Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

Eigen-Adjusted Functional Principal Component Analysis

Ci-Ren Jiang, Eardi Lila, John A. D. Aston & Jane-Ling Wang

To cite this article: Ci-Ren Jiang, Eardi Lila, John A. D. Aston & Jane-Ling Wang (2022): Eigen-Adjusted Functional Principal Component Analysis, Journal of Computational and Graphical Statistics, DOI: 10.1080/10618600.2022.2067550

To link to this article: https://doi.org/10.1080/10618600.2022.2067550

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2022, VOL. 00, NO. 0, 1–12
https://doi.org/10.1080/10618600.2022.2067550

Taylor & Francis
Taylor & Francis Group

Eigen-Adjusted Functional Principal Component Analysis

Ci-Ren Jiang[a], Eardi Lila[b], John A. D. Aston[c], and Jane-Ling Wang[d]

[a]Institute of Statistical Science, Academia Sinica, Taipei, Taiwan; [b]Department of Biostatistics, University of Washington, Seattle, WA; [c]Statistical Laboratory, University of Cambridge, Cambridge, UK; [d]Department of Statistics, University of California, Davis, Davis, CA

ABSTRACT

Functional Principal Component Analysis (FPCA) has become a widely used dimension reduction tool for functional data analysis. When additional covariates are available, existing FPCA models integrate them either in the mean function or in both the mean function and the covariance function. However, methods of the first kind are not suitable for data that display second-order variation, while those of the second kind are time-consuming and make it difficult to perform subsequent statistical analyses on the dimension-reduced representations. To tackle these issues, we introduce an eigen-adjusted FPCA model that integrates covariates in the covariance function only through its eigenvalues. In particular, different structures on the covariate-specific eigenvalues—corresponding to different practical problems—are discussed to illustrate the model's flexibility as well as utility. To handle functional observations under different sampling schemes, we employ local linear smoothers to estimate the mean function and the pooled covariance function, and a weighted least square approach to estimate the covariate-specific eigenvalues. The convergence rates of the proposed estimators are further investigated under the different sampling schemes. In addition to simulation studies, the proposed model is applied to functional Magnetic Resonance Imaging scans, collected within the Human Connectome Project, for functional connectivity investigation. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received March 2021
Accepted April 2022

KEYWORDS
Covariate-specific eigenvalue; Functional magnetic resonance imaging; Local linear smoother; Weighted least squares

1. Introduction

Principal component analysis is a classical dimension reduction tool in multivariate statistical analysis and its extension to functional data, termed Functional Principal Component Analysis (FPCA), plays a central role in the analysis of samples that are curves, functions, or surfaces as shown in the survey article by Shang (2014). However, most of the existing FPCA approaches (see, e.g., Rao 1958; Dauxois, Pousse, and Romain 1982; Rice and Silverman 1991; Cardot 2000; James, Hastie, and Suger 2000; Rice and Wu 2001; Yao, Müller, and Wang 2005; Hall, Müller, and Wang 2006; Li and Hsing 2010; Chen and Jiang 2017) assume that the observed samples arise from the same population and do not accommodate information from a set of additional covariates. In this article, motivated by the analysis of spatio-temporal brain imaging data, we introduce a novel FPCA model where additional covariates, modeling for instance spatial locations, affect both the mean and covariance of the functional samples describing the temporal component of the data.

So far, relatively little of the literature has covered FPCA methodology that adapts to the covariate information. Some of the proposed methods merely integrate the covariate information in the systematic part, that is, the mean function (see, e.g., Chiou, Müller, and Wang 2003; Jiang and Wang 2011; Zhang, Park, and Wang 2013; Zhang and Wang 2015). In order to ease the computational burden and alleviate the curse of dimensionality, in Chiou, Müller, and Wang (2003), Zhang, Park, and Wang (2013), and Zhang and Wang (2015) additive structures for the mean function are considered, while in Jiang and Wang (2011) a single index model for the mean is adopted. Here, we refer to these approaches as mean-adjusted FPCA. Alternatively, both the systematic and stochastic parts, that is, the mean function and the covariance structure, are assumed to vary with the covariates (see, e.g., Cardot 2006; Jiang and Wang 2010; Li, Staicu, and Bondell 2015). Specifically, Cardot (2006) and Jiang and Wang (2010) assume that the covariance function varies with the covariates via both its eigenfunctions and eigenvalues. Jiang, Aston, and Wang (2009) extend the covariate-adjusted FPCA model in Jiang and Wang (2010) to a multiplicative model for Positron Emission Tomography image analysis. Moreover, Li, Staicu, and Bondell (2015) propose a model that accounts for the covariate effect in skewed functional data. We refer to these methods as fully adjusted FPCA.

In some applications, such as functional brain imaging, the covariance structure of functional data describing temporal brain activity is a key element to study brain connectivity. However, mean-adjusted FPCA models are not designed to integrate covariate information in the covariance function, although they are attractive because of their computational efficiency. On the other hand, fully adjusted FPCA approaches are time-consuming if the dimension of the covariates is not low and they make subsequent statistical analyses on the dimension-reduced output difficult due to the fact that they return a set of

CONTACT John A. D. Aston j.aston@statslab.cam.ac.uk Statistical Laboratory, University of Cambridge, Cambridge, UK.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JCGS.

covariates-specific eigenvalues and eigenfunctions, which are not easily comparable across covariate values.

Therefore, in this work, we will extend the PCA model in Flury (1984, 1986) to functional data and term it eigen-adjusted FPCA. Specifically, the eigen-adjusted FPCA model assumes the covariance function varies with the covariates via its eigenvalues while the corresponding eigenfunctions remain independent of the covariates. The advantages of the proposed model are two-fold. In practical applications, FPCA is generally the first step of a subsequent statistical analysis. For instance, in the proposed application to brain imaging data in Section 5, the ultimate goal is to find a parcellation that clusters brain locations (modeled as covariates) with distinct functional connectivity patterns (i.e., distinct covariance structures). For this purpose, the covariate-independent eigenfunctions of the eigen-adjusted FPCA model offer a common reference frame that enables comparison and statistical analysis of the associated covariate-specific eigenvalues. Moreover, eigen-adjusted FPCA while being more flexible than mean-adjusted FPCA maintains comparable computational times. This is particularly important for big data applications, such as the aforementioned brain imaging studies. By means of simulations, we demonstrate that eigen-adjusted FPCA can also be useful under a fully adjusted generative model, where applying fully adjusted FPCA might not be computationally feasible. Other related work can be found in Boente, Pires, and Rodrigues (2002), Benko and Härdle (2005), Benko, Härdle, and Kneip (2009), Boente, Rodriguez, and Sued (2010), and Coffey et al. (2011).

Both functional data and longitudinal data can be modeled as observations from stochastic processes, but they are different in their sampling schemes. Specifically, functional data are densely and regularly recorded while longitudinal data are sparsely and irregularly observed. FPCA has been classically concerned with the analysis of densely observed functional samples. More recently, the methodology has been extended to handle sparsely observed longitudinal data, where only a few repeated measurements are available for each sample, invalidating approaches based on nonparametric reconstructions of individual functions (see, e.g., Yao, Müller, and Wang 2005; Hall, Müller, and Wang 2006). In this work, we develop a unified eigen-adjusted FPCA framework for both types of data.

The rest of this article is organized as follows. In Section 2 we introduce the eigen-adjusted FPCA model. The proposed estimators and their asymptotical properties are provided in Section 3. Section 4 consists of simulation studies that demonstrate the finite sample performance of the proposed method. The eigen-adjusted FPCA model is applied to analyze brain functional connectivity in Section 5. Conclusions and discussions are provided in Section 6. Moreover, the assumptions are given in Appendix A, and an alternative estimation approach for the covariate-specific eigenvalues is provided in Appendix B. The proofs are provided in the supplementary materials.

2. Covariate-Adjusted FPCA

Traditionally, FPCA handles random trajectories that are functions of time (or space). Ignoring any covariate information for the moment, let $X(t)$ be a stochastic process in L^2 with mean function $\mu(t)$ and covariance function $\Gamma(s,t)$ where $s,t \in \mathrm{T}$ and T is a finite compact interval in $\mathbb{R}$. FPCA is equivalent to a spectral decomposition of Γ and leads to the well-known Karhunen–Loève decomposition of X,

$$X(t) = \mu(t) + \sum_{j=1}^{\infty} A_j \varphi_j(t),$$

where φ_j is the eigenfunction of Γ associated to the jth largest eigenvalue, and $A_j = \langle X - \mu, \varphi_j \rangle$ is the jth functional principal component (PC) score. Here $\langle \cdot, \cdot \rangle$ stands for the inner product in the L^2 space, that is, $\langle a, b \rangle = \int_{\mathrm{T}} a(t)b(t)dt$ for $a, b \in L^2(\mathrm{T})$.

When covariate information is available, X is viewed as a random function of both time t and covariate $\boldsymbol{z}$. Specifically, $X(t,\boldsymbol{z})$ is a random function with mean $\mu(t,\boldsymbol{z})$ and covariance function $\Upsilon(s,t,\boldsymbol{z})$, where $s,t \in \mathrm{T}$ and $\boldsymbol{z}$ is in a p-dimensional compact hypercube $\mathrm{Z} \subset \mathbb{R}^p$. To accommodate $\boldsymbol{z}$ into the framework of FPCA, Jiang and Wang (2010) proposed two models: fully adjusted FPCA (fFPCA) and mean-adjusted FPCA (mFPCA). The difference between them is how the covariance structure is handled. In fFPCA, it is assumed that there exists an orthogonal expansion of $\Upsilon(s,t,\boldsymbol{z})$ in terms of eigenfunctions $\psi_j(t,\boldsymbol{z})$ and nonincreasing eigenvalues $\eta_j(\boldsymbol{z})$, that is, $\Upsilon(s,t,\boldsymbol{z}) = \sum_{j=1}^{\infty} \eta_j(\boldsymbol{z})\psi_j(s,\boldsymbol{z})\psi_j(t,\boldsymbol{z})$, and thus one can represent X as

$$X(t,\boldsymbol{z}) = \mu(t,\boldsymbol{z}) + \sum_{j=1}^{\infty} A_j(\boldsymbol{z})\psi_j(t,\boldsymbol{z}),$$

where $A_j = \langle X - \mu, \psi_j \rangle$ is the jth functional PC score with mean zero and variance $\eta_j(\boldsymbol{z})$. In mFPCA, $\boldsymbol{z}$ is treated as a realization of the random variable $\boldsymbol{Z}$. Ignoring $\boldsymbol{Z}$ after centering leads to a pooled covariance

$$\Gamma^*(s,t) = \mathrm{E}\{\Upsilon(s,t,\boldsymbol{Z})\} = \int_{\mathrm{Z}} \Upsilon(s,t,\boldsymbol{z}) dG_{\boldsymbol{z}}(\boldsymbol{z}),$$

where $G_{\boldsymbol{z}}(\cdot)$ is the distribution function of $\boldsymbol{Z}$. Assume Γ^* is a smooth function and there exists an orthogonal expansion in terms of eigenfunctions $\psi_j^*(t)$ and nonincreasing eigenvalues η_j^*; that is, $\Gamma^*(s,t) = \sum_{j=1}^{\infty} \eta_j^* \psi_j^*(s)\psi_j^*(t)$, and the random function X can be represented as

$$X(t,\boldsymbol{z}) = \mu(t,\boldsymbol{z}) + \sum_{j=1}^{\infty} A_j^* \psi_j^*(t),$$

where $A_j^* = \langle X - \mu, \psi_j^* \rangle$ is the jth functional PC score with mean zero and variance η_j^*. In mFPCA, Γ^* can be estimated with a lower-dimensional smoother and thus the estimator is of a faster convergence rate. On the contrary, fFPCA is a more flexible model as the covariate information is used in estimating the covariance function as well as the eigenvalues and the eigenfunctions.

However, FPCA rarely represents the final stage of a statistical analysis as, no matter how complex, an FPCA model will unlikely represent a faithful description of the underlying generative model. FPCA is instead commonly used as a dimension reduction tool, where the subsequent analysis is performed on the PC scores. So, despite the fFPCA model being more flexible, it introduces complications when using it as a dimension reduction tool, as the scores $A_j(\boldsymbol{z})$ do not give a complete description

of the stochastic process at the covariate value $\boldsymbol{z}$; in fact, the scores are expressed with respect to a reference frame $\psi_j(t, \boldsymbol{z})$ which is also dependent on $\boldsymbol{z}$, invalidating comparison across covariate values.

Therefore, in the following section, we propose a model that is flexible enough to handle dependencies of the covariance functions on covariate information, yet representing an effective and useful dimension reduction tool.

2.1. Eigen-Adjusted FPCA

Assume that the covariance function $\Gamma(s, t, \boldsymbol{z})$ has an orthogonal expansion in terms of eigenfunctions $\phi_k(t)$ and nonnegative eigenvalues $\lambda_k(\boldsymbol{z})$. Specifically,

$$\Gamma(s, t, \boldsymbol{z}) = \sum_{k=1}^{\infty} \lambda_k(\boldsymbol{z})\phi_k(s)\phi_k(t), \tag{2.1}$$

where $\lambda_1(\boldsymbol{z}) > \lambda_2(\boldsymbol{z}) > \cdots \geq 0$ and $\sum_{k=1}^{\infty} \lambda_k(\boldsymbol{z}) < \infty$ for $\boldsymbol{z} \in \mathrm{Z}$ and $s, t \in \mathrm{T}$. By the Karhunen–Loéve expansion, $X(t, \boldsymbol{z})$ can be represented as

$$X(t, \boldsymbol{z}) = \mu(t, \boldsymbol{z}) + \sum_{k=1}^{\infty} A_k(\boldsymbol{z})\phi_k(t), \tag{2.2}$$

where $A_k(\boldsymbol{z}) = \langle X - \mu, \phi_k \rangle$ is the jth functional PC score with mean zero and variance $\lambda_k(\boldsymbol{z})$. Compared to mFPCA, Model (2.2) is more general in that Γ in (2.1) can vary with $\boldsymbol{z}$, and compared to fFPCA, it is computationally more efficient in estimating the covariance function. The second argument will be further demonstrated in Section 3.

The model of Γ in (2.1) allows for different structural assumptions on $\lambda_k(\boldsymbol{z})$, tailored to specific problems. Roughly speaking, there are three scenarios. The first scenario is that $\lambda_k(\boldsymbol{z})$ are grouped and may not vary smoothly with $\boldsymbol{z}$. This setting was considered in most related FPCA literature (e.g., Benko, Härdle, and Kneip 2009; Boente, Rodriguez, and Sued 2010), where the goal is to test the equality of covariance functions of different known groups. The second scenario is that $\lambda_k(\boldsymbol{z})$ are continuous and vary smoothly with $\boldsymbol{z}$, such as when $\boldsymbol{z}$ models spatial coordinates and we assume spatially smooth variation. The third scenario is that $\lambda_k(\boldsymbol{z})$ are piecewise smooth functions of $\boldsymbol{z}$, that is, $\lambda_k(\boldsymbol{z})$ are smooth functions within each group. The latter is a more realistic model for the analysis of the brain imaging data in our final application. In this article, we focus on estimating $\lambda_k(\boldsymbol{z})$ under the second scenario as the third scenario can be handled with by simply exploiting the group information. When the group structure is not known, our numerical studies suggest that estimating the eigenvalues as in the second scenario and applying a clustering approach to the estimated $\lambda_k(\boldsymbol{z})$ can help retrieve the group information.

3. Estimation

Standard procedures to perform FPCA include (i) estimating the mean function, (ii) estimating the covariance function, (iii) estimating the eigenfunctions and eigenvalues, and (iv) predicting the functional PC scores. Under some regularity conditions on the mean and covariance functions, local linear smoothers (Fan and Gijbels 1996) can be applied to estimate them. Eigenfunctions and eigenvalues can be estimated via the application of an eigen-decomposition to the estimated covariance. Standard numerical approaches and PACE (Yao, Müller, and Wang 2005) can be applied to predict the functional PC scores when data are dense and sparse, respectively. In step (ii), directly estimating $\Gamma(s, t, \boldsymbol{z})$ is computational demanding especially when both T and Z are multidimensional. Exploiting the special structure of $\Gamma(s, t, \boldsymbol{z})$ in (2.1), to ease computational burden, we circumvent step (ii). Specifically, we apply an eigen-decomposition to the pooled covariance $\Gamma^* = \mathrm{E}(\Gamma)$ to obtain the eigenfunctions, and propose a Weighted Least Squares (WLS) approach, for both dense and sparse data, to estimate the eigenvalues that vary with $\boldsymbol{z}$. Below we provide the details of the proposed estimators as well as their asymptotic properties.

3.1. Mean Function

Let $\boldsymbol{z}_i \in \mathrm{Z}$ be the covariate of the ith subject, whose jth observation made at time $t_{ij} \in \mathrm{T}$ is

$$Y_{ij} = \mu(t_{ij}, \boldsymbol{z}_i) + \sum_{k=1}^{\infty} A_{ik}\phi_k(t_{ij}) + \varepsilon_{ij}, \tag{3.1}$$

where ε_{ij} is the independent measurement error with mean zero and variance σ^2, for $j = 1, \ldots, N_i$ and $i = 1, \ldots, n$. Theoretically, any $(p+1)$-dimensional smoother can be applied to estimate μ. Here, we use a $(p+1)$-dimensional local linear smoother and denote the estimator as $\hat{\mu}$. Specifically,

$$\hat{\mu}(t, \boldsymbol{z}) = \hat{b}_0, \text{ where}$$

$$(\hat{b}_0, \hat{b}_1, \hat{\boldsymbol{b}}_2)^T = \arg\min_{\boldsymbol{b}} \frac{1}{n}\sum_{i=1}^{n}\frac{1}{N_i}\sum_{j=1}^{N_i} \Big\{Y_{ij} - b_0 - b_1(t_{ij} - t) - \boldsymbol{b}_2^T(\boldsymbol{z}_i - \boldsymbol{z})\Big\}^2 \times K_{h_t}(t_{ij} - t)\Big(\prod_{k=1}^{p} K_{h_z^{(k)}}(z_i^{(k)} - z^{(k)})\Big),$$

$\boldsymbol{b} = (b_0, b_1, \boldsymbol{b}_2^T)^T$, $K_h(\cdot) = K(\cdot/h)/h$, K is a kernel function defined in Assumption A.2 in Appendix A, and h_t and $\boldsymbol{h}_z = (h_z^{(1)}, \ldots, h_z^{(p)})^T$ are the bandwidths for T and Z, respectively. For simplicity, we assume that $h_z^{(k)}$'s are all of the same order as h_z. Let $\gamma_{nk} = \left(n^{-1}\sum_{i=1}^{n} N_i^{-k}\right)^{-1}$ for $k = 1, 2$, and $\delta_n = \left[\{1 + 1/(\gamma_{n1}h_t)\}\log n/(nh_z^p)\right]^{1/2}$; denote $h_1 \approx h_2$ if h_1 is of the same order as h_2, and $h_1 \lesssim h_2$ (resp. $h_1 \gtrsim h_2$) if h_1 is of smaller (resp. larger) order than h_2. Below we provide the asymptotical properties of $\hat{\mu}$.

Theorem 3.1. Assume that Assumptions A.1–A.2 and B.1–B.2 hold. Then,

$$\sup_{t\in\mathrm{T}, \boldsymbol{z}\in\mathrm{Z}} |\hat{\mu}(t, \boldsymbol{z}) - \mu(t, \boldsymbol{z})| = O(h_t^2 + h_z^2 + \delta_n) \text{ a.s.} \tag{3.2}$$

In (3.2), $h_t^2 + h_z^2$ and δ_n^2 are the order of bias and that of variance, respectively, due to smoothing. We further elaborate on the convergence rates of $\hat{\mu}$ under two different sampling schemes in the following corollary.

Corollary 3.1. Assume that Assumptions A.1–A.2 and B.1–B.2 hold.

(a) If $\max_{1\le i\le n} N_i \le \mathcal{M}$ for some $\mathcal{M} < \infty$ and $h_t \approx h_z \approx h$,

$$\sup_{t\in\mathrm{T},\boldsymbol{z}\in\mathrm{Z}} |\hat{\mu}(t,\boldsymbol{z}) - \mu(t,\boldsymbol{z})| = O\big(h^2 + \{\log n/(nh^{p+1})\}^{1/2}\big) \text{ a.s.}$$

The optimal convergence rate is $\big(\log n/n\big)^{2/(p+5)}$.

(b) If $\max_{1\le i\le n} N_i \ge \mathcal{M}_n$, where $\mathcal{M}_n^{-1} \approx h_t \lesssim h_z \approx (\log n/n)^{1/(p+4)}$ is bounded away from zero,

$$\sup_{t\in\mathrm{T},\boldsymbol{z}\in\mathrm{Z}} |\hat{\mu}(t,\boldsymbol{z}) - \mu(t,\boldsymbol{z})| = O\big((\log n/n)^{2/(p+4)}\big) \text{ a.s.} \quad (3.3)$$

Remark 3.1. (a) represents the case of longitudinal data where N_i is finite and so it is reasonable to have h_t and h_z of the same order. (b) represents functional data where the observations are intensely recorded. Since $N_i \to \infty$, h_t is of smaller order than h_z.

3.2. Pooled Covariance and Its Eigenfunctions

In practice, the eigenfunctions are estimated by performing an eigen-decomposition on the discretized covariance function. However, directly estimating Γ may be too computationally demanding especially when both T and Z are multi-dimensional. Thus, we propose to estimate the eigenfunctions ϕ_k by applying an eigen-decomposition to the pooled covariance function,

$$\Gamma^*(s,t) = \mathrm{E}\{\Gamma(s,t,\boldsymbol{Z})\} = \sum_{k=1}^{\infty} \mathrm{E}\{\lambda_k(\boldsymbol{Z})\}\phi_k(s)\phi_k(t), \quad (3.4)$$

where $\Gamma(s,t,\boldsymbol{Z})$, $\lambda_k(\boldsymbol{Z})$, and $\phi_k(t)$ are defined in (2.1). Any two-dimensional smoother can be applied to estimate Γ^* and we pick a local linear smoother here. Specifically,

$$\hat{\Gamma}^*(s,t) = \hat{b}_0, \text{ where}$$

$$(\hat{b}_0,\hat{b}_1,\hat{b}_2)^T = \arg\min_{\boldsymbol{b}} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{N_i(N_i-1)} \sum_{j\ne k}^{N_i} \Big\{\hat{U}_{ij}\hat{U}_{ik} - b_0 - b_1(t_{ij}-s) - b_2(t_{ik}-t)\Big\}^2 \times K_{h_\Gamma}(t_{ij}-s)K_{h_\Gamma}(t_{ik}-t),$$

$\boldsymbol{b} = (b_0,b_1,b_2)^T$, $\hat{U}_{ij} = Y_{ij} - \hat{\mu}(t_{ij},\boldsymbol{z}_i)$, K is a kernel function defined as that for $\hat{\mu}$, and h_Γ is the smoothing bandwidth. The diagonal terms $\{\hat{U}_{ij}\hat{U}_{ij} \mid 1 \le j \le N_i, 1 \le i \le n\}$ are removed while estimating Γ^* since $\mathrm{cov}(Y_{ij},Y_{ik}) = \Gamma(t_{ij},t_{ik},\boldsymbol{z}_i) + \sigma^2\delta_{(t_{ij}=t_{ik})}$, where $\delta_{(t_{ij}=t_{ik})} = 1$ if $t_{ij} = t_{ik}$ and 0 otherwise. For simplicity, we let

$$\delta_{n1} = \left\{\left(1 + \frac{1}{\gamma_{n1}h_\Gamma}\right)\frac{\log n}{n}\right\}^{1/2}, \text{ and}$$

$$\delta_{n2} = \left\{\left(1 + \frac{1}{\gamma_{n1}h_\Gamma} + \frac{1}{\gamma_{n2}h_\Gamma^2}\right)\frac{\log n}{n}\right\}^{1/2}.$$

Theorem 3.2. Assume that Assumptions A.1–A.2 and B.1–B.4 hold. We can obtain

$$\sup_{s,t\in\mathrm{T}} |\hat{\Gamma}^*(s,t) - \Gamma^*(s,t)| = O\big(\delta_{n2} + h_\Gamma^2 + \delta_{n1}(h_t^2 + h_z^2 + \delta_n) + h_t^4 + h_z^4 + \delta_n^2\big) \text{ a.s.,}$$
$$= O(\delta_{n2} + h_\Gamma^2 + h_t^4 + h_z^4 + \delta_n^2) \text{ a.s.}$$

Remark 3.2. Note that $\delta_{n1}(h_t^2 + h_z^2 + \delta_n)$ should be of smaller order than δ_{n2} as h_t, h_z and δ_n go to zero as $n \to \infty$. When $p \le 3$ (in most real examples and in our data), it is reasonable to believe that $\delta_n^2 \lesssim \delta_{n2}$ and we elaborate the convergence rates in the following corollary under such a condition. Even so, some brief discussions on the convergence rates when $p \ge 4$ will be given in Section 3.4.

Corollary 3.2. Assume that Assumptions A.1–A.2 and B.1–B.4 hold.

(a) If $\max_{1\le i\le n} N_i \le \mathcal{M}$ for some $\mathcal{M} < \infty$ and $\{(\log n/n)h_\Gamma^2\}^{1/(2p+2)} \lesssim h_t, h_z \lesssim h_\Gamma^{1/2}$,

$$\sup_{s,t\in\mathrm{T}} |\hat{\Gamma}^*(s,t) - \Gamma^*(s,t)| = O\left(h_\Gamma^2 + \{\log n/(nh_\Gamma^2)\}^{1/2}\right) \text{ a.s.} \quad (3.5)$$

The restriction of bandwidths holds automatically when $p = 1$, and it leads to $h_\Gamma \gtrsim (\log n/n)^{1/(p-1)}$ when $p \ge 2$. The optimal convergence rate is $(\log n/n)^{1/3}$.

(b) If $\max_{1\le i\le n} N_i \ge \mathcal{M}_n$, where $\mathcal{M}_n^{-1} \lesssim h_\Gamma \approx h_t \lesssim (\log n/n)^{1/4}$ is bounded away from zero,

$$\sup_{s,t\in\mathrm{T}} |\hat{\Gamma}^*(s,t) - \Gamma^*(s,t)| = O\left(h_z^4 + \delta_n^2 + (\log n/n)^{1/2}\right) \text{ a.s.} \quad (3.6)$$

The optimal convergence rate is $(\log n/n)^{1/2}$ given $h_z \lesssim (\log n/n)^{1/8}$.

Once Γ^* is estimated, $\hat{\phi}_k$ can be obtained through

$$\int_{\mathrm{T}} \hat{\Gamma}^*(s,t)\hat{\phi}_k(s)ds = \hat{\lambda}_k^*\hat{\phi}_k(t).$$

Below we provide the asymptotic properties of $\hat{\phi}_k$ and those of $\hat{\lambda}_k^*$ for $1 \le k \le L$.

Theorem 3.3. Assume that Assumptions A.1–A.2 and B.1–B.4 hold, for $1 \le j \le L$:

$$\|\hat{\phi}_j - \phi_j\| = O(h_\Gamma^2 + \delta_{n1} + h_t^4 + h_z^4 + \delta_n^2) \text{ a.s.}$$
$$|\hat{\lambda}_j^* - \lambda_j^*| = O\big((\log n/n)^{1/2} + h_\Gamma^2 + \delta_{n1}(h_t^2 + h_z^2 + \delta_n) + h_t^4 + h_z^4 + \delta_n^2\big) \text{ a.s.}$$

Again, we elaborate on the convergence rates of $\hat{\phi}_j$ in regards to different sampling schemes.

Corollary 3.3. Assume that Assumptions A.1–A.4 and B.1–B.3 hold.

(a) If $\max_{1\le i\le n} N_i \le \mathcal{M}$ for some fixed $\mathcal{M}$ and $\{(\log n/n)h_\Gamma\}^{1/(2p+2)} \lesssim h_t, h_z \lesssim h_\Gamma^{1/2}$,

$$\|\hat{\phi}_j - \phi_j\| = O\left(h_\Gamma^2 + \{\log n/(nh_\Gamma)\}^{1/2}\right) \text{ a.s.} \quad (3.7)$$

The optimal convergence rate is $(\log n/n)^{2/5}$.

(b) If $\max_{1\le i\le n} N_i \ge \mathcal{M}_n$, where $\mathcal{M}_n^{-1} \lesssim h_\Gamma \approx h_t \lesssim (\log n/n)^{1/4}$ is bounded away from zero,

$$\|\hat{\phi}_j - \phi_j\| = O\left((\log n/n)^{1/2} + h_z^4 + \delta_n^2\right) \text{ a.s.} \quad (3.8)$$

The optimal convergence rate is $(\log n/n)^{1/2}$ given $h_z \lesssim (\log n/n)^{1/8}$.

3.3. Covariate-Specific Eigenvalues

Given that $E[A_{ik}^2|z_i] = \lambda_k(\boldsymbol{z}_i)$, $\lambda_k(\boldsymbol{z})$ can be estimated via applying a p-dimensional smoother to $\{(A_{ik}^2, \boldsymbol{z}_i) \mid 1 \le i \le n\}$. However, this is only feasible for dense data. When the data are sparse, such as longitudinal observations, the PC scores cannot be predicted accurately via numerical approaches for integration, and instead PACE (Yao, Müller, and Wang 2005) is usually considered. However, applying a p-dimensional smoother to the squared PC scores predicted by PACE may not be appropriate as $\text{var}\{\text{E}(A_{ik}|\boldsymbol{y}_i)\} \le \text{var}[A_{ik}|z_i] = \lambda_k(\boldsymbol{z}_i)$, where $\text{E}(A_{ik}|\boldsymbol{y}_i)$ is the PACE predictor of A_{ik}. To solve this issue, we propose a WLS procedure to estimate the eigenvalues $\lambda_k(\boldsymbol{z})$ for both sparse and dense data. An alternative, exclusively for dense data, is provided in Appendix B and is named as the PC-based approach.

Let $\mathcal{X}_i = \{\phi_\ell(t_{i,j})\phi_\ell(t_{i,k})\}$, where $j < k$, be a $[N_i(N_i-1)/2] \times L_n$ matrix, $\mathcal{Y}_i = (C_{i,12}, \ldots, C_{i,(N_i-1)(N_i)})^T$, $C_{i,jk} = U_{ij}U_{ik}$, $U_{ij} = Y_{ij} - \mu(t_{ij}, \boldsymbol{z}_i)$, and $\boldsymbol{\lambda}_z = (\lambda_1(\boldsymbol{z}), \ldots, \lambda_{L_n}(\boldsymbol{z}))^T$. Thus, $\mathcal{Y}_i$ can be represented as

$$\mathcal{Y}_i = \mathcal{X}_i\boldsymbol{\lambda}_{z_i} + \boldsymbol{\epsilon}_i,$$

where $\boldsymbol{\epsilon}_i$ is the remainder term. Given that λ_k's are smooth functions of $\boldsymbol{z}$, it is thus reasonable to estimate $\boldsymbol{\lambda}_z$ through

$$\hat{\boldsymbol{\lambda}}_z^W = (\hat{\mathcal{X}}^T\mathcal{W}_z\hat{\mathcal{X}})^{-1}(\hat{\mathcal{X}}^T\mathcal{W}_z\hat{\mathcal{Y}}), \tag{3.9}$$

where $\hat{\mathcal{Y}} = (\hat{\mathcal{Y}}_1^T, \ldots, \hat{\mathcal{Y}}_n^T)^T$, $\hat{\mathcal{X}} = (\hat{\mathcal{X}}_1^T, \ldots, \hat{\mathcal{X}}_n^T)^T$, $\hat{\mathcal{X}}_i = \{\hat{\phi}_\ell(t_{i,j})\hat{\phi}_\ell(t_{i,k})\}_{[N_i(N_i-1)/2]\times L_n}$, $\hat{\mathcal{Y}}_i = (\hat{C}_{i,12}, \ldots, \hat{C}_{i,(N_i-1),N_i})^T$, $\hat{C}_{i,jk} = \hat{U}_{i,j}\hat{U}_{i,k}$, $\mathcal{W}_z = \text{diag}(\boldsymbol{w}_1^T, \ldots, \boldsymbol{w}_n^T)$, and $\boldsymbol{w}_i$ is a vector with $N_i(N_i-1)/2$ identical elements, $\prod_{k=1}^p K_{h_\lambda^{(k)}}(z_i^{(k)} - z^{(k)})$.

Remark 3.3. Given that the order of $\hat{\lambda}_k(\boldsymbol{z})$ is determined by that of $\hat{\lambda}_k^*$, the estimated eigenvalues of the pooled covariance, the assumption $\lambda_1(\boldsymbol{z}) > \lambda_2(\boldsymbol{z}) > \cdots \ge 0$ for $\boldsymbol{z} \in \text{Z}$ in Model (2.1) implies that the correct order is expected. However, what really matters is that these estimated covariate-specific eigenvalues correspond to the same order of estimated eigenfunctions (a common reference frame) allowing the comparison and statistical analysis of these covariate-specific eigenvalues. The assumption $\lambda_1(\boldsymbol{z}) > \lambda_2(\boldsymbol{z}) > \cdots \ge 0$ for $\boldsymbol{z} \in \text{Z}$ could be slightly relaxed; specifically, the order of $\lambda_k(\boldsymbol{z})$ could vary with $\boldsymbol{z}$ as long as there are no tie for λ_k^* to ensure the identifiability of ϕ_k.

Here L_n could be a slowly divergent sequence if the following condition on $\lambda_k(\boldsymbol{z})$ is satisfied:

$$\begin{aligned}&\lambda_j(\boldsymbol{z}) > \lambda_{j+1}(\boldsymbol{z}) > 0, E[A_j^4|z] \le c\lambda_j(\boldsymbol{z}), \text{and}\\ &\lambda_j(\boldsymbol{z}) - \lambda_{j+1}(\boldsymbol{z}) > c^{-1}j^{-(a_1+1)},\end{aligned} \tag{3.10}$$

for $\boldsymbol{z} \in \text{Z}$, $a_1 > 1$, and some $c > 0$. A similar assumption can be found in the literature of functional regression (e.g., Hall and Horowitz 2007; Yao, Lei, and Wu 2015).

Theorem 3.4. Under assumptions A.1–A.2, and B.1–B.7, we have

$$\begin{aligned}\sup_{\boldsymbol{z}\in\text{Z}}|\hat{\boldsymbol{\lambda}}_z^W - \boldsymbol{\lambda}_z| = O\big(&L_n^{-a_1} + L_nh_\lambda^2 + L_n^2(h_\Gamma^2 + h_t^4 + h_z^4 + \delta_{n2} + \delta_n^2)\\ &+ h_t^2 + h_z^2 + \delta_n + 1/(nh_\lambda^p)^{1/2}\big) \text{ a.s.}\end{aligned} \tag{3.11}$$

Generally, $(h_t + h_\Gamma) \lesssim (h_z + h_\lambda)$ and we can obtain the following corollary for sparse data.

Corollary 3.4. Under assumptions A.1–A.2, and B.1–B.7,

$$\sup_{\boldsymbol{z}\in\text{Z}}|\hat{\boldsymbol{\lambda}}_z^W - \boldsymbol{\lambda}_z| = O\big(L_n^{-a_1} + L_nh_\lambda^2 + L_n^2(h_\Gamma^2 + \delta_{n2})\big) \text{ a.s.} \tag{3.12}$$

If further $h_\Gamma^2 \approx \delta_{n2}$, that is, $h_\Gamma \approx (\log n/n)^{1/6}$, $L_nh_\lambda^2 \approx L_n^2h_\Gamma^2$, and $L_n^2h_\Gamma^2 \approx L_n^{-a_1}$, the optimal convergence rate is $(\log n/n)^{a_1/3(a_1+2)}$.

3.4. When $p \ge 4$

Corollaries 3.2–3.4 provide the optimal convergence rates for $p \le 3$. The optimal convergence rates for a general p might be of interest to other examples and below we provide some discussions.

- *Extension of* Corollary 3.2
  The optimal convergence rate of (a) can only be achieved when $p \le 7$. When $p \ge 8$ and $h_\Gamma^{1/2} \lesssim h_t, h_z \lesssim \{(\log n/n)h_\Gamma^2\}^{1/(2p+2)}$, the optimal convergence rate is $(\log n/n)^{4/(p+5)}$. Under the assumptions of case (b), the optimal convergence rate is $(\log n/n)^{4/(4+p)}$ and $h_z \approx (\log n/n)^{1/(4+p)}$ when $p \ge 4$.
- *Extension of* Corollary 3.3
  Under the assumptions of case (a), the optimal convergence rate, $(\log n/n)^{2/5}$, is achieved when $p \le 5$. When $p > 5$, the optimal convergence rate is $(\log n/n)^{4/(p+5)}$ if the bandwidths satisfy $h_\Gamma^{1/2} \lesssim h_t, h_z \lesssim \{(\log n/n)h_\Gamma\}^{1/(2p+2)}$. Under the assumptions of case (b) and $h_z \approx (\log n/n)^{1/(4+p)}$, the optimal convergence rate is $(\log n/n)^{4/(4+p)}$ when $p \ge 4$.
- *Extension of* Corollary 3.4
  The optimal convergence rate is achieved when $p \le 7$. When $p \ge 8$, the order of (3.12) becomes $O(L_nh_\lambda^2 + L_n^2(h_z^4 + \delta_n^2) + L_n^{-a_1} + \delta_n)$. If further $h_z^4 \approx \delta_n^2$ (i.e., $h_z \gtrsim (\log n/n)^{1/(p+5)}$), $L_nh_\lambda^2 \approx L_n^2h_z^4$ and $L_nh_z^4 \approx L_n^{-a_1}$ and $L_n^2\delta_n^2 \gtrsim \delta_n$, $L_n \gtrsim (n/\log n)^{3/2(p+5)}$ and the convergence rate is $(\log n/n)^{3a_1/2(p+5)}$.

4. Simulations Studies

This section consists of three simulation studies. Simulation 1 compares the two eigenvalue estimators (the PC-based approach and the proposed WLS approach) under the second scenario with two sampling schemes. Simulation 2 shows that when estimating the covariance function $\Upsilon(s,t,\boldsymbol{z}) = \sum_{k=1}^\infty \tau_k(\boldsymbol{z})\psi_k(s,\boldsymbol{z})\psi_k(t,\boldsymbol{z})$ is not computationally feasible, approximating Υ with the common covariance function $\Gamma(s,t,\boldsymbol{z}) = \sum_{k=1}^\infty \lambda_k(\boldsymbol{z})\phi_k(s)\phi_k(t)$ is a better choice than doing it with the pooled covariance function $\Gamma^*(s,t) = \sum_{k=1}^\infty \lambda_k^*\phi_k(s)\phi_k(t)$. Simulation 3 comprises four settings of data generation under the third scenario and is aimed to show that the proposed WLS estimator helps reveal the latent group information of each function and thus significantly improves clustering accuracy. In the following numerical studies, we employed PACE version 2.17 (available at *http://www.stat.ucdavis.edu/PACE/*) to estimate the eigenfunctions.

Table 1. Average ISE of the estimated eigenvalues with standard errors in the parentheses when $n = 200$ and $n = 400$.

	Complete				Sparse			
	$n = 200$		$n = 400$		$n = 200$		$n = 400$	
Method	$\lambda_1(z)$	$\lambda_2(z)$	$\lambda_1(z)$	$\lambda_2(z)$	$\lambda_1(z)$	$\lambda_2(z)$	$\lambda_1(z)$	$\lambda_2(z)$
PC-based	4.28 (3.45)	0.99 (0.72)	1.99 (1.45)	0.57 (0.38)	5.63 (4.02)	2.01 (1.30)	3.29 (2.29)	1.57 (0.82)
WLS	2.93 (2.39)	0.75 (0.51)	1.49 (1.07)	0.47 (0.29)	4.46 (3.66)	1.79 (1.38)	2.49 (1.83)	0.91 (0.60)

4.1. Simulation 1

For each run we generated n (=200 and 400) curves from model (3.1). For $i = 1, \ldots, n$, we let $z_i \sim U(0,1)$, $\mu(t, z_i) = 0$, $\phi_1(t) = -\cos(\pi t/10)/\sqrt{5}$, $\phi_2(t) = \sin(\pi t/10)/\sqrt{5}$, $A_{ik}(z_i) \sim N(0, \lambda_k(z_i))$ for $k = 1, 2$, where $\lambda_1(z_i) = 4\{1 + 2\sin(0.1 + \pi z_i^2/2)\}$ and $\lambda_2(z_i) = 2\{2 + \sin(2z_i\pi)\}$, and $\varepsilon_{ij} \sim N(0,1)$. For complete data, the observations were made at 51 equally spaced time points over $[0, 10]$ per curve. To generate sparse data, we randomly selected N_i points out of the 51 equally spaced time points for complete data, where N_i were randomly selected from $\{4, 5, \ldots, 10\}$ for each curve. Each simulation consists of 100 runs.

For fair comparisons, we let the bandwidth in (3.9) be 0.2 for both types of data. In the PC-based approach, we first predicted A_{ik} via the trapezoidal rule and PACE, for functional data and longitudinal data, respectively; next, we applied a p-dimensional smoother to $\{(\hat{A}_{i,k}^2, \boldsymbol{z}_i), i = 1, \ldots, n\}$ to estimate $\lambda_k(\boldsymbol{z})$. The integrated squared error (ISE) of the PC-based approach and that of the proposed WLS approach were summarized in Table 1, indicating that the proposed WLS approach outperforms the PC-based approach under both sampling schemes. Interestingly, the proposed WLS approach outperforms the PC-based approach even in the complete data setting. Intuitively, this could be explained by the fact that, in the proposed WLS approach, "more observations" are employed to directly estimate the eigenvalues, that is, $\sum_{i=1}^{n} N_i(N_i - 1)/2$ (with $N_i = 51$) versus n. This effect will disappear as n gets much larger.

The supplementary materials provide an additional variation of simulation 1, which demonstrates the dependency of the estimation ISE on p as shown in Section 3.

4.2. Simulation 2

Understanding the covariance structure is essential in some studies, such as brain functional connectivity analyses, but directly estimating the covariance $\Upsilon(s,t,\boldsymbol{z}) = \sum_{k=1}^{\infty} \eta_k(\boldsymbol{z})\psi_k(s,\boldsymbol{z})\psi_k(t,\boldsymbol{z})$ may not be computationally feasible. Here, we demonstrate that approximating Υ with $\Gamma(s,t,\boldsymbol{z}) = \sum_{k=1}^{\infty} \lambda_k(\boldsymbol{z})\phi_k(s)\phi_k(t)$ leads to a more satisfactory output than doing it with $\Gamma^*(s,t) = \mathrm{E}\{\Upsilon(s,t,\boldsymbol{z})\}$. To mimic real data, we borrowed the *phantom* function in MATLAB to design a 2D spatial structure. Specifically, three groups were considered and shown in Figure 1, and the data were generated from the fFPCA model,

$$X(t,\boldsymbol{z}) = \mu(t,\boldsymbol{z}) + \sum_{k=1}^{2} A_k(\boldsymbol{z})\psi_k(t,\boldsymbol{z}),$$

where $\mu(t,\boldsymbol{z}) = 0$, $\psi_1(t,\boldsymbol{z}) = \sqrt{2}\sin(2\pi\|\boldsymbol{z}\|t)/2$, $\psi_2(t,\boldsymbol{z}) = \sqrt{2}\cos(2\pi\|\boldsymbol{z}\|t)/2$, $A_k(\boldsymbol{z}) \sim N(0, \lambda_k(\boldsymbol{z}))$, $\lambda_k(\boldsymbol{z})$ are listed in Table 2, $\boldsymbol{z} \in [0,1] \times [0,1]$ and $t \in [0,1]$. The observations were made at 128×128 spatial locations and each location contained 31 observations made at equi-spaced time points. The data were further contaminated with measurement errors generated from $N(0, 0.2^2)$. The experiment consists of 100 runs.

Figure 1. Areas of three groups: the gray area corresponds to $\mathcal{S}_1$, the white area corresponds to $\mathcal{S}_2$, and the rest (black) area $\mathcal{S}_0$ is white noise.

Table 2. The kth eigenvalue of group i, $\lambda_{i,k}(\boldsymbol{z})$.

cluster	1	2
$\mathcal{S}_2$	$8 + \cos(z^{(1)}2\pi)/2$	$4 + \sin((0.5 + z^{(1)})2\pi)/8$
$\mathcal{S}_1$	$3 + \cos(z^{(1)}\pi)\sin(0.5 + z^{(2)})$	$1.5 + \cos(z^{(1)}\pi)\sin(0.5 + z^{(2)})/2$
$\mathcal{S}_0$	0	0

To evaluate the performance of approximating Υ with Γ and with Γ^*, we compared their ISEs, that is, $\int\{\Upsilon(s,t,\boldsymbol{z}) - \hat{\Gamma}(s,t,\boldsymbol{z})\}^2 dsdtdz$ and $\int\{\Upsilon(s,t,\boldsymbol{z}) - \hat{\Gamma}^*(s,t)\}^2 dsdtdz$. The average ISEs (and the standard error) of $\hat{\Gamma}^*$ and $\hat{\Gamma}$ are 0.3600 (0.0002) and 0.1369 (0.0027), respectively. This shows that if the data are generated from a fFPCA model, approximating the covariance function with Γ is overall a better choice given that it is computationally feasible and has a smaller ISE.

4.3. Simulation 3

To mimic real data, we employed the *phantom* structure again. We further assumed that the eigenvalues were group-dependent (listed in Table 2 and additional simulation). Two cases of the eigenfunctions were considered: in one, different groups shared a common set of eigenfunctions while in the other, the eigenfunctions were group-dependent. The goal was to demonstrate that clustering the WLS eigenvalue estimates outperforms clus-

Table 3. The summarized statistics of Simulation 3, where Recall $= \frac{TP}{TP+FN}$, and Precision $= \frac{TP}{TP+FP}$. TP, FP, FN stand for true positive, false positive, and false negative, respectively.

Method	Cluster	3A		3B		3C		3D	
		Recall	Precision	Recall	Precision	Recall	Precision	Recall	Precision
WLS	$\mathcal{S}_0$	0.983(0.002)	0.970(0.005)	0.973(0.003)	0.987(0.002)	0.985(0.002)	0.968(0.005)	0.998(0.001)	0.974(0.002)
	$\mathcal{S}_1$	0.919(0.017)	0.954(0.005)	0.893(0.023)	0.962(0.003)	0.943(0.009)	0.963(0.005)	0.915(0.009)	0.987(0.002)
	$\mathcal{S}_2$	0.859(0.042)	0.763(0.086)	0.890(0.033)	0.926(0.016)	0.884(0.027)	0.939(0.015)	0.889(0.026)	0.987(0.007)
PC^2	$\mathcal{S}_0$	1.000(0.000)	0.667(0.011)	0.992(0.002)	0.941(0.002)	1.000(0.000)	0.663(0.028)	1.000(0.000)	0.942(0.006)
	$\mathcal{S}_1$	0.238(0.028)	0.832(0.012)	0.231(0.027)	0.939(0.004)	0.232(0.062)	0.881(0.018)	0.247(0.063)	0.977(0.005)
	$\mathcal{S}_2$	0.149(0.031)	0.469(0.076)	0.150(0.032)	0.914(0.020)	0.143(0.027)	0.638(0.380)	0.126(0.034)	0.885(0.094)

tering PC^2-based eigenvalue estimates, which is closely related to a mean-adjusted FPCA model, as it is not easy to verify this argument with real data in functional connectivity study in general. The k-means approach was employed for clustering and the performance was evaluated through the precision rates and the recall rates (defined in Table 3).

3A

The data were generated from (3.1), where $\mu(\cdot,\cdot) = 0$, $\lambda_{i,k}(\boldsymbol{z})$ were given in Table 2, $\phi_k(t) = \sin(2\pi kt) + \cos(2\pi kt)$ for $k = 1$ and 2, and ε_{ij} are iid $N(0, 0.4)$.

3B

The data generation setting was almost identical to that of 3A except that the eigenvalues were further smoothed across space. Specifically, we smoothed $\{(\lambda_k(\boldsymbol{z}_i), \boldsymbol{z}_i)|i = 1, \ldots, 128^2\}$ in 3A via a Gaussian product kernel and the standard deviation for Gaussian kernel was 0.03. The purpose was to introduce smoothness across space. Due to the smoothing procedure, the correct clustering rate of a single location was modified accordingly.

3C

The procedure of generating data is similar to that of 3A with the difference that the eigenfunctions here are group-dependent. Specifically, the data were generated from (3.1), where $\mu(\cdot,\cdot) = 0$, $\lambda_{i,k}(\boldsymbol{z})$ were given in Table 2, ε_{ij} were iid from $N(0, 0.4)$, and the group-dependent eigenfunctions were

$$\phi_k(t, \boldsymbol{z}) = \begin{cases} \phi_{1,k}(t) = \sin(2\pi kt) + \cos(2\pi kt), & \text{for } \boldsymbol{z} \in \mathcal{S}_1, \\ \phi_{2,k}(t) = \sin(2\pi kt) + \cos(4\pi kt), & \text{for } \boldsymbol{z} \in \mathcal{S}_2, \\ \phi_{3,k}(t) = 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

3D

The procedure of data generation was almost identical to that of 3C except that the eigenvalues were further smoothed across space. Specifically, we smoothed $\{(\lambda_k(\boldsymbol{z}_i), \boldsymbol{z}_i)|i = 1, \ldots, 128^2\}$ in 3C via a Gaussian product kernel and the standard deviation for Gaussian kernel was 0.03. The correct clustering rate of a single location was also modified accordingly.

Table 3 indicates that clustering the WLS estimates performs significantly better in all three regions in terms of recall rates and precision rates under these four settings. Figure 2 shows the averaged predicted class labels of both strategies, and indicates that the results of clustering based on the WLS estimates are very consistent and with small variations. As the correct clustering rate of a single location has been modified for 3B and 3D, it might not be easy to compare the results between 3B and 3A, and those between 3D and 3C directly. The intuition, however, is that around the boundaries the signals become weaker after smoothing and it is easier to separate the pixels away from the boundaries to correct groups, and simultaneously around the boundaries between groups, the cost of mis-classification becomes lower and thus the precision rates are higher given the recall remains similar. The phenomenon is observed for both approaches.

5. Data Analysis: HCP Resting State fMRI Data

Functional Magnetic Resonance Imaging (fMRI) is one of the mainstays of brain research in that it allows in vivo detection of Blood Oxygenation Level Dependent (BOLD) signals (Ogawa et al. 1990) describing cortical activation. The analysis of such signals has made it possible to determine which anatomical locations of the brain activate when a specific task is performed. More recently, it has become of central interest to use BOLD signals to determine how the different parts of the brain interact, that is, infer the brain functional network organization. In fact, brain function is characterized by long-range interactions of distinct, local, and highly specialized areas (Tononi, Sporns, and Edelman 1994; Eickhoff, Yeo, and Genon 2018). Functional imaging data are routinely adopted to infer these interacting regions, which can help us understand brain organization and function by characterizing every brain location with its functional properties. Methodologically, it is essential that we take advantage of these functional properties to estimate long-range interactions while also accommodating the presence of localized units by incorporating spatial/anatomical information. Such studies are also referred to as functional connectivity studies. In this section, we investigate functional connectivity by applying the eigen-adjusted FPCA framework introduced and we compare its results with a popular approach to connectivity analysis.

Data for this application consist of fMRI scans of 40 unrelated healthy subjects, collected within the Human Connectome Project (HCP, Van Essen et al. 2012). The minimal preprocessing pipeline has been applied to the dataset (Glasser et al. 2013). This includes artifact removal, motion correction, and registration to a standard space. The relevant fMRI signals arise from the cerebral cortex, which is the outermost layer of the brain. Here we adopt a distortion minimizing 2D planar parameterization of the cerebral cortex to apply the eigen-adjusted FPCA model, and visualize the results on the cortical surface. In our analysis, we consider the first 144 sec of the fMRI signal detected under resting-state conditions, that is, without requiring the subjects

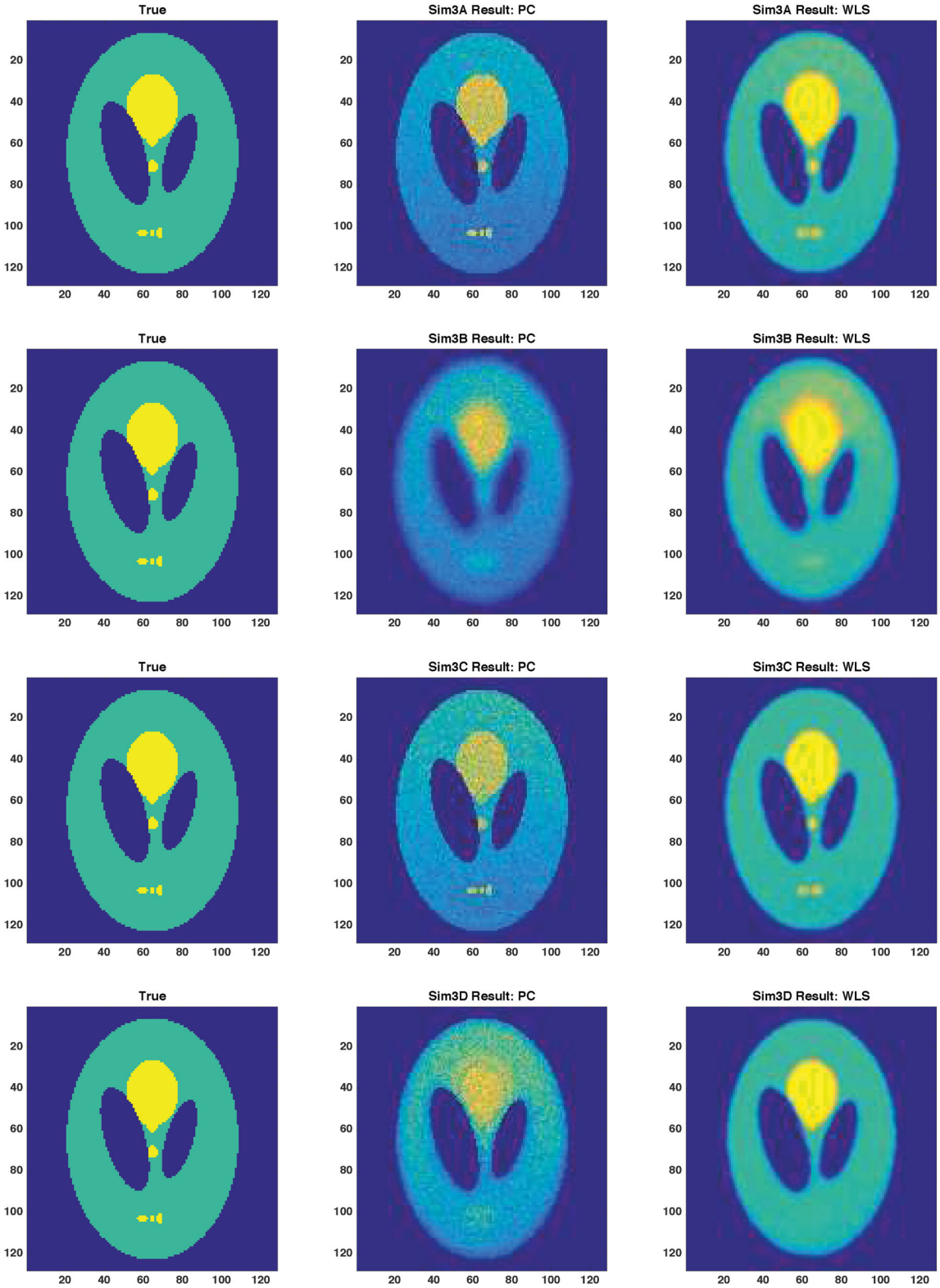

Figure 2. The true class labels and the averaged predicted labels in simulation 3.

to perform any specific task. The time samples are acquired at regular intervals of 0.72 sec. This results in a noisy signal $Y_s(t, \boldsymbol{z})$, for the sth subject, observed at a regularly spaced grid of times $t_1, \ldots, t_N$, with $N = 200$, and locations $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{N_v} \subset \mathbb{R}^2$, with $N_v = 32{,}000$.

We perform eigen-adjusted FPCA on the resting-state fMRI data as follows. We apply a 3D (2D spatial + 1D temporal) Nadaraya–Watson smoother, with product kernel, to estimate the mean function and use 5-fold cross-validation to select their bandwidths. Cross-validation suggests very large bandwidths for the time smoothing, and specifically, a zero-mean function. This is not surprising as the location-specific mean, computed across time, has been removed in the preprocessing phase.

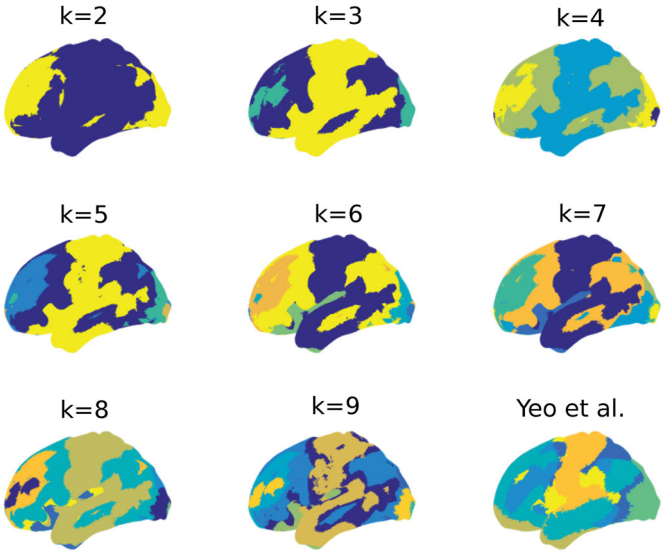

Figure 3. Here we show the results of the k-means clustering applied to the WLS estimates of the eigenvalues, for different choices of k. On the bottom-right panel, the parcellation obtained in Yeo et al. (2011). For instance, for $k = 3, \ldots, 8$, we can see that the proposed model is able to separate the motor cortex (central part of the cerebral cortex) from the rest of the cerebral cortex. The visual cortex (bottom right) is also identified as a separate cluster. For higher numbers of clusters, for example, $k = 8$, a sub-cluster isolating the primary visual cortex is also identified.

We then proceed with the estimation of the covariance functions $\Gamma(s, t, \boldsymbol{z})$, from the eigen-adjusted FPCA model in (2.1), for each spatial location $\boldsymbol{z}$. Specifically, the pooled covariance function $\Gamma^*(s, t)$ is estimated by applying a 2D local linear smoother to the associated empirical covariance. The bandwidth is chosen by 5-fold cross-validation. We then compute the eigenfunctions $\hat{\phi}_k(t)$, shown in Figure 4, by applying eigen-analysis on a discretized version of the estimated covariance function $\hat{\Gamma}^*(s, t)$. We estimate the eigenvalue maps $\hat{\lambda}_j(\boldsymbol{z})$, $j = 1, \ldots, L$ by means of WLS, as described in (3.9). We keep the first $L = 15$ eigenfunctions, which explain more than 90% of the total variance of the pooled covariance. We manually set $h_\lambda = 10$ to reflect the amount of smoothing imposed in the functional connectivity study in Yeo et al. (2011), which we use to compare our results. For each location $\boldsymbol{z}$, the resulting eigenvalues vector

$$(\hat{\lambda}_1(\boldsymbol{z}), \ldots, \hat{\lambda}_L(\boldsymbol{z}))$$

represents a multivariate summary of the temporal covariance structure $\Gamma(\cdot, \cdot, \boldsymbol{z})$. We finally identify co-activating brain regions by clustering the location-specific covariance functions through their associated eigenvalues estimates, and specifically, by applying k-means clustering on them, for different choices of k. The results are shown in Figure 3.

Resting-state time-series cannot be aligned to each other in any sensible way, thus, eigenfunctions that describe localized (in time) modes of variation should instead be ascribed to experimental artifacts. The visualization of the estimated eigenfunctions offers a diagnostic tool to assess which are the modes of variation that for this reason should be removed from the subsequent analysis. From Figure 4, we can see that the second and 15th eigenfunctions do indeed capture a localized mode of variation. However, removing them from the subsequent analysis did not substantially change the resulting parcellations.

Moreover, note that in this application, the value at each spatial location in $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_{N_v}$ can be observed for every subject. Therefore, the spatial location $\boldsymbol{z}$ is not random. Specifically, the pooled covariance function becomes $\Gamma^*(s, t) = \sum_k \bar{\lambda}_k \phi_k(s) \phi_k(t)$, where $\bar{\lambda}_k = \sum_{v=1}^{N_v} \lambda_k(\boldsymbol{z}_v)/N_v$. Therefore, if none of $\bar{\lambda}_k$'s are tied, the eigenfunctions can be consistently estimated.

In Figure 3, we also show the functional parcellation obtained in Yeo et al. (2011), where instead the following approach is adopted. A set of 1175 uniformly sampled Regions of Interest (ROI) vertices on the cerebral cortex is identified. For each subject, Pearson correlation between the fMRI time series at each spatial location $\boldsymbol{z}$ and that on the ROI vertices are computed, so that each spatial location is characterized by its first-order dependency to the ROIs. Only the top 10% correlation values for each subject are kept and these are binarized and averaged across subjects. For each vertex $\boldsymbol{z}$, this results in a multivariate descriptor

$$(\rho(\boldsymbol{z}, \text{ROI}_1), \ldots, \rho(\boldsymbol{z}, \text{ROI}_{1,175})),$$

where $\rho(\boldsymbol{z}_1, \boldsymbol{z}_2)$ is the correlation between the time-series at the two vertices $\boldsymbol{z}_1, \boldsymbol{z}_2$. Finally, spatial clustering is applied on the multivariate descriptors obtained by averaging those across subjects. In Figure 3, bottom-right panel, we show the resulting 7-networks parcellation.

Both the presented approaches aim at constructing multivariate summaries of the connectivity at each vertex, although in different fashions. To this purpose, the eigen-adjusted FPCA approach constructs a descriptor based on the temporal covariance structure at each location. Instead, the approach in Yeo et al. (2011), constructs a correlation descriptor based both on time and locations. It could be argued that in the latter approach, given that spatial information is used to construct both the multivariate descriptors and to perform spatial clustering, you would expect the clustering to be influenced by the choice of the ROIs. This cannot happen with the eigen-adjusted FPCA approach, where the multivariate descriptor exploits exclusively the temporal component, with the exception of the spatial smoothing effect introduced to contrast the low signal-to-noise ratio.

Despite the two approaches being different, and thus not immediately comparable, their results are compatible. Consider, for instance, the eigen-adjusted FPCA results in Figure 3, for $k = 7$. The proposed model is able to separate the motor cortex (central part of the cerebral cortex) from the prefrontal cortex (left part of the cerebral cortex). The visual cortex (bottom right) is also identified as a separate cluster, with a sub-cluster that seems to isolate the primary visual cortex. This separation of the brain into distinct clusters determines brain networks in a completely data-driven way, without the need for a priori explicit or implicit spatial assumptions.

6. Discussions

We have demonstrated that in a number of situations the eigen-adjusted FPCA approach is able to provide a computationally

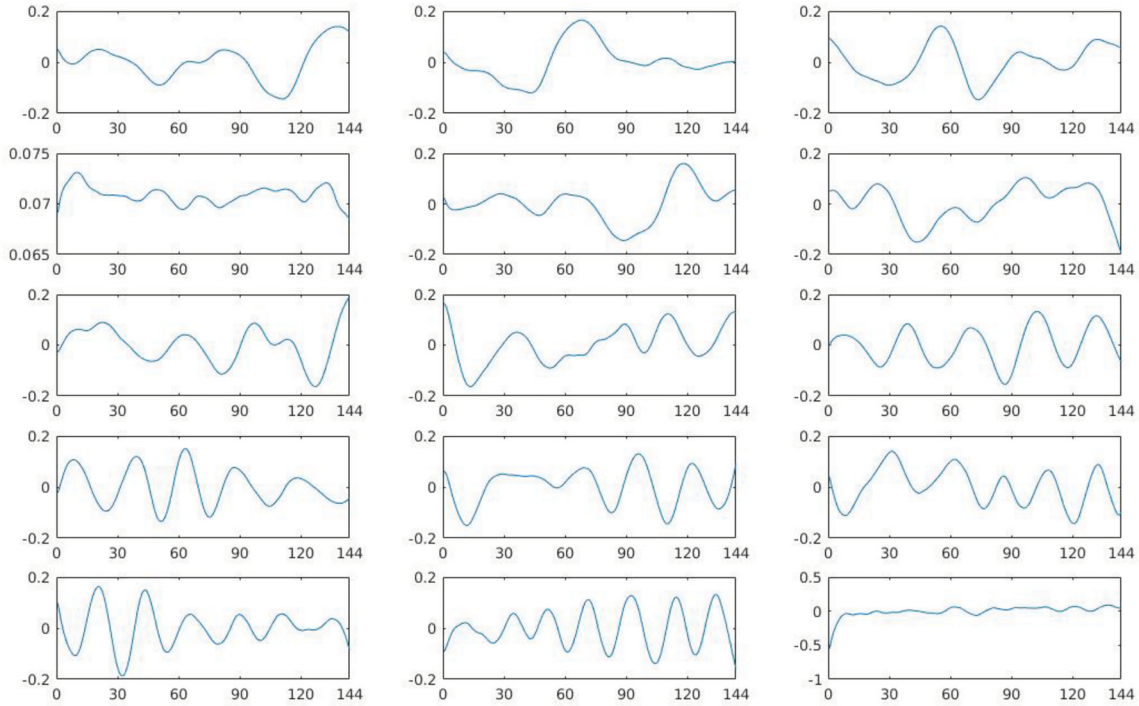

Figure 4. From left-to-right then top-to-bottom, we show the resulting first $L = 15$ eigenfunctions $\hat{\phi}_k(t)$ estimated from the pooled empirical covariance function $\hat{\Gamma}^*(s,t)$.

efficient yet flexible alternative to either a simple mean-adjusted model or a fully adjusted FPCA approach in the presence of covariates. Consistent estimates can be obtained using a WLS approach, without the need to estimate the full spatially varying covariance function but rather by a pooled version. It has been shown in simulations that the approach is effective in finite samples, particularly in relation to the kinds of data available in brain imaging applications. The application to functional connectivity shows that a comparison can be made based on the full covariate information across space, without the need to either reduce the dimension through an a priori spatial downsampling (such as definitions of ROIs) or via a seed-based approach of choosing one or two locations to compare the full data set against.

There are a number of limitations to the approach. First, should the eigenfunctions themselves directly vary with the covariates, then the model will only ever provide an approximation. However, as we have seen in the simulations, the eigen-adjusted covariance approximation is considerably better than a more simplistic approach of using a pooled covariance function in this case. In addition, it is possible that the WLS estimator could produce negative estimates, in finite samples, for positive eigenvalues. Should this be an issue, a number of possible recourses exist, including truncation or nonnegative least-squares approaches. However, for positive eigenvalues, the asymptotic properties show that this will only ever be a finite sample problem.

Overall, the eigen-adjusted FPCA approach provides a set of tools to investigate functional covariance structures that include covariate information. It is likely that as more applied questions become framed in terms of second-order structure, as seen in functional connectivity, techniques such as these will only be further needed and used.

Appendix A. Assumptions

The estimators $\hat{\mu}$ and $\hat{\Gamma}^*$ have been constructed by the local linear smoothing method. Therefore, it is natural to make the standard smoothness assumptions on the second derivatives of μ and Γ^*. Assume that the data $(\mathbf{T}_i, \mathbf{Z}_i, \mathbf{Y}_i)$, $i = 1, \ldots, n$, have the same distribution, where $\mathbf{T}_i = (T_{i1}, \ldots, T_{iN_i})$ and $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{iN_i})$. Notice that we assume $(T_{ij}, \mathbf{Z}_i)$ has marginal density $g(t, \boldsymbol{z})$. Additional assumptions and conditions are listed below.

A.1 For some constants $m_T > 0$ and $M_T < \infty$, $m_T \leq g(t, \boldsymbol{z}) \leq M_T$ for all $t \in \mathrm{T}$ and $\boldsymbol{z} \in \mathrm{Z}$. Further, $g(\cdot, \cdot)$ is differentiable with a bounded derivative.

A.2 The kernel function $K(\cdot)$ is a symmetric probability density function on $[-1, 1]$ and is of bounded variation on $[-1, 1]$. Further, we denote $\nu_2 = \int_{-1}^{1} u^2 K(u) du$.

The following assumptions are about $Y(t)$ and were also made in Li and Hsing (2010). Suppose the observation of the ith subject at time T_{ij} is $Y_{ij} = \mu(T_{ij}, \mathbf{Z}_i) + U_{ij}$, where $\mathrm{cov}(U_i(s), U_i(t)) = \Gamma(s, t, \boldsymbol{z}_i) + \sigma^2 I(s = t)$ and $\Gamma(s, t, \boldsymbol{z}_i) = \sum_{\ell} \lambda_\ell(\boldsymbol{z}_i) \phi_\ell(s) \phi_\ell(t)$. Let $\boldsymbol{h}_z = (h_z^{(1)}, \ldots, h_z^{(p)})^T$ and denote $|\boldsymbol{h}_z| = \prod_{i=1}^{p} h_z^{(i)}$. Also, $\gamma_{nk} = \left(n^{-1} \sum_{i=1}^{n} N_i^{-k} \right)^{-1}$ for $k = 1$, and 2.

B.1 μ is twice differentiable and the second derivative is bounded on $\mathrm{T} \times Z$.

B.2 $\mathrm{E}(|U_{ij}|^{\lambda_\mu}) < \infty$ and $\mathrm{E}(\sup_{t \in \mathrm{T}} |X(t)|^{\lambda_\mu}) < \infty$ for some $\lambda_\mu \in (2, \infty)$; $h_\mu \to 0$ and $(h_\mu^2 |\boldsymbol{h}_z| + h_\mu |\boldsymbol{h}_z| / \gamma_{n1})^{-1} (\log n / n)^{1 - 2/\lambda_\mu} \to 0$ as $n \to \infty$.

B.3 All second-order partial derivatives of Γ^* exist and are bounded on $\mathrm{T} \times \mathrm{T}$.

B.4 $\mathrm{E}(|U_{ij}|^{2\lambda_\Gamma}) < \infty$ and $\mathrm{E}(\sup_{t \in \mathrm{T}} |X(t)|^{2\lambda_\Gamma}) < \infty$ for some $\lambda_\Gamma \in (2, \infty)$; $h_\Gamma \to 0$ and $(h_\Gamma^4 + h_\Gamma^3 / \gamma_{n1} + h_\Gamma^2 / \gamma_{n2})^{-1} (\log n / n)^{1 - 2/\lambda_\Gamma} \to 0$ as $n \to \infty$

B.5 All second-order partial derivatives of $\lambda_k(\boldsymbol{z})$ exist and are bounded on Z for $1 \le k \le L$.
B.6 $\boldsymbol{h}_\lambda \to 0$ and $(|\boldsymbol{h}_\lambda|)^{-1}(\log n/n)^{1-2/\eta} \to 0$ as $n \to \infty$ for some $\eta \in (2, \infty)$.
B.7 For $\boldsymbol{z} \in \mathrm{Z}$ and all j, $0 < \lambda_j(\boldsymbol{z})/\lambda_j^* < \infty$.

The following assumptions are for the lemmas. Let $\boldsymbol{d}_n = (d_n^{(1)}, \ldots, d_n^{(p)})^T$.

C.1 $\mathrm{E}(|U|^\lambda) < \infty$ and $\mathrm{E}(\sup_{t\in\mathrm{T},\boldsymbol{z}\in Z} |X(t,\boldsymbol{z})|^\lambda) < \infty$ for some $\lambda \in (2, \infty)$.
C.2 Let c_n and $d_n^{(i)}$ for $i = 1, \ldots, p$ be positive sequences tending to 0, $\beta_n = c_n^2|\boldsymbol{d}_n| + c_n|\boldsymbol{d}_n|/\gamma_{n1}$ and $\beta_n^{-1}(\log n/n)^{1-2/\lambda} = o(1)$.
C.2' Let c_n be a positive sequence tending to 0, $\beta_n = c_n^2 + c_n/\gamma_{n1}$ and $\beta_n^{-1}(\log n/n)^{1-2/\lambda} = o(1)$.
C.3 $\mathrm{E}(|U|^{2\lambda}) < \infty$ and $\mathrm{E}(\sup_{t\in\mathrm{T},\boldsymbol{z}\in Z} |Y(t,\boldsymbol{z})|^{2\lambda}) < \infty$ for some $\lambda \in (2, \infty)$.
C.4 Let c_n be a positive sequence tending to 0, $\beta_n = c_n^4 + c_n^3/\gamma_{n1} + c_n^2/\gamma_{n2}$ and $\beta_n^{-1}(\log n/n)^{1-2/\lambda} = o(1)$.

Appendix B. A PC-based Approach for $\lambda_k(z)$ Estimation

Besides the WLS approach, one intuition is to first predict the PC scores and apply a p-dimensional smoother to the squared PC scores given that $\mathrm{E}\{A_k^2(\boldsymbol{z})\} = \lambda_k(\boldsymbol{z})$. When the data are dense, the PC scores can be predicted well via a numerical approach for integration. Here, we employ the trapezoidal rule for integration. Specifically,

$$\hat{A}_{ik} = \sum_{j=1}^{N_i-1} \left[\hat{U}_{ij}\hat{\phi}_k(t_{i,j}) + \hat{U}_{i,j+1}\hat{\phi}_k(t_{i,j+1})\right] \frac{(t_{i,j+1} - t_{ij})}{2}. \quad \text{(B.1)}$$

Then, a p-dimensional smoother can be applied to $\{(\hat{A}_{ik}^2, \boldsymbol{z}_i)|i = 1, \ldots, n\}$ to consistently estimate $\lambda_k(\boldsymbol{z})$ given that $\hat{A}_{ik}^2$ is a consistent estimator of $\lambda_k(\boldsymbol{z}_i)$. Let

$$\tilde{A}_{ik} = \int_T \{X_i(t) - \mu(t, \boldsymbol{z}_i)\}\hat{\phi}_k(t)dt, \text{ and}$$
$$\hat{A}_{ik}^* = \int_T \{Y_i(t) - \hat{\mu}(t, \boldsymbol{z}_i)\}\hat{\phi}_k(t)dt.$$

The asymptotics of $\hat{A}_{ik}$ can be obtained by employing the inequality,

$$|\hat{A}_{ik} - A_{ik}| \le |\hat{A}_{ik} - \hat{A}_{ik}^*| + |\hat{A}_{ik}^* - \tilde{A}_{ik}| + |\tilde{A}_{ik} - A_{ik}|.$$

First, $|\tilde{A}_{ik} - A_{ik}| = O(h_\Gamma^2 + \delta_{n1} + h_t^4 + h_z^4 + \delta_n^2)$ a.s. by applying Theorem 3.3. Second, $|\hat{A}_{ik} - \hat{A}_{ik}^*| = O(1/N_i^4)$ a.s. since $\hat{A}_{ik}^*$ is an approximation by the trapezoidal rule. Last, $|\hat{A}_{ik}^* - \tilde{A}_{ik}| = O((\log N_i/N_i)^{1/2} + \delta_n + h_t^2 + h_z^2)$ a.s. by Lemma 5 in Li and Hsing (2010) and Theorem 3.1. Therefore, $|\hat{A}_{ik} - A_{ik}| = O((\log N_i/N_i)^{1/2} + h_\Gamma^2 + \delta_{n1} + h_t^2 + h_z^2 + \delta_n)$ a.s. Since $\hat{A}_{ik}$ is a consistent estimator, $\lambda_k(\boldsymbol{z})$ can be estimated by applying a p-dimensional smoother to $\{(\hat{A}_{ik}^2, \boldsymbol{z}_i)|i = 1, \ldots, n\}$. In the numerical studies, a local linear smoother is used. Specifically,

$$\hat{\lambda}_k(\boldsymbol{z}) = \hat{b}_0, \text{ where}$$
$$(\hat{b}_0, \hat{\boldsymbol{b}}_1^T)^T = \arg\min_{\boldsymbol{b}} \frac{1}{n}\sum_{i=1}^n \left\{\hat{A}_{ik}^2 - b_0 - \boldsymbol{b}_1^T(\boldsymbol{z}_i - \boldsymbol{z})\right\}^2 \left(\prod_{k=1}^p K_{h_\lambda^{(k)}}(z_i^{(k)} - z^{(k)})\right). \quad \text{(B.2)}$$

Supplementary Materials

Supplement Assumptions, Lemmas and Proofs and additional simulation (*.pdf)

Acknowledgments

The authors would like to express gratitude for the valuable comments of an anonymous referee, the Associate Editor and the Editor. Ci-Ren Jiang's research was supported by National Science Council (Taiwan) grant NSC 101-2118-M001-013-MY2 and Ministry of Science and Technology (Taiwan) grant MOST 103-2118-M-001-001. John Aston's research was supported by the Engineering and Physical Sciences Research Council (UK) grant EP/T017961/1. The research of Jane-Ling Wang was supported by National Science Foundation grant DMS-1914917 and National Institute of Health grant UG3-0D023313 (ECHO Program).

ORCID

Ci-Ren Jiang http://orcid.org/0000-0003-2501-5543
Eardi Lila https://orcid.org/0000-0002-9199-392X
Jane-Ling Wang https://orcid.org/0000-0002-3770-9342

References

Benko, M., and Härdle, W. (2005), *Statistical Tools for Finance and Insurance*, pp. 115–134, Berlin: Springer. [2]
Benko, M., Härdle, W., and Kneip, A. (2009), "Common Functional Principal Components," *The Annals of Statistics*, 37, 1–34. [2,3]
Boente, G., Pires, A. M., and Rodrigues, I. M. (2002), "Influence Functions and Outlier Detection Under the Common Principal Components Model: A Robust Approach," *Biometrika*, 89, 861–875. [2]
Boente, G., Rodriguez, D., and Sued, M. (2010), "Inference Under Functional Proportional and Common Principal Component Models," *Journal of Multivariate Analysis*, 101, 464–475. [2,3]
Cardot, H. (2000), "Nonparametric Estimation of Smoothed Principal Components Analysis of Sampled Noisy Functions," *Journal of Nonparametric Statistics*, 12, 503–538. [1]
——— (2006), "Conditional Functional Principal Components Analysis," *Scandinavian Journal of Statistics*, 34, 317–335. [1]
Chen, L.-H., and Jiang, C.-R. (2017), "Multidimensional Functional Principal Component Analysis," *Statistics and Computing*, 27, 1181–1192. [1]
Chiou, J.-M., Müller, H.-G., and Wang, J.-L. (2003), "Functional Quasi-Likelihood Regression Models with Smooth Random Effects," *Journal of the Royal Statistical Society* Series B, 65, 405–423. [1]
Coffey, N., Harrison, A. J., Donoghue, O. A., and Hayes, K. (2011), "Common Functional Principal Components Analysis: A New Approach to Analyzing Human Movement Data," *Human Movement Science*, 30, 1144–1166. [2]
Dauxois, J., Pousse, A., and Romain, Y. (1982), "Asymptotic Theory for the Principal Component Analysis of a Vector Random Function: Some Applications of Statistical Inference," *Journal of Multivariate Analysis*, 12, 136–154. [1]
Eickhoff, S. B., Yeo, B. T., and Genon, S. (2018), "Imaging-Based Parcellations of the Human Brain," *Nature Reviews Neuroscience*, 19, 672–686. [7]
Fan, J., and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, London: Chapman and Hall. [3]
Flury, B. K. (1984), "Common Principal Components in k Groups," *Journal of the American Statistical Association*, 79, 892–898. [2]
——— (1986), "Asymptotic Theory for Common Principal Components," *The Annals of Statistics*, 14, 418–430. [2]
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., Van Essen, D. C., and Jenkinson, M. (2013), "The Minimal Preprocessing Pipelines for the Human Connectome Project," *NeuroImage*, 80, 105–124. [7]
Hall, P., and Horowitz, J. L. (2007), "Methodology and Convergence Rates for Functional Linear Regression," *The Annals of Statistics*, 35, 70–91. [5]

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. [1,2]

James, G. M., Hastie, T. J., and Suger, C. A. (2000), "Principal Components Models for Sparse Functional Data," *Biometrika*, 87, 587–602. [1]

Jiang, C.-R., Aston, J. A. D., and Wang, J.-L. (2009), "Smoothing Dynamic Positron Emission Tomography Time Courses Using Functional Principal Components," *NeuroImage*, 47, 184–193. [1]

Jiang, C.-R., and Wang, J.-L. (2010), "Covariate Adjusted Functional Principal Components Analysis for Longitudinal Data," *The Annals of Statistics*, 38, 1194–1226. [1,2]

——— (2011), "Functional Single Index Models for Longitudinal Data," *The Annals of Statistics*, 39, 362–388. [1]

Li, M., Staicu, A.-M., and Bondell, H. D. (2015), "Incorporating Covariates in Skewed Functional Data Models," *Biostatistics*, 16, 413–426. [1]

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. [1,10,11]

Ogawa, S., Lee, T. M., Kay, A. R., and Tank, D. W. (1990), "Brain Magnetic Resonance Imaging with Contrast Dependent on Blood Oxygenation," *Proceedings of the National Academy of Sciences of the United States of America*, 87, 9868–9872. [7]

Rao, C. R. (1958), "Some Statistical Methods for Comparison of Growth Curves," *Biometrics*, 14, 1–17. [1]

Rice, J., and Silverman, B. (1991), "Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves," *Journal of the Royal Statistical Society*, Series B, 53, 233–243. [1]

Rice, J. A., and Wu, C. (2001), "Nonparametric Mixed Effects Models for Uunequally Sampled Noisy Curves," *Biometrics*, 57, 253–259. [1]

Shang, H. L. (2014), "A Survey of Functional Principal Component Analysis," *AStA Advances in Statistical Analysis*, 98, 121–142. [1]

Tononi, G., Sporns, O., and Edelman, G. M. (1994), "A Measure for Brain Complexity: Relating Functional Segregation and Integration in the Nervous System," *Neurobiology*, 91, 5033–5037. [7]

Van Essen, D., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S., Della Penna, S., Feinberg, D., Glasser, M., Harel, N., Heath, A., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., and Yacoub, E. (2012), "The Human Connectome Project: A Data Acquisition Perspective," *NeuroImage*, 62, 2222–2231. [7]

Yao, F., Lei, E., and Wu, Y. (2015), "Effective Dimension Reduction for Sparse Functional Data," *Biometrika*, 102, 421–437. [5]

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. [1,2,3,5]

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R., Fischl, B., Liu, H., and Buckner, R. L. (2011), "The Organization of the Human Cerebral Cortex Estimated by Intrinsic Functional Connectivity," *Journal of Neurophysiology*, 106, 1125–1165. [9]

Zhang, X., Park, B. U., and Wang, J.-L. (2013), "Time-Varying Additive Models for Longitudinal Data," *Journal of the American Statistical Association*, 108, 983–998. [1]

Zhang, X., and Wang, J.-L. (2015), "Varying-Coefficient Additive Models for Functional Data," *Biometrika*, 102, 15–32. [1]