Published in final edited form as:

Environmetrics. 2011 December ; 22(8): 1008–1022. doi:10.1002/env.1127.

Evaluation of Bayesian spatio-temporal latent models in small area health data

Jungsoon Choi^{*,1}, Andrew B. Lawson¹, Bo Cai², and Md. Monir Hossain³

¹Division of Biostatistics and Epidemiology, College of Medicine, Medical University of South Carolina

²Department of Epidemiology and Biostatistics, University of South Carolina

³Center for Clinical and Translational Sciences, The University of Texas

Abstract

Health outcomes are linked to air pollution, demographic, or socioeconomic factors which vary across space and time. Thus, it is often found that relative risks in space-time health data have locally different temporal patterns. In such cases, latent modeling is useful in the disaggregation of risk profiles. In particular, spatio-temporal mixture models can help to isolate spatial clusters each of which has a homogeneous temporal pattern in relative risks. In mixture modeling, various weight structures can be used and two situations can be considered: the number of underlying components is known or unknown. In this paper, we compare spatio-temporal mixture models with different weight structures in both situations. In addition, spatio-temporal Dirichlet process mixture models are compared to them when the number of components is unknown. For comparison, we propose a set of spatial cluster detection diagnostics based on the posterior distribution of the weights. We also develop new accuracy measures to assess the recovery of true relative risks. Based on the simulation study, we examine the performance of various spatio-temporal mixture models in terms of proposed methods and goodness-of-fit measures. We apply our models to a county-level chronic obstructive pulmonary disease data set from the state of Georgia.

Keywords

Spatial cluster; diagnostic; Spatio-temporal mixture model; latent model; Dirichlet process mixture model

1 Introduction

The analysis of relative risk over space and time has received much attention in epidemiology studies over the last decades. Many studies often assume that relative risk is decomposed into several random components and these components explain different risk variations such as temporal effect and spatial effect (Bernardinelli *et al*, 1995; Xia *et al*, 1997; Knorr-Held and Besag, 1998; Knorr-Held, 2000; Mugglin *et al*, 2002; Dreassi *et al*, 2005; Richardson *et al*, 2006; Martinez-Beneito *et al*, 2008; Tzala and Best, 2008). For example, a well-known spatio-temporal random model proposed by Knorr-Held (2000) assumes that the number of cases, y_{ij} in the i th area ($i = 1, \dots, I$) at the j th time point ($j = 1, \dots, J$) has a Poisson distribution as $y_{ij} \sim \text{Pois}(e_{ij}\theta_{ij})$, where e_{ij} is the expected count and θ_{ij} is the relative risk. The log relative risk is specified as $\log \theta_{ij} = \mu + u_i + v_j + \zeta_j + \kappa_j + \xi_{ij}$, where μ is

*choju@musc.edu.

the overall mean parameter, u_i is the correlated spatial component, v_i is the uncorrelated spatial component, ζ_j is the correlated temporal component, κ_j is the uncorrelated temporal component, and ξ_{ij} is the space-time interaction term. In this model, the spatial components explain the overall spatial patterns over time, and the temporal components explain the overall temporal patterns over space. Such components describe global effects not local effects within their spatial and temporal domains. However, it is often found that spatio-temporal health data have several different temporal patterns in risk within the spatio-temporal domain and have a homogeneous temporal profile within a subset of geographical areas. Global models do not allow such disaggregation and they are not appropriate for the estimation of local behaviors in risk. Therefore, it is important to develop a statistical model to disaggregate risk profiles in spatio-temporal health data and then estimate temporal patterns in relative risks and identify the spatial clusters each of which has a homogeneous temporal pattern.

Mixture models provide a flexible way to model heterogeneous risk profiles. Recently, Lawson *et al.* (2010) proposed Bayesian spatio-temporal mixture (STM) models to estimate the underlying temporal patterns of relative risks in spatio-temporal disease data. They also described STM models with entry parameters when the number of temporal components is unknown. They suggested various types of weight structures in STM models and compared them for ambulatory case sensitive asthma data in the 159 counties of Georgia by using goodness-of-fit measures. In mixture models, identifying clusters as well as estimation of latent components could be a major interest, and different weight structures could provide different results in identifying clusters. However, Lawson *et al.* (2010) did not consider the fixed allocation of temporal components in STM models so the performance assessment of STM models was not assessed in terms of clustering methods. Thus the comparison of STM models with various weight structures using cluster detection methods is not only challenging but also essential for their evaluation.

There are several studies on the development of spatial cluster diagnostics in spatial health data analysis. For example, Hossain and Lawson (2006) introduced the cluster diagnostic methods for spatial models based on the residuals and posterior output. Hossain and Lawson (2010) then extended these spatial diagnostics to the spatio-temporal domain. However, cluster methods based on the estimated relative risks can be used to detect the unusual behaviour of relative risks so the use of these methods may not be appropriate in STM models where the focus is an estimation of components of risks.

In this paper, we evaluate STM models in terms of spatial cluster detection and goodness-of-fit criteria in order to investigate the effects of different weight structures. We propose a collection of spatial cluster detection diagnostics based on the posterior distribution of the weights. The spatial detection methods proposed here include individual-region diagnostics and group-of-regions diagnostics based on neighborhood information. The use of these spatial methods is appropriate for the evaluation of spatio-temporal models that have spatial clusters with distant temporal profiles. We suggest risk accuracy measures to assess the closeness of the posterior estimates of relative risks to the true values. Similarly, in the case when the number of latent components is unknown, we explore the performance of STM models with entry parameters using these measures. In addition, since the Dirichlet process mixture (DPM) model (Escobar and West, 1995; Kim *et al.*, 2006; Reich and Bondell, 2010) is a useful tool in cluster analysis, we examine a spatio-temporal Dirichlet process mixture (STDPM) model as a competitor with these STM models. We also study how well these models estimate the true number of components.

The remainder of the paper is organized as follow. In Section 2 we describe STM models with different weight structures and spatio-temporal Dirichlet process mixture models.

Section 3 introduces spatial cluster detection methods for spatio-temporal mixture models, risk accuracy measures, and goodness-of-fit measures. Section 4 presents a simulation study and Section 5 gives the real data analysis and the results. We offer a general discussion in Section 6.

2 Spatio-temporal mixture (STM) models

Following the same notation of the previous section, we assume that the observed count data are available within I small areas and J time periods and the count of disease y_{ij} in the i th area at the j th time period follows a Poisson distribution as $y_{ij} \sim \text{Pois}(e_{ij}\theta_{ij})$, where e_{ij} is the expected count and θ_{ij} is the relative risk. The log relative risk is defined as

$$\log \theta_{ij} = \mathbf{x}_{ij}^T \beta_j + \Lambda_{ij}, \quad (1)$$

where $\mathbf{x}_{ij}^T$ the vector of covariates of area i at time j with the corresponding parameter vector β_j which is time dependent. The mixture component Λ_{ij} accounts for the spatio-temporal variation in the model, and in this paper, we focus on this mixture component.

In order to disaggregate the spatial clusters each of which has a homogeneous temporal pattern in relative risk, we model Λ_{ij} as a linear combination of the underlying temporal components with the spatial weights,

$$\Lambda_{ij} = \alpha_0 + \sum_{l=1}^L w_{il} \chi_{lj}, \quad (2)$$

where α_0 is the intercept and L is the number of the mixing components. For the l th component, χ_{lj} represents the underlying temporal pattern in relative risk and w_{il} represents the corresponding weight at the i th area. Each area has a temporal pattern in relative risks, expressed by the mixture of the temporal components. The weight w_{il} is the proportion of the l th component contribution for area i . Thus, the weights have two conditions: $w_{il} \geq 0$ and $\sum_{l=1}^L w_{il} = 1$. In general, weights could have spatio-temporal dependence, but in this study we focus on the spatially dependent weights to avoid the identification problem of temporal components.

The temporal components χ_{lj} can be defined by various temporal dependency structures. In this paper, we use a Gaussian autoregressive model with order 1, AR(1), for each component, which is a commonly-used temporal structure,

$$\chi_{lj} \sim N(\rho_l \chi_{l,j-1}, \sigma_{\chi_l}^2),$$

where the autoregressive parameter ρ_l ($0 < \rho_l < 1$) and the variance $\sigma_{\chi_l}^2$ are component dependent.

2.1 Specification for the weights when the number of components is known

We consider four different weight structures when the number of components is known. We first have continuous prior distributions for the weights. Due to the additive constraint on the weights, we express w_{il} as

$$w_{il} = \frac{w_{il}^*}{\sum_{l=1}^L w_{il}^*}, \quad (3)$$

where $w_{il}^* > 0$ is the un-normalized weight. We model a Dirichlet prior distribution for the weights w_{il} by using Gamma distributions for w_{il}^* ,

$$w_{il}^* \sim \text{Gamma}(1, 1).$$

This model has no spatial dependency structure and is denoted as Model 1.

We extend Model 1 by introducing a spatial dependency structure in the weights. Model 2 assumes that the un-normalized weight w_{il}^* has a log-normal distribution with the spatially correlated mean α_{il} and the variance $\sigma_{w_l}^2$,

$$w_{il}^* \sim \text{LN}(\alpha_{il}, \sigma_{w_l}^2).$$

To account for the spatial dependency structure of the weights, the multivariate conditional autoregressive (MCAR) distribution would be appropriate for α_{il} (Mardia *et al.*, 1988; Banerjee *et al.*, 2004). In this study, for convenience, we use a multivariate intrinsic autoregressive distribution (Gelfand and Vounatsou, 2003) defined as

$$\alpha_{il} | \alpha_{i'l}, i' \neq i \sim N \left(\frac{1}{n_i} \sum_{i' \neq i} B_{ii'} \alpha_{i'l}, \frac{1}{n_i} \Sigma_{\alpha} \right),$$

where $B_{ii'}$ has the neighbor information: $B_{ii'} = 1$ if area i is adjacent to area i' , and $B_{ii'} = 0$ otherwise. The number of “neighbors” (adjacent areas) of area i is defined as $n_i = \sum_{i' \neq i} B_{ii'}$. The $L \times L$ positive definite matrix Σ_{α} represents the cross-covariance relationships between the different weights. This specification is denoted as MCAR(Σ_{α}).

As an alternative to continuous prior distributions for the weights, a discrete prior distribution that assigns one latent component to a region can be considered. For example, a singular multinomial distribution directly selects one temporal component among all the components based on the probabilities, and the selected component represents the dominant latent component of each region. While the previous models include all of the temporal components with different weight values, STM models with a singular multinomial distribution for the weights include one temporal pattern in relative risks for a given region. In this case, we model w_{il} as a singular multinomial distribution,

$$\mathbf{w}_i = \mathbf{w}_i^* = (w_{i1}^*, \dots, w_{iL}^*)^T \sim \text{Multi}(1; p_{i1}, \dots, p_{iL}), \quad \sum_{l=1}^L p_{il} = 1$$

$$p_{il} = \frac{p_{il}^*}{\sum_{l=1}^L p_{il}^*},$$

where $\mathbf{w}_i = (w_{i1}, \dots, w_{iL})^T$ and w_{il} has a value of 0 or 1. Model 3 also has a Dirichlet prior distribution for p_{il} by assigning Gamma distributions to p_{il}^* ,

$$p_{il}^* \sim \text{Gamma}(1, 1),$$

where p_{il}^* does not have any spatial dependency.

To consider the spatial structure in p_{il}^* , we model a log-normal distribution for p_{il}^* with the spatial mean α_{il} and the variance $\sigma_{p_i}^2$,

$$\begin{aligned} p_{il}^* &\sim \text{LN}(\alpha_{il}, \sigma_{p_i}^2) \\ \alpha_{il} &\sim \text{MCAR}(\Sigma_\alpha), \end{aligned}$$

where α_{il} has a MCAR distribution. This model is denoted as Model 4.

2.2 Specification for the weights when the number of components is unknown

In the previous section, we proposed four STM models with the fixed and known number of components. In general, if L is unknown, the number of components in mixture modeling must be considered to be a parameter and should be estimated. In Bayesian mixture modeling, there are many approaches for the estimation of the number of components. One common approach is to use several Bayesian goodness-of-fit criteria such as the deviance information criterion (DIC; Spiegelhalter *et al.*, 2002), the Bayesian information criterion (BIC), or the Bayes factor when comparing models with different fixed number of components. Based on these criteria, the best model is selected and the number of components is automatically estimated. This method is easy to implement, although defining the range of the number of components considered can be difficult. An alternative approach is to use Dirichlet process mixture model (Escobar and West, 1995; Teh *et al.*, 2006; Reich and Bondell, 2010) which is a mixture with infinite number of components and estimate the number of components based on the posterior distribution. This DPM model is more effective than the previous approaches often used for clustering data. A spatio-temporal DPM model is discussed in Section 2.6.

Lawson *et al.* (2010) also proposed an alternative approach that avoids fixing the number of components and was implemented in a simple way. By using entry parameters (e.g. Dellaportas *et al.*, 2002; Choi *et al.*, 2009), the weight w_{il} is modeled as

$$w_{il} = \frac{\psi_l w_{il}^*}{\sum_{l=1}^L \psi_l w_{il}^*},$$

where L is potentially infinite but it is assumed to be large enough to find the true model and ψ_l is the entry parameter that has a value of 0 or 1. When $\psi_l = 0$, the l th temporal component (χ_{lj}) is not included in the model, and when $\psi_l = 1$, the l th component is included in the model. Following Kuo and Mallick (1998), the entry parameter has a Bernoulli distribution

$$\psi_l \sim \text{Bern}(p_l),$$

where the probability p_l could have a hyperprior distribution or could be a constant. In this study, we assume $p_l = 0.5$ as this is a non-informative value.

2.3 Allocation methods

A post hoc method can be used for the allocation of the components. In the STM models with continuous prior distributions of the weights (Model 1 and 2), we use an allocation method for the estimation of the spatial clusters each of which has a homogeneous temporal pattern in risk, by defining the cluster indicator $Z_i \in \mathbb{N}$ as

$$Z_i = \arg \max_l \{w_{il}\}, \quad (4)$$

where $Z_i (= 1, \dots, L)$ becomes the label index of the temporal component having the highest weight value in the i th area. This suggests that the temporal component with the highest weight value in the i th area is the primary temporal trend of the area in relative risk. With these Z_i values, we can easily identify the spatial groups.

Since a singular multinomial prior distribution for the weights in the STM model directly selects the primary component, the cluster indicator Z_i in Model 3 and 4 becomes the label index of the component having $w_{il} = 1$.

2.4 Bayesian Estimation

In order to conduct Bayesian inference, we first derive the likelihood of the observed data $\mathbf{y}$ as

$$p(\mathbf{y}|\Theta) = \prod_{i=1}^I \prod_{j=1}^J \text{Pois}(y_{ij}|e_{ij}, \alpha_0, w_{il}, \chi_{lj}),$$

where Θ denotes a set of all the parameters of the model. The prior distributions of the intercept parameter and the variance parameters in the model are specified as

$$\alpha_0 \sim N(0, \sigma_{\alpha_0}^2), \quad \sigma_{\alpha_0}, \sigma_{\chi_l}, \sigma_{w_l^*}, \sigma_{p_l^*} \sim \text{Unif}(0, d)$$

where $\sigma_{\alpha_0}^2$ is the variance and d is a constant (Gelman, 2006). We use a Beta prior distribution, $\text{Beta}(1,1)$, for the temporal parameter p_l which is uniform on $(0,1)$. For the $L \times L$ covariance of the MCAR Σ_α , we use an inverse Wishart prior distribution, $\text{Inv-Wishart}((0.01I_L)^{-1}, L)$, where I_L is the identity matrix of size L .

For Model 2, the posterior distribution of all the parameters Θ based on the likelihood and the prior distributions is defined as

$$p(\Theta|\mathbf{y}) = p(\mathbf{y}|\Theta)p(\alpha_0|\sigma_{\alpha_0})p(\mathbf{w}|\sigma_{w^*}, \Sigma_\alpha)p(\chi|\sigma_\chi, \rho)p(\sigma_{\alpha_0})p(\sigma_{w^*})p(\sigma_\chi)p(\rho)p(\Sigma_\alpha),$$

$$\rho = (\rho_1, \dots, \rho_L)^T, \quad \mathbf{w} = (w_{11}, \dots, w_{LL})^T, \quad \chi = (\chi_{11}, \dots, \chi_{LL})^T$$

$$\sigma_{w^*} = (\sigma_{w_1^*}, \dots, \sigma_{w_L^*})^T, \quad \text{and} \quad \sigma_\chi = (\sigma_{\chi_1}, \dots, \sigma_{\chi_L})^T.$$

Posterior distributions of the parameters in the other models can be easily obtained. The estimation of the variance parameters is implemented by Gibbs sampling algorithm. For other parameters, Metropolis adaptive rejection sampling algorithm is implemented as conditional distributions are not easy to draw samples from, in general. Estimates for all the

parameters except Z_i are the posterior means because the posterior mean is the Bayes estimate under the squared error loss function and is commonly used, and it provides a natural interpretation. Since the cluster indicator Z_i is the nominal value, the posterior mode is used for the estimation of Z_i .

2.5 Computational issues and identification

In Bayesian mixture modeling, problems of identifiability of components arise since the likelihood is invariant under permutation of the component labels unless strong prior information is used (Stephens, 2000). In our models, there are two issues to consider. First, identification has influence on the estimation of the latent components. In a spatio-temporal factor analysis, the use of orthogonal components allows for identification but the interpretation of this decomposition is not easy (Wang and Wall 2003; Tzala and Best, 2008). In our STM models, we approach identification by making the assumption that latent components have only time-dependent structures while the corresponding weights have only space-dependent structures. This allows latent components to be naturally interpreted as locally temporal patterns in relative risks. Here, the temporal components have AR(1) dependence and the weights are linked with components with spatial structured distributions. The temporal components can also have stronger temporal correlation, such as AR(2), $\chi_{1j} \sim N(\rho_1 \chi_{1j-1} + \rho_2 \chi_{1j-2}, \sigma_{\chi_1}^2)$, which would allow components to separate well (Lopes *et al.*, 2008; Lawson *et al.*, 2010) However we have found that AR(1) dependence is sufficient in our application.

Second, it is possible for latent components to switch labels during posterior sampling so the averages of Markov Chain Monte Carlo (MCMC) samples of the parameters may be unreasonable estimates of the parameters. Jasra *et al.* (2005) provided a general background to the solutions previously suggested: artificial identifiability constraints (Diebolt and Robert, 1994; Richardson and Green, 1997), relabeling algorithms (Celeux, 1998; Stephens, 2000), label invariant loss functions methods (Celeux *et al.*, 2000; Hurn *et al.*, 2003), and random permutation sampling (Frühwirth-Schnatter, 2001). We have investigated when label switching problems in the STM models occurred and found that there was no label switching during MCMC simulation if a single chain was used. However, when multiple chains were used, different chains had label switched so averaging over chains was inappropriate for the estimation of the latent components. In this case, relabelling methods were required. Thus, in this study, we use a single chain to avoid label switching.

In order to guarantee convergence, a single chain with a total of 70,000 iterations is used in our simulation study and real data analysis. We discard the first 20,000 iterations as burn-in and collect every 10th iteration to obtain 5000 final samples which are used for the estimation of the parameters. We conduct MCMC convergence diagnostics using the Geweke convergence diagnostic (1992), autocorrelation functions, and trace plots. Several representative parameters and the deviance ensure acceptable MCMC convergence.

2.6 Spatio-temporal Dirichlet process mixture (STDPM) model

To provide a comparison with our proposed mixture model we also examine a spatio-temporal Dirichlet process mixture (STDPM) model. This model is commonly used in clustering analysis. We specify the STDPM model as

$$\begin{aligned} \Lambda_i &\sim G_i \\ G_i &\sim \text{DP}(\eta_i, G_0), \end{aligned} \tag{5}$$

where $\Lambda_i = (\Lambda_{i1}, \dots, \Lambda_{iJ})'$ is the vector of spatio-temporal random effects in equation (2) and has a Dirichlet process prior distribution with a scale parameter $\eta_i > 0$ and a base distribution G_0 of dimension J that becomes a multivariate normal distribution with time-dependent covariance matrix. By the stick-breaking construction of Sethuraman (1994), the Dirichlet process G_i can be represented as infinite mixtures of point masses, as $G_i = \sum_{l=1}^{\infty} w_{il} \delta_{\phi_{il}}$, with probability one. The $\delta_{\phi_{il}}$ is the mass probability at point ϕ_{il} , $\phi_{il} \stackrel{iid}{\sim} G_0$, and $\sum_{l=1}^{\infty} w_{il} = 1$. The mixture probability is modeled as $w_{il} = b_{il} \prod_{k=1}^{l-1} (1 - b_{ik}) = b_{il}(1 - b_{i1}) \dots (1 - b_{i,l-1})$ for $l > 1$, $w_{i1} = b_{i1}$, for $l = 1$, and $b_{il} \sim \text{Beta}(1, \eta_i)$.

An equivalent representation of the STDPM model using a cluster indicator $Z_i (= 1, \dots, L)$ is given by

$$\begin{aligned} \Lambda_i | Z_i, \phi_{z_i} &\sim G(\phi_{z_i}) \\ \phi_{z_i} | G_0 &\sim G_0 \\ Z_i &\sim \text{Categorical}(w_{i1}, \dots, w_{iL}), \end{aligned}$$

where G_0 denotes a Gaussian autoregressive model with order 1 so $\phi_{z_i} = (\chi_{z_i1}, \dots, \chi_{z_iJ})^T$. The number of components L should be infinite in the DPM model, but we practically assume that L is a large number in this study. Green and Richardson (2001) proved that for large L when $\eta_i/L \rightarrow 0$, the finite DPM model with the cluster indicators converges to DPM model.

Due to the fixed L , we take $w_{il} = w_{il}^* / \sum_{l=1}^L w_{il}^*$ and w_{il}^* follows the same model as w_{il} has in order to satisfy that $\sum_{l=1}^L w_{il} = 1$.

Since the mean of b_{il} is $E(b_{il}) = 1/(1 + \eta_i)$, a smaller value of η_i makes $E(b_{il})$ larger (near to 1), which produces most of the probabilities on the first few w_{il} 's and estimates a small number of components. On the other hand, a larger value of η_i value makes $E(b_{il})$ smaller (near to 0), which means the probability is dispersed and estimates a large number of

components. For the parameter η_i , a log-normal distribution is considered, $\eta_i \sim \text{LN}(\mu_{\eta_i}, \sigma_{\eta}^2)$, where $\mu_{\eta_i} = \mu_{\eta_0} + \mu_{\eta_i^*}$, μ_{η_0} has zero-mean Gaussian distribution, and $\mu_{\eta_i^*}$ has a univariate CAR prior distribution. The mean parameter of η_i (μ_{η_i}) controls the number of components: a small μ_{η_i} favors a small number of components and a larger μ_{η_i} favors a large number of components. In the STDPM model the estimated number of components is defined as

$\sum_{l=1}^L \mathbf{I}(N_l > 0)$, where $N_l = \sum_{i=1}^I \mathbf{I}(Z_i = l)$ is the number of areas having l th component and $\mathbf{I}(\cdot)$ is a indicator function. This STDPM model is denoted as Model 6 and it can be compared to Models 2 and 4 with entry parameters.

The non-spatial DPM model (Model 5) here can be defined as

$$\begin{aligned} b_{il} &\sim \text{Beta}(1, \eta) \\ \eta &\sim \text{LN}(\mu_{\eta}, \sigma_{\eta}^2) \end{aligned}$$

where μ_{η} follows a normal distribution.

3 Comparison methods

The comparison of Bayesian STM models can be conducted using a variety of criteria. In order to assess the performance of the models in recovering spatial clusters, we propose a range of spatial cluster detection diagnostics which are based on the estimates of cluster

indicators. We also develop several accuracy measures based on the posterior distributions of relative risks to examine the recovery capability of true risks. These proposed measures can be used for simulated data. In addition, we present a number of goodness-of-fit measures and prediction measures, which can be used for both real data and simulated data.

3.1 Cluster diagnostics

We suppose that Z_i^T is the true spatial cluster indicator for the i th area and $\hat{Z}_{ik}$ is the estimated cluster indicator for the i th area at the k th sample, where $k = 1, \dots, K$ and K is the number of simulated data sets. The first criteria we consider is the cluster accuracy rate of the i th area over simulations which is given by

$$A_i = \frac{\sum_{k=1}^K \mathbf{I}(Z_i^T = \hat{Z}_{ik})}{K}.$$

This measure explains how well each model recovers the true spatial cluster of an individual area. The overall cluster accuracy rate is then computed by $\bar{A} = \sum_{i=1}^I A_i / I$, which can be used as a measure of cluster accuracy. We extend this measure to incorporate spatial neighborhood information. The accuracy rate for the neighbor clusters of the i th area is defined as

$$NA_i = \frac{\sum_{k=1}^K \sum_{i' \in \delta_i} \mathbf{I}(Z_{i'}^T = \hat{Z}_{i'k})}{K \cdot n_i},$$

where δ_i is the set of neighbors of the i th area. This measure examines how exactly a model estimate the true spatial clusters of neighbors. In a similar way, the overall neighborhood

accuracy rate is calculated by $\bar{NA} = \sum_{i=1}^I NA_i / I$. Both A_i and NA_i measures show the spatial variation of cluster accuracy rates for individual areas or neighbors.

We also propose new cluster diagnostics for pairwise areas to check the ability of each model in detecting spatial clusters. We consider a binary classification test where the spatial cluster indicators of two different areas are checked for equality. Using both the true pairwise cluster output and the estimated pairwise cluster output, the pairwise accuracy rate is

$$PA = \frac{\sum_{k=1}^K \sum_{i < i'}^I [\mathbf{I}(Z_i^T = Z_{i'}^T) \mathbf{I}(\hat{Z}_{ik} = \hat{Z}_{i'k}) + \mathbf{I}(Z_i^T \neq Z_{i'}^T) \mathbf{I}(\hat{Z}_{ik} \neq \hat{Z}_{i'k})]}{KI(I-1)/2}.$$

In the binary classification test, the pairwise sensitivity is obtained by

$$P_{\text{Sen}} = \frac{\sum_{k=1}^K \sum_{i < i'}^I \mathbf{I}(Z_i^T = Z_{i'}^T) \mathbf{I}(\hat{Z}_{ik} = \hat{Z}_{i'k})}{K \sum_{i < i'}^I \mathbf{I}(Z_i^T = Z_{i'}^T)},$$

and the pairwise specificity is computed by

$$P_{\text{Spe}} = \frac{\sum_{k=1}^K \sum_{i < i'}^I \mathbf{I}(Z_i^T \neq Z_{i'}^T) \mathbf{I}(\widehat{Z}_{ik} \neq \widehat{Z}_{i'k})}{K \sum_{i < i'}^I \mathbf{I}(Z_i^T \neq Z_{i'}^T)}.$$

The pairwise sensitivity and the pairwise specificity are calculated based on the assumption that the true clusters of pairwise areas are equal and they are unequal, respectively. Thus, these measures are useful tools to investigate the cluster recovering performance for pairwise areas under the assumption that the true clusters of pairwise areas are equal or not. The pairwise accuracy rate is the overall accuracy measure for the pairwise areas.

3.2 Risk accuracy measures

In order to examine the closeness of posterior estimates for relative risks to the true values, several accuracy measures are proposed here. We define the difference of a true relative risk and its estimate as $d_{ijk} = \widehat{\theta}_{ijk} - \theta_{ijk}^T$, where θ_{ijk}^T is the true relative risk of the i th area and the j th time at the k th sample and its corresponding estimate is $\widehat{\theta}_{ijk}$. A simple measure is the average

of absolute errors for the relative risks defined as $\text{AAE}_{\text{RR}} = \frac{1}{KIJ} \sum_k \sum_{i,j} |d_{ijk}|$. The mean

square error of the relative risks is $\text{MSE}_{\text{RR}} = \frac{1}{KIJ} \sum_k \sum_{i,j} d_{ijk}^2$. Another common measure is the average of absolute relative errors, defined as

$\text{AARE}_{\text{RR}} = \frac{1}{KIJ} \sum_k \sum_{i,j} \left| \frac{d_{ijk}}{\theta_{ijk}^T} \right| = \frac{1}{KIJ} \sum_k \sum_{i,j} \left| \frac{\widehat{\theta}_{ijk} - \theta_{ijk}^T}{\theta_{ijk}^T} \right|$. We introduce an alternative measure to investigate the closeness of the estimated relative risk values to the true values by using a threshold value c ,

$$C_{ij}^{(c)} = \frac{1}{K} \sum_{k=1}^K \mathbf{I} \left(\left| \frac{\widehat{\theta}_{ijk} - \theta_{ijk}^T}{\theta_{ijk}^T} \right| < c \right).$$

This measure is a function of the threshold value c and it shows the proportion that the absolute relative errors are less than a given value c for the i th area and the j th time. Thus, this measure increases with increasing value of c and the measure with the smaller values of c is more useful to evaluate the performance of models. The overall measure over space and

time is $\overline{C}^{(c)} = \sum_{i,j} C_{ij}^{(c)} / (IJ)$ which depends on a threshold value of c . For a fixed threshold value c , the model with larger $\overline{C}^{(c)}$ is considered better. Especially, for a small value c , the model with large $\overline{C}^{(c)}$ performs well in estimating the true relative risks.

3.3 Goodness-of-fit measures to data

In this section, we present a range of measures to assess how well a model fits the data and predicts. Deviance is defined as $D(\Theta) = -2 \log p(y|\Theta)$, where $p(y|\Theta)$ is the likelihood function. The posterior mean of the deviance is $\overline{D(\Theta)} = E_{\Theta}[D(\Theta)]$ and the deviance of the posterior means is $D(\hat{\Theta})$. Based on the deviance the standard DIC is defined as

$$\text{DIC} = \overline{D(\Theta)} + pD,$$

where $\overline{D(\Theta)}$ measures the model fit, and pD represents the effective number of parameters and measures the model complexity. In Spiegelhalter *et al.* (2002), the pD is calculated by $\overline{D(\Theta)} - D(\hat{\Theta})$ so $DIC = 2\overline{D(\Theta)} - D(\hat{\Theta})$. This DIC form is a widely used model assessment criteria but the use of $D(\hat{\Theta})$ often causes a negative value for pD . Thus we consider the DIC* measure (Gelman *et al.*, 2004), which is defined as $DIC^* = \overline{D(\Theta)} + pD^*$, and pD^* is defined as half the posterior variance of the deviance so pD^* is always greater than 0. We also use an alternative DIC measure, DIC_3 (Celeux *et al.*, 2006), which uses a posterior estimate of likelihood instead of $D(\hat{\Theta})$ and is defined as $DIC_3 = \overline{D(\Theta)} + [\overline{D(\Theta)} + 2 \log \hat{p}(y|\Theta)]$. This DIC_3 measure performs well in mixture models and it is easily computed by MCMC algorithms and provides stable and reliable evaluations.

To compare models in terms of the prediction performance, we consider the Marginal Predictive-likelihood (MPL), which is obtained using the Conditional Predictive Ordinate (CPO) (Dey *et al.*, 1997),

$$MPL = \sum_{i,j} \log (CPO_{ij}),$$

where CPO_{ij} is the marginal posterior predictive density of y_{ij} given the data excluding y_{ij} so the CPO represents a cross-validation measure. Thus, the MPL explains a predictive measure for a future replication of the given data. The model with a larger value of MPL provides better model fit (Ibrahim *et al.*, 2001; Congdon, 2005).

Another criterion is the mean square prediction error (MSPE) given by

$$MSPE = \frac{1}{IJ} \sum_{i,j} (y_{ij} - \hat{y}_{ij})^2,$$

where y_{ij} is the observed value and $\hat{y}_{ij}$ is the predicted value of y_{ij} which is generated from the posterior predictive distribution.

4 Simulation Study

We conduct a simulation study to explore the performance of STM models with various weight structures in terms of a range of clustering detection methods, risk accuracy measures and goodness-of-fit measures presented in the previous section. We examine STM models when the number of components is both known and unknown. STDPM models are also compared to STM models when the number of components is unknown.

In the simulation study, we have used the 159 counties of the state of Georgia as a spatial domain. Georgia state has a large number of counties that have regular and similar spatial shapes so diverse designs for spatial clusters can be considered. In addition, this spatial domain is used by our Section 5 data set. Based on the ambulatory case sensitive asthma data analyzed by Lawson *et al.* (2010), we have used the period from 1999 to 2006 (8 years) as a temporal domain and computed the expected counts of this asthma data from the statewide population-based rates by age and gender. The expected counts ranged from 0.05 to 49.73, with a mean of 2.89. Given the spatio-temporal domain and the expected counts, we conducted simulation experiments to compare the four STM models (Model 1 - Model 4) and the two STDPM models (Model 5 - Model 6) introduced in Section 2.

4.1 Situation 1: the known number of components

In order to investigate the performance of STM models with spatial clusters of different sizes and shapes, we consider four different spatial designs for the cluster indicator Z_i and different number of components (Figure 1). Design 1 has $L = 2$ components, and the spatial pattern of the cluster indicator Z_i is defined by the population density. The first group ($Z_i = 1$) is made of counties with high population density (> 100 per square mile) while the second group ($Z_i = 2$) is made of counties with low population density (≤ 100 per square mile). Design 2 and 3 assume that the number of components is $L = 4$, but they have different spatial patterns for Z_i . Design 3 has three components each of which is assigned to two disconnected areas, and an additional component assigned to one connected area. In Design 4, the number of components is assumed to be 6.

For all the designs except Design 3, we generate simulated count y_{ijk} for county i and time j of the k th simulated set from

$$y_{ijk} \sim \text{Pois}(e_{ij}\theta_{ijk}), \quad k=1, \dots, K,$$

where $i = 1, \dots, I (= 159)$, $j = 1, \dots, J (= 8)$, and K is the number of simulated data sets. The true relative risk θ_{ijk} is modeled as a function of a temporal component,

$$\log(\theta_{ijk}) = \alpha_{0k} + \chi_{z_i, j, k}, \quad z_i = 1, \dots, l, \dots, L,$$

where α_{0k} is the intercept parameter that is chosen as an appropriate value in order to guarantee that the average of relative risks is 1 and its range is between 0 and 3.5. Each spatial cluster has the homogeneous temporal component χ_{ijk} that is independently generated as $\chi_{ijk} \sim N(\rho_{lk}\chi_{l, (j-1), k}, 1)$. The temporal parameter ρ_{lk} is independently generated from a uniform distribution with the range (0,1).

To examine the ability of recovering the true relative risks and the true temporal components, Design 3 assumes that simulated data sets have the same relative risk values over simulations but have different counts.

$$\begin{aligned} y_{ijk} &\sim \text{Pois}(e_{ij}\theta_{ij}), \quad k=1, \dots, K \\ \log(\theta_{ij}) &= \alpha_0 + \chi_{z_i, j}, \end{aligned}$$

where α_0 and $\chi_{z_i, j}$ are constant over simulations and generated from the same scheme as the previous one. The maps of the true relative risks in Design 3 are provided in Supplementary Figure 1.

For each design we generate 500 ($= K$) data sets and fit the different models (Model 1–4) of Section 2 with the fixed number of components that is the same number of components in simulated data sets. To implement this study, two software packages R (<http://www.r-project.org/>) and WinBUGS (<http://www.mrc-bsu.cam.ac.uk/bugs>) are used.

To investigate the recovery performance of the models, we need to identify the estimated temporal components $\hat{\chi}_{ij}$ with the true temporal components χ_{ij} . Label switching can cause change to the allocation of components and their labels (e.g., Stephens, 2000; Jasra *et al*, 2005). We re-label the estimated components by using the mean square error

$$\hat{\mathcal{G}} = \arg \min_{\nu} \sum_{l=1}^L \sum_{j=1}^J (\hat{\chi}_{l\nu j} - \chi_{lj})^2,$$

where $\hat{\mathcal{G}}$ is the label set for the estimated temporal components corresponding to the true components.

Table 1 shows the performance of the STM models with different weight structures in 4 designs in terms of the proposed cluster detection methods and the risk accuracy measures. In all designs except Design 3 the spatial models (Model 2 and 4) have higher cluster accuracy rates than the non-spatial models (Model 1 and 3) and cluster measures in Model 4 are slightly higher than Model 2. In Design 3, Model 4 has high cluster accuracy rates and while Model 2 has quite lower cluster accuracy values. In Model 2, the variation of the estimated weight values is smooth because of spatial priors, so the allocation method proposed in Section 2.3 could not perform well in the spatial design having isolated spatial clusters like Design 3. Thus, Model 1 provides better performance than Model 2. On the other hand, Model 4 has singular multinomial prior distributions for the weights even though a spatial prior distribution is considered. Thus, the variation of the estimated weights is not smooth and Model 4 performs well in this case. In terms of the risk accuracy measures, the spatial models have lower values than the non-spatial models and estimate the true relative risks well. We can see no difference between Model 2 and Model 4 in terms of recovering the relative risks. In addition, as the number of components increases, all the cluster detection measures except PA and P_{Spe} decrease and all risk accuracy measures increase. However, PA and P_{Spe} are stable over different number of components so it seems that the pairwise specificity is not influenced by the number of components. Overall, the spatial models are better than the non-spatial models and Model 4 is marginally better than Model 2 in some situations in terms of the cluster detection methods and the risk accuracy measures.

The maps of A_i for Model 4 in all the designs are displayed in Supplementary Figure 2. In these maps, north-west areas and south-east areas in Georgia have high accuracy rates. The maps for NA_i (not presented here) have similar spatial patterns as A_i .

Figure 2 presents the temporal plots of the true latent components and the estimated components with 95% credible intervals from Model 4 in Design 3. This suggests that Model 4 fits the true latent components well when the number of components is known.

In Figure 3, we show the plots of $C^{(c)}$ against the threshold value c for the models. As a threshold value c increases, the plots of $C^{(c)}$ in all the models tend to be similar, but when a value of c is small, the $C^{(c)}$ measure has quite different values depending on the models. For all the designs, Model 2 and 4 have almost same plots of $C^{(c)}$ and larger values of $C^{(c)}$ than the other models when c is small. In particular, Model 1 has the lowest values of $C^{(c)}$ when c is small. These results also demonstrate that the spatial models are better than the non-spatial models in terms of the risk accuracy measure.

Table 2 summarizes the results of model fitting using the average DIC (ADIC), the average DIC* (ADIC*), the average DIC₃ (ADIC₃), the average MPL (AMPL), and the average MSPE (AMSPE) over the simulations. When comparing the models, a model with smaller ADIC, ADIC*, ADIC₃ and AMSPE is better, while a model with larger AMPL is better. For all the designs, Model 2 are slightly better than Model 4 in terms of DIC and MSPE, but Model 4 has small ADIC* and ADIC₃ and large AMPL. Overall, the spatial models are better than the non-spatial models in model fitting. Model 4 performs well in terms of the goodness-of-fit measures to the data.

4.2 Situation 2: the unknown number of components

We consider the situation when the exact number of components is unknown. We compare the performance of the STM models with entry parameters by using the estimated number of components included in the model, the risk accuracy measures and the goodness-of-fit measures to the data. We also investigate the clustering detection diagnostics when the estimated number of components are the same as the true number of components. Here, the DPM models (Model 5 and 6) proposed in Section 2.6 are compared to the STM models. To produce simulated data sets, we use Design 2 for the cluster indicator Z_i with 4 latent components and define the temporal parameters of the components as $\rho = (1, 0.7, 0.4, 0.1)$ to distinguish the components. When fitting the STM models, we use 10 entry parameters which follow an independent Bernoulli distribution with probability 0.5. We also assume that the number of components in DPM models is set to 10. In model fitting, 10 components seem to be sufficient to find the true number of components and reasonable when considering the trade-off between computational time and model complexity.

For the comparison, we perform 200 simulations and include a component in the STM model if the estimated entry parameter is larger than 0.5. In Table 3, we can see that the spatial models perform well based on the estimation of the number of components while the DPM models do not fit well. For Model 1, 9.5% of the simulations only estimates the true number of components exactly and, for Model 3 and the DPM models (Model 5 and 6), none of the simulations estimate the true number of components exactly. It is shown that Model 2 and 4 have 90.5% and 73% of the simulations estimate the exact true number of components, respectively. In estimating the exact number of components, the spatial models are much better than the non-spatial models and the DPM models, and Model 2 is the best model.

In Table 4 we consider a variety of the risk accuracy measures and the goodness-of-fit measures to compare the models. For all the measures except AMSPE, the spatial mixture models are better than the non-spatial mixture models. Since Model 1 and 3 estimate the more number of components than the true number of components, they seem to be overfitting to the data and they have small AMSPE. Even though Model 4 has smaller ADIC and ADIC* than Model 2, Model 2 and Model 4 have similar ADIC₃, AMPL, and AMSPE values, which are more appropriate measures in mixture models. These spatial mixture models also have the smallest risk accuracy values. On the other hand, the non-spatial DPM model (Model 5) is better than the spatial DPM model (Model 6) in terms of the goodness-of-fit measures and the DPM models have large risk accuracy values. Overall, Model 2 and 4 are better than the DPM models and they have similar results for these measures.

Finally, we explore the performance of spatial clustering in the models only using the output when the estimated number of components is equal to the true number of components. Supplementary Table 1 presents how well the entry parameter models detect the clusters. Since Model 3, 5 and 6 have 0% for the estimation of the true number of components, there is no result. It indicates that the STM models have higher accuracy rates for the cluster detection measures than Model 1. Also, Model 2 and 4 provide similar results.

5 Georgia Chronic Obstructive Pulmonary Disease Data Analysis

Chronic obstructive pulmonary disease (COPD) is one of the most common lung diseases in the world and is currently the fourth leading cause of death in the United States (Jemal *et al.*, 2005; Berry and Wise, 2010). The number of patients with physician-diagnosed COPD in the U.S. increased from approximately 7 million in 1980 to 12 million in 2004 (Schiller *et al.*, 2005). The most important risk factors for COPD are tobacco smoking, indoor and outdoor air pollution, and socioeconomic factors, some of which vary with space and time

(Viegi *et al.*, 2006). COPD data could have spatio-temporal variation, which can allow relative risks to have different temporal patterns over space. Thus, we apply STM models to COPD data to investigate the temporal patterns in relative risks. Since the true number of components in real data analysis is generally not known, STM models with entry parameters and DPM models are considered and compared in terms of a range of goodness-of-fit measures. Unfortunately, the cluster detection methods and the risk accuracy measures proposed in Section 3 can not be employed here as we do not know the true spatial clusters and true relative risks.

We analyze county-level COPD data for the year 1999 to 2007 in Georgia, which were obtained from the state health information system OASIS (Georgia Division of Public Health: <http://oasis.state.ga.us/>). There are 159 counties and 9 years of data. The expected counts were calculated by using the internal standardization method (Banerjee *et al.*, 2004). Supplementary Figure 3 displays the maps of the standardized incidence ratios for each year and we can see the spatio-temporal variation of standardized incidence ratios. Especially, north-east areas and south-east areas in Georgia have high standardized incidence ratios of COPD over the years of study. It is an evidence that relative risks have locally different temporal patterns so this data set is appropriate in this study.

In this example, we assume $L = 10$ entry parameters in the STM models and 10 components in the DPM models to balance between computational time and model complexity. In addition, we use a small area data set (159 counties) so $L = 10$ is enough to find the true number of temporal components. Table 5 reports the estimated number of components and the results of goodness-of-fit measures. While Model 1 and 2 estimate 2 components among 10 components, Model 3 and 4 estimate 9 components and Model 5 and 6 estimate 10 components so they seem to be overfitting the data. It appears that Model 2 among the 6 models is the best fit model in terms of DIC*, DIC₃ and MPL. MSPE measure favors Model 1 but Model 2 also has small MSPE. Overall, Model 2 fits the data well and provides good prediction performance.

Figure 4 presents the temporal plots for the selected components in Model 2. Component 1 has a decreasing pattern while Component 2 has a quite stable pattern. Overall, Component 2 has larger relative risks than Component 1 over time. To examine the spatial variation of the weights in this case, the maps of the weights corresponding to the components are presented in the left two maps in Figure 5. Using our allocation method, we can identify the spatial clusters. The right map in Figure 5 shows the map of the cluster indicator Z_i from Model 2 and Atlanta areas are assigned to Component 1 and south-east areas are assigned to Component 2. We conclude that Atlanta areas have a decreasing pattern in relative risk while south-east areas have a stable pattern in risk and larger relative risks than Atlanta areas, which explain the data well (see Supplementary Figure 3).

We refit the DPM models (Model 5 and 6) with 25 components as they estimate the maximum number of components (10) considered. They also estimate 25 components so the DPM models are not appropriate for estimating the number of temporal components. To explore the effect of using the hyperprior specification for the Bernoulli entry probability (p_l) when entry parameters are used, we refit Model 2 with $p_l \sim \text{Beta}(1,1)$, and this model also estimates 2 components. However, this model has larger DIC₃ (12780) and smaller MPL (−6451) than Model 2 and the other measures are similar. It suggests that a model with fixed entry probability $p_l = 0.5$ is better than when the parameter has a hyperprior distribution and there is no effect on the posterior mean number of components. We examine the performance of Model 2 where the temporal components have a random walk Gaussian distribution (i.e. $\chi_{lj} \sim N(\chi_{lj-1}, \sigma_{\chi_l}^2)$) to assess whether the non-stationary assumption of temporal components is reasonable in this data. This model yields larger DIC measures (by

>300 units) and MSPE (475) so a stationary AR(1) distribution is more appropriate for this data. Finally, we refit Model 2 with inverse gamma prior distributions for the variance parameters and they have little effect on the estimated number of components.

6 Conclusion

In this paper, we evaluated spatio-temporal mixture models with different weight structures. When the number of mixture components was unknown, we considered mixture models with entry parameters and Dirichlet process mixture models. We developed a range of spatial cluster detection methods based on the posterior distribution of the weights in order to compare models. We also proposed several risk accuracy measures to examine the recovery of true risk. We used a variety of goodness-of-fit measures to the data in order to compare different mixture models.

The simulation study showed that spatial models perform better than non-spatial models. When the number of components is known, the spatio-temporal mixture model with a singular multinomial prior distribution of the weights performs better than the other models. But, when the exact number of components is unknown, the spatio-temporal mixture model with a spatial continuous prior distribution of the weights estimates the true number of components well and performs well. The mixture models with a singular multinomial prior distribution of the weights and Dirichlet process mixture models seems to be overfitting the data. In our real data analysis, we considered STM models with entry parameters and DP mixture models because the true number of components was unknown. We found that the mixture model with a spatial continuous distribution of the weights performs well, which was consistent with the results obtained from the simulation study.

In this study, we only focus on latent structure and its comparison was made. In many contexts it is appropriate to consider covariate adjustment in the analysis of spatio-temporal small area health data. By adding covariates in spatio-temporal mixture models, we could investigate the performance of STM models with several criteria. We could also extend the univariate spatio-temporal mixture models to the multivariate spatio-temporal mixture models.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by NIH R21 HL088654-01A2.

References

- Banerjee, S.; Carlin, BP.; Gelfand, AE. Hierarchical modeling and analysis for spatial data. New York: Chapman and Hall; 2004.
- Bernardinelli L, Clayton DG, Pascutto C, Montomoli C, Ghislandi M, Songini M. Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*. 1995; 14:2433–2443. [PubMed: 8711279]
- Berry CE, Wise RA. Mortality in COPD: Causes, Risk Factors, and Prevention. *Journal of Chronic Obstructive Pulmonary Disease*. 2010; 7:375–382. [PubMed: 20854053]
- Celeux, G. Bayesian inference for mixtures: The label switching problem. In: Payne, R.; Green, PJ., editors. *COMP-STAT 98 - Proceedings in Computational Statistics*. Physica, Heidelberg: 1998. p. 227-232.

- Celeux G, Forbes F, Robert C, Titterton M. Deviance information criteria for missing data models. *Bayesian Analysis*. 2006; 1:651–674.
- Celeux G, Hurn M, Robert CP. Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*. 2000; 95:957–970.
- Choi J, Fuentes M, Reich BJ. spatio-temporal association between fine particulate matter and daily mortality. *Computational Statistics and Data Analysis*. 2009; 53:2989–3000. [PubMed: 19652691]
- Congdon, P. *Bayesian Models for Categorical Data*. New York: John Wiley and Sons; 2005.
- Dellaportas P, Forster J, Ntzoufras I. On Bayesian model and variable selection using MCMC. *Statistics and Computing*. 2002; 12:27–36.
- Dey D, Chen MH, Chang H. Bayesian approach for nonlinear random effects models. *Biometrics*. 1997; 53:1239–1252.
- Diebolt J, Robert CP. Estimation of finite mixture distributions through Bayesian sampling. *Journal of the Royal Statistical Society B*. 1994; 56:363–375.
- Dreassi E, Biggeri A, Catelan D. Space-time models with time-dependent covariates for the analysis of the temporal lag between socioeconomic factors and lung cancer mortality. *Statistics in Medicine*. 2005; 24:1919–1932. [PubMed: 15724269]
- Escobar MD, West M. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*. 1995; 90:577–588.
- Frühwirth-Schnatter S. Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*. 2001; 96:194–209.
- Gelfand AE, Vounatsou P. Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*. 2003; 4:11–25. [PubMed: 12925327]
- Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*. 2006; 1:515–533.
- Gelman, A.; Carlin, HS.; Rubin, DB. *Bayesian data analysis*. Boca Raton: Chapman and Hall/CRC; 2004.
- Geweke, J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. Bernardo, JM.; Berger, JO.; Dawid, AP.; Smith, AFM., editors. Oxford, UK: Oxford University Press; 1992. In *Bayesian Statistics 4*
- Green J, Richardson S. Hidden Markov models and disease mapping. *Journal of the American Statistical Association*. 2002; 97:1055–1070.
- Hossain MM, Lawson AB. Cluster detection diagnostics for small area health data: With reference to evaluation of local likelihood models. *Statistics in Medicine*. 2006; 25:771–786. [PubMed: 16453370]
- Hossain MM, Lawson AB. Space-time Bayesian small area disease risk models: development and evaluation with a focus on cluster detection. *Environmental and Ecological Statistics*. 2010; 17:73–95. [PubMed: 20953293]
- Hurn M, Justel A, Robert CP. Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics*. 2003; 12:55–79.
- Ibrahim, J.; Chen, MH.; Sinha, D. *Bayesian Survival Analysis*. New York: Springer; 2001.
- Jasra A, Holmes CC, Stephens DA. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*. 2005; 20:50–67.
- Jemal A, Ward E, Hao Y, Thun M. Trends in the leading causes of death in the United States, 1970–2002. *Journal of the American Medical Association*. 2005; 294:1255–1259. [PubMed: 16160134]
- Kim S, Tadesse MG, Vannucci M. Variable selection in clustering via Dirichlet process mixture models. *Biometrika*. 2006; 93:877–893.
- Knorr-Held L. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*. 2000; 19:2555–2567. [PubMed: 10960871]
- Knorr-Held L, Besag J. Modelling risk from a disease in time and space. *Statistics in Medicine*. 1998; 17:2045–2060. [PubMed: 9789913]
- Kuo L, Mallick B. Variable selection for regression models. *Sankhya B*. 1998; 60:65–81.
- Lawson AB, Song HR, Cai B, Hossain MM, Huang K. Space-time latent component modeling of geo-referenced health data. *Statistics in Medicine*. 2010

- Lopes H, Salazar E, Gamerman D. Spatial dynamic factor analysis. *Bayesian Analysis*. 2008; 3:759–792.
- Mardia KV, Goodall C, Redfern EJ, Alonso FJ. The kriged Kalman filter (with discussion). *Test*. 1998; 7:217–285.
- Martinez-Beneito MA, Lopez-Quilez A, Botella-Rocamora P. An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine*. 2008; 27:2874–2889. [PubMed: 17979141]
- Mugglin AS, Cressie N, Gemmell I. Hierarchical statistical modelling of influenza epidemic dynamics in space and time. *Statistics in Medicine*. 2002; 21:2703–2721. [PubMed: 12228886]
- Reich BJ, Bondell HD. A spatial Dirichlet process mixture model for clustering population genetics data. *Biometrics*. 2010
- Richardson S, Green PJ. On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society B*. 1997; 59:731–792.
- Richardson S, Abellan J, Best N. Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (U.K.). *Statistical Methods in Medical Research*. 2006; 15:97–118.
- Schiller JS, Adams PF, Nelson ZC. Summary health statistics for the U.S. population: National Health Interview Survey, 2003. *Vital and Health statistics Series*. 2005; 10:1–104.
- Spiegelhalter DJ, Best N, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society B*. 2002; 64:583–639.
- Sethuraman J. A constructive definition of Dirichlet priors. *Statistica Sinica*. 1994; 4:639–650.
- Stephens M. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society B*. 2000; 62:795–809.
- Teh YW, Jordan MI, Beal MJ, David MB. Hierarchical Dirichlet Processes. *Journal of the American Statistical Association*. 2006; 101:1566–1581.
- Tzala T, Best N. Bayesian latent variable modelling of multivariate spatio-temporal variation in cancer mortality. *Statistical Methods in Medical Research*. 2008; 17:97–118. [PubMed: 17855747]
- Viegi G, Maio S, Pistelli F, Baldacci S, Carrozzi L. Epidemiology of chronic obstructive pulmonary disease: Health effects of air pollution. *Respirology*. 2006; 11:523–532. [PubMed: 16916323]
- Wang F, Wall M. Generalized common spatial factor model. *Biostatistics*. 2003; 4:569–582. [PubMed: 14557112]
- Xia H, Carlin BP, Waller LA. Hierarchical models for mapping Ohio lung cancer rates. *Environmetrics*. 1997; 8:107–120.

Figure 1.
Spatial designs of the cluster indicator (Z_i) in the simulation study.

Figure 2. Plots of the true temporal components and estimates from Model 4 in Design 3. The solid line is the true component, the dashed line is the average of the posterior estimated component, and the dotted lines are 95% intervals for the posterior estimated component.

Figure 3.
Plots of $\bar{C}^{(c)}$ against the threshold value c .

Figure 4.
Temporal plots for 2 estimated components from Model 2 in COPD data.

Figure 5. Maps of the estimated weights corresponding with the components from Model 2 and allocation results.

Table 1

Diagnostics using cluster detection measures and risk accuracy measures when the number of components is known.

Design	Model	Cluster detection measures				Risk accuracy measures			
		$\bar{\kappa}$	$\overline{NA}$	PA	P_{Sen}	P_{Spe}	AAE_{RR}	MSE_{RR}	$AARE_{RR}$
1 ($L=2$)	1	0.88	0.85	0.76	0.74	0.80	0.22	0.13	0.21
	2	0.91	0.88	0.80	0.78	0.84	0.17	0.09	0.17
	3	0.89	0.87	0.79	0.76	0.82	0.16	0.10	0.17
	4	0.91	0.88	0.81	0.79	0.84	0.15	0.09	0.16
2 ($L=4$)	1	0.77	0.78	0.82	0.64	0.88	0.33	0.23	0.42
	2	0.80	0.81	0.87	0.84	0.88	0.20	0.10	0.23
	3	0.80	0.81	0.83	0.66	0.89	0.25	0.17	0.33
	4	0.86	0.87	0.88	0.78	0.92	0.20	0.12	0.26
3 ($L=4$)	1	0.75	0.75	0.76	0.52	0.88	0.32	0.22	0.36
	2	0.58	0.58	0.71	0.59	0.77	0.24	0.13	0.26
	3	0.78	0.79	0.78	0.59	0.87	0.26	0.18	0.30
	4	0.83	0.84	0.82	0.69	0.88	0.23	0.15	0.26
4 ($L=6$)	1	0.59	0.59	0.81	0.52	0.87	0.39	0.30	0.53
	2	0.59	0.59	0.81	0.67	0.84	0.30	0.19	0.38
	3	0.69	0.70	0.84	0.54	0.90	0.31	0.23	0.42
	4	0.71	0.72	0.84	0.60	0.89	0.30	0.22	0.41

Table 2

Diagnostics using goodness-of-fit measures when the number of components is known

Design	Model	ADIC(ApD)	ADIC*(ApD*)	ADIC ₃	AMPL	AMSPE
(L = 2)	1	3732(110)	3767(145)	3706	-1877	6.46
	2	3663(44)	3712(92)	3678	-1855	6.47
	3	3690(62)	3690(62)	3677	-1855	6.54
	4	3689(61)	3689(61)	3675	-1852	6.54
(L = 4)	1	3824(166)	4046(387)	3806	-1964	7.33
	2	3679(52)	3843(216)	3722	-1886	7.30
	3	3817(163)	3817(163)	3746	-1908	7.47
	4	3770(125)	3770(125)	3721	-1884	7.46
(L = 4)	1	3833(200)	4009(377)	3783	-1948	6.29
	2	3730(116)	3849(236)	3721	-1889	6.26
	3	3802(169)	3802(169)	3732	-1903	6.39
	4	3766(141)	3766(141)	3714	-1887	6.37
(L = 6)	1	3560(-111)	4250(579)	3852	-2013	6.96
	2	3761(118)	4047(404)	3780	-1937	6.95
	3	3910(252)	3910(252)	3784	-1944	7.18
	4	3889(236)	3889(236)	3773	-1933	7.17

Table 3

Frequency table of the number of components included in the model. The true number of components is 4 and 200 simulated data sets are used.

Model	1	2	3	4	5	6	7	8	9	10
1			1	19	101	70	9			
2		3	13	181	3					
3				2	2	5	20	37	134	
4			3	146	39	11	1			
5					3	39	65	57	36	
6					1	5	9	17	41	127

Table 4

Diagnostics using risk accuracy measures and model fitting for the entry parameter models.

Model	AAE _{RR}	MSE _{RR}	AARE _{RR}	ADIC*(ApD*)	ADIC ₃	AMPL	AMSPE
1	0.34	0.23	0.43	4093(460)	3791	-1963	7.22
2	0.18	0.08	0.21	4077(456)	3707	-1879	7.40
3	0.29	0.20	0.38	4024(426)	3740	-1925	7.17
4	0.18	0.09	0.22	3790(177)	3698	-1869	7.41
5	0.27	0.19	0.35	3943(329)	3737	-1910	7.37
6	0.28	0.19	0.37	4537(893)	3758	-1921	7.61

Table 5

Diagnostic results with 10 entry parameters for COPD data.

Model	$\mathcal{L}$	DIC*(pD*)	DIC ₃	MPL	MSPE
1	2	12579(219)	12788	-6456	437
2	2	12534(193)	12764	-6447	438
3	9	13202(133)	13550	-6873	493
4	9	13225(139)	13596	-6892	488
5	10	13232(162)	13677	-7006	491
6	10	13216(156)	13609	-6932	488