Published in final edited form as:

Environ Ecol Stat. 2010 March 1; 17(1): 73–95. doi:10.1007/s10651-008-0102-z.

Space-time Bayesian small area disease risk models: development and evaluation with a focus on cluster detection

Md. Monir Hossain and Andrew B. Lawson

Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, 800 Sumter Street, Columbia, SC 29208, USA

Md. Monir Hossain: hossain@gwm.sc.edu; Andrew B. Lawson: alawson@gwm.sc.edu

Abstract

This paper extends the spatial local-likelihood model and the spatial mixture model to the space-time (ST) domain. For comparison, a standard random effect space-time (SREST) model is examined to allow evaluation of each model's ability in relation to cluster detection. To pursue this evaluation, we use the ST counterparts of spatial cluster detection diagnostics. The proposed criteria are based on posterior estimates (e.g., misclassification rate) and some are based on post-hoc analysis of posterior samples (e.g., exceedance probability). In addition, we examine more conventional model fit criteria including mean square error (MSE). We illustrate the methodology with a real ST dataset, Georgia throat cancer mortality data for the years 1994–2005, and a simulated dataset where different levels and shapes of clusters are embedded. Overall, it is found that conventional SREST models fair well in ST cluster detection and in goodness-of-fit, while for extreme risk detection the local likelihood ST model does best.

Keywords

Cluster; Space-time; Diagnostic; Local likelihood; Mixture; MCMC

1 Introduction

Space-time (ST) modeling of small area health data is a complex subject where spatial, and temporal systems and their interaction must be considered. In this paper we address the development of models that are aimed both at the description of relative risk for a disease of interest in ST, and also the ability of these models to detect unusual aggregations of disease. While these aims could be considered somewhat at odds, there is a growing literature that suggests that cluster detection can be effectively carried out by the use of sampled output from random effect smoothing models (Lawson and Denison 2002). The focus of this paper is the extension to the ST domain of two spatial models that have been proposed for risk estimation and cluster detection: the local likelihood cluster model (LLC) (Kauermann and Opsomer 2003; Hossain and Lawson 2005; Lawson 2006) and the spatial mixture Poisson (MP) model of Fernandez and Green (2002). The ST versions of these models are denoted as LLST and MPST, respectively.

The motivation for the development of spatio-temporal models is the need to provide a parsimonious description of the risk variation over spatial and temporal units when temporal

variation is also available. This scenario is commonly found when using publicly available health data: most public health departments provide temporally-varying geo-referenced count incidence data for diseases. Space-time data potentially involves two confounded dimensions of dependency: between different spatial locations and between different time points. So, any proposal of modeling to this kind of data also needs to address the interdependence between these dimensions. Although accounting for this interdependence increases the computational burden considerably, doing so may improve the identification of areas where risks of disease may be particularly notable. We also wish to develop models that can provide tools or output useful in the detection of unusual aggregations of disease or clusters. While relative risk estimation models employ smoothing components, and these could remove evidence of localized unusual risk aggregations, it is also convenient to consider these models in the context of detection of unusual risk as they can have reasonable detection behavior and they are often relatively easy to fit.

1.1 Cluster detection models and evaluation

Here we briefly describe the background of the extension of the chosen models to the ST domain. For a Gaussian model, Higdon (2007) gave a detail illustration how a ST model can be constructed for spatial and temporal dependencies. Haas (1995) proposed a ST local estimator model for kriging where the spatial and temporal dimensions of the cylinder are defined separately. The temporal range of the cylinder is a user-selected value to ensure approximate second-order stationarity along the temporal dimension.

The spatial LLC model of Hossain and Lawson (2005) appears to provide a novel approach to clustering and emphasizes the local nature of estimation. The model assumes a circular window for each data point whose radius is a lasso parameter. This lasso parameter controls the size of the window and so thence a local relative risk parameter. The data point can be the centroid of a county for count data or a residential location for case event data. The extension that we propose to ST data is based on extending the idea of a circular window for the lasso parameter to a cylindrical window where the base represents space and cylinder height represents time. The optimum values for the radius of the base and height of the cylinder are determined by using the spatial and temporal associations that exist in the data.

Using mixture models to model heterogeneity is a common approach to risk assessment (Richardson and Green 1997; Agarwal et al. 2002). Fernandez and Green (2002) proposed a spatial MP model with a variable number of mixing components which requires the implementation of a reversible jump Markov chain Monte Carlo (MCMC) algorithm for the estimation of all model parameters. The weights in their model are spatially-dependent and are estimated from a Markov random field in order to incorporate spatial neighborhood information. We extend this model to the ST domain by making the weights ST dependent. In the estimation of these weights, we employ a Markov random field to explain the spatial heterogeneity and an autoregressive-type model for temporal heterogeneity.

In addition to these two models, another ST model, a ST extension of the spatial convolution model (Knorr-Held and Besag 1998; Besag et al. 1991) is also used to check the relative performances of each model in cluster detection. As for the estimation procedure for these three models, we employ a Bayesian hierarchical modeling paradigm with joint implementation of Gibbs and Metropolis-Hasting MCMC computational methods to obtain posterior estimates of all model parameters.

The ability of models to detect unusual aggregations of disease incidence (cluster detection) can be assessed in a variety of ways. Here we extend the proposal of Hossain and Lawson (2006) for spatial diagnostics to the ST domain. Some of the criteria are based on functionals of the posterior distribution estimated from posterior sample output, while others are computed

directly from the model fit. The criteria we utilize can be used for single region checking and can be extended to groups of regions based on neighborhood information. In addition to these extensions, some further criteria are proposed based on mis-classification rates.

The remaining sections of this paper are organized as follows: Section 2 proposes the extensions of the LLC model and MP model to space-time data (LLST and MPST) and also presents the specific form of standard random effect space-time (SREST) model that will be used in comparison with the LLST and MPST models. Section 3 explains how the posterior samples can be generated from these three models for posterior inference. Section 4 extends the previously developed cluster diagnostic criteria to its ST analog and also proposes a number of new criteria based on misspecification rate found in simulated data. Section 5 illustrates the simulation of a dataset for Ohio county geographies in which clusters of different shapes and sizes are embedded and the results are reported in a subsection. Section 6 introduces the Georgian throat cancer data and reports the analysis. Section 7 gives the concluding remarks.

2 Space-time models considered in the evaluation

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 $t = 1, \dots, T$. Denote the unit counts as O_{it} to denote the observed count of disease and E_{it} to denote the expected count for area i and period t . The expected count is usually assumed to be fixed and obtained from a standard population after adjustment for case-mix, and that is assumed here.

2.1 Local likelihood space-time (LLST) model

The local LLC assumes a Poisson count data model. Hence, at the first level of hierarchy, we model the observed count as

$$O_{\delta_{it}} | \delta_{it}, \theta_{it} \sim \text{Poisson} (E_{\delta_{it}} \theta_{it}) \quad (2.1)$$

where θ_{it} is an implicit function of δ_{it} , and

$$O_{\delta_{it}} = \sum_{d_{ii'} \leq e^{\kappa_i}, d_{tt'} \leq e^{\gamma_t}} O_{i't'}, \quad E_{\delta_{it}} = \sum_{d_{ii'} \leq e^{\kappa_i}, d_{tt'} \leq e^{\gamma_t}} E_{i't'}, \quad i' = 1, \dots, n \text{ and } t' = 1, \dots, T.$$

The ST lasso can be defined as $\delta_{it} = \{e^{\kappa_i}, e^{\gamma_t}\}$. The Euclidean distance between the centroids of counties i and i' is denoted by $d_{ii'}$ and between years t and t' is denoted by $d_{tt'}$. The spatial and temporal distances are standardized to have the values between $[0, \sqrt{2}]$ and $[0, 1]$, respectively. One of the simplest ways of defining these lasso parameters, κ_i and γ_t , are as spatial and temporal structured random effects. So, the spatial lasso parameter, e^{κ_i} , is the radius and the temporal lasso parameter, e^{γ_t} , is the half-height of (i, t) -th cylinder. The estimates of radius and height are stabilized by the spatial and temporal associations. In other words, the volume of a cylinder at (i, t) -th point is determined by its spatial and temporal association with other neighboring points.

In the local likelihood formulation each Poisson component is assumed to enter independently and so the local likelihood derived from this model is given by:

$$L[O_{\delta} | E_{\delta}, \theta, \delta] = \prod_{i=1}^n \prod_{t=1}^T \left[e^{-E_{\delta_{it}} \theta_{it}} (E_{\delta_{it}} \theta_{it})^{O_{\delta_{it}}} / O_{\delta_{it}}! \right].$$

It is possible to confine the local likelihood model to the direct dependence of the θ_{it} on the (O_{it}, E_{it}) pairs resulting from a given δ_{it} . However, it is natural to specify a model at the second level of the hierarchy for the log relative risk which focuses on the link between the cylinder parameters and the risk. Hence we assume that the logarithm of relative risk is defined as

$$\log(\theta_{it}) = \rho + \kappa_i + \gamma_t + \varepsilon_i + \xi_t + \lambda_{it} \quad (2.2)$$

where ρ is an overall log-scale mean effect, κ_i and γ_t are two structured spatial and temporal random effects (as mentioned above) which control the volume of risk, ε_i and ξ_t are two unstructured spatial and temporal random effects and λ_{it} is a ST interaction random effects. In effect, the relative risk is modified by the scaling $\exp(\kappa_i + \gamma_t)$ which represents a product of the dimension parameters.

Parameter prior specification—To complete the above models (2.1) and (2.2), we need to specify prior and hyper-prior distributions for all the model parameters. First we assume a zero-mean Gaussian prior distribution for the overall rate ρ :

$$\rho \sim N(0, k_\rho),$$

where the variance k_ρ is assumed to be large but known. The other parameters assumed to have zero-mean Gaussian distributions are the temporal effect:

$$\xi_t | \sigma_\xi^2 \sim N(0, \sigma_\xi^2),$$

the unstructured spatial heterogeneity effect

$$\varepsilon_i | \sigma_\varepsilon^2 \sim N(0, \sigma_\varepsilon^2),$$

and the ST interaction effect

$$\lambda_{it} | \sigma_\lambda^2 \sim N(0, \sigma_\lambda^2).$$

Knorr-Held (2000), found that this last assumption was appropriate for the ST disease incidence data he examined. We have also found this to be the case in a range of disease mapping studies. The structured effects have conventional prior distributions often used to reflect localized spatial and temporal correlation (the conditional autoregressive (CAR) model in space and Markov random walk prior distribution in time):

$$\kappa_i | \{\kappa_{-i}\}, \sigma_\kappa^2 \sim N\left(\bar{\kappa}_i, \frac{\sigma_\kappa^2}{n_i}\right),$$

$$\gamma_t | \gamma_{t-1}, \sigma_\gamma^2 \sim N(\gamma_{t-1}, \sigma_\gamma^2),$$

where, $n_i = \sum_{j=1}^n I(j \in \{\Delta_i\})$, $\bar{\kappa}_i = \sum_{j \in \{\Delta_i\}} \frac{\kappa_j}{n_i}$ and $\{\Delta_i\}$ is the set of first-order neighbors of the i th region. The assumed first-order random walk prior distribution for the structured temporal effect, γ_t , leads to a relatively nonparametric trend in time.

At the next level of hierarchy, the choices of hyper-priors for all variance parameters $(\sigma_\kappa^2, \sigma_\gamma^2, \sigma_\varepsilon^2, \sigma_\xi^2, \sigma_\lambda^2)$ are assumed to be inverse gamma distribution: IG (a, b) . The choice of inverse gamma distribution as the prior distribution for the variance parameter is equivalent to choosing a gamma distribution for the precision parameter. The scale and shape parameters for the inverse gamma are chosen so as to have large variance. A similar idea for gamma priors is also recommended in Kelsall and Wakefield (1999). The variance parameter for ρ , k_ρ , is set to a large number, 10,000, to ensure non-informativeness.

The joint posterior distribution of all the variables $(\boldsymbol{\theta}, \rho, \boldsymbol{\kappa}, \boldsymbol{\gamma}, \boldsymbol{\varepsilon}, \boldsymbol{\xi}, \sigma_\kappa, \sigma_\gamma, \sigma_\varepsilon, \sigma_\xi, \sigma_\lambda)$ for the above LLST model is proportional to $L(\mathbf{O}_\delta | \mathbf{E}_\delta, \boldsymbol{\theta}, \boldsymbol{\delta}) p(\boldsymbol{\kappa} | \sigma_\kappa) p(\boldsymbol{\gamma} | \sigma_\gamma) p(\boldsymbol{\varepsilon} | \sigma_\varepsilon) p(\boldsymbol{\xi} | \sigma_\xi) p(\boldsymbol{\lambda} | \sigma_\lambda) p(\sigma_\kappa) p(\sigma_\gamma) p(\sigma_\varepsilon) p(\sigma_\xi) p(\sigma_\lambda) p(\rho)$, where the ST lasso, δ_{it} depends on κ_i and γ_t . We defer discussion of posterior sampling until Sect. 3.

2.2 Space-time with mixture of poissons (MPST)

The observed count for the i th county and t th year, O_{it} , is assumed to follow a mixture of poisson distributions where the total number of mixing components, k , is assumed to be fixed. Mathematically,

$$O_{it} | k, \omega, \lambda \sim \sum_{l=1}^k \omega_{itl} \text{Poisson}(O_{it} | E_{itl} \lambda_l)$$

where the weights are to satisfy two conditions, $0 < \omega_{itl} < 1$ and $\sum_{l=1}^k \omega_{itl} = 1$, and is determined by a hidden (unobserved) allocation variable, $\{z_{it}\}_{i=1, \dots, n, t=1, \dots, T}$ as $\omega_{itl} = p(z_{it} = l)$. A reversible jump (RJ) MCMC will be required when the number of mixing components, k , is a random variable. However, in this paper, we do not explore this possibility. Rather, a goodness-of-fit statistic, deviance information criteria (DIC) (Spiegelhalter et al. 2002) or mean square predictive error (MSPE) (Gelfand and Ghosh 1998), will be used to determine the optimum value for k . When we adopt this strategy, we are aware of the criticism that this approach does not estimate k and other unknown parameters jointly. None-the-less, by using an information criterion to find the value for k , we do ensure the selection of the most parsimonious model. In another paper, the above model will be implemented using variable dimension methods.

Fernandez and Green (2002) illustrated a way to calculate the location dependence weights for the spatial mixture model by using a Markov random field. These structured spatial random effects are then logistic transformed to calculate the weights. For our spatio-temporal mixture model, these ST weights can be generalized to include structured spatial effects, structured temporal effects and non-separable ST interaction effects. So, besides considering a Markov random field for spatial dependence, we can consider an autoregressive type model for temporal dependence and a model that arises as the product of these two (space and time) main effects. Thus the weights are calculated by a logistic transformed function

$$\omega_{itl} = \frac{\exp(f_{itl}/\phi)}{\sum_{l=1}^k \exp(f_{itl}/\phi)}, \quad l=1, \dots, k$$

where $\phi > 0$ is a scaling parameter, which controls the smoothness of the weight function, and f_{itl} is a function of random effect terms. The exact form of f_{itl} can be specified to suit the application. One such example could be $f_{itl} = \kappa_{il} + \gamma_{tl} + \eta_{itl}$ where different terms describe separate space and time effects and non-separable interaction effects. In the application with a real and a simulated datasets, we will consider a much simpler model as $f_{itl} = \kappa_{il} + \gamma_{tl}$, where the κ 's are the structured spatial random effects, γ 's are the structured temporal random effects. The structured spatial random effects $\kappa_l = (\kappa_{1l}, \dots, \kappa_{nl})$ have a CAR prior specification:

$$\kappa_{il} | \{\kappa_{-il}\}, \sigma_{\kappa l}^2 \sim N \left(\bar{\kappa}_{il}, \frac{\sigma_{\kappa l}^2}{n_i} \right), \quad l=1, \dots, k$$

and the structured temporal random effects $\gamma_l = (\gamma_{1l}, \dots, \gamma_{Tl})$ have the first-order autoregressive prior specification:

$$\gamma_{tl} | \gamma_{t-1l}, \sigma_{\gamma l}^2 \sim N \left(\gamma_{t-1l}, \sigma_{\gamma l}^2 \right), \quad l=1, \dots, k.$$

As for defining prior distributions for other unknown parameters in the above mixture model: the allocation variables $z_{11}, \dots, z_{nT}$ are assumed to have a multinomial prior distribution with probability $p(z_{it} = l | k, \omega, \lambda) = \omega_{il}$, for $l = 1, \dots, k$. The ϕ has a uniform prior distribution with the range $(0, \phi_{\max})$, $\phi_{\max}$ is assumed to be a small number (generally, 10). Choosing a uniform prior for ϕ allows the restriction of the estimate for ϕ to a small number by assigning a reasonably narrow range for the uniform distribution, otherwise large values for ϕ (i.e., $\phi \rightarrow \infty$) lead $\omega_{il} = 1/k$ for all i, t and l , indicating no spatial and temporal dependences in the weights. The prior distribution for the λ 's is assumed to have a gamma form and is defined as

$$\lambda_1, \dots, \lambda_k | k, \alpha, \beta \sim k! I(\lambda_1 \leq \dots \leq \lambda_k) \prod_{l=1}^k \frac{\beta^\alpha \lambda_l^{\alpha-1} e^{-\beta \lambda_l}}{\Gamma(\alpha)}.$$

The order restriction in the above gamma distribution is imposed to ensure identifiability of the estimates and a constant, $k!$, arises because of this order restriction. It has been observed for the spatial model in Fernandez and Green (2002) that this order restriction was sufficient to generate distinct posterior samples for each parameter. The hyperparameters α and β are assumed to be known.

Using the conditional independence assumption, the joint posterior distribution of all the variables is proportional to $p(\mathbf{O} | \mathbf{E}, \boldsymbol{\lambda}, k, \boldsymbol{\omega}) p(\boldsymbol{\lambda} | \alpha, \beta, k) p(\mathbf{z} | k, \boldsymbol{\omega}) p(\boldsymbol{\kappa} | \boldsymbol{\sigma}_{\kappa}, k) p(\boldsymbol{\gamma} | \boldsymbol{\sigma}_{\gamma}, k) p(\boldsymbol{\eta} | \boldsymbol{\sigma}_{\eta}, k) p(\boldsymbol{\sigma}_{\kappa} | k) p(\boldsymbol{\sigma}_{\gamma} | k) p(\boldsymbol{\sigma}_{\eta} | k)$, where the likelihood is defined as

$$p(\mathbf{O} | \mathbf{E}, \boldsymbol{\lambda}, k, \boldsymbol{\omega}) = \prod_{i=1}^n \prod_{t=1}^T \sum_{l=1}^k \omega_{il} \text{Poisson}(O_{it} | E_{it}, \lambda_l).$$

Sampling algorithms for this model are discussed in Sect. 3 and Appendix B.

2.3 Standard Random Effect Space-Time (SREST) model

The SREST model is defined by assuming the same Poisson distribution for the observed counts at the first level of hierarchy, given as

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

At the second level of the hierarchy, the logarithm of relative risk, θ_{it} , is defined as

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

This model is similar to a random effect model suggested by Knorr-Held and Besag (1998) for ST data (see also Knorr-Held 2000). To complete the above SREST model specification, the prior distributions are specified as:

$$\rho \sim N(0, k_\rho), k_\rho \text{ is assumed to be known,}$$

$$\kappa_i | \{K_{-i}\}, \sigma_\kappa^2 \sim N\left(\bar{\kappa}_i, \frac{\sigma_\kappa^2}{n_i}\right),$$

$$\gamma_t | \gamma_{t-1}, \sigma_\gamma^2 \sim N(\gamma_{t-1}, \sigma_\gamma^2),$$

$$\varepsilon_i | \sigma_\varepsilon^2 \sim N(0, \sigma_\varepsilon^2),$$

$$\xi_t | \sigma_\xi^2 \sim N(0, \sigma_\xi^2), \text{ and,}$$

$$\lambda_{it} | \sigma_\lambda^2 \sim N(0, \sigma_\lambda^2) .$$

The choice of prior distributions reflects the nature of the effects assumed. The choices of hyper-prior distributions for all variance parameters are as for the LLST model.

The joint posterior distribution of all the variables is proportional to $p(\mathbf{O}|\mathbf{E}, \boldsymbol{\theta}) p(\boldsymbol{\kappa}|\sigma_\kappa) p(\boldsymbol{\gamma}|\sigma_\gamma) p(\boldsymbol{\varepsilon}|\sigma_\varepsilon) p(\boldsymbol{\xi}|\sigma_\xi) p(\boldsymbol{\lambda}|\sigma_\lambda) p(\sigma_\kappa) p(\sigma_\gamma) p(\sigma_\varepsilon) p(\sigma_\xi) p(\sigma_\lambda) p(\rho)$ where the likelihood is defined as

$$p(\mathbf{O}|\mathbf{E}, \boldsymbol{\theta}) = \prod_{i=1}^n \prod_{t=1}^T \text{Poisson}(O_{it}|E_{it}, \theta_{it}) .$$

3 Posterior sampling

The parameters of LLST and SREST models have fixed dimension. The posterior sampling in this fixed dimension model is relatively straightforward (Gamerman and Lopes 2006). Within this hierarchy there is conditional conjugacy with IG prior distributions for Gaussian variances

and with normal prior distributions for overall mean and random effects. The posterior samples for each parameter are generated by an MCMC algorithm with joint implementation of Metropolis-Hastings and Gibbs sampling. The details and specific form of posterior distributions for LLST model are given in Appendix A. For the relatively simpler SREST model, the posterior distributions are similar as for the LLST model (Appendix A) where the pair $(\mathbf{O}_\delta, \mathbf{E}_\delta)$ is replaced by $(\mathbf{O}, \mathbf{E})$.

As for the initial values for the lasso parameters, κ and γ of LLST model, we set these two equal to a large negative number such that the exponentiation of this will be close to zero. The reason of this prior guess, the exponentiation of the estimates of κ and γ should be in the ranges $[0, \sqrt{2}]$ and $[0, 1]$, respectively, since $[0, \sqrt{2}]$ is the range for the distance between two centroids and $[0, 1]$ is for the time points. Thus, estimates close to the lower limits of these ranges imply no association whereas, close to the upper limits imply maximum association. We exploit these arguments in determining the initial values for κ and γ , and set to a value -10 , i.e., a priori we are assuming the observed data contains no cluster.

For MPST model, due to the fixed dimension of the number of mixing components k , we do not have the issue of variable dimensionality for the parameters λ , κ and γ . The implementation details and the posterior distributions for all the parameters are illustrated in Appendix B. Appendix C also presents the rationale for choice of k . We implement the Metropolis-Hastings sampling algorithm when the posterior distributions are known up to proportionality and Gibbs sampling when conditional distributions can be sampled easily.

4 Cluster diagnostics

In order to check the ability of models in the recovery of clusters we propose a set of criteria for ST model where spatial and/or spatio-temporal clusters are likely to be present. Here we assume that a model has been fitted under the Bayesian paradigm and that estimates of relative risk are available in each small area of the study. Alternative non-Bayesian estimates of relative risk could also be examined using some of the diagnostics, but some of our diagnostics address quantities only directly available from posterior sampling. Among these measures that we will propose; the exceedance-based measure is for both simulated and real data; the measures based on misspecification rates, Receiver Operating Characteristic (ROC) curves and mean square errors (MSEs) are for simulated data only.

4.1 Exceedance-based measures

The first diagnostic we consider is the *posterior exceedance probability* (PEP) of the relative risk estimate. The exceedance probability of the sampled θ_{it} can be computed as

$q_{it} = \widehat{\Pr}(\widehat{\theta}_{it} > c) = \frac{1}{G} \sum_{g=1}^G I(\widehat{\theta}_{it}^g > c)$ where $\widehat{\theta}_{it}^g$ is the estimated risk for the g th sample value from converged posterior sampling output, G is the posterior sample size and c is a threshold value. This probability estimate is commonly used to provide evidence of ‘significant’ excess risk within individual areas (see, e.g., Richardson et al. 2004). This can be regarded as a criterion for identifying ‘hot-spot’ clusters. For simulated data, the PEP for the relative risk is given by $q_{its} = \widehat{\Pr}(\widehat{\theta}_{its} > c)$ for s th simulated data. The average PEP (APEP), over the simulated data, is

defined as $\bar{q}_{it} = \frac{1}{S} \sum_s q_{its}$. In case of real data, S is set to one.

We could use PEP to examine single regions. However we could also believe that clustering should have some spatial and/or temporal integrity and so criteria that also examine neighborhoods around points could be useful. Define a set, $\{\bar{q}_{itj}; j = 0, 1, \dots, n_i\}$, of first-order neighbor $\bar{q}$ values of i th region and t th year, where $\bar{q}_{it0}$ is the $\bar{q}$ value of i th region and t th year

itself. A local measure R_i , which was defined previously in Hossain and Lawson (2006) for spatial data, can be extended to include a temporal dimension as

$$R_{it} = \frac{I(\bar{q}_{it} > 0.95) \sum_{j=0}^{n_i} I(\bar{q}_{itj} > 0.95)}{n_i + 1},$$

to calculate the proportion having APEP greater than 0.95 based on the first-order neighbors. The first indicator function in the right hand side of above equation, $I(\bar{q}_{it} > 0.95)$, is to ensure that only the regions of excess risk are used in checking the model ability in cluster detection. The measure R_{it} shows the grouping of excess risk where the average posterior probability of excess risk is greater than 0.95. Note that there is a trade off between the choice of c and the chosen critical probability value (here defined as 0.95). Higher values of c will lead to fewer areas signaling while lower critical probability values will admit more areas. For calculations of R_{it} here we have used $c = 1$.

The above measure R_{it} includes only the spatial (first-order) neighborhood information in order to check the models ability in detecting clusters. This measure can be extended to include preceding year information with an weight, ω , as

$$R_{it,t-1} = \frac{I(\bar{q}_{it} > 0.95) \sum_{j=0}^{n_i} \{ \omega I(\bar{q}_{itj} > 0.95) + (1 - \omega) I(\bar{q}_{it-1,j} > 0.95) \}}{n_i + 1}.$$

The value for the weight ($0 < \omega < 1$) can be chosen depending on the significance of preceding year information in detecting clusters of excess risks; in a typical situation it is equal to 0.5. The measure $R_{it,t-1}$ can be useful to detect spatial clusters since it adds evidence of excess risk from one preceding year with the weight $1 - \omega$. This can also be generalized over longer time periods. We do not pursue this generalization here.

4.2 Misspecification rate

Denote $\theta_{it}^{\text{true}}$ as the true relative risk that was assigned to region i and time t in the data generating process. A set of threshold values in the local relative risk can be defined to allow the definition of a cluster. Hence we can use a risk threshold as a classification tool for groups of regions to be regarded as clusters. The levels we examine here are $d = (1.0, 2.0, 3.0)$, exceeding these values means presence of clusters of excess risks. Choosing these threshold values will help to define three risk levels as: low-risk level when $d = 1$, medium-risk level when $d = 2$ and high-risk level when $d = 3$. An indicator matrix with the element, $Z_{it}^{(d)} = I(\theta_{it}^{\text{true}} \geq d)$ is also defined to show the ST location of clusters for a given threshold value d .

In a similar way, based on the mean posterior estimate $(\bar{\theta})$ that obtained for each model and for a given threshold value d , two indicator matrices $Z_1^{(d)}$ and $Z_0^{(d)}$ for *clusters* and *not-clusters*, respectively, can be defined where the elements of each matrices are

$$Z_{1it}^{(d)} = \begin{cases} 1, & \text{if } \bar{\theta}_{it} \geq d \text{ and } \theta_{it}^{\text{true}} \geq d \\ 0, & \text{if } \bar{\theta}_{it} < d \text{ and } \theta_{it}^{\text{true}} \geq d \end{cases}$$

and

$$Z_{0it}^{(d)} = \begin{cases} 1, & \text{if } \bar{\theta}_{it} < d \text{ and } \theta_{it}^{\text{true}} < d \\ 0, & \text{if } \bar{\theta}_{it} \geq d \text{ and } \theta_{it}^{\text{true}} < d \end{cases}.$$

The elements of $Z_1^{(d)}$ ($Z_0^{(d)}$) will be one if the clusters (not-clusters) for a given threshold value are correctly estimated by a specific model. Two misspecification rates based on these two indicator matrices, $Z_1^{(d)}$ and $Z_0^{(d)}$, can be defined for the purpose of checking each model's ability in detecting true clusters or not-clusters. These are

$$M R_1^{(d)} = 1 - \frac{\sum_{(i,t): \theta_{it}^{\text{true}} \geq d} Z_{1it}^{(d)}}{\sum_{(i,t)} Z_{it}^{(d)}},$$

and,

$$M R_0^{(d)} = 1 - \frac{\sum_{(i,t): \theta_{it}^{\text{true}} < d} Z_{0it}^{(d)}}{\sum_{(i,t)} (1 - Z_{it}^{(d)})}.$$

It is expected that with the increasing value of d , the direction of these two misspecification rates will be reversed.

4.3 ROC curves and AUC

The above misspecification rates are calculated for the numbers based on true clusters (true positive) and true not-clusters (true negative), but do not exploit the information in false clusters (false positive) and false not-clusters (false negative). An ROC curve can be used in this situation which is a plot of the true positive rate (sensitivity) against the false positive rate (1-specificity) for the different possible cut-off of a diagnostic test (see, Hossain and Lawson 2006 for details). In doing so, we first define the true state of "clusters" and "not-clusters" by an indicator function $Z_{it}^{(d)}$ as $I(\bar{\theta}_{it} \geq d)$ for a given value of d . We use the ROC curve (and hence the area under the curve (AUC)) to judge each model's ability in detecting true clusters.

4.4 Mean square error

When the true relative risks are known, using MSE is a common practice to check the adequacy of fit (see e.g., Green and Richardson 2002). The measure is defined as,

$$\text{MSE} = \frac{1}{NT} \sum_{it} \left(\theta_{it}^{\text{true}} - \bar{\theta}_{it} \right)^2,$$

where

$$\bar{\theta}_{it}$$

is the averaged $\hat{\theta}_{it}$ over replicated datasets. This MSE is a global measure, and shows the overall goodness-of-fit for each model over all regions and time periods. The main pitfall attached to the MSE is that it does not show any localized evidence how well the imbedded clusters (or, the true clusters) are estimated by each model. We extend the MSE to pursue this objective, as a goodness-of-fit measure only for the cluster regions for a given threshold value, d . Thus, MSE for clusters is defined as,

$$\text{MSE}^{(d)} = \frac{1}{\sum_{(i,t)} Z_{it}^{(d)}} \sum_{(i,t): \theta_{it}^{\text{true}} \geq d} \left(\theta_{it}^{\text{true}} - \bar{\theta}_{it} \right)^2.$$

In a similar way, the MSE for ‘no cluster’ can also be defined.

5 Simulation design: Ohio County geographies

We adopt the Clark and Lawson (2002) definition of spatio-temporal clusters, in the sense of persistence of clusters over space and time. We have used the 88 counties of the state of Ohio, USA, for a period of 21 years (1968–1988) to illustrate the method, as a base for simulation. This period matches the well-known lung cancer dataset analyzed by various authors (see, e.g., www.stat.unimuenchen.de/service/datenarchiv/ohio/ohio_e.html). The available variables in the lung cancer dataset are observed and expected death counts for i th county and t th year where $i=1, \dots, 88$ and $t=1, \dots, 21$. However, this dataset does not demonstrate strong localized clustering patterns and so we have resorted to the use of the Ohio geographies over a 21-year period with simulated data based on true relative risks which include known clusters. We conducted a simulation experiments for the given Ohio geography where the true risks are assigned in a way to yield spatial and/or spatio-temporal clusters. The simulated data also makes feasible a relative comparison among the LLST, MPST and SREST models, when the true risks are known.

A dataset is simulated to have a large number of clusters of different shapes and magnitudes. The underlying risks, $\{\mu_{it}\}$, are generated randomly from an Uniform (0.5, 1.5) distribution since in a normal background null situation it is expected that the relative risk at each county and year will be close to ‘one’ with some small variation. Then, clusters of different shapes and magnitudes (δ_{it}) are embedded to the desired counties and years to give $\theta_{it}^{\text{true}} = \mu_{it} + \delta_{it}$. In Fig. 1 (top-left), the maps for the embedded spatio-temporal clusters are given. For these clusters, $\theta_{it}^{\text{true}}$ ranges from 1.7 to 4.0, which is fairly high. The simulation study was intended to generate clusters of high risk areas and to check model performances for this scenario.

Given the true relative risks with the clusters added, defined under the model and the expected counts from the real data, we conditionally simulated S sets of observed counts for each county and year from a Poisson distribution with:

$$y_{its} | e_{it}, \theta_{it}^{\text{true}} \sim \text{Poisson}(e_{it} \theta_{it}^{\text{true}}); i=1, \dots, C, t=1, \dots, T \text{ and } s=1, \dots, S$$

where $\theta_{it}^{\text{true}}$ is assumed to be the true value for θ_{it} . This number of replication was fixed at $S=100$ and was chosen to reduce computation time in model fitting, but was deemed large enough to allow averaging to be representative. The results reported are for averages over the S replications.

5.1 Results

The LLST model was implemented in the software environment R (R Development Core Team 2004). The MPST and SREST models were implemented in WinBUGS (Spiegelhalter et al. 2003). The MPST model was implemented for a fixed dimension and Appendix C illustrates the reasoning how the fixed dimension was chosen. Figure 1 displays the true (top-left) and posterior simulation-averaged relative risk maps resulting from fitting the LLST model (top-right), MPST model (bottom-left) and the SREST model (bottom-right). It shows clearly that the estimates of relative risks from each model are very close to the true values. The maps for the averaged PEP ($\bar{q}$) and R are given in left and right columns of Fig. 2. The results reported are for the threshold value $c = 1$ and averages over the 100 replications. This threshold level seems to signal the clusters of high intensity well. It is also visible in these maps that all the models seem correctly signal all the spatio-temporal clusters. For the increasing threshold value, the average PEP and 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 cluster misspecification rates for each model for various threshold values. The no-cluster misspecification rates are within parenthesis. It appears that each model performs differently for each risk level. The MPST model has the lowest misspecification rate in detecting clusters of low-risk ($d = 1$), the SREST has lowest rate for clusters of medium-risk ($d = 2$) and LLST model has the lowest rate for clusters of high-risk ($d = 3$). From this table, it can be concluded that the detection rates of true clusters of low and medium risks by MPST and SREST models are better than LLST model. In case high-risk level clusters, LLST performs better than the other two models.

Figure 3 displays the ROC curve for each model with the corresponding AUC. The top three rows are for the threshold values $d = 1, 2$ and 3 . Left, middle and right columns are for the LLST, MPST and SREST models, respectively. For threshold values 1 and 2 , MPST and SREST models perform better than LLST model. For high threshold value ($d = 3$), the LLST model performs equally well to the MPST and SREST models. The ROC curves in the bottom row are for the PEP. The binary classifications rule of a true cluster is defined as before (Hossain and Lawson 2006): a true status of a ‘cluster’ is defined when the average exceedance probability, $\bar{q}_{it}$ at (i, t) th point, is greater than 0.95 since at this value the evidence of excess relative risk is substantial. The test status is defined for θ_{it}^* based on a cut-off and we call all points above the cut-off *cluster*, and all those below *not cluster*. The AUCs indicate that the LLST model performs slightly better than MPST and SREST models.

Table 2 gives the MSE for various threshold values. For clusters with low and medium risks (i.e., threshold values 1 and 2), SREST model has the lowest MSE and clusters with high risks (i.e., threshold value, 3) LLST model has the lowest MSE. These results are in coherent to the findings from Table 1. The overall MSE is given in the last row and indicates that SREST model has the lowest MSE.

6 Georgia throat cancer data and results

We analyze throat cancer mortality data for the state of Georgia for the years 1994–2005. The state of Georgia has 159 counties. The data is available from the Georgia division of Public health website <http://oasis.state.ga.us/>. In order to fit the models discussed in Sect. 2, we need to calculate the case-mix adjusted expected counts. The expected counts are obtained by the indirect method of standardization. The age-sex distribution of throat cancer death for the year 2000 of US population is considered as a reference population in the expected deaths calculation. The county specific observed SMRs (ratio of observed to expected) for the years 1994–2005 are given in Fig. 4 (top-left). The observed SMRs do not show any indication of persistent clusters of excess risks, but show some evidence of sporadic spatio-temporal clusters,

e.g., one cluster in the south for the year 1994, one cluster each near the center for 1996, 1997 and 1999, and one cluster in the south-east for 2002.

The LLST model was implemented in the software environment R. The MPST and SREST models were implemented in WinBUGS. We check the density and trace plots of posterior samples to ensure the convergence of each model parameters. The posterior mean of relative risk estimates for each county and year are given in Fig. 4: LLST model in top-right, MPST model in bottom-left and SREST model in bottom-right. The number of mixing components for the mixture model was set to three after checking the MSPE, as illustrated for simulated data in Appendix C. It appears that all three models predict the spatio-temporal clusters well. The relative risk estimates from MPST model appears to shrink near one for many counties. We suspect, an excess of zero counts for the throat cancer mortality is the main reason for this shrinkage.

Figure 5 shows the posterior estimates of exceedance probability for the LLST model (top-two rows), the MPST model (middle-two rows) and the SREST model (bottom-two rows). The probabilities are calculated for the threshold value one. It appears that all three models give almost identical signals for the regions of excess risks. There is not much evidence of clusters, group of regions of excess risks. The maps for R_{it} do not signal any clusters (the maps are not included for space limitation). It is also interesting to observe that MPST model performs equally well in comparing with LLST and SREST models in signaling those high risk regions. Overall, analyzing this real data also confirms the findings of the simulation study that the SREST model, in general, performs better, but the LLST model picks up the extreme risk counties well.

7 Conclusions

The primary focus of this paper is the proposed extension of LLST model and MPST model to ST data. The LLST model appears to provide a novel approach to clustering and emphasizes the local nature of estimation. The MCMC implementation is also straightforward. The MPST model is a semiparametric model and can be very useful in cluster analysis as it models the spatial, temporal and spatio-temporal clusters according to the mixing weights. The only difficulty lies in the MCMC implementation which could require a complicated RJ algorithm because of variable dimension of mixing components. Here we have used an approximate profile likelihood approach for inference in this model.

All three models considered in this paper, the LLST, MPST and SREST, included only the random effects corresponding to spatial, temporal and interaction effects to mimic various sources of variability. This is in agreement with the recent suggestion in Best et al. (2005) that cluster detection can be effectively carried out by the use of sampled output from random effects smoothing models. Depending on study objectives, sometimes it may be essential to detect clusters after the adjustment for covariate effects. Adding covariates to these models would not be difficult. For the spatial LLC model, Hossain and Lawson (2005) illustrated two ways of adding covariates. Along these lines, the LLST model can also be extended to include covariates. For the MPST model, the Poisson distribution can be extended to include the covariates, $\mathbf{X}$, as $f(O_{it} | \lambda_l, \beta, E_{it}, \mathbf{X}_{it}) = \text{Poisson} \left(E_{it} e^{\sum_j \beta_j X_{its}} \lambda_l \right)$. For the SREST model, the covariates can be added to calculate the logarithm of relative risk at the second level of hierarchy.

The secondary focus of this paper is the demonstration of the use of ST diagnostics, which are intended to check the models' ability in the recovery of known clusters. Among the proposed diagnostics, R_{it} is based on neighborhood information with the belief that clustering could have spatial integrity. The other proposed diagnostics: misspecification rate, MSE and ROC curve,

are risk specific, i.e., threshold value specific. This makes the model comparison feasible for various risk levels.

In general, the major results of the evaluation of the proposed methods is that we have observed that MPST and SREST models perform better for detecting clusters of low or medium risks, and LLST model performs better for detecting clusters of high risks. The misspecification rates and the MSEs for any of these models are very small. So in real applications, the SREST model may be preferable and is easy to implement and available on standard Bayesian software packages. This certainly supports the widespread use of the SREST model. Because of space limitation, we avoided providing the WinBUGS and R codes that we have developed for this paper, but they can be made available from the first author on request.

Acknowledgments

The support of National Institutes of Health NIH grant # 1 R03 CA125828-01 is gratefully acknowledged. We thank the editor and two anonymous referees for many thorough and constructive suggestions.

Appendix A

The posterior samples for each unknown parameters (θ , ρ , κ , γ , ε , ξ , σ_κ , σ_γ , σ_ε , σ_ξ , σ_λ) of LLST model are generated by the joint implementation of Gibbs and Metropolis-Hasting updating scheme. The parameter sampling schemes are given as:

$$\theta_{it} | \dots \propto \text{Poisson} (O_{\delta_{it}} | E_{\delta_{it}}, \theta_{it}) N \left(\log \theta_{it} | \rho + \kappa_i + \gamma_t + \varepsilon_i + \xi_t, \sigma_\lambda^2 \right),$$

$$\rho | \dots \sim N \left(\rho | \frac{k_\rho \sum_{i=1}^n \sum_{t=1}^T (\log \theta_{it} - \kappa_i - \gamma_t - \varepsilon_i - \xi_t)}{\sigma_\lambda^2 + nTk_\rho}, \frac{k_\rho \sigma_\lambda^2}{\sigma_\lambda^2 + nTk_\rho} \right),$$

$$\kappa_i | \dots \sim N \left(\kappa_i | \frac{\sigma_\lambda^2 \sum_{j \in \Delta_i} \kappa_j + \sigma_\kappa^2 \sum_{t=1}^T (\log \theta_{it} - \rho - \gamma_t - \varepsilon_i - \xi_t)}{n_i \sigma_\lambda^2 + T \sigma_\kappa^2}, \frac{\sigma_\kappa^2 \sigma_\lambda^2}{n_i \sigma_\lambda^2 + T \sigma_\kappa^2} \right),$$

$$\gamma_t | \dots \sim N \left(\gamma_t | \frac{\sigma_\lambda^2 \gamma_{t-1} + \sigma_\gamma^2 \sum_{i=1}^n (\log \theta_{it} - \rho - \kappa_i - \varepsilon_i - \xi_t)}{\sigma_\lambda^2 + n \sigma_\gamma^2}, \frac{\sigma_\gamma^2 \sigma_\lambda^2}{\sigma_\lambda^2 + n \sigma_\gamma^2} \right),$$

$$\varepsilon_i | \dots \sim N \left(\varepsilon_i | \frac{\sigma_\varepsilon^2 \sum_{t=1}^T (\log \theta_{it} - \rho - \kappa_i - \gamma_t - \xi_t)}{\sigma_\lambda^2 + T \sigma_\varepsilon^2}, \frac{\sigma_\varepsilon^2 \sigma_\lambda^2}{\sigma_\lambda^2 + T \sigma_\varepsilon^2} \right),$$

$$\xi_t | \dots \sim N \left(\xi_t | \frac{\sigma_\xi^2 \sum_{i=1}^n (\log \theta_{it} - \rho - \kappa_i - \gamma_t - \varepsilon_i)}{\sigma_\lambda^2 + n\sigma_\xi^2}, \frac{\sigma_\xi^2 \sigma_\lambda^2}{\sigma_\lambda^2 + n\sigma_\xi^2} \right),$$

$$\sigma_\lambda^2 | \dots \sim \text{Inv-Gamma} \left(\sigma_\lambda^2 | a_\lambda + \frac{nT}{2}, b_\lambda + \frac{1}{2} \sum_{i=1}^n \sum_{t=1}^T (\log \theta_{it} - \rho - \kappa_i - \gamma_t - \varepsilon_i - \xi_t)^2 \right),$$

$$\sigma_\kappa^2 | \dots \sim \text{Inv-Gamma} \left(\sigma_\kappa^2 | a_\kappa + \frac{n}{2}, b_\kappa + \frac{1}{2} \sum_{i=1}^n n_i (\kappa_i - \bar{\kappa}_i)^2 \right),$$

$$\sigma_\gamma^2 | \dots \sim \text{Inv-Gamma} \left(\sigma_\gamma^2 | a_\gamma + \frac{T}{2}, b_\gamma + \frac{1}{2} \sum_{t=1}^T (\gamma_t - \gamma_{t-1})^2 \right),$$

$$\sigma_\varepsilon^2 | \dots \sim \text{Inv-Gamma} \left(\sigma_\varepsilon^2 | a_\varepsilon + \frac{n}{2}, b_\varepsilon + \frac{1}{2} \sum_{i=1}^n \varepsilon_i^2 \right),$$

and,

$$\sigma_\xi^2 | \dots \sim \text{Inv-Gamma} \left(\sigma_\xi^2 | a_\xi + \frac{T}{2}, b_\xi + \frac{1}{2} \sum_{t=1}^T \xi_t^2 \right).$$

Candidate values for $(\theta'_{11}, \dots, \theta'_{nT})$ are drawn by a Metropolis-Hasting updating scheme from the posterior for θ_{it} . For all other parameters, samples are drawn directly by Gibbs sampling.

Appendix B

For the MPST model and a given value of k , the conditional distributions are:

$$\kappa_{il} | \dots \propto \prod_{t=1}^T \{ \omega_{itl} \text{Poisson} (O_{it} | E_{it}, \lambda_l) \} N \left(\kappa_{il} | \bar{\kappa}_{il}, \frac{\sigma_{\kappa l}^2}{n_i} \right),$$

$$\gamma_{it} | \dots \propto \prod_{i=1}^n \{ \omega_{itl} \text{Poisson} (O_{it} | E_{it}, \lambda_l) \} N \left(\gamma_{it} | \gamma_{t-1l}, \sigma_{\gamma l}^2 \right),$$

$$\sigma_{kl}^2 | \dots \sim \text{Inv-Gamma} \left(\sigma_{kl}^2 | a_k + \frac{n}{2}, b_k + \frac{1}{2} \sum_{i=1}^n n_i (\kappa_{il} - \bar{\kappa}_{il})^2 \right),$$

$$\sigma_{\gamma_l}^2 | \dots \sim \text{Inv-Gamma} \left(\sigma_{\gamma_l}^2 | a_\gamma + \frac{T}{2}, b_\gamma + \frac{1}{2} \sum_{t=1}^T (\gamma_{tl} - \gamma_{t-1l})^2 \right),$$

$$\phi | \dots \sim \prod_{i=1}^n \prod_{t=1}^T \left\{ \sum_{l=1}^k \omega_{itl} \text{Poisson} (O_{it} | E_{it}, \lambda_l) \right\} I (0 < \phi < \phi_{\max}),$$

$$z_{it} = l | \dots \sim \omega_{itl} \text{Poisson} (O_{it} | E_{it}, \lambda_l) I (l \in \{1, \dots, k\}),$$

and,

$$\lambda | \dots \sim \prod_{l=1}^k \text{Gamma} \left(\lambda_l | \alpha + \sum_{i,t: z_{it}=l} O_{it}, \beta + \sum_{i,t: z_{it}=l} E_{it} \right) I (\lambda_1 \leq \dots \leq \lambda_k).$$

Candidate values for $(\kappa'_{i1}, \dots, \kappa'_{ik}), (\gamma'_{i1}, \dots, \gamma'_{ik})$ and ϕ' are drawn by a Metropolis-Hasting updating scheme from the posterior distributions for $(\kappa_{i1}, \dots, \kappa_{ik}), (\gamma_{i1}, \dots, \gamma_{ik})$ and ϕ , respectively. For all other parameters, samples are drawn directly by Gibbs sampling.

Appendix C

The WinBUGS implementation is based on the idea of profile likelihood where k is fixed to an integer value since k is an integer valued function. The results reported for the MPST model is for $k = 3$ which is chosen based on the mean square prediction error (MSPE). It is calculated by

$$\text{MSPE} = \frac{1}{nTG} \sum_{g=1}^G \sum_{i=1}^n \sum_{t=1}^T (O_{it} - O_{it}^{(g, \text{pred})})^2,$$

where $O_{it}^{(g, \text{pred})}$ is the predicted O for the given estimates of all model parameters at g th iteration and G is the MCMC sample size. The predicted values are generated from the posterior predictive distribution,

$$p(O^{(\text{pred})} | O) = \int p(O^{(\text{pred})} | \beta) p(\beta | O) d\beta$$

where β includes all the model parameters. The predicted values are generated after ensuring the convergence of all model parameters.

The MSPE, like the DIC, includes a penalizing function for model complexity. However, the type of mixture models that we have considered, it appears that the MSPE keeps decreasing even for dimension $k = 5$, although the rates of decreasing reduce remarkably after $k = 3$. The MSPEs for the current simulated data for the number of mixing components 2, 3, 4 and 5 are respectively, 342.4, 167.1, 128.7 and 125.3. Adding a component to $k = 2$ reduces the MSPE by 175.3 whereas increment from 3 to 4 reduces the MSPE by 38.4 and afterwards there is not much substantial decrease. The second reduction may not be justified for the increased model complexity (i.e., the increased number of model parameters for the increase of mixing component). Note that, increase the mixing component by 1 will increase the number of parameters by 110 (one λ , eighty-eight κ 's and twenty-one γ 's).

References

- Agarwal DK, Gelfand AE, Citron-Pousty S. Zero-inflated models with application to spatial count data. *Environ Ecol Stat* 2002;9:341–355. doi:10.1023/A:1020910605990.
- Besag J, York J, Mollie A. Bayesian image restoration with two applications in spatial statistics (with discussion). *Ann Inst Stat Math* 1991;43:1–59. doi:10.1007/BF00116466.
- Best N, Richardson S, Thomas A. A comparison of Bayesian spatial models for disease mapping. *Stat Methods Med Res* 2005;14:35–59. doi:10.1191/0962280205sm388oa. [PubMed: 15690999]
- Clark, AB.; Lawson, AB. Spatio-temporal cluster modelling of small area health data, Ch 14. In: Lawson, AB.; Denison, D., editors. *Spatial cluster modelling*. London: Chapman and Hall; 2002.
- Fernandez C, Green PJ. Modelling spatially correlated data via mixtures: a Bayesian approach. *J R Stat Soc B* 2002;64:805–826. doi:10.1111/1467-9868.00362.
- Gamerman, D.; Lopes, HF. *Markov chain Monte Carlo: stochastic simulation for Bayesian inference*. 2nd edn.. Boca Raton, FL: Chapman & Hall/CRC; 2006.
- Gelfand AE, Ghosh SK. Model choice: a minimum posterior predictive loss. *Biometrika* 1998;85:1–11. doi:10.1093/biomet/85.1.1.
- Green PJ, Richardson S. Hidden Markov models and disease mapping. *J Am Stat Assoc* 2002;97:1–16. doi:10.1198/016214502388618870.
- Haas TA. Local prediction of a spatio-temporal process with an application to wet sulfate deposition. *J Am Stat Assoc* 1995;90:1189–1199. doi:10.2307/2291511.
- Higdon, H. A primer on space-time modeling from a Bayesian perspective, Ch 6. In: Finkenstadt, B.; Held, L.; Isham, V., editors. *Statistical methods for spatio-temporal systems*. Florida: Chapman & Hall/CRC; 2007.
- Hossain MM, Lawson AB. Local likelihood disease clustering: development and evaluation. *Environ Ecol Stat* 2005;12:259–273. doi:10.1007/s10651-005-1512-9.
- Hossain MM, Lawson AB. Cluster detection diagnostics for small area health data: with reference to evaluation of local likelihood models. *Stat Med* 2006;25:771–786. doi:10.1002/sim.2401. [PubMed: 16453370]
- Kauermann G, Opsomer JD. Local likelihood estimation in generalized additive models. *Scand J Stat* 2003;30:317–337. doi:10.1111/1467-9469.00333.
- Kelsall, JE.; Wakefield, JC. Discussion of “Bayesian models for spatially correlated disease and exposure data” by Best et al. In: Bernardo, JM.; Berger, JO.; Dawid, AP.; Smith, AFM., editors. *Bayesian statistics 6*. Oxford: Oxford University Press; 1999. p. 151
- Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. *Stat Med* 2000;19:2555–2567. doi:10.1002/1097-0258(20000915/30)19:17/18<2555::AID-SIM587>3.0.CO;2-#. [PubMed: 10960871]
- Knorr-Held L, Besag J. Modelling risk from a disease in time and space. *Stat Med* 1998;17:2045–2060. doi:10.1002/(SICI)1097-0258(19980930)17:18<2045::AID-SIM943>3.0.CO;2-P. [PubMed: 9789913]
- Lawson AB. Cluster detection: a critique and a Bayesian proposal. *Stat Med* 2006;25:897–916. doi:10.1002/sim.2417. [PubMed: 16453377]

- Lawson, AB.; Denison, DGT. Spatial cluster modeling: an overview, Ch 1. In: Lawson, AB.; Denison, D., editors. Spatial cluster modelling. Boca Raton, FL: Chapman and Hall; 2002.
- R Development Core Team. R Foundation for Statistical Computing. Austria: Vienna; 2004. R: A language and environment for statistical computing.
- Richardson S, Green PJ. On Bayesian analysis of mixtures with an unknown number of components. *J R Stat Soc B* 1997;59(4):731–792. doi:10.1111/1467-9868.00095.
- Richardson S, Thomson A, Best N, Elliott P. Interpreting posterior relative risk estimates in disease-mapping studies. *Environ Health Perspect* 2004;112:1016–1025. [PubMed: 15198922]
- Spiegelhalter D, Best N, Carlin B, Linde A. Bayesian measures of model complexity and fit (with discussion). *J Roy Statist Soc Ser B Methodological* 2002;64:583–639. doi: 10.1111/1467-9868.00353.
- Spiegelhalter, D.; Thomas, A.; Best, N.; Lunn, D. MRC Biostatistics Unit. Cambridge, UK: Institute of Public Health; 2003. WinBUGS user manual.

Biography

Md. Monir Hossain is a Research Assistant Professor of Biostatistics in the Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina. 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 Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina. He has a research specialization in statistical methods in spatial epidemiology. He has published extensively, including 6 books in the area of spatial epidemiology.

Fig. 1.

Thematic maps of true and the posterior estimates of relative risks for the simulated data. Top-left: true relative risks, top-right: LLST model, bottom-left: MPST model, bottom-right: SREST model

Fig. 2. Thematic maps of posterior exceedance probability $\bar{q}$ for threshold value 1 (left column) and R (right column) for simulated data. Top-row: LLST model, middle-row: MPST model and bottom-row: SREST model

Fig. 3. ROC curves and AUC. Top-three rows are for the threshold values $d = 1, 2$ and 3 , respectively, for simulated data. The bottom-row is for the exceeding probability. Left, middle and right columns are respectively, for the LLST, MPST and SREST models

Fig. 4. Thematic maps of observed SMR and the posterior estimates of relative risks of throat cancer for the state of Georgia. Top-left: observed SMR, top-right: LLST model, bottom-left: MPST model, bottom-right: SREST model

Fig. 5. Thematic maps of posterior exceedance probability $\bar{q}$ for threshold value 1 for the state of Georgia. Top-two rows: LLST model, middle-two rows: MPST model and bottom-two rows: SREST model

Table 1Cluster (not-cluster) misspecification rates for each model for a given threshold value, d

Threshold value (d)	LLST model	MPST model	SREST model
1.00	0.038(0.108)	0.010(0.042)	0.016(0.0622)
2.00	0.175(0.0)	0.051(0.0)	0.011(0.0)
3.00	0.125(0.0)	0.200(0.0)	0.175(0.0)

Table 2

Mean square error (MSE) for each model for a given threshold value, d

Threshold value (d)	LLST model	MPST model	SREST model
1.00	0.046	0.020	0.014
2.00	0.192	0.085	0.062
3.00	0.059	0.118	0.087
All	0.036	0.015	0.013