Supplementary materials for this article are available at https://doi.org/10.1007/s13253-022-00495-1.

Robust Functional Principal Component Analysis Based on a New Regression Framework

Haolun SHI and Jiguo CAO

It is of great interest to conduct robust functional principal component analysis (FPCA) that can identify the major modes of variation in the stochastic process with the presence of outliers. A new robust FPCA method is proposed in a new regression framework. An M-estimator for the functional principal components is developed based on the Huber's loss by iteratively fitting the residuals from the Karhunen–Lovève expansion for the stochastic process under the robust regression framework. Our method can naturally accommodate sparse and irregularly-sampled data. When the functional data have outliers, our method is shown to render stable and robust estimates of the functional principal components; when the functional data have no outliers, we show via simulation studies that the performance of our approach is similar to that of the conventional FPCA method. The proposed robust FPCA method is demonstrated by analyzing the Hawaii ocean oxygen data and the kidney glomerular filtration rates for patients after renal transplantation.

Supplementary materials accompanying this paper appear on-line.

Key Words: Functional data analysis; Robust statistics; M-estimation; Likelihood.

1. INTRODUCTION

Functional data analysis (FDA) is a growing statistical field for analyzing curves, surfaces, or any multidimensional functions, in which each random function is treated as a sample element (Ramsay and Silverman 2005; Ferraty and Vieu 2006). Functional data analysis is used commonly in many applications such as time-course gene expressions and brain scan images. Functional principal component analysis (FPCA) is a powerful tool for extracting the main modes of variation in the functional data. FPCA decomposes the latent stochastic process as the sum of the mean function and a linear combination of functional principal components (FPCs). The FPCs serve as a foundation for effective dimensionality reduction of the infinite-dimensional functional data, because the top few FPCs can explain

H. Shi · J. Cao (✉) Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, Canada (E-mail: *jiguo_cao@sfu.ca*).

Journal of Agricultural, Biological, and Environmental Statistics, Volume 27, Number 3, Pages 523–543
https://doi.org/10.1007/s13253-022-00495-1

the majority of the variability within the underlying stochastic process. The methodologies for conducting FPCA, as well as its theoretical properties and extensions, have been well studied in the literature, see, e.g., Ramsay and Dalzell (1991), Silverman (1996), Dauxois et al. (1982), Mas (2002), Hall and Horowitz (2007), and Yao et al. (2005). Recently, Lin et al. (2016) and Nie and Cao (2020) added a penalty function on the support of FPCs, which led to better visualization as their estimated FPCs become nonzero only in the intervals with major variation. Sang et al. (2017) proposed to conduct FPCA from a parametric perspective to improve the interpretability of the FPCs. To enhance the predictiveness of FPCs, Nie et al. (2018) developed a supervised version of FPCA that accommodates the correlation between FPCs and a response variable of interest. Dong et al. (2018) applied FPCA to analyze the estimated glomerular filtration rates (GFR) curves at multiple time points post kidney transplantation. Shi et al. (2021) considered the informative missingness of the longitudinal trajectories in FPCA.

The conventional FPCA method is based on eigendecomposition of the bivariate covariance function of the functional data. The conventional FPCA method first estimates the bivariate covariance function of the functional data by the two-dimensional smoothing approach such as the local polynomial regression (Yao et al. 2005) The FPCs are then estimated as the eigenfunctions of the estimated covariance function. However, when the functional data have outliers and contamination, the conventional FPCA method will render untrustworthy results because the estimate for the bivariate covariance function is sensitive to outliers. As a result, abnormal observations severely affect and distort the FPCs; thus the estimated FPCs may not be representative of the primary modes of variation of the uncontaminated stochastic process.

To safeguard the estimation of the FPCs against outliers, various robust extensions for FPCA have been developed. For instance, Hyndman and Ullah (2007) applied robust PCA on smoothed functional data based on a projection-pursuit approach. Sawant et al. (2012) borrowed the idea of robust PCA methods (Hubert et al. 2005; Billor et al. 2000) to estimate the spline basis coefficients and to obtain robust estimates of the FPCs. Lee et al. (2013) expressed the FPCs using the natural cubic splines and developed a fast computational algorithm based on alternating regression. Gervini (2008) developed a robust method using spherical PCA, which is resistant to contamination subject to the elliptical assumption. Bali et al. (2011) extended the robust projection-pursuit approach in PCA to functional data by combining the approach with various smoothing methods and proved the consistency of their estimators. Most of these approaches work on dense functional data by converting the functional data into the multivariate data on a common grid using smoothing or basis function expansion. For sparse and irregular functional data, Gervini (2009) proposed a robust method based on a reduced-rank t-model and established its theoretical properties. Recently, Boente and Salibian-Barrera (2015) developed an S-estimator for robust FPCA, which is based on the best lower-dimensional approximations to the functional data, and established the consistency under elliptically distributed random elements.

In this paper, we propose a new robust method for conducting FPCA based on the best approximation to the underlying stochastic process using M-estimator. We adapt an alternating robust regression approach, which circumvents the need for approximating the covariance function. The mean function and the FPCs are expanded using B-spline basis

functions. We directly estimate them under the robust regression framework. To satisfy the orthogonality constraints of the estimated FPCs, we recast the problem into a constrained optimization problem. A sequential procedure is proposed for seeking robust estimates of the FPCs.

The main contribution of our method is twofold. First, we propose a new framework for conducting FPCA based on the minimization of the objective function in the M-estimation. To the best of our knowledge, robust FPCA under such a framework is not previously available in the literature. Second, our method for extracting the FPCs is purely based on robust regression and thus does not require eigendecomposition and covariance surface estimation as in the conventional FPCA. As a result, the proposed method naturally works for sparsely and irregularly-sampled data and is shown to have superior operating characteristics than the conventional method in the simulation studies. The computing scripts for implementing our method can be obtained from the webpage https://github.com/haoluns/robustFPCA.

The rest of the paper is organized as follows. In Sect. 2, we present the proposed methodology of the robust functional principal component analysis. In Sect. 3, we conduct simulation studies to assess the empirical precision in the estimation of the underlying FPCs and compare the performance of our method with that of the conventional ones. In Sect. 4, we apply the proposed method on the Hawaii ocean oxygen data and the kidney glomerular filtration rates for patients after renal transplantation, and illustrate the robustness of the estimation in the presence of outliers. Finally, Sect. 5 concludes the paper with some remarks.

2. METHODOLOGY

We consider n independent realizations of a stochastic process $X(t)$, where $t \in \mathcal{T}$. Let y_{ij} denote the j-th observation of the i-th random function $X_i(t)$, i.e., the realization of the stochastic process $X(t)$ for the i-th individual. For $i = 1, \ldots, n$, and $j = 1, \ldots, n_i$, where n_i is the number of samples from $X_i(t)$, the data point y_{ij} is modeled as

$$y_{ij} = X_i(t_{ij}) + \epsilon_{ij},$$

where ϵ_{ij} is a random measurement error term following the normal distribution with mean zero. It is worth noting that the measurement error term ϵ_{ij} is homoscedastic, i.e., its variance is constant and does not depend on the time t_{ij}. Without loss of generality, we assume that all the n_i's are the same.

Each individual's underlying random function $X_i(t)$ can be expressed in terms of an expansion of a series of orthonormal basis functions, that is,

$$X_i(t) = \mu(t) + \sum_{k=1}^{\infty} \alpha_{ik}\psi_k(t),$$

where the basis functions should satisfy $\int_{\mathcal{T}} \psi_k^2(t)\mathrm{d}t = 1$, denoted as $\|\psi_k\|^2 = 1$, and $\int_{\mathcal{T}} \psi_k(t)\psi_l(t)\mathrm{d}t = 1$, denoted as $\langle \psi_k, \psi_l \rangle = 1$, if $k = l$, and 0 otherwise. The orthonormal basis functions $\psi_k(t), k = 1, \ldots, \infty$, are called functional principal components (FPCs), and the basis coefficients α_{ik} are called FPC scores.

We propose a functional orthonormal approximation method for estimating the first K FPCs $\psi_k(t)$. The estimation procedure seeks to locate an M-estimator that minimizes

$$\sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \rho\left(\frac{y_{ij} - \mu(t_{ij}) - \sum_{k=1}^{K} \alpha_{ik} \psi_k(t_{ij})}{\sigma} \right).$$

where $\rho(\cdot)$ is a non-negative and symmetric function that represents the contribution of each of data points to the objective function, and σ is a robust estimate for the scale. Setting $\rho(e) = e^2$ and $\sigma = 1$ results in a regular least-squared estimator and corresponds to the usual functional principal components analysis.

Various objective functions $\rho(\cdot)$ have been proposed for the robust estimation when outliers are present in the data. For example, Huber's loss (Huber and Ronchetti 2009) provides stable numerical performance due to its strong convexity, which is specified as

$$\rho(e) = \begin{cases} \frac{1}{2}e^2 & \text{for } |e| \le q \\ q|e| - \frac{1}{2}q^2 & \text{for } |e| > q \end{cases}$$

where q is a tuning constant, and its optimal value is recommended as $q = 1.345$. Alternatively, the Tukey's biweight loss function can be adopted, which is specified as

$$\rho(e) = \begin{cases} \frac{q^2}{6}\{1 - [1 - (\frac{e}{q})^2]^3\} & \text{for } |e| \le q, \\ \frac{q^2}{6} & \text{for } |e| > q, \end{cases}$$

where the tuning constant q is recommended to be $q = 4.685$.

2.1. ESTIMATING THE MEAN FUNCTION

The estimate for the mean function $\mu(t)$ is obtained by minimizing

$$\sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \rho\left(\frac{y_{ij} - \mu(t_{ij})}{\sigma_\mu} \right). \tag{1}$$

where σ_μ is a robust estimate of scale.

We express the $\mu(t)$ in terms of the B-spline basis expansion

$$\mu(t) = \sum_{s=1}^{S} \beta_{\mu,s} b_s(t),$$

where S denotes the number of spline bases and $\beta_{\mu,s}$ denotes the basis coefficient for the sth basis function $b_s(t)$, $s = 1, \ldots, S$. For notational simplicity, we denote $\boldsymbol{\beta}_\mu = (\beta_{\mu,1}, \ldots, \beta_{\mu,S})^\top$ and $\mathbf{b}(t) = (b_1(t), \ldots, b_S(t))^\top$, and rewrite the equation above as

$$\mu(t) = \boldsymbol{\beta}_\mu^\top \mathbf{b}(t).$$

The goal is to estimate the coefficient vector $\boldsymbol{\beta}_\mu$. We may recast the estimation as a weighted least square problem. Let $w(e) = \rho'(e)/e$ and let $w_{ij} = w((y_{ij} - \boldsymbol{\beta}_\mu^\top \mathbf{b}(t_{ij}))/\sigma_\mu)$. Taking derivative of (1) with respect to $\boldsymbol{\beta}_\mu$ would lead to the following estimation equation.

$$\sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} w_{ij} \left(y_{ij} - \boldsymbol{\beta}_\mu^\top \mathbf{b}(t_{ij}) \right) \mathbf{b}(t_{ij}) = \mathbf{0}. \tag{2}$$

The equation is evidently similar to that of a weighted least square problem with case weight equating w_{ij}/n_i. It should be noted that the weights w_{ij} is dependent on the value of $\boldsymbol{\beta}_\mu$, thus an iteratively reweighted least square algorithm is used for the estimation of $\boldsymbol{\beta}_\mu$, which is detailed as follows.

1. Initialize $\boldsymbol{\beta}_\mu^{(0)}$.
2. Let ℓ denote the index of iteration. Update the scale estimate $\sigma_\mu^{(\ell)}$. Let $e_{ij} = y_{ij} - \mathbf{b}(t_{ij})^\top \boldsymbol{\beta}_\mu^{(\ell)}$. The $\sigma_\mu^{(\ell)}$ updated as

$$\sigma_\mu = \text{WMED}|e_{ij} - \text{WMED}(e_{ij})|, \tag{3}$$

   where WMED is the weighted median with $1/n_{ij}$ as weight. When the weights $1/n_{ij}$ are all equal, the above formula reduces to the median absolute deviation (MAD) of the residual error e_{ij}.
3. Update the weight as $w_{ij}^{(\ell)} = w((y_{ij} - \mathbf{b}(t_{ij})^\top \boldsymbol{\beta}_\mu^{(\ell)})/\sigma_\mu^{(\ell)})$.
4. Let $\mathbf{W}^{(\ell)}$ denote the diagonal matrix where the diagonal entries are weights $w_{ij}^{(\ell)}/n_i$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. Let $\mathbf{Y}$ denote the response vector consisting of y_{ij}, and $\mathbf{X}$ denote the design matrix where the row is $\mathbf{b}(t_{ij})$, for $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. Update the estimate of $\boldsymbol{\beta}_\mu$ as

$$\boldsymbol{\beta}_\mu^{(\ell+1)} = \left\{ \mathbf{X}^\top \mathbf{W}^{(\ell)} \mathbf{X} \right\}^{-1} \mathbf{X}^\top \mathbf{W}^{(\ell)} \mathbf{Y}.$$

5. Repeat Steps 2 and 3 until convergence.

After obtaining the estimates for the mean function $\widehat{\mu}(t)$, the K functional principal components are obtained in a sequential manner, i.e., the kth FPC is approximated conditional on the estimated values of the first $k-1$ FPCs.

2.2. ESTIMATING THE FIRST FUNCTIONAL PRINCIPAL COMPONENT AND SCORE

The first FPC $\psi_1(t)$ is obtained by minimizing

$$\sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \rho\left(\frac{y_{ij} - \widehat{\mu}(t_{ij}) - \alpha_{i1}\psi_1(t_{ij})}{\sigma_1} \right), \tag{4}$$

subject to $\|\psi_1\|^2 = 1$, where σ_1 is a robust estimate of scale.

The first FPC $\psi_1(t)$ can be expressed in terms of the B-spline basis expansion

$$\begin{aligned}\psi_1(t) &= \sum_{s=1}^{S} \beta_{1,s} b_s(t) \\ &= \boldsymbol{\beta}_1^\top \mathbf{b}(t)\end{aligned}$$

where $\boldsymbol{\beta}_1 = (\beta_{1,1}, \ldots, \beta_{1,S})^\top$ is the basis coefficient vector for the first FPC.

Let $\boldsymbol{\alpha}_1 = (\alpha_{11}, \ldots, \alpha_{n1})^\top$ denote the FPC scores for all the n subjects. Our goal is to estimate both $\boldsymbol{\alpha}_1$ and the B-spline coefficient vector $\boldsymbol{\beta}_1$. The estimation is conducted in an iterative manner, i.e., conditional on the current estimate of $\boldsymbol{\alpha}_1$; the estimate of $\boldsymbol{\beta}_1$ is updated by minimizing the loss function in (4), and vice versa. The algorithm is detailed as follows.

1. Initialize $\boldsymbol{\beta}_1^{(0)}$ and hence $\psi_1^{(0)}$ subject to $\|\psi_1^{(0)}\|^2 = 1$.
2. Let ℓ denote the current index of iteration. Conditional on the current estimate $\boldsymbol{\beta}_1^{(\ell)}$, the estimated FPC is

$$\psi_1^{(\ell)}(t) = \boldsymbol{\beta}_1^{(\ell)\top} \mathbf{b}(t).$$

   Conditional on $\psi_1^{(\ell)}(t)$, for each $i = 1, \ldots, n$, we obtain the estimate for $\alpha_{i1}^{(\ell)}$ by minimizing the sum of loss functions related to the ith individual,

$$\sum_{j=1}^{n_i} \rho\left(\frac{y_{ij} - \widehat{\mu}(t_{ij}) - \alpha_{i1}\psi_1^{(\ell)}(t_{ij})}{\sigma_1}\right).$$

   where σ_1 is a robust estimate of scale, which can be computed based on (3) by setting $e_{ij} = y_{ij} - \widehat{\mu}(t_{ij}) - \alpha_{i1}\psi_1^{(\ell)}(t_{ij})$. An iteratively reweighted least square algorithm similar to Sect. 2.1 can be used for seeking the estimate of $\alpha_{i1}^{(\ell)}$.
3. Conditional on the current estimate $\boldsymbol{\alpha}_1^{(\ell)}$, update the estimate of $\boldsymbol{\beta}_1$ by minimizing the loss function

$$\sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \rho\left(\frac{y_{ij} - \widehat{\mu}(t_{ij}) - \alpha_{i1}^{(\ell)}\psi_1(t_{ij})}{\sigma_1}\right),$$

   subject to the $\|\psi_1\|^2 = 1$. To be specific, this can be formulated as

$$\boldsymbol{\beta}_1^{(\ell+1)} = \text{argmin}_{\boldsymbol{\beta}_1} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \rho\left(\frac{y_{ij} - \widehat{\mu}(t_{ij}) - \alpha_{i1}^{(\ell)}\boldsymbol{\beta}_1{}^\top \mathbf{b}(t_{ij})}{\sigma_1}\right),$$

   where σ_1 is a robust estimate of scale, which can be computed based on (3) by setting $e_{ij} = y_{ij} - \widehat{\mu}(t_{ij}) - \alpha_{i1}^{(\ell)}\boldsymbol{\beta}_1{}^\top \mathbf{b}(t_{ij})$. The minima can be found by applying the iteratively reweighted least square algorithm similar to Sect. 2.1. The norm of ψ_1 is computed as $\|\psi_1\|^2 = \boldsymbol{\beta}_1{}^\top \mathbf{R} \boldsymbol{\beta}_1$, where $\mathbf{R}$ is a matrix of integrals, $(\mathbf{R})_{ij} = \int_{\mathcal{T}} b_i(t) b_j(t) \mathrm{d}t$ The unconstrained estimate of $\boldsymbol{\beta}_1$ is obtained first and subsequently rescaled by the norm.

4. Repeat Steps 2 and 3 until the convergence criterion is reached, i.e., the maximum element in the vector $|\boldsymbol{\beta}_1^{(\ell+1)} - \boldsymbol{\beta}_1^{(\ell)}|$ is smaller than a prespecified threshold ϵ. Typically, ϵ can take a small value such as 0.0001.

In terms of the identifiability, it is worth noting that the algorithm may result in two estimates of $\psi_1(\cdot)$ having opposite signs, and will converge to either of one of them. This is an expected property of the functional principal component, because both $\psi_1(\cdot)$ and $-\psi_1(\cdot)$ satisfy the constraint $\|\psi_1\|^2 = 1$. In practice, we also find that the empirical performance of the algorithm's convergence is quite stable, and have never encountered a case of non-convergence.

2.3. ESTIMATING SUBSEQUENT FUNCTIONAL PRINCIPAL COMPONENTS AND SCORES

The subsequent functional principal components are obtained in a sequential manner. Let J denote the current index of the functional principal component of interest. From the first $(J-1)$ estimation steps, we obtain the estimates $\widehat{\boldsymbol{\beta}}_k$ and the resulting $\widehat{\psi}_k, k = 1, \ldots, J-1$. Given the estimated values of these first $(J-1)$ functional principal components, the Jth functional principal component ψ_J is obtained by minimizing the loss function

$$\sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \rho \left(\frac{y_{ij} - \widehat{\mu}(t_{ij}) - \sum_{k=1}^{J-1} \alpha_{ik} \widehat{\psi}_k(t_{ij}) - \alpha_{iJ} \psi_J(t_{ij})}{\sigma_J} \right), \tag{5}$$

subject to $\|\psi_J\|^2 = 1$ and $\langle \widehat{\psi}_k, \psi_J \rangle = 1$ if $k = J$, and 0 otherwise.

The algorithm is detailed as follows.

1. Initialize $\boldsymbol{\beta}_J^{(0)}$ and hence $\psi_J^{(0)}$ subject to $\|\psi_J^{(0)}\|^2 = 1$.

2. Let ℓ denote the current index of iteration, and let $\boldsymbol{\alpha}_i = (\alpha_{i1}, \ldots, \alpha_{iJ})^\top$ denote the FPC scores for all the FPCs in the ith subject. Conditional on the current estimate $\boldsymbol{\beta}_J^{(\ell)}$ and all the estimated $\widehat{\boldsymbol{\beta}}_k$ from the previous steps, the estimated FPCs are

$$\psi_k^{(\ell)}(t) = \widehat{\boldsymbol{\beta}}_k^\top \mathbf{b}(t), \quad k = 1, \ldots, J-1,$$
$$\psi_J^{(\ell)}(t) = \boldsymbol{\beta}_J^{(\ell)\top} \mathbf{b}(t).$$

Conditional on $\psi_1^{(\ell)}, \ldots, \psi_J^{(\ell)}$, for each $i = 1, \ldots, n$, we obtain the estimate for $\alpha_{i1}^{(\ell)}, \ldots, \alpha_{iJ}^{(\ell)}$ by minimizing the sum of the loss functions related to the ith individual,

$$\sum_{j=1}^{n_i} \rho \left(\frac{y_{ij} - \widehat{\mu}(t_{ij}) - \sum_{k=1}^{J} \alpha_{ik} \widehat{\psi}_k(t_{ij})}{\sigma_J} \right).$$

where σ_J is a robust estimate of scale, which can be computed based on (3) by setting $e_{ij} = y_{ij} - \widehat{\mu}(t_{ij}) - \sum_{k=1}^{J} \alpha_{ik} \widehat{\psi}_k(t_{ij})$. An iteratively reweighted least square algorithm similar to Sect. 2.1 can be used for seeking the estimate of $\alpha_{i1}^{(\ell)}, \ldots, \alpha_{iJ}^{(\ell)}$.

3. Denote $(\alpha_{i1}, \ldots, \alpha_{iJ})^\top$ as $\boldsymbol{\alpha}_i^{(\ell)}$. Conditional on the current estimate $\boldsymbol{\alpha}_i^{(\ell)}$, update the estimate of $\boldsymbol{\beta}_J$ by minimizing

$$\sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \rho \left(\frac{y_{ij} - \widehat{\mu}(t_{ij}) - \sum_{k=1}^{J-1} \alpha_{ik}^{(\ell)} \widehat{\psi}_k(t_{ij}) - \alpha_{iJ}^{(\ell)} \psi_J(t_{ij})}{\sigma_J} \right),$$

subject to $\|\psi_J\|^2 = 1$ and $\langle \widehat{\psi}_k, \psi_J \rangle = 1$ if $k = J$, and 0 otherwise, and σ_J is a robust estimate of scale, which can be computed based on (3) by setting $e_{ij} = y_{ij} - \widehat{\mu}(t_{ij}) - \sum_{k=1}^{J-1} \alpha_{ik}^{(\ell)} \widehat{\psi}_k(t_{ij}) - \alpha_{iJ}^{(\ell)} \psi_J(t_{ij})$. This can be formulated as a nonlinear optimization problem with equality constraints. To be specific,

$$\boldsymbol{\beta}_J^{(\ell+1)} = \operatorname{argmin}_{\boldsymbol{\beta}_J} \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \rho \left(\frac{y_{ij} - \widehat{\mu}(t_{ij}) - \sum_{k=1}^{J-1} \alpha_{ik}^{(\ell)} \widehat{\boldsymbol{\beta}}_k^\top \mathbf{b}(t_{ij}) - \alpha_{iJ}^{(\ell)} \boldsymbol{\beta}_J^\top \mathbf{b}(t_{ij})}{\sigma_J} \right)$$

subject to

$$\int_{\mathcal{T}} \left\{ \boldsymbol{\beta}_J^\top \mathbf{b}(t) \right\}^2 \mathrm{d}t = 1,$$

and

$$\int_{\mathcal{T}} \left\{ \widehat{\boldsymbol{\beta}}_k^\top \mathbf{b}(t) \right\} \left\{ \boldsymbol{\beta}_J^\top \mathbf{b}(t) \right\} \mathrm{d}t = 0, \qquad k = 1, \ldots, J-1.$$

The first constraint above can be fulfilled by rescaling the estimate of $\boldsymbol{\beta}_J$ by the norm, i.e., dividing the estimate of $\boldsymbol{\beta}_J$ by $\|\psi_J\|$, where $\|\psi_J\|^2 = \boldsymbol{\beta}_J{}^\top \mathbf{R} \boldsymbol{\beta}_J$, and $\mathbf{R}$ is a matrix of integrals, $(\mathbf{R})_{ij} = \int_{\mathcal{T}} b_i(t) b_j(t) \mathrm{d}t$. Estimation under the second constraint can be casted as a nonlinear optimization problem under the linear equality constraints $\boldsymbol{\beta}_J^\top \mathbf{C}_k = 0$, where $\mathbf{C}_k = \widehat{\boldsymbol{\beta}}_k^\top \mathbf{R}$, for $k = 1, \ldots, J-1$. The solution to such a constrained nonlinear optimization problem can be computed via the Augmented Lagrangian algorithm (Conn et al., 1991; Birgin and Martinez, 2008). The details are as follows.

(a) Initialize $\boldsymbol{\beta}_J^{(0)}$.

(b) Let h denote the index of subiteration. Update the scale estimate $\sigma_J^{(h)}$ according to (3) by setting $e_{ij} = y_{ij} - \widehat{\mu}(t_{ij}) - \sum_{k=1}^{J-1} \alpha_{ik}^{(\ell)} \widehat{\boldsymbol{\beta}}_k^\top \mathbf{b}(t_{ij}) - \alpha_{iJ}^{(\ell)} \boldsymbol{\beta}_J^{(h)\top} \mathbf{b}(t_{ij})$.

(c) Update the weight as $w_{ij}^{(h)} = w\Big(\Big(y_{ij} - \widehat{\mu}(t_{ij}) - \sum_{k=1}^{J-1} \alpha_{ik}^{(\ell)} \widehat{\boldsymbol{\beta}}_k^\top \mathbf{b}(t_{ij}) - \alpha_{iJ}^{(\ell)} \boldsymbol{\beta}_J^{(h)\top} \mathbf{b}(t_{ij}) \Big) / \sigma_J^{(h)} \Big)$.

(d) Let $\mathbf{W}^{(h)}$ denote the diagonal matrix where the diagonal entries are weights $w_{ij}^{(h)}/n_i$ for $i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. Let $\mathbf{Y}$ denote the response vector consisting of y_{ij}, and $\mathbf{X}$ denote the design matrix where the row is $\mathbf{b}(t_{ij})$, for

$i = 1, \ldots, n$ and $j = 1, \ldots, n_i$. Update the estimate of $\boldsymbol{\beta}_J$ as

$$\boldsymbol{\beta}_J^{(h+1)} = \text{argmin}_{\boldsymbol{\beta}_J} \left\{ \boldsymbol{\beta}_J^\top \left(\mathbf{X}^\top \mathbf{W}^{(h)} \mathbf{X} \right) \boldsymbol{\beta}_J - 2 \boldsymbol{\beta}_J^\top \left(\mathbf{X}^\top \mathbf{W}^{(h)} \mathbf{Y} \right) \right\},$$

subject to

$$\boldsymbol{\beta}_J^\top \mathbf{C}_k = 0.$$

The problem can be readily solved by any software package for quadratic programming, such as "quadprog" in R.

(e) Repeat the steps above until convergence.

4. Repeat Steps 2 and 3 until the convergence criterion is reached, i.e., the maximum element in the vector $|\boldsymbol{\beta}_J^{(\ell+1)} - \boldsymbol{\beta}_J^{(\ell)}|$ is smaller than a prespecified threshold ϵ. Typically, ϵ can take a small value such as 0.0001.

We select the number of basis functions that achieves the optimal smoothness control by using a cross validation procedure and loop through all the candidate values of the number of spline bases. To be specific, we divide the dataset into F folds, and perform the procedure on the $F - 1$ folds of data while holding the remaining one out for testing. Given the estimated mean function $\widehat{\mu}(\cdot)$, FPCs $\widehat{\psi}_k(\cdot)$ under a candidate value of S, we may compute the FPC scores $\hat{\alpha}_{ik}$ on the test dataset by regressing the observed $y_{ij} - \widehat{\mu}(t_{ij})$ against the $\widehat{\psi}_k(t_{ij})$. The cross validation score for such a fold can be calculated as

$$CV(S) = \sum_{i=1}^{n} \frac{1}{n_i} \sum_{j=1}^{n_i} \rho \left(\frac{y_{ij} - \widehat{\mu}(t_{ij}) - \sum_{k=1}^{J} \hat{\alpha}_{ik} \widehat{\psi}_k(t_{ij})}{\sigma} \right),$$

where $\sigma = \text{WMED}|e_{ij} - \text{WMED}(e_{ij})|$ and $e_{ij} = y_{ij} - \widehat{\mu}(t_{ij}) - \sum_{k=1}^{J} \hat{\alpha}_{ik} \widehat{\psi}_k(t_{ij})$. Subsequently, we repeat the above procedure, leaving onefold for testing, and sum up the cross validation score across all the folds as the final criterion.

Moreover, in terms of the choice of the number of FPCs, K, we define the relative proportion of variance explained as

$$\sum_{k=1}^{K-1} \widehat{\text{Var}}(\alpha_{ik}) \Big/ \sum_{k=1}^{K} \widehat{\text{Var}}(\alpha_{ik}) = 1 - \widehat{\text{Var}}(\alpha_{iK}) \Big/ \sum_{k=1}^{K} \widehat{\text{Var}}(\alpha_{ik}),$$

where $\widehat{\text{Var}}(\alpha_{ik})$ is the sample variance of the obtained estimates of α_{ik} for $k = 1, \ldots, K$. We set a threshold, e.g., 90%, for this proportion. Once the threshold is reached, no further extraction of the FPCs is needed.

To detect any potential outliers, for the ith observed trajectory, we define the following mean squared residual,

$$r_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \left(y_{ij} - \widehat{\mu}(t_{ij}) - \sum_{k=1}^{K} \alpha_{ik} \widehat{\psi}_k(t_{ij}) \right)^2.$$

Following the approach of Hubert and Vandervieren (2008) and Boente and Salibian-Barrera (2015), we first construct a skewed–adjusted boxplot of all the r_i. The trajectory i is defined as an outlier if its mean squared residual r_i exceeds the upper whisker of the skewed–adjusted boxplot.

3. SIMULATION STUDIES

3.1. SIMULATION SETUP

We perform simulation studies to assess the empirical performance of the proposed method. The goal is to compare the accuracy of estimating the FPCs using our method to that using the alternative approaches under two cases. We consider the case where the data points on the longitudinal trajectories are contaminated, i.e., subject to an additional noise with nonzero mean. In these contaminated cases, a proportion of the data points on the trajectory would be subject to contamination. In such a case, it is expected that the robust FPCA method can safeguard the estimation accuracy against the influence of the outliers or contamination.

We first set out the underlying stochastic process for simulating the trajectories. The true trajectories are simulated as the sum of a common mean function plus sum products of two component functions and FPC scores.

$$X_i(t) = \mu(t) + \alpha_{i1}\psi_1(t) + \alpha_{i2}\psi_2(t) + \alpha_{i3}\psi_3(t).$$

The mean function is

$$\mu(t) = 0.5 + \sin(6\pi t)\exp(-2t),$$

and the component functions are

$$\begin{aligned}\psi_1(t) &= \sqrt{2}\sin(2\pi t)\\ \psi_2(t) &= \sqrt{2}\cos(2\pi t)\\ \psi_3(t) &= \sqrt{2}\sin(4\pi t).\end{aligned}$$

The scores α_{i1}, α_{i2} and α_{i3} are independently sampled from the normal distributions with mean 0 and decreasing standard deviations of 9, 4 and 1, respectively, i.e., $\alpha_{i1} \sim \mathrm{N}(0, 9^2)$, $\alpha_{i2} \sim \mathrm{N}(0, 4^2)$, and $\alpha_{i3} \sim \mathrm{N}(0, 1^2)$. This is in accordance with the Mercer's theorem where the variance of the FPC scores decreases with increasing order of the FPC.

We apply the proposed robust FPCA method to the simulated trajectories. Tukey's biweight loss function is adopted as the objective function. Moreover, we use 20 B-spline basis functions to model the mean function and the FPCs. We choose the number of FPCs by setting the threshold of the relative proportion of variance to be 0.9. Our study reveals that this criterion correctly chooses the number of FPCs in over 95% simulation replicates.

Let n_c denote the number of curves in a data replication. Given the trajectory $X_i(t)$, we simulate data point from it as $y_{ij} = X_i(t_{ij}) + \epsilon_{ij}$, $i = 1, \ldots, n_c$ and $j = 1, \ldots, n_i$,

where ϵ_{ij} is a measurement error term randomly drawn from a normal distribution with mean 0 and standard deviation σ_e. The observed data points are simulated on a uniform grid $\mathcal{T} = [0, 1]$. The spacing between each design point is 0.01. For the contaminated case, we add an additional layer of noise to the uncontaminated $X_i(t_{ij})$. Let $X_i^{(c)}(t)$ denote the contaminated trajectory. We consider a complex setup as in Boente and Salibian-Barrera (2015). The contaminated trajectories are simulated as

$$\begin{aligned} X_i^{(c)}(t_{ij}) &= X_i(t_{ij}) + C \cdot V(t_{ij}), \\ C &\sim \text{Bernoulli}(\tau_c), \\ V(t_{ij}) &= B_{ij} \cdot D_{ij}, \\ B_{ij} &\sim \text{Bernoulli}(\tau_p), \\ D_{ij} &\sim \text{N}(\mu^{(c)}, \sigma_c^2) \end{aligned}$$

where C is the indicator that the curve is subject to contamination, and is independently generated from a Bernoulli distribution with probability τ_c, $V(t_{ij})$ is the amount of contamination at timepoint t_{ij}, which is contaminated with a probability τ_p and $\mu^{(c)}$ represents a directional shift resulted from the contamination. In other word, C represents the indicator of trajectory-level contamination and the probability of a curve being contaminated is τ_c, and if a curve is contaminated, $V(t_{ij})$ represents the degree of contamination, where each point on the contaminated curve is subject to a shift occurring with probability τ_p. The distance of the shift is random but tightly centered around the mean $\mu^{(c)}$.

3.2. MODEL EVALUATION

To measure how well the proposed method estimates the true FPCs, we use the integrated mean squared error (IMSE), defined as the integral of the squared difference between the estimated functional principal component function $\widehat{\psi}_k$ and the true one ψ_k,

$$\text{IMSE}(\widehat{\psi}_k) = \int_{\mathcal{T}} \left\{ \widehat{\psi}_k(t) - \psi_k(t) \right\}^2 \text{d}t, \qquad k = 1, 2, 3,$$

where $\mathcal{T} = [1, 10]$. Moreover, to assess the estimation accuracy of the FPC scores, we use the mean squared error (MSE), defined as the mean of the squared difference between the estimated kth FPC score $\hat{\alpha}_{ik}$ and the true one α_{ik},

$$\text{MSE}(\hat{\alpha}_{ik}) = \frac{1}{n_c} \sum_{i=1}^{n_c} (\hat{\alpha}_{ik} - \alpha_{ik})^2, \qquad k = 1, 2, 3.$$

In terms of the running time, as our method is primarily built upon regression, the required computational time is in fact relatively acceptable. For example, on a machine with an Intel 4-core i5-5200 CPU, using the proposed method to analyze 200 sparse curves each with 25 time points, the total running time required for the first 3 FPCs is around 30 seconds. On the other hand, the running time required to run the PACE method on the same problem is around 40 seconds. The running time scales in a roughly

linear relationship with the number of curves and the number of observations on each curve. In comparison with the higher order FPCs, the first FPC usually requires less time for the algorithm to reach the desired level of convergence. The time required for the higher order FPCs tends to be longer, but the relationship between the running time and the order of FPC is not exactly monotone. For example, the times required for obtaining each of the first 4 FPCs are 4.9, 12.6, 11.4, and 7.9 seconds, respectively.

3.3. VARYING SAMPLE SIZE

To study the asymptotical behaviour of the proposed method, we assess its estimation precision under the increasing value of n_c, the number of curves in data replication. As a default, the observed trajectories are subject to data contamination with parameter $\tau_c = 0.3$, $\tau_p = 0.4$ and $\mu^{(c)} = 10$. The standard deviation of the measurement error term is set as $\sigma_e = 0.01$. We experiment with cases with the number of curves $n_c = 1000$, 5000 and 10,000.

We compare the proposed approach with the conventional PACE method (Yao et al. 2005), in terms of the IMSE of the FPCs and MSE of the FPC scores. Table 1 shows the IMSE of the three FPCs and MSE of the three FPC scores under the robust FPCA method and the PACE method with varying values of n_c, averaged across 500 data replications. Under increasing sample sizes, we observe that both the IMSE and MSE of the FPCs and FPC scores appear to become smaller and close to zero, indicating that the estimation accuracy improves over increasing sample sizes. On the other hand, the IMSE and MSE under the PACE method do not always decrease as the number of curves becomes larger. For example, under the PACE method, the IMSE of the third FPC and the MSE of the third FPC score becomes larger even under an increased sample size.

3.4. VARYING DEGREE OF SPARSITY

Moreover, as our method is based on a regression framework, it can naturally accommodate the sparse design setup. Under the sparse design, the PACE method is also capable of handling sparsely functional data, and thus a comparison with this method is conducted.

We first simulate the full series of data points from a uniform grid $\mathcal{T} = [0, 1]$ with a spacing of 0.01. Subsequently, only p_{sparse} of the points on the grid are randomly sampled from the full series of trajectory and treated as observed sample points, and the rest of the points are regarded as missing. For other simulation parameters, we set $n_c = 200$, $\tau_c = \tau_p = 0.3$, $\mu^{(c)} = 10$, and $\sigma_e = 0.01$.

Table 2 summarizes the results for various values of p_{sparse} under the sparse design. It is evident that under the sparse setting, as the value of p_{sparse} increases, i.e., more data points are observed, the estimation accuracy improves for both the PACE method and the robust FPCA method. While both methods are not completely immune to the influence of data contamination, the IMSE of the robust FPCA method is almost consistently smaller than that under the PACE method, especially for the higher order FPCs and FPC scores.

Table 1. Comparison of integrated mean squared error of the FPCs and mean squared error of the FPC scores under the robust FPCA method and the PACE method with varying values of n_c, the number of curves in one data replication

Method	IMSE($\hat{\psi}_1$)	IMSE($\hat{\psi}_2$)	IMSE($\hat{\psi}_3$)	MSE($\hat{\alpha}_{i1}$)	MSE($\hat{\alpha}_{i2}$)	MSE($\hat{\alpha}_{i3}$)
$n_c = 1000$						
PACE	0.0004	0.0006	1.9158	0.1595	0.1047	2.8101
	(0.0005)	(0.0005)	(0.0636)	(0.1357)	(0.0410)	(0.1364)
Robust FPCA	0.0004	0.0005	0.0349	0.1113	0.0556	0.0544
	(0.0005)	(0.0005)	(0.0830)	(0.1497)	(0.0425)	(0.1110)
$n_c = 5000$						
PACE	0.0001	0.0001	1.9644	0.0782	0.0700	2.8946
	(0.0001)	(0.0001)	(0.0270)	(0.0223)	(0.0091)	(0.0628)
Robust FPCA	0.0001	0.0001	0.0017	0.0217	0.0136	0.0079
	(0.0001)	(0.0001)	(0.0027)	(0.0226)	(0.0089)	(0.0046)
$n_c = 10000$						
PACE	0.0001	0.0001	1.9761	0.0694	0.0664	2.9128
	(0.0001)	(0.0001)	(0.0196)	(0.0085)	(0.0053)	(0.0446)
Robust FPCA	0.00005	0.00005	0.0008	0.0130	0.0105	0.0063
	(0.0001)	(0.0001)	(0.0011)	(0.0108)	(0.0055)	(0.0022)

The values are averaged across 500 data replications, with standard errors shown in parentheses

Table 2. Comparison of integrated mean squared error of the FPCs and mean squared error of the FPC scores under the robust FPCA method and the PACE method with varying values of p_{sparse}, the percentage of sparse data points sampled from the full trajectory

Method	IMSE($\hat{\psi}_1$)	IMSE($\hat{\psi}_2$)	IMSE($\hat{\psi}_3$)	MSE($\hat{\alpha}_{i1}$)	MSE($\hat{\alpha}_{i2}$)	MSE($\hat{\alpha}_{i3}$)
$p_{\text{sparse}} = 0.2$						
PACE	0.0044	0.0147	1.4146	0.6875	0.5868	1.0001
	(0.0027)	(0.0079)	(0.4173)	(0.2752)	(0.1199)	(0.1885)
Robust FPCA	0.0033	0.0034	0.8725	0.7161	0.5554	1.0392
	(0.0022)	(0.0025)	(0.4973)	(0.3488)	(0.1474)	(0.2786)
$p_{\text{sparse}} = 0.3$						
PACE	0.0035	0.0111	1.5229	0.6276	0.4733	1.0974
	(0.0024)	(0.0054)	(0.3767)	(0.3039)	(0.1250)	(0.1808)
Robust FPCA	0.0025	0.0025	0.5423	0.6620	0.4518	0.7673
	(0.0021)	(0.0022)	(0.3653)	(0.3647)	(0.1460)	(0.2691)
$p_{\text{sparse}} = 0.4$						
PACE	0.0029	0.0079	1.6969	0.6029	0.4425	1.1378
	(0.0020)	(0.0043)	(0.2522)	(0.3343)	(0.1179)	(0.1254)
Robust FPCA	0.0022	0.0022	0.4467	0.6255	0.4325	0.6653
	(0.0019)	(0.0020)	(0.3299)	(0.3899)	(0.1404)	(0.2558)

The values are averaged across 500 data replications, with standard errors shown in parentheses

3.5. ADDITIONAL SIMULATION SCENARIOS

In addition to assessing and comparing the model performance under an increasing sample size, we also consider other sets of simulation scenarios. In each simulation set, we vary a specific simulation parameter and study its effects on the performance of the proposed method versus that of the PACE method. We experiment with varying values

of σ_e, the standard deviation of the measurement error. Moreover, we consider various contamination mechanisms by changing the probabilities of the curve-specific and point-specific contamination τ_c and τ_p, as well as the amount of contamination $\mu^{(c)}$. It is worth noting that by definition, the occurrence of the contamination should be relatively low. For example, if the proportion of occurrences exceeds 50%, it would be hard to distinguish the contaminated cases as opposed to simply assuming a larger variance for the measurement error. In the simulation studies, we consider the most common type of contamination while there are a few other types proposed in the literature. For example, Lee et al. (2013) examined a form of outlier that results from an overly large magnitude of the FPC scores; as our method is based upon M-regression for both FPC and FPC scores, it should also be able to handle such a type of outlier. Sawant et al. (2012) considered outliers that only occur in a certain portion of the curve, which our method should naturally be capable of handling. The additional results are summarized in the Supplementary Material.

4. APPLICATIONS

4.1. HAWAII OCEAN OXYGEN DATA

We consider a dataset obtained collected from the Hawaii Ocean Time-series Data Organization & Graphical System. This program has been in operation of extracting and collecting information of hydrographic, biological and chemical characteristics of the water column at a hydrostation in the subtropical North Pacific Ocean near Hawaii since 1988. The data that support the findings of this study are available at http://hahana.soest.hawaii.edu/hot/hot-dogs/cextraction.html. In this application, we use 11 years of data from January 1, 2008, to December 31, 2018. The variable of interest is the level of oxygen in the water (umol/kg) measured every 2 m between 0 and 200 m below the sea surface. The oxygen level can be treated as a function of depth below the sea surface, and we obtain 106 trajectories depicting the functional relationship between the oxygen level and depth from the dataset.

Each longitudinal trajectory in the dataset consists of 101 data points, with no missingness, i.e., the sampling scheme is dense. We assume that there is measurement error in the observation and both our method and the PACE method can be readily applied on dense data. To illustrate the effectiveness of the proposed robust FPCA, we compare the our method with the conventional approach via the principal components analysis through conditional expectation (PACE) method (Yao et al. 2005).

Setting the threshold for the relative proportion of the explained variance as 90%, the resulting number of selected FPCs is $K = 3$. Figure 1 shows mean function and the estimated first two FPCs of the oxygen versus depth trajectories under the two methods. The estimates of mean functions and the first FPC under the two approaches are similar, indicating that while noisy observation or extreme outliers may exist in dataset, their effect on the mean function and the first FPC is not significant. On the other hand, for the second and third FPCs, it is observed that the difference between the PACE method and the proposed method becomes more evident. For example, the estimate of the third FPC under the PACE method has a slightly deeper trough than the proposed robust FPCA method, possibly due to the influence of a few outliers.

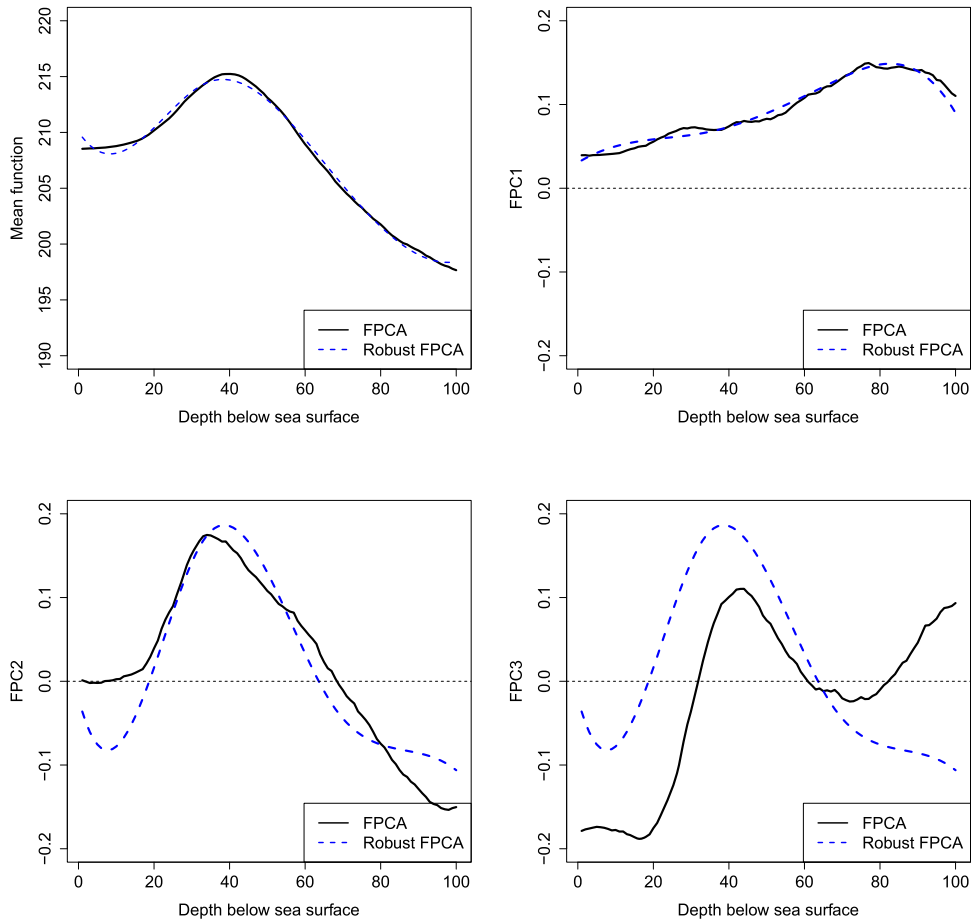

Figure 1. The estimated mean function and FPCs for the oxygen trajectories in the Hawaii ocean oxygen data.

The interpretation of the FPCs obtained with the robust FPCA method is as follows. The first FPC is constant above the zero axis, indicating the majority of the variation is the degree of horizontal shift from the mean. The second FPC is positive between 20 and 70 m and negative elsewhere, representing the difference in the oxygen level in this depth interval and the oxygen level elsewhere.

In addition, we apply the proposed approach of identifying the outliers using the skewed-adjusted boxplot of the mean squared residuals. The trajectories with mean squared residual exceeding the upper whisker of the skewed–adjusted boxplot are marked as outliers. Under this approach, two trajectories are flagged as abnormal, which is in accordance with the expectation that the number of outliers in the dataset is not high. In Fig. 2, the outlier trajectories in comparison with the rest of the trajectories are shown in black and dotted lines. It is evident that these two trajectories lie near the bottom compared with all the trajectories, one has an unusual curvature and the other has a sharp decline.

Furthermore, we compare the results of the proposed method with those under a conventional alternative. For such a time series of functions, the dynamic functional principal

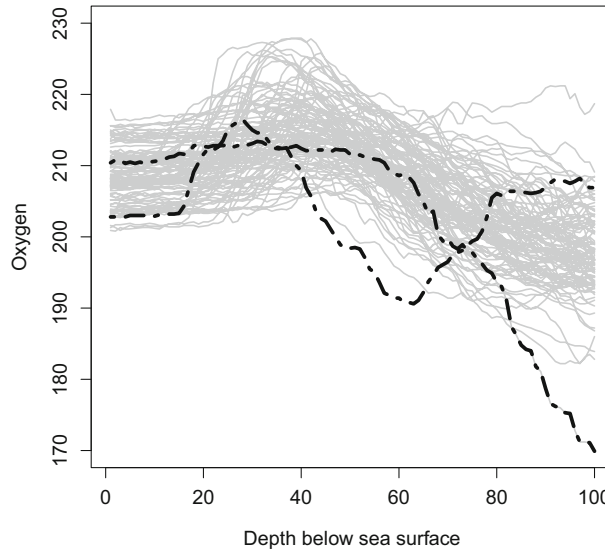

Figure 2. Detected outliers (black dotted) and the oxygen trajectories (gray) in the Hawaii ocean oxygen data .

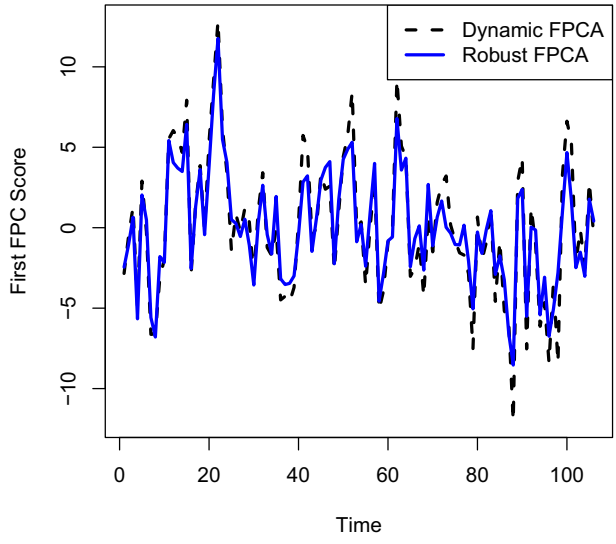

Figure 3. Comparison of the first FPC scores under the robust FPCA and the dynamic FPCA in the Hawaii ocean oxygen data .

component analysis is appropriate for decomposing the functional time series into components that are uncorrelated in time (Hormann et al. 2015). Figure 3 shows the first FPC scores obtained from the proposed robust FPCA method and from the dynamic FPCA method. We observe that the resultant FPC scores are similar in magnitude but the proposed robust method produces FPC scores that are relatively less extreme. In our future research, we will develop a robust FPCA method for the functional time series.

4.2. KIDNEY GLOMERULAR FILTRATION RATE DATA

The glomerular filtration rate (GFR) is a strong indicator of the kidney health state and risk of kidney failure, particularly for post-transplant patients. Various research has recommended GFR as a monitoring target for post kidney transplant patients. For example, the GFR of a healthy patient usually stays above 90; when the GFR falls below 60 for consecutive three months, a patient will be diagnosed as having chronic kidney disease. Kidney failure is usually characterized with a GFR value below 10. Within the dataset, there is a varying degree of variation in the patients' trajectories: a healthy subgroup of patients may have a relatively stable trajectory of GFR measurements above 90 whereas another subgroup may have downward sloping GFR curves that sharply drop below 10.

The data that support the findings of this study are available from the Organ Procurement Transplant Network/United Network for Organ Sharing (OPTN/UNOS) at https://optn.transplant.hrsa.gov/. Restrictions apply to the availability of these data, which were used under license for this study. The dataset consists of GFR curves of 252 patients; each curve contains the measurements of a patient's GFR in the follow-up visit. Certain proportion of the curve could be missing due to missed follow-ups. After receiving kidney transplant, the patients are followed up yearly for a period of 10 years in each follow-up visit, measurements of their GFR are recorded. The shape of the GFR curves is typically used as an indicator of a patient's kidney health and is used for evaluating the risk of kidney transplant failure (Salvadori et al. 2006).

Each longitudinal trajectory in the dataset usually consists of 10 data points, with possibly missing points on the trajectory. Our method is regression-based and thus is able to naturally accommodate the missing or even sparse longitudinal data. A comparison is made between the our method and the conventional approach via the PACE method. The PACE method also works for sparse and partially-missing longitudinal data, as it pools all the data points together for fitting the mean and covariance function via local smoother.

We first discuss the results when both methods are applied on the original dataset. We set the threshold for the relative proportion of the explained variance to be 90% and the resulting number of selected FPCs is $K = 3$. Figure S.1 in the Supplementary Material shows the estimated mean function and FPCs of the GFR trajectories under the two methods. The two approaches result in similar estimated mean function. Moreover, we observe that under the two methods, the first FPC is almost flat above the zero axis, indicating that the main level of variation is the horizontal shift above or below the mean function. The second FPC crosses the zero axis once. The cross points to the zero axis under the two methods are very close. In general, the second FPC is negative before the 6th year and positive afterwards, representing a change in GFR curves from the 6th year onwards. The third FPC crosses the zero axis twice; both the proposed robust FPCA and the PACE methods very close change points: the negative interval is [3.9, 8.8] and the positive intervals are [1.0,3.9] and [8.8,10], which can be interpreted as the difference in GFR values during [3.9, 8.8] and those in the other two time intervals.

Furthermore, to illustrate the effectiveness of our method against the outliers, we consider a contaminated data scenario, where a disturbance of 1000 is added to a few randomly selected curves. Through introducing such outliers, we compare the difference between

the FPC estimates under our approach and the ones obtained from the PACE method. As shown in Figure S.2 in the Supplementary Material, the outliers severely affect the mean function estimate under the PACE method, significantly shifting it upward, whereas the estimate under the robust FPCA method remains stable. This is as expected because the upward disturbance contributes significantly to a biased estimation of the mean function in the conventional approach, but is downweighted in the proposed robust framework of FPCA.

Moreover, Figure S.2 shows the estimated FPCs under the two approaches. We observe that while introducing the outliers does not seem to lead to much difference in the first FPC under the two methods, the estimate of the second and third FPC are highly distorted under the PACE method. The reason that the first FPC is undistorted can be explained as follows. The first FPC represents a relatively horizontal shift from the mean function, i.e., it depicts how far the trajectory is from the mean. The externally-introjected disturbance is essentially a uniform horizontal shift in the trajectory, which can thus be explained by the first FPC. In contrast, the second and third FPCs are severely affected by the contamination: the curves is wiggly and cross the zero axis multiple times, no longer having any meaningful interpretations as the ones under the uncontaminated scenario. It is evident that the conventional FPCA method is unable to safeguard the estimation against the influence of the outliers. In summary, the proposed method is evidently able to render stable and robust estimate of the FPCs in the presence of outliers.

5. DISCUSSION

We have developed a new method for conducting FPCA when the longitudinal data are subject to outliers and contamination. Our method is based upon a robust orthogonal approximation to the residual from the Karhunen–Lovève expansion, and has the natural capability to handle sparse and irregularly-sampled data. We have shown through simulation studies that our method can effectively identify the major sources of variability in the stochastic process and is robust to noises and outliers. It is worth mentioning that throughout the paper, we adopt the M-estimation of the functions represented by the spline basis expansion. While the asymptotical efficiency of M-estimator in linear regression has been well studied (Huber and Ronchetti 2009), further theoretical studies are needed to establish the asymptotical properties of robust M-estimation in the case of functional data.

Boente and Salibian-Barrera (2020) proposed a different method for robust functional principal component analysis. Their method is also based on the minimization of the objective function $\rho(\cdot)$. In the first step of their method, i.e., the estimation of the mean function, they adopted a robust version of local polynomial regression. Rather than the squared loss, they use $\rho(\cdot)$ as the loss function, which is known as the local M-regression. By contrast, our method directly estimates the mean function as a linear combination of B-spline basis functions by applying the robust M-regression. Subsequent steps of their method involve a robust local M-regression to estimate the covariance function. The FPCs are then estimated as the eigenfunctions of the covariance function. In comparison, our method do not need

to estimate the covariance function. We estimate FPCs directly as a linear combination of B-spline basis functions based upon the robust M-regression.

As our method is based upon M-estimation, which has a breakdown point of zero, it is not completely immune to outliers. A further simulation comparison (not shown here) indicates that under a very large proportion of contamination (e.g., 0.4) and large values of the mean directional shift in the contaminated cases (e.g., 50), the proposed method could have a large estimation error in the higher order FPCs and FPC scores. Several extensions of our method can be developed as directions for future research. For example, in each regression steps of the algorithm, we may substitute the M-estimator with the MM-estimator or the S-estimator as an alternative. The S-estimator seeks to minimize a robust estimate of the scale of the residual; it has smaller asymptotical bias and variance and is thus more robust than the M-estimator under data contamination (Rousseeuw and Leroy 2005; Pitselis 2013). The MM-estimator can be regarded as the M-estimator with Tukey's biweight loss function and initialization by a preliminary S-estimator, which leads to a higher breakdown point with better asymptotical efficiency (Yohai 1987; Salibian-Barrera 2006). Adapting the S- and MM-estimators and the corresponding iteratively reweighted least square algorithm into the alternating regression steps in our proposed method could be a worthy direction for future research. Furthermore, another direction of future research is to allow the tuning constant in Tukey's biweight loss function to be varying with t, such that it becomes a functional threshold value within the loss function. A data-driven strategy can be used to estimate the optimal threshold curve. Moreover, when the smoothness of the FPC is of interest, it might be possible to add a roughness penalty term $\lambda \int \{d^2\psi_k(t)/dt^2\}^2 \mathrm{d}t$ to the objective functions in order to achieve a smoothed FPC estimate.

ACKNOWLEDGEMENTS

The authors would like to thank the Editor, the Associate Editor, and two reviewers for their valuable comments, which are very helpful to improve this work. This work was supported by the discovery grants from the Natural Sciences and Engineering Research Council of Canada (NSERC) to J. Cao and H. Shi.

Funding Funding was provided by Canadian Network for Research and Innovation in Machining Technology, Natural Sciences and Engineering Research Council of Canada (Grand No. RGPIN-2018-06008)

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

[Received November 2020. Revised September 2021. Accepted September 2021. Published Online March 2022.]

REFERENCES

Bali JL, Boente G, Tyler DE, Wang J-L (2011) Robust functional principal components: a projection-pursuit approach. Ann Stat 39:2852–2882

Boente G, Salibian-Barrera M (2015) S-estimators for functional principal component analysis. J Am Stat Assoc 110:1100–1111

Boente G, Salibian-Barrera M (2020) Robust functional principal components for sparse longitudinal data. arXiv:2012.01540 [stat.ME]

Billor N, Hadi AS, Velleman PF (2000) BACON: blocked adaptive computationally efficient outlier nominators. Comput Stat Data Anal 34:279–298

Dauxois J, Pousse A, Romain Y (1982) Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. J Multivar Anal 12:136–154

Dong J, Wang L, Gill J, Cao J (2018) Functional principal component analysis of GFR curves after kidney transplant. Stat Methods Med Res 27:3785–3796

Ferraty F, Vieu P (2006) Nonparametric functional data analysis: methods, theory, applications and implementations. Springer-Verlag, London

Gervini D (2008) Robust functional estimation using the median and spherical principal components. Biometrika 95:587–600

Gervini D (2009) Detecting and handling outlying trajectories in irregularly sampled functional datasets. Ann Appl Stat 3:1758–1775

Hall P, Horowitz JL (2007) Methodology and convergence rates for functional linear regression. Ann Stat 35:70–91

Hormann S, Kidzinski L, Hallin M (2015) Dynamic functional principal components. J R Stat Soc Ser B (Stat Methodol) 77:319–348

Huber PJ, Ronchetti EM (2009) Robust statistics, 2nd edn. Wiley, Hoboken

Hubert M, Rousseeuw PJ, Branden KV (2005) ROBPCA: a new approach to robust principal component analysis. Technometrics 47:64–79

Hubert M, Vandervieren E (2008) An adjusted boxplot for skewed distributions. Comput Stat Data Anal 52:5186–5201

Hyndman RJ, Ullah S (2007) Robust forecasting of mortality and fertility rates: a functional data approach. Comput Stat Data Anal 51:4942–4956

Lee S, Shin H, Billor N (2013) M-type smoothing spline estimators for principal functions. Comput Stat Data Anal 66:89–100

Lin Z, Wang L, Cao J (2016) Interpretable functional principal component analysis. Biometrics 72:846–854

Mas A (2002) Weak convergence for the covariance operators of a Hilbertian linear process. Stoch Process their Appl 99:117–135

Nie Y, Cao J (2020) Sparse functional principal component analysis in a new regression framework. Comput Stat Data Anal 152:107016

Nie Y, Wang L, Liu B, Cao J (2018) Supervised functional principal component analysis. Stat Comput 28:713–723

Pitselis G (2013) A review on robust estimators applied to regression credibility. J Comput Appl Math 239:231–249

Ramsay JO, Dalzell C (1991) Some tools for functional data analysis. J R Stat Soc Ser B 53:539–572

Ramsay JO, Silverman BW (2005) Functional data analysis, 2nd edn. Springer-Verlag, New York

Rousseeuw PJ, Leroy AM (2005) Robust regression and outlier detection. Wiley, Hoboken

Salvadori M, Rosati A, Bock A, Chapman J, Dussol B, Fritsche L, Kliem V, Lebranchu Y, Oppenheimer F, Pohanka E, Tufveson G, Bertoni E (2006) Estimated one-year glomerular filtration rate is the best predictor of long-term graft function following renal transplant. Transplantation 81:202–206

Salibian-Barrera M (2006) The asymptotics of MM-estimators for linear regression with fixed designs. Metrika 63:283–294

Sang P, Wang L, Cao J (2017) Parametric functional principal component analysis. Biometrics 73:802–810

Sawant P, Billor N, Shin H (2012) Functional outlier detection with robust functional principal component analysis. Comput Stat 27:83–102

Shi H, Dong J, Wang L, Cao J (2021) Functional principal component analysis for longitudinal data with informative dropout. Stat Med 40:712–724

Silverman BW (1996) Smoothed functional principal components analysis by choice of norm. Ann Stat 24:1–24

Yao F, Müller HG, Wang JL (2005) Functional data analysis for sparse longitudinal data. J Am Stat Assoc 100:577–590

Yohai VJ (1987) High breakdown point and high efficiency robust estimates for regression. Ann Stat 20:642–656