Received: 23 November 2022 | Revised: 18 March 2024 | Accepted: 25 March 2024

DOI: 10.1002/env.2849

RESEARCH ARTICLE

WILEY

Scanner: Simultaneously temporal trend and spatial cluster detection for spatial-temporal data

Xin Wang[1] | **Xin Zhang**[2]

[1]Department of Mathematics and Statistics, San Diego State University, San Diego, California, USA

[2]Department of Statistics, Iowa State University, Ames, Iowa, USA

Correspondence
Xin Wang, Department of Mathematics and Statistics, San Diego State University, San Diego, CA, USA.
Email: xwang14@sdsu.edu

Funding information
National Science Foundation, Grant/Award Number: NSF-SES-2316353

Abstract

Identifying the underlying trajectory pattern in the spatial-temporal data analysis is a fundamental but challenging task. In this paper, we study the problem of simultaneously identifying temporal trends and spatial clusters of spatial-temporal trajectories. To achieve this goal, we propose a novel method named spatial clustered and sparse nonparametric regression (Scanner). Our method leverages the B-spline model to fit the temporal data and penalty terms on spline coefficients to reveal the underlying spatial-temporal patterns. In particular, our method estimates the model by solving a doubly-penalized least square problem, in which we use a group sparse penalty for trend detection and a spanning tree-based fusion penalty for spatial cluster recovery. We also develop an algorithm based on the alternating direction method of multipliers (ADMM) algorithm to efficiently minimize the penalized least square loss. The statistical consistency properties of Scanner estimator are established in our work. In the end, we conduct thorough numerical experiments to verify our theoretical findings and validate that our method outperforms the existing competitive approaches.

KEYWORDS

fusion penalty, precipitation data, spatial temporal data, spatial cluster detection, temporal trend detection

1 | INTRODUCTION

Spatial-temporal data analysis has attracted increasing research interests and applications in many scientific and engineering fields, such as disease spread analysis (Feng, 2022; Lawson, 2018; Torabi & Rosychuk, 2011; Ugarte et al., 2010), climate data analysis (Erhardt et al., 2015; Velarde et al., 2004; Wan et al., 2021; Zhang et al., 2016), house market study (Gong & de Haan, 2018; Wang et al., 2022), and environmental science (Fioravanti et al., 2022; Johnson et al., 2023; Jurek & Katzfuss, 2023; Rougier et al., 2023; Zhang et al., 2023). In these practical applications, it is an essential step to investigate the underlying spatial-temporal pattern. For example, Feng (2022) used a spatial-temporal generalized additive model for the COVID-19 mortality data to detect epidemic peaks over time; Zhang et al. (2016) modeled the daily precipitation data with a spatial functional linear model to understand the regionalization pattern of precipitation. In this work, we aim to both detect the temporal trends and identify the spatial clusters of the spatial-temporal data.

Suppose n spatial-temporal trajectories as $\{y(\mathbf{s}_i, t)\}_{i=1}^n$, where $y(\mathbf{s}_i, t)$ is the observation at the location $\mathbf{s}_i \in \mathbb{R}^2$, and time $t \in \mathcal{T}$. We consider the following model

$$y(\mathbf{s}_i, t) = g(\mathbf{s}_i, t) + \epsilon(\mathbf{s}_i, t), \tag{1}$$

Xin Wang and Xin Zhang contributed equally to this work.

where $g(\mathbf{s}_i, t)$ is the unknown smooth function depending on location and timestamp, and $\epsilon(\mathbf{s}_i, t)$ is the measurement error. Our goal is to detect locations with temporal trends, that is, $g(\mathbf{s}_i, t)$ is not constant, and the spatial cluster pattern such that $g(\mathbf{s}_i, t) = g(\mathbf{s}_{i'}, t)$ for all t if two locations $\mathbf{s}_i$ and $\mathbf{s}_{i'}$ are from the same spatial cluster, simultaneously.

Our study is motivated by a dataset recording the precipitation in China from 1995 to 2014. In the dataset, the precipitation data are collected from 654 meteorological stations. We show the stations' locations in Figure 1a and the observations of 20 yearly precipitation of randomly selected 50 stations in Figure 1b. In the literature, it has been widely studied to understand the temporal changes and discover the underlying local precipitation regions, which benefits climate prediction and flood zone management (Chen et al., 2009; Zhang et al., 2016). In our study, we first investigated the yearly precipitation pattern of each station. We used spline regression models to fit each station's data and evaluate the model significance with the hypothesis test in Wood (2013), which is a Wald type test to test whether the smooth term is 0 or not. The p-values are shown in Figure 2a. We found that only 53 stations have p-values smaller than 0.05, while the spline models of the other stations are not statistically significant from the testing results. The result of the hypothesis test gives us a hint that most of the stations have stationary yearly precipitation. Multiple comparisons are also an issue in this procedure. If using Bonferroni's correction to correct the multiple comparisons in this procedure, the significance level will become $0.05/660 = 7 \times 10^{-5}$, which leads to only one significant station. However, Bonferroni's correction does not consider spatial information. Besides this, the p-values are calculated in each station based on a limited number of observations, which does not borrow observations from other stations. Furthermore, to detect the local precipitation regions, we adopted the k-means clustering method to the spline coefficients across these stations (Abraham et al., 2003). We found 37 clusters based on Gap Statistic (Dudoit & Fridlyand, 2002; Tibshirani et al., 2001). Gap statistic is a widely used method to select the number of clusters in k-means. For a given number of clusters k, the pooled within-cluster sum of

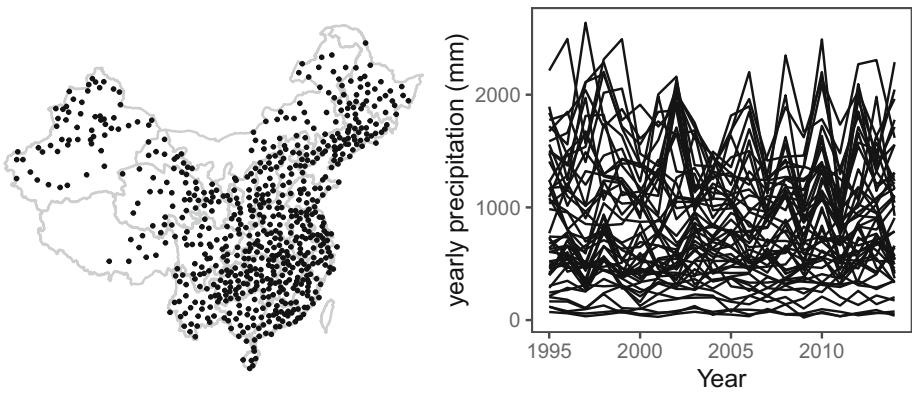


FIGURE 1 Locations and observations of meteorological stations. (a) The locations of 660 analyzed stations; (b) 50 temporal trend lines of 50 selected stations randomly selected from all 660 stations.

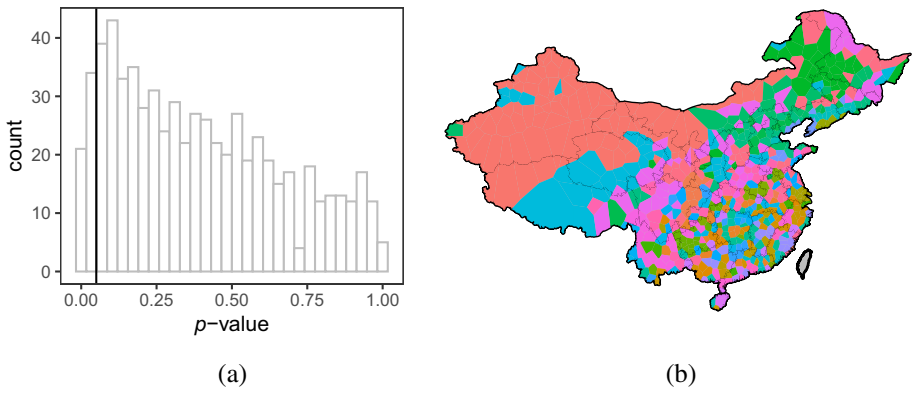


FIGURE 2 Preliminary results (a) p-value of individual models. (b) Clusters based on K-means.

squares around cluster means W_k is calculated. The Gap statistic is defined as the difference between the expectation of $\log(W_k)$ and the observed $\log(W_k)$, where the expectation is approximated based on bootstrap samples. The value of k with the minimum value of Gap statistic is selected as the optimal number of clusters. As shown in Figure 2b, some detected clusters are spatially contiguous, and some are scattered. These preliminary findings motivate our study to identify temporal trends and spatial clusters. On the one hand, we want to avoid insignificant temporal changes to improve the model estimation accuracy. On the other hand, by revealing the spatial cluster pattern, we can further enhance the model estimation efficiency by integrating the data within the same cluster.

However, it is a highly non-trivial task to develop an efficient method to automatically detect both temporal trends and spatial clusters for spatial-temporal trajectories. Up to date, there exists a bunch of work on identifying spatially contiguous clusters based on hypothesis tests, including Kulldorff and Nagarwalla (1995), Jung et al. (2007) and Cook et al. (2007). However, these methods are developed for one-time spatial observations and cannot be applied to cluster the temporal curves across spatial regions. One of the popular methods of clustering temporal curves is fitting the curves with statistical models and then clustering the model coefficients. For example, in Abraham et al. (2003) and Zhang et al. (2015), the authors estimated the temporal curves with some basis functions and used k-means on the basis coefficients to perform the clustering. James and Sugar (2003), Luan and Li (2003), and Coffey et al. (2014) used B-spline models to fit temporal data and identified the cluster pattern with linear mixed models. However, these clustering methods did not consider locations' spatial correlation. To identify the spatially contiguous clusters, Romano et al. (2013) and Haggarty et al. (2015) adopted variogram models. Jiang and Serban (2012), Liang et al. (2021), and Hu et al. (2022) used the Markov random field to incorporate spatial dependency. However, none of these mentioned works considered temporal trend detection.

Although there is no naive solution in the literature to address both the temporal trends detection and spatial clusters identification tasks, we, fortunately, find some connections between the two tasks: As no temporal trend for $\mathbf{s}_i$ means $g(\mathbf{s}_i, t) - c = 0$ for some constant c for all t, and $\mathbf{s}_i, \mathbf{s}_{i'}$ from the same cluster is equivalent to $g(\mathbf{s}_i, t) - g(\mathbf{s}_{i'}, t) = 0$ for all t, the two tasks can be reformulated as sparse structure recovering problems. The penalization method has been widely adopted to identify sparse components (Fan & Li, 2001; Simon et al., 2013; Tibshirani, 1996; Zhang, 2010). For example, Wang et al. (2008) used the smoothly clipped absolute deviation (SCAD, Fan & Li, 2001) to select significant functional components in varying-coefficient models, and Huang et al. (2010) performed variable selection in nonparametric additive models with group lasso (Simon et al., 2013). In these methods, spline models were adopted to approximate the smooth functions, and penalization terms were added to spline coefficients to select nonzero components. Also, there exist a few works using penalization methods to discover model-based clusters (Liu et al., 2023; Ma et al., 2020; Ma & Huang, 2017; Wang, 2024; Wang, Zhu, & Zhang, 2023; Yang et al., 2019; Zhu & Qu, 2018). Specifically, the fusion penalty is adopted in these methods to sparsify the model coefficient differences. With zero coefficient difference, two subjects will have the same estimated coefficients, which indicates that they come from the same cluster. However, to the best of our knowledge, there is no penalization method in the literature that can solve the temporal trends detection and spatial clusters identification tasks simultaneously.

To simultaneously address temporal trends detection and spatial cluster identification tasks, we propose a novel penalized approach named spatial clustered and sparse nonparametric regression (Scanner) method in this paper. We adopt the nonparametric additive model to estimate the temporal curves for the spatial locations. Inspired by the sparse penalization technique, our model estimation method incorporates two penalty terms: one group sparse penalty to select temporal trends and the other spanning tree-based fusion penalty to identify spatial clusters. To solve the doubly-penalized estimation problem, we develop an ADMM-based algorithm (Boyd et al., 2011), which enjoys fast convergence. We also investigate the theoretical properties of our proposed Scanner estimator. It is shown that our estimator is statistically consistent and has the capability to reveal the true temporal trends and spatial clusters with probability one. Our theoretical findings are also validated by thorough numerical experiments.

The article is organized as follows. In Section 2, we will propose our model and develop the model estimation method. In Section 3, we will establish the theoretical properties of our proposed estimator. We will provide the simulation study in Section 4, which compares our proposed method with the existing competitive approaches under different scenarios. Our proposed method will be applied to study the precipitation dataset in Section 5. We will conclude the work in Section 6.

2 | METHODOLOGY

In this section, we will first introduce our Scanner model in Section 2.1. Then, we will give the penalized estimation method in Section 2.2, which can simultaneously achieve the detection of temporal trends and spatial clusters. To solve

the penalized optimization problem, we will develop an ADMM-based algorithm in Section 2.3. Finally, we will provide the hyper-parameter selection method in Section 2.4.

2.1 | The Scanner model

Given a spatial-temporal dataset $\{\{y(\mathbf{s}_i, t_l)\}_{l=1}^{T}\}_{i=1}^{n}$, we assume the observation $y(\mathbf{s}_i, t_l)$ at the location $\mathbf{s}_i$ and time t_l follows

$$y(\mathbf{s}_i, t_l) = g(\mathbf{s}_i, t_l) + \epsilon(\mathbf{s}_i, t_l), \tag{2}$$

where $g(\mathbf{s}_i, t)$ is the unknown smooth function depending on location and timestamp, and $\epsilon(\mathbf{s}_i, t)$ is the measurement error.

Denote the set of locations with and without temporal trends as $\mathcal{S}_f$ and $\mathcal{N}_f$, respectively. Thus, $g(\mathbf{s}_i, t)$ is not a constant if $\mathbf{s}_i \in \mathcal{S}$, and $g(\mathbf{s}_i, t)$ is a constant if $\mathbf{s}_i \in \mathcal{N}_f$. Regarding spatial clustering, without loss of generality, suppose there are K underlying non-overlapped spatial clusters $\{C_k\}_{k=1}^{K}$. In the kth cluster C_k, $g^{(k)}(t)$ is the mean function. If two locations $\mathbf{s}_i$ and $\mathbf{s}_{i'}$ are from the same cluster C_k, then it holds $g(\mathbf{s}_i, t) = g(\mathbf{s}_{i'}, t) = g^{(k)}(t)$ for all t. Otherwise $g(\mathbf{s}_i, t) \neq g(\mathbf{s}_{i'}, t)$. Note that, for the kth cluster C_k, if a location $\mathbf{s}_i$ has no temporal trend, then according to our definition, $g^{(k)}(t)$ is a constant for all t, and we have all locations in C_k have no temporal trend and thus $C_k \subset \mathcal{N}_f$; Otherwise, $C_k \subset \mathcal{S}_f$ with all locations in C_k have the same temporal trend.

To investigate the temporal trend of $g(\mathbf{s}_i, t)$, we decompose it into two parts, a location-specific intercept $\mu(\mathbf{s}_i)$ and a trend function $f(\mathbf{s}_i, t)$, such that

$$y(\mathbf{s}_i, t_l) = \mu(\mathbf{s}_i) + f(\mathbf{s}_i, t_l) + \epsilon(\mathbf{s}_i, t_l). \tag{3}$$

Here, we assume the expected value $E(f(\mathbf{s}_i, t)) = 0$ with respect to t to guarantee the model identifiability. If a location $\mathbf{s}_i$ has no temporal trend, then its trend function $f(\mathbf{s}_i, t)$ would become zero, and $g(\mathbf{s}_i, t) = \mu(\mathbf{s}_i)$ for all t is a constant.

In the kth cluster C_k, $g^{(k)}(t) = \mu^{(k)} + f^{(k)}(t)$, where $\mu^{(k)}$ and $f^{(k)}(t)$ are the intercept and the trend function, respectively. If two locations $\mathbf{s}_i$ and $\mathbf{s}_{i'}$ are from the same cluster C_k, then it holds $\mu(\mathbf{s}_i) = \mu(\mathbf{s}_{i'}) = \mu^{(k)}$, and $f(\mathbf{s}_i, t) = f(\mathbf{s}_{i'}, t) = f^{(k)}(t)$ for all t. Otherwise, at least one of $\mu(\mathbf{s}_i) \neq \mu(\mathbf{s}_{i'})$ and $f(\mathbf{s}_i, t) \neq f(\mathbf{s}_{i'}, t)$ holds for two locations $\mathbf{s}_i$ and $\mathbf{s}_{i'}$ from different clusters. Note that, for the kth cluster C_k, if a location $\mathbf{s}_i$ has no temporal trend, then according to our definition $f^{(k)}(t) = f(\mathbf{s}_i, t) = 0$ for all t.

To approximate the unknown smooth function $f(\mathbf{s}_i, t)$, we leverage the standardized B-spline approach (Liu & Yang, 2010; Ma & Yang, 2011; Xue & Yang, 2006). Specifically, denote $\boldsymbol{\kappa} = \{0 = \kappa_0 < \kappa_1 < \cdots < \kappa_L < \kappa_{L+1} = 1\}$ as a partition of the interval $\mathcal{T} = [0, T]$, where L is the number of interior knots. Let $\{B_j(t)\}_{j=1}^{L+d}$ be a set of standardized B-spline basis functions of degree d with interior knots $\boldsymbol{\kappa}$ on $\mathcal{T}$ such that $E(B_j(t)) = 0$ and $E(B_j^2(t)) = 1$. With $J = L + d$ be the number of B-spline basis functions, the trend function in the kth cluster $f^{(k)}(t)$ can be approximated as

$$f^{(k)}(t) \approx \sum_{j=1}^{J} B_j(t)\beta_j^{(k)}(\mathbf{s}) = \boldsymbol{B}^{\top}(t)\boldsymbol{\beta}^{(k)}, \tag{4}$$

where $\boldsymbol{B}^{\top}(t) = (B_1(t), \ldots, B_J(t))^{\top}$ is the spline basis matrix and $\boldsymbol{\beta}^{(k)}(\mathbf{s}) = (\beta_1(\mathbf{s}), \ldots, \beta_J(\mathbf{s}))^{\top}$ is the coefficient vector. Then, we can formulate our model in a spatial-temporal semiparametric form

$$y(\mathbf{s}_i, t_l) = \begin{cases} \mu^{(k)} + \epsilon(\mathbf{s}_i, t_l), & \mathbf{s}_i \in C_k \text{ and } C_k \subset \mathcal{S}_f, \\ \mu^{(k)} + \boldsymbol{B}^{\top}(t_l)\boldsymbol{\beta}^{(k)} + \epsilon(\mathbf{s}_i, t_l), & \mathbf{s}_i \in C_k \text{ and } C_k \subset \mathcal{N}_f. \end{cases} \tag{5}$$

We name our model as spatial clustered and sparse nonparametric regression (Scanner).

2.2 | Penalized estimation

When the partitions of $\{C_k\}_{k=1}^K$ and $\{S_f, \mathcal{N}_f\}$ are known, one can easily estimate the model by solving a least square problem:

$$\min \frac{1}{2T}\sum_{l=1}^{T}\left(\sum_{C_k\subset\mathcal{N}_f}\sum_{i\in C_k}\left(y(\mathbf{s}_i,t_l)-\mu^{(k)}\right)^2+\sum_{C_k\subset S_f}\sum_{i\in C_k}\left(y(\mathbf{s}_i,t_l)-\mu^{(k)}-\boldsymbol{B}^\top(t_l)\boldsymbol{\beta}^{(k)}\right)^2\right). \tag{6}$$

However, in our problem, the partitions remain unknown and need to be recovered with the observed data. To address this issue, we introduce pseudo local coefficients at $\mathbf{s}_i$ as $\boldsymbol{\alpha}(\mathbf{s}_i)=[\mu(\mathbf{s}_i),\boldsymbol{\beta}(\mathbf{s}_i)^\top]^\top$. Following the model (5), our goal is equivalent to find the set $S_f=\{i,1\le i\le n;\|\boldsymbol{\beta}(\mathbf{s}_i)\|\neq 0\}$ and the cluster sets $\{C_k\}_{k=1}^K$ such that $\boldsymbol{\alpha}(\mathbf{s}_i)=\boldsymbol{\alpha}(\mathbf{s}_{i'})$ if $\mathbf{s}_i$ and $\mathbf{s}_{i'}\in C_k$, where $\|\cdot\|$ is the Euclidean norm.

For the first task of identifying the sparsity of $\{\boldsymbol{\beta}(\mathbf{s}_i)\}_{i=1}^n$, we adopt the group sparsity penalty $\sum_{i=1}^n\mathcal{P}(\|\boldsymbol{\beta}(\mathbf{s}_i)\|;\gamma_1,\lambda_{1i})$ based on the minimax concave penalty (MCP) (Zhang, 2010),

$$\mathcal{P}(t;\gamma,\lambda)=\begin{cases}\lambda|t|-\frac{t^2}{2\gamma}, & |t|\le\gamma\lambda,\\ \frac{1}{2}\gamma\lambda^2, & |t|>\gamma\lambda,\end{cases} \tag{7}$$

where γ_1 and λ_{1i} are two hyper-parameters. If a location $\mathbf{s}_i$ has an insignificant temporal trend, the estimated local spline coefficients $\boldsymbol{\beta}(\mathbf{s}_i)$ will shrink to zero, which recovers the model $y(\mathbf{s}_i,t)=\mu(\mathbf{s}_i)+\epsilon(\mathbf{s}_i,t)$.

In the second task of recovering spatially contiguous clusters, one can represent the locations' spatial relationship with a graph topology and construct the graph-based fusion penalty (Hallac et al., 2015):

$$\sum_{(\mathbf{s}_i,\mathbf{s}_{i'})\in\mathcal{G}}\mathcal{P}(\|\boldsymbol{\alpha}(\mathbf{s}_i)-\boldsymbol{\alpha}(\mathbf{s}_{i'})\|;\gamma_2,\lambda_2), \tag{8}$$

where $\mathcal{G}$ is an undirected graph connecting all locations $\{\mathbf{s}_i,i=1,\ldots,n\}$ and γ_2, λ_2 are hyper-parameters. In the above penalties, $\|\boldsymbol{\alpha}(\mathbf{s}_i)-\boldsymbol{\alpha}(\mathbf{s}_{i'})\|$ will shrink to zero if two locations, $\mathbf{s}_i$ and $\mathbf{s}_{i'}$, are connected in the graph and have similar coefficients. Thus, the spatial cluster pattern can be formed with the same estimated coefficients $\boldsymbol{\alpha}$'s.

In spatial data, it is reasonable to assume that locations closer to each other are similar, and a minimum spanning tree (MST) tends to connect closer locations. Graph-based and MST-based approaches are used in different works, such as Gower and Ross (1969), Eldershaw and Hegland (1997), and Grygorash et al. (2006). As discussed in Li and Sang (2019), the naive graph-based fusion penalty would introduce $O(|\mathcal{G}|)$ complexity, which might cause intensive computation with large n. They proposed using MST based on spatial locations only. As discussed by Li and Sang (2019), the constructed MST may not be fully compatible with the true cluster structure when using spatial location information only, which means that the constructed MST may not connect the locations in the same cluster properly. Lin et al. (2022) also discussed this issue for MST based on location information for the varying coefficient model. They discussed the properties of the cluster structure based on the constructed MST, which may be different from the true cluster structure. Zhang et al. (2024) improved the approach using local information to construct MSTs. They showed that by increasing the local sample size, the true cluster structure can be recovered based on the constructed MST based on spatial and local information. Wang, Zhang, and Zhu (2023) also used local information to construct MSTs and identify cluster structures for recurrent event data. Inspired by Zhang et al. (2024), we consider simplifying the spatial graph with a model-based minimum spanning tree (MST) in this work. Specifically, we do a local spline fitting at each location and obtain the location coefficient estimation $\hat{\alpha}_{init}(\mathbf{s}_i)$. Then, with the given spatial graph, we define the weight between two locations as

$$w_{ii'}=\begin{cases}\|\hat{\boldsymbol{\alpha}}_{init}(\mathbf{s}_i)-\hat{\boldsymbol{\alpha}}_{init}(\mathbf{s}_{i'})\|, & \text{if } (\mathbf{s}_i,\mathbf{s}_{i'})\in\mathcal{G}.\\ \infty, & \text{otherwise.}\end{cases} \tag{9}$$

We construct the MST, $\mathcal{T}$, based on the above weight and only penalize two locations if they are connected in the MST. And the size of $\mathcal{T}$ is $n-1$.

Thus, we propose the following doubly penalized least square problem,

$$\min \frac{1}{2T}\sum_{i=1}^{n}\sum_{l=1}^{T}\left(y(\mathbf{s}_i,t_l)-\left(\mu(\mathbf{s}_i)+\boldsymbol{B}^{\top}(t_l)\boldsymbol{\beta}(\mathbf{s}_i)\right)\right)^2 + \sum_{i=1}^{n}\mathcal{P}(\|\boldsymbol{\beta}(\mathbf{s}_i)\|;\gamma_1,\lambda_{1i}) + \sum_{(\mathbf{s}_i,\mathbf{s}_{i'})\in\mathcal{T}}\mathcal{P}(\|\boldsymbol{\alpha}(\mathbf{s}_i)-\boldsymbol{\alpha}(\mathbf{s}_{i'})\|;\gamma_2,\lambda_2). \tag{10}$$

In (10), the first term is a least square loss for model fitting, the second term is the group penalty to identify the nonzero vector $\boldsymbol{\beta}(\mathbf{s}_i)\neq\mathbf{0}$, and the third term is the tree-based fusion penalty to cluster the coefficients $\{\boldsymbol{\alpha}(\mathbf{s}_i), i=1,\ldots,n\}$.

2.3 | Optimization

For the notation simplicity, we denote $\mu(\mathbf{s}_i)=\mu_i$, $\boldsymbol{\beta}(\mathbf{s}_i)=\boldsymbol{\beta}_i$, $\boldsymbol{\alpha}(\mathbf{s}_i)=\boldsymbol{\alpha}_i$ in this section. We rewrite the problem in (10) in a matrix form:

$$\min \frac{1}{2}\sum_{i=1}^{n}\frac{1}{T}\|\boldsymbol{y}_i-\boldsymbol{X}\boldsymbol{\alpha}_i\|^2+\sum_{i=1}^{n}\mathcal{P}(\|\boldsymbol{\beta}_i\|;\gamma_1,\lambda_{1i})+\sum_{(\mathbf{s}_i,\mathbf{s}_{i'})\in\mathcal{T}}\mathcal{P}(\|\boldsymbol{\alpha}_i-\boldsymbol{\alpha}_{i'}\|;\gamma_2,\lambda_2), \tag{11}$$

where $\boldsymbol{X}=(1,\boldsymbol{B})$ with $\boldsymbol{B}=(\boldsymbol{B}(t_1),\ldots,\boldsymbol{B}(t_T))^{\top}$. To solve the above penalized minimization problem, we adopt the ADMM algorithm (Boyd et al., 2011).

First, we introduce $\boldsymbol{\delta}_{ii'}=\boldsymbol{\alpha}_i-\boldsymbol{\alpha}_{i'}$ for $(\mathbf{s}_i,\mathbf{s}_{i'})\in\mathcal{T}$ and $\boldsymbol{\psi}_i=\boldsymbol{\beta}_i$ for $i=1,\ldots,n$. Denote $\boldsymbol{\alpha}=(\boldsymbol{\alpha}_1^{\top},\ldots,\boldsymbol{\alpha}_n^{\top})^{\top}$, $\boldsymbol{\psi}=(\boldsymbol{\psi}_1^{\top},\cdots\boldsymbol{\psi}_n^{\top})^{\top}$, and $\boldsymbol{\delta}=(\boldsymbol{\delta}_{ii'}^{\top},(\mathbf{s}_i,\mathbf{s}_{i'})\in\mathcal{T})$, then the above problem is equivalent to the following constrained optimization problem with regard to $(\boldsymbol{\alpha},\boldsymbol{\psi},\boldsymbol{\delta})$,

$$\begin{aligned} &L_0(\boldsymbol{\alpha},\boldsymbol{\psi},\boldsymbol{\delta})=\frac{1}{2}\sum_{i=1}^{n}\frac{1}{T}\|\boldsymbol{y}_i-\boldsymbol{X}\boldsymbol{\alpha}_i\|^2+\sum_{i=1}^{n}\mathcal{P}(\|\boldsymbol{\psi}_i\|;\gamma_1,\lambda_{1i})+\sum_{(\mathbf{s}_i,\mathbf{s}_{i'})\in\mathcal{T}}\mathcal{P}(\|\boldsymbol{\delta}_{ii'}\|;\gamma_2,\lambda_2),\\ &\text{s.t.}\quad \boldsymbol{\alpha}_i-\boldsymbol{\alpha}_{i'}=\boldsymbol{\delta}_{ii'}\ \text{and}\ \boldsymbol{\beta}_i=\boldsymbol{\psi}_i. \end{aligned} \tag{12}$$

Let $\boldsymbol{\theta}=(\boldsymbol{\delta}^{\top},\boldsymbol{\psi}^{\top})^{\top}$, $\boldsymbol{y}=(\boldsymbol{y}_1^{\top},\ldots,\boldsymbol{y}_n^{\top})^{\top}$, and $\widetilde{\boldsymbol{X}}=\mathbf{I}_n\otimes\boldsymbol{X}$, the matrix version of the above problem can be written as,

$$\begin{aligned} &L_0(\boldsymbol{\alpha},\boldsymbol{\theta})=\frac{1}{2T}\|\boldsymbol{y}-\widetilde{\boldsymbol{X}}\boldsymbol{\alpha}\|^2+\sum_{i=1}^{n}\mathcal{P}(\|\boldsymbol{\psi}_i\|;\gamma_1,\lambda_{1i})+\sum_{(\mathbf{s}_i,\mathbf{s}_{i'})\in\mathcal{T}}\mathcal{P}(\|\boldsymbol{\delta}_{ii'}\|;\gamma_2,\lambda_2),\\ &\text{s.t.}\quad \mathbf{K}\boldsymbol{\alpha}=\boldsymbol{\theta}, \end{aligned} \tag{13}$$

where $\mathbf{K}=((\mathbf{H}\otimes\mathbf{I}_{J+1})^{\top},\mathbf{M}^{\top})^{\top}$. Here, $\mathbf{H}$ is the incident matrix of the tree such that $\boldsymbol{\delta}=(\mathbf{H}\otimes\mathbf{I}_{J+1})\boldsymbol{\alpha}$ (Li & Sang, 2019; Zhang et al., 2024). $\mathbf{M}=\boldsymbol{I}_n\otimes(\mathbf{0},\boldsymbol{I}_J)$ is the selection matrix such that $\boldsymbol{\psi}=\mathbf{M}\boldsymbol{\alpha}$.

Then, we give the augmented Lagrangian as

$$L(\boldsymbol{\alpha},\boldsymbol{\theta},\boldsymbol{\nu})=L_0(\boldsymbol{\alpha},\boldsymbol{\theta})+\langle\boldsymbol{\nu},\mathbf{K}\boldsymbol{\alpha}-\boldsymbol{\theta}\rangle+\frac{\rho}{2}\|\mathbf{K}\boldsymbol{\alpha}-\boldsymbol{\theta}\|^2, \tag{14}$$

where $\boldsymbol{\nu}$ are Lagrange multipliers and $\rho>0$ is the penalty parameter. Here, we fix $\rho=1$ as in the references (Liu et al., 2023; Ma & Huang, 2017). The minimization problem (13) is equivalent to finding the Karush–Kuhn–Tucker (KKT) point of (14) by iteratively updating as

$$\boldsymbol{\alpha}^{(m+1)}=\underset{\boldsymbol{\alpha}}{\operatorname{argmin}}\ L(\boldsymbol{\alpha},\boldsymbol{\theta}^{(m)},\boldsymbol{\nu}^{(m)}), \tag{15}$$

$$\boldsymbol{\theta}^{(m+1)}=\underset{\boldsymbol{\theta}}{\operatorname{argmin}}\ L(\boldsymbol{\alpha}^{(m+1)},\boldsymbol{\theta},\boldsymbol{\nu}^{(m)}), \tag{16}$$

$$\boldsymbol{\nu}^{(m+1)}=\boldsymbol{\nu}^{(m)}+\rho(\mathbf{K}\boldsymbol{\alpha}^{(m+1)}-\boldsymbol{\theta}^{(m+1)}). \tag{17}$$

Algorithm 1. The ADMM algorithm

Input: : Initialize $\boldsymbol{\alpha}^{(0)}$, $\boldsymbol{\delta}^{(0)}$ and $\boldsymbol{\nu}^{(0)}$.

for $m = 0, 1, 2, \ldots$ **do**

$\boldsymbol{\alpha}^{(m+1)} = (\widetilde{\boldsymbol{X}}^\top \widetilde{\boldsymbol{X}} + \rho T \mathbf{K}^\top \mathbf{K})^{-1} (\widetilde{\boldsymbol{X}}^\top \boldsymbol{y} - T\mathbf{K}^\top \boldsymbol{\nu}^{(m)} + \rho T \mathbf{K}^\top \boldsymbol{\theta}^{(m)})$.

$\tilde{\boldsymbol{\theta}}^{(m+1)} = \begin{bmatrix} \tilde{\boldsymbol{\delta}}^{(m+1)} \\ \tilde{\boldsymbol{\psi}}^{(m+1)} \end{bmatrix} = \mathbf{K}\boldsymbol{\alpha}^{(m+1)} + \frac{1}{\rho}\boldsymbol{\nu}^{(m)}$.

$\boldsymbol{\psi}_i^{(m+1)} = \begin{cases} \frac{\mathrm{ST}(\tilde{\boldsymbol{\psi}}^{(m+1)}, \lambda_{1i}/\rho)}{1 - 1/(\gamma_1 \rho)} & \text{if } \|\tilde{\boldsymbol{\psi}}_i^{(m+1)}\| \le \gamma_1 \lambda_{1i} \\ \tilde{\boldsymbol{\psi}}_i^{(m+1)} & \text{o.w.} \end{cases}$

$\boldsymbol{\delta}_{ii'}^{(m+1)} = \begin{cases} \frac{\mathrm{ST}\left(\tilde{\boldsymbol{\delta}}_{il}^{(m+1)}, \lambda_2/\rho\right)}{1 - 1/(\gamma_2 \rho)} & \text{if } \|\tilde{\boldsymbol{\delta}}_{ii'}^{(m+1)}\| \le \gamma_2 \lambda_2 \\ \tilde{\boldsymbol{\delta}}_{ii'}^{(m+1)} & \text{o.w.} \end{cases}$

$\boldsymbol{\nu}^{(m+1)} = \boldsymbol{\nu}^{(m)} + \rho(\mathbf{K}\boldsymbol{\alpha}^{(m+1)} - \boldsymbol{\theta}^{(m+1)})$.

if the convergence criterion is met **then**

Stop

end if

end for

We relegate the detailed updates and derivations to the Supporting Information and summarize the algorithm in Algorithm 1, in which $\mathrm{ST}(\boldsymbol{\omega}, t) = (1 - t/\|\boldsymbol{\omega}\|)_+ \boldsymbol{\omega}$ is the soft thresholding operation, and $(t)_+ = t$ if $t > 0$ and 0 otherwise.

Remark 1. The convergence criterion is based on the primal residual $\mathbf{K}\boldsymbol{\alpha}^{(m+1)} - \boldsymbol{\theta}^{(m+1)}$. The algorithm is stopped if $\|\mathbf{K}\boldsymbol{\alpha}^{(m+1)} - \boldsymbol{\theta}^{(m+1)}\| < \omega$, where ω is a small positive value. Here we use $\omega = 0.001$ (Boyd et al., 2011).

Remark 2. Our ADMM-based algorithm can also be used in the case of graph-based fusion penalty in (8). We can achieve this by replacing the incident matrix of the tree with that of the graph in $\mathbf{K}$ in (13). Then, the rest of the algorithmic updates will remain the same.

2.4 | Hyper-parameter selection

The hyper-parameters play an important role in our estimation. λ_{1i}'s control the sparsity of temporal trends, and λ_2 controls the homogeneity of the locations. A grid search for these hyper-parameters will lead to a high computation cost. To select proper hyper-parameters, we use a two-step procedure similar to Tang and Li (2023). First, we set $\lambda_2 = 0$ and select λ_{1i} for each location $i = 1, \ldots, n$. Then, with the selected $\{\lambda_{1i}\}_i$, we choose the optimal λ_2. We propose to use the Bayesian information criterion to select hyper-parameters in different steps. In the first step, the BIC is defined as below:

$$BIC_{\lambda_{1i}} = \log\left(\frac{1}{T}\|\boldsymbol{y}_i - \hat{\mu}_i - \boldsymbol{B}^\top \hat{\boldsymbol{\beta}}_i\|^2\right) + d_1 \frac{\log(T/J)}{T/J},$$

where $\hat{\mu}_i$ and $\hat{\boldsymbol{\beta}}_i$ are the estimates for μ_i and $\boldsymbol{\beta}_i$, respectively, and d_1 is 1 if $\hat{\boldsymbol{\beta}}_i \neq \mathbf{0}$ while 0 otherwise. In the second step, we consider the modified BIC (Ma et al., 2020),

$$BIC_{\lambda_2} = \log\left(\frac{1}{N}\|\boldsymbol{y} - \widetilde{\boldsymbol{X}}\hat{\boldsymbol{\alpha}}\|^2\right) + \log(nJ)\left(\frac{\log N}{N}\hat{K} + \frac{\log(N/J)}{N/J} d_2\right),$$

where $N = nT$, $\hat{K}$ is the number of estimated clusters, and d_2 is the number of unique estimated temporal trends. Following the references (Liu et al., 2023; Ma & Huang, 2017; Yang et al., 2019), we use fixed values of $\gamma_1 = 3$ and $\gamma_2 = 3$. The proposed two-step procedure is used in the simulation studies in Section 4 and the real data example in Section 5. In the simulation study, we evaluate the model and the procedure in different scenarios, and the proposed method can identify both temporal trends and cluster patterns well.

In the Supporting Material, we provide the code files and datasets for reproducing the results of our simulation studies and the real data analysis. Please refer to the attached README file for the details of each file.

3 | THEORETICAL PROPERTIES

In this section, the theoretical properties of the proposed estimators are studied. We first introduce some notations. Let $\mu^{(k)}$ and $f^{(k)}, k = 1, \dots, K$, be the common intercept and the temporal function for cluster k, respectively. Define $\|f(t)\|_2^2 = \int_{\mathcal{T}} f^2 dt$ as the squared L_2 norm for any square integrable function $f(t)$ on $\mathcal{T}$. For a vector $\boldsymbol{x}$, define $\|\boldsymbol{x}\|$ as the Euclidean norm. $a_T \gg b_T$ means that $a_T^{-1} b_T = o(1)$. For simplicity, we write $f_i = f_i(t) = f(\mathbf{s}_i, t)$, $\epsilon_i(t) = \epsilon(\mathbf{s}_i, t)$, and $\|f\| = \|f_i(t)\|_2$.

Let $C(\mathcal{T})$ be the class of continuous functions on $\mathcal{T}$ and $C^{(d)} = \{g(t) | g^d(t) \in C(\mathcal{T})\}$. We have the following assumptions.

(C1) The observation time points $\{t_1, \dots, t_T\}$ are chosen independently from an unknown distribution with a density $f_T(\cdot)$, which is bounded away from 0 and infinity over $\mathcal{T} = [0, 1]$.
(C2) There is a positive constant M_1 such that $E[\epsilon_i(t)^4] \le M_1 < \infty$ for $t \in \mathcal{T}$. And the random sequence $\epsilon_i(t)$ satisfies the α-mixing condition, where the α-mixing coefficient satisfying $\alpha(s) \le M_2 \rho^s$ for $0 < \rho < 1$ and M_2 is a positive constant with $0 < M_2 < \infty$.
(C3) For each $f_i, f_i \in C^{(d)}$ is a dth order continually differentiable function and $E(f_i) = 0$.
(C4) There exists a constant $c_f > 0$ such that $\min_{i \in S_f} \|f_i\| \ge c_f$.
(C5) For any two locations $\mathbf{s}_i$ and $\mathbf{s}_{i'}$ in the given connected network $\mathcal{G}$, if they are from the same cluster, then there exists a path connecting them such that all locations on the path belong to the same cluster.

Condition (C1) is commonly used in longitudinal data analysis (Huang et al., 2004; Wang et al., 2008). The α-mixing condition is generally used in longitudinal data analysis when independent error processes are not assumed; see Xue and Yang (2006) and Liu et al. (2023). (C3) is used in Liu et al. (2023) with an additional zero expectation assumption. This zero expectation assumption is a standard assumption for identifiability used in the literature, such as Xue and Yang (2006), Liu and Yang (2010) and Li et al. (2019). (C4) is a common assumption in nonparametric additive models (Huang et al., 2010; Li et al., 2019). (C5) is used in Zhang et al. (2024) to guarantee that locations in the same cluster will not be separated by other clusters. By removing inner-cluster connections, the original graph can be reduced to K subgraphs, which are the K clusters.

First, we have the following Lemma 1 to guarantee that the locations in the same cluster are connected in the MST constructed based on weights defined in (9). This Lemma is also discussed in Zhang et al. (2024). The difference is that we have B-spline regression estimators instead of OLS estimators.

Lemma 1. *Given the MST based on the weights in (9), consider any two locations $\mathbf{s}_i$ and $\mathbf{s}_{i'}$ in the same cluster. Under Condition (C5), there exists a path in the MST connecting $\mathbf{s}_i$ and $\mathbf{s}_{i'}$ such that all the locations on the path belong to the same cluster with probability approaching 1 as local sample size $T \to \infty$.*

Next, we will prove the theoretical properties of the Scanner estimator when the cluster structure $\{C\}_{k=1}^K$ is known. Let $\mathcal{S} = \{k; f^{(k)} \neq 0, 1 \le k \le K\}$ and $\mathcal{N} = \{k; f^{(k)} = 0, 1 \le k \le K\}$, which are the sets of clusters with nonzero temporal trends and zero temporal trends, respectively. When the cluster structure is known, define $\boldsymbol{W}$ as an $n \times K$ matrix such that $w_{ik} = 1$ if location i is in cluster k, 0 otherwise. Let $\widetilde{\boldsymbol{W}} = \boldsymbol{W} \otimes \boldsymbol{I}_{J+1}$, and $\boldsymbol{U} = \widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{W}} = \boldsymbol{W} \otimes \boldsymbol{X}$. We write the district cluster parameters vector as $\boldsymbol{\phi} = \left(\boldsymbol{\phi}_1^\top, \boldsymbol{\phi}_2^\top, \dots, \boldsymbol{\phi}_K^\top\right)^\top$, where $\boldsymbol{\phi}_k = \left(\mu^{(k)}, \boldsymbol{\phi}_{2k}^\top\right)^\top$. Here, $\boldsymbol{\phi}_k$'s are distinct vectors of $\boldsymbol{\alpha}_i$'s, $\mu^{(k)}$'s are the distinct intercepts of of μ_i's, and $\boldsymbol{\phi}_{2k}$'s are the distinct vectors of $\boldsymbol{\beta}_i$'s. When the cluster structure is known, the corresponding estimator is defined below,

$$\hat{\boldsymbol{\phi}} = \arg\min_{\boldsymbol{\phi}} \frac{1}{2T} \|\boldsymbol{y} - \boldsymbol{U}\boldsymbol{\phi}\|^2 + \sum_{k=1}^{K} \sum_{i \in C_k} \mathcal{P}\left(\|\boldsymbol{\phi}_{2k}\|; \gamma_1, \lambda_{1i}\right), \tag{18}$$

where the objective function in this optimization problem is a special case of (11) when the cluster structure is known.

Let $\tilde{f}^{(k)}(t) = \boldsymbol{B}^\top(t)\boldsymbol{\phi}_{2k}$, which is the temporal function in cluster k when the cluster structure is known. And the estimator based on $\hat{\boldsymbol{\phi}}$ is $\hat{f}^{(k)} = \boldsymbol{B}^\top(t)\hat{\boldsymbol{\phi}}_{2k}$. Define $r_T = \left(\frac{1}{T} + \frac{J}{T}\right)^{1/2}$, $\lambda_{\min} = \min_i \lambda_{1i}$, and $\lambda_{\max} = \max_i \lambda_{1i}$. Theorem 1 shows the asymptotic properties of $\hat{\mu}^{(k)}$ and $\hat{f}^{(k)}$ for $k = 1, \dots, K$.

Theorem 1. *Suppose that conditions (C1)–(C4) hold, n is fixed and $J = O(T^{\kappa})$ with $0 < \kappa < 0.5$, then as $T \to \infty$, $\lambda_{\max} \to 0$ and $\frac{\lambda_{\min}}{r_T + J^{-d}} \to \infty$, we have the following results*:

1. $\hat{f}^{(k)} = 0$ *for* $k \in \mathcal{N}$.
2. $|\hat{\mu}^{(k)} - \mu^{(k)}| = O_p(r_T + J^{-d})$ *for* $k = 1, \dots, K$ *and* $\|\hat{f}^{(k)} - f^{(k)}\| = O_p(r_T + J^{-d})$ *for* $k \in \mathcal{S}$.

Let $\tilde{\boldsymbol{\alpha}} = (\tilde{\boldsymbol{\alpha}}_1, \dots, \tilde{\boldsymbol{\alpha}}_n)^\top = \widetilde{\boldsymbol{W}}\hat{\boldsymbol{\phi}}$. Then, $\tilde{\boldsymbol{\alpha}}_i = \hat{\boldsymbol{\phi}}_k$ if $i \in \mathcal{C}_k$ with $\tilde{\boldsymbol{\alpha}}_i = (\tilde{\mu}_i, \tilde{\boldsymbol{\beta}}_i^\top)^\top$. Denote $\tilde{f}_i = \boldsymbol{B}^\top(t)\tilde{\boldsymbol{\beta}}_i$. Based on these definitions, we know that $\tilde{\mu}_i = \hat{\mu}^{(k)}$ and $\tilde{f}_i = \hat{f}^{(k)}$ if $i \in \mathcal{C}_k$. From the results in Theorem 1, we have the theoretical properties of $\tilde{\mu}_i$ and $\tilde{f}_i$. Let $\hat{\boldsymbol{\alpha}} = (\hat{\boldsymbol{\alpha}}_1^\top, \dots, \hat{\boldsymbol{\alpha}}_n^\top)^\top$ be the proposed estimator in (11) with $\hat{\boldsymbol{\alpha}}_i^\top = (\hat{\mu}_i, \hat{\boldsymbol{\beta}}_i^\top)$. Then, the estimated function is $\hat{f}_i = \boldsymbol{B}^\top(t)\hat{\boldsymbol{\beta}}_i$. Next, we will study the relationship between $\hat{f}_i$ and $\tilde{f}_i$, $\hat{\mu}_i$ and $\tilde{\mu}_i$. Define the minimum cluster difference $b = \min_{k \neq k'} \left[|\mu^{(k)} - \mu^{(k')}| + \|f^{(k)} - f^{(k')}\|\right]$.

Theorem 2. *Suppose that conditions (C1)–(C4) and the conditions in Theorem 1 hold, $b > \gamma_2\lambda_2$ and $\lambda_2 \gg r_T + J^{-d}$, then*

$$|\hat{\mu}_i - \tilde{\mu}_i| = O_p(r_T + J^{-d}) \text{ and } \|\hat{f}_i - \tilde{f}_i\| = O_p(r_T + J^{-d}).$$

Theorem 2 guarantees that the estimated intercept and estimated temporal functions converge to those when assuming the cluster structure is known. We have the following corollary from Theorem 1 and Theorem 2 directly.

Corollary 1. *Suppose the conditions in Theorem 2 hold, as $T \to \infty$, we have the following results*:

1. $\hat{f}_i = 0$ *for* $i \in \mathcal{N}_f$.
2. $|\hat{\mu}_i - \mu_i| = O_p(r_T + J^{-d})$ *for* $i = 1, \dots, n$ *and* $\|\hat{f}_i - f_i\| = O_p(r_T + J^{-d})$ *for* $i \in \mathcal{S}_f$.

Part 1 of Corollary 1 says that the proposed method is consistent in temporal trend selection; that is, it can identify the zero temporal trends with probability approaching 1. Part 2 provides the rate of convergence in estimating the nonzero temporal trend functions and intercepts.

The following two corollaries give us the results of recovering the cluster structure.

Corollary 2. *Suppose (C5) and all conditions in Theorem 2 hold, then the cluster structure can be recovered with probability approaching 1.*

Corollary 3. *The result in Corollary 2 holds for any graph satisfying condition (C5).*

The result of Corollary 2 depends on Lemma 1. As Zhang et al. (2024) discussed, the number of connections between clusters is $K - 1$ under Lemma 1. By removing connections between clusters, that is, identifying nonzero coefficient differences between two locations $\mathbf{s}_i$ and $\mathbf{s}_{i'}$, the original graph can be separated into K subgraphs, which correspond to K clusters. For any graphs satisfying condition (C5), we can also have K subgraphs by removing any connections between different clusters.

4 | EXPERIMENTAL STUDY

In this section, we empirically examine the performance of our proposed Scanner method. In Section 4.1, we first study the impact of the number of locations n and time length T on our estimator and compare the performance to existing competitive approaches. In Section 4.2, we evaluate and compare algorithm complexities. Another comparison with more groups is provided in Section 4.3.

In the numerical studies, the locations are uniformly located in the space $[-1, 1]^2$. The graph $\mathcal{G}$ is construed based on Delaunay triangulation (Lee & Schachter, 1980). Then, an MST can be construed based on graph $\mathcal{G}$ and the weights provided in (9). We mainly compare the following three penalization methods. The first method only considers the graph fusion penalty in the objective function, which will be the function in (11) without the sparsity penalty. This method is denoted as "SCN." The second method considers the sparsity penalty and the graph fusion penalty based on $\mathcal{G}$. This method is denoted as "Scanner.GF." The third method has the sparsity penalty and the MST fusion penalty based on

$\mathcal{T}$, denoted as "Scanner.MST." These three methods are based on our proposed algorithm and implemented in R. All simulations are implemented in server Mesxuuyan with 16-node Dual-10 Xeon CPU (2.20GHz) 12.8GB RAM per core in Computational Science Research Center at San Diego State University.

We use several metrics to evaluate the performance of the methods. First, we report the average root mean square error (ARMSE) to evaluate the model estimation performance. ARMSE is defined as $ARMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|\hat{g}_i - g_i\|^2}$, where $\|\hat{g}_i - g_i\|^2$ is approximated by 1000 values evaluated over $(0, 1)$. To compare the performance of recovering cluster, we use the estimated number of clusters ($\hat{K}$) and adjusted Rand Index (ARI) (Hubert & Arabie, 1985). ARI is widely used to measure the agreement of two partitions. The largest value is 1, and the larger, the better. As for the temporal trend detection, we report the F1 score (Sasaki, 2007). We can treat temporal trend detection as a 0-1 classification problem. F1 score measures the accuracy of classification with the highest value of 1 and the smallest value of 0. The higher the value of F1 score, the better accuracy of temporal trend detection is. The F1 score is defined as $F1 = \frac{2PR}{P+R}$, where P is precision and R is recall.

4.1 | Study on location number n and time length T

In this part, we study the performances of our proposed methods on different numbers of locations (n) and time length T. The performances are compared through the estimation performance (ARMSE), clustering performance ($\hat{K}$, ARI), and zero trend identifications (number of identified zero trends, F1 score). We consider several traditional methods besides the three penalization methods. "Global" is the B-spline regression model when all locations are pooled together, which ignores both temporal trends structure and cluster structure. "Local" refers to the method of fitting an individual B-spline regression model for each location, which ignores the spatial information for different locations. "Local.k" uses k-means to find clusters based on local B-spline regression coefficients when the number of clusters is known as the true value. "Local.gap" uses k-means based on B-spline regression coefficients when the number of clusters is selected based on Gap statistic (Dudoit & Fridlyand, 2002; Tibshirani et al., 2001). "k-means" refers to the method of using k-means to find the clusters based on the original observations when the number of clusters is fixed at the true value. Similarly, "k-means.gap" is the method based on Gap statistic. Gap statistic is implemented by R package *cluster*. These four clustering approaches do not consider spatial information. We also considered two spatial clustering methods, "ClustGeo" and "skater." "ClustGeo" (Chavent et al., 2018) is a Ward-like hierarchical clustering algorithm that uses dissimilarities in the feature space and the spatial space. Here, we use the Euclidean distance of observations as the dissimilarities in the feature space and the Euclidean distance of locations as the dissimilarities in the spatial space. R package *ClustGeo* is used. "skater" is a spatial clustering method based on graph connections. Two graph connections are 10 nearest neighbors and Delaunay triangulation, "skater.knn" and "skater.tri" are used to denote these two methods, respectively. R package "spdep" is used to implement "skater" algorithms. However, these two approaches cannot identify zero temporal trends.

We consider five underlying clusters separated by solid lines in Figure 3. Assume that $g(\mathbf{s}_i, t) = 1$ if $\mathbf{s}_i \in C_1$, $g(\mathbf{s}_i, t) = 2$ if $\mathbf{s}_i \in C_2$, $g(\mathbf{s}_i, t) = 1 - \exp(1.5t)$ if $\mathbf{s}_i \in C_3$, $g(\mathbf{s}_i, t) = \sin(2\pi t)$ if $\mathbf{s}_i \in C_4$ and $g(\mathbf{s}_i, t) = 1.5t^3$ if $\mathbf{s}_i \in C_5$. Note that the first two clusters have no temporal trend, while the other three clusters have temporal trends. Under different cluster structures, the numbers of nonzero temporal trends are 52, 101, and 255 for $n = 100$,200, and 500, respectively. The random noises are from a normal distribution with zero mean and variance $\sigma^2 = 0.4^2$. The number of locations n takes 100,200, and 500. The time stamp t_l for $l = 1, \ldots, T$ are simulated from Uniform distribution $(0, 1)$ with $T = 20$ or 50. We consider different combinations of n and T. Cubic B-spline is used, and the number of basis functions is 5 for $T = 20$ and 6 when $T = 50$.

Tables 1–3 show the average results of 200 simulations across $n = 100$,200, 500 and $T = 20, 50$. Standard deviations across 200 simulations are presented in subscripts. We observe that "Scanner.MST" outperforms other methods in terms of estimating mean functions (small ARMSE), estimating the number of clusters ($\hat{K}$), recovering spatial cluster structures (large ARI), and detecting nonzero temporal trends (large F1 score). "Global" and "Local" do not consider the sparsity and cluster structure of the data. Thus, they have large ARMSE. "Local.k", "Local.gap", "k-means" and "k-means.gap" can recover the cluster structure. But they cannot recover the structure well, which can be seen from ARI, even when the number of clusters is fixed at the true value. This is because these methods do not take time order into consideration and ignore the spatial information in the data even when there are 50 observations locally. Among the approaches with consideration of spatial information ("ClustGeo", "skater.knn" and "skater.tri"), "ClustGeo" can achieve better results in terms of recovering cluster structure compared to other traditional methods. These indicate the importance of spatial information in finding cluster structure. "SCN" uses spatial information in the model, which can recover the spatial

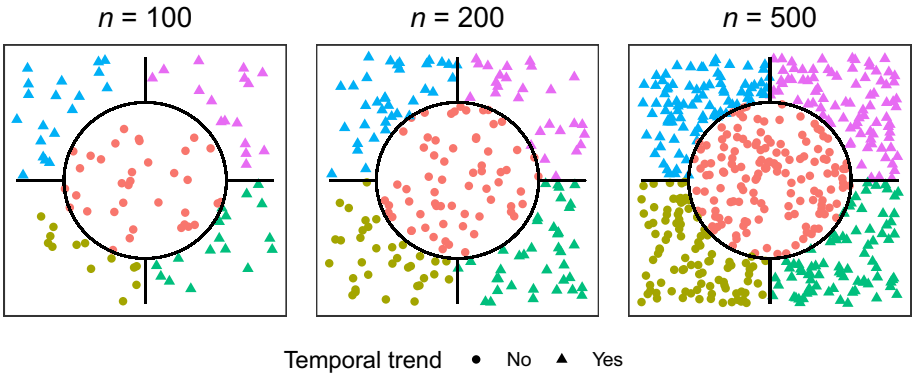


FIGURE 3 Cluster structures for $n = 100,200,500$. The solid lines and the color show the underlying cluster structure and the shape represents the status of temporal trends. A triangle represents a temporal trend, a circle represents no temporal trend.

TABLE 1 The summary results for $n = 100$ based on averages of 200 simulations.

	Method	ARMSE	$\hat{K}$	ARI	TT	F1 Score
$T = 20$	Global	$1.11_{\pm 0.0002}$	-	-	-	-
	Local	$0.26_{\pm 0.0084}$	-	-	-	-
	Local.k	$0.30_{\pm 0.1020}$	5.00	$0.73_{\pm 0.1420}$	-	-
	Local.gap	$0.56_{\pm 0.3064}$	$3.17_{\pm 1.4334}$	$0.49_{\pm 0.3076}$	-	-
	k-means	$0.22_{\pm 0.1525}$	5.00	$0.82_{\pm 0.1694}$	-	-
	k-means.gap	$0.43_{\pm 0.2877}$	$3.76_{\pm 1.3496}$	$0.63_{\pm 0.2890}$	-	-
	ClustGeo	$0.20_{\pm 0.0267}$	5.00	$0.86_{\pm 0.0451}$	-	-
	skater.knn	$0.83_{\pm 0.0469}$	5.00	$0.33_{\pm 0.0820}$	-	-
	skater.tri	$0.86_{\pm 0.0329}$	5.00	$0.41_{\pm 0.0878}$	-	-
	SCN	$0.10_{\pm 0.0305}$	$5.18_{\pm 0.3809}$	$0.96_{\pm 0.0285}$	-	-
	Scanner.GF	$0.08_{\pm 0.0327}$	$5.54_{\pm 0.6635}$	$0.97_{\pm 0.0276}$	$55.54_{\pm 5.8643}$	$0.97_{\pm 0.0481}$
	Scanner.MST	$0.07_{\pm 0.0326}$	$5.94_{\pm 0.9957}$	$0.91_{\pm 0.0949}$	$52.79_{\pm 3.4650}$	$0.99_{\pm 0.0308}$
$T = 50$	Global	$1.11_{\pm 0.0001}$	-	-	-	-
	Local	$0.18_{\pm 0.0056}$	-	-	-	-
	Local.k	$0.26_{\pm 0.1127}$	5.00	$0.74_{\pm 0.1539}$	-	-
	Local.gap	$0.60_{\pm 0.3571}$	$3.04_{\pm 1.6066}$	$0.44_{\pm 0.3334}$	-	-
	k-means	$0.25_{\pm 0.1698}$	5.00	$0.80_{\pm 0.1641}$	-	-
	k-means.gap	$0.39_{\pm 0.1922}$	$3.93_{\pm 1.1694}$	$0.66_{\pm 0.2087}$	-	-
	ClustGeo	$0.20_{\pm 0.0255}$	5.00	$0.84_{\pm 0.0520}$	-	-
	skater.knn	$0.83_{\pm 0.0532}$	5.00	$0.34_{\pm 0.0784}$	-	-
	skater.tri	$0.87_{\pm 0.0408}$	5.00	$0.42_{\pm 0.0795}$	-	-
	SCN	$0.12_{\pm 0.0258}$	$5.89_{\pm 0.5706}$	$0.92_{\pm 0.0175}$	-	-
	Scanner.GF	$0.12_{\pm 0.0234}$	$6.46_{\pm 0.9661}$	$0.93_{\pm 0.0196}$	$52.81_{\pm 1.7719}$	$0.97_{\pm 0.0153}$
	Scanner.MST	$0.04_{\pm 0.0197}$	$5.36_{\pm 0.7379}$	$0.97_{\pm 0.0617}$	$52.05_{\pm 0.8551}$	$1.00_{\pm 0.0133}$

TABLE 2 The summary results for $n = 200$ based on averages of 200 simulations.

	Method	ARMSE	$\hat{K}$	ARI	TT	F1 Score
$T = 20$	Global	$1.18_{\pm 0.0001}$	-	-	-	-
	Local	$0.23_{\pm 0.0051}$	-	-	-	-
	Local.k	$0.22_{\pm 0.1424}$	5.00	$0.84_{\pm 0.1647}$	-	-
	Local.gap	$0.38_{\pm 0.1891}$	$3.99_{\pm 1.1342}$	$0.65_{\pm 0.2510}$	-	-
	k-means	$0.19_{\pm 0.1468}$	5.00	$0.82_{\pm 0.1676}$	-	-
	k-means.gap	$0.35_{\pm 0.1846}$	$3.93_{\pm 0.9746}$	$0.68_{\pm 0.2191}$	-	-
	ClustGeo	$0.13_{\pm 0.0900}$	5.00	$0.87_{\pm 0.1407}$	-	-
	skater.knn	$0.67_{\pm 0.0610}$	5.00	$0.40_{\pm 0.1163}$	-	-
	skater.tri	$0.74_{\pm 0.0746}$	5.00	$0.39_{\pm 0.0623}$	-	-
	SCN	$0.05_{\pm 0.0189}$	$5.00_{\pm 0.0000}$	$1.00_{\pm 0.0074}$	-	-
	Scanner.GF	$0.04_{\pm 0.0194}$	$5.00_{\pm 0.0707}$	$1.00_{\pm 0.0069}$	$101.40_{\pm 3.5218}$	$1.00_{\pm 0.0148}$
	Scanner.MST	$0.04_{\pm 0.0187}$	$5.34_{\pm 0.6466}$	$0.98_{\pm 0.0467}$	$101.68_{\pm 5.5253}$	$1.00_{\pm 0.0216}$
$T = 50$	Global	$1.18_{\pm 0.0000}$	-	-	-	-
	Local	$0.17_{\pm 0.0035}$	-	-	-	-
	Local.k	$0.23_{\pm 0.1459}$	5.00	$0.81_{\pm 0.1623}$	-	-
	Local.gap	$0.38_{\pm 0.1771}$	$3.79_{\pm 0.9750}$	$0.65_{\pm 0.2345}$	-	-
	k-means	$0.24_{\pm 0.1764}$	5.00	$0.79_{\pm 0.1780}$	-	-
	k-means.gap	$0.40_{\pm 0.1676}$	$3.90_{\pm 1.0514}$	$0.64_{\pm 0.1864}$	-	-
	ClustGeo	$0.12_{\pm 0.0912}$	5.00	$0.85_{\pm 0.1483}$	-	-
	skater.knn	$0.67_{\pm 0.0579}$	5.00	$0.39_{\pm 0.1104}$	-	-
	skater.tri	$0.76_{\pm 0.0828}$	5.00	$0.40_{\pm 0.0716}$	-	-
	SCN	$0.05_{\pm 0.0182}$	$5.00_{\pm 0.0000}$	$0.99_{\pm 0.0076}$	-	-
	Scanner.GF	$0.04_{\pm 0.0191}$	$5.00_{\pm 0.0000}$	$0.99_{\pm 0.0073}$	$101.50_{\pm 0.6180}$	$1.00_{\pm 0.0024}$
	Scanner.MST	$0.03_{\pm 0.0131}$	$5.19_{\pm 0.5342}$	$0.99_{\pm 0.0322}$	$100.97_{\pm 0.3939}$	$1.00_{\pm 0.0020}$

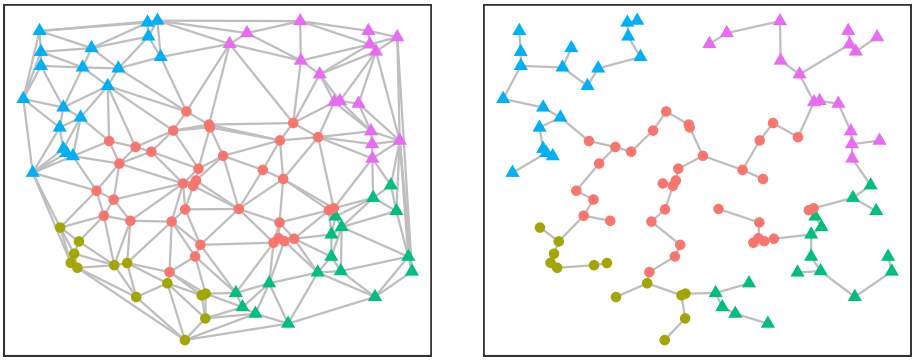
cluster pattern well. However, the ARMSE tends to be larger than "Scanner.GF" and "Scanner.MST" since it does not consider the sparsity of temporal trends. Among the methods, only "Scanner.GF" and "Scanner.MST" can identify the cluster structure and sparsity of temporal trends simultaneously. Thus, they can achieve better estimation performance (smaller ARMSE) than other approaches. When comparing these two methods, we observe that "Scanner.MST" has similar or better performances in terms of smaller ARMSE, larger ARI, and larger F1 score. Regarding the standard deviations across 200 simulations, "Scanner.MST" has slightly larger standard deviations than "Scanner.GF" for estimating the number of clusters ($\hat{K}$), and recovering the cluster structure (ARI). We also notice that, as T increases, the performances will become better. The simulation results in this section show that when spatial cluster structure and sparsity of time trends exist, ignoring either information can lead to worse performances.

4.2 | Study on computation complexity

In this section, we study the computation complexity of different penalization methods. This means that we will compare the computing time for SCN, Scanner.GF and Scanner.MST. An example of the constructed $\mathcal{G}$ and the corresponding MST is given in Figure 4 when the number of locations is 100, and the number of clusters is 5. The true functions used to construct the graph and MST in Figure 4 are provided in Section 4.1. Graph $\mathcal{G}$ in Figure 4a has 284 connections between locations, and the MST $\mathcal{T}$ in Figure 4b has 99 connections, which is $n - 1$. The number of connections is reduced by more than 50% when using a tree instead of the original graph. Then, the computational cost based on $\mathcal{T}$ is expected to significantly reduce the cost based on $\mathcal{G}$.

TABLE 3 The summary results for $n = 500$ based on averages of 200 simulations.

	Method	ARMSE	$\hat{K}$	ARI	TT	F1 Score
$T = 20$	Global	$1.16_{\pm 0.0005}$	-	-	-	-
	Local	$0.78_{\pm 0.0248}$	-	-	-	-
	Local.k	$0.28_{\pm 0.1727}$	5.00	$0.84_{\pm 0.1457}$	-	-
	Local.gap	$0.38_{\pm 0.1369}$	$4.16_{\pm 0.9374}$	$0.71_{\pm 0.1886}$	-	-
	k-means	$0.29_{\pm 0.1580}$	5.00	$0.79_{\pm 0.1496}$	-	-
	k-means.gap	$0.36_{\pm 0.1366}$	$4.30_{\pm 0.8740}$	$0.74_{\pm 0.1397}$	-	-
	ClustGeo	$0.15_{\pm 0.0648}$	5.00	$0.94_{\pm 0.0628}$	-	-
	skater.knn	$0.71_{\pm 0.0397}$	5.00	$0.32_{\pm 0.0851}$	-	-
	skater.tri	$0.75_{\pm 0.0154}$	5.00	$0.41_{\pm 0.0225}$	-	-
	SCN	$0.08_{\pm 0.0242}$	$5.04_{\pm 0.2078}$	$0.99_{\pm 0.0071}$	-	-
	Scanner.GF	$0.07_{\pm 0.0246}$	$5.05_{\pm 0.2185}$	$0.99_{\pm 0.0063}$	$255.85_{\pm 0.9912}$	$1.00_{\pm 0.0019}$
	Scanner.MST	$0.07_{\pm 0.0260}$	$5.13_{\pm 0.4171}$	$1.00_{\pm 0.0094}$	$254.96_{\pm 0.8044}$	$1.00_{\pm 0.0016}$
$T = 50$	Global	$1.16_{\pm 0.0000}$	-	-	-	-
	Local	$0.17_{\pm 0.0024}$	-	-	-	-
	Local.k	$0.25_{\pm 0.1550}$	5.00	$0.79_{\pm 0.1685}$	-	-
	Local.gap	$0.37_{\pm 0.1624}$	$4.03_{\pm 0.9321}$	$0.66_{\pm 0.2052}$	-	-
	k-means	$0.27_{\pm 0.1771}$	5.00	$0.78_{\pm 0.1673}$	-	-
	k-means.gap	$0.37_{\pm 0.1578}$	$4.15_{\pm 0.9758}$	$0.70_{\pm 0.1381}$	-	-
	ClustGeo	$0.11_{\pm 0.0742}$	5.00	$0.94_{\pm 0.0573}$	-	-
	skater.knn	$0.71_{\pm 0.0387}$	5.00	$0.31_{\pm 0.0883}$	-	-
	skater.tri	$0.75_{\pm 0.0264}$	5.00	$0.41_{\pm 0.0306}$	-	-
	SCN	$0.02_{\pm 0.0098}$	$5.01_{\pm 0.0997}$	$1.00_{\pm 0.0026}$	-	-
	Scanner.GF	$0.02_{\pm 0.0093}$	$5.01_{\pm 0.0997}$	$1.00_{\pm 0.0022}$	$255.18_{\pm 0.3852}$	$1.00_{\pm 0.0008}$
	Scanner.MST	$0.02_{\pm 0.0060}$	$5.00_{\pm 0.0707}$	$1.00_{\pm 0.0009}$	$255.00_{\pm 0.0707}$	$1.00_{\pm 0.0001}$



(a) (b)

FIGURE 4 An example of constructing graph and MST when $n = 100$. The color shows the underlying cluster structure and the shape represents the status of temporal trends. A triangle represents a temporal trend, a circle represents no temporal trend. Gray solid lines between locations represent that locations are connected in the graph. (a) an undirected graph constructed based on Delaunay triangulation; (b) MST based on the graph in (a) and the weights using (9).

We compare the computation cost of SCN, Scanner.GF and Scanner.MST when $n = 100{,}200{,}500$ for $T = 100$. The computation cost is compared based on the computation time. For each simulation, the total time based on 100 values of λ_1 and 150 values λ_2 is recorded. The range of λ_1 is from 0.0001 to 0.3, and the range of λ_2 is from 0.1 to 10. Figure 5 shows the average computation time across 200 simulations. When $n = 100$, these three methods have similar computation time. When $n = 200$ and 500, the computation cost of Scanner.MST is about 65% of SCN and Scanner.GF. We also know that Scanner.MST performs better in recovering cluster structure (larger ARI) and sparsity from Tables 1–3. Recall that the results in Section 4.1 indicate that Scanner.MST and Scanner.GF have comparable performances in estimating parameters, recovering cluster structure, and detecting temporal trends. Together with the results in this part, we can conclude that the Scanner.MST method is running faster than the Scanner.GF and SCN with similar performances or better performances than Scanner.GF and SCN.

4.3 | Study on complexity cluster structure

In this part, we study the performances of the proposed methods under a more complicated scenario with 10 clusters for $n = 100$ and $T = 20, 50, 100$. The functions of the ten clusters are $g(\mathbf{s}_i, t) = 1$, $g(\mathbf{s}_i, t) = -1$, $g(\mathbf{s}_i, t) = 2$, $g(\mathbf{s}_i, t) = -2$, $g(\mathbf{s}_i, t) = 1 - \exp(1.5t)$, $g(\mathbf{s}_i, t) = -1 + \exp(1.5t)$, $g(\mathbf{s}_i, t) = \sin(2\pi t)$, $g(\mathbf{s}_i, t) = -\sin(2\pi t)$, $g(\mathbf{s}_i, t) = 1.5t^3$ and $g(\mathbf{s}_i, t) = -1.5t^3$. Figure 6 shows the cluster structure. The number of nonzero temporal trends is 75. Table 4 shows the average values of

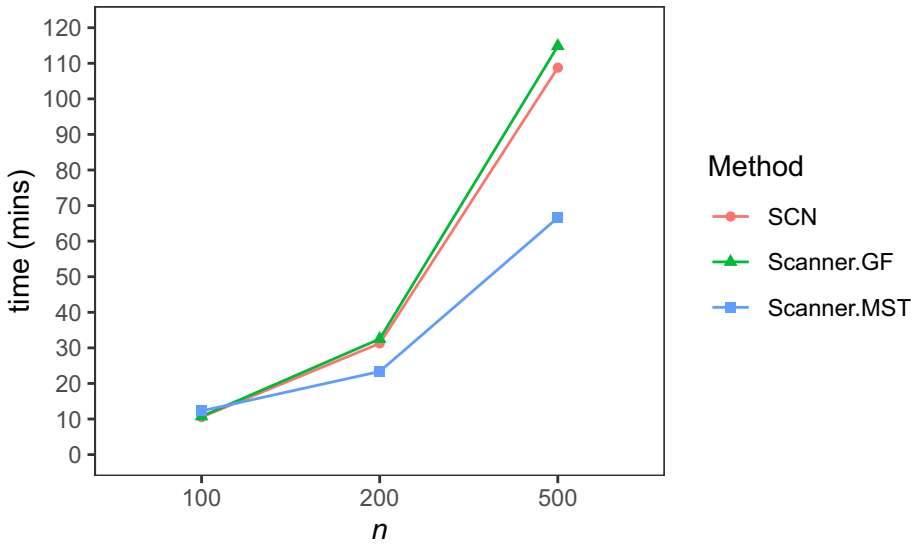


FIGURE 5 Averaged computation time when $n = 100{,}200{,}500$ for $T = 100$ across 200 simulations.

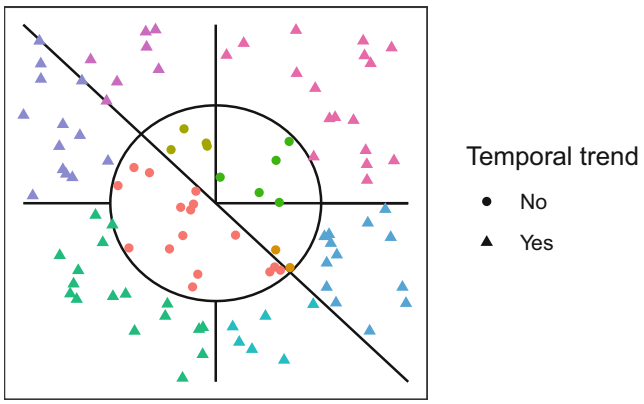


FIGURE 6 Cluster structure for $n = 100$ when the number of clusters is 10. The solid lines and the color show the underlying cluster structure and the shape represents the status of temporal trends. A triangle represents a temporal trend, a circle represents no temporal trend.

TABLE 4 The summary results for 10 clusters based on averages of 200 simulations.

	Method	ARMSE	$\hat{K}$	ARI	TT	F1 Score
$T = 20$	SCN	$0.18_{\pm 0.0345}$	$9.22_{\pm 0.5531}$	$0.95_{\pm 0.0173}$	-	-
	Scanner.GF	$0.17_{\pm 0.0379}$	$9.49_{\pm 0.7017}$	$0.95_{\pm 0.0223}$	$88.68_{\pm 8.7721}$	$0.90_{\pm 0.0485}$
	Scanner.MST	$0.22_{\pm 0.0899}$	$9.80_{\pm 1.7557}$	$0.89_{\pm 0.0693}$	$93.70_{\pm 8.3563}$	$0.88_{\pm 0.0440}$
$T = 50$	SCN	$0.10_{\pm 0.0292}$	$10.27_{\pm 0.6852}$	$0.97_{\pm 0.0150}$	-	-
	Scanner.GF	$0.10_{\pm 0.0322}$	$10.46_{\pm 0.8554}$	$0.97_{\pm 0.0197}$	$79.11_{\pm 3.2698}$	$0.96_{\pm 0.0197}$
	Scanner.MST	$0.06_{\pm 0.0159}$	$10.58_{\pm 0.7980}$	$0.98_{\pm 0.0280}$	$78.06_{\pm 4.0445}$	$0.97_{\pm 0.0254}$
$T = 100$	SCN	$0.04_{\pm 0.0136}$	$10.06_{\pm 0.2471}$	$1.00_{\pm 0.0065}$	-	-
	Scanner.GF	$0.04_{\pm 0.0133}$	$10.08_{\pm 0.2720}$	$1.00_{\pm 0.0062}$	$77.48_{\pm 3.1383}$	$0.97_{\pm 0.0198}$
	Scanner.MST	$0.03_{\pm 0.0031}$	$10.02_{\pm 0.1404}$	$1.00_{\pm 0.0054}$	$76.41_{\pm 2.9726}$	$0.98_{\pm 0.0190}$

different measures based on 200 simulations. We observe that the proposed method can recover the cluster structure well with $\hat{K}$ close to 10 and ARI close to 1. The F1 score of "Scanner.GF" increases by about 5% for $T = 50{,}100$ compared to $T = 20$, and the F1 score of "Scanner.MST" increases by about 10% for $T = 50{,}100$ compared to $T = 20$. These indicate that the performance of identifying nonzero temporal trends improves a lot when $T = 50{,}100$ compared to $T = 20$. And when $T = 50{,}100$, "Scanner.MST" and "Scanner.GF" have similar standard deviations across 200 simulations.

5 | REAL DATA ANALYSIS

In this section, we apply our proposed Scanner approach to study our motivating dataset described in Section 1. The original data were collected daily by the National Meteorological Information Center, China Meteorological Administration. Precipitation regions are studied in different literature (Gomes et al., 2018; Liu & Xu, 2016; Roushangar & Alizadeh, 2018; Satyanarayana & Srinivas, 2011; Xiao et al., 2013). Claps et al. (2022) gave a review of rainfall regionalization techniques. Finding homogeneous precipitation regions is essential in studying the precipitation trends, estimation of precipitation quantiles in ungauged locations (Claps et al., 2022), investigating spatial distributions of precipitation (Satyanarayana & Srinivas, 2011) and effective management of water resources (Roushangar & Alizadeh, 2018). Studying annual precipitation is one of the approaches (Roushangar & Alizadeh, 2018; Satyanarayana & Srinivas, 2011). After having precipitation regions, people can study the features of different regions separately. In traditional hydrology methodology, model-free approaches are used, which cannot identify the temporal trends.

We study the cluster and temporal pattern for yearly precipitation data in 660 basic stations from 1995 to 2014. As shown in Figure 1a, these stations are spatially spread, and closer stations could have similar precipitation patterns. For each location, we have local observations to fit individual spline regression models, and the regression coefficients are used to construct the MST to represent the spatial connections among stations. Besides this, from the 50 randomly temporal trend lines of 50 selected stations in Figure 1b, some trends could be constants over the considered year period. These observations motivate us to find spatial clusters and detect zero and nonzero temporal trends together.

Our goal is to find the precipitation region and estimate the temporal trends simultaneously. In the analysis, we have $n = 660$ (660 stations) and $T = 20$ (number of observations in each station). As discussed in Section 2, $\rho = 1$ is used in the algorithm. Similar to the simulation studies, the procedure introduced in Section 2.4 is used. That is, in the first step, tuning parameter $\lambda'_{1i}s$ are selected from a grid of 100 values with a range from 0.0001 to 0.3. In the second step, a grid of 1000 values of λ_2 ranging from 0.1 to 20 is used. $\gamma_1 = \gamma_2 = 3$ are used following the literature as discussed in Section 2.4. The graph constructed based on the Delaunay triangulation has 1959 connections. And the MST has 659 connections. Figure 7 shows the constructed undirected graph and the MST. The MST only has about 30% connections of the graph in Figure 7a. As shown in the simulation study, Scanner based on MST has similar model performances to the method based on the graph but has less computation cost. Thus, we only use Scanner based on MST to analyze the yearly precipitation data.

We show the estimated results in Figure 8. In the figure, colors indicate different clusters, and the pattern indicates whether a temporal trend exists. There are 23 spatial clusters in total, and 5 of them have temporal trends. The map has

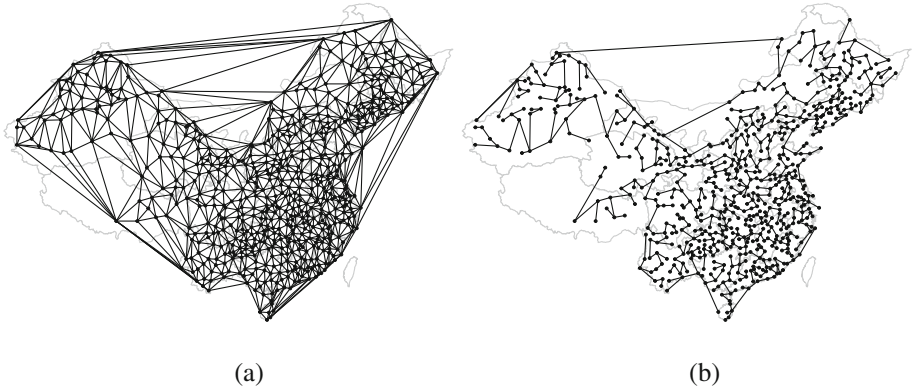


FIGURE 7 Constructed undirected graph and MST. The solid lines between stations represent connections in the graph. (a) Undirected graph. (b) MST.

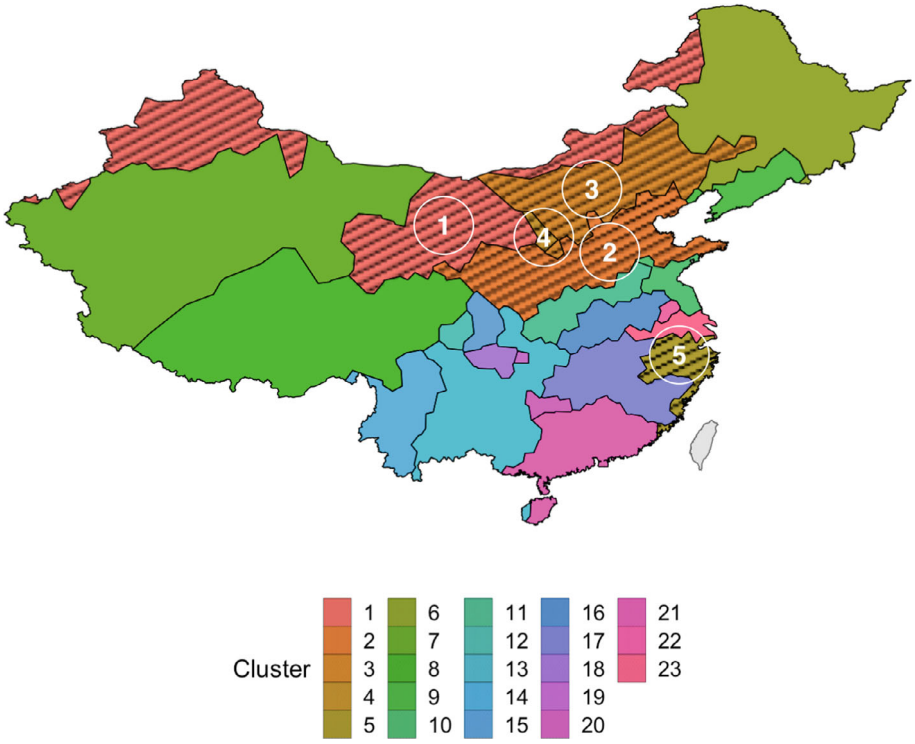


FIGURE 8 Maps of estimated clusters and trends based on Scanner.MST. The color shows different clusters, and the pattern shows the existence of temporal trends.

a similar cluster pattern to the regions in Chen et al. (2009) and Zhang et al. (2016) based on daily precipitation data, which are related to the stepwise manner of East Asian monsoon (Ding, 2004), where the stepwise manner refers to the fact that the regions are not continuous and jumps exist among different regions. We also apply the spatial clustering method,*ClustGeo*, to the data. The estimated clusters are shown in Figure 9 when the number of clusters is fixed at 23. The cluster structure based on *ClustGeo* in Figure 9 has more isolated stations than that in Figure 8. There are also similar cluster patterns between these two results. For example, both clusters are detected in the northeast of China, Xizang province area, and the east coast area.

Figure 10a shows the fitted lines and 95% confidence bands of the five clusters with temporal trends. The separable confidence bands indicate that these five clusters are different. Figure 10b shows the estimated mean and the 95% confidence intervals for clusters without temporal trends. Besides these, it can be seen that Cluster 1 has a little temporal trend

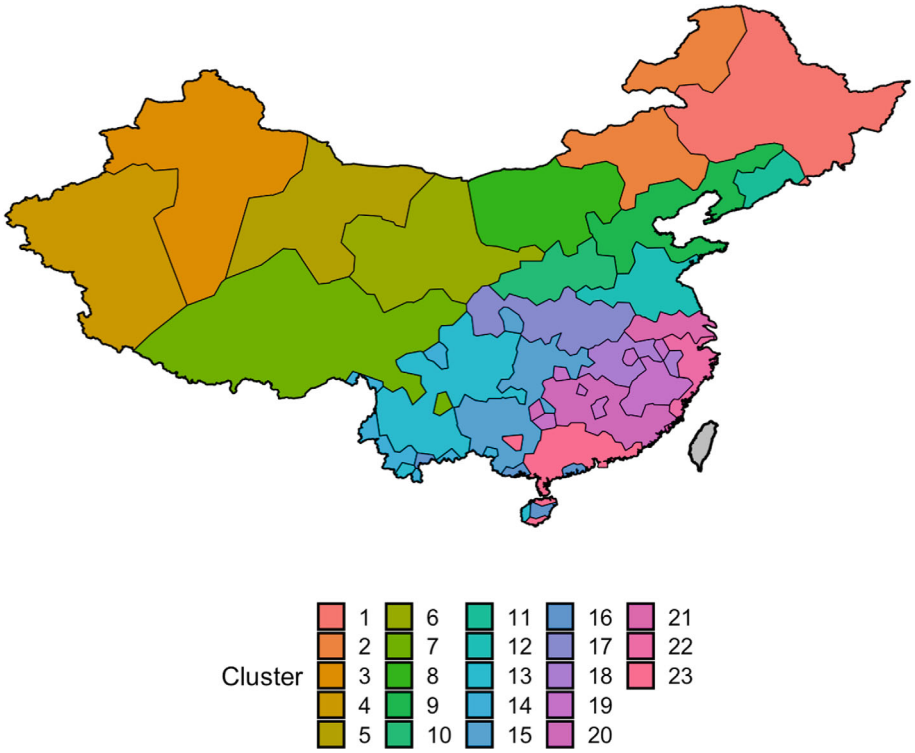


FIGURE 9 Maps of estimated clusters based on ClustGeo. The color shows different clusters.

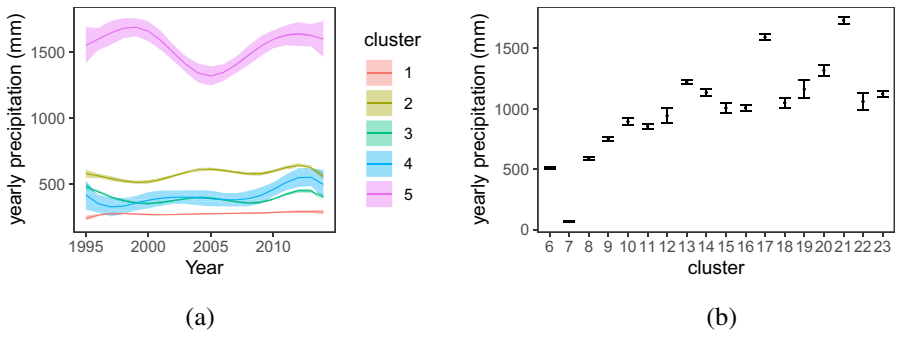


FIGURE 10 Estimated clusters with time trends and without time trends. (a) Estimated clusters with nonzero time trend. (b) 95% confidence intervals for clusters with zero time trend.

around year 1995, which corresponds to the northern parts in Figure 8. The yearly precipitation change is relatively small compared to other clusters. Cluster 4 is close to Cluster 3, which shows an increasing trend. This means that yearly precipitation has increased in these two clusters, especially in Cluster 4, which also indicates that attention should be given to this area about the precipitation change. Cluster 4 and Cluster 3 are also close on the map and correspond to the central north. Cluster 2 is the central cluster with larger precipitation than clusters 1, 2, and 3, which has a slightly decreasing trend in the recent five years. This also indicates that there is a potential drought in this area. Cluster 5 has a different cycle than other clusters, which corresponds to the southern cluster in Figure 8. This area is close to the ocean, which has larger precipitation than other clusters but also has a larger variability among different years. For clusters without temporal trends, the yearly precipitations are stationary over these 20 years. Different clusters have different estimated yearly precipitations and may have different variations due to different numbers of stations. In clusters 12, 19, and 22, there are 3, 1, and 2 stations, respectively. In other clusters without temporal trends, there are at least 5 stations, resulting in small variations. Compared to the regions in Chen et al. (2009) and Zhang et al. (2016) based on monthly data, there are some similarities and differences due to different research perspectives (monthly data vs yearly data) and statistical

approaches. In particular, Cluster 5 is smaller than a similar cluster in the literature. In the literature, the cluster around the area in cluster 5 also contains more western areas. Our result can make sense because stations in western areas are far from the ocean, which could have different patterns than those closer. Cluster 2 is similar to the northern part of cluster 1 in Chen et al. (2009), which corresponds to the areas around the Yellow River. Cluster 3 is similar to cluster 2 in Chen et al. (2009) and Zhang et al. (2016), which corresponds to a dry region. Cluster 3 and Cluster 4 are similar but show some differences in temporal trends. For the Hainan island, we got a similar pattern to that in Zhang et al. (2016), where the western part separated from other areas in Hainan.

Within one cluster, we can assume stations are homogeneous. And among different clusters, different patterns exist. Then, cluster features can be studied in different clusters. For example, studying spatial correlations of precipitation is important for surface data generation, data collection, and interpretation, as discussed in Fan et al. (2021). They studied spatial correlations of daily precipitation over different precipitation regions, where these regions are based on model-free approaches, which may not show the differences among different regions. Our clusters are data-driven, which can better reveal the heterogeneity of different areas.

6 | CONCLUSION

In this article, we propose a new method called spatial clustered and sparse nonparametric regression (Scanner) models to find spatial clusters and select nonzero temporal trends for spatial temporal data. To achieve the goal, we use B-spline regression models to model longitudinal curves and design a doubly penalized least squared approach. In Scanner, a group sparse penalty is used to detect temporal trends, and a spanning tree-based fusion penalty is used for identifying spatial clusters. A new algorithm based on the ADMM algorithm is developed to find the cluster structure of locations, identify the sparsity of temporal trends, and estimate parameters automatically. In the simulation study, we used several numerical examples to investigate and compare the performance of Scanner. The numerical results show that Scanner can recover the spatial structure and detect temporal trends well. We also found that the tree-based fusion penalty has a similar or better performance compared to the graph-based fusion penalty in most scenarios. Furthermore, the tree-based approach can greatly reduce the computational cost. When the number of clusters is larger, the tree structure might hurt the performance of recovering cluster structure a little bit when local data is not enough. But the results are still reasonable. Besides these, we established the theoretical properties of the Scanner estimator in terms of its consistency in recovering cluster structure and identifying temporal trends.

In our simulation studies and the real data analysis, all locations have the same number of observations and are observed at the same time points. Theoretically, the algorithm can handle missing values as long as each location has observations since we can still use B-spline to approximate the smooth functions, that is $f(\boldsymbol{s}_i, t) = \boldsymbol{B}^\top(t)\boldsymbol{\beta}(\boldsymbol{s}_i)$, which does not require that all locations have the same observed time. However, the performances still need to be investigated under different missing scenarios.

The idea of the proposed approach can be extended to varying coefficient models (Wang et al., 2008) and semiparametric varying coefficient models Tian et al. (2014). In varying coefficient models with multiple covariates, the corresponding regression coefficients are modeled as smooth functions of time. In traditional varying coefficient models, for each covariate, the smooth function is assumed to be the same across all locations. This can be relaxed by assuming smooth functions have clusters across different locations. Furthermore, we can also select variables and find clusters of spatial locations simultaneously by using doubly penalty functions used in our work. The advantages of the extended model are that a more flexible model may have better estimation performance, and the double penalty approach can find clusters and identify the sparsity of effects of covariates together. However, there are still some disadvantages that need to be addressed in the future. As discussed here and in the literature, location-based MST may not accurately identify the true clusters. We need to explore the approaches of using local information to construct MST and evaluate the performances. The computation efficiency is another potential issue since multiple smooth functions are needed to be estimated together. Besides these, the proposed approach can also be extended to other types of models to study different kinds of spatial temporal data, such as binary data and count data.

FUNDING INFORMATION STATEMENT

The work of Xin Wang is supported in part by the National Science Foundation grant NSF SES-2316353.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the Supplementary Material of this article.

ORCID

Xin Wang https://orcid.org/0000-0001-7801-1728

REFERENCES

Abraham, C., Cornillon, P.-A., Matzner-Løber, E., & Molinari, N. (2003). Unsupervised curve clustering using B-splines. *Scandinavian Journal of Statistics*, *30*(3), 581–595.

Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, *3*(1), 1–122.

Chavent, M., Kuentz-Simonet, V., Labenne, A., & Saracco, J. (2018). Clustgeo: An *R* package for hierarchical clustering with spatial constraints. *Computational Statistics*, *33*(4), 1799–1822.

Chen, L., Chen, D., Wang, H., & Yan, J. (2009). Regionalization of precipitation regimes in China. *Atmospheric and Oceanic Science Letters*, *2*(5), 301–307.

Claps, P., Ganora, D., & Mazzoglio, P. (2022). *Rainfall regionalization techniques*. In *Rainfall* (pp. 327–350). Elsevier.

Coffey, N., Hinde, J., & Holian, E. (2014). Clustering longitudinal profiles using *P*-splines and mixed effects models applied to time-course gene expression data. *Computational Statistics & Data Analysis*, *71*, 14–29.

Cook, A. J., Gold, D. R., & Li, Y. (2007). Spatial cluster detection for censored outcome data. *Biometrics*, *63*(2), 540–549.

Ding, Y. (2004). *Seasonal march of the East-Asian summer monsoon*. In *East Asian monsoon* (pp. 3–53). World Scientific.

Dudoit, S., & Fridlyand, J. (2002). A prediction-based resampling method for estimating the number of clusters in a dataset. *Genome Biology*, *3*(7), 1–21.

Eldershaw, C., & Hegland, M. (1997). *Cluster analysis using triangulation*. In *Computational techniques and applications* (pp. 201–208). World Scientific.

Erhardt, T. M., Czado, C., & Schepsmeier, U. (2015). *R*-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, *71*(2), 323–332.

Fan, C., Yin, S., & Chen, D. (2021). Spatial correlations of daily precipitation over mainland China. *International Journal of Climatology*, *41*(14), 6350–6365.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, *96*(456), 1348–1360.

Feng, C. (2022). Spatial-temporal generalized additive model for modeling COVID-19 mortality risk in Toronto. *Canada. Spatial Statistics*, *49*, 100526.

Fioravanti, G., Cameletti, M., Martino, S., Cattani, G., & Pisoni, E. (2022). A spatiotemporal analysis of NO_2 concentrations during the Italian 2020 COVID-19 lockdown. *Environmetrics*, *33*(4), e2723.

Gomes, E. P., Blanco, C. J. C., & Pessoa, F. C. L. (2018). Regionalization of precipitation with determination of homogeneous regions via fuzzy c-means. *RBRH*, *23*, e51.

Gong, Y., & de Haan, J. (2018). Accounting for spatial variation of land prices in hedonic imputation house price indices: A semi-parametric approach. *Journal of Official Statistics*, *34*(3), 695–720.

Gower, J. C., & Ross, G. J. (1969). Minimum spanning trees and single linkage cluster analysis. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, *18*(1), 54–64.

Grygorash, O., Zhou, Y., & Jorgensen, Z. (2006). *Minimum spanning tree based clustering algorithms*. In *2006 18th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'06)* (pp. 73–81). IEEE.

Haggarty, R., Miller, C., & Scott, E. (2015). Spatially weighted functional clustering of river network data. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, *64*(3), 491–506.

Hallac, D., Leskovec, J., & Boyd, S. (2015). *Network lasso: Clustering and optimization in large graphs*. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 387–396). ACM.

Hu, G., Geng, J., Xue, Y., & Sang, H. (2022). Bayesian spatial homogeneity pursuit of functional data: An application to the US income distribution. *Bayesian Analysis*, *1*(1), 1–27.

Huang, J., Horowitz, J. L., & Wei, F. (2010). Variable selection in nonparametric additive models. *Annals of Statistics*, *38*(4), 2282–2313.

Huang, J. Z., Wu, C. O., & Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, *14*(3), 763–788.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*(1), 193–218.

James, G. M., & Sugar, C. A. (2003). Clustering for sparsely sampled functional data. *Journal of the American Statistical Association*, *98*(462), 397–408.

Jiang, H., & Serban, N. (2012). Clustering random curves under spatial interdependence with application to service accessibility. *Technometrics*, *54*(2), 108–119.

Johnson, S., Heaps, S., Wilson, K., & Wilkinson, D. (2023). A Bayesian spatio-temporal model for short-term forecasting of precipitation fields. *Environmetrics*, *34*(8), e2824.

Jung, I., Kulldorff, M., & Klassen, A. C. (2007). A spatial scan statistic for ordinal data. *Statistics in Medicine*, *26*(7), 1594–1607.

Jurek, M., & Katzfuss, M. (2023). Scalable spatio-temporal smoothing via hierarchical sparse Cholesky decomposition. *Environmetrics*, *34*(1), e2757.

Kulldorff, M., & Nagarwalla, N. (1995). Spatial disease clusters: Detection and inference. *Statistics in Medicine*, *14*(8), 799–810.

Lawson, A. B. (2018). *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. Chapman and Hall/CRC.

Lee, D.-T., & Schachter, B. J. (1980). Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer and Information Sciences*, *9*(3), 219–242.

Li, F., & Sang, H. (2019). Spatial homogeneity pursuit of regression coefficients for large datasets. *Journal of the American Statistical Association*, *114*(527), 1050–1062.

Li, X., Wang, L., & Nettleton, D. (2019). Additive partially linear models for ultra-high-dimensional regression. *Stat*, *8*(1), e223.

Liang, D., Zhang, H., Chang, X., & Huang, H. (2021). Modeling and regionalization of China's PM2.5 using spatial-functional mixture models. *Journal of the American Statistical Association*, *116*(533), 116–132.

Lin, F., Tang, Y., Zhu, H., & Zhu, Z. (2022). Spatially clustered varying coefficient model. *Journal of Multivariate Analysis*, *192*, 105023. https://doi.org/10.1016/j.jmva.2022.105023

Liu, L., & Xu, Z. (2016). Regionalization of precipitation and the spatiotemporal distribution of extreme precipitation in Southwestern China. *Natural Hazards*, *80*, 1195–1211.

Liu, M., Yang, J., Liu, Y., Jia, B., Chen, Y.-F., Sun, L., & Ma, S. (2023). A fusion learning method to subgroup analysis of Alzheimer's disease. *Journal of Applied Statistics*, *50*(8), 1686–1708.

Liu, R., & Yang, L. (2010). Spline-backfitted kernel smoothing of additive coefficient model. *Econometric Theory*, *26*(1), 29–59.

Luan, Y., & Li, H. (2003). Clustering of time-course gene expression data using a mixed-effects model with b-splines. *Bioinformatics*, *19*(4), 474–482.

Ma, S., & Huang, J. (2017). A concave pairwise fusion approach to subgroup analysis. *Journal of the American Statistical Association*, *112*(517), 410–423.

Ma, S., Huang, J., Zhang, Z., & Liu, M. (2020). Exploration of heterogeneous treatment effects via concave fusion. *The International Journal of Biostatistics*, *16*(1). https://www.degruyter.com/document/doi/10.1515/ijb-2018-0026/html

Ma, S., & Yang, L. (2011). Spline-backfitted kernel smoothing of partially linear additive model. *Journal of Statistical Planning and Inference*, *141*(1), 204–219.

Romano, E., Balzanella, A., & Verde, R. (2013). *A regionalization method for spatial functional data based on Variogram models: an application on environmental data*. In *Advances in theoretical and applied statistics* (pp. 99–108). Springer.

Rougier, J., Brady, A., Bamber, J.L., Chuter, S., Royston, S., Vishwakarma, D.J., Westaway, R., & Ziegler, Y. (2023). The scope of the Kalman filter for spatio-temporal applications in environmental science. *Environmetrics*, *34*(1), e2773.

Roushangar, K., & Alizadeh, F. (2018). A multiscale spatio-temporal framework to regionalize annual precipitation using k-means and self-organizing map technique. *Journal of Mountain Science*, *15*(7), 1481–1497.

Sasaki, Y. (2007). The truth of the F-measure. *Teach Tutor Mater*, *1*(5), 1–5.

Satyanarayana, P., & Srinivas, V. (2011). Regionalization of precipitation in data sparse areas using large scale atmospheric variables–A fuzzy clustering approach. *Journal of Hydrology*, *405*(3-4), 462–473.

Simon, N., Friedman, J., Hastie, T., & Tibshirani, R. (2013). A sparse-group lasso. *Journal of Computational and Graphical Statistics*, *22*(2), 231–245.

Tang, X., & Li, L. (2023). Multivariate temporal point process regression. *Journal of the American Statistical Association*, *118*(542), 830–845.

Tian, R., Xue, L., & Liu, C. (2014). Penalized quadratic inference functions for semiparametric varying coefficient partially linear models with longitudinal data. *Journal of Multivariate Analysis*, *132*, 94–110.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B: Methodological*, *58*(1), 267–288.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, *63*(2), 411–423.

Torabi, M., & Rosychuk, R. J. (2011). Spatio-temporal modelling using B-spline for disease mapping: analysis of childhood cancer trends. *Journal of Applied Statistics*, *38*(9), 1769–1781.

Ugarte, M. D., Goicoa, T., & Militino, A. F. (2010). Spatio-temporal modeling of mortality risks using penalized splines. *Environmetrics*, *21*(3-4), 270–289.

Velarde, L. G. C., Migon, H. S., Pereira, B., & de Braganca Pereira, B. (2004). Space–time modeling of rainfall data. *Environmetrics: The official journal of the International Environmetrics Society*, *15*(6), 561–576.

Wan, Y., Xu, M., Huang, H., & Xi Chen, S. (2021). A spatio-temporal model for the analysis and prediction of fine particulate matter concentration in Beijing. *Environmetrics*, *32*(1), e2648.

Wang, L., Li, H., & Huang, J. Z. (2008). Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. *Journal of the American Statistical Association*, *103*(484), 1556–1569.

Wang, L., Wang, G., Yu, H., & Wang, F. (2022). Prediction and analysis of residential house price using a flexible spatiotemporal model. *Journal of Applied Economics*, *25*(1), 503–522.

Wang, X. (2024). Clustering of longitudinal curves via a penalized method and EM algorithm. *Computational Statistics*, *39*, 1485–1512.

Wang, X., Zhang, X., & Zhu, Z. (2023). Clustered coefficient regression models for poisson process with an application to seasonal warranty claim data. *Technometrics*, *65*(4), 514–523.

Wang, X., Zhu, Z., & Zhang, H. H. (2023). Spatial heterogeneity automatic detection and estimation. *Computational Statistics & Data Analysis*, *180*, 107667. https://doi.org/10.1016/j.csda.2022.107667

Wood, S. N. (2013). On *p*-values for smooth components of an extended generalized additive model. *Biometrika*, *100*(1), 221–228.

Xiao, M., Zhang, Q., Singh, V. P., & Chen, X. (2013). Regionalization-based spatiotemporal variations of precipitation regimes across China. *Theoretical and Applied Climatology*, *114*, 203–212.

Xue, L., & Yang, L. (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica*, *16*(4), 1423–1446.

Yang, X., Yan, X., & Huang, J. (2019). High-dimensional integrative analysis with homogeneity and sparsity recovery. *Journal of Multivariate Analysis*, *174*, 104529. https://doi.org/10.1016/j.jmva.2019.06.007

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, *38*(2), 894–942.

Zhang, H., Zhu, Z., & Yin, S. (2016). *Identifying precipitation regimes in China using model-based clustering of spatial functional data*. In *Proceedings of the 6th International Workshop on Climate Informatics: CI 2016* (pp. 117–120). National Center for Atmospheric Research.

Zhang, X., Gao, W., & Yun, S. U. (2015). Electricity consumer archetypes study based on functional data analysis and k-means algorithm. *Power System Technology*, *39*(2), 3153–3162.

Zhang, X., Liu, J., & Zhu, Z. (2024). Learning coefficient heterogeneity over networks: A distributed spanning-tree-based Fused-Lasso regression. *Journal of the American Statistical Association*, *119*(545), 485–497.

Zhang, Y., Chen, S. X., & Bao, L. (2023). Air pollution estimation under air stagnation—A case study of Beijing. *Environmetrics*, *34*(6), e2819.

Zhu, X., & Qu, A. (2018). Cluster analysis of longitudinal profiles with subgroups. *Electronic Journal of Statistics*, *12*(1), 171–193.

SUPPORTING INFORMATION

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

How to cite this article: Wang, X., & Zhang, X. (2024). **Scanner**: Simultaneously temporal trend and spatial cluster detection for spatial-temporal data. *Environmetrics*, e2849. https://doi.org/10.1002/env.2849