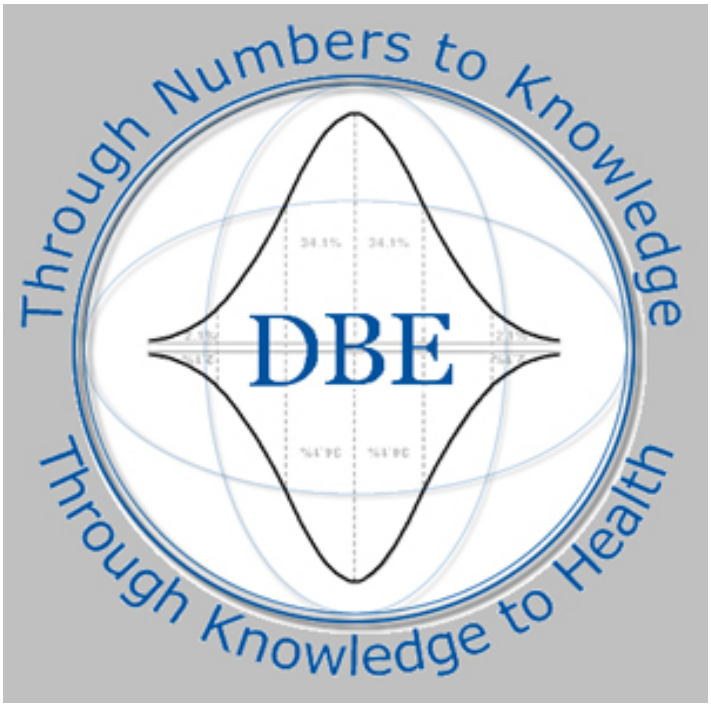
2011

Evaluation of Bayesian Spatial-Temporal Latent Models in Small Area Health Data

Jungsoon Choi
Medical University of South Carolina

Andrew B. Lawson
Medical University of South Carolina



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

Jungsoon Choi[1,*], Andrew B. Lawson[1], Bo Cai[2], and Md. Monir Hossain[3]

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

[2]Department of Epidemiology and Biostatistics, University of South Carolina

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

[*]email: choju@musc.edu

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 spatial health data have locally different patterns. In such cases, latent modeling is useful in the disaggregation of risk profiles. In particular, spatial-temporal mixture models can help to isolate spatial clusters each of which has a homogeneous temporal pattern in relative risks. Mixture models are assumed as they have various weight structures and considered in two situations: the number of underlying components is known or unknown. In this paper, we compare spatial-temporal mixture models with different weight structures in both situations. For comparison, we propose a set of spatial cluster detection diagnostics which are based on the posterior distribution of weights. We also develop new accuracy measures to assess the recovery of true relative risk. Based on the simulation study, we examine the performance of various spatial-temporal mixture

models in terms of proposed methods and goodness-of-fit measures. We examine two real data sets: low birth weight data and chronic obstructive pulmonary disease data.

Key words: Spatial cluster; diagnostic; Spatial temporal mixture model; latent model; small area health

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). In general, such components describe global effects not local effects within their spatial and temporal domains. For instance, a spatial random component explains the overall spatial pattern over time periods, and a temporal random component explains the overall temporal pattern over areas. However, it is often found that spatial-temporal health data have several different temporal patterns in risk within their spatial-temporal domain and have a homogeneous temporal profile within 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 spatial-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 a Bayesian spatial-temporal mixture (STM) model to estimate the underlying temporal patterns of relative risks in spatial-temporal disease data. They also described STM models with entry parameters when the number of temporal com-

ponents is unknown. They developed STM models with various types of weight priors for the latent components and compared these models 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 allocation of components in STM models and the performance of STM models in terms of clustering methods. Thus the comparison of STM models with various weight structures by 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 the posterior output. Hossain and Lawson (2010) then extended these spatial diagnostics to the spatial-temporal methods, which are based on the estimated relative risks. However, these cluster methods are used to detect the unusual behaviour of relative risks so the use of these methods may not be appropriate in STM models.

In this paper, we evaluate various 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 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 spatial-temporal models that have spatial clusters which have distant temporal profiles. We present risk accuracy measures to assess the closeness of posterior estimates of relative risks to the true values. Similarly, in the case when the number of components is unknown, we explore the performance of STM models with entry parameters by using these measures. 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. Section 3 introduces spatial cluster detection methods for spatial-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 Models

We assume that the observed count data are available within I small areas and J time periods. Denote the count of disease in the ith area at the jth time period as y_{ij}, where $i = 1, \cdots, I$ and $j = 1, \cdots, J$. We make the conventional assumption that y_{ij} follows a Poisson distribution as

$$y_{ij} \sim Pois(e_{ij}\theta_{ij}),$$

where e_{ij} is the observed expected count and θ_{ij} is the relative risk. The log relative risk is defined as

$$\log \theta_{ij} = \mathbf{x}'_{ij}\boldsymbol{\beta}_j + \Lambda_{ij}, \tag{1}$$

where $\mathbf{x}'_{ij}$ is the vector of covariates of area i at time j with the corresponding parameter vector $\boldsymbol{\beta}_j$ which is time dependent. The mixture component Λ_{ij} accounts for the spatial-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,

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

where α_0 is the intercept and L is the number of the temporal components. For the lth component, χ_{lj} represents the underlying temporal pattern in relative risk and w_{il} represents the corresponding weight at the ith area. Each area has a temporal pattern in relative risks expressed by the mixture of temporal components. The weight w_{il} is the proportion of lth component contribution for area i. Thus, weights have two conditions: $w_{il} \geq 0$ and $\sum_{l=1}^{L} w_{il} = 1$. Here, weights can generally be spatial-temporal random effects, but in this study we focus on the spatially dependent weights because of 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 for each component, which is a commonly-used temporal structure,

$$\chi_{lj} \sim N(\rho_l \chi_{lj-1}, \sigma^2_{\chi_l}),$$

where the temporal parameter ρ_l $(0 < \rho_l < 1)$ and the variance $\sigma^2_{\chi_l}$ change with components.

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

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

$$w_{il} = \frac{w^*_{il}}{\sum_{l=1}^{L} w^*_{il}}, \tag{3}$$

where $w^*_{il} > 0$ is the un-normalized weight. We model a Dirichlet prior distribution for the weights w_{il} by using a Gamma distribution in the un-normalized weights 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 adding 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 $\sigma^2_{w_l^*}$,

$$w^*_{il} \sim \mathrm{LN}(\alpha_{il}, \sigma^2_{w_l^*}).$$

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

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

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 ith area is defined as $n_i = \sum_{i'\neq i} B_{ii'}$. The $L \times L$ positive definite matrix $\mathbf{\Sigma_\alpha}$ represents the cross-covariance relationships between the different weights. This specification is denoted as MCAR($\mathbf{\Sigma_\alpha}$).

As an alternative to continuous prior distributions for the weights, a discrete prior distribution that assigns one latent component to a region can be assumed. For example, a singular multinomial distribution directly allocates one temporal component among all the components based on the probabilities. This distribution easily allows a flat classification of regions into components and the selected component represents the dominant latent component of each region. While the previous models include all temporal components with different weight values, STM models with a singular multinomial distribution for weights

include one temporal pattern in relative risks for a region. We model w_{il} as a singular multinomial distribution,

$$\begin{aligned} w_{il} = w^*_{il} &\sim \text{Multi}(1, p_{il}), \qquad \sum_{l=1}^{L} p_{il} = 1 \\ p_{il} &= \frac{p^*_{il}}{\sum_{l=1}^{L} p^*_{il}}, \end{aligned}$$

where w_{il} has a value of 0 or 1. Model 3 also has a Dirichlet prior distribution for p_{il} by assigning a Gamma distribution for p^*_{il},

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

where p^*_{il} does not involve any spatial dependency.

To enable the spatial dependency effects to p^*_{il}, we model a log-normal distribution for p^*_{il} with the spatial mean α_{il} and the variance $\sigma^2_{p^*_l}$,

$$\begin{aligned} p^*_{il} &\sim \text{LN}(\alpha_{il}, \sigma^2_{p^*_l}) \\ \alpha_{il} &\sim \text{MCAR}(\boldsymbol{\Sigma}_{\boldsymbol{\alpha}}), \end{aligned}$$

where α_{il} has a MCAR distribution. This 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 the mixture model 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 simple, but defining the range of the number of components considered can be difficult. An alternative approach is to use reversible jump MCMC (RJMCMC; Green, 1995) which estimates the number of components based on the posterior distribution. RJMCMC is more effective than the previous approaches, but it has the problem of computational cost and complexity.

Lawson *et al.* (2010) proposed an alternative approach that avoids fixing the number of components and is simply implemented. By using entry parameters (e.g. Dellaportas *et al.*, 2002; Choi *et al.*, 2009), the weight is modeled as

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

where L 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 lth latent component is not included in the model, and when $\psi_l = 1$, the lth component is included in the model. Following Kuo and Mallick (1998), the entry parameter has a Bernoulli distribution

$$\psi_l \sim Bern(p_l),$$

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 to provide the allocation of components based on weight prior distributions. In the STM models with continuous prior distributions for the weights (Model 1 and 2), we can 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

$$Z_i = \arg\max_l \{w_{il}\},$$

where $Z_i (= 1, \cdots, L)$ becomes the label index of the temporal component having the highest weight value in the ith area. This suggests that the temporal component with the highest weight value in the ith area is the primary temporal trend of the area. 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 with $w_{il} = 1$.

2.4 Bayesian Estimation

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

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

The prior distributions of the intercept parameter and variance parameters in the model are specified as

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

where $\sigma_{\alpha_0}^2$ is the variance and d is a constant (Gelman, 2006). We use a Beta prior distribution, Beta(1,1), for the temporal parameter ρ_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, Inv-Wishart$((0.01I_L)^{-1}, L)$, where I_L is the identity matrix of size L.

For Model 2, the posterior distribution for all the parameters Θ based on the likelihood

and the prior distributions is defined as

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

$$\boldsymbol{\rho} = (\rho_1, \cdots, \rho_L)^T, \qquad \mathbf{w} = (w_{11}, \cdots, w_{IL})^T, \qquad \boldsymbol{\chi} = (\chi_{11}, \cdots, \chi_{LJ})^T$$

$$\boldsymbol{\sigma}_{w^*} = (\sigma_{w_1^*}, \cdots, \sigma_{w_L^*})^T, \qquad \text{and} \quad \boldsymbol{\sigma}_\chi = (\sigma_{\chi_1}, \cdots, \sigma_{\chi_L})^T.$$

Posterior distributions of the other models can be easily obtained. The estimation of the parameters is implemented by hybrid Gibbs and Metropolis-Hasting sampling algorithms. Estimates for all the parameters except Z_i are the posterior means as the posteior mean is the Bayes estimator under quadratic loss. Since the cluster indicator Z_i is the nominal value, the posterior mode is used for the estimation of Z_i.

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 capability of recovery 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 in Bayesian models, 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 ith area and $\hat{Z}_{ik}$ is the estimated cluster indicator for the ith area at the kth sample, where $k = 1, \cdots, K$, and K is the number of simulated data sets. The first criteria we consider is the cluster accuracy rate of the ith area over simulations which is given by

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

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 $\overline{A} = \sum_{i=1}^{I} A_i/I$, which can be used as the measure of the cluster accuracy for each model. We extend this measure to incorporate spatial neighborhood information. The accuracy rate for the neighbor clusters of the ith area is defined as

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

where δ_i is the set of neighbors of the ith area. This measure examines the performance of the cluster detection for neighbors. In a similar way, the overall neighborhood accuracy rate is calculated by $\overline{NA} = \sum_{i=1}^{I} NA_i/I$. 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 outputs and the estimated pairwise cluster outputs, the pairwise accuracy rate is

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

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

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

The pairwise sensitivity and the pairwise specificity are calculated based on the assumption that true clusters of pairwise areas are equal and they are unequal, respectively. Thus, these measures are useful tools to investigate the performance of cluster recovering for pairwise areas under the assumption that the true clusters of 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 true values, several accuracy measures are proposed here. We define the difference of a true relative risk

and its estimate as $d_{ijk} = \hat{\theta}_{ijk} - \theta^T_{ijk}$, where θ^T_{ijk} is the true relative risk of the ith area and the jth time at the kth sample and its corresponding estimate is $\hat{\theta}_{ijk}$. 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 for the relative risks is $\text{MSE}_{\text{RR}} = \frac{1}{KIJ}\sum_k \sum_{i,j} d^2_{ijk}$. Another common measure is the average of absolute relative errors, defined as $\text{AARE}_{\text{RR}} = \frac{1}{KIJ}\sum_k \sum_{i,j} |\frac{d_{ijk}}{\theta^T_{ijk}}| = \frac{1}{KIJ}\sum_k \sum_{i,j} |\frac{\hat{\theta}_{ijk}-\theta^T_{ijk}}{\theta^T_{ijk}}|$. 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,

$$C^{(c)}_{ij} = \frac{1}{K}\sum_{k=1}^{K} I\Big(\Big|\frac{\hat{\theta}_{ijk} - \theta^T_{ijk}}{\theta^T_{ijk}}\Big| < c\Big).$$

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 ith area and the jth 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 $\overline{C}^{(c)} = \sum_{(i,j)} C^{(c)}_{ij}/(IJ)$ 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 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

$$D(\Theta) = -2 \log p(\mathbf{y}|\Theta),$$

where $p(y|\Theta)$ is the likelihood function for the data given the parameters Θ. 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

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

where $p_D = \overline{D(\Theta)} - D(\hat{\Theta})$ represents the effective number of parameters. Here, $\overline{D(\Theta)}$ measures the model fit and p_D measures the model complexity. The standard DIC is a widely used model assessment criteria but it may not be easy to obtain the correct p_D in mixture models. Thus, we use an alternative DIC measure, DIC_3 suggested by Celeux *et al.* (2006), which performs well in mixture models. This measure uses a posterior estimate of likelihood instead of $D(\hat{\theta})$. We have $\text{DIC}_3 = \overline{D(\Theta)} + [\overline{D(\Theta)} + 2 \log \hat{p}(\mathbf{y}|\Theta)]$, which is easily computed by Markov Chain Monte Carlo (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 by using the Conditional Predictive Ordinate

(CPO) (Dey *et al.*, 1997)

$$\text{MPL} = \sum_{i,j} \log\left(\text{CPO}_{ij}\right),$$

where CPO_{ij} is the marginal posterior predictive density of y_{ij} given the data excluding y_{ij}. The CPO represents a cross-validation measure for each observation given the remainder of the data. 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

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

where y_{ij} is the observed value and $\hat{y}_{ij}$ is a value of y_{ij} for 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 and goodness-of-fit measures presented in the previous section. We examine STM models when the number of components is both known and unknown.

In the simulation study, we have used the 159 counties of the state of Georgia as a

spatial domain. Georgia state has a fair number of counties and and is a common spatial layout so diverse designs for spatial clusters can be considered. This spatial domain is also used in our Section 5 data sets. 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 spatial-temporal domain and the expected count data, we conducted simulation experiments to compare the four STM models (Model 1 - Model 4) 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 spatial designs for the cluster indicator Z_i and different number of components (Figure 1). Design 1 has $L = 2$ components. The spatial pattern of the cluster indicator Z_i is defined by the population density. The first group ($Z_i = 1$) is made of the counties of high population density (> 100 per square mile) and the second group ($Z_i = 2$) is made of the counties of low population density (≤ 100 per square mile). Design 2 and 3 assume the number of components is $L = 4$, but they have different spatial patterns for the cluster. Design 3 has the clusters with the distant areas. 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 kth simulated set from

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

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

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

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_{ljk}, which is generated independently from a normal distribution with $N(\rho_{lk}\chi_{l,(j-1),k}, 1)$. The temporal parameter ρ_{lk} is generated independently from a uniform distribution with the range $[0, 1]$.

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

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

where α_0 and $\chi_{z_i,j}$ are constant over simulations and generated from the same scheme as the previous one. Figure 2 shows the maps of the true relative risks for the simulated data

in Design 3.

For each design we generate 500(= K) data sets and fit the different models (Model 1-4) of Section 2 with the same number of components with simulated data. To implement this study, two softwares R (`http://www.r-project.org/`) and WinBUGS (`http://www.mrc-bsu.cam.ac.uk/bugs`) are used. For the estimation of the posterior distributions of the parameters we discard the first 20000 iterations as burn-in and collect every 10th iteration to obtain 5000 final samples.

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

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

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

Table 1 shows the performance of different models in 4 designs in terms of the proposed cluster detection methods and 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 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 with isolated spatial clusters like Design 3. Thus, Model 1 provides better performance than Model 2. On the other hand, Model 4 has a singular multinomial prior distribution 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 for both Model 2 and Model 4 in terms of recovering the relative risk. 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 from Model 4 in all the designs are displayed in Figure 3. 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 4 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.

In Figure 5, we show the plots of $\overline{C}^{(c)}$ against the threshold value c for the models. As a threshold value c increases, the plots of $\overline{C}^{(c)}$ in all the models tend to be similar, but when c is small, $\overline{C}^{(c)}$ measure has quite different values depending on the models. For all the designs, Model 2 and 4 have almost same plots of $\overline{C}^{(c)}$ and larger values of $\overline{C}^{(c)}$ than the other models when c is small. In particular, Model 1 has the lowest $\overline{C}^{(c)}$ values 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 model comparison results of model fitting by the average DIC(ADIC), the average DIC_3 (ADIC_3), the average MPL (AMPL), and the average MSPE (AMSPE) over the simulations. For the calculation of the pD in the standard DIC, we used two different ways in WinBUGS and R. In WinBUGS, the pD is computed by the difference between the posterior mean of the deviance and the deviance of the posterior means of the parameters, but, in R, the pD is computed by the variance of deviance (pD*=var(D)/2). Thus, the average DIC with the former pD is denoted as ADIC and the average DIC with the latter pD* is denoted as ADIC*. When comparing the models, a model with smaller ADIC, ADIC*, ADIC_3 and AMSPE is better, but 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 ADIC_3 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 also consider the situation when the exact number of components is unknown. In this case, we use the entry parameters proposed in Section 2.2. 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. To produce simulated data sets, we use Design 2 for the cluster indicator Z_i with 4 latent components and define the temporal parameter of the components as $\rho = (1, 0.7, 0.4, 0.1)$ to distinguish the components. When fitting the models, we use 10 entry parameters which follow an independent Bernoulli distribution with probability 0.5.

For the comparison, we perform 200 simulations and include a component in the 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. For Model 1, 9.5% of the simulations only estimates the true number of components exactly and, for Model 3, 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 Model 2 is better than Model 4.

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 models are better than the non-spatial 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. Model 4 is marginally better than Model 2 in terms of model fitting, but Model 2 and Model 4 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. Table 5 presents how well the entry parameter models detect the clusters. Model 3 has 0% for the estimation of the true number of components, and we have no results here. It indicates that the spatial models have higher accuracy rates for the cluster detection measures than Model 1. Also, Model 2 and 4 provide similar results.

5 Data Analysis

For the assessment of the performance of STM models with real data, we analyze two health data sets in the state of Georgia: low birth weight (LBW, $< 2500gm$) data and chronic obstructive pulmonary disease (COPD) data. LBW is one of important child's health problems and is affected by demographic and socioeconomic factors which vary with space and time. COPD is one of the most common lung diseases and is linked with particulates and indoor air contaminants which also vary with space and time. Thus, these data sets could show spatial-temporal variation, which can allow relative risks to have different temporal patterns over space. With these real data sets it is interesting to estimate several STM

models and evaluate their performance by using various criteria. In addition, we conduct a comparison of the allocation estimates from different models. In real data analysis, we do not know the true spatial clusters and true relative risks so the cluster detection methods and the risk accuracy measures proposed in Section 3 can not be employed here. However, we can compare four different STM models in terms of a range of goodness-of-fit measures and their allocation results. To see the performance of the STM models in the situation when the number of temporal components is known or unknown, we apply STM models with fixed L to LBW data and STM models with entry parameters to COPD data.

5.1 Georgia low birth weight data (fixed L)

We apply STM models to LBW data for the years 1994 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 14 years of data. The expected counts were calculated by using the internal standardization method (Banerjee *et al.*, 2004), where the population is the number of infants. Figure 6 presents a selection of standardized incidence maps for low birth weight births and shows the spatial-temporal variation of standardized incidence ratios. South-west areas have high standardized incidence ratios of LBW for the year 2007. It is an evidence that relative risks have locally different temporal patterns so this data set is appropriate in this study.

Using this LBW data, we first fitted four STM models with 10 entry parameters to determine the number of components. Model 1 and 2 estimate 2 temporal components among

10 components. But, Model 3 and 4 estimate 8 components among which 6 components look negligible and redundant even though they are included in the model. We conclude that two temporal components in relative risks are sufficient in this data. We again fit STM models with 2 components and compare them by using goodness-of-fit measures (Table 6). All measures favor Model 2 so Model 2 is the best model. In this example, Model 3 and 4 provide poor fitting. We also fitted STM models with up to 8 components and found that Model 2 is overall the best model.

Figure 7 shows the temporal plots of 2 components from Model 2 and 4. As you can see, component values for both models are different but the temporal patterns look similar. Figure 8 displays the maps of the estimated allocation indicator Z_i from Model 2 and 4. These maps indicate that Model 2 and 4 have similar allocation results. The percentage that two models have the same component is 89.31%. South-west areas have Component 1 that tends to increase after the year 2004.

5.2 Georgia chronic obstructive pulmonary disease data (unknown L)

To investigate the performance of STM models when the number of components is unknown, we analyze county-level COPD data for a period of 9 years (1999-2007) in Georgia, which were obtained from OASIS. The expected counts were computed by using the internal standardization method. Figure 9 displays the maps of the standardized incidence ratios for each year and we can see the spatial and temporal variation. Especially, north areas

and south-east areas in Georgia have high standardized incidence ratios of COPD over the years of study.

In this example, we assume that the number of latent components is unknown and fit four STM models with $L = 10$ entry parameters. Table 7 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 they seem to be overfitting the data. It appears that Model 2 is the best fit model in terms of several DICs and MPL. MSPE measure favors Model 1 but Model 2 also has small MSPE. Overall, Model 2 fits the data well and provide good prediction performance.

Figure 10 presents the temporal plots for the components included in the model 2 based on the entry parameters. Component 1 has a decreasing pattern and Component 2 has a quite stable pattern. To examine the spatial variation of the weights in this case, the maps of the weights corresponding with the components are presented in the left two maps in Figure 11. Using our allocation method, we can identify the spatial clusters. The right map in Figure 11 shows the map of the cluster indicator Z_i from Model 2 and atlanta areas have Component 1 and south-east areas have Component 2.

6 Conclusion

In this paper, we evaluated spatial-temporal mixture models with different weight structures. We considered mixture models with entry parameters when the number of components was

unknown. For the comparison of these models, we developed a range of spatial cluster detection methods based on the posterior distribution of the weights. 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 STM models with different spatial prior distributions for weights have similar results. When the exact number of components is unknown, the mixture model with a non-spatial singular multinomial prior distribution for weights could be overfitting the data. The STM model with a spatial continuous prior distribution for weights estimates the true number of components well. In our real data analysis, the STM model with a spatial continuous distribution for weights performs well when the number of components is both known and unknown. However, the STM models with a singular multinomial distribution for weights tend to be overfitting the data when the number of latent components is unknown. From our simulation study and real data analysis, we found that the STM model with a spatial continuous distribution for weights works well in terms of various criteria. The STM model with a spatial singular multinomial distribution for weights performs well in the simulation study but does not work well in real data analysis.

Our spatial cluster detection measures and risk accuracy measures can be only used for a simulation study. In real data analysis, a comparison of the estimated allocation indicator and the estimated latent components is an alternative way. By adding covariates in spatial-

temporal mixture models, we could investigate the performance of STM models with several criteria. We could also study the performance of multivariate STM models.

Acknowledgements

This work was supported by NIH R21 R21HL088654-01A2.

References

Banerjee S, Carlin BP, Gelfand AE. 2004. *Hierarchical modeling and analysis for spatial data.* Chapman and Hall: New York.

Bernardinelli L, Clayton DG, Pascutto C, Montomoli C, Ghislandi M, Songini M. 1995. Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine* **14**: 2433–2443.

Celeux G, Forbes F, Robert C, Titterington M. 2006. Deviance information criteria for missing data models. *Bayesian Analysis* **1**: 651–674.

Choi J, Fuentes M, Reich BJ. 2009. Spatial-temporal association between fine particulate matter and daily mortality. *Computational Statistics and Data Analysis* **53**: 2989–3000.

Congdon P. 2005. *Bayesian Models for Categorical Data.* John Wiley and Sons: New York.

Dellaportas P, Forster J, Ntzoufras I. 2002. On Bayesian model and variable selection using MCMC. *Statistics and Computing* **12**: 27–36.

Dey D, Chen MH, Chang H. 1997. Bayesian approach for nonlinear random effects models. *Biometrics* **53**: 1239–1252.

Dreassi E, Biggeri A, Catelan D. 2005. Space-time models with time-dependent covariates for the analysis of the temporal lag between socioeconomic factors and lung cancer mortality. *Statistics in Medicine* **24**: 1919–1932.

Gelfand AE, Vounatsou P. 2003. Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* **4**: 11–25.

Gelman A. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* **1**: 515–533.

Green PJ. 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**: 711–732.

Hossain MM, Lawson AB. 2006. Cluster detection diagnostics for small area health data: With reference to evaluation of local likelihood models. *Statistics in Medicine* **25**: 771–786.

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

Ibrahim J, Chen MH, Sinha D. 2001. *Bayesian Survival Analysis*. Springer: New York.

Jasra A, Holmes CC, Stephens DA. 2005. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science* **20**: 50–67.

Knorr-Held L. 2000. Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* **19**: 2555–2567.

Knorr-Held L, Besag J. 1998. Modelling risk from a disease in time and space. *Statistics in Medicine* **17**: 2045–2060.

Kuo L, Mallick B. 1998. Variable selection for regression models. *Sankhya B* **60**: 65–81.

Lawson AB, Song HR, Cai B, Hossain MM, Huang K. 2010. Space-time latent component modeling of geo-referenced health data. *Statistics in Medicine*. DOI: 10.1002/sim.3917.

Mardia KV, Goodall C, Redfern EJ, Alonso FJ. 1998. The kriged Kalman filter (with discussion). *Test* **7**: 217–285.

Martinez-Beneito MA, Lopez-Quilez A, Botella-Rocamora P. 2008. An autoregressive approach to spatio-temporal disease mapping. *Statistics in Medicine* **27**: 2874–2889.

Mugglin AS, Cressie N, Gemmell I. 2002. Hierarchical statistical modelling of influenza epidemic dynamics in space and time. *Statistics in Medicine* **21**: 2703–2721.

Richardson S, Abellan J, Best N. 2006. Bayesian spatio-temporal analysis of joint patterns of male and female lung cancer risks in Yorkshire (U.K.). *Statistical Methods in Medical Research* **15**: 97–118.

Spiegelhalter DJ, Best N, Carlin BP, van der Linde A. 2002. Bayesian deviance, the effective number of parameters and the comparison of arbitrarily 35 complex models. *Journal of the Royal Statistical Society B* **64**: 583–640.

Stephens M. 2000. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society B* **62**: 795–809.

Tzala T, Best N. 2008. Bayesian latent variable modelling of multivariate spatio-temporal variation in cancer mortality. *Statistical Methods in Medical Research* **17**: 97–118.

Xia H, Carlin BP, Waller LA. 1997. Hierarchical models for mapping Ohio lung cancer rates. *Environmetrics* **8**: 107–120.

Table 1: Diagnostics using cluster detection measures and risk accuracy measures.

		Cluster detection measures					Risk accuracy measures		
Design	Model	$\overline{A}$	$\overline{NA}$	PA	P_{Sen}	P_{Spe}	AAE_{RR}	MSE_{RR}	AARE_{RR}
1	1	0.879	0.854	0.763	0.738	0.799	0.222	0.134	0.213
($L = 2$)	2	0.906	0.881	0.804	0.781	0.836	0.166	0.091	0.168
	3	0.894	0.868	0.787	0.764	0.819	0.162	0.095	0.173
	4	0.910	0.884	0.809	0.787	0.840	0.154	0.087	0.162
2	1	0.772	0.780	0.816	0.640	0.877	0.332	0.228	0.423
($L = 4$)	2	0.804	0.805	0.865	0.839	0.875	0.201	0.095	0.227
	3	0.799	0.807	0.828	0.658	0.888	0.248	0.170	0.331
	4	0.861	0.865	0.879	0.777	0.915	0.199	0.117	0.259
3	1	0.746	0.753	0.755	0.519	0.867	0.317	0.215	0.355
($L = 4$)	2	0.575	0.582	0.712	0.588	0.770	0.238	0.128	0.261
	3	0.783	0.792	0.782	0.588	0.874	0.260	0.178	0.295
	4	0.833	0.840	0.820	0.690	0.882	0.232	0.152	0.261
4	1	0.585	0.592	0.807	0.524	0.870	0.393	0.299	0.526
($L = 6$)	2	0.585	0.590	0.806	0.674	0.836	0.304	0.194	0.378
	3	0.688	0.697	0.837	0.540	0.904	0.314	0.232	0.424
	4	0.711	0.720	0.839	0.602	0.892	0.300	0.216	0.408

Table 2: Model fitting.

Design	Model	ApD	ADIC	ApD*	ADIC*	ADIC_3	AMPL	AMSPE
1	1	110.396	3732.444	144.765	3766.813	3705.581	-1877.034	6.459
($L = 2$)	2	43.580	3663.441	92.401	3712.261	3678.456	-1855.111	6.473
	3	61.622	3689.709	61.622	3689.709	3677.046	-1854.524	6.540
	4	60.808	3689.305	60.808	3689.305	3675.433	-1852.001	6.539
2	1	165.646	3824.385	387.219	4045.959	3806.081	-1963.502	7.325
($L = 4$)	2	51.620	3678.755	215.859	3842.993	3721.737	-1886.486	7.299
	3	162.918	3817.006	162.918	3817.006	3746.266	-1907.680	7.466
	4	124.924	3769.504	124.924	3769.504	3721.195	-1884.251	7.460
3	1	200.055	3832.518	376.761	4009.224	3783.352	-1947.835	6.288
($L = 4$)	2	116.010	3729.698	235.753	3849.441	3720.598	-1888.835	6.258
	3	168.628	3801.892	168.628	3801.892	3732.270	-1902.656	6.394
	4	140.842	3765.994	140.842	3765.994	3714.058	-1887.202	6.371
4	1	-110.885	3560.379	578.534	4249.798	3851.909	-2012.938	6.961
($L = 6$)	2	117.530	3760.724	403.916	4047.111	3780.235	-1937.141	6.950
	3	251.651	3909.611	251.651	3909.611	3784.418	-1943.686	7.175
	4	235.966	3889.429	235.966	3889.429	3773.440	-1933.226	7.173

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			

Table 4: Diagnostics using risk accuracy measures and model fitting for the entry parameter models.

Model	AAE_{RR}	MSE_{RR}	$AARE_{RR}$	ApD	ADIC	ApD*	ADIC*	$ADIC_3$	AMPL	AMSPE
1	0.338	0.233	0.425	459.728	4093.090	459.728	4093.090	3790.593	-1963.247	7.217
2	0.180	0.084	0.213	456.188	4076.694	456.188	4076.694	3706.834	-1879.265	7.396
3	0.290	0.195	0.384	426.390	4024.191	426.390	4024.191	3739.544	-1925.120	7.167
4	0.177	0.087	0.216	177.054	3789.762	177.054	3789.762	3697.751	-1868.999	7.405

Table 5: Cluster diagnostics only when the estimated number of components is equal to the true number of components.

Model	$\overline{A}$	$\overline{NA}$	PA	P_{Sen}	P_{Spe}
1	0.743	0.754	0.793	0.597	0.861
2	0.931	0.933	0.937	0.886	0.955
3					
4	0.938	0.939	0.941	0.885	0.961

Table 6: Diagnostic results for LBW data.

Model	ApD	ADIC	ApD*	ADIC*	$ADIC_3$	AMPL	AMSPE
1	176.395	15141.400	304.032	15269.037	15212.922	-7632.256	244.710
2	177.180	15078.100	202.147	15103.067	15142.573	-7598.050	243.817
3	44.280	16576.106	44.280	16576.106	16688.370	-8355.617	297.236
4	48.113	16580.343	48.113	16580.343	16686.765	-8354.261	299.955

Table 7: Diagnostic results with 10 entry parameters for COPD data.

Model	$\hat{L}$	ApD	ADIC	ApD*	ADIC*	$ADIC_3$	AMPL	AMSPE
1	2	218.850	12578.637	218.850	12578.637	12788.335	-6456.497	436.830
2	2	193.130	12534.080	193.130	12534.080	12763.503	-6447.497	437.864
3	9	132.882	13201.955	132.882	13201.955	13550.268	-6872.754	492.641
4	9	138.852	13224.942	138.852	13224.942	13595.930	-6891.754	488.464

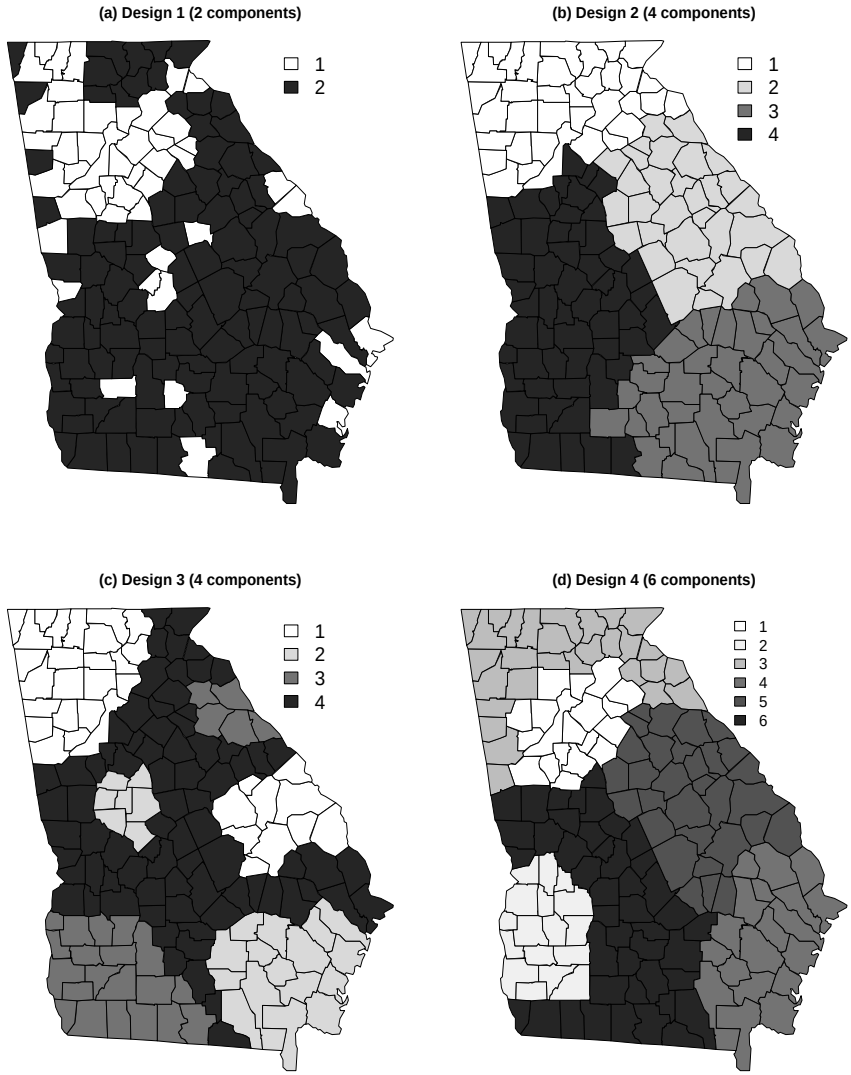


Figure 1: Spatial designs of the cluster indicator (Z_i) in the simulation study.

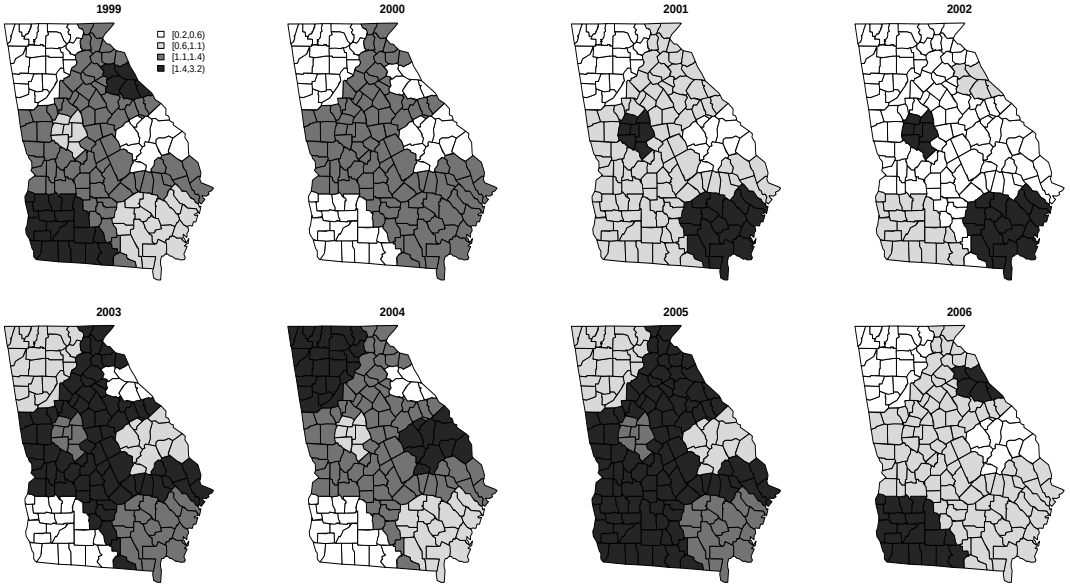


Figure 2: Maps of the true relative risks in Design 3.

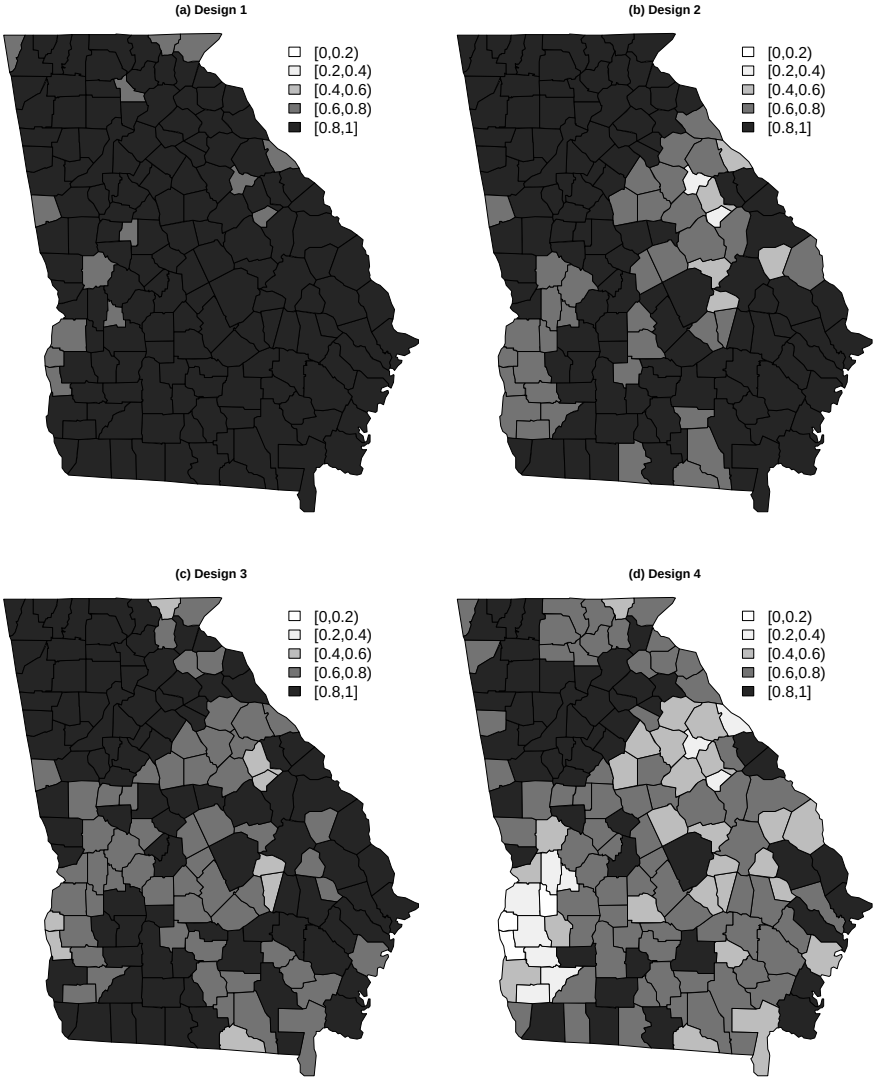


Figure 3: Maps of A_i for Model 4.

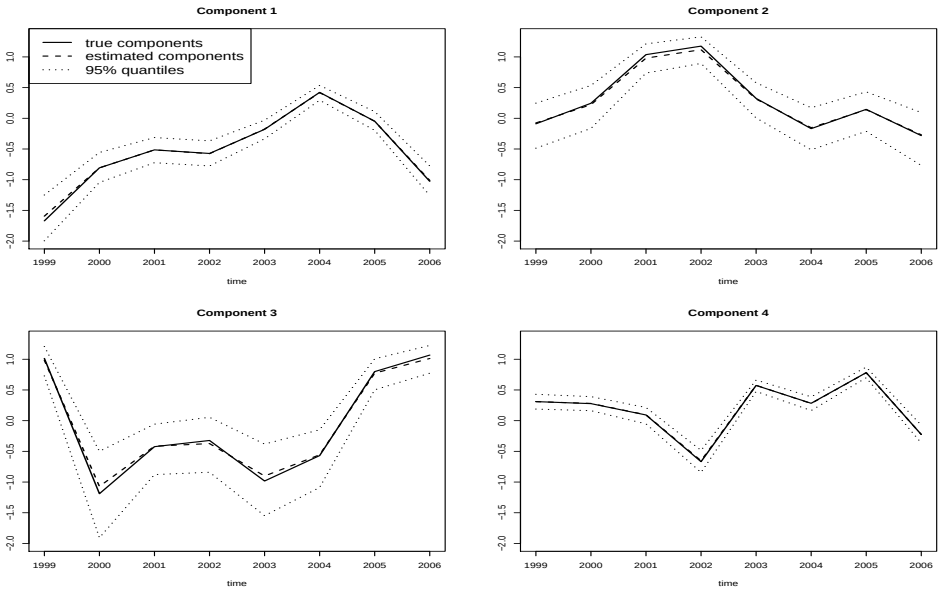


Figure 4: Plots of the true temporal components and estimates from Model 4 in Design 3. The solid line is the true component, the dashed line is the average of the posterior estimated component, and the dotted lines are 95% intervals for the posterior estimated component.

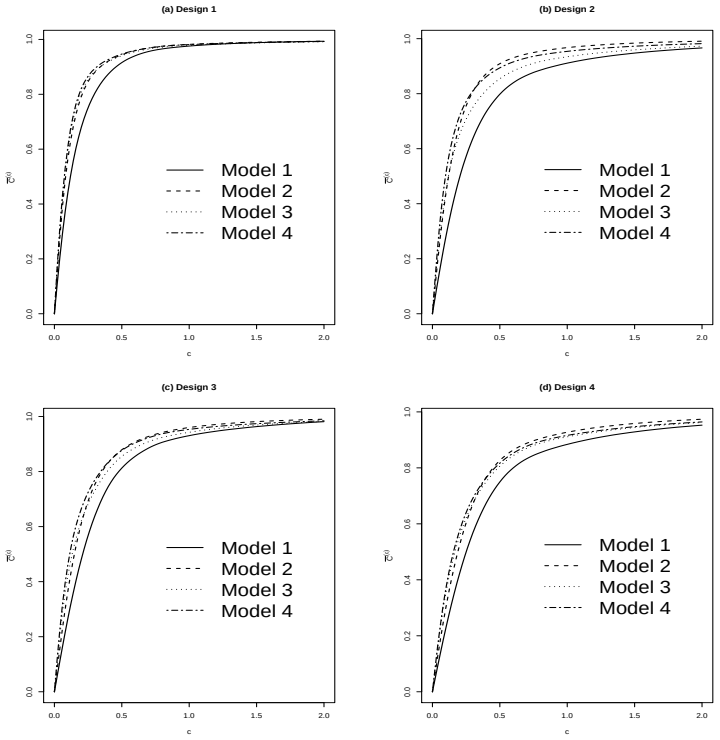


Figure 5: Plots of $\overline{C}^{(c)}$ against the threshold value c.

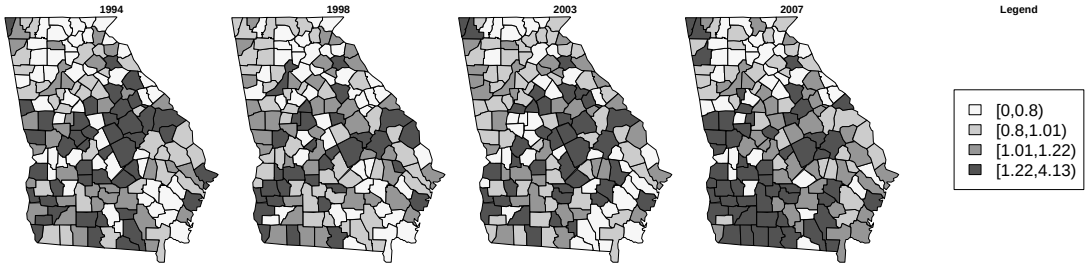


Figure 6: A selection of four years of standardized incidence maps for county-level LBW data in Georgia (1994, 1998, 2003, and 2007).

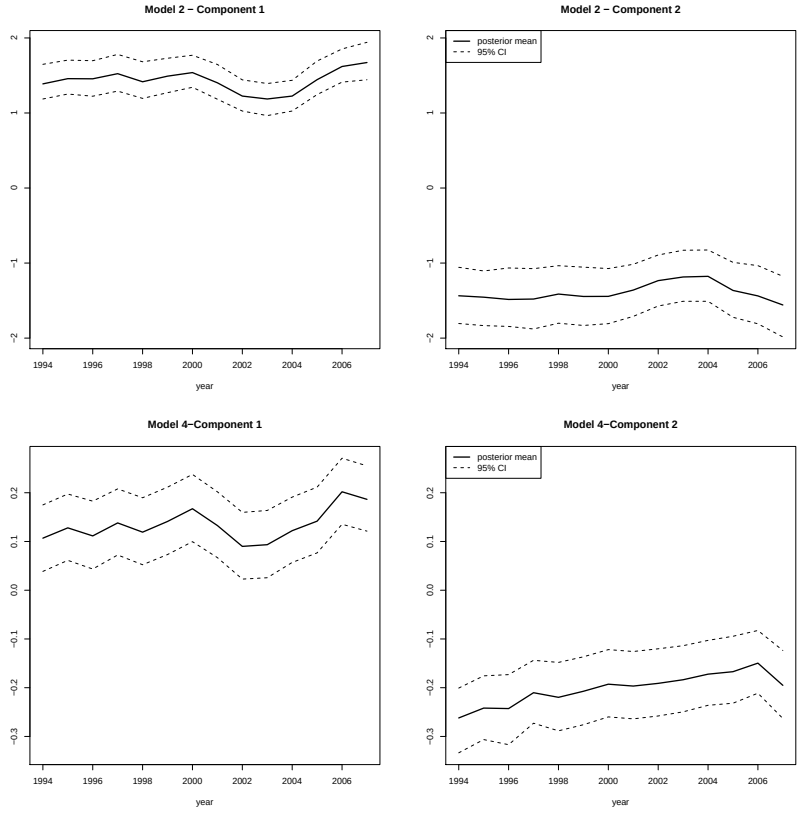


Figure 7: Temporal plots for 2 components from Model 2 and Model 4 in LBW data.

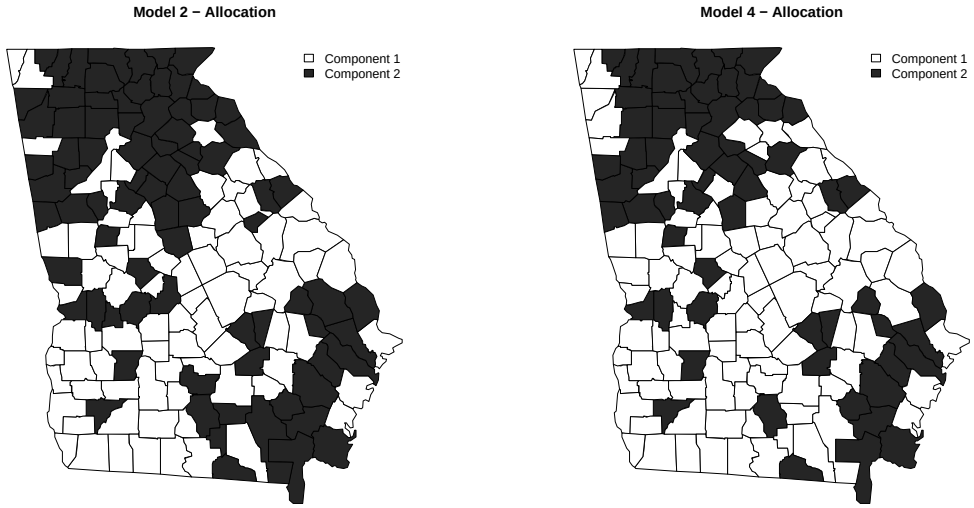


Figure 8: Maps of the allocations from Model 2 and 4 with 2 components in LBW data.

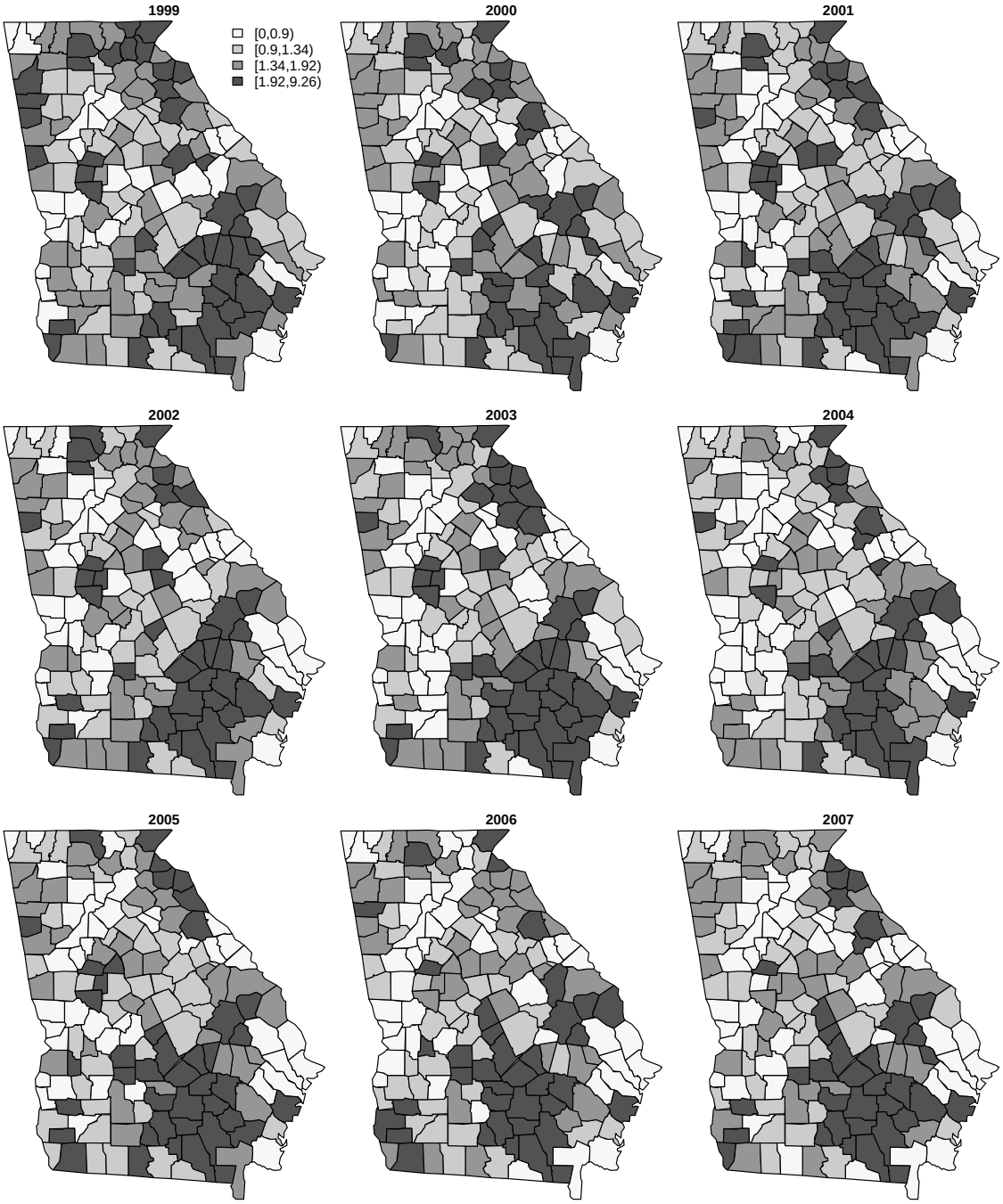


Figure 9: Standardized incidence maps of county-level COPD data in Georgia.

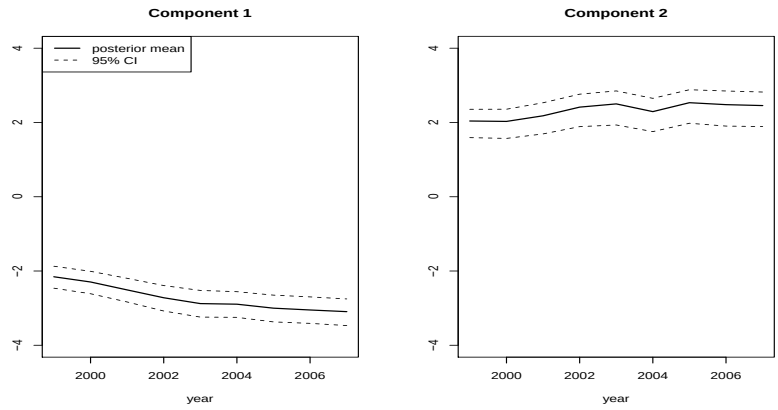


Figure 10: Temporal plots for 2 estimated components from Model 2 in COPD data.

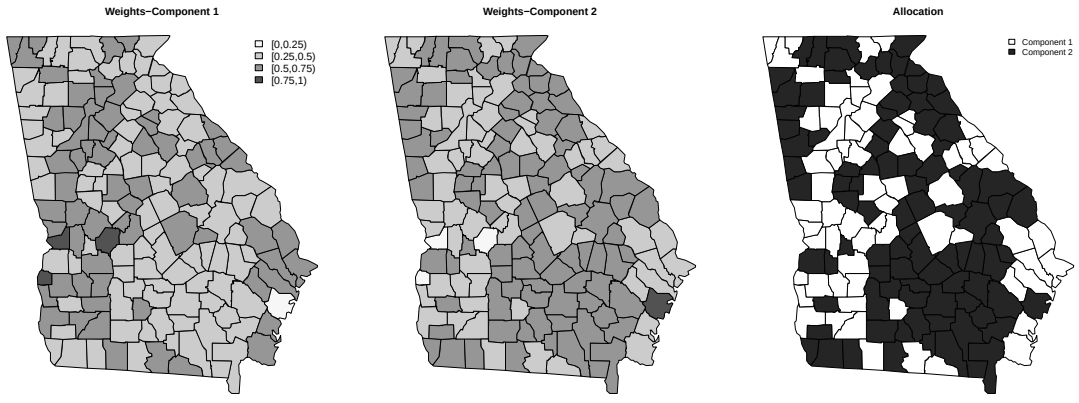


Figure 11: Maps of the estimated weights corresponding with the components from Model 2 and allocation results.