Straive	JTSA	70002	WILEY	Dispatch: 18-06-2025	CE:
	Journal Name	Manuscript No		No. of Pages: 17	PE:

Journal of Time Series Analysis

WILEY

SPECIAL ISSUE ARTICLE

Spatiotemporal Heterogeneity Learning: Generalized SpatioTemporal Semi-Varying Coefficient Models With Structure Identification

Zhiling Gu[1] | Xinyi Li[2] | Guannan Wang[3] | Lily Wang[4]

[1]Yale University, New Haven, Connecticut, USA | [2]Clemson University, Clemson, South Carolina, USA | [3]William & Mary, Williamsburg, Virginia, USA | [4]Department of Statistics, George Mason University, Fairfax, Virginia, USA

Correspondence: Lily Wang (lwang41@gmu.edu)

Received: 6 July 2024 | **Revised:** 6 June 2025 | **Accepted:** 9 June 2025

Funding: This work was supported by the National Institutes of Health (P20 GM139769, R01 AG085616), the Simonsen Foundation (963447), and the Directorate for Mathematical and Physical Sciences (2210658, 2426173).

Keywords: penalization | prismatic partition | semi-varying coefficient models | spatiotemporal data | splines

ABSTRACT

This paper proposes a class of Generalized SpatioTemporal Semi-Varying Coefficient Models (GST-SVCMs) with structure identification to enhance the detection and interpretation of spatiotemporal heterogeneity in factors influencing response variables. The proposed framework effectively distinguishes between spatiotemporally varying and constant effects, addressing a key limitation of current modeling approaches. By identifying and separating these components, the GST-SVCM structure identification method improves both computational efficiency and the statistical power of downstream analyses. The estimators of constant coefficients and varying coefficient functions are consistent, and the estimators of the constant coefficients are asymptotically normal, facilitating reliable statistical inference. Extensive Monte Carlo simulations demonstrate that the proposed method accurately identifies the true model structure and significantly improves prediction accuracy compared to purely varying coefficient models that do not incorporate structure identification. To further refine model granularity, we extend GST-SVCMs by introducing the Hierarchical Spatiotemporal Varying Coefficient Model (HSTVCM) with automatic structure identification, which decomposes effects into spatial, temporal, and spatiotemporal components for more precise structure identification. The practical utility of the proposed methodologies is validated through an application to particulate matter (PM) data, providing insights into the influence of meteorological factors on PM levels and determining whether these effects exhibit true spatiotemporal variation.

MSC2020 Classification: 62G08, 62M10, 62H11

1 | Introduction

Varying coefficient models (VCMs) are a class of statistical models that extend classical linear models by allowing coefficients to change as a function of one or more variables (Hastie and Tibshirani 1993; Fan and Zhang 1999; Chiang et al. 2001; Gelfand et al. 2003). This adaptability makes VCMs particularly effective in capturing dynamic relationships between predictors and response variables that traditional models with fixed/constant coefficients fail to describe adequately. Expanding on the capabilities of VCMs, spatiotemporal varying coefficient models (STVCMs) incorporate spatial and temporal information into the coefficient functions. This integration is crucial to understanding complex phenomena characterized by both spatial and temporal variability. Traditional models, which assume constant effects across space and time, often miss these complex variations, leading to biased estimates and inaccurate predictions. This limitation could be particularly critical in fields such as environmental science, epidemiology, and economics, where the interaction between spatial and temporal factors plays a significant role in affecting outcomes.

For instance, in the analysis of particulate matter (PM) data, it is key to account for spatiotemporal heterogeneity in the effects of various meteorological factors such as temperature, humidity, wind speed, and atmospheric pressure on PM levels (Keller et al. 2015; Xue et al. 2017). Models with fixed coefficients assume a uniform effect of these factors in different regions and times, leading to the overlooking of critical variations. The STVCM addresses this issue by allowing coefficients to change dynamically across space and time, providing a more accurate and detailed representation of the underlying processes. In the context of PM studies, this flexibility provides valuable insights into how and why PM levels fluctuate, revealing patterns such as higher sensitivity to temperature changes in urban areas during summer months or varying impacts of wind speed in coastal versus inland regions. This detailed understanding enables more accurate predictions of PM levels and supports the development of targeted interventions to reduce pollution and protect public health.

In many applied fields, data are frequently collected as count or binary responses associated with geographic locations and temporal points. In this article, we consider the Generalized SpatioTemporal Varying Coefficient Model (GSTVCM), which encompasses various existing semiparametric models. Let $\mathcal{T}$ and Ω represent the one-dimensional (1D) time domain and the two-dimensional (2D) spatial domain that can have arbitrary shapes. Suppose that there are n space-time observations $\{(\mathbf{S}_i, T_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$ from the joint distribution of $(\mathbf{S}, T, \mathbf{X}, Y)$. For the ith observed quadruplet, $T_i \in \mathcal{T}$ and $\mathbf{S}_i \equiv (S_{i1}, S_{i2})^\top \in \Omega$ are the time and spatial location of the ith observation, $\mathbf{X}_i$ represents the observed explanatory variables, and Y_i is the response of interest. In particular, $\mathbf{X}_i$ and Y_i are observations at $(\mathbf{S}_i, T_i)$. For simplicity, we denote $\mathbf{X}_i \equiv \mathbf{X}_i(\mathbf{S}_i, T_i)$ and $Y_i \equiv Y_i(\mathbf{S}_i, T_i)$, unless emphasizing their spatiotemporal characteristics. Similarly, we use $\mathbf{x} \equiv \mathbf{x}(\mathbf{s}, t)$ and $y \equiv y(\mathbf{s}, t)$ unless otherwise stated.

We focus on the exponential dispersion family of distributions, including binomial, Poisson, and negative binomial, with a fixed number of parameters for modeling purposes. We assume that the conditional density of Y given $(\mathbf{S}, T, \mathbf{X}) = (\mathbf{s}, t, \mathbf{x})$ belongs to the exponential family $f_{Y|\mathbf{S},T,\mathbf{X}}(y|\mathbf{s}, t, \mathbf{x}) = \exp[y\xi(\mathbf{s}, t, \mathbf{x}) - \mathcal{B}\{\xi(\mathbf{s}, t, \mathbf{x})\} + \mathcal{C}(y)]$, for known functions $\mathcal{B}$ and $\mathcal{C}$, where ξ is the so-called natural parameter and is related to the unknown mean response by $\mu(\mathbf{s}, t, \mathbf{x}) = \mathrm{E}(Y|\mathbf{S} = \mathbf{s}, T = t, \mathbf{X} = \mathbf{x}) = \mathcal{B}'\{\xi(\mathbf{s}, t, \mathbf{x})\}$. In GSTVCM, $\mu(\mathbf{s}, t, \mathbf{x})$ is modeled via a link function g in the following form:

$$g\{\mu(\mathbf{s}, t, \mathbf{x})\} = \beta_0(\mathbf{s}, t) + \sum_{\ell=1}^{p} \beta_\ell(\mathbf{s}, t) x_\ell(\mathbf{s}, t) \tag{1}$$

where $\beta_0, \beta_1, \ldots, \beta_p$ are unknown trivariate functions, varying w.r.t. location $\mathbf{s}$ and time t, indicating the relationship between $\mathbf{X}$ and Y can vary along time and across different spatial locations.

Our work on GSTVCM draws inspiration from previous research on spatial varying coefficient models (SVCMs; see Kim and Wang 2021) and STVCMs. In the spatial regression context, Gelfand et al. (2003) introduced a Bayesian hierarchical SVCM that employs Gaussian processes to model coefficient functions. Another prominent approach is the geographically weighted regression (GWR) method (Fotheringham et al. 2002), which uses a weighted least squares approach to estimate the surface of the coefficient, with the bandwidth parameter determined through domain knowledge or cross-validation.

One of the main challenges with the (G)STVCM is that, while they account for spatiotemporal heterogeneity, they often sacrifice model parsimony. This increased complexity arises from the large number of parameters required to capture the varying effects across both space and time, making the models more difficult to interpret and manage. Furthermore, when dealing with limited sample sizes, traditional nonparametric models are prone to overfitting, fitting the noise in the data rather than the underlying trend. This overfitting leads to overly optimistic predictions that do not generalize well to new data. There have been several recent attempts to address this issue for VCMs by detecting whether coefficients are varying or constant.

Traditional spatiotemporal models have contributed extensively to capturing spatial and temporal heterogeneity. Bayesian hierarchical models, such as those introduced by Wikle et al. (1998), Stroud et al. (2001), Gelfand et al. (2003), and Paez et al. (2008), offer extensive flexibility by incorporating random effects, functional effects, and mixtures of spatially, temporally and spatiotemporally varying coefficients. These frameworks excel in handling non-Gaussian data and accounting for covariate measurement errors. However, they present significant challenges, including the careful specification of prior distributions and substantial computational costs for high-dimensional datasets due to iterative sampling techniques such as Markov Chain Monte Carlo. These limitations often preclude their application in real-time analysis or in scenarios requiring rapid computation. On the other hand, frequentist approaches, such as the geographically and temporally weighted regression (GTWR) method (Huang et al. 2010) provide a computationally more efficient alternative for modeling spatiotemporal dynamics. GTWR excels in estimating varying coefficients across space and time using weighted local regression; however, it often lacks the hierarchical and probabilistic flexibility of Bayesian methods.

In this paper, we propose a more efficient learning approach for Generalized SpatioTemporal Semi-varying Coefficient Models (GST-SVCMs) with structure identification, which addresses these gaps by adopting a frequentist perspective. Our proposed framework automatically identifies a parsimonious model structure by distinguishing between spatiotemporally varying and constant covariate effects. Once we correctly identify the varying and fixed effect of the covariates, the original model reduces to a partially varying coefficient form. This approach offers a balance between flexibility and simplicity, enabling researchers to capture complex relationships without overcomplicating the model structure.

The proposed GST-SVCM structure identification framework is designed to separate these components, reducing the number of parameters and enhancing the interpretability and parsimony of the model without compromising the ability to capture complex spatiotemporal dynamics. By avoiding unnecessary complexity, this approach mitigates risk overfitting, ensuring that the model remains robust even with limited data. Our proposed workflow is divided into two stages: structure identification and refitting. The structure identification phase is crucial as it sets the foundation for a more efficient and accurate model, which is then refined in the refitting stage.

Identifying constant coefficients is a crucial task, even in VCMs, and this has been extensively discussed by various research studies. Different methods have been proposed to distinguish constant coefficients from varying ones effectively; see (Noh et al. 2012; Wang and Kulasekera 2012; Lian et al. 2015, 2013; Chen et al. 2017; Li et al. 2015) and SVCM; see (Mu et al. 2020; Li et al. 2021). However, they all studied the VCM with a time index, spatial index, or other univariate indices. In contrast, our methodology is developed under the STVCM framework with both spatial and time indices, which requires more advanced tools to deal with the spatiotemporal index and corresponding irregular domain.

When data are collected over complex spatial domains, conventional nonparametric methods often suffer from the "leakage" problem (Ramsay 2002; Wood et al. 2008), which refers to poor inference performance when smoothing over boundaries. To address this issue, we utilize the tensor product spline on the triangular prismatic partitions (Yu et al. 2022) as described in Section 2.1. Compared to other kernel smoothing-based methods or traditional tensor product smoothing, our approach effectively handles the intricacies of irregular data distributed across complex domains. Furthermore, our method facilitates the convenient application of regularization techniques for model identification, which remains a challenge for adaptive or sequential smoothing approaches.

To achieve structure identification, we propose a penalized approach for model structure identification (i.e., determination of spatially varying vs. constant coefficients) followed by model estimation with identified sparse structure. Our proposed framework includes an automatic model identification method that balances flexibility and efficiency by considering both spatiotemporally varying and constant effects of various factors affecting response variables. This enables a more accurate understanding of the heterogeneity and dynamics of these effects, as it efficiently identifies constant and spatiotemporally varying components. To support this methodology, we establish theoretical guarantees for the consistency of the identified model structure. Additionally, we prove that the estimators of constant coefficients and varying coefficient functions are consistent, with the former exhibiting asymptotic normality, facilitating reliable statistical inference.

The contributions of this paper are threefold. First, we propose an estimation method for GST-SVCMs using tensor product splines over triangular prismatic partitions and demonstrate the theoretical properties of both constant and varying coefficients. To the best of our knowledge, this is the first work developed within the generalized spatiotemporal framework. Second, we introduce a structure identification method for GST-SVCMs, which enables automatic model selection through penalization, with theoretical guarantees. The practicality and effectiveness of GST-SVCM with structure identification are validated through extensive simulation studies. Third, we extend GST-SVCMs by developing the Hierarchical Spatiotemporal Varying Coefficient Model (HSTVCM), which further refines structural identification by distinguishing spatially varying, temporally varying, and fully spatiotemporal effects, enhancing both model interpretability and estimation accuracy.

The rest of the article is organized as follows. In Section 2, we introduce GST-SVCMs and our proposed estimation method. We also developed theories regarding the convergence of estimations and the asymptotic distribution of linear coefficients. In Section 3, we describe the penalized spline framework to identify the structure of a GST-SVCM using nonparametric approximation and present a theorem on the accuracy of model structure identification. Section 4 discusses the details of implementation, and Section 5 evaluates the performance of the proposed method through simulation studies. Section 6 illustrates how the proposed method can be extended to achieve a more granular model identification. In Section 7, we present our empirical analysis of the PM data. Finally, Section 8 provides concluding remarks. Proofs of the theorems, technical lemmas, and additional simulation studies are included in the Supporting Information.

2 | Estimation of GST-SVCMs

This section investigates GST-SVCMs, a class of semi-varying coefficient models where some explanatory variables have constant coefficients, while others have spatiotemporally varying coefficients. The GST-SVCM is defined as:

$$g\{\mu(\mathbf{s},t,\mathbf{x})\} \equiv \eta(\mathbf{s},t,\mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta},\mathcal{A}) = \beta_0(\mathbf{s},t) + \sum_{\ell\in\mathcal{A}^c}\alpha_\ell x_\ell(\mathbf{s},t) + \sum_{\ell\in\mathcal{A}}\beta_\ell(\mathbf{s},t)x_\ell(\mathbf{s},t) \tag{2}$$

where $(\mathbf{s},t)\in\Omega\times\mathcal{T}$, and $\mathcal{A}$ and $\mathcal{A}^c$ are the index sets, such that x_ℓ has spatiotemporally varying coefficient function β_ℓ, or only constant coefficients α_ℓ, respectively. If all coefficient functions are constants, that is, $\mathcal{A}=\emptyset$, model (2) reduces to a classical linear regression model. On the other hand, if all coefficients vary spatiotemporally, that is, $\mathcal{A}^c=\emptyset$, model (2) becomes a special case of GSTVCM in (1), representing the most complex form of the model.

To facilitate the discussion, we introduce the following notation. For a two-dimensional domain Ω and any function $f:\Omega\to\mathbb{R}$, its supremum norm is defined as $\|f\|_{\infty,\Omega}=\sup_{\mathbf{s}\in\Omega}|f(\mathbf{s})|$. We also define its semi-norm as $|f|_{k,\infty,\Omega}=\max_{i+j=k}\|\nabla^i_{s_1}D^j_{s_2}f(s_1,s_2)\|_{\infty,\Omega}$, where s_1 and s_2 denote the coordinates, and $\nabla^i_{s_1}$ represents the partial derivative of degree i in the direction of s_1. We consider the function space:

$$\mathcal{F}=\Big\{\eta(\mathbf{s},t,\mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta},\mathcal{A}):\boldsymbol{\alpha}\in\mathbb{R}^{|\mathcal{A}^c|};\ \beta_\ell(\mathbf{s},t)\in\mathbb{W}^{d+1}(\Omega)\otimes\mathbb{C}^{\varrho-2}(\mathcal{T}),\ell\in\{0\}\cup\mathcal{A}\Big\},$$

where $\mathbb{W}^{d+1}(\Omega)=\{f\in\Omega:|f|_{k,\infty,\Omega}<\infty, 0\le k\le d+1\}$ is the standard Sobolev space of bivariate functions over Ω with degree $d+1$; $\otimes$ denotes the tensor product; and $\mathbb{C}^{\varrho-2}(\mathcal{T})$ is the space consisting of all continuous univariate functions whose $(\varrho-2)$th order derivatives exist and are continuous over $\mathcal{T}$.

2.1 | Estimation Method

We start with the GST-SVCM estimation procedure under the true model structure, where the index sets $\mathcal{A}$ and $\mathcal{A}^c\}$ are known. The key idea is to approximate the varying coefficient function β_ℓ in (2) using tensor product splines over prismatic partitions (Yu et al. 2022) based on B-splines and Bernstein basis polynomials (Lai and Schumaker 2007), followed by the standard quasi-likelihood approach. We introduce these concepts in detail.

2.1.1 | B-Splines and Bernstein Basis Polynomials

For the time domain $\mathcal{T}=[t_1,t_2]$, we use univariate B-splines on $\mathcal{T}$ of degree ϱ with N_1 interior knots. In particular, we consider interior knots $\boldsymbol{\pi}=\{\pi_1,\ldots,\pi_{N_1}\}$, such that $t_1=\pi_{1-\varrho}=\cdots=\pi_0<\pi_1<\cdots\pi_{N_1}<\pi_{N_1+1}=\cdots=\pi_{N_1+\varrho}=t_2$. Polynomial splines of order ϱ are polynomial functions with $(\varrho-1)$-degree on subintervals $I_b=[\pi_b,\pi_{b+1})$, $b=0,\ldots,N_1-1$, and $I_{N_1}=[\pi_{N_1},\pi_{N_1+1}]$, and have $\varrho-2$ continuous derivatives globally. Let $\mathbb{U}^{\varrho}(\boldsymbol{\pi})$ stand for the space of such polynomial splines, whose bases can be formed as B-splines, which are denoted as $\mathbf{U}(t)=\{U_q(t),q\in\mathcal{N}\}^\top$, where $\mathcal{N}=\{1,\ldots,N_1+\varrho\}$ is the index set of the basis with cardinality $|\mathcal{N}|=N_1+\varrho$. Let $h_b=\pi_{b+1}-\pi_b$ be the distance between two adjoint knots and $h=\max_{0\le b\le N_1}h_b$ be the maximum distance. We have $h\asymp N_1^{-1}$ due to the constant length of $\mathcal{T}$.

For the spatial domain $\Omega\subset\mathbb{R}^2$, we approximate it using a triangulation $\triangle=\{\tau_j,1\le j\le N_2\}$, a collection of N_2 triangles such that any pair of triangles, τ_j and $\tau_{j'}$, either share an edge, a vertex, or do not intersect, and $\Omega=\cup_{j=1}^{N_2}\tau_j$. Given a triangulation $\triangle$ and degree $d>0$, we define a collection of *bivariate Bernstein-Bézier polynomials*, $\{B_m,m\in\mathcal{M}\}$, which form a basis for the function space of degree d and smoothness r, denoted $\mathbb{S}_d^r(\triangle)$. In particular, $\mathbb{S}_d^r(\triangle)$ is defined as $\mathbb{S}_d^r(\triangle)=\{g\in\mathbb{C}^r(\Omega),g|_\tau\in\mathbb{P}_d,\tau\in\triangle\}$, where $\mathbb{P}_d=\{f(s_1,s_2)=\sum_k c_k s_1^{a_k}s_2^{b_k},a_k+b_k=d,c_k\in\mathbb{R}\}$ is the set of homogeneous bivariate polynomials of degree d, $\mathbb{C}^r(\Omega)$ is the space of rth continuously differentiable functions, and the cardinality of $\mathcal{M}$ is $|\mathcal{M}|=N_2(d+2)(d+1)/2$. Therefore, for any function $g\in\mathbb{S}_d^r(\triangle)$, we can write its expansion as $g(\mathbf{s})=\sum_{m\in\mathcal{M}}\gamma_m B_m(\mathbf{s})$, with linear constraints; see Lai and Wang (2013). We denote the triangulation size of $\triangle$, defined as the longest edge of all triangles in $\triangle$, by $|\triangle|$. Due to the constant area of Ω, we have $|\triangle|\asymp N_2^{-1/2}$.

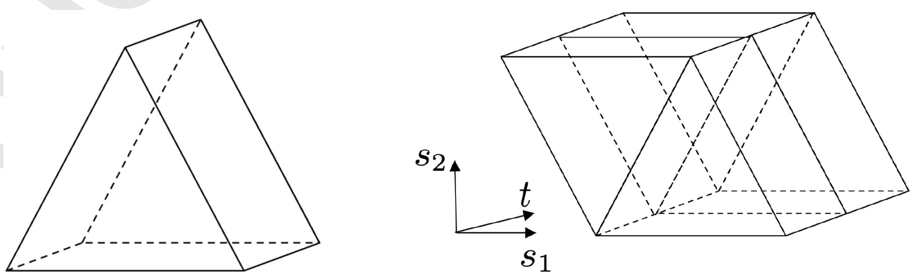
2.1.2 | Tensor Product Splines and Quasi-Likelihood Approach

For a spatiotemporal domain $\Omega\times\mathcal{T}$, we construct a triangular prismatic partition $\mathcal{E}$ as follows. First, we construct triangulation $\triangle$ with N_2 triangles on Ω, and interval partitions with N_1 interior knots $\boldsymbol{\pi}$ over $\mathcal{T}$. For a triangle $\tau_a\in\triangle$ and an interval I_b, $0\le b\le N_1$, we define their Cartesian product $\mathbb{A}_{a,b}=\tau_a\times I_b$ as a triangular prism, as illustrated in Figure 1a. Then we define a face-to-face triangular prismatic partition of $\Omega\times\mathcal{T}$, $\mathcal{E}=\{\mathbb{A}_{a,b}:1\le a\le N_2,0\le b\le N_1\}$, such that each pair of prisms either shares a common vertex, edge, or face or does not overlap. By construction, $\mathbb{A}_{a,b}$'s are right triangular prisms with six vertices, nine edges, and five faces; see Figure 1b.

Based on a triangular prismatic partition $\mathcal{E}$ of domain $\Omega\times\mathcal{T}$, we consider the function space: $\mathbb{T}^{(\varrho,d,r)}(\mathcal{E})=\{\sum_{q\in\mathcal{N}}\sum_{m\in\mathcal{M}}c_{q,m}U_q(t)B_m(\mathbf{s}):\mathcal{H}\mathbf{c}=\mathbf{0},$ for $\mathbf{c}=(c_{q,m},q\in\mathcal{N},m\in\mathcal{M})^\top\}$, where $\mathcal{H}$ is a constraint matrix to enforce smoothness conditions on the boundaries of each $\Delta_{a,b}$; see the supplementary material of Yu et al. (2020) for an example of $\mathcal{H}$. Next, we define the corresponding tensor product basis: $\boldsymbol{\psi}(\mathbf{s},t)=\{\psi_j(\mathbf{s},t),j\in\mathcal{J}\}=(U_1(t)B_1(\mathbf{s}),\ldots,U_{|\mathcal{N}|}(t)B_1(\mathbf{s}))^\top,U_1(t)B_2(\mathbf{s}),\ldots,U_{|\mathcal{N}|}(t)B_2(\mathbf{s}),\ldots,U_{|\mathcal{N}|}(t)B_{|\mathcal{M}|}(\mathbf{s}))^\top$. Given a model structure $\mathcal{A}$, we define the nonparametric approximation of any function $\beta(\mathbf{s},t)\in\mathbb{W}^{d+1}(\Omega)\otimes\mathbb{C}^{\varrho-2}(\mathcal{T})$ as $\beta_\ell(\mathbf{s},t)\approx\boldsymbol{\psi}(\mathbf{s},t)^\top\boldsymbol{\gamma}_\ell$, and denote $\eta(\mathbf{s},t,\mathbf{x};\boldsymbol{\alpha},\boldsymbol{\gamma},\mathcal{A})=\sum_{\ell\in\mathcal{A}^c}\alpha_\ell x_\ell+\sum_{\ell\in\{0\}\cup\mathcal{A}}\boldsymbol{\psi}(\mathbf{s},t)^\top\boldsymbol{\gamma}_\ell x_\ell$ as the nonparametric approximation of $\eta_0(\mathbf{s},t,\mathbf{x})$. Correspondingly, we define the approximation space:

$$\begin{aligned}\mathcal{G}=\Big\{&\eta(\mathbf{s},t,\mathbf{x};\boldsymbol{\alpha},\boldsymbol{\gamma},\mathcal{A}):\boldsymbol{\alpha}\in\mathbb{R}^{|\mathcal{A}^c|};\\ &\beta_\ell(\mathbf{s},t)=\boldsymbol{\psi}(\mathbf{s},t)^\top\boldsymbol{\gamma}\in\mathbb{T}^{(\varrho,d,r)}(\mathcal{E}),\ell\in\{0\}\cup\mathcal{A}\Big\}\end{aligned}\tag{3}$$

To address the constraint on $\mathcal{H}$, we employ a QR decomposition. Specifically, by the QR decomposition of $\mathcal{H}$, $\mathcal{H}^\top=(\mathcal{Q}_1,\mathcal{Q}_2)\begin{pmatrix}\mathcal{R}_1\\ \mathbf{0}\end{pmatrix}$, where $(\mathcal{Q}_1,\mathcal{Q}_2)$ is an orthogonal matrix and $\mathcal{R}_1$ is a full rank matrix with the same rank as $\mathcal{H}$. Applying the reparameterization $\boldsymbol{\gamma}_\ell=\mathcal{Q}_2\boldsymbol{\gamma}_\ell^*$, $\boldsymbol{\psi}_\ell^*(\mathbf{s},t)=\mathcal{Q}_2^\top\boldsymbol{\psi}_\ell(\mathbf{s},t)$, the constraint $\mathcal{H}\boldsymbol{\gamma}_\ell=\mathbf{0}$ would be automatically satisfied.



(a) (b)

FIGURE 1 | Example of (a) one triangular prism element $\mathbb{A}_{a,b}$ and (b) a triangular prismatic partition $\mathcal{E}$. (a) $\mathbb{A}_{a,b}$, (b) $\mathcal{E}=\cup_{a,b}\mathbb{A}_{a,b}$.

If $\mathrm{var}(Y) = \sigma^2 V\{g^{-1}(\eta(\mathbf{S}, T, \mathbf{X}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathcal{A})\}$ for some known positive function V, and σ^2 is a dispersion parameter, then estimation of the mean can be achieved by replacing the conditional log-likelihood function $\log\{f_{Y|\mathbf{S},T,\mathbf{X}}(y|\mathbf{s}, t, \mathbf{x})\}$ with a quasi-likelihood function $Q\{g^{-1}(\eta), y\}$, which satisfies $\nabla_\mu Q(\mu, y) = (y-\mu)/\{\sigma^2 V(\mu)\}$.

Now we define the penalized negative log quasi-likelihood $L_{n,\mathcal{A}}$:

$$
\begin{aligned}
L_{n,\mathcal{A}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = &-\frac{1}{n}\sum_{i=1}^{n} Q\left[g^{-1}\{\eta(\mathbf{S}_i, T_i, \mathbf{X}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathcal{A})\}, Y_i\right] \\
&+ \sum_{\ell \in \mathcal{A}} \{\lambda_{1,\ell} f_1(\beta_\ell) + \lambda_{2,\ell} f_2(\beta_\ell)\}
\end{aligned} \tag{4}
$$

where $f_1(\beta_\ell) = \int_{\Omega\times\mathcal{T}} (\nabla_t^2 \beta_\ell)^2 ds_1 ds_2 dt$ and $f_2(\beta_\ell) = \int_{\Omega\times\mathcal{T}} \left\{(\nabla_{s_1}^2 \beta_\ell)^2 + (\nabla_{s_2}^2 \beta_\ell)^2\right\} ds_1 ds_2 dt$ are functions measuring the roughness of β_ℓ w.r.t. time and space respectively (Yu et al. 2022); $\nabla_{s_j}^q$ is the qth partial derivative in the direction s_j, $j = 1, 2$; ∇_t^q is the qth derivative w.r.t. t; $\lambda_{1,\ell}$ and $\lambda_{2,\ell}$ are the penalty parameters controlling the smoothness of β_ℓ w.r.t. t and $\mathbf{s}$, respectively.

Note that $f_1(\beta_\ell) = \boldsymbol{\gamma}_\ell^\top \mathbf{P}_1 \boldsymbol{\gamma}_\ell$ and $f_2(\beta_\ell) = \boldsymbol{\gamma}_\ell^\top \mathbf{P}_2 \boldsymbol{\gamma}_\ell$, where $\mathbf{P}_1$ and $\mathbf{P}_2$ are matrices that store the second-order derivatives of a tensor product spline function $\sum_{j\in\mathcal{J}} \psi_j(s_1, s_2, t)\gamma_j$ w.r.t. t and $\mathbf{s}$, respectively. Specifically, for $\boldsymbol{\gamma} = \{\gamma_j\}_{j\in\mathcal{J}}$, $f_1(\sum_{j\in\mathcal{J}} \psi_j \gamma_j) = \boldsymbol{\gamma}^\top \mathbf{P}_1 \boldsymbol{\gamma} = \boldsymbol{\gamma}^\top \mathbf{P}_U \otimes \mathbf{M}_B \boldsymbol{\gamma}$ and $f_2(\sum_{j\in\mathcal{J}} \psi_j \gamma_j) = \boldsymbol{\gamma}^\top \mathbf{P}_2 \boldsymbol{\gamma} = \boldsymbol{\gamma}^\top \mathbf{M}_U \otimes \mathbf{P}_B \boldsymbol{\gamma}$, where $\mathbf{M}_U$ and $\mathbf{P}_U$ are $|\mathcal{N}| \times |\mathcal{N}|$ matrices with $(\mathbf{M}_U)_{q,q'} = \int_\mathcal{T} U_q(t) U_{q'}(t) dt$ and $(\mathbf{P}_U)_{q,q'} = \int_\mathcal{T} \nabla_t^2 U_q(t) \nabla_t^2 U_{q'}(t) dt$, and $\mathbf{M}_B$ and $\mathbf{P}_B$ are $|\mathcal{M}| \times |\mathcal{M}|$ matrices with

$$
\begin{aligned}
(\mathbf{M}_B)_{m,m'} &= \int_\Omega B_m(s_1, s_2) B_{m'}(s_1, s_2) ds_1 ds_2, \\
(\mathbf{P}_B)_{m,m'} &= \int_\Omega \Big\{ \nabla_{s_1}^2 B_m(s_1, s_2) \nabla_{s_1}^2 B_{m'}(s_1, s_2) \\
&\quad + \nabla_{s_2}^2 B_m(s_1, s_2) \nabla_{s_2}^2 B_{m'}(s_1, s_2) \Big\} ds_1 ds_2.
\end{aligned}
$$

Thus, (4) can be simplified as an unconstrained minimization problem:

$$
(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}}^*) = \operatorname*{arg\,min}_{\boldsymbol{\alpha}, \boldsymbol{\gamma}^*} L_{n,\mathcal{A}}(\boldsymbol{\alpha}, \boldsymbol{\gamma}^*) \tag{5}
$$

where

$$
\begin{aligned}
&L_{n,\mathcal{A}}(\boldsymbol{\alpha}, \boldsymbol{\gamma}^*) \\
&= -\frac{1}{n}\sum_{i=1}^{n} Q\left[g^{-1}\left\{ \sum_{\ell\in\mathcal{A}^c} \alpha_\ell X_{i\ell} \right.\right. \\
&\quad \left.\left. + \sum_{\ell\in\{0\}\cup\mathcal{A}} X_{i\ell} \boldsymbol{\psi}^*(\mathbf{S}_i, T_i)^\top \boldsymbol{\gamma}_\ell^* \right\}, Y_i \right] \\
&\quad + \sum_{\ell\in\mathcal{A}} \lambda_{1,\ell} \boldsymbol{\gamma}_\ell^{*\top} \mathcal{Q}_2^\top \mathbf{P}_1 \mathcal{Q}_2 \boldsymbol{\gamma}_\ell^* \\
&\quad + \sum_{\ell\in\mathcal{A}} \lambda_{2,\ell} \boldsymbol{\gamma}_\ell^{*\top} \mathcal{Q}_2^\top \mathbf{P}_2 \mathcal{Q}_2 \boldsymbol{\gamma}_\ell^*.
\end{aligned}
$$

2.2 | Theoretical Properties

This section presents the asymptotic properties of the proposed estimators for the components of the GST-SVCM with the true model structure $\mathcal{A}_0$.

For real-valued vectors $\mathbf{v}, \mathbf{v}_1, \mathbf{v}_2 \in \mathbb{R}^p$, we define the inner product as $\langle \mathbf{v}_1, \mathbf{v}_2 \rangle = \mathbf{v}_1^\top \mathbf{v}_2$, and the Euclidean norm as $\|\mathbf{v}\|_2 = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle} = \sqrt{\sum_{\ell=1}^p v_\ell^2}$. We denote $|\mathbf{v}| = \max_{1\le\ell\le p} v_\ell$ as the vector supremum norm. For any two functions $f_1, f_2 \in \mathcal{F}$, we consider the theoretical inner product $\langle f_1, f_2 \rangle = \mathrm{E}\{f_1(\mathbf{S}, T, \mathbf{X}) f_2(\mathbf{S}, T, \mathbf{X})\}$, and the empirical inner product $\langle f_1, f_2 \rangle_n = n^{-1}\sum_{i=1}^n \{f_1(\mathbf{S}_i, T_i, \mathbf{X}_i) f_2(\mathbf{S}_i, T_i, \mathbf{X}_i)\}$. Consequently, for any function $f \in \mathcal{F}$, the theoretical norm is $\|f\|^2 = \langle f, f \rangle$, and the empirical norm is $\|f\|_n^2 = \langle f, f \rangle_n$. Let $q_1(\eta, y)$ and $q_2(\eta, y)$ be the first and second partial derivatives of the quasi-likelihood function $Q\{g^{-1}(\eta), y\}$, respectively, where $q_1(\eta, y) = \frac{\partial}{\partial\eta} Q\{g^{-1}(\eta), y\} = \{y - g^{-1}(\eta)\}\rho_1(\eta)$, $q_2(\eta, y) = \frac{\partial^2}{\partial\eta^2} Q\{g^{-1}(\eta), y\} = \{y - g^{-1}(\eta)\}\rho_1'(\eta) - \rho_2(\eta)$, and $\rho_j(\eta) = \{\frac{\partial}{\partial\eta} g^{-1}(\eta)\}^j / [\sigma^2 V\{g^{-1}(\eta)\}]$, $j = 1, 2$.

Throughout the rest of the paper, we use the subscript "0" to denote the underlying true parameter and space; for example, $\mathcal{A}_0$, $\boldsymbol{\alpha}_0 = (\alpha_{0\ell}, \ell \in \mathcal{A}_0^c)^\top$, and $\boldsymbol{\beta}_0 = (\beta_{00}, \{\beta_{0\ell}\}_{\ell\in\mathcal{A}_0})^\top$. We define $\mathcal{F}_0$ and $\mathcal{G}_0$ similarly to $\mathcal{F}$ and $\mathcal{G}$, respectively, by assigning $\mathcal{A} = \mathcal{A}_0$. We denote the oracle estimator as $\boldsymbol{\theta}_0 = (\boldsymbol{\alpha}_0^\top, \boldsymbol{\gamma}_0^{*\top})^\top$ and $\tilde{\eta}(\mathbf{s}, t, \mathbf{x}; \boldsymbol{\alpha}_0, \boldsymbol{\gamma}_0^*, \mathcal{A}_0) = \mathbf{x}_{\mathcal{A}_0}^\top \boldsymbol{\alpha}_0 + \sum_{\ell\in\mathcal{A}_0} \boldsymbol{\psi}_\ell^*(\mathbf{s}, t)^\top \boldsymbol{\gamma}_{0\ell}^*$. We define the random noise $\varepsilon_i = \varepsilon(\mathbf{S}_i, T_i, \mathbf{X}_i) = Y_i - g^{-1}(\eta_0(\mathbf{S}_i, T_i, \mathbf{X}_i))$ as the deviance of Y_i from the true mean.

In the following theoretical analysis, we adopt an infill asymptotic framework, where the number of observations increases within a fixed domain. We first state the technical assumptions.

- (A1) For $\ell \in \mathcal{A}_0$, $\beta_{0\ell} \in \mathbb{W}^{d+1,\infty}(\Omega) \times \mathbb{C}^{\varrho-2}(\mathcal{T})$.
- (A2) The density function $f(\mathbf{s}, t)$ of $(\mathbf{S}, T)$ is bounded away from zero and infinity on $\Omega \times \mathcal{T}$.
- (A3) The function $q_2(x, y) < 0$, $c_1 < |q_2(x, y)| < C_1$ and $c_2 < |\frac{\partial}{\partial x} q_2(x, y)| < C_2$ for $x \in \mathbb{R}$ and y in the range of the response variable. The functions $V(\cdot)$, $g^{-1}(\cdot)$, the first-order derivative of $g^{-1}(\cdot)$ are continuous, and there exist positive constants c_ρ and C_ρ such that $c_\rho \le \rho_2(\cdot) \le C_\rho$. For each $(\mathbf{s}, t, \mathbf{x})$, $\mathrm{Var}(Y|\mathbf{S} = \mathbf{s}, T = t, \mathbf{X} = \mathbf{x})$ and $g'(\mu(\mathbf{s}, t, \mathbf{x}))$ are nonzero.
- (A4) $\{(\mathbf{S}_i, T_i, \mathbf{X}_i, Y_i, \varepsilon_i)\}_{i=1}^n$ are independently and identically distributed, where the errors satisfy $\mathrm{E}\{\varepsilon_i | \mathbf{S}_i = \mathbf{s}, T_i = t, \mathbf{X}_i = \mathbf{x}\} = 0$ and $\mathrm{E}(|\varepsilon_i|^{2+\delta} | \mathbf{S}_i = \mathbf{s}, T_i = t, \mathbf{X}_i = \mathbf{x}) < \infty$ for some $\delta \in (1/2, \infty)$.
- (A5) For any $\ell = 1, \ldots, p$, there exists a positive constant C_ℓ such that $|X_\ell| \le C_\ell$; Denote $\mathbf{Q}(\mathbf{s}, t) = \mathrm{E}\left\{(1, \mathbf{X}^\top)^\top (1, \mathbf{X}^\top) \middle| \mathbf{S} = \mathbf{s}, T = t\right\}$. The eigenvalues of $\mathbf{Q}(\mathbf{s}, t)$ are bounded away from 0 and infinity uniformly for all $\mathbf{s} \in \Omega$ for all $(\mathbf{s}, t) \in \Omega \times \mathcal{T}$.
- (A6) Assume there exists some constant $0 < c < C < \infty$, such that $c \le \max_b h_b / \min_b h_b \le C$.
- (A7) The triangulation $\triangle$ is π-quasi-uniform, that is, there exists a positive constant π such that $|\triangle|/r_\triangle \le \pi$, where $|\triangle| = \max\{|\tau|, \text{for any triangle } \tau \in \triangle\}$ and $r_\triangle = \min\{r_\tau, \tau \in \triangle\}$. Note here $|\tau|$ is the length of the longest edge of triangle τ, and r_τ is the radius of the largest disk that can be inscribed in triangle τ.

(A8) The size of the triangular prismatic partition $\triangle$ and h satisfy $h^{-1/2}|\triangle|^{-1}\left(\frac{\log n}{n}\right)^{1/2} \to 0, (h^{\varrho} + |\triangle|^{d+1}) \to 0, h^{-5/2} |\triangle|^{-5} \log n/n^{1/2} \to 0, \quad h^{2\varrho-3/2}|\triangle|^{-3}n^{1/2} \to 0, \quad h^{-3/2} |\triangle|^{2d-1}n^{1/2} \to 0$. The roughness penalty parameter vectors Λ_1 and Λ_2 satisfy $|\Lambda_1|h^{-2} \to 0$, $|\Lambda_2||\triangle|^{-4} \to 0$, $|\Lambda_1|^2|\triangle|^{-3}h^{-11/2}n^{1/2} \to 0$, $|\Lambda_1|^2|\triangle|^{-11}h^{-3/2}n^{1/2} \to 0$.

(A9) The matrix $\boldsymbol{\Sigma} = \mathrm{E}\left[\rho_2\{\eta^0(\mathbf{S},T,\mathbf{X})\}\widetilde{\mathbb{D}}^{\mathrm{c}}(\mathbf{S},T,\mathbf{X})\widetilde{\mathbb{D}}^{\mathrm{c}}(\mathbf{S},T,\mathbf{X})^{\top}\right]$ is positive definite, where $\widetilde{\mathbb{D}}^{\mathrm{c}}(\mathbf{S},T,\mathbf{X}) = \mathbb{D}^{\mathrm{c}}(\mathbf{X}_{\mathcal{A}_0^{\mathrm{c}}}) - \Gamma(\mathbf{S},T,\mathbf{X}_{\mathcal{A}_0})$, and $\Gamma(\mathbf{s},t,\mathbf{x}_{\mathcal{A}_0})$ is a projection matrix in (8).

(A10) As $n \to \infty$, $\omega_n \to 0$, $r_n/\omega_n \to 0$, where r_n is the L_2 convergence rate of oracle estimator $\hat{\theta}$ defined in Theorem S1 in the Supporting Information.

Remark 1. These are mild and reasonable assumptions that can be satisfied in many practical situations. Assumption (A1) indicates the true coefficient functions $\beta_{0\ell}, \ell \in \mathcal{A}_0$ are reasonably smooth; see Lai and Wang (2013). Assumption (A2) ensures that realized observations of $(\mathbf{S}, T)$ are randomly scattered within $\Omega \times \mathcal{T}$. Assumption (A3) lists conditions that enable the development of convergence and asymptotic normality under quasi-likelihood framework. Assumption (A4) is a regularity condition for regression. Assumption (A5) ensures there is no multicollinearity among covariates. Assumptions (A6) and (A7) of h and $\triangle$ are common assumptions in the spline approximation literature. Assumption (A8) lists the requirement for triangular prismatic partition through h and $\triangle$, as well as the roughness penalty parameters Λ_1 and Λ_2. Assumption (A9) ensures $\widetilde{\mathbb{D}}^{\mathrm{c}}$ and $\mathbb{D}^{\mathrm{v}}$ are functionally unrelated, contributing to the asymptotic normality of linear coefficients $\hat{\boldsymbol{\alpha}}$ in Theorem 2. Assumption (A10) facilitates the consistency of structure identification in Theorem S1.

Theorem 1 (Convergence rate of $\hat{\theta}$). *Assume Assumptions (A1) to (A8) in the Supporting Information hold, under the true structure $\mathcal{A}_0$, then*

$$\|\hat{\theta} - \theta_0\|_2 = O_p(r_n), \quad \|\hat{\eta} - \eta_0\| = O_p(r_n h^{-1/2}|\triangle|^{-1}) \tag{6}$$

where $r_n = h^{-1/2}|\triangle|^{-1}(\log n/n)^{1/2} + (h^{\varrho} + |\triangle|^{d+1}) + |\Lambda_1|h^{-2} + |\Lambda_2||\triangle|^{-4}$, *and* $|\Lambda_b| = \max_{\ell} \lambda_{b,\ell}$, $b = 1,2$ *represent the maximum roughness penalty parameter*.

Theorem 1 establishes the convergence of the GST-SVCM estimators. The convergence rate is determined by the fineness of the triangular prismatic partition ($|\triangle|$ and h), the number of observations (n), the degree of bivariate spline (d), and the order of the univariate spline (ϱ).

Next, Theorem 2 demonstrates that the constant coefficients $\hat{\alpha}_{\ell}, \ell \in \mathcal{A}_0^{\mathrm{c}}$ follow a normal distribution asymptotically.

Theorem 2 (Asymptotic Normality of $\hat{\boldsymbol{\alpha}}$). *When $\mathcal{A}_0$ is known, under Assumptions (A1) to (A9) in the Supporting Information, the constant coefficients $\hat{\boldsymbol{\alpha}}$ in the refitting process satisfy that* $\sqrt{n}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}^{-1})$, *where* $\boldsymbol{\Sigma} = \mathrm{E}\left[\rho_2\{\eta^0(\mathbf{S},T,\mathbf{X})\}\widetilde{\mathbb{D}}^{\mathrm{c}}(\mathbf{S},T,\mathbf{X})\widetilde{\mathbb{D}}^{\mathrm{c}}(\mathbf{S},T,\mathbf{X})^{\top}\right]$, *and*

$$\begin{aligned}
\mathbb{D}_{\mathcal{A}_0}^{\mathrm{v}\top} &= \{(1, \mathbf{X}_{i,\mathcal{A}_0}^{\top})^{\top} \otimes \psi^*(\mathbf{S}_i, T_i)\}_{i=1}^n \\
\mathbb{D}_{\mathcal{A}_0}^{\mathrm{c}\top} &= (X_{i\ell}, 1 \le i \le n, \ell \in \mathcal{A}_0^{\mathrm{c}})
\end{aligned} \tag{7}$$

$$\begin{aligned}
\widetilde{\mathbb{D}}_{\mathcal{A}_0}^{\mathrm{c}}(\mathbf{S},T,\mathbf{X}) &= \mathbb{D}^{\mathrm{c}}(\mathbf{X}_{\mathcal{A}_0^{\mathrm{c}}}) - \Gamma(\mathbf{S},T,\mathbf{X}_{\mathcal{A}_0}) \\
\Gamma^{\top}(\mathbf{S},T,\mathbf{X}_{\mathcal{A}_0}) &= \mathbb{D}_{\mathcal{A}_0}^{\mathrm{v}\top}\mathrm{E}\{\mathbb{D}_{\mathcal{A}_0}^{\mathrm{v}}\mathbb{D}_{\mathcal{A}_0}^{\mathrm{v}\top}\}^{-1}\mathrm{E}\{\mathbb{D}_{\mathcal{A}_0}^{\mathrm{v}}\mathbb{D}_{\mathcal{A}_0}^{\mathrm{c}\top}\}
\end{aligned} \tag{8}$$

Theorem 2 establishes the asymptotic normality of $\hat{\boldsymbol{\alpha}}$, with mean zero and a covariance matrix related to the covariance of $\{X_{\ell}, \ell \in \mathcal{A}_0^{\mathrm{c}}\}$ that is orthogonal to the space spanned by $\{X_{\ell}, \ell \in \mathcal{A}_0\}$.

3 | GST-SVCM With Structure Identification

In this section, we propose an automatic structure identification method for GST-SVCMs. We address the challenge of identifying the unknown true model structure $\mathcal{A}_0$ in (2) by formulating the structure identification as a penalized quasi-likelihood problem, followed by reestimating the model in tensor product splines as described in Section 2.1.

3.1 | Structure Identification Methods

To identify $\mathcal{A}_0$, we employ a regularization approach to detect the varying signal of β_{ℓ}. We first decompose $\beta_{\ell}(\mathbf{s},t) = \alpha_{\ell} + \beta_{\ell}^{v}(\mathbf{s},t)$, where β_{ℓ} is divided into a possibly nonzero constant α_{ℓ} and a centered varying coefficient function $\beta_{\ell}^{v}(\mathbf{s},t)$. Correspondingly, we decompose the approximation space $\mathcal{G}$ into two parts:

$$\begin{aligned}
\mathcal{G}_c &= \{\eta(\mathbf{s},t,\mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta},\mathcal{A}) : \boldsymbol{\alpha} \in \mathbb{R}^{|\mathcal{A}^{\mathrm{c}}|}; \beta_{\ell}(\mathbf{s},t) \equiv 0\}, \\
\mathcal{G}_v &= \big\{\eta(\mathbf{s},t,\mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta},\mathcal{A}) : \boldsymbol{\alpha} \in \mathbb{R}^{|\mathcal{A}^{\mathrm{c}}|}; \\
&\qquad \beta_{\ell}(\mathbf{s},t) \in \mathbb{T}_v^{(\varrho,d,r)}(\mathcal{E}), \ell \in \{0\} \cup \mathcal{A}\big\},
\end{aligned}$$

where $\mathbb{T}_v^{(\varrho,d,r)}(\mathcal{E}) = \{f \in \mathbb{T}^{(\varrho,d,r)}(\mathcal{E}) : \mathrm{E}(f) = 0\}$ represents the space spanned by standardized tensor product spline basis $\psi^N(\mathbf{s},t)$, constructed as follows

$$\begin{aligned}
\psi_j^N(\mathbf{s},t) &= \frac{\psi_j^0(\mathbf{s},t)}{\sqrt{\mathrm{E}\{\psi_j^0(\mathbf{S},T)^2\}}} \\
\psi_j^0(\mathbf{s},t) &= \psi_j(\mathbf{s},t) - \mathrm{E}\{\psi_j(\mathbf{S},T)\}, \quad j \in \mathcal{J}
\end{aligned} \tag{9}$$

By construction, $\mathrm{E}\{\psi_j^N(\mathbf{S},T)\} = 0$ and $\mathrm{E}\{\psi_j^N(\mathbf{S},T)\}^2 = 1$, for $j \in \mathcal{J}$. For simplicity, we use ψ_j instead of ψ_j^N for the remainder of the paper. We use the superscript "I" to represent the identification of the structure.

Next, we assume that all covariates have spatiotemporally varying coefficients $\alpha_{\ell} + \beta_{\ell}(\mathbf{s},t), 1 \le \ell \le p$, and that the intercept $\beta_0(\mathbf{s},t)$ is spatiotemporally varying. For model identifiability, we enforce the constraint $\mathrm{E}\{\beta_{\ell}(\mathbf{S},T) = 0\}, 1 \le \ell \le p$. Thus, we arrive at the working model for η,

$$\eta^{\mathrm{I}}(\mathbf{s},t,\mathbf{x};\boldsymbol{\alpha},\boldsymbol{\beta}) = \beta_0(\mathbf{s},t) + \sum_{1 \le \ell \le p}\{\alpha_{\ell} + \beta_{\ell}(\mathbf{s},t)\}x_{\ell}(\mathbf{s},t). \tag{10}$$

If $\ell \in \mathcal{A}_0$, we expect β_{ℓ} to significantly differ from a zero constant function. Otherwise, if $\ell \in \mathcal{A}_0^{\mathrm{c}}$, we expect β_{ℓ} to be negligible and potentially penalized to zero during training.

Our method for model structure identification minimizes the penalized negative log quasi-likelihood:

$$L_n^{\mathrm{I}}(\boldsymbol{\alpha},\boldsymbol{\beta}) = -\frac{1}{n}\sum_{i=1}^{n} Q\left[g^{-1}\{\eta^{\mathrm{I}}(\mathbf{S}_i, T_i, \mathbf{X}_i; \boldsymbol{\alpha}, \boldsymbol{\beta})\}, Y_i\right] + \sum_{\ell=1}^{p} p_{\omega_n}(\|\beta_\ell\|_n) + \sum_{\ell=1}^{p}\{\lambda_{1,\ell} f_1(\beta_\ell) + \lambda_{2,\ell} f_2(\beta_\ell)\},$$

where p_{ω_n} is the group SCAD penalty (Xue 2009) to identify whether β_ℓ is in the true model, satisfying $p_{\omega_n}(0) = 0$ and

$$p'_{\omega_n}(b) = \omega_n\left\{ I(b \le \omega_n) + \frac{(a\omega_n - b)_+}{(a-1)\omega_n} I(b > \omega_n)\right\},$$

for some $a > 2$ and $b > 0$.

Here, ω_n is the tuning parameter that controls the complexity of the selected model and a is a fixed constant. Let

$$(\widehat{\boldsymbol{\alpha}}^{\mathrm{I}}, \widehat{\boldsymbol{\beta}}^{\mathrm{I}}) = \underset{\alpha_\ell \in \mathbb{R}, \beta_\ell \in \mathcal{G}_v, 1\le \ell \le p}{\arg\min}\ L_n^{\mathrm{I}}(\boldsymbol{\alpha},\boldsymbol{\beta}) \tag{11}$$

The estimated model structure is then:

$$\widehat{\mathcal{A}} = \left\{\ell = 1, \ldots, p : \widehat{\beta}_\ell^{\mathrm{I}} \not\equiv 0\right\}, \quad \widehat{\mathcal{A}}^c = \left\{\ell = 1, \ldots, p : \widehat{\beta}_\ell^{\mathrm{I}} \equiv 0\right\} \tag{12}$$

By approximating $\beta(\cdot,\cdot)$ using tensor product splines, we can reformulate the structure identification step of GST-SVCMs defined in (10–12) as follows:

$$\eta^{\mathrm{I}}(\mathbf{s}, t, \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\gamma}) = \sum_{1\le \ell \le p}\{\alpha_\ell + \boldsymbol{\psi}_\ell(\mathbf{s}, t)^{*\top}\boldsymbol{\gamma}_\ell^*\} x_\ell$$

$$L_n^{\mathrm{I}}(\boldsymbol{\alpha}, \boldsymbol{\gamma}^*) = -\frac{1}{n}\sum_{i=1}^{n} Q\left[g^{-1}\left\{\sum_{1\le \ell\le p}\alpha_\ell X_{i\ell} + \sum_{0\le \ell \le p} X_{i\ell}\boldsymbol{\psi}^*(\mathbf{S}_i, T_i)^\top \boldsymbol{\gamma}_\ell^*\right\}, Y_i\right] + \sum_{\ell=1}^{p} p_{\omega_n}(\|\boldsymbol{\gamma}_\ell^*\|) + \sum_{\ell\in\mathcal{A}} \lambda_{1,\ell}\boldsymbol{\gamma}_\ell^{*\top} Q_2^\top \mathbf{P}_1 Q_2 \boldsymbol{\gamma}_\ell^* + \sum_{\ell\in\mathcal{A}} \lambda_{2,\ell}\boldsymbol{\gamma}_\ell^{*\top} Q_2^\top \mathbf{P}_2 Q_2 \boldsymbol{\gamma}_\ell^* \tag{13}$$

$$\widehat{\mathcal{A}} = \left\{\ell = 1, \ldots, p : \widehat{\boldsymbol{\gamma}}_\ell^{\mathrm{I}} \not\equiv \mathbf{0}\right\}$$

$$\widehat{\mathcal{A}}^c = \left\{\ell = 1, \ldots, p : \widehat{\boldsymbol{\gamma}}_\ell^{\mathrm{I}} \equiv \mathbf{0}\right\} \tag{14}$$

Remark 2. The definition of L_n^{I} and $L_{n,\mathcal{A}}$ differs in whether the model structure $\mathcal{A}$ is known. The former assumes a varying coefficient function for all covariates X_ℓ, while the latter assumes that only $\{X_\ell, \ell \in \mathcal{A}\}$ has a varying coefficient function and abandons the penalty for structure identification.

3.2 | Theoretical Properties

We illustrate in Theorem 3 that with a proper choice of penalty parameter λ, the model structure will be correctly identified in probability.

Theorem 3 (Structure Identification). *Under Assumptions (A1)–(A8), (A10), as $n \to \infty$, $P(\widehat{\mathcal{A}} = \mathcal{A}_0) \to 1$ and $P(\widehat{\mathcal{A}}^c = \mathcal{A}_0^c) \to 1$.*

Theorem 3 shows that as the number of observations n increases, the probability of accurately identifying the unknown model structure converges to one. Therefore, for spatiotemporal data with an unknown heterogeneity structure, we can first perform the model identification outlined in Section 3 followed by the model estimation outlined in Section 2. This provides a comprehensive framework for handling GST-SVCM with an unknown model structure.

4 | Implementation

In this section, we discuss the practical implementation of the estimation and structure identification of GST-SVCMs. Without loss of generality and robustness, we divide the entire procedure into two stages: identification and refitting, as detailed in Algorithm 1.

In the identification stage, the model structure is determined using the group SCAD penalty with a relatively coarse triangulation and lower spline degrees. This step is crucial for accurately distinguishing between spatiotemporally varying and constant covariate effects, thereby identifying a parsimonious model structure. Once the structure is identified, the refitting stage estimates the model with this identified structure using a finer triangulation and higher spline degrees to obtain more accurate parameter estimates. Refitting is necessary because the coarse triangulation and lower spline degrees used in the identification stage, while reducing the number of parameters to estimate, can introduce bias in the coefficient estimates. By refitting with a refined model structure, we mitigate this bias, enhancing both reliability and precision. This two-stage approach ensures an optimal balance between model complexity and interpretability while maintaining robustness and generality.

Group SCAD penalty parameters for tensor product splines. The structure identification of GST-SVCMs incorporates several tuning parameters that require careful consideration. In Stage 1, we employ the group SCAD penalty parameter ω_n to regulate the sparseness of β_ℓ, which is implemented using the GRPREG package in R. To select the optimal penalty parameter ω_n's, we use the Extended Bayesian Information Criterion (EBIC) proposed by Chen and Chen (2008) as follows:

$$\mathrm{EBIC}(\omega_n|\widehat{\theta}) = -2L_n(\widehat{\theta}, \widehat{Y}_i) + [\widehat{\theta}]\log(n) + 2\log\binom{|\widehat{\theta}|}{[\widehat{\theta}]}, \tag{15}$$

where $L_n(\widehat{\theta}, \widehat{Y}_i) = \sum_{i=1}^{n} Q\left[g^{-1}\{\eta^{\mathrm{I}}(\mathbf{S}_i, T_i, \mathbf{X}_i; \widehat{\theta})\}, \widehat{Y}_i\right]$ is the log quasilikelihood; $\widehat{Y}_i$ is estimated observation using penalized estimator that minimizes (14); $[\theta]$ denotes the number of nonzero entries of θ, and $|\theta|$ denotes the length of θ. The optimal ω_n is selected to minimize $\mathrm{EBIC}(\omega_n|\widehat{\theta})$.

Roughness penalty parameters $\lambda_{1,\ell}, \lambda_{2,\ell}$. During the identification stage, we omit the roughness penalty terms $\lambda_{1,\ell} f_1(\beta_\ell)$ and $\lambda_{2,\ell} f_2(\beta_\ell)$, for $\ell = 1, \ldots, p$, in the penalized least squares

ALGORITHM 1 | The estimation and structure identification procedure for GST-SVCMs.

Input: Dataset $\mathbf{O} = \left\{(\mathbf{S}_i, T_i, \mathbf{X}_i, Y_i)\right\}_{i=1}^{n}$ for both model identification and refitting.
Output: Estimated of model structure $\widehat{\mathcal{A}}, \widehat{\mathcal{A}}^{\mathrm{c}}$; estimated parameters $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\alpha}}^{\top}, \widehat{\boldsymbol{\gamma}}^{*\top})^{\top}$.
Stage 1: Identification.
  Initialize with full model structure: $\mathcal{A} = \{1, \dots, p\}, \mathcal{A}^{\mathrm{c}} = \emptyset$.
  Standardize $\{Y_i\}_{i=1}^{n}$ and $\{X_{i\ell}\}_{i=1}^{n}$ to obtain standardized dataset $\mathbf{O}^* = \{(\mathbf{S}_i, T_i, \mathbf{X}_i^*, Y_i^*)\}_{i=1}^{n}$, where $\mathbf{X}^*$ and Y^* are standardized covariates and responses.
  for $\omega_n \in \mathcal{W}^*$ **do**
    (i) Compute $\widehat{\boldsymbol{\theta}}_{\omega_n}$ by minimizing (14) with $\mathbf{O}^*$.
    (ii) Calculate corresponding extended BIC score, $\mathrm{EBIC}(\widehat{\boldsymbol{\theta}}^{\mathrm{I}}_{\omega_n})$ defined in (16).
  end
  Select optimal penalty parameter $\omega^* = \arg\min_{\omega_n} \{\mathrm{EBIC}(\widehat{\boldsymbol{\theta}}^{\mathrm{I}}_{\omega_n})\}$.
  The estimated model structure, that is, $\widehat{\mathcal{A}}, \widehat{\mathcal{A}}^{\mathrm{c}}$, is then deduced by $\widehat{\boldsymbol{\theta}}^{\mathrm{I}}_{\omega^*}$ as defined in (15).
Stage 2: Refitting.
  Use selected model structure $\widehat{\mathcal{A}}, \widehat{\mathcal{A}}^{\mathrm{c}}$.
  Estimate $\widehat{\boldsymbol{\theta}}$ by minimizing (5) with original dataset $\mathbf{O}$.

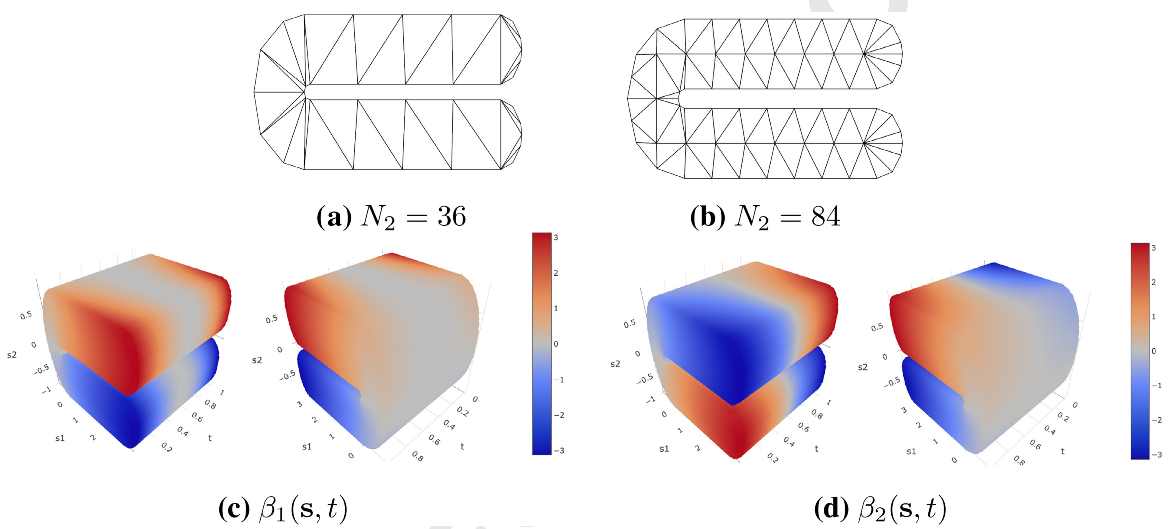


FIGURE 2 | Illustrations of horseshoe-shaped domain Ω for spatial locations $\mathbf{s}$ with (a) coarse triangulation with 32 triangles and (b) fine triangulation with 86 triangles; Spatiotemporally varying coefficient functions (c) $\beta_0(\mathbf{s}, t)$, (d) $\beta_1(\mathbf{s}, t)$ and (e) $\beta_2(\mathbf{s}, t)$ from two viewpoints.

problem to simplify the computation and avoid technical difficulties. In the refitting stage, these penalties are reintroduced to enhance model accuracy and stability. The optimal values of $\lambda_{1,\ell}$ and $\lambda_{2,\ell}$ can be determined using k-fold cross-validation (CV), minimizing the cross-validated mean squared prediction error (CV-MSPE): $n^{-1}\sum_{k=1}^{K}\sum_{i\in i[k]}(g^{-1}(\widehat{\eta}_i^{-[k]}) - Y_i)^2$, where $i[k]$ indicates the index set of kth fold, and $g^{-1}(\widehat{\eta}_i^{-[k]})$ represents the predicted mean evaluated at $(\mathbf{s}_i, t_i, \mathbf{x}_i)$ with the identified and refitted model trained on all but the kth fold. The unique spatial locations are randomly partitioned into k equal-sized folds, $\mathcal{S}_k$, $k = 1, \dots, K$, and with $i[k] = \{i : \mathbf{s}_i \in \mathcal{S}_k\}$ being the index set of the kth fold. This sampling strategy is consistently applied for all CV procedures in this paper.

Nonparametric settings. In our proposed framework, we employ the following rules of thumb to balance computational efficiency and accuracy when selecting the degree of univariate splines ϱ, the number of interior knots N_1, the smoothness for bivariate splines r, the degree of bivariate splines d and the triangulation $\triangle$. In Stage 1, to allow for a more stable and faster computation during the identification step, we recommend fixing $\triangle$ so that, on average, at least $d(d+1)/2$ observed locations within each triangle in $\triangle$, and $r = 0$ for smoothness. In addition, we employ a CV procedure to select the optimal nonparametric setting from a pool of nonparametric settings with lower model complexity features: $N_1 \in \{3, 5\}, \varrho \in \{2, 3\}, d \in \{0, 2\}$. In Stage 2, with the identified model structure, users can consider a refined triangulation $\triangle$, a greater number of interior knots N_1, and an increased degree of tensor product splines d and ϱ to enhance accuracy. In the numerical examples, we consider $N_1 = 3, \varrho = 3, d = 2, r = 1$.

5 | Simulation Studies

In this section, we conduct simulation studies to evaluate the finite-sample performance of the proposed estimation and structure identification procedure for GST-SVCMs. Across all simulations, we consider the following conditional mean function: $\mu_i = \sum_{\ell\in\mathcal{A}}\beta_\ell(\mathbf{s}_i, t_i)x_{i\ell}(\mathbf{s}_i, t_i) + \sum_{\ell\in\mathcal{A}^{\mathrm{c}}}\alpha_\ell x_{i\ell}(\mathbf{s}_i, t_i)$, where $\beta_\ell(\mathbf{s}, t)$ for $\ell = 1, 2$ are spatiotemporally varying coefficient functions defined as: $\beta_1(\mathbf{s}, t) = 3m_0(\mathbf{s})(t - 0.5)^2$ and

$\beta_2(\mathbf{s},t) = 0.75m_0(\mathbf{s})(2t-1)$. Here, $m_0(\mathbf{s})$ is the function defined in `mgcv::fs.test` with parameters `r0 = 0.1, r = 0.5, l = 3, b = 1`. These coefficient functions are illustrated in Figure 2c,d. The coefficients α_ℓ for $\ell = 3,4,5,6$ are constants, with $\alpha_3 = 1, \alpha_4 = -1, \alpha_5 = -0.5, \alpha_6 = -0.1$. For $1 \le \ell \le p$ and $1 \le i \le n$, the independent continuous covariates $X_{i\ell}$'s are generated from standard normal distributions. Within the horseshoe-shaped domain shown in Figure 2, we consider two types of response distributions:

- *Case I (Gaussian)*: $Y_i \sim \text{Normal}(\mu_i, 2^2)$, where $g(\mu) = \mu$;
- *Case II (Bernoulli)*: $Y_i \sim \text{Bernoulli}(\mu_i)$, where $g(\mu) = \log\{\mu/(1-\mu)\}$.

We evaluate the performance of the proposed method for different numbers of unique spatial locations ($n_S = 60, 120, 240, 360$) and unique time points ($n_T = 60, 120, 240, 360$). The total number of observations is given by $n = n_S \times n_T$.

We evaluate the proposed procedure in terms of the accuracy of model structure identification and the prediction precision with the following criteria:

(I) Percentage of identified vArying component (PA) for x_ℓ: the average percentage of x_ℓ detected as having a spatiotemporally varying coefficient function: $\text{PA}_\ell = L^{-1}\sum_{b=1}^{L} I(\ell \in \hat{\mathcal{A}}_b)$, where $b = 1, \ldots, L$ is the index of Monte Carlo replications, $L = 100$ is the number of replications, and $\hat{\mathcal{A}}_b$ represents the set of covariates indices identified as having a spatiotemporally varying coefficient function at the bth replication.

(II) Mean Squared Errors (MSE) for coefficient estimates for x_ℓ:

$$\text{MSE}_\ell = \frac{1}{L}\sum_{b=1}^{L}\text{MSE}_{\ell,b},$$

$$\text{MSE}_{\ell,b} = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}\{\hat{\beta}_{\ell,b}(\mathbf{s}_i,t_i) - \beta_\ell(\mathbf{s}_i,t_i)\}^2, & \ell \in \mathcal{A}_0, \ell \in \hat{\mathcal{A}}_b, \\ \frac{1}{n}\sum_{i=1}^{n}\{\hat{\alpha}_{\ell,b} - \beta_\ell(\mathbf{s}_i,t_i)\}^2, & \ell \in \mathcal{A}_0, \ell \in \hat{\mathcal{A}}_b^{c}, \\ \frac{1}{n}\sum_{i=1}^{n}\{\hat{\beta}_{\ell,b}(\mathbf{s}_i,t_i) - \alpha_\ell\}^2, & \ell \in \mathcal{A}_0^{c}, \ell \in \hat{\mathcal{A}}_b, \\ (\hat{\alpha}_{\ell,b} - \alpha_\ell)^2, & \ell \in \mathcal{A}_0^{c}, \ell \in \hat{\mathcal{A}}_b^{c}, \end{cases}$$

where $\hat{\beta}_{\ell,b}$ and $\hat{\alpha}_{\ell,b}$ are the refitting estimates for the bth random sample.

(III) Mean Integrated Squared Errors (MISE) for coefficient estimates, defined as

$$\text{MISE}_\ell = \frac{1}{L}\sum_{b=1}^{L}\text{MISE}_{\ell,b},$$

$$\text{MISE}_{\ell,b} = \begin{cases} \frac{1}{N_g}\sum_{j=1}^{N_g}\{\hat{\beta}_{\ell,b}(\mathbf{s}_j,t_j) - \beta_\ell(\mathbf{s}_j,t_j)\}^2, & \ell \in \mathcal{A}_0, \ell \in \hat{\mathcal{A}}_b, \\ \frac{1}{N_g}\sum_{j=1}^{N_g}\{\hat{\alpha}_{\ell,b} - \beta_\ell(\mathbf{s}_j,t_j)\}^2, & \ell \in \mathcal{A}_0, \ell \in \hat{\mathcal{A}}_b^{c}, \\ \frac{1}{N_g}\sum_{j=1}^{N_g}\{\hat{\beta}_{\ell,b}(\mathbf{s}_j,t_j) - \alpha_\ell\}^2, & \ell \in \mathcal{A}_0^{c}, \ell \in \hat{\mathcal{A}}_b, \\ (\hat{\alpha}_{\ell,b} - \alpha_\ell)^2, & \ell \in \mathcal{A}_0^{c}, \ell \in \hat{\mathcal{A}}_b^{c}, \end{cases}$$

where $\{(\mathbf{s}_j,t_j) : 1 \le j \le N_g\}$ are uniform lattices over $\Omega \times \mathcal{T}$. In our simulation studies, we use $N_g = 80 \times 50 \times 50$ lattice points over $\Omega \times \mathcal{T}$.

(IV) Cross-validated Mean Squared Prediction Error (CV-MSPE) for response Y:

$$\text{CV-MSPE}_Y = \frac{1}{L}\sum_{b=1}^{L}\text{CV-MSPE}_{Y,b},$$

$$\text{CV-MSPE}_{Y,b} = \frac{1}{n}\sum_{k=1}^{K}\sum_{i\in[k]}\{g^{-1}(\hat{\eta}_{i,b}^{-[k]}) - Y_i\}^2.$$

Details of the CV procedure are provided in Section 4.

For Criterion (I), PA_ℓ is reported for all ℓ to see how well the proposed method identifies the structure of each component in the model. In practice, misidentifying a spatially varying coefficient function as constant may be more detrimental to subsequent analyses, as it fails to capture the spatial signal. In contrast, with limited data, incorrectly identifying a constant coefficient as spatially varying can lead to a challenging and unreliable estimation.

Table 1 presents the consistency in model structure identification as n_T and n_S increase. The values of PAℓ represent the

TABLE 1 | Percentage of β_ℓ, $\ell = 1, \ldots, 6$, detected as spatiotemporally varying functions (PA_ℓ).

Gaussian family								Binomial family							
n_S	n_T	PA_1	PA_2	PA_3	PA_4	PA_5	PA_6	n_S	n_T	PA_1	PA_2	PA_3	PA_4	PA_5	PA_6
60	60	0.47	0.98	0.00	0.00	0.00	0.00	120	120	0.66	1.00	0.00	0.00	0.00	0.00
	120	1.00	1.00	0.00	0.00	0.00	0.00		240	1.00	1.00	0.00	0.00	0.00	0.00
	240	1.00	1.00	0.00	0.00	0.00	0.00		360	1.00	1.00	0.00	0.00	0.00	0.00
120	60	1.00	1.00	0.00	0.00	0.00	0.00	240	120	1.00	1.00	0.00	0.00	0.00	0.00
	120	1.00	1.00	0.00	0.00	0.00	0.00		240	1.00	1.00	0.00	0.00	0.00	0.00
	240	1.00	1.00	0.00	0.00	0.00	0.00		360	1.00	1.00	0.00	0.00	0.00	0.00
240	60	1.00	1.00	0.00	0.00	0.00	0.00	360	120	1.00	1.00	0.00	0.00	0.00	0.00
	120	1.00	1.00	0.00	0.00	0.00	0.00		240	1.00	1.00	0.00	0.00	0.00	0.00
	240	1.00	1.00	0.00	0.00	0.00	0.00		360	1.00	1.00	0.00	0.00	0.00	0.00

empirical percentages of X_ℓ being detected to have varying coefficient functions across 100 iterations. Given that the true model structure is $\mathcal{A}_0 = \{1,2\}, \mathcal{A}_0^c = \{3,4,5,6\}$, higher PA_ℓ for $\ell \in \mathcal{A}_0$ and the lower PA_ℓ for $\ell \in \mathcal{A}_0^c$ correspond to better model identification accuracy. As n_S and n_T increase, the detection accuracy improves significantly. A moderate number of observations, specifically $n_T \geq 120$ or $n_S \geq 120$ for the Gaussian family and $n_T \geq 240$ or $n_S \geq 240$ for the binomial family, yield a highly accurate model structure identification. These results confirm the effectiveness of the proposed structure identification procedure for GST-SVCMs, even when applied to a subset of large-scale data, making the model determination process more efficient and scalable. To illustrate the impact of hyperparameters, we conduct additional simulation studies on different choices of $\triangle$, N_1, d, and ϱ, and the results are reported in Section S1 in the Supporting Information.

For each setting, we report MISE_ℓ for $\ell = 1,2$ and MSE_ℓ for $\ell = 3,4,5,6$ under both the identified model and the full model, where the latter assumes all covariates have spatiotemporally varying coefficient functions. Results for the identified and full models are denoted as $[\cdot]^{\mathrm{I}}$ and $[\cdot]^{\mathrm{F}}$, respectively. As shown in Table 2, $\mathrm{MSE}_\ell^{\mathrm{I}}$ are strictly smaller than $\mathrm{MSE}_\ell^{\mathrm{F}}$ for all estimators of the constant coefficients, demonstrating that structure identification significantly enhances estimation accuracy. Moreover, when the model structure is correctly identified, the MISEs for the spatiotemporally varying coefficient estimators remain comparable between the identified and full models. Additionally, the computing time of GST-SVCM with structure identification, even when accounting for refitting, is significantly shorter than that of the full model.

Table 3 presents empirical coverage rates and average standard errors for constant coefficients across varying sample sizes under both scenarios. The empirical coverage rates measure the proportion of times the 95% confidence interval contains the true parameter value over 100 replications, while the values in parentheses represent the average standard errors of the estimators. For both scenarios, the standard error decreases as the sample size increases. With a moderate sample size, the empirical coverage rates are close to the nominal 95% level, regardless of the simulation setting.

6 | Granular Model Identification

In this section, we extend the proposed method to achieve more granular model identification by introducing an alternative formulation of the GST-SVCM in (2), referred to as the hierarchical spatiotemporal varying coefficient model (HSTVCM):

$$\begin{aligned}\eta(\mathbf{s},t,\mathbf{x}) = g\{\mu(\mathbf{s},t,\mathbf{x})\} &= \beta_0^{s,t}(\mathbf{s},t) + \sum_{\ell=1}^{p}\alpha_\ell x_\ell(\mathbf{s},t) \\ &+ \sum_{\ell\in\mathcal{A}^t}\beta_\ell^t(t)x_\ell(\mathbf{s},t) + \sum_{\ell\in\mathcal{A}^s}\beta_\ell^s(\mathbf{s})x_\ell(\mathbf{s},t) \\ &+ \sum_{\ell\in\mathcal{A}^{s,t}}\beta_\ell^{s,t}(\mathbf{s},t)x_\ell(\mathbf{s},t) \qquad (16)\end{aligned}$$

TABLE 2 | Mean Squared Errors (MSEs) and Mean Integrated Squared Errors (MISEs) in estimating β_ℓ under identified models ($\mathrm{MSE}_\ell^{\mathrm{I}}$) and full models ($\mathrm{MSE}_\ell^{\mathrm{F}}$) in Simulation Studies with BIC.

n_T	n_S	$\mathrm{MSE}_1^{\mathrm{I}}$	$\mathrm{MSE}_1^{\mathrm{F}}$	$\mathrm{MSE}_2^{\mathrm{I}}$	$\mathrm{MSE}_2^{\mathrm{F}}$	$\mathrm{MSE}_3^{\mathrm{I}}$	$\mathrm{MSE}_3^{\mathrm{F}}$	$\mathrm{MSE}_4^{\mathrm{I}}$	$\mathrm{MSE}_4^{\mathrm{F}}$	$\mathrm{MSE}_5^{\mathrm{I}}$	$\mathrm{MSE}_5^{\mathrm{F}}$	$\mathrm{MSE}_6^{\mathrm{I}}$	$\mathrm{MSE}_6^{\mathrm{F}}$
							Gaussian						
60	60	0.4177	0.1680	0.2045	0.1809	0.0017	0.0031	0.0011	0.0026	0.0013	0.0032	0.0016	0.0027
	120	0.0856	0.0857	0.0952	0.0953	0.0005	0.0016	0.0005	0.0011	0.0008	0.0015	0.0005	0.0012
	240	0.0442	0.0441	0.0478	0.0479	0.0003	0.0006	0.0003	0.0007	0.0003	0.0007	0.0002	0.0007
120	60	0.1088	0.1088	0.1283	0.1283	0.0006	0.0016	0.0007	0.0011	0.0005	0.0014	0.0005	0.0010
	120	0.0499	0.0499	0.0551	0.0552	0.0003	0.0007	0.0003	0.0005	0.0003	0.0007	0.0003	0.0007
	240	0.0250	0.0250	0.0271	0.0271	0.0002	0.0004	0.0001	0.0003	0.0002	0.0003	0.0001	0.0004
240	60	0.0721	0.0722	0.0877	0.0879	0.0003	0.0006	0.0003	0.0007	0.0003	0.0007	0.0002	0.0006
	120	0.0295	0.0294	0.0314	0.0316	0.0002	0.0004	0.0001	0.0002	0.0001	0.0003	0.0001	0.0003
	240	0.0139	0.0139	0.0152	0.0152	0.0001	0.0002	0.0001	0.0001	0.0001	0.0002	0.0001	0.0002
							Binomial						
120	120	0.2769	0.0911	0.1167	0.1019	0.0023	0.0017	0.0026	0.0016	0.0007	0.0014	0.0004	0.0013
	240	0.0474	0.0475	0.0528	0.0529	0.0003	0.0008	0.0004	0.0010	0.0003	0.0006	0.0002	0.0006
	360	0.0322	0.0322	0.0352	0.0353	0.0002	0.0004	0.0002	0.0005	0.0002	0.0005	0.0002	0.0004
240	120	0.0535	0.0536	0.0621	0.0623	0.0002	0.0009	0.0004	0.0008	0.0003	0.0007	0.0002	0.0006
	240	0.0263	0.0264	0.0296	0.0297	0.0002	0.0004	0.0001	0.0003	0.0001	0.0004	0.0002	0.0003
	360	0.0174	0.0174	0.0194	0.0194	0.0001	0.0003	0.0001	0.0003	0.0001	0.0002	0.0001	0.0002
360	120	0.0391	0.0392	0.0448	0.0448	0.0002	0.0005	0.0002	0.0006	0.0002	0.0004	0.0002	0.0004
	240	0.0188	0.0188	0.0210	0.0210	0.0001	0.0003	0.0001	0.0003	0.0001	0.0002	0.0001	0.0002
	360	0.0123	0.0124	0.0138	0.0138	0.0001	0.0002	0.0001	0.0002	0.0001	0.0001	0.0001	0.0001

TABLE 3 | Standard errors and empirical coverage rates of 95% CIs for constant coefficients.

		Gaussian family						Binomial family			
n_S	n_T	α_3	α_4	α_5	α_6	n_S	n_T	α_3	α_4	α_5	α_6
60	60	0.89 (0.035)	0.97 (0.035)	0.94 (0.035)	0.91 (0.035)	120	120	0.67 (0.024)	0.65 (0.024)	0.92 (0.022)	0.96 (0.021)
	120	0.95 (0.024)	0.96 (0.024)	0.92 (0.024)	0.97 (0.024)		240	0.95 (0.017)	0.91 (0.017)	0.93 (0.016)	0.94 (0.015)
	240	0.92 (0.017)	0.94 (0.017)	0.95 (0.017)	0.98 (0.017)		360	0.97 (0.010)	0.93 (0.010)	0.96 (0.009)	0.97 (0.009)
120	60	0.96 (0.024)	0.92 (0.024)	0.94 (0.024)	0.98 (0.024)	240	120	0.97 (0.017)	0.92 (0.017)	0.87 (0.016)	0.96 (0.015)
	120	0.94 (0.017)	0.96 (0.017)	0.94 (0.017)	0.94 (0.017)		240	0.95 (0.012)	0.96 (0.012)	0.94 (0.011)	0.93 (0.011)
	240	0.93 (0.012)	0.91 (0.012)	0.91 (0.012)	0.95 (0.012)		360	0.97 (0.010)	0.93 (0.010)	0.96 (0.009)	0.97 (0.009)
240	60	0.96 (0.017)	0.97 (0.017)	0.91 (0.017)	0.97 (0.017)	360	120	0.94 (0.014)	0.94 (0.014)	0.95 (0.013)	0.94 (0.012)
	120	0.90 (0.012)	0.94 (0.012)	0.95 (0.012)	0.95 (0.012)		240	0.95 (0.010)	0.93 (0.010)	0.93 (0.009)	0.94 (0.009)
	240	0.92 (0.008)	0.95 (0.008)	0.95 (0.008)	0.95 (0.008)		360	0.94 (0.008)	0.95 (0.008)	0.96 (0.007)	0.95 (0.007)

where $\mathcal{A}^s = \{\ell : \beta_\ell^s \not\equiv 0\}$, $\mathcal{A}^t = \{\ell : \beta_\ell^t \not\equiv 0\}$ and $\mathcal{A}^{s,t} = \{\ell : \beta_\ell^{s,t} \not\equiv 0\}$ denote the sets of covariates with spatial-only, temporal-only, and spatiotemporal interaction effects, respectively. Unlike the GST-SVCM model, which assumes covariates have either constant or fully spatiotemporally varying effects, the HSTVCM differentiates among these distinct sources of variation. This distinction enhances structure identification and yields a more parsimonious representation, which is particularly advantageous in high-dimensional applications where isolating spatial, temporal, and interaction effects is essential for interpretability and predictive accuracy.

We define the approximation function space of $\eta(\mathbf{s}, t, \mathbf{x})$ as $\widetilde{\mathcal{G}}$ as follows

$$\widetilde{\mathcal{G}} = \Big\{ \eta(\mathbf{s}, t, \mathbf{x}; \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathcal{A}^s, \mathcal{A}^t, \mathcal{A}^{s,t}) : \boldsymbol{\alpha} \in \mathbb{R}^p;$$
$$\beta_\ell^s(\mathbf{s}) \in \widetilde{\mathbb{S}}_d^r(\triangle), \ell \in \mathcal{A}^s; \beta_\ell^t(t) \in \widetilde{\mathbb{U}}^\varrho(\mathcal{T}), \ell \in \mathcal{A}^t;$$
$$\beta_\ell^{s,t}(\mathbf{s}, t) \in \widetilde{\mathbb{T}}_v^{(\varrho,d,r)}(\mathcal{E}), \ell \in \{0\} \cup \mathcal{A}^{s,t} \Big\},$$

where $\widetilde{\mathbb{S}}_d^r(\triangle) = \{f(\mathbf{s}) \in \mathbb{S}_d^r(\triangle) : \mathrm{E}f(\mathbf{S}) = 0\}$ denotes the function space of centered spatially varying functions; $\widetilde{\mathbb{U}}^\varrho(\mathcal{T}) = \{f(t) \in \mathbb{U}^\varrho(\mathcal{T}) : \mathrm{E}f(T) = 0\}$ denotes the function space of centered temporally varying functions; and $\widetilde{\mathbb{T}}_v^{(\varrho,d,r)}(\mathcal{E}) = \{f(\mathbf{s}, t) \in \mathbb{T}_v^{(\varrho,d,r)}(\mathcal{E}) : \mathrm{E}f(\mathbf{S}, T) = 0 \cap f(\mathbf{s}, t) \neq f_1(\mathbf{s}) + f_2(t), f_1 \in \mathbb{U}^\varrho(\mathcal{T}), f_2 \in \mathbb{S}_d^r(\triangle)\}$ denotes the function space of centered spatiotemporal varying functions that are not purely spatially varying nor purely temporally varying.

Algorithm 2 outlines the procedure for identifying $\mathcal{A}^s, \mathcal{A}^t$ and $\mathcal{A}^{s,t}$. To enhance computational efficiency, we first fit $g(\mu(\mathbf{s}, t, \mathbf{x})) = \beta_0(\mathbf{s}, t)$ without applying an identification penalty and define the adjusted response as $\widetilde{Y}_i = Y_i - g^{-1}(\widehat{\beta}_0(\mathbf{s}, t))$, referred to as Stage 0. Subsequently, in Stage 1, we determine the model structure by minimizing the penalized negative quasi-likelihood

$$\widehat{\eta} = \mathop{\arg\min}_{\eta \in \widetilde{\mathcal{G}}} -\frac{1}{n}\sum_{i=1}^n Q\Bigg[g^{-1}\Bigg\{\sum_{\ell=1}^p \{\alpha_\ell + \beta_\ell^t(t) + \beta_\ell^s(\mathbf{s}) + \beta_\ell^{s,t}(\mathbf{s}, t)\} x_\ell(\mathbf{s}, t)\Bigg\}, \widetilde{Y}_i\Bigg] + \sum_{\ell=1}^p p_{\omega_{n,s}}(\|\beta_\ell^s\|_n) + \sum_{\ell=1}^p p_{\omega_{n,t}}(\|\beta_\ell^t\|_n) + \sum_{\ell=1}^p p_{\omega_n}(\|\beta_\ell^{s,t}\|_n) \quad (17)$$

where $p_{\omega_{n,s}}(\cdot), p_{\omega_{n,t}}(\cdot)$ and $p_{\omega_{n,st}}(\cdot)$ are group SCAD penalty functions, and $\omega_{n,t}, \omega_{n,s}, \omega_{n,st}$ are corresponding penalty parameters for identifying $\mathcal{A}^s, \mathcal{A}^t$ and $\mathcal{A}^{s,t}$, respectively (Li et al. 2019; Li, Wang, and Wang 2021).

During refitting (Stage 2), we simplify the model (16), such that as long as a covariate has space-time interaction term, that is, $\ell \in \mathcal{A}^{s,t}$, any existing additive temporal or spatial-only varying terms are absorbed by $\beta_\ell^{s,t}$ as follows

$$\begin{aligned} &\eta(\mathbf{s}, t, \mathbf{x}; \mathcal{A}^s, \mathcal{A}^t, \mathcal{A}^{s,t}) \\ &= g\{\mu(\mathbf{s}, t, \mathbf{x}; \mathcal{A}^s, \mathcal{A}^t, \mathcal{A}^{s,t})\} = \beta_0^{s,t}(\mathbf{s}, t) + \sum_{\ell=1}^p \alpha_\ell x_\ell(\mathbf{s}, t) \\ &\quad + \sum_{\ell \in \mathcal{A}^t \setminus \mathcal{A}^{s,t}} \beta_\ell^t(t) x_\ell(\mathbf{s}, t) + \sum_{\ell \in \mathcal{A}^s \setminus \mathcal{A}^{s,t}} \beta_\ell^s(\mathbf{s}) x_\ell(\mathbf{s}, t) \\ &\quad + \sum_{\ell \in \mathcal{A}^{s,t}} \beta_\ell^{s,t}(\mathbf{s}, t) x_\ell(\mathbf{s}, t), \end{aligned}$$

where $\beta_\ell^s(\mathbf{s}) \in \widetilde{\mathbb{S}}_d^r(\triangle)$ are spatially varying coefficient functions for $\ell \in \mathcal{A}^s \setminus \mathcal{A}^{s,t}$; $\beta_\ell^t(t) \in \widetilde{\mathbb{U}}^\varrho(\mathcal{T})$ are temporally varying coefficient functions for $\ell \in \mathcal{A}^t$; and $\beta_\ell^{s,t}(\mathbf{s}, t) \in \mathbb{T}_v^{(\varrho,d,r)}(\mathcal{E}), \ell \in \{0\} \cup \mathcal{A}^{s,t}$ are spatiotemporally varying coefficients.

Similar to (4), the model refitting (Stage 2) is carried out by minimizing the following negative loglikelihood function with g defined (16):

$$\begin{aligned} &L_{n,\mathcal{A}^s,\mathcal{A}^t,\mathcal{A}^{s,t}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \\ &= -\frac{1}{n}\sum_{i=1}^n Q\big[g^{-1}\{\eta(\mathbf{S}_i, T_i, \mathbf{X}_i; \boldsymbol{\alpha}, \boldsymbol{\beta}, \mathcal{A}^s, \mathcal{A}^t, \mathcal{A}^{s,t})\}, Y_i\big] \\ &\quad + \sum_{\ell \in \mathcal{A}^{s,t}} \{\lambda_{1,\ell} f_1(\beta_\ell^{s,t}) + \lambda_{2,\ell} f_2(\beta_\ell^{s,t})\} + \sum_{\ell \in \mathcal{A}^s} \lambda_{3,\ell} f_3(\beta_\ell^s), \end{aligned}$$

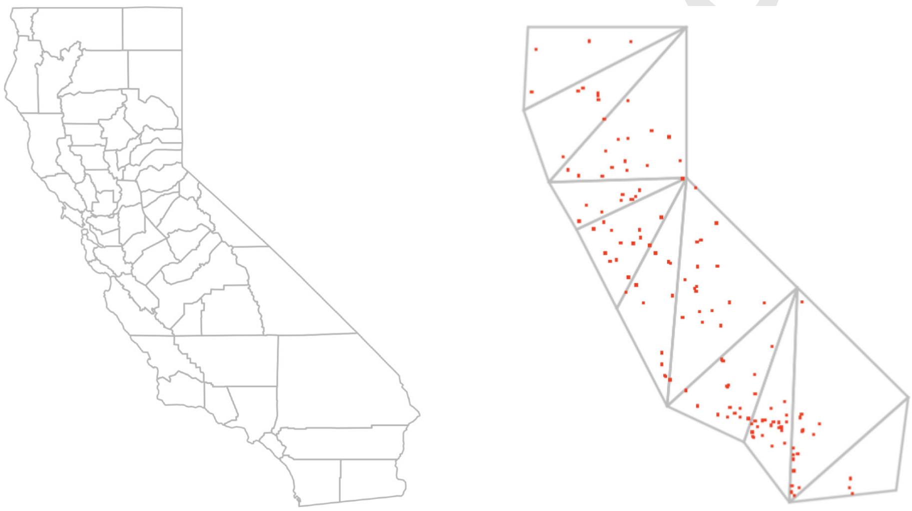
where f_1 and f_2 is defined along with (4); $f_3(\beta_\ell) = \int_\Omega \left\{(\nabla_{s_1}^2 \beta_\ell)^2 + (\nabla_{s_2}^2 \beta_\ell)^2\right\} ds_1 ds_2$ is a function measuring the roughness of β_ℓ^s with respect to spatial locations (Wang et al. 2022).

7 | Data Application

Particulate matter (PM) has consistently shown an adverse influence on public health (Harrison and Yin 2000). Fine particles

ALGORITHM 2 | Granular structure identification for HSTVCMs in (17).

Input: Dataset $\mathbf{O} = \left\{(\mathbf{S}_i, T_i, \mathbf{X}_i, Y_i)\right\}_{i=1}^{n}$ for both model identification and refitting.
Output: Identified structure set $\widehat{\mathcal{A}}^{s,t}$ and estimated coefficient functions
Stage 0: Initial Intercept Estimation
  Fit a generalized model with only the spatiotemporal intercept term, $\mathbb{E}[Y_i|\mathbf{s}_i, t_i] = g^{-1}(\beta_0(\mathbf{s}_i, t_i))$, and obtain the estimate $\widetilde{\beta}_0(\mathbf{s}, t)$.Define an adjusted response to remove the estimated intercept: $\widetilde{Y}_i = Y_i - g^{-1}(\widetilde{\beta}_0(\mathbf{s}_i, t_i))$.
Stage 1: Structure Identification of Spatiotemporal Interaction
  Consider dataset $\widetilde{\mathbf{O}} = \{\mathbf{S}_i, T_i, \mathbf{X}_i, \widetilde{Y}_i\}_{i=1}^{n}$. Identify $\mathcal{A}^t, \mathcal{A}^t, \mathcal{A}^{s,t}$ in the HSTVCM by minimizing the penalized negative log quasi-likelihood function (18). Then $\widehat{\mathcal{A}}^s = \{\ell : \widehat{\beta}_\ell^s \not\equiv 0\}, \widehat{\mathcal{A}}^t = \{\ell : \widehat{\beta}_\ell^t \not\equiv 0\}$, $\widehat{\mathcal{A}}^{s,t} = \{\ell : \widehat{\beta}_\ell^{s,t} \not\equiv 0\}$. Detailed implementation can be found in Algorithm 1.
Stage 2: Model Refitting
  Use the $\widehat{\mathcal{A}}^s, \widehat{\mathcal{A}}^t, \widehat{\mathcal{A}}^{s,t}$ to refit the HSTVCM by minimizing the negative loglikelihood function as in (5) by replacing g to be consistent with (17).



(a) **(b)**

FIGURE 3 | (a) Map of California generated using the `ggplot2` package in R. (b) Triangulation of California with $N_2 = 11$ triangles, overlaid with observation sites from 2011.

(PM2.5), smaller than 2.5μ m in diameter, are of particular concern. Recent studies have noted that PM2.5 is affected by various meteorological conditions (Wang and Ogawa 2015; Chen et al. 2020). In this section, we apply the proposed method to environmental data to investigate the relationship between PM2.5 and different meteorological factors.

We examine the association of the daily mean of surface concentrations of PM2.5 with various meteorological drivers, including daily total gridded precipitation (PPTN), surface wind speed (WS), surface daily minimum air temperature (Tmin), surface daily maximum air temperature (Tmax) (Livneh et al. 2013), air relative humidity (RH), and total column cloud cover (TCDC) (Mesinger et al. 2006). The daily PM2.5 for 2011 is obtained from the US Environmental Protection Agency, and the meteorological drivers are provided by the National Oceanic and Atmospheric Administration (http://www.esrl.noaa.gov/psd/). In this analysis, we use all $n_S = 134$ distinct spatial locations across California and $n_T = 364$ temporal observations throughout 2011. The distribution of observation sites is shown in Figure 3b.

We first apply the proposed GST-SVCM with structure identification to analyze how meteorological variables affect PM2.5 concentrations, capturing spatial and temporal variations. Let $\{(\mathbf{S}_i, T_i, \mathbf{X}_i, Y_i)\}_{i=1}^{n}$ be the observations, where $Y_i = \text{PM2.5}_i$ is the response variable and $\mathbf{X}_i = \{\text{PPTN}, \text{WS}, \text{Tmin}, \text{Tmax}, \text{RH}, \text{TCDC}\}_i$ are covariates. The model is specified as:

$$
\begin{aligned}
\text{PM2.5}_i = {} & \beta_0(\mathbf{S}_i, T_i) + \beta_1(\mathbf{S}_i, T_i)\text{PPTN}(\mathbf{S}_i, T_i) \\
& + \beta_2(\mathbf{S}_i, T_i)\text{WS}(\mathbf{S}_i, T_i) + \beta_3(\mathbf{S}_i, T_i)\text{Tmin}(\mathbf{S}_i, T_i) \\
& + \beta_4(\mathbf{S}_i, T_i)\text{Tmax}(\mathbf{S}_i, T_i) + \beta_5(\mathbf{S}_i, T_i)\text{RH}(\mathbf{S}_i, T_i) \\
& + \beta_6(\mathbf{S}_i, T_i)\text{TCDC}(\mathbf{S}_i, T_i) + \epsilon_i, \ 1 \le i \le n.
\end{aligned}
$$

TABLE 4 | Identified model structure for HSTVCM and estimated constant coefficients (with 95% confidence intervals) for each covariate.

Model	Covariate	Type of variation	C-Coeff. (α_ℓ)	95% CI
GST-SVCM	Intercept	Spatiotemporal	2.2196	[2.2077, 2.2315]
	PPTN	Spatiotemporal	−0.0873	[−0.1046,−0.0700]
	WS	Spatiotemporal	−0.0542	[−0.0625,−0.0459]
	Tmin	Spatiotemporal	0.1577	[0.1435,0.1719]
	Tmax	Constant	0.1492	[0.1326,0.1658]
	RH	Constant	0.0073	[−0.0024,0.0170]
	TCDC	Spatiotemporal	−0.0513	[−0.0596,−0.0430]
HSTVCM	Intercept	Spatiotemporal	2.2376	[2.2248,2.2503]
	PPTN	Additive Spatial + Temporal	0.4223	[0.2573,0.5874]
	WS	Additive Spatial + Temporal	−0.0120	[−0.0444,0.0205]
	Tmin	Spatiotemporal	0.1413	[0.1265,0.1561]
	Tmax	Spatiotemporal	0.1467	[0.1293,0.1640]
	RH	Spatiotemporal	0.0041	[−0.0059,0.0142]
	TCDC	Temporal Only	−0.0556	[−0.0652,−0.0461]

Note: (1) "Type of Variation" indicates whether the coefficient for each covariate is purely spatial, purely temporal, additive spatial-plus-temporal, or fully spatiotemporal. (2) The "C-Coeff." column shows the baseline constant effect estimated for each covariate, while the 95% CI gives its uncertainty bounds. (3) For visualizations of spatial or temporal effects; see Figure 5 for spatial maps, time-series plots.

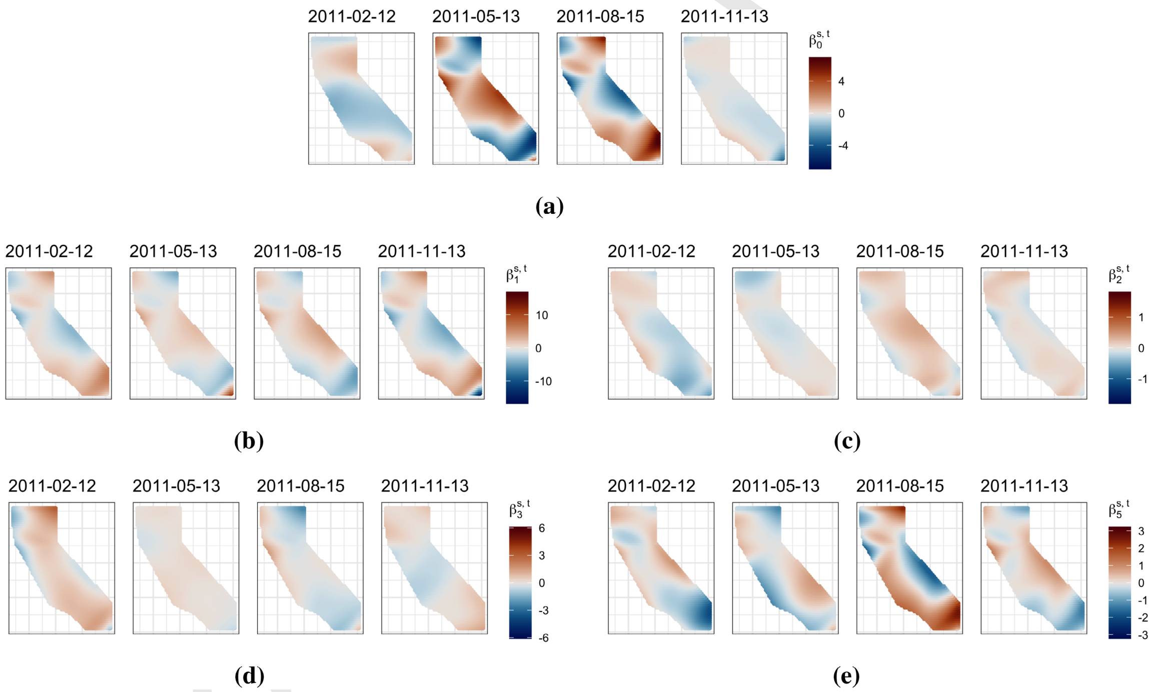


FIGURE 4 | Estimated spatiotemporally varying coefficient functions based on the identified GST-SVCM. (a) $\beta_0(\mathbf{s}, t)$ for intercept, (b) $\beta_1(\mathbf{s}, t)$ for PPTN, (c) $\beta_2(\mathbf{s}, t)$ for WS, (d) $\beta_3(\mathbf{s}, t)$ for Tmin, (e) $\beta_5(\mathbf{s}, t)$ for RH.

Following Algorithm 1, we conduct the GST-SVCM model structure identification and model fitting. For the initial structure identification (Stage 1), we implement a triangular prismatic partition consisting of $N_1 = 3$ interior knots, $N_2 = 11$ triangles, and $d = 2, r = 1, \varrho = 2$. Figure 3b illustrates the spatial triangulation used for model structure identification and refitting. The proposed structure identification method identifies $\hat{\mathcal{A}}^c = \{4, 6\}$ and $\hat{\mathcal{A}} = \{1, 2, 3, 5\}$, suggesting that Tmax and TCDC exhibit linear effects on PM2.5, whereas PPTN, WS, Tmin, and RH demonstrate spatiotemporally varying effects. Following structure identification, we refit the model (Stage 2) with the same smoothing parameter settings ($N_1 = 3, N_2 = 11, d = 2, r = 1, \varrho = 2$). Estimates of the baseline constant coefficients α_ℓ's for PPTN, WS, Tmin, Tmax, RH, TCDC along with their 95% confidence intervals are reported in the upper panel of Table 4. The spatiotemporal varying effects of the covariates are visualized in Figure 4.

Next, we apply the proposed refined HSTVCM in (16) and perform structure identification following Algorithm 2. We employ

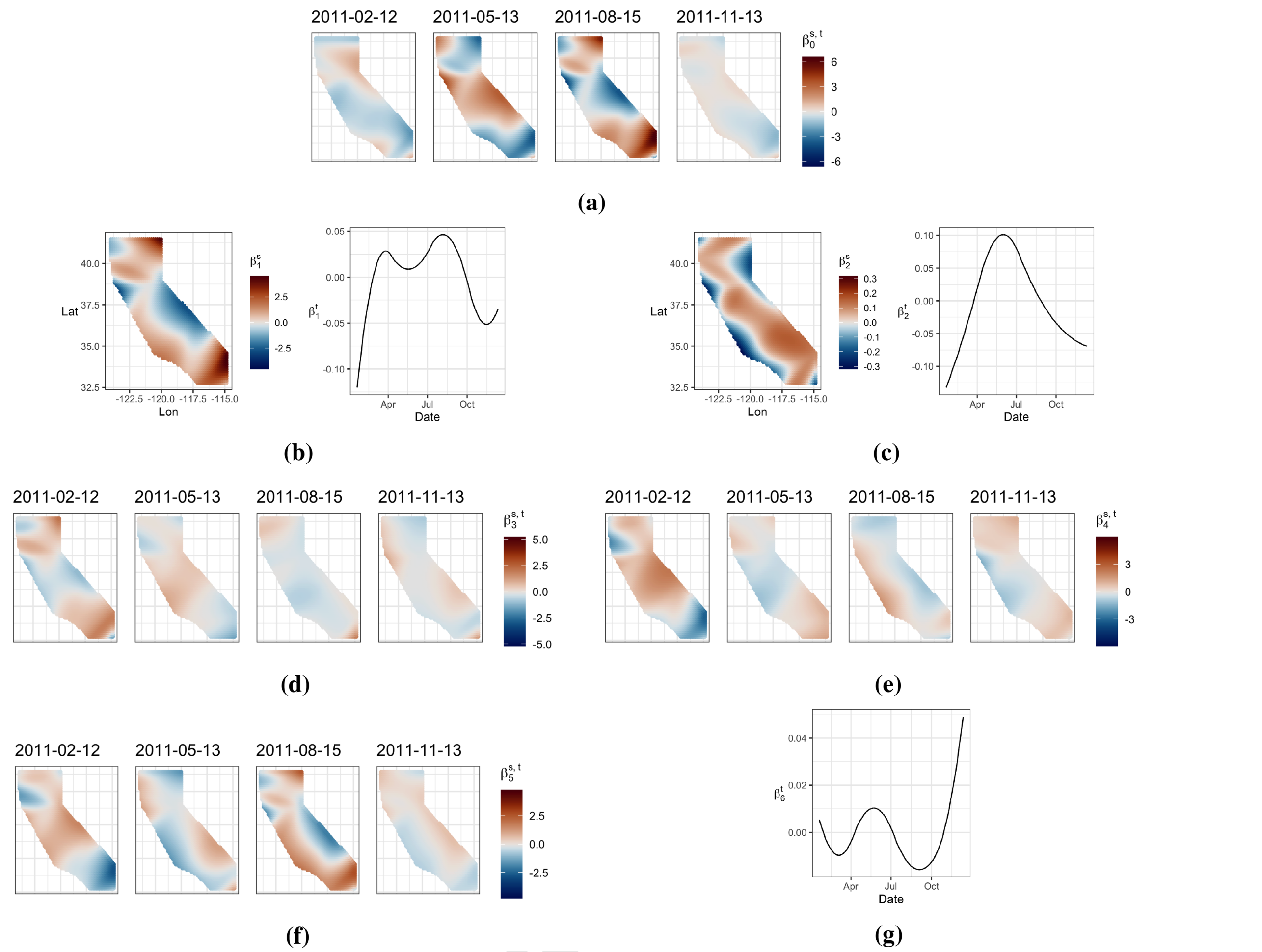


FIGURE 5 | Varying coefficient functions for identified HSTVCM with $\hat{\mathcal{A}}^s = \{1, 2, 3, 5\}$, $\hat{\mathcal{A}}^t = \{1, 2, 4, 5, 6\}$, $\hat{\mathcal{A}}^{s,t} = \{3, 4, 5\}$. (a) $\beta_0(\mathbf{s}, t)$ for intercept, (b) $\beta_1^s(\mathbf{s})$ and $\beta_1^t(t)$ for PPTN, (c) $\beta_2^s(\mathbf{s})$ and $\beta_2^t(t)$ for WS, (d) $\beta_3^{s,t}(\mathbf{s}, t)$ for Tmin, (e) $\beta_4^{s,t}(\mathbf{s}, t)$ for Tmax, (f) $\beta_5^{s,t}(\mathbf{s}, t)$ for RH, (g) $\beta_6^t(t)$ for TCDC.

TABLE 5 | Cross-validated mean squared prediction errors (CV-MSPEs) and mean squared errors (MSEs) for the identified models.

Identified model for HSTVCM		Identified model for GST-SVCM		Pure linear model		Full model	
CV-MSPE	MSE	CV-MSPE	MSE	CV-MSPE	MSE	CV-MSPE	MSE
0.2751	0.2523	0.2991	0.2537	0.3072	0.3002	0.2990	0.2459

a consistent parameterization for initial fitting (Stage 0), model identification (Stage 1), and refitting (Stage 2) stages, with $d = 2, r = 1, \varrho = 2, N_1 = 3, N_2 = 11$. The identified model structure is characterized by $\hat{\mathcal{A}}^s = \{1, 2, 3, 5\}$, $\hat{\mathcal{A}}^t = \{1, 2, 4, 5, 6\}$, and $\hat{\mathcal{A}}^{s,t} = \{3, 4, 5\}$. The identified structure indicates that PPTN and WS ($\ell = 1, 2$) have additive spatial and temporal varying coefficient functions; Tmin, Tmax and RH ($\ell = 3, 4, 5$) have spatiotemporally varying coefficient functions; and TCDC has a temporally varying effect on PM2.5. The estimated baseline constant coefficients for all explanatory variables, along with their 95% confidence intervals, are reported in the bottom panel of Table 4. Figure 5 visualizes the varying components of the coefficient functions, highlighting the spatial, temporal, and spatiotemporal variations.

In general, our analysis reveals distinct spatiotemporal patterns in the relationship between meteorological conditions and PM2.5 concentrations across California, which align with the discussion in Jacob and Winner (2009). Specifically, Figures 4a and 5a present the spatiotemporally varying intercept function, highlighting the consistent overall trend of PM2.5 between the two identified models. The results indicate that PM2.5 exhibits less spatial variation during winter compared to non-winter periods. The temporal varying coefficient function $\beta_1^t(t)$ in Figure 5b shows precipitation events during the wet season (late fall to early spring) effectively reduce PM2.5 concentrations, whereas dry periods (late spring to early fall) promote particulate matter accumulation. This finding aligns with the state's characteristic precipitation pattern and supports the critical role of precipitation in removing PM2.5 from the atmosphere. Figures 4c and 5c show that the impact of wind speed demonstrates clear spatial heterogeneity, with a notable coastal-to-inland gradient. This effect is likely influenced by altitude, which impacts the ventilation and transport of pollutants (Chow et al. 2006). Figures 4d–e and 5d–f show that the temperature and relative humidity's influence on PM2.5 exhibits spatial variability, with both positive and negative associations observed across different regions. In Figure 5g, total column cloud cover shows increasing temporal effects toward the end of the year, reflecting seasonal atmospheric patterns that influence pollutant dispersion.

Table 5 reports the MSEs and the 10-fold CV-MSPEs for the identified models based on GST-SVCM and the refined HSTVCM, along with the pure linear regression model and the full model. The results demonstrate that the proposed structure identification approach yields a more parsimonious model while maintaining strong predictive accuracy. In contrast, the full model shows signs of overfitting, as shown by the discrepancy between its MSE and MSPE. Notably, the identified HSTVCM achieves the lowest MSE and CV-MSPE among all models, highlighting the efficiency gain achieved by incorporating a more granular model structure.

In addition, to assess model adequacy, we conduct model diagnostics using extended versions of Moran's I test adapted for spatiotemporal data. Let $e_i \equiv e_i(\mathbf{s}_i, t_i) = Y_i(\mathbf{s}_i, t_i) - \widehat{Y}_i(\mathbf{s}_i, t_i)$ denote the residuals. We test the null hypothesis $H_0 : e_i(\mathbf{s}_i, t_i)$ are mutually independent against the alternative hypothesis H_a : $e_i(\mathbf{s}_i, t_i)$ exhibit spatiotemporal dependence. Rejection of H_0 indicates that the identified model does not adequately capture the spatiotemporal relationship between covariates and PM2.5 concentrations. Following Wikle et al. (2019) and Dubé and Legros (2013), we employ the Moran's I statistics:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\left(e_i - \bar{e}\right)\left(e_j - \bar{e}\right)}{\left(\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right)\left\{\sum_{i=1}^{n}\left(e_i - \bar{e}\right)^2\right\}},$$

where $\bar{e} = \sum_{i=1}^{n} e_i/n$ represents the mean residual, and $w_{ij} = w_{s,ij} \times w_{t,ij}$ are weights that indicate the relationship between $(\mathbf{s}_i, t_i)$ and $(\mathbf{s}_j, t_j)$. Specifically, $w_{s,ij} = s_{ij}^{-1} I(s_{ij} < \bar{s}_i)$, $s_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|_2$ represents the spatial relation, and $w_{t,ij} = t_{ij}^{-1} I(0 < t_{ij} < \bar{t}) + I(i \neq j, t_{ij} = 0)$, $t_{ij} = |t_i - t_j|$ represents the temporal relation, where the spatial neighborhood threshold $\bar{s}_i$ is set to be the tenth percentile of all pairwise spatial distances $\|\mathbf{s}_i - \mathbf{s}_j\|_2$, and temporal neighborhood threshold $\bar{t} = 1/10$ corresponds to one month memory effect. For computation efficiency, we randomly sample 1000 instances $\{(\mathbf{s}_i, t_i, e_i)\}_{i=1}^{1000}$ and calculate the p-value. The procedure is repeated 100 times, and the averaged p-values, $\overline{p}$, are calculated. The pure linear model exhibits significant spatiotemporal dependence ($\overline{p} = 0.02$), but we fail to reject the null hypothesis for the full model ($\overline{p} = 0.32$) and identified models by HSTVCM ($\overline{p} = 0.34$) and GST-SVCM ($\overline{p} = 0.36$). These results indicate the models identified by HSTVCM and GST-SVCM effectively account for spatiotemporal dependence, despite their sparse structures.

8 | Conclusion

In this article, we introduced a class of flexible and parsimonious models for spatiotemporal regression with constant and varying coefficients, termed Generalized Spatiotemporal Semi-varying Coefficient Models (GST-SVCMs), and proposed an efficient estimation method. In addition, we proposed a model structure identification approach for GST-SVCMs, which enables users to automatically identify which coefficients are constant and which are spatiotemporally varying, thereby enhancing estimation efficiency and prediction accuracy. We demonstrated that the estimators of constant coefficients and varying coefficient functions in the GST-SVCM estimation are consistent, and the estimators of the constant coefficients are asymptotically normal. Furthermore, we showed that the proposed structure identification for GST-SVCMs can correctly identify the model structure with probability approaching one. Through extensive simulation studies, we illustrated the robust asymptotic behavior of the method. We further validated the method by applying it to a PM2.5 dataset, where both simulation and empirical results highlighted the efficiency gains of utilizing the identified sparse model structure compared to a more complex full model. The proposed method proved particularly advantageous when the sample size was moderate, allowing accurate structure identification and reliable estimation without overfitting the data. This approach not only enhanced computational efficiency, but also significantly improved the interpretability and predictive performance of spatiotemporal models.

Despite these advancements, our model assumes the independence of errors once the deterministic regression function, accounting for spatiotemporal variations, is extracted. While this simplifies the modeling process, it may overlook residual spatiotemporal correlations present in real-world data, potentially impacting the accuracy and validity of the proposed method. To address this limitation, future work could explore the integration of a spatiotemporal autoregressive varying coefficient model, based on methodologies such as those proposed in Yu et al. (2022). However, incorporating this extension presents significant methodological challenges that require careful theoretical and computational development. We plan to address these complexities in future research.

Our simulation studies, conducted under the assumption of moderately smooth spatiotemporal processes, demonstrate that the two-stage identification procedure reliably detects covariates with varying effects for sufficiently large sample sizes. However, in scenarios characterized by extreme non-stationarity or highly localized patterns, additional observations or more adaptive modeling strategies may be necessary. Future research should examine the impact of irregular triangulations and greater variability in spatially and temporally varying effects through additional simulations and empirical studies. We also recommend that researchers assess stationarity assumptions in their data and refine smoothing parameter selection when abrupt or localized changes in spatiotemporal structure are suspected.

While recent methodological and computational advances have significantly improved the efficiency of Bayesian spatiotemporal models (Gelfand et al. 2003) for handling moderate-to-large datasets, the proposed GST-SVCM structure identification framework distinguishes itself with superior computational efficiency, making it particularly well-suited for even larger spatiotemporal data. To handle "big" data, a compelling future extension involves integrating sampling techniques into our model structure identification process. In this scenario, a stratified sampling approach could be particularly effective. The data could be divided into different temporal strata (e.g., seasons or months) and spatial strata (e.g., geographical regions), followed by random sampling within each stratum. The proposed framework would then be applied to the sampled data to identify the model structure. Although sampling reduces the amount of data processed simultaneously, managing and integrating results from multiple samples can introduce additional complexity. We leave the investigation of this sampling-based approach and its implications for computational efficiency and model accuracy to future research direction.

Acknowledgments

Research reported in this publication was partially supported by the National Institute of General Medical Sciences of the National Institutes of Health (NIH) under Award Number P20 GM139769 (Xinyi Li), NIH grant Award Number R01 AG085616 (Li Wang and Guannan Wang), National Science Foundation awards DMS-2210658 (Xinyi Li) and DMS-2426173 (Li Wang), and Simons Foundation Mathematics and Physical Sciences-Collaboration Grant for Mathematicians #963447 (Guannan Wang). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the National Science Foundation.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data used in the application are publicly available from U.S. federal agency websites. Daily PM2.5 concentration data for the year 2011 were obtained from the U.S. Environmental Protection Agency (https://www.epa.gov/outdoor-air-quality-data/download-daily-data). Meteorological variables were sourced from the National Oceanic and Atmospheric Administration (https://psl.noaa.gov/).

References

Chen, J., and Z. Chen. 2008. "Extended Bayesian Information Criteria for Model Selection With Large Model Spaces." *Biometrika* 95: 759–771.

Chen, Y., Y. Bai, and W. Fung. 2017. "Structural Identification and Variable Selection in High-Dimensional Varying-Coefficient Models." *Journal of Nonparametric Statistics* 29: 258–279.

Chen, Z., D. Chen, C. Zhao, et al. 2020. "Influence of Meteorological Conditions on PM2. 5 Concentrations Across China: A Review of Methodology and Mechanism." *Environment International* 139: 105558.

Chiang, C.-T., J. A. Rice, and C. O. Wu. 2001. "Smoothing Spline Estimation for Varying Coefficient Models With Repeatedly Measured Dependent Variables." *Journal of the American Statistical Association* 96: 605–619.

Chow, J. C., L.-W. A. Chen, J. G. Watson, et al. 2006. "PM2.5 Chemical Composition and Spatiotemporal Variability During the California Regional PM10/PM2.5 Air Quality Study (CRPAQS)." *Journal of Geophysical Research: Atmospheres* 111.

Dubé, J., and D. Legros. 2013. "A Spatio-Temporal Measure of Spatial Dependence: An Example Using Real Estate Data." *Papers in Regional Science* 92: 19–31.

Fan, J., and W. Zhang. 1999. "Statistical Estimation in Varying Coefficient Models." *Annals of Statistics* 27: 1491–1518.

Fotheringham, A. S., C. Brunsdon, and E. M. Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley.

Gelfand, E. A., H. J. Kim, C. F. Sirmans, and S. Banerjee. 2003. "Spatial Modeling With Spatially Varying Coefficient Processes." *Journal of the American Statistical Association* 98: 387–396.

Harrison, R. M., and J. Yin. 2000. "Particulate Matter in the Atmosphere: Which Particle Properties Are Important for Its Effects on Health?" *Science of the Total Environment* 249: 85–101.

Hastie, T., and R. Tibshirani. 1993. "Varying-Coefficient Models." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 55: 757–779.

Huang, B., B. Wu, and M. Barry. 2010. "Geographically and Temporally Weighted Regression for Modeling Spatio-Temporal Variation in House Prices." *International Journal of Geographical Information Science* 24: 383–401.

Jacob, D. J., and D. A. Winner. 2009. "Effect of Climate Change on Air Quality." *Atmospheric Environment* 43: 51–63.

Keller, J. P., C. Olives, S.-Y. Kim, et al. 2015. "A Unified Spatiotemporal Modeling Approach for Predicting Concentrations of Multiple Air Pollutants in the Multi-Ethnic Study of Atherosclerosis and Air Pollution." *Environmental Health Perspectives* 123: 301–309.

Kim, M., and L. Wang. 2021. "Generalized Spatially Varying Coefficient Models." *Journal of Computational and Graphical Statistics* 30: 1–10.

Lai, M. J., and L. L. Schumaker. 2007. *Spline Functions on Triangulations*. 1st ed. Cambridge University Press.

Lai, M. J., and L. Wang. 2013. "Bivariate Penalized Splines for Regression." *Statistica Sinica* 23: 1399–1417.

Li, D., Y. Ke, W. Zhang, et al. 2015. "Model Selection and Structure Specification in Ultra-High Dimensional Generalised Semi-Varying Coefficient Models." *Annals of Statistics* 43: 2676–2705.

Li, X., L. Wang, and D. Nettleton. 2019. "Simultaneous Sparse Model Identification and Learning for Ultra-High-Dimensional Additive Partially Linear Models." *Journal of Multivariate Analysis* 173: 204–228.

Li, X., L. Wang, and H. J. Wang. 2021. "Sparse Learning and Structure Identification for Ultrahigh-Dimensional Image-On-Scalar Regression." *Journal of the American Statistical Association* 116: 1994–2008.

Li, X., L. Wang, H. J. Wang, and Alzheimer's Disease Neuroimaging Initiative. 2021. "Sparse Learning and Structure Identification for Ultrahigh-Dimensional Image-On-Scalar Regression." *Journal of the American Statistical Association* 116: 1994–2008.

Lian, H., P. Lai, and H. Liang. 2013. "Partially Linear Structure Selection in Cox Models With Varying Coefficients." *Biometrics* 69: 348–357.

Lian, H., J. Meng, and K. Zhao. 2015. "Spline Estimator for Simultaneous Variable Selection and Constant Coefficient Identification in High-Dimensional Generalized Varying-Coefficient Models." *Journal of Multivariate Analysis* 141: 81–103.

Livneh, B., E. A. Rosenberg, C. Lin, et al. 2013. "A Long-Term Hydrologically Based Dataset of Land Surface Fluxes and States for the Conterminous United States: Update and Extensions." *Journal of Climate* 26: 9384–9392.

Mesinger, F., G. DiMego, E. Kalnay, et al. 2006. "North American Regional Reanalysis." *Bulletin of the American Meteorological Society* 87: 343–360.

Mu, J., G. Wang, and L. Wang. 2020. "Spatial Autoregressive Partially Linear Varying Coefficient Models." *Journal of Nonparametric Statistics* 32: 428–451.

Noh, H., K. Chung, and I. Van Keilegom. 2012. "Variable Selection of Varying Coefficient Models in Quantile Regression." *Electronic Journal of Statistics* 6: 1220–1238.

Paez, M. S., D. Gamerman, F. M. Landim, and E. Salazar. 2008. "Spatially Varying Dynamic Coefficient Models." *Journal of Statistical Planning and Inference* 138: 1038–1058.

Ramsay, T. 2002. "Spline Smoothing Over Difficult Regions." *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 64: 307–319.

Stroud, J. R., P. Müller, and B. Sansó. 2001. “Dynamic Models for Spatiotemporal Data.” *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 63: 673–689.

Wang, D., and K. Kulasekera. 2012. “Parametric Component Detection and Variable Selection in Varying-Coefficient Partially Linear Models.” *Journal of Multivariate Analysis* 112: 117–129.

Wang, J., and S. Ogawa. 2015. “Effects of Meteorological Conditions on PM2. 5 Concentrations in Nagasaki, Japan.” *International Journal of Environmental Research and Public Health* 12: 9089–9101.

Wang, Y., M. Kim, S. Yu, X. Li, G. Wang, and L. Wang. 2022. “Nonparametric Estimation and Inference for Spatiotemporal Epidemic Models.” *Journal of Nonparametric Statistics* 34: 683–705.

Wikle, C. K., L. M. Berliner, and N. Cressie. 1998. “Hierarchical Bayesian Space-Time Models.” *Environmental and Ecological Statistics* 5: 117–154.

Wikle, C. K., A. Zammit-Mangion, and N. Cressie. 2019. *Spatio-Temporal Statistics With R*. Chapman and Hall/CRC.

Wood, S. N., M. V. Bravington, and S. L. Hedley. 2008. “Soap Film Smoothing.” *Journal of the Royal Statistical Society, Series B: Statistical Methodology* 70: 931–955.

Q4 Xue, L. 2009. “Consistent Variable Selection in Additive Models.” *Statistica Sinica*: 1281–1296.

Xue, T., Y. Zheng, G. Geng, et al. 2017. “Fusing Observational, Satellite Remote Sensing and Air Quality Model Simulated Data to Estimate Spatiotemporal Variations of PM2. 5 Exposure in China.” *Remote Sensing* 9: 221.

Q5 Yu, S., G. Wang, L. Wang, C. Liu, and L. Yang. 2020. “Estimation and Inference for Generalized Geoadditive Models.” *Journal of the American Statistical Association*: 1–27.

Yu, S., Y. Wang, L. Wang, and L. Gao. 2022. “Spatiotemporal Autoregressive Partially Linear Varying Coefficient Models.” *Statistica Sinica* 32: 1–28.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1.** Supporting Information.

WILEY

Author Query Form

Journal: JTSA

Article: 70002

Dear Author,

During the copyediting of your manuscript, the following queries arose.

Please refer to the query reference callout numbers in the page proofs and respond to each by marking the necessary comments using the PDF annotation tools.

Please remember illegible or unclear comments and corrections may delay publication.

Many thanks for your assistance.

AUTHOR: Please note that missing content in references have been updated where we have been able to match the missing elements without ambiguity against a standard citation database, to meet the reference style requirements of the journal. It is your responsibility to check and ensure that all listed references are complete and accurate.

Query References	Query	Remarks
Q1	AUTHOR: Figure 2 labels are a mismatch between artwork and caption, in art work has only ('a to d') part labels, but in caption part label 'e' has been described. Also the part label 'c' and 'd' descriptions have been mismatched between artwork and caption. Please check.	
Q2	AUTHOR: We have provided this standard statement as per journal style. Please modify only if you have any conflict to declare, else retain the same.	
Q3	AUTHOR: Please provide the "page range" for reference Chow et al., 2006.	
Q4	AUTHOR: Please provide the "volume number" for reference Xue, 2009.	
Q5	AUTHOR: Please provide the "volume number" for reference Yu, Wang, Wang, Liu & Yang, 2020.	