

## Space-time stick-breaking processes for small area disease cluster estimation

Md. Monir Hossain · Andrew B. Lawson · Bo Cai ·  
Jungsoon Choi · Jihong Liu · Russell S. Kirby

Received: 19 April 2011 / Revised: 3 January 2012 / Published online: 15 June 2012  
© Springer Science+Business Media, LLC 2012

**Abstract** We propose a space-time stick-breaking process for the disease cluster estimation. The dependencies for spatial and temporal effects are introduced by using space-time covariate dependent kernel stick-breaking processes. We compared this model with the space-time standard random effect model by checking each model's ability in terms of cluster detection of various shapes and sizes. This comparison was made for simulated data where the true risks were known. For the simulated data, we have observed that space-time stick-breaking process performs better in detecting medium- and high-risk clusters. For the real data, county specific low birth weight incidences for the state of South Carolina for the years 1997–2007, we have illustrated how the proposed model can be used to find grouping of counties of higher incidence rate.

**Keywords** Cluster · Dependence · Dirichlet process mixture · Space-time · Stick-breaking processes

---

Md. M. Hossain (✉)  
Division of Biostatistics and Epidemiology, Cincinnati Children's Hospital Medical Center,  
Cincinnati, Ohio  
e-mail: md.hossain@cchmc.org

A. B. Lawson · J. Choi  
Division of Biostatistics, Bioinformatics and Epidemiology, Medical University of South Carolina,  
Charleston, SC, USA

B. Cai · J. Liu  
Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC, USA

R. S. Kirby  
Department of Community and Family Health, University of South Florida, Tampa, FL, USA

## 1 Introduction

Health data are now routinely available for many spatial locations for successive years. This repeated year information introduces a further dimension that is often important because of changes in socio-demographic structure or other health factors during the time period of the study. It has also been observed that any significant changes close in space can be close in time too. This observation underscores the importance of space-time modeling and emphasizes the consideration of two potential dimensions of dependency: between spatially neighbor locations, and between successive time points. Including these dimensions in the analysis adjusts the parameter estimates for spatial, temporal and spatial-temporal dependences, and contributes to the better understanding of disease etiology. Central to small area health investigations, an important application of these models is the identification of areas associated with high rates of disease incidence, i.e. cluster detection.

Recent works by [Gelfand et al. \(2005\)](#), [Griffin and Steel \(2006\)](#) and [Duan et al. \(2007\)](#) have proposed to use Dirichlet process (DP) prior distribution for better risk estimation of spatially reference health data. In these works, the spatial dependence was introduced either by defining spatial model for the mixing weights or for the mixing components. In [Gelfand et al. \(2005\)](#), dependent DP prior distributions are assumed for the random mixing components, and stick-breaking prior distributions are assumed for the weight components. The weights are not indexed by spatial locations. For the consideration of dependency among the neighboring spatial locations, a zero-mean stationary Gaussian process model was considered for the base distribution from which the random mixing components are drawn. A generalization of the above spatial dependence DP was proposed by [Duan et al. \(2007\)](#). In their proposal, instead of a common surface selection (i.e., common weights) for all spatial locations, a latent covariate was introduced to determine a random surface selection (i.e., the weights are spatial location dependent). The latent covariates are generated from a Gaussian random field. In these works ([Gelfand et al. 2005](#); [Duan et al. 2007](#)), an error process of Gaussian distributions was used for the mixing components to have continuous supports for the observed data, and by this way the discreteness of the DP prior distribution was resolved. In addition to this, the dependent DP was introduced through the mixing components by assigning zero-mean stationary Gaussian process prior distribution, i.e., by allowing the mixing parameter to be drawn from a random field. In a different approach, [Griffin and Steel \(2006\)](#) proposed a dependent DP by introducing an ordered stick-breaking prior distribution for the mixing weights. The ordering depends on the closeness of a covariate value, i.e., distributions for similar covariate values will be assigned to similar orderings. In a non-spatial setup, [Dunson and Park \(2008\)](#) proposed kernel based stick-breaking processes to smooth the mixing weights. [Reich and Fuentes \(2007\)](#) extended these kernels to include spatial processes in order to model the spatial dependence where the weights are spatially indexed.

The development of space-time dependent Dirichlet Processes is very limited. [Gelfand et al. \(2005\)](#) used the temporal observations as the replications at each spatial location in their development of a spatial dependence DP. Instead of considering this temporal information as independent replications in spatial dependence DP, [Kottas et al. \(2007\)](#) viewed it as a temporally evolving spatial process. The proposal is based

on the decomposition of mixing components with temporal index into a first-order autoregressive term and a random innovation term. Dependent DP prior distributions are assumed for the random innovations, and a zero-mean stationary Gaussian process is considered for the base distribution.

In this paper, we propose a different space-time dependence DP mixture model for the modeling of incidence count data that are observed at spatial locations for successive time points. In our proposal, we introduce the dependencies for spatial and temporal effects by using space-time covariate dependent kernel stick-breaking processes. The space-time covariate dependent kernel can include spatially varying regression coefficients and separable/non-separable space-time effects. The spatially varying regression coefficients are modeled by a multivariate conditional autoregressive (MCAR) type model, and the temporal effects are modeled by a first order autoregressive type model. Our proposal is different from Kottas et al. (2007) in at least two ways: i) we use the mixing weights instead of mixing components to introduce space-time dependency; and ii) our space-time dependent kernel includes separate effects for space, time, and their interaction, instead of considering a dynamic spatial process.

The rest of this paper is structured as follows. First, we give a brief description of DP nonparametrics. Following that, we introduce the proposed space-time stick-breaking process (Sect. 2). Section 3 describes a standard space-time random effect model, employed here for comparison with the proposed stick-breaking process. Implementation of the posterior sampling algorithm and the prior distribution specifications are discussed in Sect. 4. In sect. 5 we describe a simulation study with the aim of comparative evaluation of the two modeling approaches. The comparisons are made by using a cluster detection diagnostics by checking these models’ ability in terms of cluster detection of various shapes and sizes. An application to county level low birth weight incidences in the state of South Carolina is described in Sect. 6. With the real data, we describe how the grouping information of space-time stick-breaking process can be used for finding clusters of similar risk areas. Section 7 provides the concluding remarks.

### 1.1 Dirichlet process mixture background

The data  $y_1, \dots, y_n$  are assumed to be generated from an unknown density  $r$ , i.e.  $y_i \stackrel{iid}{\sim} r(y_i), i = 1, \dots, n$ . Then, the unknown  $r$  in Dirichlet process mixture (DPM) model (Ferguson 1983; Lo 1984) framework can be written as

$$r(y) = \int_{\Theta} f(y|\theta)dG(\theta) \tag{1.1}$$

where  $f(\cdot|\theta)$  is a conditional density for a given cluster membership, i.e. indexed by  $\theta$  which can be a scalar or a vector, and  $\theta \in \Theta$ .  $G(\cdot)$  in (1.1) is the mixing distribution. In Bayesian nonparametric,  $G(\cdot)$  is a random probability measure (RPM), i.e., a probability measure over a set of probability distributions. The other way of seeing (1.1),

marginally, the distribution of  $y$  is a mixture distribution over density  $f(y|\theta)$ . The RPM  $G(\theta)$  can be discrete or continuous. In order to achieve grouping in the observed data, we are to introduce discreteness in  $G(\theta)$ . We select a DP prior distribution for this RPM,  $G(\cdot)$ .

The stick-breaking construction of Sethuraman (1994) confirms that the measures drawn from a DP are discrete with probability one. The probability measure  $G$  can be represented as infinite mixtures of point masses, as  $G = \sum_{j=1}^{\infty} \omega_j \delta_{\theta_j}$ , where the locations  $\theta_j \stackrel{iid}{\sim} G_0$ ,  $\delta_{\theta}$  denotes the point mass located at  $\theta$  and the stick-breaking probabilities  $\omega_j = \gamma_j \prod_{k=1}^{j-1} (1 - \gamma_k)$  with  $\gamma_j \stackrel{iid}{\sim} \text{Beta}(1, \alpha)$ .

Using this constructive definition of the DP, Eq. (1.1) can be written as  $r(y) = \sum_{j=1}^{\infty} \omega_j f(y|\theta_j)$ . It reveals the fact that the prior distribution for the unknown  $r$  can be defined by infinite mixtures of density  $f$ . The above can also be written as  $y_i \stackrel{iid}{\sim} \sum_{j=1}^{\infty} \omega_j f(\cdot|\theta_j)$ .

The DPM is defined by writing the above model in a Bayesian hierarchical fashion

$$\begin{aligned} y_i &| \theta_i^* \sim f(\cdot|\theta_i^*), \quad i = 1, \dots, n \\ \theta_i^* &| G \stackrel{iid}{\sim} G \\ G &\sim DP(\alpha, G_0) \end{aligned} \tag{1.2}$$

In words,  $y_i$  is drawn from  $f(\cdot|\theta_i^*)$ , parameterized by  $\theta_i^*$ ,  $\theta_i^*$  is drawn independently from an identical RPM  $G$  and  $G$  has a DP prior distribution with a base distribution  $G_0$  and a scale (or, concentration) parameter  $\alpha$ . DPM is a mixture model since a number of  $\theta_i^*$ 's can take the same value because of discrete  $G$ , and  $y_i$ 's with the same value of  $\theta_i^*$  belong to the same cluster. Hence, the  $\theta_i^*$ 's give rise to a partition of the set of  $y_i$ 's and  $\alpha$  controls the distribution of the number of partitions.

Another way of writing the above DPM models for a finite number of clusters (say,  $K$ ) with a cluster assignment variable  $z_i$  and  $\theta_i^* = \theta_{z_i}$  as

$$\begin{aligned} y_i | z_i, \theta &\sim f(\cdot|\theta_{z_i}), \quad i = 1, \dots, n \\ z_i | \omega_1, \dots, \omega_K &\sim \text{Discrete}(\omega_1, \dots, \omega_K) \\ \omega_1, \dots, \omega_K &\sim D(\alpha/K, \dots, \alpha/K) \\ \theta_l &\sim G_0, \quad l = 1, \dots, K \end{aligned}$$

In this representation, the cluster assignment variable  $z_i$  loses the meaning of discreteness as the space for  $z_i$  becomes continuous in the limit  $K \rightarrow \infty$  since the number of clusters becomes countably infinite and the prior expectation of any specific cluster tend to zero. Hence, the assignment variable  $z_i$  can be ignored and  $\theta_{z_i}$  can be replaced by  $\theta_j$ . The parameter  $\theta_j$  is drawn from a DP with base distribution instead. Green and Richardson (2001) showed that for large  $K$  when  $\alpha/K \rightarrow 0$ , the finite mixture model with multinomial prior distributions for the assignment variables and Dirichlet distribution for the weight (DMA in their terminology) converges to DPM.

## 2 Stick-breaking process space-time (SBPST) model

We assume that disease incidence data are observed in a fixed set of small areas and for discrete time periods. Denote the small area units as  $i = 1, \dots, n$ , and the temporal units as  $t = 1, \dots, T$ . The observed count of incidence is denoted by  $O_{it}$ , and the expected count by  $E_{it}$  for location  $i$  and time  $t$ . The expected count is usually assumed to be fixed and obtained from a standard population after adjustment for case-mix. The observed count can be modeled as a mixture of Poisson distributions as

$$O_{it}|K, \omega, \theta \sim \sum_{j=1}^K \omega_{itj} \text{Poisson}(O_{it}|E_{it}\theta_j), \tag{2.1}$$

where  $\theta = (\theta_1, \dots, \theta_K)$  is a vector of mixing components, and non-negative  $\omega_{itj}$ 's are the mixing weights and satisfy the condition  $\sum_{j=1}^K \omega_{itj} = 1$ . The number of mixing components  $K$  is not fixed beforehand and considered a random variable.

Following the De Finetti theorem, if  $(O_{11}, O_{12}, \dots)$  are infinitely exchangeable, then the joint probability  $p(O_{11}, O_{12}, \dots, O_{nT})$  has a representation as a mixture:

$$p(O_{11}, O_{12}, \dots, O_{nT}) = \int \left( \prod_{i=1}^n \prod_{t=1}^T f(O_{it}|\theta) \right) dG(\theta)$$

for RPM,  $G(\theta)$ . A general representation of this can be written in the DPM framework (1.2) as

$$\begin{aligned} O_{it}|E_{it}, \theta_j &\sim f(\cdot|E_{it}\theta_j), \quad i = 1, \dots, n \text{ and } t = 1, \dots, T \\ \theta_j|G &\stackrel{iid}{\sim} G \\ G &\sim DP(\alpha, G_0). \end{aligned}$$

The functional form for  $f(\cdot|\cdot)$  can assume any probability distribution that belongs to exponential family, and defined for positive random variable. Depending on this specific distribution,  $\theta_j$  could be a scalar or a vector of parameters. We assume a Poisson distribution for  $f(\cdot|E_{it}\theta_j)$  which is a reasonable assumption for count data and because of this specification,  $\theta_j$  can be interpreted as a relative-risk parameter. This interpretation of  $\theta_j$  potentially contributes to the specification of the baseline distribution,  $G_0$ . We assume a gamma distribution of the form  $\text{Gamma}(\zeta, \xi)$  for the baseline measure  $G_0$ , where  $\zeta$  and  $\xi$  are respectively the shape and scale parameters.

The stick-breaking construction of the RPM is  $G = \sum_{j=1}^\infty \omega_{itj} \delta_{\theta_j}$ , where  $\delta_{\theta_j}$  is the mass probability at point  $\theta_j$ . The mixing weight,  $\omega_{itj}$  is indexed by space and time so that the component specific weights will be able to incorporate spatial and temporal dependence into the mixture model. We define stick-breaking representation for the weights by  $\omega_{itj} = B_{itj} \prod_{k=1}^{j-1} (1 - B_{itk})$ , where  $B_{itj} = q_{itj} p_j$  with  $q_{itj}$  is a covariate dependent kernel function and  $p_j \sim \text{Beta}(1, \alpha)$ . The above stick-breaking representation tells, the unit probability stick is broken off by the amount  $q_{it1} p_{t1}$  for the first basis,

by the amount  $q_{it2}p_{t2}$  from the remaining probability  $(1 - q_{it1}p_{t1})$  for the second basis, and it continues until using up the unit stick in the limit. Although there are many possible kernel functions that can be used for spatially correlated data, e.g. uniform or squared-exponential kernel (Reich and Fuentes 2007), or logistic normal or grouped continuous kernel (Fernandez and Green 2002), we preferred to use logistic normal kernel since Fernandez and Green (2002) already demonstrated its performance in disease mapping and also it is reasonably flexible to introduce spatially varying regression coefficients with the space-time covariates to this kernel. The covariate dependent kernel is specified by a logistic normal form  $q_{itj} = \exp(h_{itj}/\phi) / \sum_{l=1}^K \exp(h_{itl}/\phi)$ , where the scaling parameter,  $\phi (> 0)$ , controls the smoothness of  $h_{itj}$ , and the specific form of it will depend on the application. Instead of logit link function, other link functions such as probit are also possible but here we did not explore this possibility. The  $\phi$  has a uniform prior distribution with the range  $(0, \phi_{\max})$ ,  $\phi_{\max}$  is assumed to be a small number (generally, 10) (Green and Richardson 2002).

The dependency on covariates can be introduced by defining a function  $h_{itj} = \beta_{ij}^T \mathbf{X}_{it}$ , where  $\mathbf{X}_{it} = (1, X_{2it}, \dots, X_{pit})^T$  and  $\beta_{i1}, \beta_{i2}, \dots, \beta_{iK}$  are  $K$  vectors of unknown regression parameters and each of which has  $p$  elements. These parameters are component specific, even though  $\mathbf{X}_{it}$  is constant across mixtures. This model in addition of heterogeneous covariate effect, also considers the spatial dependence between the covariates. In Hossain and Lawson (2010), we have implemented a simpler model as  $h_{itj} = \kappa_{ij} + \gamma_{tj}$ , where the  $\kappa$ 's are the structured spatial random effects,  $\gamma$ 's are the structured temporal random effects. However, the function  $h_{itj}$  can also include unstructured spatial and temporal effects, and separable and non-separable space-time interaction effects. In order to define the exact form for  $h_{itj}$ , we need to remember the fact in many applications with real data, including the non-separable interaction effects does not improve the goodness-of-fit statistics (e.g., Knorr-Held 2000). In the current applications with South Carolina law birth weight data, we will illustrate two other functional forms for  $h_{itj}$  and also explain how to model non-separable space-time interaction effects.

To ensure that the above space-time stick-breaking prior is proper, i.e.,  $\sum_{j=1}^K \omega_{itj} = 1$ , according to Ishwaran and James (2001) (see also, Ishwaran and Zarepour 2002) we must show for infinite  $K$  that  $\sum_{j=1}^{\infty} E(\log(1 - B_{itj})) = -\infty$ . The proof is straightforward since  $E(q_{itj})$  and  $E(p_j)$  are both positive. It is relevant to mention that  $q_{itj}$  is restricted to the interval  $[0, 1]$ . For finite  $K$ ,  $\sum_{j=1}^K \omega_{itj} = 1$  can be proved by setting  $B_{itK} = 1$ . Setting the condition  $B_{itK} = 1$  means, the infinite mixture is truncated to a finite  $K$  by putting all of the masses for the components with  $j \geq K$  to  $\omega_{itK}$ .

### 3 Standard random effect space-time (SREST) model

We compare the above SBPST model to the standard random effect space-time (SREST) model since it is one of the best performing model in disease mapping for space-time data (Knorr-Held and Besag 1998; Knorr-Held 2000). Because of its simple hierarchical structure, it is relatively easier to implement in any statistical software (specifically, in WinBUGS). Although SREST model was initially introduced for smoothing risks, this model can also be used to find clusters of excess risk regions.

The SREST model can be regarded as a hierarchical generalized mixed model, which allows any probability distribution that belongs to the exponential family to be specified at the first level of hierarchy. For example [Knorr-Held and Besag \(1998\)](#) used Bernoulli distribution for modeling the Ohio lung cancer mortality data. Since, the outcome variable of our interest is a count data, we assume a Poisson distribution at the first level of hierarchy, specified as

$$O_{it}|\theta_{it} \sim \text{Poisson}(E_{it}\theta_{it}).$$

The specification for the model for relative-risk,  $\theta_{it}$ , will depend on study objective. For example, the logarithm of relative risk could have the form  $\log(\theta_{it}) = \beta_i^T \mathbf{X}_{it}$ , where  $\mathbf{X}_{it} = (1, X_{2it}, \dots, X_{pit})^T$ , and  $\beta_{i1}, \beta_{i2}, \dots, \beta_{ip}$  are  $p$  unknown regression coefficients and the joint probability distribution for these coefficients can be specified by multivariate CAR model ([Assunção 2003](#)). Since, our interest is to get a better area specific relative-risk estimate for the observed cancer mortality for the simulated dataset in order to compare each model’s ability in terms of cluster detection of various shapes and sizes; it will be useful to include random effects to represent each source of variation. Considering random effects model is also common ([Best et al. 2005](#)) in small area health investigation since most often the disease under study is rare and the area population sizes are small. These conditions increase the random variation associated with risk estimates and including random effects to include each source of these variations may improve the relative risk estimates. We consider a nearly saturated model for the logarithm of relative-risk, defined as

$$\log(\theta_{it}) = \rho + \kappa_i + \gamma_t + \varepsilon_i + \xi_t + \lambda_{it},$$

where  $\rho$  is an overall mean parameter,  $\kappa_i$  and  $\gamma_t$  are spatial and temporal random effects for structured heterogeneity,  $\varepsilon_i$  and  $\xi_t$  are for unstructured heterogeneity, and  $\lambda_{it}$  is a space-time separable random effect.

#### 4 Posterior sampling and the prior distribution specification

Recent research ([Ohlssen et al. 2007](#)) suggested that from computational aspect a finite approximation of the stick-breaking representation of the full DP model is more feasible to apply. The SBPST model in (2.1) can be regarded as a finite approximation of the full model where the number of mixing component,  $K$  could be fixed to a number for which very minimal information is compromised. This finite approximation for SBPST model can be implemented in WinBUGS ([Spiegelhalter et al. 2003](#)) by introducing space-time latent variables  $\mathbf{Z} = (Z_{11}, Z_{12}, \dots, Z_{nT})$ . These latent variables indicate the group membership for the unobserved variables  $\theta_j$  where the mixing weights are defined as  $\omega_{itj} = \text{pr}(Z_{it} = j)$ . WinBUGS uses Gibbs sampling with necessary Metropolis-Hastings steps to obtain samples from the posterior distributions. The WinBUGS implementation of the SREST model is much easier and straightforward.

#### 4.1 Prior distribution specification and the identifiability condition for the SBPST model

The prior distribution for the spatial dependence between the  $\beta_{ij}$ 's will be specified by a proper multivariate CAR model (Gelfand and Vounatsou 2003) which takes the form  $(\beta_{ij} | \{\beta_{i'j'}, i' \neq i\}, \rho_j, \Sigma_j) \sim \text{MCAR}(\rho_j, \Sigma_j)$ , where  $\Sigma_j$  is a variance-covariance matrix of order  $p \times p$  and  $\rho_j$  is a vector of spatial autocorrelation of order  $p$ . In the case of independent covariates, this multivariate specification will be reduced to the proper univariate CAR model. The prior distributions for spatial autocorrelation and covariance parameters can be simplified by assuming common prior distributions for all mixture components so that information can be pooled over all components. We implemented the latter approach. A standard choice of hyperprior distribution for the precision matrix,  $\Sigma^{-1}$ , is a Wishart distribution with  $p$  degrees of freedom,  $\Sigma^{-1} \sim W(\Omega, p)$ . A common practice is to set  $\Omega$  to a unit matrix (Gelman et al. 2004). The hyperprior distributions for the scale and shape parameters of baseline measure are assumed as, respectively,  $\exp(0.1)$  and  $\text{Gamma}(0.001, 0.001)$  (Ohlssen et al. 2007). The spread parameter  $\alpha$  is assumed to have uniform distribution with the range 0.3–10 (Ohlssen et al. 2007).

We are aware that when the covariates are introduced in the finite mixture model and the parameters are space and component dependent, the issues of label switching and potential overfitting become more delicate. Hennig (2000) showed that the regression parameters are identifiable if the number  $K$  of clusters is smaller than the number of distinct  $(p - 1)$ -dimensional hyperplanes generated by the covariates, i.e., if the covariates show too little variability. We propose to imposing the restriction  $\sum_{i=1}^n \beta_{i..} = C$  as Assunção (2003) suggested, for the Bayesian spatially varying parameter model to ensure identifiability.

#### 4.2 Prior distribution specification for the SREST model

The prior distribution specifications for the structured spatial effect is assigned to be a CAR model with the hyperprior distribution for the precision parameter as  $\text{Gamma}(0.5, 0.0005)$  (Bernardinelli et al. 1995), and for the structured temporal effect is assigned to be a first-order autoregressive model with the hyperprior distributions for the autocorrelation parameter as  $\text{Beta}(1, 1)$  (Lawson et al. 2010) and the temporal standard deviation as  $\text{Uniform}(0, 20)$  (Lambert et al. 2005). All other random effects  $\varepsilon_i$ ,  $\xi_t$  and  $\lambda_{it}$  are assumed to have prior distributions as, respectively,  $N(0, \sigma_\varepsilon^2)$ ,  $N(0, \sigma_\xi^2)$  and  $N(0, \sigma_\lambda^2)$ . A flat prior distribution can be assumed for the overall mean parameter  $\rho$  as  $N(0, k_\rho)$ , where  $k_\rho$  is a large quantity (say, 10,000). All the standard deviation parameters are assumed to have the same prior distribution specifications as  $\text{Uniform}(0, 20)$  (Lambert et al. 2005).

### 5 Simulation design: Ohio County geographies

We conduct a simulation study to assess the performance of the SBPST model in comparison to the SREST model. The aim was to check how these two models perform in



estimating clusters of different shapes and sizes. We use the Ohio lung cancer mortality dataset for our simulation since it is a freely available dataset ([www.stat.uni-muenchen.de/service/datenarchiv/ohio/ohio\\_e.html](http://www.stat.uni-muenchen.de/service/datenarchiv/ohio/ohio_e.html)), and have been analyzed previously in many studies (e.g., Xia and Carlin 1998; Knorr-Held and Besag 1998; Waller et al. 1997).

The state of Ohio in USA has 88 counties. We use the Ohio geographies and, the county and year specific expected lung cancer mortality for the period 1968–1988 as a base for our simulation. The simulation is as described in Hossain and Lawson (2010). In short, the steps of simulation are as follows.

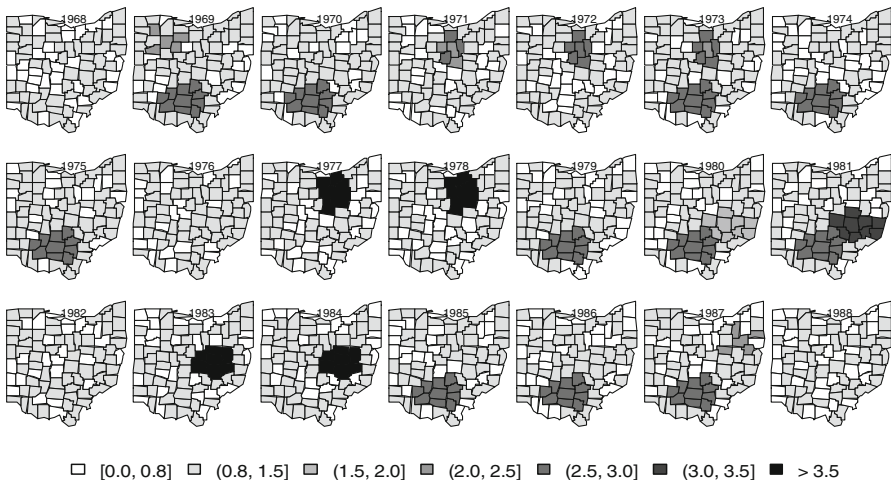
- i. The underlying risks are generated by

$$\mu_{it} \sim \text{Uniform}(0.5, 1.5); i = 1, \dots, 88, \text{ and } t = 1, \dots, 21,$$

since in a normal situation it is expected that the relative risk at each county and year will be close to ‘one’ with some small variation,

- ii. The clusters of different shapes and magnitudes ( $\delta_{it}$ ) are embedded to the desired counties and years to give  $\theta_{it}^{\text{true}} = \mu_{it} + \delta_{it}$ , where  $\theta_{it}^{\text{true}}$  is assumed to be the true value of  $\theta_{it}$ . The values for  $\delta$ 's are chosen in the range of 0.0–3.39 such that the  $\theta^{\text{true}}$  vary in the range of 0.5–4.0, and
- iii. The county and year specific lung cancer mortality count was generated from a Poisson model with:  $y_{it}^s | e_{it} \theta_{it}^{\text{true}} \sim \text{Poisson}(e_{it} \theta_{it}^{\text{true}})$ , where  $s = 1, \dots, S$  is the size of replication.

In our simulation experiment, the number of replications was set at  $S = 100$  and was chosen as a balance between computation time and accuracy. We assume that the size 100 is reasonable to capture all the variations in the replicated dataset. In Fig. 1, the maps for the  $\theta_{it}^{\text{true}}$  are given. The darker colors show the higher risk regions. Some of the clusters are embedded with fairly high values, intended to check model performances for this scenario.



**Fig. 1** Thematic maps of true relative risks of mortality from lung cancer incidences that have been assumed for simulating the datasets for Ohio geographies for the years 1968–1988

## 5.1 SBPST model

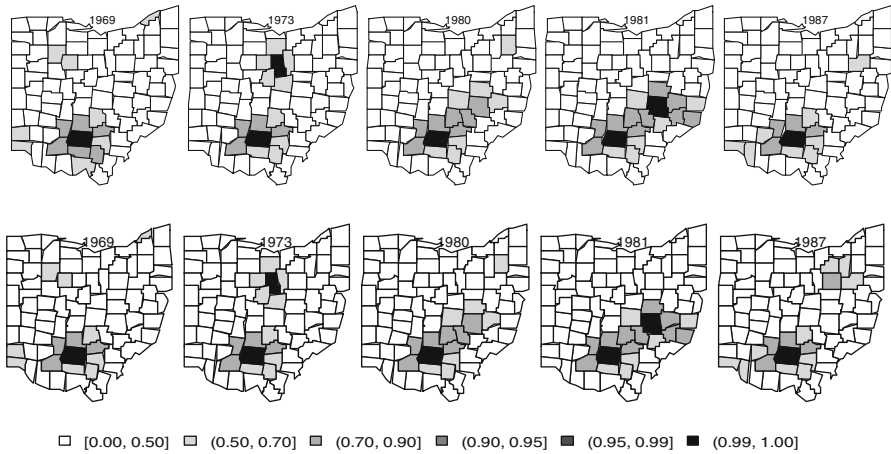
In the assessment of SBPST model for the simulated datasets, we have considered the function,  $h_{itj} = \kappa_{ij} + \gamma_{tj}$ , where the  $\kappa$ 's are the structured spatial random effects,  $\gamma$ 's are the structured temporal random effects.

The structured spatial random effects  $\kappa_j = (\kappa_{1j}, \dots, \kappa_{nj})$  have a CAR prior distribution, specified by:  $\kappa_{ij} | \{\kappa_{-ij}\}, \sigma_{\kappa_j}^2 \sim N(\bar{\kappa}_{ij}, \frac{\sigma_{\kappa_j}^2}{n_i})$ ,  $j = 1, \dots, K$ , where,  $n_i = \sum_{k=1}^n I(k \in \{\Delta_i\})$ ,  $\bar{\kappa}_{ij} = \sum_{i \in \{\Delta_i\}} \frac{\kappa_{ij}}{n_i}$ ,  $\{\Delta_i\}$  is the set of first-order neighbors of the  $i$ th region and  $\{\kappa_{-ij}\}$  is the set of all  $\kappa$ 's excluding  $\kappa_{ij}$ . A vague prior distribution for the spatially hyperparameter  $\sigma_{\kappa_j}^{-2}$  ( $j = 1, \dots, K$ ) is assigned and is specified as Gamma(0.5, 0.0005) (Bernardinelli et al. 1995).

The structured temporal random effects  $\gamma_j = (\gamma_{1j}, \dots, \gamma_{Tj})$  have the first-order autoregressive prior distribution:  $\gamma_{1j} | \chi_j, \sigma_{\gamma_j}^2 \sim N(0, \sigma_{\gamma_j}^2 / (1 - \chi_j^2))$ , and  $\gamma_{tj} | \gamma_{t-1j}, \chi_j, \sigma_{\gamma_j}^2 \sim N(\chi_j \gamma_{t-1j}, \sigma_{\gamma_j}^2)$  for  $t = 2, \dots, T$ , and  $j = 1, \dots, K$ . The first-order autocorrelation coefficient  $\chi_j \in (0, 1)$  and  $\gamma_{tj}$  reduces to temporal independence when  $\chi_j = 0$ . When estimating  $\gamma_{1j}$  ( $j = 1, \dots, K$ ), we restricted  $\chi_j^2$  to be less than one to ensure identifiability. We assign a noninformative prior distribution for  $\chi_j$ , specified as Beta(1, 1) (Lawson et al. 2010). The choice Beta(1, 1) gives the standard uniform distribution. A noninformative prior distribution was assigned for the temporal standard deviation  $\sigma_{\gamma_j}$  ( $j = 1, \dots, K$ ), and is specified as Uniform(0, 20) (Lambert et al. 2005).

## 5.2 Results

For the simulated data where the true excess risk regions are known, it is possible to check the relative performances of SBPST and SREST models in the recovery of clusters. We use previously developed spatial-temporal diagnostic criteria in Hossain and Lawson (2010) for this comparison. The SBPST and SREST models were implemented in WinBUGS. The WinBUGS code developed for the SBPST model is available from the first author on request. The SBPST model was implemented for a finite number of components; here it was set to 7. The number was chosen to balance between the model complexity and computation time. To estimate all the model parameters, we ran two parallel chains with widely different initial values for 15,000 iterations. After discarding the first 10,000 iterations as burn-in samples, 5,000 from each chain were considered for posterior inference. The reported results are the posterior mean over these 10,000 MCMC samples. We checked the Gelman–Rubin statistic, kernel density plot and the trace plot of each parameter to ensure convergence. In addition, we performed a limited sensitivity analysis for the SBPST model for the choice of number of components. We compared the posterior exceedance probability that was obtained for the number of components 7 with that for the number of components 12, only for the first simulated dataset, and observed that the estimates were consistent with these numbers. Hence, the results reported for the SBPST model hereafter are for the number of component, 7. The thematic maps for  $R$  ( $R$  is defined as the average number of realizations where posterior exceedance probability is  $>0.95$ ) for the selected years



**Fig. 2** Thematic maps of  $R$  for the selected years (1969, 1973, 1980, 1981, and 1987) of Ohio simulated data for the threshold value ‘one’. *Top row* stick-breaking process space-time (SBPST) model, and *bottom row* standard random effects space-time (SREST) model

**Table 1** Cluster misspecification rates and mean square error for the stick-breaking process space-time (SBPST) model and the standard random effects space-time (SREST) model for the threshold values,  $c = 1, 2,$  and  $3$

Threshold value ( $c$ )	Cluster misspecification rate		Mean square error	
	SBPST	SREST	SBPST	SREST
1.00	0.013	0.016	0.015	0.014
2.00	0.006	0.011	0.036	0.062
3.00	0.100	0.175	0.021	0.087

1969, 1973, 1980, 1981, and 1987 are given in Fig. 2 where the top row is the result for SBPST model and the bottom rows is for the SREST model. For space limitation we report results only for those years where the clusters vary in shape and size. The posterior exceedance probability (Richardson et al. 2004) was calculated for the threshold value ‘1’, i.e.,  $\sum_{g=1}^G I(\hat{\theta}_{it}^g > 1)/G$ , where  $G$  is the posterior sample size and  $\hat{\theta}_{it}^g$  is the estimated risk for the  $g$ th sample value from converged posterior sampling output. The reported results are averages over the 100 replicates. This threshold level seems to signal the clusters of high intensity well. It is also visible in these maps that these two models seem to correctly signal all the spatio-temporal clusters. For the increasing threshold value (i.e., 2–3), the  $R$  maps appear to become ‘cleaned out’, in that lower risk areas no longer signal (the maps are not included here for space limitation).

Table 1 gives the threshold value specific cluster misspecification rates and the mean square error (MSE) for each model. The threshold values are set at  $c = 1, 2,$  and  $3$ . Choosing these threshold values will help to define three risk levels as: low-risk level when  $c = 1$ , medium-risk level when  $c = 2$  and high-risk level when  $c = 3$ . It appears that in detecting clusters of medium-risk ( $c = 2$ ) and high-risk ( $c = 3$ ), SBPST model performs better than SREST models. In detecting clusters of low-risk ( $c = 1$ ), SREST model performs slightly better than SBPST model.

## 6 South Carolina low birth weight data and results

Low birth weight (LBW) is an important indicator of women's reproductive health and general health status of population (Goldenberg and Culhane 2007). It is defined as baby weighting <2,500 g at live-birth. The causes of LBW may include individual level behavioral and psychosocial factors, neighborhood characteristics, environmental exposures, access to prenatal care, and biological factors. Recent studies on LBW have focused on the effects of neighborhood factors such as environmental health factors, residential segregation, and income inequalities using ecological or multilevel study design (Grady 2006; Janevic et al. 2010). In our current study, we sought to find grouping of areas where rates of LBW incidence are relatively higher by modeling the spatio-temporal patterns of LBW according to SBPST methods while considering space- and time-varying covariates and the potential interactions between the spatio-temporal factors.

The state of South Carolina has 46 counties, and the study period is considered as years 1997–2007. We have considered three models for the county and year specific South Carolina LBW incidences. The models are different based on their specification of the function  $h_{itj}$ . Model 1 considers no covariates, specified only by the structured spatial and structured temporal random effects, i.e.  $h_{itj} = \kappa_{ij} + \gamma_{tj}$ . Model 2 considers covariates, and we consider the county and year specific covariates: population density (PD), proportion of African-American (PAA), median household income (MHI), proportion of poverty (PP), and unemployment rate (UR). These covariates are considered by following the previous studies by Fang et al. (1999) and Pearl et al. (2001). Thus, we use the form,  $h_{itj} = \beta_{0j} + \beta_{1j}PD_{it} + \beta_{2j}PAA_{it} + \beta_{3j}MHI_{it} + \beta_{4j}PR_{it} + \beta_{5j}UR_{it} + \kappa_{ij} + \gamma_{tj}$ . This model assumes that the regression parameters are component specific and has the flexibility to capture any effect of discontinuity or clustering pattern in covariates on LBW. Model 3 considers a general approach by defining a function,  $h_{itj} = \beta_{0ij} + \beta_{1ij}PD_{it} + \beta_{2ij}PAA_{it} + \beta_{3ij}MHI_{it} + \beta_{4ij}PR_{it} + \beta_{5ij}UR_{it} + \beta_{6ij}t$ . This model illustrates a simpler way to model a space-time interaction effect, and specifying a multivariate prior distribution for  $\beta_{ij}$ 's ensures a nonseparable covariance for this effect. Using a multivariate prior distribution for the joint modeling of covariate effects is very practical in a sense that counties with a higher UR are likely to have a higher PP and a lower MHI.

As an aid to model selection we use the deviance information criteria (DIC) (Spiegelhalter et al. 2003) and mean square prediction error (MSPE) (Gelfand and Ghosh 1998). Both of these criteria, DIC and MSPE, are based on the predictive ability of the model while the DIC captures the model complexity by the effective number of parameters. The DIC for model  $M$  is defined as  $DIC_M = \overline{D(\Theta_M)} + p_M$ , where  $\Theta_M$  is the set of all parameters under model  $M$ ,  $\overline{D(\Theta_M)}$  is the posterior mean deviance and  $p_M$  is the effective number of parameters, which is a measure of model complexity. The  $p_M$  is commonly measured by  $\overline{D(\Theta_M)} - D(\hat{\Theta}_M)$ , where  $D(\hat{\Theta}_M)$  is the deviance of the posterior means. In our mixture model, the latent variables to indicate the group membership are discrete, and this membership indicator changes over the MCMC iterations. Thus, it is not clear which group membership information will be used to calculate  $D(\hat{\Theta}_M)$ . Instead, we follow the Gelman et al. (2004) proposal to measure  $p_M$  which is defined as half the posterior variance of the deviance. The MSPE

**Table 2** The deviance information criteria (DIC) and the mean square predictive error (MSPE) values for the three competing SBPST models for the South Carolina low birth weight (LBW) incidences for the years 1997 to 2007

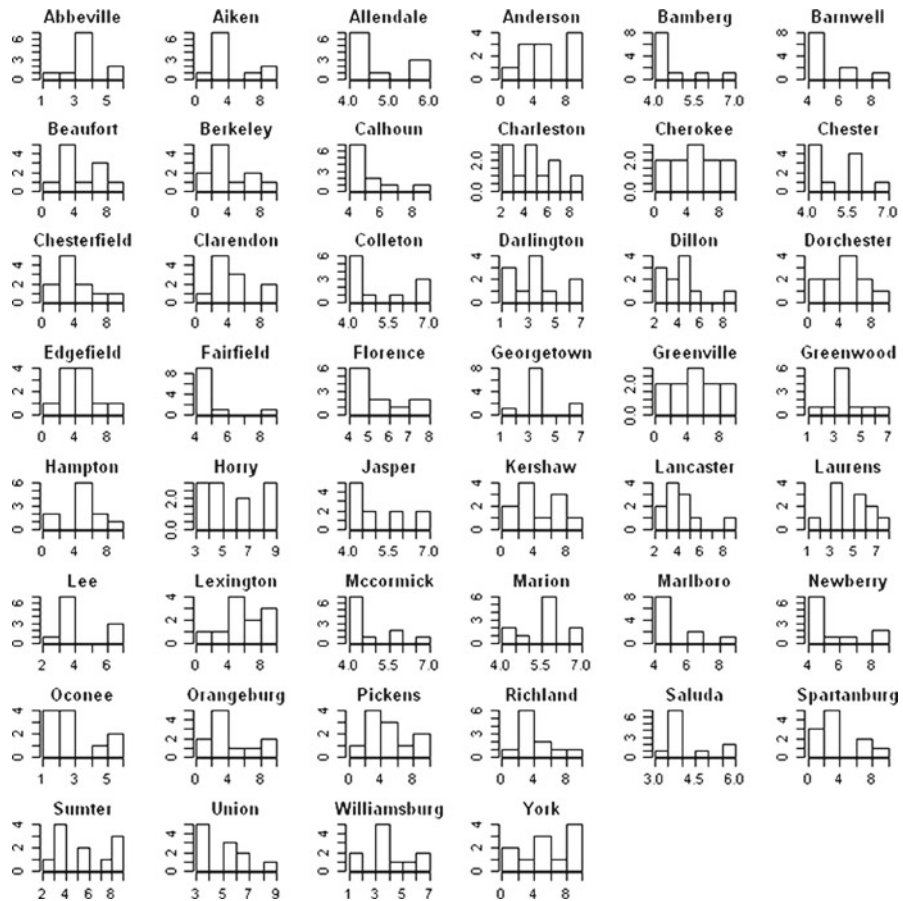
Model	Mean of deviance	Variance of deviance	DIC	MSPE
Model 1	3636.0	27.71	3649.86	224.2
Model 2	3676.0	28.66	3690.33	256.0
Model 3	3664.0	20.76	3674.38	244.6

for model  $M$  is calculated by  $MSPE_M = \sum_{g=1}^G \sum_{i=1}^n \sum_{t=1}^T (O_{it} - O_{it}^{(g, M)})^2 / nTG$ , where  $O_{it}^{(g, M)}$  is the predicted  $O$  for the model  $M$  at  $g$ th iteration and  $G$  is the MCMC sample size. The predicted values are generated from the posterior predictive distribution,  $p(O^{(pred)}|O) = \int p(O^{(pred)}|\Psi)p(\Psi|O)d\Psi$ , after ensuring the convergence of all model parameters,  $\Psi$ . As it appears, MSPE does not separate out any measure for model complexity, although Gelfand and Ghosh (1998) illustrated that for increasing model complexity MSPE increases.

The data sources for this study are: LBW data acquired from the South Carolina Department of Health and Environmental Control (<http://www.scdhec.gov/>), and population, income, poverty and unemployment data are from the US Census Bureau (<http://www.census.gov>). The county and year specific expected LBW incidences,  $E_{it}$ , is calculated by  $n_{it}D$ , where  $n_{it}$  is the total birth in county  $i$  and year  $t$ , and  $D$  is the South Carolina overall LBW rate calculated by the ratio of total LBW to total birth over the entire spatial-temporal domain (Banerjee et al. 2004). The county specific standardized incidence ratios (SIRs) for the years 1997, 2001, 2004, and 2007 are given in Fig. 4 (top row). The observed SIRs are showing a presence of a small cluster of excess risks in the east in the year 2001. This cluster appears again in the year 2004. Although there is no spatial cluster that is persistent over the study period, there are some sporadic regions of excess risks.

### 6.1 Results

The three models as specified in above are fitted in WinBUGS for a fixed number of components. The number of components was set to 10 in order to balance between computational time and model complexity. We preferred to choose the number of components for the real data a little higher than the simulated data since the number of counties for SC is much smaller than the state of Ohio. The posterior inference is based on 10,000 samples from two parallel chains, after a burn-in period of 10,000 samples for each chain. The initial values for each chain were set to widely different values. We ensured the convergence of each estimate by checking the Gelman–Rubin statistic, kernel density plot and the trace plot. The DIC and the MSPE values for these models are reported in Table 2. We have observed similar trends in DIC and MSPE values for the three models specified in above. The smallest values are obtained for Model 1 which included only the structured spatial and structured temporal components. The second smallest values are obtained for Model 3. This model considered a

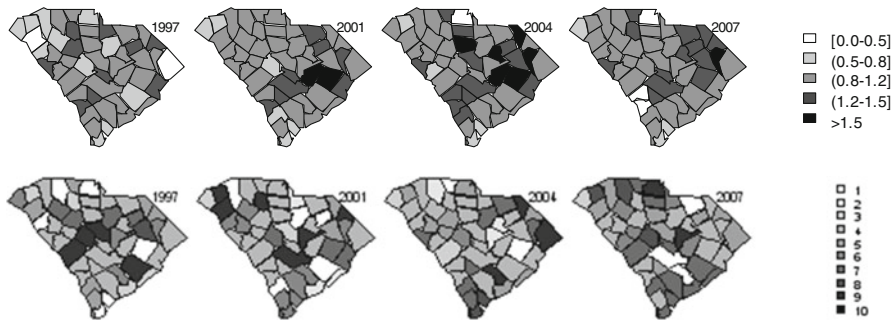


**Fig. 3** County specific histogram plot of group membership labels over the 11 year study period (1997–2007) of South Carolina low birth weight (LBW) incidences from SBPST Model 1

MCAR model as a prior distribution for the spatially varying regression parameters, and also includes a space-time nonseparable effect. The DIC and MSPE values for Model 3 are lower comparing to Model 2 indicates the presence of spatial correlation within the covariates, although this evidence is not substantial since Model 1 has the smallest DIC and MSPE values. In the following we report the results for Model 1.

In cluster analysis, we intend to find the group in which observations are similar to each other than observations in other groups. One way of assigning the appropriate group label to each observation is by calculating the frequency distribution of the latent variable  $Z_{it}$  over all posterior samples, and then assigns the label for which the frequency is maximized. The frequency value for each  $l$  can be calculated by  $n_{it}^{(l)} = \sum_{g=1}^G I(Z_{it}^{(g)} = l)$ , where  $Z_{it}^{(g)}$  is the  $g$ th posterior latent variable  $Z_{it}$  which indicate the group membership,  $l = 1, \dots, K$  and  $G$  is the posterior sample size. The group label for the  $(i, t)$ th observation will be  $l'$  for which  $n_{it}^{(l')} = \max_{l \in K}(n_{it}^{(l)})$ . In Fig. 3, we report the histogram plot of group membership labels over 11 years





**Fig. 4** Thematic maps of standardized incidence ratios (*top row*) and the group membership labels of South Carolina low birth weight (LBW) incidences from SBPST Model 1 (*bottom row*), for the selected years (1997, 2001, 2004, and 2007)

(1997–2007) for each county specific South Carolina LBW incidences. The membership labels are in the range of 1–10, since the number of mixing components  $K$  in the computation was set to 10. It appears that the number of grouping for each county was varying over the 11 year study period and it was varying in the range of 3–6 levels. The maximum number of grouping occurred for the counties Charleston and Greenwood. The thematic maps of this membership label for the years 1997, 2001, 2004 and 2007 are presented in Fig. 4 (bottom). In both maps, top and bottom of Fig. 4, the darker colors may not bear the same meaning. In the top, darker the colors indicate high risk counties and in the bottom, it distinguishes the group membership.

### 7 Conclusions

In this paper, we propose a space-time stick-breaking process for the modeling of incidence count data that are observed at spatially varying locations and for successive time points, and illustrated how this model can be used for finding clusters of high risk areas. We have used the Poisson distribution as an obvious choice of modeling the count data at the first level of hierarchy. The dependencies for spatial and temporal effects are introduced by using space-time covariate dependent kernel stick-breaking processes for the weight component. The space-time dependent kernel can include component specific spatially varying regression coefficients and non-separable space-time effects. The spatially varying regression coefficients are modeled by a MCAR type model.

The proposed model has been extensively validated by using the simulated data and compared with the SREST model by checking each model’s ability in terms of detecting clusters of various shapes and sizes. In generating artificial geo-referenced lung cancer incidences, we have used Ohio geographies and 21 years of county specific expected lung cancer incidences for the years 1968–1988. Clusters of various shapes and sizes are then embedded to the simulated data. The finding is mixed; neither of these two models can be singled out as the best performing model to be recommended for all levels of cluster detection. We have observed that SBPST model performs better than SREST model in detecting medium- and high-risks clusters. The SREST model performs better in detecting low risk clusters. After the validation, we applied the

SBPST model to a real dataset, county specific low birth weight incidences for the state of South Carolina for the years 1997–2007. By using the latent variable labels in posterior samples, we reported the county specific group membership labels over the 11 year study period by histogram plot (Fig. 3). We have observed some similarities of these group membership labels with observed SMR (Fig. 4).

In our current implementation of the SBPST model, we fixed the number of components to a finite number. In our future work we will consider setting this number as an unknown constant and estimate this number by using reversible jump MCMC (RJMCMC) method.

**Acknowledgments** The support of NIH grants UL1 RR024148 (CTSA) and R21 HL088654-01A2 are gratefully acknowledged. Our sincere thanks go to the editor for many constructive comments which contributed to the further improvement of this manuscript.

## References

- Assunção RM (2003) Space varying coefficient models for small area data. *Environmetrics* 14:453–473
- Banerjee S, Carlin B, Gelfand AE (2004) Hierarchical modeling and analysis for spatial data. Chapman and Hall, New York
- Bernardinelli L, Clayton D, Montomoli C (1995) Bayesian estimates of disease maps: how important are priors?. *Stat Med* 14:2411–2431
- Best N, Richardson S, Thomas A (2005) A comparison of Bayesian spatial models for disease mapping. *Stat Methods Med Res* 14:35–59
- Duan JA, Guindani M, Gelfand AE (2007) Generalized spatial Dirichlet process models. *Biometrika* 94: 809–825
- Dunson DB, Park JH (2008) Kernel stick-breaking processes. *Biometrika* 95:307–323
- Fang J, Madhavan S, Alderman MH (1999) Low birthweight: race and maternal nativity-impact of community income. *Pediatrics* 103:E5
- Ferguson TS (1983) Bayesian density estimation by mixtures of normal distributions. In: Rizvi MH, Rustagi JS, Siegmund D (eds) Recent advances in statistics. Academic Press, New York pp 287–302
- Fernandez C, Green PJ (2002) Modelling spatially correlated data via mixtures: a Bayesian approach. *J R Stat Soc Ser B* 64:805–826
- Gelfand AE, Ghosh SK (1998) Model choice: a minimum posterior predictive loss. *Biometrika* 85:1–11
- Gelfand AE, Kottas A, MacEachern SN (2005) Bayesian nonparametric spatial modeling with Dirichlet process mixing. *J Am Stat Assoc* 100:1021–1035
- Gelfand AE, Vounatsou P (2003) Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* 4:11–25
- Gelman A, Carlin JB, Stern HS, Rubin DB (2004) Bayesian data analysis. Chapman & Hall, Boca Raton
- Goldenberg RL, Culhane JF (2007) Low birth weight in the United States. *Am J Clin Nutr* 85:584S–590S
- Grady S (2006) Racial disparities in low birthweight and the contribution of residential segregation: a multilevel analysis. *Soc Sci Med* 63:3013–3029
- Green J, Richardson S (2002) Hidden Markov models and disease mapping. *J Am Stat Assoc* 97:1055–1070
- Green PJ, Richardson S (2001) Modelling heterogeneity with and without the Dirichlet process. *Scand J Stat* 28:355–375
- Griffin JE, Steel MF (2006) Order-based dependent Dirichlet processes. *J Am Stat Assoc* 101:179–194
- Hennig C (2000) Identifiability of models for clusterwise linear regression. *J Classif* 17:273–296
- Hossain MM, Lawson AB (2010) Space-time Bayesian small area disease risk models: development and evaluation with a focus on cluster detection. *Environ Ecol Stat* 17:73–95
- Ishwaran H, James LF (2001) Gibbs sampling methods for stick-breaking priors. *J Am Stat Assoc* 96: 161–173
- Ishwaran H, Zarepour M (2002) Exact and approximate sum representations for the Dirichlet process. *Can J Stat* 30:269–283
- Janevic T, Stein CR, Savitz DA, Kaufman JS, Mason SM, Herring AH (2010) Neighborhood deprivation and adverse birth outcomes among diverse ethnic groups. *Ann Epidemiol* 20:445–451



- Knorr-Held L (2000) Bayesian modelling of inseparable space-time variation in disease risk. *Stat Med* 19:2555–2567
- Knorr-Held L, Besag J (1998) Modelling risk from a disease in time and space. *Stat Med* 17:2045–2060
- Kottas A, Duan JA, Gelfand AE (2007) Modeling disease incidence data with spatial and spatial-temporal Dirichlet process mixtures. *Biom J* 49:1–14
- Lambert PC, Sutton AJ, Burton PR, Abrams KR, Jones DR (2005) How vague is vague? A simulation impact of the use of vague prior distributions in MCMC using winbugs. *Stat Med* 24:2401–2428
- Lawson AB, Song HR, Cai B, Hossain MM, Huang K (2010) Space-time latent component modeling of geo-referenced health data. *Stat Med* 29:2012–2027
- Lo AY (1984) On a class of Bayesian nonparametric estimates: I. Density estimates. *Ann Stat* 12:351–357
- Ohlssen DI, Sharples LD, Spiegelhalter DJ (2007) Flexible random-effects models using Bayesian semi-parametric models: application to institutional comparisons. *Stat Med* 26:2088–2112
- Pearl M, Braveman P, Abrams B (2001) The relationship of neighborhood socioeconomic characteristics to birthweight among 5 ethnic groups in California. *Am J Public Health* 91:1808–1814
- Reich BJ, Fuentes M (2007) A multivariate semiparametric Bayesian spatial modeling framework for hurricane surface wind fields. *Ann Appl Stat* 1:249–264
- Richardson S, Thomas A, Best N, Elliott P (2004) Interpreting posterior relative risk estimates in disease-mapping studies. *Environ Health Perspect* 112:1016–1025
- Sethuraman J (1994) A constructive definition of Dirichlet priors. *Stat Sin* 4:639–650
- Spiegelhalter D, Thomas A, Best N, Lunn D (2003) WinBUGS user manual [1.4.]. MRC Biostatistics Unit, Institute of Public Health, Cambridge
- Waller LA, Carlin BP, Xia H, Gelfand AE (1997) Hierarchical spatio-temporal mapping of disease rates. *J Am Stat Assoc* 92:S.607–S.617
- Xia H, Carlin BP (1998) Spatio-temporal models with errors in covariates: mapping Ohio lung cancer mortality. *Stat Med* 17:2025–2043

## Author Biographies

**Md. Monir Hossain** is an Assistant Professor of Biostatistics in the Center for Clinical and Translational Sciences, University of Texas Health Science Center at Houston, USA. His research interests include developing models for spatial and spatial-temporal data, and developing method for cluster diagnostics.

**Andrew B. Lawson** is a Professor of Biostatistics in the Division of Biostatistics, Bioinformatics and Epidemiology, Medical University of South Carolina at Charleston, USA. He has a research specialization in statistical methods in spatial epidemiology. He has published extensively, including 6 books in the area of spatial epidemiology.

**Bo Cai** is an Assistant Professor of Biostatistics in the Department of Epidemiology and Biostatistics, University of South Carolina at Columbia, USA. He has research interests in Semiparametric modeling, latent variable modeling and spatial modeling. He has published extensively in statistical methodological journals.

**Jungsoon Choi** is a Post-Doctoral Scholar in Biostatistics in the Division of Biostatistics, Bioinformatics and Epidemiology, Medical University of South Carolina at Charleston, USA. Her postdoctoral training is in the area of spatial statistics and latent structure modeling.

**Jihong Liu** is an Associate Professor of Epidemiology in the Department of Epidemiology and Biostatistics, University of South Carolina at Columbia, USA. Her research interests include perinatal epidemiology and reproductive health.

**Russell S. Kirby** is a Professor of Epidemiology in the Department of Community and Family Health, University of South Florida at Tampa, USA. His research specializations include perinatal epidemiology, birth defects and developmental disabilities epidemiology and prevention, GIS and spatial analysis.