



Spatially Varying Coefficient Models with Sign Preservation of the Coefficient Functions

Myungjin KIM, Li WANG[✉], and Yuyu ZHOU

This paper considers the estimation and inference of spatially varying coefficient models, while preserving the sign of the coefficient functions. In practice, there are various situations where coefficient functions are assumed to be in a certain subspace. For example, they should be either nonnegative or nonpositive on a domain by their nature. However, optimization on a global space of coefficient functions does not ensure that estimates preserve meaningful features in their signs. In this paper, we propose sign-preserved and efficient estimators of the coefficient functions using novel bivariate spline estimators under their smoothness conditions. Our algorithm, based on the alternating direction method of multipliers, yields estimated coefficient functions that are nonnegative or nonpositive, consistent, and efficient. Simulation studies are conducted to address the advantages of the sign preservation method for a specific situation, where coefficient functions have sign constraints. Furthermore, we propose residual bootstrap-based confidence intervals for sign preserving coefficient functions over the domain of interest after adjusting the inherent bias of penalized smoothing spline techniques. Finally, we evaluate our method in a case study using air temperature, land surface temperature, and elevation in the USA.

Supplementary materials accompanying this paper appear online.

Key Words: Bernstein basis; Inequality constraints; Quadratic programming; Triangulation.

1. INTRODUCTION

Spatially varying coefficient models (SVCs) have been widely used in spatial regression problems since they are quite flexible and interpretable with various local features on a specific location. Various methods have been proposed to estimate the coefficients in the SVC. There are essentially two approaches: Bayesian approaches and frequentist approaches. Examples of the former are the SVC in the Bayesian (Gelfand et al. 2003; Banerjee et al. 2014; Hamm et al. 2015; Finley and Banerjee 2020) and Bayesian spatiotem-

M. Kim · L. Wang (✉) Department of Statistics, Iowa State University, Ames, IA 50011, USA
(E-mail: lilywang@iastate.edu). Y. Zhou, Department of Geological and Atmospheric Sciences, Iowa State University, Ames, IA 50011, USA.

poral SVCM (Bakar et al. (2015, 2016)). Examples of the latter include the geographically weighted regression (GWR) proposed by Brunson et al. (1996), the local polynomial maximum likelihood method suggested by Sun et al. (2014), and the bivariate splines over the triangulation approach of Mu et al. (2018).

In many applications, it has often been assumed that coefficient functions are nonnegative over the domain of study by their nature. For example, there is a strong positive relationship between the land surface temperature (LST) and the air temperature so that the LST is a valuable proxy for predicting the air temperature. Although impact differs through season/time (strongest during late summer and fall while weakest during winter and early spring) (Mutibwa et al. 2015) as well as other factors, such as land cover, sun exposure, and distance from the coast (Flores and Lillo 2010), an increase in the LST is expected to be related to the rise in air temperature. Also, the higher the elevation, the lower the air temperature is likely to be since high altitudes have low pressure, which results in a decrease in the air temperature.

While different estimation methods have been proposed for estimating SVCMs, due to limitations in the estimation procedure, the coefficient functions are not guaranteed to preserve their sign over the domain even though the true coefficient functions are positive or negative over their entire domain. To tackle this problem, we aim to provide an efficient estimation method for the class of SVCMs with sign preservation of the coefficient functions. Some research has been conducted concerning a sign preservation smoothing method; for example, Schumaker and Speleers (2010) for a condition for the non-negativity of the cubic Clough–Tocher macroelement and Chan and Ong (2001) for the non-negativity of a cubic Bézier triangular patch. These pure math methods are not suitable for an error based model. Most importantly, their methods and conditions are only for a single bivariate function, not for multiple bivariate functions in the SVCM, which can represent local features across different locations in a model.

In this paper, we propose novel bivariate spline estimators of the spatially varying coefficient functions that are neither complex, nor computationally intensive, under either non-negative or nonpositive constraint of coefficient functions. The spline formulation has other computational advantages as well. After finding the coefficients that minimize the penalized least squared errors, the spline function can be evaluated easily and efficiently at any desired location over the domain.

We conduct a simulation to address how the proposed sign preserving spline method works in terms of accuracy and interpretation when the true coefficient functions are non-negative. We also construct residual bootstrap-based confidence intervals for sign preservation coefficient functions on some locations after reducing the inherent bias of penalized smoothing spline methods.

The rest of the paper is organized as follows. Section 2 introduces details of the model specification and the proposed estimation method. Section 3 describes algorithm and selection of tuning parameters. In Sect. 4, we report the finite sample performance of the method using simulation studies. An application of estimating air temperature using land surface temperature and elevation in the USA is illustrated in Sect. 5. In addition, modified bootstrap-based global and individual tests are conducted to verify whether sign preservation coefficient functions are spatially varying or not in the study area.

2. MODEL AND ESTIMATION METHOD

2.1. SPATIALLY VARYING COEFFICIENT MODELS

Suppose that there are n randomly selected points, and let $\mathbf{S}_i = (S_{i1}, S_{i2})^\top$ be the i th location, $i = 1, \dots, n$, which ranges over a domain Ω . Let Y_i be the response of interest and $\mathbf{X}_i = (X_{i0}, X_{i1}, \dots, X_{ip})^\top$ with $X_{i0} \equiv 1$ being features on a certain location. Assume that $\{(\mathbf{S}_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$ satisfies the following model (Brunsdon et al. 1996)

$$Y_i = \mathbf{X}_i^\top \boldsymbol{\beta}_0(\mathbf{S}_i) + \sigma(\mathbf{S}_i)\epsilon_i = \sum_{k=0}^p X_{ik}\beta_{0,k}(\mathbf{S}_i) + \sigma(\mathbf{S}_i)\epsilon_i, \quad i = 1, \dots, n, \quad (1)$$

where $\beta_{0,k}(\cdot)$ are unknown coefficients that vary across different locations, $\sigma(\cdot)$ is the conditional standard deviation function, and ϵ_i are the random noises with $E(\epsilon_i) = 0$ and $\text{Var}(\epsilon_i) = 1$, and independent of $\mathbf{X}_i$. Let $\boldsymbol{\beta}_0(\cdot) = (\boldsymbol{\beta}_{0,1}^\top(\cdot), \boldsymbol{\beta}_{0,2}^\top(\cdot), \boldsymbol{\beta}_{0,3}^\top(\cdot))^\top$, where $\boldsymbol{\beta}_{0,1}(\cdot) = (\beta_{0,1}(\cdot), \dots, \beta_{0,p_1}(\cdot))^\top$, $\boldsymbol{\beta}_{0,2}(\cdot) = (\beta_{0,p_1+1}(\cdot), \dots, \beta_{0,p_2}(\cdot))^\top$, and $\boldsymbol{\beta}_{0,3}(\cdot) = (\beta_{0,p_2+1}(\cdot), \dots, \beta_{0,p}(\cdot))^\top$, for $0 < p_1 < p_2 < p$. Without loss of generality, we assume that in model (1), the coefficient functions, $\beta_{0,p_1+1}(\cdot), \dots, \beta_{0,p_2}(\cdot)$, are nonnegative, the coefficient functions, $\beta_{0,p_2+1}(\cdot), \dots, \beta_{0,p}(\cdot)$, are nonpositive, and the rest of them are not sign constrained. We are interested in estimating $\boldsymbol{\beta}_0(\cdot) = (\beta_{0,0}(\cdot), \beta_{0,1}(\cdot), \dots, \beta_{0,p}(\cdot))^\top$ from the observations $\{(\mathbf{S}_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$, while preserving their signs. We also would like to make inferences for $\beta_0(\cdot)$.

2.2. BIVARIATE SPLINE BASIS APPROXIMATION ON TRIANGULATION

Suppose a domain $\Omega = \cup_{i=1}^N T_i$ can be partitioned to N triangles, $T_1, \dots, T_N$. Denote a triangulation of Ω by $\Delta := \{T_1, \dots, T_N\}$. For any given triangle $T \in \Delta$, we can write $T := \langle \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \rangle$, where $\mathbf{v}_i := (v_{1i}, v_{2i})^\top$, $i = 1, 2, 3$, are three vertices of T . For any fixed location $s \in \Omega$, let b_1, b_2 , and b_3 be the barycentric coordinates of point s , which ensures that $s = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + b_3\mathbf{v}_3$ with $b_1 + b_2 + b_3 = 1$. Then, for $s \in \Omega$, we can define the Bernstein basis polynomials of degree $d \geq 1$ relative to a triangle T as $B_{ijk}^d(s) = d!/(i!j!k!)b_1^i b_2^j b_3^k$, $i + j + k = d$, $s \in T$.

Let $\mathbb{P}_d(T)$ denote the space of all polynomials of degree less than or equal to d on $T \in \Delta$. Then, any polynomial $\mathcal{P}(s) \in \mathbb{P}_d(T)$ can be uniquely represented by $\sum_{i+j+k=d} \gamma_{ijk} B_{ijk}^d(s)$ for $s \in T$ and spline coefficient $\{\gamma_{ijk}, i + j + k = d\}$; see Lai and Schumaker (2007) for more details. Let $\mathcal{C}^r(\Omega)$ be the space of the r th continuously differentiable functions over Ω . For smooth piecewise polynomials across intersection (either a shared edge or vertex) between a pair of triangles in Δ , we introduce the spline space of degree d and smoothness r over Δ in the following:

$$\mathbb{S}_{d,r}(\Delta) = \{\mathcal{P}(\cdot) \in \mathcal{C}^r(\Omega) : \mathcal{P}|_T(\cdot) \in \mathbb{P}_d(T), T \in \Delta\}, \quad (2)$$

where $\mathcal{P}|_T(\cdot)$ is the polynomial piece of spline $\mathcal{P}(\cdot)$ restricted on a triangle T .

Let $\mathcal{M}$ be an index set of the basis functions on triangulation Δ with N number of triangles, and denote $|\mathcal{M}|$ as the cardinality of $\mathcal{M}$, then, $|\mathcal{M}| = N(d+1)(d+2)/2$. For

notational convenience in the rest of paper, we use $\{B_{ijk}^d(s), i + j + k = d\} = \{B_m(s), m \in \mathcal{M}\}$ and $\{\gamma_{ijk}, i + j + k = d\} = \{\gamma_m, m \in \mathcal{M}\}$ from now on. Then, denote $\{B_m(s), m \in \mathcal{M}\}$ as a set of bivariate Bernstein basis polynomials for $\mathbb{S}_{d,r}(\Delta)$ in (2). Then, we can approximate the bivariate functions $\beta_0(\cdot) \in \mathbb{S}_{d,r}(\Delta)$ by $\mathbf{B}(s)^\top \boldsymbol{\gamma}$, where $\mathbf{B}(s) = \{B_m(s), m \in \mathcal{M}\}^\top$ and $\boldsymbol{\gamma} = \{\gamma_m, m \in \mathcal{M}\}^\top$. To ensure the global smoothness constraint in (2), we introduce the matrix $\mathbf{H}$ associated with smoothness requirement on the spline coefficients $\boldsymbol{\gamma}$ across shared edges or vertices of triangles, to be more specific, we require that $\mathbf{H}\boldsymbol{\gamma} = \mathbf{0}$; see Section S.2 in Supplementary Materials for an example of $\mathbf{H}$.

2.3. PENALIZED CONSTRAINED LEAST SQUARES METHOD

In a typical spatial analysis, data are often sparsely distributed in some regions. Appropriate penalization on spatially varying coefficient functions can help tackle such sparsity issues by borrowing features from other points. In this section, we introduce the penalized least squares method with smoothness and inequality constraints for sign preservation.

For a bivariate function $\beta(\cdot)$, we define its energy functional as follows:

$$\mathcal{E}(\beta) = \int_{\Omega} \left\{ (D_{s_1}^2 \beta(s))^2 + 2(D_{s_1} D_{s_2} \beta(s))^2 + (D_{s_2}^2 \beta(s))^2 \right\} ds,$$

where $D_{s_j}^v \beta(s)$ is the v th-order derivative in the direction s_j at any point $s = (s_1, s_2)^\top$. Next, let $\mathbb{S}_{d,r}^+(\Delta_k) = \mathbb{S}_{d,r}(\Delta_k) \cap \{\mathcal{P}(\cdot) : 0 < \mathcal{P}(\cdot) < \infty\}$ and $\mathbb{S}_{d,r}^-(\Delta_k) = \mathbb{S}_{d,r}(\Delta_k) \cap \{\mathcal{P}(\cdot) : -\infty < \mathcal{P}(\cdot) < 0\}$. Further, we denote

$$\mathcal{J}_{d,r}(\Delta_k) = \begin{cases} \mathbb{S}_{d,r}(\Delta_k), & \text{if } k = 0, \dots, p_1, \\ \mathbb{S}_{d,r}^+(\Delta_k), & \text{if } k = p_1 + 1, \dots, p_2, \\ \mathbb{S}_{d,r}^-(\Delta_k), & \text{if } k = p_2 + 1, \dots, p. \end{cases}$$

To estimate the coefficient functions in model (1) under the sign preservation constraint as described in Sect. 2.1, we solve the following constrained minimization problem:

$$\min_{\beta_{0,k}(\cdot) \in \mathcal{J}_{d,r}(\Delta_k), k=0, \dots, p} \sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p X_{ik} \beta_{0,k}(\mathbf{S}_i) \right\}^2 + \frac{1}{2} \sum_{k=0}^p \lambda_k \mathcal{E}(\beta_{0,k}), \quad (3)$$

where λ_k are the penalty parameters and can be selected based on some criteria; see Sect. 3.2 for how to select λ_k . Since the coefficient functions $\beta_{0,k}(\cdot)$'s, $k = 0, \dots, p$, are intrinsically infinite dimensional, dimension reduction is necessary. Different approaches can be applied to reduce the dimensionality, and here, we consider the bivariate spline approximation technique discussed in Sect. 2.2. Suppose that $\mathbf{Y} = (Y_1, \dots, Y_n)^\top$ and $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_n)^\top$, where $\mathbf{Z}_i = \mathbf{X}_i \otimes \mathbf{B}(\mathbf{S}_i)$ and $\mathbf{X}_i = (1, X_{i1}, \dots, X_{ip})^\top$. In addition, let $\mathbf{D}_\Lambda = \Lambda \otimes \mathbf{P}$, and $\Lambda = \text{diag}(\lambda_0, \dots, \lambda_p)$, where $\mathbf{P}$ is the block diagonal penalty matrix ensuring that $\mathcal{E}(\mathbf{B}\boldsymbol{\gamma}) = \boldsymbol{\gamma}^\top \mathbf{P}\boldsymbol{\gamma}$; see Section B.2.2 in Supplementary Materials of Yu et al. (2020) for an example of penalty matrix $\mathbf{P}$.

To reflect the smoothness conditions in (2), the minimization problem in (3) is changed into the following constrained minimization problem:

$$\begin{aligned} \min_{\boldsymbol{\gamma}} \|\mathbf{Y} - \mathbf{Z}\boldsymbol{\gamma}\|^2 + \boldsymbol{\gamma}^\top \mathbf{D}_\Lambda \boldsymbol{\gamma}, \quad \text{subject to } \mathbf{H}\boldsymbol{\gamma}_k = \mathbf{0}, \quad k = 0, \dots, p, \\ \boldsymbol{\gamma}_k > \mathbf{0}, \quad k = p_1 + 1, \dots, p_2, \quad \text{and } \boldsymbol{\gamma}_k < \mathbf{0}, \quad k = p_2 + 1, \dots, p, \end{aligned}$$

where $\mathbf{H}$ is a matrix associated with smoothness condition.

Note that we have both equality and inequality constraints in the above optimization problem. To simplify the problem, we first reparametrize the spline coefficient by $\boldsymbol{\gamma}_k = \mathbf{Q}_2 \boldsymbol{\gamma}_k^*$ based on the following QR decomposition: $\mathbf{H}^\top = \mathbf{Q}\mathbf{R} = (\mathbf{Q}_1 \mathbf{Q}_2) \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{pmatrix}$, where $\mathbf{Q}$ is an orthogonal matrix, $\mathbf{R}$ is an upper triangle matrix, $\mathbf{Q}_1$ represents the first $r = \text{rank}(\mathbf{H})$ columns of $\mathbf{Q}$, and $\mathbf{R}_2$ is a matrix of zeros. Then, it ensures that $\mathbf{H}\boldsymbol{\gamma}_k = \mathbf{0}$, $k = 0, \dots, p$. Let $\mathbf{B}^*(\mathbf{S}_i) = \mathbf{Q}_2^\top \mathbf{B}(\mathbf{S}_i)$ and $\mathbf{Z}^* = (\mathbf{Z}_1^*, \dots, \mathbf{Z}_n^*)^\top$, where $\mathbf{Z}_i^* = \mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{S}_i)$. Denote $\mathbf{D}_\Lambda^* = \mathbf{A} \otimes \mathbf{Q}_2^\top \mathbf{P} \mathbf{Q}_2$. Then, the constraint optimization problem reduces to an penalized problem only with inequality constraints:

$$\begin{aligned} \min_{\boldsymbol{\gamma}^*} \|\mathbf{Y} - \mathbf{Z}^* \boldsymbol{\gamma}^*\|^2 + \boldsymbol{\gamma}^{*\top} \mathbf{D}_\Lambda^* \boldsymbol{\gamma}^*, \\ \text{subject to } \mathbf{Q}_2 \boldsymbol{\gamma}_k^* > \mathbf{0}, \quad k = p_1 + 1, \dots, p_2, \quad \text{and } \mathbf{Q}_2 \boldsymbol{\gamma}_k^* < \mathbf{0}, \quad k = p_2 + 1, \dots, p. \end{aligned} \quad (4)$$

Then, we estimate $\beta_{0,k}(s)$ by $\hat{\beta}_k(s) = \mathbf{B}(s)^\top \hat{\boldsymbol{\gamma}}_k = \sum_{m \in \mathcal{M}} B_m(s) \hat{\gamma}_{km}$, where $\hat{\boldsymbol{\gamma}}_k = \{\hat{\gamma}_{km}, m \in \mathcal{M}\}^\top = \mathbf{Q}_2 \hat{\boldsymbol{\gamma}}_k^*$, and $\hat{\boldsymbol{\gamma}}_k^*$ is the minimizer of (4), for $k = 0, \dots, p$.

3. IMPLEMENTATION

3.1. QUADRATIC PROGRAMMING WITH INEQUALITY CONSTRAINTS

The optimization problem in (4) is the quadratic programming with inequality conditions. The main idea is based on the alternating direction method of multipliers (ADMM) algorithm (Boyd et al. 2011; Nishihara et al. 2015; Stellato et al. 2020a). Let $\mathbf{W} = \mathbf{Z}^{*\top} \mathbf{Z}^* + \mathbf{D}_\Lambda^*$, $\mathbf{A} = \mathbf{I} \otimes \mathbf{Q}_2$, and $\boldsymbol{\xi} = -\mathbf{Z}^\top \mathbf{Y}$. Then, (4) can be converted to the following constrained minimization problem:

$$\min_{\boldsymbol{\gamma}^*} \frac{1}{2} \boldsymbol{\gamma}^{*\top} \mathbf{W} \boldsymbol{\gamma}^* + \boldsymbol{\xi}^\top \boldsymbol{\gamma}^*, \quad \mathbf{A} \boldsymbol{\gamma}^* \in \mathcal{C}, \quad (5)$$

where $\mathcal{C} = \{\mathbf{x} \in \mathbf{R}^{(p+1)|\mathcal{M}|} : x_\ell \in \mathcal{C}_\ell, \ell = 1, \dots, (p+1)|\mathcal{M}|\}$, and

$$\mathcal{C}_\ell = \begin{cases} -\infty < x_\ell < \infty, & \ell = 1, \dots, (p_1 + 1)|\mathcal{M}|, \\ 0 < x_\ell < \infty, & \ell = (p_1 + 1)|\mathcal{M}| + 1, \dots, (p_2 + 1)|\mathcal{M}|, \\ -\infty < x_\ell < 0, & \ell = (p_2 + 1)|\mathcal{M}| + 1, \dots, (p + 1)|\mathcal{M}|. \end{cases}$$

Using a decision variable $\boldsymbol{y}_D^*$, and auxiliary variables $\tilde{\boldsymbol{y}}^*$ and $\tilde{\boldsymbol{y}}_D^*$, the problem in (5) can be re-expressed as:

$$\min_{(\tilde{\boldsymbol{y}}^*, \tilde{\boldsymbol{y}}_D^*)} \frac{1}{2} \tilde{\boldsymbol{y}}^{*\top} \mathbf{W} \tilde{\boldsymbol{y}}^* + \boldsymbol{\xi}^\top \tilde{\boldsymbol{y}}^* + \mathcal{I}_1(\tilde{\boldsymbol{y}}^*, \tilde{\boldsymbol{y}}_D^*) + \mathcal{I}_2(\tilde{\boldsymbol{y}}_D^*), \text{ such that } (\tilde{\boldsymbol{y}}^*, \tilde{\boldsymbol{y}}_D^*) = (\boldsymbol{y}^*, \boldsymbol{y}_D^*), \quad (6)$$

where

$$\mathcal{I}_1(\tilde{\boldsymbol{y}}^*, \tilde{\boldsymbol{y}}_D^*) = \begin{cases} 0, & \text{if } \mathbf{A} \tilde{\boldsymbol{y}}^* = \tilde{\boldsymbol{y}}_D^* \\ \infty, & \text{otherwise} \end{cases}, \text{ and } \mathcal{I}_2(\tilde{\boldsymbol{y}}_D^*) = \begin{cases} 0, & \text{if } \tilde{\boldsymbol{y}}_D^* \in \mathcal{C} \\ \infty, & \text{otherwise} \end{cases}.$$

To solve the problem (6), we consider an iteration of the ADMM, in which we update the variables $\tilde{\boldsymbol{y}}^{*(k+1)}$ and $\tilde{\boldsymbol{y}}_D^{*(k+1)}$ by solving the following problem:

$$\begin{aligned} & (\tilde{\boldsymbol{y}}^{*(k+1)}, \tilde{\boldsymbol{y}}_D^{*(k+1)}) \\ &= \arg \min_{(\tilde{\boldsymbol{y}}^*, \tilde{\boldsymbol{y}}_D^*): \mathbf{A} \tilde{\boldsymbol{y}}^* = \tilde{\boldsymbol{y}}_D^*} \frac{1}{2} \tilde{\boldsymbol{y}}^{*\top} \mathbf{W} \tilde{\boldsymbol{y}}^* + \boldsymbol{\xi}^\top \tilde{\boldsymbol{y}}^* + \frac{\sigma}{2} \left\| \tilde{\boldsymbol{y}}^* - \boldsymbol{y}^{*(k)} + \sigma^{-1} \boldsymbol{\zeta}^{(k)} \right\|_2^2 \\ & \quad + \frac{\rho}{2} \left\| \tilde{\boldsymbol{y}}_D^* - \boldsymbol{y}_D^{*(k)} + \rho^{-1} \boldsymbol{\phi}^{(k)} \right\|_2^2, \end{aligned} \quad (7)$$

$$\boldsymbol{y}^{*(k+1)} = \alpha \tilde{\boldsymbol{y}}^{*(k+1)} + (1 - \alpha) \boldsymbol{y}^{*(k)} + \sigma^{-1} \boldsymbol{\zeta}^{(k)}, \quad (8)$$

$$\boldsymbol{y}_D^{*(k+1)} = \Pi \left\{ \alpha \tilde{\boldsymbol{y}}_D^{*(k+1)} + (1 - \alpha) \boldsymbol{y}_D^{*(k)} + \rho^{-1} \boldsymbol{\phi}^{(k)} \right\}, \quad (9)$$

$$\boldsymbol{\zeta}^{(k+1)} = \boldsymbol{\zeta}^{(k)} + \sigma \left\{ \alpha \tilde{\boldsymbol{y}}^{*(k+1)} + (1 - \alpha) \boldsymbol{y}^{*(k)} - \boldsymbol{y}^{*(k+1)} \right\}, \quad (10)$$

$$\boldsymbol{\phi}^{(k+1)} = \boldsymbol{\phi}^{(k)} + \rho \left\{ \alpha \tilde{\boldsymbol{y}}_D^{*(k+1)} + (1 - \alpha) \boldsymbol{y}_D^{*(k)} - \boldsymbol{y}_D^{*(k+1)} \right\}, \quad (11)$$

where $\Pi(\cdot)$ is the Euclidean projection onto $\mathcal{C}$, $\boldsymbol{\zeta}^{(k)}$ and $\boldsymbol{\phi}^{(k)}$ are related to the dual variables of the constraints $\tilde{\boldsymbol{y}}^* = \boldsymbol{y}^*$ and $\tilde{\boldsymbol{y}}_D^* = \boldsymbol{y}_D^*$.

All updates related to (7)–(11) can be done via Algorithm 1. Note that an update of $\boldsymbol{\zeta}^{(k+1)}$ in (10) can be disregarded due to (8) and (10). The details of termination and tuning parameters are discussed in Remarks 1 and 2. The *solve_osqp* function in the *osqp* R package (Stellato et al. 2020b) is used to implement Algorithm 1.

Remark 1. For Algorithm 1, we consider the following termination criteria: $\|r_{\text{prime}}^{(k)}\|_\infty \leq \epsilon_{\text{prime}}$ and $\|r_{\text{dual}}^{(k)}\|_\infty \leq \epsilon_{\text{dual}}$, where

$$r_{\text{prime}}^{(k)} = \mathbf{A} \boldsymbol{y}^{*(k)} - \boldsymbol{y}_D^{*(k)}, \quad r_{\text{dual}}^{(k)} = \mathbf{W} \boldsymbol{y}^{*(k)} + \boldsymbol{\xi} + \mathbf{A}^\top \boldsymbol{\phi}^{(k)},$$

$$\epsilon_{\text{prime}} = \epsilon_{\text{abs}} + \epsilon_{\text{rel}} \max\{\|\mathbf{A} \boldsymbol{y}^{*(k)}\|_\infty, \|\boldsymbol{y}_D^{*(k)}\|_\infty\},$$

$$\epsilon_{\text{dual}} = \epsilon_{\text{abs}} + \epsilon_{\text{rel}} \max\{\|\mathbf{W} \boldsymbol{y}^{*(k)}\|_\infty, \|\mathbf{A}^\top \boldsymbol{\phi}^{(k)}\|_\infty, \|\boldsymbol{\xi}\|_\infty\}.$$

In our numerical studies, the absolute tolerance ϵ_{abs} and relative tolerance ϵ_{rel} are set to be 10^{-6} and 10^{-6} .

Step 1. Initialize variables $\mathbf{y}^{(0)}, \mathbf{y}_D^{(0)}, \boldsymbol{\phi}^{(0)}$ and tuning parameters $\alpha \in (0, 2), \rho > 0, \sigma > 0$.
Step 2. Based on information, update $\mathbf{W}, \mathbf{A}$, and $\boldsymbol{\xi}$.
Step 3. Set $j = 0$.
while $(\mathbf{y}^*, \mathbf{y}_D^*, \boldsymbol{\phi})$ does not converge **do**
 (i) Find $(\tilde{\mathbf{y}}^{*(j+1)}, \mathbf{v}^{(j+1)})$ by solving the following KKT matrix

$$\begin{pmatrix} \mathbf{W} + \sigma \mathbf{I} & \mathbf{A}^\top \\ \mathbf{A} & -\rho^{-1} \mathbf{I} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{y}}^{*(j+1)} \\ \mathbf{v}^{(j+1)} \end{pmatrix} = \begin{pmatrix} \sigma \mathbf{y}^{*(j)} - \boldsymbol{\xi} \\ \mathbf{y}_D^{*(j)} - \rho^{-1} \boldsymbol{\phi}^{(j)} \end{pmatrix}$$

 (ii) Update $\tilde{\mathbf{y}}_D^{*(j+1)} = \mathbf{y}_D^{*(j)} + \rho^{-1}(\mathbf{v}^{(j+1)} - \boldsymbol{\phi}^{(j)})$.
 (iii) Set $\mathbf{y}^{*(j+1)} = \alpha \tilde{\mathbf{y}}^{*(j+1)} + (1 - \alpha) \mathbf{y}^{*(j)}$.
 (iv) Obtain $\mathbf{y}_D^{*(j+1)} = \Pi \left\{ \alpha \tilde{\mathbf{y}}_D^{*(j)} + (1 - \alpha) \mathbf{y}_D^{*(j)} + \rho^{-1} \boldsymbol{\phi}^{(j)} \right\}$, where $\Pi(\cdot)$ is the Euclidean projection onto $\mathcal{C}$.
 (v) Update $\boldsymbol{\phi}^{(j+1)} = \boldsymbol{\phi}^{(j)} + \rho \left\{ \alpha \tilde{\mathbf{y}}_D^{*(j+1)} + (1 - \alpha) \mathbf{y}_D^{*(j)} - \mathbf{y}_D^{*(j+1)} \right\}$.
 (vi) Set $j = j + 1$.
end

Algorithm 1: A sign preserving optimization algorithm.

Remark 2. The ADMM algorithm has additional parameters α, σ , and ρ . The selection of these parameters is important for the number of iterations required to converge, and it is an open research area (Nishihara et al. 2015; Ghadimi et al. 2015; Giselsson and Boyd 2016). Algorithm 1 becomes a simple ADMM without an over-relaxation step in the case that $\alpha = 1$. However, the relaxation parameter α in the range from 1.5 to 1.8 has been shown to reduce the number of iteration for the convergence (Eckstein 1994; Eckstein and Ferris 1998; Ghadimi et al. 2015). If step size ρ is assumed to become fixed after a finite number of iterations, the fixed- ρ theory applied (Boyd et al. 2011). However, to improve the convergence in practice, various types of an adaptive scheme for varying step size parameters through each iteration are considered (He et al. 2000; Wang and Liao 2001; Stellato et al. 2020a). The small value of σ is commonly considered to speed up the convergence with numerical stability (Stellato et al. 2020a). Throughout simulation and case study in this paper, we choose the regularization parameter $\sigma = 10^{-6}$, the relaxation parameter $\alpha = 1.6$, and follow the adaptive scheme to update the step size automatically ρ suggested by Stellato et al. (2020a).

3.2. SMOOTHNESS PARAMETERS AND TRIANGULATION SELECTION

Given the parameters related to the ADMM algorithm in Sect. 3.1, the above optimization procedure also requires the choice of roughness penalty parameters $\boldsymbol{\lambda} = (\lambda_0, \dots, \lambda_p)^\top$ to balance between accuracy of fitting and smoothness of bivariate functions $\beta_{0,k}(\mathbf{s})$, $k = 0, \dots, p$. For roughness parameters, we use the generalized cross-validation (GCV) measure defined as

$$\text{GCV}(\boldsymbol{\lambda}) = \frac{n \|\mathbf{Y} - \mathbf{S}(\boldsymbol{\lambda}) \mathbf{Y}\|^2}{\text{tr} \{\mathbf{I} - \mathbf{S}(\boldsymbol{\lambda})\}^2}, \quad (12)$$

where $\mathbf{S}(\boldsymbol{\lambda}) = \mathbf{Z}^* \mathbf{W}^{-1} \mathbf{Z}^{*\top}$.

For a triangulation, we consider a well-balanced triangulation concerning the shape and size of triangles inside the triangulation. Lai and Schumaker (2007) suggested an optimal triangulation in terms of the maxmin angle criterion to generate such triangulation. An example of the maxmin angle method is discussed in Section S.1 of Supplementary Materials. The Delaunay triangulation algorithm is used to implement the maxmin angle method. In both the simulation studies and application analysis below, we use the `TriMesh` function within `Triangulation R` package (Wang and Lai 2019) to create triangulations.

According to Lai and Schumaker (2007), for a given smoothness $r \geq 1$, degree $d \geq 3r+2$ ensures that the bivariate spline achieves the optimal estimation power asymptotically; for example, $d = 5$ and $r = 1$. However, too high degree and smoothness would lead to unnecessary computational burden since many parameters need to be estimated. Thus, if it is not necessary to increase the smoothness, we suggest using $r = 1$ and $d = 3$, which usually provide enough accuracy for smooth functions in practice, or higher up to $d = 5$ for the best approximation power. For all numerical studies in this paper, we consider a degree up to $d = 3$ to reduce the computational cost.

3.3. POINTWISE CONFIDENCE INTERVALS FOR THE VARYING COEFFICIENT FUNCTIONS

Note that in general estimation procedures, confidence regions provide more information than a single estimate can. With the sign-preserved smoothing, however, building confidence intervals for the coefficient functions is a challenging problem, and the existing literature is rather limited and narrow in scope. In this section, we propose a wild bootstrap method to construct the pointwise confidence intervals (PCIs) for the coefficient functions; see Algorithm 2.

Due to the spline estimator's inherent bias problem, the coverage probabilities of the confidence intervals of the coefficient functions often fall below the nominal level at some locations. Instead of directly applying the penalized estimator to residual bootstrap in Algorithm 2, we consider $\lambda^* = \zeta\lambda$, $\zeta \in (0, 1)$ in order to reduce the effect of the penalty parameter. The values of ζ ranged from 0.1 to 0.2 are suggested by Dai (2017) to choose balanced performance in terms of both better coverage probability and smoothness of PCIs through locations. We suggest taking $\zeta = 0.1$ in both simulation and application.

4. SIMULATION

In this section, we conduct simulations to evaluate the performance of the proposed methodology, the spatially varying coefficient models with sign preservation (SVC-M-SP). On a horseshoe domain Ω in $[-1, 3.5] \times [-1, 1]$ rectangle, we consider two bivariate functions, $\beta_{0,0}(s)$ and $\beta_{0,1}(s)$; see Fig. 2. Note that $\beta_{0,0}(s)$ is modified from a test function (Wood et al. 2008) as described in Section S.3 of Supplementary Materials and $\beta_{0,1}(s)$ is defined as $\pi(s_1^2 + s_2^2)/10$ so that their values are all positive across the domain Ω . We consider the following model: for any $i = 1, \dots, n$,

$$Y_i = \beta_{0,0}(\mathbf{S}_i) + \beta_{0,1}(\mathbf{S}_i)X_i + \epsilon_i, \quad \beta_{0,k}(\cdot) \in \mathcal{R}^+, \quad k = 0, 1,$$

- Step 1.** Using the sample $\{(Y_i, \mathbf{X}_i, \mathbf{S}_i)\}_{i=1}^n$, fit the model and obtain the estimators of the coefficient functions $\hat{\beta}_k(\cdot)$, $k = 0, \dots, p$. Calculate the following residuals $\hat{\epsilon}_i = Y_i - \mathbf{X}_i^\top \hat{\beta}(\mathbf{S}_i)$, $i = 1, \dots, n$.
- Step 2.** Generate the bootstrap residual $\{\epsilon_i^*\}_{i=1}^n$ from the empirical distribution of the residuals $\hat{\epsilon}_i$ in Step 1, and set $Y_i^* = \mathbf{X}_i^\top \hat{\beta}(\mathbf{S}_i) + \epsilon_i^*$, $i = 1, \dots, n$.
- Step 3.** Using the bootstrap sample $\{(Y_i^*, \mathbf{X}_i, \mathbf{S}_i)\}_{i=1}^n$, refit the SVCM-SP, and obtain the bootstrap estimators, $\hat{\beta}_k^*(\cdot)$, for $k = 0, \dots, p$.
- Step 4.** Repeat Steps 2 and 3 B times and obtain a set of bootstrap estimates $\{\hat{\beta}_k^{*(b)}(\cdot)\}_{b=1}^B$. For a given location $s \in \Omega$ and $0 < \alpha < 1$, estimate $q_{\alpha/2}^*(s)$ and $q_{1-\alpha/2}^*(s)$ from the bootstrap samples, where $q_{\alpha/2}^*(s)$ and $q_{1-\alpha/2}^*(s)$ are the $\alpha/2$ th and $(1 - \alpha/2)$ th quantiles of the bootstrap distribution of $n^{1/2}\{\hat{\beta}_k^*(s) - \hat{\beta}_k(s)\}$, respectively.
- Step 5.** An asymptotic $100(1 - \alpha)\%$ bootstrap PCI is given by

$$\left\{ \hat{\beta}_k(s) - n^{-1/2} q_{1-\alpha/2}^*(s), \hat{\beta}_k(s) - n^{-1/2} q_{\alpha/2}^*(s) \right\}, s \in \Omega.$$

Algorithm 2: Algorithm for constructing bootstrap based $100(1 - \alpha)\%$ PCIs.

where $X_i \sim \text{Unif}(0, 1)$, and $\epsilon_i \sim N(0, 1)$.

For the SVCM and SVCM-SP, we consider the bivariate spline basis functions with degree $d = 2$ and 3 and smoothness $r = 1$. The roughness parameters λ_0 and λ_1 are selected by the GCV measure in (12). To investigate the effect of different triangulations, Δ_1 (40 vertices, 38 triangles), Δ_2 (71 vertices, 86 triangles), and Δ_3 (90 vertices, 112 triangles) are considered as shown in Fig. 1. For the accuracy of the estimators of the true coefficient functions, we report the approximated mean integrated squared error (AMISE), which is calculated as follows: for $k = 0, 1$, $\text{AMISE} = N_s^{-1} \sum_{i=1}^{N_s} \left\{ \hat{\beta}_k(s_i) - \beta_{0,k}(s_i) \right\}^2$, where $s_i \in \Omega$ and $i = 1, \dots, N_s$. The AMISEs of the SVCM and SVCM-SP estimators for $\beta_{0,0}(\cdot)$ and $\beta_{0,1}(\cdot)$ are calculated, and their averages over 500 simulations are given by Table 1. In general, as triangulation gets finer, the performance becomes better. However, when the triangulation is fine enough, further refining the triangulation does not significantly improve the result but increases the computation burden. Larger sample size and higher degree provide better accuracy based on the AMISE for both SVCM and SVCM-SP. Through all scenarios, SVCM-SP shows better results.

Also, we check whether the SVCM-SP and SVCM estimators are nonnegative. The average proportion of negative values for $\hat{\beta}_0(\cdot)$ and $\hat{\beta}_1(\cdot)$ according to the SVCM and SVCM-SP are given by Table 1. We can see that the proportion of negative values for $\hat{\beta}_1(\cdot)$ based on the SVCM method is between 0.12 and 0.15 for scenarios while that based on the SVCM-SP method is zero. Thus, the SVCM is not able to preserve the non-negativity, while estimates of bivariate functions using the SVCM-SP method are nonnegative. As discussed in Sect. 1, the SVCM does not have any non-negativity preservation constraints when solving the optimization problem so that the estimators of $\beta_{0,0}(\cdot)$ and $\beta_{0,1}(\cdot)$ may have negative values, even though the true values are not. In contrast, the SVCM-SP is guaranteed to be nonnegative among all the simulation scenarios under the sign preservation constraints.

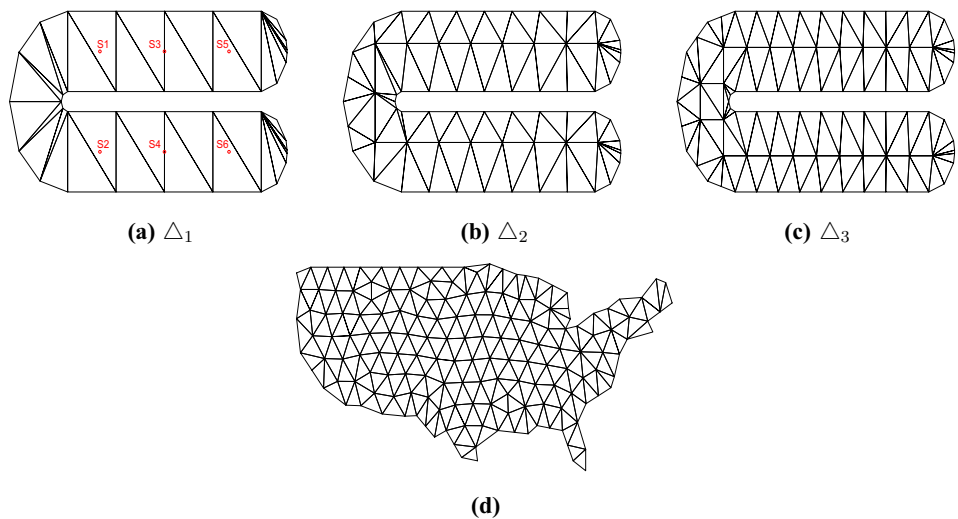


Figure 1. Triangulations considered in the simulation study and application, where the red circle in **a** represents the location selected to evaluate the 95% bootstrap PCIs; **d** triangulation used in modeling the Ta in the application (Color figure online).

Table 1. Results of the AMISE and proportion of negative values of $\widehat{\beta}_0(\cdot)$ and $\widehat{\beta}_1(\cdot)$, tenfold cross-validation MSPEs and the median computational time (in seconds) of estimation through 500 simulation iterations for the SVCN-SP and SVCN methods

	Triangulation	<i>d</i>	SVCN-SP				SVCN			
			<i>n</i> = 1000		<i>n</i> = 2000		<i>n</i> = 1000		<i>n</i> = 2000	
			$\widehat{\beta}_{0,0}$	$\widehat{\beta}_{0,1}$	$\widehat{\beta}_{0,0}$	$\widehat{\beta}_{0,1}$	$\widehat{\beta}_{0,0}$	$\widehat{\beta}_{0,1}$	$\widehat{\beta}_{0,0}$	$\widehat{\beta}_{0,1}$
AMISE	Δ_1	2	0.043	0.059	0.026	0.033	0.050	0.087	0.030	0.047
		3	0.037	0.059	0.021	0.033	0.044	0.086	0.025	0.047
	Δ_2	2	0.044	0.059	0.027	0.034	0.052	0.087	0.031	0.050
		3	0.036	0.060	0.021	0.033	0.043	0.085	0.024	0.046
	Δ_3	2	0.040	0.059	0.023	0.033	0.047	0.087	0.027	0.047
		3	0.036	0.060	0.021	0.033	0.042	0.085	0.024	0.046
Proportion of negative values	Δ_1	2	0.000	0.000	0.000	0.000	0.004	0.148	0.002	0.125
		3	0.000	0.000	0.000	0.000	0.004	0.150	0.002	0.131
	Δ_2	2	0.000	0.000	0.000	0.000	0.004	0.148	0.002	0.131
		3	0.000	0.000	0.000	0.000	0.004	0.151	0.002	0.132
	Δ_3	2	0.000	0.000	0.000	0.000	0.004	0.151	0.002	0.132
		3	0.000	0.000	0.000	0.000	0.004	0.151	0.002	0.132
MSPE	Δ_1	2	1.042		1.027		1.044		1.028	
		3	1.035		1.021		1.037		1.022	
	Δ_2	2	1.041		1.026		1.043		1.027	
		3	1.034		1.020		1.035		1.021	
	Δ_3	2	1.037		1.023		1.039		1.024	
		3	1.033		1.020		1.035		1.021	
Time	Δ_1	2	1.64		10.84		1.39		9.19	
		3	17.78		38.02		8.40		27.79	
	Δ_2	2	6.36		14.78		3.43		12.13	
		3	53.91		85.30		16.47		50.84	
	Δ_3	2	9.98		20.55		4.31		15.56	
		3	101.24		132.29		23.44		68.33	

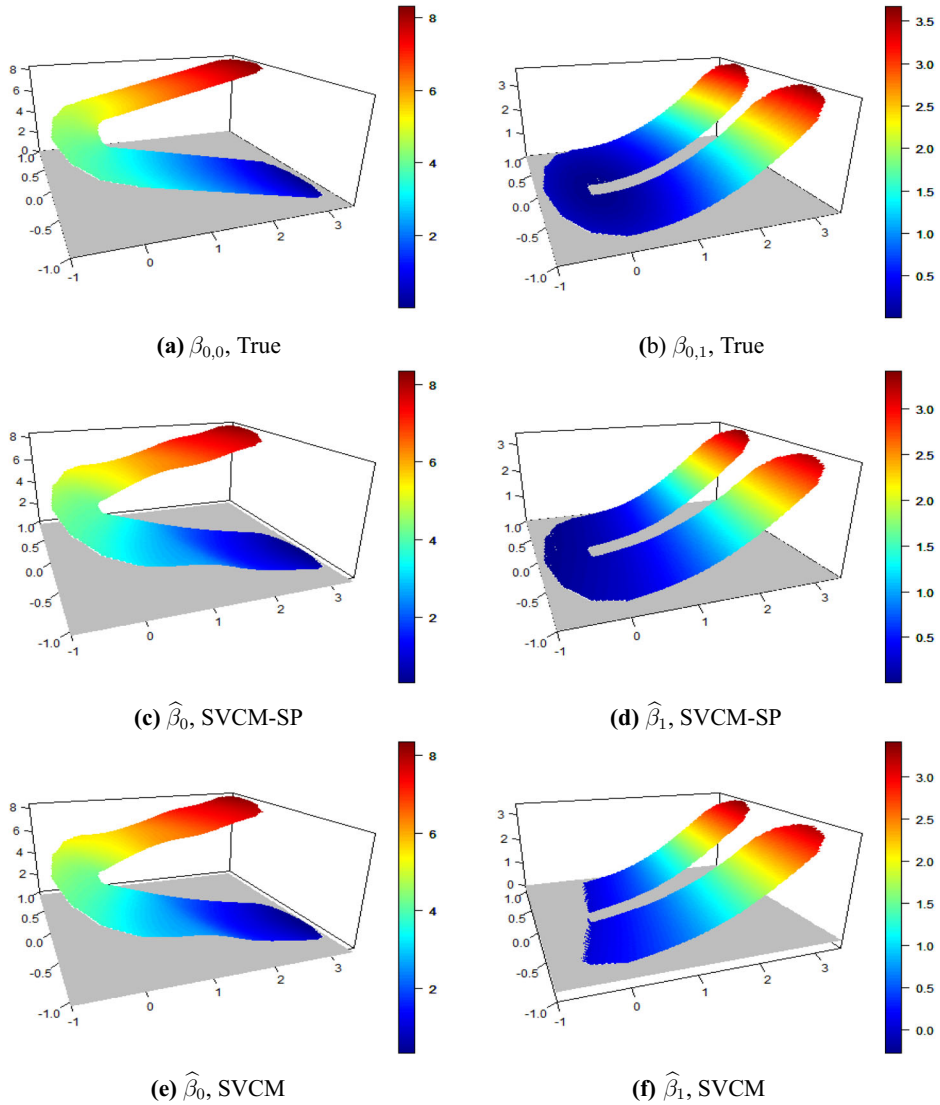


Figure 2. Surfaces of the true coefficient functions and their SVC and SVC-SP estimates based on triangulation Δ_1 , degree $d = 2$ and sample size $n = 2000$. Note that the surface colored in gray is zero, indicating values below this are negative .

Figure 2 shows that the surfaces of estimates based on SVC and SVC-SP have a similar pattern all across the domain. Especially for $\hat{\beta}_1(\cdot)$ on the left arms of the horse domain, most values based on the SVC method in Fig. 2f are negative, which is physically impossible in $\beta_{0,1}(\cdot)$ in our simulation setup. However, the SVC-SP method produces all nonnegative estimates across the domain Ω .

Next, to investigate the predictive ability for the response and computing speed of two methods, we list the results of the tenfold cross-validation mean squared prediction error (MSPE) and a median computational time of 500 simulations for estimation in Table 1. In

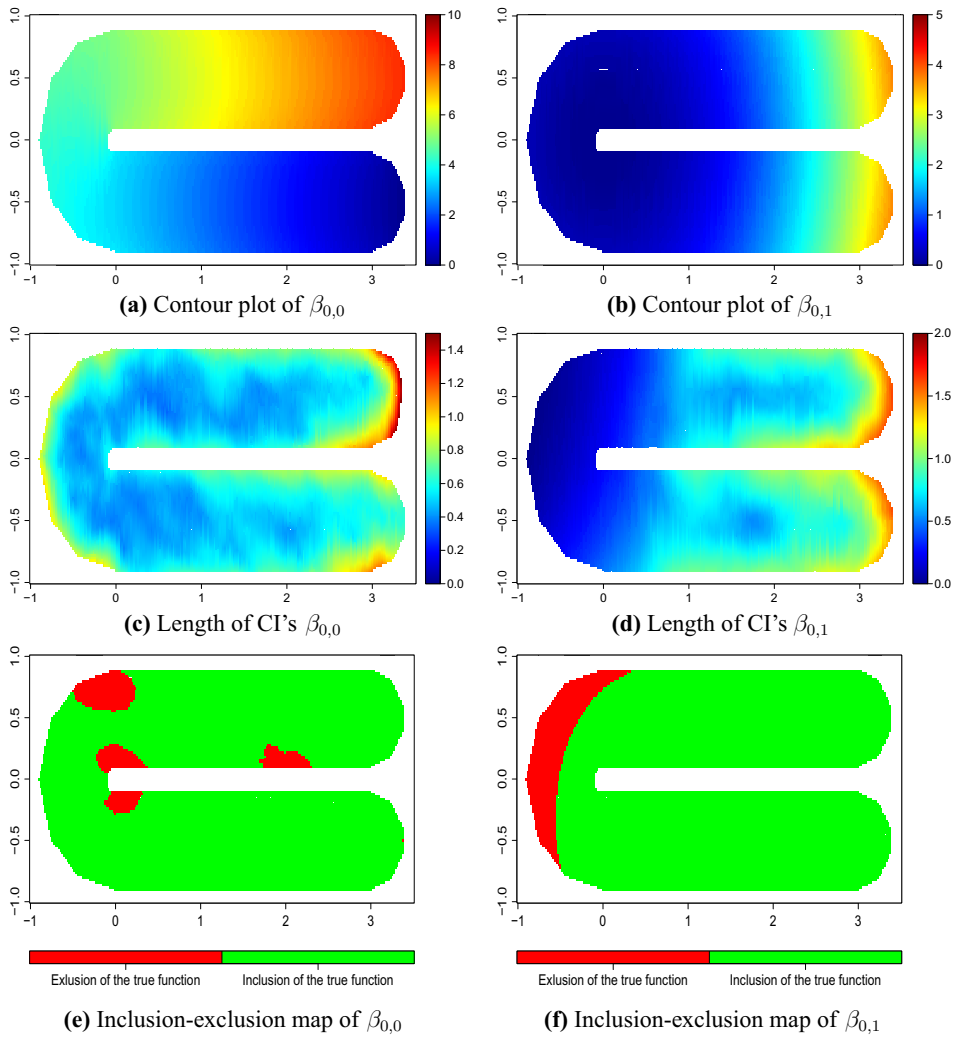


Figure 3. Contour plots of the true coefficient functions $\beta_{0,k}(\cdot)$, $k = 0, 1$, lengths of 95% PCIs and binary maps for inclusion–exclusion of the truth from 95% PCIs based on triangulation Δ_1 , degree $d = 2$ and sample size $n = 2000$.

this simulation, the SVCM-SP and the SVCM provide quite similar MSPEs. For computing time, it is reported based on a personal computer with Intel(R) Core(TM) i7-1065G7 CPU @ 1.30GHz and 8.00GB RAM. A higher degree and a finer triangulation lead to heavier computing burden compared to a lower degree and a rough triangulation. We also observe that, as n increases, the SVCM-SP and SVCM show an almost linear complexity of the sample size, although the SVCM-SP requires more computational time for the same situation due to solving inequality constraints.

Lastly, we investigate the coverage probabilities of the proposed bootstrap PCIs described in Sect. 3.3. We select six locations from the domain and build the 95% PCIs for $\beta_{0,k}(\mathbf{S}_i)$, $k = 0, 1$. See the locations of these selected points $\mathbf{S}_1$ – $\mathbf{S}_6$ on Fig. 1a. Based on a total of

Table 2. Coverage probabilities of the 95% bootstrap confidence intervals of $\beta_{0,0}(\mathbf{S}_i)$ and $\beta_{0,1}(\mathbf{S}_i)$, $i = 1, 2, \dots, 6$

Coefficient	n	Triangulation	d	$\mathbf{S}_1$	$\mathbf{S}_2$	$\mathbf{S}_3$	$\mathbf{S}_4$	$\mathbf{S}_5$	$\mathbf{S}_6$
$\beta_{0,0}(\mathbf{S}_i)$	1000	Δ_1	2	0.954	0.908	0.944	0.948	0.958	0.894
			3	0.966	0.946	0.962	0.982	0.968	0.966
		Δ_2	2	0.954	0.900	0.920	0.968	0.958	0.970
			3	0.960	0.956	0.972	0.984	0.962	0.972
		Δ_1	2	0.970	0.954	0.950	0.892	0.958	0.868
			3	0.982	0.986	0.980	0.986	0.986	0.972
	2000	Δ_2	2	0.964	0.960	0.938	0.976	0.958	0.944
			3	0.986	0.986	0.990	0.994	0.988	0.976
		Δ_1	2	0.918	0.920	0.944	0.944	0.966	0.960
			3	0.900	0.868	0.970	0.956	0.966	0.950
		Δ_2	2	0.904	0.860	0.956	0.946	0.968	0.952
			3	0.896	0.846	0.972	0.958	0.958	0.952
$\beta_{0,1}(\mathbf{S}_i)$	1000	Δ_1	2	0.924	0.940	0.964	0.940	0.960	0.950
			3	0.930	0.896	0.974	0.940	0.958	0.944
		Δ_2	2	0.914	0.882	0.980	0.944	0.958	0.950
			3	0.928	0.882	0.970	0.950	0.962	0.954
	2000	Δ_1	2	0.924	0.940	0.964	0.940	0.960	0.950
			3	0.930	0.896	0.974	0.940	0.958	0.944
		Δ_2	2	0.914	0.882	0.980	0.944	0.958	0.950
			3	0.928	0.882	0.970	0.950	0.962	0.954

100 bootstrap samples, we calculate the percentage of replications that the proposed PCIs cover the true value of $\beta_{0,k}(\mathbf{S}_i)$. The results of the coverage percentage on each selected point are given in Table 2. Although a little under-coverage is observed on some points (e.g., $\mathbf{S}_2$ for $\beta_{0,1}(\cdot)$ and $n = 1,000$), the PCIs achieve the specific coverage probability in general. Figure 3a, b displays the contour plot of the true coefficient functions. For a typical replication, Fig. 3c–f provides maps of the length of the 95% PCIs, and binary maps for inclusion–exclusion of the truth based on the corresponding 95% PCIs. Longer PCIs are observed primarily on the upper arm of the right side for $\beta_{0,0}(\mathbf{s})$, as well as on the upper and lower arms of the right side for $\beta_{0,1}(\mathbf{s})$. This is because the values of true coefficient functions are relatively larger compared to those in other locations.

5. AN APPLICATION TO THE AIR TEMPERATURE STUDY

This section illustrates the application of the proposed method to estimate air temperature using land surface temperature (LST) and elevation (Elevation).

High-spatial-resolution gridded air temperature (Ta) is essential to the study of urban climate such as urban heat islands and their impacts on building energy use (Li et al. 2019). However, current gridded air temperature data are limited and cannot capture the variation of urban air temperature due to the high heterogeneity in urban areas. Statistical models are used to estimate or interpolate gridded Ta at weather stations (Li et al. 2018b). LST is a useful auxiliary dataset in the estimation or interpolation. In this study, we test our method using a dataset of the seamless 1km resolution daily LST (Li et al. 2018a) and air temperature at weather stations in the USA.

The dataset contains the maximum Ta as the response variable, observed from the 6,969 weather stations, $\mathbf{S}_i$, on August 07, 2010. Two covariates include the standardized LST

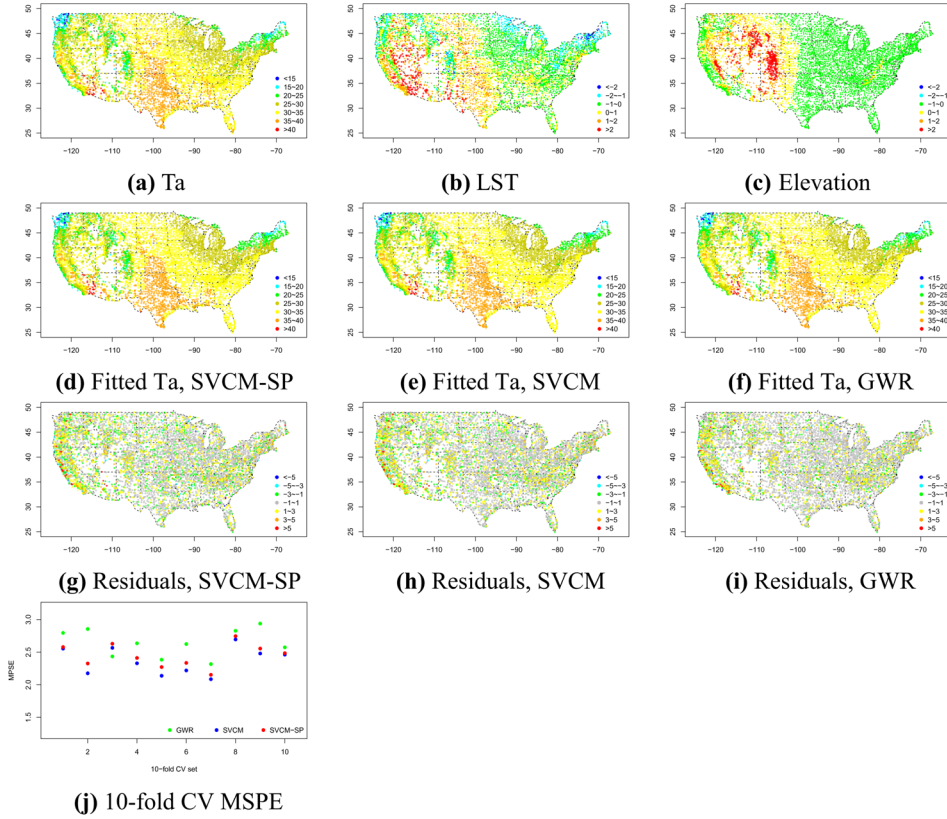


Figure 4. Scatter plots of Ta, LST, Elevation, the fitted Ta, residuals, and tenfold CV MSPEs of the SVCm-SP, SVCm, GWR.

and Elevation at the stations. As we discussed in Sect. 1, assuming that nonnegative and nonpositive constraints for $\beta_{0,1}(\cdot)$ and $\beta_{0,2}(\cdot)$, respectively, we consider the following model:

$$\text{Ta}_i = \beta_{0,0}(\mathbf{S}_i) + \beta_{0,1}(\mathbf{S}_i)\text{LST}_i + \beta_{0,2}(\mathbf{S}_i)\text{Elevation}_i + \epsilon_i, \quad (13)$$

where $\mathbf{S}_i$ is the i th weather station, $i = 1, \dots, 6969$, and $\beta_{0,0}(\cdot) \in \mathcal{R}$, $\beta_{0,1}(\cdot) \in \mathcal{R}^+$, $\beta_{0,2}(\cdot) \in \mathcal{R}^-$ are the coefficient functions for LST and Elevation. Based on the value, we classify the Ta into seven groups: (1) less than 15, (2) 15–20, (3) 20–25, (4) 25–30, (5) 30–35, (6) 35–40, and (7) greater than 40. These groups are illustrated in Fig. 4a. Similarly, Fig. 4b, c illustrates the values of LST and Elevation.

For the SVCm and the SVCm-SP, we apply the bivariate spline smoothing method over triangulation. We consider degree 3 and smoothness 1. Figure 1d displays the triangulation adopted, which contains 255 triangles and 161 vertices. For the GWR, the Gaussian kernel is considered and the cross-validation bandwidth is selected by the function `gwr.sel` of `spgwr` R package (Bivand et al. 2020). The function `gwr` in the same R package is used to fit the GWR model.

Histograms in Figure S.4 of Supplementary Materials clearly show the difference between the SVCMM and the SVCMM-SP. Based on the SVCMM, $\hat{\beta}_1(s)$ has negative values as low as -4.02 , while the SVCMM-SP completely satisfies the nonnegative property. Also, the estimator $\hat{\beta}_2(s)$ in the SVCMM ranges from -10.21 to 10.25 , while in the SVCMM-SP, it ranges from -7.56 to a small negative value close to zero. We see that the SVCMM-SP estimators $\hat{\beta}_1(s)$ and $\hat{\beta}_2(s)$, with sign preservation, behave much better than the SVCMM. Thus, the proposed SVCMM-SP method give a sensible picture of `LST` and `Elevation`, which implies that `Ta` increases with `LST` and decreases with `Elevation` on the domain of interest, even though the amount of their effects are different through locations. Scatter plots of the fitted `Ta` and residual values based on three methods are presented in Fig. 4d–i. From these plots, we see that the fitted values and residuals do not seem to show a big difference between the three methods. To compare the prediction performance of different methods, we report the tenfold cross-validation (CV) mean squared prediction error (MSPE) for different methods: the GWR (2.64), SVCMM (2.37), and SVCMM-SP (2.45). It is clear that the SVCMM and SVCMM-SP provide better performance in the tenfold CV MSPE than the GWR; see Fig. 4j for the MSPE of each fold based on three methods.

The estimates for the coefficient functions, and the upper and lower bounds of the 95% PCIs for the coefficient functions of the SVCMM-SP are given in Fig. 5. From Fig. 5e, we see that the effect of `LST` is the largest in California. From Fig. 5f, we see that in Mountain states (Colorado, Wyoming, Utah, New Mexico, Nevada, Idaho, Arizona, Montana) as well as Oregon, and Washington, `Elevation` has a significant negative effect on `Ta`. From Fig. 5d, we also notice that, for the northeastern region and the west coast of the USA, the intercept term is smaller than that in other states.

One interesting statistical question is whether some coefficients are really spatially varying or not over a domain on a subspace with sign constraints. To answer this question, we propose a bootstrap-based global test for spatial stationarity of coefficient functions, which can be conducted by comparing the residual sum of squares (RSS) from both parametric and nonparametric fittings according to reduced and full models, respectively. The null hypothesis for a global test with sign constraints is as follows:

$$H_0 : \beta_{0,k}(s) = \beta_{0,k}, \quad 0 \leq k \leq 2, \quad \beta_{0,0}(\cdot) \in \mathcal{R}, \quad \beta_{0,1}(\cdot) \in \mathcal{R}^+, \quad \beta_{0,2}(\cdot) \in \mathcal{R}^-. \quad (14)$$

The corresponding test statistic is given by $T_n = (\text{RSS}_0 - \text{RSS}_1)/\text{RSS}_1$, where $\text{RSS}_0 = \sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^2 X_{ik} \hat{\beta}_k \right\}^2$, and $\text{RSS}_1 = \sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^2 X_{ik} \hat{\beta}_k(\mathbf{S}_i) \right\}^2$ are the residual sum of squares under H_0 and under the model (13), respectively.

The null hypothesis (14) is rejected for large values of the test statistic T_n . Since the distribution of the error in the model is less straightforward, bootstrap is a possible method to test the hypothesis in (14) for the non-stationarity of coefficient functions. Algorithm 3 describes a nonparametric bootstrap procedure modified based on Cai et al. (2000) to calculate the p value of the above global test. The corresponding p value based on T_n is less than 0.05 in this application. Thus, at a significance level of 0.05, there is no significant evidence to conclude that at least one of the coefficients in the model (13) is not spatially varying through the USA.

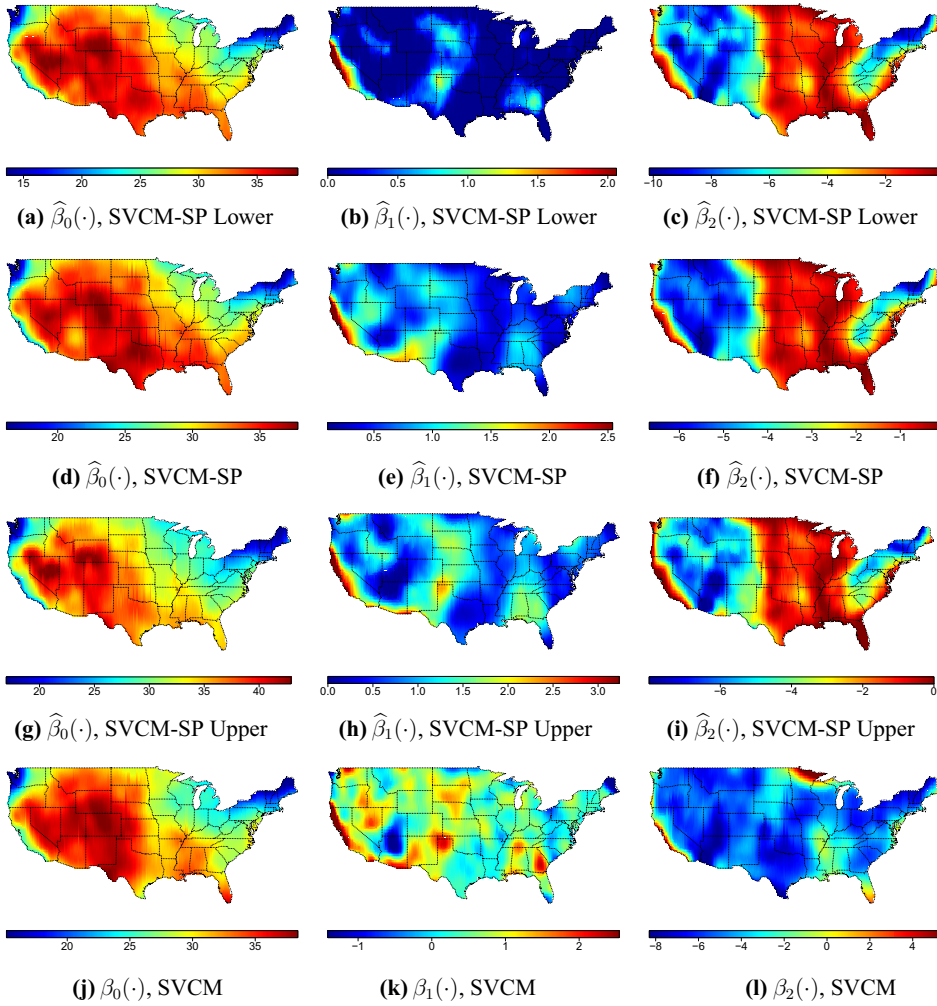


Figure 5. SVCMM-SP and SVCMM estimates for the coefficient functions, and the upper and lower bounds of the 95% confidence intervals for the coefficient functions in the application.

Step 1. Based on the data $\{(S_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$, calculate the residuals $\hat{\varepsilon}_i = Y_i - \sum_{k=0}^p X_{ik} \hat{\beta}_k(S_i)$, $i = 1, \dots, n$, from the model (13) and obtain the centered residuals $\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}$, where $\bar{\hat{\varepsilon}} = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i$;

Step 2. From the empirical distribution of the centered residuals $\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}$ defined in Step 1, generate the bootstrap residuals $\{\varepsilon_i^*\}_{i=1}^n$, and set $Y_i^* = \sum_{k=0}^p X_{ik} \hat{\beta}_k + \varepsilon_i^*$;

Step 3. Using the bootstrap sample $\{(S_i, \mathbf{X}_i, Y_i^*)\}_{i=1}^n$, evaluate the bootstrap test statistic T_n^* ;

Step 4. Repeat Steps 2 and 3 B times and obtain a set of a bootstrap test statistic $\{T_{nb}^*\}_{b=1}^B$. Then, the corresponding p value is estimated by $\hat{p} = \sum_{b=1}^B I(T_{nb}^* \geq T_n)/B$, where $I(\cdot)$ is the indicator function,

Algorithm 3: Algorithm of global test for the spatial stationarity of coefficient functions.

However, a significant p value of the above global test does not indicate which coefficients vary through a station. Thus, a natural subsequent test is an individual stationarity test for each particular coefficient function. Related hypotheses are as follows:

$$H_{0k} : \beta_{0,k}(s) = \beta_{0,k}, \quad \text{v.s.} \quad H_{1k} : \beta_{0,k}(s) \neq \beta_{0,k}, \quad \text{for } k = 0, 1, 2, \quad (15)$$

where $\beta_{0,0}(\cdot) \in \mathcal{R}$, $\beta_{0,1}(\cdot) \in \mathcal{R}^+$ and $\beta_{0,2}(\cdot) \in \mathcal{R}^-$. To conduct a hypothesis test, the test statistics are defined by $T_{nk} = (\text{RSS}_{0k} - \text{RSS}_1) / \text{RSS}_1$, $k = 0, 1, 2$, where the residual sum of squares under H_{0k} , $k = 0, 1, 2$, and H_1 in (15) are

$$\text{RSS}_{0k} = \sum_{i=1}^n \left\{ Y_i - \sum_{k' \neq k} X_{ik} \hat{\beta}_{k'}(\mathbf{S}_i) - X_{ik} \hat{\beta}_k \right\}^2, \quad \text{RSS}_1 = \sum_{i=1}^n \left\{ Y_i - \sum_{k'=0}^2 X_{ik'} \hat{\beta}_{k'}(\mathbf{S}_i) \right\}^2.$$

Algorithm 4 provides how to estimate the p value of the individual hypothesis test for $k = 0, 1, 2$. At a significance level of 0.05, for $k = 0, 1, 2$, we reject the null hypotheses H_{0k} in (15) and conclude that all the coefficient functions with sign preservation in the model (13) are spatially varying through weather stations in the USA.

Step 1. Based on the data $\{(\mathbf{S}_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$, calculate the residuals $\hat{\varepsilon}_i = Y_i - \sum_{k'=0}^2 X_{ik'} \hat{\beta}_{k'}(\mathbf{S}_i)$, $i = 1, \dots, n$, from the model (13) and obtain the centered residuals $\hat{\varepsilon}_i - \bar{\varepsilon}$, where $\bar{\varepsilon} = n^{-1} \sum_{i=1}^n \hat{\varepsilon}_i$;

Step 2. From the empirical distribution of the centered residuals $\hat{\varepsilon}_i - \bar{\varepsilon}$ defined in Step 1, generate the bootstrap residuals $\{\varepsilon_i^*\}_{i=1}^n$, and set $Y_{i,k}^* = X_{ik} \hat{\beta}_k + \sum_{k' \neq k} X_{ik'} \hat{\beta}_{k'}(\mathbf{S}_i) + \varepsilon_i^*$ for $k = 0, 1, 2$;

Step 3. Using the bootstrap sample $\{(\mathbf{S}_i, \mathbf{X}_i, Y_{i,k}^*)\}_{i=1}^n$, evaluate the bootstrap test statistic $T_{n,k}^*$ for $k = 0, 1, 2$;

Step 4. Repeat Steps 2 and 3 B times and obtain sets of a bootstrap test statistic $\{T_{nbk}^*\}_{b=1, k=0}^{B, 2}$, and the corresponding p values for k th individual tests are estimated by $\hat{p}_k = \sum_{b=1}^B I(T_{nbk}^* \geq T_{nk}) / B$, for $k = 0, 1, 2$, where $I(\cdot)$ is the indicator function.

Algorithm 4: Algorithm of individual tests for the spatial stationarity.

6. CONCLUSION AND DISCUSSION

In this paper, we propose a class of nonparametric regression models with sign-preserved spatially varying coefficient functions. The SVCN-SP method is applicable for studies where coefficient functions in a model have sign constraints, and their solutions should be in a specific subspace for the appropriate direction of the sign for functions. The proposed penalized bivariate spline method on the triangulation technique for coefficient functions is flexible and efficient. It allows different degrees and smoothness on an irregular domain with a complex boundary.

Furthermore, other approaches to estimate the coefficient function under some constraint could be more fully explored. For example, Bayesian methods with a log-Gaussian or transformation of prior may also achieve the sign preservation of coefficient functions. Note that Bayesian methods require an appropriate choice of the prior distribution of coefficients. Prior beliefs could be controversial due to subjective selection, which cannot be easily determined using a mechanical way. In practice, if prior is based on poor knowledge or experience, the Bayesian approach could generate misleading results. Compared to the hierarchical structure of Bayesian methods, our method provides an easier formulation to implement based on penalized ordinary least squares, and thus more computationally efficient. It is interesting to have a thorough comparison of our proposed method with the Bayesian approaches in future work.

This study opens several avenues for future research. First, our methods can be applied to spatiotemporal data analysis. For example, in the air temperature study, we can investigate the spatiotemporal dynamics of air temperature using land surface temperature and elevation. The method with sign preservation would be more appropriate in terms of both interpretation and accuracy for the case. Second, as inference for the coefficient function with the sign-preserved smoothing is challenging, we suggest making an inference about coefficient functions based on bootstrap procedure, which might impose a heavy computational burden. It is interesting to develop asymptotic properties of sign-preserved coefficient functions to build a more efficient inference procedure.

ACKNOWLEDGEMENTS

Funding was provided by National Science Foundation awards CCF-1934884, DMS-1916204, Laurence H. Baker Center for Bioinformatics and Biological Statistics at Iowa State University, and College of Liberal Arts and Science's (LAS) Dean's Emerging Faculty Leaders award at Iowa State University. The authors would like to thank the editor, the associate editor, and two anonymous referees for their constructive comments.

DATA AVAILABILITY STATEMENT

Data used in the simulation and application are available in electronic Supplementary Material.

[Received August 2020. Revised January 2021. Accepted February 2021. Published Online February 2021.]

REFERENCES

- Bakar KS, Kokic P, Jin H (2015) A spatiodynamic model for assessing frost risk in South-Eastern Australia. *J R Stat Soc Ser C* 64:755–778
- Bakar KS, Kokic P, Jin H (2016) Hierarchical spatially varying coefficient and temporal dynamic process models using spTDyn. *J Stat Comput Simul* 86:820–840
- Banerjee S, Carlin BP, Gelfand AE (2014) Hierarchical modeling and analysis for spatial data. CRC Press, New York

- Bivand R, Yu D, Nakaya T, Garcia-Lopez MA (2020) spgwr: geographically weighted regression. R package version 0.6-34. <https://cran.r-project.org/web/packages/spgwr/index.html>
- Boyd S, Parikh N, Chu E, Peleato B, Eckstein J (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found Trends® Mach Learn* 3:1–122
- Brunsdon C, Fotheringham AS, Charlton EM (1996) Geographically weighted regression: a method for exploring spatial nonstationarity. *Geogr Anal* 28:281–298
- Cai Z, Fan J, Yao Q (2000) Functional-coefficient regression models for nonlinear time series. *J Am Stat Assoc* 95:941–956
- Chan E, Ong B (2001) Range restricted scattered data interpolation using convex combination of cubic Bézier triangles. *J Comput Appl Math* 136:135–147
- Dai N (2017) Inference for penalized spline regression: improving confidence intervals by reducing the penalty. arXiv preprint [arXiv:1706.00865](https://arxiv.org/abs/1706.00865)
- Eckstein J (1994) Parallel alternating direction multiplier decomposition of convex programs. *J Optim Theory Appl* 80:39–62
- Eckstein J, Ferris MC (1998) Operator-splitting methods for monotone affine variational inequalities, with a parallel application to optimal control. *INFORMS J Comput* 10:218–235
- Finley AO, Banerjee S (2020) Bayesian spatially varying coefficient models in the spBayes R Package. *Environ Model Softw* 125:104608
- Flores F, Lillo M (2010) Simple air temperature estimation method from MODIS satellite images on a regional scale. *Chil J Agric Res* 70:436–445
- Gelfand AE, Kim H-J, Sirmans C, Banerjee S (2003) Spatial modeling with spatially varying coefficient processes. *J Am Stat Assoc* 98:387–396
- Ghadimi E, Teixeira A, Shames I, Johansson M (2015) Optimal parameter selection for the alternating direction method of multipliers (ADMM): quadratic problems. *IEEE Trans Autom Control* 60:644–658
- Giselsson P, Boyd S (2016) Linear convergence and metric selection for Douglas–Rachford splitting and ADMM. *IEEE Trans Autom Control* 62:532–544
- Hamm N, Finley A, Schaap M, Stein A (2015) A spatially varying coefficient model for mapping PM10 air quality at the European Scale. *Atmos Environ* 102:393–405
- He B, Yang H, Wang S (2000) Alternating direction method with self-adaptive penalty parameters for monotone variational inequalities. *J Optim Theory Appl* 106:337–356
- Lai MJ, Schumaker LL (2007) Spline functions on triangulations. Cambridge University Press, Cambridge
- Li X, Zhou Y, Asrar GR, Zhu Z (2018a) Creating a seamless 1 km resolution daily land surface temperature dataset for urban and surrounding areas in the conterminous United States. *Remote Sens Environ* 206:84–97
- Li X, Zhou Y, Asrar GR, Zhu Z (2018b) Developing a 1 km resolution daily air temperature dataset for urban and surrounding areas in the conterminous United States. *Remote Sens Environ* 215:74–84
- Li X, Zhou Y, Yu S, Jia G, Li H, Li W (2019) Urban heat island impacts on building energy consumption: a review of approaches and findings. *Energy* 174:407–419
- Mu J, Wang G, Wang L (2018) Estimation and inference in spatially varying coefficient models. *Environmetrics* 29:e2485
- Mutiibwa D, Strachan S, Albright T (2015) Land surface temperature and surface air temperature in complex terrain. *IEEE J Sel Top Appl Earth Obs Remote Sens* 8:4762–4774
- Nishihara R, Lessard L, Recht B, Packard A, Jordan M (2015) A general analysis of the convergence of ADMM. In: *Proceedings of the 32nd international conference on machine learning*, pp 343–352
- Schumaker LL, Speleers H (2010) Nonnegativity preserving macro-element interpolation of scattered data. *Comput Aided Geom Des* 27:245–261
- Stellato B, Banjac G, Goulart P, Bemporad A, Boyd S (2020a) OSQP: an operator splitting solver for quadratic programs. *Math Program Comput* 12:637–672
- Stellato B, Banjac G, Goulart P, Boyd S, Anderson E (2020b) osqp: quadratic programming solver using the 'OSQP' Library. R package version 0.6.0.3. <https://cran.r-project.org/web/packages/osqp/index.html>

- Sun Y, Yan H, Zhang W, Lu Z et al (2014) A semiparametric spatial dynamic model. *Ann Stat* 42:700–727
- Wang L, Lai MJ (2019) Triangulation. R package version 1.0. <https://github.com/funstatpackages/Triangulation>
- Wang S, Liao L (2001) Decomposition method with a variable parameter for a class of monotone variational inequality problems. *J Optim Theory Appl* 109:415–429
- Wood NS, Bravington VM, Hedley LS (2008) Soap film smoothing. *J R Stat Soc B* 70:931–955
- Yu S, Wang G, Wang L, Liu C, Yang L (2020) Estimation and inference for generalized geoaddivitive models. *J Am Stat Assoc* 115:761–774

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.