Biometrics 64, 1–22 DOI: 000
November 2018

# Simultaneous Confidence Corridors for Mean Functions in Functional Data Analysis of Imaging Data

**Yueying Wang[1], Guannan Wang[2], Li Wang[1,*] and R. Todd Ogden[3]**

**for the Alzheimer's Disease Neuroimaging Initiative**

[1]Department of Statistics, Iowa State University, Ames, Iowa, U.S.A.

[2]Department of Mathematics, College of William and Mary, Williamsburg, Virginia, U.S.A.

[3]Department of Biostatistics, Columbia University, New York, New York, U.S.A.

*email:* lilywang@iastate.edu

Summary: 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 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 (PET) data in two different studies.

Key words: Bivariate splines; Functional principal component analysis; Image analysis; Semiparametric efficiency triangulation.

This paper has been submitted for consideration for publication in *Biometrics*

## 1. Introduction

In recent years, as digital technology has 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 effectiveness of treatment. For example, radiographic imaging is one of the effective and clinically useful tools for examining various body tissues to identify various conditions. Large-scale imaging data offers an incredibly rich data resource for scientific and medical discovery.

Functional data analysis 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 $\boldsymbol{z} = (z_1, z_2)$ be a point in $\Omega$. The model we consider is:

$$Y_i(\boldsymbol{z}) = \mu(\boldsymbol{z}) + \eta_i(\boldsymbol{z}) + \sigma(\boldsymbol{z})\varepsilon_i(\boldsymbol{z}),\ i = 1, \ldots, n,\ \boldsymbol{z} \in \Omega, \tag{1}$$

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

For biomedical imaging data, the objects (e.g., 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, i.e., poor estimation over difficult regions that is a result of smoothing inappropriately across boundaries of features.

In this paper, 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 $\mu(\boldsymbol{z})$ 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, $\mu(\boldsymbol{z})$, together with quantifying the estimation uncertainty and making comparisons between populations. The prevailing analytic technique, termed the "mass univariate" approach, involves regarding each pixel/voxel as a unit, and for each unit, performing a test of hypotheses. The obvious multiple comparisons issue can be dealt with in many ways; popular approaches involve random field theory and false discovery rate. An alternative approach is to regard 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 and towards simultaneous, 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.

Developing separately from the brain imaging analysis literature, SCCs have long been recognized as vital tools for inference on the global behavior of functions, see for example, Wang and Yang (2009), Wang and Yang (2010), Krivobokova et al. (2010), Song et al. (2014), Wang et al. (2013), Wang et al. (2014) and Cai et al. (2019). Existing SCC work for functional data analysis (FDA) has concentrated on the one-dimensional case. For example,

papers investigating SCCs of mean curves of functional data include Degras (2011), Cao et al. (2012), Chang et al. (2017), Cao (2014), Zheng et al. (2014), Degras (2017) and Cao and Wang (2018). Cao et al. (2016) proposed a SCC for the covariance function for dense functional data. Choi and Reimherr (2018) proposed geometric approach to confidence regions for functional parameters. In functional regression, Zhu et al. (2012) proposed SCCs for the regression coefficient functions for multivariate varying coefficient model for functional responses. Gu et al. (2014) proposed the SCC for varying coefficient regression with sparse functional data. However, there is scant literature on SCCs for imaging data or other more general 2D functions.

In this paper, we derive SCCs with exact coverage probability for the mean functions $\mu(\cdot)$ in (1) via extreme value theory of Gaussian processes and approximating mean functions with bivariate splines. Our simulation studies indicate 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 they would be if all the $n$ random images had been observed without noise.

Motivated by the need to statistically quantify the difference between two imaging datasets that arise in medical imaging studies, we further consider two-sample inference for imaging data 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 the analysis of imaging data 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 paper 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 datasets. In Section 7, we conclude the article with some discussions. Technical proofs of the theoretical results are provided in the supplementary materials.

## 2. Models and Estimation Method

### 2.1 *Models*

Let $\boldsymbol{z}_j \in \Omega$ indicate the $j$th pixel, $j = 1, \ldots N$, in the domain $\Omega$, and let $Y_{ij}$ be the imaging response of subject $i$ at location $j$, and the actual data set consists of $\{(Y_{ij}, \boldsymbol{z}_j)\}$, $i = 1, \ldots, n$, $j = 1, \ldots, N$, which can be modeled as

$$Y_{ij} = \mu(\boldsymbol{z}_j) + \eta_i(\boldsymbol{z}_j) + \sigma(\boldsymbol{z}_j)\varepsilon_{ij}. \tag{2}$$

Denote the eigenvalue and eigenfunction sequences of the covariance operator $G_\eta(\boldsymbol{z}, \boldsymbol{z}')$ as $\{\lambda_k\}_{k=1}^{\infty}$ and $\{\psi_k(\boldsymbol{z})\}_{k=1}^{\infty}$, in which $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant 0$, $\sum_{k=1}^{\infty} \lambda_k < \infty$, and $\{\psi_k\}_{k=1}^{\infty}$ form an orthonormal basis of $L_2(\Omega)$. It follows from spectral theory that $G_\eta(\boldsymbol{z}, \boldsymbol{z}') = \sum_{k=1}^{\infty} \lambda_k \psi_k(\boldsymbol{z})\psi_k(\boldsymbol{z}')$. The $i$th stochastic process $\{\eta_i(\boldsymbol{z}), \boldsymbol{z} \in \Omega\}$ allows the Karhunen-Loéve $L_2$ representation Sang and Huang (2012): $\eta_i(\boldsymbol{z}) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k(\boldsymbol{z})$, where $\phi_k(\boldsymbol{z}) = \sqrt{\lambda_k}\psi_k(\boldsymbol{z})$, 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 functional principal component score of the $i$th subject in classical functional principal component analysis (FPCA). Thus, the response measurements

in (2) can be represented as follows

$$Y_{ij} = \mu(\boldsymbol{z}_j) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(\boldsymbol{z}_j) + \sigma(\boldsymbol{z}_j)\varepsilon_{ij}. \tag{3}$$

2.2 *Bivariate spline basis approximation over triangulations*

Medical imaging data are typically observed on an irregular domain $\Omega$, and thus triangulation is an effective strategy to handle such type of data. We approximate the mean function in (3) by the bivariate splines that are piecewise polynomial functions over a 2D triangulated domain; see Lai and Wang (2013). In this section, 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 $\triangle = \{T_1, \ldots, T_M\}$ 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 $\triangle$ is either a shared vertex or a shared edge. In the rest of the paper, we assume that all points $\boldsymbol{z}_j$'s lie in the interior of some triangle in $\triangle$, i.e., they are not on edges or vertices of triangles in $\triangle$.

Given a triangle $T \in \triangle$, let $|T|$ be its longest edge length, and $\varrho_T$ be the radius of the largest disk which can be inscribed in $T$. Define the shape parameter of $T$ as the ratio $\pi_T = |T|/\varrho_T$. When $\pi_T$ is small, the triangles are relatively uniform in the sense that all angles of triangles in the triangulation $\triangle$ are roughly the same. Denote the size of $\triangle$ by $|\triangle| := \max\{|T|, T \in \triangle\}$, i.e., the length of the longest edge of all triangles in $\triangle$.

For an integer $r \geqslant 0$, let $\mathcal{C}^r(\Omega)$ be the collection of all $r$-th continuously differentiable functions over $\Omega$. Given a triangulation $\triangle$, 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 triangulation $\triangle$, 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 $d$. We use Bernstein basis polynomials to represent the bivariate splines. Let $\{B_m\}_{m\in\mathcal{M}}$ be the set of degree-$d$ bivariate Bernstein basis polynomials for $\mathcal{S}_d^r(\triangle)$, 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 $\mathbf{B}^\top(\boldsymbol{z}_j) = \{B_m(\boldsymbol{z}_j), m \in \mathcal{M}\}$. We approximate the mean function $\mu(\boldsymbol{z})$ by $\mu(\boldsymbol{z}) \approx \mathbf{B}^\top(\boldsymbol{z})\boldsymbol{\gamma}$, where $\boldsymbol{\gamma}^\top = (\gamma_m, m \in \mathcal{M})$ is the spline coefficient vector.

To define the penalized spline method, for any function $g(\boldsymbol{z})$ and direction $z_h$, $h = 1, 2$, let $\nabla_{z_h}^v g(\boldsymbol{z})$ denote the $v$-th order derivative in the direction $z_h$ at the point $\boldsymbol{z}$. We consider the following penalized least squares problem:

$$\min_{g\in S_d^r(\triangle)} \sum_{i=1}^{n}\sum_{j=1}^{N} \{Y_i(\boldsymbol{z}_j) - g(\boldsymbol{z}_j)\}^2 + \rho_n \mathcal{E}(g),$$

where $\mathcal{E}(s) = \sum_{T\in\triangle}\int_T \sum_{i+j=2}\binom{2}{i}(\nabla_{z_1}^i \nabla_{z_2}^j s)^2 dz_1 dz_2$ 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. Thus, we have to minimize

$$\sum_{i=1}^{n}\sum_{j=1}^{N} \left\{Y_i(\boldsymbol{z}_j) - \mathbf{B}^\top(\boldsymbol{z}_j)\boldsymbol{\gamma}\right\}^2 + \rho_n \boldsymbol{\gamma}^\top \mathbf{P}\boldsymbol{\gamma}, \text{ subject to } \mathbf{H}\boldsymbol{\gamma} = 0,$$

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

We first remove the constraint via QR decomposition of $\mathbf{H}^\top$: $\mathbf{H}^\top = \mathbf{Q}\mathbf{R} = (\mathbf{Q}_1\ \mathbf{Q}_2)\binom{\mathbf{R}_1}{\mathbf{R}_2}$, where $\mathbf{Q}$ is orthogonal and $\mathbf{R}$ is upper triangular, the submatrix $\mathbf{Q}_1$ is the first $r$ columns of $\mathbf{Q}$, where $r$ 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 $\mathbf{H}\boldsymbol{\gamma} = 0$. The minimization problem is thus converted to a conventional unrestricted penalized regression problem:

$$\sum_{i=1}^{n}\sum_{j=1}^{N} \left\{Y_i(\boldsymbol{z}_j) - \mathbf{B}^\top(\boldsymbol{z}_j)\mathbf{Q}_2\boldsymbol{\theta}\right\}^2 + \rho_n \boldsymbol{\theta}^\top \mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\boldsymbol{\theta}. \tag{4}$$

Denote $\bar{Y}_{\cdot,j} = \frac{1}{n}\sum_{i=1}^n Y_{ij}$, $\bar{\mathbf{Y}} = (\bar{Y}_{\cdot,1}, \ldots, \bar{Y}_{\cdot,N})^\top$, $\mathbf{U} = \mathbf{B}\mathbf{Q}_2$, and $\mathbf{D} = \mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2$. Then,

minimizing (4) is equivalent to minimizing

$$\left\|\bar{\mathbf{Y}} - \mathbf{B}\mathbf{Q}_2\boldsymbol{\theta}\right\|^2 + \frac{\rho_n}{n}\boldsymbol{\theta}^\top\mathbf{Q}_2^\top\mathbf{P}\mathbf{Q}_2\boldsymbol{\theta} = \left\|\bar{\mathbf{Y}} - \mathbf{U}\boldsymbol{\theta}\right\|^2 + \frac{\rho_n}{n}\boldsymbol{\theta}^\top\mathbf{D}\boldsymbol{\theta},$$

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

### 2.3 *Convergence of the Penalized Estimators*

In this section, we investigate the asymptotic properties for the proposed bivariate spline estimators. To discuss these properties, we introduce some notation of norms. For any function $g$ over the closure of domain $\Omega$, denote $\|g\|^2_{L^2(\Omega)} = \int_\Omega g^2(\boldsymbol{z})d\boldsymbol{z}$ the regular $L_2$ norm of $g$, and $\|g\|_{\infty,\Omega} = \sup_{\boldsymbol{z}\in\Omega}|g(\boldsymbol{z})|$ the supremum norm of $g$. Let $|g|_{v,\infty,\Omega} = \max_{i+j=v}\|\nabla^i_{z_1}\nabla^j_{z_2}g\|_{\infty,\Omega}$ be the maximum norms of all the $v$th order derivatives of $g$ over $\Omega$.

Let $\mathcal{W}^{d,\infty}(\Omega) = \{g : |g|_{k,\infty,\Omega} < \infty, 0 \leqslant k \leqslant d\}$ be the standard Sobolev space. Given random variables $S_n$ for $n \geqslant 1$, we write $S_n = O_P(b_n)$ if $\lim_{c\to\infty}\limsup_n P(|S_n| \geqslant cb_n) = 0$. Similarly, we write $S_n = o_P(b_n)$ if $\lim_n P(|S_n| \geqslant cb_n) = 0$, for any constant $c > 0$.

We next introduce some technical conditions.

(A1) The bivariate function $\mu(\cdot) \in \mathcal{W}^{d+1,\infty}(\Omega)$ for an integer $d \geqslant 1$.

(A2) For any $k \geqslant 1$, $\xi_{ik}$'s are i.i.d. random variables with mean 0, variance 1 and $E|\xi_{ik}|^{4+\delta_1} < +\infty$ for some constant $\delta_1 > 0$. For any $i = 1,\ldots,n$, $j = 1,\ldots,N$, $\varepsilon_{ij}$'s are i.i.d with mean 0, variance 1, and $E|\varepsilon_{ij}|^{4+\delta_2} < +\infty$ for some constant $\delta_2 > 0$.

(A3) The function $\sigma \in \mathcal{C}^{(1)}(\Omega)$ with $0 < c_\sigma \leqslant \sigma(\boldsymbol{z}) \leqslant C_\sigma \leqslant \infty$ for any $\boldsymbol{z} \in \Omega$; for any $k$, $\psi_k \in \mathcal{C}^{(1)}(\Omega)$ and the variance function $0 < c_G \leqslant G_\eta(\boldsymbol{z},\boldsymbol{z}) \leqslant C_G \leqslant \infty$, for any $\boldsymbol{z} \in \Omega$.

(A4) The triangulation is $\pi$-quasi-uniform, that is, there exists a positive constant $\pi$ such that $(\min_{T\in\triangle}\varrho_T)^{-1}|\triangle| \leqslant \pi$.

(A5) As $n \to \infty$, $N^{-1}n^{1/(d+1)}\log(n) \to 0$, the triangulation size satisfies that $N^{-1}\log(n) \ll$

$|\triangle|^2 \ll \min\{n^{(2+\delta_2)/(4+\delta_2)}N^{-1}\log^{-1}(n), n^{-1/(d+1)}\}$, and the smoothing penalty parameter $\rho_n$ satisfies $\rho_n = o\{\min(n^{1/2}N|\triangle|^3, nN^{3/2}|\triangle|^6, nN|\triangle|^5)\}$.

(A6) For $k \in \{1, \ldots, \kappa\}$ and a nonnegative integer $s$, $\phi_k(\boldsymbol{z}) \in \mathcal{W}^{s+1,\infty}(\Omega)$, $\sum_{k=1}^{\kappa} \|\phi_k\|_\infty < \infty$. $\frac{\rho_n}{nN|\triangle|^3}\sum_{k=1}^{\kappa_n}\|\phi_k\|_{2,\infty} = o(1)$, $\left(1+\frac{\rho_n}{nN|\triangle|^5}\right)\sum_{k=1}^{\kappa_n}|\triangle|^{s+1}\|\phi_k\|_{s+1,\infty} = o(1)$ for a sequence $\{\kappa_n\}_{n=1}^{\infty}$ of increasing integers, with $\lim_{n\to\infty}\kappa_n = \kappa$, as $n \to \infty$. Meanwhile, $\sum_{k=\kappa_n+1}^{\kappa}\|\phi_k\|_\infty = o(1)$. The number $\kappa$ of nonzero eigenvalues is finite or $\kappa$ is infinite.

(A7) As $n \to \infty$, for some $0 < \delta_3 < 1$, $N^{-1}n^{1/(d+1)+\delta_3} \to 0$, $N|\triangle_\eta|^2 \to \infty$, $n^2|\triangle_\eta|^4/\log n \to \infty$.

The above assumptions are mild conditions that can be satisfied in many practical situations. Assumption (A1) is typically assumed about the true underlying functions in the nonparametric estimation literature. Assumption (A1) can be relaxed by only requiring $\mu(\cdot) \in \mathcal{C}^{(0)}(\Omega)$ if the imaging data has sharp edges. Assumptions (A2) and (A3) are common conditions used in the literature; see for example, Cao et al. (2012). Assumption (A4) suggests the use of more uniform triangulations with smaller shape parameters. Assumption (A5) describes the requirement of the growth rate of the dimension of the spline spaces relative to the sample size and the image resolution. Assumption (A5) also implies that the number of pixels for each image $N$ diverges to infinity as $n \to \infty$, a well-developed asymptotic scenario for dense functional data, see Li and Hsing (2010) and Zhu et al. (2014).

The following theorem provides the $L_2$ convergence rate and uniform convergence rate of $\widehat{\mu}(\cdot)$. The detailed proofs of this theorem are given in the supplemental materials.

THEOREM 1: *Suppose Assumptions (A1), (A2), (A3), (A4) hold and* $N^{1/2}|\triangle| \to \infty$ *as* $N \to \infty$. *Then the bivariate penalized spline estimator* $\mu(\cdot)$ *is consistent and satisfies*

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

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

## 3. Simultaneous Confidence Corridors

### 3.1 *One sample*

In this section, we develop the simultaneous confidence corridors for the mean function $\mu(\cdot)$.

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

Define the "oracle" estimator $\bar{\mu}(\boldsymbol{z}) = \mu(\boldsymbol{z}) + \frac{1}{n}\sum_{i=1}^{n}\eta_i(\boldsymbol{z})$. Of course this is infeasible due to the finite pixel grid $\{\boldsymbol{z}_j : j = 1,\ldots,N\}$ and the measurement error. The following theorem presents the asymptotic properties of $\bar{\mu}(\boldsymbol{z})$ and shows that the difference between the BPS estimator $\widehat{\mu}(\boldsymbol{z})$ and the "oracle" smoother $\bar{\mu}(\boldsymbol{z})$ is uniformly bounded at an $o_P(n^{1/2})$ rate.

THEOREM 2: *Under Assumptions (A1)–(A6), for any $\alpha \in (0,1)$, as $n \to \infty$,*

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

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

Based on Theorems 1 and 2, we obtain the following asymptotic SCCs for $\mu(\boldsymbol{z})$, $\boldsymbol{z} \in \Omega$.

COROLLARY 1: *Under Assumptions (A1)–(A6), for any $\alpha \in (0,1)$, as $n \to \infty$, an asymptotic $100(1-\alpha)\,\%$ exact SCC for $\mu(\boldsymbol{z})$ is*

$$P\left\{\mu(\boldsymbol{z}) \in \widehat{\mu}(\boldsymbol{z}) \pm n^{-1/2} q_{1-\alpha} G_\eta(\boldsymbol{z},\boldsymbol{z})^{1/2},\ \boldsymbol{z} \in \Omega\right\} \to 1-\alpha.$$

### 3.2 *Extension to two-sample case*

While one-sample confidence bands are of primary interest in many situations, in some brain imaging analysis, interest lies in comparing two groups, e.g., patients and normal control

subjects. In this section, 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. With these two-sample SCCs, we can assess differences of images with quantified uncertainty.

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 $G_{H\eta}(\boldsymbol{z}, \boldsymbol{z}') = \sum_{k=1}^{\kappa} \phi_{Hk}(\boldsymbol{z}) \phi_{Hk}(\boldsymbol{z}')$ be positive definite function and $\widehat{\mu}_H$ be the spline estimates for the group mean function $\mu_H$.

Let $V(\boldsymbol{z}, \boldsymbol{z}') = G_{1\eta}(\boldsymbol{z}, \boldsymbol{z}') + rG_{2\eta}(\boldsymbol{z}, \boldsymbol{z}')$, where $r = \lim_{n_1 \to \infty} n_1/n_2$. Denote by $W(\boldsymbol{z})$, $\boldsymbol{z} \in \Omega$ a standardized Gaussian process such that $EW(\boldsymbol{z}) = 0$, $EW^2(\boldsymbol{z}) = 1$ with covariance

$$E[W(\boldsymbol{z}) W(\boldsymbol{z}')] = \{V(\boldsymbol{z}, \boldsymbol{z})\}^{-1/2} V(\boldsymbol{z}, \boldsymbol{z}') \{V(\boldsymbol{z}', \boldsymbol{z}')\}^{-1/2}.$$

Denoted by $q_{12,\alpha}$ the $(1-\alpha)$-th quantile of the absolute maximal distribution of $W(\boldsymbol{z})$, $\boldsymbol{z} \in \Omega$.

THEOREM 3: *Under Assumptions (A1)–(A6), for any* $\alpha \in (0, 1)$, *as* $n_1 \to \infty$, $\widehat{r} \to r > 0$,

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

Theorem 3 suggests that an asymptotic $100(1-\alpha)\%$ exact SCC for $(\mu_1 - \mu_2)(\boldsymbol{z})$ can be constructed as $(\widehat{\mu}_1 - \widehat{\mu}_2)(\boldsymbol{z}) \pm n_1^{-1} q_{12,\alpha} \{V(\boldsymbol{z}, \boldsymbol{z})\}^{1/2}$.

## 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 *Estimation of FPCA*

In practice, to construct SCCs as derived in Section 3, it is necessary to estimate the FPCA, such as the variance-covariance function $G_\eta(\boldsymbol{z}, \boldsymbol{z}')$ and its eigenvalues and eigenfunctions.

For any $i = 1, \ldots, n$, $j = 1, \ldots, N$, let $\widehat{R}_{ij} = Y_{ij} - \widehat{\mu}(\boldsymbol{z}_j)$ be the residual. We estimate $\eta_i(\boldsymbol{z})$ individually by employing the bivariate spline smoothing method to $\{(\widehat{R}_{ij}, \boldsymbol{z}_j)\}_{j=1}^N$. To be more specific, for each $i = 1, \ldots, n$, we define the spline estimator of $\eta_i(\boldsymbol{z})$ as

$$\widehat{\eta}_i(\boldsymbol{z}) = \underset{g_i \in \mathcal{S}_d^r(\triangle^*)}{\arg\min} \sum_{j=1}^{N} \left\{ \widehat{R}_{ij} - g_i(\boldsymbol{z}_j) \right\}^2 + \rho_n^* \mathcal{E}(g_i),$$

where the triangulation $\triangle^*$ and smoothness penalty $\rho_n^*$ may be different from those introduced in Section 2 when estimating $\mu(\boldsymbol{z})$. Next, define the estimator of $G_\eta(\boldsymbol{z}, \boldsymbol{z}')$ as

$$\widehat{G}_\eta(\boldsymbol{z}, \boldsymbol{z}') = n^{-1} \sum_{i=1}^{n} \widehat{\eta}_i(\boldsymbol{z}) \widehat{\eta}_i(\boldsymbol{z}'), \tag{5}$$

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

$$\int_\Omega \widehat{G}_\eta(\boldsymbol{z}, \boldsymbol{z}') \widehat{\psi}_k(\boldsymbol{z}) d\boldsymbol{z} = \widehat{\lambda}_k \widehat{\psi}_k(\boldsymbol{z}'), \tag{6}$$

where $\widehat{\psi}_k$ are subject to $\int_\Omega \widehat{\psi}_k^2(\boldsymbol{z})\, d\boldsymbol{z} = 1$ and $\int_\Omega \widehat{\psi}_k(\boldsymbol{z}) \widehat{\psi}_{k'}(\boldsymbol{z}) d\boldsymbol{z} = 0$ for $k' < k$. If $N$ is sufficiently large, the left hand side of (6) can be approximated by $\sum_{j=1}^N \widehat{G}(\boldsymbol{z}_j, \boldsymbol{z}_{j'}) \widehat{\psi}_k(\boldsymbol{z}_j) A(\boldsymbol{z}_j)$, where $A(\boldsymbol{z}_j)$ is the area of the pixel $\boldsymbol{z}_j$.

Theorem 4 below characterizes the uniform weak convergence of $\widehat{G}_\eta(\boldsymbol{z}, \boldsymbol{z}')$ and the convergence of $\widehat{\psi}_k$ and $\widehat{\lambda}_k$.

THEOREM 4: *Under Assumptions (A1)–(A7), we have*

*(i) The spline estimator $\widehat{G}_\eta(\boldsymbol{z}, \boldsymbol{z}')$ in (5) uniformly converges to $G_\eta(\boldsymbol{z}, \boldsymbol{z}')$ in probability, i.e.,* $\sup_{(\boldsymbol{z}, \boldsymbol{z}') \in \Omega^2} |\widehat{G}_\eta(\boldsymbol{z}, \boldsymbol{z}') - G_\eta(\boldsymbol{z}, \boldsymbol{z}')| = o_P(1)$.

*(ii)* $\|\widehat{\psi}_k - \psi_k\| = o_P(1)$, $|\widehat{\lambda}_k - \lambda_k| = o_P(1)$, *for* $k = 1, \ldots, \kappa$.

Although, in theory, the Karhunen-Loéve representation of the covariance function consists of infinite number of terms. In applications, it is typical to truncate the spectral decomposition at an integer chosen so as to account for some predetermined proportion of the variance Hall et al. (2006); Li et al. (2013). One can select the number of principal component using the

Akaike information criterion (AIC) suggested by Yao et al. (2005) or Bayesian information criterion (BIC) proposed by Li et al. (2013).

### 4.2 *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 $\zeta_b(\boldsymbol{z}) = \widehat{G}_\eta^{-1/2}(\boldsymbol{z},\boldsymbol{z})\sum_{k=1}^{\kappa}\widehat{\lambda}_k^{1/2}Z_{k,b}\widehat{\psi}_k(\boldsymbol{z})$, where $Z_{k,b}$ are i.i.d standard normal variables with $1 \leqslant k \leqslant \kappa$ and $b = 1,\ldots,B$ for a preset large integer $B$. Next, 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 $\zeta_b(\boldsymbol{z})$.

To construct the SCC for the two sample case, we simulate

$$\widehat{W}_b(\boldsymbol{z}) = \{\widehat{V}(\boldsymbol{z},\boldsymbol{z})\}^{-1/2}\sum_{k=1}^{\kappa}\{\widehat{\lambda}_{1k}^{1/2}Z_{1k,b}\widehat{\psi}_{1k}(\boldsymbol{z}) - (n_1/n_2)^{1/2}\widehat{\lambda}_{2k}^{1/2}Z_{2k,b}\widehat{\psi}_{2k}(\boldsymbol{z})\},\ \boldsymbol{z}\in\Omega.$$

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

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 can not 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

$$\mathrm{GCV}(\rho_n) = \frac{\left\|\bar{\mathbf{Y}} - \mathbf{S}(\rho_n)\bar{\mathbf{Y}}\right\|^2}{N\{1-\mathrm{tr}\{\mathbf{S}(\rho_n)\}/N\}^2}$$

over a grid of values of $\rho_n$, where $\mathbf{S}(\rho_n) = \mathbf{U}(\mathbf{U}^\top\mathbf{U} + n^{-1}\rho_n\mathbf{D})^{-1}\mathbf{U}^\top$.

### 4.3 *Triangulation selection*

To construct the SCC, we need to choose the triangulation used in the BPS, a notoriously difficult task for constructing nonparametric pointwise confidence intervals or simultaneous confidence bands. 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. Other criteria include the density control (adaptivity) and optimal size (number of triangles), etc. For a fixed number of triangles, Lai and Schumaker (2007) recommend selecting the triangulation according to "max-min" criterion which maximizes the minimum angle of all the angles of the triangles in the triangulation.

In this section, we focus on how to select the size of the triangulation, $|\triangle|$, for quasi-uniform triangulations. We choose $|\triangle|$ so as to minimize an estimator of coverage error of the SCCs. Here we describe our selection scheme for the one sample case.

- Step 1. Estimators of $\mu(\boldsymbol{z})$ and $\eta_i(\boldsymbol{z})$. Based on $\{Y_{ij}\}_{j=1,i=1}^{N,n}$, obtain $\widehat{\mu}(\boldsymbol{z})$ using the BPS method. Define $\widehat{R}_{ij} = Y_{ij} - \widehat{\mu}(\boldsymbol{z}_j)$. Then obtain $\widehat{\eta}_i(\boldsymbol{z})$ by smoothing $\widehat{R}_{ij}$ via penalized bivariate splines. Let $\widehat{\varepsilon}_{ij} = \widehat{R}_{ij} - \widehat{\eta}_i(\boldsymbol{z}_j)$.
- Step 2. Estimators of $G_\eta(\boldsymbol{z},\boldsymbol{z})$ and $q_{1-\alpha}$. We estimate $G_\eta(\boldsymbol{z},\boldsymbol{z})$ and $q_{1-\alpha}$ by $\widehat{G}_\eta(\boldsymbol{z},\boldsymbol{z})$ in (5) and $\widehat{q}_{1-\alpha}$ described in Section 3, respectively.
- Step 3. We generate an independent random sample $\delta_i^{(b)}$ and $\delta_{ij}^{(b)}$ from the standard normal distribution, and define $Y_{ij}^{*(b)} = \widehat{\mu}(\boldsymbol{z}_j) + \delta_i^{(b)}\widehat{\eta}_i(\boldsymbol{z}_j) + \delta_{ij}^{(b)}\widehat{\varepsilon}_{ij}$. For any fixed $\alpha \in (0,1)$, based on $\left\{Y_{ij}^{*(b)}\right\}_{j=1,i=1}^{N,n}$, construct $100(1-\alpha)\%$ SCCs for resampled data $\left\{Y_{ij}^{*(b)}\right\}_{j=1,i=1}^{N,n}$: $\mathcal{B}^{*(b)}(\alpha),\ b=1,\cdots,B$, $\mathcal{B}^{*(b)}(\alpha) = \widehat{\mu}^{*(b)}(\boldsymbol{z}) \pm n^{-1/2} q_{1-\alpha}^{*(b)} G_\eta^{*(b)}(\boldsymbol{z},\boldsymbol{z})^{1/2}$.
- Step 4. We choose $|\triangle|$ by minimizing the objective function

$$\int_{\alpha-\delta}^{\alpha+\delta} \left\{ \frac{1}{B}\sum_{b=1}^{B} I(\widehat{\mu} \in \mathcal{B}^{*(b)}(\alpha); |\triangle|) - (1-\alpha) \right\}^2 d\alpha,$$

for some constant $0 < \delta < \alpha$, which is taken to be 0.005 in our simulation studies.

Once $|\triangle|$ is chosen, one can build the triangulated meshes using typical triangulation construction methods such as Delaunay Triangulation De Loera et al. (2010).

## 5. Simulation Studies

In this section, we describe two Monte Carlo simulations to examine the finite sample performance of the proposed method. In our simulation studies, we use the 15-point grid where the values of $\log_{10}(\lambda)$ are equally spaced between $-6$ and 3.

### 5.1 *One Sample SCC*

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

$$Y_{ij} = \mu(\boldsymbol{z}_j) + \sum_{k=1}^{\kappa} \sqrt{\lambda_k}\xi_{ij}\psi_k(\boldsymbol{z}_j) + \sigma(\boldsymbol{z}_j)\varepsilon_{ij},$$

where $\boldsymbol{z}_j = (z_{1j}, z_{2j}) \in \Omega \subset [0,1]^2$. To demonstrate the practical performance of our theoretical results, we consider the following three mean functions:

- (quadratic) $\mu(\boldsymbol{z}) = 20\left\{(z_1 - 0.5)^2 + (z_2 - 0.5)^2\right\}$,
- (cubic) $\mu(\boldsymbol{z}) = 3.2(-z_1^3 + z_2^3) + 2.4$,
- (exponential) $\mu(\boldsymbol{z}) = 5\exp\left[-15\left\{(z_1 - 0.5)^2 + (z_2 - 0.5)^2\right\}\right] + 0.5$,

and the corresponding images are shown in the first column of Figures 1 – 5.

To simulate the within-image dependence, we generate $\xi_{ik} \overset{\text{i.i.d}}{\sim} N(0,1)$ for $i = 1, \ldots, n,\ j = 1, \ldots, N,$ and $k = 1, \ldots, \kappa$, where $\kappa = 2$ and $\psi_1(\boldsymbol{z}) = a_1 \sin(2\pi z_1),\ \psi_2(\boldsymbol{z}) = a_2(\cos(2\pi z_2) + a_0)$. Here, we set $a_0 = -0.039$, $a_1 = 1.588$, $a_2 = 2.157$, such that $\int_\Omega \psi_1^2(\boldsymbol{z})d\boldsymbol{z} = \int_\Omega \psi_2^2(\boldsymbol{z})d\boldsymbol{z} = 1$ and $\int_\Omega \psi_1(\boldsymbol{z})\psi_2(\boldsymbol{z})d\boldsymbol{z} = 0$. For the eigenvalues, we set (i) $\lambda_1 = 0.2$, $\lambda_2 = 0.05$, (ii) $\lambda_1 = 0.5$, $\lambda_2 = 0.2$. The noise level is set to be $\sigma = 0.1$. We consider $n = 50,\ 100,\ 200$ and for each image, the number of pixels is set to be $N = 921,\ 2104$.

To apply our method, we consider three different triangulations which are also shown in the first column of Figures 1 – 5. The first triangulation ($\triangle_1$) contains 49 triangles and 38 vertices; the second triangulation ($\triangle_2$) contains 80 triangles and 54 vertices; while the third triangulation ($\triangle_3$) contains 144 triangles and 87 vertices. The estimated mean function based on these three triangulations are shown in the second columns of Figures 1 – 5, and the SCCs are given in the last two columns of each figure. Although we construct the SCCs based on three different confidence levels: $1-\alpha = 0.90,\ 0.95$ and 0.99, for the sake of space, we only present the plots of SCCs when $\alpha = 0.01$ since it provides the widest SCCs. And all the plots of SCCs with different confidence levels look quite similar. From Figures 1 – 5, one can see that all three triangulations result in almost the same estimates and SCCs. Based on Figures 1, 4 and 5, one can also see that even when the number of images is moderate, the estimation is very accurate and the coverage rate is close to the nominal confidence level regardless of the type of underling mean functions.

[Figure 1 about here.]

[Figure 2 about here.]

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

Table 1 summarizes the estimated coverage rate based on 1000 simulated datasets. The number in parenthesis represents the average bandwidth. This table also confirms 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 $n$.

[Table 1 about here.]

### 5.2 *Two Sample SCC*

In this simulation study, we examine the power of detecting a difference in mean images based on the proposed two samples simultaneous confidence corridor. The data generation process is very similar as in Simulation 1. Two group of images are generated from the model:

$$Y_{H,ij} = \mu_H(\boldsymbol{z}_j) + \sum_{k=1}^{\kappa} \sqrt{\lambda_k}\xi_{ij}\psi_k(\boldsymbol{z}_j) + \sigma(\boldsymbol{z}_j)\varepsilon_{ij},\ H = 1, 2,$$

and $\psi_k$'s are generated as in Example 1. The mean functions for two groups considered here are $\mu_1(\boldsymbol{z}) = 20\{(z_1 - 0.5)^2 + (z_2 - 0.5)^2\}$ and $\mu_2(\boldsymbol{z}) = \mu_1(\boldsymbol{z}) + \delta(-z_1^3 + z_2^3)$. The value of $\delta$ controls the difference between the two groups. Table 2 and Figure 6 summarize the estimated probability of rejecting $H_0$ with nominal level $\alpha = 0.10,\ 0.05$ and $0.01$. When $\delta = 0$, the probability should be close to the nominal level, while when $\delta$, is large, the estimated power should be close to 1.

[Table 2 about here.]

[Figure 6 about here.]

From Table 2 and Figure 6, one can see even when the number of the images is moderate, the size of the test is very close to the nominal level, and the estimated power increases quickly as $n$ increases. 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 for mean functions 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 1A 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). 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, previously shown to be affected by Alzheimer's disease (Marcus et al., 2014).

In each case, we consider the following hypotheses between images taken from two groups:

$$H_0 : \ \mu_1(\boldsymbol{z}) = \mu_2(\boldsymbol{z}), \text{ for all } \boldsymbol{z} \in \Omega \quad \text{v.s.} \quad H_a : \ \mu_1(\boldsymbol{z}) \neq \mu_2(\boldsymbol{z}) \text{ for some } \boldsymbol{z} \in \Omega.$$

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 7 displays the results of the application of the proposed procedure to these data. In (a) can be seen the portions of the SCC that do not contain 0; the estimation of the mean difference between the two groups is shown in (b), and the lower and upper SCCs are shown in (c) and (d).

[Figure 7 about here.]

Next, we illustrate these procedures by applying them to PET data from the Alzheimer's Disease Neuroimaging Initiative (ADNI; `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 Alzheimers disease (AD). This dataset consists of 112 subjects with normal cognitive functions (control group; CON), 213 subjects with mild cognitive impairment (MCI), and 122 subjects who have been diagnosed with Alzheimer's Disease (AD).

Among the three triangulation candidates ($\triangle_1$–$\triangle_3$) considered in simulation studies, we choose $\triangle_1$ using the bootstrap method proposed in Subsection 4.3. The results of this application are displayed in Figure 8. The first row of Figure 8 displays the areas in which

0 is not contained within the SCC comparing each pair of diagnostic groups. This suggests that the AD group has widespread mean differences from each of the other two groups. Since this dataset is relatively large, we also stratify the data according to sex and age (greater or less than 75 years) and within each stratum we examine the SCC for the difference between all pairs of diagnostic groups. The breakdowns of these data in terms of these variables are given in Table 3.

[Figure 8 about here.]

[Table 3 about here.]

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 sub-group 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 semiparametrical efficient, 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.

In this paper, we approximate the bivariate function of the spatial effect using the bivariate splines over triangulations. We prefer the bivariate penalized splines (BPS) due to their (i) convenient representations with flexible degrees and various smoothness, (ii) computational efficiency, and (iii) great ability of handling the sparse designs. An R package "Basis" has been developed and is available from the second author's website.

A few more issues still merit further research. For instance, the triangulation is selected using the wild bootstrap, which works well in practice, but a stronger theoretical justification for their use is still needed in the FDA context. 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. Lastly, 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 Materials

Supplementary materials are available with this paper at the Biometrics website on Wiley Online Library. In the supplemental materials, we provide the technical proofs for the main theorems.

### Acknowledgements

Li Wang's research was supported in part by National Science Foundation award DMS-1542332. Todd Ogden's work was partially supported by NIH grants 5 R01 EB024526 and 2 P50 MH090964. We are grateful to Ming-Jun Lai for providing the code of bivariate spline basis functions. Data used in preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (`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`.

## References

Cai, L., Liu, R., Wang, S., and Yang, L. (2019). Simultaneous confidence bands for mean and variance functions based on deterministic design. *Statistica Sinica* **29,** 000–000.

Cao, G. (2014). Simultaneous confidence bands for derivatives of dependent functional data. *Electronic Journal of Statistics* **8,** 2639–2663.

Cao, G. and Wang, L. (2018). Simultaneous inference for the mean of repeated functional data. *Journal of Multivariate Analysis* **165,** 279–295.

Cao, G., Wang, L., Li, Y., and Yang, L. (2016). Oracle-efficient confidence envelopes for covariance functions in dense functional data. *Statistica Sinica* **26,** 359–383.

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.

Chang, C., Lin, X., and Ogden, R. T. (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, J. A., Rambau, J., and Santos, F. (2010). *Triangulations Structures for algorithms and applications*. Springer.

Degras, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statistica Sinica* **21,** 1735–1765.

Degras, D. A. (2017). Simultaneous confidence bands for the mean of functional data. *Wiley Interdisciplinary Reviews: Computational Statistics* **9,** e1397.

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., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics* **34,** 1493–1517.

Krivobokova, T., Kneib, T., and Claeskens, G. (2010). Simultaneous confidence bands for penalizedsplineestimators. *Journal of the American Statistical Association* **105,** 852–863.

Lai, M. J. and Schumaker, L. L. (2007). *Spline functions on triangulations.* Cambridge University Press.

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

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.

Li, Y., Wang, N., and Carroll, R. J. (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, R. M. (2014). Brain pet in the diagnosis of alzheimer's disease. *Clinical Nuclear Medicine* **39,** e413–22.

Parsey, R. V., Ogden, R. T., Miller, J. M., Tin, A., Hesselgrave, N., Goldstein, E., Mikhno, A., Milak, M., Zanderigo, F., Sullivan, G. M., Oquendo, M. A., and Mann, J. J. (2010). Higher serotonin 1a binding in a second major depression cohort: modeling and reference region considerations. *Biological Psychiatry* **68,** 170–178.

Parsey, R. V., Oquendo, M. A., Ogden, R. T., Olvet, D. M., Simpson, N., Huang, Y. Y., Van Heertum, R. L., Arango, V., and Mann, J. J. (2006). Altered serotonin 1a binding in major depression: a [carbonyl-c-11] way100635 positron emission tomography study. *Biological Psychiatry* **59,** 106–113.

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.

Song, Q., Liu, R., Shao, Q., and Yang, L. (2014). A simultaneous confidence band for dense longitudinal regression. *Communications in Statistics - Theory and Methods* **43,** 5195–5210.

Wang, J., Cheng, F., and Yang, L. (2013). Smooth simultaneous confidence bands for cumulative distribution functions. *Journal of Nonparametric Statistics* **25,** 395–407.

Wang, J., Liu, R., Cheng, F., and Yang, L. (2014). Oracally efficient estimation of autoregressive error distribution with simultaneous confidence band. *The Annals of Statistics* **42,** 654–668.

Wang, J. and Yang, L. (2009). Polynomial spline confidence bands for regression curves. *Statistica Sinica* **19,** 325–342.

Wang, L. and Yang, L. (2010). Simultaneous confidence bands for time-series prediction function. *Journal of Nonparametric Statistics* **22,** 999–1018.

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.

Zheng, S., Yang, L., and Härdle, W. K. (2014). A smooth simultaneous confidence corridor for the mean of sparse functional data. *Journal of the American Statistical Association* **109,** 661–673.

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.

*Received November* 2018. *Revised November* 2018. *Accepted November* 2018.

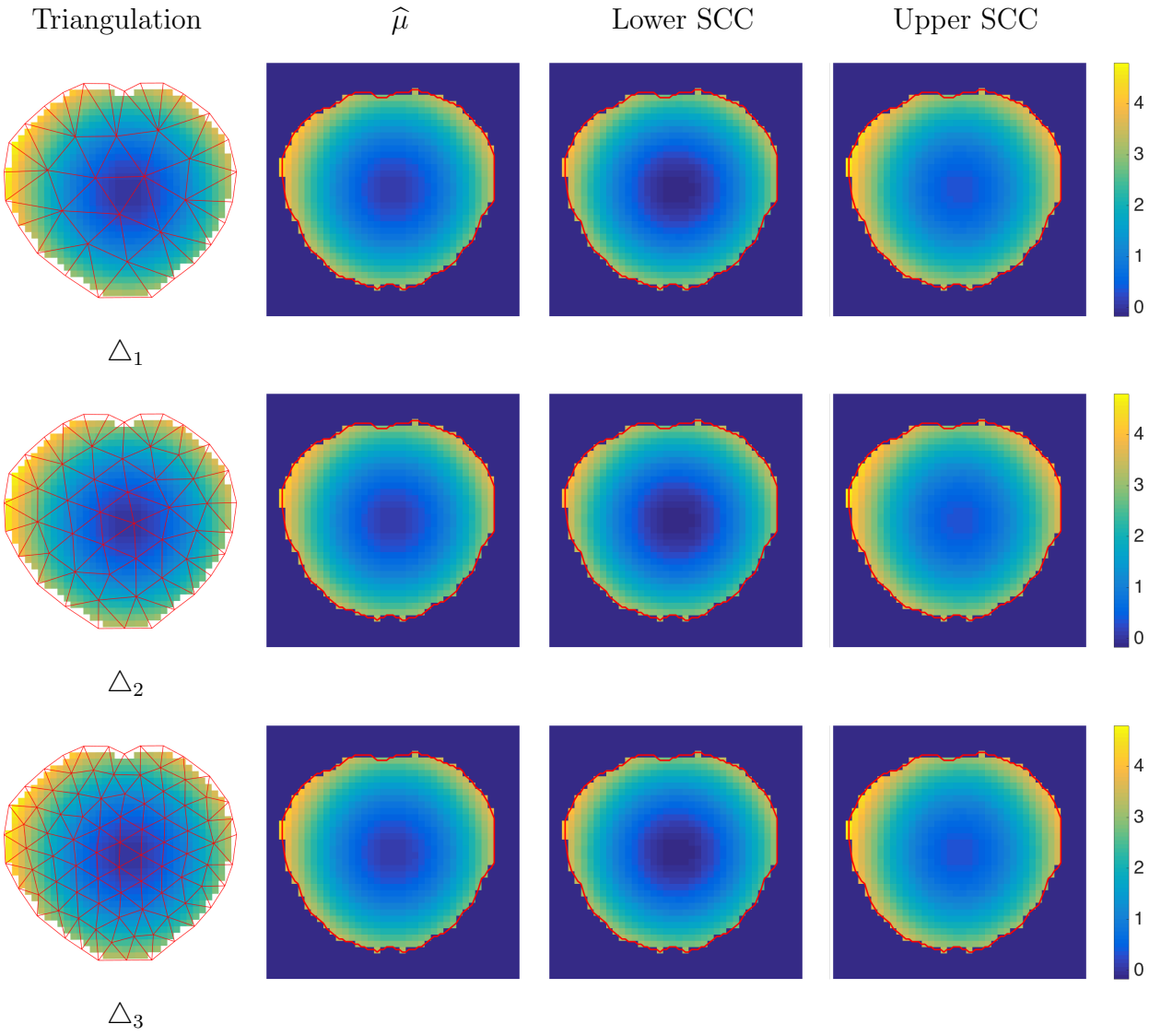



**Figure 1.** SCCs for Quadratic Function with $n = 50$ and $\alpha = 0.01$.

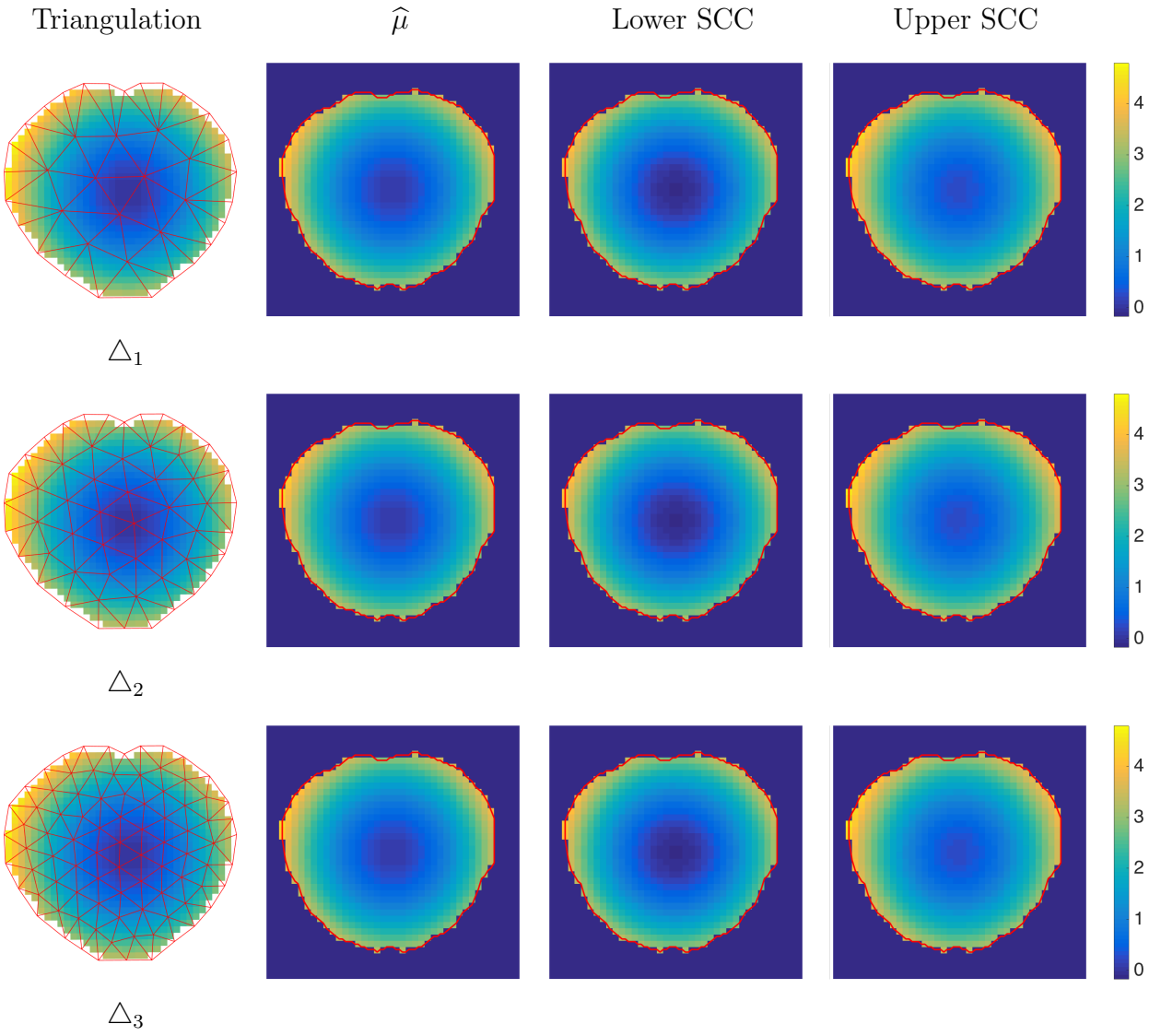



**Figure 2.** SCCs for Quadratic Function with $n = 100$ and $\alpha = 0.01$.

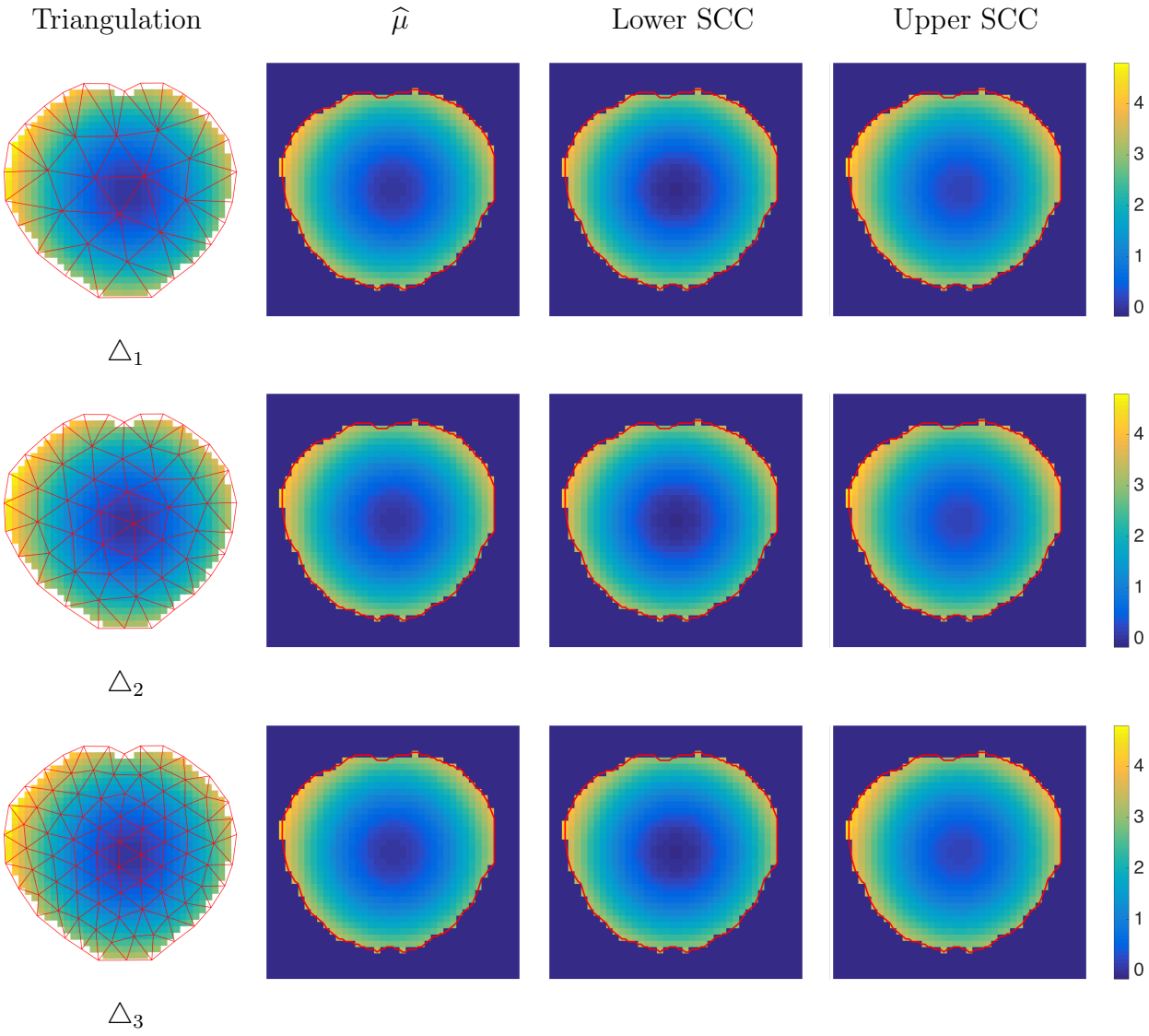



**Figure 3.** SCC for Quadratic Function with $n = 200$ and $\alpha = 0.01$.

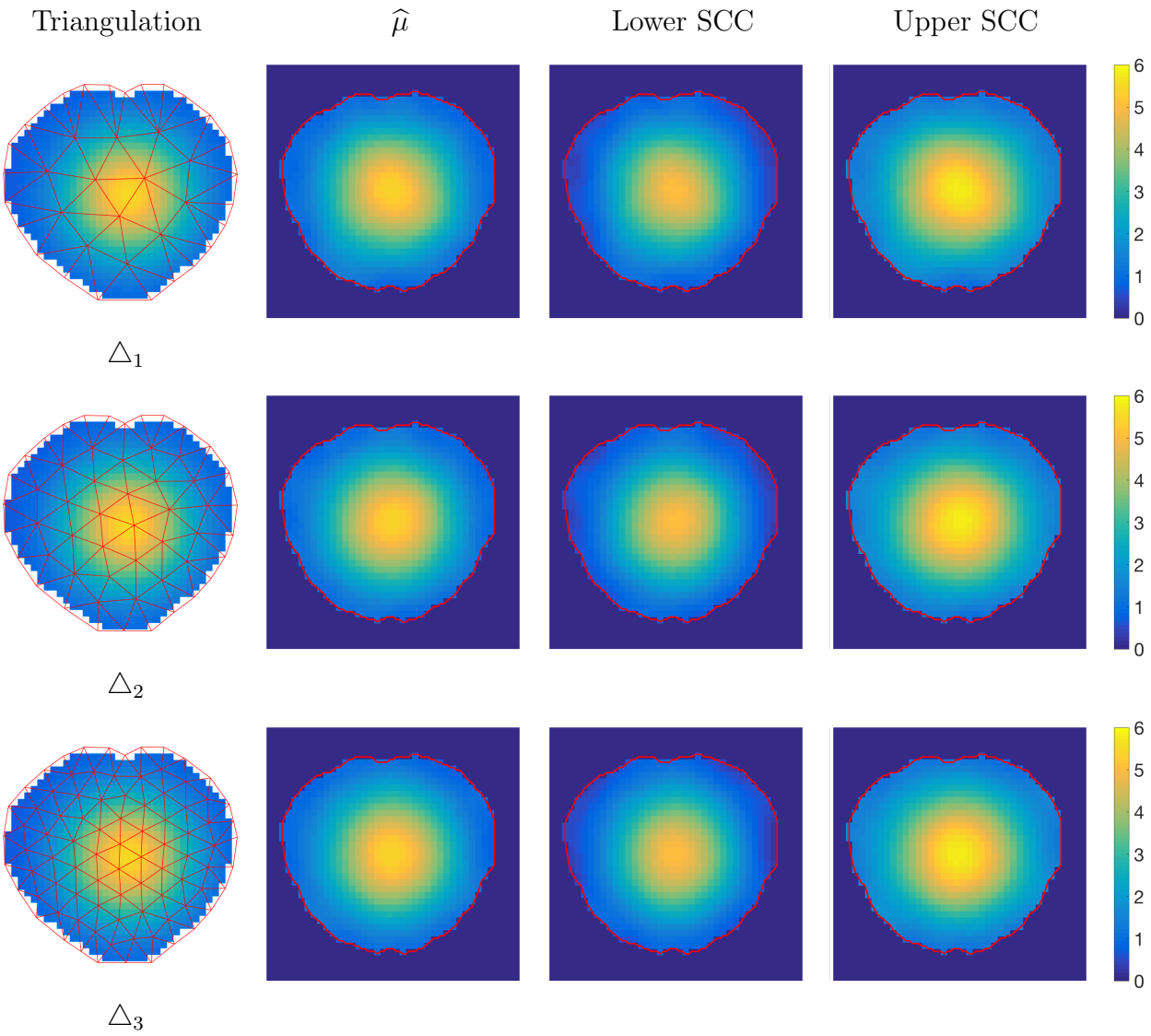



**Figure 4.** SCCs for Cubic Function with $n = 50$ and $\alpha = 0.01$.

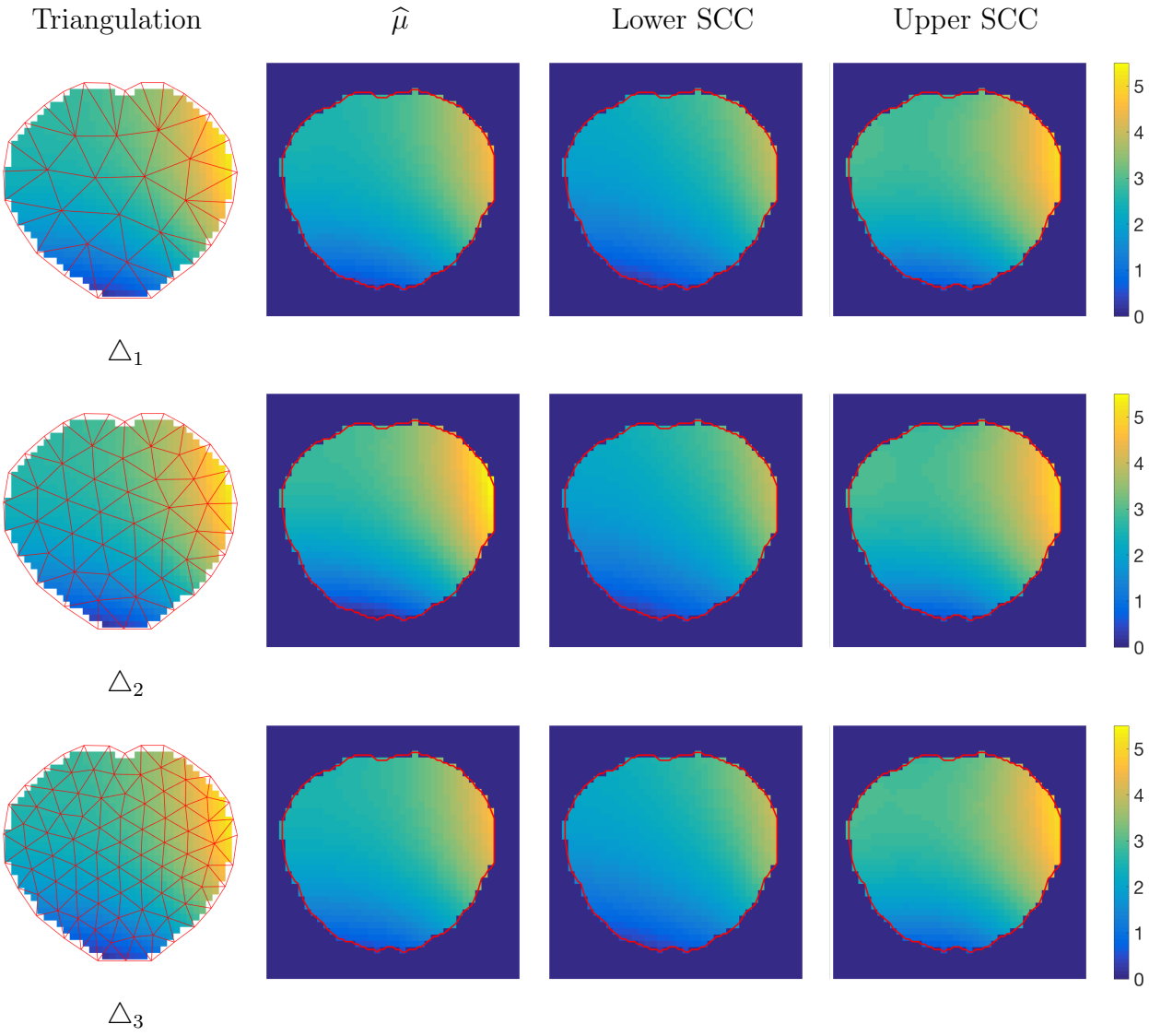



**Figure 5.** SCCs for Exponential Function with $n = 50$ and $\alpha = 0.01$.

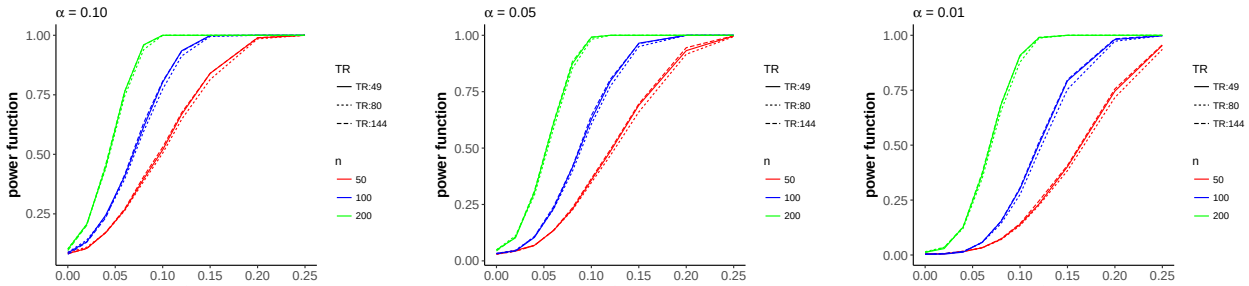
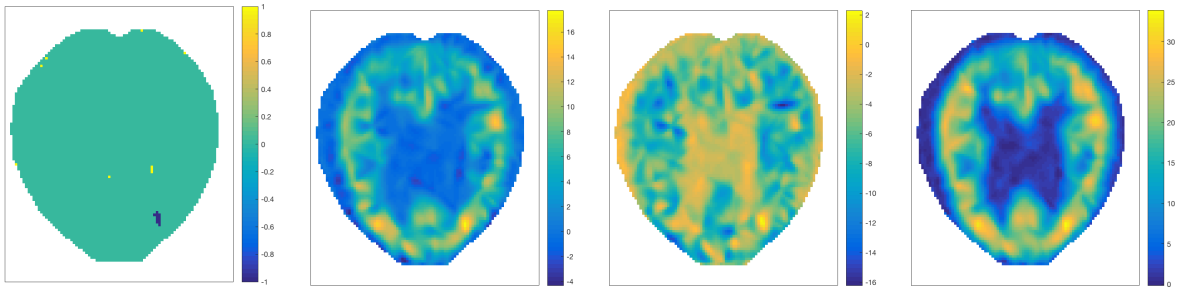



**Figure 6.** Type I error and empirical power of two sample test for different $\alpha$'s.



(a) Coverage of 0 (b) $\widehat{\mu}_{\text{MDD}} - \widehat{\mu}_{\text{CON}}$ (c) Lower SCC (d) Upper SCC

**Figure 7.** SCC for comparison between CON and MDD.

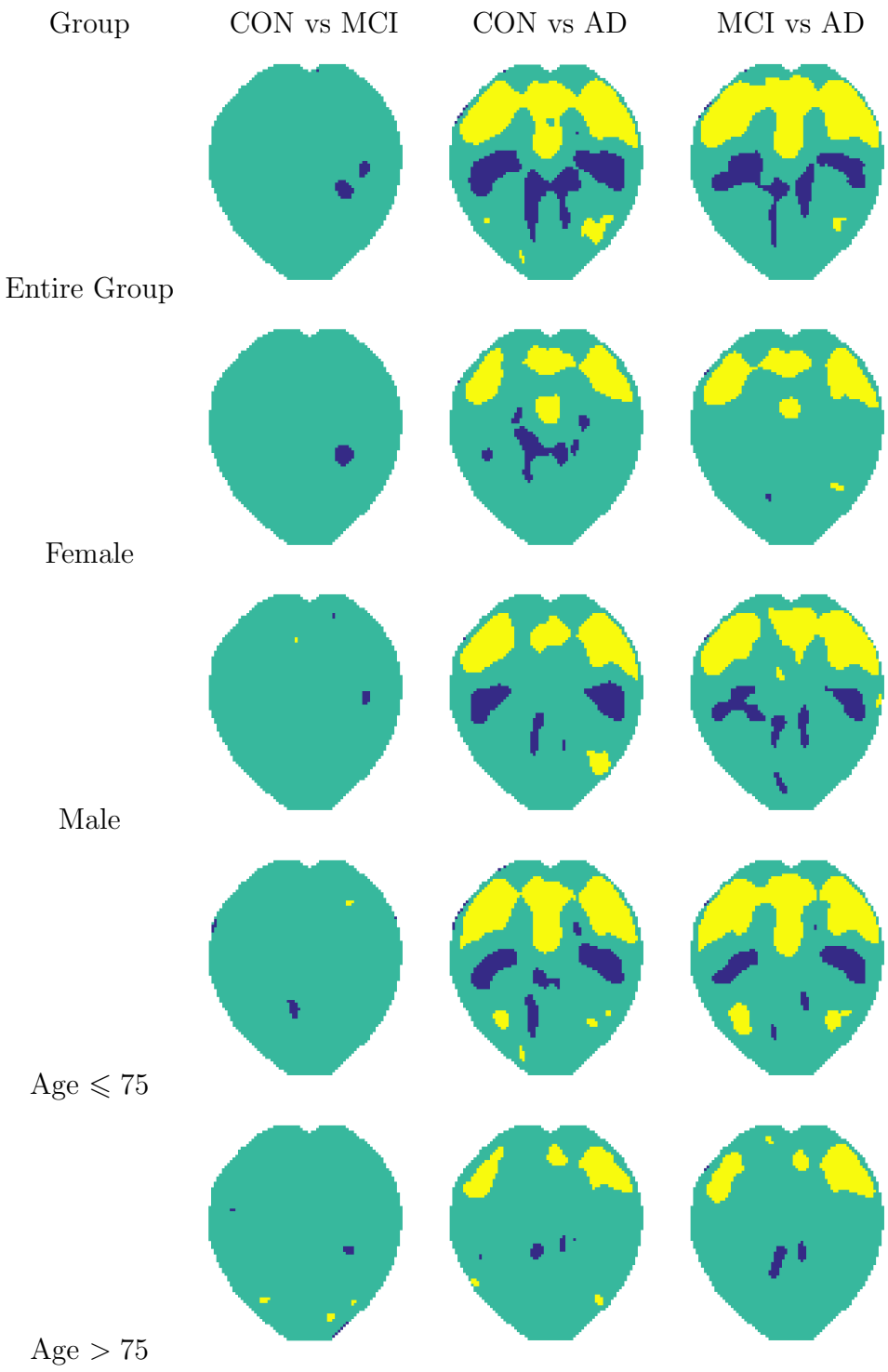



**Figure 8.** Coverage of 0 of SCC for pairwise comparisons among CON, MCI and AD. (Yellow color indicates 0 falls above the upper band and blue color indicates 0 falls beneath the lower band.)

**Table 1**
*Empirical coverage rates of the SCCs.*

| $n$ | $\alpha = 0.10$ | | | $\alpha = 0.05$ | | | $\alpha = 0.01$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $\triangle_1$ | $\triangle_2$ | $\triangle_3$ | $\triangle_1$ | $\triangle_2$ | $\triangle_3$ | $\triangle_1$ | $\triangle_2$ | $\triangle_3$ |
| $\mu = 20\{(x-0.5)^2 + (y-0.5)^2\}$ | | | | | | | | | |
| 50 | 0.875 | 0.882 | 0.878 | 0.936 | 0.938 | 0.940 | 0.980 | 0.980 | 0.982 |
| | (0.618) | (0.619) | (0.619) | (0.706) | (0.707) | (0.707) | (0.872) | (0.872) | (0.873) |
| 100 | 0.895 | 0.892 | 0.899 | 0.943 | 0.943 | 0.945 | 0.980 | 0.976 | 0.978 |
| | (0.441) | (0.441) | (0.441) | (0.503) | (0.503) | (0.503) | (0.623) | (0.623) | (0.623) |
| 200 | 0.908 | 0.899 | 0.907 | 0.954 | 0.952 | 0.957 | 0.987 | 0.987 | 0.987 |
| | (0.313) | (0.313) | (0.313) | (0.357) | (0.357) | (0.357) | (0.441) | (0.441) | (0.441) |
| $\mu = 5\exp[-15\{(x-0.5)^2 + (y-0.5)^2\}] + 0.5$ | | | | | | | | | |
| 50 | 0.828 | 0.840 | 0.836 | 0.909 | 0.913 | 0.914 | 0.969 | 0.968 | 0.972 |
| | (0.619) | (0.619) | (0.619) | (0.706) | (0.707) | (0.707) | (0.872) | (0.873) | (0.873) |
| 100 | 0.844 | 0.853 | 0.854 | 0.910 | 0.914 | 0.915 | 0.970 | 0.972 | 0.974 |
| | (0.441) | (0.441) | (0.441) | (0.503) | (0.503) | (0.503) | (0.623) | (0.623) | (0.623) |
| 200 | 0.868 | 0.862 | 0.853 | 0.930 | 0.928 | 0.929 | 0.983 | 0.979 | 0.982 |
| | (0.313) | (0.313) | (0.313) | (0.357) | (0.357) | (0.357) | (0.441) | (0.441) | (0.442) |
| $\mu = 3.2(-x^3 + y^3) + 2.4$ | | | | | | | | | |
| 50 | 0.828 | 0.853 | 0.852 | 0.910 | 0.926 | 0.921 | 0.972 | 0.972 | 0.973 |
| | (0.619) | (0.619) | (0.619) | (0.706) | (0.707) | (0.707) | (0.872) | (0.873) | (0.873) |
| 100 | 0.836 | 0.849 | 0.847 | 0.921 | 0.932 | 0.926 | 0.971 | 0.973 | 0.974 |
| | (0.441) | (0.441) | (0.441) | (0.503) | (0.503) | (0.503) | (0.623) | (0.623) | (0.623) |
| 200 | 0.865 | 0.880 | 0.876 | 0.934 | 0.938 | 0.937 | 0.979 | 0.984 | 0.979 |
| | (0.313) | (0.313) | (0.313) | (0.357) | (0.357) | (0.357) | (0.441) | (0.441) | (0.441) |

**Table 2**
*Type I Error and Empirical Power of Two Sample Test*

| $n$ | $d$ | $\triangle$ | $\delta$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0.00 | 0.02 | 0.04 | 0.06 | 0.08 | 0.10 | 0.12 | 0.15 | 0.20 | 0.25 |
| $\alpha = 0.10$ | | | | | | | | | | | | |
| 50 | 2 | 49 | 0.083 | 0.105 | 0.171 | 0.267 | 0.397 | 0.519 | 0.668 | 0.840 | 0.989 | 1.000 |
| | | 80 | 0.088 | 0.11 | 0.171 | 0.264 | 0.387 | 0.505 | 0.647 | 0.814 | 0.985 | 1.000 |
| | | 144 | 0.083 | 0.104 | 0.173 | 0.271 | 0.407 | 0.530 | 0.675 | 0.840 | 0.990 | 1.000 |
| 100 | 2 | 49 | 0.086 | 0.133 | 0.245 | 0.409 | 0.616 | 0.803 | 0.935 | 0.999 | 1.000 | 1.000 |
| | | 80 | 0.088 | 0.140 | 0.240 | 0.396 | 0.597 | 0.774 | 0.913 | 0.995 | 1.000 | 1.000 |
| | | 144 | 0.080 | 0.131 | 0.234 | 0.415 | 0.632 | 0.807 | 0.933 | 0.998 | 1.000 | 1.000 |
| 200 | 2 | 49 | 0.101 | 0.207 | 0.448 | 0.764 | 0.960 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 80 | 0.103 | 0.206 | 0.441 | 0.745 | 0.940 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 144 | 0.096 | 0.202 | 0.458 | 0.764 | 0.959 | 1.000 | 1.000 | 1.000 | 1.000 | 1.000 |
| $\alpha = 0.05$ | | | | | | | | | | | | |
| 50 | 2 | 49 | 0.032 | 0.046 | 0.069 | 0.134 | 0.229 | 0.358 | 0.485 | 0.690 | 0.932 | 0.997 |
| | | 80 | 0.029 | 0.043 | 0.069 | 0.134 | 0.225 | 0.347 | 0.467 | 0.660 | 0.916 | 0.994 |
| | | 144 | 0.030 | 0.043 | 0.068 | 0.135 | 0.235 | 0.365 | 0.493 | 0.696 | 0.945 | 0.998 |
| 100 | 2 | 49 | 0.033 | 0.044 | 0.104 | 0.231 | 0.414 | 0.627 | 0.797 | 0.965 | 1.000 | 1.000 |
| | | 80 | 0.033 | 0.047 | 0.107 | 0.228 | 0.397 | 0.606 | 0.774 | 0.950 | 1.000 | 1.000 |
| | | 144 | 0.031 | 0.044 | 0.108 | 0.239 | 0.416 | 0.644 | 0.805 | 0.963 | 1.000 | 1.000 |
| 200 | 2 | 49 | 0.049 | 0.102 | 0.304 | 0.607 | 0.874 | 0.991 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 80 | 0.049 | 0.111 | 0.291 | 0.585 | 0.853 | 0.982 | 1.000 | 1.000 | 1.000 | 1.000 |
| | | 144 | 0.044 | 0.106 | 0.310 | 0.612 | 0.883 | 0.992 | 1.000 | 1.000 | 1.000 | 1.000 |
| $\alpha = 0.01$ | | | | | | | | | | | | |
| 50 | 2 | 49 | 0.005 | 0.006 | 0.015 | 0.033 | 0.072 | 0.139 | 0.232 | 0.402 | 0.746 | 0.954 |
| | | 80 | 0.005 | 0.006 | 0.018 | 0.032 | 0.070 | 0.134 | 0.226 | 0.384 | 0.717 | 0.936 |
| | | 144 | 0.003 | 0.005 | 0.016 | 0.033 | 0.074 | 0.143 | 0.246 | 0.408 | 0.756 | 0.957 |
| 100 | 2 | 49 | 0.003 | 0.004 | 0.013 | 0.058 | 0.153 | 0.301 | 0.504 | 0.793 | 0.981 | 0.998 |
| | | 80 | 0.004 | 0.006 | 0.015 | 0.056 | 0.143 | 0.275 | 0.472 | 0.754 | 0.973 | 0.998 |
| | | 144 | 0.004 | 0.007 | 0.016 | 0.057 | 0.154 | 0.306 | 0.514 | 0.798 | 0.984 | 0.999 |
| 200 | 2 | 49 | 0.012 | 0.029 | 0.127 | 0.365 | 0.684 | 0.909 | 0.989 | 1.000 | 1.000 | 1.000 |
| | | 80 | 0.015 | 0.035 | 0.121 | 0.347 | 0.656 | 0.878 | 0.987 | 1.000 | 1.000 | 1.000 |
| | | 144 | 0.012 | 0.034 | 0.128 | 0.373 | 0.688 | 0.905 | 0.991 | 1.000 | 1.000 | 1.000 |

**Table 3**
*Two way table of diagnosis vs. gender and age group.*

| | | Diagnosis | | | |
|---|---|---|---|---|---|
| | | CON | MCI | AD | Total |
| Gender | Female | 42 | 77 | 50 | 169 |
| | Male | 70 | 136 | 72 | 278 |
| Age | Age $\leqslant 75$ | 54 | 107 | 60 | 221 |
| | Age $> 75$ | 58 | 106 | 62 | 226 |
| Total | | 112 | 213 | 122 | 447 |