Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: 1061-8600 (Print) 1537-2715 (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

Generalized Spatially Varying Coefficient Models

Myungjin Kim & Li Wang

To cite this article: Myungjin Kim & Li Wang (2020): Generalized Spatially Varying Coefficient Models, Journal of Computational and Graphical Statistics, DOI: 10.1080/10618600.2020.1754225

To link to this article: https://doi.org/10.1080/10618600.2020.1754225

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2020, VOL. 00, NO. 0, 1–10
https://doi.org/10.1080/10618600.2020.1754225

Generalized Spatially Varying Coefficient Models

Myungjin Kim and Li Wang

Department of Statistics, Iowa State University, Ames, IA

ABSTRACT

In this article, we introduce a new class of nonparametric regression models, called generalized spatially varying coefficient models (GSVCMs), for data distributed over complex domains. For model estimation, we propose a nonparametric quasi-likelihood approach using the bivariate penalized spline approximation technique. We show that our estimation procedure is able to handle irregularly-shaped spatial domains with complex boundaries. Under some regularity conditions, the estimator for the coefficient function is proved to be consistent in the L_2 sense and its convergence rate is established. We develop a numerically stable algorithm using penalized iteratively reweighted least squares method to estimate the coefficient functions in GSVCMs. To gain efficiency in the computation for large-scale data, we further propose a QR decomposition-based algorithm, which requires only sub-blocks of the design matrix to be computed at a time, so that it allows efficient estimation of GSVCMs for large datasets with modest computer hardware. The finite sample performance of the GSVCM and its estimation method is examined by simulations studies. The proposed method is also illustrated by an analysis of the crash data in Florida. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received November 2018
Revised March 2020

KEYWORDS
Bivariate penalized spline;
Generalized cross-validation;
QR decomposition;
Quasi-likelihood;
Triangulation

1. Introduction

Varying coefficient models (VCMs) are very important tools in nonparametric regression analysis. Due to their flexibility and interpretability, VCMs have been widely applied to many scientific areas in the past three decades (see, e.g., Hastie and Tibshirani 1993; Yang et al. 2006; Fan and Zhang 2008; Park et al. 2015). In many practical fields, data are collected on a count or binary response which are generally related with the geographical locations. The spatially varying coefficient model (SVCM) allows the coefficients of the covariates to change with the locations, so it helps to efficiently investigate the spatial nonstationarity of the data (see, e.g., Brunsdon, Fotheringham, and Charlton 1996; Fotheringham, Brunsdon, and Charlton 2002; Mu, Wang, and Wang 2018). In this article, we consider the generalized spatially varying coefficient model (GSVCM), which encompasses various existing semiparametric models.

Let $\{(Y_i, \mathbf{X}_i, \mathbf{S}_i)\}_{i=1}^n = \{(Y_i, X_{i0}, X_{i1}, \ldots, X_{ip}, S_{i1}, S_{i2})\}_{i=1}^n$ (with $X_{i0} \equiv 1$) be an independent copy of size n from the joint distribution of $(Y, \mathbf{X}, \mathbf{S})$, where Y is the response variable, $\mathbf{X} = (X_0, X_1, \ldots, X_p)^\top$ is a $(p+1)$-dimensional vector of the explanatory variables with $X_0 \equiv 1$, and $\mathbf{S} = (S_1, S_2)^\top$ is the location of the observation distributed on a bounded domain $\Omega \subseteq \mathbb{R}^2$ of arbitrary shape.

In this article, we focus on the exponential dispersion families of distributions, such as binomial, Poisson and negative binomial with a fixed number of parameters for modeling purposes. The conditional density of Y given $(\mathbf{X}, \mathbf{S}) = (\boldsymbol{x}, \boldsymbol{s})$ can be represented as

$$f_{Y|\mathbf{X},\mathbf{S}}\left(y \middle| \boldsymbol{x}, \boldsymbol{s}\right) = \exp\left[\frac{1}{\sigma^2}\left\{y\xi\left(\boldsymbol{x}, \boldsymbol{s}\right) - \mathcal{B}\left\{\xi\left(\boldsymbol{x}, \boldsymbol{s}\right)\right\}\right\} + \mathcal{C}\left(y, \sigma^2\right)\right], \quad (1)$$

for some known functions $\mathcal{B}$ and $\mathcal{C}$, dispersion parameter σ^2 and the canonical parameter ξ. Let $\mu\left(\boldsymbol{x}, \boldsymbol{s}\right)$ be the conditional expectation of Y given $(\mathbf{X}, \mathbf{S}) = (\boldsymbol{x}, \boldsymbol{s})$. In the following, we assume $\mu\left(\boldsymbol{x}, \boldsymbol{s}\right)$ is modeled via a link function g as follows:

$$g\left\{\mu\left(\boldsymbol{x}, \boldsymbol{s}\right)\right\} = \boldsymbol{x}^\top \boldsymbol{\beta}(\boldsymbol{s}) = \sum_{k=0}^{p} x_k \beta_k(\boldsymbol{s}), \quad (2)$$

where $\boldsymbol{\beta}(\boldsymbol{s}) = \left\{\beta_0(\boldsymbol{s}), \ldots, \beta_p(\boldsymbol{s})\right\}^\top$ is a vector of coefficient functions to be estimated, and the conditional variance function of Y is $\mathrm{var}\left(Y|\mathbf{X} = \boldsymbol{x}, \mathbf{S} = \boldsymbol{s}\right) = \sigma^2 V\left\{\mu\left(\boldsymbol{x}, \boldsymbol{s}\right)\right\}$ for some variance function V.

The present work was motivated by the studies of road traffic crashes in Florida; see details in Section 5. The frequency of crashes on the state highway system within each census block group (see Figure 1) in the year 2014 were collected from the Florida Department of Transportation. In addition, explanatory variables, including vehicle-miles traveled and demographic variables, were collected from the Florida Department of Transportation and the U.S. Census Bureau, and many of them exhibit pronounced spatially nonstationary relationships with the response variable. For example, densely populated areas are exposed to relatively higher crash frequency compared to sparsely populated areas. It would clearly be more suitable to consider spatially varying coefficients in the model.

CONTACT Li Wang lilywang@iastate.edu Department of Statistics, Iowa State University, Ames, IA 50011.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JCGS.

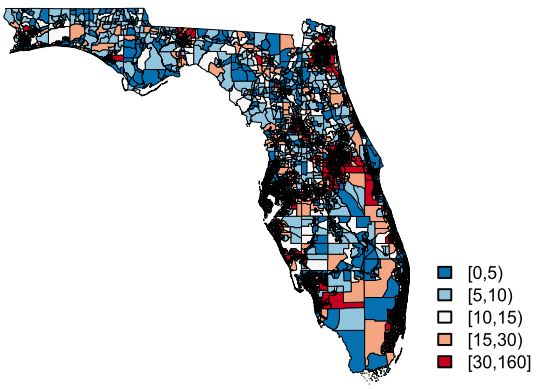

Figure 1. Crashes within census block group in Florida.

The proposed GSVCM is quite flexible and it includes many other existing models as special cases, such as the generalized linear regression model (GLM) when $\beta_k(\cdot)$'s are assumed to be constant or linear function; generalized varying coefficient model when the index variable is a scalar such as time; the SVCM when the link function is an identity function; and the generalized additive model (GAM) by Wang and Yang (2007), Liu, Yang, and Härdle (2013), and Zheng et al. (2016).

In this article, we aim to provide an efficient estimation method for the class of GSVCMs described in (2). A variety of smoothing methods are considered to estimate the coefficient functions $\beta_k(\cdot)$'s in SVCMs. A popular estimation method is the local kernel-based approach, for example, the geographically weighted regression (GWR) proposed by Brunsdon, Fotheringham, and Charlton (1996) and the local polynomial maximum likelihood method by Sun et al. (2014). Taking into account the distance between data points, the kernel-based approach incorporates the local spatial information concerning that data points nearer to the regression point have a larger weight than those farther away. However, as the sample size grows—and it is not uncommon in spatial data (e.g., image analysis and remote sensing) to have a sample size of tens of thousands to millions—the kernel method can be a computationally prohibitive process.

In reality, the spatial domain can have complex boundaries as well as interior holes. There is a vast literature devoted to handling data distributed over complex domains, such as the spatial regression models with differential regularization described in Ramsay (2002), Sangalli, Ramsay, and Ramsay (2013), Azzimonti et al. (2015), and Wilhelm and Sangalli (2016), the geodesic low-rank thin plate splines introduced by Wang and Ranalli (2007), the complex region spatial smoother described in Scott-Hayward et al. (2014), and the soap film smoothing (SOAP) proposed by Wood, Bravington, and Hedley (2008). Although those methods are shown to work well in situations where the geostatistical information is represented by a single bivariate function, they do not allow the coefficient functions of covariates to vary over the locations.

To make the estimation feasible for large datasets distributed over complex domains, in this article we propose estimating the functional coefficients using the series approximation technique. We consider triangulation technique which is known as a computationally efficient tool to handle such a domain. Specifically, we approximate $\beta_k(\cdot)$, $k = 0, \ldots, p$, using the bivariate penalized splines over triangulations (BPST) and employ the quasi-likelihood method to obtain the estimator for the coefficient functions. The BPST method has been proved to have better efficiency for data distributed over irregularly shaped domains and provide optimal rates of convergence for the nonparametric components in a standard nonparametric regression model (see, e.g., Lai and Wang 2013). We show in this article that using the bivariate spline approximation in the GSVCM solves the problem of "leakage" across the complex domains where many conventional smoothing tools suffer; that is, this is very effective at accounting for complex domain boundaries. In the situation where the effect of covariates is spatially associated with a response, we demonstrate that the proposed method results in better accuracy and predictive capabilities through numerical studies.

Other contributions of this article can be demonstrated as follows: First, the article broadens the field of applications of the VCM and SVCM. For example, Mu, Wang, and Wang (2018) conducted the analysis for the SVCM based on the BPST method, which suggested a competing method for estimation and inference for the SVCM. However, their method is limited to the case of continuous responses and linear models. Our proposed method can be applied to a variety of data types with the flexible relationship between the mean and linear predictor. Second, we support the methodology with accompanying theory, and to be more specific, we derive the L_2 convergence rate of the proposed estimators under some regularity conditions. Third, we develop an efficient algorithm using a QR decomposition of a partitioned covariate matrix to handle large datasets containing hundreds of thousands or millions of data.

The rest of the article is organized as follows. In Section 2, we introduce our estimation method and present the asymptotic properties of the proposed estimators. In Section 3, we present our estimation algorithms and discuss the details of the implementation. Section 3.1 starts with a description of a penalized iteratively reweighted least squares algorithm, which is reasonably efficient when applied to datasets with a few thousands of observations. Section 3.2 extends the method to deal with much larger datasets. We evaluate the finite sample performance of the method in Section 4 with simulation studies. An empirical analysis of the road traffic crashes study is illustrated in Section 5. The details of the proofs are deferred to the supplementary materials.

2. Methodology

For the GSVCM described in (1) and (2), the estimation of the coefficient functions can be achieved by maximizing the quasi-likelihood function, $\ell_Q(\mu, y)$, which satisfies $\frac{\partial}{\partial \mu}\ell_Q(\mu, y) = \frac{y-\mu}{\sigma^2 V(\mu)}$. Note that estimation of the varying coefficient functions requires a degree of statistical smoothing. In the following, we introduce a bivariate spline smoothing method to incorporate spatial information. We refer to Lai and Schumaker (2007) and Zhou and Pan (2014) for a detailed introduction of the triangulation technique and how to construct the bivariate spline basis functions over triangulation.

2.1. Bivariate Spline Approximation Over Triangulation

Suppose that any domain Ω can be fully covered by finitely many N triangles, $T_1, \ldots, T_N$, that is, $\Omega = \cup_{i=1}^N T_i$. A collection of these triangles, $\triangle := \{T_1, \ldots, T_N\}$, is called a triangulation of the domain Ω. For a nondegenerate triangle $T \in \triangle$, denote $|\triangle| := \max\{|T|, T \in \triangle\}$ and $\rho_\triangle := \min\{\rho_T, T \in \triangle\}$, where $|T|$ is length of the longest edge of T and ρ_T is the radius of the largest circle inscribed in T.

For a triangle $T \in \triangle$ in $\mathbb{R}^2$ with vertices $\boldsymbol{s}_i := (s_{1i}, s_{2i})^\top$, for $i = 1, 2, 3$, numbered in counter-clockwise order, we can write $T := \langle \boldsymbol{s}_1, \boldsymbol{s}_2, \boldsymbol{s}_3 \rangle$. Then, any point $\boldsymbol{s} \in \mathbb{R}^2$ can be uniquely represented as $\boldsymbol{s} = b_1\boldsymbol{s}_1 + b_2\boldsymbol{s}_2 + b_3\boldsymbol{s}_3$ such that $b_1 + b_2 + b_3 = 1$ (see Lai and Schumaker 2007, Lemma 2.1), and the coefficients (b_1, b_2, b_3) are called the "barycentric coordinates" of point $\boldsymbol{s} \in T$. The Bernstein basis polynomials of degree $d \geq 1$ relative to T are defined as $B^d_{T;ijk}(\boldsymbol{s}) = \frac{d!}{i!j!k!} b_1^i b_2^j b_3^k$, $i + j + k = d$.

Given a nonnegative integer d, write $\mathbb{P}_d(T)$ for the space of all polynomials of degree less than or equal to d on T. Then, the set of Bernstein basis polynomials forms a basis for $\mathbb{P}_d(T)$ since the barycentric coordinates b_1, b_2, b_3 of $\boldsymbol{s} \in T$ are all linear functions of the Cartesian coordinates. It follows that for nondegenerate triangle T and coefficients $\{\gamma_{T;ijk}\}$, any polynomial $\mathcal{P} \in \mathbb{P}_d(T)$ can be uniquely written as $\mathcal{P}(\boldsymbol{s})|_T = \sum_{i+j+k=d} \gamma_{T;ijk} B^d_{T;ijk}(\boldsymbol{s})$ called the B-form of $\mathcal{P}$ relative to T. Given $0 \leq r < d$ and a triangulation $\triangle$, the spline space of degree d and smoothness r over $\triangle$ is defined as $\mathbb{S}^r_d(\triangle) = \{\mathcal{P} \in \mathbb{C}^r(\Omega) : \mathcal{P}|_{T_i} \in \mathbb{P}_d(T_i), T_i \in \triangle, i = 1, \ldots, N\}$, where $\mathbb{C}^r(\Omega)$ is the space of rth continuously differentiable functions over the domain Ω.

For triangulations $\triangle_k$ with N_k triangles, $k = 0, \ldots, p$, define $\{B_{km}\}_{m \in \mathcal{M}}$ a set of bivariate Bernstein basis polynomials for $\mathbb{S}^r_d(\triangle_k)$, where $\mathcal{M}_k$ is an index set for basis functions on triangulation $\triangle_k$ with cardinality $|\mathcal{M}_k| = N_k(d+1)(d+2)/2$. Then, the bivariate functions $\beta_k \in \mathbb{S}^r_d(\triangle_k)$ in model (2) can be approximated by $\sum_{m \in \mathcal{M}_k} B_{km}(\boldsymbol{s})\gamma_{km} = \mathbf{B}_k(\boldsymbol{s})^\top \boldsymbol{\gamma}_k$, where $\mathbf{B}_k(\boldsymbol{s}) = \{B_{km}(\boldsymbol{s}), m \in \mathcal{M}_k\}^\top$ and $\boldsymbol{\gamma}_k = \{\gamma_{km}, m \in \mathcal{M}_k\}^\top$ are the vector of bivariate basis functions evaluated at a location $\boldsymbol{s}$ and the corresponding spline coefficient vector, respectively. A common strategy to reflect global smoothness restrictions in $\mathbb{S}^r_d(\triangle)$ is to introduce the constraint matrix $\mathbf{H}_k$ which satisfies $\mathbf{H}_k\boldsymbol{\gamma}_k = \mathbf{0}$, $k = 0, \ldots, p$, where $\mathbf{H}_k$ depends on smoothness parameter of the basis functions and the structure of the triangulation, and it enforces smoothness across shared edges of triangles.

2.2. Penalized Quasi-Likelihood Method

In many spatial studies, the observations are sparsely and irregularly over a domain Ω. The penalized spline method is well known as an efficient tool to deal with such a problem. In this section, we consider the penalized quasi-likelihood method, in which the energy functional is defined as follows: $\mathcal{E}(f) = \int_\Omega \left\{(\nabla^2_{s_1} f)^2 + 2(\nabla_{s_1}\nabla_{s_2} f)^2 + (\nabla^2_{s_2} f)^2\right\} ds_1 ds_2$, where $\nabla^q_{s_j} f(\boldsymbol{s})$ is the qth order derivative in the direction s_j, $j = 1, 2$, at any location $\boldsymbol{s} = (s_1, s_2)^\top$. For nonnegative smoothness parameters λ_k for $k = 0, \ldots, p$, we consider the following penalized quasi-likelihood problem:

$$\max_{\boldsymbol{\beta}} \sum_{i=1}^n \ell_Q \left[g^{-1} \left\{ \sum_{k=0}^p X_{ik}\beta_k(\mathbf{S}_i) \right\}, Y_i \right] - \frac{1}{2} \sum_{k=0}^p \lambda_k \mathcal{E}(\beta_k), \quad (3)$$

where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_p)^\top$ and $\beta_k \in \mathbb{S}^r_d(\triangle_k)$.

Note that the bivariate functions in (3) suffer from the "curse of dimensionality." To overcome this obstacle, we adopt the bivariate spline approximation technique described in Section 2.1. Although one can use different triangulations, Bernstein basis functions and constraint matrices for each bivariate function, for simplicity, we assume that $\triangle_k = \triangle$, $\mathbf{B}_k = \mathbf{B} = (\mathbf{B}(\mathbf{S}_1), \ldots, \mathbf{B}(\mathbf{S}_n))^\top$ and $\mathbf{H}_k = \mathbf{H}$ where $\mathbf{B}_k(\mathbf{S}_i) = \mathbf{B}(\mathbf{S}_i) = \{B_m(\mathbf{S}_i), m \in \mathcal{M}\}^\top$ for $i = 1, \ldots, n$ and $k = 0, \ldots, p$. By the basis expansion, the energy functional $\mathcal{E}(\beta_k)$ can be approximated by $\mathcal{E}(\mathbf{B}^\top\boldsymbol{\gamma}_k) = \boldsymbol{\gamma}_k^\top \mathbf{P}\boldsymbol{\gamma}_k$, where $\mathbf{P}$ is the block diagonal penalty matrix.

Considering the basis expansion and constraints on $\mathbf{H}_k$, $k = 0, \ldots, p$, the maximization problem (3) is changed to minimize

$$-\sum_{i=1}^n \ell_Q \left[g^{-1} \left\{ \sum_{k=0}^p X_{ik}\mathbf{B}(\mathbf{S}_i)^\top \boldsymbol{\gamma}_k \right\}, Y_i \right] + \frac{1}{2} \sum_{k=0}^p \lambda_k \boldsymbol{\gamma}_k^\top \mathbf{P}\boldsymbol{\gamma}_k \text{ subject to } \mathbf{H}\boldsymbol{\gamma}_k = \mathbf{0}. \quad (4)$$

A key to solving this constrained minimization problem is to remove the constraint via QR decomposition of $\mathbf{H}^\top$ whose rank is r, as follows: $\mathbf{H}^\top = \mathbf{QR} = (\mathbf{Q}_1\ \mathbf{Q}_2)\binom{\mathbf{R}_1}{\mathbf{R}_2}$, where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{R}$ is an upper triangle matrix, the submatrix $\mathbf{Q}_1$ is the first r columns of $\mathbf{Q}$, and $\mathbf{R}_2$ is a matrix of zeros. We reparametrize using $\boldsymbol{\gamma}_k = \mathbf{Q}_2\boldsymbol{\gamma}_k^*$ for some $\boldsymbol{\gamma}_k^*$, then it is guaranteed that $\mathbf{H}\boldsymbol{\gamma}_k = \mathbf{0}$. The constrained minimization problem (4) now reduces to an unconstrained penalized problem:

$$\min_{\boldsymbol{\gamma}^*} -\sum_{i=1}^n \ell_Q \left[g^{-1} \left\{ \sum_{k=0}^p X_{ik}\mathbf{B}(\mathbf{S}_i)^\top \mathbf{Q}_2\boldsymbol{\gamma}_k^* \right\}, Y_i \right] + \frac{1}{2} \sum_{k=0}^p \lambda_k \boldsymbol{\gamma}_k^{*\top} \mathbf{Q}_2^\top \mathbf{P}\mathbf{Q}_2\boldsymbol{\gamma}_k^*, \quad (5)$$

where $\boldsymbol{\gamma}^* = (\boldsymbol{\gamma}_0^{*\top}, \ldots, \boldsymbol{\gamma}_p^{*\top})^\top$. Let $\widehat{\boldsymbol{\gamma}}^* = (\widehat{\boldsymbol{\gamma}}_0^{*\top}, \ldots, \widehat{\boldsymbol{\gamma}}_p^{*\top})^\top$ be the minimizers of (5), and for $k = 0, \ldots, p$, let $\widehat{\boldsymbol{\gamma}}_k = \{\widehat{\gamma}_{km}, m \in \mathcal{M}\}^\top = \mathbf{Q}_2\widehat{\boldsymbol{\gamma}}_k^*$. Then, the bivariate spline estimators of $\beta_k(\boldsymbol{s})$ are $\widehat{\beta}_k(\boldsymbol{s}) = \mathbf{B}(\boldsymbol{s})^\top\widehat{\boldsymbol{\gamma}}_k = \sum_{m \in \mathcal{M}} B_m(\boldsymbol{s})\widehat{\gamma}_{km}$, for $k = 0, \ldots, p$.

2.3. Asymptotic Results

Let $\eta(\boldsymbol{x}, \boldsymbol{s}) = \sum_{k=0}^p x_k\beta_k(\boldsymbol{s})$, $\eta_i^0 = \eta(\mathbf{X}_i, \mathbf{S}_i)$, and let $\varepsilon_i = Y_i - g^{-1}(\eta_i^0)$ be the error term. For the quasi-likelihood function, $\ell_Q\{g^{-1}(x), y\}$, denote $q_1(\eta, y) = \frac{\partial}{\partial\eta}\ell_Q\{g^{-1}(\eta), y\} = \{y - g^{-1}(\eta)\}\rho_1(\eta)$, and $q_2(\eta, y) = \frac{\partial^2}{\partial\eta^2}\ell_Q\{g^{-1}(\eta), y\} = \{y - g^{-1}(\eta)\}\rho_1'(\eta) - \rho_2(\eta)$, where $\rho_j(\eta) = \{\frac{\partial}{\partial\eta}g^{-1}(\eta)\}^j/[\sigma^2 V\{g^{-1}(\eta)\}] = \left[\{g'(g^{-1}(\eta)\}^j\sigma^2 V\{g^{-1}(\eta)\}\right]^{-1}$, $j = 1, 2$. For a vector valued function $\boldsymbol{\phi} = (\phi_0, \ldots, \phi_p)^\top$, let $\|\boldsymbol{\phi}\|_{L_2} = \left\{\sum_{k=0}^p \|\phi_k\|^2_{L_2}\right\}^{1/2}$ and $\|\boldsymbol{\phi}\|_\infty = \max_{0 \leq k \leq p} \|\phi_k\|_\infty$. Define $|\phi|_{\upsilon,\infty} = \max_{i+j=\upsilon} \|\nabla^i_{s_1}\nabla^j_{s_2}\phi\|_\infty$ for a nonnegative integer υ.

The following are the technical assumptions needed to facilitate the technical details, though they may not be the weakest conditions.

(A1) For any $k = 0, 1, \ldots, p$, the bivariate function $\beta_k \in \mathcal{W}^{d+1,\infty}(\Omega)$ for an integer $d \geq 1$, where $\mathcal{W}^{d+1,\infty}(\Omega) = \{\phi : |\phi|_{k,\infty} < \infty, 0 \leq k \leq d+1\}$ is the Sobolev space.

(A2) The density function $f_{\boldsymbol{s}}(\cdot)$ of $\mathbf{S}$ is bounded away from zero and infinity on Ω.

(A3) The function $q_2(x, y) < 0$ 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 $c_\rho \leq \rho_2(\cdot) \leq C_\rho$. For each $(\boldsymbol{x}, \boldsymbol{s})$, $\text{var}(Y|\mathbf{X} = \boldsymbol{x}, \mathbf{S} = \boldsymbol{s})$ and $g'(\mu(\boldsymbol{x}, \boldsymbol{s}))$ are nonzero.

(A4) The errors satisfy $\mathrm{E}\{\varepsilon_i|\mathbf{X}_i = \boldsymbol{x}, \mathbf{S}_i = \boldsymbol{s}\} = 0$ and $\mathrm{E}(|\varepsilon_i|^{2+\iota}|\mathbf{X}_i = \boldsymbol{x}, \mathbf{S}_i = \boldsymbol{s}) < \infty$ for some $\iota \in (1/2, \infty)$. $(\mathbf{X}_i^\top, \mathbf{S}_i, Y_i)$ are independently and identically distributed.

(A5) For any $k = 0, \ldots, p$, there exists a positive constant C_k such that $|X_k| \leq C_k$. The eigenvalues $\psi_0(\boldsymbol{s}) \leq \psi_1(\boldsymbol{s}) \leq \cdots \leq \psi_p(\boldsymbol{s})$ of $\boldsymbol{\Sigma}(\boldsymbol{s}) = \mathrm{E}(\boldsymbol{X}\boldsymbol{X}^\top|\mathbf{S} = \boldsymbol{s})$ are bounded away from 0 and infinity uniformly for all $\boldsymbol{s} \in \Omega$; that is, there are positive constants C_1 and C_2 such that $C_1 \leq \psi_0(\boldsymbol{s}) \leq \psi_1(\boldsymbol{s}) \leq \cdots \leq \psi_p(\boldsymbol{s}) \leq C_2$ for all $\boldsymbol{s} \in \Omega$.

(A6) The triangulation $\triangle$ is π-quasi-uniform, that is, there exists a positive constant π such that $|\triangle|/\rho_\triangle \leq \pi$, where $|\triangle| = \max\{|T|, T \in \triangle\}$ and $\rho_\triangle = \min\{\rho_T\}$.

(A7) The triangulation size satisfies that $|\triangle| \to 0$, $|\triangle|^{-2} \log n/n \to 0$, and the smoothness penalty parameter satisfies that $\lambda_{\max} n^{-1}|\triangle|^{-4} \to 0$, where $\lambda_{\max} = \max\{\lambda_0, \ldots, \lambda_p\}$.

The above assumptions are mild conditions that can be satisfied in many practical situations. Assumption (A1) describes the smoothness requirement on the true coefficient functions, which are frequently used in the literature of nonparametric estimation. Assumption (A2) is a standard condition on the distribution of spatial sites. Assumption (A3) is a commonly used condition in the quasi-likelihood method (see, e.g., Carroll et al. 1997; Wang et al. 2011; Wang and Cao 2018). Particularly, the condition $q_2(x, y) < 0$ in (A3) is to ensure the uniqueness of the solution. Assumptions (A2) and (A4) are similar to Assumptions (A3) and (A4) in Liu, Yang, and Härdle (2013). The condition on eigenvalues in (A5) is essentially a requirement that the vector $(1, X_{i1}, \ldots, X_{ip})^\top$ is not multicolinear. Assumptions (A1), (A2), and (A5) are regularity conditions in the nonparametric regression literature, for example, they are similar to conditions used in Mu, Wang, and Wang (2018). Assumption (A6) is widely used in the triangulation based literature (see, e.g., Lai and Wang 2013; Wang et al. 2020). For example, the class of all triangulations whose smallest angles are bounded away from zero by a positive constant have this property. Assumption (A7) gives the conditions on smoothing parameters and triangulation size in the BPST estimation.

The following theorem provides the L_2 convergence rate of the spline estimators, $\widehat{\beta}_k(\boldsymbol{s})$, for $k = 0, 1, \ldots, p$. The detailed proof is illustrated in the supplementary materials.

Theorem 1. Under Assumptions (A1)–(A7), the bivariate spline estimators, $\widehat{\beta}_k(\boldsymbol{s})$, $k = 0, \ldots, p$, satisfy that

$$\sum_{k=0}^{p} \|\widehat{\beta}_k - \beta_k\|_{L_2} = O_{\text{a.s.}}\left\{|\triangle|^{-1}\left(\frac{\log n}{n}\right)^{1/2} + |\triangle|^{d+1} + \frac{\lambda_{\max}}{n|\triangle|^4}\right\},$$

as $n \to \infty$, where $\lambda_{\max} = \max\{\lambda_0, \lambda_1, \ldots, \lambda_p\}$.

3. Implementation

3.1. A Penalized Iteratively Reweighted Least Squares Algorithm

Let $\mathbf{Y} = (Y_1, \ldots, Y_n)^\top$ and $\mathbf{X}_i = (1, X_{i1}, \ldots, X_{ip})^\top$ be the vector of the response variable and the ith row vector of the design matrix. In addition, denote $\mathbf{Z} = (\mathbf{Z}_1, \ldots, \mathbf{Z}_n)^\top$, where $\mathbf{Z}_i = \mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{S}_i)$ and $\mathbf{B}^*(\mathbf{S}_i) = \mathbf{Q}_2^\top \mathbf{B}(\mathbf{S}_i)$. Let $\boldsymbol{\eta}(\boldsymbol{\gamma}^*) = \{\eta_i\}_{i=1}^n = \left\{\sum_{k=0}^p X_{ik}\mathbf{B}(\mathbf{S}_i)^\top \mathbf{Q}_2 \boldsymbol{\gamma}_k^*\right\}_{i=1}^n$. Furthermore, denote the mean vector $\boldsymbol{\mu}(\boldsymbol{\gamma}^*) = \{\mu_i\}_{i=1}^n = \left\{g^{-1}(\eta_i)\right\}_{i=1}^n$, the variance function matrix $\mathbf{V} = \text{diag}\{V(\mu_i)\}_{i=1}^n$, the diagonal matrix with elements of the derivative of a link function $\mathbf{G} = \text{diag}\{g'(\mu_i)\}_{i=1}^n$, and the weight matrix $\mathbf{W} = \mathbf{V}^{-1}\mathbf{G}^{-2} = \text{diag}\left[\{V(\mu_i)g'(\mu_i)^2\}^{-1}\right]_{i=1}^n$.

As described in Section 2, the estimators of the coefficient functions $\beta_k(\cdot)$ can be obtained via the minimization in (5). Instead of solving the minimization directly, we propose the following penalized iteratively reweighted least squares (PIRLS) procedure. Suppose that we have estimates $\boldsymbol{\gamma}^{*(j)}$, $\boldsymbol{\mu}^{(j)} = \boldsymbol{\mu}(\boldsymbol{\gamma}^{*(j)})$, $\boldsymbol{\eta}^{(j)} = \boldsymbol{\eta}(\boldsymbol{\gamma}^{*(j)})$, and $\mathbf{V}^{(j)}$ at the current jth iteration. Then, one can consider the following objective function at $(j+1)$th iteration: $L_P^{(j+1)} = \left\|\{\mathbf{V}^{(j)}\}^{-1/2}\{\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\gamma}^*)\}\right\|^2 + \frac{1}{2}\sum_{k=0}^p \lambda_k \boldsymbol{\gamma}_k^{*\top}\mathbf{Q}_2^\top \mathbf{P}\mathbf{Q}_2\boldsymbol{\gamma}_k^*$.

Note that the expectation function $\boldsymbol{\mu}(\boldsymbol{\gamma}^*)$ can be replaced by with its first-order Taylor expansion around $\boldsymbol{\gamma}^{*(j)}$ so that

$$\begin{aligned} L_P^{(j+1)} &\approx \left\|\{\mathbf{V}^{(j)}\}^{-1/2}\left[\mathbf{Y} - \boldsymbol{\mu}^{(j)} - \{\mathbf{G}^{(j)}\}^{-1}\mathbf{Z}\left(\boldsymbol{\gamma}^* - \boldsymbol{\gamma}^{*(j)}\right)\right]\right\|^2 \\ &\quad + \frac{1}{2}\sum_{k=0}^{p} \lambda_k \boldsymbol{\gamma}_k^{*\top}\mathbf{Q}_2^\top \mathbf{P}\mathbf{Q}_2\boldsymbol{\gamma}_k^* \\ &= \left\|\left\{\mathbf{W}^{(j)}\right\}^{1/2}\left(\widetilde{\mathbf{Y}}^{(j)} - \mathbf{Z}\boldsymbol{\gamma}^*\right)\right\|^2 \\ &\quad + \frac{1}{2}\sum_{k=0}^{p} \lambda_k \boldsymbol{\gamma}_k^{*\top}\mathbf{Q}_2^\top \mathbf{P}\mathbf{Q}_2\boldsymbol{\gamma}_k^*, \end{aligned} \tag{6}$$

where $\widetilde{\mathbf{Y}}^{(j)} = (\widetilde{Y}_1^{(j)}, \ldots, \widetilde{Y}_n^{(j)})^\top$ with $\widetilde{Y}_i^{(j)} = g'(\mu_i^{(j)})(Y_i - \mu_i^{(j)}) + \eta_i^{(j)}$. The PIRLS procedure is represented in Algorithm 1 below. In the numerical analysis below, we set $\mu_i^{(0)} = Y_i + 0.1$ and $\eta_i^{(0)} = g(\mu_i^{(0)})$ as the initial values to start the iteration.

3.2. An Algorithm for Large-Scale Data

Although our method works efficiently when applied to datasets containing up to a few thousands of observations, it tends to be slow for large datasets containing hundreds of thousands or millions of data. In this section, we develop an efficient algorithm using a QR decomposition of a partitioned covariate matrix

Step 1. Starting with initial values $\boldsymbol{\eta}^{(0)}$ and $\boldsymbol{\mu}^{(0)}$, for $i = 1, \ldots, n$, calculate $g'(\mu_i^{(0)})$ and $V(\mu_i^{(0)})$ to have initial weight matrix $\mathbf{W}^{(0)}$ and working variable $\widetilde{\mathbf{Y}}^{(0)}$.
Step 2. Let J be the number of iteration that the sequence of $\boldsymbol{\gamma}^*$'s converges.
foreach $1 \le j \le J-1$ **do**
(i) Using the current $\boldsymbol{\eta}^{(j)}$ and $\boldsymbol{\mu}^{(j)}$, for $i = 1, \ldots, n$, update $g'(\mu_i^{(j)})$ and $V(\mu_i^{(j)})$ for current weight matrix $\mathbf{W}^{(j)}$ and working variable $\widetilde{\mathbf{Y}}^{(j)}$.
(ii) Find the solution by minimizing the objective function in (6) with respect to $\boldsymbol{\gamma}^*$ and set $\boldsymbol{\gamma}^{*(j+1)}$ to the solution and then update $\boldsymbol{\eta}^{(j+1)} = \boldsymbol{\eta}(\boldsymbol{\gamma}^{*(j+1)})$ and $\boldsymbol{\mu}^{(j+1)} = \boldsymbol{\mu}(\boldsymbol{\gamma}^{*(j+1)})$.
(iii) Set j to $j+1$.
end

Algorithm 1: The PIRLS algorithm.

to reduce the computing burden especially for large datasets. Suppose that the design matrix in our model is decomposed into $\mathbf{W}^{1/2}\mathbf{Z} = \mathbf{Q}^*\mathbf{R}^*$, where $\mathbf{Q}^*$ is a $n \times (p+1)|\mathcal{M}|$ column orthogonal matrix and $\mathbf{R}^*$ is a $(p+1)|\mathcal{M}| \times (p+1)|\mathcal{M}|$ upper triangular matrix. If $\mathbb{Y} = \mathbf{Q}^{*\top}\mathbf{W}^{1/2}\widetilde{\mathbf{Y}}$ and $\|\mathbf{v}\|^2 = \|\mathbf{W}^{1/2}\widetilde{\mathbf{Y}}\|^2 - \|\mathbb{Y}\|^2$, the expression in (6) becomes

$$\left\|\mathbb{Y} - \mathbf{R}^*\boldsymbol{\gamma}^*\right\|^2 + \|\mathbf{v}\|^2 + \frac{1}{2}\sum_{k=0}^{p}\lambda_k\boldsymbol{\gamma}_k^{*\top}\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\boldsymbol{\gamma}_k^*. \quad (7)$$

Algorithm 2 presents the modified PIRLS algorithm of our proposed method based on the QR-updating method, referred to as the QRGSVCM. It is very efficient for large datasets when the number of parameters to be estimated is relatively small. Because this algorithm relies on performance-oriented iteration in Step 3 (∗), its convergence is not guaranteed especially for small- or moderate-sized datasets. However, the kind of ill-conditioning that leads to convergence issues tends to decrease as the sample increases; see the discussions in Wood, Goude, and Shaw (2015).

Step 1. Construct M non-overlapping subsets $\pi_1, \ldots, \pi_M$ such that $\cup_m \pi_m = \{1, \ldots, n\}$ and $\pi_m \cap \pi_{m'} = \phi$ for all $m \ne m'$.
Step 2. Set $\boldsymbol{\eta}^{(0)}$ as an initial value and $r = 0$. Also, set $\mathbf{R}^*$ and $\mathbb{Y}$ to be the null matrix and vector to stack up the following sub-matrix and vector, respectively.
Step 3. Let J be the number of iteration that the sequence of $\boldsymbol{\gamma}^*$'s converges.
foreach $1 \le j \le J-1$ **do**
foreach $1 \le m \le M$ **do**
(i) Set $\mathbf{R}_0^* = \mathbf{R}^*$, $\mathbb{Y}_0 = \mathbb{Y}$ and $\mathbb{Z}_m = \{\mathbf{Z}_i : i \in \pi_m\}$.
(ii) Using the current $\widehat{\boldsymbol{\eta}} = \mathbb{Z}_m\widehat{\boldsymbol{\gamma}}^* = \{\widehat{\eta}_i\}_{i\in\pi_m}$ and $\widehat{\boldsymbol{\mu}} = g^{-1}(\widehat{\boldsymbol{\eta}})$, for $i \in \pi_m$, update $g'(\widehat{\mu}_i)$, and $V(\widehat{\mu}_i)$ for current weight matrix $\mathbf{W} = \text{diag}\left[\{V(\mu_i)g'(\mu_i)^2\}^{-1}\right]_{i\in\pi_m}$ and $\widetilde{\mathbb{Y}} = \{g'(\widehat{\mu}_i)(y_i - \widehat{\mu}_i) + \widehat{\eta}_i\}_{i\in\pi_m}$.
(iii) Set $r = r + \|\mathbf{W}^{1/2}\widetilde{\mathbb{Y}}\|^2$ and construct $\begin{pmatrix}\mathbf{R}_0^* \\ \mathbf{W}^{1/2}\mathbb{Z}_m\end{pmatrix} = \mathbf{Q}^*\mathbf{R}^*$, and $\mathbb{Y} = \mathbf{Q}^{*\top}\begin{pmatrix}\mathbb{Y}_0 \\ \mathbf{W}^{1/2}\widetilde{\mathbb{Y}}\end{pmatrix}$.
end
(∗) Set $\|\boldsymbol{v}\|^2 = r - \|\mathbb{Y}\|^2$ and estimate the smoothing parameter $\boldsymbol{\lambda}$. Find the solution by minimizing the objective function in (7) w.r.t. $\boldsymbol{\gamma}^*$ and set $\widehat{\boldsymbol{\gamma}}^*$ to be the solution.
end

Algorithm 2: QR decomposition-based PIRLS algorithm.

3.3. Smoothing Parameter and Triangulation Selections

Note that, in the PIRLS procedure above, selecting appropriate smoothing parameters $\boldsymbol{\lambda} = (\lambda_0, \ldots, \lambda_p)^\top$ is important because they control the balance between data fitting and the variability of the functions, $\beta_k(\boldsymbol{s})$, $k = 0, \ldots, p$, so that the algorithm does not lead to too rough or too smooth surfaces. The cross-validation (CV) is a popular method to choose the smoothing parameters $\boldsymbol{\lambda}$. However, due to its computational burden and tendency to under-smooth the data, the generalized cross-validation (GCV) criterion is commonly considered in a wide range of applications. Let $\widetilde{\mathbf{Y}} = \left\{g'(\mu_i)(Y_i - \mu_i) + \eta_i\right\}_{i=1}^n$ and $\mathbf{D}_\Lambda = \Lambda \otimes \mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2$, where $\Lambda = \text{diag}(\lambda_0, \lambda_1, \ldots, \lambda_p)$. Let $\mathbf{M}(\boldsymbol{\lambda}) = \mathbf{Z}^\top\mathbf{W}\mathbf{Z} + \mathbf{D}_\Lambda$ and denote the smoothing matrix as $\mathbf{S}(\boldsymbol{\lambda}) = \mathbf{Z}\{\mathbf{M}(\boldsymbol{\lambda})\}^{-1}\mathbf{Z}^\top\mathbf{W}$. The smoothing parameters $\boldsymbol{\lambda}$ can be chosen by minimizing the following the GCV criterion

$$\text{GCV}(\boldsymbol{\lambda}) = \frac{n^{-1}\|\mathbf{W}^{1/2}\{\widetilde{\mathbf{Y}} - \mathbf{S}(\boldsymbol{\lambda})\widetilde{\mathbf{Y}}\}\|^2}{\left[n^{-1}\text{tr}\{\mathbf{I} - \mathbf{S}(\boldsymbol{\lambda})\}\right]^2} \quad (8)$$

over a grid of values to select $\boldsymbol{\lambda}$.

For the QRGSVCM method in Section 3.2, the following two modified measures based on QR decomposition can be used for smoothness parameters $\boldsymbol{\lambda}$ selection:

$$\text{GCV}(\boldsymbol{\lambda}) = \frac{n(\|\mathbb{Y} - \mathbf{R}^*\widehat{\boldsymbol{\gamma}}^*\|^2 + \|\mathbf{v}\|^2)}{[n - \text{tr}\{\mathbf{H}(\boldsymbol{\lambda})\}]^2} \text{ or}$$
$$C_p(\boldsymbol{\lambda}) = \|\mathbb{Y} - \mathbf{R}^*\widehat{\boldsymbol{\gamma}}^*\|^2 + \|\mathbf{v}\|^2 + 2\widehat{\sigma}^2\text{tr}\{\mathbf{H}(\boldsymbol{\lambda})\}, \quad (9)$$

where $\widehat{\sigma}^2 = \left(\|\mathbb{Y} - \mathbf{R}^*\widehat{\boldsymbol{\gamma}}^*\|^2 + \|\mathbf{v}\|^2\right) / [n - \text{tr}\{\mathbf{H}(\boldsymbol{\lambda})\}]^2$ and $\mathbf{H}(\boldsymbol{\lambda}) = (\mathbf{R}^{*\top}\mathbf{R}^* + \mathbf{D}_\Lambda)^{-1}\mathbf{R}^{*\top}\mathbf{R}^*$.

Selecting an appropriate triangulation is also important as few triangles may not capture the features of the functions and too many triangles require quite intensive computation. There are a variety of ways to construct triangulations. An optimal triangulation is a partition of the domain into triangles, that is best according to some criterion that measures the shape, size or number of triangles. For example, a "good" triangulation usually refers to those with well-shaped triangles, no small angles, or/and no obtuse angles. Lai and Schumaker (2007) suggested the maxmin-angle criterion, that is, we choose a triangulation whose minimum angle is as large as possible; see Section S.1 of the supplementary materials for a simple example of maxmin-angle triangulation. Based on the maxmin-angle criterion and Assumption (A6), we seek triangulations with a collection of well-balanced triangles in terms of the shape and size. In practice, one can use the Delaunay triangulation algorithm, and many existing software packages provide a way to generate a Delaunay triangulation; see, for example, Matlab program *delaunay.m*, *distmesh2d.m* or Mathematica function *DelaunayTriangulation*. In our simulation and application studies below, we use *distmesh2d.m* to create the triangulations.

Finally, we discuss how to choose the smoothness and degrees for the spline basis functions. According to Lai and Schumaker (2007), for a fixed smoothness $r \ge 1$, when degree $d \ge 3r + 2$, the bivariate spline achieves full estimation power asymptotically. For example, when $r = 1$ and $d = 5$, the bivariate spline has the optimal approximation order. However, when the degree or smoothness is too high, there are too many parameters to be estimated, which can result in unnecessary

computation burden. Thus, if it is not necessary to increase the smoothness, we suggest choosing $r = 1$ and degree $d = 5$ for the computational efficiency. In practice, for some smooth function without sharp edges as in the simulation studies, we suggest employing a smooth parameter $r = 1$ with degree $d = 3$ or higher. On the other hand, if some coefficient functions have sharp edges, we suggest using smooth parameter $r = 0$ with a lower degree $d = 1$ or 2.

4. Simulation

In this section, we conduct simulation studies to evaluate the performance of the proposed methodology. We consider a modified horseshoe domain Ω in a $[-1, 3.5] \times [-1, 1]$ rectangle constructed as in Ramsay (2002) and Wood, Bravington, and Hedley (2008); see Figure 2 for the domain. The rectangle is evenly divided into $N_s = 901 \times 401$ grid points, and the functions are evaluated at only the locations falling inside the horseshoe domain. We simulate data according to the GSVCM using corresponding links with the following three scenarios:

- Scenario 1: For $i = 1, \ldots, n$, $Y_i \sim \text{Normal}(\mu_i, \sigma^2)$, where $\sigma^2 = 0.25$, $\mu_i = \beta_0(\mathbf{S}_i)X_{1i} + \beta_1(\mathbf{S}_i)X_{2i}$, and $X_{ki} \sim \text{Uniform}(0, 1), k = 0, 1$.
- Scenario 2: For $i = 1, \ldots, n$, $Y_i \sim \text{Poisson}(\mu_i)$, where $\log(\mu_i) = \beta_0(\mathbf{S}_i)X_{1i} + \beta_1(\mathbf{S}_i)X_{2i}$, and $X_{ki} \sim \text{Uniform}(0, 1), k = 0, 1$.
- Scenario 3: For $i = 1, \ldots, n$, $Y_i \sim \text{NB}(\theta, \mu_i/(\theta + \mu_i))$, where $\theta = 6$, $\text{E}(Y_i) = \mu_i$ and $\text{var}(Y_i) = \mu_i + \mu_i^2/\theta$, with $\log(\mu_i) = \beta_0(\mathbf{S}_i)X_{1i} + \beta_1(\mathbf{S}_i)X_{2i}$, and $X_{ki} \sim \text{Uniform}(0, 1), k = 0, 1$.

The true varying coefficient functions in the above models are depicted in Figure 3, where $\beta_0(\cdot)$ is the same function used in Wood, Bravington, and Hedley (2008), and $\beta_1(\cdot)$ is a modified function from $\beta_0(\cdot)$; see the function defined in Section S.2 of the supplementary materials.

For Scenario 1, we consider the GSVCM and GWR methods only for small samples as the GWR method is not feasible for large samples due to its heavy computational requirements. Note that the GWR method is not available for noncontinuous responses as mentioned in Section 1. Thus, for Scenarios 2 and 3, we only consider the GSVCM, and we also investigate the large-scale sample performance of the proposed QRGSVCM method, and compare the performance and computing speed of the QRGSVCM with the GSVCM. For the bivariate spline basis functions used in the GSVCM estimation, we set degree $d = 2$ and 3, smoothness $r = 1$, and the smoothing parameters λ_0 and λ_1 are chosen by minimizing the GCV criterion introduced in (8). We adopt triangulations $\triangle_0$ (35 triangles, 37 vertices), $\triangle_1$ (77 triangles, 65 vertices), and $\triangle_2$ (183 triangles, 130 vertices) given in Figure 2. The additional parameter θ for the negative binomial is selected in a way such that the Pearson estimate of the scale parameter is close to 1. The function *gwr* in R package `spgwr` is used for the GWR method.

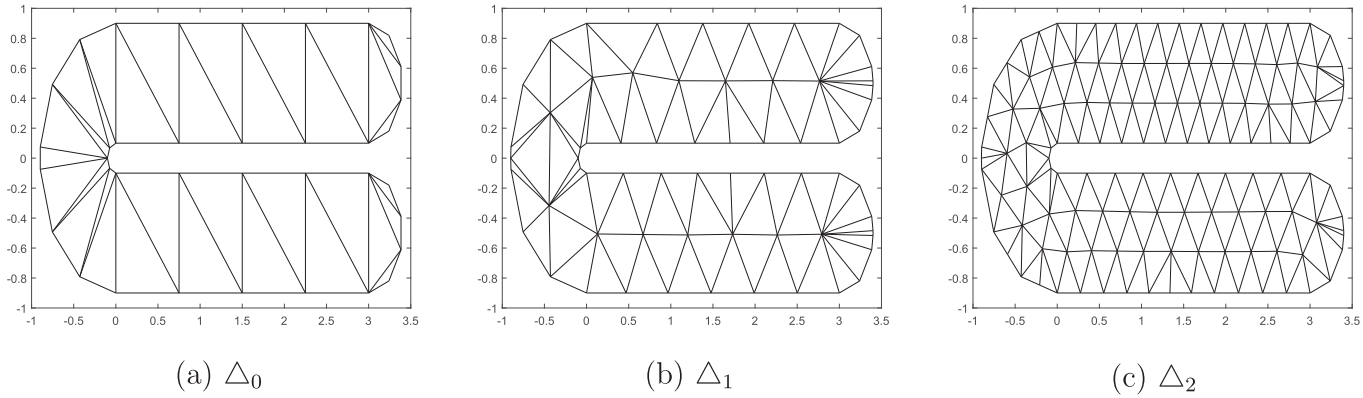

Figure 2. Triangulations, $\triangle_0$, $\triangle_1$, and $\triangle_2$, on the horseshoe domain.

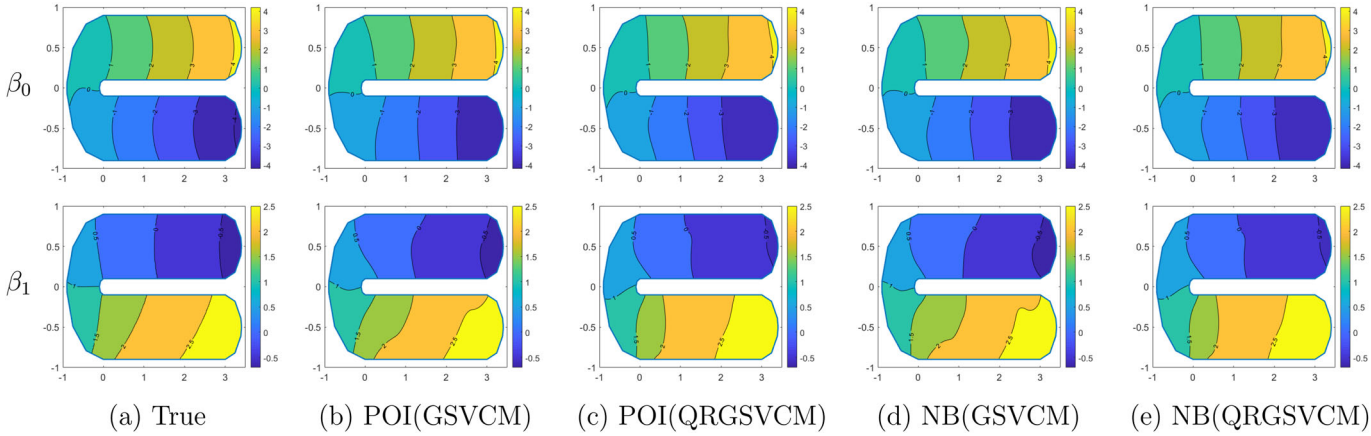

Figure 3. Contour plots of the true coefficient functions, and their GSVCM and QRGSVCM estimates under the Poisson (POI) and negative binomial (NB).

Table 1. MISEs of the estimators of the coefficient functions and running time (seconds per replication).

Family	n	Method	d	β_0			β_1			Time		
				$\triangle_0$	$\triangle_1$	$\triangle_2$	$\triangle_0$	$\triangle_1$	$\triangle_2$	$\triangle_0$	$\triangle_1$	$\triangle_2$
Gaussian	1000	GSVCM	2	0.29	0.26	0.28	0.22	0.21	0.22	0.77	1.23	2.83
			3	0.24	0.23	0.23	0.20	0.19	0.19	1.44	2.36	6.87
	2000	GSVCM	2	0.18	0.16	0.18	0.15	0.14	0.14	1.26	1.82	3.64
			3	0.15	0.14	0.14	0.13	0.12	0.12	2.62	3.95	9.57
	1000	GWR	–		0.59			0.42			2669.17	
	2000	GWR	–		0.32			0.24			10,259.27	
Poisson	1000	GSVCM	2	0.73	0.71	0.70	0.35	0.34	0.33	0.83	1.33	2.99
			3	0.62	0.59	0.59	0.31	0.30	0.29	1.69	2.90	8.39
	2000	GSVCM	2	0.45	0.42	0.44	0.22	0.21	0.21	1.38	2.00	4.01
			3	0.38	0.36	0.36	0.19	0.18	0.18	3.05	5.01	12.52
	50,000	GSVCM	2	0.07	0.04	0.05	0.04	0.02	0.03	165.97	202.44	249.43
			3	0.03	0.03	0.03	0.02	0.02	0.02	346.56	526.70	925.79
	50,000	QRGSVCM	2	0.06	0.03	0.05	0.03	0.02	0.03	18.59	22.94	35.88
			3	0.03	0.03	0.03	0.02	0.02	0.02	51.24	76.70	248.05
	100,000	QRGSVCM	2	0.05	0.02	0.03	0.03	0.01	0.02	37.95	46.34	71.24
			3	0.02	0.02	0.02	0.01	0.01	0.01	103.73	195.68	526.30
NB	1000	GSVCM	2	0.74	0.71	0.71	0.46	0.45	0.45	1.50	2.21	4.33
			3	0.66	0.67	0.67	0.43	0.44	0.45	3.80	7.10	21.40
	2000	GSVCM	2	0.48	0.46	0.47	0.29	0.29	0.29	2.99	4.18	7.62
			3	0.41	0.40	0.41	0.27	0.27	0.29	7.87	14.85	37.89
	50,000	GSVCM	2	0.07	0.04	0.06	0.04	0.03	0.04	542.45	648.04	834.30
			3	0.04	0.04	0.04	0.03	0.03	0.03	1115.72	1748.18	3010.95
	50,000	QRGSVCM	2	0.07	0.04	0.06	0.04	0.03	0.04	25.62	33.64	55.29
			3	0.04	0.03	0.04	0.03	0.03	0.03	89.77	171.80	735.34
	100,000	QRGSVCM	2	0.06	0.02	0.04	0.03	0.02	0.03	50.97	67.20	106.66
			3	0.02	0.02	0.02	0.02	0.02	0.02	169.80	441.47	1495.43

To compare the accuracy of the estimators, we quantify the errors of the estimators of coefficient functions by mean integrated squared error (MISE) given as:

$$\text{MISE}(\widehat{\beta}_k) = \frac{1}{\sum_{i=1}^{N_s} I\{\mathbf{s}_i \in \Omega\}} \sum_{\mathbf{s}_i \in \Omega, i=1,\ldots,N_s} \left\{\widehat{\beta}_k(\mathbf{s}_i) - \beta_k(\mathbf{s}_i)\right\}^2, \quad k = 0, 1, \tag{10}$$

where $\mathbf{s}_i$'s are the grid points over the rectangle. The MISEs of β_0 and β_1 based on three triangulations and different sample sizes are given in Table 1. From this table, one sees that all estimators tend to lead to better performance as the sample size increases. For every scenario, the GSVCM with higher degrees tends to result in higher accuracy in terms of the MISE. With regard to the triangulation, refining the triangulation from $\triangle_0$ to $\triangle_1$ reduces the MISE. However, triangulation $\triangle_1$ seems to be fine enough to reach great approximation power as MISEs based on $\triangle_1$ and $\triangle_2$ are similar for every scenario. For Scenario 1, Table 1 shows that the GSVCM estimates are far more accurate than the GWR estimates.

For Scenarios 2 and 3, we illustrate the estimated surfaces for a typical replication with $\triangle_1$ and degree $d = 3$ and sample size $n = 50{,}000$ in Figure 3, which shows that the estimated surfaces for β_0 and β_1 are visually similar to true surfaces of β_0 and β_1. See Section S.2 of the supplementary materials for the estimated contour plots of the coefficient functions using the GSVCM and GWR methods for Scenario 1 with 2000 observation.

Next, we report the computing speed of different methods based on a cluster with Intel®Xeon® CPU E5-2640 v3 @ 2.60GHz and 128 GB of RAM per node. For Scenario 1, the GSVCM ($\triangle_0$) takes only 2.6 sec per iteration to fit the model with 2000 observations. GSVCM ($\triangle_1$) and GSVCM ($\triangle_2$) take just a few seconds longer than GSVCM ($\triangle_0$). However, the GWR usually has to spend more than 2.8 hr to complete one iteration with 2000 observations. For Scenarios 2 and 3 with small sample sizes ($n = 1000$ and 2000), the GSVCM method takes a few seconds per iteration for fitting the model. However, when we increase the sample size to 50,000, the GSVCM requires heavy computation as shown in Table 1. In contrast, the QRGSVCM dramatically reduces the computational time while provides a similar accuracy compared with the GSVCM method. Even for a larger sample size with $n = 100{,}000$, its computing time is still much shorter than that of the GSVCM for $n = 50{,}000$.

Finally, we compare our proposed GSVCM with the following GAM: $g\{\mu(\mathbf{X}_i, \mathbf{S}_i)\} = \zeta(\mathbf{S}_i) + f_1(X_{1i}) + f_1(X_{2i})$, $i = 1, \ldots, n$, where $\zeta(\cdot)$ is an unknown bivariate smooth function and $f_k(\cdot)$'s are unknown univariate functions. To fit the GAM, we focus on two methods: (i) thin-plate splines (TPS) described in Wahba (1990) and Wood (2003) and (ii) soap film smoothing (SOAP) proposed by Wood, Bravington, and Hedley (2008). The function *gam* in R package `mgcv` is used for the GAM methods.

Because our proposed and the competing methods have different structural functions in the systematic component, direct comparisons based on MISE in (10) are not possible. Instead, we compare the 10-fold cross-validation (CV) mean squared prediction error (MSPE) in the sense of predictive ability. The result is illustrated in Table 2. Due to the heavy computation requirement and the collinearity issue, the cross-validation calculation for the GWR is infeasible so that it is not included in Table 2. From this table, the SOAP shows better performance than the TPS, as expected on complex domains. In addition, one sees that the GSVCM outperforms both the SOAP and TPS in terms of MSPE. The main reason is that both of the bivariate coefficients are spatially varying in the underlying true model; however, there is only a single bivariate function to account for

Table 2. Results of the 10-fold CV MSPEs.

Family	n	GSCVM ($d = 2$)		GSCVM ($d = 3$)		TPS		SOAP	
		MSPE	Time	MSPE	Time	MSPE	Time	MSPE	Time
Normal	1000	0.260	2.13	0.259	6.86	0.868	1.32	0.837	16.91
	2000	0.256	4.44	0.255	14.74	0.853	3.88	0.830	18.63
Poisson	1000	3.245	3.01	3.222	11.75	9.729	1.89	9.155	20.08
	2000	3.113	6.04	3.104	24.03	9.415	5.47	8.940	26.75
	50,000	2.973*	57.38*	2.973*	458.73*	9.011	128.26	8.555	471.15
NB	1000	8.128	10.88	8.122	50.24	17.835	1.68	17.686	18.91
	2000	7.697	24.35	7.716	110.23	17.380	4.88	17.286	26.53
	50,000	7.369*	154.78*	7.372*	1227.02*	17.067	126.51	16.899	1160.53

NOTE: The GSVCM is based on $\triangle_1$.
*The results obtained using the QRGSVCM method.

the spatial information in the GAMs. Especially, for negative binomial scenario, our proposed method shows a significant improvement over the TPS and SOAP in the sense of MSPE. In general, the GSVCM with a higher degree tends to improve the predictive capability.

Table 2 also reports the computing time of different method based on the same computing environment as described before. For $n = 1000$ and 2000, the GSVCM with degree 2 and the TPS take a few seconds to implement one 10-fold CV in Scenario 1, and the GSVCM with degree 2 takes a little bit longer time than the TPS in Scenarios 2 and 3. However, the GSVCM with degree 2 is faster than the SOAP method for all the scenarios. When $n = 50{,}000$, the QRGSVCM with degree 2 spends 57.38 and 154.78 sec per iteration for Poisson and negative binomial random components, respectively, which is computationally very competitive compared to the TPS and SOAP methods.

5. Application

This section presents the analysis of the crash dataset described in Section 1. The dataset includes 11,249 census block groups with six covariates collected from the Florida Department of Transportation and the U.S. Census Bureau. A couple of outliers and areas with no population were excluded from this study since their off-roadway crash frequencies of interest are mostly zeros. For each census block group, vehicle miles traveled ("VMT"), total population ("POP"), proportion of males ("MR"), proportion of Hispanics in population ("HR"), proportion of people age 65 and older ("OR"), and unemployed rate ("UR") are measured as covariates. The off-roadway crash frequency is considered as the response of our interest.

Assume that the crash frequency follows the negative binomial random component, that is, $Y_i \sim \text{NB}\left(\theta, \frac{\mu_i}{\theta+\mu_i}\right)$, where θ is estimated in a similar way as in the simulation study. In addition, suppose that, for the location $i = 1, \ldots, 11{,}249$,

$$\begin{aligned}\log(\mu_i) = {} & \beta_0(\mathbf{S}_i) + \beta_1(\mathbf{S}_i)\text{VMT}_i^* + \beta_2(\mathbf{S}_i)\text{POP}_i^* + \beta_3(\mathbf{S}_i)\text{MR}_i \\ & + \beta_4(\mathbf{S}_i)\text{HR}_i + \beta_5(\mathbf{S}_i)\text{OR}_i + \beta_6(\mathbf{S}_i)\text{UR}_i,\end{aligned}$$

where $\beta_k(\cdot)$'s are unknown bivariate functions, for $k = 0, \ldots, 6$. The covariates with "$*$" are transformed from the original value by $X^* = \log(X + 0.01)$, and we centralize all the covariates in the analysis. For the GSVCM estimation, the triangulation with 174 triangles and 119 vertices is adopted for the bivariate splines; see Figure 4(a). For the spline basis functions, we consider each degree $d = 2$ to 4, and smoothness $r = 1$. The penalty parameters, $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_6)^\top$, are found based on GCV defined in (8).

It is interesting to compare our model to some other approaches in the literature. We choose the GAM as the competing model, which is the typical model to analyze such data. To be more specific, we assume that

$$\begin{aligned}\log(\mu_i) = {} & \zeta(\mathbf{S}_i) + \beta_1^*(\text{VMT}_i^*) + \beta_2^*(\text{POP}_i^*) + \beta_3^*(\text{MR}_i) \\ & + \beta_4^*(\text{HR}_i) + \beta_5^*(\text{OR}_i) + \beta_6^*(\text{UR}_i),\end{aligned}$$

where $\beta_k^*(\cdot)$'s are unknown univariate functions, and $\zeta(\cdot)$ is an unknown bivariate function. The function *gam* in R package `mgcv` is used for the GAM methods. To compare the accuracy and predictive capability of models, we calculate the MSE and 10-fold CV MSPE for TPS, SOAP, and GSVCM with degrees 2 to 4, and the results are illustrated in Table 3. The SOAP method results in a better performance than the TPS method since TPS does not take into account the irregular shape of the domain when smoothing a bivariate function so that it leads to the leakage problem. Our proposed method outperforms TPS and SOAP methods in terms of both the MSE and 10-fold CV MSPE. Their inferior performance may be due to the lack of accounting for the spatially nonstationarity in relationship between the car crash and its corresponding covariates.

To investigate the spatial stationarity, we focus on the nonstationarity tests of individual bivariate coefficient functions, and consider the following hypotheses:

$$H_{0k} : \beta_k(\boldsymbol{s}) = \beta_k \text{ versus } H_{1k} : \beta_k(\boldsymbol{s}) \neq \beta_k, \tag{11}$$

where $\beta_k(\boldsymbol{s})$'s are true functions for $k = 0, \ldots, p$. We consider a generalized quasi-likelihood ratio (GQLR) and define the test statistic as the difference of the quasi-likelihoods under the full and reduced models as follows:

$$\begin{aligned}\lambda_n(H_0) = \sum_{i=1}^{n} \Bigg[& \ell_Q\left\{g^{-1}\left(\mathbf{X}_i^\top \widehat{\boldsymbol{\beta}}_F\right), Y_i\right\} \\ & - \ell_Q\left\{g^{-1}\left(\mathbf{X}_i^\top \widehat{\boldsymbol{\beta}}_R\right), Y_i\right\}\Bigg],\end{aligned} \tag{12}$$

where $\widehat{\boldsymbol{\beta}}_F$ and $\widehat{\boldsymbol{\beta}}_R$ are the estimators under the full and reduced models, respectively.

Motivated by the GQLR test described in Tang, Li, and Guan (2016), we propose a similar procedure to estimate the p-value of the test; see Algorithm 3. Applying the GQLR test for the hypothesis in (11), we find that all coefficient functions are indeed statistically significant with p-value less than 0.05.

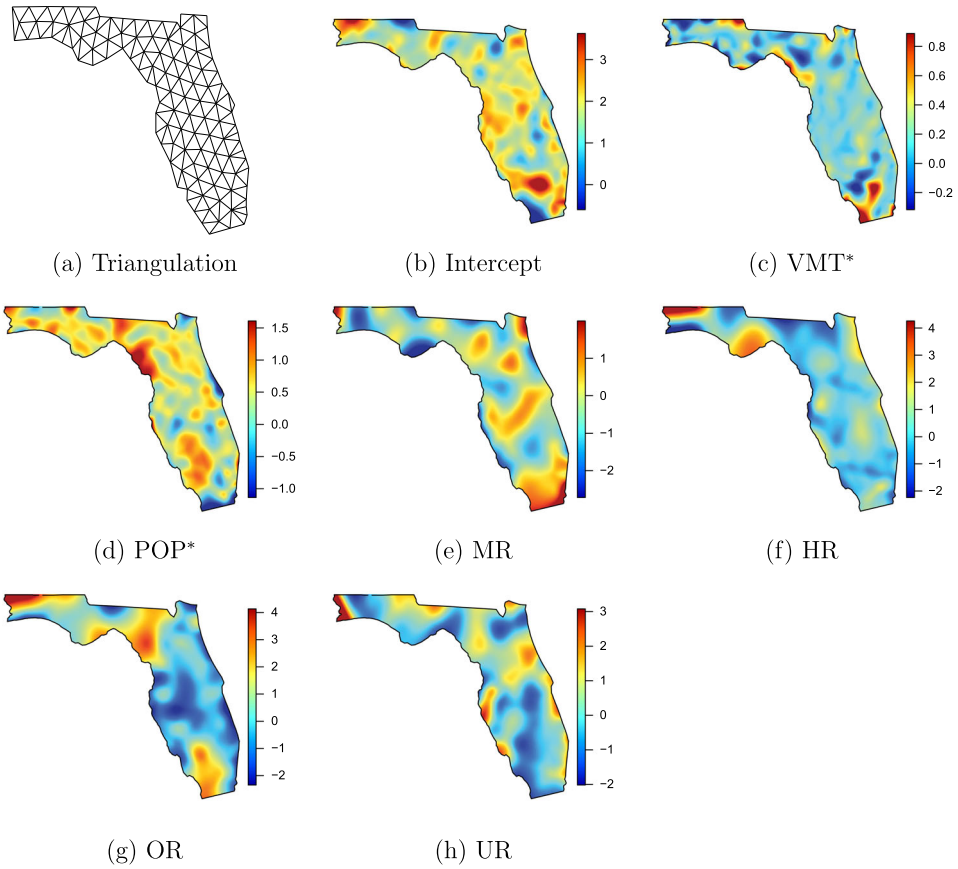

Figure 4. Triangulation of Florida and the estimated coefficient maps.

Step 1. Based on the data $\{(Y_i, \mathbf{X}_i, \mathbf{S}_i)\}_{i=1}^n$, calculate the GQLR test statistic, $\lambda_n(H_0)$, defined in (12).
Step 2. For $i = 1, \ldots, n$, generate $Y_i^* = g^{-1}(\mu_i^*) + \epsilon_i^*$, where $\mu_i^* = \mathbf{X}_i^\top \widehat{\boldsymbol{\beta}}_R(\mathbf{S}_i)$, $\epsilon_i^* = \xi_i \widehat{\epsilon}_i$, $\widehat{\epsilon}_i$'s are residuals from the full model, and ξ_i's follow a Mammen's two-point distribution with probability $P\{\xi_i = -(\sqrt{5}-1)/2\} = (\sqrt{5}+1)/(2\sqrt{5})$, and $P\{\xi_i = (\sqrt{5}+1)/2\} = (\sqrt{5}-1)/(2\sqrt{5})$.
Step 3. For $b = 1, \ldots, B$, calculate the GQLR test statistic $\lambda_n^{(b)}(H_0)$ based on the bootstrap sample $\{Y_i^*, \mathbf{X}_i, \mathbf{S}_i\}$ as the same argument defined as (12).
Step 4. Repeat Steps 2 and 3 B times and obtain the estimated p-value which is calculated by $\widehat{p} = \sum_{b=1}^B I\left\{\lambda_n^{(b)}(H_0) \geq \lambda_n(H_0)\right\}/B$, where $I(\cdot)$ is the indicator function.

Algorithm 3: GQLR test for GSVCM.

The estimated bivariate surfaces of $\beta_k(\boldsymbol{s})$, $k = 0, \ldots, 6$, are depicted in Figures 4(b)–(h), respectively. The effect of intercept is large for the locations with large crash frequency in Figure 1, except for the northern area of Escambia county and Santa Rosa county. The effect of "VMT*" on crash frequency is between the levels exp(0) and exp(0.5) in most areas. Everglades National Park and the border of three states (Florida, Alabama, and Georgia) have a relatively large effect of "VMT*" on crash frequency compared to other areas. Interestingly, the largest effect of "POP*" on crash frequency happens in Dixie county and Levy county, whose populations are very small. In addition, males seem to be more exposed to the risk of crash in Miami. Holding other effects of covariates fixed, Hispanics are more likely to be involved in a car crash in Apalachicola National Forest, Jacksonville, and the north of Escambia and Santa Rosa while the effect of "OR" is prominent in Taylor county, Everglades National Park, and the north of Escambia and Santa Rosa. The eminently large effects of "HR," "OR," as well as "UR" are observed in some areas of Escambia and Santa Rosa, whose northern areas have the large intercept effect even though crash frequencies in those areas are small.

Table 3. MSEs and 10-fold CV MSPEs for TPS, SOAP, and GSVCM.

	TPS	SOAP	GSVCM		
			$d = 2$	$d = 3$	$d = 4$
MSE	182.03	177.46	174.31	157.73	147.54
MSPE	184.64	181.05	182.27	174.53	172.31

6. Conclusion

In this article, we propose a class of generalized nonparametric regression models with spatially varying coefficients, which is flexible enough to cover a variety of models. The GSVCM can be applied to various types of data with discrete or categorical response variables. The proposed BPST method works well for both regularly and irregularly spaced data over a domain with complex boundary. Allowing different degrees and smoothness parameters, our method provides flexible and efficient estimators for the varying coefficient functions and mitigates the adverse effect of the collinearity problem.

This study can be further extended in the following ways. First, we can add a partially linear part in the proposed model. In many applications, some of covariates might have homogeneous linear effects over the domain. By considering this, we can further explore the estimation in a generalized partially linear spatially varying coefficient models and make the estimators even more efficient. Second, we only considered the spatial dependence for the crash data study at a certain time point. In recent years, there has been increasing attention to spatial and temporal data analysis, which has opened a broad range of new applications in Environmental Sciences, Geography, Resilience, Energy, and more. It is interesting to extend the proposed method to such kind of spatiotemporal data analysis. As suggested by one of anonymous referees, it is also interesting to develop inferential tools to identify the regions where the coefficient is significant. It will not only help to narrow down the region of interest, but also assist decision-makers in efficiently focusing on the problem. Developing an efficient and automatic region identification criterion for GSVCMs is challenging and warrants future study.

Supplementary Materials

Technical material: The file GSVCM_sub_final.pdf (PDF file) provides the detailed notations and detailed proof of Theorem 1 in the article.

Codes and data: The file contains the codes and data for Sections 4 and 5. A README file describes the contents. (Code.zip)

R package BPST: R package "BPST" used to perform the proposed method in the article is available in the following website: *https://github.com/funstatpackages/BPST*.

Acknowledgments

The authors would like to thank the editor, Dr. Tyler McCormick, an associate editor, and three referees for their constructive comments that significantly improved an earlier version of this article.

Funding

Li Wang's research was supported in part by National Science Foundation awards DMS-1542332 and DMS-1916204.

References

Azzimonti, L., Sangalli, L. M., Secchi, P., Domanin, M., and Nobile, F. (2015), "Blood Flow Velocity Field Estimation via Spatial Regression With PDE Penalization," *Journal of the American Statistical Association*, 110, 1057–1071. [2]

Brunsdon, C., Fotheringham, A. S., and Charlton, E. M. (1996), "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity," *Geographical Analysis*, 28, 281–298. [1,2]

Carroll, R. J., Fan, J., Gijbels, I., and Wand, M. P. (1997), "Generalized Partially Linear Single-Index Models," *Journal of the American Statistical Association*, 92, 477–489. [4]

Fan, J., and Zhang, W. (2008), "Statistical Methods With Varying Coefficient Models," *Statistics and Its Interface*, 1, 179–195. [1]

Fotheringham, A. S., Brunsdon, C., and Charlton, E. M. (2002), *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*, Chichester: Wiley. [1]

Hastie, T., and Tibshirani, R. (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society*, Series B, 55, 757–796. [1]

Lai, M. J., and Schumaker, L. L. (2007), *Spline Functions on Triangulations*, Cambridge: Cambridge University Press. [2,3,5]

Lai, M. J., and Wang, L. (2013), "Bivariate Penalized Splines for Regression," *Statistica Sinica*, 23, 1399–1417. [2,4]

Liu, R., Yang, L., and Härdle, W. K. (2013), "Oracally Efficient Two-Step Estimation of Generalized Additive Model," *Journal of the American Statistical Association*, 108, 619–631. [2,4]

Mu, J., Wang, G., and Wang, L. (2018), "Estimation and Inference in Spatially Varying Coefficient Models," *Environmetrics*, 29, e2485. [1,2,4]

Park, B. U., Mammen, E., Lee, Y. K., and Lee, E. R. (2015), "Varying Coefficient Regression Models: A Review and New Developments," *International Statistical Review*, 83, 36–64. [1]

Ramsay, T. (2002), "Spline Smoothing Over Difficult Regions," *Journal of the Royal Statistical Society*, Series B, 64, 307–319. [2,6]

Sangalli, L., Ramsay, J., and Ramsay, T. (2013), "Spatial Spline Regression Models," *Journal of the Royal Statistical Society*, Series B, 75, 681–703. [2]

Scott-Hayward, L. A. S., MacKenzie, M. L., Donovan, C. R., Walker, C., and Ashe, E. (2014), "Complex Region Spatial Smoother (CReSS)," *Journal of Computational and Graphical Statistics*, 23, 340–360. [2]

Sun, Y., Yan, H., Zhang, W., and Lu, Z. (2014), "A Semiparametric Spatial Dynamic Model," *The Annals of Statistics*, 42, 700–727. [2]

Tang, J., Li, Y., and Guan, Y. (2016), "Generalized Quasi-Likelihood Ratio Tests for Semiparametric Analysis of Covariance Models in Longitudinal Data," *Journal of the American Statistical Association*, 111, 736–747. [8]

Wahba, G. (1990), *Spline Models for Observational Data*, Philadelphia, PA: SIAM. [7]

Wang, H., and Ranalli, M. G. (2007), "Low-Rank Smoothing Splines on Complicated Domains," *Biometrics*, 63, 209–217. [2]

Wang, L., and Cao, G. (2018), "Efficient Estimation for Generalized Partially Linear Single-Index Models," *Bernoulli*, 24, 1101–1127. [4]

Wang, L., Liu, X., Liang, H., and Carroll, R. (2011), "Estimation and Variable Selection for Generalized Additive Partial Linear Models," *The Annals of Statistics*, 39, 1827–1851. [4]

Wang, L., Wang, G., Lai, M. J., and Gao, L. (2020), "Efficient Estimation of Partially Linear Models for Data on Complicated Domains by Bivariate Penalized Splines Over Triangulations," *Statistica Sinica*, 30, 347–369. [4]

Wang, L., and Yang, L. (2007), "Spline-Backfitted Kernel Smoothing of Nonlinear Additive Autoregression Model," *The Annals of Statistics*, 35, 2474–2503. [2]

Wilhelm, M., and Sangalli, L. M. (2016), "Generalized Spatial Regression With Differential Regularization," *Journal of Statistical Computation and Simulation*, 86, 2497–2518. [2]

Wood, S. N. (2003), "Thin Plate Regression Splines," *Journal of the Royal Statistical Society*, Series B, 65, 95–114. [7]

Wood, S. N., Bravington, M. V., and Hedley, S. L. (2008), "Soap Film Smoothing," *Journal of the Royal Statistical Society*, Series B, 70, 931–955. [2,6,7]

Wood, S. N., Goude, Y., and Shaw, S. (2015), "Generalized Additive Models for Large Data Sets," *Journal of the Royal Statistical Society*, Series C, 64, 139–155. [5]

Yang, L., Park, B. U., Xue, L., and Härdle, W. (2006), "Estimation and Testing for Varying Coefficients in Additive Models With Marginal Integration," *Journal of the American Statistical Association*, 101, 1212–1227. [1]

Zheng, S., Liu, R., Yang, L., and Härdle, W. K. (2016), "Statistical Inference for Generalized Additive Models: Simultaneous Confidence Corridors and Variable Selection," *Test*, 25, 607–626. [2]

Zhou, L., and Pan, H. (2014), "Smoothing Noisy Data for Irregular Regions Using Penalized Bivariate Splines on Triangulations," *Computational Statistics*, 29, 263–281. [2]