Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

A Bayesian multi-stage spatio-temporally dependent model for spatial clustering and variable selection

Shaopei MA[1] | Keming YU[2] | Man-lai TANG[3] | Jianxin PAN[4] | Wolfgang Karl HÄRDLE[5] | Maozai TIAN*[1]

[1]Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing, China
[2]Mathematical Sciences, Brunel University, Uxbridge, London, United Kingdom

[3]Mathematical Sciences, Brunel University, Uxbridge, London, United Kingdom
[4]School of Mathematics, University of Manchester, Manchester, United Kingdom
[5]School of Business and Economics, Humboldt-Universität zu Berlin , Berlin, Germany

Correspondence
*Maozai TIAN. Email: mztian@ruc.edu.cn

Summary

In spatio-temporal epidemiological analysis, it is of critical importance to identify the significant covariates and estimate the associated time-varying effects on the health outcome. Due to the heterogeneity of spatio-temporal data, the subsets of important covariates may vary across space and the temporal trends of covariate effects could be locally different. However, many spatial models neglected the potential local variation patterns, leading to inappropriate inference. Thus, this paper proposes a flexible Bayesian hierarchical model to simultaneously identify spatial clusters of regression coefficients with common temporal trends, select significant covariates for each spatial group by introducing binary entry parameters and estimate spatio-temporally varying disease risks. A multi-stage strategy is employed to reduce the confounding bias caused by spatially structured random components. A simulation study demonstrates the outperformance of the proposed method, compared with several alternatives based on different assessment criteria. The methodology is motivated by two important case studies. The first concerns the low birth weight incidence data in 159 counties of Georgia, USA, for the years 2007-2018 and investigates the time-varying effects of potential contributing covariates in different cluster regions. The second concerns the circulatory disease risks across 323 local authorities in England over 10 years and explores the underlying spatial clusters and associated important risk factors.

KEYWORDS:
Bayesian hierarchical model; Spatial clustering; Spatial confounding problem ; Spatio-temporal modeling; Variable selection.

1 | INTRODUCTION

Over the last decades, there has been an increasing interest in spatio-temporal data, with the consequent development of numerous statistical methodologies for modeling public health outcomes across space and time. Especially, the association between covariates and health outcomes is getting much more attention.[1,2,3] Many existing spatial methods are based on global modeling, with restrictive assumptions of constant covariate effects across the entire domain of interest.[4] However, the impacts of risk factors on the disease are likely to be spatio-temporally varying. Therefore, discovering the spatio-temporal patterns of the covariate effects is essential and helpful to understand disease aetiology and further monitor healthcare access.

Our work is motivated by two important public health case studies. The first is the county-level infant low birth weight (LBW) data in Georgia between 2007-2018 and the second is the circulatory disease data in local authorities of England over a 10-year period, both of which have been previously studied by some researchers.[5,6,7] It is reported that both diseases have obvious adverse impacts on the development of public health. Therefore, it is of great importance to analyze the possible risk factors and identify significant variables for the region of interest. However, the effects of a covariate can differ in different regions due to varying geographical characteristics, concentration levels of the risk factors and so on. For example, as stated by Pearl et al,[8] the unemployment rate plays a role only in the areas with relatively high joblessness for the LBW data analysis. To our knowledge, most approaches assume that the covariate effects on health outcomes vary with small areas, that is, different small areas have different regression coefficients. This assumption apparently ignores the spatial correlation among the adjacent areas. It is noted that the temporal patterns of covariate effects may be locally similar and there are spatial clusters in each of which covariate effects have homogeneous temporal trends.[9] What's more, some covariates in the clusters may have little effects on the outcomes, that is, the subset of significant covariates could vary with spatial clusters. Thus, in both data cases, the aim is to cluster small areas together that share common significant covariates and estimate the spatio-temporally varying covariate effects on the health outcomes. The results have important potential implications for a reduction in disease risks through targeted public health interventions.

Some researchers have studied Bayesian models with varying coefficients, such as Cai et al,[10] Gelfand and Vounatsou,[11] Luis[12] and so on. Recently, Lee et al[13] proposed detecting clusters of units based on hypotheses testing. But they assumed independent observations, which are not necessarily applicable to spatial data. As an extension, Lee et al[14] proposed a mixed-effects model for spatial cluster detection of regression coefficients. However, the method mainly focused on spatial data, and did not take the temporal trends of covariate effects into consideration. Thus, it is necessary to develop a statistical method to identify spatial groups of areas exhibiting homogeneous temporal profiles of covariate effects by borrowing information across space and time.

In epidemiologic analysis, it is of great importance to select covariates that are closely related to the health outcomes. Many variable selection approaches have been developed to choose significant covariates for the whole domain of interest, with fixed subset of risk factors across space. However, utilizing categorical parameters to identify spatial groups, Choi et al[9] found that covariate effects are significantly different depending on areas. Thus, in this article, we introduce binary entry parameters first proposed by Kuo and Mallick[15] to allow for variable selection for each spatial cluster. A flexible Bayesian hierarchical model is proposed with spatio-temporally varying coefficients, which simultaneously identifies spatial clusters with distinct temporal patterns of covariate effects, as well as selects group-specific important covariates for the health outcomes.

In spatio-temporal health studies, both risk factors and random effects are important for estimating the health outcomes. However, many studies[16] only took spatial and temporal random components into account while some others[9,12] only considered covariate effects, where extra source of information in the random components may be omitted. In this paper, both space-time varying covariates and random effects are incorporated into the model. According to the previous studies, spatial confounding problems may arise due to the possible correlation between risk factors and random effects, making it difficult to obtain accurate estimates.[17] Lawson et al[18] proposed a multi-stage spatial mixture model for $PM_{2.5}$ concentration data, which was proved by Baer and Lawson[19] to be rather effective for reducing the estimation bias. Since this problem has not been fully discussed for spatio-temporal data, it is meaningful to take it into consideration in this paper to provide better estimates for the parameters of interest.

In this article, a Bayesian spatio-temporally dependent model is proposed to simultaneously identify spatial clusters in each of which covariates have homogeneous temporal effects, select significant covariates for each spatial group by introducing entry parameters, as well as better estimate locally varying covariate effects and spatio-temporally varying disease risks. To reduce confounding bias, the multi-stage estimation strategy is employed. In the first stage, we only incorporate covariate information into the hierarchical model. An obvious difference from previous studies is that inspired by Choi et al[9] and Boulieri et al,[16] the areas in the same cluster are not restricted to being spatially adjacent, since the temporal behaviors of coefficients in non-neighboring areas may also be the same. Thus, a set of spatially structured weights is employed to capture the grouping patterns. To achieve locally varying variable selection, the coefficient in each group is hierarchically modeled as the product of a binary entry parameter and a group-specific temporal process. The resulting residuals may contain much variability and are modeled in the second stage with extra space-time random components for overdispersion. In the third stage, both the covariates and the random component estimates are incorporated as inputs into the model, and the spatio-temporally dependent covariate effects on the health outcome are finally estimated using the same model structure as the first stage.

The novelty of this article lies in identifying spatial clusters with distinct temporal profiles of covariate effects and selecting significant covariates varying with different groups simultaneously within a statistical model. While many studies investigate spatial grouping defined in terms of the response,[2] the proposed method focuses on the locally varying covariate effects. To our knowledge, it is the first attempt to consider variable selection within the framework of spatially dependent clustering for space-time health data. Besides, the proposed model has good adaptability to employ various spatial dependence structures. Also, by introducing a multi-stage strategy, it can reduce the confounding bias neglected by many previous studies and thus, better estimate the spatio-temporally varying coefficients and disease relative risks.

The remainder of the article is organized as follows. Section 2 describes the Bayesian multi-stage hierarchical latent model for space-time data. Section 3 explores a variety of model assessment criteria. A simulation study is conducted in Section 4 to compare the proposed method with several alternatives. Section 5 analyzes the real data of low birth weight incidence in Georgia and circulatory disease data in England. Section 6 presents the discussion.

2 | BAYESIAN MULTI-STAGE SPATIO-TEMPORAL MODELING

2.1 | Proposed model

In this section, we propose a Bayesian multi-stage hierarchical model with spatio-temporally varying coefficients. A latent Poisson regression model is employed for health count data in small areas where relative risk is assumed to be a function of possible risk factors and space-time random components. However, spatial confounding problems may arise when spatio-temporally varying risk factors and space-time random components are both included in the model.[17,18] For example, median household income in the LBW data may be correlated with the space-time random components so it can be difficult to exactly estimate both the effects of median household income and the underlying random components. Therefore, following Lawson et al,[18,19] a multi-stage process is utilized to reduce the confounding bias. The proposed method could simultaneously achieve spatial clustering in the coefficients as well as variable selection for each spatial group.

Let $\{Y_{it}\}$ denote the observed count of disease in the ith spatial unit($i = 1, \dots, N$) and the tth time point($t = 1, \dots, T$) and E_{it} be the expected count, which is usually obtained by applying a standard table of sex-specific and age group-specific rates to the population count in area i at time t.[20] Y_{it} is assumed to follow a Poisson distribution as

$$Y_{it} \mid \theta_{it} \sim \text{Poisson}\left(E_{it}\theta_{it}\right), \quad i = 1, \cdots, N, \quad t = 1, \cdots, T,$$

where θ_{it} is the relative risk.

In the first stage, the logarithm of relative risk is modeled only by the covariate information with spatio-temporally varying coefficients as follows:

$$\log(\theta_{it}) = \beta_0 + \mathbf{X}_{it}'\boldsymbol{\beta}_{it}, \tag{1}$$

where β_0 is the intercept parameter to capture the overall relative risk effect across space and time, $\mathbf{X}_{it} = \left(X_{it1}, \dots, X_{itP}\right)'$ denotes a $P \times 1$ vector of covariates in area i at time t and $\boldsymbol{\beta}_{it} = \left(\beta_{it1}, \dots, \beta_{itP}\right)'$ denotes the corresponding $P \times 1$ coefficients vector depending on space and time. Equation (1) is called as the covariates-only model. In small area health studies, many approaches assume that the covariate effects are area-specific, that is, different small areas have different regression coefficients, sometimes leading to neglect of the similarity between spatial units. According to Lee et al,[13] the changing patterns of covariate effects may be locally homogeneous owing to spatial correlation. Thus, it is important to consider the identification of spatial clusters in the regression coefficients.

In this paper, following Chol et al,[9] it is assumed that there are spatial clusters in each of which covariate effects have homogeneous temporal patterns. Thus, varying with spatial groups, not small areas, the covariate effect $\boldsymbol{\beta}_{it}$ is modeled as

$$\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_{C(i),t},$$

where $C(i)(= 1, \dots, S)$ is the spatial cluster indicator and S is the number of clusters.

To identify spatial clusters, $C(i)$ is assumed to follow a categorical distribution as

$$C(i) \sim \text{Cat}(w_{i1}, \dots, w_{iS}),$$

where $w_{is}, s = 1, \ldots, S$ is the probability of area i belonging to cluster s, thus $\sum_{s=1}^{S} w_{is} = 1$ and $w_{is} \geq 0$. The probability w_{is} is expressed using unstandardized weight $w_{is}^* \geq 0$:

$$w_{is} = \frac{w_{is}^*}{\sum_{s=1}^{S} w_{is}^*}.$$

Considering the spatial correlation in the small areas, the weight parameter w_{is}^* is modeled as the following spatial dependence structure

$$\log\left(w_{is}^*\right) \sim N\left(\eta_{is}, \sigma_s^2\right),$$

where η_{is} is the spatially dependent mean and σ_s^2 is the variance parameter. To account for spatial correlation in the mean, the parameter η_{is} is assumed to follow an intrinsic conditional autoregressive (ICAR) distribution proposed by Besag et al,[21] which is widely used in space-time studies, such as Boulieri et al.[16] Its full conditional distribution is specified as

$$\eta_{is} \mid \eta_{i's}, i' \neq i \sim N\left(\frac{1}{n_i}\sum_{i' \sim i} \eta_{i's}, \frac{\sigma_{\eta_s}^2}{n_i}\right)$$

independently for each s, where $i' \sim i$ denotes that area i' is a neighbor of area i, n_i is the number of neighbors of area i, and $\sigma_{\eta_s}^2$ controls the magnitude of spatial variation. This specification is denoted as $\eta_{is} \sim \text{ICAR}(\sigma_{\eta_s}^2)$. The constraint $\sum_{i=1}^{N} \eta_{is} = 0$ is needed to ensure the identifiability of ICAR prior.

Further, considering that some covariates in the spatial groups may have little effects on the outcomes, we propose the development of spatial variable selection approach to allow the subset of significant covariates to vary across spatial groups. To be exact, denoting the coefficient vector within the spatial group $C(i) = s$ by $\boldsymbol{\beta}_{st} = \left(\beta_{st1}, \ldots, \beta_{stP}\right)'$, the pth coefficient parameter β_{stp} is hierarchically modeled as

$$\beta_{stp} = \delta_{sp} \times \lambda_{stp}, \quad s = 1, \ldots, S, \quad p = 1, \cdots, P,$$

where δ_{sp} is a binary selection parameter with a value of 0 or 1, and λ_{stp} denotes the pth covariate effect with temporal dependence when the pth covariate is selected in cluster s. If the selection indicator $\delta_{sp} = 0$, then the pth covariate is not correlated with the outcome in the areas of cluster s. Otherwise, the indicator $\delta_{sp} = 1$ means that the pth covariate is significant and has a certain temporal dependence in cluster s, with the corresponding coefficient $\beta_{stp} = \lambda_{stp}$.

The subset of covariates with significant effects on the outcomes varies with different spatial clusters depending on the value of δ_{sp}. And in each cluster s, a significant covariate effect has a specific temporal pattern λ_{stp} which is distinguishing from the patterns of the covariate effects in other clusters. There are many alternative temporal structures for the parameter λ_{stp} to consider, such as autoregressive models or random walk models. In this paper, an AR(1) process is considered for the temporal pattern of λ_{stp} as

$$\lambda_{stp} | \lambda_{s,t-1,p} \sim N(\rho_{\lambda_{sp}} \lambda_{s,t-1,p}, \sigma_{\lambda_{sp}}^2), \quad s = 1, \ldots, S, \quad p = 1, \cdots, P,$$

where $\rho_{\lambda_{sp}}$ is the autoregression coefficient and $\sigma_{\lambda_{sp}}^2$ is the corresponding variance. It is worth noting that generalizations of the AR model for multiple lags are available, yet less parsimonious.

The binary selection parameter δ_{sp} is assumed to follow a Bernoulli distribution as

$$\delta_{sp} | \phi_{sp} \sim \text{Ber}(\phi_{sp}),$$

where the probability ϕ_{sp} explains the behaviour of the pth covariate in cluster s. According to Adin et al,[22] in general, there is little evidence to show spatial dependence between different clusters. Thus, an unstructured prior distribution can be used to model the selection probability ϕ_{sp} as

$$\text{logit}(\phi_{sp}) = \mu_p + \xi_{sp}, \tag{2}$$

where μ_p is the intercept parameter and $\xi_{sp} \sim N(0, \sigma_{\xi_{sp}}^2)$ is the spatially uncorrelated parameter.

From the hierarchical models in the first stage, we can obtain the posterior distribution of the relative risk, θ_{it}, and compute the posterior mean as its estimate by Bayesian approaches. The residuals can be calculated as

$$\hat{r}_{it} = \log(Y_{it}/E_{it}) - \log(\hat{\theta}_{it}).$$

Then we come to the second stage. Since the covariates-only model (1) does not include random effects, and there may be extra source of space-time structure in the random components, it is essential to further model this residual as

$$\begin{aligned} \hat{r}_{it}|\psi_{it} &\sim N(\psi_{it}, \sigma_r^2), \\ \psi_{it} &= u_i + v_i + \gamma_t + \zeta_t, \end{aligned} \tag{3}$$

where ψ_{it} explains the spatio-temporal random effects and σ_r^2 is the variance. It is assumed that the spatially uncorrelated component u_i follows a normal distribution as $u_i \sim N(0, \sigma_u^2)$ and the spatially correlated component v_i follows an ICAR distribution as $v_i \sim \text{ICAR}(\sigma_v^2)$. A normal distribution is also considered for the temporally uncorrelated component γ_t as $\gamma_t \sim N(0, \sigma_\gamma^2)$ and an AR(1) distribution is given to the temporally correlated component ζ_t as $\zeta_t \sim N(\rho_\zeta \zeta_{t-1}, \sigma_\zeta^2)$ with $0 < \rho_\zeta < 1$. Note that there are many alternative dependence structures for ψ_{it} to consider. By fitting the residual model in Equation (3), the estimation of ψ_{it} can be obtained using its posterior mean and then is used as input in the following stage.

In the third stage, the covariate information along with the estimated space-time component $\hat{\psi}_{it}$ from the residual model (3) is incorporated into the following restricted model

$$\log(\theta_{it}) = \beta_0 + \mathbf{X}_{it}'\boldsymbol{\beta}_{it} + \hat{\psi}_{it} + \varepsilon_{it}, \tag{4}$$

where the coefficient parameter $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_{C(i),t}$ has the same hierarchical structure as that of the covariates-only model in the first stage. The random component $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$ is the uncorrelated space-time interaction term, representing the structure not captured by the estimated space-time component, $\hat{\psi}_{it}$. We mainly focus on the estimation of $\beta_0, \boldsymbol{\beta}_{it}, C(i), \delta_{sp}$, and λ_{stp}, which can be obtained from this final model in the third stage. As illustrated by Lawson et al,[18] the multi-stage method could yield lower error in the estimation of regression parameters than single-stage model, which is also shown in our simulation study.

2.2 | Bayesian inference

For the covariates-only model (Eq. 1) in the first stage and the restricted model (Eq. 4) in the third stage, the prior distributions of the intercept parameters β_0 and μ_p (Eq. 2) are given by $\beta_0 \sim N(0, \sigma_0^2)$ and $\mu_p \sim N(0, \sigma_{\mu_p}^2)$ respectively. Following Gelman[23] who evaluated the choice of prior distributions for variance parameters, we specify the standard deviation parameters in the model to follow a uniform distribution as $\sigma_s, \sigma_{\eta_s}, \sigma_{\lambda_{sp}}, \sigma_{\xi_{sp}} \sim \text{Uniform}(0, c)$, where c is a constant. The parameter $\rho_{\lambda_{sp}}$ in the temporal structure of λ_{stp} is considered to be $\rho_{\lambda_{sp}} \sim \text{Uniform}(0, 1)$. For both models, the likelihood functions of the observed counts $\mathbf{Y} = \{Y_{it}, i = 1, \ldots, N, t = 1, \ldots, T\}$ are expressed respectively as

$$\begin{aligned} p(\mathbf{Y}|\Theta_1) &= \prod_{i=1}^{N}\prod_{t=1}^{T} \text{Poisson}(Y_{it}|E_{it}, \mathbf{X}_{it}, \beta_0, \boldsymbol{\beta}_{C(i),t}), \\ p(\mathbf{Y}|\Theta_3) &= \prod_{i=1}^{N}\prod_{t=1}^{T} \text{Poisson}(Y_{it}|E_{it}, \mathbf{X}_{it}, \beta_0, \boldsymbol{\beta}_{C(i),t}, \hat{\psi}_{it}), \end{aligned}$$

where Θ_1 and Θ_3 are respectively the sets of parameters included in the covariates-only model in the first stage and the restricted model in the third stage. Then the posterior inference for the parameters Θ_1 and Θ_3 are seperately obtained based on the likelihood function and the prior distribution by

$$p(\Theta_i|\mathbf{Y}) \propto p(\mathbf{Y}|\Theta_i) \times f(\Theta_i), i = 1, 3,$$

where $f(\Theta_i), i = 1, 3$ are the joint prior distributions of the parameters $\Theta_i, i = 1, 3$.

In the residual model (Eq. 3) of the second stage, the standard deviation parameters $\{\sigma_r, \sigma_u, \sigma_v, \sigma_\gamma, \sigma_\zeta\}$ are considered to follow uniform distributions. The parameter ρ_ζ in the temporal structure of ζ_t is specified as $\rho_\zeta \sim \text{Uniform}(0, 1)$. The likelihood function is expressed as

$$p(\hat{r}_{it}|\Theta_2) = \prod_{i=1}^{N}\prod_{t=1}^{T} \text{Normal}(\hat{r}_{it}|\sigma_r^2, u_i, v_i, \gamma_t, \zeta_t),$$

where $\Theta_2 = \{\sigma_r, \sigma_u, \sigma_v, \sigma_\gamma, \sigma_\zeta, \rho_\zeta\}$ includes all the parameters in the residual model. Similarly, the posterior distribution of Θ_2 is obtained based on the likelihood and the prior distribution as

$$p(\Theta_2|\hat{r}_{it}) \propto p(\hat{r}_{it}|\Theta_2) \times f(\Theta_2),$$

where $f(\Theta_2)$ is the joint prior distribution of the parameter Θ_2.

Bayesian inference for the parameters $\Theta_i, i = 1, 2, 3$ is separately conducted in the three stages, using MCMC algorithms via statistical software WinBUGS and R package Nimble. As stated before, our primary interest is spatial clustering patterns and covariate effects, involving the estimation of parameters $\beta_0, \boldsymbol{\beta}_{it}, C(i), \delta_{sp}$, and λ_{stp}, which are finally drawn from the third stage. In the simulation and real data analysis, the convergence of MCMC sampling is checked by trace plots, autocorrelation funtions and Geweke convergence diagnostics. In addition, a sensitivity analysis for the priors is done in Section 4.3 to examine how sensitive the posterior distributions are to the choice of priors.

Considering that the spatial cluster indicator $C(i)$ is a nominal value, we use its posterior mode for estimation. To determine the subset of significant covariates in each cluster, the estimate of marginal posterior selection probability, $P(\delta_{sp} = 1|\mathbf{Y})$, is employed. If the estimate is greater than or equal to a cutoff value, then the corresponding covariate is selected into the model for cluster s. Otherwise, $\hat{\delta}_{sp} = 0$ and the corresponding pth covariate is not selected in cluster s. In this article, different cutoff values are considered to conduct sensitivity analysis in the simulation study of Section 4. Other parameters in the model are estimated by the posterior means.

3 | MODEL ASSESSMENT CRITERIA

For model comparison, a variety of assessment criteria are considered in this section. To evaluate the performance of the models in recovering spatial clusters, we develop several measures depending on the estimated cluster indicators. The binary entry parameter δ_{sp} is used to examine the accuracy of variable selection and coefficient estimation. Also, several kinds of goodness-of-fit measures are presented based on the posterior distributions. The proposed criteria are used in the simulation study and real data analysis.

3.1 | Cluster detection accuracy

Let C_i^T denote the true spatial cluster indicator for area i and $\hat{C}_{ik}$ denote the estimated indicator for area i in the kth simulation, where $k = 1, \ldots, K$. The accuracy rate of cluster identification for area i is computed over simulations as

$$AT_i = \frac{1}{K} \sum_{k=1}^{K} I(\hat{C}_{ik} = C_i^T),$$

where $I(\cdot)$ is an indicator function. The measure AT_i evaluates the capability of spatial cluster recovery for area i. In order to obtain an overall assessment of the cluster detection accuracy rate, the average value over areas is defined as: $\overline{T} = \frac{1}{N} \sum_{i=1}^{N} AT_i$, which can be viewed as a representative measure for cluster identification.

To further investigate the performance in each cluster, the detection accuracy rate is developed for each spatial group as follows:

$$GT_s = \frac{1}{N_s} \sum_{\{i: C_i^T = s\}} AT_i,$$

where N_s denotes the number of areas in cluster $s, s = 1, \ldots, S$. Thus, GT_s measures the rate of correctly classified areas in cluster s. Both $\overline{T}$ and GT_s are used in the simulation study to compare the clustering performance of different models.

3.2 | Variable selection and coefficient estimation accuracy

In this paper, the binary entry parameter δ_{sp} controls the selection of the pth covariate in cluster s. To assess the spatial performance of variable selection, the local selection accuracy rate for pth covariate is proposed as

$$\mathrm{SA}_p = \frac{1}{SK} \sum_{s,k} I(\hat{\delta}_{spk} = \delta_{sp}),$$

where $I(\cdot)$ is an indicator function. $\hat{\delta}_{spk}$ is the estimation of the selection parameter δ_{sp} in the kth simulation. As discussed before, if the marginal posterior inclusion probability is greater than the cutoff value, then $\hat{\delta}_{spk} = 1$; otherwise, $\hat{\delta}_{spk} = 0$. This measure reflects the behavior of spatial variable selection for each covariate over clusters and simulations.

For the coefficient estimation performance, we employ the posterior means of the regression parameters and their 95% credible intervals. Mean absolute error(MAE) and mean squared error(MSE) for each covariate are computed as

$$\text{MAE}_p = \frac{1}{STK}\sum_{s,t,k}|\hat{\beta}_{stpk} - \beta_{stp}|,$$
$$\text{MSE}_p = \frac{1}{STK}\sum_{s,t,k}(\hat{\beta}_{stpk} - \beta_{stp})^2,$$

where $\hat{\beta}_{stpk}$ is the posterior mean of coefficient β_{stp} in the kth simulation. The average coverage probability(CP) for 95% intervals is also utilized for evaluation.

3.3 | Goodness-of-fit measures

Based on the posterior distribution of the deviance function, DIC is a widely used criterion measuring the fitness and complexity of a Bayesian model. Celeux et al[24] also suggested the use of DIC_3 as an alternative of DIC in mixture models. DIC and DIC_3 can be easily computed by MCMC algorithms and provide stable evaluations for mixture models.

In terms of the prediction performance of the models, we examine the marginal predictive likelihood (MPL) defined as

$$\text{MPL} = \sum_{i=1}^{N}\sum_{t=1}^{T}\log\left(\text{CPO}_{it}\right),$$

where conditional predictive ordinate CPO_{it} is the marginal posterior predictive density of y_{it} given the data except y_{it} and can be easily obtained by one-time estimation run based on the posterior samples.

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

$$\text{MSPE} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Y_{it} - \hat{Y}_{it}\right)^2,$$

where Y_{it} and $\hat{Y}_{it}$ are respectively the observed value and predicted value generated from the posterior predictive distribution.

In general, the model with a smaller value of DIC, DIC_3, MSPE and a larger value of MPL provides better fitness for the data.

4 | SIMULATION STUDY

In this section, we explore the performance of the proposed multi-stage hierarchical model and several alternatives based on a range of assessment criteria presented in the previous section.

4.1 | Data generation

The simulated data are generated under six design schemes with different spatio-temporal domains and correlation patterns.

In Designs 1 and 2, the state of Georgia, USA is used as the spatial domain of interest, which contains $N = 159$ counties and has been studied in terms of different small area health data.[25,26] The motivating low birth weight data in Georgia is considered in the real data analysis of Section 5. In both designs, the outcomes are generated from the following Possion distribution:

$$y_{itk} \sim \text{Poisson}\left(e_{itk}\theta_{itk}\right),$$
$$\log\left(\theta_{itk}\right) = \beta_0 + \sum_{p=1}^{P}\beta_{itp}x_{itpk} + u_{ik} + v_{ik},$$

where y_{itk} is the simulated count in area i at time t in the kth simulation, $i = 1,\ldots,N$, $t = 1,\ldots,T$ and $k = 1,\ldots,K$. K is the number of simulations. There are $T = 8$ time points and $P = 4$ covariates. The expected count e_{itk} is generated independently from a Uniform distribution, Uniform(5,10) and $\beta_0 = 1$. There are spatial components u_{ik} and v_{ik} in the random effects, which are generated respectively as $u_{ik} \sim N(0, 0.05^2)$ and $v_{ik} \sim \text{ICAR}(0.05^2)$. In Disign 1, the covariates x_{itpk} for $p = 1, 2, 3$ are independently generated from the normal distribution $N(0, 0.2^2)$, while in Design 2, the covariates are simulated from multivariate normal distribution $N(0, \Sigma)$ with $\Sigma = (\sigma_{ij})_{3\times3}$, $\sigma_{ij} = 0.2^2[I(i = j) + r^{|i-j|}I(i \neq j)]$. The parameter r is set to be 0.5 or 0.8 to explore the influence of moderate or high variable correlation. Following Choi and Lawson,[27] the fourth covariate in both designs is set as $x_{it4k} = v_{ik} + z_{itk}$ with $z_{itk} \sim N(0, 0.2^2)$ to consider the spatial confounding problem.

The regression coefficient β_{itp} is group-specific for both designs. If area i is assigned into group s, i.e. $C(i) = s$, then $\beta_{itp} = \beta_{stp} = \delta_{sp} \times \lambda_{stp}$. Figure 1 displays the spatial design of the cluster indicator C_i and the significant covariates in each cluster in

the domain of Georgia. There are $S = 6$ spatial clusters for Designs 1 and 2 in the left panel of Figure 1. The areas in the same cluster are not always set to be adjacent to explore the cluster detection performance of the proposed model. And it is easy to find that the subset of significant covariates varies across spatial groups. For example, cluster 1 is at the upper-left corner of the map with significant covariates X_1 and X_2, that is, $\delta_{11} = \delta_{12} = 1$ and $\delta_{13} = \delta_{14} = 0$, while cluster 2 at the upper-right corner has significant covariates X_1 and X_3.

Each significant covariate effect has a certain temporal pattern in the cluster, with an AR(1) structure $\lambda_{stp}|\lambda_{s,t-1,p} \sim N(\rho_{\lambda_s}\lambda_{s,t-1,p}, 1)$. The temporal parameters are specified as $(\rho_{\lambda_1}, \ldots, \rho_{\lambda_S}) = (0.8, 0.7, 0.6, 0.5, 0.4, 0.3)$ for the covariates to obtain distinct temporal patterns in the clusters. For example, the first covariate X_1 is designed to have significant effects in cluster (1,2,4,6), and the solid lines in Figure 3 present the designed variation of $\lambda_{st1}, s = 1, 2, 4, 6$, the true temporal profiles of the first covariate effect in the selected clusters 1,2,4 and 6. The temporal patterns of the other three covariate effects for X_2, X_3 and X_4 have similar dynamic structures and are omitted for brevity. Therefore, the regression coefficients have spatio-temporally dependent structures in the sense that the temporal patterns of covariate effects vary with spatial groups.

In Designs 3 and 4, the data are simulated with more complicated correlation structures and cluster patterns based on the Georgia state map. The counts are obtained from the following model with constant relative risks over simulations:

$$y_{itk} \sim \text{Poisson}\left(e_{it}\theta_{it}\right),$$
$$\log\left(\theta_{it}\right) = \beta_0 + \sum_{p=1}^{P} \beta_{itp}x_{itp} + u_i + v_i + \gamma_t + \zeta_t,$$

where temporally random component $\gamma_t \sim N(0, 0.1^2)$ and $\zeta_t \sim N(0.4\zeta_{t-1}, 0.1^2)$. In this case, we set $T = 12$ and $P = 5$ with the first four covariates as well as $\{e_{it}, \beta_0, u_i, v_i\}$ generated from the same scheme as Design 1. Following from Lawson et al,[18] the fifth covariate X_{it5} is generated with spatial or temporal dependence. Specifically, X_{it5} is simulated from an AR(1) distribution with temporal parameter 0.3 in Design 3, while in Design 4, X_{it5} varies over both space and time, generated from a normal distribution with mean 0 and covariance $\Sigma_{X_5} = 0.1\Sigma_S \otimes \Sigma_T$, where Σ_S and Σ_T are respectively the covariance matrices of the ICAR distribution with overall variance 0.1 and the AR(1) distribution with temporal parameter 0.7. For both designs, the spatial clusters and significant covariates are allocated as the right panel of Figure 1 with S=9 groups. The cluster patterns are rather irregular with many scattered areas. The temporal trends of the covariate effects in the first six clusters are set to follow the same AR(1) patterns as Design 1. In clusters 7-9, an AR(2) structure is given by $\lambda_{stp}|\lambda_{s,t-1,p}, \lambda_{s,t-2,p} \sim N(\rho^{(1)}_{\lambda_s}\lambda_{s,t-1,p} + \rho^{(2)}_{\lambda_s}\lambda_{s,t-2,p}, 1)$ for $p = 1, \ldots, 5$, $s = 7, 8, 9$. The autoregression coefficient pairs $(\rho^{(1)}_{\lambda_s}, \rho^{(2)}_{\lambda_s})$ for clusters 7-9 are respectively specified as $(0, 0.6), (0.3, 0.5), (-0.2, 0.4)$.

In Designs 5 and 6, the spatial domain of interest is further enlarged with a much larger sample size to evaluate the scalability of the proposed method. Specifically, the data are generated for 616 counties in seven US states: Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina and Tennessee. The responses are generated from the Poisson distribution similar to Design 1 with $T = 12$ and $P = 4$, except that the spatial cluster patterns are respectively designed with $S = 7$ for Design 5 and $S = 10$ for Design 6. The spatial allocation and its corresponding significant covariates are displayed in Figure 2. For the temporal trends of the covariate effects, an AR(1) pattern is specified for Design 5 with autocorrelation parameter $(\rho_{\lambda_1}, \ldots, \rho_{\lambda_S}) = (0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2)$. For Design 6, the coefficient trends in the first seven clusters are the same as above, while in clusters 8-10, AR(2) structures are respectively given with parameter pairs $(\rho^{(1)}_{\lambda_s}, \rho^{(2)}_{\lambda_s}) = (-0.2, 0.5), (0.3, -0.4), (0.6, 0.3)$, $s = 8, 9, 10$.

We generate $K = 200$ data sets and for each simulated data, different competing models are fitted. The determination of the number of the spatial clusters is an open question in general. Many previous studies are devoted to solving this problem, but often have computational burdens in the analysis of small areas.[28,29] In this paper, we overcome this problem in virtue of various model assessment criteria, which is convenient and user-friendly.

The models under consideration are as follows:

(1) **Model 1**: Bayesian multi-stage hierarchical model proposed in this paper.

(2) **Model 2**: Bayesian full model with spatial clustering and variable selection which does not employ the multi-stage strategy. (i.e. single-stage model: $\log(\theta_{it}) = \beta_0 + \mathbf{X}'_{it}\boldsymbol{\beta}_{it} + \psi_{it} + \varepsilon_{it}$ and $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_{C(i),t}$ as above)

(3) **Model 3**: Bayesian multi-stage hierarchical model with variable selection,[27] in which there is no spatial cluster indicators. (i.e. in the first stage, $\log(\theta_{it}) = \beta_0 + \mathbf{X}'_{it}\boldsymbol{\beta}_{it}$, $\beta_{itp} = \delta_{ip} \times \lambda_{tp}$, where δ_{ip} is the binary entry parameter for each small area.)

(4) **Model 4**: Bayesian multi-stage hierarchical model with spatial cluster indicators, which does not consider variable selection in the clusters.[9] (i.e. in the first stage, $\log(\theta_{it}) = \beta_0 + \mathbf{X}'_{it}\boldsymbol{\beta}_{it}$, $\boldsymbol{\beta}_{it} = \boldsymbol{\beta}_{C(i),t}$, $\beta_{stp} \sim N(\rho_{sp}\beta_{s,t-1,p}, \sigma^2_{sp})$)

We fit these models using Bayesian inference based on the posterior samples of 40,000 iterations with a single chain via MCMC. After the burn-in period of 10,000 samples, a thinning rate of 10 is used to avoid high autocorrelation of the samples. The chains from the trace plots are stationary and well-mixed. Effective sample sizes are also large enough to ensure convergency. Finally, 3000 iterations are obtained to make parameter estimation. As for the binary entry parameter δ_{sp} for covariate selection, we first consider the cutoff value as 0.5 following the previous studies.[18,27,30] Sensitivity analysis on the cutoff value will be given later in this section to verify the reasonability of the choice. The results presented below are averaged over simulations.

4.2 | Simulation results

To evaluate the performance in selecting the ture number of spatial clusters by assessment criteria, the paper first uses a range of alternative cluster numbers to fit the model. Table 1 summarizes the percentage of the cluster numbers selected by DIC_3 criterion for Models 1,2,4 in 100 simulations. Evidently, in all the designs, the true numbers of clusters could be correctly selected in most cases, indicating the reliability of DIC_3 criterion. The results of Model 1 (the proposed model) are consistently good, unaffected by the complex temporal or spatial patterns in Designs 3-6. The high correlation of the covariates has a mild influence on the cluster selection, which is still within reason. Thus, the following simulation results are displayed under the right setting of S for the considered designs.

Table 2 presents the results of the four models in terms of goodness-of-fit measures, including the average of DIC, DIC_3, MPL and MSPE over the simulations. In Design 1, Model 4 has the smallest DIC, but larger MSPE value and much smaller MPL value. Only with a slight difference, Model 1 has the second smallest DIC. And along with the DIC_3 criterion, both MPL and MSPE favor Model 1. In Design 2, the model performance decreases to a certain extent with the gradual increase of variable correlation (r), but the proposed model (Model 1) still behaves well especially at the moderate correlation. In Design 3, Model 1 outcompetes the alternatives, indicating that the proposed multi-stage hierarchical model fits the data well. In Design 4, the selection results of the criteria are a little inconsistent, with MPL favoring Model 2 and MSPE favoring Model 4, but the performance of Model 1 is still good, since the difference between the models is rather small. The results are similar in Designs 5 and 6, though the spatial domain and cluster allocation are much more complicated, showing the good scalability of the proposed method.

Table 3 shows the performance of Models 1, 2 and 4 in terms of the cluster detection accuracy proposed in Section 3.1, where $GT_s, s = 1, \dots, S$ measures the rate of correctly classified areas in cluster s. From Table 3, we can see that the cluster identification results of Model 2 are the worst among the three models, since it is based on single-stage strategy, not taking the spatial confounding problem into consideration and may have inevitable estimation bias. The performance of Model 4 is a little better, but still much inferior to Model 1. The results in Design 2 with $r = 0.8$ show that highly correlated covariates may have a slightly negative effect on the cluster detection. The distributions of spatial clusters in Design 6 could influence the performance of the alternative methods, but the proposed model consistently produces stable results. The cluster detection measures prefer Model 1 in all the designs. For illustration, the map of AT_i obtained from Model 1 in Design 1 is displayed in Figure 4, which shows that most counties in Georgia have relatively high cluster detection accuracy rates and the overall identification of spatial clusters is satisfactory.

Tables 4-10 respectively display the coefficients estimation performance of the considered models in Designs 1-6 based on MAE, MSE and CP measures. In all the designs, Model 1 overall behaves well with smaller MAE and MSE values than the competing models. In Design 1, MAE picks Model 4 for the estimation of β_2 and MSE picks Model 3 for β_3, but the slight differences do not have significant effects. In Design 2, the performance of the proposed method is relatively robust to the variable correlation, though there is a slight possibility to be affected by highly correlated covariates. It's worth noting that in Designs 1-6, the covariate X_4 is set to be highly correlated with the spatial random component, for which Model 2 (the single-stage model) has poor parameter estimation performance. In Designs 3-4, the spatial and temporal dependence in X_5 also has a negative impact on the competing models. By contrast, Model 1 obviously improves the estimation accuracy, especially for β_4 and β_5. Thus, the proposed multi-stage method really reduces the confounding bias. In Designs 5-6, where the spatial domain is much larger ($N = 616$), the coefficient estimation results of the method are still comparable to those in Design 1, which shows the broad applicability of the proposed model. Therefore, Model 1 generally provides outstanding estimation performance in the considered simulation designs.

Taking one simulation result in Design 1 for example, Figure 3 presents the true temporal trends of the first covariate effects in the selected four clusters and the estimated ones with 95% credible intervals from Model 1. Clearly, the true profiles are exactly contained in the credible intervals of the estimated temporal trends and the interval widths are overall small. Similarly, plots for the other regression coefficients also indicate that Model 1 fits the true temporal trends of coefficient parameters well.

Table 11 displays the accuracy rates of spatial variable selection for Models 1-3 in Design 1. The measure $SA_p, p = 1, \ldots, P$, defined in Section 3.2, depends on the choice of the cutoff value. Thus, sensitivity analysis is conducted based on different cutoff values: (0.5, 0.65, 0.8). The "Average" value in the table is obtained by Average= $\sum_{p=1}^{P} SA_p/P$. The bigger values of the measures indicate the better performance of the model for selecting the true covariates. It can be seen that the selection accuracy rates decrease with the larger cutoff values. All the models have poor performance when the cutoff value is set as 0.8. This suggests that the value of 0.5 we considered before is a reasonable choice, which is consistent with the previous findings of Barbieri and Berger.[30] With the cutoff value fixed at 0.5, the variable selection results in Designs 2-6 are presented in Table 12. Overall, the proposed method behaves well, with high adaptability to different design scenarios, including the moderate or high variable correlation in Design 2, complex spatiao-temporal correlation structure in Designs 3-4 and much larger spatial domain in Designs 5-6. The accuracy rates of Model 1 are consistently higher than the other two models, reflecting that both the spatial cluster identification and the multi-stage estimation strategy are necessary for spatio-temporal data modeling.

A common problem with MCMC simulation is computation time. Table 13 displays the run times for the various simulation settings, which are recorded using a desktop PC with Intel(R) Core (TM) i9 process (3.70 GHz) and 128 GB RAM. It is seen that the proposed method is quite efficient. The computation time is mainly influenced by the sample size and the number of time points while the correlation structure of the variables and the number of clusters do not have an obvious effect on it. The computation efficiency is relatively high and all of the run times are within a reasonable range, indicating the good scalability of the proposed method.

4.3 | Prior sensitivity analysis

A sensitivity study on priors is performed to examine the robustness of the above results. In this section, the focus is on the priors set for the parameters of most interest. Specifically, the inverse gamma distributions IG(1,1) and IG(1,0.5) are considered as alternative priors for the variance parameters in Section 2.2, where IG(1,0.5) is more informative than the original prior and IG(1,1) is more diffuse. The temporal parameters $\rho_{\lambda_{sp}}$ and ρ_{ζ} are also reconsidered to follow Beta priors Beta(2,2) and Beta(2,4). In addition, combinations of these alternative priors are specified to explore the comprehensive influence of different priors on model results.

Compared with the original setting, it is found that the alternative sets of priors produce a bit different posterior estimates for the parameters of interest. According to the previous studies, a visual inspection of the posterior density plots is a useful tool for comparation, which reveals that in this case, only slight shifts exist and the results are substantively comparable. Therefore, the priors do not largely influence the actual conclusions and the method is relatively robust to different prior settings.

5 | REAL DATA ANALYSIS

In this section, the proposed method is respectively applied to the motivating public health data: the low birth weight (LBW) incidence data in Georgia, USA and the circulatory disease data in England. For these real data, we also compare the proposed model with the other three competing models mentioned in Section 4. The Bayesian posterior samples of 30,000 iterations were obtained after a burn-in period of 10,000 samples. The convergence of samples is checked through trace plots and Geweke diagnostic tools. Using a thinning rate of 10 to reduce sample dependence, we finally conducted parameter estimation based on a total of 3000 samples.

5.1 | Low birth weight data

5.1.1 | Data description

Low birth weight (infant weight at birth of less than 2500g) is a widely-used measure for population health. We obtained the LBW incidence data in 159 counties of Georgia for the years 2007-2018. The counts are publicly available from the Georgia health information system Oasis(http://oasis.state.ga.us/). Figure 5 displays the spatio-temporal distributions of standardized incidence ratios(SIR), defined as the number of LBW births divided by the number of expected cases calculated by the internal standardization method.[31] From Figure 5, we can see that overall, the incidence ratios in central and southern areas of Georgia are higher than other areas. And the number of counties with low incidence ratios is increasing over time. There is a significant reduction for the risk ratios in northern areas across the years. For southeastern areas, the incidence ratios decrease at first but

then increase, with lower ratios in the years 2013-2015. Thus, the standardized incidence ratios have an obvious spatio-temporal variation in the domain of Georgia, which motivates us to explore the underlying correlation structures and the contributing risk factors.

According to the previous studies[5,10] and data availability, we considered the county-level population density (defined as population divided by the total land area in square miles), social development index (SDI), median household income, and unemployment rate as socioeconomic predictors of low birth weights, which were obtained from the US Census Bureau and the US Bureau of Labor Statistics. The social development index provides a comprehensive assessment of the neighborhood environment where women live and work, encompassing household composition, minority status, housing type and transportation.[32] Each county receives a SDI percentile ranking value ranging from 0 to 1, with higher values indicating better social environments. For other socio-demographic and behavioral risk factors, we adopted the proportion of mothers with less than 12th grade education and the proportion of mothers smoking during pregnancy as the possible related covariates. Preliminary analysis showed that among the above covariates, median household income, unemployment rate and social development index are moderately correlated with each other, which could be well handled by the proposed method according to the simulation results. The considered covariates also have spatio-temporal variations and Figure 6 displays the maps of the predictors for the year of 2018, showing significant spatial patterns. Thus, it is reasonable to consider the spatio-temporally varying coefficients of the predictors which can be used for spatial clustering and variable selection.

5.1.2 | Analysis results

To determine the best number of spatial clusters, we conducted a series of comparisons with the proposed model based on different number of clusters. Considering the regional characteristics of Georgia and spatial variation of the covariates, the number of clusters is set to range from 3 to 12. The most suitable choice is made according to the model assessment criteria. Figure 7 displays the variation of DIC, DIC_3, MPL and MSPE for the proposed model with different number of clusters. Although the values of DIC and DIC_3 are similar, DIC measure picks the model with 6 groups while DIC_3 reaches its minimum value with the number of 5. Among the considered models, MPL prefers the model with 5 or 7 clusters, while MSPE picks $S = 5$ as the optimal. We can see that the maximum number of clusters (12) does not provide the best performance for the model, and additional exploration shows that much larger number of clusters (>12) will only increase model complexity, but reduce fitness to the data. So it is necessary to make use of the goodness-of-fit measures to select a balance and here we choose $S = 5$ according to the model criteria above.

Then the proposed model is compared with the other three competing models mentioned in Section 4 with the number of spatial groups fixed at 5. Table 14 shows the performance of the models in terms of the goodness-of-fit measures for the LBW incidence data. Model 1 and Model 4 have similar DIC values and both are smaller than the other two models. Model 3 has the largest DIC and DIC_3 values, reflecting that the spatial clustering of covariate effects really exist and ignoring the underlying spatial structure has great negative impact on data fitting. With the smallest MPL and largest MSPE values, Model 2 behaves badly due to the possible confounding problems. Overall, DIC_3, MPL and MSPE measures prefer Model 1 (the proposed multi-stage model), showing that both spatial clustering and variable selection are worth considering for the LBW data.

The map of spatial cluster indicators from the best-fit Model 1 is displayed in Figure 8. The first group includes 32 counties, most of which are located in the Atlanta area. The second group has two separated parts with 33 counties in the northeast and southwest areas. The third group contains the largest number of counties, most of which are in the southeast corner of Georgia and a few are in the northernmost areas. The fourth and fifth groups include relatively scattered counties in different regions. From the map we can see that the proposed model provides a good spatial clustering for the counties and allows for discontinuity in each cluster at the same time.

Figure 9 shows the map of the estimated binary entry parameters $\delta_{sp}, p = 1, \ldots, 6$, i.e. variable selection results for the LBW data. It is seen that the effects of most covariates on the outcome are spatially varying with clusters except for median household income and the proportion of maternal smoking during pregnancy, which are globally significant in the domain of Georgia. Combining the cluster indicators in Figure 8, it is easy to find that population density is selected in clusters (1,5), where most large and medium-sized cities are situated. The social development index is chosen to be a significant factor in 3 clusters: (1,4,5), showing that SDI, the composite measure of neighborhood environments, plays a more important role in developed regions. Unemployment rate is only related to birthweight in clusters with relatively high concentration (generally, >8% unemployment rate), which is consistent with the finding of Pearl et al.[8] Low level of maternal education is another risk factor for low birthweight in clusters (3,4) mainly located in rural Georgia. Thus, the subset of significant covariates is spatially varying across clusters due to different regional conditions, indicating the necessity of variable selection for each cluster.

The estimated coefficient parameters β_{stp} are displayed in Figure 10. There are six plots, each of which represents the temporal trends of the corresponding covariate effects in the clusters. The spatial clusters are distinguished by the lines in different colors. It is obvious that the coefficients of the variables in the selected clusters are significantly larger than those in the unselected ones, since the coefficient lines corresponding to the unselected clusters are much closer to the horizontal axis in the plots. This shows that the proposed model can correctly identify the important risk factors for each cluster.

From Figure 10, it is known that higher population density, social development index and median household income are connected with decreased disease relative ratios, while the remaining covariates are positively associated with the LBW ratios in Georgia. In addition, there are significant temporal patterns for the covariate effects, which vary across clusters and covariates. In particular, the increasing effects of social development index highlight the importance of community environments in shaping pregnancy outcomes. The degree to which a community exhibits certain social conditions, including low percentage of vehicle access, crowded households, or limited English proficiency, may differentially influence the living conditions of pregnant women, such as food security and healthcare utilization, manifesting in higher risks of preterm delivery or low birth weights.[33,34] Median household income is a global risk factor for low birth weight with greater influence in rural areas. It is reported that low-income women are more likely to delay initiation of prenatal care and have inadequate nutritional consumption during pregnancy, which often negatively affects the fetal weight gain. The cumulative effects of poverty over time undoubtedly have played a role in increasing the risk of low birth weight.[8,35] High unemployment in some regions indicates the deterioration of social environments and may further contribute to problems of social organizations, such as community support for health-promoting behaviors and access to health care or nutritious food. Community joblessness may also serve as the external stressor in the environment, resulting in high levels of psychological distress especially for pregnant women.[35] Mothers with less than a high school education tend to have a higher incidence of low birth weight in rural Georgia since they are more likely to be young, exhibit unwholesome individual behaviors, and have less access to prenatal care, which interact with each other and together affect the gain of adequate weight during pregnancy. Maternal smoking is associated with a reduction in birthweight, and has maintained a relatively constant relationship to birthweight over the study period. The different magnitudes in the clusters may be affected by the average frequency of smoking during pregnancy and other related anthropometry characteristics in the specific areas.[36] It has been studied that smoking hampers delivery of nutrients or oxygen to the uteroplacental circulation, which will slow down the normal fetal growth.[35] Therefore, a variety of factors are clearly linked to low birth weight with distinct covariate effects, which are helpful in identifying high-risk groups and mobilizing resources for intervention.

The maps of the estimated standardized incidence ratios (SIR) for the LBW data based on the proposed model are presented in Figure 11 for the year of 2008, 2012 and 2016. Clearly, the estimates have captured the main geographical patterns and temporal behaviors in the domain of Georgia. The incidence ratios in the southern areas are overall higher than those in the north. And the risks of low birth weight are decreasing across the years, especially in the central and northern counties.

5.2 | England circulatory disease data

5.2.1 | Data description

This section studies the circulatory disease in England, which is one of the largest causes of death worldwide according to the World Health Organization (www.who.int/mediacentre/factsheets/fs310/en/). The Health and Social Care Information Centre records the yearly counts of hospital admissions by local and unitary authorities (LUAs), which are considered as the responses in the model, i.e. $\{Y_{it}\}$ for $i = 1, \ldots, N(N = 323)$ LUAs in England in year $t = 1, \ldots, T(T = 10)$. The expected numbers of patients $\{E_{it}\}$ are calculated for each LUA and each year to adjust for the different population sizes and demographic structures.

The standardized incidence ratio $\text{SIR}_{it} = Y_{it}/E_{it}$ could serve the exploratory analysis of the disease risk. Figure 12a displays the averaged standardized incidence ratios of the circulatory disease over all years. In most adjacent areas, the disease risks are smoothly changing while there are also evident spatial step changes with higher risks, especially around Newcastle city in the north and Manchester city in the west. This spatial structure motivates us to explore the underlying risk factors and possible cluster patterns. Following from Rushworth et al,[7] the paper focuses on the social determinants of the circulatory disease, including socio-economic deprivation (poverty), air pollution and urbanization level as covariates of interest. Poverty is measured by the percentage of working age population who are in receipt of Jobseekers Allowance from the Health and Social Care Information Centre. The particulate matter PM_{10} concentrations are collected from the Department for Environment, Food and Rural Affairs on the LUA scale, while the urbanization level of each LUA is measured by the proportion of super-output areas within each LUA. Figure 12b-d displays the spatial variations of the covarates which are averaged over years. There are

obvious spatial clustering patterns in the maps, which may contribute to the incidence of circulatory disease. Therefore, spatio-temporal analysis is needed to capture the differential exposures and health outcomes in different areas so as to assess efforts for mitigating risks and targeting interventions towards the high-risk areas.

5.2.2 | Analysis results

The number of spatial clusters is selected based on the model assessment criteria, with alternative numbers ranging from 4 to 13. Both DIC and DIC_3 measures prefer the model with 7 clusters, while MPL picks the one with 6 clusters and MSPE reaches its relatively low points when $S = 7$ and 8. It is found that if the number of clusters is further increased, the goodness of model fit will decrease. Thus the best fit number of clusters is chosen to be $S = 7$.

The proposed model is also compared with the forementioned competing models to explore the importance of certain model components. Only using single-stage modeling process, Model 2 has larger DIC_3 and MSPE values. The performance of Model 3 is poor in terms of the measure of DIC, possibly because it does not take into account spatial clustering patterns of the variable coefficients. Model 4 behaves a little better, but is still inferior to Model 1 based on MPL and MSPE criteria. On the whole, the spatio-temporal structure and variable selection technique in the proposed model are rather necessary for the analysis of circulatory disease data.

Based on the proposed method, the estimated spatial clustering pattern is displayed in Figure 13. It suggests the presence of substantial heterogenity in terms of the covariate effects. In particular, there is evident septation in the central band of northern England, which is somewhat consistent with the distribution of standardized incidence ratios in Figure 12a. According to the cluster map, the three largest cities in England are classified into different clusters, that is, London in cluster 6, Birmingham in cluster 4 and Manchester in cluster 3, thus all of which are associated with different contributing factors for the incidence of circulatory disease.

The variable selection results and the corresponding coefficient variation patterns of the significant covariates are respectively shown in Figures 14 and 15. The influence of socio-economic poverty is significant in the majority of England (clusters 1,3,4,5,6), where more than 13% of the population receives Jobseekers Allowance, showing a high proportion of people living in poverty in these areas. It is seen from Figure 15 that severer economic deprivation is associated with increased risk of hospital admissions. Overall poverty rates in the selected clusters have been fairly persistent for years, showing poverty in these regions is deeply entrenched. The coefficients of poverty have upward temporal trends in most of the clusters, which indicates the gradually enhanced effects of socio-economic deprivation over years. The increased circulatory disease prevalence in people with higher poverty is attributable to a constellation of social, behavioral and psychosocial factors that are more widespread in disadvantaged individuals.[37] For example, poor areas generally have fewer food outlets and supermarkets, resulting in limited access to fresh vegetables and higher cost of healthy food. In addition, lower-income individuals have more stress, including insecurity in housing, safety and so on, while also have fewer resources to deal with these challenges. The mismatch results in increased distress and risk behaviours, which may lead to circulatory disease.[38] The particulate matter (PM_{10}) mainly affects the eastern England in clusters (4,6,7) with relatively higher concentrations of air pollution. During the whole study period, the overall PM_{10} concentrations show a slight decline and then a significant increase. The middle panel in Figure 15 displays the dynamic effects of PM_{10} concentrations, indicating that serious air pollution will increase people's susceptibility to circulatory disease in various degrees. There is substantial evidence that the adverse effects of inhaled PM are associated with oxidative stress and a subsequent systemic inflammatory response, which could accelerate the occurrence of circulatory diseases.[39] The severity of marginal effects of PM_{10} concentrations can differ due to locale, length of exposure, and weather conditions experienced by varying populations. The effect of urbanization is globally significant in England with higher level of urbanization substantially facilitating the disease incidence. It's seen that the share of urban population is growing rapidly. The pathways through which urbanization affects health are complex and multifactorial.[40] There is a marked increase in consumption of energy-rich foods and occupational physical activity is less common in urbanized areas.[41] Urbanization has led to changes in patterns of human activity, diet, and social structures, with profound implications for chronic diseases. In general, the effect magnitudes increase with higher urbanization levels, which may also be influenced by social demographic structures.

From the results above, it is known that the significant covariates are spatially varying across clusters and the covariate effects are not constant, but temporally varying. The proposed method is successfully applied to the disease data, which allows the identification of geographical clusters that exhibit similar risk patterns. The analysis results could help public health experts to recognize the underlying risk factors for different spatial groups and provide detailed insight into the spatial structure, which enables health resources and public interventions to be targeted appropriately.

6 | DISCUSSION

In this paper, we develop a flexible Bayesian multi-stage hierarchical model with spatio-temporally varying coefficients for the space-time data. The proposed model simultaneously allows for identification of spatial clusters in each of which covariate effects have homogeneous temporal patterns as well as selection of significant covariates for each cluster. Different from the previous spatial clustering studies which primarily focus on the variation of response, the paper is interested in the distinct covariate effects. A set of spatially structured weights is employed to help determine the spatial grouping patterns, which relax the restriction that the areas in the same cluster must be adjacent. To achieve locally varying variable selection, a binary entry parameter is incorporated into the model which can flexibly switch in/out of the corresponding covarite.

Many existing variable selection approaches regard the important covariates as fixed across space. One of the major advantages of the proposed method is that the subset of significant risk factors is modeled to vary with different spatial clusters. The temporal trends of the covariate effects in the selected clusters are structured by an autoregression process, which could be easily replaced by a random walk or other dynamic processes if needed. The spatial dependence structures also have many alternative distributions to employ, which shows a good adaptability of the proposed model.

In the simulation study, the proposed model is compared with several competing models based on a series of assessment criteria. It is shown that the multi-stage method is very helpful to reduce the confounding bias and it is necessary to consider both spatial clustering and variable selection to better estimate the spatio-temporally varying coefficients and relative risks in the model. In the real data analysis, the best number of clusters needs to be specified in advance according to the goodness-of-fit measures. Despite its computational convenience, this strategy does not consider the uncertainty of the number of clusters when conducting statistical inference for the parameters of interest. Thus, future extensions may be to develop a new approach that enables simultaneous inference for the posterior distributions of the number of clusters and grouping patterns under the Bayesian framework.

The proposed model is applied to two motivating health data sets, both of which are found to simultaneously exist spatial dependence and heterogeneity. The clusters estimated from the model indicate that the associations between covariates and health outcomes are not constant over the whole space domain, but varying with groups. Therefore, in order to control the disease incidence, the government should play a leadership role in stimulating necessary discussion and action based on local conditions. For the low birth weight in Georgia, subsidized support is supposed to be provided for low-income expectant mothers to address their necessary personal needs. In northern areas, improvements of public services are important, including convenient vehicle access, early prenatal care enrollment and good nutritional status. In southeastern and southwestern areas, supplying more high-quality job and educational opportunities could make a difference to decrease the incidence of low birth weight. Considering the universal effects of smoking on birthweight, the state government should promote effective interventions to reduce such behavioral risks. Many of the established risk factors are amenable to prevention, and of these, many can be recognized before pregnancy occurs. For the circulatory disease in England, urbanization is an issue that needs special attention. Although urban settings have long been associated with human development and progress, they can also lead to significant health problems. Policies to mitigate the adverse health effects of urbanization need to develop structural interventions, such as urban planning to promote active lifestyles and agricultural projects facilitating healthy dietary changes. In the central and eastern regions, it is important to improve the socio-economic situation of the low-income groups. Clinicians and policymakers are supposed to devise targeted measures to achieve equity in healthcare access across the spectrum of income-strata. Essential strategies include promoting health education, reducing the cost of healthy foods, and other public policy efforts to provide targeted care to marginalized populations. Considering the burden of air pollution in the south, policies are needed to address the environmental issues, such as promoting more new energy vehicles and encouraging the use of public transportation. In addition, it is necessary to remind people to avoid high-intensity exercise in heavily polluted environments, which could be a simple strategy to mitigate the harmful effects of air pollution.

ACKNOWLEDGEMENTS

The work was partially supported by the National Natural Science Foundation of China (No.11861042), and the China Statistical Research Project (No.2020LZ25).

CONFLICT OF INTEREST

The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

We use publicly available data and the link to the data source is provided in the paper.

References

1. Lee D, Mukhopadhyay S, Rushworth A, Sahu SK. A rigorous statistical framework for spatio-temporal pollution prediction and estimation of its long-term impact on health. *Biostatistics.* 2017;18:370–385.

2. Napier G, Lee D, Robertson C, Lawson A. A Bayesian space-time model for clustering areal units based on their disease trends. *Biostatistics.* 2019;20:681–697.

3. Dai X, Yan Z, Tian M, Tang M. Quantile Regression for General Spatial Panel Data Models with Fixed Effects. *J Appl Stat.* 2020;47:45–60.

4. Dai X, Jin L, Tian M, Shi L. Bayesian Local Influence for Spatial Autoregressive Models with Heteroscedasticity. *Stat Pap.* 2019;60:1423-1446.

5. Kirby R, Liu J, Lawson A, Choi J, Cai B, Hossain M. Small area low birth weight incidence and socio-economic predictors: a latent spatial structure approach. *Spat Spatiotemporal Epidemiol.* 2011; 2(4):265-271.

6. Choi J, Lawson AB. A Bayesian two-stage spatially dependent variable selection model for spacetime health data. *Stat Methods Med Res.* 2019,28(9):2570-2582.

7. Rushworth A, Lee D, Sarran C. An adaptive spatiotemporal smoothing model for estimating trends and step changes in disease risk. *J R Stat Soc Ser C Appl Stat.* 2017;66(1):141-157.

8. Pearl M, Braveman P, Abrams B. The relationship of neighborhood socioeconomic characteristics to birthweight among 5 ethnic groups in California. *Am. J. Public Health.* 2001;91(11):1808-1814.

9. Choi J, Lawson AB, Cai B, Hossain MM, Kirby RS, Liu J. A Bayesian latent model with spatio-temporally varying coefficients in low birth weight incidence data. *Stat Methods Med Res.* 2012;21(5):445-456.

10. Cai B, Lawson AB, Hossain M, Choi J, Kirby RS, Liu J. Bayesian semiparametric model with spatially-temporally varying coefficients selection. *Stat Med.* 2013;32(21):3670-3685.

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

12. Nieto-Barajas LE. Bayesian regression with spatiotemporal varying coefficients. *Biom J.* 2020;62(5):1245-1263.

13. Lee J, Gangnon RE, Zhu J. Cluster detection of spatial regression coefficients. *Stat Med.* 2017;36:1118-1133.

14. Lee J, Sun Y, Chang HH. Spatial cluster detection of regression coefficients in a mixed-effects model. *Environmetrics.* 2020;31:e2578.

15. Kuo L, Mallick B. Variable selection for regression models. *Sankhya Ser B.* 1998;60:65-81.

16. Boulieri A, Bennett JE, Blangiardo MA. Bayesian mixture modeling approach for public health surveillance. *Biostatistics.* 2020;21:369-383.

17. Reich B, Hodges J, Zadnik V. Effects of residual smoothing on the posterior of the fixed effects in disease-mapping models. *Biometrics.* 2006;62:1197-1206.

18. Lawson AB, Choi J, Cai B, Hossain M, Kirby RS, Liu J. Bayesian 2-stage space-time mixture modeling with spatial misalignment of the exposure in small area health data. *J Agric Biol Environ Stat.* 2012;17:417-441.

19. Baer DR, Lawson AB. Evaluation of Bayesian multiple stage estimation under spatial CAR model variants, *J Stat Comput Simul.* 2019;89(1):98-144.

20. Bernardinelli L, Montomoli C. Empirical Bayes versus fully Bayesian analysis of geographical variation in disease risk. *Stat Med.* 1992;11:983-1007.

21. Besag J, York J, Mollie J. Bayesian image restoration with two applications in spatial statistics. *Ann Inst Stat Math.* 1991;43:1-59.

22. Adin A, Lee D, Goicoa T, Ugarte MD. A two-stage approach to estimate spatial and spatio-temporal disease risks in the presence of local discontinuities and clusters. *Stat Methods Med Res.* 2019;28(9):2595-2613.

23. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal.* 2006; 1: 515-533

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

25. Carroll R, Lawson AB, Faes C, Kirby RS, Aregay M, Watjou K. Spatio-temporal Bayesian model selection for disease mapping. *Environmetrics.* 2016;27(8):466-478.

26. Shannon J, Abraham A, Bagwell Adams G, Hauer M. Racial disparities for COVID19 mortality in Georgia: Spatial analysis by age based on excess deaths. *Soc Sci Med.* 2022;292:114549.

27. Choi J, Lawson AB. Bayesian spatially dependent variable selection for small area health modeling. *Stat Methods Med Res.* 2018;27:234-249.

28. Escobar MD, West M. Bayesian density estimation and inference using mixtures. *J Am Stat Assoc.* 1995;90:577-588.

29. Reich BJ, Bondell HD. A spatial Dirichlet process mixture model for clustering population genetics data. *Biometrics.*2011;67:381-390.

30. Barbieri MM, Berger J. Optimal predictive model selection. *Ann Stat.* 2004;32:870-897.

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

32. How Does Place Affect Health? Agency for Toxic Substances and Disease Registry. https://www.atsdr.cdc.gov/ placeandhealth/howdoesPlaceaffectHealth.html. Accessed June 7, 2023.

33. Givens M, Teal E N, Patel V, Manuck T A. Preterm birth among pregnant women living in areas with high social vulnerability. *Am J Obstet Gynecol MFM.* 2021;3(5):100414.

34. Cogswell R, Masotti M, Morris AA, Hart A, Murray T, Yancy C. Assessment of US heart transplantation equity as a function of race: Observational analyses of the OPTN database. *Lancet Reg Health Am.* 2022;13:100290.

35. Dolapo R. Raji. Health Outcomes in a Foreign Land: A role for epigenomic and environmental interaction. *J Food Sci Nutr Res.* 2021;4:267-269.

36. Wang R, Sun T, Yang Q, et al. Low birthweight of children is positively associated with mothers prenatal tobacco smoke exposure in Shanghai: a cross-sectional study. *BMC Pregnancy Childbirth.* 2020;20(1):1-9.

37. Yusuf S, Rangarajan S, Teo K, et al. Cardiovascular risk and events in 17 low-, middle-, and high-income countries. *N Engl J Med.* 2014;371(9):818-827.

38. Schultz WM, Kelli HM, Lisko JC, et al. Socioeconomic status and cardiovascular outcomes: challenges and interventions. *Circulation.* 2018;137(20):2166-2178.

TABLE 1 Percentage table of the number of clusters selected by DIC_3 criterion over simulations (%). The true numbers of spatial clusters are respectively 6, 6, 9, 9, 7, 10 in Designs 1-6. The number of simulations is 100.

		S						
Design	Method	3	4	5	**6**	7	8	9
1	Model 1	0	2	1	94	3	0	0
	Model 2	1	3	5	86	4	1	0
	Model 4	2	3	4	83	6	2	0
		3	4	5	**6**	7	8	9
2	Model 1	1	1	3	92	2	1	0
$r = 0.5$	Model 2	0	3	8	82	5	2	0
	Model 4	0	4	6	80	8	1	1
		3	4	5	**6**	7	8	9
2	Model 1	0	1	6	87	4	2	0
$r = 0.8$	Model 2	0	2	11	79	7	1	0
	Model 4	1	2	6	75	10	4	2
		6	7	8	**9**	10	11	12
3	Model 1	0	2	2	93	3	0	0
	Model 2	2	2	4	85	5	2	0
	Model 4	2	3	3	86	4	2	0
		6	7	8	**9**	10	11	12
4	Model 1	0	2	3	92	2	1	0
	Model 2	1	2	6	83	4	4	0
	Model 4	1	4	5	84	3	3	0
		4	5	6	**7**	8	9	10
5	Model 1	0	1	3	94	2	0	0
	Model 2	0	1	6	88	4	1	0
	Model 4	1	1	5	84	6	2	1
		7	8	9	**10**	11	12	13
6	Model 1	0	1	4	91	3	1	0
	Model 2	0	3	5	84	5	2	1
	Model 4	1	2	6	85	4	2	0

39. Gong P, Liang S, Carlton EJ, et al. Urbanisation and health in China. *Lancet.* 2012;379(9818):843-852.

40. Mensah GA, Roth GA, Fuster V. The global burden of cardiovascular diseases and risk factors: 2020 and beyond. *J Am Coll Cardiol.* 2019;74(20):2529-2532.

41. Kuddus MA, Tynan E, McBryde E. Urbanization: a problem for the rich and the poor? *Public Health Rev.* 2020;41:1-4.

TABLE 2 Comparison of the considered models in terms of the goodness-of-fit measures in the simulation study. DIC and DIC_3 refer to the deviance information criterion. MPL refers to the marginal predictive likelihood. MSPE refers to the mean square prediction error.

Design	Method	DIC	DIC_3	MPL	MSPE
1	Model 1	4177.5	4018.2	-2138.5	2.41
	Model 2	4239.6	4132.7	-2176.3	3.58
	Model 3	4306.2	4197.3	-2195.4	3.29
	Model 4	4174.6	4086.5	-2188.9	3.72
2	Model 1	4195.1	4042.7	-2156.3	2.74
$r = 0.5$	Model 2	4286.5	4175.8	-2199.8	3.92
	Model 3	4331.3	4254.9	-2231.6	3.67
	Model 4	4297.9	4145.6	-2245.5	4.05
2	Model 1	4483.7	4358.2	-2438.6	3.78
$r = 0.8$	Model 2	4574.2	4463.3	-2509.4	4.59
	Model 3	4651.8	4576.4	-2516.3	4.62
	Model 4	4632.5	4519.6	-2538.9	4.86
3	Model 1	5342.1	5214.7	-2786.1	10.47
	Model 2	5457.2	5387.6	-2814.3	12.52
	Model 3	5513.6	5429.8	-2833.5	13.66
	Model 4	5335.9	5298.4	-2820.4	11.48
4	Model 1	5512.4	5488.2	-2825.7	11.32
	Model 2	5616.2	5546.3	-2818.4	12.98
	Model 3	5759.1	5523.8	-2856.3	13.72
	Model 4	5528.3	5507.6	-2839.1	11.03
5	Model 1	5439.3	5351.2	-2628.5	9.35
	Model 2	5517.2	5426.8	-2697.3	10.81
	Model 3	5596.8	5481.5	-2755.7	11.59
	Model 4	5478.1	5397.4	-2729.6	11.14
6	Model 1	6138.8	6102.4	-3086.9	14.45
	Model 2	6246.9	6195.7	-3122.1	16.71
	Model 3	6281.5	6203.5	-3110.6	16.53
	Model 4	6223.1	6158.3	-3128.4	16.02

TABLE 3 Comparison of the considered models in terms of the cluster detection accuracy measures in the simulation study. There are respectively 6, 6, 9, 9, 7, 10 clusters in Designs 1-6

Design	Method	GT_1	GT_2	GT_3	GT_4	GT_5	GT_6	GT_7	GT_8	GT_9	GT_{10}	$\overline{T}$
1	Model 1	0.91	0.93	0.92	0.88	0.90	0.95	–	–	–	–	0.92
	Model 2	0.81	0.84	0.86	0.76	0.83	0.90	–	–	–	–	0.82
	Model 4	0.83	0.90	0.89	0.79	0.85	0.93	–	–	–	–	0.85
2	Model 1	0.90	0.88	0.89	0.90	0.87	0.92	–	–	–	–	0.89
$r = 0.5$	Model 2	0.81	0.80	0.78	0.82	0.81	0.79	–	–	–	–	0.80
	Model 4	0.82	0.81	0.80	0.84	0.81	0.83	–	–	–	–	0.82
2	Model 1	0.81	0.78	0.82	0.79	0.83	0.82	–	–	–	–	0.81
$r = 0.8$	Model 2	0.72	0.71	0.67	0.74	0.71	0.69	–	–	–	–	0.71
	Model 4	0.73	0.68	0.69	0.68	0.72	0.66	–	–	–	–	0.69
3	Model 1	0.89	0.88	0.86	0.84	0.87	0.89	0.85	0.86	0.91	–	0.87
	Model 2	0.82	0.79	0.76	0.80	0.83	0.85	0.80	0.77	0.75	–	0.80
	Model 4	0.84	0.82	0.81	0.83	0.84	0.83	0.82	0.79	0.81	–	0.82
4	Model 1	0.90	0.85	0.82	0.85	0.86	0.87	0.86	0.84	0.88	–	0.86
	Model 2	0.83	0.75	0.73	0.78	0.82	0.84	0.81	0.80	0.83	–	0.80
	Model 4	0.81	0.83	0.76	0.80	0.82	0.78	0.82	0.84	0.83	–	0.81
5	Model 1	0.92	0.89	0.91	0.89	0.88	0.90	0.89	–	–	–	0.90
	Model 2	0.81	0.83	0.78	0.77	0.84	0.80	0.79	–	–	–	0.80
	Model 4	0.85	0.82	0.84	0.81	0.78	0.79	0.83	–	–	–	0.82
6	Model 1	0.85	0.87	0.83	0.84	0.85	0.82	0.85	0.87	0.81	0.86	0.85
	Model 2	0.77	0.72	0.75	0.71	0.70	0.73	0.71	0.68	0.67	0.69	0.71
	Model 4	0.76	0.78	0.75	0.79	0.77	0.81	0.75	0.79	0.74	0.72	0.77

$GT_s, s = 1, \ldots, S$: the rate of correctly classified areas in cluster s.

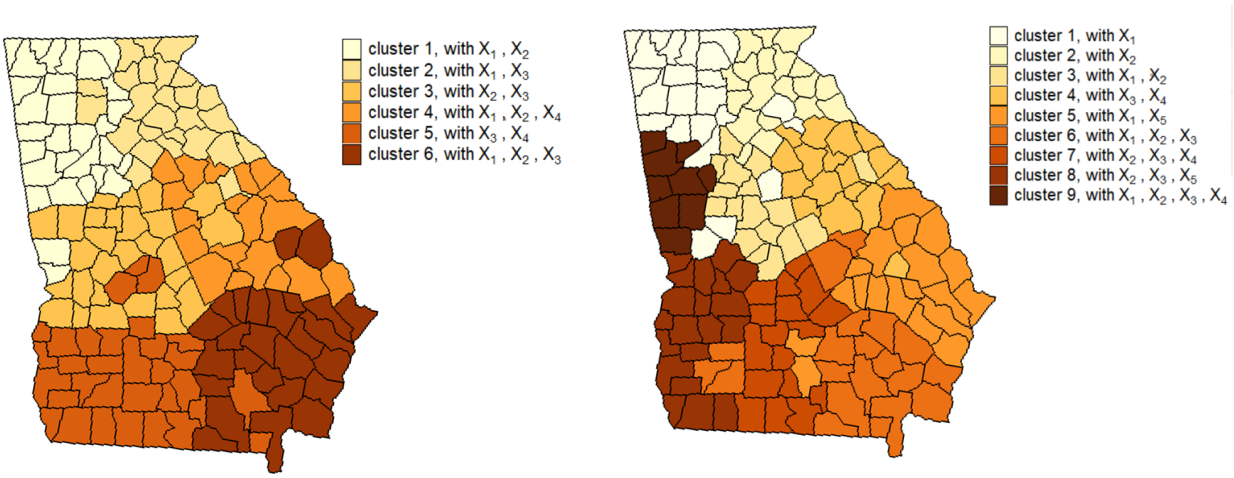


FIGURE 1 Maps of the spatial cluster patterns in Designs 1-4 of the simulation study. The left panel is used in Designs 1 and 2 with $S = 6$ clusters while the right is used in Designs 3 and 4 with $S = 9$ clusters. The legend indicates the significant covariates in each cluster.

TABLE 4 Comparison results of the considered models in terms of the coefficient estimation measures in Design 1 of the simulation study

Parameter	Model	MAE	MSE	CP
β_1	Model 1	0.28	0.22	0.89
	Model 2	0.31	0.35	0.86
	Model 3	0.34	0.36	0.85
	Model 4	0.36	0.39	0.85
β_2	Model 1	0.55	0.46	0.83
	Model 2	0.59	0.57	0.81
	Model 3	0.60	0.53	0.81
	Model 4	0.54	0.50	0.79
β_3	Model 1	0.47	0.53	0.92
	Model 2	0.52	0.55	0.90
	Model 3	0.59	0.51	0.86
	Model 4	0.53	0.58	0.89
β_4	Model 1	0.32	0.27	0.93
	Model 2	0.54	0.58	0.89
	Model 3	0.36	0.33	0.91
	Model 4	0.35	0.37	0.92

Abbreviations: MAE, mean absolute error; MSE, mean squared error; CP, coverage probability.

TABLE 5 Comparison results of the considered models in terms of the coefficient estimation measures with $r = 0.5$ in Design 2 of the simulation study

Parameter	Model	MAE	MSE	CP
β_1	Model 1	0.30	0.20	0.88
	Model 2	0.35	0.38	0.84
	Model 3	0.32	0.39	0.82
	Model 4	0.38	0.42	0.80
β_2	Model 1	0.53	0.49	0.84
	Model 2	0.62	0.55	0.80
	Model 3	0.64	0.52	0.81
	Model 4	0.57	0.58	0.80
β_3	Model 1	0.49	0.56	0.91
	Model 2	0.57	0.62	0.88
	Model 3	0.61	0.56	0.85
	Model 4	0.55	0.56	0.87
β_4	Model 1	0.30	0.29	0.92
	Model 2	0.57	0.60	0.86
	Model 3	0.39	0.36	0.89
	Model 4	0.38	0.41	0.88

Abbreviations: MAE, mean absolute error; MSE, mean squared error; CP, coverage probability.

TABLE 6 Comparison results of the considered models in terms of the coefficient estimation measures with $r = 0.8$ in Design 2 of the simulation study

Parameter	Model	MAE	MSE	CP
β_1	Model 1	0.42	0.29	0.83
	Model 2	0.46	0.47	0.77
	Model 3	0.44	0.49	0.78
	Model 4	0.49	0.51	0.75
β_2	Model 1	0.62	0.57	0.79
	Model 2	0.69	0.72	0.75
	Model 3	0.67	0.69	0.74
	Model 4	0.64	0.62	0.77
β_3	Model 1	0.58	0.64	0.82
	Model 2	0.65	0.67	0.77
	Model 3	0.69	0.63	0.79
	Model 4	0.62	0.65	0.80
β_4	Model 1	0.35	0.33	0.84
	Model 2	0.68	0.66	0.79
	Model 3	0.48	0.41	0.83
	Model 4	0.45	0.47	0.84

Abbreviations: MAE, mean absolute error; MSE, mean squared error; CP, coverage probability.

TABLE 7 Comparison results of the considered models in terms of the coefficient estimation measures in Design 3 of the simulation study

Parameter	Model	MAE	MSE	CP
β_1	Model 1	0.24	0.27	0.92
	Model 2	0.31	0.35	0.88
	Model 3	0.33	0.39	0.82
	Model 4	0.28	0.32	0.90
β_2	Model 1	0.27	0.25	0.91
	Model 2	0.26	0.34	0.87
	Model 3	0.31	0.37	0.82
	Model 4	0.35	0.32	0.86
β_3	Model 1	0.43	0.38	0.88
	Model 2	0.46	0.52	0.85
	Model 3	0.48	0.42	0.83
	Model 4	0.55	0.51	0.90
β_4	Model 1	0.26	0.29	0.93
	Model 2	0.49	0.55	0.79
	Model 3	0.28	0.34	0.84
	Model 4	0.33	0.27	0.86
β_5	Model 1	0.28	0.31	0.90
	Model 2	0.52	0.48	0.80
	Model 3	0.43	0.45	0.82
	Model 4	0.41	0.49	0.84

Abbreviations: MAE, mean absolute error; MSE, mean squared error; CP, coverage probability.

TABLE 8 Comparison results of the considered models in terms of the coefficient estimation measures in Design 4 of the simulation study

Parameter	Model	MAE	MSE	CP
β_1	Model 1	0.26	0.31	0.93
	Model 2	0.28	0.37	0.90
	Model 3	0.30	0.35	0.82
	Model 4	0.33	0.42	0.84
β_2	Model 1	0.25	0.23	0.90
	Model 2	0.29	0.21	0.91
	Model 3	0.32	0.39	0.87
	Model 4	0.30	0.36	0.89
β_3	Model 1	0.33	0.25	0.92
	Model 2	0.36	0.32	0.86
	Model 3	0.31	0.42	0.82
	Model 4	0.40	0.48	0.88
β_4	Model 1	0.23	0.27	0.89
	Model 2	0.57	0.51	0.76
	Model 3	0.29	0.36	0.81
	Model 4	0.31	0.29	0.87
β_5	Model 1	0.32	0.36	0.91
	Model 2	0.56	0.61	0.83
	Model 3	0.41	0.37	0.86
	Model 4	0.38	0.44	0.88

Abbreviations: MAE, mean absolute error; MSE, mean squared error; CP, coverage probability.

TABLE 9 Comparison results of the considered models in terms of the coefficient estimation measures in Design 5 of the simulation study

Parameter	Model	MAE	MSE	CP
β_1	Model 1	0.25	0.27	0.93
	Model 2	0.29	0.33	0.88
	Model 3	0.31	0.28	0.87
	Model 4	0.28	0.31	0.90
β_2	Model 1	0.28	0.32	0.91
	Model 2	0.33	0.37	0.85
	Model 3	0.31	0.29	0.88
	Model 4	0.34	0.41	0.84
β_3	Model 1	0.21	0.26	0.94
	Model 2	0.26	0.34	0.90
	Model 3	0.29	0.37	0.87
	Model 4	0.24	0.28	0.91
β_4	Model 1	0.35	0.31	0.92
	Model 2	0.57	0.64	0.82
	Model 3	0.39	0.34	0.87
	Model 4	0.42	0.37	0.85

Abbreviations: MAE, mean absolute error; MSE, mean squared error; CP, coverage probability.

TABLE 10 Comparison results of the considered models in terms of the coefficient estimation measures in Design 6 of the simulation study

Parameter	Model	MAE	MSE	CP
β_1	Model 1	0.31	0.36	0.91
	Model 2	0.37	0.42	0.85
	Model 3	0.35	0.44	0.83
	Model 4	0.33	0.41	0.87
β_2	Model 1	0.34	0.41	0.90
	Model 2	0.39	0.45	0.87
	Model 3	0.42	0.37	0.84
	Model 4	0.33	0.45	0.88
β_3	Model 1	0.28	0.22	0.92
	Model 2	0.34	0.45	0.83
	Model 3	0.37	0.42	0.86
	Model 4	0.31	0.39	0.88
β_4	Model 1	0.39	0.43	0.90
	Model 2	0.53	0.62	0.81
	Model 3	0.45	0.40	0.86
	Model 4	0.47	0.41	0.83

Abbreviations: MAE, mean absolute error; MSE, mean squared error; CP, coverage probability.

TABLE 11 Comparison of the considered models in terms of the variable selection measures in Design 1 of the simulation study, based on different cutoff values

Cutoff value	Model	SA_1	SA_2	SA_3	SA_4	Average
0.50	Model 1	0.846	0.829	0.845	0.879	0.850
	Model 2	0.831	0.808	0.816	0.812	0.817
	Model 3	0.825	0.819	0.803	0.826	0.818
0.65	Model 1	0.712	0.685	0.765	0.743	0.726
	Model 2	0.697	0.661	0.734	0.720	0.703
	Model 3	0.693	0.667	0.739	0.722	0.705
0.80	Model 1	0.574	0.483	0.766	0.530	0.588
	Model 2	0.535	0.462	0.739	0.491	0.557
	Model 3	0.526	0.460	0.734	0.523	0.561

TABLE 12 Comparison of the considered models in terms of the variable selection measures in Designs 2-6 of the simulation study, with the cutoff value fixed at 0.5

Design	Model	SA_1	SA_2	SA_3	SA_4	SA_5	Average
2	Model 1	0.842	0.831	0.839	0.858	–	0.843
$r = 0.5$	Model 2	0.823	0.815	0.812	0.796	–	0.812
	Model 3	0.806	0.813	0.801	0.819	–	0.809
2	Model 1	0.802	0.791	0.804	0.825	–	0.806
$r = 0.8$	Model 2	0.773	0.784	0.786	0.739	–	0.771
	Model 3	0.760	0.753	0.764	0.785	–	0.766
3	Model 1	0.871	0.823	0.836	0.845	0.818	0.839
	Model 2	0.829	0.844	0.817	0.796	0.802	0.818
	Model 3	0.833	0.837	0.798	0.809	0.825	0.820
4	Model 1	0.856	0.822	0.814	0.789	0.793	0.815
	Model 2	0.824	0.782	0.801	0.753	0.722	0.776
	Model 3	0.831	0.815	0.784	0.773	0.766	0.794
5	Model 1	0.869	0.846	0.862	0.833	–	0.853
	Model 2	0.820	0.814	0.816	0.771	–	0.805
	Model 3	0.832	0.829	0.835	0.781	–	0.819
6	Model 1	0.827	0.824	0.833	0.818	–	0.826
	Model 2	0.774	0.795	0.801	0.763	–	0.783
	Model 3	0.782	0.791	0.794	0.786	–	0.788

TABLE 13 The run times (mins) for various simulation settings

Settings	Design 1	Design 2	Design 3	Design 4	Design 5	Design 6
N	159	159	159	159	616	616
T	8	8	12	12	12	12
P	4	4	5	5	4	4
S	6	6	9	9	7	10
Time	5.35	5.41	12.87	12.93	15.42	15.73

TABLE 14 Comparison of the considered models in terms of the goodness-of-fit measures in the real data analysis

Model	DIC	DIC_3	MPL	MSPE
Model 1	15018	15003	-7275	105
Model 2	15170	15114	-7347	144
Model 3	15213	15185	-7321	137
Model 4	15014	15156	-7339	133

Abbreviations: DIC, deviance information criterion; MPL, marginal predictive likelihood; MSPE, mean square prediction error.

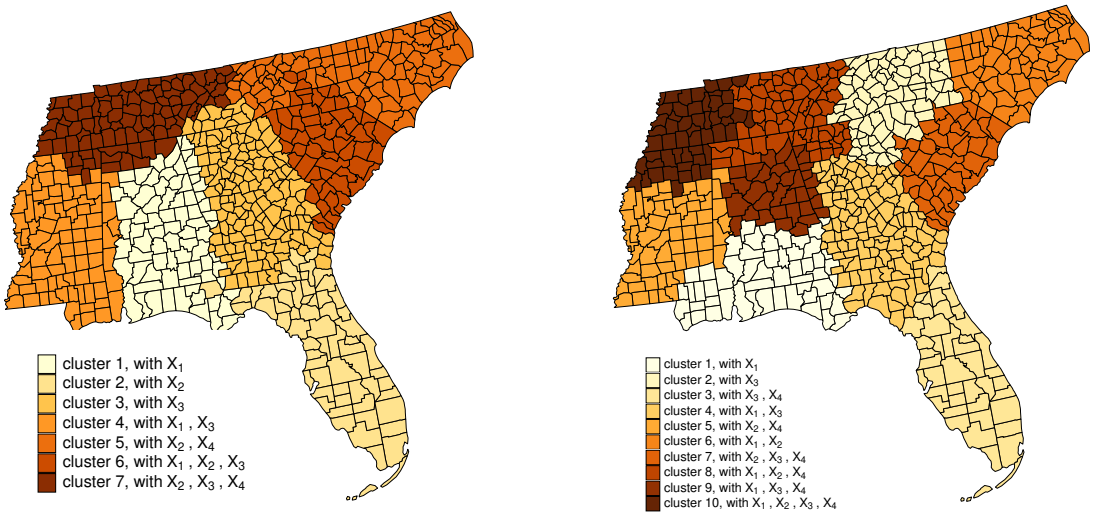


FIGURE 2 Maps of the spatial cluster patterns in Designs 5-6 of the simulation study. The left panel is used in Design 5 with $S = 7$ clusters while the right is used in Design 6 with $S = 10$ clusters. The legend indicates the significant covariates in each cluster.

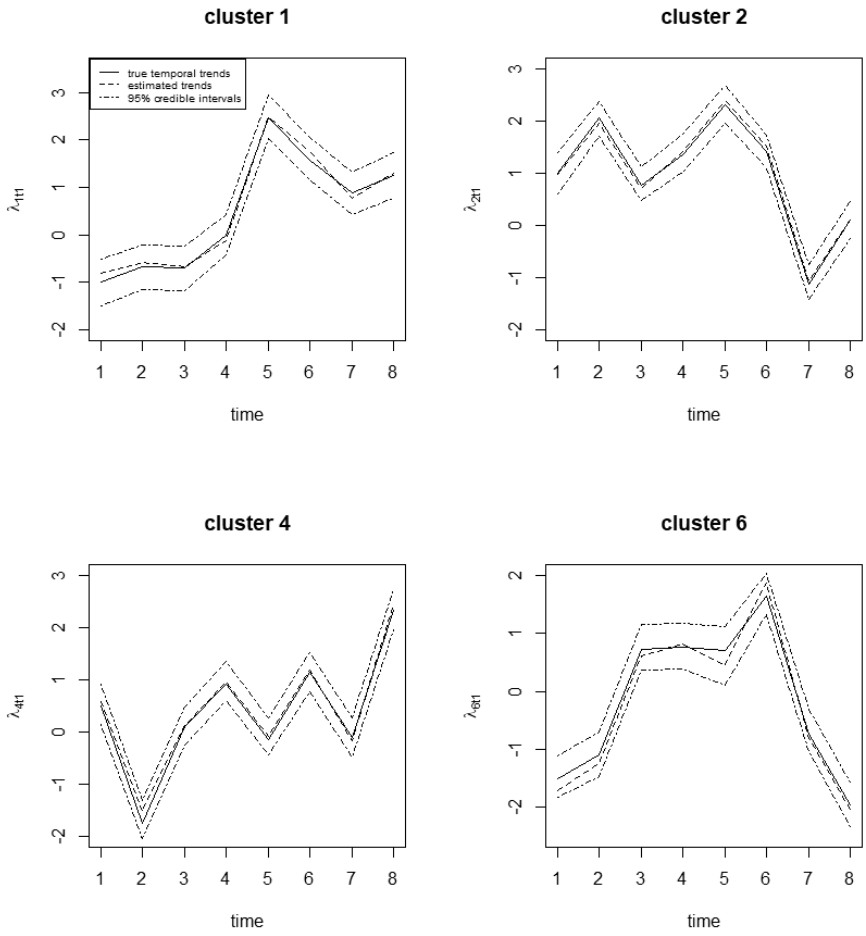


FIGURE 3 Plots of the true significant temporal trends and estimated ones from Model 1 for the first covariate effect in the selected four clusters in Design 1. The solid line is the true temporal trend, the dashed line is the average of the posterior estimates, and the dot-dashed lines are 95% credible intervals for the posterior estimated profiles.

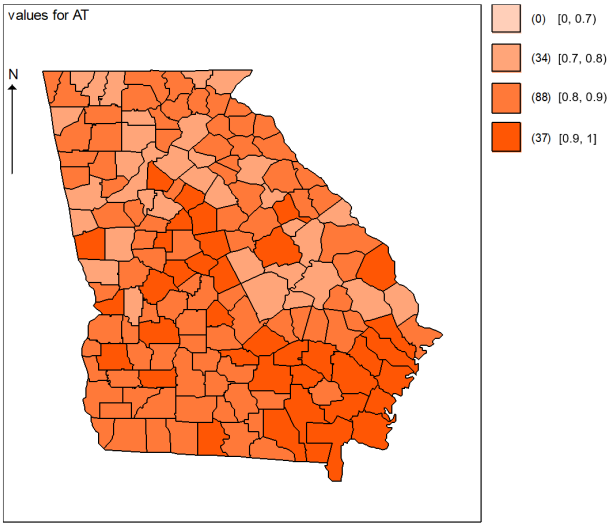


FIGURE 4 The map of AT_i from Model 1 in Design 1 of the simulation study. The measure AT_i evaluates the capability of spatial cluster recovery for area i. The areas with darker colors have relatively higher cluster detection accuracy rates.

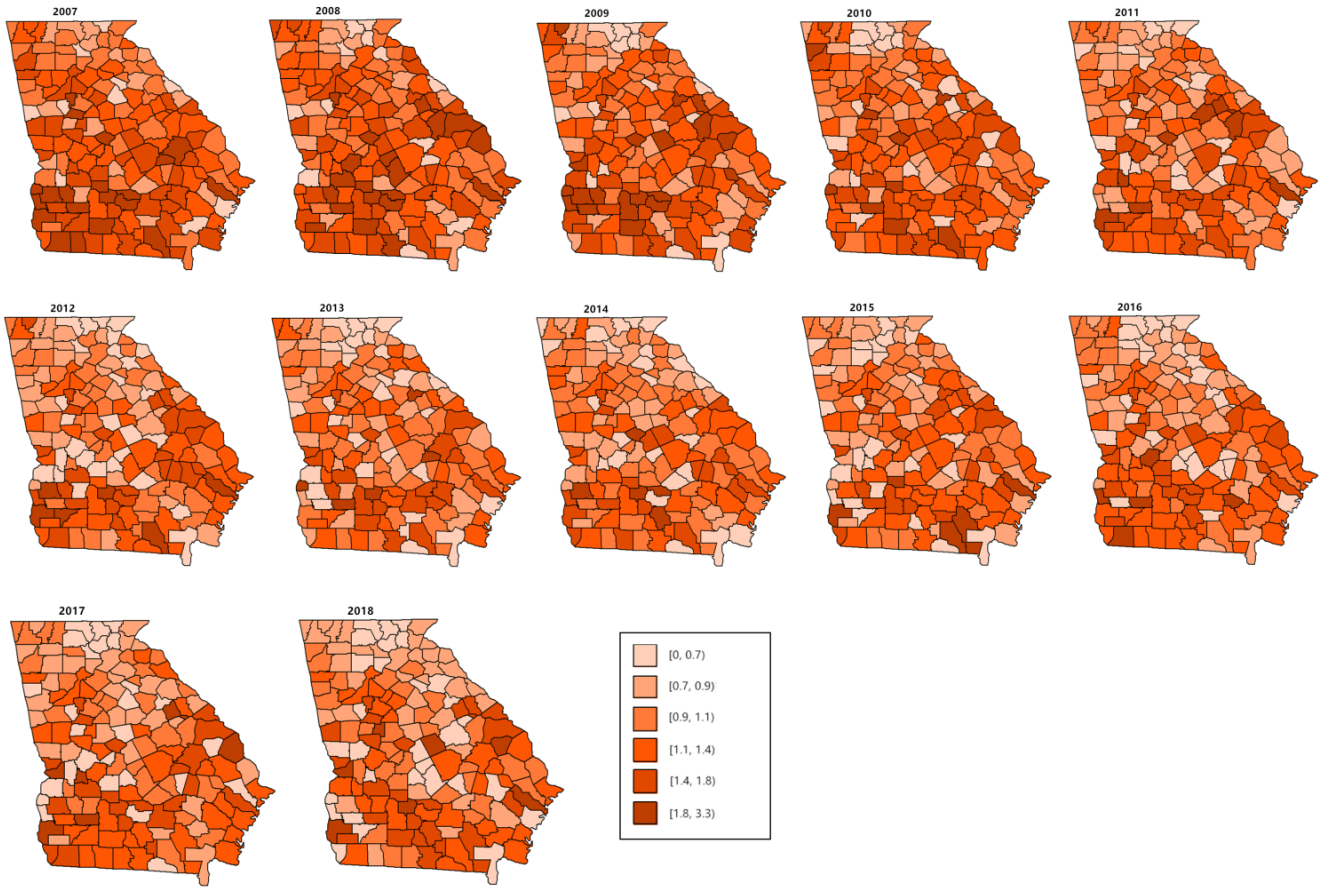


FIGURE 5 Maps of the standardized incidence ratios for the low birth weight incidence data in Georgia through 2007-2018

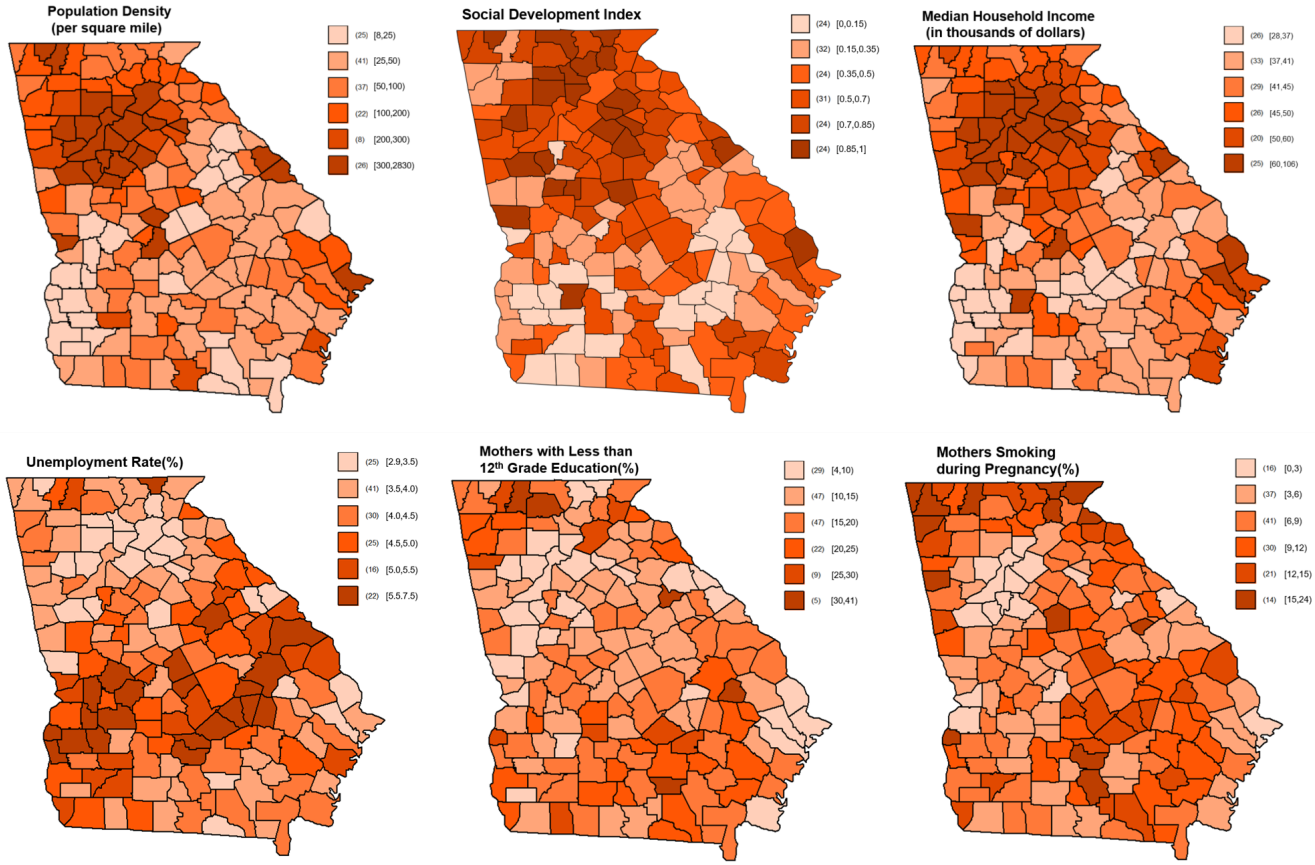


FIGURE 6 Maps of the six potential risk factors for LBW data in the year of 2018 in Georgia

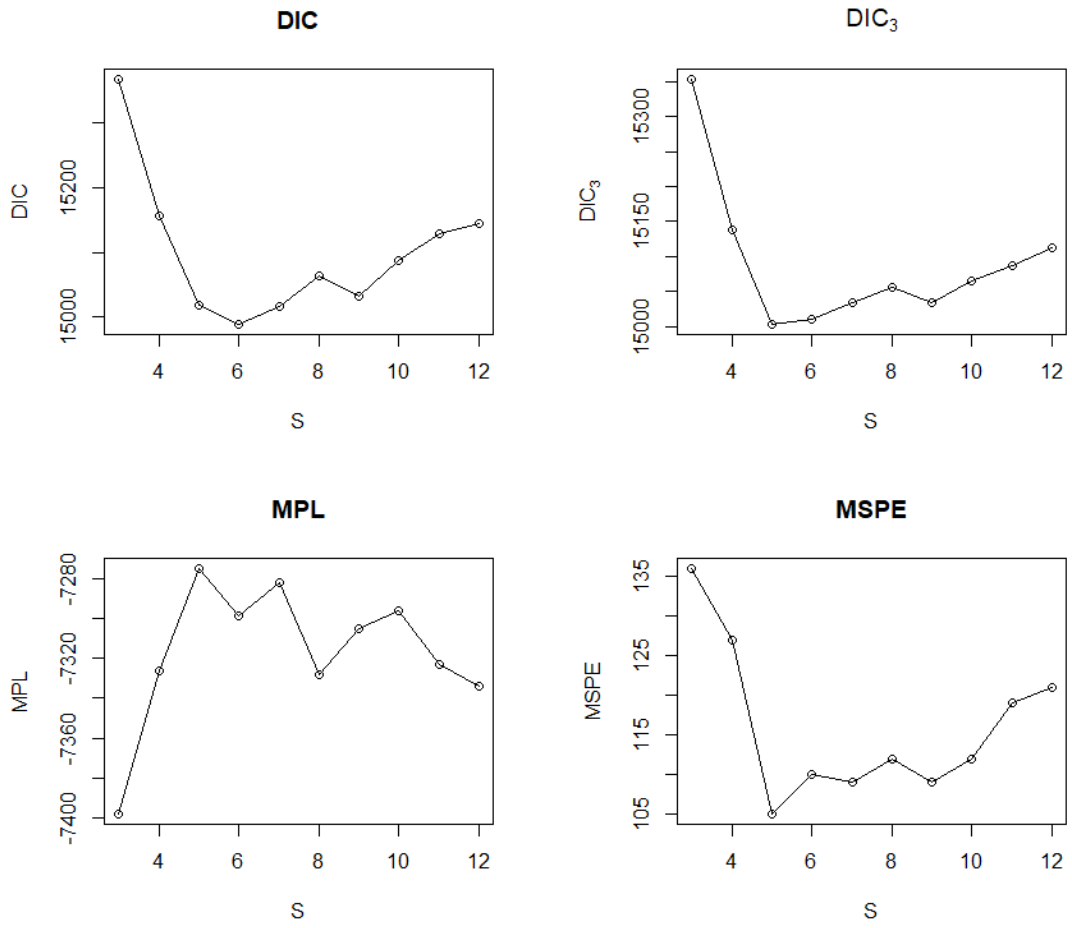


FIGURE 7 The goodness-of-fit measures for the proposed model with different number of clusters S

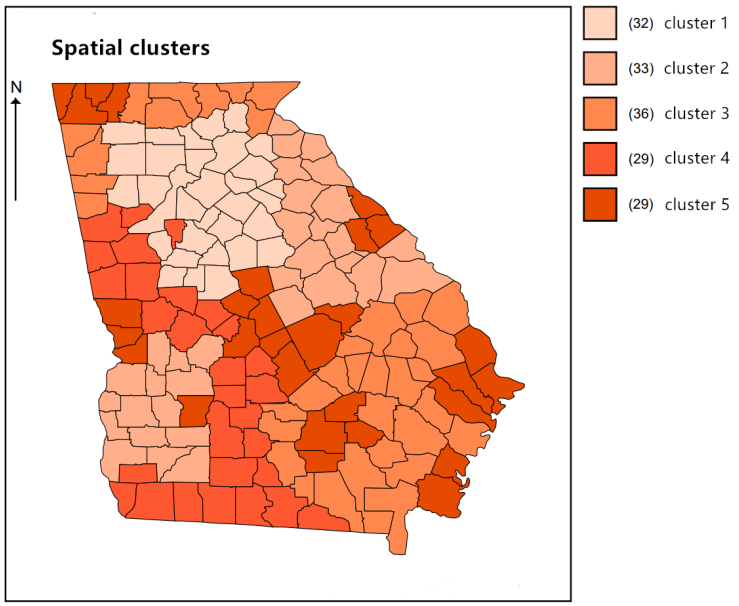


FIGURE 8 The map of estimated spatial clusters for the LBW incidence data

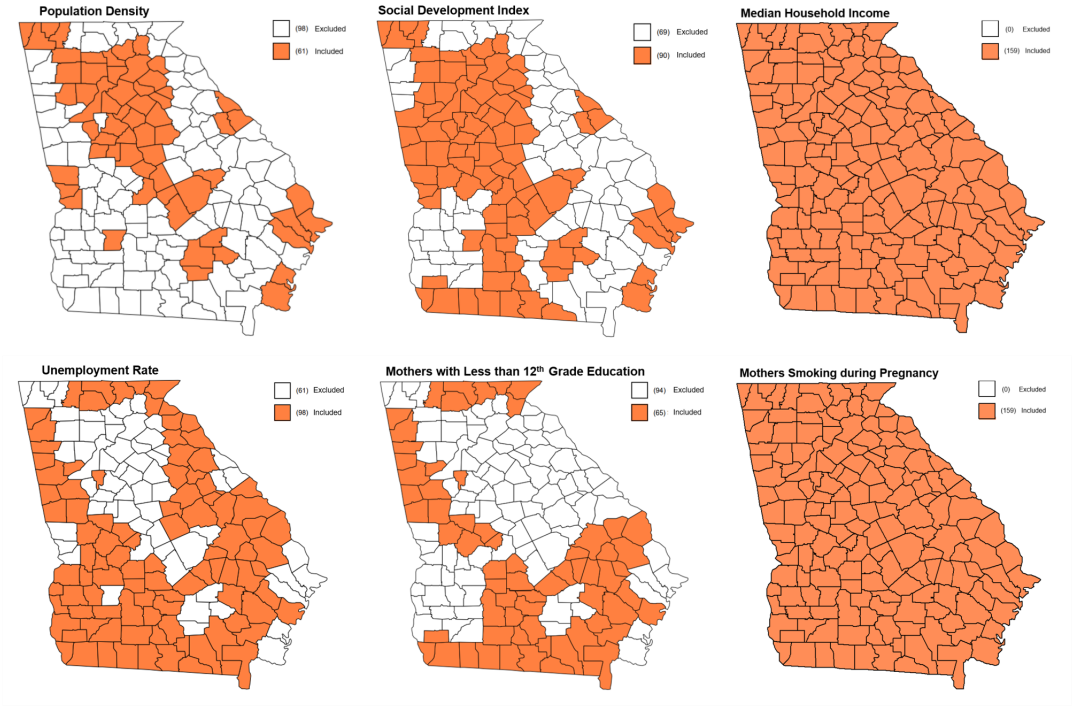


FIGURE 9 Maps of the estimated selection indicators for the six considered covariates for the LBW data in Georgia. The colored areas represent that the corresponding covariates are significant in the region; otherwise, the covariates are not correlated with the outcome in the corresponding clusters.

FIGURE 10 The temporal trends of the estimated coefficients for the six considered covariates for the LBW data. Different linetypes and colors represent the regression coefficients in different clusters.

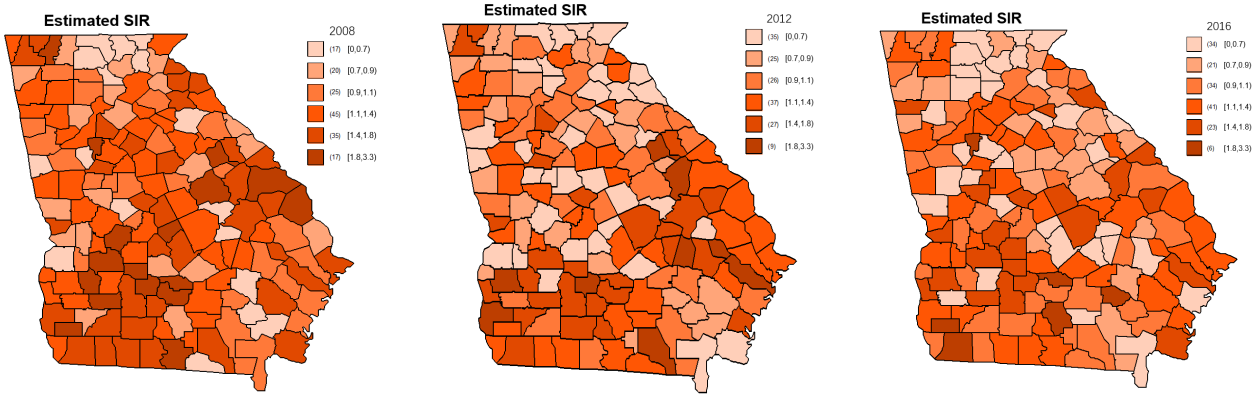
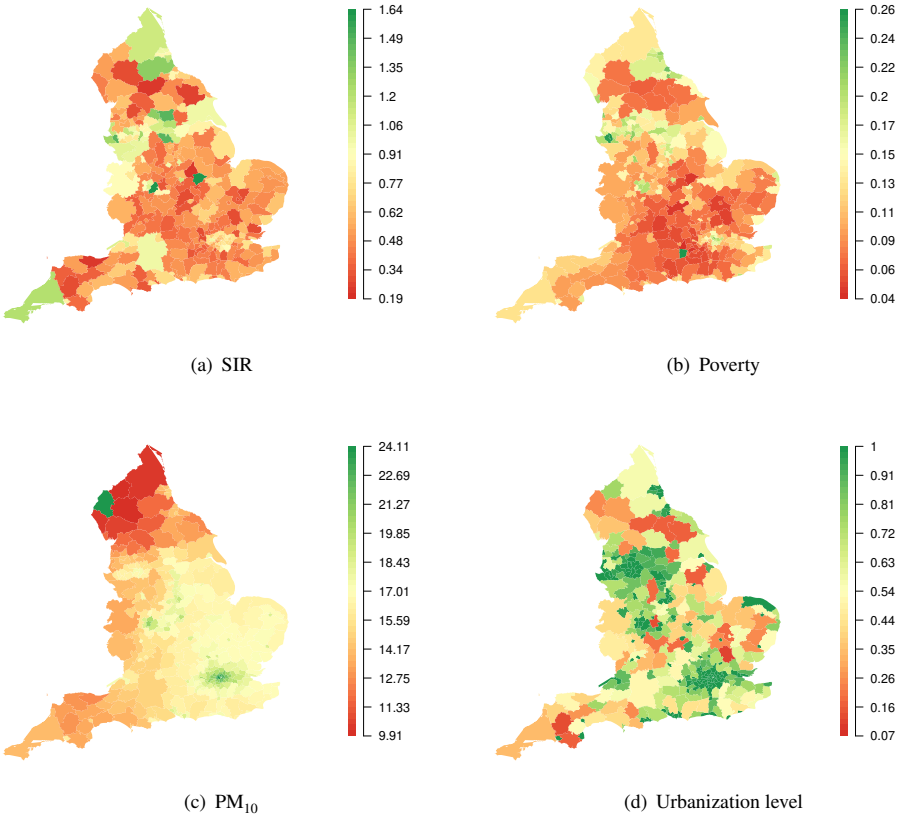


FIGURE 11 Maps of the estimated standardized incidence ratios (SIR) of LBW data based on the proposed method in years 2008, 2012 and 2016



(a) SIR (b) Poverty

(c) PM_{10} (d) Urbanization level

FIGURE 12 Maps of the England circulatory data. (a) Averaged standardized incidence ratios (SIR) for circulatory hospital admissions across England local authorities between 2001 and 2010. (b)-(d) Average values of the covariates over all years for poverty, PM_{10} and urbanization level, respectively.

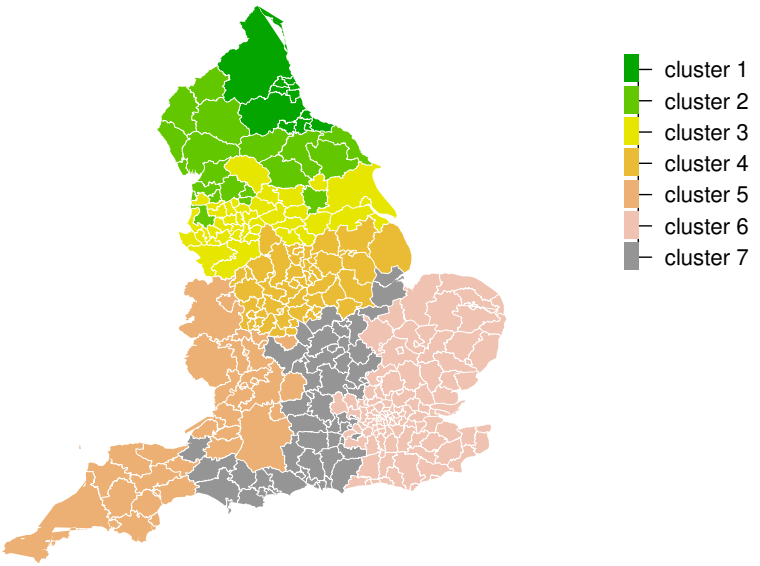


FIGURE 13 The map of estimated spatial clusters for the circulatory disease data in England

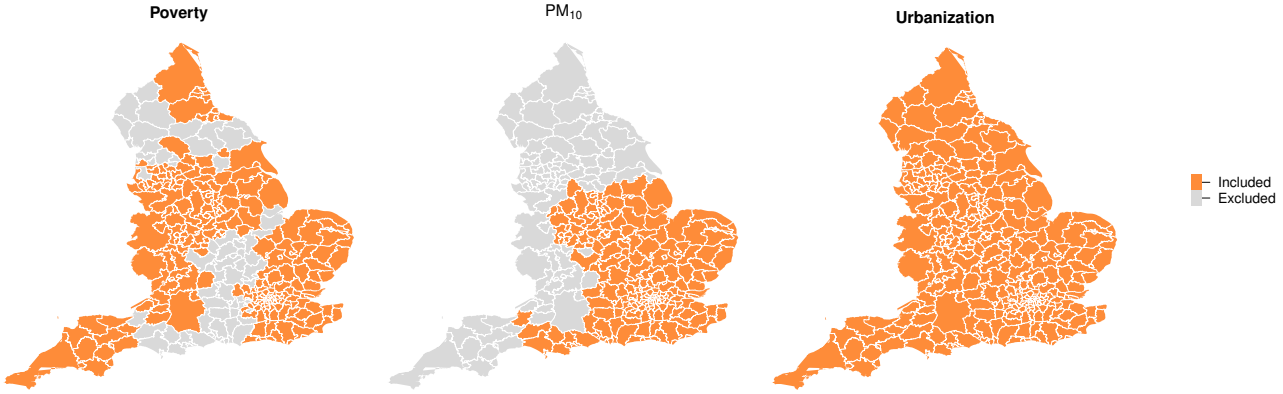


FIGURE 14 Maps of the estimated selection indicators for the covariates in the circulatory disease data in England. The colored areas represent that the corresponding covariates are significant in the region; otherwise, the covariates are not correlated with the outcome in the corresponding clusters.

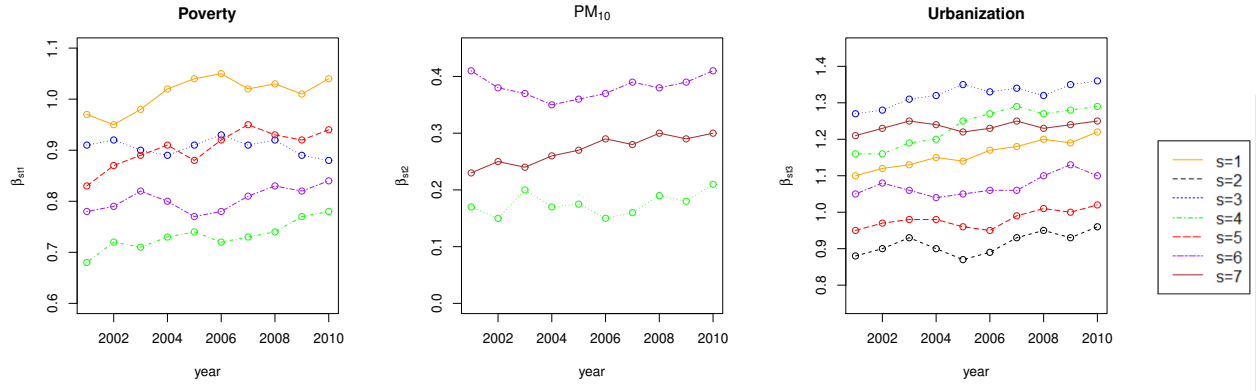


FIGURE 15 The temporal trends of the estimated coefficients for the significant covariates in the circulatory disease data. Different linetypes and colors represent the regression coefficients in different clusters.