



RESEARCH ARTICLE

Clustered spatio-temporal varying coefficient regression model

Junho Lee¹ | Maria E. Kamenetsky² | Ronald E. Gangnon^{2,3} | Jun Zhu⁴

¹Statistics Program, CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

²Department of Population Health Sciences, University of Wisconsin-Madison, Madison, Wisconsin, USA

³Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, Wisconsin, USA

⁴Department of Statistics, University of Wisconsin-Madison, Madison, Wisconsin, USA

Correspondence

Ronald E. Gangnon, Department of Biostatistics and Medical Informatics and Department of Population Health Sciences, University of Wisconsin-Madison, Madison, WI 53726, USA.

Email: ronald@biostat.wisc.edu

Funding information

King Abdullah University of Science and Technology, Grant/Award Number: OSR-2019-CRG7-3800; University of Wisconsin-Madison, Grant/Award Number: a USGS CESU project and a pilot project from CDHA

In regression analysis for spatio-temporal data, identifying clusters of spatial units over time in a regression coefficient could provide insight into the unique relationship between a response and covariates in certain subdomains of space and time windows relative to the background in other parts of the spatial domain and the time period of interest. In this article, we propose a varying coefficient regression method for spatial data repeatedly sampled over time, with heterogeneity in regression coefficients across both space and over time. In particular, we extend a varying coefficient regression model for spatial-only data to spatio-temporal data with flexible temporal patterns. We consider the detection of a potential cylindrical cluster of regression coefficients based on testing whether the regression coefficient is the same or not over the entire spatial domain for each time point. For multiple clusters, we develop a sequential identification approach. We assess the power and identification of known clusters via a simulation study. Our proposed methodology is illustrated by the analysis of a cancer mortality dataset in the Southeast of the U.S.

KEYWORDS

regression, spatial cluster detection, spatial scan statistic, spatio-temporal cluster detection, spatio-temporal varying coefficient, varying coefficient regression

1 | INTRODUCTION

Spatio-temporal regression is one of the most used methods to explore the relationship between a response and covariates for spatial data repeatedly sampled over time. In particular, spatial varying coefficient regression approaches consider different regression coefficients for different locations in a study region. The purpose of this article is to develop a new spatio-temporal varying coefficient regression method that detects clusters in space and/or time with distinctive patterns relative to the background variation.

In spatio-temporal regression, it is a common assumption that covariate effects are homogeneous across the studied domain. However, real applications often show that relationships between a response and covariates vary across the region or time. That is, the homogeneous regression coefficients assumption is questionable in practice. Thus, there have been many studies to take spatio-temporal variations in estimated covariate effects into account. Geographically weighted regression (GWR)^{1,2} and its variants^{3,4} are well-known approaches to spatial varying coefficient regression for spatial-only data. GWR provides locally weighted regression coefficients, which vary geographically across space. There are numerous real applications from the literature using GWR, which show spatial heterogeneity in estimated covariate effects.⁵⁻²⁴ Assunção²⁵ and Gamerman et al²⁶ studied Bayesian varying coefficients models for areal data. Recently, Wang and Sun²⁷ proposed a penalized local polynomial model for the estimation of spatially varying coefficient models. Alternatively, there are some studies for varying coefficient regression models based on spatial cluster frameworks. Lawson et al²⁸ proposed an approach which provides the grouping of regression coefficients directly when the number of groups is known a priori. Lee et al^{29,30} proposed a spatial cluster detection method for regression coefficients which allows the identification of an unknown number of spatial clusters in the regression coefficients directly via hypothesis testing and the construction of spatially varying coefficient regression based on detected spatial clusters. Lee et al²⁹ illustrated that, within a unified modeling framework for spatial clusters of covariates in relation to the response, this methodology was more rigorous to discern heterogeneous spatial patterns than GWR. Furthermore, Lee et al³⁰ introduced a spatial blockwise random effect to take spatial dependency into account. More recently, Lagona et al³¹ proposed to estimate space-varying effects on the regression coefficients by exploiting a multivariate hidden Markov field and using an EM algorithm and composite likelihood methods. However, the aforementioned methods are applicable to spatial only data, and none are directly applicable to spatio-temporal data.

For cluster detection in space and/or time, under a frequentist framework, the spatial scan statistic^{32,33} and spatio-temporal scan statistic^{34,35} as well as their many variants³⁶⁻⁵⁰ are popular approaches. Especially, Jung⁴² proposed a covariate-adjusted spatial scan statistic based on generalized linear models for the spatial cluster detection in the intercepts only, while the slopes associated with the covariates are assumed identical across the spatial domain. An alternative framework uses Bayesian models to detect spatial clusters⁵¹⁻⁵⁶ and spatio-temporal cluster identification.⁵⁷⁻⁵⁹ Besides the cluster detection framework, Lawson et al⁶⁰ and Napier et al⁶¹ proposed spatial clustering approaches based on temporal trends for spatio-temporal data. To date, however, all of the spatio-temporal cluster detection or clustering approaches focus on the response or the intercepts only, but not the relationship between the response and covariates across space and time.

There are some studies on varying coefficient models for spatio-temporal data. Dreassi et al⁶² constructed a space-time hierarchical Bayesian model with time-dependent covariates for areal data. Cai et al⁶³ developed a Bayesian regression model with multivariate linear splines for spatio-temporal data. However, the coefficients are not spatially clustered in those models, so that there appears to be very limited work for spatio-temporal cluster analysis focusing on the regression coefficients.

Here, we propose a clustered spatio-temporal varying coefficient regression model. The motivating dataset comprises county-level cancer mortality rates in the Southeast of the United States in three time periods: 2000-2004, 2005-2009, and 2010-2014. Of particular interest is cancer mortality in relation to urbanization for each county. In a regression model for the motivating data, a subdomain may exist in space and time which provides a distinct association between cancer mortality and the extent of urbanization relative to the background. Such a cluster of distinctive relationships may be captured by clusters of regression coefficients in space and time, and may provide important insight into public health and policy-making. Thus, in this article, we consider the development of new spatio-temporal varying coefficient regression models and novel statistical methodology for identification of clustered spatio-temporal varying coefficients. This may be viewed as an extension of our previous work with spatial only data²⁹ to spatio-temporal data, but the extension is substantive due to a multitude of challenges in modeling, testing, and computation. For a spatio-temporal cluster, we define a cylindrical cluster by a circular window for a temporal interval. To identify a single cluster, we develop hypothesis testing for potential cylindrical clusters, and for multiple clusters, we propose two methods to identify multiple clusters in the intercepts and slopes sequentially. In addition to introducing our method, we also demonstrate that models using a basis of circular windows can effectively identify noncircular windows.

The remainder of the article is organized as follows. In Section 2, we define cylindrical spatio-temporal clusters and propose a new spatio-temporal varying coefficient regression model. In Section 3, we develop hypothesis testing for spatio-temporal cluster effects in a simplified setting and propose single cluster identification method as well as multiple

clusters identification method via a sequential scheme. In Section 4, we evaluate our proposed method via simulation studies for power and identification of multiple known clusters. An illustration with the cancer mortality dataset is presented in Section 5. Some technical details are provided as supplementary materials.

2 | VARYING COEFFICIENT REGRESSION WITH CLUSTERS

For the spatial cluster, we follow the notation in Lee et al²⁹ Let \mathcal{D} denote a spatial domain of interest in \mathbb{R}^2 and be partitioned into N cells. For cell $i = 1, \dots, N$, let $\mathbf{s}_i = (s_{1i}, s_{2i})'$ denote the coordinates of the geographical centroid. Then, a circular spatial cluster C^s is defined as a set of cells that are within the radius r from the center \mathbf{c} such that

$$C^s = \{i \mid d(\mathbf{s}_i, \mathbf{c}) \leq r\}, \quad (1)$$

where $d(\cdot, \cdot)$ is the distance between two locations. Although we consider the circular window because it is a common practice in the cluster detection or scan statistic approaches, it can be modified with other shapes, such as ellipses and squares.³⁸⁻⁴⁰

Furthermore, let \mathcal{T} denote a temporal domain of interest in \mathbb{R}^1 . Let T denote the number of time points that partition the temporal domain \mathcal{T} . Consider a spatio-temporal domain $\mathcal{D} \times \mathcal{T}$ in $\mathbb{R}^2 \times \mathbb{R}^1$. That is, we have cell $i = 1, \dots, N$ at multiple time points $t = 1, \dots, T$. Then, a cylindrical spatio-temporal cluster $\{(i, t) \mid d(\mathbf{s}_i, \mathbf{c}) \leq r, l \leq t \leq u\}$ is an expansion of the circular spatial cluster C^s in (1) and has a circular window for the time interval between l and u .^{34,57}

Let $\mathbf{x}_{it} = (1, x_{it})'$ denote the i th covariate vector at the time point t , and let $\boldsymbol{\beta}_t$ denote the regression coefficient vector for the background (i.e., noncluster) at the time point t . Also let \mathbf{c}_j and r_j denote the center and the radius of a circle, in metric d , defining the spatial extent of the j th cluster for $j = 1, \dots, J$.

We model the response variable in cell i and at time t as $y_{it} = \mu_{it} + \varepsilon_{it}$, where the random error ε_{it} 's are iid $\mathcal{N}(0, \sigma^2)$ with a variance component $\sigma^2 > 0$. Furthermore, the mean response μ_{it} follows a spatio-temporal varying coefficient regression model as follows:

$$\begin{aligned} \mu_{it} &= \mathbf{x}_{it}' \boldsymbol{\beta}_{it}, \\ \boldsymbol{\beta}_{it} &= \boldsymbol{\beta}_t + \sum_{j=1}^J \boldsymbol{\theta}_j \cdot \mathbf{I}\{d(\mathbf{s}_i, \mathbf{c}_j) \leq r_j, l_j \leq t \leq u_j\}, \end{aligned} \quad (2)$$

$$\text{or } \boldsymbol{\beta}_{it} = \boldsymbol{\beta}_t + \sum_{j=1}^J \boldsymbol{\theta}_{jt} \cdot \mathbf{I}\{d(\mathbf{s}_i, \mathbf{c}_j) \leq r_j\}. \quad (3)$$

In (2), l_j and u_j are the lower and upper limits of the time interval defining the temporal extent of the j th cluster, and $\boldsymbol{\theta}_j$ is the corresponding cluster effect. In (3), while there are no limits of the temporal extent of each cluster, the j th cluster effects are defined at each time point t as $\boldsymbol{\theta}_{jt}$. The cluster $\{i \mid d(\mathbf{s}_i, \mathbf{c}) \leq r\}$ in (3) also can be thought of as a cylindrical spatio-temporal cluster which is a temporal expansion of the circular spatial cluster C^s in (1) for all the time points $t = 1, \dots, T$. Gangnon⁵⁸ showed that the setting (3) was effective in identifying flexible temporal patterns.

A more sophisticated model, which has three different types of cluster in time, in space, and in both space and time, is possible but not discussed here since (2) and (3) are more practical. Details regarding the interrelationships between these models are described in Appendix A in the Supplementary Materials.

For the j th cluster, l_j and u_j limit the temporal extent of the cluster associated with $\boldsymbol{\theta}_j$ in (2). By contrast, there is no limit in the temporal extent of the cluster, although $\boldsymbol{\theta}_{jt}$ allows separate cluster effects at each time point, in (3). Thus, if a total of J clusters are given, the numbers of parameters are $2(T + J)$ and $2(T + JT)$ in (2) and (3), respectively. Thus, (2) is more appealing since it has fewer parameters than (3).

In practice, we do not know where the true clusters are located and thus, we should estimate them as well as their corresponding parameters. Our approach is to begin with the model with a single cluster (i.e., $J = 1$) which is unknown and, estimate the cluster by evaluating all potential clusters and picking the most significant one. Let M be the number of potential circular clusters of the form (1) in the spatial domain \mathcal{D} . Then, the number of potential spatio-temporal clusters of the form in (2) is $MT(T + 1)/2$, while it is still M if we adopt the cylindrical clusters of

the form in (3). Since, to detect a cluster, we estimate $2(T + 1)$ parameters as much as $MT(T + 1)/2$ times in (2) and $4T$ parameters as much as M times in (3), the total numbers of parameters to estimate are $MT(T + 1)^2$ and $4MT$ in (2) and (3), respectively. The ratio of these two is $\{MT(T + 1)^2\}/(4MT) = (T + 1)^2/4 = O(T^2)$, which means (3) is less computationally intensive than (2) yet more flexible for identifying temporal patterns. Thus, we develop our method based on (3).

3 | IDENTIFICATION OF CLUSTERED SPATIO-TEMPORAL VARYING COEFFICIENTS

Let C_j denote the j th cluster and be defined as

$$C_j = \{i \mid d(\mathbf{s}_i, \mathbf{c}_j) \leq r_j\} \quad \text{for all } t = 1, \dots, T,$$

where $j = 1, \dots, J$. Then, the spatio-temporal varying coefficient regression model (3) can be expressed as

$$\begin{aligned} \mu_{it} &= \mathbf{x}'_{it} \boldsymbol{\beta}_{it}, \\ \boldsymbol{\beta}_{it} &= \boldsymbol{\beta}_t + \sum_{j=1}^J \boldsymbol{\theta}_{jt} \cdot \mathbf{I}\{i \in C_j\} = (\beta_{0,t}, \beta_{1,t})' + \sum_{j=1}^J (\theta_{j,0,t}, \theta_{j,1,t})' \cdot \mathbf{I}\{i \in C_j\}, \end{aligned} \quad (4)$$

where, respectively, $\beta_{0,t}$ and $\beta_{1,t}$ are the intercept and the slope for the background, and $\theta_{j,0,t}$ and $\theta_{j,1,t}$ are the cluster C_j effect in the intercepts and in the slopes at time point t .

3.1 | Single cluster identification

Let $\mathcal{C} = \{C_1, C_2, \dots\}$ denote the set of all potential clusters. Now, we simplify (4) with a single cluster $C_k \in \mathcal{C}, k = 1, 2, \dots$, as

$$\mu_{it} = \begin{cases} \beta_{0,t} + \beta_{1,t}x_{it} & \text{if } i \notin C_k \\ (\beta_{0,t} + \theta_{C_k,0,t}) + (\beta_{1,t} + \theta_{C_k,1,t})x_{it} & \text{if } i \in C_k \end{cases}, \quad (5)$$

where $\theta_{C_k,0,t}$ and $\theta_{C_k,1,t}$ are the cluster effect in the intercepts and in the slopes, respectively, of the cluster C_k at time point t .

For the k th candidate cluster $C_k \in \mathcal{C}$, we develop a cluster specific hypothesis testing:

$$H_{0_k} : \boldsymbol{\theta}_{C_k} = \mathbf{0} \quad \text{vs} \quad H_{A_k} : \boldsymbol{\theta}_{C_k} \neq \mathbf{0}, \quad (6)$$

where $\boldsymbol{\theta}_{C_k} = (\boldsymbol{\theta}'_{C_k,1}, \dots, \boldsymbol{\theta}'_{C_k,T})'$ is the cluster effect vector of the cluster C_k such that $\boldsymbol{\theta}_{C_k,t} = (\theta_{C_k,0,t}, \theta_{C_k,1,t})'$ for $t = 1, \dots, T$. That is, $\boldsymbol{\theta}_{C_k,t} = (0, 0)'$ for all $t = 1, \dots, T$ under H_{0_k} , while $\boldsymbol{\theta}_{C_k,t} \neq (0, 0)'$ for at least one $t = 1, \dots, T$ under H_{A_k} . Define $\boldsymbol{\beta} = (\boldsymbol{\beta}'_1, \dots, \boldsymbol{\beta}'_T)'$ as the background regression coefficient vector. Then, a likelihood ratio test (LRT) statistic for (6) is defined as

$$\lambda(C_k) = \frac{\mathcal{L}(\hat{\boldsymbol{\beta}}_{A_k}, \hat{\boldsymbol{\theta}}_{A_k}, \hat{\sigma}_{A_k}^2)}{\mathcal{L}(\hat{\boldsymbol{\beta}}_{0_k}, \hat{\boldsymbol{\theta}}_{0_k}, \hat{\sigma}_{0_k}^2)}, \quad (7)$$

where $(\hat{\boldsymbol{\beta}}'_{0_k}, \hat{\boldsymbol{\theta}}'_{0_k}, \hat{\sigma}_{0_k}^2)'$ and $(\hat{\boldsymbol{\beta}}'_{A_k}, \hat{\boldsymbol{\theta}}'_{A_k}, \hat{\sigma}_{A_k}^2)'$ denote the maximum likelihood estimates (MLEs) of $(\boldsymbol{\beta}', \boldsymbol{\theta}', \sigma^2)'$ under H_{0_k} and H_{A_k} , respectively, and $\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)$ is the likelihood evaluated at $(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2)$.

Next, we take an unknown generic cluster C into account to find it among all potential clusters in \mathcal{C} . For the unknown $C \in \mathcal{C}$, we consider a global hypothesis testing:

$$H_0 : \boldsymbol{\theta}_C = \mathbf{0} \text{ for all } C \in \mathcal{C} \quad \text{vs} \quad H_A : \boldsymbol{\theta}_C \neq \mathbf{0} \text{ for some } C \in \mathcal{C}. \quad (8)$$

For the test statistic for (8), we take the largest value of the LRT statistics (7) for all the potential clusters in C , and define the corresponding cluster as the cluster estimate \hat{C} :

$$\nu = \max_{C \in \mathcal{C}} \{\lambda(C)\}, \quad \hat{C} = \arg \max_{C \in \mathcal{C}} \{\lambda(C)\}. \quad (9)$$

Furthermore, to compute a P -value, we consider a Monte Carlo method in the spirit of a parametric bootstrap^{29,30,64} because the null distribution of the test statistic ν in (9) does not exist in a closed form. First, compute $(\hat{\beta}'_0, \hat{\theta}'_0, \hat{\sigma}^2_0)'$, the MLEs of the parameters under H_0 in (8). Second, generate S Monte Carlo samples under H_0 with $(\hat{\beta}'_0, \hat{\theta}'_0, \hat{\sigma}^2_0)'$. Third, compute the test statistic ν in (9) for each Monte Carlo sample and the original dataset. Denote them ν_1, \dots, ν_S for the Monte Carlo samples and ν_{orig} for the original dataset. Then, the P -value is $R/(S+1)$, where R is the rank of ν_{orig} among $\{\nu_{\text{orig}}, \nu_1, \dots, \nu_S\}$ and the largest number gets a rank of 1.

The computational complexity and algorithm can follow Lee et al²⁹ by extending to the spatio-temporal domain. We provide the detailed algorithm for the spatio-temporal data in Appendix B in the Supplementary Materials.

3.2 | Multiple clusters identification

For the multiple clusters identification, we develop a sequential detection approach with the following steps:

- (i) Predefine C with N cells on the spatial lattice and the maximum radius r_{max} .
- (ii) Estimate the background coefficients $\hat{\beta}_t$ for $t = 1, \dots, T$ under H_0 in (8), and compute the residuals $e_{0it} = y_{it} - \mathbf{x}'_{it} \hat{\beta}_t$.
- (iii) Obtain the cluster $\hat{C} = \arg \max_{C \in \mathcal{C}} \{\lambda(C)\}$ with the residuals as the responses, its P -value, and corresponding coefficients $\hat{\theta}_t$ for $t = 1, \dots, T$. In addition, update the background coefficients estimates $\hat{\beta}_t$.
- (iv) Update the residuals by removing the cluster effect as $e_{jit} = e_{(j-1)it} - \mathbf{x}'_{it} \hat{\theta}_t \cdot \mathbf{I}\{i \in \hat{C}\}$, where e_{jit} 's are the residuals from the model fit with the j th cluster.
- (v) Repeat steps (iii) and (iv) until P -value $> \alpha$. That is, stop only if the P -value in step (iii) is greater than the significance level α .

The P -value corresponding to each cluster estimate is also obtained in a sequential fashion since we detect an unknown number of clusters one by one.^{29,30,65}

The multiple clusters identification method proposed above is based on the hypothesis tests for the cluster effect in both the intercepts and the slopes. Thus, in case we have more interest in the slopes than the intercepts, it is not clear to distinguish whether the significance of the cluster results from the intercepts or the slopes. To address such a problem and to provide multiple spatio-temporal clusters in the slopes and the intercepts separately, we also adopt the two-stage identification scheme.^{29,30} The details about two-stage multiple clusters identification with the spatio-temporal data are provided in Appendix C in the Supplementary Materials. Furthermore, we will refer to the approach proposed in Section 3 as the single-stage identification to distinguish from the two-stage identification.

4 | SIMULATION STUDY

In this section, we evaluate our proposed methodology via simulation studies. We set the unit square $[0, 1]^2 \in \mathbb{R}^2$ and $[0, 3] \in \mathbb{R}^1$ to be a spatial domain and a temporal domain, respectively. Furthermore, we consider 25×25 grid of cells ($N = 625$) with the center of cells $\{0.02, 0.06, \dots, 0.98\} \times \{0.02, 0.06, \dots, 0.98\}$ over the three time points $t = 1, 2, 3$ ($T = 3$). The covariate, x_{it} for cell i at time point t , is generated from the standard normal distribution $\mathcal{N}(0, 1)$. The regression coefficients in the background β_t and the variance component of the random error ε_{it} are set to be $\beta_t = (0, 0)'$ for $t = 1, 2, 3$ and $\sigma^2 = 1$, respectively. On this simulation setting, we consider a single cluster for the power evaluation and two overlapping clusters to assess the strength of our method to identify the true clusters.

Besides the spatial setting above, we also set the Southeast of the U.S. and its counties to be another spatial domain and spatial units, respectively. This study area covers 616 counties in seven states (Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, and Tennessee). Except for these spatial domain and units, the other settings for the simulation are the same as above: three time points, $\beta_t = (0, 0)'$, and $\sigma^2 = 1$. We consider three clusters defined with noncircular spatial windows. And then, we apply the proposed method to see how effectively circular windows work when the true clusters are noncircular.

4.1 | Power evaluation

For the power evaluation, we consider three single cluster settings where each cluster is defined by different center but the same radius of $3/25$ unit on the spatio-temporal domain $[0, 1]^2 \times [0, 3]$. Clusters are located at the center with 29 cells, the bottom with 18 cells and the corner with 11 cells, respectively, for each time point t . These cluster settings are shown in Figure D.1. in the Supplementary Materials.

We denote $\theta^{\text{int}} = (\theta_{C,0,1}, \theta_{C,0,2}, \theta_{C,0,3})'$ and $\theta^{\text{slp}} = (\theta_{C,1,1}, \theta_{C,1,2}, \theta_{C,1,3})'$ the cluster effect vectors in the intercepts and the slopes, respectively. Furthermore, we set two different scenarios for the cluster effect, one such that the cluster effect in the slopes is the same as in the intercepts ($\theta^{\text{int}} = \theta^{\text{slp}} = \theta^*$), and the other such that the cluster effect is in the intercepts only ($\theta^{\text{int}} = \theta^*$ and $\theta^{\text{slp}} = \mathbf{0}$). For the temporal patterns, we also consider three different settings: *Set1* : $\theta^* = (\theta^*, 0, 0)'$, *Set2* : $\theta^* = (\theta^*, \theta^*, 0)'$, and *Set3* : $\theta^* = (\theta^*, \theta^*, \theta^*)'$. That is, the cluster effect is at the first time point in *Set1*, at both the first and the second time points in *Set2* and at all three time points in *Set3*. The value of θ^* is set to be 0, $1/8$, $1/4$, $1/2$, or 1, representing five different levels of signal to noise ratio. For each combination of three circular clusters, two cluster effects ($\theta^{\text{int}}, \theta^{\text{slp}}$), three temporal patterns, and five signal θ^* levels, we simulate 1,000 datasets.

We define power to be the proportion of simulations in which the global null hypothesis (8) is rejected at the significance level $\alpha = 0.05$.^{29,30,37,44,66,67} Furthermore, we use the critical value of the test statistic v in (9) to test the identified cluster in each simulated dataset. The critical value is achieved by the empirical null distribution of v , which is based on 10,000 null simulations, at the significance level $\alpha = 0.05$ with the maximum radius $1/5$ unit.

Table 1 shows the empirical power for each simulation setting. As expected, power increases as the signal-to-noise ratio (SNR: θ/σ) increases or as the number of time points with nonzero cluster effect increases (from *Set1* to *Set3*). When the cluster effects are both in the slopes and the intercepts ($\theta^{\text{int}} = \theta^{\text{slp}} = \theta^*$) with SNR 1, it shows at least 96.1% power except when the first time point only has a nonzero cluster effect for the quarter circular cluster (*Set1* in the panel of “corner”). When the cluster effect is in the intercepts only ($\theta^{\text{slp}} = \mathbf{0}$), powers are lower than when the cluster effect in the slopes is not zero ($\theta^{\text{slp}} = \theta^*$).

For comparisons, we also conduct a power evaluation of the spatial varying coefficient model²⁹ for a single time point at the significance level $\alpha = 0.05$. We consider the same cluster settings (center, bottom, and corner) and the cluster effect both in the slopes and the intercepts with $\theta^* = 1$. The empirical powers are 99.0%, 77.5%, and 48.9% for the center, bottom, and corner, respectively. However, the spatio-temporal approach provides higher powers even when the cluster effect is only at a single time point: the empirical powers are 100.0%, 96.1%, and 55.3% for the center, bottom, and corner, respectively, when $\theta^{\text{int}} = \theta^{\text{slp}} = (1, 0, 0)'$ in Table 1. That is, there is evidence that the proposed spatio-temporal method is better at detecting spatial clusters than the spatial-only approach.

4.2 | True clusters identification

We evaluate how well our proposed method identifies two overlapping true clusters. We consider two circular windows which overlap with each other and have the same radius of $3/25$ unit. We also set two different scenarios for the cluster effects, one such that the cluster effects are in the slopes and the intercepts for each cluster and the other such that the cluster effects are in the slopes and the intercepts for one cluster, while there is the cluster effect in the intercepts only for the second cluster. That is, $\theta_{C_1}^{\text{int}} = \theta_{C_1}^{\text{slp}} = \theta_{C_2}^{\text{int}} = \theta_{C_2}^{\text{slp}} = \theta^*$ in the first scenario, and $\theta_{C_1}^{\text{int}} = \theta_{C_1}^{\text{slp}} = \theta_{C_2}^{\text{int}} = \theta^*$ and $\theta_{C_2}^{\text{slp}} = \mathbf{0}$ in the second scenario. Furthermore, similar to the power evaluation in Section 4.1, we set three different settings for the temporal patterns with $\theta^* = (\theta^*, 0, 0)$, $\theta^* = (\theta^*, \theta^*, 0)$, and $\theta^* = (\theta^*, \theta^*, \theta^*)$. A total of six cluster settings are shown in Figure D.2. The signal θ^* is set to be 2, 1, or $1/2$, corresponding to strong, medium, or weak cluster effect, respectively, relative to the error SD $\sigma = 1$. We simulate 1,000 datasets for a total of eighteen different combinations of six cluster settings and three signal θ^* levels. For each simulated dataset, we estimate the regression coefficients for the identified

TABLE 1 Power (in percentage) for cluster identification on 25×25 square grid and three time points with the maximum cluster radius $r_{\max} = 1/5$

	θ^*	Center (29 cells)			Bottom (18 cells)			Corner (11 cells)		
		Set1	Set2	Set3	Set1	Set2	Set3	Set1	Set2	Set3
(1) $\theta^{\text{int}} = \theta^{\text{slp}} = \theta^*$	0	5.1	5.1	5.1	5.4	5.4	5.4	5.3	5.3	5.3
	1/8	5.5	5.1	5.6	5.6	5.6	5.8	5.7	5.5	5.4
	1/4	7.8	11.0	16.8	6.1	6.6	7.6	5.6	6.2	6.0
	1/2	42.0	80.7	95.5	14.7	40.5	50.2	7.1	16.9	22.7
	1	100.0	100.0	100.0	96.1	99.9	100.0	55.3	97.7	99.4
(2) $\theta^{\text{int}} = \theta^*$, $\theta^{\text{slp}} = \mathbf{0}$	0	5.1	5.1	5.1	5.4	5.4	5.4	5.3	5.3	5.3
	1/8	5.2	5.0	5.5	5.4	5.4	5.5	5.3	5.5	5.4
	1/4	5.6	6.2	7.0	5.5	5.8	6.0	5.3	5.7	5.7
	1/2	8.4	20.1	43.4	5.9	10.3	17.8	5.5	7.2	9.9
	1	62.8	98.7	100.0	30.2	81.2	97.5	12.4	41.1	76.1

Note: (1) The cluster effect in the slopes is the same as in the intercepts ($\theta^{\text{int}} = \theta^{\text{slp}} = \theta^*$). (2) The cluster effect is in the intercepts only ($\theta^{\text{int}} = \theta^*$, $\theta^{\text{slp}} = \mathbf{0}$). $\theta^* = (\theta^*, 0, 0)'$ in Set1, $\theta^* = (\theta^*, \theta^*, 0)'$ in Set2, and $\theta^* = (\theta^*, \theta^*, \theta^*)'$ in Set3, respectively. The error SD is $\sigma = 1$, and θ^* is set to be 0, 1/8, 1/4, 1/2, or 1.

clusters, and we map the mean coefficient estimates in comparison with the true values. To identify clusters, we apply the sequential identification approach for multiple clusters proposed in Section 3.2 with the critical value of the test statistic (9) at the significance level $\alpha = 0.05$ which we achieved in Section 4.1.

Figure 1 provides the maps for the simulated data with the medium cluster effect ($\theta^* = 1$). The maps of true coefficients are provided in left subfigure, and the maps of the mean coefficient estimates are illustrated in right subfigure. For each subfigure, the first three columns of maps are for the intercepts, whereas the last three columns of maps are for the slopes at each time point. Furthermore, for each subfigure, a total of six rows represents the six cluster settings. From Figure 1, we see that our method identifies the true clusters well enough except for Setting2 in row 2. That is, when the cluster effect is only in the intercepts at the first time point, it is generally a challenge to identify that cluster.

The results with the strong cluster effects ($\theta^* = 2$) are given in Figure D.3 in Appendix D. The findings are similar to or even better than the case with the medium cluster effect ($\theta^* = 1$). The corresponding mean coefficient estimates are close to the true clusters and the true regression coefficients. The results with the weak cluster effects ($\theta^* = 1/2$) are given in Figure D.4 in Appendix D. It fails to identify all of the true clusters, as expected with a small SNR (1/2), especially when the cluster effect is only at the first time point.

We also conduct simulation studies for the two-stage identification method. Those results from simulation studies are similar to the findings from the single-stage approach and provided in Appendix E in the Supplementary Materials.

4.3 | Noncircular windows identification

In practice, especially with the irregular grid data (e.g., county-level data), the choice of the shape for spatial windows acts as a penalty on identifying clusters rather than forces the detected clusters to be in a particular form. Thus, the circular window is commonly used in practice because it can be simply defined even in irregular grid data.

Here, we evaluate how effectively models using a basis of circular windows can identify noncircular windows. We consider three clusters in Tennessee, Alabama, and Florida, respectively. Those clusters comprise counties that longitudes $\geq -86.41721^\circ$ in Tennessee, longitudes $\geq -86.72663^\circ$ and latitudes $\leq 32.85406^\circ$ in Alabama, and latitudes $\leq 28.05903^\circ$ in Florida, respectively. Furthermore, the cluster effects are set to be $\theta_{\text{TN}}^{\text{int}} = \theta_{\text{TN}}^{\text{slp}} = (-2, 0, 2)$ in Tennessee, $\theta_{\text{AL}}^{\text{int}} = \theta_{\text{AL}}^{\text{slp}} = (2, 2, 2)$ in Alabama, and $\theta_{\text{FL}}^{\text{int}} = \theta_{\text{FL}}^{\text{slp}} = (2, 0, -2)$ in Florida, respectively. We simulate 1,000 datasets and apply the sequential identification approach for each dataset to identify multiple clusters based on circular windows. The critical value of the test statistic (9) is achieved from 10,000 null simulations, at the significance level $\alpha = 0.05$ with $r_{\max} = 300$ km.

The maps of the mean slope estimates across the simulations are provided in Figure 2 with the true slopes maps. The true slopes maps are illustrated in the top row, while the maps of the mean slope estimates are displayed in the

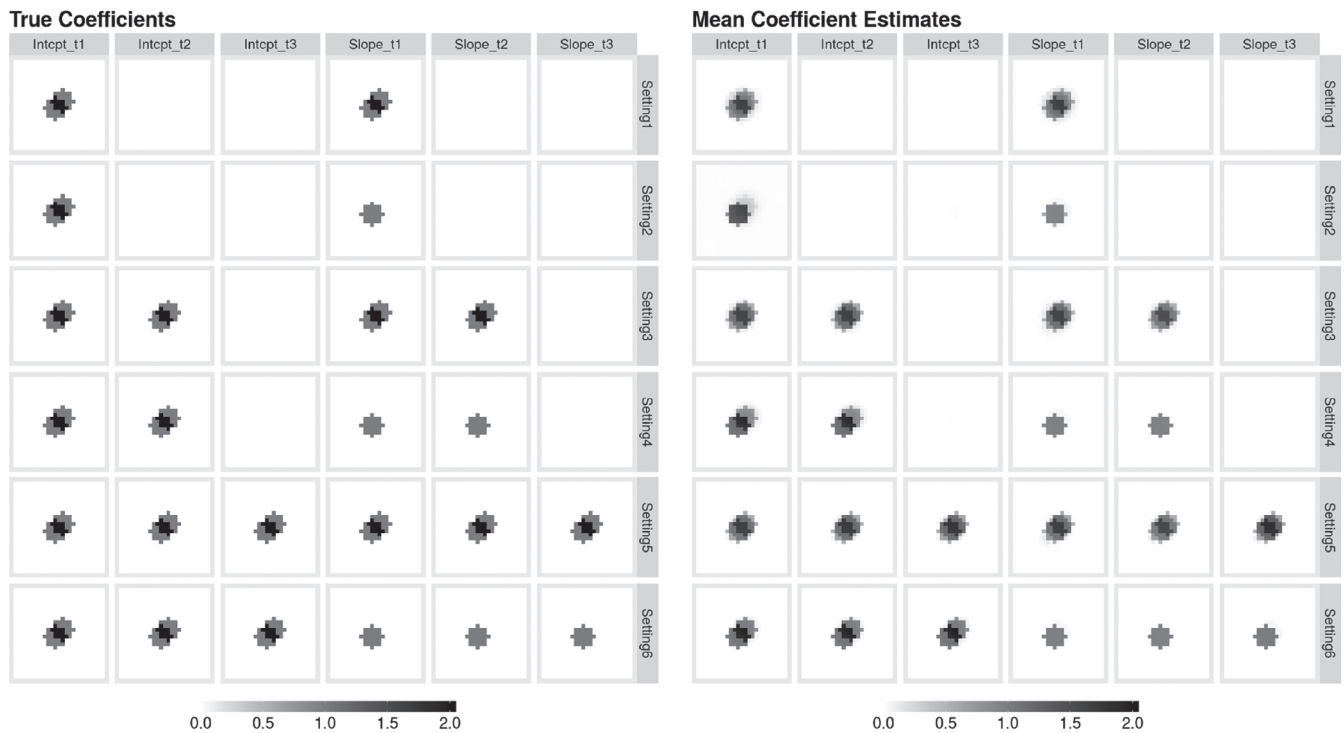


FIGURE 1 Maps of the mean coefficient estimates for each cell from the 1,000 simulated datasets with two overlapping clusters. The error SD is $\sigma = 1$, and the cluster effect is $\theta^* = 1$: the signal-to-noise ratio (θ^*/σ) is 1

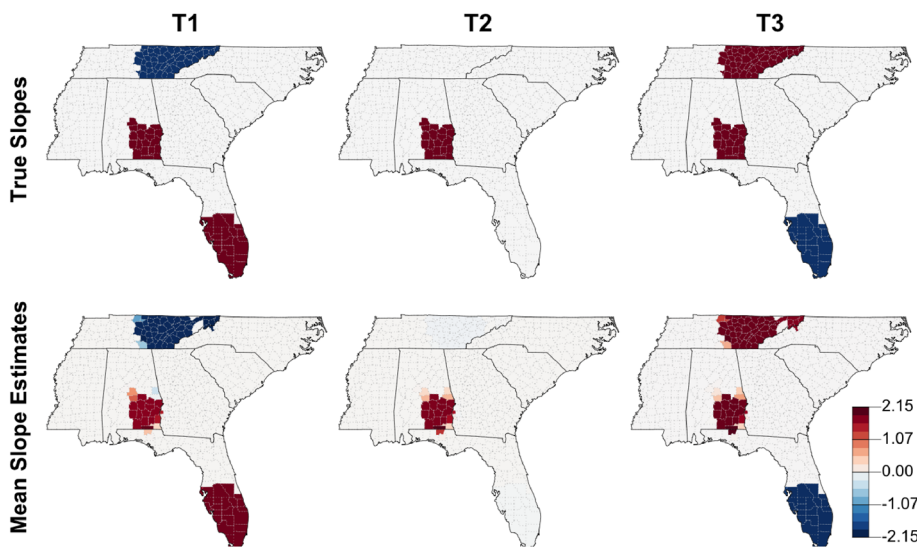


FIGURE 2 Maps of the slopes for each county with three clusters of noncircular spatial windows. Top row: Map of true slopes for each time point. Bottom row: Map of the mean slope estimates across the 1,000 simulations for each time point [Colour figure can be viewed at wileyonlinelibrary.com]

bottom row. We see that all three true clusters are far from the circular shapes. However, the mean slope estimates maps show that those noncircular true clusters are identified effectively, albeit not parsimoniously, by using circular windows.

5 | DATA EXAMPLE

We illustrate our approach to the identification of clustered spatio-temporal varying coefficients with a cancer mortality dataset in the Southeast of the U.S. For each county, the mortality rate is defined as the number of patient deaths due

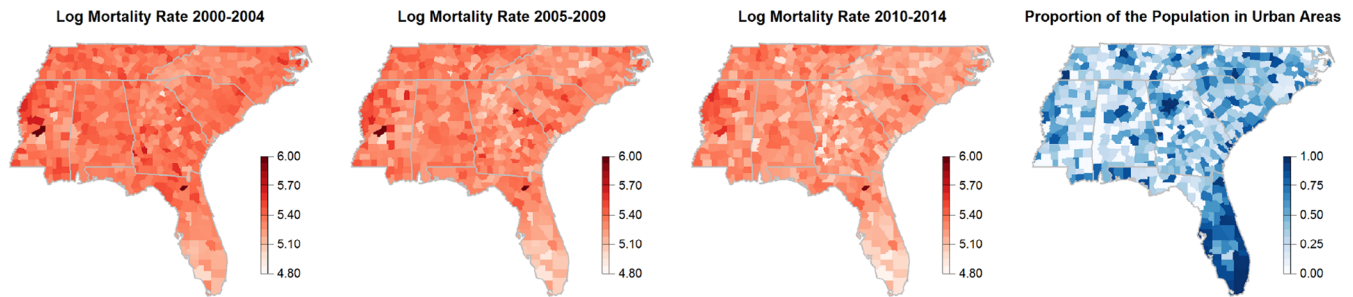


FIGURE 3 The log cancer mortality rate and the proportion of the population in urban areas for each county in the states of Alabama, Florida, Georgia, Mississippi, North Carolina, South Carolina, and Tennessee [Colour figure can be viewed at wileyonlinelibrary.com]

to malignant neoplasms (simply known as cancers, ICD-10 Codes: C00-C97) per 100,000 population per year over three 5-year periods 2000-2004, 2005-2009, and 2010-2014. These rates are age-adjusted to the 2000 U.S. standard population. The extent of urbanization is defined to be the proportion of the population in urban areas in the census year 2000. That is, we have the cancer mortality rates for three-time intervals while the extent of urbanization is based on the one census year because the levels of urbanization are relatively stable over the 15-year period 2000 to 2014.

We take the logarithm of the cancer mortality rate and consider regression models with the log cancer mortality rate and the extent of urbanization as the response variable and the covariate, respectively. It can be shown that $\text{Var}(y_{it}) = \log r_{it} \approx (n_{it}\rho_{it})^{-1} + \sigma^2$, where r_{it} is the rate, n_{it} is the county population and $\rho_{it} = E(r_{it})$ for the i th county at the time t , respectively. Since $(n_{it}\rho_{it})^{-1}$ is negligible with county populations in the thousands, we assume a constant variance. Furthermore, the residuals do not provide evidence for nonnormality, additional spatial clusters based on the spatial scan statistics, or temporal dependency based on the autocorrelation function. Thus, the assumption of independent errors seems reasonable.

The maps of the log cancer mortality rate and of the urbanization are shown in Figure 3. The log cancer mortality rate maps show an overall decreasing patterns over time. However, it is challenging to find geographical clusters of the cancer mortality rate in relation to the extent of urbanization via eyeballing. Thus, we identify possible spatio-temporal clusters using both the single-stage approach and the two-stage approach which we proposed. Here, we only illustrate the results from the single-stage identification in the main article. We relegate the details about the results from the two-stage identification to Appendix E in the Supplementary Materials.

The covariate, the extent of urbanization, is centered to have a zero mean. We set the maximum radius for a potential cluster to be $r_{\max} = 300$ km, since the largest circular spatial window with r_{\max} is large enough to cover all or the majority of each of the seven states. Furthermore, the P -value for each identified cluster is obtained from 1,000 Monte Carlo samples. For comparison, we also apply the existing clustered spatial varying coefficient regression model for each time period separately.²⁹

Table 2 provides a total of ten significant spatio-temporal clusters which are identified via the single-stage identification method at $\alpha = 0.05$. The maps of the slope estimates and the intercept estimates with the identified clusters are given in Figures 4 and 5. In each figure, the top row displays the spatio-temporal varying coefficients whereas the bottom row displays the results from the separate spatial varying coefficient models for each time period. The identified spatio-temporal clusters are qualitatively the same as the identified spatial-only clusters. Spatio-temporal approach finds more clusters in Mississippi and Tennessee which are missed by the spatial-only method. Furthermore, by considering the same spatial windows for every time periods, we could discern the temporal patterns in a certain geographical subdomain.

Figure 6 illustrates the significant spatio-temporal clusters identified via the single-stage method at $\alpha = 0.05$ and the corresponding coefficient estimates. Columns 1 and 2 are the maps of the slope and the intercept estimates for the identified clusters and the background. Among the total of ten identified clusters, two clusters in Mississippi overlap with each other and a small cluster in Georgia is nested in a bigger one. Thus, we could consider a total of eleven nonoverlapping clusters which show different patterns against the background. The maps of representative counties for these nonoverlapping clusters are illustrated in the third column of Figure 6, and we name the eleven nonoverlapping clusters based on those representative counties. Scatter plots with fitted regression lines for each nonoverlapping

C	Centroid	Radius	Counties	P-value
\hat{C}_1	Sunflower, MS	131	24	.001
\hat{C}_2	Baker, FL	31	2	.001
\hat{C}_3	Glades, FL	214	28	.001
\hat{C}_4	Franklin, GA	144	67	.001
\hat{C}_5	Dooly, GA	127	55	.001
\hat{C}_6	Davie, NC	125	34	.001
\hat{C}_7	Neshoba, MS	88	15	.001
\hat{C}_8	Coffee, TN	34	2	.018
\hat{C}_9	Wilkes, GA	42	7	.012
\hat{C}_{10}	Union, MS	37	4	.035

TABLE 2 Identified clusters via the single-stage identification method at $\alpha = 0.05$

Note: The response is the log cancer mortality rate and the covariate is the extent of the urbanization in a county.

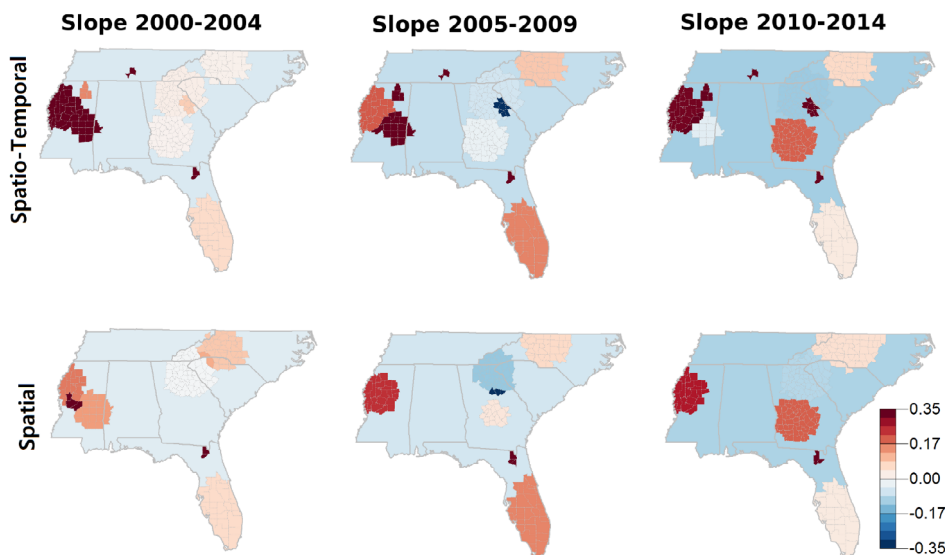
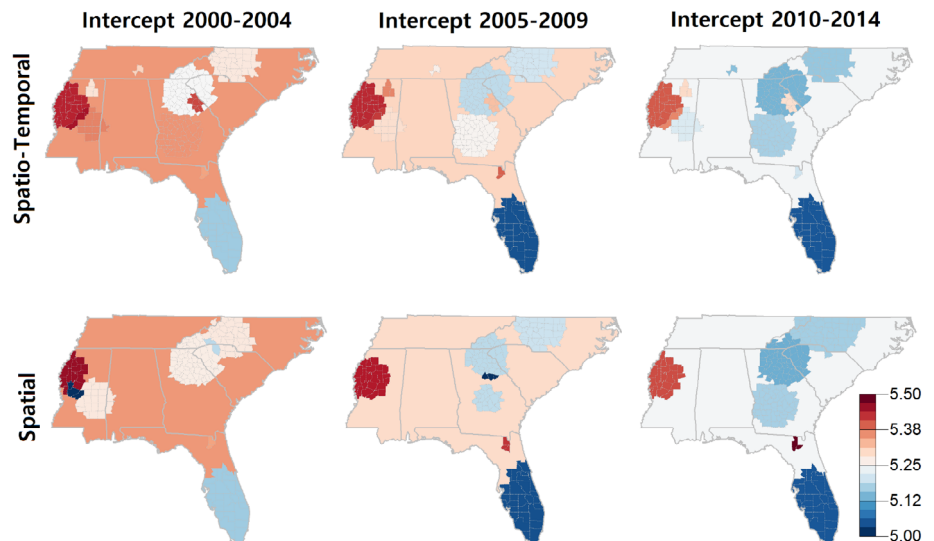


FIGURE 4 Slope estimates for overlapping clusters identified via the single-stage identification method at $\alpha = 0.05$. Top row: Spatio-temporal varying coefficient model via the single-stage identification. Bottom row: Spatial varying coefficient model via the single-stage identification for each time period [Colour figure can be viewed at wileyonlinelibrary.com]

cluster are provided at the bottom of Figure 6. Table 3 provides the coefficient estimates for each nonoverlapping cluster.

The identified clusters all show features different from the background with distinct estimated slopes and estimated intercepts. Notably, the Sunflower, MS cluster can be seen as one of the more distinct regions since it has positive estimated slopes while the background has negative estimated slopes for all three time periods. This cluster also always provides higher intercepts than the background, possibly because there is inadequate accessibility to the health care services while it is more likely to be exposed to the causes of cancers in a more urbanized area of the Sunflower, MS cluster. However, for better understanding, we would need to consult population health professionals. In addition, the background has negative estimated slopes for all three time periods, with a decreasing trend over time, suggesting a negative association between the log cancer mortality and the extent of urbanization throughout all time periods, and this negative association is becoming stronger over time. Some clusters also have a decreasing temporal trend as observed in the background, such as the Attala, MS cluster, the Neshoba, MS cluster, and the Franklin, GA cluster. In particular, the Neshoba, MS cluster has positive estimated slopes in the first two time periods but a negative estimated slope in the last time period. This is possibly due to the more urbanized counties (Oktibbeha and Lauderdale), with higher mortality in the early 2000s but the lowest mortality rates post 2010. Another interesting temporal trend in the estimated slopes is a decrease from the first to the second time period, but an increase again from the second to the third time period, such as the Sunflower, MS, the Wilkes, GA, the Dooly, GA and the Baker, FL clusters, possibly because the mortality decreases more in the more urbanized

FIGURE 5 Intercept estimates for overlapping clusters identified via the single-stage identification method at $\alpha = 0.05$. Top row: Spatio-temporal varying coefficient model via the single-stage identification. Bottom row: Spatial varying coefficient model via the single-stage identification for each time period [Colour figure can be viewed at wileyonlinelibrary.com]



counties than in the less urbanized counties from the first to the second time period. By contrast, the mortality decreases more in the less urbanized counties than in the more urbanized counties from the second to the third time period. This pattern appears clearly in the Wilkes, GA and the Dooly, GA clusters. Finally, the estimated slopes have increased from the first to the second time period, but have decreased again from the second to the third time period such as the Union, MS, the Coffee, TN, the Davie, NC and the Glades, FL clusters, possibly because that the mortality decreases more in the less urbanized counties than in the more urbanized counties from the first to the second time period. By contrast, the mortality decreases more in the more urbanized counties than in the less urbanized counties from the second to the third time period.

6 | CONCLUSIONS AND DISCUSSION

In this article, we have proposed a new approach to the identification of clustered spatio-temporal varying coefficients in a regression model. It is a substantive extension from the clustered spatial varying coefficient regression model with spatial only data as we have had a multitude of challenges to address. With the capacity to capture flexible temporal patterns comes a larger number of parameters to estimate as well as a larger number of potential clusters, from which the most significant one is to be chosen among them. Both the increased number of parameters and a large number of potential clusters require a large number of matrix manipulations and thus, potentially very high computational cost. Furthermore, if we extend the proposed method to the multiple covariates model, the number of clusters may increase with the number of covariates. However, we have tackled the computational challenges and found ways to reduce the computational complexity with multiple covariates, as provided in Appendix B. Therefore, it is plausible that the proposed method can be readily extended to more than one covariate.

Our proposed methodology can be used to locate spatial subdomains which have different relationships between a response variable and a covariate in a varying coefficient regression setting. In addition, by allowing for cylindrical clusters with the same circular spatial windows for all time points, the temporal patterns of the regression coefficients can be quantified for certain subdomains. Although we consider the circular window, adaptations to other shapes are straightforward. However, as we can see in the simulation studies, noncircular windows can be effectively, albeit not parsimoniously, represented by using circular windows. That is, the choice to use circular windows effectively acts as a penalty on the shape of the spatial window rather than forcing windows to be circular.

The simulation studies support the satisfactory performance of our method regarding the power as well as the identification of multiple true clusters. In the data example, we observe the proposed clustered spatio-temporal varying coefficient regression model reflects well the spatial heterogeneity which was identified via the spatial-only approach. Furthermore, the proposed method identified more spatial heterogeneity which cannot be found in the spatial-only approach. The companion software for our methodology R package `coefclust`, and an illustrative example are available at <https://mkamenet3.github.io/coefclust/>.

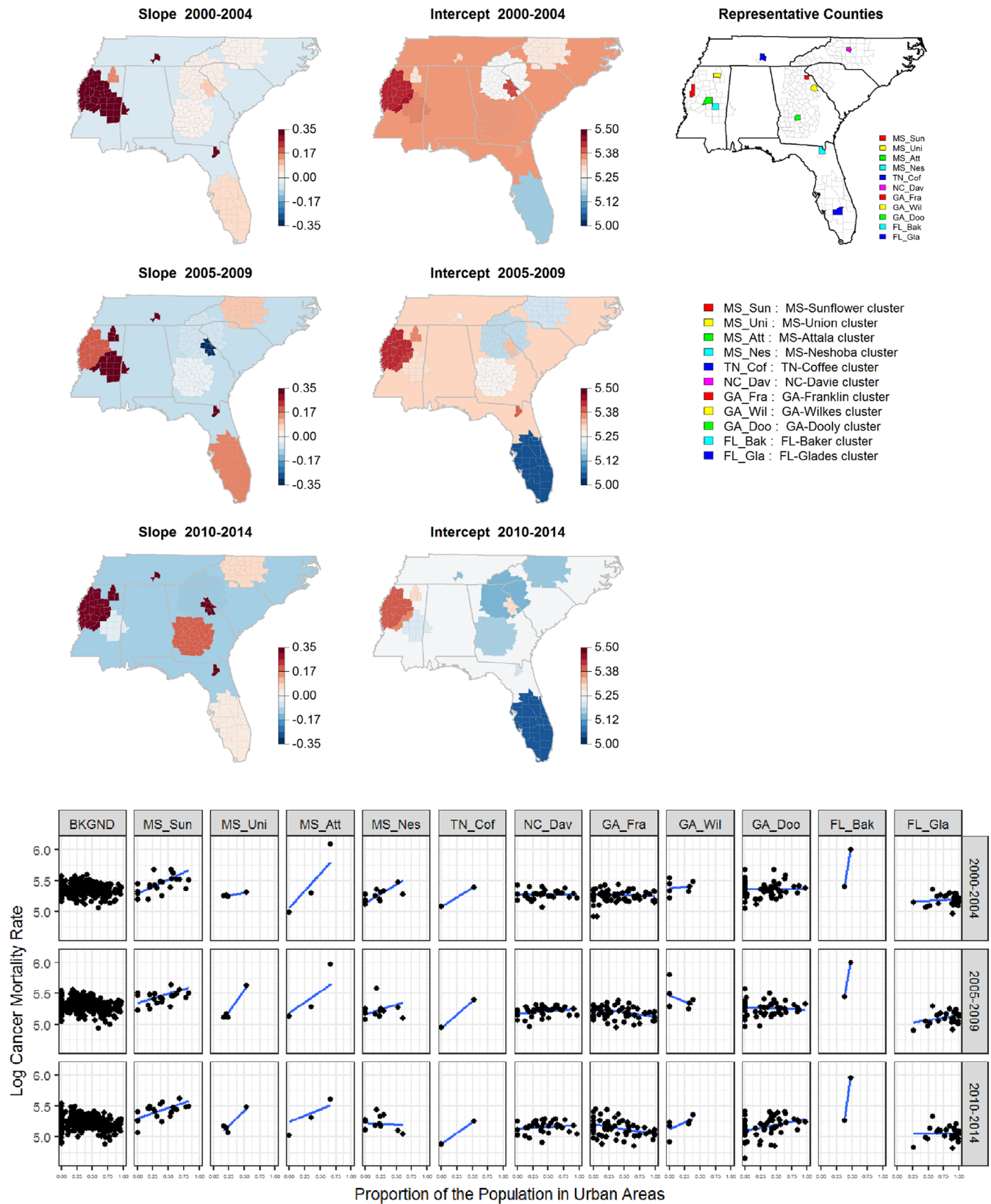


FIGURE 6 Significant overlapping clusters at $\alpha = 0.05$ via the single-stage identification method. Column 1: Slope estimates for each county. Column 2: Intercept estimates for each county. Column 3: Representative counties for nonoverlapping clusters. Bottom row: Scatter plots with fitted regression lines for each nonoverlapping cluster [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 3 Coefficient estimates for each nonoverlapping cluster from the single-stage identification method

Cluster						
Code	Name	Counties	Coeff	2000-2004	2005-2009	2010-2014
BKGND	Background	388	Intercept	5.354	5.296	5.240
			Slope	-0.047	-0.075	-0.110
MS_Sun	Sunflower, MS cluster	21	Intercept	5.461	5.458	5.426
			Slope	0.462	0.294	0.339
MS_Uni	Union, MS cluster	4	Intercept	5.283	5.372	5.298
			Slope	0.169	1.442	0.964
MS_Att	Attala, MS cluster	3	Intercept	5.471	5.442	5.394
			Slope	1.103	0.688	0.392
MS_Nes	Neshoba, MS cluster	12	Intercept	5.364	5.281	5.208
			Slope	0.595	0.319	-0.056
TN_Cof	Coffee, MS cluster	2	Intercept	5.304	5.261	5.145
			Slope	0.601	0.865	0.693
NC_Dav	Davie, NC cluster	34	Intercept	5.277	5.203	5.156
			Slope	0.005	0.087	0.058
GA_Fra	Franklin, GA cluster	60	Intercept	5.269	5.214	5.150
			Slope	-0.022	-0.138	-0.161
GA_Wil	Wilkes, GA cluster	7	Intercept	5.416	5.325	5.291
			Slope	0.086	-0.390	0.444
GA_Doo	Dooly, GA cluster	55	Intercept	5.367	5.262	5.173
			Slope	0.001	-0.036	0.194
FL_Bak	Baker, FL cluster	2	Intercept	5.371	5.422	5.229
			Slope	6.189	5.692	7.152
FL_Gla	Glades, FL cluster	28	Intercept	5.174	5.049	5.060
			Slope	0.043	0.160	-0.009

Note: The response is the log cancer mortality rate and the covariate is the extent of urbanization.

The existing methods including our proposed method for cluster detection do not quantify the uncertainty of the identified cluster yet. Although Lee et al⁶⁴ proposed an approach to the uncertainty quantification for a spatial cluster, it is restricted to the one-dimensional space and only considers a cluster in the response. However, the test statistic v in (9) is defined in the three-dimensional space and its sampling distribution is unknown. Since an analytic distribution is not available, it is challenging to make general inference about the unknown spatial or spatio-temporal cluster. Furthermore, it would be interesting to develop the inference of cluster-specific coefficients. We have not pursued inference here, such as standard errors or P -values, since each cluster was sequentially detected under the single cluster assumption. We leave this and others for future research.

ACKNOWLEDGEMENTS

Funding has been provided by a USGS CESU project and a pilot project from the Center for Demography of Health and Aging at the University of Wisconsin-Madison and by King Abdullah University of Science and Technology (KAUST), Office of Sponsored Research (OSR) under Award No: OSR-2019-CRG7-3800.

CONFLICT OF INTEREST


The authors declare no potential conflict of interests.

DATA AVAILABILITY STATEMENT

The source of the cancer mortality dataset is Centers for Disease Control and Prevention (<http://wonder.cdc.gov>), and the source of the urbanization dataset is Census Bureau (<http://www.census.gov>).

ORCID

Junho Lee  <https://orcid.org/0000-0002-8001-700X>

Maria E. Kamenetsky  <https://orcid.org/0000-0002-5401-7786>

Ronald E. Gangnon  <https://orcid.org/0000-0003-2587-6714>

REFERENCES

1. Brunson C, Fotheringham AS, Charlton ME. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geogr Anal.* 1996;28(4):281-298. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>.
2. Fotheringham AS, Brunson C, Charlton ME. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. New York, NY: Wiley; 2002.
3. Wang N, Mei CL, Yan XD. Local linear estimation of spatially varying coefficient models: an improvement on the geographically weighted regression technique. *Environ Plann A Economy Space.* 2008;40(4):986-1005. <https://doi.org/10.1068/a3941>.
4. Leong YY, Yue JC. A modification to geographically weighted regression. *Int J Health Geogr.* 2017;16(1):11. <https://doi.org/10.1186/s12942-017-0085-9>.
5. Windle MJS, Rose GA, Devillers R, Fortin MJ. Exploring spatial non-stationarity of fisheries survey data using geographically weighted regression (GWR): an example from the Northwest Atlantic. *ICES J Mar Sci.* 2009;67(1):145-154. <https://doi.org/10.1093/icesjms/fsp224>.
6. Shoff C, Yang TC, Matthews SA. What has geography got to do with it? using GWR to explore place-specific associations with prenatal care utilization. *GeoJournal.* 2012;77(3):331-341. <https://doi.org/10.1007/s10708-010-9405-3>.
7. Wang C, Zhang J, Yan X. The use of geographically weighted regression for the relationship among extreme climate indices in China. *Math Probl Eng.* 2012;2012:369539. <https://doi.org/10.1155/2012/369539>.
8. Yang TC, Matthews SA. Understanding the non-stationary associations between distrust of the health care system, health conditions, and self-rated health in the elderly: a geographically weighted regression approach. *Health Place.* 2012;18(3):576-585. <https://doi.org/10.1016/J.HEALTHPLACE.2012.01.007>.
9. Black NC. An ecological approach to understanding adult obesity prevalence in the united states: a county-level analysis using geographically weighted regression. *Appl Spat Anal Policy.* 2014;7(3):283-299. <https://doi.org/10.1007/s12061-014-9108-0>.
10. Kang D, Dall'Erba S. Exploring the spatially varying innovation capacity of the US counties in the framework of Griliches' knowledge production function: a mixed GWR approach. *J Geogr Syst.* 2016;18(2):125-157. <https://doi.org/10.1007/s10109-016-0228-8>.
11. Benassi F, Naccarato A. Households in potential economic distress. a geographically weighted regression model for Italy, 2001-2011. *Spatial Stat.* 2017;21:362-376. <https://doi.org/10.1016/J.SPASTA.2017.03.002>.
12. Lee HJ, Kim EJ, Lee SW. Examining spatial variation in the effects of Japanese red pine (*Pinus densiflora*) on burn severity using geographically weighted regression. *Sustain For.* 2017;9(5). <https://doi.org/10.3390/su9050804>.
13. Wang C, Du S. Analyzing explanatory factors of urban pluvial floods in Shanghai using geographically weighted regression. *Stoch Env Res Risk A.* 2017;31(7):1777-1790. <https://doi.org/10.1007/s00477-016-1242-6>.
14. Wu C, Liu G, Huang C. Prediction of soil salinity in the yellow river delta using geographically weighted regression. *Arch Agron Soil Sci.* 2017;63(7):928-941. <https://doi.org/10.1080/03650340.2016.1249475>.
15. Wu SS, Yang H, Guo F, Han RM. Spatial patterns and origins of heavy metals in Sheyang River catchment in Jiangsu, China based on geographically weighted regression. *Sci Total Environ.* 2017;580:1518-1529. <https://doi.org/10.1016/J.SCITOTENV.2016.12.137>.
16. Acharya BK, Cao C. Modeling the spatially varying risk factors of dengue fever in Jhapa district, Nepal, using the semi-parametric geographically weighted regression model. *Int J Biometeorol.* 2018;62(11):1973-1986. <https://doi.org/10.1007/s00484-018-1601-8>.
17. Chao L, Zhang K, Li Z, Zhu Y, Wang J, Yu Z. Geographically weighted regression based methods for merging satellite and gauge precipitation. *J Hydrol.* 2018;558:275-289. <https://doi.org/10.1016/J.JHYDROL.2018.01.042>.
18. Mao L, Yang J, Deng G. Mapping rural-urban disparities in late-stage cancer with high-resolution rurality index and GWR. *Spatial Spatio-temporal Epidemiol.* 2018;26:15-23. <https://doi.org/10.1016/J.SSTE.2018.04.001>.
19. Ribeiro MC, Pereira MJ. Modelling local uncertainty in relations between birth weight and air quality within an urban area: combining geographically weighted regression with geostatistical simulation. *Environ Sci Pollut Res.* 2018;25(26):25942-25954. <https://doi.org/10.1007/s11356-018-2614-x>.
20. Sultana S, Pourebrahim N, Kim H. Household energy expenditures in North Carolina: a geographically weighted regression approach. *Sustainability.* 2018;10(5):1511. <https://doi.org/10.3390/su10051511>.
21. Xu Z, Ouyang A. The factors influencing china's population distribution and spatial heterogeneity: a prefectural-level analysis using geographically weighted regression. *Appl Spat Anal Policy.* 2018;11(3):465-480. <https://doi.org/10.1007/s12061-017-9224-8>.
22. Wende D. Spatial risk adjustment between health insurances: using GWR in risk adjustment models to conserve incentives for service optimisation and reduce MAUP. *Eur J Health Econ.* 2019;20(7):1079-1091. <https://doi.org/10.1007/s10198-019-01079-6>.
23. Runadi T, Widyaningsih Y, Lestari D. Modeling total crime and the affecting factors in Central Java using geographically weighted regression. *J Phys Conf Ser.* 2020;1442:12026. <https://doi.org/10.1088/1742-6596/1442/1/012026>.

24. Zhao J, Yang L, Li L, Wang L, Hu Q, Wang Y. Analysis of the lake-effect on precipitation in the Taihu Lake Basin based on the GWR merged precipitation. *Water*. 2020;12(1):180. <https://doi.org/10.3390/w12010180>.
25. Assunção RM. Space varying coefficient models for small area data. *Environmetrics*. 2003;14(5):453-473. <https://doi.org/10.1002/env.599>.
26. Gamerman D, Moreira AR, Rue H. Space-varying regression models: specifications and simulation. *Comput Stat Data Anal*. 2003;42(3):513-533. [https://doi.org/10.1016/S0167-9473\(02\)00211-6](https://doi.org/10.1016/S0167-9473(02)00211-6).
27. Wang W, Sun Y. Penalized local polynomial regression for spatial data. *Biometrics*. 2019;75(4):1179-1190. <https://doi.org/10.1111/biom.13077>.
28. Lawson AB, Choi J, Zhang J. Prior choice in discrete latent modeling of spatially referenced cancer survival. *Stat Methods Med Res*. 2014;23(2):183-200. <https://doi.org/10.1177/0962280212447148>.
29. Lee J, Gangnon RE, Zhu J. Cluster detection of spatial regression coefficients. *Stat Med*. 2017;36(7):1118-1133. <https://doi.org/10.1002/sim.7172>.
30. Lee J, Sun Y, Chang HH. Spatial cluster detection of regression coefficients in a mixed-effects model. *Environmetrics*. 2020;31(2):e2578. <https://doi.org/10.1002/env.2578>.
31. Lagona F, Ranalli M, Barbi E. A model with space-varying regression coefficients for clustering multivariate spatial count data. *Biom J*. 2020;62(2):1508-1524. <https://doi.org/10.1002/bimj.201900229>.
32. Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Stat Med*. 1995;14(8):799-810. <https://doi.org/10.1002/sim.4780140809>.
33. Kulldorff M. A spatial scan statistic. *Commun Stat Part A*. 1997;26:1481-1496. <https://doi.org/10.1080/03610929708831995>.
34. Kulldorff M, Athas W, Feuer E, Miller B, Key C. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos. *New Mexico Am J Publ Health*. 1998;88:1377-1380. <https://doi.org/10.2105/AJPH.88.9.1377>.
35. Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. *J Royal Stat Soc A Stat Soc*. 2001;164(1):61-72. <https://doi.org/10.1111/1467-985X.00186>.
36. Duczmal L, Assunção R. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Comput Stat Data Anal*. 2004;45(2):269-286. [https://doi.org/10.1016/S0167-9473\(02\)00302-X](https://doi.org/10.1016/S0167-9473(02)00302-X).
37. Gangnon RE, Clayton MK. Likelihood-based tests for localized spatial clustering of disease. *Environmetrics*. 2004;15(8):797-810. <https://doi.org/10.1002/env.662>.
38. Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *Int J Health Geogr*. 2005;4:11. <https://doi.org/10.1186/1476-072X-4-11>.
39. Assunção R, Costa M, Tavares A, Ferreira S. Fast detection of arbitrarily shaped disease clusters. *Stat Med*. 2006;25(5):723-742. <https://doi.org/10.1002/sim.2411>.
40. Kulldorff M, Huang L, Pickle L, Duczmal L. An elliptic spatial scan statistic. *Stat Med*. 2006;25(22):3929-3943. <https://doi.org/10.1002/sim.2490>.
41. Takahashi K, Kulldorff M, Tango T, Yih K. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. *Int J Health Geogr* 2008; 7: 14. <https://doi.org/10.1186/1476-072X-7-14>
42. Jung I. A generalized linear models approach to spatial scan statistics for covariate adjustment. *Stat Med*. 2009;28(7):1131-1143. <https://doi.org/10.1002/sim.3535>.
43. Kulldorff M, Huang L, Konty K. A scan statistic for continuous data based on the normal probability model. *Int J Health Geogr*. 2009;8(1):58. <https://doi.org/10.1186/1476-072X-8-58>.
44. Gangnon RE. Local multiplicity adjustments for spatial cluster detection. *Environ Ecol Stat*. 2010;17(1):55-71. <https://doi.org/10.1007/s10651-008-0101-0>.
45. Jung I, Kulldorff M, Richard OJ. A spatial scan statistic for multinomial data. *Stat Med*. 2010;29(18):1910-1918. <https://doi.org/10.1002/sim.3951>.
46. Jung I, Lee H. Spatial cluster detection for ordinal outcome data. *Stat Med*. 2012;31(29):4040-4048. <https://doi.org/10.1002/sim.5475>.
47. Neill DB. Fast subset scan for spatial pattern detection. *J Royal Stat Soc Ser B (Stat Methodol)*. 2012;74(2):337-360. <https://doi.org/10.1111/j.1467-9868.2011.01014.x>.
48. Shu L, Jiang W, Tsui KL. A standardized scan statistic for detecting spatial clusters with estimated parameters. *Naval Res Logist (NRL)*. 2012;59(6):397-410. <https://doi.org/10.1002/nav.21493>.
49. Xu J, Gangnon RE. Stepwise and stagewise approaches for spatial cluster detection. *Spatial Spatio-temporal Epidemiol*. 2016;17:59-74. <https://doi.org/10.1016/J.SSTE.2016.04.007>.
50. Lin PS, Kung YH, Clayton M. Spatial scan statistics for detection of multiple clusters with arbitrary shapes. *Biometrics*. 2016;72(4):1226-1234. <https://doi.org/10.1111/biom.12509>.
51. Gangnon RE, Clayton MK. Bayesian detection modeling of spatial disease clustering. *Biometrics*. 2000;56(3):922-935. <https://doi.org/10.1111/j.0006-341X.2000.00922.x>.
52. Gangnon RE, Clayton MK. A hierarchical model for spatially clustered disease rates. *Stat Med*. 2003;22(20):3213-3228. <https://doi.org/10.1002/sim.1570>.
53. Gangnon RE, Clayton MK. Cluster detection using Bayes factors from overparameterized cluster models. *Environ Ecol Stat*. 2007;14:69-82. <https://doi.org/10.1007/s10651-006-0007-7>.
54. Lawson AB. Cluster modelling of disease incidence via RJMCMC methods: a comparative evaluation. *Stat Med*. 2000;19:2361-2375.
55. Clark AB, Lawson AB. Spatio-temporal cluster modelling of small area health data. In: Lawson AB, Denison D, eds. *Spatial Cluster Modelling*. Boca Raton, FL: Chapman & Hall/CRC Press; 2002:235-258.

56. Wakefield J, Kim A. A bayesian model for cluster detection. *Biostatistics*. 2013;14(4):752-765. <https://doi.org/10.1093/biostatistics/kxt001>.
57. Yan P, Clayton MK. A cluster model for space-time disease counts. *Stat Med*. 2006;25(5):867-881. <https://doi.org/10.1002/sim.2424>.
58. Gangnon RE. A model for space-time cluster detection using spatial clusters with flexible temporal risk patterns. *Stat Med*. 2010;29(22):2325-2337. <https://doi.org/10.1002/sim.3984>.
59. Hossain MM, Lawson AB. Space-time Bayesian small area disease risk models: development and evaluation with a focus on cluster detection. *Environ Ecol Stat*. 2010;17(1):73-95. <https://doi.org/10.1007/s10651-008-0102-z>.
60. Lawson AB, Song HR, Cai B, Hossain MM, Huang K. Space-time latent component modeling of geo-referenced health data. *Stat Med*. 2010;29(19):2012-2027. <https://doi.org/10.1002/sim.3917>.
61. Napier G, Lee D, Robertson C, Lawson A. A Bayesian space-time model for clustering areal units based on their disease trends. *Biostatistics*. 2018;20(4):681-697. <https://doi.org/10.1093/biostatistics/kxy024>.
62. Dreassi E, Biggeri A, Catelan D. Space-time models with time-dependent covariates for the analysis of the temporal lag between socioeconomic factors and lung cancer mortality. *Stat Med*. 2005;24(12):1919-1932. <https://doi.org/10.1002/sim.2063>.
63. Cai B, Lawson AB, Hossain MM, Choi J. Bayesian latent structure models with space-time dependent covariates. *Stat Model*. 2012;12(2):145-164. <https://doi.org/10.1177/1471082X1001200202>.
64. Lee J, Gangnon RE, Zhu J, Liang J. Uncertainty of a detected spatial cluster in 1D: quantification and visualization. *Stat*. 2017;6(1):345-359. <https://doi.org/10.1002/sta4.161>.
65. Zhang Z, Assunção R, Kulldorff M. Spatial scan statistic adjusted for multiple clusters. *J Probab Stat*. 2010;11:642379. <https://doi.org/10.1155/2010/642379>.
66. Waller LA, Hill EG, Rudd RA. The geography of power: statistical performance of tests of clusters and clustering in heterogeneous populations. *Stat Med*. 2006;25(5):853-865. <https://doi.org/10.1002/sim.2418>.
67. Gangnon RE. Local multiplicity adjustment for the spatial scan statistic using the gumbel distribution. *Biometrics*. 2012;68(1):174-182. <https://doi.org/10.1111/j.1541-0420.2011.01643.x>.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Lee J, Kamenetsky ME, Gangnon RE, Zhu J. Clustered spatio-temporal varying coefficient regression model. *Statistics in Medicine*. 2021;40:465–480. <https://doi.org/10.1002/sim.8785>