

Estimation and inference in spatially varying coefficient models

Jingru Mu^a Guannan Wang^b and Li Wang^{a*}

Summary: Spatially varying coefficient models (SVCMs) are a classical tool to explore the spatial nonstationarity of a regression relationship for spatial data. In this paper, we study the estimation and inference in SVCMs for data distributed over complex domains. We use bivariate splines over triangulations to represent the coefficient functions. The estimators of the coefficient functions are consistent, and rates of convergence of the proposed estimators are established. A penalized bivariate spline estimation method is also introduced, in which a roughness penalty is incorporated to balance the goodness-of-fit and smoothness. In addition, we propose hypothesis tests to examine if the coefficient function is really varying over space or admits a certain parametric form. The proposed method is much more computationally efficient than the well-known geographically weighted regression technique and thus usable for analyzing massive datasets. The performance of the estimators and the proposed tests are evaluated by simulation experiments. An environmental data example is used to illustrate the application of the proposed method.

Keywords: bivariate splines; bootstrap test; penalized splines; permutation test; spatial data; triangulation.

1. INTRODUCTION

In spatial data analysis, a common problem is to identify the nature of the relationship that exists between variables. In many situations, a simple “global” model often cannot explain the relationships between some sets of variables, which is referred to “spatial nonstationarity”. To handle such nonstationarity, the model needs to reflect the spatially varying

^a Department of Statistics, Iowa State University, IA 50011, USA

^b Department of Mathematics, College of William & Mary, Williamsburg, VA 23187, USA

* Correspondence to: Li Wang, Department of Statistics, Iowa State University, Ames, IA 50011, USA. E-mail: lilywang@iastate.edu

structure within the data. In this paper, we investigate a class of spatially varying coefficient models (SVCs) to explore the spatial non-stationarity of a regression relationship. The data in our study need not be evenly distributed, instead, we assume the observations are randomly distributed over two-dimensional domain $\Omega \subseteq \mathbb{R}^2$ of arbitrary shape, for example, a polygonal domain with interior holes. Suppose there are n random selected locations, and let $\mathbf{U}_i = (U_{i1}, U_{i2})^\top$ be the location of i -th point, $i = 1, \dots, n$, which ranges over Ω . Let Y_i be the response variable and $\mathbf{X}_i = (X_{i0}, X_{i1}, \dots, X_{ip})^\top$ with $X_{i0} \equiv 1$ being the explanatory variables. Suppose that $\{(\mathbf{U}_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$ satisfies the following model (Brunsdon et al., 1996, 1998; Fotheringham et al., 2002; Shen et al., 2011)

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

where $\beta_k(\cdot)$'s are unknown varying-coefficient functions, and ϵ_i 's are i.i.d random noises with $E(\epsilon_i) = 0$ and $\text{Var}(\epsilon_i) = \sigma^2$, and independent of \mathbf{X}_i . Our primary interest is to estimate and make inferences for $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^\top$ based on the given observations $\{(\mathbf{U}_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$.

When $\beta_k(\cdot)$'s are univariate functions, model (1) is the typical varying coefficient model which has been extensively studied in the literature (Hastie and Tibshirani, 1993; Fan and Zhang, 1999; Xue, 2006; Ferguson et al., 2009; Tang and Cheng, 2009; Lian, 2012). In this paper, $\beta_k(\cdot)$'s are bivariate functions of locations, and model (1) allows the regression coefficients to vary over space and therefore can be used to explore spatial non-stationarity of the regression relationship via the spatial variation patterns of the estimated coefficients.

In the past decade, SVCs have been widely applied to a variety of fields including geography (Su et al., 2017), ecology (Finley, 2011), econometrics (Bitter et al., 2007; Helbich and Griffith, 2016; Al-Sulami et al., 2017), epidemiology (Nakaya et al., 2005), meteorology (Lu et al., 2009), and environmental science (Waller et al., 2007; Hu et al., 2013; Tang, 2014; Huang et al., 2017).

There is a rich literature on how to estimate SVCs. Two competing methods are the

Bayesian approach, see Assunção (2003); Gelfand et al. (2003), and the local approach, such as the geographically weighted regression (GWR) technique (Brunsdon et al., 1996, 1998; Fotheringham et al., 2002). The Bayesian procedure is carried out by assuming a certain prior distribution of the coefficients and computing their posterior distribution on which the estimation and inference are performed. However, there is no correct way of choosing a prior. In practice, misleading results will be generated if one does not choose prior distributions with caution. In addition, for a large dataset with many variables being estimated, the Bayesian method may be prohibitively computationally intensive. The GWR method estimates the coefficients in the traditional regression framework of kernel smoothing. It incorporates local spatial relationships into the regression framework in an intuitive and explicit manner. While this local kernel-based approach is very nice and useful, it becomes very computationally intensive for large datasets as it requires solving an optimization problem at every sample location. Typically, the GWR model fitting and spatial prediction require $O(n^2)$ operations for a data set of size n . Recent evolutions in technology provide increasing volumes of spatial data (Banerjee et al., 2008; Zhang et al., 2013), which is beyond the computing limit of the traditional Bayesian and GWR method. It is urgent to develop a more computationally expedient tool for analyzing spatial data.

Tang and Cheng (2009) and Lu et al. (2014) proposed a B-spline approximation of the coefficient functions. The method is fast and efficient since it inherits many advantages of spline-based techniques. However, the data are required to be regularly spaced over a rectangular domain. In practice, spatial data are often collected over complex domains with irregular boundaries, peninsulas and interior holes. Many smoothing methods, such as kernel smoothing, tensor product smoothing and wavelet smoothing, suffer from the problem of “leakage” across the complex domains, which refers to the poor estimation over difficult regions by smoothing inappropriately across boundary features; see the discussions in Ramsay (2002); Wood et al. (2008); Sangalli et al. (2013).

In this paper, we develop a powerful and efficient method to estimate SVCMs for data distributed over two-dimensional complex domains. Our method tackles the estimation problem differently from the local approach, and the coefficient functions $\beta_k(\cdot)$'s are approximated using the bivariate splines over triangulations in Lai and Schumaker (2007) and Lai and Wang (2013). The proposed estimator solves the problem of “leakage” across the complex domains. Another advantage of this approach is that it can formulate a global penalized least squares problem, thus it is sufficiently fast and efficient for the user to analyze large datasets within seconds. In addition, under the independence error condition, which is not uncommon in the GWR literature (Brunsdon et al., 1996, 1998; Shen et al., 2011; Huang et al., 2017; Su et al., 2017, for example), we show the proposed coefficient estimators converge to the true coefficient functions.

An important statistical question in fitting SVCMs is whether the coefficient function is really varying over space (Brunsdon et al., 1999; Leung et al., 2000), which amounts to testing if the coefficient functions are constant or in a certain parametric form. In the pioneering work of GWR by Brunsdon et al. (1996), two kinds of permutation test are proposed for global stationarity and individual stationarity, respectively. For the individual test, the variability of the estimated coefficient function is used to describe the plausibility of a constant coefficient. Brunsdon et al. (1999) developed a test via comparing the residual sum of squares (RSS) from the GWR estimation with that from the ordinary least square estimation for the null hypothesis of global spatial stationarity. Moreover, Leung et al. (2000) introduced another RSS based statistics to test for the global stationarity of the regression relationship. Motivated by many sophisticated statistical inferential problems in a variety of areas and fueled by the power of modern computing techniques, bootstrap methods get increasingly popular during the past two decades. For example, Mei et al. (2006) used the bootstrap test to investigate the zero coefficients in a mixed GWR model and Cai et al. (2000) proposed a new wild bootstrap test for the goodness of fit of the varying

coefficient models for nonlinear time series. In this work, we adopt the idea in Cai et al. (2000) and employ a bootstrap test for testing a globally stationary regression relationship in an SVCM. For individual stationarity test, we suggest the permutation test, which is an easily understandable and generally applicable approach to testing problems. Our simulation shows that the resulting testing procedure is indeed powerful and the bootstrap method does give the right null distribution.

The rest of the paper is organized as follows. In Section 2 we give a short review of the triangulations and propose our estimation method based on bivariate splines. Section 3 is devoted to the asymptotic analysis of the proposed estimators. Section 4 extends the bivariate splines to the penalized bivariate splines, in which smoothing parameters are used to balance the goodness-of-fit and smoothness. Section 5 describes the bootstrap goodness of fit test to examine the global stationarity, and the permutation test for each coefficient functions to check individual stationarity. Section 6 presents simulation results comparing our method with its competitors. An illustration of the proposed approach is provided in Section 7 by an analysis of the particle pollution data. Section 8 concludes the paper. Technical details and more numerical studies are provided in Supplemental Materials.

2. TRIANGULATIONS AND BIVARIATE SPLINE ESTIMATORS

Our estimation is based on bivariate splines over triangulations (BST). Below we briefly introduce the techniques of triangulations and the bivariate spline smoothing for SVCMs.

2.1. Triangulations

Triangulation is an effective tool to handle data distributed on irregular regions with complex boundaries and/or interior holes. In the following we use τ to denote a triangle which is a convex hull of three points not located in one line. A collection $\Delta = \{\tau_1, \dots, \tau_K\}$ of K triangles

is called a triangulation of $\Omega = \cup_{j=1}^K \tau_j$ provided that if a pair of triangles in Δ intersect, then their intersection is either a common vertex or a common edge. Without loss of generality, we assume that all \mathbf{U}_i 's are inside triangles of Δ , that is, they are not on edges or vertices of triangles in Δ . Otherwise, we can simply count them twice or multiple times if any observation is located on an edge or at a vertex of Δ .

There are quite a few packages available that can be used to construct a triangulation. For example, one can use the ‘‘Delaunay’’ algorithm to find a triangulation; see MATLAB program *delaunay.m* or MATHEMATICA function *DelaunayTriangulation*. ‘‘DistMesh’’ is another method to generate unstructured triangular and tetrahedral meshes; see the *DistMesh* generator on <http://persson.berkeley.edu/distmesh/>. A detailed description of the program is provided by Persson and Strang (2004). In all the simulation studies and real data analysis below, we used the ‘‘DistMesh’’ to generate the triangulations.

2.2. Bivariate spline estimators

For a nonnegative integer r , let $\mathbb{C}^r(\Omega)$ be the collection of all r -th continuously differentiable functions over Ω . Given a triangulation Δ , let $\mathcal{S}_d^r(\Delta) = \{s \in \mathbb{C}^r(\Omega) : s|_\tau \in \mathbb{P}_d(\tau), \tau \in \Delta\}$ be a spline space of degree d and smoothness r over triangulation Δ , where $s|_\tau$ is the polynomial piece of spline s restricted on triangle τ , and \mathbb{P}_d is the space of all polynomials of degree less than or equal to d . Given $\{(\mathbf{U}_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$, we consider the following minimization problem:

$$\min_{s_k \in \mathcal{S}_d^r(\Delta), k=0, \dots, p} \sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p X_{ik} s_k(\mathbf{u}_i) \right\}^2.$$

We use Bernstein basis polynomials to represent the bivariate splines. For any $k = 0, \dots, p$, let $\{B_j\}_{j \in \mathcal{J}_k}$ be the set of degree- d bivariate Bernstein basis polynomials for $\mathcal{S}_d^r(\Delta)$ constructed in Lai and Schumaker (2007), where \mathcal{J}_k denotes the index set of the basis functions. Then we can write the function $s_k(\mathbf{u}) = \sum_{j \in \mathcal{J}_k} B_{kj}(\mathbf{u}) \gamma_{kj} = \mathbf{B}_k(\mathbf{u})^\top \boldsymbol{\gamma}_k$, where

$\boldsymbol{\gamma}_k = (\gamma_{kj}, j \in \mathcal{J}_k)^\top$ is the spline coefficient vector. Using the above approximation, we have the following minimization:

$$\sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p X_{ik} \mathbf{B}_k(\mathbf{U}_i)^\top \boldsymbol{\gamma}_k \right\}^2. \quad (2)$$

To meet the smoothness requirement of the bivariate splines, we need to impose some linear constraints on the spline coefficients to enforces smoothness across shared edges of triangles. Denote \mathbf{H}_k the constraint matrix on the coefficients $\boldsymbol{\gamma}_k$, which depends on the smoothness r and the structure of the triangulation. Putting all smoothness conditions together yields $\mathbf{H}_k \boldsymbol{\gamma}_k = \mathbf{0}$. We first remove the constraint via the following QR decomposition: $\mathbf{H}_k^\top = (\mathbf{Q}_{1,k} \ \mathbf{Q}_{2,k}) \begin{pmatrix} \mathbf{R}_{1,k} \\ \mathbf{0} \end{pmatrix}$, where $(\mathbf{Q}_{1,k} \ \mathbf{Q}_{2,k})$ is an orthogonal matrix, and $\begin{pmatrix} \mathbf{R}_{1,k} \\ \mathbf{0} \end{pmatrix}$ is an upper triangle matrix. We then reparametrize using $\boldsymbol{\gamma}_k = \mathbf{Q}_{2,k} \boldsymbol{\theta}_k$ for some $\boldsymbol{\theta}_k$, then it is guaranteed that $\mathbf{H}_k \boldsymbol{\gamma}_k = \mathbf{0}$. Thus, the minimization problem in (2) is now converted to a conventional regression problem without any restriction:

$$\sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p X_{ik} \mathbf{B}_k^\top(\mathbf{U}_i) \mathbf{Q}_{2,k} \boldsymbol{\theta}_k \right\}^2. \quad (3)$$

For simplicity, we assume $\mathbf{B}(\mathbf{u}) = \mathbf{B}_0(\mathbf{u}) = \mathbf{B}_1(\mathbf{u}) = \cdots = \mathbf{B}_p(\mathbf{u}) = \{B_j(\mathbf{u})\}_{j \in \mathcal{J}}$, then $\mathbf{H}_0 = \mathbf{H}_1 = \cdots = \mathbf{H}_p$ and $\mathbf{Q}_2 = \mathbf{Q}_{2,0} = \mathbf{Q}_{2,1} = \cdots = \mathbf{Q}_{2,p}$. In practice, if the coefficients are of very different degrees of smoothness, one can choose different bivariate spline basis functions with variable triangulations for different coefficient functions to guarantee sufficient smoothness. Denote $\boldsymbol{\theta} = (\boldsymbol{\theta}_0^\top, \boldsymbol{\theta}_1^\top, \cdots, \boldsymbol{\theta}_p^\top)^\top$ and let $\mathbf{X}_i = (1, X_{i1}, \dots, X_{ip})^\top$. Let $\mathbf{B}^*(\mathbf{U}_i) = \mathbf{Q}_2^\top \mathbf{B}(\mathbf{U}_i)$. Then the minimization problem in (3) can be written as

$$\sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p X_{ik} \mathbf{B}^*(\mathbf{U}_i)^\top \boldsymbol{\theta}_k \right\}^2. \quad (4)$$

Let “ \otimes ” denote the Kronecker product. Solving the least squares problem in (4), we obtain

$$\widehat{\boldsymbol{\theta}} = \left[\sum_{i=1}^n \{\mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i)\} \{\mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i)\}^\top \right]^{-1} \sum_{i=1}^n \{\mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i)\} Y_i. \quad (5)$$

The BST estimator of $\beta_k(\mathbf{u})$ is $\widehat{\beta}_k(\mathbf{u}) = \mathbf{B}(\mathbf{u})^\top \widehat{\boldsymbol{\gamma}}_k$, where $\widehat{\boldsymbol{\gamma}}_k = \mathbf{Q}_2 \widehat{\boldsymbol{\theta}}_k$, for $k = 0, \dots, p$.

3. ASYMPTOTIC RESULTS

This section studies the asymptotic properties of the proposed estimators. To discuss these properties, we introduce some notation of norms. For any function g over the closure of domain Ω , denote $\|g\|_{L^2(\Omega)}^2 = \int_{\mathbf{u} \in \Omega} g^2(\mathbf{u}) du_1 du_2$ the regular L_2 norm of g , and $\|g\|_{\infty, \Omega} = \sup_{\mathbf{u} \in \Omega} |g(\mathbf{u})|$ the supremum norm of g . For directions u_j , $j = 1, 2$, let $D_{u_j}^q g(\mathbf{u})$ denote the q -th order derivative in the direction u_j at the point \mathbf{u} . Let $|g|_{v, \infty, \Omega} = \max_{i+j=v} \|D_{u_1}^i D_{u_2}^j g(\mathbf{u})\|_{\infty, \Omega}$ be the maximum norms of all the v th order derivatives of g over Ω .

Let $W^{\ell, \infty}(\Omega) = \{g : |g|_{k, \infty, \Omega} < \infty, 0 \leq k \leq \ell\}$ be the standard Sobolev space. Given random variables T_n for $n \geq 1$, we write $T_n = O_P(b_n)$ if $\lim_{c \rightarrow \infty} \limsup_n P(|T_n| \geq cb_n) = 0$. Similarly, we write $T_n = o_P(b_n)$ if $\lim_n P(|T_n| \geq cb_n) = 0$, for any constant $c > 0$. Also, we write $a_n \asymp b_n$ if there exist two positive constants c_1, c_2 such that $c_1 |a_n| \leq |b_n| \leq c_2 |a_n|$, for all $n \geq 1$.

For a triangle $\tau \in \Delta$ defined in Section 2.1, let $|\tau|$ be its longest edge length, and ρ_τ be the radius of the largest disk which can be inscribed in τ . Define the shape parameter of τ as the ratio $\pi_\tau = |\tau|/\rho_\tau$. When π_τ is small, the triangles are relatively uniform in the sense that all angles of triangles in the triangulation τ are relatively the same. Denote the size of Δ by $|\Delta| := \max\{|\tau|, \tau \in \Delta\}$, i.e., the length of the longest edge of Δ .

In the following we introduce some technical conditions.

(C1) The joint density function of $\mathbf{U} = (U_1, U_2)$, $f_{\mathbf{U}}(\cdot)$, is bounded away from 0 and infinity.

- (C2) For any $k = 0, \dots, p$, there exists a positive constant C_k such that $|X_k| \leq C_k$. The eigenvalues $\phi_0(\mathbf{u}) \leq \phi_1(\mathbf{u}) \leq \dots \leq \phi_p(\mathbf{u})$ of $\Sigma(\mathbf{u}) = E(\mathbf{X}\mathbf{X}^\top | \mathbf{U} = \mathbf{u})$ are bounded away from 0 and infinity uniformly for all $\mathbf{u} \in \Omega$; that is, there are positive constants C_1 and C_2 such that $C_1 \leq \phi_0(\mathbf{u}) \leq \phi_1(\mathbf{u}) \leq \dots \leq \phi_p(\mathbf{u}) \leq C_2$ for all $\mathbf{u} \in \Omega$.
- (C3) For any $k = 0, \dots, p$, the bivariate function $\beta_k \in W^{\ell+1, \infty}(\Omega)$ for an integer $\ell \geq 1$.
- (C4) For every $s \in \mathcal{S}_d^r(\Delta)$ and every $\tau \in \Delta$, there exists a positive constant F_1 , independent of s and τ , such that $F_1 \|s\|_{\infty, \tau} \leq \left\{ \sum_{\mathbf{U}_i \in \tau, i=1, \dots, n} s(\mathbf{U}_i)^2 \right\}^{1/2}$, for all $\tau \in \Delta$.
- (C5) Let F_2 be the largest among the numbers of observations in triangles $\tau \in \Delta$. That is, $\left\{ \sum_{\mathbf{U}_i \in \tau, i=1, \dots, n} s(\mathbf{U}_i)^2 \right\}^{1/2} \leq F_2 \|s\|_{\infty, \tau}$, for all $\tau \in \Delta$, where $\|s\|_{\infty, \tau}$ denotes the supremum norm of s over triangle τ . The constants F_1 and F_2 satisfy $F_2/F_1 = O(1)$.
- (C6) The triangulation Δ is π -quasi-uniform, that is, there exists a positive constant π such that the triangulation Δ satisfies $|\Delta|/\rho_\tau \leq \pi$, for all $\tau \in \Delta$.

Conditions (C1) and (C2) are common in the nonparametric regression literature, specifically, they are similar to Conditions (C1) and (C2) in Xue (2006) and Conditions (C1)-(C3) in Huang et al. (2004). Condition (C3) describes the requirement for the coefficient functions as usually used in the literature of nonparametric estimation. Condition (C4) ensures the existence of a discrete least squares spline. In practice, it requires that within each triangle, the number of data points should not be too small. Condition (C5) suggests that we should not put too many observations in one triangle. In Section 4, we describe the penalized least squares spline fitting so that Conditions (C4) and (C5) can be relaxed in the application, for example, F_1 can be zero for some triangles. Condition (C6) suggests the use of more uniform triangulations with smaller shape parameters and this condition can be automatically handled via delaunay and distmesh triangulation program in MATLAB/MATHEMATICA.

The following theorem provides the convergence rate of $\widehat{\beta}_k(\cdot)$. The detailed proofs of this theorem are given in the supplemental materials.

Theorem 1 *Suppose Conditions (C1)-(C6) hold, then for any $k = 0, \dots, p$, the bivariate penalized estimator $\widehat{\beta}_k(\cdot)$ is consistent and satisfies that $\|\widehat{\beta}_k - \beta_k\|_{L^2(\Omega)} = O_P\left(\frac{F_2}{F_1}|\Delta|^{\ell+1} + \frac{1}{\sqrt{n}|\Delta|}\right)$.*

Theorem 1 implies, if $F_2/F_1 = O(1)$ and the number of triangles K_n and the sample size n satisfy that $K_n \asymp n^{1/(\ell+2)}$, then the BST estimator $\widehat{\beta}_k$ has the convergence rate $\|\widehat{\beta}_k - \beta_k\|_{L^2(\Omega)}^2 = O_P(n^{-(\ell+1)/(\ell+2)})$, which is the optimal convergence rate in Stone (1982).

4. BIVARIATE PENALIZED SPLINE ESTIMATORS

When we have regions of sparse data, bivariate penalized splines, as a direct ridge regression shrinkage type global smoothing method, provide a more convenient tool for data fitting than the BST approach presented in Section 2.2. In this section, we introduce a computationally efficient and stable method to estimate the regression coefficients based on the bivariate penalized splines over triangulations (BPST). In this approach, roughness penalty parameters are used to balance the goodness-of-fit and smoothness. The number and shape of the triangles in triangulation are no longer crucial, compared with the BST in Section 2.2, as long as the minimum number of triangles is reached. To define the BPST method, let

$$\mathcal{E}(g) = \int_{\Omega} \{(D_{u_1}^2 g)^2 + 2(D_{u_1} D_{u_2} g)^2 + (D_{u_2}^2 g)^2\} du_1 du_2, \quad (6)$$

which is similar to the thin-plate spline penalty (Green and Silverman, 1994) except the latter is integrated over the entire plane \mathbb{R}^2 . An advantage of this penalty is that it is invariant with respect to Euclidean transformations of spatial coordinates, thus, the bivariate smoothing does not depend on the choice of the coordinate system.

Let $\lambda_k \geq 0$ be the penalty parameter for coefficient function β_k , $k = 0, 1, \dots, p$. Given $\{(\mathbf{U}_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$, we consider the following regularized minimization problem:

$$\min_{s_k \in \mathcal{S}_d^r(\Delta), k=0, \dots, p} \sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p \mathbf{X}_{ik} s_k(\mathbf{u}_i) \right\}^2 + \sum_{k=0}^p \lambda_k \mathcal{E}(s_k),$$

where separate penalty parameters are used to allow different smoothness for different coefficient functions. Using the bivariate splines approximation, we have the following minimization:

$$\sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p \mathbf{X}_{ik} \mathbf{B}_k(\mathbf{U}_i)^\top \boldsymbol{\gamma}_k \right\}^2 + \sum_{k=0}^p \lambda_k \boldsymbol{\gamma}_k^\top \mathbf{P}_k \boldsymbol{\gamma}_k. \quad (7)$$

where \mathbf{P}_k is the diagonally block penalty matrix satisfying that $\boldsymbol{\gamma}_k^\top \mathbf{P}_k \boldsymbol{\gamma}_k = \mathcal{E}(\mathbf{B}_k^\top \boldsymbol{\gamma}_k)$.

Similar to what has been done in Section 2.2, we remove the constraint via QR decomposition of \mathbf{H}_k^\top , then the minimization problem in (7) is converted to:

$$\sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p X_{ik} \mathbf{B}_k^\top(\mathbf{U}_i) \mathbf{Q}_{2,k} \boldsymbol{\theta}_k \right\}^2 + \sum_{k=0}^p \lambda_k \boldsymbol{\theta}_k^\top \mathbf{Q}_{2,k}^\top \mathbf{P}_k \mathbf{Q}_{2,k} \boldsymbol{\theta}_k. \quad (8)$$

Assuming $\mathbf{B}(\mathbf{u}) = \mathbf{B}_0(\mathbf{u}) = \mathbf{B}_1(\mathbf{u}) = \dots = \mathbf{B}_p(\mathbf{u}) = \{B_j(\mathbf{u})\}_{j \in \mathcal{J}}$, the minimization problem in (8) can be written as

$$\sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p X_{ik} \mathbf{B}(\mathbf{U}_i)^\top \mathbf{Q}_2 \boldsymbol{\theta}_k \right\}^2 + \sum_{k=0}^p \lambda_k \boldsymbol{\theta}_k^\top \mathbf{Q}_2^\top \mathbf{P} \mathbf{Q}_2 \boldsymbol{\theta}_k. \quad (9)$$

Let $\boldsymbol{\Lambda} = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_p)$ and $\mathbf{D}_\Lambda = \boldsymbol{\Lambda} \otimes (\mathbf{Q}_2^\top \mathbf{P} \mathbf{Q}_2)$. Solving the penalized least squares problem in (9), we obtain

$$\widehat{\boldsymbol{\theta}}^* = \left[\sum_{i=1}^n \{ \mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i) \} \{ \mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i) \}^\top + \mathbf{D}_\Lambda \right]^{-1} \sum_{i=1}^n \{ \mathbf{X}_i \otimes \mathbf{B}^*(\mathbf{U}_i) \} Y_i.$$

Therefore, the penalized estimators of $\widehat{\beta}_k^*(\mathbf{u}) = \mathbf{B}(\mathbf{u})^\top \widehat{\boldsymbol{\gamma}}_k^*$, where $\widehat{\boldsymbol{\gamma}}_k^* = \mathbf{Q}_2 \widehat{\boldsymbol{\theta}}_k^*$, $k = 0, \dots, p$.

A crucial issue for the implementation of the above penalized smoothing is the selection of the penalty parameters λ_k 's, which control the trade-off between the goodness of fit and smoothness. A standard possibility is to select the penalty parameters using the cross-validation approach, for example, multi-fold cross-validation (MCV) and generalized cross-validation (GCV). Based on our simulation studies and real data applications, we find that MCV and GCV usually yield similar results. In this work, we choose a common λ for all coefficient functions, nevertheless, separate smoothing parameters for coefficient functions can be adopted with heavier computing; see the discussions in Ruppert (2002).

5. TESTS FOR NONSTATIONARITY

5.1. Goodness-of-fit test

Statistical tests for examining if some of the coefficients vary over the space are fundamental in achieving a valid interpretation of spatial non-stationarity of the regression relationship. To test whether model (1) holds with a specified parametric form such as linear regression models, we propose a bootstrap goodness-of-fit test based on the comparison of the residual sum of squares (RSS) from both parametric and nonparametric fittings.

Consider the null hypothesis

$$H_0 : \beta_k(\mathbf{u}) = \beta_k(\mathbf{u}; \boldsymbol{\rho}), \quad 0 \leq k \leq p, \quad (10)$$

where $\beta_k(\cdot; \boldsymbol{\rho})$ is a given family of functions indexed by unknown parameter vector $\boldsymbol{\rho}$. Let $\hat{\boldsymbol{\rho}}$ be an estimator of $\boldsymbol{\rho}$. The RSS under H_0 (RSS_0) and the RSS corresponding to model (1) (RSS_1) are

$$\text{RSS}_0 = \sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p X_{ik} \beta_k(\mathbf{U}_i; \hat{\boldsymbol{\rho}}) \right\}^2, \quad \text{RSS}_1 = \sum_{i=1}^n \left\{ Y_i - \sum_{k=0}^p X_{ik} \hat{\beta}_k(\mathbf{U}_i) \right\}^2.$$

The test statistic is defined as

$$T_n = (\text{RSS}_0 - \text{RSS}_1)/\text{RSS}_1 = \text{RSS}_0/\text{RSS}_1 - 1, \quad (11)$$

and we reject the null hypothesis (10) for large values of T_n . Due to the feature of less assumption on the distribution of the error term of the model, bootstrap is a good technique for testing the nonstationarity for the spatially varying coefficient models. We use the following nonparametric bootstrap approach (Cai et al., 2000) to evaluate the p -value of the test.

Step 1. Based on the data $\{(\mathbf{U}_i, \mathbf{X}_i, Y_i)\}_{i=1}^n$, obtain the following residuals $\hat{\varepsilon}_i = Y_i - \sum_{k=0}^p X_{ik}\hat{\beta}_k(\mathbf{U}_i)$, $i = 1, \dots, n$, and calculate 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. Generate the bootstrap residuals $\{\varepsilon_i^*\}_{i=1}^n$ from the empirical distribution function of the centered residuals $\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}$ in Step 1, and define $Y_i^* = \sum_{k=0}^p X_{ik}\beta_k(\mathbf{U}_i; \hat{\boldsymbol{\rho}}) + \varepsilon_i^*$;

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

Step 4. Repeat Steps 2 and 3 B times and obtain a bootstrap sample of the test statistic T_n as $\{T_{nb}^*\}_{b=1}^B$, and the p -value is estimated by $\hat{p} = \sum_{b=1}^B I(T_{nb}^* \geq T_{obs})/B$, where $I(\cdot)$ is the indicator function, and T_{obs} is the observed value of the test statistic T_n by (11); or reject the null hypothesis H_0 when T_n is greater than the upper- α quantile of $\{T_{nb}^*\}_{b=1}^B$.

Note that the nonparametric estimate is always consistent, no matter the null or the alternative hypothesis is correct. So here we bootstrap the centralized residuals from the nonparametric fit instead of the parametric fit, and this should provide a consistent estimator of the null hypothesis even when the null hypothesis does not hold. As proved in Kreiss et al. (2008), which considered nonparametric bootstrap tests in a general nonparametric regression setting, asymptotically the conditional distribution of the bootstrap test statistic

is indeed the distribution of the test statistic under the null hypothesis, as long as $\widehat{\boldsymbol{\rho}}$ converges to $\boldsymbol{\rho}$ at the root- n rate.

5.2. Testing individual function stationarity

One important question arose in varying coefficient literature is that “Does a particular set of local parameter estimates exhibit significant spatial variation?”. To answer this question, we focus on testing the following null hypothesis

$$H_{0k} : \beta_k(\mathbf{u}) = \beta_k, \quad \text{v.s.} \quad H_{1k} : \beta_k(\mathbf{u}) \neq \beta_k, \quad \text{for a fixed } k = 0, \dots, p. \quad (12)$$

To conduct a hypothesis test, the variability of the local estimates can be used to examine the plausibility of the stationarity assumption held in traditional regression (Brunsdon et al., 1996, 1999). Specifically, for a given covariate function β_k at location i , suppose $\widehat{\beta}_k(\mathbf{u}_i)$ is the BST or BPST estimate of $\beta_k(\mathbf{u}_i)$. If we take n values of this parameter estimate (one for each location point within the region), an estimate of variability of the parameter is given by the variance of the n parameter estimates.

The test statistic is defined as

$$V_{nk} = \frac{1}{n-1} \sum_{i=1}^n \left(\widehat{\beta}_k(\mathbf{U}_i) - \overline{\widehat{\beta}_k} \right)^2, \quad (13)$$

and we reject the null hypothesis (12) for large values of V_{nk} .

The next stage is to determine the sampling distribution of V_{nk} under the null hypothesis. Under H_{0k} , any permutation of \mathbf{U}_i amongst the data points is equally likely. Thus, the observed value of V_{nk} could be compared to the values obtained from randomly rearranging the data in space and repeating the BPST procedure. The comparison between the observed V_{nk} value and those obtained from a large number of randomized distributions can then form the basis of the significance test.

Step 1. Randomly shuffle the n locations and obtain $\{\mathbf{U}_i^*\}_{i=1}^n$;

Step 2. Calculate the test statistic V_{nk}^* based on the sample $\{(\mathbf{U}_i^*, \mathbf{X}_i, Y_i)\}_{i=1}^n$;

Step 3. Repeat Steps 1 and 2 B times and obtain a sample of the test statistic V_{nk}^* as $\{V_{nk,b}^*\}_{b=1}^B$, and the p -value is estimated by $\hat{p} = \sum_{b=1}^B I(V_{nk,b}^* \geq V_{k,obs})/B$, where $V_{k,obs}$ is the observed value of the test statistic V_{nk} by (13), or reject the null hypothesis when V_{nk} is greater than the upper- α quantile of $\{V_{nk,b}^*\}_{b=1}^B$.

6. SIMULATION

In this section, we analyze synthetic data generated from the model to assess the validity of the proposed estimation and inference procedure based on BST and BPST smoothing methods. We also implement the GWR method to each of these artificial data and compare the estimator with our proposed ones.

To obtain the BST and BPST estimators, we set degree $d = 2$ and smoothness $r = 1$ when generating the bivariate spline basis functions. The supplementary materials provide more simulation results with different values of d and different triangulations. For BPST, a common penalty λ for all coefficient functions is selected using 5-fold cross-validation from a 9-point grid, where the values of $\log_{10}(\lambda)$ are equally spaced between -2 and 2 . For the GWR method, we use the “`spgwr`” R package to obtain the GWR estimator. In all simulation studies, the total number of replications is 500.

6.1. Simulation study 1

Following Shen et al. (2011), the spatial layout in this example is designated as a $[0, 6]^2$ domain, and the population is collected at $N = 100 \times 100$ lattice points with equal distance between any two neighboring points along the horizontal and vertical directions. At each location, the response variable is generated by $Y_i = \beta_0(\mathbf{U}_i) + X_i\beta_1(\mathbf{U}_i) + \varepsilon_i$, $i = 1, \dots, N$,

where X_i is generated randomly from a Uniform(0,2) distribution. The random error, ε_i , $i = 1, \dots, n$, are generated independently from $N(0, 1)$, and the coefficient functions are

$$\beta_0(u_1, u_2) = 2\sin\left(\frac{\pi u_1}{6}\right), \quad \beta_1(u_1, u_2) = \frac{2}{81}\{9 - (3 - u_1)^2\}\{9 - (3 - u_2)^2\}. \quad (14)$$

See Figure 1 (a) and (b) for the contour plots of the two true coefficient functions. We randomly sample $n = 500, 1000$ and 2000 points from the 100×100 points in each Monte Carlo experiment.

[Figure 1 about here.]

Figure 1 (c) shows the triangulation used to obtain the BST and BPST estimators, and there are 13 triangles and 12 vertices in this triangulation. The mean squared estimation error (MSE) and mean squared prediction error (MSPE) for the estimators of the coefficient functions, as well as the MSPE of the response variable Y , are computed using:

$$\begin{aligned} \text{MSE}(\hat{\beta}_k) &= \frac{1}{n} \sum_{i=1}^n \left\{ \hat{\beta}_k(\mathbf{U}_i) - \beta_k(\mathbf{U}_i) \right\}^2, \quad \text{MSPE}(\hat{\beta}_k) = \frac{1}{N} \sum_{i=1}^N \left\{ \hat{\beta}_k(\mathbf{U}_i) - \beta_k(\mathbf{U}_i) \right\}^2, \quad k = 0, 1, \\ \text{MSPE}(\hat{Y}) &= \frac{1}{N} \sum_{i=1}^N (\hat{Y}_i - Y_i)^2. \end{aligned}$$

In addition, we report the bias (BIAS) of $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ in estimating the true variance of errors σ^2 . All the results are summarized in Table 1.

[Table 1 about here.]

We compare the proposed BST and BPST with GWR by evaluating the estimation accuracy and their predictive accuracy of the spatial pattern. As the sample size increases, all three estimators tend to result in better performance in terms of MSE, MSPE and BIAS. Moreover, regardless of the sample size, both the BST and BPST estimators outperform the GWR estimator. Compared with the BST estimators, the BPST estimators are more

stable, especially when the sample size is small, and the difference between the BST and the BPST estimators is getting smaller as the sample size increases. The supplementary materials provide more simulation results with different triangulations, which suggest that the number of triangles or the number of basis functions only have a little effect on the BPST estimator, especially when there is a sufficient number of triangles.

Figure 2 visualizes the estimated surfaces of $\beta_0(\cdot)$ and $\beta_1(\cdot)$ using BST, BPST and GWR, which are based on one typical replication with $n = 2000$. These plots suggest that the BPST method is able to estimate the spatial pattern with the greatest accuracy, followed by the BST method, and the GWR is not able to capture the spatial pattern very accurately.

[Figure 2 about here.]

In terms of the computing, since the GWR technique is largely based on kernel regression, a locally weighted regression is required at every single point in the dataset, which results in a great computational complexity. In contrast, both BST and BPST can be formulated as one single least squares problem, thus, the computing is very easy and fast. The last column (Time) in Table 1 summarizes the average computing time per iteration in seconds for each method. All the methods are implemented using a personal computer with Intel(R) Core(TM) i5 CPU dual core @ 2.90GHz and 8.00GB RAM. Specifically, one can see that as the sample size increases, the computational time for GWR method increases dramatically while BST and BPST almost provide a linear complexity of the sample size when the number of triangles is much less than the sample size.

As one of the most important inferences in the GWR literature, the test for a globally stationary regression relationship can provide the information that a SVCM is really necessary for the given data. Next we investigate the performance of the proposed bootstrap test described in Section 5.1. We consider the following hypothesis:

$$H_0 : \beta_k(\mathbf{u}) = \beta_k, \quad k = 0, 1 \quad \text{v.s.} \quad H_1 : \beta_k(\mathbf{u}) \neq \beta_k, \quad \text{for at least one } k.$$

Note that H_0 corresponds to the ordinary linear regression model. The power function is evaluated under a sequence of the alternative models indexed by δ :

$$H_1 : \beta_k(\mathbf{u}) = \bar{\beta}_k^0 + \delta(\beta_k^0(\mathbf{u}) - \bar{\beta}_k^0), \quad k = 0, 1, \quad (0 \leq \delta \leq 1), \quad (15)$$

where $\beta_k^0(\mathbf{u})$, $k = 0, 1$, are given in (14), and $\bar{\beta}_k^0$ is the average height of $\beta_k^0(\mathbf{u})$, and in our simulation, $\bar{\beta}_0^0 = 1.2604$, $\bar{\beta}_1^0 = 0.8710$. The parameter δ is designated different values to evaluate the power of the test. The null hypothesis corresponds to $\delta = 0$ in the coefficients.

We apply the bootstrap goodness-of-fit test in a simulation with 500 replications of sample size $n = 500, 1000$, and record the relative frequencies of rejecting H_0 under the significance level $\alpha = 0.05$. For each realization, we repeat bootstrap sampling 100 times. Figure 4 illustrates the empirical frequencies of rejecting H_0 against δ using BPST and GWR methods. For GWR, we use the “`BFC02.gwr.test`” function within the “`spgwr`” package in R, where the test result is obtained by test statistics based on the RSS described in Fotheringham et al. (2002). When $\delta = 0$, these relative frequencies represent the size of the test. For BPST, the relative frequency is 0.060, which is fairly close to the given significant level 5%. This demonstrates that bootstrap estimate of the null distribution is approximately correct. However, the GWR gives 0.000, which is much smaller than the significant level. For our BPST based bootstrap test, regardless of the sample size, the power increases rapidly to one when $\delta = 0.2$, suggesting that the proposed test is quite powerful. Overall, the proposed test is of a higher power in identifying the varying coefficients than the GWR based test.

[Figure 3 about here.]

Next we conduct an individual stationarity test with $H_{0k} : \beta_k(\mathbf{u}) = \beta_k$ v.s. $H_{1k} : \beta_k(\mathbf{u}) \neq \beta_k$, for $k = 0, 1$. Similar to (15), the power function is evaluated under a sequence of the alternative models indexed by δ , specifically, $H_{1k} : \beta_k(\mathbf{u}) = \bar{\beta}_k^0 + \delta(\beta_k^0(\mathbf{u}) - \bar{\beta}_k^0)$, $(0 \leq \delta \leq 1)$.

We apply the individual stationarity test described in Section 5.2 at significant level $\alpha =$

0.05 with $n = 500, 1000$, respectively. Similar as in the global stationarity, 500 replications with $B = 100$ bootstrap samples in each replication are conducted to compute the p -value. The individual test for the GWR is based on the test statistic proposed in Leung et al. (2000) and implemented using the R function “`LMZ.F3GWR.test`”.

When $\delta = 0$, the rejection frequencies of BPST are reasonably close to the given significance level for both coefficient functions. Consider the specific alternative with $\delta = 0.4$, where the functions $\{\beta_k(\mathbf{u})\}$ under H_{1k} are shown in Figure 4. Even with such a small difference, we can correctly detect the alternative over 80% of the 500 simulations. With the value of δ increasing, the rejection frequencies increase rapidly, and thus, the rejection frequency of BPST is definitely high if coefficient functions are indeed spatially varying. However, for GWR, the rejection frequencies are much higher than the given significance level, which indicates a much larger Type I error. Although the results improve with the sample size increasing, the rejection rates are still twice or three times higher than the significance level.

[Figure 4 about here.]

In summary, the simulation study demonstrates that the proposed bootstrap tests can well approximate the null distribution of the test statistic even for moderate sample size.

6.2. Simulation study 2

In this simulation study, we consider a modified horseshoe domain constructed in Wood et al. (2008). In particular, we divide the entire horseshoe domain evenly into $N = 401 \times 901$ grid points, which is considered as the population. We adopt the coefficient functions shown in Figure 5 (a) and (b), where $\beta_0(\cdot)$ is the same function used in Wood (2003) and $\beta_1(u_1, u_2) = 4\sin\{0.05\pi(u_1^2 + u_2^2)\}$. The response variable Y_i are generated from the following model: $Y_i = \beta_0(\mathbf{u}_i) + X_i\beta_1(\mathbf{u}_i) + \varepsilon_i$, $i = 1, \dots, N$, where X_i is generated randomly from a Uniform(0,2) distribution. The random error, ε_i , $i = 1, \dots, n$, are generated independently

from $N(0, 0.5^2)$ distribution. For each of the 500 Monte Carlo experiment, we randomly sample $n = 2000$ and 5000 locations uniformly on the domain.

[Figure 5 about here.]

Figure 5 (c) shows the triangulation used to obtain the BST and BPST estimators, and there are 77 triangles and 65 vertices in this triangulation. The results of MSE, MSPE and BIAS based on 500 replications are summarized in Table 2, and the predicted surfaces from one iteration when $n = 2000$ are demonstrated in Figure 6. These results highlight, on the irregular domain, that both BST and BPST methods can provide more accurate and efficient estimation than the GWR method.

[Table 2 about here.]

[Figure 6 about here.]

Another computational issue worth mentioning is that when GWR is used for prediction, a locally weighted regression is also required at every single point in the predicting dataset. In this example, the prediction is conducted over the entire $401 \times 901 = 361,301$ grid points. Therefore, the entire process cannot be completed using personal computers, instead, the GWR algorithm is conducted via cluster using a parallel computing of 24 general-purpose compute nodes with 128GB RAM associated to each node and on average each iteration takes more than 5 hours to complete for $n = 2000$. Unfortunately, when $n = 5000$, we are not able to obtain any results using the same cluster with the same setting within one week. On the other hand, for BST and BPST, the prediction can be done with one simple matrix multiplication, and thus it is very fast even when the dataset is huge. For example, one BPST-based iteration of size 5000 only takes about one minute on a personal computer with Intel(R) Core(TM) i5 CPU dual core @ 2.90GHz and 8.00GB RAM.

7. APPLICATION TO AIR POLLUTION DATA ANALYSIS

Recently, fine particles (PM_{2.5}, particulate matter with diameter of 2.5 micrometers or less) has become a major air quality concern since it poses significant risks to human health, such as asthma, chronic bronchitis, lung cancer, atherosclerosis, etc. Many recent research works (Tai et al., 2010; Hu et al., 2013; Russell et al., 2017, for example) suggest that the PM_{2.5} concentrations depend on meteorological conditions. To improve the current pollution control strategies, there is an urgent need for a more comprehensive understanding of PM_{2.5} and a more accurate quantification between the meteorological drivers and the levels of PM_{2.5}.

In this section, we show the applicability of the proposed SVCM on a meteorological dataset to study the effect of meteorological characteristics on air quality. In our study, daily mean surface concentrations of total PM_{2.5} for the year 2011 are obtained from the United States Environmental Protection Agency; meteorological drivers are provided by the National Oceanic and Atmospheric Administration (<http://www.esrl.noaa.gov/psd/>); daily total gridded precipitation (PPTN), surface wind speed (WS), surface daily minimum air temperature (T_{min}) and surface daily maximum air temperature (T_{max}) are acquired from Livneh data (Livneh et al., 2013); air relative humidity (RH) and total column cloud cover (TCDC) are obtained from North American Regional Reanalysis. See Table 3 for details.

[Table 3 about here.]

Noting that there are some missing values in RH, TCDC and total PM_{2.5}, we aggregate the data by season and focus on the most severe polluted season in a year — winter (December, January, February). We predict the PM_{2.5} for winter season using the proposed SVCM:

$$\begin{aligned} \text{PM}_{2.5} = & \beta_0(\mathbf{u}) + \beta_1(\mathbf{u})\text{PPTN} + \beta_2(\mathbf{u})\text{RH} + \beta_3(\mathbf{u})T_{min} + \beta_4(\mathbf{u})T_{max} \\ & + \beta_5(\mathbf{u})\text{WS} + \beta_6(\mathbf{u})\text{TCDC} + \varepsilon. \end{aligned} \tag{16}$$

We also consider the multiple linear regression (MLR) without using the spatial information:

$$\text{PM}_{2.5} = \beta_0 + \beta_1 \text{PPTN} + \beta_2 \text{RH} + \beta_3 T_{\min} + \beta_4 T_{\max} + \beta_5 \text{WS} + \beta_6 \text{TCDC} + \varepsilon. \quad (17)$$

To evaluate different methods, we examine both the estimation accuracy and the prediction accuracy of the MLR in (17), the GWR and the BPST in (16). The out-of-sample prediction errors of each method are calculated by 10-fold cross validation. The MSE and MSPE of three methods are summarized in Table 5. It is obvious that BPST estimator provides a much more accurate estimation and prediction. Figure 7 (b)-(h) summarize the coefficient estimation results via BPST.

[Table 4 about here.]

[Figure 7 about here.]

A natural question is if the coefficients are really varying over space in model (16). We now use our proposed test procedure in Section 5 to answer this question. The p -values of these tests are listed in Table 5. For the global stationary test, the p -value is smaller than 0.001, so under the significance level $\alpha = 0.05$ we conclude that at least one of the coefficient is nonstationary over the entire United States. For the individual stationarity test, the resulting p -values for the intercept, T_{\max} , WS and TCDC are all smaller than the significant level $\alpha = 0.05$, indicating that the coefficients $\beta_0(\mathbf{u})$, $\beta_4(\mathbf{u})$, $\beta_5(\mathbf{u})$ and $\beta_6(\mathbf{u})$ are really varying over space. However, for PPTN, RH and T_{\min} , the p -value $\gg \alpha$, so we don't have enough evidence to reject the null hypothesis. The estimated coefficient function plots in Figure 7 confirm the conclusion of the proposed tests.

[Table 5 about here.]

8. CONCLUDING REMARKS

In summary, the proposed method has the following advantages in analyzing the spatial nonstationarity of a regression relationship for spatial data. First, comparing with GWR, the proposed method is much more computationally efficient to deal with large datasets. Specifically, the computational complexity of BST and BPST is $O(nK_n^2)$, which indicates that it is almost linear in terms of the sample size. In addition, as a global estimation with an explicit model expression, the proposed spline approach enables easy-to-implement prediction compared to the local approaches. Second, the proposed method can overcome the problem of “leakage” across the complex domains that many conventional tools suffer. Third, by introducing the roughness penalty into the BST, the BPST can alleviate the adverse effect of the collinearity problem in GWR (Wheeler and Tiefelsdorf, 2005) and provide more accurate estimators of the coefficient functions. The BPST also easily allows different smoothness for different functional coefficients, which is enabled by assigning different penalty parameters accordingly. Finally, with increasing volumes of data being collected on the environment through remote sensing platforms, complex sensor networks and GPS movement, this work provides one feasible approach to study large scale environmental spatial data.

The proposed method in this article can be easily extended to semiparametric varying-coefficient partially linear models (Brunsdon et al., 1999; Fotheringham et al., 2002; Fan and Huang, 2005), where some coefficients in the model are assumed to be constant and the remaining coefficients are allowed to spatially vary across the studied region.

In spatial data analysis, there are mainly two issues: spatial dependence and spatial heterogeneity, and our paper focuses on the later one. To understand the theoretical behavior of the estimators, we assume the errors are independent. Although this assumption is not uncommon in the GWR literature, it is more realistic to relax the independence assumption. The spatial dependence can be alleviated by choosing the optimal triangulation, it may not fully vanish, and certainly there is more future work ahead to investigate this issue. Sun

et al. (2014) proposed a semiparametric spatial autoregressive varying coefficient model:

$$Y_i = \alpha \sum_{j \neq i} w_{ij} Y_j + \mathbf{X}_i^\top \boldsymbol{\beta}(\mathbf{U}_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (18)$$

where w_{ij} is the impact of Y_j on Y_i , for example, a specified physical or economic distance. This model considers the neighboring effect and is thus able to take care of the spatial dependence issue. We are interested in extending our work to this class of models for irregularly spaced data over complex domains. However, it is challenging to define w_{ij} (distance) for our method due to the “leakage” problem across complex boundary features. Another interesting future work is the spatio-temporal extension to analyze data collected across time as well as space. We might be able to establish some promising theoretical results under some dependent error assumptions if letting the number of time points go to infinity. We believe more careful and intensive future work is necessary in these directions.

Acknowledgements

Guannan Wang’s research was partially supported by the Faculty Summer Research Grant from College of William & Mary. Li Wang’s research was partially supported by the National Science Foundation grant DMS-1542332. The authors are very grateful to Ming-Jun Lai for providing us with the Matlab code on bivariate spline basis constructions. The authors would also like to thank the editor, associate editor and two reviewers for their valuable comments and suggestions to improve the quality of the paper.

REFERENCES

- Al-Sulami, D., Jiang, Z., Lu, Z., and Zhu, J. (2017). Estimation for semiparametric nonlinear regression of irregularly located spatial time-series data. *Econometrics and Statistics*, 2:22–35.
- Assunção, M. R. (2003). Space varying coefficient models for small area data. *Environmetrics*, 14:453–473.
- Banerjee, S., Gelfand, E. A., Finley, O. A., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70:825–848.
- Bitter, C., Mulligan, F. G., and Dall’erba, S. (2007). Incorporating spatial variation in housing attribute prices: a comparison of geographically weighted regression and the spatial expansion method. *Journal of Geographical Systems*, 9:7–27.
- Brunsdon, C., Fotheringham, A. S., and Charlton, E. M. (1996). Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis*, 28:281–298.
- Brunsdon, C., Fotheringham, A. S., and Charlton, E. M. (1998). Geographically weighted regression. *Journal of the Royal Statistical Society: Series D (The Statistician)*, 47:431–443.
- Brunsdon, C., Fotheringham, A. S., and Charlton, E. M. (1999). Some notes on parametric significance tests for geographically weighted regression. *Journal of Regional Science*, 39:497–524.
- Cai, Z., Fan, J., and Yao, Q. (2000). Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95:941–956.
- Fan, J. and Huang, T. (2005). Profile likelihood inferences on semiparametric varying-coefficient partially linear models. *Bernoulli*, 11:1031–1057.
- Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Annals of Statistics*, 27:1491–1518.
- Ferguson, C. A., Bowman, A. W., Scott, E. M., and Carvalho, L. (2009). Multivariate varying-coefficient models for an ecological system. *Environmetrics*, 20:460–476.
- Finley, O. A. (2011). Comparing spatially-varying coefficients models for analysis of ecological data with non-stationary and anisotropic residual dependence. *Methods in Ecology and Evolution*, 2:143–154.
- Fotheringham, A. S., Brunsdon, C., and Charlton, E. M. (2002). *Geographically weighted regression: the analysis of spatially varying relationships*. John Wiley & Sons.
- Gelfand, E. A., Kim, H. J., Sirmans, C. F., and Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, 98:387–396.

- Green, P. J. and Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*. Chapman and Hall, London.
- Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 55:757–796.
- Helbich, M. and Griffith, A. D. (2016). Spatially varying coefficient models in real estate: eigenvector spatial filtering and alternative approaches. *Computer, Environment and Urban Systems*, 57:1–11.
- Hu, X., Waller, A. L., Al-Hamdan, Z. M., Crosson, L. W., Estes, G. M., Estes, M. S., Quattrochi, A. D., Sarnat, A. J., and Liu, Y. (2013). Estimating ground-level pm_{2.5} concentrations in the southeastern u.s. using geographically weighted regression. *Environmental Research*, 121:1–10.
- Huang, J., Wu, O. C., and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, 14:763–788.
- Huang, Y., Yuan, M., and Lu, Y. (2017). Spatially varying relationships between surface urban heat islands and driving factors across cities in china. *Environment and Planning B: Urban Analytics and City Science*, 0:1–18.
- Kreiss, J. P., Neumann, H. M., and Yao, Q. (2008). Bootstrap tests for simple structures in nonparametric time series regression. *Statistics and its Interface*, 1:367–380.
- Lai, M. J. and Schumaker, L. L. (2007). *Spline functions on triangulations*. Cambridge University Press.
- Lai, M. J. and Wang, L. (2013). Bivariate penalized splines for regression. *Statistica Sinica*, 23:1399–1417.
- Leung, Y., Mei, C. L., and Zhang, W. X. (2000). Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environment and Planning A*, 32:9–32.
- Lian, H. (2012). Variable selection for high-dimensional generalized varying-coefficient models. *Statistica Sinica*, 22:1563–1588.
- Livneh, B., Rosenberg, A. E., Lin, C., Nijssen, B., Mishra, V., Andreadis, M. K., Maurer, P. E., and Lettenmaier, P. D. (2013). A long-term hydrologically based dataset of land surface fluxes and states for the conterminous united states: update and extensions. *Journal of Climate*, 26:9384–9392.
- Lu, Z., Steinskog, D. J., Tjøstheim, D., and Yao, Q. (2009). Adaptively varying-coefficient spatiotemporal models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71:859–880.
- Lu, Z., Tang, Q., and Cheng, L. (2014). Estimating spatial quantile regression with functional coefficients: a robust semiparametric framework. *Bernoulli*, 20:164–189.

- Mei, C. L., Wang, N., and Zhang, W. X. (2006). Testing the importance of the explanatory variables in a mixed geographically weighted regression model. *Environment and Planning A*, 38:587–598.
- Nakaya, T., Fotheringham, A. S., Brunsdon, C., and Charlton, E. M. (2005). Geographically weighted poisson regression for disease association mapping. *Statistics in Medicine*, 24:2695–2717.
- Persson, P. O. and Strang, G. (2004). A simple mesh generator in matlab. *SIAM Review*, 46:329–345.
- Ramsay, T. (2002). Spline smoothing over difficult regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64:307–319.
- Ruppert, D. (2002). Selecting the number of knots for penalized splines. *Journal of computational and graphical statistics*, 11:735–757.
- Russell, T. B., Wang, D., and McMahan, S. C. (2017). Spatially modeling the effects of meteorological drivers of $pm_{2.5}$ in the eastern united states via a local linear penalized quantile regression estimator. *Environmetrics*, 28:e2448–n/a.
- Sangalli, L., Ramsay, J., and Ramsay, T. (2013). Spatial spline regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 75:681–703.
- Shen, S. L., Mei, C. L., and Zhang, Y. J. (2011). Spatially varying coefficient models: testing for spatial heteroscedasticity and reweighting estimation of the coefficients. *Environment and Planning A*, 43:1723–1745.
- Stone, J. C. (1982). Optimal global rates of convergence for nonparametric regression. *The annals of statistics*, 10:1040–1053.
- Su, S., Lei, C., Li, A., J., P., and Z., C. (2017). Coverage inequality and quality of volunteered geographic features in chinese cities: analyzing the associated local characteristics using geographically weighted regression. *Applied Geography*, 78:78–93.
- Sun, Y., Yan, H., Zhang, W., and Lu, Z. (2014). A semiparametric spatial dynamic model. *Annals of Statistics*, 42:700–727.
- Tai, A., Mickley, J. L., and Jacob, J. D. (2010). Correlations between fine particulate matter ($pm_{2.5}$) and meteorological variables in the united states: Implications for the sensitivity of $pm_{2.5}$ to climate change. *Atmospheric Environment*, 44:3976–3984.
- Tang, Q. (2014). Robust estimation for functional coefficient regression models with spatial data. *Statistics*, 48:388–404.

- Tang, Q. and Cheng, L. (2009). B-spline estimation for varying coefficient regression with spatial data. *Science in China Series A: Mathematics*, 11:2321–2340.
- Waller, A. L., Zhu, L., Gotway, A. C., Gorman, M. D., and Gruenewald, J. P. (2007). Quantifying geographic variations in associations between alcohol distribution and violence: a comparison of geographically weighted regression and spatially varying coefficient models. *Stochastic Environmental Research and Risk Assessment*, 21:573–588.
- Wheeler, D. and Tiefelsdorf, M. (2005). Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7:161–187.
- Wood, N. S., Bravington, V. M., and Hedley, L. S. (2008). Soap film smoothing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70:931–955.
- Wood, S. N. (2003). Thin plate regression splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 65:95–114.
- Xue, L. and Yang, L. (2006). Additive coefficient modeling via polynomial spline. *Statistica Sinica*, 16:1423–1446.
- Zhang, Q., Deng, X., Qian, P., and Wang, X. (2013). Spatial modeling for refining and predicting surface potential mapping with enhanced resolution. *Nanoscale*, 5:921–926.

FIGURES

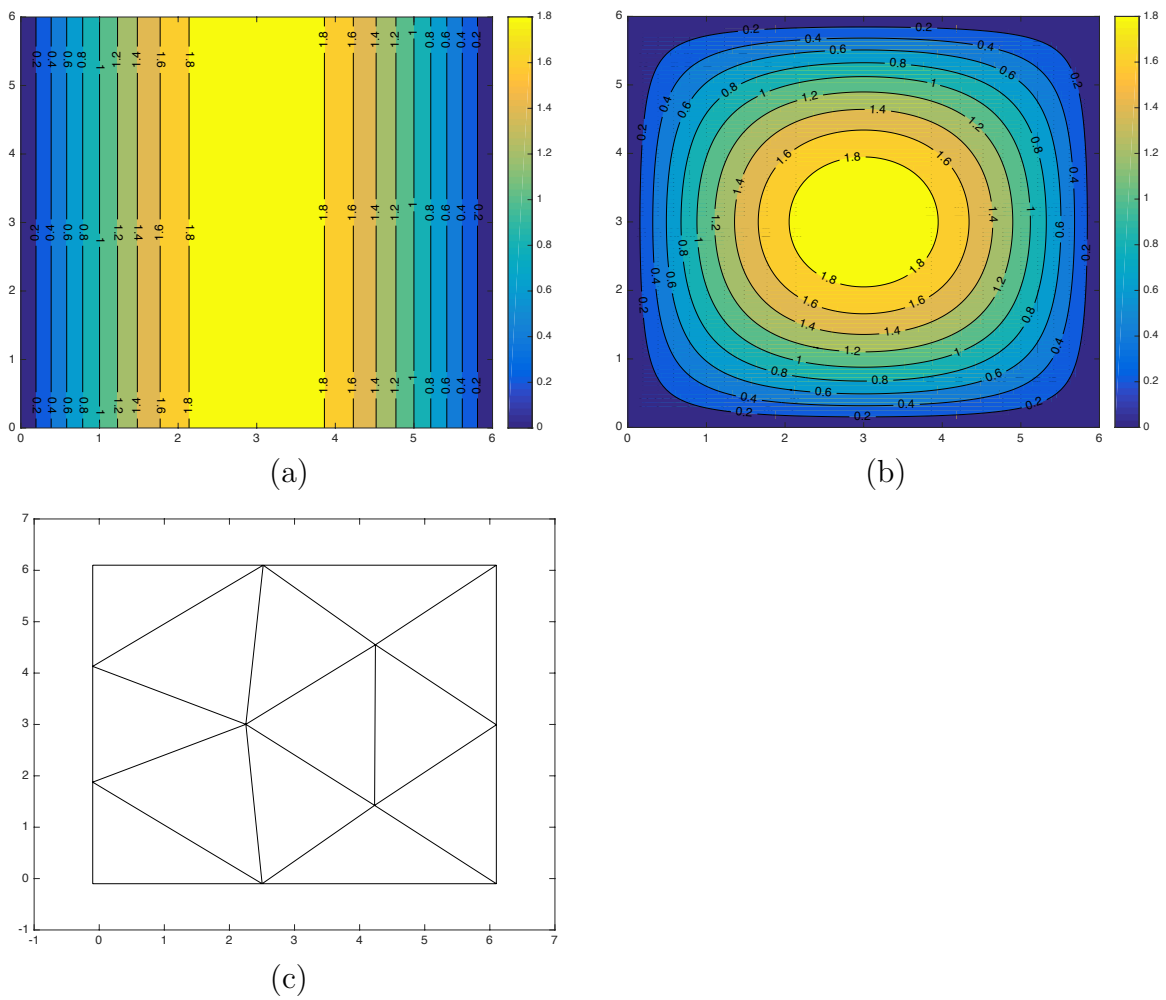


Figure 1. Simulation study 1: plots of (a) true β_0 ; (b) true β_1 ; (c) triangulation.

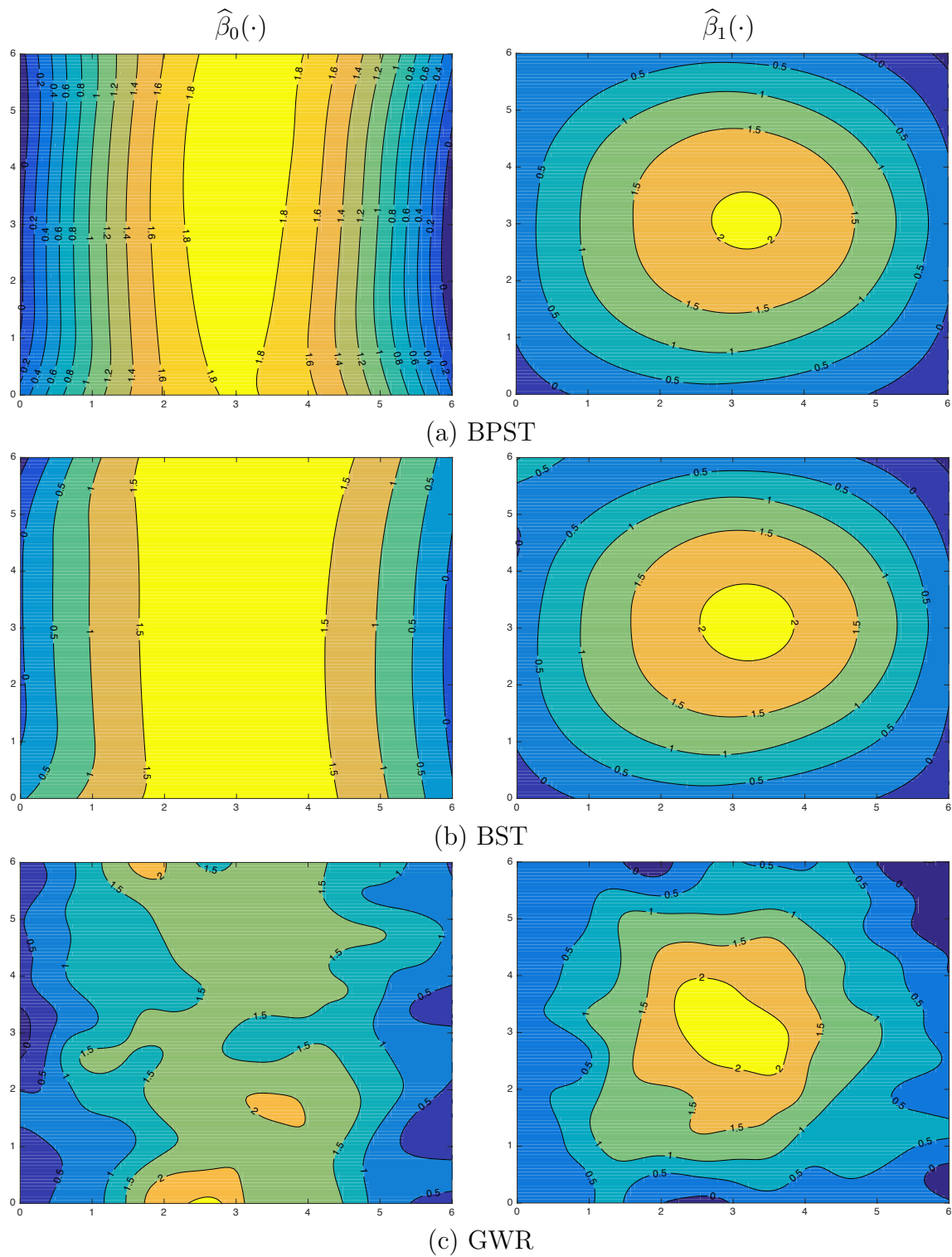


Figure 2. Simulation study 1: estimated surface via (a) BPST; (b) BST; (c) GWR based on sample size $n = 2000$.

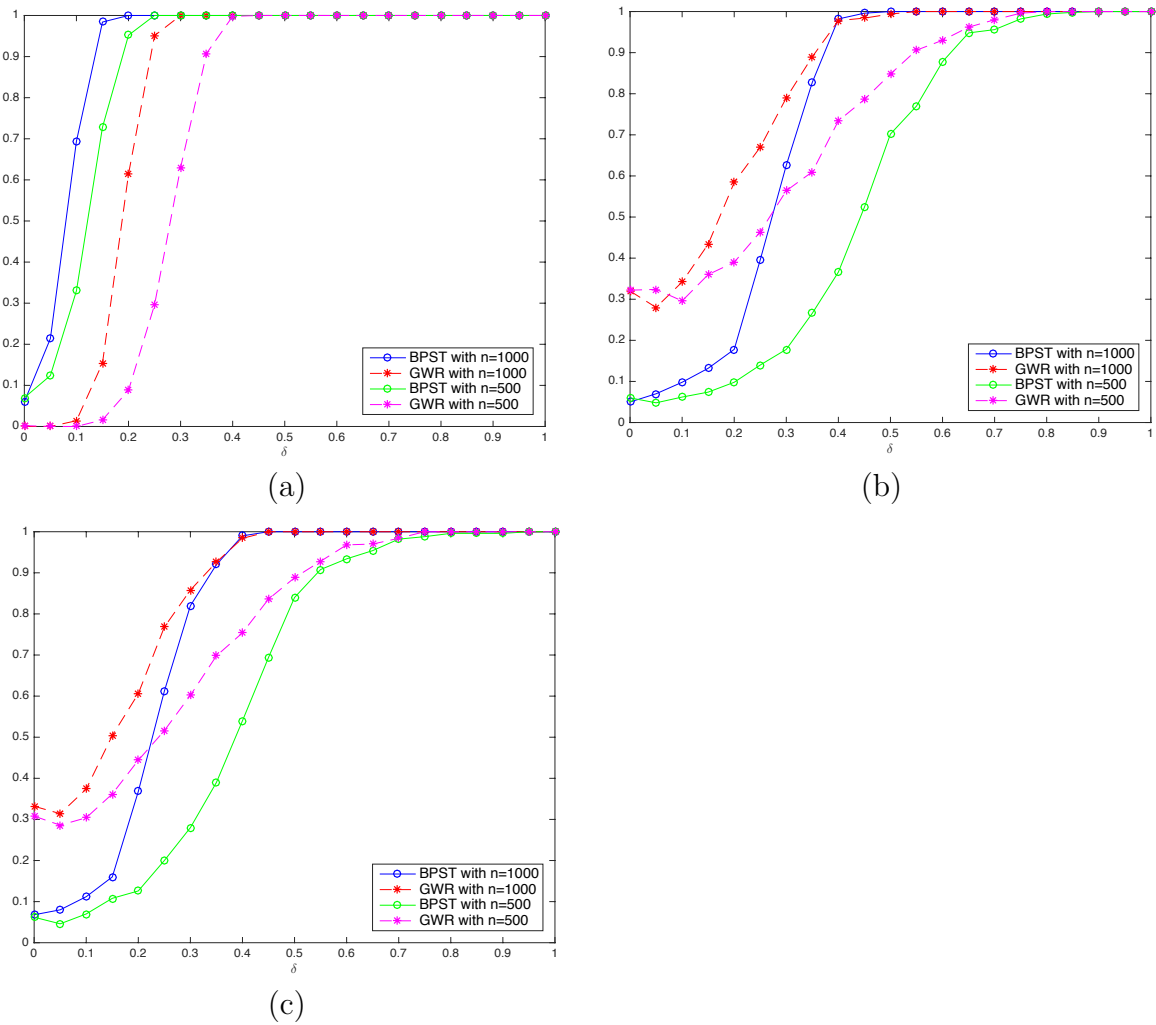


Figure 3. Type I Errors and Power for Bootstrap Tests in Simulation study 1: (a) $H_0 : \beta_k(\mathbf{u}) = \beta_k, k = 1, 2$ vs $H_1 : \beta_k(\mathbf{u}) = \bar{\beta}_k^0 + \delta(\beta_k^0(\mathbf{u}) - \bar{\beta}_k^0), k = 1, 2$; (b) $H_0 : \beta_0(\mathbf{u}) = \beta_0$ vs $H_1 : \beta_0(\mathbf{u}) = \bar{\beta}_0^0 + \delta(\beta_0^0(\mathbf{u}) - \bar{\beta}_0^0)$; (c) $H_0 : \beta_1(\mathbf{u}) = \beta_1$ vs $H_1 : \beta_1(\mathbf{u}) = \bar{\beta}_1^0 + \delta(\beta_1^0(\mathbf{u}) - \bar{\beta}_1^0)$.

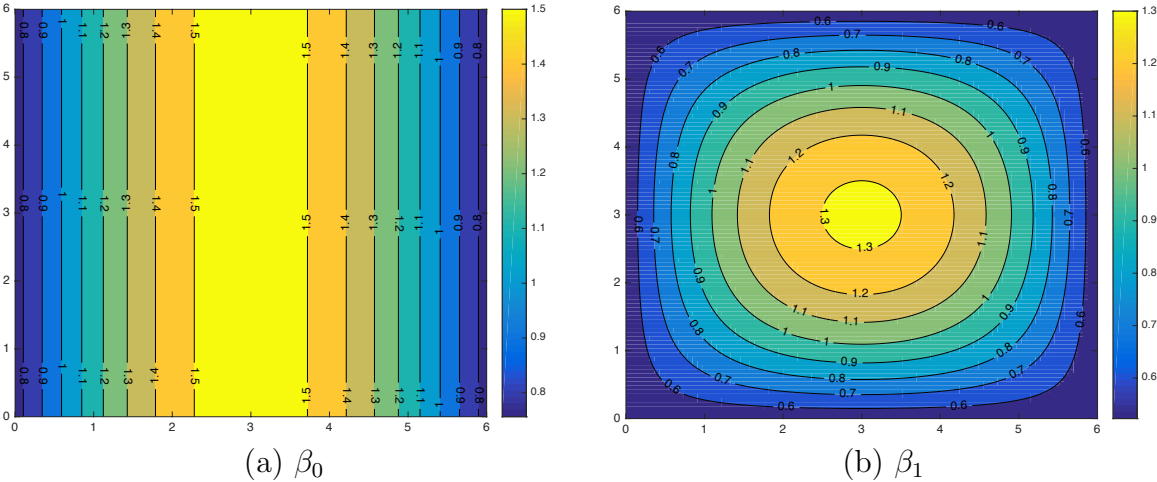


Figure 4. Simulation study 1: coefficient functions under H_1 when $\delta = 0.4$.

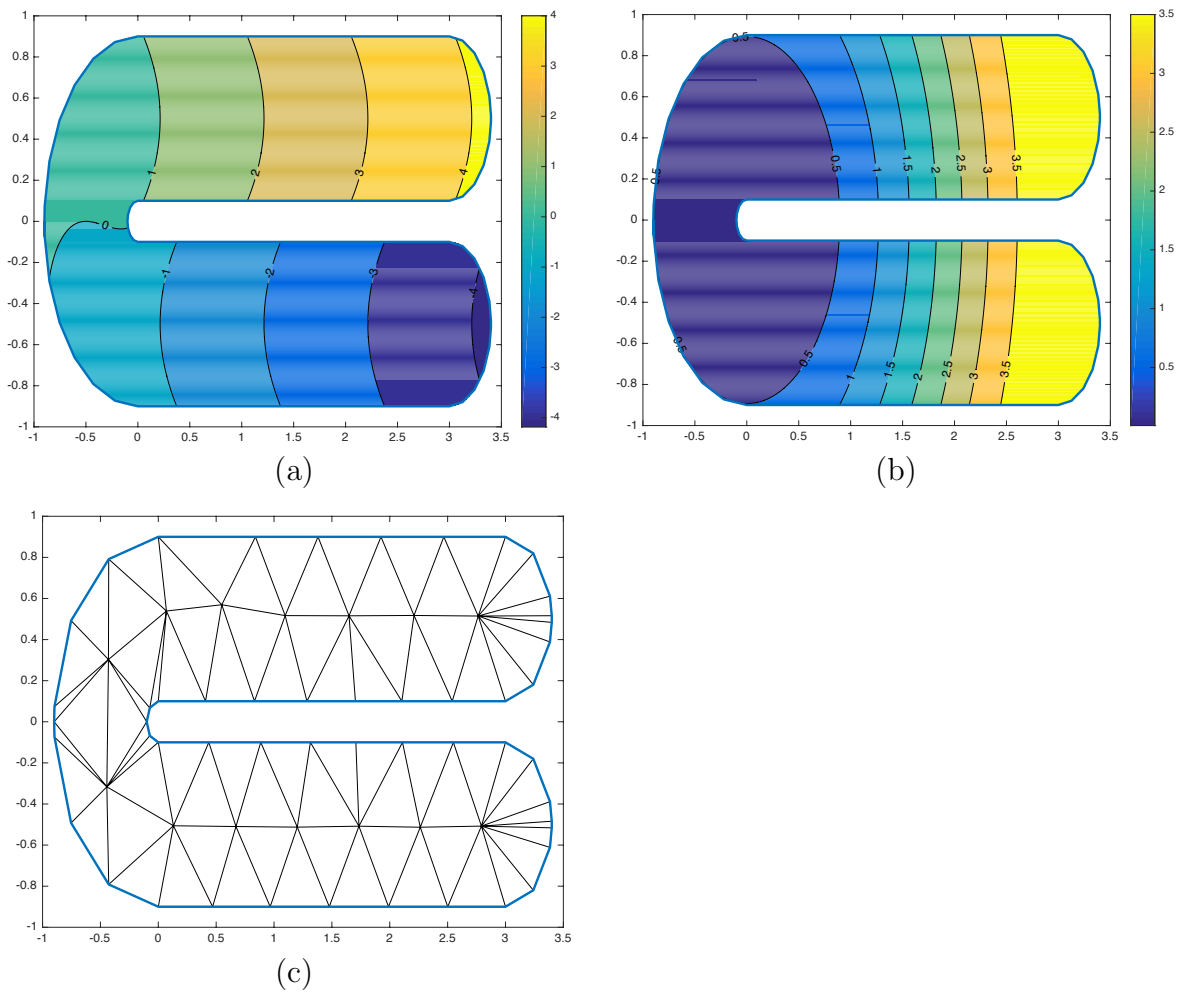


Figure 5. Simulation study 2: plots of (a) true β_0 ; (b) true β_1 ; (c) triangulation.

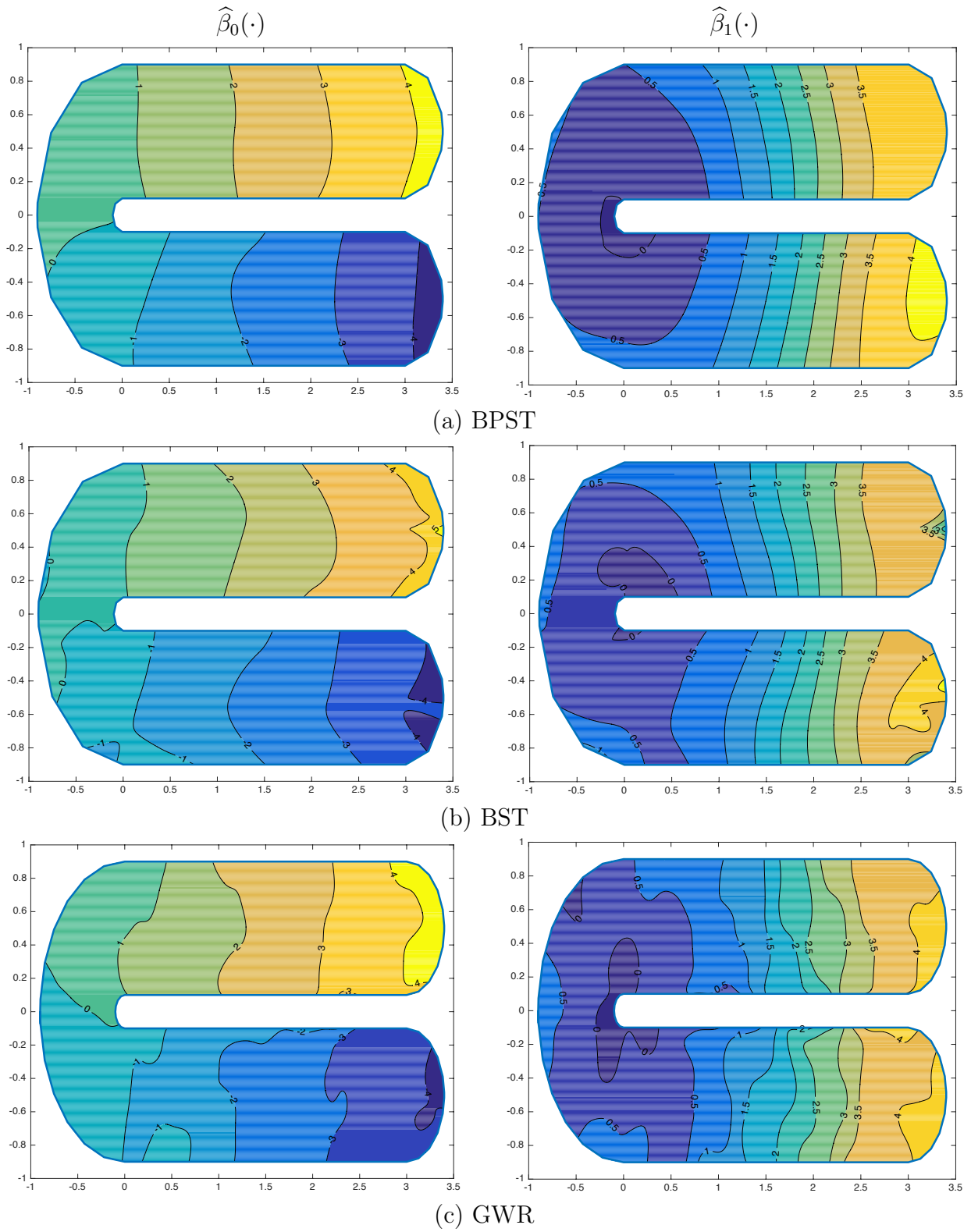


Figure 6. Simulation study 2: estimated surface via (a) BPST; (b) BST; (c) GWR based on sample size $n = 2000$.

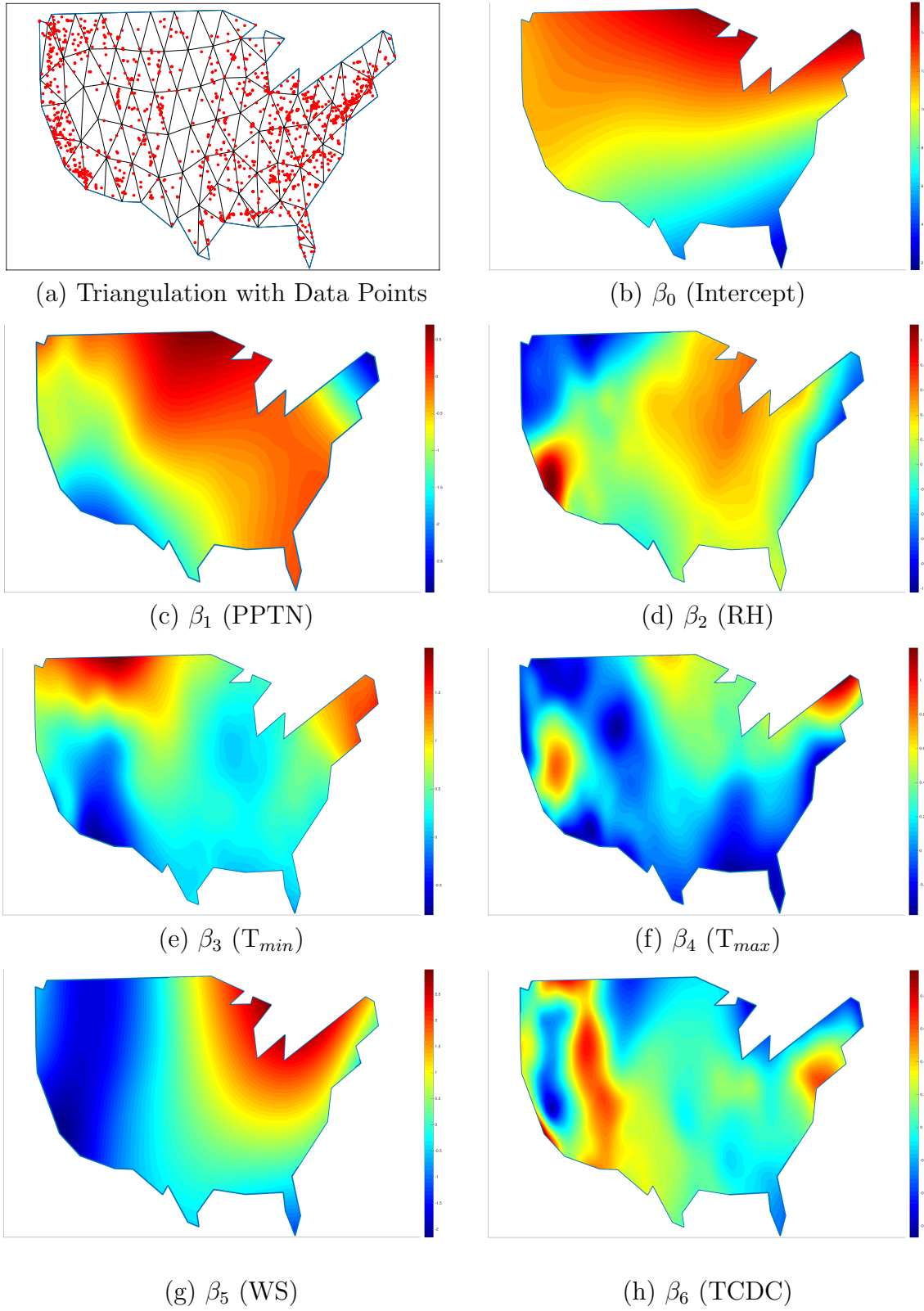


Figure 7. Estimates of the coefficient functions of the SVCM for $PM_{2.5}$ data.

TABLES

Table 1. Estimation and prediction results for Simulation study 1.

n	Method	β_0		β_1		σ^2	Y	Time
		MSE	MSPE	MSE	MSPE	BIAS	MSPE	
500	BST	0.1046	0.1128	0.0840	0.0904	0.0642	1.0345	0.087
	BPST	0.0628	0.0637	0.0548	0.0559	0.0527	1.0165	1.102
	GWR	0.1708	0.1763	0.1379	0.1434	0.1439	1.1063	28.152
1000	BST	0.0479	0.0496	0.0410	0.0421	0.0427	1.0030	0.121
	BPST	0.0340	0.0345	0.0326	0.0331	0.0381	0.9984	2.138
	GWR	0.1127	0.1144	0.0889	0.0907	0.1097	1.0505	102.666
2000	BST	0.0241	0.0244	0.0219	0.0222	0.0300	0.9897	0.186
	BPST	0.0204	0.0205	0.0205	0.0207	0.0266	0.9888	4.902
	GWR	0.0726	0.0733	0.0564	0.0571	0.0798	1.0122	441.230

Table 2. Estimation and prediction results for Simulation study 2.

n	Method	β_0		β_1		σ^2	Y	Time
		MSE	MSPE	MSE	MSPE	BIAS	MSPE	
2000	BST	0.0312	0.0348	0.0237	0.0269	0.0129	0.2647	12.376 [#]
	BPST	0.0130	0.0132	0.0080	0.0081	0.0060	0.2588	29.589 [#]
	GWR	0.0325	0.0337	0.0240	0.0248	0.0306	0.2697	18280.320 [*]
5000	BST	0.0117	0.0120	0.0085	0.0088	0.0051	0.2560	14.033 [#]
	BPST	0.0070	0.0070	0.0042	0.0042	0.0028	0.2545	64.016 [#]
	GWR	— ^{**}	— ^{**}	— ^{**}	— ^{**}	— ^{**}	— ^{**}	— ^{**}

[#] The average computational time is measured using personal computer with Intel(R) Core(TM) i5 CPU dual core @ 2.90GHz and 8.00GB RAM.

^{*} The average computational time is measured by cluster using a parallel computing of 24 general-purpose compute nodes with 128GB RAM associated to each node.

—^{**} We don't have results here because the computing time for 500 iterations in total is more than 168 hours even using cluster with 24 cores parallelly.

Table 3. Meteorological parameters.

Variable	Meteorological Parameter
PPTN	Daily total precipitation (mm)
RH	Air relative humidity at 2m (%)
T_{min}	Surface daily minimum air temperature ($^{\circ}C$)
T_{max}	Surface daily maximum air temperature ($^{\circ}C$)
WS	Surface wind speed (m/s)
TCDC	Total column cloud cover (%)

Table 4. Estimation and prediction accuracy for air pollution data.

	OLS	GWR	BPST
MSE	18.68	8.44	7.03
MSPE	18.95	13.20	12.30

Table 5. Hypothesis tests with their p -values of the tests for PM_{2.5} data.

Null hypothesis	Corresponding variables	p -Value
$\beta_0(\mathbf{u}) = \beta_0$	Intercept	< 0.001
$\beta_1(\mathbf{u}) = \beta_1$	PPTN	0.406
$\beta_2(\mathbf{u}) = \beta_2$	RH	0.688
$\beta_3(\mathbf{u}) = \beta_3$	T _{min}	0.430
$\beta_4(\mathbf{u}) = \beta_4$	T _{max}	0.020
$\beta_5(\mathbf{u}) = \beta_5$	WS	< 0.001
$\beta_6(\mathbf{u}) = \beta_6$	TCDC	< 0.001
$\beta_k(\mathbf{u}) = \beta_k, k = 0, 1, \dots, 6$		< 0.001