The Annals of Applied Statistics
2016, Vol. 10, No. 4, 1854–1879
DOI: 10.1214/16-AOAS975

SMOOTH PRINCIPAL COMPONENT ANALYSIS OVER TWO-DIMENSIONAL MANIFOLDS WITH AN APPLICATION TO NEUROIMAGING

BY EARDI LILA[*,†], JOHN A. D. ASTON[*,1] AND LAURA M. SANGALLI[†]

University of Cambridge[*] and Politecnico di Milano[†]

Motivated by the analysis of high-dimensional neuroimaging signals located over the cortical surface, we introduce a novel Principal Component Analysis technique that can handle functional data located over a two-dimensional manifold. For this purpose a regularization approach is adopted, introducing a smoothing penalty coherent with the geodesic distance over the manifold. The model introduced can be applied to any manifold topology, and can naturally handle missing data and functional samples evaluated in different grids of points. We approach the discretization task by means of finite element analysis, and propose an efficient iterative algorithm for its resolution. We compare the performances of the proposed algorithm with other approaches classically adopted in literature. We finally apply the proposed method to resting state functional magnetic resonance imaging data from the Human Connectome Project, where the method shows substantial differential variations between brain regions that were not apparent with other approaches.

1. Introduction. The recent growth of data arising from neuroimaging has led to profound changes in the understanding of the brain. Neuroimaging is a multidisciplinary activity and the role of statistics in its success should not be underestimated. Much of the work to date has been to determine how to use statistical models in high-dimensional settings that arise out of such imaging modalities as functional Magnetic Resonance Imaging (fMRI) and Electroencephalography (EEG). However, it is becoming increasingly clear that there is now a need to incorporate more and more complex information about brain structure and function into the statistical analysis to enhance our present understanding of the brain.

Considerable amounts of the brain signal captured, for example, by fMRI arise from the cerebral cortex. The cerebral cortex is the highly convoluted thin sheet where most neural activity is focused. It is natural to represent this thin sheet as a 2D surface embedded in a 3D space, structured with a 2D geodesic distance, rather than the 3D Euclidean distance within the volume. In fact, functionally distinct areas may be close to each other if measured with Euclidean distance, but, due

Received December 2015; revised August 2016.

[1]Supported in part by the Engineering and Physical Sciences Research Council (Grants EP/K021672/2, EP/N014588/1).

Key words and phrases. Functional data analysis, principal component analysis, differential regularization, functional magnetic resonance imaging.

to the highly convoluted morphology of the cerebral cortex, their 2D geodesic distance along the cortical surface can be far greater. While early approaches to the analysis of hemodynamic signals ignore the morphology of the cortical surface, it has now been well established [Glasser et al. (2013) and references therein] that it is beneficial to analyze neuroimaging data through the processing of the signals on the cortical surface using surface-constrained techniques. Classical tools such as nonparametric smoothing models have already been adapted to deal with this kind of data; see, for example, Chung, Hanson and Pollak (2014).

The goal of the present paper is to introduce a novel Principal Component Analysis (PCA) technique suitable for working with functional signals distributed over curved domains and specifically over two-dimensional smooth Riemannian manifolds, such as the cortical surface. The cortical surface can be extracted from structural Magnetic Resonance Imaging (MRI), a noninvasive scanning technique used to visualize the internal structure of the brain, rendering it as a 3D image with high spatial resolution. The signal of interest, which we want to analyze with respect to the surface, comes from fMRI, which detects a Blood Oxygen Level Dependent (BOLD) signal [Ogawa et al. (1990)] as a series of repeated measurements in time, yielding a time series of 3D images. An increased neural activity in a particular area of the brain causes an increased demand for oxygen. As the fMRI signal is related to changes in the relative ratio of oxy- to deoxy-hemoglobin, due to their differing magnetic properties, the signal captured within an fMRI scan is considered to be a surrogate for neural activity and is used to produce activation maps or investigate brain functional connectivity. The fMRI signal of each individual related to the neural activity in the cerebral cortex is generally mapped on a common template cortical surface to allow multi-subject statistical analysis.

In this paper, in particular, we will focus our attention on functional connectivity (FC). FC maps, on the cortical surface, can be constructed computing the pairwise correlation between all vertex's fMRI time-series and the mean time-series of a region of interest. The resulting FC map for each subject provides a clear view of areas to which the region of interest is functionally connected.

In practice, the template cortical surface is represented by a triangulated surface that can be considered a discrete approximation of the underlying smooth compact two-dimensional Riemannian manifold $\mathcal{M} \subset \mathbb{R}^3$ modeling the cortical surface. Each resting state FC map can be represented by a function $x_i : \mathcal{M} \to \mathbb{R}$. Once we have the correlation maps on the cortical surface we want to study how the phenomena varies from subject to subject. A statistical technique for this study is PCA. It is natural to contextualize this task in the framework of Functional Data Analysis [Ramsay and Silverman (2005)].

In Section 2 we establish the formal theoretical properties of Functional PCA (FPCA) in the case of random functions whose domain is a manifold $\mathcal{M}$. In Section 3 we introduce a novel FPCA model and propose an algorithm for its resolution. We then give some simulation results in Section 4, indicating the performance of our methodology as compared to other methods in literature. We then return to

the FC maps example in Section 5 to consider how the surface based PCA analysis might be used in this case, and draw some concluding remarks in Section 6.

2. Functional principal component analysis. Consider the space of square integrable functions on $\mathcal{M}$: $L^2(\mathcal{M}) = \{f : \mathcal{M} \to \mathbb{R} : \int_{\mathcal{M}} |f(p)|^2\, dp < \infty\}$ with the inner product $\langle f, g\rangle_{\mathcal{M}} = \int_{\mathcal{M}} f(p)g(p)\, dp$ and norm $\|f\|_{\mathcal{M}} = \int_{\mathcal{M}} |f(p)|^2\, dp$. Consider the random variable X with values in $L^2(\mathcal{M})$, mean $\mu = \mathbb{E}[X]$ and a finite second moment, that is, $\int_{\mathcal{M}} \mathbb{E}[X^2] < \infty$, and assume that its covariance function $K(p,q) = \mathbb{E}[(X(p) - \mu(p))(X(q) - \mu(q))]$ is square integrable. Mercer's lemma [Riesz and Sz.-Nagy (1955)] guarantees the existence of a nonincreasing sequence (κ_j) of eigenvalues of K and an orthonormal sequence of corresponding eigenfunctions (ψ_j), such that

$$\int_{\mathcal{M}} K(p,q)\psi_j(p)\, dp = \kappa_j \psi_j(q), \qquad \forall q \in \mathcal{M} \tag{2.1}$$

and that $K(p,q)$ can be written as $K(p,q) = \sum_{j=1}^{\infty} \kappa_j \psi_j(p)\psi_j(q)$ for each $p, q \in \mathcal{M}$. Thus, X can be expanded as $X = \mu + \sum_{j=1}^{\infty} \varepsilon_j \psi_j$, where the random variables $\varepsilon_1, \varepsilon_2, \ldots$ are uncorrelated and are given by $\varepsilon_j = \int_{\mathcal{M}} \{X(p) - \mu(p)\}\psi_j(p)\, dp$. This is also known as the Karhunen–Loève (KL) expansion of X.

The collection (ψ_j) defines the strongest modes of variation in the random function X, and these are called Principal Component (PC) functions. In fact, ψ_1 is such that

$$\psi_1 = \operatorname*{argmax}_{\phi : \|\phi\|_{\mathcal{M}} = 1} \int_{\mathcal{M}} \int_{\mathcal{M}} \phi(p)K(p,q)\phi(q)\, dp\, dq,$$

while ψ_m, for $m > 1$, solves an analogous problem with the added constraint of ψ_m being orthogonal to the previous $m-1$ functions $\psi_1, \ldots, \psi_{m-1}$, that is,

$$\psi_m = \operatorname*{argmax}_{\substack{\phi : \|\phi\|_{\mathcal{M}} = 1 \\ \langle \phi, \psi_j \rangle_{\mathcal{M}} = 0,\, j = 1, \ldots, m-1}} \int_{\mathcal{M}} \int_{\mathcal{M}} \phi(p)K(p,q)\phi(q)\, dp\, dq.$$

The random variables $\varepsilon_1, \varepsilon_2, \ldots$ are called PC scores.

Another important property of PC functions is the best M basis approximation. In fact, for any fixed $M \in \mathbb{N}$, the first M PC functions of X satisfy

$$(\psi_i)_{m=1}^{M} = \operatorname*{argmin}_{(\{\phi_m\}_{m=1}^{M} : \langle \phi_m, \phi_l \rangle = \delta_{ml})} \mathbb{E} \int_{\mathcal{M}} \left\{ X - \mu - \sum_{m=1}^{M} \langle X, \phi_m \rangle \phi_m \right\}^2, \tag{2.2}$$

where δ_{ml} is the Kronecker delta; that is, $\delta_{ml} = 1$ for $m = l$ and 0 otherwise.

Suppose $x_1, \ldots, x_n$ are n smooth samples from X. Usually, for each of these functions, only noisy evaluations $x_i(p_j)$ on a fixed discrete grid of points $p_1, \ldots, p_s$ are given. In this setting, we will now recall the two standard approaches to FPCA: the presmoothing approach and the regularized PCA approach.

The presmoothing approach is based on the two following steps. In the first step, the observations associated to each function are smoothed in order to obtain smooth representations of $x_1, \ldots, x_n$. Then the sample mean $\bar{x} = n^{-1} \sum_i x_i$ and the sample covariance $\hat{K}(p,q) = \frac{1}{n} \sum_{i=1}^n (x_i(p) - \bar{x}(p))(x_i(q) - \bar{x}(q))$ are used to estimate μ and K, respectively. Finally, the estimates of the PC functions $\hat{\psi}_1, \hat{\psi}_2, \ldots$ are computed through the characterization $\int_{\mathcal{M}} \hat{K}(p,q)\hat{\psi}_j(p)\,dp = \hat{\kappa}_j \hat{\psi}_j(q)$, which is solved by the discretization of the problem on a fine grid or by the basis expansion of estimated smooth functions. In the case where the domain is an interval of the real line, a theoretical study on the accuracy of $\hat{\psi}_j$ as an estimate of ψ_j is offered, for example, in Hall and Hosseini-Nasab (2006).

Define the $n \times s$ matrix $\mathbf{X} = (x_i(p_j))$, the column vector $\boldsymbol{\mu} = (\frac{1}{n} \sum_{i=1}^n x_i(p_j))$ of length s, the $n \times M$ matrix $\mathbf{A} = (\langle X_i, \phi_m \rangle)$ and the $s \times M$ matrix $\boldsymbol{\Phi} = (\phi_m(p_j))$. Let $\mathbf{1}$ denote the column vector of length n with all entries equal to 1. The empirical counterpart of the objective function in (2.2) becomes

$$\frac{1}{n} \| \mathbf{X} - \mathbf{1}\boldsymbol{\mu}^T - \mathbf{A}\boldsymbol{\Phi}^T \|_F^2, \tag{2.3}$$

where $\| \cdot \|_F$ is the Frobenius norm, defined as the square root of the sum of the squares of its elements. This last formulation gives a natural way to deal with the fact that only pointwise and noisy evaluations $x_i(p_j), i = 1, \ldots, n, j = 1, \ldots, s$, of the underlying functional samples are usually available. However, it does not incorporate any information on the smoothness of the functional data. In fact, considering the Singular Value Decomposition (SVD) of $\mathbf{X} - \mathbf{1}\boldsymbol{\mu}^T = \mathbf{U}\mathbf{D}\mathbf{V}^T$, it can be shown that the minimizing arguments of (2.3) are $\hat{\boldsymbol{\Phi}} = \mathbf{V}$ and $\hat{\mathbf{A}} = \mathbf{U}\mathbf{D}$, thus the obtained formulation is a multivariate PCA applied to the data matrix $\mathbf{X}$.

The regularized PCA approach consists of adding a penalization term to the classic formulation of the PCA in order to recover a desired feature of the estimated underlying functions. In particular, the formulation (2.3) has shown a great flexibility for this purpose. Examples of models where a sparseness property is assumed on the data are offered, for instance, in Jolliffe, Trendafilov and Uddin (2003), Shen and Huang (2008), Zou and Hastie (2005). In the specific case of functional data analysis, the penalization term usually encourages the PC functions to be smooth. Examples of PCA models that explicitly incorporate a smoothing penalization term are given by Huang, Shen and Buja (2008), Rice and Silverman (1991), Silverman (1996). The cited works deal with functions whose domain is a limited interval in $\mathbb{R}$, and, in particular, our proposal can be seen as an extension of Huang, Shen and Buja (2008) to the case of functions whose domain is a two-dimensional manifold. Zhou and Pan (2014) recently proposed a smooth FPCA for two-dimensional functions on irregular planar domains; their approach is based on a mixed effects model that specifies the PC functions as bivariate splines on triangulations and the PC scores as random effects. Here we propose a FPCA model that can handle real functions observable on a two-dimensional manifold. We shall

consider a smoothing penalty operator, coherent with the 2D geodesic distances on the manifold. This leads to the definition of a model that can fully exploit the information about the geometry of the manifold.

3. Smooth FPCA over two-dimensional manifolds.

3.1. *Geometric concepts.* We first introduce the essential geometric concepts that allow the definition of the Laplace–Beltrami operator, which plays a central role in the proposed model. In detail, let the bijective and smooth function $\varphi : U \subset \mathbb{R}^2 \to \mathbb{R}^3$ be a local parametrization of $\mathcal{M}$ around the point $p \in \mathcal{M}$, as depicted in Figure 1. Let $\theta \in U$ be such that $\theta = \varphi^{-1}(p)$; then

$$\left\{ \frac{\partial \varphi}{\partial \theta_i}(\theta) \right\}_{i=1,2} \tag{3.1}$$

defines a basis for the tangent space $T_p\mathcal{M}$ at the point p.

The Riemannian manifold $\mathcal{M}$ can be equipped with a metric by defining a scalar product g_p on the tangent space $T_p\mathcal{M}$. This enables, for instance, the computation of the lengths of curves or integrals on the surface. Fixing the reference system on the tangent plane with the basis (3.1), we can represent g_p as the matrix $G = (g_{ij})_{i,j=1,2}$ such that

$$g_p(v, w) = \sum_{i,j=1}^{2} g_{ij} v_i w_j$$

for all $v = \sum v_i \frac{\partial \varphi}{\partial \theta_i}(\theta)$ and $w = \sum w_i \frac{\partial \varphi}{\partial \theta_i}(\theta)$. In our case it is natural to consider the scalar product induced by the Euclidean embedding space $\mathbb{R}^3$, that is, the first

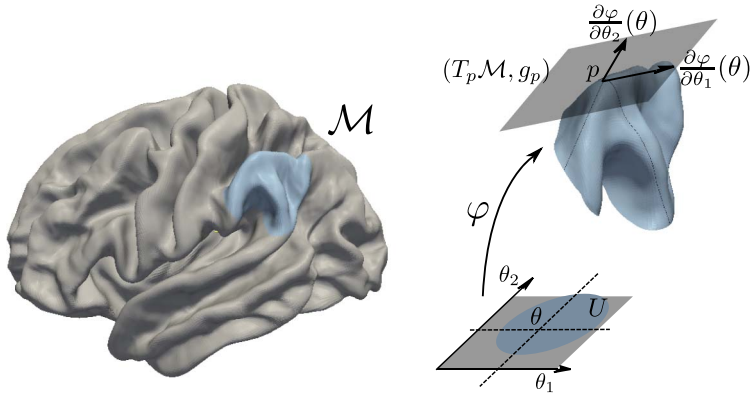

FIG. 1. *A pictorial representation of the geometric objects modeling the idealized cortical surface* $\mathcal{M}$.

fundamental form

$$g_{ij}(\theta) = \frac{\partial \varphi}{\partial \theta_i}(\theta) \cdot \frac{\partial \varphi}{\partial \theta_j}(\theta),$$

where $\cdot$ denotes the inner product in $\mathbb{R}^3$. Moreover, we denote by $G^{-1} = (g^{ij})_{i,j=1,2}$ the inverse of the matrix G and by $g = \det(G)$ the determinant of the matrix G.

Let now $f : \mathcal{M} \to \mathbb{R}$ be a real-valued and twice differentiable function on the manifold $\mathcal{M}$. Let $F = f \circ \varphi$, then the gradient $\nabla_{\mathcal{M}} f$ is defined as

$$(\nabla_{\mathcal{M}} f)(p) = \sum_{i,j=1}^{2} g^{ij}(\theta) \frac{\partial F}{\partial \theta_j}(\theta) \frac{\partial \varphi}{\partial \theta_j}(\theta).$$

In the case of a flat manifold $\mathcal{M}$, the last expression reduces to the expression of the gradient in $\mathbb{R}^2$, that is, $\nabla = (\frac{\partial}{\partial \theta_1}, \frac{\partial}{\partial \theta_2})$.

The Laplace–Beltrami operator $\Delta_{\mathcal{M}}$ is a generalization to the case of surfaces of the standard Laplacian defined on $\mathbb{R}^n$, that is, $\Delta = \sum_{i=1}^{n} \frac{\partial^2}{\partial^2 \theta_i}$. It is related to the second partial derivatives of f on $\mathcal{M}$, that is, its local curvature, and it is defined as

$$(\Delta_{\mathcal{M}} f)(p) = \frac{1}{\sqrt{g(\theta)}} \sum_{i,j=1}^{2} \frac{\partial}{\partial \theta_j} g^{ij} \sqrt{g(\theta)} \frac{\partial F}{\partial \theta_j}(\theta).$$

The defined operator is invariant with respect to rigid transformations of the reference system on U.

3.2. *Model.* Suppose now the sample of n functions $x_i : \mathcal{M} \to \mathbb{R}$ is observed at a fixed set of points $p_1, \ldots, p_s$ in $\mathcal{M}$ (this will be relaxed later). Let $\mathbf{u} = \{u_i\}_{i=1,\ldots,n}$ be an n-dimensional real column vector. We propose to estimate the first PC function $\hat{f} : \mathcal{M} \to \mathbb{R}$ and the associated PC scores vector $\hat{\mathbf{u}}$ by solving the equation

$$(\hat{\mathbf{u}}, \hat{f}) = \underset{\mathbf{u}, f}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{s} \big(x_i(p_j) - u_i f(p_j)\big)^2 + \lambda \mathbf{u}^T \mathbf{u} \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f, \tag{3.2}$$

where the Laplace–Beltrami operator is integrated over the manifold $\mathcal{M}$, enabling a global roughness penalty on f, while the empirical term encourages f to capture the strongest mode of variation. The parameter λ controls the trade-off between the empirical term of the objective function and the roughness penalizing term. The $\mathbf{u}^T \mathbf{u}$ term is justified by some invariance considerations on the objective function, as done in the case of one-dimensional domains in Huang, Shen and Buja (2008). Consider the transformation $(\mathbf{u} \to c\mathbf{u}, f \to \frac{1}{c} f)$, with c a constant, and the transformation $(\mathbf{X} \to c\mathbf{X}, \mathbf{u} \to c\mathbf{u})$, where $\mathbf{X} = (x_i(p_j))$. Then the objective function

in (3.2) is invariant with respect to the first transformation, while the empirical and the smoothness terms are rescaled by the same coefficient with the second transformation.

The subsequent PCs can be extracted sequentially by removing the preceding estimated components from the data matrix $\mathbf{X}$. This allows the selection of a different penalization parameter λ for each PC estimate. We will refer to the model introduced as the Smooth Manifold FPCA (SM-FPCA).

3.3. *Iterative algorithm.* Here we present the numerical algorithm for the resolution of the model introduced above. Our approach for the minimization of the functional (3.2) can be summarized in the following two steps:

- Splitting the optimization in a finite-dimensional optimization in $\mathbf{u}$ and an infinite-dimensional optimization in f;
- Approximating the infinite-dimensional solution thanks to a Surface Finite Element discretization.

Let $\mathbf{f}_s$ be the vector of length s such that $\mathbf{f}_s = (f(p_1), \ldots, f(p_s))^T$. The expression in (3.2) can be rewritten as

$$(\hat{\mathbf{u}}, \hat{f}) = \underset{\mathbf{u}, f}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{u}\mathbf{f}_s^T\|_F^2 + \lambda \mathbf{u}^T \mathbf{u} \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f. \tag{3.3}$$

A normalization constraint must be considered in this minimization problem to make the representation unique, as in fact multiplying $\mathbf{u}$ by a constant and dividing f by the same constant does not change the objective function (3.3). In particular, we set the constraint $\|\mathbf{u}\|_2 = 1$, as this allows us to leave the infinite-dimensional optimization in f unconstrained.

Our proposal for the minimization of the criterion (3.3) is to alternate the minimization of $\mathbf{u}$ and f in an iterative algorithm:

Step 1. Estimation of $\mathbf{u}$ given f. For a given f, the minimizing $\mathbf{u}$ of the objective function in (3.3) is

$$\mathbf{u} = \frac{\mathbf{X}\mathbf{f}_s}{\|\mathbf{f}_s\|_2^2 + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f}, \tag{3.4}$$

and the minimizing unitary-norm vector $\mathbf{u}$ is

$$\mathbf{u} = \frac{\mathbf{X}\mathbf{f}_s}{\|\mathbf{X}\mathbf{f}_s\|_2}. \tag{3.5}$$

Step 2. Estimation of f given $\mathbf{u}$. For a given $\mathbf{u}$, solving (3.3) with respect to f is equivalent to finding the f that minimizes

$$J_{\lambda,\mathbf{u}}(f) = \mathbf{f}_s^T \mathbf{f}_s + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f - 2\mathbf{f}_s^T \mathbf{X}^T \mathbf{u}. \tag{3.6}$$

Step 1 is basically the classical expression of the score vector given the loadings vector, where in this case the loading vector is given by $\mathbf{f}_s$, the evaluations of the PC function in $p_1, \ldots, p_s$. The problem in Step 2 is not trivial, consisting in an infinite-dimensional minimization problem. Let z_j denote the jth element of the vector $\mathbf{X}^T\mathbf{u}$; then minimizing the functional in (3.6) is equivalent to minimizing

$$\sum_{j=1}^{s}(z_j - f(p_j))^2 + \lambda \int_{\mathcal{M}} \Delta^2_{\mathcal{M}} f. \tag{3.7}$$

This problem involves estimating a smooth field f defined on a manifold, starting from noisy observations z_j at points p_j. In the case of real functions defined on the real line, adopting a penalty of the form $\lambda \int f''$, the minimization problem turns out to have a finite-dimensional closed-form solution that is a cubic spline [Green and Silverman (1993)]. For real functions defined on an Euclidean space, cubic splines are generalized by thin-plate splines. In this case, for an opportune smoothing penalty, the solution of the minimization problem can be expressed in terms of a finite linear combination of radial basis functions [Duchon (1977)].

However, the case of real functions defined on a non-Euclidean domain $\mathcal{M}$ is more involved. In the special case where $\mathcal{M}$ is a sphere or a sphere-like surface, that is, $\mathcal{M} = \{\sigma(v) = \rho(v)v : v \in S\}$ where $S \subset \mathbb{R}^3$ is the unit sphere centered at the origin, this smoothing problem has been considered, among others, by Wahba (1981) and Alfeld, Neamtu and Schumaker (1996). Moreover, the functional (3.7) is considered, among others, by Ettinger, Perotto and Sangalli (2016) and Dassi et al. (2015). Here $\mathcal{M}$ is respectively a manifold homeomorphic to an open-ended cylinder and a manifold homeomorphic to a sphere. In the latter two works the field f is estimated by first conformally recasting the problem to a planar domain and then discretizing it by means of planar finite elements, generalizing the planar smoothing model in Ramsay (2002). Our approach is also based on a Finite Element (FE) discretization, but, differently from Ettinger, Perotto and Sangalli (2016) and Dassi et al. (2015), we construct here an FE space directly on the triangulated surface $\mathcal{M}_\mathcal{T}$ that approximates the manifold $\mathcal{M}$, that is, we use surface FE, avoiding any flattening step and thereby allowing the formulation to be applicable to any manifold topology.

3.4. *Surface Finite Element discretization.* Assume, for clarity of exposition only, that $\mathcal{M}$ is a closed surface, as in our motivating application. The case of non-closed surfaces can be handled by considering some appropriate boundary conditions as done, for instance, in the planar case in Sangalli, Ramsay and Ramsay (2013). Consider the linear functional space $H^2(\mathcal{M})$, the space of functions in $L^2(\mathcal{M})$ with first and second weak derivatives in $L^2(\mathcal{M})$. The infinite-dimensional part of the estimation problem can be reformulated as follows: find $\hat{f} \in H^2(\mathcal{M})$ such that

$$\hat{f} = \operatorname*{argmin}_{f \in H^2(\mathcal{M})} J_{\lambda,\mathbf{u}}(f). \tag{3.8}$$

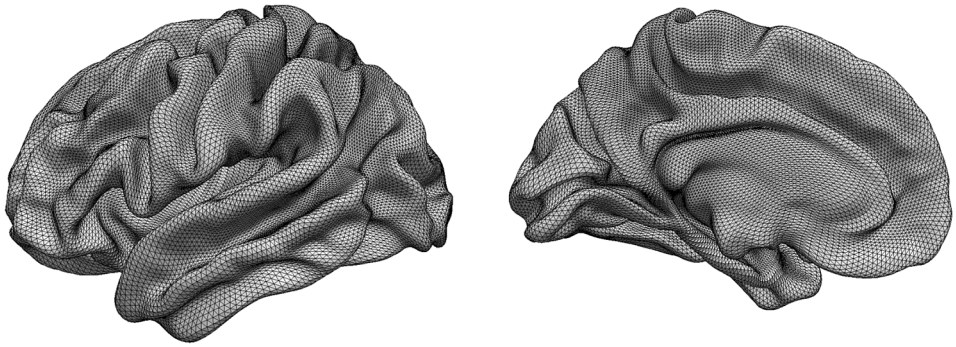

FIG. 2. *The triangulated surface approximating the left hemisphere of the template brain. The mesh is composed by* $32K$ *nodes and by* $64K$ *triangles.*

PROPOSITION 1. *The solution* $\hat{f} \in H^2(\mathcal{M})$ *exists and is unique and is such that*

$$\sum_{j=1}^{s} \varphi(p_j)\hat{f}(p_j) + \lambda \int_{\mathcal{M}} \Delta_{\mathcal{M}}\varphi \Delta_{\mathcal{M}}\hat{f} = \sum_{j=1}^{s} \varphi(p_j) \sum_{i=1}^{n} x_i(p_j)u_i \tag{3.9}$$

for every $\varphi \in H^2(\mathcal{M})$.

As detailed in the Supplementary Material [Lila, Aston and Sangalli (2017)], the key idea is to minimize $J_{\lambda,\mathbf{u}}(f)$ by differentiating this functional with respect to f. This leads to (3.9), which characterizes the estimate $\hat{f}$ as the solution of a linear fourth-order problem.

Consider now a triangulated surface $\mathcal{M}_{\mathcal{T}}$, the union of the finite set of triangles $\mathcal{T}$, giving an approximated representation of the manifold $\mathcal{M}$. Figure 2, for instance, shows the triangulated surface approximating the left hemisphere of a template brain. We then consider the linear finite element space V consisting in a set of globally continuous functions over $\mathcal{M}_{\mathcal{T}}$ that are linear affine where restricted to any triangle τ in $\mathcal{T}$, that is,

$$V = \{v \in C^0(\mathcal{M}_{\mathcal{T}}) : v|_\tau \text{ is linear affine for each } \tau \in \mathcal{T}\}.$$

This space is spanned by the nodal basis $\psi_1, \ldots, \psi_K$ associated to the nodes $\xi_1, \ldots, \xi_K$, corresponding to the vertices of the triangulation $\mathcal{M}_{\mathcal{T}}$. Such basis functions are lagrangian, meaning that $\psi_i(\xi_j) = 1$ if $i = j$ and $\psi_i(\xi_j) = 0$ otherwise. Setting $\mathbf{f} = (f(\xi_1), \ldots, f(\xi_K))^T$ and $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_K)^T$, every function $f \in V$ has the form

$$f(p) = \sum_{k=1}^{K} f(\xi_k)\psi_k(p) = \mathbf{f}^T\boldsymbol{\psi}(p) \tag{3.10}$$

for each $p \in \mathcal{M}_{\mathcal{T}}$. The surface finite element space provides a finite-dimensional subspace of $H^1(\mathcal{M})$ [Dziuk (1988)]. To use this finite element space to discretize

the infinite-dimensional problem (3.9), that is well posed in $H^2(\mathcal{M})$, we first need a reformulation of (3.9) that involves only first-order derivatives. This can be obtained by introducing an auxiliary function g that plays the role of $\Delta_{\mathcal{M}} f$, splitting the equation (3.9) into a coupled system of second-order problems and finally integrating by parts the second-order terms. The details of this derivation can be found in the Supplementary Material. The discrete estimators $\hat{f}_h, \hat{g}_h \in V$ are then obtained by solving

$$\begin{cases} \displaystyle\int_{\mathcal{M}_\mathcal{T}} \nabla_{\mathcal{M}_\mathcal{T}} \hat{f}_h \nabla_{\mathcal{M}_\mathcal{T}} \varphi_h - \int_{\mathcal{M}_\mathcal{T}} \hat{g}_h \varphi_h = 0, \\ \displaystyle\lambda \int_{\mathcal{M}_\mathcal{T}} \nabla_{\mathcal{M}_\mathcal{T}} \hat{g}_h \nabla_{\mathcal{M}_\mathcal{T}} v_h + \sum_{j=1}^{s} \hat{f}_h(p_j) v_h(p_j) = \sum_{j=1}^{s} v_h(p_j) \sum_{i=1}^{n} x_i(p_j) u_i \end{cases} \tag{3.11}$$

for all $\varphi_h, v_h \in V$. Define the $s \times K$ matrix $\boldsymbol{\Psi} = (\psi_k(p_j))$ and the $K \times K$ matrices $\mathbf{R}_0 = \int_{\mathcal{M}_\mathcal{T}} (\boldsymbol{\psi}\boldsymbol{\psi}^T)$ and $\mathbf{R}_1 = \int_{\mathcal{M}_\mathcal{T}} (\nabla_{\mathcal{M}_\mathcal{T}} \boldsymbol{\psi})(\nabla_{\mathcal{M}_\mathcal{T}} \boldsymbol{\psi})^T$. Then, exploiting the representation (3.10) of functions in V, we can rewrite (3.11) as a linear system. Specifically, the Finite Element solution $\hat{f}_h(p)$ of the discrete counterpart (3.11) is given by $\hat{f}_h(p) = \boldsymbol{\psi}(p)^T \hat{\mathbf{f}}$, where $\hat{\mathbf{f}}$ is the solution of

$$\begin{bmatrix} \boldsymbol{\Psi}^T\boldsymbol{\Psi} & \lambda \mathbf{R}_1 \\ \lambda \mathbf{R}_1 & -\lambda \mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{f}} \\ \hat{\mathbf{g}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi}^T \mathbf{X}^T \mathbf{u} \\ \mathbf{0} \end{bmatrix}. \tag{3.12}$$

Solving (3.12) leads to

$$\hat{\mathbf{f}} = \left(\boldsymbol{\Psi}^T\boldsymbol{\Psi} + \lambda \mathbf{R}_1 \mathbf{R}_0^{-1} \mathbf{R}_1\right)^{-1} \boldsymbol{\Psi}^T \mathbf{X}^T \mathbf{u}. \tag{3.13}$$

Although this last formula is a compact expression of the solution, it is preferable to compute the solution from the linear system (3.12) due to the sparsity property of the matrix in the left-hand side. As an example, in the simulations and the application shown in Sections 4–5, respectively, less then 1% and less then 0.1% of the elements in the matrix in the left-hand side of (3.12) are different from zero, allowing a very efficient solution of the linear system.

In the model introduced, we assume that all the observed functions x_i are sampled on the common set of points $p_1, \ldots, p_s \in \mathcal{M}$. Suppose, moreover, $p_1, \ldots, p_s \in \mathcal{M}$ coincide with the vertices of the triangulated surface $\mathcal{M}_\mathcal{T}$. In this particular case, an alternative approach could consist of interpreting the points $p_1, \ldots, p_s \in \mathcal{M}_\mathcal{T}$ as the nodes of a graph linked by the edges of the triangulation and considering the model (3.2) with a discrete smoothness operator term instead of the Laplace–Beltrami operator [see, e.g., Belkin and Niyogi (2001) for the choice of the penalization term and Cai et al. (2011) for an application to matrix decomposition]. However, thanks to its functional nature, the formulation (3.2) can be easily extended to the case of missing data or sparsely sampled functional data. Specifically, suppose now that each function x_i is observable on a set of points

$p_1^i, \ldots, p_{s_i}^i$; then the natural extension of the model (3.2) becomes

$$(\hat{\mathbf{u}}, \hat{f}) = \underset{\mathbf{u}, f}{\operatorname{argmin}} \sum_{i=1}^{n} \sum_{j=1}^{s_i} \left(x_i(p_j^i) - u_i f(p_j^i)\right)^2 + \lambda \mathbf{u}^T \mathbf{u} \int_{\mathcal{M}} \Delta_{\mathcal{M}}^2 f. \tag{3.14}$$

Following the same procedure, we can define an analogous algorithm based on the following two steps:

Step 1. For a given f, the unitary-norm vector $\mathbf{u}$ minimizing (3.14) is given by

$$\mathbf{u} \text{ such that } u_i = \frac{\sum_{j=1}^{s_i} x_i(p_j^i) f(p_j^i)}{\sqrt{\sum_{i=1}^{n} (\sum_{j=1}^{s_i} x_i(p_j^i) f(p_j^i))^2}}.$$

Step 2. For a given $\mathbf{u}$, the function f minimizing (3.14) is given by $f = \mathbf{f}^T \boldsymbol{\psi}$ with $\mathbf{f}$ such that

$$\begin{bmatrix} \mathbf{L} & \lambda \mathbf{R}_1 \\ \lambda \mathbf{R}_1 & -\lambda \mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} \mathbf{D}^T \mathbf{u} \\ \mathbf{0} \end{bmatrix},$$

where

$$\mathbf{L} = \begin{bmatrix} \sum_{i=1}^{n} \sum_{j=1}^{s_i} u_i^2 \psi_1(p_j^i) \psi_1(p_j^i) & \cdots & \sum_{i=1}^{n} \sum_{j=1}^{s_i} u_i^2 \psi_1(p_j^i) \psi_K(p_j^i) \\ & \cdots & \\ \sum_{i=1}^{n} \sum_{j=1}^{s_i} u_i^2 \psi_K(p_j^i) \psi_1(p_j^i) & \cdots & \sum_{i=1}^{n} \sum_{j=1}^{s_i} u_i^2 \psi_K(p_j^i) \psi_K(p_j^i) \end{bmatrix},$$

$$\mathbf{D} = \begin{bmatrix} \sum_{j=1}^{s_1} \psi_1(p_j^1) x_1(p_j^1) & \cdots & \sum_{j=1}^{s_n} \psi_1(p_j^n) x_n(p_j^n) \\ & \cdots & \\ \sum_{j=1}^{s_1} \psi_K(p_j^1) x_1(p_j^1) & \cdots & \sum_{j=1}^{s_n} \psi_K(p_j^n) x_n(p_j^n) \end{bmatrix}.$$

3.5. *SM-FPCA algorithm*. The algorithm for the resolution of the model SM-FPCA (3.2) can be summarized in the following steps.

Algorithm 1 SM-FPCA algorithm

1: Initialization:

(a) Computation of $\boldsymbol{\Psi}$, $\mathbf{R}_0$ and $\mathbf{R}_1$.
(b) Perform the SVD: $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}^T$.
(c) $\mathbf{f}_s \leftarrow \mathbf{V}[:, 1]$, where $\mathbf{V}[:, 1]$ are the loadings of the first PC.

2: Scores estimation:

$$\mathbf{u} \leftarrow \frac{\mathbf{X}\mathbf{f}_s}{\|\mathbf{X}\mathbf{f}_s\|_2}.$$

3: PC function's estimation: $\mathbf{f}$ such that

$$\begin{bmatrix} \boldsymbol{\Psi}^T\boldsymbol{\Psi} & \lambda\mathbf{R}_1 \\ \lambda\mathbf{R}_1 & -\lambda\mathbf{R}_0 \end{bmatrix} \begin{bmatrix} \mathbf{f} \\ \mathbf{g} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Psi}^T\mathbf{X}^T\mathbf{u} \\ \mathbf{0} \end{bmatrix}.$$

4: PC function's evaluation:

$$\mathbf{f}_s \leftarrow \boldsymbol{\Psi}^T\mathbf{f}.$$

5: Repeat Steps 2–4 until convergence.
6: Normalization:

$$\hat{f}(p) \leftarrow \frac{\mathbf{f}^T\boldsymbol{\psi}(p)}{\|\mathbf{f}^T\boldsymbol{\psi}\|_{L^2(\mathcal{M}_\mathcal{T})}}.$$

The problems (3.2)–(3.14) are nonconvex minimization problems in $(\mathbf{u}, f)$. However, in the previous section we proved the existence and uniqueness of the minimizing f given $\mathbf{u}$ and vice-versa. This implies that the objective function is nonincreasing under the update rules of Algorithm 1. Since the first guess of the PC function, given by the SVD, is usually a good starting point, in all our simulations no convergence problem has been detected.

3.6. *Parameters selection.* The SM-FPCA model has a smoothing parameter $\lambda > 0$ that adjusts the trade-off between the fidelity of the estimate to the data, via the sum of the squared errors, and the smoothness of the solution, via the penalty term. The problem of choosing the smoothing parameter is common to all smoothing problems.

The flexibility given by the smoothing parameter can be seen as an advantageous feature; by varying the smoothing parameter, the data can be explored on different scales. However, in many cases a data-driven automatic method is necessary. In the following simulations we consider two different criteria. The first approach consists of a K-fold cross-validation. The data matrix $\mathbf{X}$ is partitioned by rows into K roughly equal groups. For each group of data $k = 1, \ldots, K$, the dataset

can be split into a validation set $\mathbf{X}^k$, composed of the elements of the kth group, and a training set, composed of the remaining elements. For different smoothing parameters, the loading function f^{-k} is estimated from the training dataset. Given the estimated loading function f^{-k}, the associated score vector $\mathbf{u}^k$ is computed on the validation dataset. Since f^{-k} has been computed on the training dataset, $\mathbf{u}^k$ should be computed on the validation dataset via the formula (3.4), where $\int_{\mathcal{M}} \Delta_{\mathcal{M}}^2$ can be approximated by $\mathbf{g}^T\mathbf{R}_0\mathbf{g}$, with $g_h(p) = \boldsymbol{\psi}(p)^T\mathbf{g}$ the auxiliary function approximating $\Delta_{\mathcal{M}} f$. Finally, we select the value of the parameter λ that minimizes the following score:

$$\mathrm{CV}(\lambda) = \sum_{k=1}^{K} \frac{\sum_{i=1}^{n}\sum_{j=1}^{s}(x_i(p_j) - u_i^k f^{-k}(p_j))^2}{np}. \tag{3.15}$$

The second approach is based on the minimization of a generalized cross-validation (GCV) criteria integrated on the regression step of the iterative algorithm. Setting $\mathbf{S}(\lambda) = \boldsymbol{\Psi}^T(\boldsymbol{\Psi}^T\boldsymbol{\Psi} + \lambda\mathbf{R}_1\mathbf{R}_0^{-1}\mathbf{R}_1)^{-1}\boldsymbol{\Psi}^T$, the GCV score is defined as

$$\mathrm{GCV}(\lambda) = \frac{1}{s}\frac{\|(\mathbf{I} - \mathbf{S}(\lambda))(\mathbf{X}^T\mathbf{u})\|^2}{(1 - \frac{1}{s}\operatorname{tr}\{\mathbf{S}(\lambda)\})^2}.$$

The GCV score represents the average misfit of the regression model with a leave-one-out cross-validation strategy on the observations' vector $\mathbf{X}^T\mathbf{u}$. However, excluding the ith element from the vector $\mathbf{X}^T\mathbf{u}$ can be interpreted as removing the ith column from the data matrix $\mathbf{X}$. Thus, in terms of the data-matrix, this strategy can be interpreted as a leave-one-column-out cross-validation strategy, as opposed to the K-fold, where the data matrix $\mathbf{X}$ is partitioned by rows. The GCV approach is generally faster than the K-fold approach. However, the K-fold does not require the inversion of any matrix. This is an advantageous feature, since generally the inverse of a sparse matrix is not sparse. It is thus applicable also to datasets $\mathbf{X}$ with a large number of columns s.

3.7. *Total explained variance.* Another parameter that must be chosen is the number of PCs that satisfactorily reduces the dimension of the data. A classical approach consists of selecting this parameter on the basis of cumulated explained variance of the PC. While in the ordinary PC the scores vectors are uncorrelated and their loadings are orthogonal, in our formulation neither the loadings are explicitly imposed to be orthogonal nor the PC scores to be uncorrelated. It is nevertheless possible to define an index of explained variance as follows. Let $\hat{\mathbf{U}}$ be the $n \times k$ matrix such that the columns of $\hat{\mathbf{U}}$ are the first k PC scores vectors. Since in our estimation procedure the PC scores are normalized to have unitary norm, the variance of the PCs is captured by the PC functions. It is thus necessary to consider here the unnormalized PC scores, obtained by multiplying each score vector by the norm of the associated PC function. Without the uncorrelation assumption,

it is meaningless to compute the total variance explained by the first k PCs by $\operatorname{tr}(\hat{\mathbf{U}}^T\hat{\mathbf{U}})$. To overcome this problem, Zou, Hastie and Tibshirani (2006) propose to remove linear dependence between correlated PC scores vectors by regression projection. Thus, they compute the QR decomposition of $\hat{\mathbf{U}}$ as $\hat{\mathbf{U}} = \mathbf{QR}$ and define the *adjusted total variance* as $\sum_{j=1}^{k} \mathbf{R}_{jj}^2$, where $\mathbf{R}_{jj}$ represents the variance explained by the jth PC that is not already explained by the previous $j-1$ components.

4. Simulation studies. In this section we conduct simulations to assess the performance of the SM-FPCA algorithm compared to other methods.

We consider as domain of the functional observations a triangulated surface $\mathcal{M}_{\mathcal{T}}$ with 642 nodes that approximates the brainstem. On this triangulated surface we generate the orthonormal functions $\{v_l\}_{l=1,2,3}$, consisting in three eigenfunctions of the Laplace–Beltrami operator, as shown in Figure 3. These functions represent the first three PC functions. We then generate $n = 50$ smooth functions $x_1, \ldots, x_{50}$ on $\mathcal{M}_{\mathcal{T}}$ by

$$x_i = u_{i1}v_1 + u_{i2}v_2 + u_{i3}v_3, \qquad i = 1, \ldots, n, \tag{4.1}$$

where u_{i1}, u_{i2}, u_{i3} are independent random variables that represent the scores and are distributed as $u_{il} \sim \mathcal{N}(0, \sigma_l^2)$, with $\sigma_1 = 5$, $\sigma_2 = 3$ and $\sigma_3 = 1$. The smooth functions x_i are then sampled at locations $p_j \in \mathbb{R}^3$ with $j = 1, \ldots, s$ coinciding with the nodes of the triangulated surface. Moreover, at each of these points we add to the functions a Gaussian noise with mean zero and standard deviation $\sigma = 0.1$ to obtain the noisy observations denoted with $x_i(p_j)$. We are thus interested in recovering the smooth PC functions $\{v_l\}_{l=1,2,3}$ from these noisy observations over $\mathcal{M}_{\mathcal{T}}$. We compare the proposed SM-FPCA technique to two alternative approaches.

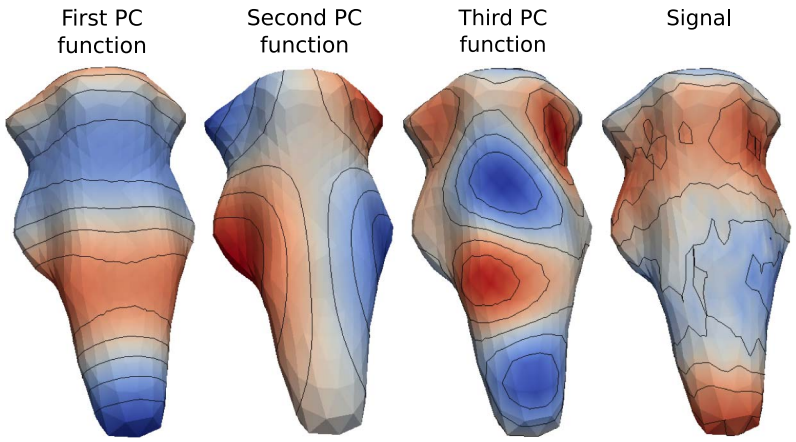

FIG. 3. *From left to right, a plot of the true first, second and third PC functions and a plot of a noisy observation on the brainstem generated from these three PC functions.*

The first basic approach we consider is a simple multivariate PCA (MV-PCA) applied to the data matrix $\mathbf{X}$. The PC functions are thus obtained by piecewise linear interpolation over the mesh $\mathcal{M}_\mathcal{T}$. Finally, they are normalized to have unitary norm in $L^2(\mathcal{M}_\mathcal{T})$.

A second natural approach is based on a presmoothing of the noisy observations that tries to recover the smooth functions $x_i, i = 1, \ldots, n$, from their noisy observations $x_i(p_j)$, followed by a MV-PCA on the denoised evaluations of the functions on $p_j, j = 1, \ldots, s$. The smoothing problem for a field defined on a Riemannian manifold is not trivial. In this case the smoothing technique applied is Iterated Heat Kernel (IHK) smoothing [Chung et al. (2005)]. The heat kernel smoothing of the noisy observation $x_i(p_j)$ is given by $K_\eta \times x_i(p_j) = \int_\mathcal{M} K_\eta(p, q)x_i(p_j)\,dq$, where η is the smoothing parameter and K_η is the heat kernel, whose analytic expression can be extracted from the eigenfunctions of the Laplace–Beltrami operator. However, for numerical approximation, it can be shown that, for η small and for q close to p, we have

$$K_\eta(p, q) \approx \frac{1}{(2\pi\eta)^{\frac{1}{2}}} \exp\left[-\frac{d^2(p, q)}{2\eta^2}\right].$$

The desired level of smoothing can be reached after k iterations, thanks to the following property: $K_\eta^k \times f = K_\eta \times \cdots \times K_\eta \times f = K_{\sqrt{k}\eta}$. For a fixed bandwidth η, the level of smoothing is determined by an optimal number of iterations selected via the F-test criterion outlined in Chung et al. (2005). In these simulations, the bandwidth has been set at $\eta = 2.5$, heuristically selecting the one with the best performance after some initial pilot studies. We refer to this approach as IHK-PCA.

The proposed SM-FPCA technique is implemented as follows. For each PC we run Algorithm 1 with 15 iterations of steps 2–4. For the choice of the optimal smoothing parameter λ, both K-fold, with $K = 5$, and GCV approaches have been applied.

The reconstructed PC functions using the three different approaches are shown in Figure 4. It is evident that applying the MV-PCA yields to a reconstruction far from the truth because of the absence of any spatial information. The reconstruction through the IHK-PCA approach and the SM-FPCA model is considerably more satisfactory. In Figure 5 we show the plots with the cumulative percentage of explained variance, where, in the case of SM-FPCA, the explained variance has been computed as detailed in Section 3.6.

While the poor performance of the MV-PCA is evident, to assess the performance of the other two methods, we apply them to 100 datasets generated as previously detailed. The quality of the estimated individual surfaces is then measured using the mean square error (MSE) over all the locations $p_j, j = 1, \ldots, s$. MSEs are also used to evaluate the reconstruction of the PC scores vectors. Another performance measure used is the principal angle between the subspace spanned by

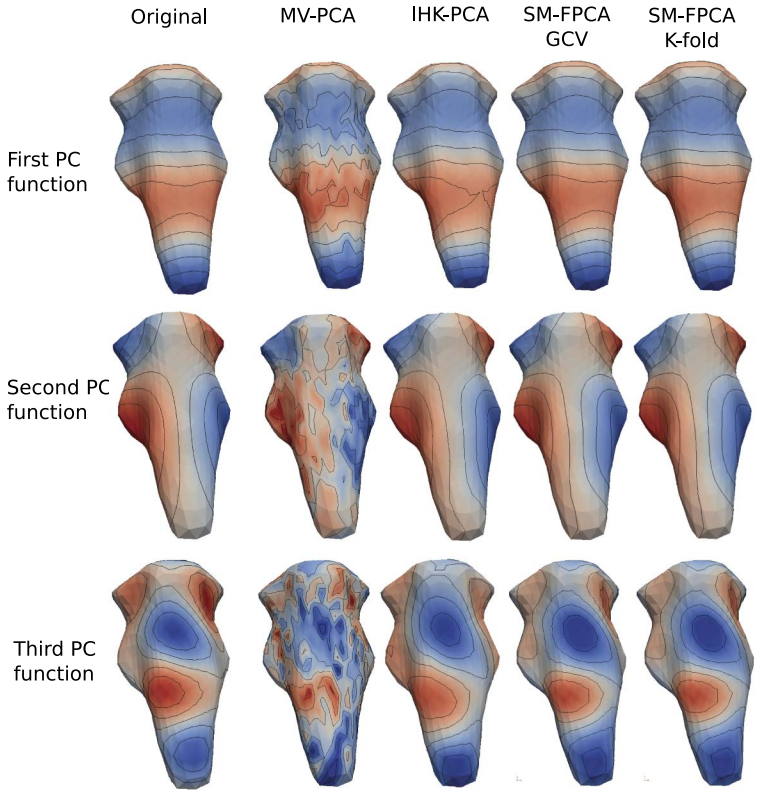

FIG. 4. *From left to right, contours of the original PC functions and their estimates, respectively, with MV-PCA, IHK-PCA, SM-FPCA GCV and SM-FPCA K-fold. From a visual inspection, MV-PCA shows unsatisfactory results, while a better estimation is achieved by IHK-PCA and SM-FPCA. In particular, SM-FPCA is able to better capture details that IHK-PCA ignores. This is apparent, for instance, in the third PC function reconstruction in the top-left and top-right corners.*

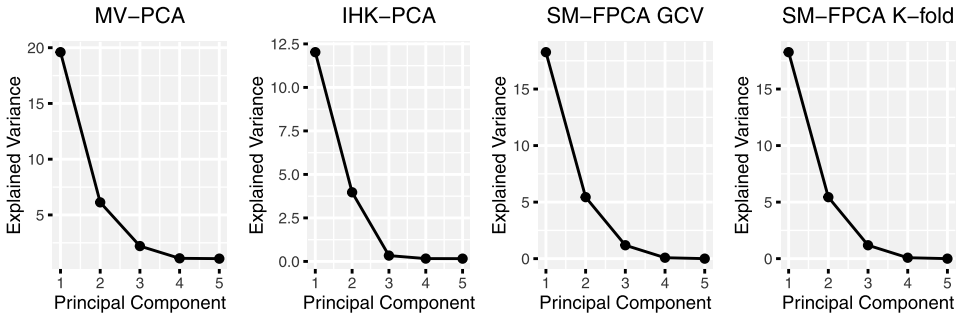

FIG. 5. *From left to right, plot of the empirical variances explained by the first* 5 *PCs computed with MV-PCA, IHK-PCA, SM-FPCA GCV and SM-FPCA K-fold.*

the estimated PC functions and the subspace spanned by the true PC functions, as used in Shen and Huang (2008). Intuitively, the principal angle measures how similar the two subspaces are. For this purpose we construct the $s \times 3$ matrices $\mathbb{V} = (v_i(p_j))$ and $\hat{\mathbb{V}} = (\hat{v}_i(p_j))$, where $\hat{v}_i$ is the ith estimate of the true PC function v_i. Then we compute the orthonormal set of basis $\mathbf{Q}_{\mathbb{V}}$ and $\mathbf{Q}_{\hat{\mathbb{V}}}$ from the QR decomposition of $\mathbb{V}$ and $\hat{\mathbb{V}}$. The principal angle is defined as the angle $\cos^{-1}(\rho)$, where ρ is the minimum singular value of $\mathbf{Q}_{\hat{\mathbb{V}}}^T \mathbf{Q}_{\mathbb{V}}$. The results are summarized in the boxplots in Figure 6, which compares the MV-PCA, IHK-PCA and SM-FPCA algorithms with respect to the reconstruction's errors of the PC functions $\{v_l\}_{l=1,2,3}$, the PC scores $\{\mathbf{u}_l\}_{l=1,2,3}$ where $\mathbf{u}_l = (u_{il})$, the reconstructed signals

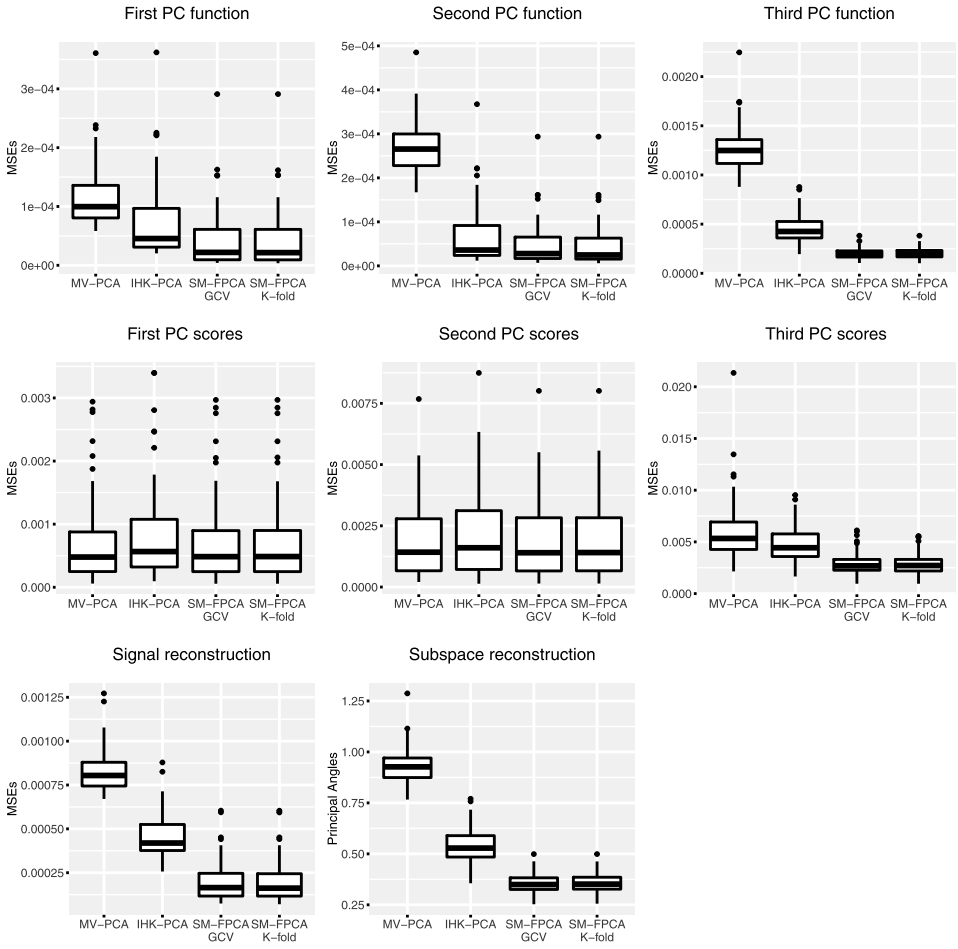

FIG. 6. *Boxplots summarizing the performance of IHK-PCA and SM-FPCA. For the SM-FPCA, both GCV and K-fold have been applied for the selection of the smoothing parameter.*

$x_i = u_{i1}v_1 + u_{i2}v_2 + u_{i3}v_3$ for $i = 1, \ldots, 50$, and the principal angles between the subspaces spanned by the true and estimated PC functions.

The boxplots highlight the fact that SM-FPCA provides the best estimates of the PC functions, corresponding scores vectors, signals and subspace reconstruction.

5. Application. The dataset which we consider in this paper arises from the Human Connectome Project Consortium [HCP, Essen et al. (2012)], which is collecting data such as structural scans, resting-state and task-based functional MRI scans, and diffusion-weighted MRI scans from a large number of healthy volunteers to help elucidate normal brain function. Many preprocessing considerations have already been resolved in the so-called minimally preprocessed dataset. Among the various preprocessing pipelines applied to the HCP original data, of particular interest for us is the one named *fMRISurface* [Glasser et al. (2013)]. This pipeline provides a transformation of the 3D structural MRI and 4D signal from the functional MRI scan so as to enable the application of statistical analysis techniques on brain surfaces. For each subject, the personal cortical surface is extracted as a triangulated surface from the structural MRI and to each vertex of this mesh is associated a BOLD time-series derived from the BOLD signal of the underlying gray-matter ribbon. The extracted cortical surfaces are aligned to a template cortical surface generated from the cortical surfaces of 69 healthy adults. In practice, this cortical surface is represented by two triangulated surfaces with 32K vertices, one for each hemisphere. In Figure 2 the left hemisphere is shown. Through this anatomical transformation map, the patients' BOLD time-series, on the cortical surface, are coherently located to the vertices of the template cortical surface. This, of course, raises questions about the implications of anatomical alignment, and a small simulation study in the supplementary material investigates this issue. The fMRI signal used for our analysis has been acquired in the absence of any task, and for this reason is also called the resting state fMRI. Finally, each time-series is filtered to the band of frequencies $[0.009, 0.08]$ Hz. Summarizing, the data considered are fMRI filtered time-series on a common triangulated template mesh.

As already mentioned in Section 1, a classic approach in the study of the resting state fMRI is to exploit the time dimension of the data for the extraction of a connectivity measure among the different parts of the cortical surface. A standard choice for this purpose is the computation of the temporal correlation. It first consists of identifying a Region of Interest (ROI) on the cortical surface. This is the area whose behavior, as compared to the rest of the cortical surface, is of interest for the investigator. Within each subject, a cross-sectional average of all the time-series in the ROI is used to find a representative mean time-series. To each vertex of the cortical surface we associate the pairwise correlation of the time-series located in that vertex with the subject-specific time-series representative of the ROI. Finally, each correlation value is transformed using Fisher's r-to-z transformation,

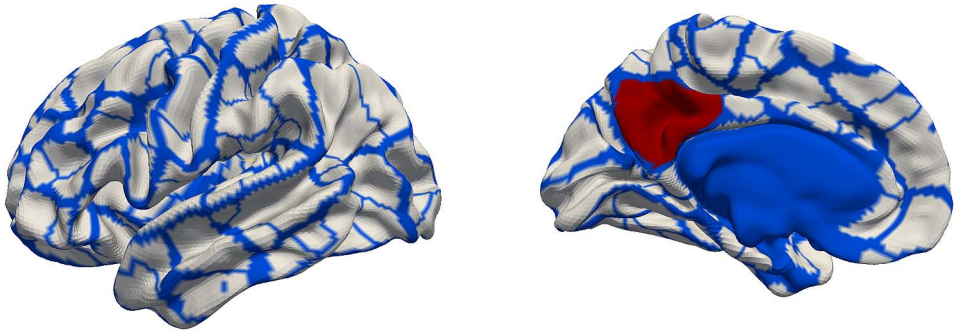

FIG. 7. *Parcellation of the cortical surface derived in Gordon et al. (2014). In red the Region of Interest chosen for the computation of the RSFC maps. This region is localized on an area of the cerebral cortex called the precuneus. The blue colors indicate the parcellated regions, with the major blue area being the part of the brain that connects the two brain hemispheres, which is not part of the cortical surface and which is therefore excluded from the analysis.*

yielding a resting state functional connectivity (RSFC) map for each subject. The total number of subjects considered for this analysis is 491.

For the choice of the ROI, we consider the cortical parcellation derived in Gordon et al. (2014), where a group-average boundary map of the cortical surface is derived from resting state fMRI (Figure 7). The identified cortical areas are unlikely to correspond to the individual parcellation of each subject since they are derived from a group average study. However, they can serve as reasonable ROIs in individual subjects. The parcel that served as ROI in the following analysis is highlighted in red in Figure 7. For the chosen ROI, a snapshot of the RSFC map of one subject is shown in Figure 8.

The mean RSFC map is shown in Figure 9. As expected, high correlation values are visible inside the ROI. The mean RSFC over 491 subjects shows a variability coherent with the parcellation, in the sense that the vertices inside each parcel show similar values. We wish now to understand which are the main modes of variation of these RSFC maps among the different subjects by applying a PCA.

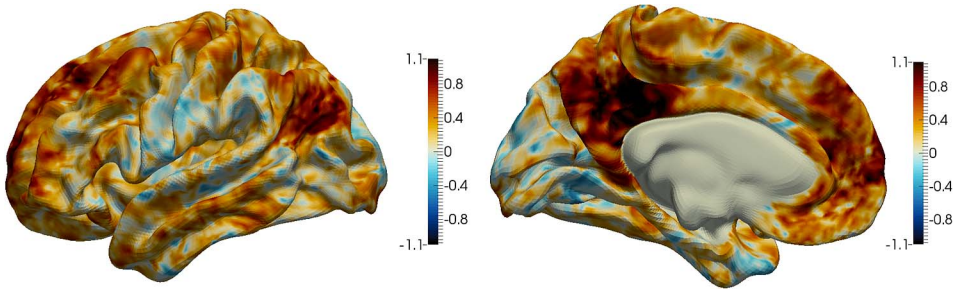

FIG. 8. *A snapshot of the RSFC map of one subject.*

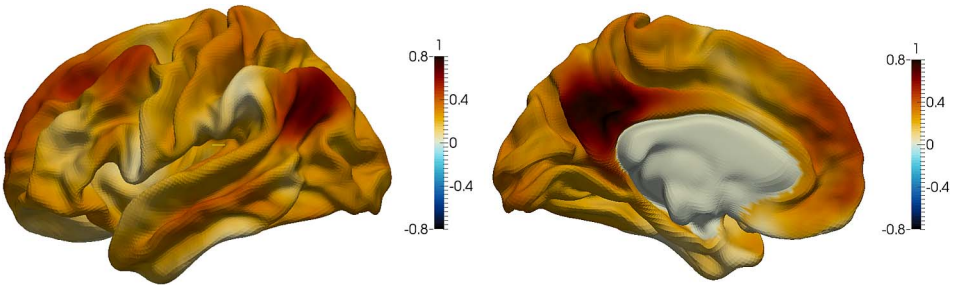

FIG. 9. *The mean RSFC map computed over* 491 *subject. As expected, high correlation values are visible inside the ROI.*

The first three PC functions, estimated with SM-FPCA, are shown in Figures 10–12 as compared to the PC functions derived from MV-PCA and IHK-PCA. The choice of the smoothing parameter for the SM-FPCA is based on the K-fold cross-validation, with $K = 5$.

The PC functions estimated from the MV-PCA show an excessive variability since the sample size is not sufficiently large to deal with the extremely high dimensionality of the data, and the spatial information is completely ignored by this model. In fact, even recent attempts to model the subject variability from resting state fMRI lead to the conclusion that spatial mismatches, introduced by the alignment problem, are one of the biggest sources of currently observable differences between subjects [Harrison et al. (2015)]. This registration process can

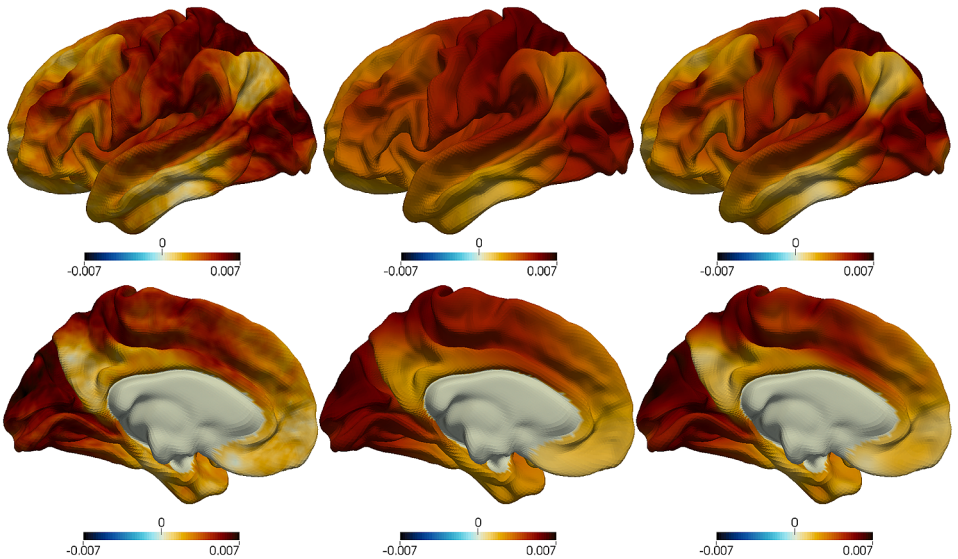

FIG. 10. *From left to right, two views of the first PC function computed respectively with MV-PCA, IHK-PCA and SM-FPCA.*

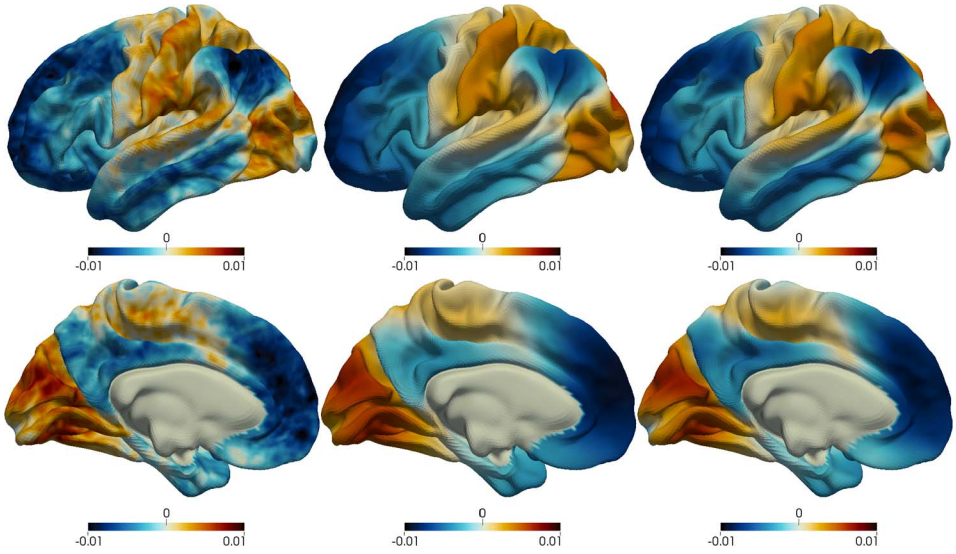

FIG. 11. *From left to right, two views of the second PC function computed respectively with MV-PCA, IHK-PCA and SM-FPCA.*

result in misalignments, due to the lack of functional regions being perfectly coincident or due to situations where the local topology is strongly different among subjects. These misalignments can introduce fictitious effects on the computed

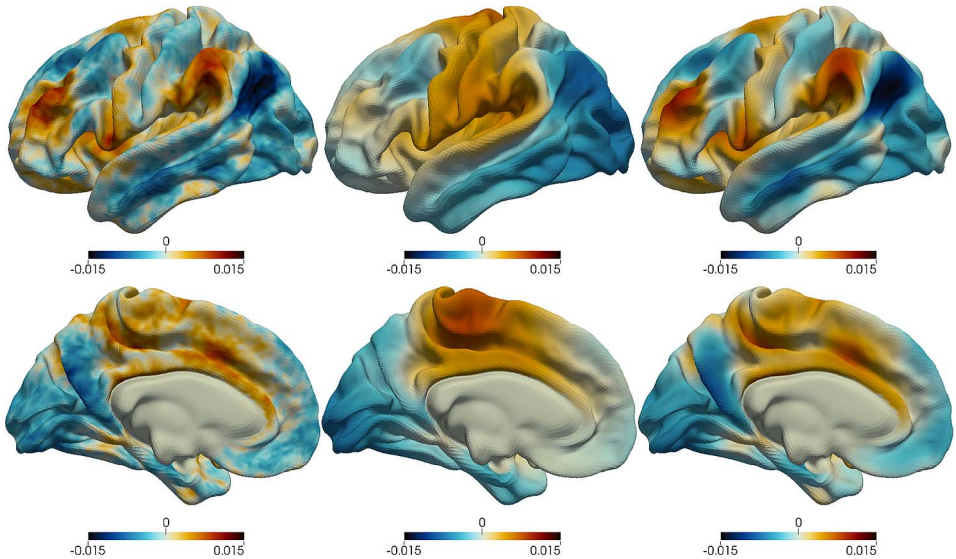

FIG. 12. *From left to right, two views of the third PC function computed respectively with MV-PCA, IHK-PCA and SM-FPCA.*

PC functions. Data misalignment is a well-known problem in FDA [Marron et al. (2015)]. For functional data with one-dimensional domains, typical approaches are based on shifting or (monotone) transformations of the domain of each function. But neither shifting nor monotonic transformations make sense on a generic non-Euclidean domain, and so it is not clear how to generalize the standard FDA approaches. The introduction of a smoothing penalty in the PCA model should reduce the variability effects due to misalignment. In fact, the smoothing parameter in the SM-FPCA algorithm can be seen as a further degree of freedom that allows a multiscale analysis, meaning that, by increasing the smoothing penalty parameter, it is possible to constrain the results to show only the macroscopical effects of the phenomena and to remove the artifacts introduced by the preprocessing steps.

Both IHK-PCA and SM-FPCA return smooth PC functions. A visual inspection of the estimated PC functions, though, highlights that IHK-PCA completely smooths out sharper changes in the modes of variations, missing some localized features that are apparent in MV-PCA and are also very well captured by the proposed SM-FPCA. Comparing, for instance, the estimated third PC functions, in the top views of Figure 12, one can see for both MV-PCA and SM-PCA corresponding localized areas with very high values (in red) and very low values (in blue) that are instead missing in the IHK-PCA estimate. By contrast, the presmoothing approach appears to introduce some artifacts: looking at the bottom views in Figure 12, one can, for instance, notice that IHK-PCA estimated the third PC function has high values in the higher part of the plot, which do not have a match on the MV-PCA or on the SM-FPCA estimate.

For the purpose of interpretation of the PC functions, we might prefer to plot the functions $\mu \pm 2\sigma f$, where μ denotes the mean RSFC map, σ denotes the standard deviation of the PC scores vector and f denotes the associated PC function. In Figure 13 we show the described plot for the first PC function. We can observe that while the high correlation value in the ROI and inferior parietal are in the first approximation preserved from subject to subject, a high variability between subjects can be observed in the areas surrounding the ROI and the inferior parietal, which is understood due to individual inter-subject differences [Buckner, Andrews-Hanna and Schacter (2008) and references therein]. However, it should be noted that variability can be both somewhat localized as well as more spatially smooth, indicating that, even in resting state data, brain regions have a differential response which is not simply a result of noise in the data.

6. Discussion. In this paper we introduced a novel PCA technique that can handle functional data located over a two-dimensional manifold. The adopted approach is based on a regularized PCA model. In particular, a smoothness penalty term that measures the curvature of a function over a manifold is considered, and the estimation problem is solved via an iterative algorithm that uses finite elements. The motivating application is the analysis the RSFC maps over the cortical

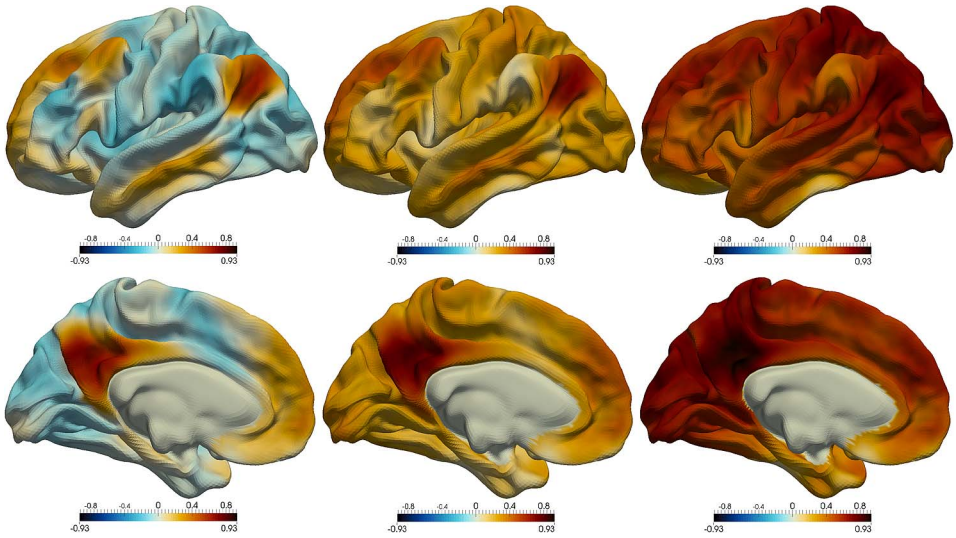

FIG. 13. *From left to right, two views of* $\mu - 2\sigma f$, μ, $\mu + 2\sigma f$, *where* μ *denotes the mean RSFC map,* σ *denotes the standard deviation of the first PC scores vector and* f *denotes the first PC function.*

surface, derived from the fMRI. In this setting the adoption of a MV-PCA suffers from the high-dimensionality of the data with respect to the relatively small sample size. The adoption of an approach based on individual presmoothing of the functional samples, followed by a MV-PCA, gives smooth estimates of the PC functions. However, this presmoothing step tends to remove useful information from the original data. The proposed SM-FPCA instead returns smooth PC functions that nevertheless are able to capture localized features of the estimated PC functions. It could also be imagined that, in more complex study designs (such as patient versus control studies), these PC functions, along with the associated scores, could be used to investigate diverse differences between groups or covariate effects.

A further important feature of SM-FPCA is its computational efficiency. The most computationally intensive operation is the resolution of the linear system in the iterative algorithm. However, this linear system enjoys two important properties. The first is the independence between its dimensions, related to the number of nodes of the triangular mesh, and the number of pointwise observations available for each functional sample as well as the sample size. In fact, since its resolution time depends mostly on the mesh size, a mesh simplification approach [Dassi et al. (2015)] could be adopted to speed up the algorithm. The second and most fundamental property is the sparsity of the linear system. The use of a sparse solver allows an efficient computation of the solution. For instance, in the final application, the dimension of the linear system is 64K $\times$ 64K. Despite its dimension, the solving time is less than a second. The application of the entire algorithm, for a

fixed smoothing parameter, with 15 iterations is less than 15 seconds on a Intel Core i5-3470 3.20 GHz workstation with 4 GB of RAM.

Acknowledgments. The authors would very much like to thank the Editor, Associate Editor and referees for their really encouraging comments throughout the process. Data Access: The data is available from the Human Connectome Project Consortium [HCP, Essen et al. (2012)].

SUPPLEMENTARY MATERIAL

Supplement to "Smooth Principal Component Analysis over two-dimensional manifolds with an application to neuroimaging" (DOI: 10.1214/16-AOAS975SUPP; .pdf). The online supplementary material contains the theoretic details of the Finite Element discretization approach. Moreover, it includes further simulations on the sphere investigating both the methodology and its robustness to alignment issues.

REFERENCES

ALFELD, P., NEAMTU, M. and SCHUMAKER, L. L. (1996). Fitting scattered data on sphere-like surfaces using spherical splines. *J. Comput. Appl. Math.* **73** 5–43. MR1424867

BELKIN, M. and NIYOGI, P. (2001). Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Advances in Neural Information Processing Systems* 14 (T. G. Dietterich, S. Becker and Z. Ghahramani, eds.) 585–591.

BUCKNER, R. L., ANDREWS-HANNA, J. R. and SCHACTER, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Ann. N. Y. Acad. Sci.* **1124** 1–38.

CAI, D., HE, X., HAN, J. and HUANG, T. S. (2011). Graph regularized nonnegative matrix factorization for data representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **33** 1548–1560.

CHUNG, M. K., HANSON, J. L. and POLLAK, S. D. (2014). Statistical analysis on brain surfaces. Technical report, University of Wisconsin–Madison.

CHUNG, M. K., ROBBINS, S. M., DALTON, K. M., DAVIDSON, R. J., ALEXANDER, A. L. and EVANS, A. C. (2005). Cortical thickness analysis in autism with heat kernel smoothing. *NeuroImage* **25** 1256–1265.

DASSI, F., ETTINGER, B., PEROTTO, S. and SANGALLI, L. M. (2015). A mesh simplification strategy for a spatial regression analysis over the cortical surface of the brain. *Appl. Numer. Math.* **90** 111–131. MR3300898

DUCHON, J. (1977). Splines minimizing rotation-invariant semi-norms in Sobolev spaces. In *Constructive Theory of Functions of Several Variables* (*Proc. Conf., Math. Res. Inst., Oberwolfach*, 1976) 85–1000. Springer, Berlin. MR0493110

DZIUK, G. (1988). Finite elements for the Beltrami operator on arbitrary surfaces. In *Partial Differential Equations and Calculus of Variations* (S. Hildebrandt and R. Leis, eds.). *Lecture Notes in Math.* **1357** 142–155. Springer, Berlin. MR0976234

ESSEN, D. C. V., UGURBIL, K., AUERBACH, E., BARCH, D., BEHRENS, T. E. J., BUCHOLZ, R., CHANG, A., CHEN, L., CORBETTA, M., CURTISS, S. W., PENNA, S. D., FEINBERG, D., GLASSER, M. F., HAREL, N., HEATH, A. C., LARSON-PRIOR, L., MARCUS, D., MICHALAREAS, G., MOELLER, S., OOSTENVELD, R., PETERSEN, S. E., PRIOR, F., SCHLAGGAR, B. L., SMITH, S. M., SNYDER, A. Z., XU, J. and YACOUB, E. (2012). The Human Connectome Project: A data acquisition perspective. *NeuroImage* **62** 2222–2231.

ETTINGER, B., PEROTTO, S. and SANGALLI, L. M. (2016). Spatial regression models over two-dimensional manifolds. *Biometrika* **103** 71–88. MR3465822

GLASSER, M. F., SOTIROPOULOS, S. N., WILSON, J. A., COALSON, T. S., FISCHL, B., ANDERSSON, J. L., XU, J., JBABDI, S., WEBSTER, M., POLIMENI, J. R., ESSEN, D. C. V. and JENKINSON, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80** 105–124.

GORDON, E. M., LAUMANN, T. O., ADEYEMO, B., HUCKINS, J. F., KELLEY, W. M. and PETERSEN, S. E. (2014). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cereb. Cortex*.

GREEN, P. J. and SILVERMAN, B. W. (1993). *Nonparametric Regression and Generalized Linear Models*. CRC Press, Boca Raton.

HALL, P. and HOSSEINI-NASAB, M. (2006). On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **68** 109–126. MR2212577

HARRISON, S. J., WOOLRICH, M. W., ROBINSON, E. C., GLASSER, M. F., BECKMANN, C. F., JENKINSON, M. and SMITH, S. M. (2015). Large-scale probabilistic functional modes from resting state fMRI. *NeuroImage* **109** 217–231.

HUANG, J. Z., SHEN, H. and BUJA, A. (2008). Functional principal components analysis via penalized rank one approximation. *Electron. J. Stat.* **2** 678–695. MR2426107

JOLLIFFE, I. T., TRENDAFILOV, N. T. and UDDIN, M. (2003). A modified principal component technique based on the LASSO. *J. Comput. Graph. Statist.* **12** 531–547. MR2002634

LILA, E., ASTON, J. A. D. and SANGALLI, L. M. Supplement to "Smooth Principal Component Analysis over two-dimensional manifolds with an application to neuroimaging." DOI:10.1214/16-AOAS975SUPP.

MARRON, J. S., RAMSAY, J. O., SANGALLI, L. M. and SRIVASTAVA, A. (2015). Functional data analysis of amplitude and phase variation. *Statist. Sci.* **30** 468–484. MR3432837

OGAWA, S., LEE, T. M., KAY, A. R. and TANK, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proc. Natl. Acad. Sci. USA* **87** 9868–9872.

RAMSAY, T. (2002). Spline smoothing over difficult regions. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 307–319. MR1904707

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer, New York. MR2168993

RICE, J. A. and SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B* **53** 233–243. MR1094283

RIESZ, F. and SZ.-NAGY, B. (1955). *Functional Analysis*. Frederick Ungar Publishing Co., New York. MR0071727

SANGALLI, L. M., RAMSAY, J. O. and RAMSAY, T. O. (2013). Spatial spline regression models. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75** 681–703. MR3091654

SHEN, H. and HUANG, J. Z. (2008). Sparse principal component analysis via regularized low rank matrix approximation. *J. Multivariate Anal.* **99** 1015–1034. MR2419336

SILVERMAN, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.* **24** 1–24. MR1389877

WAHBA, G. (1981). Spline interpolation and smoothing on the sphere. *SIAM J. Sci. Statist. Comput.* **2** 5–16. MR0618629

ZHOU, L. and PAN, H. (2014). Principal component analysis of two-dimensional functional data. *J. Comput. Graph. Statist.* **23** 779–801. MR3224656

ZOU, H. and HASTIE, T. (2005). Regularization and variable selection via the elastic net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **67** 301–320. MR2137327

ZOU, H., HASTIE, T. and TIBSHIRANI, R. (2006). Sparse principal component analysis. *J. Comput. Graph. Statist.* **15** 265–286. MR2252527

E. LILA
CAMBRIDGE CENTRE FOR ANALYSIS
UNIVERSITY OF CAMBRIDGE
WILBERFORCE ROAD
CAMBRIDGE CB3 0WB
UNITED KINGDOM
E-MAIL: e.lila@maths.cam.ac.uk

J. A. D. ASTON
STATISTICAL LABORATORY
UNIVERSITY OF CAMBRIDGE
WILBERFORCE ROAD
CAMBRIDGE CB3 0WB
UNITED KINGDOM
E-MAIL: j.aston@statslab.cam.ac.uk

L. M. SANGALLI
MOX—DIPARTIMENTO DI MATEMATICA
POLITECNICO DI MILANO
PIAZZA LEONARDO DA VINCI, 32
20133 MILAN
ITALY
E-MAIL: laura.sangalli@polimi.it