

Article

# A Method of L1-Norm Principal Component Analysis for Functional Data

Fengmin Yu <sup>1,2</sup>, Liming Liu <sup>1,\*</sup>, Nanxiang Yu <sup>2</sup>, Lianghao Ji <sup>3</sup>  and Dong Qiu <sup>2</sup> 

<sup>1</sup> School of Statistics, Capital University of Economics and Business, Beijing 100071, China; yufm@cqupt.edu.cn

<sup>2</sup> School of Science, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; yunx@cqupt.edu.cn (N.Y.); qiudong@cqupt.edu.cn (D.Q.)

<sup>3</sup> School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China; jilh@cqupt.edu.cn

\* Correspondence: llm5609@163.com; Tel.: +86-159-1084-6465

Received: 15 December 2019; Accepted: 10 January 2020; Published: 20 January 2020



**Abstract:** Recently, with the popularization of intelligent terminals, research on intelligent big data has been paid more attention. Among these data, a kind of intelligent big data with functional characteristics, which is called functional data, has attracted attention. Functional data principal component analysis (FPCA), as an unsupervised machine learning method, plays a vital role in the analysis of functional data. FPCA is the primary step for functional data exploration, and the reliability of FPCA plays an important role in subsequent analysis. However, classical L2-norm functional data principal component analysis (L2-norm FPCA) is sensitive to outliers. Inspired by the multivariate data L1-norm principal component analysis methods, we propose an L1-norm functional data principal component analysis method (L1-norm FPCA). Because the proposed method utilizes L1-norm, the L1-norm FPCs are less sensitive to the outliers than L2-norm FPCs which are the characteristic functions of symmetric covariance operator. A corresponding algorithm for solving the L1-norm maximized optimization model is extended to functional data based on the idea of the multivariate data L1-norm principal component analysis method. Numerical experiments show that L1-norm FPCA proposed in this paper has a better robustness than L2-norm FPCA, and the reconstruction ability of the L1-norm principal component analysis to the original uncontaminated functional data is as good as that of the L2-norm principal component analysis.

**Keywords:** functional data; L1-norm; outliers; principal component analysis; robust

## 1. Introduction

In recent years, with the rapid popularization of intelligent terminals and sensors, massive data have been rapidly accumulated, and the processing technology of intelligent big data has attracted more and more attention. Among these data, kinds of intelligent big data with function characteristics, such as physiological indicator data, growth curve data, air quality data, and temperature data, has also attracted people's attention. In fact, these data are discrete samples of a continuous function, so such data are known in the literature as functional data [1–10]. The difference between functional data and traditional multivariate data is that the former regards the observed discrete data as a whole and as a realization of a random process. Therefore, the first step of statistical analysis is to fit the discrete data into smooth curves; this can solve the problems of missing data and inconsistent sampling intervals, which are difficult issues for multivariate data. Moreover, if the fitting curve is smooth enough, we can get more information from its derivatives, which is impossible for traditional multivariate data. As a nonparametric statistical method, functional data analysis is not limited by a model and its parameters,

so it can better reflect real laws in nature. At present, statistical analysis methods for functional data have been widely used in the fields of biology, medicine, economics and meteorology [11–18].

Functional data principal component analysis (FPCA), as an unsupervised machine learning method, plays a vital role in the analysis of functional data. The central idea of FPCA is to use a few orthogonal dimensions to express most of the information of the original functional data. Through dimensionality reduction, the analysis of the original functional data can be transformed into the analysis of the characteristic functions of a few dimensions, thus greatly reducing the complexity of the functional data and allowing for the better interpretation of the function data. Since J.Q. Ramsay proposed the idea of functional principal component analysis in 1991 [19], various pieces of research on functional principal component analysis have emerged one after another. Classical functional principal components are the characteristic functions of the symmetric empirical covariance operator [20]. As early as 1982, Pousse and Romain studied the asymptotic properties of the characteristic functions of the empirical covariance operator: the empirical functional principal components [21]. In order to avoid the violent oscillation of the obtained principal component weight function, Rice and Silverman (1991) proposed a smooth functional principal component estimation method that smoothed the principal component weight function by adding penalties to the variance after projection [22]. The consistency of the estimate of the smooth functional principal component was then confirmed by Pezzulli and Silverman (1993) [23]. Silverman (1996) proposed another method of smooth functional principal components. Unlike the methods of Rice and Silverman (1991), the new method achieved the smoothness of the principal component function by penalizing the norm of the projected variance [24]. Gareth (2000) studied principal component analysis for sparse function data [25]. Boente (2000) studied the functional principal components-based kernel [26] Hall (2006) studied the properties of functional principal components [27]. Benko (2009) studied common functional principal components [28], and Hormann (2015) studied dynamic functional principal components [29].

Functional data principal component analysis (FPCA) is an important research subject of machine learning and artificial intelligence, and it is the primary step for functional data exploration. Therefore, the reliability of FPCA plays an important role in subsequent analysis. The aforementioned principal component methods for functional data were established in L2-norm framework. However, because the L2-norm enlarges the influence of outliers, the traditional functional principal components analysis method is sensitive to outliers. On the other hand, in regard to multivariate data, relevant research of principal component analysis methods [30–37] has shown that the principal component analysis method of L1-norm for multivariate data has a better robustness than that of the L2-norm. In [30], Kwak (2008) proposed an L1-PCA optimization model based on L1-norm maximization for multivariable data, i.e.,  $W_{L1} = \underset{W \in R^{D \times K}, W^T W = I}{\operatorname{argmax}} \|W^T X\|_1$ . The algorithm in [30] gives an approximate solver for  $W_{L1} = \underset{W \in R^{D \times K}, W^T W = I}{\operatorname{argmax}} \|W^T X\|_1$  through a sequence of deflating nullspace projections with cost  $O(N^2 DM)$ , and it is robust to outliers and invariant to rotations. In [31], Nie et al. (2011) simultaneously approximated all M L1-PCs of X with complexity  $O(N^2 DM + NM^3)$ ; however, the principal components obtained by [31] were highly dependent on the the finding of the dimension M of a subspace. For example, the projection vector obtained when M = 1 may not be in a subspace obtained when M = 2. The optimal algorithm in [33] introduced a bit-flipping-based approximate solver for  $W_{L1} = \underset{W \in R^{D \times K}, W^T W = I}{\operatorname{argmax}} \|W^T X\|_1$  with complexity  $O(ND \min\{N, D\} + N^2(M^4 + dM^2) + NdM^3)$ , where  $d = \operatorname{rank}(X)$ ; this solution has a low performance degradation, and is close to L2-PCA, but the cost is that it is not as robust as that in [30]. The work in [32] offered an algorithm for exact calculation  $W_{L1} = \underset{W \in R^{D \times K}, W^T W = I}{\operatorname{argmax}} \|W^T X\|_1$  with complexity  $O(2^{NM})$ ; however, when X is big data of large N and/or large dimension D, the cost is prohibitive. The authors of [34] studied the relationship of independent component analysis (ICA) and L1-PCA, and they proved that independent component analysis (ICA) can be performed by L1-norm PCA under the assumption of whitening. The authors of [36] computed L1-PCA by an incremental algorithm, in which only one measurement was processed at a time, and the

changes in the nominal signal subspace could be tracked. Instead of maximizing the L1-norm deviation of the projected data, the authors of [35,37] focused on minimizing the L1-norm reconstruction error. However, in contrast to the conventional L2-PCA, the solutions of the minimization of the L1-norm reconstruction error might not be same as the solutions of the maximization of the L1-norm deviation of projected data.

Inspired by these pieces of research on L1-PCA for multivariable data, in this paper, we try to construct a robust L1-norm principal component analysis method for functional data (L1-norm FPCA). Firstly, we build a functional data L1-norm maximized principal component optimization model, and then a corresponding algorithm for solving the L1-norm maximized optimization model is extended to functional data based on the idea of a multivariate data L1-norm principal component analysis method [30]. Numerical experiments show that the L1-norm functional principal component analysis method provides a more robust estimation of principal components than the traditional L2-norm functional principal component analysis method (L2-norm FPCA). Finally, by comparing the reconstruction errors of the L1-norm FPCA and L2-norm FPCA, it is found that the reconstruction ability of the L1-norm principal components to the original uncontaminated functional data is as good as that of the L2-norm functional principal components.

## 2. Problem Description

### 2.1. L2-Norm Functional Principal Component Analysis (L2-Norm FPCA)

Suppose  $x_1(t), x_2(t), \dots, x_n(t), t \in \tau \subset R$  are implementations of the square integrable random process  $X(t) \in L_2(\tau)$ . Without a loss of generality, we assume that  $x_1(t), x_2(t), \dots, x_n(t), t \in \tau \subset R$  are centralized. The purpose of functional principal component analysis (FPCA) is to express as much information as possible of the original functional data with as few dimensions as possible. Firstly, the case of only one principal component is considered. At this point, the task of FPCA is to find a “projection direction” in infinite dimensional space so that the variance of projection of  $x_1(t), x_2(t), \dots, x_n(t), t \in \tau$  to that direction is maximum. Assuming that the projection direction is  $\xi_1(t)$ , which is called the first functional principal component weight function of functional data  $x_1(t), x_2(t), \dots, x_n(t), t \in \tau$ , then  $\xi_1(t)$  should be the solution of the following optimization problem:

$$\begin{aligned} \max_{\xi_1(t)} \frac{1}{n} \sum_{i=1}^n \left( \int \xi_1(t) x_i(t) dt \right)^2 \\ \text{s.t.} \int \xi_1^2(t) dt = 1 \end{aligned} \quad (1)$$

If the information that is expressed by one principal component is insufficient, a second projection direction  $\xi_2(t)$ , which is orthogonal to the first principal component direction  $\xi_1(t)$  and maximizes the variance of the functional data  $x_1(t), x_2(t), \dots, x_n(t), t \in \tau$  under the orthogonality condition, is necessary. This is the second functional principal component weight function. And so on, this process continues until the obtained principal components can express enough information. Therefore, the subsequent principal component weight functions need to satisfy the following optimization model:

$$\begin{aligned} \max_{\xi_j(t)} \frac{1}{n} \sum_{i=1}^n \left( \int \xi_j(t) x_i(t) dt \right)^2 \\ \text{s.t.} \int \xi_j^2(t) dt = 1, j = 2, 3, \dots, m \\ \int \xi_j(t) \xi_k(t) dt = 0, k = 1, 2, \dots, j-1 \end{aligned} \quad (2)$$

J.Q. Ramsay proved that the principal component weight functions  $\xi_1(t), \xi_2(t), \dots, \xi_m(t)$  of functional data  $x_1(t), x_2(t), \dots, x_n(t), t \in \tau$  are the eigenfunctions that correspond to the first  $m$  largest eigenvalues of sample covariance function of functional data  $x_1(t), x_2(t), \dots, x_n(t), t \in \tau$ , i.e.,

$\int \hat{C}(t,s)\xi_i(t)dt = \rho_i \xi_i(s), i = 1, 2, \dots, m$ , where  $\rho_1 \geq \rho_2 \geq \dots \geq \rho_m$  are the eigenvalues of the covariance function  $\hat{C}(t,s)$ .

From the optimization Formulas (1) and (2), it is easy to find that the above L2-norm functional principal components enlarge the influence of outliers and are sensitive to outliers. Therefore, L1-norm functional principal components are constructed in this paper. Compared with the traditional L2-norm, the L1-norm weakens the influence of outliers. It can be expected that the L1-norm functional principal components have a good anti-noise ability.

### 2.2. L1-Norm Functional Principal Component Analysis (L1-Norm FPCA)

Suppose  $x_1(t), x_2(t), \dots, x_n(t)$  are the implementations of square integrable stochastic process  $x(\cdot)$ . Without a loss of generality, suppose that  $x_1(t), x_2(t), \dots, x_n(t)$  have been centralized. Now we want to find an m-dimensional linear subspace so that the L1-norm dispersion of the projection of  $x_1(t), x_2(t), \dots, x_n(t)$  in this subspace is the largest. Assume that the subspace is spanned by  $\beta_1(t), \beta_2(t), \dots, \beta_m(t)$ , and the optimization problem corresponding to Formulas (1) and (2) can be obtained:

$$\begin{aligned} \max_{\beta_1, \beta_2, \dots, \beta_m} \sum_{i=1}^n \sum_{j=1}^m \left| \int \beta_j(t)x_i(t)dt \right| \tag{3} \\ \text{s.t. } \int \beta_j^2(t)dt = 1, j = 1, 2, \dots, m \\ \int \beta_j(t)\beta_k(t)dt = 0, j, k = 1, 2, \dots, m; j \neq k \end{aligned}$$

where  $\beta_1(t), \beta_2(t), \dots, \beta_m(t)$  are called L1-norm principal component weight functions for  $x_1(t), x_2(t), \dots, x_n(t)$ .

It is not easy to solve the Optimization Problem (3) because the objective function is non-differentiable, non-convex, and contains an absolute value operation. Next, we try to find the solution of Optimization Problem (3) from the perspective of orthogonal basis expansion.

Assuming that  $x_1(t), x_2(t), \dots, x_n(t); \beta_1(t), \beta_2(t), \dots, \beta_m(t), t \in \tau$  can be linearly represented by the same standard orthogonal basis functions  $\phi_1(t), \phi_2(t), \dots$  with the same number of basis functions  $K$ , i.e.,

$$\begin{aligned} x_i(t) &= \sum_{v=1}^K c_{iv}\phi_v(t) = c_i^T \phi(t) \quad i = 1, 2, \dots, n \\ \beta_j(t) &= \sum_{v=1}^K b_{jv}\phi_v(t) = b_j^T \phi(t) \quad j = 1, 2, \dots, m \end{aligned}$$

where  $K$  is a positive integer.  $c_i = (c_{i1}, c_{i2}, \dots, c_{iK})^T; b_i = (b_{j1}, b_{j2}, \dots, b_{jK})^T; \phi(t) = (\phi_1(t), \phi_2(t), \dots, \phi_K(t))^T$ ;

Under the above assumptions, we get:

$$\max_{\beta_1, \beta_2, \dots, \beta_m} \sum_{i=1}^n \sum_{j=1}^m \left| \int \beta_j(t)x_i(t)dt \right| = \max_{b_1, b_2, \dots, b_m} \sum_{i=1}^n \sum_{j=1}^m \left| \int b_j^T \phi(t)\phi^T(t)c_i dt \right| = \max_{b_1, b_2, \dots, b_m} \sum_{i=1}^n \sum_{j=1}^m |b_j^T c_i|$$

Since  $\int \beta_j^2(t)dt = \int b_j^T \phi(t)\phi^T(t)b_j dt = b_j^T b_j$ , the constraints  $\int \beta_j^2(t)dt = 1, j = 1, 2, \dots, m$  can be expressed as  $b_j^T b_j = 1, j = 1, 2, \dots, m$ , and the constrains  $\int \beta_j(t)\beta_k(t)dt = 0, j, k = 1, 2, \dots, m; j \neq k$  can be expressed as  $b_j^T b_k = 0, j, k = 1, 2, \dots, m; j \neq k$ . Therefore, the Optimization Problem (3) can be transformed into the following Optimization Problem (4):

$$\max_{b_1, b_2, \dots, b_m} \sum_{i=1}^n \sum_{j=1}^m |b_j^T c_i| = \max_{b_1, b_2, \dots, b_m} \sum_{i=1}^n \sum_{j=1}^m \left| \sum_{k=1}^K b_{jk}c_{ik} \right| \tag{4}$$

$$\begin{aligned} \text{s.t. } & b_j^T b_j = 1, j = 1, 2, \dots, m \\ & b_j^T b_k = 0, k = 1, 2, \dots, m; j \neq k \end{aligned}$$

If we can get the solution of Optimization Problem (4), according to  $\beta_j(t) = b_j^T \phi(t)$ ,  $j = 1, 2, \dots, m$ , we can get the solution of Optimization Problem (3). There are several algorithms to solve Optimization Problem (4), such as those in [30,31,33], each of which has its own advantages. According to the goal of building robust principal components for functional data, we finally choose the algorithm in [30], because the principal components calculated in [30] are more robust to outliers, and this algorithm is relatively low-complexity when the data number, data dimension, and the principal components number are large.

Next, based on the orthogonal basis expansion of functional data, we employ the L1-norm PCA algorithm of multivariate data [30] to get the solving algorithm of the L1-norm functional principal component weight functions (Abbreviation: L1-FPCA algorithm). The algorithm is rewritten in the next section.

### 3. The Solving Algorithm of L1-Norm Functional Principal Component Weight Functions (L1-FPCA Algorithm)

#### 3.1. Only One Principal Component

First, we discuss the case where there is only one principal component, namely  $m = 1$ . In this case, the Optimization Problems (3) and (4) are, respectively, simplified as follows:

$$\max_{\beta(t)} \sum_{i=1}^n \left| \int \beta(t) x_i(t) dt \right| \quad (5)$$

$$\text{s.t. } \int \beta^2(t) dt = 1$$

and

$$\max_b \sum_{i=1}^n |b^T c_i| \quad (6)$$

$$\text{s.t. } b^T b = 1$$

Next, we construct L1-FPCA algorithm to solve the Optimization Problems (5) and (6).

L1-FPCA Algorithm:

**Step 1:** Arbitrarily choose the initial projection direction  $\beta^0(t)$ , get  $b^0$  by  $\beta^0(t) = (b^0)^T \phi(t)$ , normalize  $b^0: b^0 = \frac{b^0}{\|b^0\|_2}$ , and set the iteration number  $k$  to be 0.

**Step 2:** For all  $i \in (1, 2, \dots, n)$ , if  $\int \beta^k(t) x_i(t) dt < 0$ , i.e.,  $(b^k)^T c_i < 0$ , let  $p_i^k = -1$ ; otherwise  $p_i^k = 1$ .

**Step 3:** Let  $b^k = \sum_{i=1}^n p_i^{k-1} c_i$ , normalize  $b^k: b^k = \frac{b^k}{\|b^k\|_2}$ , and get the corresponding  $\beta^k(t)$  by  $\beta^k(t) = (b^k)^T \phi(t)$ .

**Step 4:** If  $\beta^k(t) \neq \beta^{k-1}(t)$ , return to step 2. If there is  $i$  such that  $\int \beta^k(t) x_i(t) dt = 0$ , i.e.,  $(b^k)^T c_i = 0$ , then let  $b^k = \frac{(b^k + \Delta b)}{\|b^k + \Delta b\|_2}$  and get the corresponding  $\beta^k(t)$ , then return to step 2, where  $\Delta b$  is a small non-zero vector. Otherwise, let  $\beta^*(t) = \beta^k(t)$ ,  $b^* = b^k$  and  $\beta^*(t) = (b^*)^T \phi(t)$ , stop.

**Theorem 1.** The L1-FPCA algorithm is convergent, and its convergence point  $b^*$  is the local maximum point of the Optimization Problem (6) and  $\beta^*(t)$  is the local maximum point of Optimization Problem (5).

**Proof.** First, we prove that the objective function  $\sum_{i=1}^n \left| \int \beta^k(t)x_i(t)dt \right|$  and  $\sum_{i=1}^n |b^T c_i|$  are nondecreasing in the iteration process of L1-FPCA algorithm, i.e.,

$$\begin{aligned} \sum_{i=1}^n \left| \int \beta^k(t)x_i(t)dt \right| &= \sum_{i=1}^n \left| (b^k)^T c_i \right| = \sum_{i=1}^n (b^k)^T p_i^k c_i = (b^k)^T \sum_{i=1}^n p_i^k c_i \geq (b^k)^T \sum_{i=1}^n p_i^{k-1} c_i \\ &\geq (b^{k-1})^T \sum_{i=1}^n p_i^{k-1} c_i = \sum_{i=1}^n (b^{k-1})^T p_i^{k-1} c_i = \sum_{i=1}^n \left| (b^{k-1})^T c_i \right| \\ &= \sum_{i=1}^n \left| \int \beta^{k-1}(t)x_i(t)dt \right| \end{aligned}$$

Therefore, the objective function  $\sum_{i=1}^n \left| \int \beta^k(t)x_i(t)dt \right|$  and  $\sum_{i=1}^n |b^T c_i|$  are nondecreasing. Additionally, because there are only finite number of data points, the convergence points  $\beta^*(t)$  and  $b^*$  of the L1-FPCA algorithm exist.

Next, we prove that  $b^*$  and  $\beta^*(t)$  are the local maxima of the corresponding optimization problem.

Suppose that  $b^* = b^k$ , that is the convergence point  $b^*$  is found after  $k$  iterations. Because for any  $i \in (1, 2, \dots, n)$ ,  $b^T p_i^k c_i \rightarrow (b^*)^T p_i^k c_i$ , there is a neighborhood  $N(b^*)$  of  $b^*$  so that for  $b \in N(b^*)$ ,  $b^T p_i^k c_i \geq 0$  and

$$\sum_{i=1}^n \left| \int \beta^*(t)x_i(t) \right| = \sum_{i=1}^n \left| (b^*)^T c_i \right| = \sum_{i=1}^n (b^*)^T p_i^k c_i = (b^*)^T \sum_{i=1}^n p_i^k c_i$$

Because  $b^*$  is the convergence point,  $b^*$  is parallel to  $\sum_{i=1}^n p_i^k c_i$ ; therefore,  $(b^*)^T \sum_{i=1}^n p_i^k c_i \geq b^T \sum_{i=1}^n p_i^k c_i = \sum_{i=1}^n |b^T c_i|$ , so for  $b \in N(b^*)$ ,  $\sum_{i=1}^n \left| \int \beta^*(t)x_i(t) \right| = \sum_{i=1}^n \left| (b^*)^T c_i \right| \geq \sum_{i=1}^n |b^T c_i| = \sum_{i=1}^n \left| \int \beta(t)x_i(t) \right|$ ; that is,  $b^*$  is the local maximum of  $\sum_{i=1}^n |b^T c_i|$  and  $\beta^*(t)$  is the local maximum of  $\sum_{i=1}^n \left| \int \beta(t)x_i(t)dt \right|$ .

Therefore, the L1-FPCA procedure finds a local maximum point  $b^*$  of  $\sum_{i=1}^n |b^T c_i|$  and  $\beta^*(t)$  of  $\sum_{i=1}^n \left| \int \beta(t)x_i(t)dt \right|$ . □

Since the L1-FPCA algorithm obtains a local optimal solution, we expect to find the global optimal solution with great probability by appropriately setting the initial projection direction  $\beta^0(t)$ , e.g., by setting  $\beta^0(t) = \operatorname{argmax}_{x_i(t)} \int x_i^2(t)dt$  or by setting it to be the solution of L2-FPCA. In practice, we usually select several different initial projection directions  $\beta^0(t)$  and calculate the respective local optimal solutions, and the solution with maximized the objective function  $\sum_{i=1}^n \left| \int \beta(t)x_i(t)dt \right|$  is selected as the optimal solution.

### 3.2. Multiple Principal Components

Suppose that  $m$  principal components ( $m > 1$ ) are needed, and the L1-FPCA algorithm needs to sequentially find  $m$  principal component projection directions  $b_1, b_2, \dots, b_m$  and corresponding  $\beta_1(t), \beta_2(t), \dots, \beta_m(t)$ . The specific algorithm is as follows:

**Step 1:** Let  $\beta_0(t) = 0$ , i.e.,  $b_0 = 0, \{c_i^0 = c_i\}_{i=1}^n$ .

**Step 2:** For all  $i \in (1, 2, \dots, n)$ , let  $c_i^1 = c_i^0 - b_0(b_0^T c_i^0)$  and apply the L1-FPCA algorithm to  $c^1 = (c_1^1, c_2^1, \dots, c_n^1)$  to obtain the projection vector  $b_1$  and the corresponding  $\beta_1(t)$ .

**Step 3:** For all  $i \in (1, 2, \dots, n)$ , let  $c_i^j = c_i^{j-1} - b_{j-1}(b_{j-1}^T c_i^{j-1})$  and apply the L1-FPCA algorithm to  $c^j = (c_1^j, c_2^j, \dots, c_n^j)$  to obtain the projection vector  $b_j$  and the corresponding  $\beta_j(t)$ .

**Step 4:** Repeat Step 3 until  $m$  projection vectors  $b_1, b_2, \dots, b_m$  and corresponding  $\beta_1(t), \beta_2(t), \dots, \beta_m(t)$  are obtained.

Since  $b_1, b_2, \dots, b_m$  are standard orthogonal dimensions in  $R^K$  space [38], the principal component weight functions  $\beta_1(t), \beta_2(t), \dots, \beta_m(t) t \in \tau$  are also standard orthogonal dimensions because:

$$\langle \beta_j(t), \beta_k(t) \rangle = \int \beta_j(t) \beta_k(t) dt = \int b_j^T \phi(t) \phi^T(t) b_k dt = b_j^T b_k = 0, k = 1, 2, \dots, m; j \neq k$$

$$\langle \beta_j(t), \beta_j(t) \rangle = \int \beta_j(t) \beta_j(t) dt = \int b_j^T \phi(t) \phi^T(t) b_j dt = b_j^T b_j = 1, j = 1, 2, \dots, m$$

As with the L2-norm functional principal component analysis, it is necessary to consider how many principal components are appropriate. This problem needs to be determined by the cumulative variance contribution rate. That is, according to the variance of the  $j$  projection direction,  $v_j = \frac{1}{n} \sum_{i=1}^n (\int \beta_j(t) x_i(t) dt)^2, j = 1, 2, \dots, m$ , the total variance of the first  $k$  projection directions can be calculated as  $S(k) = \sum_{j=1}^k v_j$ , and the total variance of the original functional data is  $S = \frac{1}{n} \sum_{i=1}^n (\int x_i^2(t) dt)^2$ . Thus, for the actual problems, the number of final principal component weight functions can be determined when  $\frac{S(k)}{S}$  is more than 80% or 85%.

## 4. Numerical Examples

### 4.1. Simulation

In order to compare the robustness to outliers of L1-norm functional principal components (L1-FPCs) that are proposed in this paper and the classical L2-norm functional principal components (L2-FPCs), we performed this simulation. We referred to the simulation setting given by Fraiman and Muniz (2001) [38]. Here, we considered that functional data  $x_1(t), x_2(t), \dots, x_n(t)$  are the implementations of squared integrable stochastic process  $X(\cdot)$ , and the function curves were generated from different model. There was no contamination in Model 1, and several other models suffered from different types contamination based on Model 1.

**Model 1** (no contamination):  $x_i(t) = m(t) + \varepsilon_i(t), i = 1, 2, \dots, n$ , where error term  $\varepsilon_i(t)$  is a stochastic Gaussian process with zero mean and covariance function  $\text{cov}(s, t) = (1/2)(1/2)^{0.9|t-s|}$  and  $m(t) = 4t, t \in [0, 1]$ .

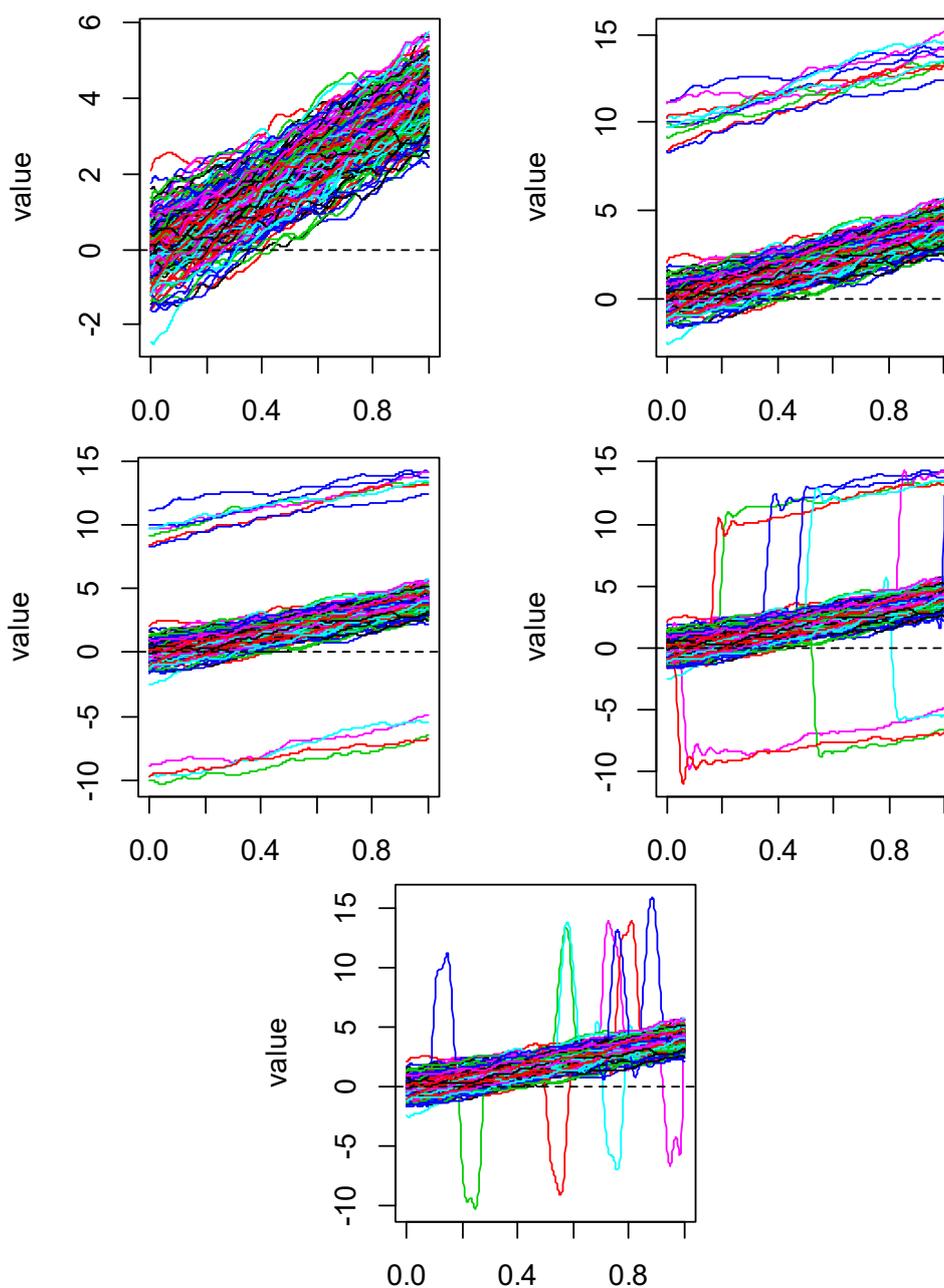
**Model 2** (asymmetric contamination):  $y_i(t) = x_i(t) + c_i M, i = 1, 2, \dots, n$ , where  $c_i$  is the sample of the 0–1 distribution with the parameter  $q$ , and  $M$  is the contamination constant.

**Model 3** (symmetric contamination):  $y_i(t) = x_i(t) + c_i \sigma_i M, i = 1, 2, \dots, n$ , where  $c_i$  and  $M$  are defined as in Model 2 and  $\sigma_i$  is a sequence of random variables with values of 1 and  $-1$  with a probability of 1/2 that is independent of  $c_i$ .

**Model 4** (partially contaminated):  $y_i(t) = \begin{cases} x_i(t) + c_i \sigma_i M, & t \geq T_i \\ x_i(t), & t < T_i \end{cases}, i = 1, 2, \dots, n$ , where  $T_i$  is a random number generated from a uniform distribution on  $[0, 1]$ .

**Model 5** (peak contamination):  $y_i(t) = \begin{cases} x_i(t) + c_i \sigma_i M, & T \leq t \leq T_i + l \\ x_i(t), & t \notin [T_i, T_i + l] \end{cases}, i = 1, 2, \dots, n$ , where  $l = 1/15$  and  $T_i$  is a random number generated from a uniform distribution on  $[0, 1 - l]$ .

Figure 1 shows the simulated curves of these five models. For each model, we set 100 equal-interval sampling points in  $[0, 1]$  and generated 200 replications. For Model 1, the parameter  $q$  was 0 and the contamination constant  $M$  was 0. For several other contaminated models, we considered several levels of contamination, with  $q = 5\%$  and  $10\%$  and contamination constants  $M = 5$  and  $10$ . When fitting function curves, we use generalized cross validation (GCV) to obtain the number of bases. The results showed that the number of bases of Model 1–3 were the same, while those of Models 4 and 5 were different. However, due to the need of calculating the change of principal component coefficient, we had to calculate it on the same basis. Therefore, for comparison purposes, in Models 4 and 5, we selected the same number of bases as that of Model 1.



**Figure 1.** Curves generated from Model 1 (without contamination), Model 2 (asymmetric contamination), Model 3 (symmetric contamination), Model 4 (partial contamination) and Model 5 (peak contamination) with  $n = 200$ ,  $p = 100$ ,  $q = 5\%$  and  $M = 10$ .

Classical L2-norm FPCA and L1-norm FPCA were used for the simulated functional data corresponding to these five models. We focused on their robust to various abnormal disturbances. When implementing L1-norm FPCA on Model 1, by comparing the value of objective function, the initial value was chosen as the first L2-norm functional principal component weight function, i.e.,  $\beta^0(t) = \xi(t)$ , where  $\xi(t)$  is the eigenfunction corresponding to the largest eigenvalue of the sample covariance function of the functional data in Model 1. Because the L1-norm FPCA of the following several disturbance models should be compared with Model 1, in order to ensure the consistency of conditions when calculating the L1-norm FPCA of the following several disturbance models, the initialization values also adopted the eigenfunction corresponding to the largest eigenvalue of the sample covariance function of the corresponding functional data.

The sums of absolute values of the coefficient differences of several principal components under non-contamination and contamination were compared and analyzed. The sum of the absolute values of the corresponding coefficient changes are given in Tables 1–4. Since the variance contribution rate of the first principal component reached 80%, only the first principal component function was taken in Models 1, 2, 3 and 4. However, in order to achieve a similar variance contribution rate, at least four principal components were needed in Model 5. Thus, for Models 1, 2, 3 and 4, we only show the changes of the first principal component function. For Model 5, we show the changes of the first four principal component functions.

**Table 1.** The sum of the absolute values of the first principal component weight function coefficient changes for no contamination and asymmetric contamination (5% and 10%).

q	M = 10				M = 5			
	L1-norm FPC		L2-norm FPC		L1-norm FPC		L2-norm FPC	
	1st FPC		1st FPC		1st FPC		1st FPC	
5%	0.17		1.13		0.13		0.91	
10%	0.22		1.24		0.18		1.16	

**Table 2.** The sum of the absolute values of the first principal component weight function coefficient changes for no contamination and symmetric contamination (5% and 10%).

q	M = 10				M = 5			
	L1-norm FPC		L2-norm FPC		L1-norm FPC		L2-norm FPC	
	1st FPC		1st FPC		1st FPC		1st FPC	
5%	0.2		1.13		0.1		0.96	
10%	0.25		1.17		0.23		1.04	

**Table 3.** The sum of the absolute values of the first principal component weight function coefficient changes for no contamination and partial contamination (5% and 10%).

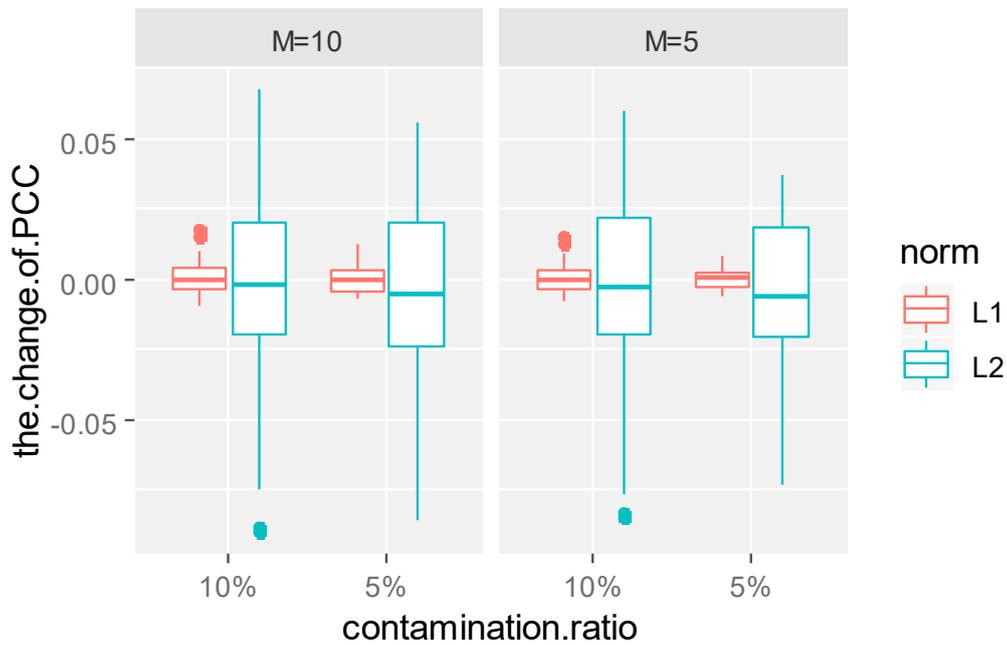
q	M = 10				M = 5			
	L1-norm FPC		L2-norm FPC		L1-norm FPC		L2-norm FPC	
	1st FPC		1st FPC		1st FPC		1st FPC	
5%	1.17		13.47		0.81		10.67	
10%	1.70		14.77		1.24		12.29	

**Table 4.** The sum of the absolute values of the first four principal component weight functions. Coefficient changes for no contamination and peak contamination (5% and 10%).

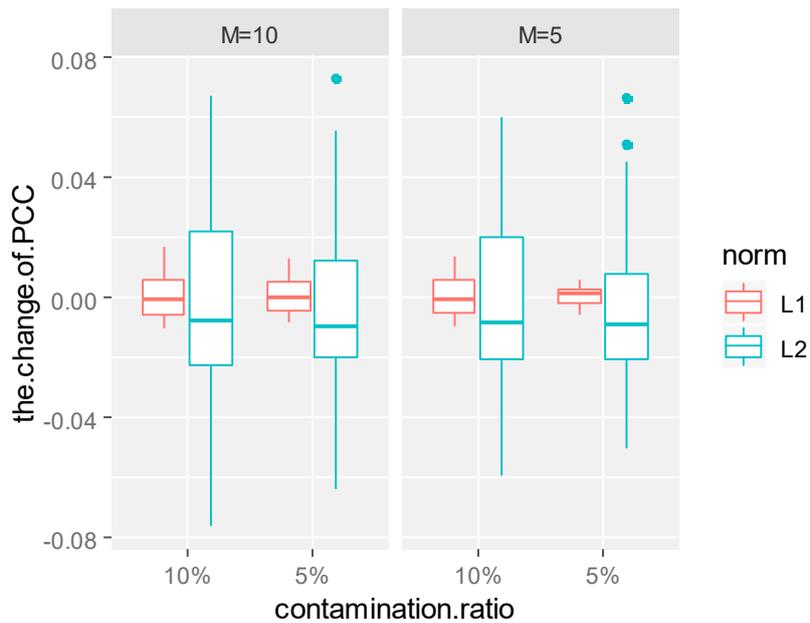
q	M = 10								M = 5							
	L1-norm FPC				L2-norm FPC				L1-norm FPC				L2-norm FPC			
	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th	1st	2nd	3rd	4th
5%	0.5	1.5	3.9	9.8	3.9	66.3	39.2	45.2	0.2	0.7	1.3	10.1	3.9	66.8	39.9	45.7
10%	0.8	1.8	4.1	8.4	5.4	24.2	48.4	54.3	0.4	1.2	2.2	3.4	2.3	10.9	22.3	42.2

It can be seen from Tables 1–4 that under the same contamination ratio and contamination size, the coefficient changes of the principal component weight functions of the L1-norm were significantly smaller than those of the L2-norm, which shows that the functional principal components of the L1-norm were more stable than those of the L2-norm, no matter which form of contamination was received. This conclusion can also be confirmed from the boxplots of the coefficient changes of the principal component weight functions.

As can be seen from Figures 2–5, in the same contamination ratio and size, the changes of L1-norm principal component coefficient are more concentrated near zero compared with the changes of the L2-norm principal component coefficient, which shows that under the same contamination mode, L1-norm functional principal components were more robust to outliers and more reliable.

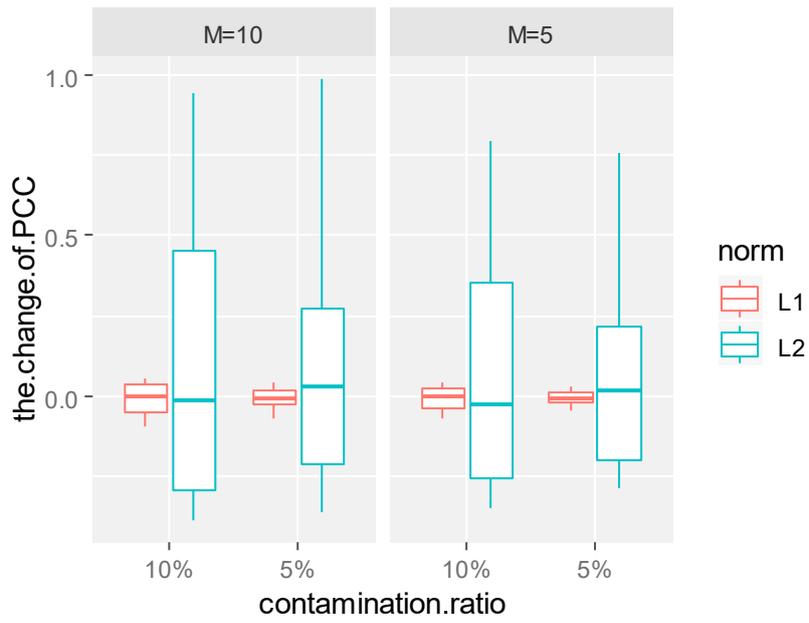


**Figure 2.** The boxplots of the change of the first principal component coefficient for asymmetric contamination ( $q = 5\%$  and  $q = 10\%$ ;  $M = 5$  and  $M = 10$ ).

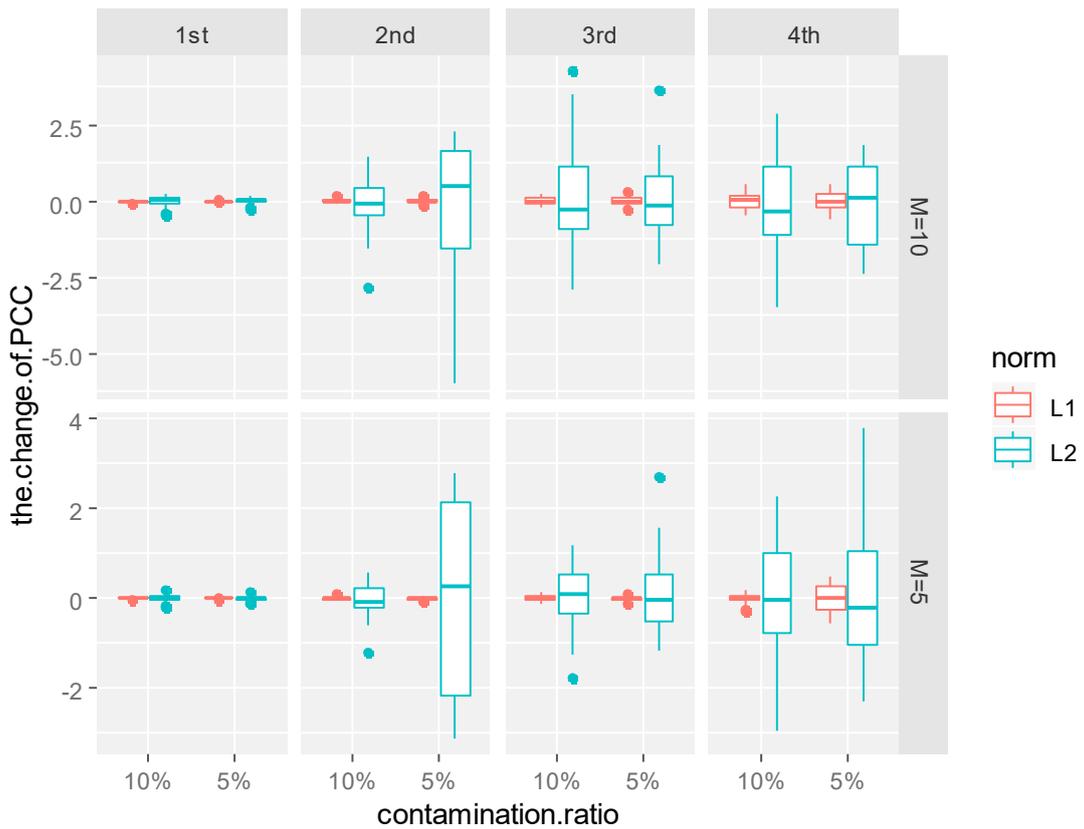


**Figure 3.** The boxplots of the change of the first principal component coefficient for symmetric contamination ( $q = 5\%$  and  $q = 10\%$ ;  $M = 5$  and  $M = 10$ ).

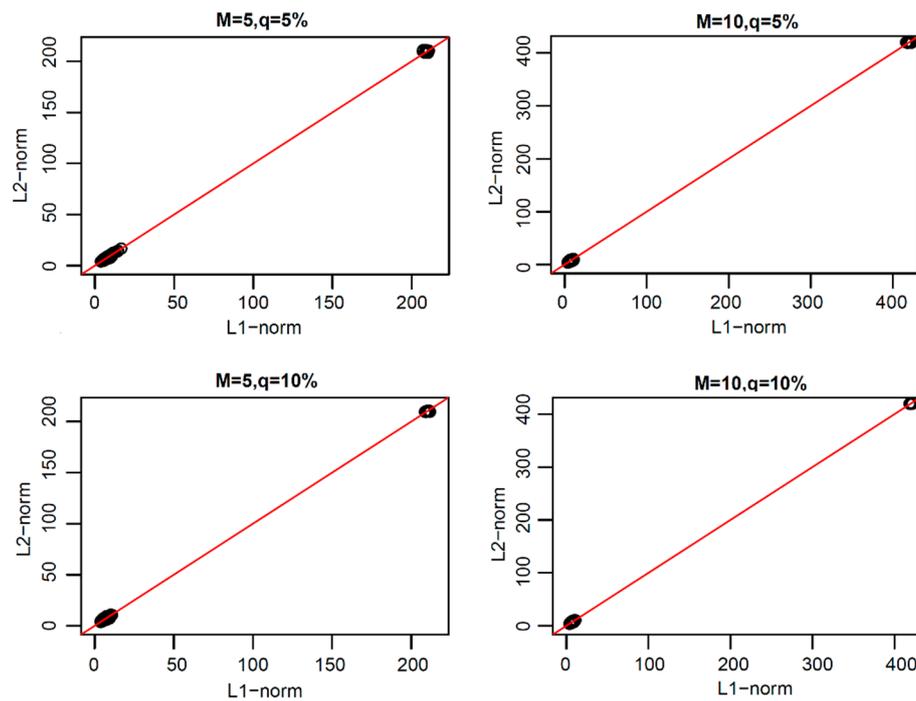
From the above research, we found that the L1-norm functional principal components were more robust than L2-norm functional principal components. Thus, how can one reconstruct the original functional data with these two types of principal components? In order to study this problem, we reconstructed the original uncontaminated functional data with the same number of functional principal components of L1-norm and L2-norm under each model. The scatter plots of the coefficients of the two types of reconstructed error curves are shown in Figures 6–9.



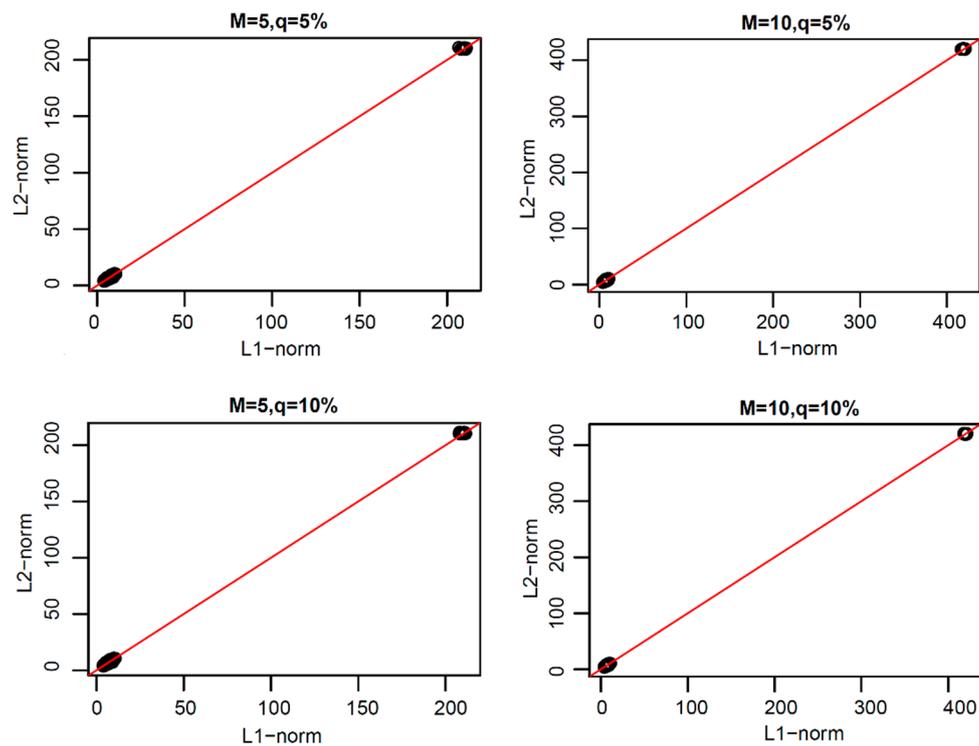
**Figure 4.** The boxplots of the change of the first principal component coefficient for partial contamination ( $q = 5\%$  and  $q = 10\%$ ;  $M = 5$  and  $M = 10$ ).



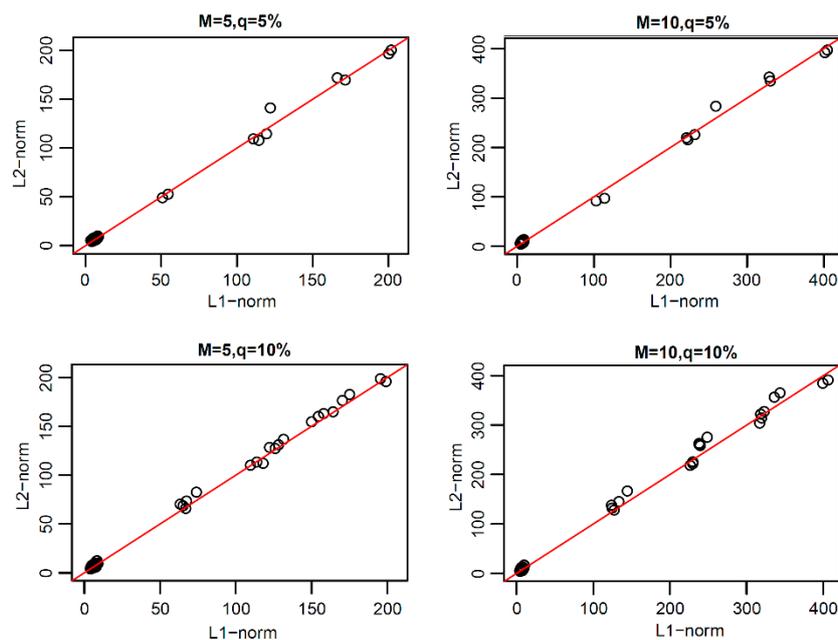
**Figure 5.** The boxplots of the change of the first four principal component coefficient for peak contamination ( $q = 5\%$  and  $q = 10\%$ ;  $M = 5$  and  $M = 10$ ).



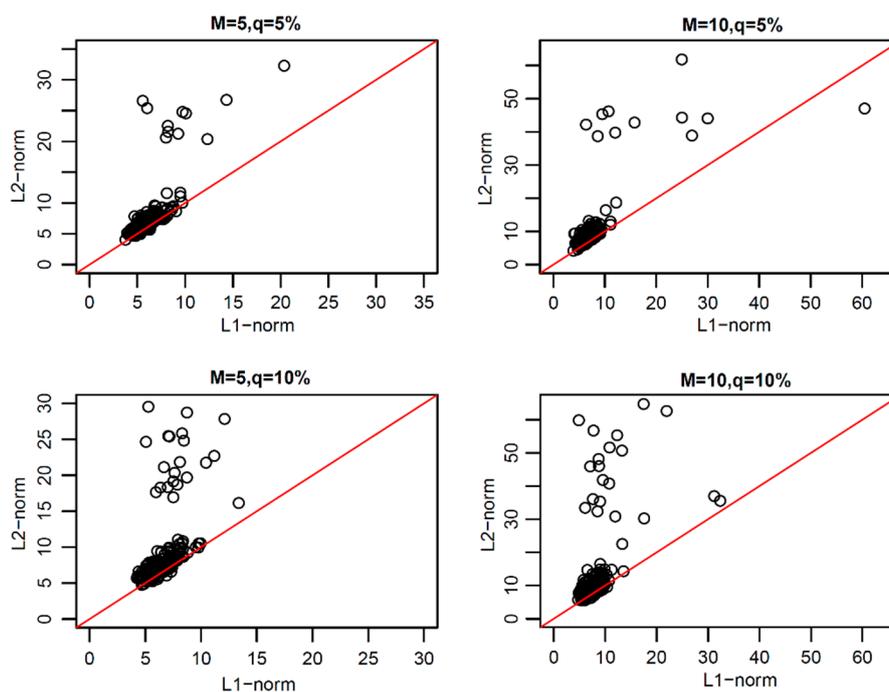
**Figure 6.** Scatter plots of the coefficients of the reconstruction error curves of L1-norm and L2-norm under asymmetric contamination.



**Figure 7.** Scatter plots of the coefficients of the reconstruction error curves of L1-norm and L2-norm under symmetric contamination.



**Figure 8.** Scatter plots of the coefficients of the reconstruction error curves of L1-norm and L2-norm under partial contamination.



**Figure 9.** Scatter plots of the coefficients of the reconstruction error curves of L1-norm and L2-norm under peak contamination.

In Figures 6–9, we can see that the scatter plots of the reconstruction error curve coefficients of L1-norm and L2-norm were always near the line  $y = x$  under the first three pollution models, and under peak pollution, the reconstruction error of the L1-norm was smaller than that of the L2-norm. When using the paired one-sided  $T$ -test, the  $p$ -values were found to all be close to 1, indicating that the reconstruction error curve coefficients of the L1-norm were not greater than those of the L2-norm. Thus, the reconstruction ability of the L1-norm principal components to the original uncontaminated functional data was not worse than that of the L2-norm principal components. The results of the paired one-sided  $T$ -test are shown in Tables 5–8.

**Table 5.** The table of the one-sided paired  $T$ -test of the coefficients of the reconstruction error curves of L1-norm and L2-norm under asymmetric contamination (Alternative hypothesis: The true difference of the reconstruction error curve coefficients of L1-norm and L2-norm in means was greater than 0.).

q	M = 5			M = 10		
	t	df	p-Value	t	df	p-Value
5%	−2.8447	199	0.9975	−2.1651	199	0.9842
10%	−2.2484	199	0.9872	−2.5843	199	0.9948

**Table 6.** The table of the one-sided paired  $T$ -test of the coefficients of the reconstruction error curves of L1-norm and L2-norm under symmetric contamination (Alternative hypothesis: The true difference of the reconstruction error curve coefficients of L1-norm and L2-norm in means was greater than 0.).

q	M = 5			M = 10		
	t	df	p-Value	t	df	p-Value
5%	−3.8761	199	0.9999	−3.34	199	0.9995
10%	−4.7628	199	1	−3.5293	199	0.9997

**Table 7.** The table of the one-sided paired  $T$ -test of the coefficients of the reconstruction error curve of L1-norm and L2-norm under partial contamination (Alternative hypothesis: The true difference of the reconstruction error curve coefficients of L1-norm and L2-norm in means was greater than 0.).

q	M = 5			M = 10		
	t	df	p-Value	t	df	p-Value
5%	−5.2373	199	1	−4.9371	199	1
10%	−7.7896	199	1	−5.033	199	1

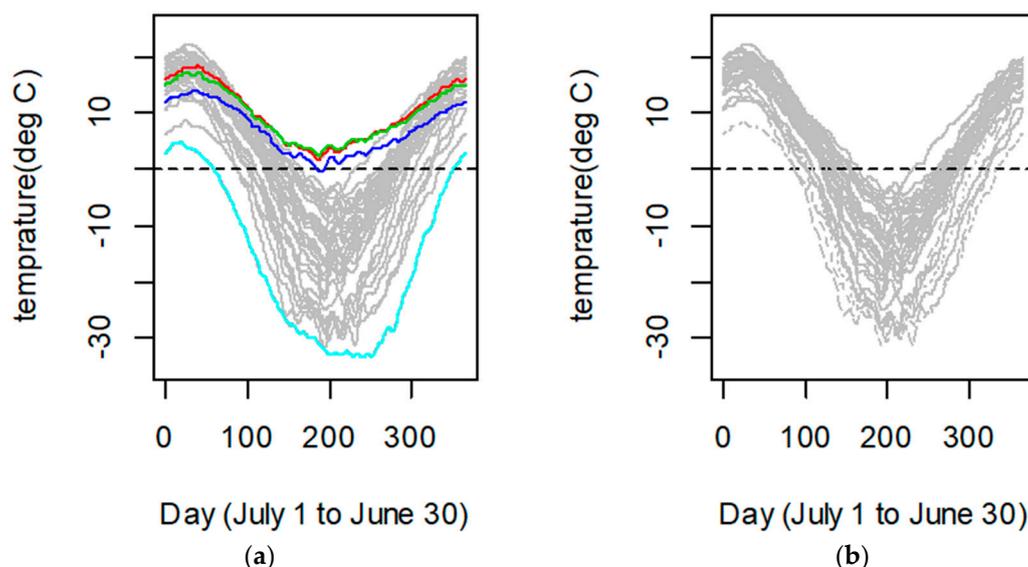
**Table 8.** The table of the one-sided paired  $T$ -test of the coefficients of the reconstruction error curves of L1-norm and L2-norm under peak contamination (Alternative hypothesis: The true difference of the reconstruction error curve coefficients of L1-norm and L2-norm in means was greater than 0.).

q	M = 5			M = 10		
	t	df	p-Value	t	df	p-Value
5%	−6.6502	199	1	−6.6212	199	1
10%	−7.6313	199	1	−6.8564	199	1

The above experiments showed that the functional principal component of the L1-norm was not just stable and reliable, it also had the same reconstruction ability as the L2-norm.

#### 4.2. Canadian Weather Data

We used Canadian weather data, which provide daily temperatures at 35 different locations in Canada averaged over 1960–1994, in order to compare the robust to outliers of the L1-norm functional principal components and L2-norm functional principal components when the functional data were contaminated by abnormal data. Firstly, by considering the periodic characteristics of the data, the discrete temperature observation data were fitted into 35 functional curves by a Fourier basis function, and the number of the basis functions was 65. The fitting curves are shown in Figure 10a. As can be seen when using the function data outlier detection method [39], the temperature modes of the four stations of Vancouver, Victoria, Pr. Rupert and Resolute were different from those of the other stations. Figure 10b shows this function after removing the data from these four observatories. The functional data of the 35 observatories were called the whole data, and the functional data after removing Vancouver, Victoria, Pr. Rupert and Resolute were normal data, so the whole data can be understood as the addition of abnormalities to the normal data.



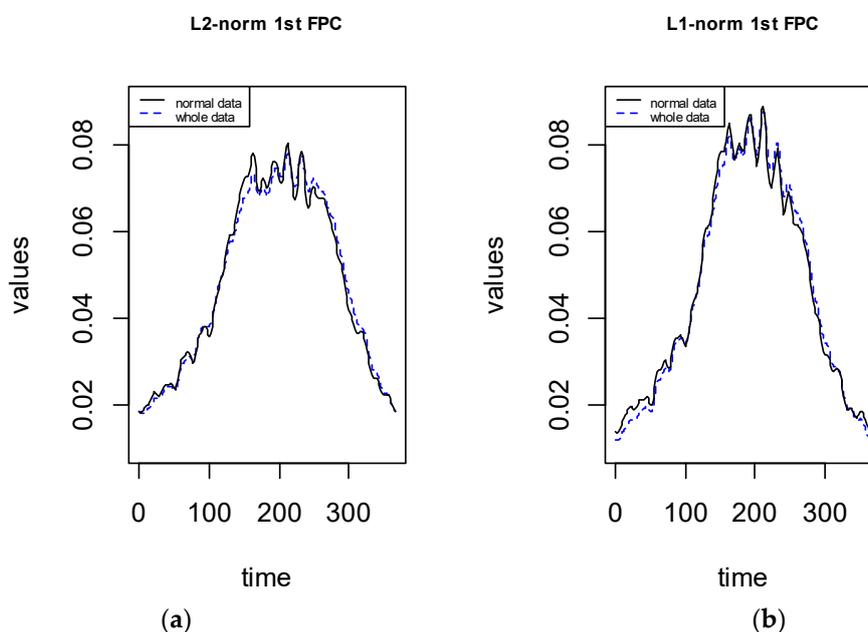
**Figure 10.** Daily mean temperature curves of 35 observatories in Canada ((a) whole data, (b) normal data).

In order to compare the robustness between the L2-norm functional principal component weighting functions and the L1-norm functional principal component weighting functions to outliers, the L2-norm functional principal components and L1-norm functional principal components were, respectively, used for normal data and data added with outliers. For each method, the results of the two cases were compared, because the variance contribution rate of the first two principal components reached 90%, though the latter analysis only focused on the first two functional principal components.

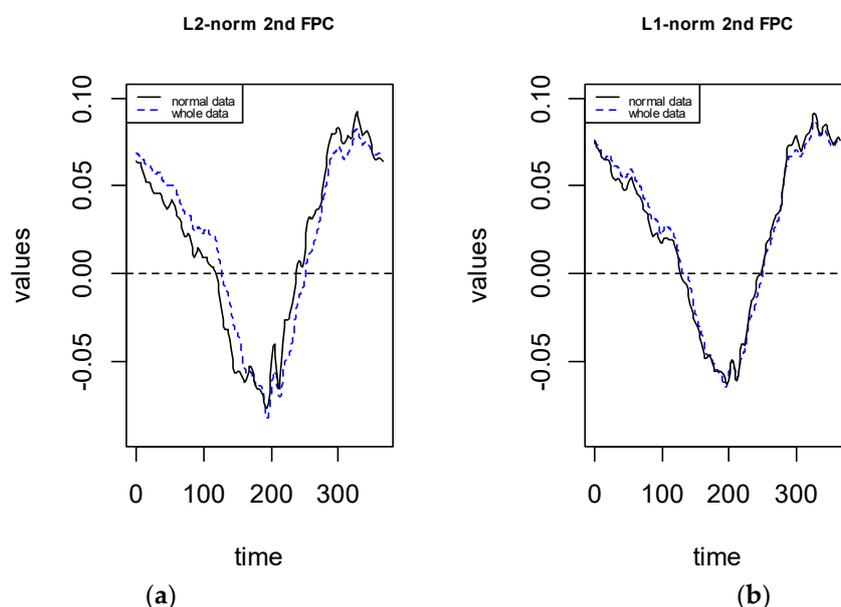
Figure 11 shows the change of the first principal component weight function before and after adding outliers by using two functional principal component analysis methods. Figure 11a is a graph of the first principal component weight function that was obtained by using the L2-norm functional principal component method. The solid line is the result of normal data, and the dashed line is the result of adding four abnormal curves. Figure 11b is a graph of the first principal component weight function that was obtained by using the proposed L1-norm functional principal component method. After comparing the objective function, the initial value was chosen as the first L2-norm functional principal component weight function, i.e.,  $\beta^0(t) = \xi(t)$ , where  $\xi(t)$  is the eigenfunction corresponding to the largest eigenvalue of the sample covariance function of normal functional data and the same method for whole functional data. The solid line is the result of normal data, and the dashed line is the result of adding four abnormal curves. By comparing the coefficients of the two first functional principal component weighting functions, it was found that the sum of the absolute change of the coefficients of the first principal component weighting functions that were obtained by the L1-norm method before and after adding abnormal values was 0.16, which was less than the 0.18 corresponding to the L2-norm. Next, the performance of the second principal component weight function is discussed.

Figure 12 shows the change of the second principal component weight function before and after the addition of outliers by using two functional principal component analysis methods. Figure 12a is a graph of the second principal component weight function that was obtained by using the L2-norm functional principal component method. The solid line is the result of normal data, and the dashed line is the result of adding four abnormal curves. Figure 12b is a graph of the second principal component weight function that was obtained by using the proposed L1-norm functional principal component method. The solid line is the result of normal data, and the dashed line is the result of adding four abnormal curves. By comparing the coefficients of the two second function principal component weighting functions, it was found that the sum of absolute change of the coefficients of the second principal component weighting functions that were obtained by the L1-norm method before and after

adding abnormal values was 0.33, which was less than the 0.76 corresponding to the L2-norm. the sums of the absolute values of the coefficient change of the principal component weight functions under the two methods are shown in Table 5.



**Figure 11.** The first principal component weight function for normal data and whole data. (a) L2-norm, (b) L1-norm.



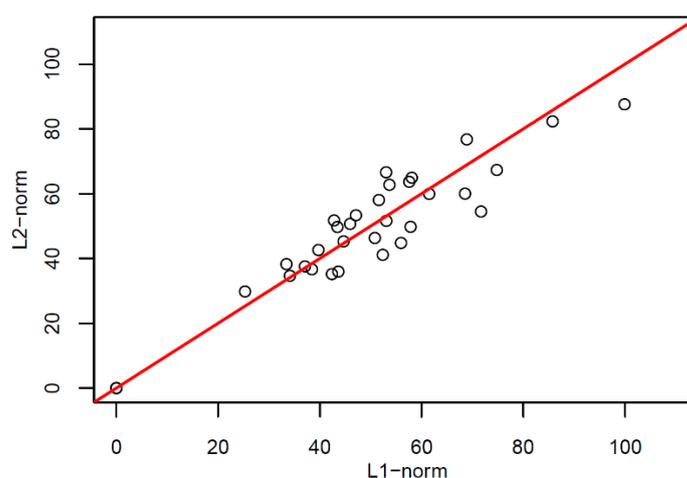
**Figure 12.** The second principal component weight function for normal data and whole data. (a) L2-norm, (b) L1-norm

Table 9 shows that the classical L2-norm principal components weight functions greatly changed before and after removing outliers, reflecting its sensitivity to outliers. However, the L1-norm functional principal components weight functions presented in this paper had little change before and after adding abnormal values. Therefore, that the L1-norm principal component weight function proposed in this paper has a strong anti-noise ability and a good stability.

**Table 9.** The sum of absolute change of the coefficients of the first two principal component weighting functions.

The Sum of Absolute Change of the Coefficients	The 1st Function Principal Component Weighting Function	The 2nd Function Principal Component Weighting Function
L2-norm	0.18	0.76
L1-norm	0.16	0.33

We also compared the reconstruction ability of two types of principal components to normal data. The scatter plots of the coefficients of the two types of reconstructed error curves are shown in Figure 13.



**Figure 13.** Scatter plot of reconstruction error curve of L1-norm and L2-norm to normal data.

From Figure 13, it can be seen the scatter plots of the reconstruction error curve coefficients of L1-norm and L2-norm were always near the line  $y = x$ ; When we performed a paired one-sided  $T$ -test for the two groups of reconstruction error curve coefficients, the  $t$  value was found to 1.0323, the degree of freedom for the  $t$ -statistic was 33, and the  $p$ -value was 0.1547, which indicates that the reconstruction error curve coefficients of L1-norm were not greater than those of L2-norm. Thus, the reconstruction ability of the L1-norm principal components to the original uncontaminated functional data was not worse than the L2-norm principal components.

## 5. Concluding Remarks

FPCA is a primary step for functional data exploration, and the reliability of FPCA plays an important role in subsequent analysis. The existing principal component methods of functional data were established in an L2-norm framework. However, because the L2-norm enlarges the influence of outliers, the traditional functional principal components analysis method is sensitive to outliers. On the other hand, in regard to multivariate data, the relevant research on the principal component analysis method [30–37] have shown that the principal component analysis method of L1-norm for multivariate data has a better robustness than that of the L2-norm. Motivated by this research, in this paper, we tried to construct an L1-norm principal component analysis method for functional data. Firstly, we built a functional data L1-norm maximized principal component optimization model. Then, a corresponding algorithm for solving the L1-norm maximized optimization model was constructed based on the idea of multivariate data L1-norm principal component analysis method [30]. An extensive simulation study was conducted, and a real dataset of Canadian weather was employed to assess the robustness of the L1-norm functional principal component analysis. From the simulation study that considered different contamination configurations (symmetric, asymmetric, partial and peak), we found that the L1-norm functional principal component analysis method provides a more robust estimation of

principal components than the traditional L2-norm principal component analysis method. Finally, by comparing the reconstruction errors of the L1-norm FPCA and L2-norm FPCA, it was found that the reconstruction ability of the L1-norm principal components to the original uncontaminated functional data is as good as that of the L2-norm principal components. Therefore, when functional data contain outliers, the estimation given by the L1-norm functional principal component analysis method is more reliable. The proposed L1-norm FPCA may prove to be an useful addition to functional data analysis.

**Author Contributions:** Individual contributions to this article: conceptualization, F.Y. and L.L.; methodology, F.Y. and L.L.; software, L.J. and D.Q.; validation, F.Y. and N.Y.; writing—original draft preparation, F.Y. and L.L.; writing—review and editing, F.Y. and L.L.; supervision, L.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported in part by the Beijing Natural Science Foundation under Grant No.9172003, in part by the National Natural Science Foundation of China under Grant No. 61876200, in part by the Natural Science Foundation Project of Chongqing Science and Technology Commission under Grant No. cstc2018jcyjAX0112.

**Acknowledgments:** The authors thank the anonymous referees for their careful reading and helpful suggestions, which help to improve the quality of this paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Kowal, D.R. Integer-valued functional data analysis for measles forecasting. *Biometric* **2019**, in press. [[CrossRef](#)] [[PubMed](#)]
2. Wagner-Muns, I.M.; Guardiola, I.G.; Sama-ranayke, V.A.; Kayani, W.I. A functional data analysis approach to traffic volume forecasting. *IEEE Trans. Intell. Transp. Syst.* **2017**, *19*, 878–888. [[CrossRef](#)]
3. Ramsay, J.O.; Silverman, B.W. Applied functional data analysis. *J. Educ. Behav. Stat.* **2008**, *24*, 5822–5828.
4. Yao, F.; Müller, H.G.; Wang, J.L. Functional data analysis for sparse longitudinal data. *J. Am. Stat. Assoc.* **2005**, *100*, 577–590. [[CrossRef](#)]
5. Auton, T. Applied functional data analysis: Methods and case studies. *J. R. Stat. Soc.* **2010**, *167*, 378–379. [[CrossRef](#)]
6. Zambom, A.Z.; Collazos, J.A.; Dias, R. Functional data clustering via hypothesis testing *k*-means. *Comput. Stat.* **2019**, *34*, 527–549. [[CrossRef](#)]
7. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice*; Springer Science & Business Media: Berlin, Germany, 2006.
8. Horváth, L.; Kokoszka, P. *Inference for Functional Data with Applications*; Springer Science & Business Media: Berlin, Germany, 2012; Volume 200.
9. Tarpey, T.; Kinatader, K.K. Clustering functional data. *J. Classif.* **2003**, *20*, 93–114. [[CrossRef](#)]
10. Ramsay, J.O.; Silverman, B.W. *Applied Functional Data Analysis: Methods and Case Studies*; Springer: Berlin/Heidelberg, Germany, 2007.
11. Estévez-Pérez, G.; Vilar, J.A. Functional anova starting from discrete data: An application to air quality data. *Environ. Ecol. Stat.* **2013**, *20*, 495–517. [[CrossRef](#)]
12. Ignaccolo, R.; Ghigo, S.; Giovenali, E. Analysis of air quality monitoring networks by functional clustering. *Environmetrics* **2010**, *19*, 672–686. [[CrossRef](#)]
13. Ferraty, F.; Vieu, P. Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. *Nonparametr. Stat.* **2004**, *16*, 111–125. [[CrossRef](#)]
14. Febrero, M.; Galeano, P.; González-Manteiga, W. Outlier detection in functional data by depth measures, with application to identify abnormal nox levels. *Environmetrics* **2010**, *19*, 331–345. [[CrossRef](#)]
15. Ratcliffe, S.J.; Heller, G.Z.; Leader, L.R. Functional data analysis with application to periodically stimulated foetal heart rate data ii functional logistic regression. *Stat. Med.* **2002**, *21*, 1103–1114. [[CrossRef](#)] [[PubMed](#)]
16. Giraldo, R.; Delicado, P.; Mateu, J. Continuous time-varying kriging for spatial prediction of functional data: An environmental application. *J. Agric. Biol. Environ. Stat.* **2010**, *15*, 66–82. [[CrossRef](#)]
17. Ferraty, F.; Rabhi, A.; Vieu, P. Conditional quantiles for dependent functional data with application to the climatic “el niño” phenomenon. *Sankhyā Indian J. Stat.* **2005**, *67*, 378–398.

18. Baladandayuthapani, V.; Mallick, B.K.; Young Hong, M.; Lupton, J.R.; Turner, N.D.; Carroll, R.J. Bayesian hierarchical spatially correlated functional data analysis with application to colon carcinogenesis. *Biometrics* **2008**, *64*, 64–73. [[CrossRef](#)]
19. Ramsay, J.O.; Dalzell, C.J. Some tools for functional data analysis. *J. R. Stat. Soc.* **1991**, *53*, 539–572. [[CrossRef](#)]
20. Ramsay, J.O. Functional data analysis. *Encycl. Stat. Sci.* **2004**, *4*. [[CrossRef](#)]
21. Dauxois, J.; Pousse, A.; Romain, Y. Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivar. Anal.* **1982**, *12*, 136–154. [[CrossRef](#)]
22. Rice, J.A.; Silverman, B.W. Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Stat. Soc.* **1991**, *53*, 233–243. [[CrossRef](#)]
23. Levy, A.; Rubinstein, J. Some properties of smoothed principal component analysis for functional data. *J. Opt. Soc. Am.* **1999**, *16*, 28–35. [[CrossRef](#)]
24. Silverman, B.W. Smoothed functional principal components analysis by choice of norm. *Ann. Stat.* **1996**, *24*, 1–24. [[CrossRef](#)]
25. James, G.M.; Hastie, T.J.; Sugar, C.A. Principal component models for sparse functional data. *Biometrika* **2000**, *87*, 587–602. [[CrossRef](#)]
26. Boente, G.; Fraiman, R. Kernel-based functional principal components. *Stat. Probab. Lett.* **2000**, *48*, 335–345. [[CrossRef](#)]
27. Hall, P.; Hosseini-Nasab, M. On properties of functional principal components analysis. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2006**, *68*, 109–126. [[CrossRef](#)]
28. Benko, M.; Härdle, W.; Kneip, A. Common functional principal components. *Ann. Stat.* **2009**, *37*, 1–34. [[CrossRef](#)]
29. Hörmann, S.; Kidziński, Ł.; Hallin, M. Dynamic functional principal components. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2015**, *77*, 319–348. [[CrossRef](#)]
30. Kwak, N. Principal component analysis based on l1-norm maximization. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *30*, 1672–1680. [[CrossRef](#)]
31. Nie, F.; Huang, H.; Ding, C.; Luo, D.; Wang, H. Robust principal component analysis with non-greedy  $\ell_1$ -norm maximization. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011.
32. Markopoulos, P.P.; Karystinos, G.N.; Pados, D.A. Optimal algorithms for L1-subspace signal processing. *IEEE Trans. Signal Process.* **2014**, *62*, 5046–5058. [[CrossRef](#)]
33. Markopoulos, P.P.; Kundu, S.; Chamadia, S.; Pados, D.A. Efficient L1-norm principal-component analysis via bit flipping. *IEEE Trans. Signal Process.* **2017**, *65*, 4252–4264. [[CrossRef](#)]
34. Martin-Clemente, R.; Zarzoso, V. On the link between L1-PCA and ICA. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 515–528. [[CrossRef](#)]
35. Park, Y.W.; Klabjan, D. Iteratively reweighted least squares algorithms for L1-norm principal component analysis. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 12–15 December; IEEE: Piscataway, NJ, USA, 2016.
36. Markopoulos, P.P.; Dhanaraj, M.; Savakis, A. Adaptive L1-norm principal-component analysis with online outlier rejection. *IEEE J. Sel. Top. Signal Process.* **2018**, *12*, 1131–1143. [[CrossRef](#)]
37. Tsagkarakis, N.; Markopoulos, P.P.; Pados, D.A. On the L1-norm approximation of a matrix by another of lower rank. In Proceedings of the 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), Anaheim, CA, USA, 18–20 December 2016; IEEE: Piscataway, NJ, USA, 2016.
38. Fraiman, R.; Muniz, G. Trimmed means for functional data. *Test* **2001**, *10*, 419–440. [[CrossRef](#)]
39. Yu, F.; Liu, L.; Jin, L.; Yu, N.; Shang, H. A method for detecting outliers in functional data. In Proceedings of the IECON 2017-43rd Annual Conference of the IEEE Industrial Electronics Society, Beijing, China, 29 October–1 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 7405–7410.

