Statistica Sinica **31** (2021), 1463-1487
doi:https://doi.org/10.5705/ss.202019.0188

MULTIVARIATE SPLINE ESTIMATION AND INFERENCE FOR IMAGE-ON-SCALAR REGRESSION

Shan Yu, Guannan Wang, Li Wang and Lijian Yang

University of Virginia, College of William & Mary, Iowa State University and Tsinghua University

Abstract: Motivated by recent analyses of data in biomedical imaging studies, we consider a class of image-on-scalar regression models for imaging responses and scalar predictors. We propose using flexible multivariate splines over triangulations to handle the irregular domain of the objects of interest on the images, as well as other characteristics of images. The proposed estimators of the coefficient functions are proved to be root-n consistent and asymptotically normal under some regularity conditions. We also provide a consistent and computationally efficient estimator of the covariance function. Asymptotic pointwise confidence intervals and data-driven simultaneous confidence corridors for the coefficient functions are constructed. Our method can simultaneously estimate and make inferences on the coefficient functions, while incorporating spatial heterogeneity and spatial correlation. A highly efficient and scalable estimation algorithm is developed. Monte Carlo simulation studies are conducted to examine the finite-sample performance of the proposed method, which is then applied to the spatially normalized positron emission tomography data of the Alzheimer's Disease Neuroimaging Initiative.

Key words and phrases: Coefficient maps, confidence corridors, image analysis, multivariate splines, triangulation.

1. Introduction

Medical and public health studies collect massive amount of imaging data using methods such as functional magnetic resonance imaging (fMRI), positron emission tomography (PET) imaging, computed tomography (CT), and ultrasonic imaging. Much of these data can be characterized as functional data. Compared with traditional one-dimensional (1D) functional data, these imaging data are complex, high-dimensional, and structured, which poses challenges to traditional statistical methods.

We propose a unifying approach to characterize the varying associations between imaging responses and a set of explanatory variables. Three types of sta-

Corresponding author: Li Wang, Department of Statistics and the Statistical Laboratory, Iowa State University, Ames, IA 50011, USA. E-mail: lilywang@iastate.edu.

tistical methods are widely used to investigate such associations. The first category includes the univariate approaches and pixel-/voxel-based methods (Worsley et al. (2004); Stein et al. (2010); Hibar et al. (2015)), which take each pixel/voxel as a basic analytic unit. Because all pixels/voxels are treated as independent, a major drawback of these methods is that they ignore correlation between the pixels/voxels. The second category is the tensor regression. This approach considers an image as a multi-dimensional array (Zhou, Li and Zhu (2013); Li and Zhang (2017)), which is then changed to a vector to perform the regression. However, doing so naively yields an ultra-high dimensionality and requires a novel dimension-reduction technique and highly scalable algorithms (Li and Zhang (2017)). The third category is the functional data analysis (FDA) approach, in which an image is viewed as the realization of a function defined on a given domain (Zhu, Li and Kong (2012); Zhu, Fan and Kong (2014); Reiss et al. (2017)). Using an FDA, we are able to combine information both across and within functions.

We adopt the FDA approach in this study. Functional linear models (FLMs) are widely used to model the regression relationship between a response and some set of predictors from multiple subjects. In the literature (Ramsay and Silverman (2005); Müller (2005); Morris (2015); Wang, Chiou and Muller (2016)), FLMs are often categorized based on whether the outcome, the predictor, or both are functional: (i) functional predictor regression (scalar-on-function) (Cardot, Ferraty and Sarda (1999, 2003); Hall and Horowitz (2007)); (ii) functional response regression (function-on-scalar) (Morris and Carroll (2006); Reiss, Huang and Mennes (2010); Staicu, Crainiceanu and Carroll (2010); Zhu, Fan and Kong (2014); Zhang and Wang (2015); Chen, Delicado and Müller (2017)); and (iii) function-on-function regression (Ramsay and Dalzell (1991); Yao, Müller and Wang (2005); Sentürk and Müller (2010); Wu and Müller (2011)).

Motivated by the structure of brain imaging data, we propose a novel image-on-scalar regression model with spatially varying coefficients that captures the varying associations between imaging phenotypes and a set of explanatory variables. Figure 1 shows a schematic diagram of the proposed modeling approach. Specifically, let Ω be a two-dimensional bounded domain, and let $\boldsymbol{z} = (z_1, z_2)$ be the location point on Ω. For the ith subject, $i = 1, \ldots, n$, let $Y_i(\boldsymbol{z})$ be the imaging measurement at location $\boldsymbol{z} \in \Omega$, and let $X_{i\ell}$, for $\ell = 0, 1, \ldots, p$, with $X_{i0} \equiv 1$, be scalar predictors, for example, clinic variables (such as age and sex) and genetic factors. The spatially varying coefficient regression characterizes the associations between imaging measures and covariates, and is given by the following model:

$$Y_i(\boldsymbol{z}) = \widetilde{\mathbf{X}}_i^\top \boldsymbol{\beta}^o(\boldsymbol{z}) + \eta_i(\boldsymbol{z}) + \sigma(\boldsymbol{z})\varepsilon_i(\boldsymbol{z}),\ i = 1, \ldots, n,\ \boldsymbol{z} \in \Omega,$$

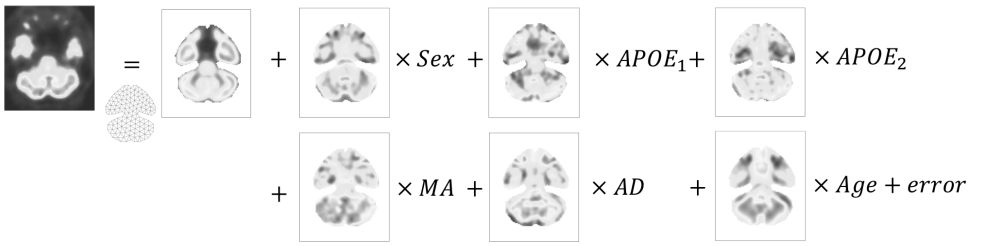

Figure 1. A schematic diagram of proposed modeling approach.

where $\widetilde{\mathbf{X}}_i = (X_{i0}, X_{i1}, \ldots, X_{ip})^\top$, $\boldsymbol{\beta}^o = (\beta_0^o, \beta_1^o, \ldots, \beta_p^o)^\top$ is a vector of some unknown bivariate functions, $\eta_i(\boldsymbol{z})$ characterizes the individual image variations, $\varepsilon_i(\boldsymbol{z})$ represents additional measurement errors, and $\sigma(\boldsymbol{z})$ is a positive deterministic function. In the following, we assume that $\eta_i(\boldsymbol{z})$ and $\varepsilon_i(\boldsymbol{z})$ are mutually independent. Moreover, we assume that $\eta_i(\boldsymbol{z})$, for $i = 1, \ldots, n$, are independent and identically distributed (i.i.d.) copies of an L_2 stochastic process with mean zero and covariance function $G_\eta(\boldsymbol{z}, \boldsymbol{z}') = \mathrm{cov}\{\eta_i(\boldsymbol{z}), \eta_i(\boldsymbol{z}')\}$. Furthermore, $\varepsilon_i(\boldsymbol{z})$, for $i = 1, \ldots, n$, are i.i.d. copies of a stochastic process with zero mean. and covariance function $G_\varepsilon(\boldsymbol{z}, \boldsymbol{z}') = \mathrm{cov}\{\varepsilon_i(\boldsymbol{z}), \varepsilon_i(\boldsymbol{z}')\} = I(\boldsymbol{z} = \boldsymbol{z}')$.

For a 1D function-on-scalar regression, Chapter 13 of Ramsay and Silverman (2005) provides a common model-fitting strategy, in which the coefficient functions are expanded using some sets of basis functions, and the basis coefficients are estimated using the ordinary least squares method. However, it is not trivial to extend this to an image-on-scalar regression, particularly with biomedical imaging responses. For biomedical images, the objects (e.g., organs) on the images are usually irregularly shaped (e.g., breast tumors). Another example is that of brain images, as shown in Figure 1, especially slices from the bottom and the top of the brain. Even though some images seem to be rectangular, the true signal comes only from the domain of an object, and the image contains only noise outside the boundary of the object. Many smoothing methods, such as, tensor product smoothing (Reiss et al. (2017); Chen, Delicado and Müller (2017)), kernel smoothing (Zhu, Fan and Kong (2014)), and wavelet smoothing (Morris and Carroll (2006)), provide poor estimations over difficult regions because they smooth inappropriately across boundary features, referred to as the "leakage" problem in the smoothing literature; see Ramsay (2002) and Sangalli, Ramsay and Ramsay (2013). Next, for technical reasons, imaging data often have different visual qualities. The general characteristics of medical images are determined and limited by the technology for each specific modality. As a result, there is a great interest in developing a flexible method with varying smoothness

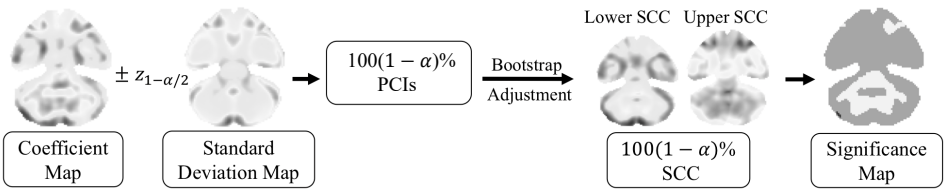

Figure 2. A schematic diagram of proposed inferential approach.

to adaptively smooth biomedical imaging data.

In this study, we tackle the above challenges using bivariate splines on triangulations (Lai and Wang (2013)) to effectively model the spatially nonstationary relationship and preserve the important features (shape, smoothness) of the imaging data. A triangulation can represent any two-dimensional (2D) geometric domain effectively because any polygon can be decomposed into triangles. We study the asymptotic properties of the bivariate spline estimators of the coefficient functions, and show that our spline estimators are root-n consistent and asymptotically normal. The asymptotic results are used as a guideline to construct pointwise confidence intervals (PCIs) and simultaneous confidence corridors (SCCs; also referred to as "simultaneous confidence bands/regions") for the true coefficient functions. Figure 2 shows the proposed inferential approach. Our method is statistically more efficient than the tensor regression (Li and Zhang (2017)) and the three-stage estimation (Zhu, Fan and Kong (2014)), because it is able to accommodate complex domains of arbitrary shape and adjust the individual smoothing needs of different coefficient functions using multiple smoothing parameters. In addition, our method does not rely on estimating the spatial similarity and adaptive weights repeatedly, as in Zhu, Fan and Kong (2014); thus, it is much simpler.

The remainder of the paper is structured as follows. Section 2 describes the spline estimators for the coefficient functions, and establishes their asymptotic properties. Section 3 describes the bootstrap method used to construct the SCC and how to estimate the unknown variance functions involved in the SCC. Section 4 presents the implementation of the proposed estimation and inference. Section 5 reports our findings from two simulation studies. In Section 6, we illustrate the proposed method using PET data provided by the Alzheimer's Disease Neuroimaging Initiative (ADNI). Section 7 concludes the paper. All technical proofs of the theoretical results and additional numerical results are deferred to the online Supplementary Material.

2. Models and Estimation Method

2.1. Image-on-scalar regression model

Let $\boldsymbol{z}_j \in \Omega$ be the center point of the jth pixel in the domain Ω, and let Y_{ij} be the imaging response of subject i at location j. The actual data set consists of $\{(Y_{ij}, \widetilde{\mathbf{X}}_i, \boldsymbol{z}_j), i = 1, \ldots, n, j = 1, \ldots, N\}$, which can be modeled as follows:

$$Y_{ij} = \sum_{\ell=0}^{p} X_{i\ell}\beta_\ell^o(\boldsymbol{z}_j) + \eta_i(\boldsymbol{z}_j) + \sigma(\boldsymbol{z}_j)\varepsilon_{ij}. \tag{2.1}$$

Denote the eigenvalues and eigenfunctions of the covariance operator $G_\eta(\boldsymbol{z}, \boldsymbol{z}')$ as $\{\lambda_k\}_{k=1}^{\infty}$ and $\{\psi_k(\boldsymbol{z})\}_{k=1}^{\infty}$, respectively, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, $\sum_{k=1}^{\infty}\lambda_k < \infty$, and $\{\psi_k\}_{k=1}^{\infty}$ forms an orthonormal basis of $L^2(\Omega)$. It follows from spectral theory that $G_\eta(\boldsymbol{z}, \boldsymbol{z}') = \sum_{k=1}^{\infty}\lambda_k\psi_k(\boldsymbol{z})\psi_k(\boldsymbol{z}')$. The ith trajectory $\{\eta_i(\boldsymbol{z}), \boldsymbol{z} \in \Omega\}$ allows the Karhunen–Loéve L^2 representation (Li and Hsing (2010); Sang and Huang (2012)): $\eta_i(\boldsymbol{z}) = \sum_{k=1}^{\infty}\lambda_k^{1/2}\xi_{ik}\psi_k(\boldsymbol{z})$, $\lambda_k^{1/2}\xi_{ik} = \int_{\boldsymbol{z}\in\Omega}\eta_i(\boldsymbol{z})\psi_k(\boldsymbol{z})d\boldsymbol{z}$, where the random coefficients ξ_{ik} are uncorrelated random variables with mean zero and $E(\xi_{ik}\xi_{ik'}) = I(k = k')$, referred to as the kth functional principal component score (FPCA) of the ith subject. Thus, the response measurements in (2.1) can be represented as follows:

$$Y_{ij} = \sum_{\ell=0}^{p}\beta_\ell^o(\boldsymbol{z}_j)X_{i\ell} + \sum_{k=1}^{\infty}\lambda_k^{1/2}\xi_{ik}\psi_k(\boldsymbol{z}_j) + \sigma(\boldsymbol{z}_j)\varepsilon_{ij}. \tag{2.2}$$

2.2. Spline approximation over triangulations and penalized regression

Note that the objects of interest on many biomedical images are often distributed over an irregular domain Ω. Triangulation is an effective strategy to handle such data. For example, the spatial smoothing problem over difficult regions in Ramsay (2002) and Sangalli, Ramsay and Ramsay (2013) was solved using the finite element method (FEM) on triangulations, which was developed primarily to solve partial differential equations. Here, we approximate each coefficient function in (2.2) using bivariate splines over triangulations (Lai and Schumaker (2007)). The idea is to approximate each function $\beta_\ell(\cdot)$ using Bernstein basis polynomials that are piecewise polynomial functions over a 2D triangulated domain. Compared with the FEM, the proposed approach is appealing in the sense that its spline functions are more flexible and it uses various smoothness settings to better approximate the coefficient functions. In this section, we briefly introduce the triangulation technique and describe the bivariate penalized spline

smoothing (BPST) method used to approximate the spatial data.

Triangulation is an effective tool to deal with data distributed over difficult regions with complex boundaries and/or interior holes. In the following, we use T to denote a triangle that is a convex hull of three points not located on one line. A collection $\triangle = \{T_1, \ldots, T_H\}$ of H triangles is called a triangulation of $\Omega = \cup_{h=1}^{H} T_h$, provided that any nonempty intersection between a pair of triangles in $\triangle$ is either a shared vertex or a shared edge. Given a triangle $T \in \triangle$, let $|T|$ be its longest edge length and ϱ_T be the radius of the largest disk inscribed in T. Define the shape parameter of T as the ratio $\pi_T = |T|/\varrho_T$. When π_T is small, the triangles are relatively uniform in the sense that all angles of the triangles in $\triangle$ are relatively the same. Denote the size of $\triangle$ by $|\triangle| = \max\{|T|, T \in \triangle\}$, that is, the length of the longest edge of $\triangle$. For an integer $r \geq 0$, let $\mathcal{C}^r(\Omega)$ be the collection of all rth continuously differentiable functions over Ω. Given $\triangle$, let $\mathcal{S}_d^r(\triangle) = \{s \in \mathcal{C}^r(\Omega) : s|_T \in \mathbb{P}_d(T), T \in \triangle\}$ be a spline space of degree d and smoothness r over $\triangle$, where $s|_T$ is the polynomial piece of spline s restricted on triangle T, and $\mathbb{P}_d$ is the space of all polynomials of degree less than or equal to d. Note that the major difference between the FEM and the BPST is the flexibility of the smoothness, r, and the degree of the polynomials, d. Specifically, the FEM in Sangalli, Ramsay and Ramsay (2013) requires that $r = 0$ and $d = 1$ or 2, whereas the BPST allows smoothness $r \geq 0$ and various degrees of polynomials.

We use Bernstein basis polynomials to represent the bivariate splines. For any $\ell = 0, 1, \ldots, p$, denote by $\triangle_\ell$ the triangulation of the ℓth component. Define

$$\mathcal{G}^{(p+1)} \equiv \mathcal{G}^{(p+1)}(\triangle_0 \times \cdots \times \triangle_p) = \left\{ \boldsymbol{g} = (g_0, \ldots, g_p)^\top, g_\ell \in \mathcal{S}_d^r(\triangle_\ell), \ell = 0, \ldots, p \right\},$$

and let $\{B_{\ell m}\}_{m \in \mathcal{M}_\ell}$ be the set of degree-d bivariate Bernstein basis polynomials for $\mathcal{S}_d^r(\triangle_\ell)$, where $\mathcal{M}_\ell$ is an index set of Bernstein basis polynomials. Denote by $\mathbf{B}_\ell$ the evaluation matrix of the Bernstein basis polynomials for the ℓth component, and let the jth row of $\mathbf{B}_\ell$ is given by $\mathbf{B}_\ell^\top(\boldsymbol{z}_j) = \{B_{\ell m}(\boldsymbol{z}_j), m \in \mathcal{M}_\ell\}$. We approximate each $\beta_\ell(\cdot)$ using $\beta_\ell(\boldsymbol{z}_j) \approx \mathbf{B}_\ell^\top(\boldsymbol{z}_j)\boldsymbol{\gamma}_\ell$, for $\ell = 0, 1, \ldots, p$, where $\boldsymbol{\gamma}_\ell^\top = (\gamma_{\ell m}, m \in \mathcal{M}_\ell)$ is the spline coefficient vector.

Penalized spline smoothing has gained in popularity over the last two decades; see Hall and Opsomer (2005); Claeskens, Krivobokova and Opsomer (2009); Schwarz and Krivobokova (2016). To define the penalized spline method, for any direction z_q, $q = 1, 2$, let $\nabla_{z_q}^v s(\boldsymbol{z})$ denote the vth–order derivative in the direction z_q at the point $\boldsymbol{z}$. We consider the following penalized least squares

problem:

$$\min_{(\beta_0,\ldots,\beta_p)^\top\in\mathcal{G}^{(p+1)}} \sum_{i=1}^{n}\sum_{j=1}^{N}\left\{Y_{ij}-\sum_{\ell=0}^{p}X_{i\ell}\beta_\ell(\boldsymbol{z}_j)\right\}^2+\sum_{\ell=0}^{p}\rho_{n,\ell}\mathcal{E}(\beta_\ell),$$

where $\mathcal{E}(s)=\sum_{T\in\triangle}\int_T\sum_{i+j=2}\binom{2}{i}(\nabla^i_{z_1}\nabla^j_{z_2}s)^2dz_1dz_2$ is the roughness penalty, and $\rho_{n,\ell}$ is the penalty parameter for the ℓth function.

To satisfy the smoothness condition of the splines, we need to impose some linear constraints on the spline coefficients $\boldsymbol{\gamma}_\ell$: $\mathbf{H}_\ell\boldsymbol{\gamma}_\ell=\mathbf{0}$, for $\ell=0,1,\ldots,p$. Thus, we have to minimize the following constrained least squares:

$$\sum_{i=1}^{n}\sum_{j=1}^{N}\left\{Y_{ij}-\sum_{\ell=0}^{p}X_{i\ell}\mathbf{B}_\ell^\top(\boldsymbol{z}_j)\boldsymbol{\gamma}_\ell\right\}^2+\sum_{\ell=0}^{p}\rho_{n,\ell}\boldsymbol{\gamma}_\ell^\top\mathbf{P}_\ell\boldsymbol{\gamma}_\ell,\ \text{subject to}\ \mathbf{H}_\ell\boldsymbol{\gamma}_\ell=0,$$

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

We first remove the constraint using a QR decomposition of the transpose of the constraint matrix $\mathbf{H}_\ell$. Applying a QR decomposition on $\mathbf{H}_\ell^\top$, we have $\mathbf{H}_\ell^\top=\mathbf{Q}_\ell\mathbf{R}_\ell=(\mathbf{Q}_{\ell,1}\ \mathbf{Q}_{\ell,2})\binom{\mathbf{R}_{\ell,1}}{\mathbf{R}_{\ell,2}}$, where $\mathbf{Q}_\ell$ is an orthogonal matrix and $\mathbf{R}_\ell$ is an upper triangular matrix. The submatrix $\mathbf{Q}_{\ell,1}$ represents the first r columns of $\mathbf{Q}_\ell$, where r is the rank of matrix $\mathbf{H}_\ell$, and $\mathbf{R}_{\ell,2}$ is a matrix of zeros. We reparametrize this using $\boldsymbol{\gamma}_\ell=\mathbf{Q}_{\ell,2}\boldsymbol{\theta}_\ell$, for some $\boldsymbol{\theta}_\ell$. Then, it is guaranteed that $\mathbf{H}_\ell\boldsymbol{\gamma}_\ell=\mathbf{0}$. Thus, the minimization problem is converted to the following conventional penalized regression problem, without restrictions:

$$\sum_{i=1}^{n}\sum_{j=1}^{N}\left\{Y_{ij}-\sum_{\ell=0}^{p}X_{i\ell}\mathbf{B}_\ell^\top(\boldsymbol{z}_j)\mathbf{Q}_{\ell,2}\boldsymbol{\theta}_\ell\right\}^2+\sum_{\ell=0}^{p}\rho_{n,\ell}\boldsymbol{\theta}_\ell^\top\mathbf{D}_\ell\boldsymbol{\theta}_\ell, \tag{2.3}$$

where $\mathbf{D}_\ell=\mathbf{Q}_{\ell,2}^\top\mathbf{P}_\ell\mathbf{Q}_{\ell,2}$.

Let $\widetilde{\mathbf{Y}}_i=(Y_{i1},Y_{i2},\ldots,Y_{iN})^\top$, $\mathbf{B}_\ell(\boldsymbol{z})=\{B_{\ell m}(\boldsymbol{z}),m\in\mathcal{M}_\ell\}^\top$, $\mathbb{Y}=(\widetilde{\mathbf{Y}}_1^\top,\ldots,\widetilde{\mathbf{Y}}_n^\top)^\top$, and $\mathbb{U}=(\mathbf{U}_{11},\mathbf{U}_{12},\ldots,\mathbf{U}_{nN})^\top$, where

$$\mathbf{U}_{ij}=\{X_{i0}\mathbf{B}_0(\boldsymbol{z}_j)^\top\mathbf{Q}_{0,2},X_{i1}\mathbf{B}_1(\boldsymbol{z}_j)^\top\mathbf{Q}_{1,2},\ldots,X_{ip}\mathbf{B}_p(\boldsymbol{z}_j)^\top\mathbf{Q}_{p,2}\}^\top. \tag{2.4}$$

Let $\boldsymbol{\theta}=(\boldsymbol{\theta}_0^\top,\boldsymbol{\theta}_1^\top,\ldots,\boldsymbol{\theta}_p^\top)^\top$ and $\mathbb{D}(\rho_{n,0},\ldots,\rho_{n,p})=\operatorname{diag}\{\rho_{n,0}\mathbf{D}_0,\ldots,\rho_{n,p}\mathbf{D}_p\}$. Minimizing (2.3) is then equivalent to minimizing $\|\mathbb{Y}-\mathbb{U}\boldsymbol{\theta}\|^2+\boldsymbol{\theta}^\top\mathbb{D}(\rho_{n,0},\ldots,\rho_{n,p})\boldsymbol{\theta}$. Hence,

$$\widehat{\boldsymbol{\theta}}=(\widehat{\boldsymbol{\theta}}_0^\top,\widehat{\boldsymbol{\theta}}_1^\top,\ldots,\widehat{\boldsymbol{\theta}}_p^\top)^\top=\{\mathbb{U}^\top\mathbb{U}+\mathbb{D}(\rho_{n,0},\ldots,\rho_{n,p})\}^{-1}\mathbb{U}^\top\mathbb{Y}.$$

Thus, the estimators of $\boldsymbol{\gamma}_\ell$ and $\beta_\ell(\cdot)$ are

$$\widehat{\boldsymbol{\gamma}}_\ell = \mathbf{Q}_{\ell,2}\widehat{\boldsymbol{\theta}}_\ell, \quad \widehat{\beta}_\ell(\boldsymbol{z}) = \mathbf{B}_\ell(\boldsymbol{z})^\top \widehat{\boldsymbol{\gamma}}_\ell. \tag{2.5}$$

2.3. Asymptotic properties of the BPST estimators

This section examines the asymptotics of the proposed estimators. Given random variables U_n for $n \geq 1$, we write $U_n = O_P(b_n)$ if $\lim_{c\to\infty} \limsup_n P(|U_n| \geq cb_n) = 0$. Similarly, we write $U_n = o_P(b_n)$ if $\lim_n P(|U_n| \geq cb_n) = 0$, for any constant $c > 0$. Next, to facilitate discussion, we introduce some notation of norms. For any function g over the closure of domain Ω, denote $\|g\|^2_{L^2(\Omega)} = \int_\Omega g^2(\boldsymbol{z})d\boldsymbol{z}$ as the regular L_2 norm of g, and $\|g\|_{\infty,\Omega} = \sup_{\boldsymbol{z}\in\Omega} |g(\boldsymbol{z})|$ as the supremum norm of g. Further denote $\|\boldsymbol{g}\|_{\upsilon,\infty,\Omega} = \max_{0\leq\ell\leq p} |g_\ell|_{\upsilon,\infty,\Omega}$, where $|g|_{\upsilon,\infty,\Omega} = \max_{i+j=\upsilon} \|\nabla^i_{z_1}\nabla^j_{z_2} g\|_{\infty,\Omega}$ is the maximum norm of all υth–order derivatives of g over Ω. Let $\mathcal{W}^{d,\infty}(\Omega) = \{g : |g|_{k,\infty,\Omega} < \infty, 0 \leq k \leq d\}$ be the standard Sobolev space. Next, we introduce some technical conditions.

(A1) For any $\ell = 0, \ldots, p$, $\beta^o_\ell(\cdot) \in \mathcal{W}^{d+1,\infty}(\Omega)$, for an integer $d \geq 1$.

(A2) For any $i = 1, \ldots, n$, $j = 1, \ldots, N$, ε_{ij}'s are independent with mean zero and variance one, and for any $k \geq 1$, ξ_{ik} are uncorrelated random variables with mean zero and variance one.

(A3) For any $\ell = 0, 1, \ldots, p$, there exists a positive constant C_ℓ, such that $E|X_\ell|^8 \leq C_\ell$. The eigenvalues of $\boldsymbol{\Sigma}_X = E(\boldsymbol{X}\boldsymbol{X}^\top)$ are bounded away from zero and infinity.

(A4) The function $\sigma(\boldsymbol{z}) \in \mathcal{C}^{(1)}(\Omega)$, with $0 < c_\sigma \leq \sigma(\boldsymbol{z}) \leq C_\sigma \leq \infty$, for any $\boldsymbol{z} \in \Omega$; for any k, $\psi_k(\boldsymbol{z}) \in \mathcal{C}^{(1)}(\Omega)$ and $0 < c_G \leq G_\eta(\boldsymbol{z}, \boldsymbol{z}) \leq C_G \leq \infty$, for any $\boldsymbol{z} \in \Omega$.

(A5) Let $|\underline{\triangle}| = \min_{0\leq\ell\leq p} |\triangle_\ell|$ and $|\overline{\triangle}| = \max_{0\leq\ell\leq p} |\triangle_\ell|$. The triangulations $\triangle_\ell$ satisify that $\limsup_n(|\overline{\triangle}|/|\underline{\triangle}|) < \infty$. The triangulations are π-quasi-uniform; that is, there exists a positive constant π, such that $\max_{0\leq\ell\leq p}\{(\min_{T\in\triangle_\ell} \varrho_T)^{-1}|\triangle_\ell|\} \leq \pi$.

(A6) As $N \to \infty$, $n \to \infty$, for some $0 < \kappa < 1$, $N^{-1}n^{1/(d+1)+\kappa} \to 0$, $n^{1/2}|\overline{\triangle}|^{d+1} \to 0$, $N^{1/2}|\underline{\triangle}| \to \infty$, and the smoothing parameters satisfy that $n^{-1/2}N^{-1}|\underline{\triangle}|^{-3}$ $\rho_n \to 0$, where $\rho_n = \max_{0\leq\ell\leq p} \rho_{n,\ell}$.

The above assumptions are mild conditions that are satisfied in many practical situations. Assumption (A1) describes the usual requirement on the coefficient

functions described in the literature on nonparametric estimation. Assumption (A1) can be relaxed to Assumption (A1′) in Section 2.4, which only requires $\beta_\ell^o(\cdot) \in \mathcal{C}^{(0)}(\Omega)$ when dealing with imaging data with sharp edges; see Section 2.4. Assumptions (A1) and (A2) are similar to Assumptions (A1) and (A2) in Gu et al. (2014) and Assumptions (A1)–(A3) in Huang, Wu and Zhou (2004). Assumption (A3) is analogous to Assumption (A5) in Gu et al. (2014), ensuring that $X_{i\ell}$ is not multicollinear. Assumption (A5) requires that $\triangle_\ell$ be of similar size, and suggests the use of more uniform triangulations with smaller shape parameters. Assumption (A6) implies that the number of pixels for each image N diverges to infinity and the sample size n grows as $N \to \infty$, a well-developed asymptotic scenario for dense functional data (Li and Hsing (2010)). Assumption (A6) also describes the requirement of the growth rate of the dimension of the spline spaces relative to the sample size and the image resolution. This assumption is easily satisfied because images measured using current technology are usually of sufficiently high resolution.

The following theorem provides the L_2 convergence rate of $\widehat{\beta}_\ell(\cdot)$, for $\ell = 0, 1, \ldots, p$. A detailed proof is given in Appendix 1 in the Supplementary Material.

Theorem 1. *Suppose Assumptions* (A1)–(A5) *hold and* $N^{1/2}|\underline{\triangle}| \to \infty$ *as* $N \to \infty$. *Then, for any* $\ell = 0, 1, \ldots, p$, *the BPST estimator* $\widehat{\beta}_\ell(\cdot)$ *is consistent and satisfies* $\|\widehat{\beta}_\ell - \beta_\ell^o\|_{L^2(\Omega)} = O_P\{\rho_n/(nN|\underline{\triangle}|^3)\|\boldsymbol{\beta}^o\|_{2,\infty} + (1 + \rho_n/(nN|\underline{\triangle}|^5))|\overline{\triangle}|^{d+1} \|\boldsymbol{\beta}^o\|_{d+1,\infty} + n^{-1/2}\}$.

Theorem 2 states the asymptotic normality of $\widehat{\beta}_\ell$ at any given point $\boldsymbol{z} \in \Omega$, for $\ell = 0, 1, \ldots, p$. See Appendix 1 in the Supplementary Material for a detailed proof. Denote

$$\boldsymbol{\Xi}_n(\boldsymbol{z}) = \widetilde{\mathbb{B}}(\boldsymbol{z})^\top E\left\{\boldsymbol{\Gamma}_{n,\rho}^{-1}\frac{1}{n^2N^2}\sum_{i=1}^{n}\sum_{j,j'=1}^{N}\mathbf{U}_{ij}\mathbf{U}_{ij'}^\top G_\eta(\boldsymbol{z}_j, \boldsymbol{z}_{j'})\boldsymbol{\Gamma}_{n,\rho}^{-1}\right\}\widetilde{\mathbb{B}}(\boldsymbol{z}), \quad (2.6)$$

where $\mathbf{U}_{ij}$ and $\boldsymbol{\Gamma}_{n,\rho}$ are given in (2.4) and (S1.17), respectively, in Appendix 1, $\widetilde{\mathbb{B}}(\boldsymbol{z}) = \mathrm{diag}\{\widetilde{\mathbf{B}}_0(\boldsymbol{z}), \ldots, \widetilde{\mathbf{B}}_p(\boldsymbol{z})\}$, and $\widetilde{\mathbf{B}}_\ell(\boldsymbol{z}) = \mathbf{Q}_{2,\ell}^\top\mathbf{B}_\ell(\boldsymbol{z})$ for $\ell = 0, \ldots, p$.

Theorem 2. *Suppose Assumptions* (A1)–(A6) *hold. If for any* $\ell = 0, 1, \ldots, p$, $|X_{i\ell}| \le C_\ell < \infty$, *then* $\boldsymbol{\Xi}_n^{-1/2}(\boldsymbol{z})\{\widehat{\boldsymbol{\beta}}(\boldsymbol{z}) - \boldsymbol{\beta}^o(\boldsymbol{z})\} \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \mathbf{I}_{(p+1)\times(p+1)}\right)$ *as* $N \to \infty$ *and* $n \to \infty$, *where* $\boldsymbol{\Xi}_n(\boldsymbol{z})$ *is given in* (2.6). *Furthermore, there exist positive constants* $c_V < C_V < +\infty$, *such that* $c_V n^{-1}(1+\rho_n/(nN|\underline{\triangle}|^4))^{-2} \le \mathrm{Var}\{\widehat{\beta}_\ell(\boldsymbol{z})\} \le C_V n^{-1}$, *for any* $\ell = 0, 1, \ldots, p$.

2.4. Piecewise constant spline over triangulation smoothing

Many imaging data can be regarded as a noisy version of a piecewise-smooth function of $\boldsymbol{z} \in \Omega$ with sharp edges, which often reflect the functional or structural changes. The penalized bivariate spline smoothing method introduced, in Section 2.2, assumes some degree of smoothness over the entire image. To relax this assumption while preserving the features of sharp edges, we make the following less stringent assumption on the smoothness of the coefficient functions:

(A1′) For any $\ell = 0, \ldots, p$, the bivariate function $\beta_\ell^o(\cdot) \in \mathcal{C}^{(0)}(\Omega)$.

For the estimation, we consider the piecewise constant spline over triangulation (PCST) method. For any $\ell = 1, \ldots, p$, denote by $\mathcal{PC}(\triangle_\ell)$ the space of piecewise constant functions over each T_m, for $m \in \mathcal{M}_\ell$. The bivariate spline basis functions of $\mathcal{PC}(\triangle_\ell)$ are denoted as $\{B_{\ell m}(\boldsymbol{z})\}_{m \in \mathcal{M}_\ell}$, which are simply indicator functions over triangle T_m, $B_{\ell m}(\boldsymbol{z}) = I(\boldsymbol{z} \in T_m)$, $m \in \mathcal{M}_\ell$. Assumption (A1′) controls the bias of the piecewise constant spline estimator for β_ℓ^o and leads to the estimation consistency.

When using the constant bivariate spline basis functions, we have $\mathcal{E}(s) = 0$ for all $s \in \mathcal{PC}(\triangle)$, and for any $\boldsymbol{z} \in \Omega$, $\mathbf{B}_\ell(\boldsymbol{z})\mathbf{B}_\ell(\boldsymbol{z})^\top = \mathrm{diag}\{B_{\ell m}^2(\boldsymbol{z}), m \in \mathcal{M}_\ell\}$. Then, $\widehat{\boldsymbol{\gamma}}_m = (\widehat{\gamma}_{0m}, \widehat{\gamma}_{1m}, \ldots, \widehat{\gamma}_{pm})^\top = \widehat{\mathbf{V}}_m^{-1} \left\{ (nN)^{-1} \sum_{i=1}^n \sum_{j=1}^N B_{\ell m}(\boldsymbol{z}_j) X_{i\ell} Y_{ij} \right\}_{\ell=0}^p$, where

$$\widehat{\mathbf{V}}_m = \frac{1}{nN} \sum_{j=1}^N B_{\ell m}^2(\boldsymbol{z}_j) \sum_{i=1}^n \widetilde{\mathbf{X}}_i \widetilde{\mathbf{X}}_i^\top = \left\{ \frac{1}{nN} \sum_{i=1}^n \sum_{j=1}^N B_{\ell m}^2(\boldsymbol{z}_j) X_{i\ell} X_{i\ell'} \right\}_{\ell,\ell'=0}^p . \quad (2.7)$$

By simple linear algebra, for any $\ell = 0, \ldots, p$, the PCST estimator is given by

$$\widehat{\beta}_\ell^{\mathrm{c}}(\boldsymbol{z}) = \sum_{m \in \mathcal{M}_\ell} \widehat{\gamma}_{\ell m} B_{\ell m}(\boldsymbol{z}). \quad (2.8)$$

For any $\boldsymbol{z} \in \Omega$, define the index of the triangle containing $\boldsymbol{z}$ as $m(\boldsymbol{z})$; that is, $m(\boldsymbol{z}) = m$ if $\boldsymbol{z} \in T_m$. Then, $\widehat{\beta}_\ell(\boldsymbol{z}) = \widehat{\boldsymbol{\gamma}}_{\ell m(\boldsymbol{z})}$ and $\widehat{\boldsymbol{\beta}}^{\mathrm{c}}(\boldsymbol{z}) = (\widehat{\beta}_0^{\mathrm{c}}(\boldsymbol{z}), \ldots, \widehat{\beta}_p^{\mathrm{c}}(\boldsymbol{z}))^\top = (\widehat{\gamma}_{0m(\boldsymbol{z})}, \ldots, \widehat{\gamma}_{pm(\boldsymbol{z})})^\top = \widehat{\boldsymbol{\gamma}}_{m(\boldsymbol{z})}$. For any $\boldsymbol{z} \in \Omega$, denote

$$\boldsymbol{\Sigma}_n(\boldsymbol{z}) = n^{-1} \boldsymbol{\Sigma}_X^{-1} G_\eta(\boldsymbol{z}, \boldsymbol{z}). \quad (2.9)$$

Theorem 3 shows the asymptotic normality of the piecewise constant estimators $\widehat{\boldsymbol{\beta}}(\boldsymbol{z})$. See the Supplementary Material for detailed proofs. To obtain the asymptotic variance-covariance function, we also need the following assumption:

(C1) The variables ξ_{ik} and ε_{ij} are independent and satisfy $E|\xi_{ik}|^{4+\delta_1} < +\infty$ for

some $\delta_1 > 0$, and $E\,|\varepsilon_{ij}|^{4+\delta_2} < \infty$ for some $\delta_2 > 0$.

Theorem 3. *Under Assumptions* (A1′), (A2)–(A5), *and* (C1), *as* $N \to \infty$ *and* $n \to \infty$, *if for some* $0 < \kappa < 1$, $N^{-1}n^{1+\kappa} \to 0$, $N^{-1/2} \ll |\underline{\triangle}| \le |\overline{\triangle}| \ll n^{1/4}N^{-1/2}$, *and* $\|\sum_{k=1}^{\infty}\lambda_k^{1/2}\psi_k\|_\infty < \infty$, *then for any* $\boldsymbol{z} \in \Omega$, $\boldsymbol{\Sigma}_n^{-1/2}(\boldsymbol{z})\{\widehat{\boldsymbol{\beta}}^{\mathrm{c}}(\boldsymbol{z}) - \boldsymbol{\beta}^{o}(\boldsymbol{z})\} \xrightarrow{\mathcal{L}} N\left(\mathbf{0}, \mathbf{I}_{(p+1)\times(p+1)}\right)$, *where* $\boldsymbol{\Sigma}_n(\boldsymbol{z})$ *is* (2.9); $\mathrm{pr}\{(\sigma^{\mathrm{c}}_{n,\ell\ell})^{-1}(\boldsymbol{z})\big|\widehat{\beta}_\ell(\boldsymbol{z}) - \beta^o_\ell(\boldsymbol{z})\big| \le Z_{1-\alpha/2}\} \to 1-\alpha$, *for any* $\alpha \in (0,1)$, *as* $N \to \infty$, $n \to \infty$, *where* $\sigma^{\mathrm{c}}_{n,\ell\ell}(\boldsymbol{z})$ *is the square root of the* (ℓ,ℓ)*th entry of the matrix* $\boldsymbol{\Sigma}_n(\boldsymbol{z})$, *and* $Z_{1-\alpha/2}$ *is the* $100\,(1-\alpha/2)$*th percentile of the standard normal distribution.*

3. Variance Function Estimation and Simultaneous Confidence Corridors

3.1. Estimation of the variance function

Define the estimated residual $\widehat{R}_{ij} = Y_{ij} - \sum_{\ell=0}^{p} X_{i\ell}\widehat{\beta}_\ell(\boldsymbol{z}_j)$ or $Y_{ij} - \sum_{\ell=0}^{p} X_{i\ell}\,\widehat{\beta}^{\mathrm{c}}_\ell(\boldsymbol{z}_j)$, for any $i = 1,\ldots,n$, $j = 1,\ldots,N$. We apply the bivariate spline smoothing method to $\{(\widehat{R}_{ij}, \boldsymbol{z}_j)\}_{j=1}^{N}$. Specifically, we define

$$\widehat{\eta}_i(\boldsymbol{z}) = \underset{g_i \in \mathcal{S}_d^r(\triangle_\eta)}{\operatorname{argmin}} \sum_{j=1}^{N}\left\{\widehat{R}_{ij} - g_i(\boldsymbol{z}_j)\right\}^2,\ i = 1,\ldots,n, \tag{3.1}$$

as the spline estimator of $\eta_i(\boldsymbol{z})$, where the triangulation $\triangle_\eta$ may differ from that introduced in Section 2 when estimating $\beta^o_\ell(\boldsymbol{z})$. Next, let $\widehat{\epsilon}_{ij} = \widehat{R}_{ij} - \widehat{\eta}_i(\boldsymbol{z}_j)$. Define the estimators of $G_\eta(\boldsymbol{z},\boldsymbol{z}')$ and $\sigma^2(\boldsymbol{z}_j)$ as

$$\widehat{G}_\eta(\boldsymbol{z},\boldsymbol{z}') = n^{-1}\sum_{i=1}^{n}\widehat{\eta}_i(\boldsymbol{z})\widehat{\eta}_i(\boldsymbol{z}') \text{ and } \widehat{\sigma}^2(\boldsymbol{z}_j) = n^{-1}\sum_{i=1}^{n}\widehat{\epsilon}_{ij}\widehat{\epsilon}_{ij}, \tag{3.2}$$

respectively. In general, for spline estimators ($d \ge 0$), denote $\widehat{\boldsymbol{\Xi}}_n(\boldsymbol{z}) = \{\widehat{\sigma}^2_{n,\ell\ell'}(\boldsymbol{z})\}_{\ell,\ell'=0}^{p}$, where

$$\begin{aligned}\widehat{\boldsymbol{\Xi}}_n(\boldsymbol{z}) = \frac{1}{n^2N^2}\widetilde{\mathbb{B}}(\boldsymbol{z})^\top\sum_{i=1}^{n}\Bigg\{&\sum_{j,j'=1}^{N}\boldsymbol{\Gamma}_{n,\rho}^{-1}\mathbf{U}_{ij}\mathbf{U}_{ij'}^\top\widehat{G}_\eta(\boldsymbol{z}_j,\boldsymbol{z}_{j'})\boldsymbol{\Gamma}_{n,\rho}^{-1}\\ &+\sum_{j=1}^{N}\mathbf{U}_{ij}\mathbf{U}_{ij}^\top\widehat{\sigma}^2(\boldsymbol{z}_j)\Bigg\}\widetilde{\mathbb{B}}(\boldsymbol{z}).\end{aligned} \tag{3.3}$$

Note that the estimation can be much simplified if PCST smoothing is applied. In this case, the variance-covariance matrix $\boldsymbol{\Sigma}_n(\boldsymbol{z})$ can be simply estimated using

$$\widehat{\mathbf{\Sigma}}_n(\boldsymbol{z}) = \{(\widehat{\sigma}^{\mathrm{c}}_{n,\ell\ell'})^2(\boldsymbol{z})\}_{\ell,\ell'=0}^{p} = \frac{1}{n}\left(n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_i^{\top}\right)^{-1}\left\{\widehat{G}_\eta(\boldsymbol{z},\boldsymbol{z}) + \frac{\widehat{\sigma}^2(\boldsymbol{z})}{NA_{m(\boldsymbol{z})}}\right\},$$

where $A_{m(\boldsymbol{z})}$ is the area of triangle $T_{m(\boldsymbol{z})}$ divided by the area of the domain. The following conditions (C2)–(C3) are required for the bivariate spline approximation in the covariance estimation and to establish the estimation consistency. The proofs of the results in this section are provided in the Supplementary Material.

(C2) For any $k \geq 1$, $\psi_k(\boldsymbol{z}) \in \mathcal{W}^{s+1,\infty}$ for an integer $s \geq 0$, and for a sequence $\{K_n\}_{n=1}^{\infty}$ of increasing positive integers with $\lim_n K_n \to \infty$, $|\triangle_\eta|^{s+1}\sum_{k=1}^{K_n}\lambda_k^{1/2}\|\psi_k\|_{s+1,\infty} \to 0$ as $N \to \infty$, $n \to \infty$.

(C3) As $N \to \infty$, $n \to \infty$, for some $0 < \kappa < 1$, $N^{-1}n^{1/(d+1)+\kappa} \to 0$, $N|\triangle_\eta|^2 \to \infty$, and $n|\triangle_\eta|^2/(\log n)^{1/2} \to \infty$.

Assumption (C2) concerns the bounded smoothness of the principal components that bound the bias terms in the spline covariance estimator.

Theorem 4. *Under Assumptions* (A1)–(A6) *and* (C1)–(C3), $\widehat{G}_\eta(\boldsymbol{z},\boldsymbol{z}')$ *uniformly converges to* $G_\eta(\boldsymbol{z},\boldsymbol{z}')$ *in probability; that is,* $\sup_{(\boldsymbol{z},\boldsymbol{z}')\in\Omega^2}|\widehat{G}_\eta(\boldsymbol{z},\boldsymbol{z}') - G_\eta(\boldsymbol{z},\boldsymbol{z}')| = o_P(1)$.

Corollary 1. *Under Assumptions* (A1)–(A6), (C1)–(C3), *the estimator of* $\widehat{\mathbf{\Sigma}}_n(\boldsymbol{z})$ *uniformly converges to to* $\mathbf{\Sigma}_n(\boldsymbol{z})$ *in probability; that is,* $\sup_{\boldsymbol{z}\in\Omega}|\widehat{\mathbf{\Sigma}}_n(\boldsymbol{z}) - \mathbf{\Sigma}_n(\boldsymbol{z})| = o_P(1)$.

Denote

$$\widehat{\sigma}^{\mathrm{c}}_{n,\ell\ell}(\boldsymbol{z}) = n^{-1/2}\left[\mathbf{e}_\ell^{\top}\left(n^{-1}\sum_{i=1}^{n}\widetilde{\mathbf{X}}_i\widetilde{\mathbf{X}}_i^{\top}\right)^{-1}\mathbf{e}_\ell\left\{\widehat{G}_\eta(\boldsymbol{z},\boldsymbol{z}) + \frac{\widehat{\sigma}^2(\boldsymbol{z})}{NA_{m(\boldsymbol{z})}}\right\}\right]^{1/2}. \quad (3.4)$$

From Corollary 1, $\widehat{\sigma}^{\mathrm{c}}_{n,\ell\ell}(\boldsymbol{z})$ is a consistent estimator of $\sigma^{\mathrm{c}}_{n,\ell\ell}(\boldsymbol{z})$ in (2.9).

3.2. Bootstrap simultaneous confidence corridors (SCCs)

From Theorems 2 and 3 and Slutzky's Theorem, we have the following asymptotic PCIs.

Corollary 2.

(a) For the BPST estimators, under Assumptions (A1)–(A6), *for any* $\ell = 0,\ldots,p$, $\alpha \in (0,1)$, *as* $N \to \infty$, $n \to \infty$, *an asymptotic* $100(1-\alpha)\%$ *PCI for* $\beta_\ell^o(\boldsymbol{z})$, *is* $\widehat{\beta}_\ell(\boldsymbol{z}) \pm \sigma_{n,\ell\ell}(\boldsymbol{z})Z_{1-\alpha/2}$, *for any* $\boldsymbol{z} \in \Omega$, *where* $\sigma^2_{n,\ell\ell}(\boldsymbol{z})$ *is*

the (ℓ,ℓ)*th entry of the matrix* $\boldsymbol{\Xi}_n^{-1/2}(\boldsymbol{z})$*, and* $Z_{1-\alpha/2}$ *is the* $100\,(1-\alpha/2)$*th percentile of the standard normal distribution.*

(b) For the PCST estimators, under Assumptions (A1′) *and* (A2)–(A6)*, if for some* $0<\kappa<1$*,* $N^{-1}n^{1+\kappa}\to 0$*, an asymptotic* $100(1-\alpha)\%$ *PCI for* $\beta_\ell^o(\boldsymbol{z})$ *is* $\widehat{\beta}_\ell^{\mathrm{c}}(\boldsymbol{z})\pm\sigma_{n,\ell\ell}^{\mathrm{c}}(\boldsymbol{z})Z_{1-\alpha/2}$*, for any* $\boldsymbol{z}\in\Omega$*, where* $\sigma_{n,\ell\ell}^{c}(\boldsymbol{z})$ *is the standard deviation function of* $\widehat{\beta}_\ell^{\mathrm{c}}(\boldsymbol{z})$ *in Theorem* 3*.*

Next, we introduce a simple bootstrap approach to extend the PCIs to the SCCs. Our approach is based on the nonparametric bootstrap method used in Hall and Horowitz (2013). We triangulate the domain Ω using quasi-uniform triangles, obtaining a set of approximate $100(1-\alpha)\%$ PCIs. In the following, α_0 denotes the nominal confidence level of the desired SCCs. We recalibrate the PCIs using the following bootstrap method.

Step 1. Based on $\{(\widetilde{\mathbf{X}}_i,Y_{ij})\}_{j=1,i=1}^{N,n}$, obtain the coefficient functions $\beta_\ell^o(\boldsymbol{z})$ using the BPST estimators $\widehat{\beta}_\ell(\boldsymbol{z})$ in (2.5) or the PCST estimators $\widehat{\beta}_\ell^{\mathrm{c}}(\boldsymbol{z})$ in (2.8), for $\ell=0,\dots,p$. Let $\widehat{\mu}(\boldsymbol{z})=\sum_{\ell=0}^p X_{i\ell}\widehat{\beta}_\ell(\boldsymbol{z})$ or $\sum_{\ell=0}^p X_{i\ell}\widehat{\beta}_\ell^{\mathrm{c}}(\boldsymbol{z})$.

Step 2. Obtain $\widehat{\eta}_i(\boldsymbol{z})$ and $\widehat{\varepsilon}_{ij}$ presented in (3.1)–(3.2), and estimate $G_\eta(\boldsymbol{z},\boldsymbol{z})$, $\sigma^2(\boldsymbol{z})$, and $\sigma_{n,\ell\ell}^2(\boldsymbol{z})$ using $\widehat{G}_\eta(\boldsymbol{z},\boldsymbol{z})$ and $\widehat{\sigma}^2(\boldsymbol{z})$ in (3.2) and $\widehat{\sigma}_{n,\ell\ell}^2(\boldsymbol{z})$ in (3.3) or (3.4), respectively.

Step 3. Obtain an adjusted nominal confidence level $\widehat{\alpha}_\ell(\alpha_0)$.

(i) Generate an independent random sample $\delta_i^{(b)}$ and $\delta_{ij}^{(b)}$ from $\{-1,1\}$ with probability 0.5 each, and define $Y_{ij}^{*(b)}=\widehat{\mu}(\boldsymbol{z}_j)+\delta_i^{(b)}\widehat{\eta}_i(\boldsymbol{z}_j)+\delta_{ij}^{(b)}\widehat{\varepsilon}_{ij}$.

(ii) Based on $\{(\widetilde{\mathbf{X}}_i,Y_{ij}^{*(b)})\}_{j=1,i=1}^{N,n}$, obtain $\widehat{\beta}_\ell^{*(b)}(\boldsymbol{z})$ using (2.5) or (2.8), and calculate $\widehat{\sigma}_{n,\ell\ell}^{*(b)}$ using (3.3) or (3.4).

(iii) Construct SCCs for the resampled data $\{(\widetilde{\mathbf{X}}_i,Y_{ij}^{*(b)})\}_{j=1,i=1}^{N,n}$: $\mathcal{B}_{(b)}^*(\alpha)$, $b=1,\dots,B$,

$$\mathcal{B}_{(b)}^*(\alpha)=\{(\boldsymbol{z},y):\boldsymbol{z}\in\Omega,\\ \widehat{\beta}_\ell^{*(b)}(\boldsymbol{z})-\widehat{\sigma}_{n,\ell\ell}^{*(b)}(\boldsymbol{z})Z_{1-\alpha/2}\le y\le\widehat{\beta}_\ell^{*(b)}(\boldsymbol{z})+\widehat{\sigma}_{n,\ell\ell}^{*(b)}(\boldsymbol{z})Z_{1-\alpha/2}\}.$$

(iv) Estimate the coverage rate $\tau_\ell(\boldsymbol{z}_j,\alpha)=P\{(\boldsymbol{z}_j,\widehat{\beta}_\ell(\boldsymbol{z}_j))\in\mathcal{B}^*(\alpha)|\mathbb{X}\}$ using $\widehat{\tau}_\ell(\boldsymbol{z}_j,\alpha)=(1/B)\sum_{b=1}^B I\{(\boldsymbol{z}_j,\widehat{\beta}_\ell(\boldsymbol{z}_j))\in\mathcal{B}_{(b)}^*(\alpha)\}$.

(v) Find the root of the equation $\widehat{\tau}_\ell(\boldsymbol{z}_j,\alpha)=1-\alpha_0$, for $j=1,\dots,N$, and denote it as $\{\widehat{\alpha}_\ell(\boldsymbol{z}_j,\alpha_0)\}_{j=1}^N$. The root can be found using the grid method by repeating the last two steps for different values of α.

(vi) Take the minimum of $\{\widehat{\alpha}_\ell(\boldsymbol{z}_j, \alpha_0)\}_{j=1}^N$ and denote it as $\widehat{\alpha}_\ell \equiv \widehat{\alpha}_\ell(\alpha_0)$.

Step 4. Construct the final SCCs: $\mathcal{B}(\widehat{\alpha}_\ell) = \{(\boldsymbol{z}, y) : \boldsymbol{z} \in \Omega, \widehat{\beta}_\ell(\boldsymbol{z}) - \widehat{\sigma}_{n,\ell\ell}(\boldsymbol{z}) Z_{1-\widehat{\alpha}_\ell/2} \leq y \leq \widehat{\beta}_\ell(\boldsymbol{z}) + \widehat{\sigma}_{n,\ell\ell}(\boldsymbol{z}) Z_{1-\widehat{\alpha}_\ell/2}\}$.

4. Implementation

The proposed procedure can be implemented using our R package "FDAimage" (Yu, Wang and Wang (2019)), in which the bivariate spline basis is generated using the R package "BPST" (Wang, Wang and Lai (2019)). When the response imaging seems to be a realization from some smooth function, we suggest using the smoothing parameter $r = 1$ and degree $d \geq 5$, which achieves full estimation power asymptotically (Lai and Schumaker (2007)). In contrast, if there are sharp edges on the images, we suggest considering the PCST presented in Section 2.4.

Selecting suitable values for the smoothing parameters is important to good model fitting. To select $\rho_{n,\ell}$, for $\ell = 0, \ldots, p$, we used K-fold cross-validation (CV). The individuals are randomly partitioned into K groups, where one group is retained as a test set, and the remaining $K - 1$ groups are used as training sets. The CV process is repeated K times (the folds), with each of the K groups used exactly once as the validation data. Then, the K-fold CV score is

$$\mathrm{CV}(\rho_{n,0}, \ldots, \rho_{n,p}) = K^{-1} \sum_{k=1}^{K} (|\mathcal{V}_k|N)^{-1} \sum_{i \in \mathcal{V}_k} \sum_{j=1}^{N} \left\{ Y_{ij} - \widetilde{\mathbf{X}}_i^\top \widehat{\boldsymbol{\beta}}_{-k}(\boldsymbol{z}_j) \right\}^2,$$

where $\mathcal{V}_k$ is the kth testing set for $k = 1, \ldots, K$, and $\widehat{\boldsymbol{\beta}}_{-k}$ is the corresponding estimator after removing the kth testing set. We use $K = 5$ in our numerical examples.

To determine an optimal triangulation, the criterion usually considers the shape, size, or number of triangles. In terms of shape, a "good" triangulation usually refers to one with well-shaped triangles without small angles and/or obtuse angles. Therefore, for a given number of triangles, Lai and Schumaker (2007) and Lindgren, Rue and Lindström (2011) recommended selecting the triangulation according to "max-min" criterion, which maximizes the minimum angle of all the angles of the triangles in the triangulation. With respect to the number of triangles, our numerical studies show that a lower limit of the number of triangles is necessary to capture the features of the images. However, once this minimum number has been reached, refining the triangulation further usually has little effect on the fitting process. In practice, when using higher-order BPST smoothing, we suggest taking the number of triangles as $H_n = \min\{\lfloor c_1 n^{1/(2d+2)} N^{1/2} \rfloor, N/10\}$,

where c_1 is a tuning parameter. We find that $c_1 \in [0.3, 2.0]$ works well in our numerical studies. When using the PCST, we suggest taking the number of triangles as $H_n = \min\{\lfloor c_2 n^{-1/4} N \rfloor, N/2\}$, with $c_2 \in [0.3, 2.0]$. Once H_n is chosen, we can build the triangulation using typical triangulation construction methods, such as Delaunay triangulation and DistMesh (Persson and Strang (2004)).

5. Simulation Studies

In this section, we conduct two Monte Carlo simulation studies using our R package "FDAimage" (Yu, Wang and Wang (2019)) to examine the finite-sample performance of the proposed methodology. The triangulations used here can be found in the data set in the "FDAimage" package. To illustrate the performance of our estimation method, we compare the proposed spline method with the kernel method proposed by Zhu, Fan and Kong (2014) (Kernel) and the tensor regression method of Li and Zhang (2017) (Tensor). To implement the kernel method, we use the R Package *SVCM*, which is publicly available at `https://github.com/BIG-S2/SVCM`. For the tensor method, the accompanying MATLAB code at `https://ani.stat.fsu.edu/~henry/TensorEnvelopes_html.html` is used. We compare the proposed method with the tensor regression approach in Li and Zhang (2017) and the three-stage FDA approach in Zhu, Fan and Kong (2014).

5.1. Example 1

To illustrate the advantage of the proposed method over a complex domain, we study the horseshoe domain in Sangalli, Ramsay and Ramsay (2013). The response images are generated from the following model: $Y_{ij} = \beta_0^o(\boldsymbol{z}_j) + X_i\beta_1^o(\boldsymbol{z}_j) + \eta_i(\boldsymbol{z}_j) + \sigma\varepsilon_{ij}$, for $i = 1, \ldots, n$, $j = 1, \ldots, N$, and $\boldsymbol{z}_j \in \Omega$. To understand the advantages and disadvantages of different methods, we consider two types of coefficient functions in the above image-on-scalar regression model: (I) functions with jumps; and (II) smooth functions. The true coefficient functions are shown in Figure 3.

For each image, we set the resolution as 100×50 (pixels). The true signal falls only within the horseshoe domain (3,182 pixels); outside the domain is pure noise. We generate the scalar covariate $X_i \sim N(0, 1)$, and then truncate it by $[-3, +3]$. We set $\eta_i(\boldsymbol{z}) = \sum_{k=1}^{2} \lambda_k^{1/2} \xi_{ik} \psi_k(\boldsymbol{z})$, where $(\lambda_1, \lambda_2) = (0.1, 0.02)$ or $(0.2, 0.05)$ and ξ_{i1} and $\xi_{i2} \sim N(0, 1)$, $\psi_1(\boldsymbol{z}) = c_1 \sin(2\pi z_1)$, and $\psi_2(\boldsymbol{z}) = c_2 \cos(2\pi z_2)$. Let $c_1 = 0.56$ and $c_2 = 0.61$, such that ψ_1 and ψ_2 are orthonormal functions on Ω. The measurement error ε_{ij} is independently generated from $N(0, 1)$ and $\sigma = 1.0,\ 2.0$.

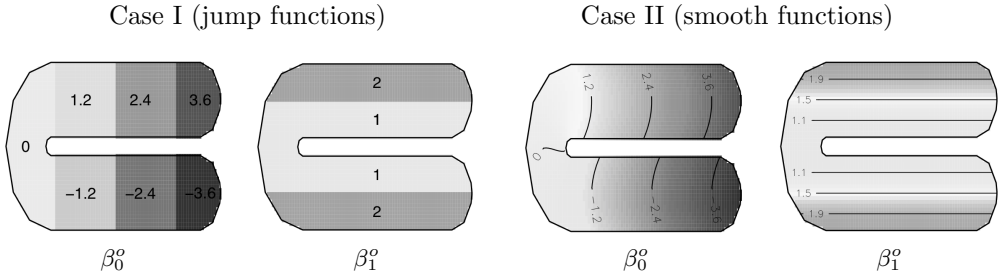

Figure 3. The true coefficient functions in Simulation Example 1.

Table 1. Estimation errors of the coefficient estimators, $\sigma = 2.0$.

Function Type	n	Method	$\lambda_1 = 0.03,\ \lambda_2 = 0.006$		$\lambda_1 = 0.2,\ \lambda_2 = 0.05$	
			β_0	β_1	β_0	β_1
Jump	50	BPST	0.0139	0.0182	0.0145	0.0189
		PCST	0.0088	0.0090	0.0094	0.0097
		Kernel	0.0801	0.0819	0.0807	0.0826
		Tensor	0.0799	0.0248	0.0799	0.0254
	100	BPST	0.0090	0.0118	0.0093	0.0122
		PCST	0.0044	0.0044	0.0047	0.0047
		Kernel	0.0400	0.0405	0.0403	0.0409
		Tensor	0.0395	0.0166	0.0399	0.0171
Smooth	50	BPST	0.0026	0.0032	0.0032	0.0041
		PCST	0.0088	0.0090	0.0119	0.0139
		Kernel	0.0801	0.0819	0.0807	0.0826
		Tensor	0.0799	0.0256	0.0806	0.0271
	100	BPST	0.0016	0.0019	0.0019	0.0022
		PCST	0.0070	0.0086	0.0073	0.0090
		Kernel	0.0400	0.0405	0.0403	0.0409
		Tensor	0.0399	0.0168	0.0402	0.0179

To fit the model, we consider the BPST and PCST methods presented in Section 2. To obtain the BPST estimators, we set $d = 5$ and $r = 0$ when generating the bivariate spline basis functions. Figure S2.1 in the Supplementary Material illustrates the triangulations used for the BPST and PCST. The triangulation used for the BPST ($\triangle_1$) contains 90 triangles (73 vertices), and the triangulation used for the PCST ($\triangle_2$) contains 346 triangles (226 vertices).

We quantify the estimation accuracy of the coefficient functions using the mean squared error (MSE). Table 1 provides the average MSE (across 500 Monte Carlo experiments) for two types of coefficient functions. To save space, we present the results for $\sigma = 2.0$ only; the results for $\sigma = 1.0$ are presented in Table S2.1 in the Supplementary Material. As expected, the estimation accuracy of all

the methods improves as the sample size increases or the noise level decreases. In both scenarios, the BPST and PCST outperform the other two competitors, reflecting the advantage of our method over a complex domain. When the true coefficient functions are smooth, the BPST provides the best estimation, followed by the PCST. On the other hand, when the true coefficient function contains jumps, the PCST provides a better result. For the tensor regression, the estimator of $\beta_1^o(\cdot)$ is much more accurate than that of $\beta_0^o(\cdot)$, owing to the design of the coefficient function. Figure 3 shows that, in contrast to the intercept function of $\beta_0^o(\cdot)$, the true slope function of $\beta_1^o(\cdot)$ is still smooth across the complex boundary. Moreover, when the coefficient function is smooth across the boundary, the estimation accuracy is also affected by the domain of the true signal. The performance of the kernel method is not affected by the design of the coefficient functions. Instead, it depends heavily on the noise level, owing to the three-stage structure.

5.2. Example 2

In this example, we simulate the data by considering the domains of the fifth and 35th slices of the brain images illustrated in Section 6 as the domain Ω. We generate response images based on a set of smooth coefficient functions from the following model: $Y_{ij} = \sum_{\ell=0}^{2} X_{i\ell}\beta_\ell^o(\boldsymbol{z}_j) + \eta_i(\boldsymbol{z}_j) + \sigma\varepsilon_{ij}$, for $i = 1, \ldots, n$, $j = 1, \ldots, N$, and $\boldsymbol{z}_j \in \Omega$, where $\beta_0^o(\boldsymbol{z}) = 5\{(z_1 - 0.5)^2 + (z_2 - 0.5)^2\}$, $\beta_1^o(\boldsymbol{z}) = -1.5z_1^3 + 1.5z_2^3$ and $\beta_2^o(\boldsymbol{z}) = 2 - 2\exp[-8\{(z_1 - 0.5)^2 + (z_2 - 0.5)^2\}]$. The true coefficient images are shown in the first columns of Figures S2.5 and S2.6 in the Supplementary Material for the fifth and 35th slices, respectively. For each image, we simulate the data at all 79×95 pixels. To mimic real brain images, the true signals are generated only on the pixels/voxels (3,476 or 5,203 pixels in total) within the brain domain; outside the boundary of the brain, the image contains only noise. We set $X_{i0} = 1$ and generate $\widetilde{\mathbf{X}}_i = (X_{i1}, X_{i2})^\top \sim N(\mathbf{0}, \mathbf{\Sigma})$, with $\mathbf{\Sigma} = \left(\begin{smallmatrix} 1.0 & 0.5 \\ 0.5 & 1.0 \end{smallmatrix}\right)$ and $X_{i\ell}$ truncated by $[-3, +3]$. For the error terms, we set $\eta_i(\boldsymbol{z}) = \sum_{k=1}^{2} \lambda_k^{1/2}\xi_{ik}\psi_k(\boldsymbol{z})$, where ξ_{i1} and $\xi_{i2} \sim N(0, 1)$, $\psi_1(\boldsymbol{z}) = 1.488\{\sin(\pi z_1) - 1.5\}$, $\psi_2(\boldsymbol{z}) = 1.939\cos(2\pi z_2)$, and $(\lambda_1, \lambda_2) = (0.1, 0.02)$ or $(0.2, 0.05)$. The measurement error ε_{ij} is independently generated from $N(0, 1)$ and $\sigma = 0.5,\ 1.0$. To conserve space, we show only the results for the domain of the fifth slice for $\sigma = 1.0$ here. The results for $\sigma = 0.5$ and those based on the domain of the 35th slice are shown in Section S2 of the Supplementary Material.

Because the functions in this example are smooth, for the bivariate spline approach, we consider only the BPST method. To further study the effect of different triangulations, we consider $\triangle_3$ and $\triangle_4$; see Figure S2.4 in the Supple-

Table 2. Estimation errors of the coefficient function estimators, $\sigma = 1.0$.

n	Method	$\lambda_1 = 0.1,\ \lambda_2 = 0.02$			$\lambda_1 = 0.2,\ \lambda_2 = 0.05$		
		β_0	β_1	β_2	β_0	β_3	β_2
50	BPST($\triangle_3$)	0.003	0.005	0.005	0.007	0.011	0.010
	BPST($\triangle_4$)	0.003	0.005	0.005	0.007	0.010	0.009
	Kernel	0.023	0.032	0.032	0.026	0.037	0.037
	Tensor	0.023	0.013	0.019	0.026	0.017	0.024
100	BPST($\triangle_3$)	0.002	0.002	0.002	0.003	0.005	0.005
	BPST($\triangle_4$)	0.002	0.002	0.002	0.003	0.004	0.004
	Kernel	0.011	0.015	0.015	0.013	0.018	0.018
	Tensor	0.011	0.007	0.011	0.013	0.009	0.013

mentary Material. Similarly to Section 5.1, we summarize the MSE for different coefficient functions based on 500 Monte Carlo experiments in Table 2. Columns 2–5 in Figure S2.5 in the Supplementary Material show the estimated coefficient functions using the kernel, tensor and BPST methods, respectively. Table 2 and Figure S2.5 in the Supplementary Material show that the estimation accuracy improves for all methods as the sample size increases or the noise level decreases. In all settings, the BPST method has the smallest MSE compared with the kernel and tensor methods, reflecting the advantage of our method in estimating the coefficient functions and, hence, the regression function. Because the kernel and tensor methods are both designed for a rectangle domain, the estimation accuracy can be affected by the noise outside the domain. Futhermore, the MSE is invariable across two triangulations, thus, $\triangle_3$ might be sufficient to capture the feature in the data set. This also implies that when this minimum number of triangles is reached, further refining the triangulation has little effect on the fitting process, but makes the computational burden unnecessarily heavy.

Finally, we illustrate the finite-sample performance of the proposed SCCs for the coefficient functions described in Section 3. In particular, we report the empirical coverage probabilities of the nominal 95% SCCs using triangulation $\triangle_3$. We evaluate the coverage of the proposed SCCs over all pixels on the interior of Ω, and test whether the true functions are entirely covered by the SCCs at these pixels. Table 3 summarizes the empirical coverage rate (ECR) for 500 Monte Carlo experiments of the 95% SCCs and the average width of the SCCs. The results clearly show that the ECRs of the SCCs are well approximated to 95%, particularly as the sample size increases. Table 3 also reveals that the SCCs tend to be narrower when the sample size becomes larger or the noise level decreases.

Table 3. The coverage rate of the 95% SCCs for the coefficient functions.

n	λ	σ	Coverage			Width		
			β_0	β_1	β_2	β_0	β_1	β_2
50	(0.1,0.02)	0.5	0.976	0.928	0.938	0.332	0.362	0.377
		1.0	0.976	0.940	0.952	0.358	0.392	0.413
	(0.2,0.05)	0.5	0.962	0.918	0.932	0.445	0.497	0.513
		1.0	0.970	0.930	0.940	0.478	0.527	0.544
100	(0.1,0.02)	0.5	0.970	0.956	0.956	0.234	0.250	0.267
		1.0	0.978	0.968	0.978	0.262	0.285	0.297
	(0.2,0.05)	0.5	0.956	0.958	0.936	0.313	0.348	0.357
		1.0	0.966	0.964	0.954	0.344	0.378	0.389

6. ADNI Data Analysis

To illustrate the proposed method, we consider the spatially normalized FDG (fludeoxyglucose) PET data of the Alzheimer's Disease Neuroimaging Initiative (ADNI). As pointed out in Marcus, Mena and Subramaniam (2014), FDG-PET images have been shown to be a promising modality for detecting functional brain changes in Alzheimer's Disease (AD). The data can be obtained from the ADNI database at `http://adni.loni.usc.edu/`. The database contains spatially normalized PET images of 447 subjects. Of these 447 subjects, 112 have normal cognitive functions, considered to be the control group, 213 are diagnosed as mild cognitive impairment (MCI), and 122 are diagnosed as AD. Table S2.5 in the Supplementary Material summarizes the distribution of patients by diagnosis status and sex.

In this study, we examine several patient-level features: (i) demographical features, such as age (Age) and sex (Sex); (ii) a dummy variable for the abnormal diagnosis status "MA" (1 = "AD" or "MCI", zero otherwise); (iii) a dummy variable for "AD" (1 = "AD," zero otherwise); and (iv) dummy variables for the APOE genotype, the strongest genetic risk factor for "AD"; see Corder et al. (1993). We code APOE_1 as a dummy variable for subjects with one epsilon 4 allele, and APOE_2 as subjects who have two alleles.

Noting that the PET images are 3D, we select the 5th, 8th, 15th, 35th, 55th, 62nd, and 6fifth horizontal slices (bottom to up) of the brain from a total of 68 slices to illustrate our method. Each slice of the image contains 79×95 pixels, but the domains of different brain slices are quite different. Specifically, the domain boundary for the bottom slices and upper slices are much more complex than the slices in the middle; more examples can be found in Figure S2.7 in the Supplementary Material. For each slice, we consider the following image-on-scalar

Table 4. 10-fold CV results for the ADNI dataset. ($\times 10^{-2}$)

Method	Slice 5	Slice 8	Slice 15	Slice 35	Slice 55	Slice 62	Slice 65
BPST	1.4508	1.4809	1.5013	1.5633	2.0693	2.3020	2.6239
Kernel	1.4533	1.4828	1.5021	1.5638	2.0715	2.3060	2.6303
Tensor	1.5010	1.5260	1.5400	1.5900	2.1000	2.3340	2.6400

regression:

$$Y_i(\boldsymbol{z}_j) = \beta_0(\boldsymbol{z}_j) + \beta_1(\boldsymbol{z}_j)\mathrm{MA}_i + \beta_2(\boldsymbol{z}_j)\mathrm{AD}_i + \beta_3(\boldsymbol{z}_j)\mathrm{Age}_i + \beta_4(\boldsymbol{z}_j)\mathrm{Sex}_i + \beta_5(\boldsymbol{z}_j)\mathrm{APOE}_{1i} + \beta_6(\boldsymbol{z}_j)\mathrm{APOE}_{2i} + \eta_i(\boldsymbol{z}_j) + \sigma(\boldsymbol{z}_j)\varepsilon_i(\boldsymbol{z}_j),\ i = 1, \ldots, n.$$

We fit the above model using the BPST method for each slice; see Figure S2.7 in the Supplementary Material for the set of triangulations used for the BPST method. The image maps in Figure 4 and Figures S2.8 and S2.9 in the Supplementary Material present the estimated coefficient functions using the BPST ($d = 5$, $r = 1$) method. To evaluate the predictive performance, Table 4 reports the 10-fold CV (parts of the images are left out as training sets) MSPE results for the BPST method, kernel method in Zhu, Fan and Kong (2014), and tensor regression method in Li and Zhang (2017). The table shows that the MSPEs of the BPST method are uniformly smaller than those of the kernel method and tensor regression methods.

Next, we construct the 95% SCCs to check whether the covariates are significant. The yellow and blue colors on the "significance" map in Figure 4 indicate the regions in which zero is below the lower SCC or above the upper SCC, respectively. Using these estimated coefficient functions and the 95% SCCs, we can assess the impact of the covariates on the response images. Taking the fifth slice as an example, the main impact of "AD" on in the PET images is an increase in activity in the cerebellum compared with a normal individual. The cerebellum obtains information from the sensory systems, spinal cord, and other parts of the brain, and then regulates motor movements, resulting in smooth and balanced muscular activities. The significance map of "Age" also shows an increase in activity in the cerebellum, and "Sex" shows different effects in the male and female brain images. The significance maps of the covariates for all other slices of the PET image are shown in Figures S2.10 – S2.11 in the Supplementary Material. From these figures, we can see that the effect of the covariates on the brain activity level varies between slices, depending on the location of the slice; see the Supplementary Material for further details.

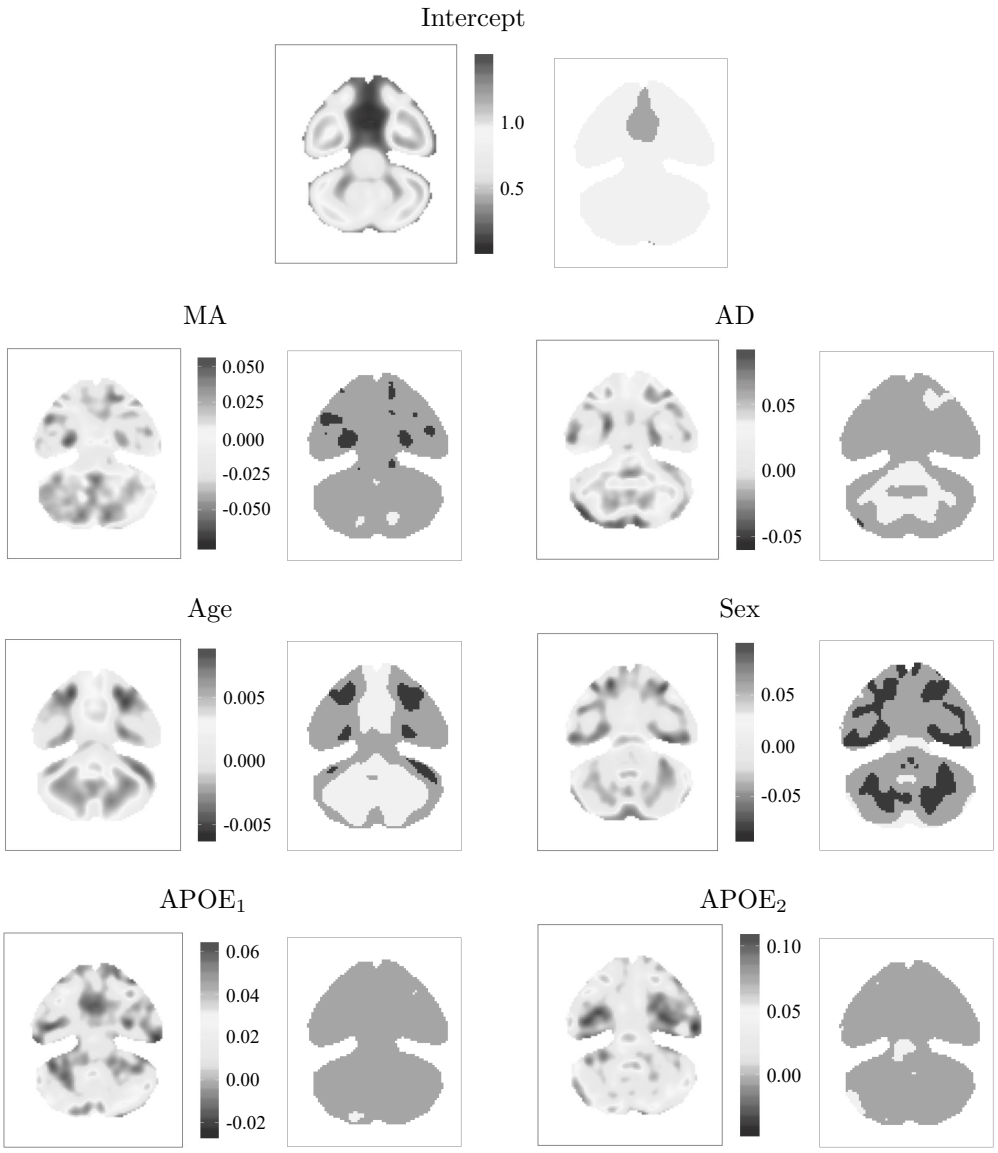

Figure 4. The BPST estimate and significance map of the coefficient functions for the fifth slice of the PET images. The yellow and blue colors in the significance map indicate the regions in which zero is below the lower SCC or above the upper SCC, respectively.

7. Conclusion

We examine a class of image-on-scalar regression models to efficiently explore the spatial nonstationarity of a regression relationship between imaging responses and scalar predictors, allowing the regression coefficients to change with the pixels. We have proposed an efficient estimation procedure to carry out statistical

inference. We have developed a fast and accurate method for estimating the coefficient images, while consistently estimating their standard deviation images. Our method provides coefficient maps and significance maps that highlight and visualize the associations with brain and the potential risk factors, adjusted for other patient-level features, as well as permitting inference. In addition, it allows an easy implementation of piecewise polynomial representations of various degrees and smoothness over an arbitrary triangulation, and therefore can handle irregular-shaped 2D objects with different visual qualities. This provides enormous flexibility, accommodating various types of nonstationarity that are commonly encountered in imaging data analysis. Our methodology is extendable to 3D images to fully realize its potential usefulness in biomedical imaging. Instead of using bivariate splines over triangulation, the trivariate splines over tetrahedral partitions introduced in Lai and Schumaker (2007) could be well suited, because they have many properties in common with the bivariate splines over triangulation. However, this is a nontrivial task, because the compuation is much more challenging for high-resolution 3D images than it is for 2D images, and thus warrants further investigation.

Supplementary Material

In the online Supplementary Material, we provide technical proofs for the main theorems and additional results from the simulation studies and ADNI data analysis.

Acknowledgments

The authors wish to thank the editor, associate editor, and two reviewers for their constructive comments and suggestions. Shan Yu's research was partially supported by the Iowa State University Plant Sciences Institute Scholars Program. Li Wang's research was partially supported by National Science Foundation grants DMS-1542332 and DMS-1916204. Lijian Yang's research was partially supported by National Natural Science Foundation of China award 11771240. The study data were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (`ADNI.loni.usc.edu`). As such, the investigators within the ADNI contributed to the design and implementation of the ADNI and/or provided data, but did not participate in the analysis or the writing of this report. A complete listing of ADNI investigators can be found at: `http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf`.

References

Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters* **45**, 11–22.

Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica* **13**, 571–591.

Chen, K., Delicado, P. and Müller, H.-G. (2017). Modelling function-valued stochastic processes, with applications to fertility dynamics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79**, 177–196.

Claeskens, G., Krivobokova, T. and Opsomer, J. D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika* **96**, 529–544.

Corder, E., Saunders, A., Strittmatter, W., Schmechel, D., Gaskell, P., Small, G. et al. (1993). Gene dose of apolipoprotein e type 4 allele and the risk of alzheimer's disease in late onset families. *Science* **261**, 921–923.

Gu, L., Wang, L., Härdle, W. K. and Yang, L. (2014). A simultaneous confidence corridor for varying coefficient regression with sparse functional data. *Test* **23**, 806–843.

Hall, P. and Horowitz, J. (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *The Annals of Statistics* **41**, 1892–1921.

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics* **35**, 70–91.

Hall, P. and Opsomer, J. D. (2005). Theory for penalised spline regression. *Biometrika* **92**, 105–118.

Hibar, D. P., Stein, J. L., Renteria, M. E., Arias-Vasquez, A., Desrivières, S., Jahanshad, N. et al. (2015). Common genetic variants influence human subcortical brain structures. *Nature* **520**, 224–229.

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

Lai, M.-J. and Schumaker, L. L. (2007). *Spline Functions on Triangulations*. Cambridge University Press, Cambridge.

Lai, M. J. and Wang, L. (2013). Bivariate penalized splines for regression. *Statistica Sinica* **23**, 1399–1417.

Li, L. and Zhang, X. (2017). Parsimonious tensor response regression. *Journal of the American Statistical Association* **112**, 1131–1146.

Li, Y. and Hsing, T. (2010). Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics* **38**, 3321–3351.

Lindgren, F., Rue, H. and Lindström, J. (2011). An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 423–498.

Marcus, C., Mena, E. and Subramaniam, R. M. (2014). Brain pet in the diagnosis of alzheimer's disease. *Clinical Nuclear Medicine* **39**, e413–e422.

Morris, J. S. (2015). Functional regression. *Annual Review of Statistics and Its Application* **2**, 321–359.

Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **68**, 179–199.

Müller, H.-G. (2005). Functional modelling and classification of longitudinal data. *Scandinavian*

Journal of Statistics **32**, 223–240.

Persson, P. O. and Strang, G. (2004). A simple mesh generator in matlab. *SIAM Review* **46**, 329–345.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Methodological)* **53**, 539–572.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer, New York.

Ramsay, T. (2002). Spline smoothing over difficult regions. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **64**, 307–319.

Reiss, P. T., Goldsmith, J., Shang, H. L. and Ogden, R. T. (2017). Methods for scalar-on-function regression. *International Statistical Review* **85**, 228–249.

Reiss, P. T., Huang, L. and Mennes, M. (2010). Fast function-on-scalar regression with penalized basis expansions. *The International Journal of Biostatistics* **6**, Article 28.

Sang, H. and Huang, J. Z. (2012). A full scale approximation of covariance functions for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **74**, 111–132.

Sangalli, L. M., Ramsay, J. O. and Ramsay, T. O. (2013). Spatial spline regression models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **75**, 681–703.

Schwarz, K. and Krivobokova, T. (2016). A unified framework for spline estimators. *Biometrika* **103**, 121–131.

Sentürk, D. and Müller, H.-G. (2010). Functional varying coefficient models for longitudinal data. *Journal of the American Statistical Association* **105**, 1256–1264.

Staicu, A. M., Crainiceanu, C. M. and Carroll, R. J. (2010). Fast methods for spatially correlated multilevel functional data. *Biostatistics* **11**, 177–194.

Stein, J. L., Hua, X., Lee, S., Ho, A. J., Leow, A. D., Toga, A. W. et al. (2010). Voxelwise genome-wide association study (vgwas). *Neuroimage* **53**, 1160–1174.

Wang, G., Wang, L. and Lai, M. J. (2019). Package 'BPST'. R package version 1.0. `https://github.com/funstatpackages/BPST`.

Wang, J.-L., Chiou, J.-M. and Muller, H.-G. (2016). Functional data analysis. *Annal Review of Statistics and Its Applications* **3**, 257–295.

Worsley, K. J., Taylor, J. E., Tomaiuolo, F. and Lerch, J. (2004). Unified univariate and multivariate random field theory. *NeuroImage* **23**, S189–S195.

Wu, S. and Müller, H.-G. (2011). Response-adaptive regression for longitudinal data. *Biometrics* **67**, 852–860.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics* **33**, 2873–2903.

Yu, S., Wang, G. and Wang, L. (2019). Package 'FDAimage'. R package version 1.0. `https://github.com/funstatpackages/FDAimage`.

Zhang, X. and Wang, J.-L. (2015). Varying-coefficient additive models for functional data. *Biometrika* **102**, 15–32.

Zhou, H., Li, L. and Zhu, H. (2013). Tensor regression with applications in neuroimaging data analysis. *Journal of the American Statistical Association* **108**, 540–552.

Zhu, H., Fan, J. and Kong, L. (2014). Spatially varying coefficient model for neuroimaging data with jump discontinuities. *Journal of the American Statistical Association* **109**, 1084–1098.

Zhu, H., Li, R. and Kong, L. (2012). Multivariate varying coefficient model for functional responses. *The Annals of statistics* **40**, 2634–2666.

Shan Yu
Department of Statistics, University of Virginia, Charlottesville, VA 22904, USA.
E-mail: sy5jx@virginia.edu
Guannan Wang
Department of Mathematics, College of William & Mary, Williamsburg, VA 23187, USA.
E-mail: gwang01@wm.edu
Li Wang
Department of Statistics and the Statistical Laboratory, Iowa State University, Ames, IA 50011, USA.
E-mail: lilywang@iastate.edu
Lijian Yang
Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing, 100084, China.
E-mail: yanglijian@tsinghua.edu.cn

(Received May 2019; accepted November 2019)