



Published in final edited form as:

*Biometrics*. 2020 June ; 76(2): 427–437. doi:10.1111/biom.13156.

## Simultaneous confidence corridors for mean functions in functional data analysis of imaging data

Yueying Wang<sup>1</sup>, Guannan Wang<sup>2</sup>, Li Wang<sup>1</sup>, R. Todd Ogden<sup>3</sup>

<sup>1</sup>Department of Statistics, Iowa State University, Ames, Iowa

<sup>2</sup>Department of Mathematics, College of William and Mary, Williamsburg, Virginia

<sup>3</sup>Department of Biostatistics, Columbia University, New York, New York

### Abstract

Motivated by recent work involving the analysis of biomedical imaging data, we present a novel procedure for constructing simultaneous confidence corridors for the mean of imaging data. We propose to use flexible bivariate splines over triangulations to handle an irregular domain of the images that is common in brain imaging studies and in other biomedical imaging applications. The proposed spline estimators of the mean functions are shown to be consistent and asymptotically normal under some regularity conditions. We also provide a computationally efficient estimator of the covariance function and derive its uniform consistency. The procedure is also extended to the two-sample case in which we focus on comparing the mean functions from two populations of imaging data. Through Monte Carlo simulation studies, we examine the finite sample performance of the proposed method. Finally, the proposed method is applied to analyze brain positron emission tomography data in two different studies. One data set used in preparation of this article was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

### Keywords

bivariate splines; functional principal component analysis; image analysis; semiparametric efficiency; triangulation

## 1 | INTRODUCTION

In recent years, as digital technology advanced significantly, valuable imaging data of body structures and organs can be easily collected during routine clinical practice. This new paradigm presents new opportunities to innovate in both research and clinical settings. Medical imaging technology has revolutionized health care over the past three decades, allowing doctors to find or detect tumors and other abnormalities and evaluate the

---

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

### SUPPORTING INFORMATION

Web Appendices, Tables, and Figures referenced in Sections 1, 5, and 6 are available with this article at the *Biometrics* website on Wiley Online Library. The R package for the proposed method ImageSCC is available at <https://github.com/funstatpackages/ImageSCC>. For the implementations of R packages of Triangulation and BPST, see <https://github.com/funstatpackages/Triangulation> and <https://github.com/funstatpackages/BPST>.

effectiveness of treatment. Functional data analysis (FDA) provides modern analytical tools for imaging data, which can be viewed as realizations of random functions. Let  $\Omega$  be a two-dimensional bounded domain, and  $\mathbf{z} = (z_1, z_2)$  be a point in  $\Omega$ . The model we consider is

$$Y_i(\mathbf{z}) = \mu(\mathbf{z}) + \eta_i(\mathbf{z}) + \sigma(\mathbf{z})\varepsilon_i(\mathbf{z}), \quad i = 1, \dots, n, \quad \mathbf{z} \in \Omega, \quad (1)$$

which is one instance of the general *function-on-scalar regression model*. In model (1),  $Y_i(\mathbf{z})$  denotes the imaging measurement at location  $\mathbf{z} \in \Omega$ ;  $\eta_i(\mathbf{z})$  is a stochastic process indexed by  $\mathbf{z}$ , which characterizes subject-level image variations; and  $\sigma(\mathbf{z})$  is a positive deterministic function. We assume that  $\eta_i(\mathbf{z})$  and  $\varepsilon_i(\mathbf{z})$  are mutually independent,  $\eta_i(\mathbf{z})$  are i.i.d. copies of a  $L_2$  stochastic process  $\eta(\mathbf{z})$  with mean zero and covariance function  $G_\eta(\mathbf{z}, \mathbf{z}')$ ,  $\varepsilon_i(\mathbf{z})$  are i.i.d. instances of a stochastic process of  $\varepsilon(\mathbf{z})$  with mean zero, and covariance function  $\text{Cov}\{\varepsilon(\mathbf{z}), \varepsilon(\mathbf{z}')\} = I(\mathbf{z} = \mathbf{z}')$ .

For biomedical imaging data, the objects (eg, tumor tissues, brain regions, etc) appearing in the images are typically irregularly shaped. Many smoothing methods in the literature, such as tensor product smoothing, kernel smoothing, and wavelet smoothing, suffer from the problem of “leakage” across the complex domains, that is, poor estimation over difficult regions as a result of smoothing inappropriately across boundaries of features.

In this article, we endeavor to address these challenges by applying bivariate splines over triangulations (Lai and Wang, 2013) to preserve important features (shape, smoothness) of imaging data. Spline functions defined this way offer more flexibility and varying amounts of smoothness, allowing us to better approximate the mean functions. We study the asymptotic properties of the spline estimators of  $\mu(\mathbf{z})$  by using bivariate penalized splines (BPS) defined on triangulations and show that our estimator is consistent and asymptotically normal.

In addition, when analyzing biomedical imaging data, such as brain images, typical questions lie in estimating the mean function,  $\mu(\mathbf{z})$ , together with quantifying the estimation uncertainty and making comparisons between populations. However, making a statistically rigorous inference for imaging data is challenging, and one of the main obstacles is the complicated spatial correlation structure. The prevailing analytic technique, termed the “mass univariate” approach, involves regarding each pixel/voxel as a unit, and for each unit, making a traditional univariate statistical inference, such as a simple  $t$  test. The obvious multiple comparisons issue can be dealt with in many ways; popular approaches include the Bonferroni correction, the random field theory (Worsley *et al.*, 2004; Adler and Taylor, 2007; Siegmund *et al.*, 2011), and the cluster threshold-based approach (Forman *et al.*, 1995).

However, many of the multiple testing methods are ad hoc methods, which involve setting the threshold by eye, based on the practitioner’s experience and knowledge. Our simulation study in Supporting Information Appendix A also demonstrates that those ad hoc methods heavily depend on the choice of the threshold. In this article, we propose an alternative approach that treats the imaging data as an instance of functional data, regarded as being continuously defined but observed on a regular grid. If we consider the imaging data as

being functional, attention naturally turns from considering each pixel/voxel as the basic analytical unit toward analyzing the entire image simultaneously, for instance, calculating simultaneous confidence corridors (SCCs; also called “simultaneous confidence bands” or “uniform confidence band/region”). As pointed out in Choi and Reimherr (2018) and Degras (2017), conventional multiple comparison methods are less useful in the functional data setup because the infinite cardinality of the domain would lead to unbounded confidence regions.

In statistics, SCCs are vital and fundamental tools for inference on the global behavior of functions (Degras, 2017). However, they have received relatively little attention in the literature of FDA. Moreover, existing SCC work for FDA has concentrated on the one-dimensional case. For the development of SCCs for mean curves of functional data, see the simulation-based techniques (Degras, 2011; Cao *et al.*, 2012; Zheng *et al.*, 2014; Cao and Wang, 2018), the functional principal component (FPC) decomposition-based approach (Goldsmith *et al.*, 2013), and the geometric approach by Choi and Reimherr (2018) in Hilbert spaces. Zhu *et al.* (2012) proposed SCCs for the regression coefficient functions for multivariate varying coefficient model for functional responses. Gu *et al.* (2014) and Chang *et al.* (2017) proposed the SCC for coefficient functions in the function-on-scalar regression model. However, there is scant literature on SCCs for imaging data or other more general two-dimensional (2D) functions. Although the geometric method in Choi and Reimherr (2018) can be used to construct SCCs in Hilbert spaces over rectangular domains, it doesn't work well for objects over complex domains with arbitrary shape, which are very common in biomedical imaging studies. In addition, the geometric method is conservative because it is essentially based on a modification of Scheffé's method.

In this article, we derive SCCs with exact coverage probability for the 2D functional mean function  $\mu(z)$ ,  $z \in \Omega$ , in (1) via the extreme value theory of Gaussian processes (Adler, 1990) and approximating mean functions with bivariate splines. Our simulation studies indicate that the proposed SCCs are computationally efficient and have the correct coverage probability for finite samples. We also show that the spline estimator and the accompanying SCC are asymptotically the same as if all the images are observed without noise.

Motivated by the need to statistically quantify the difference between two imaging data sets arising in medical imaging studies, we further consider two-sample inference and extend our SCC construction procedure to a two-sample problem. Specifically, we focus on constructing SCC for the difference of the mean functions from two independent samples. The comparison of mean functions is particularly useful for imaging analysis in some biomedical settings such as comparing imaging outcomes for groups randomized either to placebo or to active treatment. Any mean differences may be localized and irregularly shaped, and so an estimation method should be flexible enough to allow for such differences. The approach developed here allows comparison of treatments simultaneously across the entire domain of interest.

We organize our article as follows. Section 2 describes the BPS estimators, and establishes their asymptotic properties for imaging data. Section 3 proposes asymptotic pointwise confidence intervals and SCCs that are constructed based on the BPS estimators. In Section

4, we discuss how to estimate the unknown components involved in the SCC construction and other issues of implementation. Section 5 reports findings from a simulation study. In Section 6, we apply the proposed methods to two real brain imaging data sets. In Section 7, we conclude the article with some discussions. Proofs of the theoretical results and additional numerical results are provided in Supporting Information.

## 2 | MODELS AND ESTIMATION METHOD

In practice, the functional imaging response variable,  $Y_i(\cdot)$ , is only measured on a regular grid of pixels,  $z_j \in \Omega, j = 1, \dots, N$ . For notational simplicity, we let  $Y_{ij} = Y_i(z_j)$  be the imaging response of subject  $i$  at location  $j$ , and the actual data set consists of  $\{(Y_{ij}, z_j)\}, i = 1, \dots, n, j = 1, \dots, N$ , which can be modeled as

$$Y_{ij} = \mu(z_j) + \eta_i(z_j) + \sigma(z_j)\epsilon_{ij}. \tag{2}$$

### 2.1 | Bivariate spline basis approximation over triangulations

For model (2), we first consider the estimation of the mean function,  $\mu(\cdot)$ . Medical imaging data are typically observed on an irregular domain  $\Omega$ . We approximate the mean function in (2) by the bivariate splines that are piecewise polynomial functions over a 2D triangulated domain; see Lai and Wang (2013). In the following, we briefly introduce the techniques of triangulations and describe the BPS smoothing method.

Triangulation is an effective tool for handling data distributed on irregular regions with complex boundaries and/or interior holes. In the following, we use  $T$  to denote a triangle which is a convex hull of three points that are not collinear. A collection  $\mathcal{T} = \{T_1, \dots, T_M\}$  of  $M$  triangles is called a triangulation of  $\Omega = \cup_{m=1}^M T_m$  if any nonempty intersection between a pair of triangles in  $\mathcal{T}$  is either a shared vertex or a shared edge. Given a triangle  $T \in \mathcal{T}$ , let  $|T|$  be its longest edge length, then the size of  $\mathcal{T}$  is defined as  $|\mathcal{T}| := \max\{|T|, T \in \mathcal{T}\}$ , that is, the length of the longest edge of all triangles in  $\mathcal{T}$ .

For an integer  $r \geq 0$ , let  $\mathcal{C}^r(\Omega)$  be the collection of all  $r$ -th continuously differentiable functions over  $\Omega$ . Given a triangulation  $\mathcal{T}$ , let  $\mathcal{S}_d^r(\Delta) = \{s \in \mathcal{C}^r(\Omega) : s|_T \in \mathbb{P}_d(T), T \in \Delta\}$  be a spline space of degree  $d$  and smoothness  $r$  over triangulation  $\mathcal{T}$ , 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  $d$ . We use Bernstein basis polynomials to represent the bivariate splines. For any triangle  $T \in \mathcal{T}$  and any fixed point  $z \in \Omega$ , let  $b_1, b_2$  and  $b_3$  be the barycentric coordinates of  $z$  relative to  $T$ . Then, the Bernstein basis polynomials of degree  $d$  relative to triangle  $T$  are defined as  $B_{ijk}^{T;d}(z) = (i!j!k!)^{-1}d!b_1^i b_2^j b_3^k, i + j + k = d$ . Let  $\{B_m\}_{m \in \mathcal{M}}$  be the set of degree- $d$  bivariate Bernstein basis polynomials for  $\mathcal{S}_d^r(\Delta)$ , where  $\mathcal{M}$  stands for an index set of Bernstein basis polynomials. Denote by  $\mathbf{B}$  the evaluation matrix of Bernstein basis polynomials, where the  $j$ th row of  $\mathbf{B}$  is given by  $\mathbf{B}^T(z_j) = \{B_m(z_j), m \in \mathcal{M}\}$ , for  $j = 1, \dots, N$ . We can approximate the mean function  $\mu(z)$  by  $\mu(z) \approx \mathbf{B}^T(z)\boldsymbol{\gamma}$ , where  $\boldsymbol{\gamma}^T = (\gamma_m, m \in \mathcal{M})$  is

the spline coefficient vector. The above bivariate spline basis can be easily constructed via the R package BPST.

To define the penalized spline method, for any function  $g(\mathbf{z})$  and direction  $z_h$ ,  $h = 1, 2$ , let  $\nabla_{z_h}^v g(\mathbf{z})$  denote the  $v$ -th order derivative in the direction  $z_h$  at the point  $\mathbf{z}$ . We consider the following penalized least squares problem:  $\min_g \in S_d^r(\Delta) \sum_{i=1}^n \sum_{j=1}^N \{Y_{ij} - g(\mathbf{z}_j)\}^2 + \rho_n \mathcal{E}(g)$ , where  $\mathcal{E}(s) = \sum_{T \in \Delta} \int_T \sum_{i+j=2} \binom{2}{i} (\nabla_{z_1}^i \nabla_{z_2}^j s)^2 dz_1 dz_2$  is the roughness penalty, and  $\rho_n$  is the roughness penalty parameter. To meet the smoothness requirement of the splines, we need to impose some linear constraints on the spline coefficients  $\boldsymbol{\gamma}$ .  $\mathbf{H}\boldsymbol{\gamma} = \mathbf{0}$  to be specific. See Section B.2 of the Supplementary Material of Yu *et al.* (2019) for a simple example of  $\mathbf{H}$ . Thus, we have to minimize  $\sum_{i=1}^n \sum_{j=1}^N \{Y_{ij} - \mathbf{B}^T(\mathbf{z}_j)\boldsymbol{\gamma}\}^2 + \rho_n \boldsymbol{\gamma}^T \mathbf{P}\boldsymbol{\gamma}$ , subject to  $\mathbf{H}\boldsymbol{\gamma} = \mathbf{0}$ , where  $\mathbf{P}$  is the block diagonal penalty matrix satisfying  $\boldsymbol{\gamma}^T \mathbf{P}\boldsymbol{\gamma} = \varepsilon(\mathbf{B}\boldsymbol{\gamma})$ .

We first remove the constraint via QR decomposition of  $\mathbf{H}^T$ :  $\mathbf{H}^T = \mathbf{Q}\mathbf{R} = (\mathbf{Q}_1 \mathbf{Q}_2) \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{pmatrix}$ , where  $\mathbf{Q}$  is orthogonal and  $\mathbf{R}$  is upper triangular, the submatrix  $\mathbf{Q}_1$  is the first  $p$  columns of  $\mathbf{Q}$ , where  $p$  is the rank of  $\mathbf{H}$ , and  $\mathbf{R}_2$  is a matrix of zeros. Next, we reparametrize using  $\boldsymbol{\gamma} = \mathbf{Q}_2 \boldsymbol{\theta}$  for some  $\boldsymbol{\theta}$ , then it is guaranteed that  $\mathbf{H}\boldsymbol{\gamma} = \mathbf{0}$ . The minimization problem is thus converted to a conventional unrestricted penalized regression problem:

$$\sum_{i=1}^n \sum_{j=1}^N \{Y_{ij} - \tilde{\mathbf{B}}^T(\mathbf{z}_j)\mathbf{Q}_2 \boldsymbol{\theta}\}^2 + \rho_n \boldsymbol{\theta}^T \mathbf{Q}_2^T \mathbf{P}\mathbf{Q}_2 \boldsymbol{\theta}, \tag{3}$$

where  $\tilde{\mathbf{B}}(\mathbf{z}) = \mathbf{Q}_2^T \mathbf{B}(\mathbf{z})$ . Denote  $\bar{Y}_{\cdot, j} = n^{-1} \sum_{i=1}^n Y_{ij}$ ,  $\bar{\mathbf{Y}} = (\bar{Y}_{\cdot, 1}, \dots, \bar{Y}_{\cdot, N})^T$ ,  $\mathbf{U} = \mathbf{B}\mathbf{Q}_2$ , and  $\mathbf{D} = \mathbf{Q}_2^T \mathbf{P}\mathbf{Q}_2$ . Then, minimizing (3) is equivalent to minimizing

$$\begin{aligned} & \|\bar{\mathbf{Y}} - \mathbf{B}\mathbf{Q}_2 \boldsymbol{\theta}\|^2 + n^{-1} \rho_n \boldsymbol{\theta}^T \mathbf{Q}_2^T \mathbf{P}\mathbf{Q}_2 \boldsymbol{\theta} \\ & = \|\bar{\mathbf{Y}} - \mathbf{U}\boldsymbol{\theta}\|^2 + n^{-1} \rho_n \boldsymbol{\theta}^T \mathbf{D}\boldsymbol{\theta}, \end{aligned}$$

and the solution is given by  $\hat{\boldsymbol{\theta}} = \{\mathbf{U}^T \mathbf{U} + n^{-1} \rho_n \mathbf{D}\}^{-1} \mathbf{U}^T \bar{\mathbf{Y}}$ . Thus, the estimator of  $\boldsymbol{\gamma}$  and  $\mu(\cdot)$  are:  $\hat{\boldsymbol{\gamma}} = \mathbf{Q}_2 \hat{\boldsymbol{\theta}}$ ,  $\hat{\mu}(\mathbf{z}) = \mathbf{B}(\mathbf{z})^T \hat{\boldsymbol{\gamma}}$ .

### 2.2 | Functional principal component analysis

For the second component,  $\eta_i(\mathbf{z})$ , in model (2), we consider a spectral decomposition of its covariance function  $G_{\eta}(\mathbf{z}, \mathbf{z}')$ . Denote the eigenvalue and eigenfunction sequences of the covariance operator  $G_{\eta}(\mathbf{z}, \mathbf{z}')$  as  $\{\lambda_k\}_{k=1}^{\infty}$  and  $\{\psi_k(\mathbf{z})\}_{k=1}^{\infty}$ , in which  $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ ,  $\sum_{k=1}^{\infty} \lambda_k < \infty$ , and  $\{\psi_k\}_{k=1}^{\infty}$  form an orthonormal basis of  $L_2(\Omega)$ . It follows from spectral theory that  $G_{\eta}(\mathbf{z}, \mathbf{z}') = \sum_{k=1}^{\infty} \lambda_k \psi_k(\mathbf{z}) \psi_k(\mathbf{z}')$ . The  $i$ th stochastic process  $\{\eta_i(\mathbf{z}), \mathbf{z} \in \Omega\}$  allows the Karhunen-Loève  $L_2$  representation:  $\eta_i(\mathbf{z}) = \sum_{k=1}^{\infty} \xi_{ik} \phi_k(\mathbf{z})$ , where

$\phi_k(z) = (\lambda_k)^{1/2} \psi_k(z)$ , and the coefficients  $\xi_{ik}$ 's are uncorrelated random variables with mean 0 and  $E(\xi_{ik} \xi_{ik'}) = I k(= k')$ , referred to as the  $k$ th FPC score of the  $i$ th subject in classical functional principal component analysis (FPCA). Thus, the response measurements in (2) can be represented as follows:

$$Y_{ij} = \mu(z_j) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(z_j) + \sigma(z_j) \varepsilon_{ij}.$$

Next, we describe the method of estimating the FPCA: the variance-covariance function  $G_{\eta}(z, z')$  and its eigenvalues and eigenfunctions. For any  $i = 1, \dots, n, j = 1, \dots, N$ , let  $\hat{R}_{ij} = Y_{ij} - \hat{\mu}(z_j)$  be the residual. We estimate  $\eta_i(z)$  individually by employing the bivariate spline smoothing method to  $\{(\hat{R}_{ij}, z_j)\}_{j=1}^N$ . To be more specific, for each  $i = 1, \dots, n$ , we define the spline estimator of  $\eta_i(z)$  as  $\hat{\eta}_i(z) = \operatorname{argmin}_{g_i \in \mathcal{S}_{\Delta}^r(\Delta^*)} \sum_{j=1}^N \{\hat{R}_{ij} - g_i(z_j)\}^2 + \rho_n^* \mathcal{E}(g_i)$ , where the triangulation  $\Delta^*$  and smoothness penalty  $\rho_n^*$  may be different from those introduced in Section 2 when estimating  $\mu(z)$ . Next, define the estimator of  $G_{\eta}(z, z')$  as

$$\hat{G}_{\eta}(z, z') = n^{-1} \sum_{i=1}^n \hat{\eta}_i(z) \hat{\eta}_i(z'), \tag{4}$$

and we estimate the eigenfunctions  $\psi_k(\cdot)$  using the following eigenequations:

$$\int_{\Omega} \hat{G}_{\eta}(z, z') \hat{\psi}_k(z) dz = \hat{\lambda}_k \hat{\psi}_k(z'), \tag{5}$$

where  $\hat{\psi}_k$ 's are subject to  $\int_{\Omega} \hat{\psi}_k^2(z) dz = 1$  and  $\int_{\Omega} \hat{\psi}_k(z) \hat{\psi}_{k'}(z) dz = 0$  for  $k' < k$ . If  $N$  is sufficiently large, the left hand side of can be (5) approximated by  $\sum_{j=1}^N \hat{G}(z_j, z_j) \hat{\psi}_k(z_j) A(z_j)$ , where  $A(z_j)$  is the area of the pixel  $z_j$ .

### 2.3 | Theoretical properties of the estimators

We investigate the asymptotic properties of the proposed spline estimators. To discuss these properties, we introduce some notation first. For any function  $g$  over the closure of domain  $\Omega$ , denote by  $\|g\|_{L_2(\Omega)}^2 = \int_{\Omega} g^2(z) dz$  the regular  $L_2$  norm of  $g$ , and by  $\|g\|_{\infty, \Omega} = \sup_{z \in \Omega} |g(z)|$  the supremum norm of  $g$ . Let  $\|g\|_{v, \infty, \Omega} = \max_{i+j=v} \|\nabla_{z_1}^i \nabla_{z_2}^j g\|_{\infty, \Omega}$  be the maximum norms of all the  $v$ th order derivatives of  $g$  over  $\Omega$ . For notational simplicity, we suppress the subscript  $\Omega$  below. Given random variables  $S_n$  for  $n \geq 1$ , we write  $S_n = O_p(b_n)$  if  $\lim_{c \rightarrow \infty} \limsup_n P(|S_n| \geq cb_n) = 0$ . Similarly, we write  $S_n = o_p(b_n)$  if  $\lim_n P(|S_n| \geq cb_n) = 0$ , for any constant  $c > 0$ .

The following theorem provides the  $L_2$  and uniform convergence rate of  $\hat{\mu}(\cdot)$ . The detailed proofs of this theorem are given in Web Appendix B.3 of Supporting Information.

**Theorem 1.** *Suppose Assumptions (A1) to (A4) in Web Appendix B of Supporting Information hold, and  $N^{1/2}|\Delta| \rightarrow \infty$  as  $N \rightarrow \infty$ . Then the bivariate penalized spline estimator  $\hat{\mu}(\cdot)$  is consistent and satisfies*

$$\begin{aligned} \|\hat{\mu} - \mu\|_{L_2} = O_P \left\{ \frac{\rho_n}{nN|\Delta|^3} \|\mu\|_{2,\infty} \right. \\ \left. + \left( 1 + \frac{\rho_n}{nN|\Delta|^5} \right) |\Delta|^{d+1} \|\mu\|_{d+1,\infty} \right. \\ \left. + \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{nN}|\Delta|} \right\}. \end{aligned}$$

*In addition, if Assumptions (A1) to (A5) hold, we have  $\|\hat{\mu} - \mu\|_\infty = o_P\{(n^{-1}\log(n))^{1/2}\}$  and  $\|\hat{\mu} - \mu\|_{L_2} = O_P(n^{-1/2})$ .*

Theorem 2 characterizes the uniform weak convergence of  $\hat{G}_\eta(\mathbf{z}, \mathbf{z}')$  and the convergence of  $\hat{\psi}_k$  and  $\hat{\lambda}_k$ .

**Theorem 2.** *Under Assumptions (A1) to (A7) in Web Appendix B of Supporting Information, we have the following results: (a) The spline estimator  $\hat{G}_\eta(\mathbf{z}, \mathbf{z}')$  in (4) uniformly converges to  $G_\eta(\mathbf{z}, \mathbf{z}')$  in probability, that is,*

*$\sup_{(\mathbf{z}, \mathbf{z}') \in \Omega^2} |\hat{G}_\eta(\mathbf{z}, \mathbf{z}') - G_\eta(\mathbf{z}, \mathbf{z}')| = o_P(1)$ ; (b)  $\|\hat{\psi}_k - \psi_k\| = o_P(1)$ ,  $|\hat{\lambda}_k - \lambda_k| = o_P(1)$ , for  $k = 1, \dots, \kappa$ .*

Although, in theory, the Karhunen-Loève representation of the covariance function consists of an infinite number of terms. In applications, it is typical to truncate the spectral decomposition to an integer chosen so as to account for some predetermined proportion of the variance. One can select the number of principal components using the Akaike information criterion (AIC; Yao *et al.*, 2005) or Bayesian information criterion (BIC; Li *et al.*, 2013).

### 3 | SIMULTANEOUS CONFIDENCE CORRIDORS

#### 3.1 | One sample

Let  $G_\eta(\cdot, \cdot)$  be a positive definite function defined as  $G_\eta(\mathbf{z}, \mathbf{z}') = \sum_{k=1}^{\kappa} \lambda_k \psi_k(\mathbf{z}) \psi_k(\mathbf{z}')$ ,  $\mathbf{z}, \mathbf{z}' \in \Omega$ . Denote by  $\zeta(\mathbf{z})$ ,  $\mathbf{z} \in \Omega$  a standardized Gaussian process such that  $E\zeta(\mathbf{z}) = 0$ ,  $E\zeta^2(\mathbf{z}) = 1$  with covariance function  $E\zeta(\mathbf{z})\zeta(\mathbf{z}') = G_\eta(\mathbf{z}, \mathbf{z}')\{G_\eta(\mathbf{z}, \mathbf{z})G_\eta(\mathbf{z}', \mathbf{z}')\}^{-1/2}$ ,  $\mathbf{z}, \mathbf{z}' \in \Omega$ . Denote by  $q_{1-\alpha}$  the  $100(1-\alpha)$ th percentile of the distribution of the absolute maximum of  $\zeta(\mathbf{z})$ ,  $\mathbf{z} \in \Omega$ , that is  $P\{\sup_{\mathbf{z} \in \Omega} |\zeta(\mathbf{z})| \leq q_{1-\alpha}\} = 1 - \alpha$ ,  $\alpha \in (0, 1)$ .

Define the ‘‘oracle’’ estimator  $\bar{\mu}(\mathbf{z}) = \mu(\mathbf{z}) + n^{-1} \sum_{i=1}^n \eta_i(\mathbf{z})$ , which is infeasible due to the finite pixel grid  $\{\mathbf{z}_j; j = 1, \dots, N\}$  and the measurement error. The following theorem presents the asymptotic properties of  $\bar{\mu}(\mathbf{z})$  and shows that the difference between the BPS estimator  $\hat{\mu}(\mathbf{z})$  and the ‘‘oracle’’ smoother  $\bar{\mu}(\mathbf{z})$  is uniformly bounded at an  $o_P(n^{1/2})$  rate.



**Theorem 3.** Under Assumptions (A1) to (A6) in Web Appendix B of Supporting Information, for any  $\alpha \in (0,1)$ , as  $N \rightarrow \infty$ ,  $n \rightarrow \infty$ ,

$$P \left\{ \sup_{z \in \Omega} n^{1/2} |\bar{\mu}(z) - \mu(z)| G_{\eta}(z, z)^{-1/2} \leq q_{1-\alpha} \right\} \rightarrow 1 - \alpha,$$

$$\text{and } \sup_{z \in \Omega} |\bar{\mu}(z) - \hat{\mu}(z)| = o_P(n^{-1/2}).$$

Based on Theorems 1 and 3, we obtain the following asymptotic SCCs for  $\mu(z)$ ,  $z \in \Omega$ .

**Corollary 1.** Under the assumptions of Theorem 3, for any  $\alpha \in (0,1)$ , as  $N \rightarrow \infty$ ,  $n \rightarrow \infty$ , an asymptotic  $100(1 - \alpha)\%$  exact SCC for  $\mu(z)$  is  $\hat{\mu}(z) \pm n^{-1/2} q_{1-\alpha} G_{\eta}(z, z)^{1/2}$ .

### 3.2 | Extension to two-sample case

While one-sample SCCs are of primary interest in many situations, in some brain imaging analysis, interest lies in comparing two groups, for example, patients and normal control subjects. Next, we extend our method to two-sample problems, constructing SCCs for the difference between mean functions from two independent groups, analogous to a two-sample  $t$  test.

Given two groups of imaging observations with sample sizes  $n_1$  and  $n_2$ , respectively, defined on a common region  $\Omega$ . For  $H = 1, 2$ , let  $G_{H\eta}(z, z') = \sum_{k=1}^{KH} \phi_{Hk}(z) \phi_{Hk}(z')$  be a positive definite function and  $\hat{\mu}_H$  be the spline estimates for the group mean function  $\mu_H$ . Let  $V(z, z') = G_{1\eta}(z, z') + \tau G_{2\eta}(z, z')$ , where  $\tau = \lim_{n_1 \rightarrow \infty} n_1/n_2$ . Denote by  $W(z)$ ,  $z \in \Omega$ , a standardized Gaussian process such that  $EW(z) = 0$ ,  $EW^2(z) = 1$  with covariance  $E[V(z, z')^{-1/2} V(z, z') \{V(z', z')\}^{-1/2}] = \{V(z, z')\}^{-1/2} V(z, z') \{V(z', z')\}^{-1/2}$ . Denote  $q_{12,\alpha}$  the  $(1 - \alpha)$ th quantile of the absolute maximal distribution of  $W(z)$ ,  $z \in \Omega$ .

**Theorem 4.** Under Assumptions (A1) to (A6) in Web Appendix B of Supporting Information, for any  $\alpha \in (0,1)$ , as  $N \rightarrow \infty$ ,  $n_1 \rightarrow \infty$ ,

$$P \left\{ \sup_{z \in \Omega} \frac{n_1^{1/2} |(\hat{\mu}_1 - \hat{\mu}_2)(z) - (\mu_1 - \mu_2)(z)|}{\sqrt{V(z, z)}} \leq q_{12,\alpha} \right\} \rightarrow 1 - \alpha.$$

Theorem 4 suggests that an asymptotic  $100(1 - \alpha)\%$  exact SCC for  $(\mu_1 - \mu_2)(z)$  can be constructed as  $(\hat{\mu}_1 - \hat{\mu}_2)(z) \pm n_1^{-1/2} q_{12,\alpha} \{V(z, z)\}^{1/2}$ .

## 4 | IMPLEMENTATION

Without loss of generality, we describe the implementation of the proposed SCCs for the one-sample case. The procedure can be similarly adopted to the two-sample mean cases.



#### 4.1 | Quantile estimation and smoothing parameter selection

The quantile  $q_{1-\alpha}$  used to construct the SCCs in Corollary 1 cannot be obtained analytically; however, it can be approximated by numerical simulation as follows: first, we simulate

$\zeta_b(z) = \hat{G}_\eta^{-1/2}(z, z) \sum_{k=1}^{\kappa} \hat{\lambda}_k^{1/2} Z_{k,b} \hat{\psi}_k(z)$ , where  $Z_{k,b}$  are i.i.d standard normal variables with  $1 \leq k \leq \kappa$  and  $b = 1, \dots, B$  for a preset large integer  $B$ . Then, we estimate the quantile  $q_{1-\alpha}$  by the corresponding empirical quantile of these maximum values by taking the maximal absolute value for each copy of  $\zeta_b(z)$ . To construct the SCC for the two-sample case, denote  $\hat{V}(z, z') = \hat{G}_{1\eta}(z, z') + \tau \hat{G}_{2\eta}(z, z')$ . We simulate

$$\hat{W}_b(z) = \left\{ \hat{V}(z, z) \right\}^{-1/2} \left\{ \sum_{k=1}^{\kappa_1} \hat{\lambda}_{1k}^{1/2} Z_{1k,b} \hat{\psi}_{1k}(z) - (n_1/n_2)^{1/2} \sum_{k=1}^{\kappa_2} \hat{\lambda}_{2k}^{1/2} Z_{2k,b} \hat{\psi}_{2k}(z) \right\}, z \in \Omega.$$

Then,  $q_{12,\alpha}$  can be estimated by the empirical quantile of the  $B$  simulated  $\|\hat{W}_b\|_\infty$ 's,  $b = 1, \dots, B$ .

Next, for a good fit of the data, it is necessary to choose a suitable value of the smoothing parameter  $\rho_n$ . A large value of  $\rho_n$  enforces a smoother fitted function with larger fitting errors, while a small  $\rho_n$  may result in overfitting of the data. Since the in-sample fitting errors cannot gauge the prediction accuracy of the fitted function, we select a criterion function that attempts to measure the out-of-sample performance of the fitted model. Minimizing the generalized cross-validation (GCV) criterion is one computationally efficient approach to selecting smoothing parameters that also has good theoretical properties. We choose the smoothing parameter by minimizing the following

$$\text{GCV}(\rho_n) = \|\bar{Y} - \mathbf{S}(\rho_n)\bar{Y}\|^2 / \left[ N \{1 - \text{tr}\{\mathbf{S}(\rho_n)\}/N\}^2 \right] \text{ over a grid of values of } \rho_n, \text{ where } \mathbf{S}(\rho_n) = \mathbf{U}(\mathbf{U}^T\mathbf{U} + n^{-1}\rho_n\mathbf{D})^{-1}\mathbf{U}^T.$$

#### 4.2 | Spline basis and triangulation selection

To construct the SCC, we need to choose the spline basis functions and triangulation used in the BPS, a notoriously difficult task for constructing nonparametric pointwise confidence intervals or simultaneous confidence bands.

When the resolution of the imaging is relatively high and the mean imaging seems to be a realization from some smooth function without sharp edges, we suggest using smooth parameter  $r = 1$  with degree  $d \geq 4$ . When  $d \geq 5$ , the proposed spline achieves full estimation power asymptotically (Lai and Wang, 2013). It is generally believed that subject-level image variation  $\eta_S$  are less smooth than the mean function. Thus, we suggest considering lower order splines, such as  $d = 2$ , when estimating the  $\eta_S$ .

An optimal triangulation is a partition of the domain which is best according to some criterion that measures the shape, size or number of triangles. For example, a ‘‘good’’ triangulation usually refers to those with well-shaped triangles, no small angles or/and no obtuse angles. We suggest building the triangulated meshes using typical triangulation construction methods such as Delaunay Triangulation (De Loera *et al.*, 2010). The Matlab code DistMesh and R package Triangulation can be used to construct the triangulation. When estimating the mean function  $\mu(\cdot)$ , we suggest choosing the triangulation  $\mu$  based on leave-images-out  $k$ -fold cross-validation (CV). In the estimation of the  $\eta_j(\cdot)$ 's, we suggest

choosing the triangulation  $\eta$  so as to minimize a bootstrap estimator of the coverage error of the SCCs. In Algorithm A1 in Supporting Information, we describe our selection scheme for the one-sample case, which can be extended straightforwardly to the two-sample case.

### 4.3 | Variance estimation for measurement errors and SCC adjustment

For certain imaging types and modalities, our Assumptions (A2) and (A3) about the measurement errors may not be completely satisfied. We propose a modification to the SCC procedure in Section 3 to deal with images with relatively large measurement errors.

For the one-sample SCC, for any  $j = 1, \dots, N$ , let  $\hat{\epsilon}_{ij} = \hat{R}_{ij} - \hat{\eta}_i(z_j)$ , and we estimate  $\sigma^2(z_j)$  by  $\hat{\sigma}^2(z_j) = n^{-1} \sum_{i=1}^n \hat{\epsilon}_{ij} \hat{\epsilon}_{ij}$ . Next, denote  $\hat{\epsilon}(z) = (nN)^{-1} \tilde{\mathbf{B}}(z)^T \Gamma_{N,\rho}^{-1} \sum_{i=1}^n \sum_{j=1}^N \tilde{\mathbf{B}}(z_j) \sigma(z_j) \epsilon_{ij}$ . We estimate the variance-covariance function of  $\hat{\epsilon}(z)$ ,  $\tilde{G}_\epsilon(z, z') = \text{Cov}\{\hat{\epsilon}(z), \hat{\epsilon}(z')\}$ , by

$$\hat{G}_\epsilon(z, z') = n^{-1} N^{-2} \tilde{\mathbf{B}}(z)^T \Gamma_{N,\rho}^{-1} \left\{ \sum_{j=1}^N \tilde{\mathbf{B}}(z_j) \hat{\sigma}^2(z_j) \tilde{\mathbf{B}}(z_j)^T \right\} \times \Gamma_{N,\rho}^{-1} \tilde{\mathbf{B}}(z'),$$

where  $\Gamma_{N,\rho}$  is given in (B.5) in Supporting Information.

Denote  $\hat{\Sigma}(z, z') = \hat{G}_\eta(z, z') + n \hat{G}_\epsilon(z, z')$ . We adjust the approximation procedure of quantile  $q_{1-\alpha}$  as follows: first, we simulate

$$\zeta_b(z) = \hat{\Sigma}^{-1/2}(z, z) \left\{ \sum_{k=1}^{\kappa} \hat{\lambda}_k^{1/2} \hat{\psi}_k(z) Z_{k,\xi}^{(b)} + N^{-1} \tilde{\mathbf{B}}(z)^T \Gamma_{N,\rho}^{-1} \sum_{j=1}^N \tilde{\mathbf{B}}(z_j) \hat{\sigma}(z_j) Z_{j,\epsilon}^{(b)} \right\},$$

where  $Z_{k,\xi}^{(b)}$  and  $Z_{j,\epsilon}^{(b)}$  are i.i.d standard normal variables with  $1 \leq k \leq \kappa$ ,  $1 \leq j \leq N$ ; next, we estimate the quantile  $q_{1-\alpha}$  by the corresponding empirical quantile of the  $B$  simulated  $\|\zeta_b\|_\infty$ ; finally, we construct the SCC as  $\hat{\mu}(z) \pm n^{-1/2} q_{1-\alpha} \hat{\Sigma}(z, z)^{1/2}$ ,  $z \in \Omega$ .

For the two-sample case, we can similarly modify the procedure by defining  $\hat{\Sigma}_H(z, z') = \hat{G}_{\eta,H} + n_H \hat{G}_{\epsilon,H}$ , for  $H = 1, 2$ , and  $\hat{\Xi}(z, z') = \hat{\Sigma}_1(z, z') + n_1/n_2 \hat{\Sigma}_2(z, z')$ . Let  $\hat{\sigma}_H(z)$  be the estimator of  $\sigma_H(z)$ , for  $H = 1, 2$ . To estimate  $q_{12,\alpha}$ , we simulate

$$\begin{aligned} \widehat{W}_b(\mathbf{z}) = & \left\{ \widehat{\Xi}(\mathbf{z}, \mathbf{z}) \right\}^{-1/2} \left\{ \sum_{k=1}^{\kappa_2} \widehat{\lambda}_{1k}^{1/2} Z_{1k, \xi}^{(b)} \widehat{\psi}_{1k}(\mathbf{z}) \right. \\ & - \left( \frac{n_1}{n_2} \right)^{1/2} \sum_{k=1}^{\kappa_2} \widehat{\lambda}_{2k}^{1/2} Z_{2k, \xi}^{(b)} \widehat{\psi}_{2k}(\mathbf{z}) \\ & + \widetilde{\mathbf{B}}(\mathbf{z})^T \Gamma_{N, \rho_1}^{-1} \frac{1}{N} \sum_{j=1}^N \widetilde{\mathbf{B}}(\mathbf{z}_j) \widehat{\sigma}_1(\mathbf{z}_j) Z_{1j, \varepsilon}^{(b)} \\ & \left. - \left( \frac{n_1}{n_2} \right)^{1/2} \widetilde{\mathbf{B}}(\mathbf{z})^T \Gamma_{N, \rho_2}^{-1} \frac{1}{N} \sum_{j=1}^N \widetilde{\mathbf{B}}(\mathbf{z}_j) \widehat{\sigma}_2(\mathbf{z}_j) Z_{2j, \varepsilon}^{(b)} \right\}, \end{aligned}$$

where  $Z_{Hk, \xi}^{(b)}$  and  $Z_{Hj, \varepsilon}^{(b)}$  are i.i.d standard normal variables with  $1 \leq k \leq \kappa_H$ ,  $1 \leq j \leq N$  for  $H = 1, 2$ . Then,  $q_{12, \alpha}$  can be estimated by the empirical quantile of the  $B$  simulated  $\|\widehat{W}_b\|_\infty$ 's,  $b = 1, \dots, B$ . A modified SCC for  $\mu_1(\mathbf{z}) - \mu_2(\mathbf{z})$  can thus be constructed as  $(\widehat{\mu}_1 - \widehat{\mu}_2)(\mathbf{z}) \pm n_1^{-1/2} q_{12, \alpha} \left\{ \widehat{\Xi}(\mathbf{z}, \mathbf{z}) \right\}^{1/2}$ .

## 5 | SIMULATION STUDIES

In this section, we describe two Monte Carlo simulations to examine the finite sample performance of the proposed method.

### 5.1 | One sample SCC

In this simulation study, the measurements on the images are generated from the model:

$$Y_{ij} = \mu(\mathbf{z}_j) + \sum_{k=1}^2 \lambda_k^{1/2} \xi_{ij} \psi_k(\mathbf{z}_j) + \sigma(\mathbf{z}_j) \varepsilon_{ij}, \quad i = 1, \dots, n, \\ j = 1, \dots, N,$$

where  $\mathbf{z}_j = (z_{1j}, z_{2j}) \in \Omega \subset [0, 1]^2$ , and  $\Omega$  is the same as the domain of the brain images shown in Section 6. To demonstrate the practical performance of our theoretical results, we consider the following four mean functions:

- (quadratic)  $\mu(\mathbf{z}) = 20\{(z_1 - 0.5)^2 + (z_2 - 0.5)^2\}$ ,
- (exponential)  $\mu(\mathbf{z}) = 5 \exp[-15\{(z_1 - 0.5)^2 + (z_2 - 0.5)^2\}] + 0.5$ ,
- (cubic)  $\mu(\mathbf{z}) = 3.2(-z_1^3 + z_2^3) + 2.4$ ,
- (sine)  $\mu(\mathbf{z}) = -10[\sin\{5\pi(z_1 + 0.22)\} - \sin\{5\pi(z_2 - 0.18)\}] + 2.8$ ,

shown in the first column of Figures A1 to A6 in Supporting Information.

To simulate the within-image dependence, we generate  $\xi_{ik} \stackrel{\text{i.i.d}}{\sim} N(0, 1)$ , for  $k = 1, 2$ . For the eigenvalues, we set  $\lambda_1 = 0.5$ ,  $\lambda_2 = 0.2$ . For the eigenfunctions, we let  $\psi_1(\mathbf{z}) = c_1 \sin(\pi z_1) + c_2$ ,  $\psi_2(\mathbf{z}) = c_3 \cos(\pi z_2) + c_4$ , where  $c_1 = 0.988$ ,  $c_2 = 0.5$ ,  $c_3 = 2.157$ , and  $c_4 = -0.084$  to guarantee that the eigenfunctions are orthonormal. We generate heterogenous measurement errors with  $\sigma(\mathbf{z}) = 0.25\{1 - (z_1 - 0.5)^2 - (z_2 - 0.5)^2\}$ . We consider  $n = 50, 100, 200$ , and for

each image, we consider two types of resolution:  $40 \times 40$  and  $79 \times 79$  with  $N = 921$  and 3682 pixels falling inside the domain, respectively.

To apply our method, we consider three different triangulations which are also shown in the first column of Figures A1 to A6 in Supporting Information. The first triangulation (  $\tau_1$  ) contains 49 triangles and 38 vertices; the second triangulation (  $\tau_2$  ) contains 80 triangles and 54 vertices; while the third triangulation (  $\tau_3$  ) contains 144 triangles and 87 vertices. The estimated mean function based on these three triangulations are shown in the second columns of Figures A1 to A6, and the corresponding 99% SCCs are given in the last two columns. From these figures, one can see that all three triangulations result in almost the same estimates and SCCs. One can also see that even when the number of images is moderately large, the estimation is very accurate regardless of the type of underlying mean functions.

Table 1 and Table A1 in Supporting Information summarize the estimated coverage rate of the SCCs based on 1000 replications for  $N = 921$  and 3682, respectively. The number in parenthesis represents the average width of the SCCs. These two tables also confirm that there is little difference among the three triangulations and that the coverage rate is closer to the nominal confidence level for larger values of  $n$ .

## 5.2 | Two-sample simultaneous confidence corridor

In this simulation study, we examine the power of detecting a difference in mean images based on the proposed two-sample SCC. Two group of images are generated from the model

$$Y_{H,ij} = \mu_H(\mathbf{z}_j) + \sum_{k=1}^K \lambda_k^{1/2} \xi_{ij} \psi_k(\mathbf{z}_j) + \sigma(\mathbf{z}_j) \varepsilon_{ij}, \quad H = 1, 2,$$

where  $\psi_k$ 's are generated as in the simulation in Section 5.1. We consider the following:

$$\begin{aligned} H_0: \mu_1(\mathbf{z}) &= \mu_2(\mathbf{z}), \quad \text{for all } \mathbf{z} \in \Omega \quad \text{vs} \\ H_a: \mu_1(\mathbf{z}) &\neq \mu_2(\mathbf{z}) \quad \text{for some } \mathbf{z} \in \Omega. \end{aligned} \quad (6)$$

The mean functions for two groups considered here are  $\mu_1(\mathbf{z}) = 20\{(z_1 - 0.5)^2 + (z_2 - 0.5)^2\}$ , and  $\mu_2(\mathbf{z}) = \mu_1(\mathbf{z}) + \delta(-z_1^3 + z_2^3)$ . The value of  $\delta$  controls the difference between the two groups. The eigenvalues  $\lambda_k$ 's, eigenfunctions  $\psi_k$ 's and the measurement errors  $\varepsilon_{ij}$ 's are generated in the same way as in the simulation presented in Section 5.1, and we set  $\sigma(\mathbf{z}) = 0.1$ .

Figure 1 and Table A2 in Supporting Information summarize the estimated probability of rejecting  $H_0$  in (6) with nominal level  $\alpha = 0.10, 0.05, \text{ and } 0.01$ . When  $\delta = 0$ , the probability should be close to the nominal level, and when  $\delta$  is large, the estimated power should be close to 1. From Figure 1 and Table A2, one can see even when the numbers of the images  $n_1$  and  $n_2$  are moderately large, the size of the test is very close to the nominal level. The estimated power increases quickly as  $n_1$  and  $n_2$  increase. The performance of the procedure is similar and consistent for different triangulations.

## 6 | APPLICATIONS TO BRAIN IMAGING DATA

In this section, we implement the proposed SCCs to analyze brain imaging data. In particular, we consider data taken from positron emission tomography (PET) studies with two different settings: one using the tracer  $[C^{11}]WAY100635$  that has an affinity for the serotonin 1 A receptor in a study of major depressive disorder (MDD); and one using the fluorodeoxyglucose tracer  $[F^{18}]FDG$ , a glucose analog, in a study of dementia. The imaging data are naturally three-dimensional in each case, but we focus here on one strategically selected slice in each setting. For the MDD study, we select the horizontal slice which passes through the midbrain and the amygdala, two regions implicated in MDD (Parsey *et al.*, 2010). As pointed out by Marcus *et al.* (2014), within the brain, the anatomical regions that are commonly affected by Alzheimer diseases are the bilateral superior medial frontal, anterior, middle cingulate and bilateral parietal cortices, while the regions such as the bilateral medial temporal lobes are usually less affected. Therefore, for the  $[F^{18}]FDG$  study, we focus on the 48th horizontal slice of the brain since it passes through the frontal and parietal lobes. In each case, we consider the hypotheses in (6) for the difference between two mean functions.

For the  $[C^{11}]WAY100635$  data, we have 40 subjects who are classified as normal controls and 26 who have been diagnosed with MDD (Parsey *et al.*, 2006). Figure 2 displays the results of the application of the proposed procedure to these data. The portions of the SCCs not containing zero can be seen in (A); the estimation of the mean difference between towards the two groups is shown in (B), and the lower and upper SCCs are shown in (C) and (D).

Next, we illustrate these procedures using the PET data from the Alzheimer's Disease Neuroimaging Initiative (ADNI; [adni.loni.usc.edu](http://adni.loni.usc.edu)). One of the primary goals of the ADNI study is to test whether PET and some other biological markers can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

We use the proposed method in Section 4.2 to choose the triangulation and spline basis functions. Among the three triangulations (1–3) considered in simulation studies, we choose 3 when estimating the mean functions, and 1 when estimating the covariance functions. We use smooth parameter  $r = 1$  with degree  $d = 5$  for the estimation of mean function and  $d = 2$  for the estimation of  $\eta_j$ 's. The first row of Figure 3 displays the areas in which zero is not contained within the 95% SCC comparing each pair of diagnostic groups. This suggests that the AD group has widespread mean differences from each of the other two groups. We also stratify the data according to sex and age, and the breakdowns of these data in terms of these variables are given in Table A1 in Supporting Information. Within each stratum, we examine the SCC for the difference between all pairs of diagnostic groups, and the results are also shown in Figure 3. The large apparent differences in the full group analysis can be seen (but to a lesser extent) in the comparisons among the males and among the relatively younger population, but are less pronounced in the other subgroup analyses.

## 7 | DISCUSSION

We develop SCCs for mean functions of imaging data in the functional data framework. We show that the proposed procedure has desirable statistical properties: the estimators are asymptotically efficient as if all images were observed with no error. One main advantage of our method is its computational efficiency and feasibility for large-scale imaging data. It greatly enhances the application of SCCs to imaging data in biomedical studies.

A few more issues still merit further research. For instance, the triangulation selection using the CV and wild bootstrap works well in practice, but a stronger theoretical justification for their use is still needed. In recent years, there has been a great deal of work on functional regression. It is interesting to extend the proposed methodology to functional regression models. The construction of SCCs in such models is a significant challenge and requires more in-depth investigation. Last but not least, it is also interesting to develop SCCs for large-scale longitudinal imaging data, in which accounting for the dependence within the subject as well as for the longitudinal design is crucial for making inference.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### ACKNOWLEDGMENTS

The authors are truly grateful to the editor, the associate editor, and two reviewers for their constructive suggestions that led to significant improvement of the article. Li Wang's research was supported in part by NSF awards DMS-1916204 and DMS-1542332. Todd Ogden's work was partially supported by NIH grants 5 R01 EB024526 and 2 P50 MH090964. Data used in preparation of this article were obtained from the ADNI database ([adni.loni.usc.edu](http://adni.loni.usc.edu)). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or 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](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

Funding information

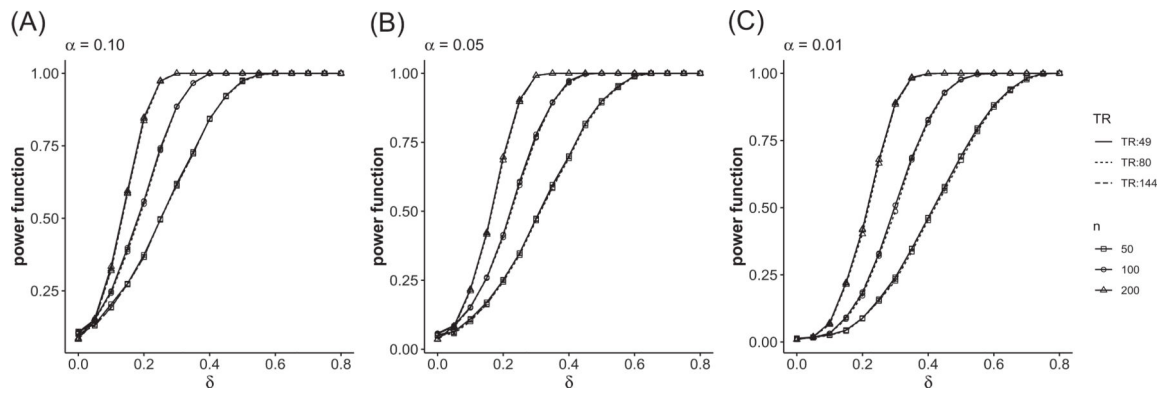
National Institute of Biomedical Imaging and Bioengineering, Grant/Award Number: EB024526; National Institute of Mental Health, Grant/Award Number: MH090964; Division of Mathematical Sciences, Grant/Award Numbers: 1542332, 1916204

### REFERENCES

- Adler RJ (1990) An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes. Hayward, CA: Institute of Mathematical Statistics.
- Adler RJ and Taylor JE (2007) Random Fields and Geometry. New York: Springer.
- Cao G and Wang L (2018) Simultaneous inference for the mean of repeated functional data. *Journal of Multivariate Analysis*, 165, 279–295.
- Cao G, Yang L and Todem D (2012) Simultaneous inference for the mean function based on dense functional data. *Journal of Nonparametric Statistics*, 24, 359–377. [PubMed: 22665964]
- Chang C, Lin X and Ogden RT (2017) Simultaneous confidence bands for functional regression models. *Journal of Statistical Planning and Inference*, 188, 67–81.
- Choi H and Reimherr M (2018) A geometric approach to confidence regions and bands for functional parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80, 239–260.

- De Loera JA, Rambau J and Santos F (2010) *Triangulations Structures for Algorithms and Applications*. Berlin: Springer.
- Degras DA (2011) Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica*, 21, 1735–1765.
- Degras DA (2017) Simultaneous confidence bands for the mean of functional data. *Wiley Interdisciplinary Reviews: Computational Statistics*, 9, e1397.
- Forman SD, Cohen JD, Fitzgerald M, Eddy WF, Mintun MA and Noll DC (1995) Improved assessment of significant activation in functional magnetic resonance imaging (fMRI): use of a clustersize threshold. *Magnetic Resonance in medicine*, 33, 636–647. [PubMed: 7596267]
- Goldsmith J, Greven S and Crainiceanu C (2013) Corrected confidence bands for functional data using principal components. *Biometrics*, 69, 41–51. [PubMed: 23003003]
- Gu L, Wang L, Härdle WK and Yang L (2014) A simultaneous confidence corridor for varying coefficient regression with sparse functional data. *Test*, 23, 806–843.
- Lai MJ and Wang L (2013) Bivariate penalized splines for regression. *Statistica Sinica*, 23, 1399–1417.
- Li Y, Wang N and Carroll RJ (2013) Selecting the number of principal components in functional data. *Journal of the American Statistical Association*, 108, 1284–1294.
- Marcus C, Mena E and Subramaniam RM (2014) Brain pet in the diagnosis of Alzheimer’s disease. *Clinical Nuclear Medicine*, 39, e413. [PubMed: 25199063]
- Parsey RV, Ogden RT, Miller JM, Tin A, Hesselgrave N and Goldstein E (2010) Higher serotonin 1a binding in a second major depression cohort: modeling and reference region considerations. *Biological Psychiatry*, 68, 170–178. [PubMed: 20497898]
- Parsey RV, Oquendo MA, Ogden RT, Olvet DM, Simpson N, Huang YY et al. (2006) Altered serotonin 1a binding in major depression: a [carbonyl-c-11] way100635 positron emission tomography study. *Biological Psychiatry*, 59, 106–113. [PubMed: 16154547]
- Siegmund D, Zhang N and Yakir B (2011) False discovery rate for scanning statistics. *Biometrika*, 98, 979–985.
- Worsley KJ, Taylor JE, Tomaiuolo F and Lerch J (2004) Unified univariate and multivariate random field theory. *Neuroimage*, 23, S189–S195. [PubMed: 15501088]
- Yao F, Müller H-G and Wang J-L (2005) Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100, 577–590.
- Yu S, Wang G, Wang L, Liu C and Yang L (2019) Estimation and inference for generalized geoadditive models. *Journal of the American Statistical Association*, 1–27.
- Zheng S, Yang L and Härdle WK (2014) A smooth simultaneous confidence corridor for the mean of sparse functional data. *Journal of the American Statistical Association*, 109, 661–673.
- Zhu H, Li R and Kong L (2012) Multivariate varying coefficient model for functional responses. *The Annals of Statistics*, 40, 2634–2666. [PubMed: 23645942]





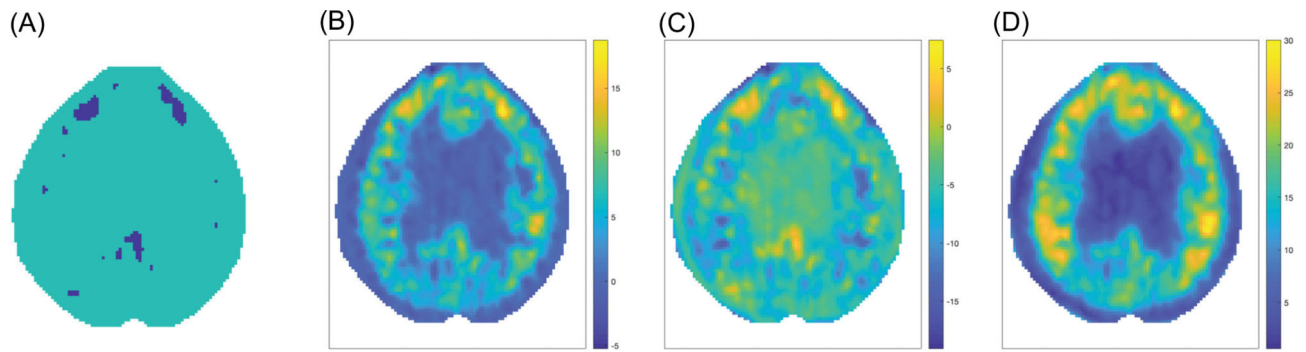
**FIGURE 1.** Type I error and empirical power of two-sample tests for different  $\alpha$ 's: (A)  $\alpha = 0.10$ , (B)  $\alpha = 0.05$ , and (C)  $\alpha = 0.01$

Author Manuscript

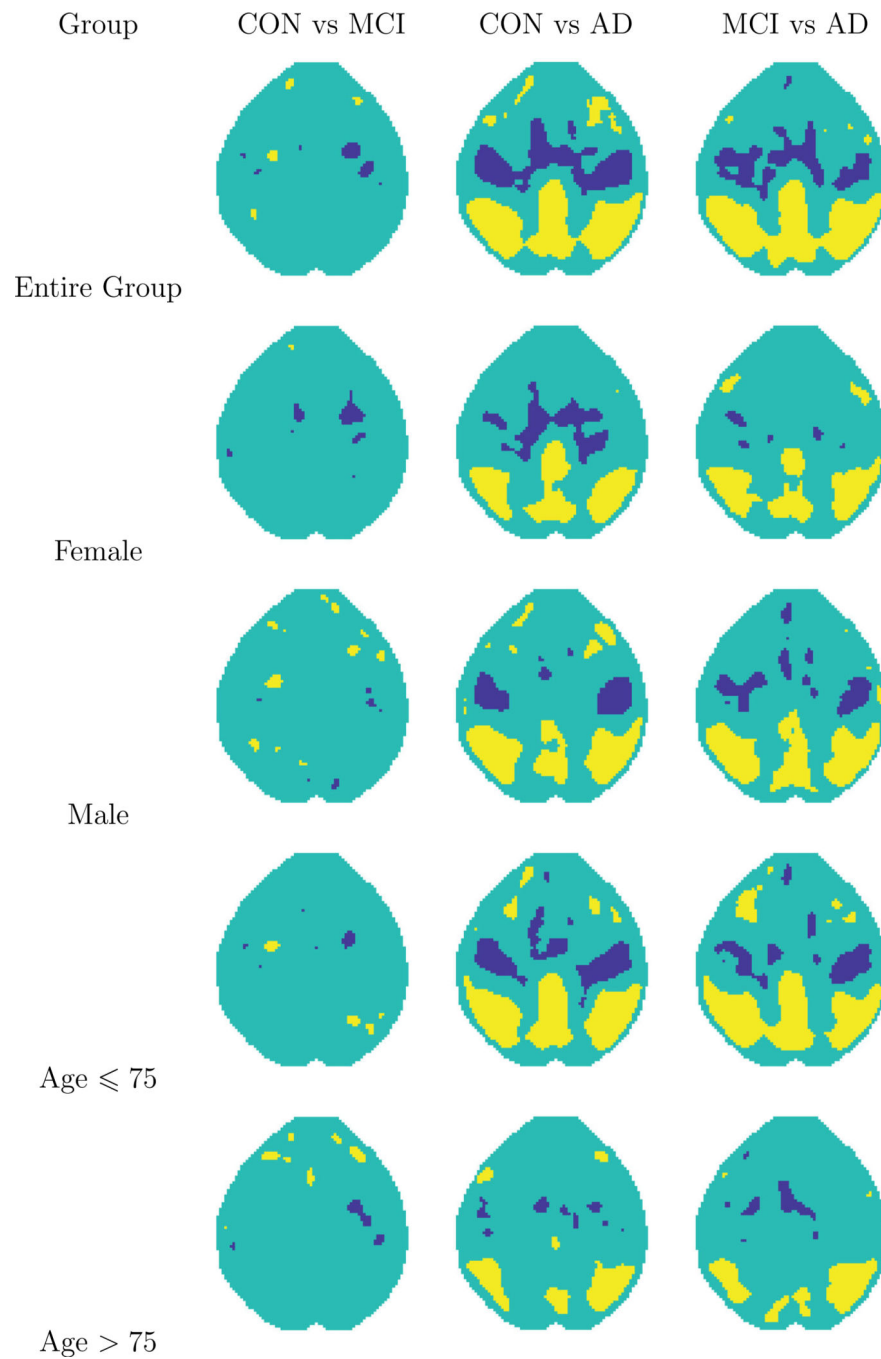
Author Manuscript

Author Manuscript

Author Manuscript



**FIGURE 2.** Simultaneous confidence corridor (SCC) for comparison between normal control (CON) and major depressive disorder (MDD): (A) coverage of zero, (B)  $\hat{\mu}_{MDD} - \hat{\mu}_{CON}$ , (C) lower simultaneous confidence corridor (SCC), and (D) upper SCC. In (A), yellow color indicates zero falls above the upper SCC and blue color indicates zero falls beneath the lower SCC



**FIGURE 3.**

Coverage of zero of simultaneous confidence corridor (SCC) for pairwise comparisons among CON, mild cognitive impairment (MCI) and Alzheimer's disease (AD). Yellow color indicates zero falls above the upper SCC and blue color indicates zero falls beneath the lower SCC

Empirical coverage rates and average widths of the simultaneous confidence corridors ( $N = 921$ )

TABLE 1

$n$	$\alpha = 0.10$			$\alpha = 0.05$			$\alpha = 0.01$		
	1	2	3	1	2	3	1	2	3
$\mu(z) = 20\{(z_1 - 0.5)^2 + (z_2 - 0.5)^2\}$									
50	0.858 (0.651)	0.860 (0.651)	0.874 (0.659)	0.928 (0.739)	0.929 (0.739)	0.935 (0.747)	0.977 (0.908)	0.981 (0.908)	0.981 (0.916)
100	0.891 (0.473)	0.893 (0.473)	0.897 (0.474)	0.944 (0.535)	0.947 (0.535)	0.949 (0.537)	0.979 (0.657)	0.979 (0.657)	0.980 (0.659)
200	0.896 (0.335)	0.897 (0.336)	0.897 (0.337)	0.942 (0.379)	0.949 (0.380)	0.948 (0.381)	0.987 (0.465)	0.988 (0.466)	0.988 (0.467)
$\mu(z) = 5 \exp[-15\{(z_1 - 0.5)^2 + (z_2 - 0.5)^2\}] + 0.5$									
50	0.877 (0.664)	0.879 (0.666)	0.879 (0.667)	0.939 (0.752)	0.941 (0.754)	0.937 (0.755)	0.983 (0.921)	0.983 (0.923)	0.982 (0.924)
100	0.888 (0.473)	0.892 (0.474)	0.892 (0.474)	0.942 (0.535)	0.944 (0.536)	0.945 (0.537)	0.979 (0.657)	0.980 (0.658)	0.980 (0.659)
200	0.904 (0.341)	0.890 (0.336)	0.902 (0.342)	0.947 (0.385)	0.942 (0.381)	0.949 (0.386)	0.986 (0.470)	0.986 (0.466)	0.986 (0.472)
$\mu(z) = 3.2(-z_1^3 + z_2^3) + 2.4$									
50	0.876 (0.639)	0.879 (0.639)	0.880 (0.639)	0.934 (0.727)	0.937 (0.728)	0.938 (0.728)	0.980 (0.896)	0.981 (0.896)	0.981 (0.897)
100	0.870 (0.455)	0.876 (0.455)	0.884 (0.457)	0.929 (0.517)	0.935 (0.517)	0.938 (0.519)	0.979 (0.639)	0.980 (0.640)	0.980 (0.642)
200	0.890 (0.326)	0.889 (0.325)	0.906 (0.329)	0.941 (0.370)	0.942 (0.370)	0.953 (0.373)	0.984 (0.456)	0.986 (0.456)	0.985 (0.459)
$\mu(z) = -10[\sin\{5\pi(z_1 + 0.22)\} - \sin\{5\pi(z_2 - 0.18)\}] + 2.8$									
50	0.882 (0.734)	0.869 (0.740)	0.879 (0.754)	0.937 (0.821)	0.930 (0.828)	0.939 (0.843)	0.981 (0.989)	0.976 (0.996)	0.980 (1.011)
100	0.886 (0.522)	0.901 (0.534)	0.880 (0.536)	0.938 (0.584)	0.946 (0.596)	0.935 (0.598)	0.982 (0.705)	0.983 (0.718)	0.982 (0.721)
200	0.877 (0.370)	0.891 (0.378)	0.887 (0.384)	0.937 (0.414)	0.951 (0.423)	0.947 (0.429)	0.985 (0.499)	0.986 (0.508)	0.984 (0.514)