - n_{\boldsymbol{\kappa}}),$$

where $n_{\boldsymbol{\kappa}}$ denotes the number of excluded predictors in the model, i.e., $\sum_{l=2}^{p} \kappa_l$.

Step 7: When link function is identity, update τ from its full conditional distribution

$$\text{Gamma}\left\{c_\tau + \frac{nT}{2}, d_\tau + \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T}(y_{it} - \eta_{it})^2\right\},$$

where $\pi(\tau) \sim \text{Gamma}(c_\tau, d_\tau)$.

References

Assunção RM (2003) Space varying coefficient models for small area data. *Environmetrics*, **14**, 453–73.

Assunção RM, Potter JE and Cavenaghi SM (2002) A Bayesian space varying parameter model applied to estimating fertility schedules. *Statistics in Medicine*, **21**, 2057–2075.

Banerjee S and Johnson GA (2006) Coregionalized single and multi-resolution spatially varying growth-curve modelling. *Biometrics*, **62**, 864–76.

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

Bigelow J and Dunson DB (2007) Bayesian adaptive regression splines for hierarchical data. *Biometrics*, **63**, 724–32.

Cowles MK and Carlin BP (1995) Markov Chain Monte Carlo diagnostics: A comparative review. *Journal of the American Statistical Association*, **91**, 883–04.

Cui W and George EI (2008) Empirical Bayes vs. fully Bayes variable selection. *Journal of Statistical Planning and Inference*, **138**, 888–900.

Denison DGT, Holmes CC, Mallick BK and Smith AFM (2002) *Bayesian Methods for Nonlinear Classification and Regression*. Chichester: John Wiley.

Dey D Chen MH and Change H (1997) Bayesian approach for nonlinear random effects models. *Biometrics*, **53**, 1239–1252.

Draper D and Krnjajić M (2006) *Bayesian model specification.* Technical report, Department of Applied Mathematics and Statistics, Baskin School of Engineering, University of California, Santa Cruz.

Dreassi E, Biggeri A and Catelan D (2005) Space-time models with time dependent covariates for the analysis of the temporal lag between socio-economic factors and mortality. *Statistics in Medicine*, **24**, 1919–32.

Friedman JH (1991) Multivariate adaptive regression splines. *Annals of Statistics*, **19**, 1–141.

Gamerman D, Moreira ARB and Rue H (2003) Space-varying regression models: specifications and simulation. *Computational Statistics and Data Analysis*, **42**, 513–33.

Geisser S (1984) On prior distribution for binary trials. *American Statistician*, **38**, 244–51 (with discussion).

Geisser S (1993) Predictive inference: An introduction. Chapman & Hall, London.

Gelfand AE, Dey D and Chang H (1992) Model determination using predictive distributions with implementation via sampling based methods (with discussion). In J Bernado *et al.* (eds), *Bayesian Statistics 4*. Oxford University Press, 147–67.

Gelfand A, Banerjee S and Gamerman D (2005) Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, **16**, 465–79.

George EI and McCulloch RE (1993) Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, **88**, 881–89.

Gilks WR, Best NG and Tan KKC (1995) Adaptive rejection metropolis sampling within Gibbs sampling. *Journal of Applied Statistics*, **44**, 455–72.

Green P (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–32.

Hastie TJ and Tibshirani RJ (1990) *Generalized additive models*. London: Chapman & Hall.

Holmes C and Mallick B (2001) Bayesian regression with multivariate linear splines. *Journal of the Royal Statistical Society, Series B*, **63**, 3–17.

Holmes C and Mallick B (2003) Generalized nonlinear modelling with multivariate free-knot regression splines. *Journal of the American Statistical Association*, **98**, 352–68.

Hossain MM and Lawson AB (2010) Space-time Bayesian small area disease risk models: development and evaluation with a focus on cluster detection. *Environmental and Ecological Statistics*, **17**, 73–95.

Kneib T and Fahrmeir L (2006) Structured additive regression for categorical space-time data: a mixed model approach. *Biometrics*, **62**, 109–18.

Knorr-Held L (2000) Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, **19**, 2555–67.

Knorr-Held L and Besag J (1998) Modelling risk from a disease in time and space. *Statistics in Medicine*, **17**, 2045–60.

Lagazio C, Biggeri A and Dreassi E (2003) Age-period-cohort models and disease mapping. *Environmetrics*, **14**, 475–90.

Lagazio C, Dreassi E and Biggeri A (2001) A hierarchical Bayesian model for space-time variation of disease risk. *Statistical Modelling*, **1**, 17–29.

Lunn DJ, Thomas A, Best N and Spiegelhalter D (2000) WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–37.

Paez MS, Gamerman D, Landim FMP and Salazar E (2008) Spatially varying dynamic coefficient models. *Journal of Statistical Planning and Inference*, **138**, 1038–58.

Pintore A, Speckman P and Holmes C (2006) Spatially adaptive smoothing splines. *Biometrika*, **93**, 113–25.

Plummer M (2008) Penalized loss functions for Bayesian model comparison. *Biostatistics*, **9**, 523–39.

Plummer M, Best NG, Cowles K and Vines K (2006) CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, **6**, 7–11.

Richardson S, Abellán JJ and Best N (2006) Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire. *Statistical Methods in Medical Research*, **15**, 385–407.

Ruppert D (2002) Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, **11**, 735–57.

Ruppert D, Wand MP and Caroll RJ (2003) *Semiparametric regression*. Cambridge: Cambridge University Press.

Schmid V and Held L (2004) Bayesian extrapolation of space–time trends in cancer registry data. *Biometrics*, **60**, 1034–42.

Scott J and Berger J (2006) An exploration of aspects of Bayesian multiple testing. *Journal of Statistical Planning and Inference*, **136**, 2144–62.

Scott J and Berger J (2008) *Multiple testing, empirical bayes, and the variable-selection problem*. Discussion Paper 2008-10, Department of Statistical Science, Duke University.

Sinha D and Dey DK (1997) Semiparametric Bayesian analysis of survival data. *Journal of the American Statistics Association*, **92**, 1195–1212.

Spiegelhalter D, Thomas A, Best N and Gilks W (1996) *BUGS 0.5: Bayesian inference using Gibbs sampling manual*. Cambridge, UK: MRC Biostatistics Unit, Institute of Public Health.

Spiegelhalter DJ, Best NG, Carlin BP and Linde AVD (2002) Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, **64**, 1–34.

Ugarte MD, Goicoa T, Ibáñez B and Militino AF (2009) Evaluating the performance of spatio-temporal Bayesian models in disease mapping. *Environmetrics*, **20**, 647–65.

Waller LA, Carlin BP, Xia H and Gelfand AE (1997) Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, **92**, 607–17.

Zhao Y, Staudenmayer J, Coull BA and Wand MP (2006) General design Bayesian generalized linear mixed models. *Statistical Science*, **21**, 35–51.