Robust Methods for Multivariate Functional Data Analysis

by

Pallavi Sawant

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 3, 2013

Keywords: Multivariate Functional Data, Robust Eigenfunction , Robust Functional Regression

Approved by

Nedret Billor, Chair, Associate Professor of Mathematics and Statistics
Asheber Abebe, Associate Professor of Mathematics and Statistics
Peng Zeng, Associate Professor of Mathematics and Statistics
George T. Flowers, Dean, Graduate School

Abstract

In the present work we have proposed a robust multivariate functional principal component analysis (RMFPCA) method, that is efficient in estimation and fast in computation, to achieve dimension reduction of dataset and to develop tools for detection of functional outliers. We intend to develop smooth principal functions as M-type smoothing spline estimators by using penalized M-regression with a bounded loss function. The proposed method is more efficient since it makes maximal use of the normally observed measurements by separately downweighing abnormally observed measurements in a single curve. Using natural cubic splines formulation the computation of the proposed method becomes fast for functional data. We have described accompanying diagnostic plots that can be used to detect possible outliers. Simulations are conducted to investigate the effectiveness of the proposed robust multivariate functional principal component analysis (PCA) based on MM estimation, in which we compare our proposed methodology with classical multivariate functional PCA.

The estimation of a functional coefficient in a regression setting where the response is a scalar and explanatory variables are sampling points of a continuous process is considered. Multivariate linear model is considered for several applications, for instance chemometrics where some chemical variables have to be predicted by a digitized signal such as the Near Infrared Reflectance (NIR) spectroscopic information. These methods do not really take into account the functional nature of the data. Further it is typical to have outlying observations in such datasets. Fitting functional parameter by using functional regression is vulnerable to unusual data. Therefore from functional point of view a robust functional principal component regression (RFPCR) is proposed for regressing scalar response on the space, spanned by small number of eigenfunctions of the functional predictor. Before running a

regression the outlying trajectories in this space are down weighted by using re-weighted least squares approach. Several simulation results and the analysis of a real data set indicate the robustness of the proposed method.

Acknowledgments

I express my sincere gratitude to my advisor, Dr. Nedret Billor, who suggested this dissertation and generously gave, continued guidance and support during the course of this work. I highly appreciate her insights and valuable suggestions for this work, without which I could not have completed this dissertation. I would also like to acknowledge my committee members Dr. Asheber Abebe and Dr. Peng Zeng for their valuable support.

Many thanks to the faculty, staff and fellow students at the Department of Mathematics and Statistics for making my study at Auburn such a wonderful experience. I would like to express my special gratitude to my parents for their never-ending love and support throughout my life. I feel lucky to have Rajesh as my life partner, whose companionship, and ever lasting support and commitment during all these years of studying at the Auburn University have been a source of strength and encouragement. Finally it's the grace and blessings of the God that has helped me fulfill my aspirations.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

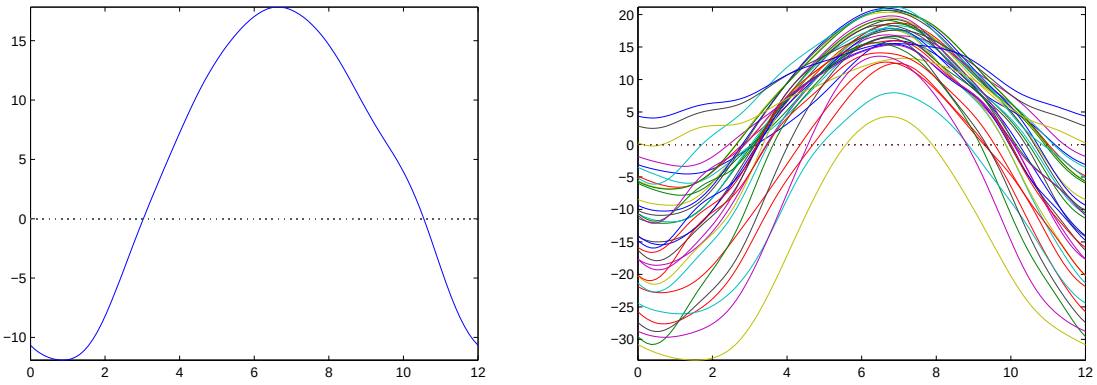
Functional Data Analysis (FDA) or "data analysis with curves" is a popular subject in statistics with a wide range of applications. FDA is an assemblage of different methods in statistical analysis for analyzing curves or functional data. With advances in technologies high volumes of complex data are generated. For example, in climatology many weather parameters have been recorded over decades, in the field of chemistry and physics chemometric analysis is done for spectroscopy. Complexity and large size of databases mandate use of new tools for analysis such as FDA [64, 47].

Functional data analysis helps to extract additional information from densely sampled observations over a time or space. In standard statistical methodology the focus is on the set of data vectors whereas, in FDA focus is on the type of data structure such as curves, shapes, images, or set of functional observations.

In FDA, each observed curve is thought of as a single observation rather than a collection of individual observations. A curve can be regarded as an infinite-dimensional vector, whose dimensions may not be countable (Figure 1.1(a),(b)).

In a traditional statistical methodology, the usual data types are univariate and multivariate. A univariate dataset contains numbers as its observations; while a multivariate dataset contains vectors as its observations. A number is one-dimensional while a vector is multi-dimensional. Multivariate Data Analysis (MDA) is an extension of Univariate Data Analysis and FDA is an extension of multivariate analysis, where the random vectors are of infinite dimensions.

In number of situations functional data can be treated as multivariate data. However, treating data directly as multivariate data may pose difficulty, such as when design points



(a) (b)

Figure 1.1: Example of (a)Functional Observation and (b)Functional dataset

are not equal in subjects. So, direct multivariate treatment may not be possible in this case. This calls for the development of functional data analysis.

When each functional observation is sampled at a same set of design points, the functional data we get may look like multivariate data. But functional data is in general different from the multivariate data in the following aspects: 1. For a functional observation, the observed data is sampled from an underlying smooth function, whereas in a multivariate dataset for an observed vector there is no such structure. 2. The dimension of a functional observation is so large that it is regarded as a continuous function. This can be seen in Figure 1.1(a). This dimension is often larger than the sample size. 3. The time points can be different from one data point to another. All of these different aspects necessitate development of functional data analysis [77].

There are three advantages in treating data in functional forms. First, by representing data in functional form with small number of parameters reduces its size considerably. Second, since FDA deals with continuous functions; information between observed points is not lost. For finite sets of observations, FDA first estimates functions from the observed data, and then discretizes the function at any suitable choice of time points for further analysis. The free choice of analyzed points is attractive when observational points are different in

each subject. Thirdly, it is very useful to have particularly interesting features in some time interval to have more closely spaced points.

In functional data framework, the random variables are defined on the functional space. To model the population of these random functions we think of a functional data observation as a realization of a stochastic process, $X(t)$, $t \in T$, where T is a bounded interval in $\Re$.

Some mathematical concepts used for FDA are explained here. In FDA, we work with a functional Hilbert space L_2 (e.g., inner product space) which is determined by an inner product $\langle x, y \rangle$ [45]. In a finite dimension with $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$, the Euclidean inner product is defined in the following way:

$$\langle x, y \rangle_{\Re^n} = \Sigma x_i y_i.$$

In a functional space where $x = x(t)$ and $y = y(t)$ are functions, the L_2 inner product is defined as:

$$\langle x, y \rangle_{L_2} = \int x(t)y(t)dt,$$

where $x, y \in L_2$. For the convenience we drop the subscripts $\langle x, y \rangle_{L_2}$ and just use $\langle x, y \rangle$.

The statistical model underlying functional data, which is represented by curves, is a stochastic process defined on a given infinite dimensional function space. While dealing with functional data mostly univariate cases are considered i.e. $X(t) \in \Re$, $\forall\ t = \{t_1, t_2, \ldots, t_q\} \in T$, $T \in [t_{min}, t_{max}]$, a path of X is represented by a single curve. The multidimensional case or multivariate functional data, $X = X(t)$, $t = \{t_1, t_2, \ldots, t_q\} \in T$, $T \in [t_{min}, t_{max}]$ with $X(t) = (X_1(t), \ldots, X_p(t))' \in \Re^p$, $p \geq 2$, is rarely considered in literature. In multivariate functional data each observation is a finite dimension vector whose elements are functions and X is represented by a set of p curves. The dependency between p curves provides the structure of X.

Consider the functions as processes in continuous time defined over an interval, say $T \in [t_{min},\ t_{max}]$. The i^{th} replication of functional observation is denoted as $x_i(t) \in L_2[T]$,

$i = 1, \ldots, n$. In practice, it is impossible to observe the functional values in continuous time. We usually obtain the data only on a finite and discrete grid $t = \{t_1, t_2, \ldots, t_q\} \in T$ in the following manner:

$$y_i = x_i(t) + \epsilon_i, \qquad 1 \leq i \leq n,$$

where ϵ_i is a random error or noise with zero mean and variance function $\sigma_i^2(t)$. For simplicity, we assume that all processes are observed at the same time points, which are equally spaced on T and is denoted by $t = \{t_1, t_2, \ldots, t_q\}$, but in reality $t_j \in T$ can be different, i.e., t_{ij} depending on i where $1 \leq j \leq n_i$, $1 \leq i \leq n$ and n_i is number of time points for i^{th} replication of functional observation.

For any data analysis in the FDA framework first step is functional data smoothing. It is done to convert raw discrete data points into smooth functions i.e., to convert data to functional form. Smoothing method is used to minimize noise in raw data for calculations and analysis. There are different types of smoothers that can be applied to functional data. In this thesis we use smoothing based on basis-function method. By the use of basis function discrete data is represented as a smooth function this is also known as functional data smoothing. In the basis expansion method; the function x_i can be represented as a linear combination of first k known basis functions ϕ_K, $K = 1, \ldots, k$, where k is large enough, $k < q$. In this approach, a functional observation x_i is expressed as:

$$x_i(t) = \sum_{K=1}^{k} c_{iK}\phi_K(t),$$

where ϕ is vector-valued function having components $\phi_1, \ldots, \phi_{k_l}$. The C is $n \times k$ coefficient matrix of the expansion, where $C = [c_{iK}]$, $1 \leq i \leq n$, $1 \leq K \leq k$. The simultaneous expansion of all n curves can be expressed in matrix notation as:

$$x = C\phi,$$

where x is a vector-valued function with x_i, $1 \leq i \leq n$, as its components. This approach is preferred since it makes good approximation of the data with a relatively small number of parameters. This may be considered to be a dimension reduction operation.

There are many basis functions possible. Fourier and B-spline bases are most frequently used bases functions. Since no basis is universally good, choosing one is a complex issue. However, there are guidelines for specific situations as each candidate function for the basis has the unique characteristics; for example if the data are periodic then a Fourier basis is used and for non-periodic data or data that have a lot of local features B-spline works better. The selection of the basis function $\phi_K(t)$ is done by observing the data. Selecting proper order of expression k (the number of basis functions) is important question in the basis expansion. There are many ways to decide the number of basis functions like Cross-Validation (CV), Generalized Cross Validation (GCV) or other similar criteria. In this thesis we use GCV developed by Craven and Wahba [15], which is described in Section 4.2.

High dimensional data occurrence is natural in some practical applications such as studies involving image analysis and microarray datasets in genomic studies. In such applications dimension q is greater than n, sample size. High dimensionality problem has two distinct features: first the dimension q depends on the discretization order. This is not given in advance and can be arbitrarily increased. Second, the data from the discretized functions are likely to be highly correlated. Thus; they create difficulty in estimation of the covariance matrices.

The notion behind dimension reduction methods is to transform original variables into a few new variables so that new variables contain most of the information in the original variables. Principal component analysis (PCA) has been widely used for dimension reduction and visualization of multivariate and high-dimensional data. In recent decades, multivariate PCA has been extended to functional PCA. FPCA is a useful tool for data reduction, which is achieved by identifying main modes of variability of a given dataset.

For the set of multivariate curves, one can be interested in optimal representation of curves in a functional space of reduced dimension. Multivariate Functional Principal Component Analysis (MFPCA) is a useful statistical technique for understanding the structure of multivariate functional data (MFD). They are effective dimension reduction tools for MFD. Ramsay and Silverman [64] have given an example of bivariate functional data, $X(t) = (X_1(t), X_2(t))' \in \Re^2$, as a model for gait data (knee and hip measures) used in the context of multivariate functional principal component analysis (FPCA) as an extension of the univariate case. Principal components obtained from MFPCA have the same interpretation as in the functional univariate case. MFPCA aims to explain the covariance structure of data by means of small number of functional components. These functional components are linear combinations of the original functional variables. This gives better interpretation of the different sources of variation. Thus MFPCA for data reduction is useful tool for analyzing high dimensional data.

Majority of statistical techniques used in multivariate and functional data analyses assume that the dataset is free of outliers. However, outliers occur very frequently in functional and multivariate functional data. Possible sources of outliers are errors in recording and measurement, incorrect distribution assumption, unknown data structure, or novel phenomenon [39]. Since traditional FPCA and multivariate FPCA are sensitive to outliers, presence of outliers in a functional dataset makes the resulting principal function unreliable.

In the presence of outliers, dimension reduction via MFPCA would yield untrustworthy results since MFPCA is also known to be sensitive to outliers. Although several robust FPCA methods have been proposed for univariate functional data, to our knowledge, there has been no study on the robustness of MFPCA. This necessitates to develop robust MFPCA for multivariate FD.

The main contribution of our work is to construct a robust MFPCA method to achieve dimension reduction of data and to develop tools for detection of functional outliers. This research work is organized as follows. Chapter 2 reviews univariate functional PCA and

multivariate functional PCA. Chapter 3 describes two robust dimension reduction methods for univariate functional data. In Chapter 4 we explore the sensitivity of MFPCA to outliers by using two different approaches of robust MFPCA and finally propose a functional outlier detection procedure based on MM estimation. We describe accompanying diagnostic plots that can be used to detect possible outliers. Simulations are conducted to investigate the effectiveness of the proposed robust multivariate functional PCA based on MM estimation, in which we compare our procedure with classical multivariate functional PCA. In Chapter 5, we propose a robust technique for estimating the parameter function in a functional linear model with scalar response by using robust functional principal component analysis. Finally, conclusions and proposed future work conclude the dissertation in Chapter 6.

Chapter 2

Functional Principal Component Analysis

2.1 Introduction

In various areas such as chemometrics, biometrics, engineering, genetics, and e-commerce the data come from the observation of continuous phenomenons of time or space known as functional data. Due to advancement of new techniques it is now possible to record large number of variables simultaneously. The nature of this data in many applications is high dimensional where the number of variables (q) is greater than the number of observations (n) ($n \ll q$). The focus of researchers is on analysis of such data due to the emergence of statistical problems while applying various statistical tools for data analysis.

Principal component analysis (PCA) has been widely used for dimension reduction and visualization of multivariate and high-dimensional data. In recent decades, multivariate PCA has been extended to functional PCA.

Functional principal component analysis (FPCA) is a useful tool to reduce dimension of functional data. The first step in a FDA is to represent the data in a lower dimensional space in order to have better interpretation. This is done by performing FPCA to capture the main modes of variability of the data by means of small number of components which are linear combinations of original variables that allow for better interpretation of various sources of variation.

A number of recent papers have investigated different versions and properties of functional principal component analysis. Ramsay and Silverman [64] and Ferraty and Vieu [24] described the general approach to reformulate PCA in terms of functions. Early work on techniques for functional PCA includes that of Rice and Silverman [66] and Silverman [73] and more recent research includes Cardot [8], James *et al.* [40] and Huang *et al.* [34]. Rice

and Silverman [66] proposed a method based on projecting functional data onto finite dimensional basis and applying a classical multivariate analysis with a roughness penalty on the weight functions to achieve smoothness. Silverman [73] considered estimators based on penalizing the norm rather than the sample variance. Asymptotic properties were studied by Dauxois *et al.* [17] but their approach may produce rough principal components. Boente and Fraiman [7] provided smooth estimators by considering a kernel approach by regularizing the trajectories. FPCA methods for sparsely sampled functional data or longitudinal data have been developed, among others, by James *et al.* [40], Müller [58] and Yao *et al.* [76]. FPCA was considered by Li and Chiou [49], to determine the number of clusters in the problem of functional data clustering. More recent work on estimation of the principal components and the covariance function includes Hall and Hosseini-Nasab [31], Hall *et al.* [32] and Yao and Lee [75].

For the set of multivariate curves, a researcher may be interested in optimal representation of curves in a functional space of reduced dimension. Multivariate Functional Principal Component Analysis (MFPCA) is a useful statistical technique for understanding the structure of multivariate functional data (MFD). MFPCA is an effective dimension reduction tool for MFD. Ramsay and Silverman [64] have given an example of bivariate functional data, $X(t) = (X_1(t), X_2(t))' \in \Re^2$, as a model for gait data (knee and hip measures) used in the context of multivariate functional principal component analysis (FPCA) as an extension of the univariate case. Principal components in MFPCA have the same interpretation as in the functional univariate case. MFPCA aims to explain the covariance structure of data by means of small number of functional components. These functional components are linear combinations of the original variables. This gives better interpretation of the different sources of variation.

In literature, very limited work has been done on multivariate functional data. To deal with multivariate functional data Ramsay and Silverman [64] proposed to concatenate the observations of the functions on a fine grid of points into a single vector and then to perform

FPCA for the concatenated functions. The final results of the Ramsay and Silverman [64] are scalar whereas, the method proposed by Berrendero *et al.* [3] summarizes the curves with functions. Berrendero *et al.* [3] built functional principal components by carrying out classical multivariate PCA for each value of the domain on which the functions are observed and suggested the integrated variance as a suitable criterion. Kayano *et al.* [44] proposed regularized functional principal component procedure based on Gaussian basis functions for multivariate functional data. Jacques and Preda [42] considered MFPCA that take into account the possible use of non orthonormal basis and use of different basis for each dimension of the multivariate curves.

Section 2.2 describes classical principal component analysis in which the principal components are linear combinations of the variables that represent the most significant modes of variation in the data. The weights of these linear combinations are obtained by solving an optimization problem that can be expressed in terms of the eigenvalues and eigenvectors of the covariance matrix with constraints involving the Euclidean norm of the vector of weights. The natural extension of classical PCA to functional data is to replace the Euclidean norm by the L_2-norm and the covariance matrix by the covariance function of the process generating the data are described in Section 2.3. This general approach is described by Ramsay and Silverman [64] and Ferraty and Vieu [24]. In Section 2.4, MFPCA is described which is an extension of the univariate FPCA.

2.2 Classical Principal Component Analysis (CPCA)

Principal Component Analysis (PCA) is a useful tool for data reduction, which is achieved by identifying main modes of variability of a given dataset and is used for understanding the structure of a multivariate dataset.

In multivariate data, the central notion is to find weight vectors $\gamma_j \in \Re^p$ for which a linear combination of centered variables

$$z_i = \sum_{j=1}^{p} \gamma_j x_{ij} = \gamma' x_i \qquad i = 1, \ldots, n \tag{2.1}$$

that have maximal variance subject to constraints $\gamma'_m \gamma_r = \mathcal{I}(m = r)$ for $m < r$, where $\gamma = [\gamma_1, \ldots, \gamma_p]'$ and $x_i = [x_{i1}, \ldots, x_{ip}]'$. The solution is obtained by the means of spectral decomposition of the variance-covariance matrix [43, 63].

2.3 Univariate Functional Principal Component Analysis

When the dataset is in the form of a curve, the procedure for classical PCA can be generalized to functional principal component analysis to obtain main modes of variability for the curves. Instead of variable values x_{ij}, used in PCA, functional values $x_i(t)$ are used in FPCA, so that the discrete index j in the multivariate context is replaced by continuous index t. Unlike multivariate PCA, components in functional PCA are functions rather than vectors. So summations over j are replaced by integrations over t.

Let $\{X(t), t \in T\}$ be a stochastic process where T is some index set which is a bounded interval on $\Re$. The principal component scores corresponding to weight γ is generalized to an integral form,

$$z_i = \int \gamma_j(t) x_i(t) dt. \tag{2.2}$$

The weight function $\gamma_j(t)$ is obtained by solving

$$\max_{\langle \gamma_{\mathbf{j}}, \gamma_{\mathbf{m}} \rangle = \mathcal{I}(\mathbf{j}=\mathbf{m}),\, \mathbf{j} \leq \mathbf{m}} n^{-1} \sum (\int \gamma_j x_i)^2 \tag{2.3}$$

or equivalent to solving the functional eigenequation

$$\int \psi(s,t)\gamma(t)dt = \lambda\gamma(s), \qquad \gamma \in L_2, \tag{2.4}$$

where $\psi(s,t)$ is the covariance function of the $x(t)$. The sequence of eigenfunctions γ_i, $i = 1, 2, \ldots$, sorted with respect to the corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$ solves the FPCA problem 2.3. The eigenequation is the same general equation as in PCA, except here γ is now an eigenfunction rather than an eigenvector. There is a major difference between the multivariate and functional eigenanalysis. In multivariate case the eigenvalue-eigenfunction pairs are p (number of variables) whereas, in functional case they are infinite (number of functional values). In practice, the unknown covariance function ψ needs to be estimated by the sample values $x_i(t)$, $1 \leq i \leq n$, where for each i, $x_i(t)$ is observed on a discrete set of points $t = \{t_1, \ldots, t_q\}$ for finite q.

FPCA problem can be represented in terms of basis function approach. In which, first k bases functions in a basis $\{\phi_1, \ldots, \phi_k\}$ are used, where k is large enough, so that these functions will be able to describe most of the features of the data. The bases are selected based on the nature of the data; for example if the data are smooth and periodic then a Fourier basis might be ideal and for data that have a lot of local features then B-splines might work better. We approximate each x_i by:

$$\hat{x}_i(t) = \sum_{K=1}^{k} c_{iK}\phi_K(t). \tag{2.5}$$

We can express all n curves simultaneously by defining the vector-valued function x to have components $x_1, x_2, \ldots, x_n$ and the vector valued function ϕ to have components $\phi_1, \ldots, \phi_k$ as:

$$x = C\phi, \tag{2.6}$$

where the coefficient matrix C is $n \times k$. In matrix terms, the variance-covariance function, $\psi(s,t)$, is:

$$\psi(s,t) = n^{-1}\phi(s)'C'C\phi(t). \tag{2.7}$$

Let W be a symmetric matrix of order k given as

$$W = \int \phi\phi'. \tag{2.8}$$

Suppose that the weight function γ has the expansion

$$\gamma(s) = \sum b_K \phi_K(s) \tag{2.9}$$

and in matrix notation, $\gamma(s) = \phi(s)'b$. Using equations (2.6-2.9) the left hand side of eigen equation in 2.4 becomes

$$\int \psi(s,t)\gamma(t)dt = \int n^{-1}\phi(s)'C'C\phi(t)\phi(t)'bdt$$

$$= \phi(s)'n^{-1}C'CW'b.$$

The eigenequation can be written as:

$$\phi(s)'n^{-1}C'CWb = \lambda\phi(s)'b. \tag{2.10}$$

As this equation holds true for all s, it can be written in matrix form in the following manner:

$$n^{-1}C'CWb = \lambda b. \tag{2.11}$$

As $\| \gamma \| = 1$ implies $b'Wb = 1$ and similarly, two functions γ_1 and γ_2 will be orthogonal if and only if the corresponding vectors of coefficients satisfy $b_1'Wb_2 = 0$. We define $u = W^{1/2}b$ to get the required principal components by solving equivalent symmetric eigenvalue problem

$$n^{-1}W^{1/2}C'CW^{1/2}u = \lambda u \tag{2.12}$$

and compute $b = W^{-1/2}u$ for each eigenvector. If the basis is orthonormal then $W = I$. The functional PCA problem reduces to the standard multivariate PCA of the coefficient array C.

2.4 Multivariate Functional Principal Component Analysis

Consider data represented by curves is a stochastic process with continuous time, $X = \{X(t)\}_{t\in[0,T]}$ with $X(t) = (X_1(t), \ldots, X_p(t))' \in \Re^p$, $p \geq 2$. Let $x_1, \ldots, x_n$, with $x_i = (x_{i1}, \ldots, x_{ip})$, be the observation of the sample $X_1, \ldots, X_n$. The i^{th} discrete data set observed at t_{ij} for X_l is represented by $(t_{ij}, x_{ilj}); j = 1, \ldots, n_i$.

It is assumed that each discrete data $\{(t_{ij}, x_{ilj}); j = 1, \ldots, n_i\}$ is generated from the nonlinear regression model

$$y_{ilj} = x_{il}(t_{ij}) + \epsilon_{ilj}, \qquad j = 1, \ldots, n_i, i = 1, \ldots, n, l = 1, \ldots, p,$$

where the errors ϵ_{ilj} are independently normally distributed with mean 0 and variance σ_{il}^2. Each discrete data set $(t_{ij}, x_{ilj}); j = 1, \ldots, n_i$ is converted to functional data by using a smoothing method. The nonlinear functions x_{il}(t) are assumed to be expressed as a linear combinations of basis functions ϕ_{lK}(t) $(K = 1, \ldots, k_l)$. The estimated nonlinear functions for each i and l are given by

$$\hat{x}_{il}(t) = \sum_{K=1}^{k_l} c_{ilK}\phi_K^l(t), \qquad t \in [t_{min}, t_{max}], \tag{2.13}$$

and its matrix formulation is

$$x_i(t) = \Phi c_i',$$

where $c_i = (c_{i11}, \ldots, c_{i1k_1}, c_{i21}, \ldots, c_{i2k_2}, \ldots, c_{ip1}, \ldots, c_{ipk_p},)$ being the vector of the basis expansion coefficients, $x_i(t) = (x_{i1}(t), \ldots, x_{ip}(t))$, $i = 1, \ldots, n$ and

$$\Phi(t) = \begin{pmatrix} \phi_1^1(t) & \cdots & \phi_{k_1}^1(t) & 0 & & & \cdots & & & 0 \\ 0 & \cdots & 0 & \phi_1^2(t) & \cdots & \phi_{k_2}^2(t) & 0 & \cdots & & 0 \\ & & & & & & & & & \\ 0 & & & \cdots & & & 0 & \phi^p(t)_1 & \cdots & \phi_{k_p}^p(t) \end{pmatrix}_{(p\times k),}$$

where $k = \sum_{l=1}^{p} k_l$.

Let $Q(t) = (X_1(t), \ldots, X_p(t))' \in \Re^p$, $p \geq 2$ and $\tilde{C}$ be the $n \times \sum_{l=1}^{k_l}$ - matrix, whose rows are the vectors $c_{i.}$. Using these notations, we have

$$Q(t)_{(n\times p)} = \tilde{C}_{(n\times \sum k_l)} \Phi'(t)_{(\sum k_l \times p)}.$$

Under the basis expansion assumption in (2.13), the covariance estimator $\hat{\psi}$ of ψ, for all $s, t \in [t_{min}, t_{max}]$, is given by

$$\hat{\psi}(s,t) = \frac{1}{n}(Q(s) - \hat{\mu}'(s))'(Q(t) - \hat{\mu}'(t)) = \frac{1}{n}\Phi(s)C'C\Phi'(t), \tag{2.14}$$

where $Q(s) - \hat{\mu}'(s)$ means that $Q(s)$ is mean centered, $\hat{\mu}(s) = \frac{1}{n}\sum_{i=1}^{n} x_i(t)$ and $C = (I_n - 1_n(1/n, \ldots, 1/n))\tilde{C}$ where I_n and 1_n are the identity $n \times n$ -matrix and the unit column vector of size n, respectively.

It is assumed that the weight function γ^m can be expressed in terms of the same basis functions as the functional data sets $(x_{i1j}, \ldots, x_{ipj})$,

$$\gamma^m(t) = \Phi(t)b'^m, \quad m \geq 1, \tag{2.15}$$

with $b_m = (b_{m11}, \ldots, b_{m1k_1}, b_{j21}, \ldots, b_{m2k_2}, \ldots, b_{mp1}, \ldots, b_{mpk_p})$.

Let z_i be an inner product for a p-dimensional weight function $\gamma(t) = (\gamma_1(t), \ldots, \gamma_p(t))'$ $(t \in T)$ and i^{th} p-dimensional functional data $x_i(t) = (x_{i1}(t), \ldots, x_{ip}(t))'$,

$$z_i =< \gamma, x_i >_p = \sum_{l=1}^{p} < \gamma_l, x_{il} >= \sum_{l=1}^{p} \int_T \gamma_l(t)x_{il}(t)dt, \quad i = 1, \ldots, n.$$

The objective in principal component analysis of functional data is to estimate the principal component (PC) curves by the p-dimensional weight function $\gamma(t)$ that maximizes the sample variance of the inner products subject to the orthonormal constrains. The sample variance $var(z)$ of $\{z_i; i = 1, \ldots, n\}$ can be written as

$$var(z) = n^{-1} \sum_{i=1}^{n} z_i^2 = n^{-1} \sum_{i=1}^{n} (\sum_{l=1}^{p} \int_T \gamma_l(t) x_{il}(t) dt)^2.$$

In analogy with the multivariate case, the functional PCA problem leads to the maximum problem of the sample variance which is equivalent to solve the eigenvalue problem from the functional eigenequation system

$$V\gamma^m = \lambda^m \gamma^m, \tag{2.16}$$

where γ is a vector-function and the covariance operator V of X is an integral operator with kernel ψ.

$$V\gamma = \int_0^T \psi(\cdot, t)\gamma(t) dt,$$

and spectral analysis of V provides a countable set of positive eigenvalues λ^m, $m \geq 1$, associated with an orthonormal basis of eigenfunctions $\gamma^m = (\gamma_1^m, \ldots, \gamma_p^m)$.
Using the estimation $\hat{\psi}$ of ψ, the eigen problem in (2.16) becomes equivalent to

$$\int_T \hat{\psi}(s, t)\gamma^m(t) dt = \lambda^m \gamma^m(s). \tag{2.17}$$

Using (2.14) and (2.15) in (2.17), this becomes

$$\int_0^T \frac{1}{n} \Phi(s) C' C \Phi'(t) \gamma^{m\prime}(t) ds = \lambda^m \Phi(s) b^{m\prime}. \tag{2.18}$$

$$\Leftrightarrow \frac{1}{n}\Phi(s)C'C\int_0^T \Phi'(t)\Phi(t)dtb^{m\prime} = \lambda^m\Phi(s)b^{m\prime}, \tag{2.19}$$

where $W = \int_0^T \Phi'(t)\Phi(t)dt$ is $\sum_{l=1}^p k_l \times \sum_{l=1}^p k_l$ is the symmetric block-diagonal matrix of the inner products between the basis functions. Since (2.19) is true for all s, it can be written as

$$\frac{1}{n}C'CWb^{m\prime} = \lambda^m b^{m\prime}.$$

Let $u^m = b^m W^{1/2}$, the multivariate functional principal component analysis is reduced to the usual PCA of the matrix $CW^{1/2}$, then we obtain

$$\frac{1}{n}W^{1/2\prime}C'CW^{1/2}u^{m\prime} = \lambda^m u^{m\prime}.$$

The coefficient b^m, $m \geq 1$, of the eigenfunctions γ^m are obtained by $b^m = (W^{1/2\prime})^{-1}u^{m\prime}$ and the principal components scores, $z^m = \int_0^T \sum_{l=1}^p (Q_l(t) - \mu_l(t))\gamma_l^m(t))dt$, are given by

$$z^m = CWb^{m\prime}, \quad m \geq 1.$$

Let $\lambda^1 \geq \lambda^2, \ldots$ be the eigenvalues of equation 2.4 and $\gamma^1, \gamma^2, \ldots$ be the orthonormal eigenvectors corresponding to the eigenvalues $\lambda^1, \lambda^2, \ldots,$ respectively. First eigenfunctions $\gamma^1(t) = (\gamma_1^1(t), \ldots, \gamma_p^1(t))'$ the p-dimensional weight function $\gamma^1(t)$ that maximizes sample variance subject to $\| \gamma^1 \|_p^2 = 1$. Second eigenfunctions $\gamma^2(t) = (\gamma_1^2(t), \ldots, \gamma_p^2(t))'$ the p-dimensional weight function $\gamma^2(t)$ that maximizes sample variance subject to $\| \gamma^2 \|_p^2 = 1$ and $< \gamma^1, \gamma^2 >= 0$. The other eigenfunctions are computed likewise, $m(\geq 3)th$ eigenfunctions $\gamma^m(t) = (\gamma_1^m(t), \ldots, \gamma_p^m(t))'$ the p-dimensional weight function $\gamma^m(t)$ that maximizes sample variance subject to $\| \gamma^m \|_p^2 = 1$ and $< \gamma^m, \gamma^r >= \sum_{l=1}^p < \gamma_l^m, \gamma_l^r >= \sum_{l=1}^p \int_T \gamma_l^m(t)\gamma_l^r(t)dt = 0$

$(r \leq m)$. The m^{th} PC score is defined by

$$z_i^m =< \gamma^m, xi >_p = \sum_{l=1}^{p} < \gamma_l^m, x_{il} >= \sum_{l=1}^{p} \int_T \gamma_l^m(t) x_{il}(t) dt, \qquad i = 1, \ldots, n,$$

where $m = 1, \ldots, M$ where M is the number of PCs for which the cumulative contribution rate is over 90%. The first few M PCs contain almost complete information about individual variations. The p-dimensional m^{th} PC curves $\gamma^m(t)$ and PC scores $\{z_i^m =< \gamma^m, x_i >_p; i = 1, \ldots, n\}$ can then be obtained.

In practice, to deal with multivariate functional data Ramsay and Silverman [64] carried out the calculation for functional PCA by replacing each function $x_{il}; l = 1, \ldots, p$ with a vector of values at a discrete grid of points. For each i these vectors are concatenated into single long vector x_i. The covariance matrix of x_i is a discretized version of operator V as defined in (2.16). PCA is performed on the vectors x_i and the results of principal component vectors corresponding to each variable are separated. The analysis is completed by applying a suitable inverse transformation to each of these parts if necessary.

In this chapter, we examined univariate FPCA and MFPCA as a dimension reduction tool for functional data (FD) and MFD. Although, FPCA and MFPCA solves dimensionality problem, it fails to deal with data containing outliers since covariance function is sensitive to outliers. In the next Chapters 3 and 4 robust methods are discussed to overcome this problem.

Chapter 3

Robust Univariate Functional Principal Component Analysis

3.1 Introduction

The notion behind dimension reduction methods is to transform original variables into new few variables so that new variables contain most of the information of the original variables. Functional principal component analysis (FPCA) is a useful tool to reduce the dimension of the functional data.

In presence of outlying observations the classical FPCA is unreliable resulting in inaccurate principal function estimates. Since it is based on the empirical covariance function. Functional outlier can be defined in different ways. For instance, a shape functional outlier is a curve with a different pattern from the other curves, e.g. very irregular in a set of smooth curves or increasing when the remaining ones are decreasing; a magnitude functional outlier is a curve which is very distant from the mean. To deal with outlying observations researchers have proposed many robust methods for functional PCA. Locantore *et al.* [50] described a robust procedure to deal complex data sets of human eye images. Hyndman and Ullah [38] applied robust projection pursuit approach to smoothed trajectories. Sawant *et al.* [71] proposed a robust approach of principal components based on a robust eigenanalysis of the coefficients of the observed data. All of these approaches reduce the functional problem to the multivariate by using the coefficients of a basis expansion. The smoothing step may incur smoothing bias. However, this smoothing bias may not be ignorable unless the number of trajectories are larger than the size of the grid (Zhang and Chen [78]). Hubert *et al.* [35, 36, 37] proposed ROBPCA is based on the minimum covariance determinant (MCD) estimator [67, 68] of multivariate location vector and scatter matrix. Billor *et al.* [5] proposed BACONPCA is based on the estimator of the location vector and scatter matrix

obtained from the basic subset which is found algorithmically by utilizing BACON approach [4].

Gervini [26] considered a functional version of the estimators for spherical PCA. Although the spherical PCA is resistant for any contamination model having elliptical properties, it is susceptible to other types of contamination (Boente and Fraiman [7], Maronna [54]). Gervini [27] proposed robust functional principal component estimators for sparsely and irregularly observed functional data based on model approach and basis expansion. It uses multivariate t-distribution to downweigh outlying observations. This approach is found to be attractive as it deals with sparsely observed data in the longitudinal studies. Since this approach employs EM algorithm it is computationally demanding thus, limiting its usage for typical functional data. For sparsely and irregularly observed functional data Gervini [27] developed robust functional principal component estimators. Hyndman and Ullah [38] considered robust estimators of the functional principal directions using a projection-pursuit approach. Recently, Bali et al. [2] considered robust estimators of the functional principal components by using a fully functional robust projection pursuit approach and derived their consistency and qualitative robustness. Recently, Lee *et al.* [46] proposed M-type smoothing spline estimators for principal functions. This method is efficient since it makes maximal use of the normally observed measurements by separately downweighing abnormally observed measurements in a single curve.

In this chapter, we review previous two approaches for outlier detection methods via robust functional PCA for univariate functional data. The robust FPCA method is to obtain functional principal components that are less influenced by outliers. Robust FPCA method and outlier detection method proposed by Sawant *et al.* [71] and Lee *et al.* [46] are described in Section 3.2 and Section 3.3, respectively.

3.2 Robust Functional Principal Component Analysis Based on Multivariate Approach

Outlier detection procedure in functional data using robust functional principal component analysis is developed by Sawant *et al.* [71], which is based on multivariate principal component analysis. Let $x_i(t) \in L_2[T]$, $i = 1, \ldots, n$, be a function measured over a continuous variable $T \in [t_{min}, t_{max}]$. These n curves are measured on a fine grid of q equally spaced points $t = \{t_1, t_2, \ldots, t_q\} \in T$ in the following manner:

$$y_i = x_i(t) + \epsilon_i, \qquad 1 \leq i \leq n,$$

subject to random error ϵ_i with zero mean and constant variance function $\sigma_i^2(t)$.

The function x_i is represented as a linear combination of the first k orthonormal basis functions ϕ_K, $K = 1, \ldots, k$, where k is large enough, $k < q$ using basis expansion method as:

$$x_i(t) = \sum_{K=1}^{k} c_{iK}\phi_K(t),$$

where ϕ is vector-valued function having components $\phi_1, \ldots, \phi_k$. The C is n x k coefficient matrix of the expansion, where $C = [c_{iK}]$, $1 \leq i \leq n$, $1 \leq K \leq k$.

The optimal number of basis functions, k, is obtained by GCV developed by Craven and Wahba [15], is used and is given in Subsection 4.2.2. On observed functions the coefficients c_{iK}, for $i = 1, \ldots, n$ and $K = 1, \ldots, k$ are computed by using the least squares approach. Due to orthonormal basis function the functional PCA problem reduces to the standard multivariate PCA of the coefficient array C (see Section 2.3 and Ramsay and Silverman [64]).

Robust PCA method such as ROBPCA [37] or BACONPCA [5] is applied on C and orthogonal-score plot [37] is used for identification of outliers. After using robust PCA method robust scores are obtained as $Z = C \times V$, where Z is $n \times k_1$ matrix, C is $n \times k$ matrix of the coefficients, V is $k \times k_1$ robust eigenvectors and $k_1 \leq k$. The optimal number

of components k_1 is the minimal value for which the cumulative percentage of total variance is greater than or equal to 90%. Robust coefficients are obtained by transforming the data back to $\Re^k$ as:$\hat{C} = Z \times V^T$ (Sawant *et al.* [71]).

3.3 Robust Functional Principal Component Analysis

Robust functional PCA developed by Lee *et al.* [46] based on M-estimation with roughness penalty is discussed in this section. Consider a collection of n sample curves recorded on a discrete grid of sampling points $t_1, \ldots, t_q$, which are not equally spaced. Denote the underlying function for the ith sample curve by $x_i(t)$, $i = 1, \ldots, n$. All observations x_{ij}, which is a measurement for subject i at time t_j, are arranged into a single data matrix X.

The first SVD-layer gives the best rank-one approximation of x with respect to the squared Frobenius norm, where $\| \ . \ \|_F^2$ indicates the squared Frobenius norm, which is the sum of squared elements of the matrix. Instead of using SVD, principal components can be achieved by the unregularized LS criterion for rank-one approximation as given in the following optimization problem

$$min_{z,\gamma} \parallel x - z\gamma^T \parallel_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{q}(x_{ij} - z_i\gamma_j)^2, \tag{3.1}$$

where $z = (z_1, \ldots, z_n)^T$ and $\gamma = (\gamma_1, \ldots, \gamma_q)^T$ and any rank one matrix of size $n \times q$ can be written as $z\gamma^T$ with $\gamma^T\gamma = 1$.

To promote two objectives, robustness and smoothness on principal components, criteria (3.1) is modified and some regularization penalty and robust regression technique are employed and the modified criterion for finding γ is given by

$$Q(\gamma) = \sum_{i=1}^{n}\sum_{j=1}^{q}\hat{\sigma}_j^2\rho(\frac{x_{ij} - z_i\gamma(t_j)}{\hat{\sigma}_j}) + P_\lambda(\gamma). \tag{3.2}$$

Here $P_\lambda(\gamma)$ is a roughness penalty on γ and $\rho(.)$ is a bounded loss function, such as biweight function. The error, $x_{ij} - z_i\gamma_j$, is the orthogonal distance between data point and projected point and $\hat{\sigma}_j$ is a preliminary scale estimate of errors and λ is a smoothing parameter.

The data matrix X is mean-centered by the functional 10% trimmed mean suggested by Fraiman and Muniz [21]. For the robust estimation of γ and making the resulting estimator of γ scale invariant the $\hat{\sigma}_j$ should be estimated robustly, e.g. the normalized median absolute deviation about the median.

The roughness penalty, $P_\lambda(\gamma)$, is a flexible and effective way to enhance the smoothness of principal function and is given by:

$$P_\lambda(\gamma) = \frac{\lambda}{2}\int \gamma''(t)^2 dt = \frac{\lambda}{2}\gamma^T\Omega\gamma,$$

where $\lambda > 0$ is a smoothing parameter controls the trade-off between residual error and local variation. Here the robust cross validation for a choice of λ is used which is provided in Subsection 4.3.1. By the property of γ to be a natural cubic spline (NCS), the $q \times q$ penalty matrix $\Omega = QR^{-1}Q^T$ consists of two banded matrices $Q_{q\times(q-2)}$ and $R_{(q-2)\times(q-2)}$, which help solving the equation fast. The form of biweight function is:

$$\rho(x) = \begin{cases} 1 - \{1 - (x/k)^2\}^3 & \text{if} \quad |x| \le k \\ 1 & \text{if} \quad |x| > k, \end{cases} \tag{3.3}$$

where k is a tuning constant and the 95% efficiency is obtained with the tuning constant $k = 4.68$ [55]. Let $\psi(.) = \rho'$ be the derivative of ρ. Differentiating the objective function in equation (3.2) with respect to γ produces estimating equations as: $S + \lambda\Omega\gamma = 0$, where $S = [S_1, \ldots, S_q]^T$ and $S_j = -\sum_{i=1}^n \psi(\frac{x_{ij}-z_i\gamma_j}{\hat{\sigma}_j})z_i\hat{\sigma}_j)$. Define the weight function $\omega_{ij} = \psi(e_{ij})/e_{ij}$. Then the estimating equations may be written as:

$$Z + \lambda\Omega\gamma = 0, \tag{3.4}$$

where $Z = [Z_1, \ldots, Z_q]^T$ and $Z_j = -\sum_{i=1}^{n} \omega_{iq} z_i (x_{iq} - z_i \gamma_q)$. The weights ω_{ij} depend upon the residuals, the residuals depend upon the estimated γ, and the estimated γ depends upon the weights. So the weights ω_{ij} are updated by using the previous estimates. The robust loss function downweighs large residuals by providing small weights for outliers and so severe outliers has no impact on γ estimation. The weight ω_{ij} is applied to a single point x_{ij} rather than to whole curve. Thus, unusual measurements partially observed inside a single curve receive low weights and are controlled and only other points on the same curve properly contribute to the estimation of principal functions. This enables to make use of maximal information on all trajectories except for outlying measurements locally appearing on curves.

Let $b = (\sum_i^n \omega_{i1} z_i x_{i1}, \ldots, \sum_i^n \omega_{iq} z_i x_{iq})^T$ and $M = diag(\sum_i^n \omega_{ij} z_i^2)_{j=1,\ldots,q}$. By using $\Omega = QR^{-1}Q^T$ and solution obtained by solving the estimating equation (3.2) is:

$$b = (M + \lambda QR^{-1}Q^T)\gamma. \tag{3.5}$$

Finding γ from the equation (3.5) will need $O(q^3)$ arithmetic operators due to $q \times q$ matrix inversion. Solving the equation (3.5) for large q becomes a computation burden. Due to NCS formulation of γ and by using banded matrix properties the equation (3.5) can be solved in $O(q)$ arithmetic operations.

Let $\upsilon = (\upsilon_2, \ldots, \upsilon_{q-2})^T$ be a $(q-2)$ vector evaluated on time points $t_2, \ldots, t_{q-1}$, where $\upsilon_j = \gamma''(t_j)$. From Green and Silverman [29] a NCS with knots $t_1, \ldots, t_q$ satisfies:

$$Q^T\gamma = R\upsilon. \tag{3.6}$$

Using the equation (3.6), equation (3.5) can be rearranged as:

$$M\gamma = b - \lambda QR^{-1}Q^T\gamma = b - \lambda Q\upsilon. \tag{3.7}$$

By pre-multiplying $Q^T M^{-1}$ on the both sides and using (3.6) the equation (3.7) is rewritten as:

$$(R + \lambda Q^T M^{-1} Q)\upsilon = Q^T M^{-1} b.$$

The matrix $(R+\lambda Q^T M^{-1} Q)$ has bandwidth 5, it has a Cholesky decomposition of the form:

$$R + \lambda Q^T M^{-1} Q = LDL^T, \tag{3.8}$$

where D is a strictly positive diagonal matrix and L is a lower triangular band matrix. Matrices D and L are calculated by Cholesky decomposition of $B = R + \lambda Q^T M^{-1} Q$. By using the Cholesky decomposition υ is obtained without matrix inversion and then $\gamma = (\gamma_1, \ldots, \gamma_q)^T$ is computed elementwise using $\gamma_j = \{b - \lambda Q\upsilon\}_j / M_j$, as in (3.7), where $\{.\}_j$ is the jth element of vector in argument and M_j is the jth diagonal element of M. Finding γ by this algorithm is computationally faster than invoking the inversion to get $\gamma = (M+\lambda Q)^{-1} b$, especially when q is large.

The next step is to normalize γ by $\gamma \leftarrow \gamma / \parallel \gamma \parallel$ and then estimate $z = x\gamma$, which is the orthogonal projection of data matrix x onto the principal function γ. Two estimation steps for γ and z are performed in an alternating fashion, starting with an initial value of γ estimated from the existing robust multivariate PCA methods. In the estimation procedure for robust principal function, the principal function γ and the associated PC scores z are alternately estimated in iterative procedure as σ_j keeps updating. After obtaining γ and z, the subsequent principal components are acquired from the rank-1 deflated matrix $x - z\gamma^T$.

Algorithm for Robust FPCA

Input: $x^*_{n\times q} = (x^*_{ij})$ containing n curves observed at q time points.

Step 1: Compute the functional trimmed mean $m(t)$.

Step 2: Set $x_{n\times q} = (x_{ij})$ with $x_{ij} = x^*_{ij} - m(t_j)$.

Step 3: For $k = 1, 2, \ldots, K$, run the following steps for a set of grid points of λ.

1: Set an initial value for γ.

2: Update the PC score $z = x\gamma$.

3: Compute the necessary quantities: $r_{ij} = x_{ij} - z_j\gamma_j$, M-scale estimate $\hat{\sigma}_j$, $e_{ij} = r_{ij}/\hat{\sigma}_j$, and $w_{ij} = \psi(e_{ij})/e_{ij}$.

4: Compute $b = (\sum_i^n \omega_{i1} z_i x_{i1}, \ldots, \sum_i^n \omega_{iq} z_i x_{iq})^T$ and $M = diag(\sum_i^n \omega_{ij} z_i^2)_{j=1,\ldots,q}$.

5: Compute υ and $\gamma = (\gamma_j)_{j=1,\ldots,p}$ with $\gamma_j = \{b - \lambda Q\upsilon\}_j/M_j$. Then normalize it by $\gamma \leftarrow \gamma/\|\gamma\|$.

6: Repeat *1–5* until the convergence is met. At the convergence, calculate $CV(\lambda)$.

Step 4: Find z and γ corresponding to λ which achieves the minimum CV(λ) and set $z_k \leftarrow z$ and $\gamma^k \leftarrow \gamma$. Replace $x \leftarrow x - z\gamma^T$ and go back to the *Step 3* with the deflated x. Repeat the same procedure until the desired number, say K, of principal components are acquired.

Output: Robust PC functions, $\gamma = (\gamma^1, \ldots, \gamma^K)$ and PC scores, $z = (z_1, \ldots, z_K)$.

Chapter 4

Robust Multivariate Functional Principal Component Analysis

4.1 Introduction

For the set of multivariate curves, one may be interested in optimal representation of curves in a functional space of reduced dimension. Multivariate Functional Principal Component Analysis (MFPCA) is a useful statistical technique for understanding the structure of multivariate functional data (MFD). They are effective dimension reduction tools for MFD. Principal components from MFPCA have the same interpretation as in the functional univariate case. MFPCA aims to explain the covariance structure of data by means of small number of functional components. These functional components are linear combinations of the original variables. This gives better interpretation of the different sources of variation. Thus effectiveness of MFPCA in data reduction is useful in analysis of high dimensional data.

In literature very limited work has been done on multivariate FPCA for multivariate functional data. Ramsay and Silverman [64] and Berrendero *et al.* [3] have suggested principal component analysis for multivariate functional data. Kayano and Sadanori [44] proposed regularized functional principal component procedure based on Gaussian basis functions for multivariate functional data. Jacques and Preda [42] considered MFPCA that takes into account the possible use of non orthonormal basis and use of different basis for each dimension of the multivariate curves.

In the presence of outliers, dimension reduction via MFPCA would yield untrustworthy results since MFPCA is known to be sensitive to outliers. Although several robust FPCA methods have been proposed for univariate functional data, to our knowledge, there has been no study on the robustness of MFPCA for multivariate functional data. This necessitates

to develop robust MFPCA for multivariate FD. The main contribution of our work is to construct a robust MFPCA method to achieve dimension reduction of data and to develop tools for detection of functional outliers.

In this chapter, we introduce two approaches for outlier detection methods via robust functional PCA for multivariate functional data. The robust MFPCA method is to obtain functional principal components that are less influenced by outliers. Section 4.2 describes robust MFPCA method based on multivariate approach. In Section 4.3, we extend univariate robust FPCA based on MM estimation to multivariate FD. Simulation study is conducted to investigate the effectiveness of the both proposed robust multivariate functional PCA based on multivariate and MM estimation in Sections 4.4.2 and 4.4.3, respectively.

4.2 Proposed Method I: Robust Multivariate Functional Principal Component Analysis Based on Multivariate Approach (RMFPCA1)

Suppose n independent discrete observations $\{t_{ij}, (x_{i1j}, \ldots, x_{ipj});\ j = 1, \ldots, q\}$, $(i = 1, \ldots, n)$, where each t_{ij} is finite and discrete for p variables $X_1, \ldots, X_p$.

Each discrete data set $\{(t_{ij}, x_{ilj}; j = 1, \ldots, q)\}$ is converted to functional data $\hat{x}_{il}(t)$ by using a smoothing method in following manner. The function x_{il} can be represented as a linear combination of the first k_l basis functions ϕ_{lK}, $K = 1, \ldots, k_l$, where k_l is large enough, $k_l < q$ using basis expansion method as:

$$\hat{x}_{il}(t) = \sum_{K=1}^{k_l} c_{ilK}\phi_K^l(t),$$

where ϕ^l is vector-valued function having components $\phi_1^l, \ldots, \phi_{k_l}^l$. The C_l is n x k_l coefficient matrix of the expansion, where $C_l = [c_{ilK}]$, $1 \leq i \leq n$, $1 \leq l \leq p$ $1 \leq K \leq k_l$. The simultaneous expansion of all n curves can be expressed in matrix notation as

$$x_i(t) = \Phi c_i',$$

where $c_i = (c_{i11}, \ldots, c_{i1k_1}, c_{i21}, \ldots, c_{i2k_2}, \ldots, c_{ip1}, \ldots, c_{ipk_p},)$ being the vector of the basis expansion coefficients and

$$\Phi(t) = \begin{pmatrix} \phi_1^1(t) & \cdots & \phi_{k_1}^1(t) & 0 & & & \cdots & & & 0 \\ 0 & \cdots & 0 & \phi_1^2(t) & \cdots & \phi_{k_2}^2(t) & 0 & \cdots & & 0 \\ & & & & & & & & & \\ 0 & & & \cdots & & & 0 & \phi^p(t)_1 & \cdots & \phi_{k_p}^p(t) \end{pmatrix}_{(p \times k)},$$

where $k = \sum_{l=1}^{p} k_l$. To select optimal number of basis functions, k_l, GCV developed by Craven and Wahba [15], is used and is given in Subsection 4.2.2. On partially observed functions the coefficients $C_l = c_{ilK}$ are computed by using the least squares approach, for $i = 1, \ldots, n$, $l = 1, \ldots, p$ and $K = 1, \ldots, k_l$. For multivariate functional data Ramsay

and Silverman [64] suggested to concatenate the coefficients in a suitable basis expansion into a single long vector for each observation and then perform PCA on $CW^{1/2}$, where $W_{k\times k} = \int_0^T \Phi(t)'\Phi(t)dt$ is the symmetric block-diagonal matrix of the inner products between the basis functions and $C = [C_1 C_2 \dots C_p]$ (Section 2.4). C is a $n \times k$, where $k = \sum_{l=1}^{p} k_l$, matrix obtained by concatenating coefficient matrix for each variable. Since we deal with basis function that is orthonormal the multivariate functional PCA problem reduces to the standard multivariate PCA of the coefficient matrix C. Therefore in our procedure instead of dealing with classical PCA on $CW^{1/2}$ we apply robust PCA on $CW^{1/2}$.

Applying robust PCA method like ROBPCA [37] or BACONPCA [5] on C would provide the equivalent information about the structure of the covariance function of functional data $x(t)$. Outliers in C will be equivalent to the outliers in functional data $x(t)$. Therefore, the diagnostic plots developed to detect outliers by using multivariate PCA method can also be used to detect functional outliers. Orthogonal-score plot [37], which is a by-product the robust PCA method is used for identification of outliers. By using ROBPCA [37] or BACONPCA [5] we obtain robust scores z in the following manner:

$$z = C \times \gamma',$$

where z is $n \times k^1$ matrix, C is $n \times k$ matrix of the coefficients, γ is $k \times k^1$ robust eigenvectors and $k^1 \leq k$. The selection criteria to choose the components k^1 is based on the eigenvalues. The predetermined threshold value is 90%. The optimal number of components k^1 is the minimal value for which the cumulative percentage of total variance is greater than or equal to 90%.

4.2.1 Diagnostic Plot for Detection of Outliers

The diagnostic plot developed to detect outliers by using PCA method for multivariate data can be used to detect functional outliers for multivariate functional data. Orthogonal

score plot proposed by Hubert *et al.* [37] is used to distinguish between regular observations and the outliers. This diagnostic plot is a scatter plot of the orthogonal distance Od_i versus the robust score distance Sd_i. The score distance is defined as

$$Sd_i = \sqrt{\sum_{K=1}^{k} (z_i^K)^2/\lambda^K}, \quad i = 1, \ldots, n,$$

The orthogonal distance which measures the distance between an observation x_i and its projection in the k-dimensional PCA-subspace, Od_i, is given by

$$Od_i = \| \, x_i - m - \sum_{K=1}^{k} z_i^K \gamma^K \, \|, \quad i = 1, \ldots, n,$$

where z_i^K is the i^{th} score for K^{th} component and λ^K is the eigenvalue for K^{th} component and m is the mean function and γ^K is the K^{th} eigenfunction.

If Sd_i and Od_i are both large, then the i^{th} observation is far away from the homogeneous observations (right corner of the plot) is identified as functional outliers. Observations having large Sd_i or Od_i and well separated from the homogeneous observations are also identified as a functional outliers.

Two cutoff lines as in Hubert *et al.* [37] are used to separate functional outliers from the regular observations. Observations found in right top corner or observations separated from regular observations are identified as outliers.

4.2.2 Selecting Number of Basis

Selecting optimal number of bases, k_l, is important because if k_l is too large it may introduce small variation with large bias and if k_l is too small then we may miss some aspects of smooth function x_l that we want to estimate. This will also introduce less bias with large variance. To choose the appropriate number of basis functions a popular measure in the smoothing methods known as generalized cross validation (GCV) developed by Craven

and Wahba [15] is used. This criterion is defined as:

$$k_l = \arg\min_{\mathbf{j}}(GCV(j)),$$

where

$$GCV(j) = \frac{n \times SSE}{(n-j)^2}, \quad j = 3, \ldots, q-1,$$

$$SSE = \sum_{i=1}^{n}(x_{il} - \hat{x_{il}})^2, \quad \hat{x_{il}} = \sum_{K=1}^{k_l} c_{ilK}\phi_K^l.$$

There is another technique, cross-validation (CV) based on minimizing mean squared error (MSE). Minimizing CV can lead to under-smoothing the data by introducing large variation. However, GCV has advantage over CV technique since it has less tendency to undersmooth the data.

The choice of number of bases relies on $\hat{x_{il}}$. The coefficients $C_l = [c_{ilK}]$ are computed by using least squares method and then $\hat{x_{il}}$ are estimated. Since, least squares method is sensitive to outliers, the choice of number of bases is also affected by outliers. Robust version of this criteria for selecting number of bases can be obtained by estimating the coefficients robustly.

4.3 Proposed Method II: Robust Multivariate Functional Principal Component Analysis (RMFPCA2)

The data $x = \{x(t)\}_{t\in[t_{min},t_{max}]}$ represented by a set of p curves with $x(t) = (x_1(t), \ldots, x_p(t))' \in \Re^p, \quad p \geq 2$. Saporta [70] shows that for multivariate functional data the Karhunen-Loeve expansion holds in following manner

$$x(t) = \mu(t) + \sum_{m\geq 1} z^m \gamma^m, \quad t \in [t_{min}, t_{max}],$$

where $\mu_l = \{\mu_l(t) = E[x_l(t)]\}_{t\in[t_{min},t_{max}]}$ denotes the mean function of x_l, $l = 1,\ldots,p$ and $\mu = (\mu_1,\ldots,\mu_p)'$, denotes the mean function of x. The covariance operator ν of x

$$\nu\gamma = \int_T V(\cdot,t)\gamma(t)dt,$$

is an integral operator with kernel V. For any $s,t \in [t_{min},t_{max}] = T$, $V(s,t)$ is a $p\times p$ matrix with elements

$$V(s,t)[j,l] = Cov(x_j(s),x_l(t)), \quad j,l = 1,\ldots,p.$$

The spectral analysis of V provides a countable set of positive eigenvalues $\{\lambda^m\}_{m\geq 1}$, associated to an orthonormal basis of eigenfunctions $\gamma^m = (\gamma_1^m,\ldots,\gamma_p^m)$ and are the solutions of

$$V\gamma^m = \lambda^m\gamma^m,$$

with $\lambda^1 \geq \lambda^2 \geq \ldots$ are eigenvalues and $\int_T \sum_{l=1}^p \gamma_l^m(t)\gamma_l^{m'}(t)dt = 1$ if $m = m'$ and 0 otherwise. The principal components z^m of x are zero-mean random variables defined as the projections of x on the eigenfunctions of V

$$z^m = \int_T \langle x(t)-\mu(t),\gamma^m(t)\rangle_{\Re^p}dt = \int_T \sum_{l=1}^p (x_l(t)-\mu_l(t))\gamma_l^m(t)dt.$$

For multivariate functional data, Ramsay and Silverman [64] suggested to concatenate the observations of the functions on a fine grid of points into a single vector and then perform PCA on these concatenated vector. In our procedure instead of dealing with coefficients we extend robust FPCA described in Section 3.3 on x, where $x = [x_1x_2\ldots x_p]$, $x_l = x_{ilj}$, $1\leq i\leq n$, $1\leq l\leq p$, $1\leq j\leq q$.

For enhancing the smoothness of principal function the roughness penalty is introduced and to achieve robust components robust regression technique is applied. The penalty term can be written in a quadratic form as $\int \gamma_l''(t)^2 dt$. The penalized sum of squares is given by:

$$PSS(\gamma) = \sum_{i=1}^{n}\sum_{j=1}^{pq} s_j^2 \rho(\frac{x_{ij} - z_i\gamma(t_j)}{s_j}) + \frac{\lambda}{2}\int \gamma''(t)^2 dt, \tag{4.1}$$

where $pq = p \times q$, $z = (z_1, \ldots, z_n)^T$, $\gamma = (\gamma_1, \ldots, \gamma_{(pq)})^T$ with $\gamma^T\gamma = 1$ and $\lambda > 0$ is a smoothing parameter controls the trade-off between residual error and local variation. In this situation the robust cross validation for a choice of λ is used which is provided in subsection 4.3.1.

For simplicity the data matrix x is mean-centered by the functional 10% trimmed mean suggested by Fraiman and Muniz [21]. For the robust estimation of γ a robust scale parameter s_j is employed and is given as $s_j = MAR/0.6745$ for all j, where $MAR = med_i(|r_{ij} - med_{i\prime}(r_{i'j})|)$ is the median absolute residual.

By the property of γ to be a NCS we have $\int \gamma''(t)^2 dt = \gamma^T\Omega\gamma$, where the $(pq) \times (pq)$ penalty matrix $\Omega = QR^{-1}Q^T$ consists of two matrices $Q_{pq\times(pq-2)}$ and $R_{(pq-2)\times(pq-2)}$ with bandwidth 5.

Compute γ by minimizing equation (4.1) (Lee *et al.* [46]). Details for calculation is given in Section 3.3. After obtaining γ, the next step is to normalize it by $\gamma \leftarrow \gamma/ \parallel \gamma \parallel$ and then estimate $z = x\gamma$, which is the orthogonal projection of data matrix x onto the principal function γ.

In the estimation procedure for robust principal function, the principal function γ and the associated PC scores z are alternately estimated in iterative procedure where s_j keeps updating, starting with an initial value of γ, which can be estimated from the existing robust multivariate PCA methods.

The second principal component and its scores are obtained by applying the same alternating procedure to the rank-1 subtracted data matrix $x^1 = x - z\gamma^T$. Similarly the

other subsequent principal components can be obtained sequentially by removing the effect of preceding pairs. We continue the procedure until the desired number, say K, of principal components is acquired where robust PC functions are $\gamma = (\gamma^1, \ldots, \gamma^K)$ and the corresponding PC scores are $z = (z^1, \ldots, z^K)$.

4.3.1 Robust leave-out-one-column cross validation

In the implementation of estimating algorithm in Sections 3.3 and 4.3, tuning parameter λ is obtained by the weighted leave-out-one-column cross validation (CV). This robust CV method (Lee *et al.* [46]) is modification of the leave-out-one-column CV method proposed by Huang *et al.* [34] under robust FPCA framework. CV based on row deletion involves actual computation of a large number of leave-out-one-row estimates and takes much longer time than the CV method based on column deletion. In robust CV method instead of deleting a single curve an observation at time t_j, the j^{th} column of x is removed for assessing prediction.

Let $y = vec(x) = (x_1^T, \ldots, x_{(pq)}^T)^T$, where y is an $n(pq)$-sized vector, with x_j, the j^{th} column of x, which corresponds to observations at time t_j, and $\hat{\gamma}^{(-j)} = (\hat{\gamma}^{1(-j)}, \ldots, \hat{\gamma}^{(pq)(-j)})^T$ be the solution of penalized sum of squares after removing x_j.

Define the leave-out-one-column CV as

$$CV(\lambda) = \frac{1}{(pq)} \sum_{j=1}^{(pq)} (z\hat{\gamma}_j^{(-j)} - x_j)^T W_j^{(-j)} (z\hat{\gamma}_j^{(-j)} - x_j), \tag{4.2}$$

where $W_j^{(-j)} = diag(w_{ij}^{(-j)})_{i=1,\ldots,n}$ and $w_{ij}^{(-j)} = \psi((z_i\hat{\gamma}_j^{(-j)} - x_{ij})/s_j)/((z_i\hat{\gamma}_j^{(-j)} - x_{ij})/s_j)$.
$W_j^{(-j)}$ is computed from overall fitting results and the CV criterion is simplified as

$$CV(\lambda) = \frac{1}{pq} \sum_{j=1}^{pq} (z\hat{\gamma}_j - x_j)^T K_j (z\hat{\gamma}_j - x_j), \tag{4.3}$$

where the symmetric matrix K_j is given as

$$K_j = W_j^{(-j)} + \frac{S_{jj}}{1 - c_j S_{jj}} \{a_j (a_j^{(-j)})^T \ + (a_j^{(-j)}) a_j^T\} + \frac{c_j^{(-j)} S_{jj}^2}{(1 - c_j S_{jj})^2} a_j a_j^T,$$

with A_{jj} being the j^{th} diagonal element of $A = (M + \lambda\Omega)^{-1}$, $c_j = z^T W_j z$, $a_j = W_j z$ and $c_j^{(-j)} = z^T W_j^{(-j)} z$ and $a_j^{(-j)} = W_j^{(-j)} z$. Formula (4.3) enables us to use overall fitting result for the weighted CV without fitting (pq) times.

4.4 Numerical Examples

In this chapter a real data and simulation study are given to demonstrate the performance of the proposed method for multivariate FD.

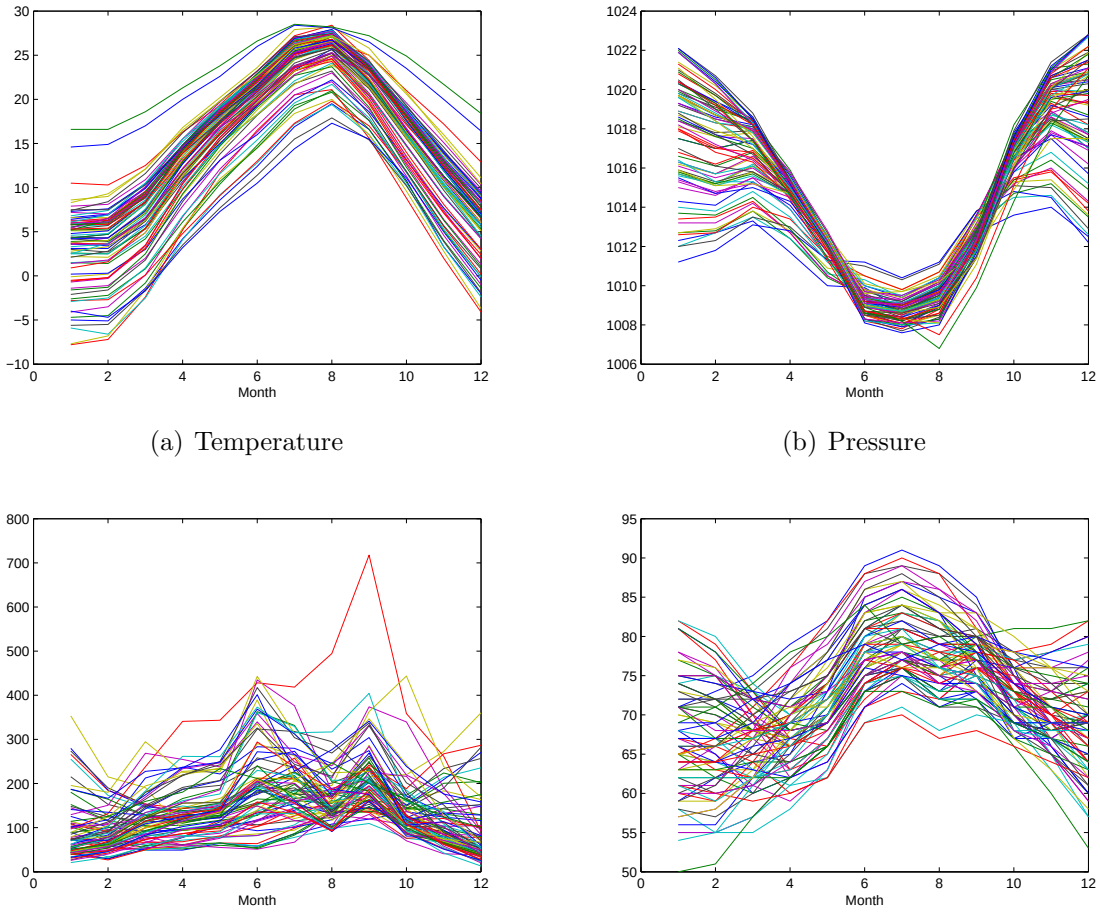
4.4.1 Dataset

The aim of the analysis is to illustrate the performance of the RMFPCA1 on the weather data, available in Chronological Scientific Tables 2005, which was used by Matsui and Konishi [57].

The weather data observed at 79 stations in Japan is a monthly data, observed at 12 points averaged from 1971 to 2000. The dataset includes four variables: monthly observed average temperatures, average atmospheric pressure, precipitation and average humidity. The Figures 4.1(a)-(d) exhibit the sample curves for monthly observed average temperatures, average atmospheric pressure, precipitation and average humidity, respectively. In Figures 4.1(a)-(d), the group of curves shows presence of a few trajectories that are in some way different from the rest.

We apply RMFPCA2 (robust method) to weather data as well as MFPCA (classical method) for the comparison purpose. The robust multivariate PCA is used to find the number of principal components which count 95% of total variability.

The first and second principal function estimates by MFPCA and RMFPCA2 for temperature, pressure, precipitation and humidity are given in Figures 4.2-4.5, respectively. Figures 4.2 and 4.3 the RMFPCA displays similar patterns but produces slightly different estimates with those from MFPCA for temperature and pressure variable, respectively. Temperature variable shows two outliers (78 and 79) (Figure 4.7 (a)). The first two principal functions computed by classical method clearly shows effect of outliers (Figure 4.2(a) and (b)). Figure 4.7 (b) shows pressure variable contains outlying observations (78 and 79). For the pressure variable first and second principal function computed by MFPCA are affected due to partial outliers at center. Ends of first and second principal function computed by



(a) Temperature

(b) Pressure

(c) Precipitation

(d) Humid

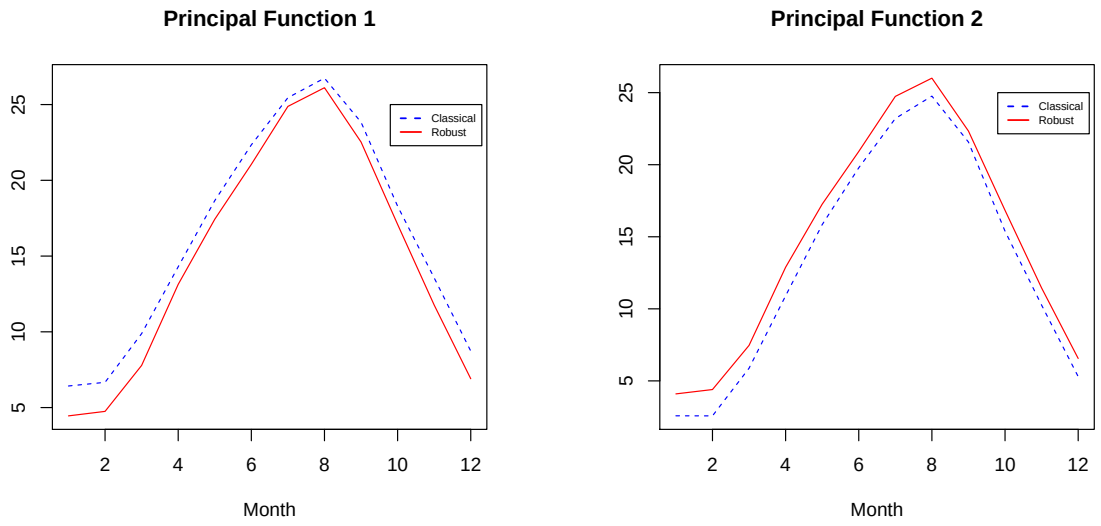
Figure 4.1: Weather Data.

MFPCA are also affected due to many observations having different shape at end points (Figure 4.3).

The results from Figure 4.4 for precipitation variable and Figure 4.5 for humidity variable show that principal function estimates by RMFPCA2 have departure from MFPCA (Ramsay and Silverman [64]). This can be attributed to observations with different shapes in the precipitation and humidity variable affecting major variabilities, explained by first and second principal functions in MFPCA. Precipitation variable contains few outlying curves (45, 47 and 48) (Figure 4.7 (c)). First principal function computed by MFPCA has large

value at center when compared with first principal function RMFPCA2, which clearly shows influence of outliers.

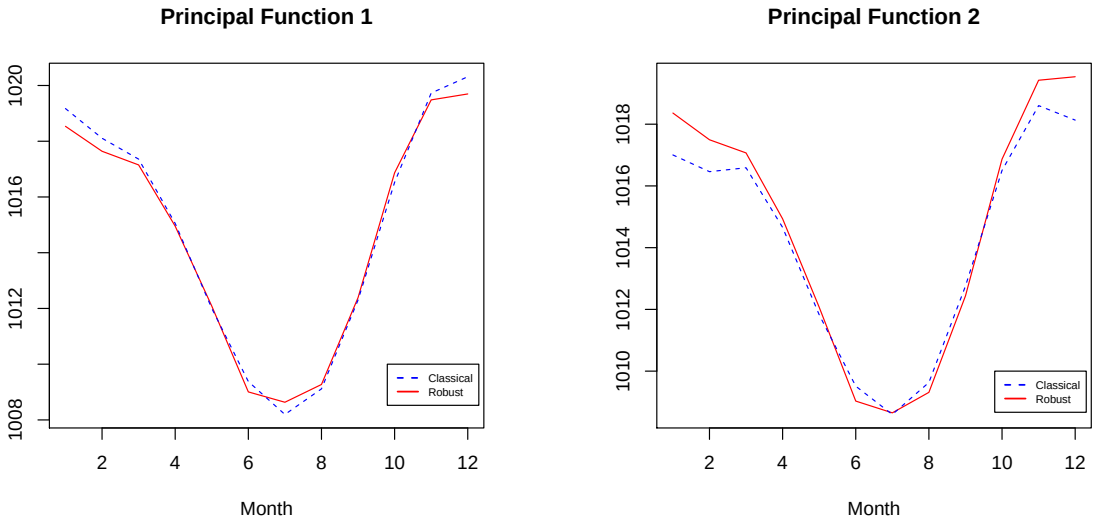
To detect the outliers for the four variables the resulting diagnostic plot is given by classical and robust method (Figure 4.6).



(a) (b)

Figure 4.2: Temperature Variable.

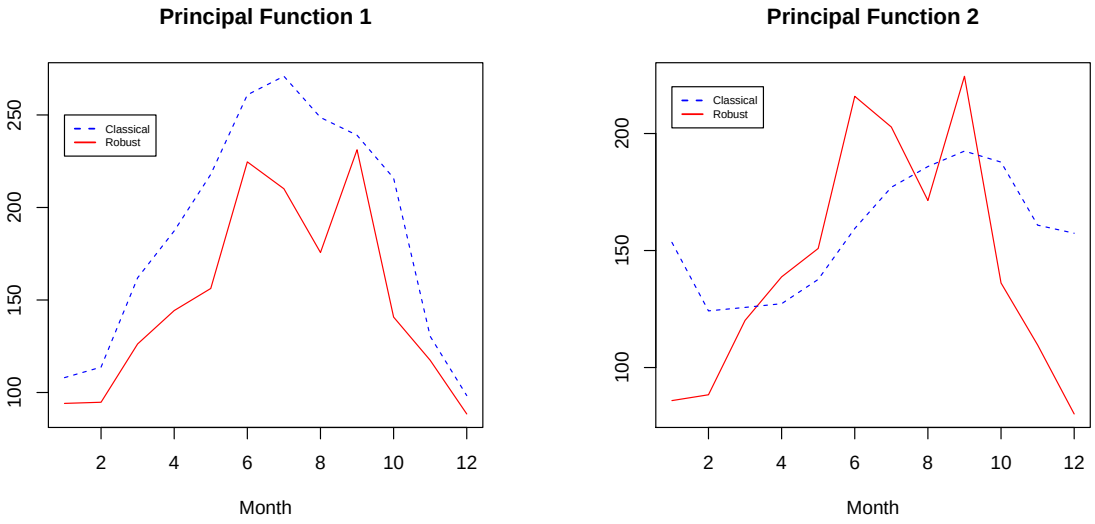
The orthogonal score plot indicates that both MFPCA (Figure 4.6(a)) and RMFPCA2 (Figure 4.6(b)) methods detected three similar outliers(45, 47 and 48). MFPCA detected additional one outlier (27) and RMFPCA2 detected two additional outliers (78 and 79). But, outliers detected by classical method does not conform with outliers detected by robust method for four variables. The one additional outlier (27) detected by MFPCA has shape almost same as regular observations and observation number 27 is not far from other observations of data for all four variables. Figure 4.7(a)) shows observations 78 and 79 are far from actual data for temperature variable. Observations 78 and 79 have different shape than regular observations for pressure variable (Figure 4.7(b)). Abnormal curves, 78 and 79, are identified by robust method but classical method fails to identify these outlier.



(a) (b)

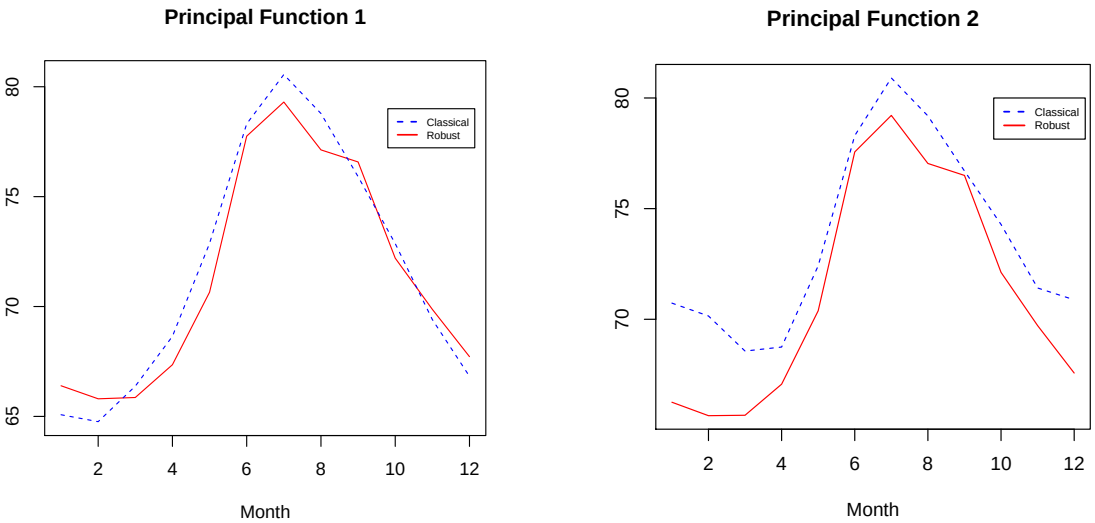
Figure 4.3: Pressure Variable.

Humidity variable contains group of observations of different shapes which are shown in Figure 4.8. For humidity variable there are no severe outliers.



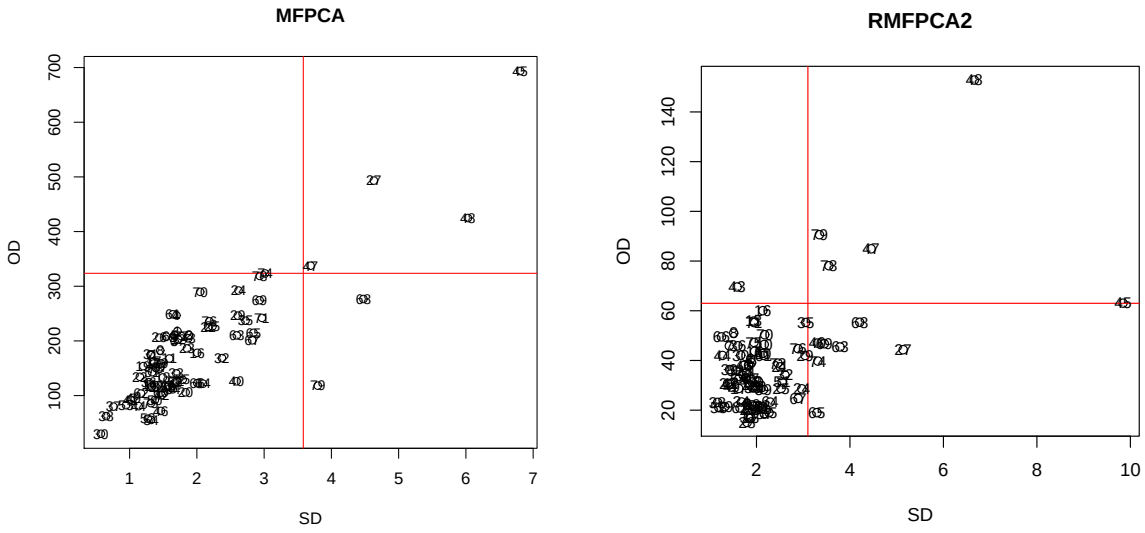
(a) (b)

Figure 4.4: Precipitation Variable.



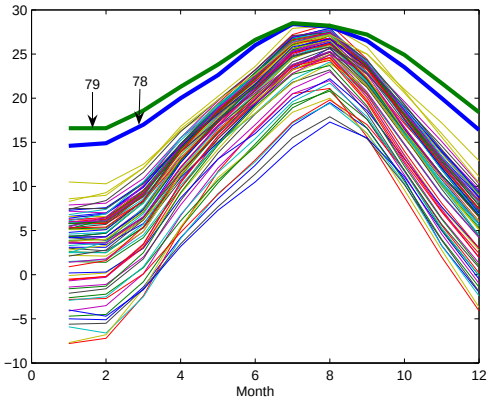
(a) (b)

Figure 4.5: Humid Variable.

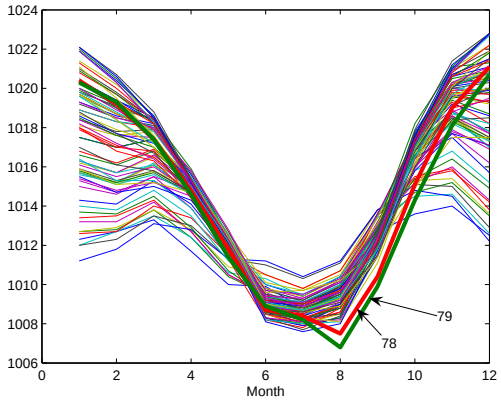


(a) (b)

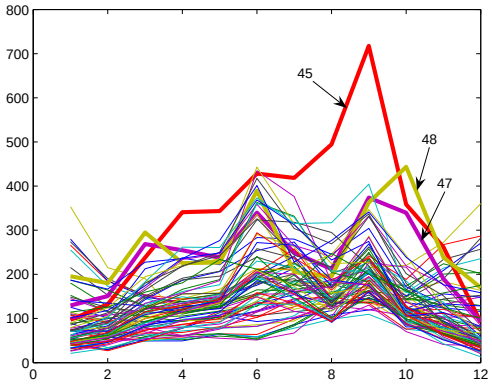
Figure 4.6: Outliers in Weather Data.



(a) Temperature



(b) Pressure



(c) Precipitation

Figure 4.7: Weather Data with outliers by robust method.

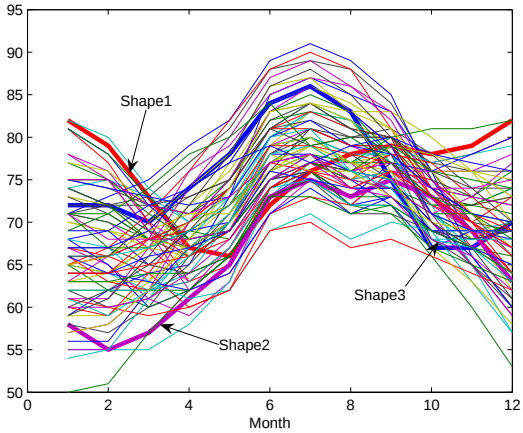


Figure 4.8: Humid variable with observations with different shape.

4.4.2 Simulation for Robust Multivariate Functional Principal Component Analysis Based on Multivariate Approach (RMFPCA1)

In this section simulation study is given to check the optimality of the robust functional PCA for outlier detection in multivariate functional data. The simulation setting given by Fraiman and Muniz [21, 51], with few changes, is used here. For simulation we consider multivariate functional data with two variables i.e. $(p = 2)$, $\{(x_{1l}, \ldots, x_{nl});\ l = 1, \ldots, p, \}$ obtained as realizations from a stochastic process $X(\cdot)$. This bivariate functional data has continuous paths on $[0, 1]$. Curves are generated from different models. The following models are considered for this simulation study.

Model 1 (no contamination): $X_i(t) = g(t) + e_i(t), 1 \leq i \leq n$, where model error term $e_i(t)$ is a stochastic Gaussian process with zero mean and covariance function $\vartheta(s,t) = (1/2)(1/2)^{(0.9)|t-s|}$ and (1) g(t)=4t, (2) g(t)=3-4t with $t \in [0, 1]$.

Model 2 (asymmetric contamination): $Y_i(t) = X_i(t) + c_i E$, $1 \leq i \leq n$, where c_i is 1 with probability q and 0 with probability $1 - q$; E is the contamination size constant.

Model 3 (symmetric contamination): $Y_i(t) = X_i(t) + c_i \sigma_i E$, $1 \leq i \leq n$, where c_i and E are defined as in model 2 and σ_i is a sequence of random variables independent of c_i taking values 1 and -1 with probability 1/2.

Model 4 (partially contaminated): $Y_i(t) = X_i(t) + c_i \sigma_i E$, if $t \geq Ti$, $1 \leq i \leq n$, and $Y_i(t) = X_i(t)$, if $t < T_i$, where T_i is a random number generated from a uniform distribution on $[0, 1]$.

Model 5 (Peak contamination): $Y_i(t) = X_i(t) + c_i \sigma_i E$, if $T_i \leq t \leq T_i + \ell$, $1 \leq i \leq n$, and $Y_i(t) = X_i(t)$, if $t \notin [T_i, T_i + \ell]$, where $\ell = 2/30$ and T_i is a random number from a uniform distribution in $[0, 1 - \ell]$.

Figure 4.9 exhibits curves simulated from these models. For each model, we generated 100 replications, with one setting each for low and high dimensional data. For low dimensional data we consider $n = 100,\ q = 12$ setting and for high dimensional data we consider setting with $n = 50,\ q = 100$. Noncontaminated bivariate functional data are obtained by

generating X_1 and X_2 variables from model 1, (i.e. contamination % (cp)= 0 and $E = 0$). Contaminated bivariate functional data are obtained by generating X_1 and X_2 variables each from contaminated models (2, 3, 4 and 5), where we considered $cp = 5, 10$ and $E = 10$. The number of basis used for functional curves simulated from these five models are obtained from GCV method. We estimate coefficients, C_l, $l = 1, 2$, for both variables by using the least squares method. MFPCA based on classical PCA (CPCA) and RMFPCA based on ROBPCA (Section 4.2) are applied on the simulated functional data according to the five models .

GCV method finds the same number of basis for contaminated and uncontaminated data for symmetric and asymmetric models. For partial and peak contamination cases GCV method finds different number of basis for contaminated and uncontaminated models. The reason for this is the shape or pattern of the contaminated curves is different than the uncontaminated data.

Two quantitative measures of the goodness of the methods are considered. The first one is mean proportion of variability (MPV) :

$$MPV = 1/N \sum_{r=1}^{N} \frac{\hat{\lambda}_1^r + \hat{\lambda}_2^r + \ldots + \hat{\lambda}_k^r}{\lambda_1 + \lambda_2 + \ldots + \lambda_k + \ldots + \lambda_q},$$

where N denotes the number of iterations, λ_j is the true j^{th} eigenvalue of the covariance function, $\hat{\lambda}_j^r$ is the estimated value of λ_j at the r^{th} replication. The $\hat{\lambda}_j^r$ is obtained by using classical or robust multivariate techniques on concatenated coefficient matrix of contaminated or uncontaminated model. For each setting, the optimal value for the mean proportion of explained variability is taken as 90% which corresponds to $k = 2$ (k is number of principal components).

The second quantitative measure is the norm, that is, the square root of sum of squared error of $\hat{\lambda}_1$ given by $\sqrt{\sum_{r=1}^{N}(\hat{\lambda}_1^{(r)} - \lambda_1)^2}$, where λ_1 is the largest true eigenvalue of the covariance function and $\hat{\lambda}_1$ is the estimated largest eigenvalue of λ_1 at the r^{th} replication. The $\hat{\lambda}_1^r$

is obtained by using classical or robust multivariate techniques on concatenated coefficient matrix of contaminated or uncontaminated model. The optimal value is zero or near zero.

Simulated data with no contamination is compared with simulated data with contamination, which is introduced by models 2,3,4 and 5. From Table 4.1, it is clear that CPCA provides the best mean proportion of explained variability when there is no contamination in the data as expected. For the uncontaminated data robust methods also yield comparable results. However, when contamination is introduced to the data (models 2-5) the eigenvalues obtained with CPCA are overestimated. Since estimated percentages of MPV are larger than 100%. In ROBPCA we obtain MPV of 80% for low dimensional data without and with contamination. For high dimensional data the mean percentage of explained variability is similarly 80% for without and with contamination. The main reason behind this is the optimal direction obtained by ROBPCA are robust to outliers. CPCA clearly fails and provides the worst possible result because mean proportion of variability is above 100%. The simulation results of mean proportion of variability for contaminated data versus uncontaminated data under low and high dimensional settings for other comparisons yielded very similar results observed in Table 4.1 and therefore they are not repeated here. As we increase the contamination level (cp) by 15% ROBPCA gives MPV value larger than 100% for high and low dimensional setting.

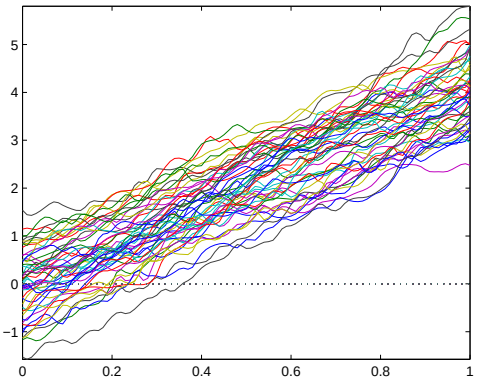
Simulation results for the norm with $N = 100$ iterations and $cp = 5, 10, 15$ percent contamination level for comparison of contaminated versus uncontaminated are summarized in Figure 4.10. For this comparison, we used one high and one low dimensional settings with the value of $E = 10$. The ideal value of norm must be very small or near zero. We conclude that the norm is near zero when there is no contamination for all methods. This is an indication of ROBPCA being also effective methods for uncontaminated data. The norm based on CPCA tends to increase as contamination level increases. For contaminated data, norms corresponding to ROBPCA method yield minimum value which is near zero for high and low dimensional settings. For contamination level cp=15% the norm value is

Table 4.1: Simulation results of the MPV (mean proportion of explained variability) for no contamination (both variables) (0%) and contaminated data (symmetric contamination for both variables (5%, 10%, 15%)) for low and high dimensional cases.

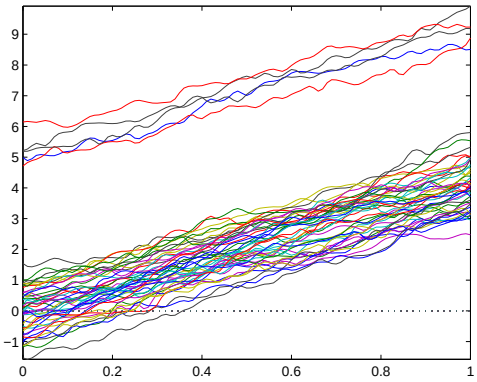
Contamination	High dimension:n=50, p=100	
	CPCA	ROBPCA
0%	0.837	0.786
5%	12.273	0.779
10%	19.360	0.718
15%	31.173	2.438
Contamination	Low dimension:n=100, p=12	
	CPCA	ROBPCA
0%	0.876	0.823
5%	11.247	0.795
10%	21.118	0.742
15%	32.132	1.343

above 0 for ROBPCA for low and high dimensional setting. When contamination level is increased above 10% the eigenvalues obtained by robust method are overestimated, resulting large norm value. For other simulation settings under low and high dimensional data similar results were observed as in Figure 4.10 therefore they are not reported here.

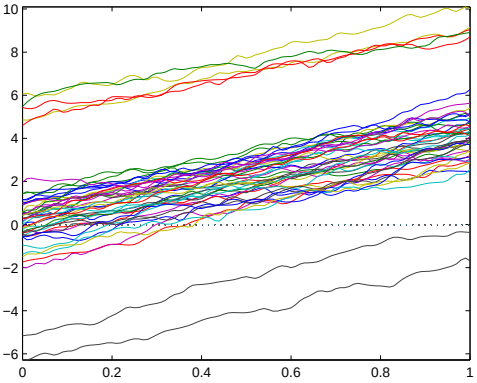
From the results of Figure 4.10 and Table 4.1 we can deduce that ROBPCA outperforms the CPCA. But ROBPCA method does not yield satisfactory results when percentage of outliers in data is above 10%.



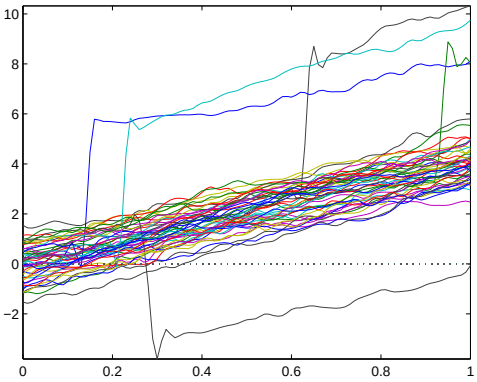
(a) Model 1



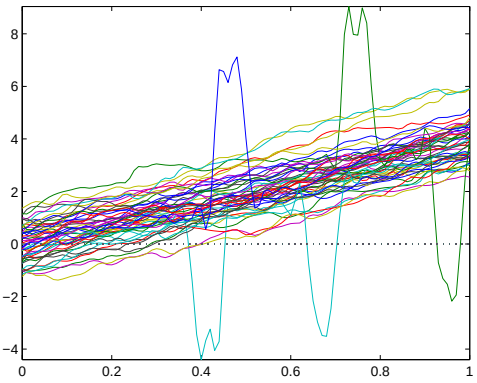
(b) Model 2



(c) Model 3

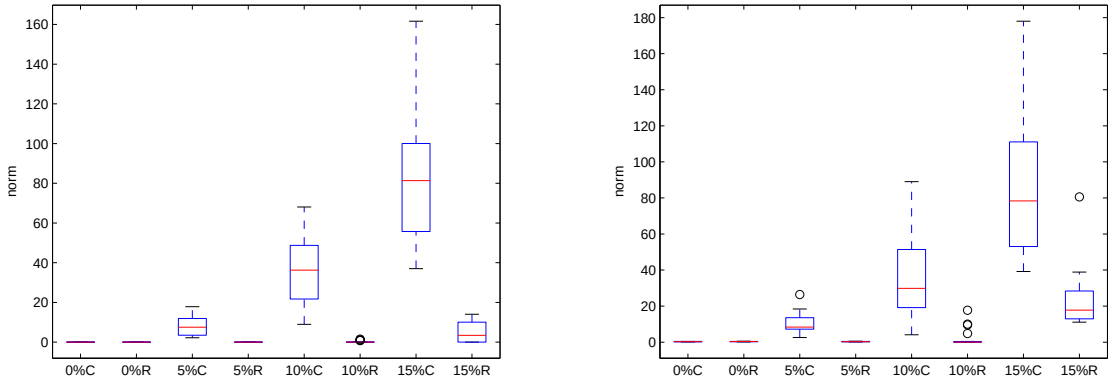


(d) Model 4



(e) Model 5

Figure 4.9: Curves generated from model 1 (cp=0%), model 2 (asymmetric contamination), model 3 (symmetric contamination), model 4 (partial contamination) and model 5 (peak contamination) with n=50, q=100, E=10 and cp=0.1.



(a) low dimension

(b) high dimension

Figure 4.10: Boxplots of norm when there is no contamination (0%) for both variables and symmetric contamination (5%,10%,15%) for both variables for CPCA(C) and ROBPCA(R).

4.4.3 Simulation for Proposed Robust Multivariate Functional Principal Component Analysis (RMFPCA2)

The proposed method is tested using simulated datasets under various scenarios. The performance is compared with existing classical multivariate functional PCA method [64]. The simulation setting inspired by Kayano and Sadanori [44], with few changes, is used here. For simulation we consider multivariate functional data $\{(x_{1l}, \ldots, x_{nl}); l = 1, \ldots, p; p = 3\}$ obtained as realizations from a stochastic process $X(\cdot)$. This multivariate functional data has continuous paths on $[0, 1]$. Let $x_i = (x_{ilj}, x_{i2j}, x_{i3j}) \in \Re^3$, $i = 1, \ldots, n$, $j = 1, \ldots, q$, be data on a finite and discrete grid at t_{ij} for 3-variable X_1, X_2, X_3.

Functional observations at the ordered time points $t_j \in (0, 1)$ with 10 principal components are generated from the underlying multivariate functional model:

$$x_{ilj} = x_{il}(t_{ij}) + \epsilon_{ilj} = m_l(t_{ij}) + \sum_{k=1}^{10} u_{ilk}\phi_{lk}(t_{ij}) + \epsilon_{ilj},$$

where the mean function $m(t)$ is assumed to be the following functions: 1. $m_1(t) = 0.3t^2 + 2t + 10$, 2. $m_2(t) = -4.5 * t^2 - 2 * t + 10$, 3. $m_3(t) = cos(-3 * t) * sin(-3 * pi * t)$ and $\phi_{lk}(t) = \sqrt{2}cos(k\pi t)$ is orthonormal principal functions. The errors ϵ_{ilj} are assumed to be generated from 0-mean Gaussian distributions of standard deviations 2 and 1 for $k = 1, 2$ and 0.05 for $k = 3, \ldots, 10$. For this functional model, some noise is added to create such cases where the conventional functional PCA suffers. Simulated curves are generated from different models. Model 0 was generated without contamination and several other models were generated with different types of contaminations.

Model 0 (No noise intervention): Model without outliers is used to test performance of robust multivariate functional principal component analysis (MFPCA) under absence of abnormal curves.

Model 1 (Some curves with unusual PC scores): For this model PC scores for the second principal function is contaminated. This can be done by multiplying u_{il2} by 10 for randomly selected 10% curves.

Model 2 (Some curves with irregular noise intervention): The noise intervention is introduced in randomly selected 10% curves. This is done by randomly selecting a time point t from a uniform distribution $(0, 1-\frac{1}{15})$ and then adding $c \times M$ to all observed values evaluated at $[t, t+\frac{1}{15}]$, where c takes value -1 or 1 with probability 1/2 and M is a random value from $N(15, 0.1^2)$.

Model 3 (All curves with irregular noise intervention): In this situation all curves are subjected to the noise intervention such that it is difficult to identify the outliers when compared with other curves. For all curves, a time point t is randomly selected from a uniform distribution $(0, 1-\frac{1}{15})$ and then $c \times M$ term is added to all observed values evaluated at $[t, t+\frac{1}{15}]$, where c takes value -1 or 1 with probability 1/2 and M is a random value from $N(15, 0.1^2)$.

Model 4 (Curves with large noise): Randomly selected 10% curves are replaced by $m(t_j) + z_{ij}$ where z_{ij} has Normal distribution with mean 0 and variance 1^2.

Model 5(Some curves with partial contamination):In this scenario few curves are partially contaminated. For 10% randomly selected curves, a time point t is randomly selected from a uniform distribution Y on $[0, 1]$. If $t \geq Y$, and then the term $c \times M$ term is added, where c and M are defined as in model 3.

Figures 4.11 and 4.12 and Figures 4.14 and 4.16 exhibit curves simulated from these models. For each simulation set, we generated 100 replications, with one setting each for low and high dimensional data and all methods were applied to the simulated data sets to estimate the first 2 principal functions. Let $X_1(t)$, $X_2(t)$ and $X_3(t)$ denote the three variables for multivariate functional curves. Three simulation settings are considered with different types of contamination. The purpose of these different simulation settings is to assess the robustness of the proposed algorithm under the different types of contamination. For low

dimensional data we consider $n = 100,\ q = 20$ setting and for high dimensional data we consider setting with $n = 50,\ q = 100$ for three variables in all simulation settings. On x we apply RMFPCA2 (Section 4.3) and MFPCA (Section 2.4) on the simulated 3-functional variable data and find PC functions $\gamma = (\gamma^1, \ldots, \gamma^K)_{(pq)\times K}$. $K = 2$ are number of PCs for which the cumulative percentage of total variance is over 95%.

Quantitative measure: To quantify and compare the performance of methods, the average of total mean squared errors (TMSE) between the first true principal function, $\phi_{l1}(t)$, for l^{th} variable and its estimate, $\gamma_l^1(t)$ is used

$$\sum_{l=1}^{p} \parallel \gamma_l^1 - \phi_{l1} \parallel^2 = \sum_{l=1}^{p} \int (\gamma_l^1(t) - \phi_{l1}(t))^2 dt.$$

Since multivariate functional data are observed on a finite grid of time points, a discrete version of the TMSE is computed.

Average of 100 total mean squared errors for classical and robust methods are given in Tables 4.2 and 4.3 for high and low dimensional data, respectively. Datasets from all models are contaminated in some ways so that conventional methods will be affected by such contamination. TMSE from Tables 4.2 and 4.3 indicates the RMFPCA2 (robust method) remarkably outperforms the MFPCA (classical method) in all scenarios.

In simulation setting 1, $X_1(t)$ and $X_2(t)$ are generated from model without contamination (model 0) and $X_3(t)$ contains some contamination and this variable is obtained from model 3 (Figure 4.12). First principal function of contaminated third variable is given in Figure 4.13, which clearly shows that first principal function computed by classical method is deteriorated from the true principal function. The principal function computed by classical method clearly shows effect of irregular noise intervention. Figure 4.14 shows the simulation setting 2, where $X_1(t)$ is generated from model 0, $X_2(t)$ and $X_3(t)$ contain contaminations and are obtained by using model 1 and model 2 respectively. For the variable $X_2(t)$ the second principal function is selected as the first principal function in the conventional method

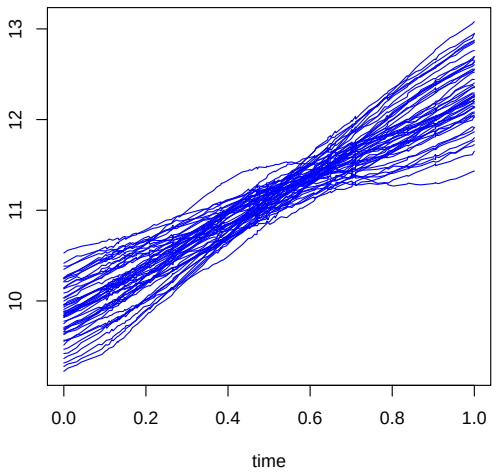
Table 4.2: Simulation results of the TMSE (total mean squared errors) of the estimated first principal function for simulated datasets for high dimensional case.

High dimension:n=50, p=100			
Variable	Contamination	MFPCA	RMFPCA
1	Model 0		
2	Model 0	0.9751	0.0766
3	Model 3		
1	Model 0		
2	Model 1	0.9587	0.0765
3	Model 2		
1	Model 5		
2	Model 4	0.9348	0.0733
3	Model 0		

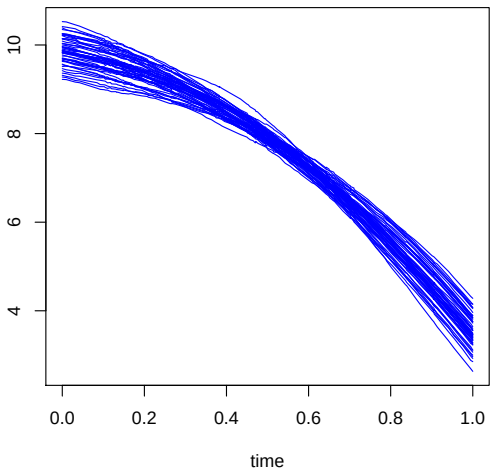
because its PC scores have large variance due to few outlying curves (Figure 4.15(a)). The classical method is influenced by the outlying curves and so they have large PC2 scores which leads to the estimation of the second principal function as the first principal function. $X_3(t)$ is generated from model 2 where noise partly intervenes in the few outlying curves, principal function estimates from classical method deteriorated while those from robust method found principal function almost correctly (Figure 4.15(b)). For the simulation setting 3 model 5 and model 4 are used to contaminate $X_1(t)$ and $X_2(t)$, respectively (Figure 4.16). Third variable contains no outliers and is generated by using model 0. Figure 4.17 shows the first principal functions by both methods for contaminated variables. The shape of the first principal function of $X_1(t)$ produced by classical method is affected due to partial outliers (Figure 4.17(a)). And for $X_2(t)$ due to large noise the first principal function gave the devastating results, while robust method resulted in the reasonably good estimates under such severe condition (Figure 4.17(b)). Under all the scenarios of different contamination, robust method found principal functions almost correctly.

Table 4.3: Simulation results of the TMSE (total mean squared errors) of the estimated first principal function for simulated datasets for low dimensional case.

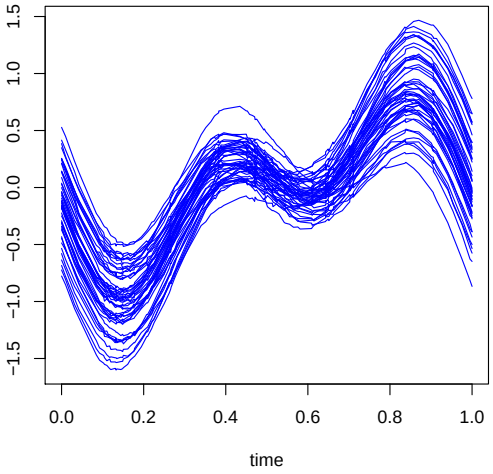
Low dimension:n=100, p=20			
Variable	Contamination	MFPCA	RMFPCA
1	Model 0		
2	Model 0	0.9988	0.0779
3	Model 3		
1	Model 0		
2	Model 1	0.8751	0.0767
3	Model 2		
1	Model 5		
2	Model 4	0.8880	0.0781
3	Model 0		



(a) Variable-1

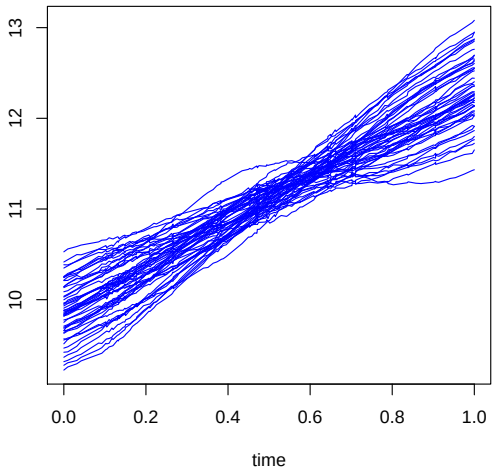


(b) Variable-2

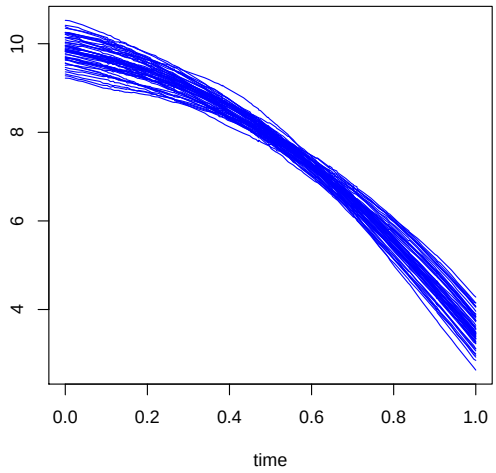


(c) Variable-3

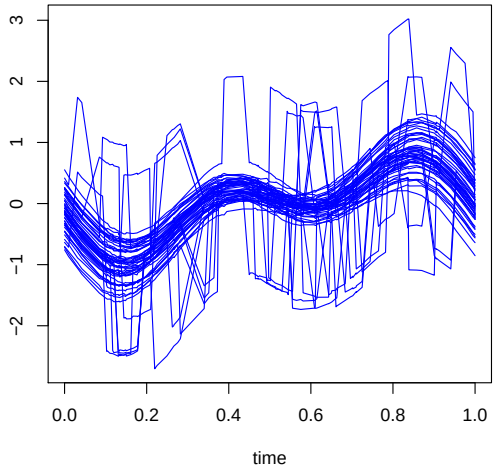
Figure 4.11: Curves generated from model 0 for $X_1(t)$, $X_2(t)$ and $X_3(t)$ with n=50, q=100.



(a) Model 0

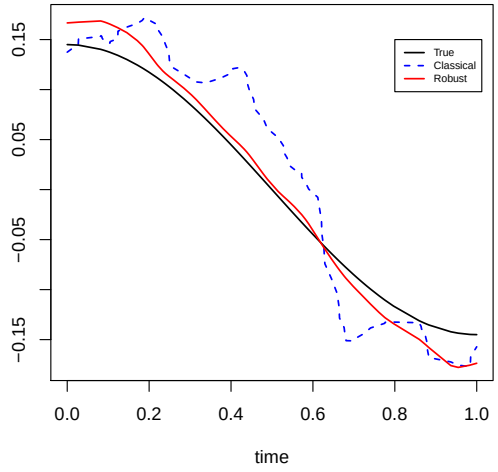


(b) Model 0



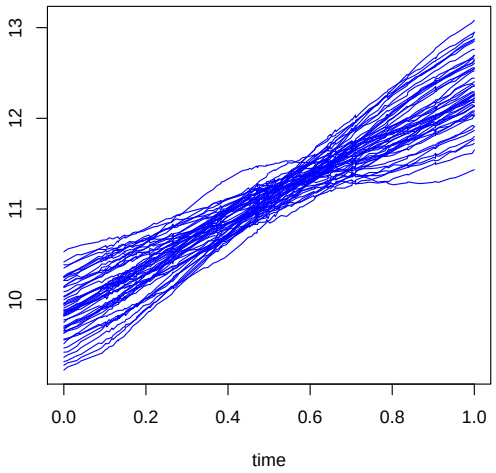
(c) Model 3

Figure 4.12: Simulation setting-1:Curves generated from model 0 for $X_1(t)$ and $X_2(t)$ and from model 3 for $X_3(t)$ with n=50, q=100.

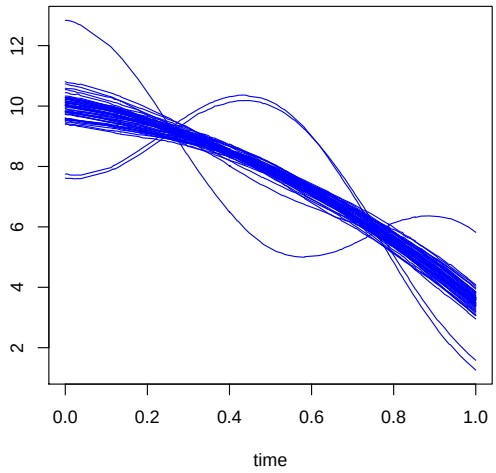


(a) Variable-3

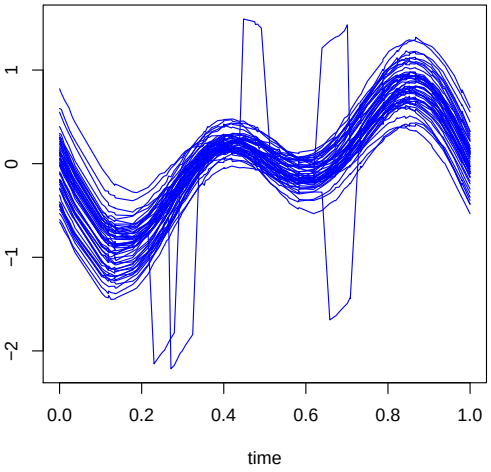
Figure 4.13: First principal function for Simulation setting-1 for contaminated variables.



(a) Model 0

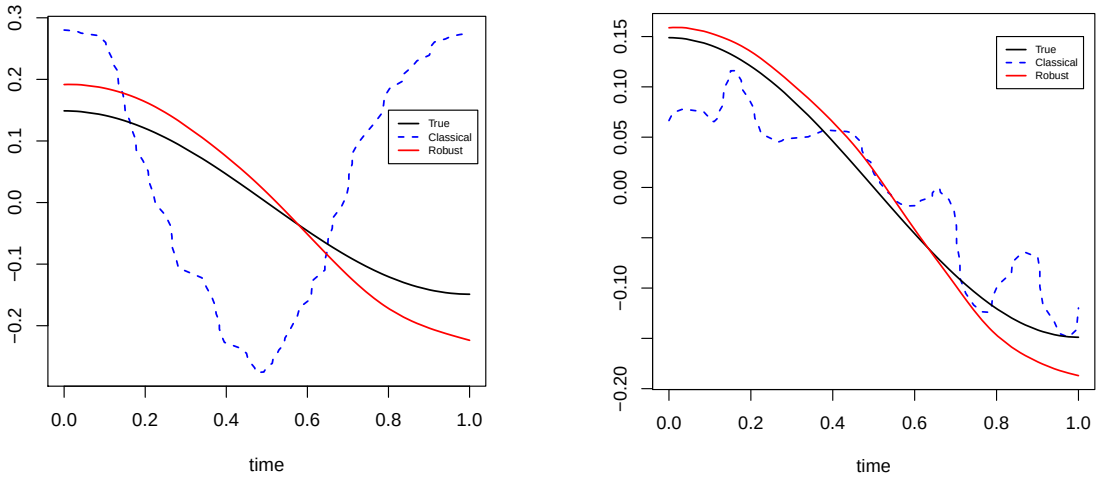


(b) Model 1



(c) Model 2

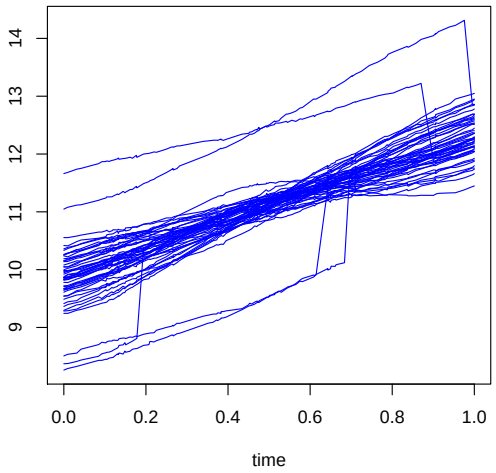
Figure 4.14: Simulation setting-2:Curves generated from model 0,1,2 for $X_1(t)$, $X_2(t)$ and $X_3(t)$, respectively with n=50, q=100.



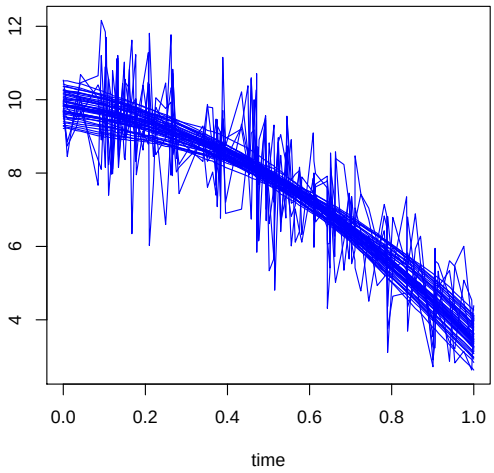
(a) Variable-2

(b) Variable-3

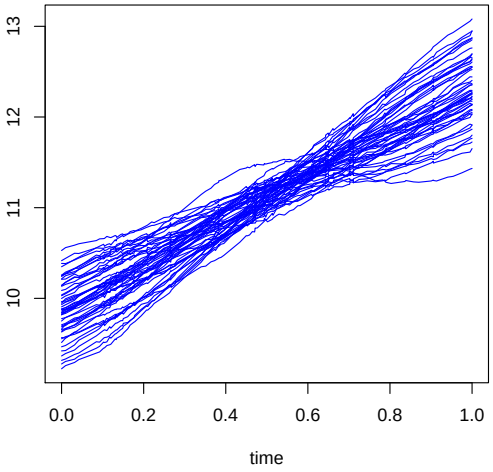
Figure 4.15: First principal function for Simulation setting-2 for contaminated variables.



(a) Model 5

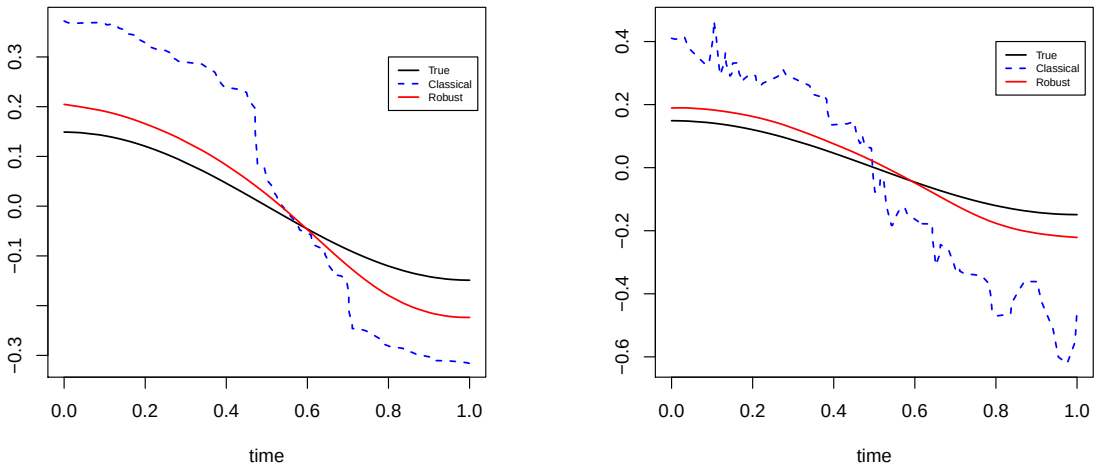


(b) Model 4



(c) Model 0

Figure 4.16: Simulation setting-3:Curves generated from model 5,4,0 for $X_1(t)$, $X_2(t)$ and $X_3(t)$, respectively with n=50, q=100.



(a) Variable-1

(b) Variable-2

Figure 4.17: First principal function for Simulation setting-3 for contaminated variables.

Chapter 5

Robust Functional Linear Model

5.1 Introduction

The consequence of development in technology and computation has resulted in an increase in the number of applications where observations are functions or images. These observations which can be expressed by functional regression models are often times seen in various fields such as climatology, chemometrics, linguistics (Ramsay and Silverman [64] and Ferraty and Vieu [24]). Recently researchers have put more emphasis on functional linear models in which the regressors and/or the response are of a functional nature and proposed several methods for estimating the functional parameter [16, 18, 22, 23, 62].

The functional versions of the diagnostic measures based on Cook's distance [13] is introduced by Chiou and M*ü*ller [12] and Shen and Xu [72] for the models where the regressors are real or curves and the responses are functional. Febrero *et al.* [23] reviewed estimation based on the classical functional principal components method and then analyzed influence in the functional linear model with scalar response. They have proposed three measures of influence by generalizing the measures proposed for the standard regression model by Cook [13] and Peña [61]. The functional regression framework developed by M*ü*ller and Stadt*ü*ller [59] uses regularization based on the Karhunen-Loeve expansion which leads to regression on functional principal components. In this method observations are projected on a finite dimensional space spanned by eigenfunctions of the (empirical) covariance operator. Ramsay and Dalzell [62] and Cardot *et al.* [10] proposed regularization through a penalized least squares approach after expanding functional parameter in some basis (such as splines). The generalized functional linear model is given by Cardot and Sarda [11].

Functional regressors are infinite in nature. Problem with infinite dimensionality of the regressor is that, it results into infinitely many sets of solutions or suffers from multicollinearity. Traditional methods for functional regression are based on an L_2 norm of the residuals and are sensitive to outliers or influential observations, which has a serious effect on the estimation and prediction of the functional linear model. Influential observations in a given dataset can have a strong impact on analysis. If these outlying or influential observations are removed from the data then this may substantially affect the statistical inference. An alternative approach to classical methods is robust estimation which is not affected by the presence of outliers.

Approach of Hastie and Mallows [33] is based on a smooth B-spline expansion for functional parameter. Marx and Eilers [52] considered a smooth basis expansion procedure with roughness penalty in a least squares criterion. For linear regression with a functional predictor and scalar response, Cardot *et al.* [9] provided consistency results and discussed inference for the regression function. Marx and Eilers [53] proposed overcoming the multicollinearity problem by using B-spline expansion of functional parameter and adding a roughness penalty. James [41] and Reiss and Ogden [65] considered regression model with functional covariates. Functional principal component regression is developed by Cardot *et al.* [9] and Reiss and Ogden [65] where scalar response is regressed on first few PC loadings of the functional regressors. There is limited literature on robust functional regression. Robust estimator for nonparametric models was considered by Crambes *et al.* [14]. Recently, Maronna and Yohai [56] proposed robust functional regression for pre-smoothed curves and Gervini [28] proposed functional robust regression for longitudinal data.

Our goal is to propose a robust version of the estimator for functional linear regression with a functional predictor and scalar response. We assume that functional predictors contain abnormal observations and scalar response is free of outliers. The outline of this chapter is as follows. Section 5.2 provides the details of the proposed robust functional principal component regression (RFPCR). Diagnostic plots used for principal component regression

(PCR) are described and implemented to diagnose functional outliers in functional regression. Real and simulated data sets are utilized to demonstrate the performance of the proposed method in Section 5.3.

5.2 Proposed Method: Robust Functional Principal Component Regression (RFPCR)

The functional linear model with a scalar response is a regression model with the regressor which is a random curve and the response which is real random variable defined on the same probability space. We assume that (X, y) is a pair of random sample where $X = (X(t))$, $X \in L_2(T)$, $t \in T = [t_{min}, t_{max}] \subset \Re$ and y is a real random variable. For easy computation we assume that both X and y are centered; i.e. $E[X(t)] = 0$, and $E[y] = 0$. Assuming $E(\| X \|^2) < \infty$, the dependence between the scalar response y and the functional random variable X is written as:

$$y = \alpha + \langle X, \beta \rangle + \epsilon = \alpha + \int_T X(t)\beta(t)dt + \epsilon, \tag{5.1}$$

where $\langle ., . \rangle$, denotes the $L_2(T)$ inner product, β is a square integrable function defined on T and errors, ϵ, is a real random variable with $E[\epsilon] = 0$, $E[X(t)\epsilon] = 0$ and finite variance equal to σ^2. In practice $X_i(t)$ is observed discretely on a finite grid $t_{ij} \in T \subset \Re, i = 1, \ldots, n, j = 1, \ldots, q$ as a random sample of pairs $(X_i(t_{ij}), y_i)$, $i = 1, \ldots, n$. We assume that functional predictors contain abnormal observations and scalar response is free of outliers. Dimension of $X_i(t)$ is reduced by using a finite series expansion

$$X_i(t) = \sum_{j=1}^{K_x} c_{ij}\psi_j(t), \tag{5.2}$$

where $\psi_j(t)$, $j = 1, \ldots, K_x$ are K_x basis functions and $C = [c_{ij}]$ is $n \times K_x$ coefficient matrix of the basis expansion (Chapter 1). Equation (5.2) can be written by approximating $X_i(t)$ in terms of eigenfunction expansion based on a truncated Karhunen-Loève decomposition ([1]) as

$$X_i(t) = \sum_{j=1}^{K_x} c_{ij}\gamma_j(t), \tag{5.3}$$

where the coefficient matrix, $C = [c_{ij}]$, $i = 1, \ldots, n$, $j = 1, \ldots, K_x$, is the scores and γ is the collection of the first K_x eigenfunctions of the covariance matrix $V(s,t) = cov[X_i(s), X_i(t)]$ (M*ü*ller and Stadtmuller [59], chapter 8 in [64]). In the proposed method, matrix of robust eigenfunctions γ of size $K_x \times q$ is obtained by applying robust FPCA by Lee *et al.* [46] (Section 3.3) on $X_i(t)$. The optimal number of components K_x is the minimal number of principal components needed to explain 99% of the total variation in the discretized versions of the random functions $X_i(t)$.

Estimation of $\beta(t)$ is done by using a truncated power series spline basis $\phi(t) = \{\phi_1(t), \ldots, \phi_{K_b}(t)\}$ such that

$$\beta(t) = \sum_{k=1}^{K_b} b_k \phi_k(t) = \phi(t) b, \tag{5.4}$$

where $b = \{b_1, \ldots, b_{K_b}\}'$. This approach imposes differentiability and allows simple control of smoothness (Goldsmith *et al.* [30]). The constraint on choice of number of eigenfunctions is $K_x \geq K_b$, where $K_b = min(K_x, 35)$ is taken large enough to prevent under smoothing (Rupert [69]). Now the model (5.1) can be expressed as

$$\alpha + \int_T X_i(t)\beta(t)dt = \alpha + \int_T C\gamma(t)\phi(t)'bdt = \alpha + CJ_{\gamma\phi}b, \tag{5.5}$$

where $K_x \times K_b$ matrix $J_{\gamma\phi}$ is defined by

$$J_{\gamma\phi} = \int \gamma(t)\phi'(t)dt. \tag{5.6}$$

Let $\zeta = (\alpha, b_1, \ldots, b_{K_b})'$ be $(K_b + 1)$ vector and the coefficient matrix $Q = [1 \;\; CJ_{\gamma\phi}]$ be the $n \times (K_b + 1)$ matrix. The matrix notation of model (5.5) is

$$\hat{y}_{(n\times 1)} = Q_{(n\times(K_b+1))}\hat{\zeta}_{((K_b+1)\times 1)}, \tag{5.7}$$

The least squares estimate of the $\hat{\zeta}$ can be obtained by solving the equation (5.7). Since the scores matrix, C, contains the information of outliers, this estimator is not resistant to outliers ([28]). To obtain robust ζ a mechanism to downweight outlying scores is implemented. Weighted least squares method by Billor *et al.* [6] is used in a following manner

Obtain robust Mahalanobis distance of Q_i, $i = 1, \ldots, n$, as $D_i = \sqrt{(\sum_{j=1}^{q} Q_{ij}^2/\lambda_j)}$ using ROBPCA. Q_i is i^{th} row of Q and λ_j are the eigenvalues of covariance matrix of Q. Initial weights are obtained as $w_i^0 = w^*(D_i)$, where

$$w^*(a_i) = min\left(1, \frac{1}{max(|a_i|, median_i(|a_i|))}\right). \tag{5.8}$$

Low weights are assigned to the observations with large robust distances. Normalized distances, d_i, are calculated by

$$d_i = \frac{D_i^2}{\sum_{i=1}^{n} D_i^2} \tag{5.9}$$

Algorithm

Input: Data matrix, Q, of size $n \times (K_b + 1)$, $n \times 1$ response variable, y, initial weight vector, w^0 and normalized distance vector, d.

Output: Coefficent function, $\hat{\zeta}$, and the corresponding residual, r.

Step 0: Let $W = w^0 = diag\{\sqrt{w_i^0}\}, i = 1, \ldots, n$.

Step 1: Obtain weighted Q and y by multiplying with W as $Q_w = WQ$ and $y_w = Wy$. Regress y_w on Q_w to obtain $\hat{\zeta}$ and fitted values $\hat{y}$.

Step 2:Calculate the residual vector, $r = y - \hat{y}$, $R_i = \frac{r_i}{mad_i(r_i)}$, where mad=median absolute deviation and obtain new weights as

$$w_i = (1 - d_i)w^*(R_i). \tag{5.10}$$

Redefine $W = diag\{\sqrt{w_i}\}, i = 1, \ldots, n$.

Step 3: Return to step 1 until the convergence of $\hat{\zeta}$.

5.2.1 Diagnostic Plot for Detection of Outliers

Diagnostic plots used to detect outliers in principal component regression are used to find functional outliers in functional principal component regression analysis. Orthogonal-score plot (Section 4.2.1) and Residual-score plot proposed by Hubert *et al.* [37] are used to distinguish regular observations from outliers. The Residual-score plot is a scatter plot of the robust residual distance Rd_i versus the robust score distance Sd_i. The score distance is defined as

$$Sd_i = \sqrt{\sum_{K=1}^{K_x} (z_i^K)^2/\lambda^K}, \quad i = 1, \ldots, n,$$

where z_i^K is the i^{th} score for K^{th} component and λ^K is the eigenvalue for K^{th} component. The scaled residuals Rd_i, is given by

$$Rd_i = \left| \frac{r_i}{mad_i(r_i)} \right| = |R_i|, \quad i = 1, \ldots, n,$$

where r_i is the robust residual value. Two cutoff lines based on classical nonparametric threshold are used to classify the observations. The cutoff value for horizontal line is tr(Rd) and for vertical line is tr(Sd), which help to flag outliers from regular observations. The classical nonparametric threshold [74] is given as

$$tr(a) = median_i(a_i) + (2.5)mad_i(a_i),$$

where $\mathbf{a} \geq 0$ is any vector with positive entries.

Residual-score plot classifies observations as homogeneous observations (lower left corner) and observations in upper right corner or far away from the homogeneous observations (having large score or residual distances) are known to be functional outliers.

5.3 Numerical Examples

In this section, benchmark data set and several simulation studies are used to explore the performance of the proposed algorithm, RFPCR. The proposed method is compared with, one of the existing classical functional regression method (Goldsmith *et al.* [30])

5.3.1 Simulation

Simulation configurations are conducted with two different beta functions. Both settings aim to assess the robustness of the proposed algorithms under different error distributions.

A similar simulation setting described by Goldsmith *et al.* [30] is employed in this section. Consider functional regressor model and a continuous outcome with the grid $\{t_j = (j/10) : j = 0, 1, \ldots, 100\}$ on the interval [0,10]. Scalar outcomes Y_i and regressor functions $X_i(t)$ are generated from the following model

$$Simulation \;\; setting - 1 : Y_i = \frac{1}{K}\sum_{j=0}^{100} X_i(t_j)\beta^1(t_j) + \epsilon_i, \quad i = 1, \ldots, n, \tag{5.11}$$

$$Simulation \;\; setting - 2 : Y_i = \frac{1}{K}\sum_{j=0}^{100} X_i(t_j)\beta^2(t_j) + \epsilon_i, \quad i = 1, \ldots, n, \tag{5.12}$$

$$X_i(t_j) = \omega_{i_1} + \omega_{i_2}t_j + \sum_{k=1}^{10}\left\{\vartheta_{ik_1}sin\left(\frac{2\pi k}{10}t_j\right) + \vartheta_{ik_2}cos\left(\frac{2\pi k}{10}t_j\right)\right\}, \tag{5.13}$$

where $\epsilon_i \sim N[0, \sigma_\epsilon^2]$, $\omega_{i_1} \sim N[0, 25]$, $\omega_{i_2} \sim N[0, 0.04]$, and $\vartheta_{ik_1}, \vartheta_{ik_2} \sim N[0, 1/k^2]$. Figure 5.5 gives a boxplot of scalar response and Figure 5.1(a) displays a sample of 50 random functions $X_i(t)$.

The values that have been chosen for the different parameters for this simulation study are the following: the observed outcomes Y_i, are generated by considering $\sigma_\epsilon^2 \in \{0.5, 1\}$,

two true coefficient functions $\beta^l(t)$, $l = 1, 2$ and functional regressor $X_i(t)$ is generated from no contamination model (Model 0) and then contaminated by using Model 1-3. The choices of the true coefficient functions considered in simulations are $\beta^1(t) = sin(\pi t/5)$ and $\beta^2(t) = (t/2.5)^2$. The observed outcomes Y_i are assumed to be free of outliers.

Curves $X_i(t)$ are generated from different models. Model 0 was generated without contamination and several other models were generated with different types of contaminations.
Model 0 *(no contamination)*: Equation 5.13
Model 1 *(partially contaminated)*: $Y_i(t) = X_i(t) + c_i\sigma_i E$, if $t \geq Ti$, $1 \leq i \leq n$, and $Y_i(t) = X_i(t)$, if $t < T_i$, where T_i is a random number generated from a uniform distribution on $[0, 1]$, E is a random value from $N(15, 0.1^2)$ and c_i is 1 with probability q and 0 with probability $1 - q$.
Model 2 *(Peak contamination)*: $Y_i(t) = X_i(t) + c_i\sigma_i E$, if $T_i \leq t \leq T_i + \ell$, $1 \leq i \leq n$, and $Y_i(t) = X_i(t)$, if $t \notin [T_i, T_i + \ell]$, where $\ell = 2/30$ and T_i is a random number from a uniform distribution in $[0, 1 - \ell]$.
Model 3 *(asymmetric contamination)*: $Y_i(t) = X_i(t) + c_i E$, $1 \leq i \leq n$, where c_i is 1 with probability q and 0 with probability $1 - q$; E is the contamination size constant.

For each combination of the parameter values σ^2_ϵ, $\beta^l(t)$ and $X_i(t)$, 100 datasets $[Y_i, X_i(t_j) : i = 1, \ldots, n]$ are simulated for high dimension (n=50, q=101) and low dimension (n=200, q=101) setting. Proposed approach RFPCR (Section 5.2) to estimate $\beta^l(t)$ is compared with classical method proposed by Goldsmith *et al.* [30].

The average mean square error (AMSE) of $\hat{\beta}(t)^l$ over the 100 samples is utilized as a quantitative measure to access the performance of both the methods

$$AMSE(\hat{\beta}(\cdot)) = \frac{1}{100}\sum_{r=1} 100 \left[\frac{1}{q}\sum_{j=1}^{q}\left\{\hat{\beta}^l_r(t_j) - \beta^l(t_j)\right\}^2\right], \quad l = 1, 2,$$

where $\hat{\beta}^l_r(t)$ is the estimated coefficient function of true beta function $\beta^l_r(t)$ from the r^{th} simulated data set.

Table 5.1: Simulation results of the average MSE for each combination of the true coefficient function and the measurement error variance for low dimensional case.

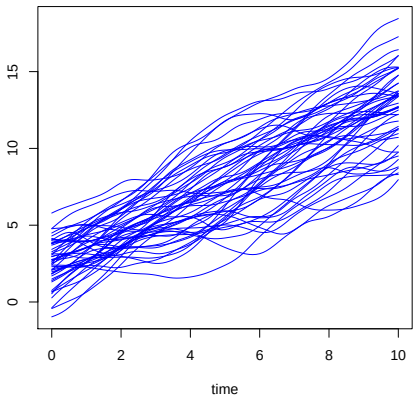
Low dimension:n=200, q=101					
Beta function	Contamination	σ^2=0.5		σ^2=1	
		PFR	RFPCR	PFR	RFPCR
β^1	Model 0	0.00173	0.00176	0.00021	0.0012
	Model 1	0.4252	0.0014	0.2550	0.0969
	Model 2	0.0106	0.0009	32.545	0.0030
β^2	Model 0	0.00006	0.0002	0.0035	0.0036
	Model 1	28.3246	0.0109	0.3224	0.0022
	Model 3	0.4431	0.0349	0.4313	0.0038

Tables 5.1- 5.2 summarized to compare the AMSE for each set of the parameters for classical method and robust method (RFPCR). When there is no contamination in functional regressor, the performance of classical and robust method is similar for all possible parameter combinations and for low and high dimension datasets. For contaminated functional regressor in high and low dimension setting the proposed method has smaller AMSE value than classical method for every parameter combination.

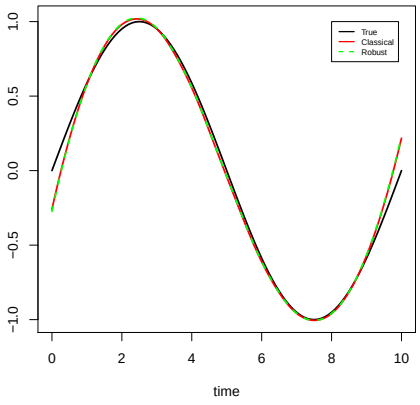
Figures 5.1- 5.4 displays the functional regressor in the left panel and respective estimated beta functions by both the methods in right panel for $\sigma_\epsilon^2 = 0.5$. For the smooth β^2 (Figures 5.2 and 5.4), the proposed method provides the closest estimates for low and high dimensional setting. Proposed method performs slightly worse for β^1 for low and high dimensional data (Figures 5.1 and 5.3). When functional regressor is contaminated the performance of classical method is relatively poor for low and high dimensional setting (Figures 5.1- 5.4). For $\sigma_\epsilon^2 = 1$ yielded very similar results which are not reported here.

Table 5.2: Simulation results of the average MSE for each combination of the true coefficient function and the measurement error variance for high dimensional case.

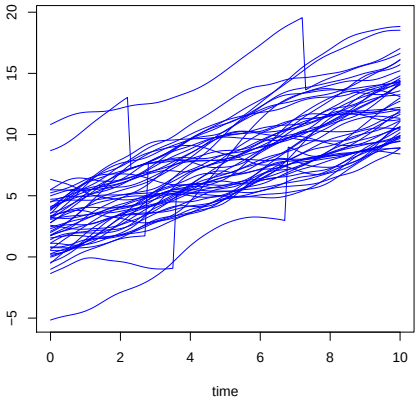
High dimension:n=50, q=101					
Beta function	Contamination	σ^2=0.5		σ^2=1	
		PFR	RFPCR	PFR	RFPCR
β^1	Model 0	0.0077	0.0082	0.0153	0.0158
	Model 1	0.0797	0.00910	31.0956	0.0043
	Model 2	0.6843	0.0103	11.2888	0.0592
β^2	Model 0	0.0014	0.0007	0.0006	0.0041
	Model 1	32.0457	0.3218	94.3803	2.5476
	Model 3	13.348	0.2741	20.6127	0.0074



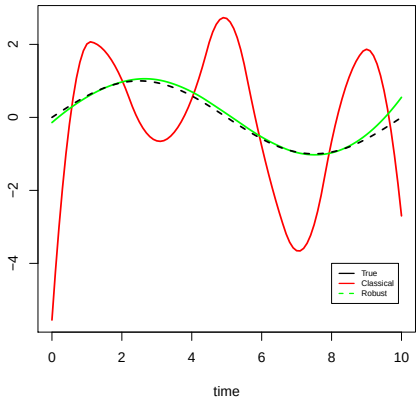
(a) Model 0



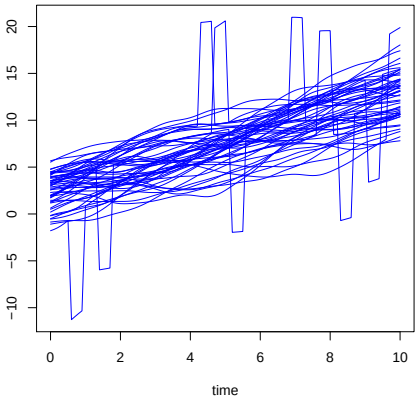
(b) $\hat{\beta}^1$



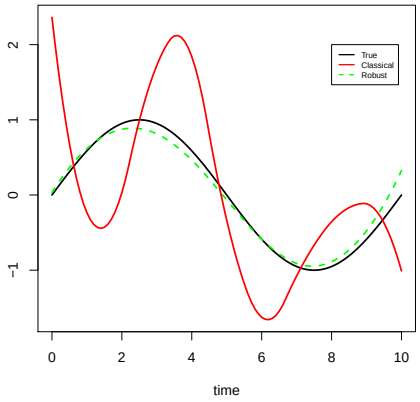
(c) Model 1



(d) $\hat{\beta}^1$

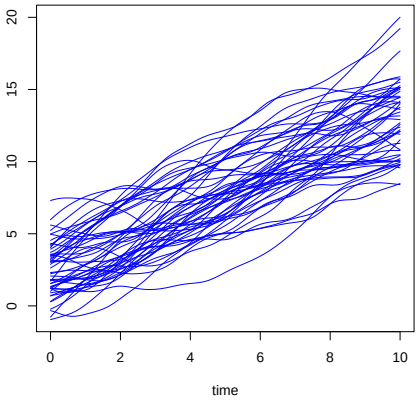


(e) Model 2

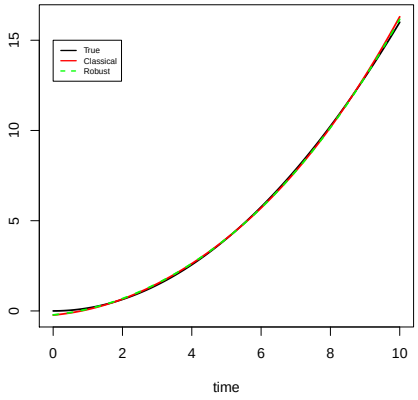


(f) $\hat{\beta}^1$

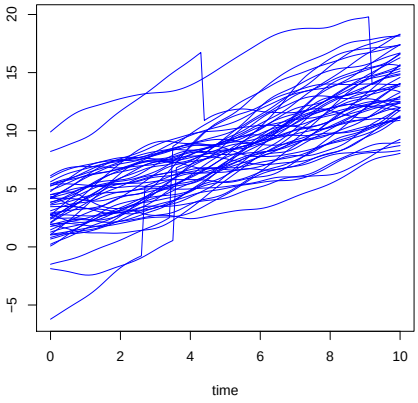
Figure 5.1: The left panel displays sample of random functions generated from simulation setting 1 with n=50, q=101. The right panel displays estimated beta function for contaminated functional predictor in left panel.



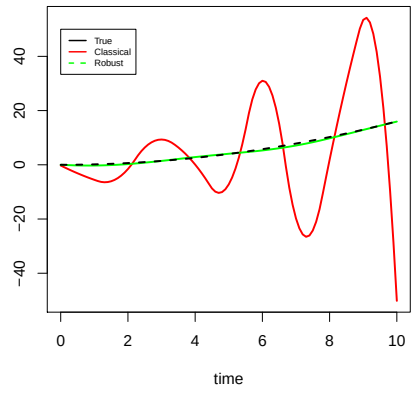
(a) Model 0



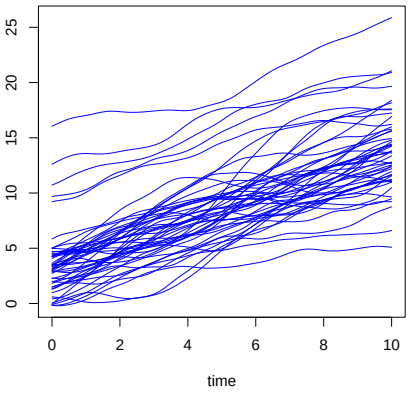
(b) $\hat{\beta}^2$



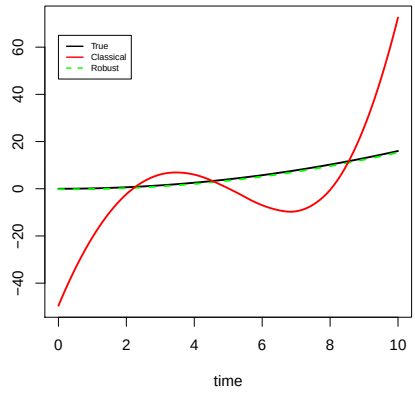
(c) Model 1



(d) $\hat{\beta}^2$

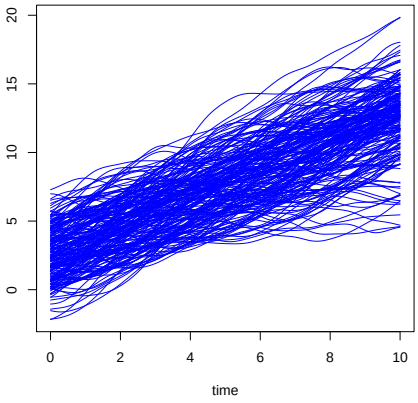


(e) Model 3

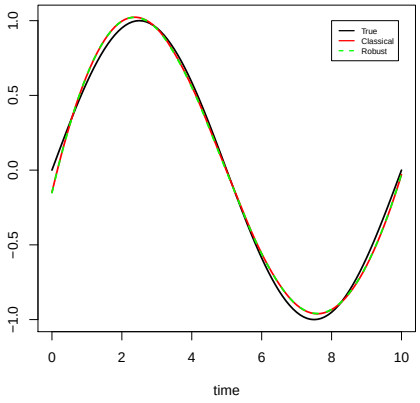


(f) $\hat{\beta}^2$

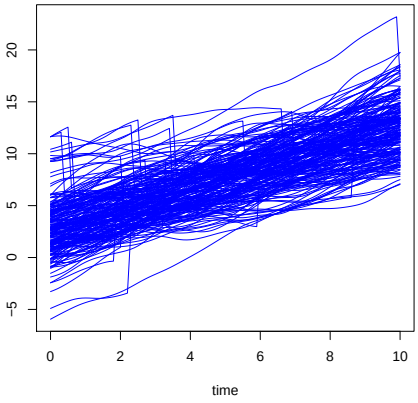
Figure 5.2: The left panel displays sample of random functions generated from simulation setting 2 with n=50, q=101. The right panel displays estimated beta function for contaminated functional predictor in left panel.



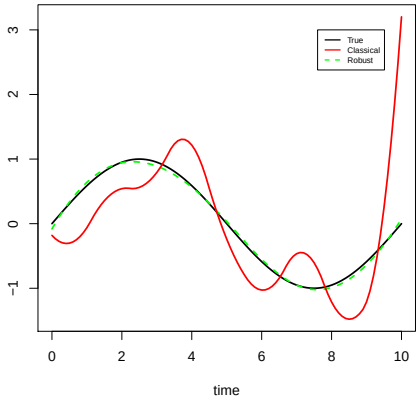
(a) Model 0



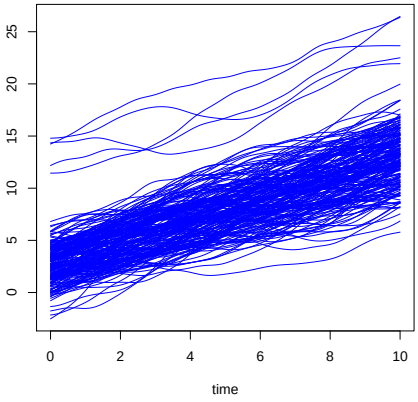
(b) $\hat{\beta}^1$



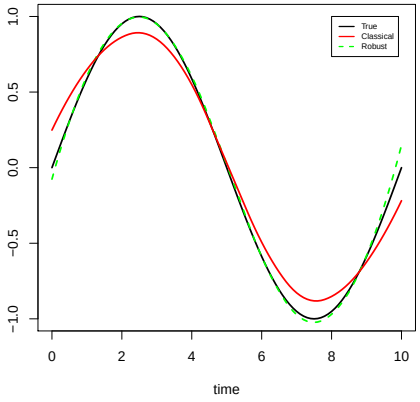
(c) Model 1



(d) $\hat{\beta}^1$

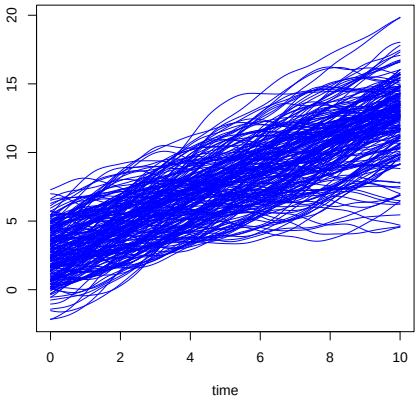


(e) Model 2

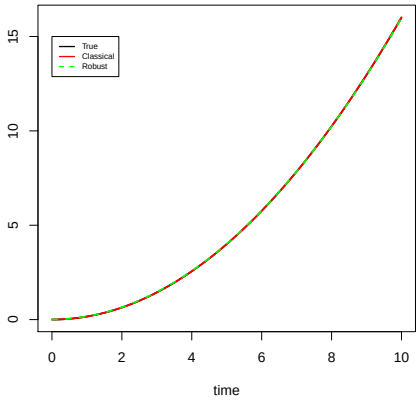


(f) $\hat{\beta}^1$

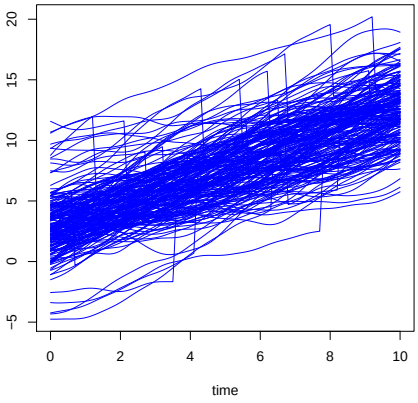
Figure 5.3: The left panel displays sample of random functions generated from simulation setting 1 with n=200, q=101. The right panel displays estimated beta function for contaminated functional predictor in left panel.



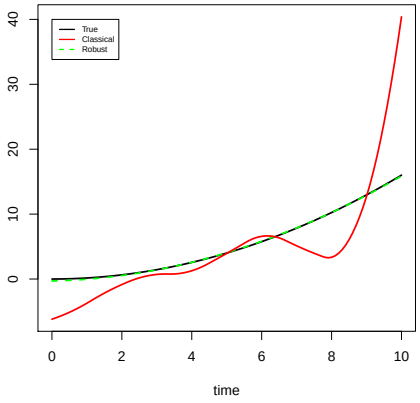
(a) Model 0



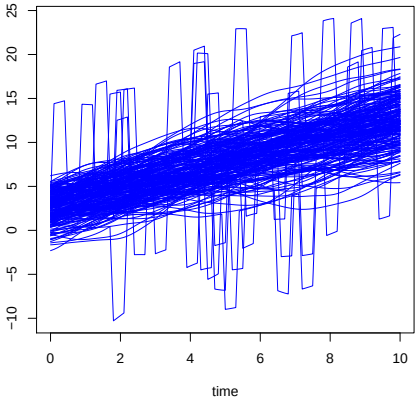
(b) $\hat{\beta}^2$



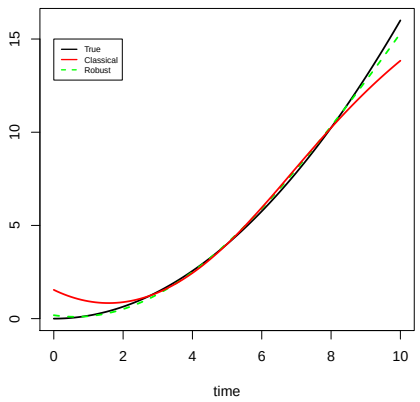
(c) Model 1



(d) $\hat{\beta}^2$



(e) Model 3



(f) $\hat{\beta}^2$

Figure 5.4: The left panel displays sample of random functions generated from simulation setting 2 with n=200, q=101. The right panel displays estimated beta function for contaminated functional predictor in left panel.

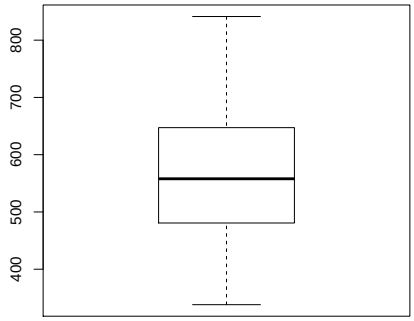


Figure 5.5: Scalar response used in simulation setting-1 and 2 for high and low dimension.

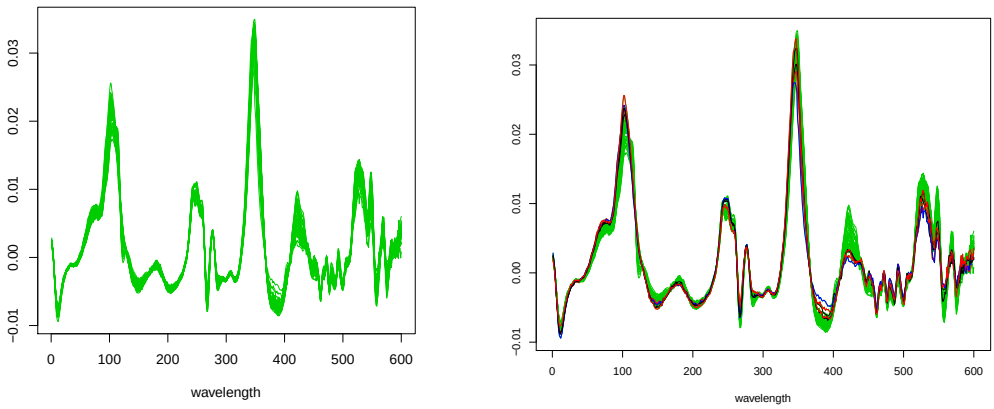
5.3.2 Data Sets

Preprocessed Biscuit-Dough Data

In this section, classical method (Goldsmith *et al.* [30]) and RFPCR (robust method) are applied on the well-known chemometrics example from Osborne *et al.* [60]. Biscuit-dough data set consists of 40 NIR spectra of biscuit dough and four response variables (percentages of fat, sucrose, flour and water). The range of original spectra is of 1100nm to 2498nm in steps of 2nm, i.e q = 700. Since the channels at the ends are known to be less reliable only q = 601 (1200nm to 2400nm in steps of 2 nm) wavelengths are used. The aim of this analysis is to predict percentage of water, based on the 40 biscuit dough samples with q = 601 wavelengths. Biscuit-dough (X-data) is displayed in Figure 5.10 (a), which consists of n = 40 curves and each curve represents q = 601 wavelengths and from the figure it is clear that the spectra have shifted due to unequal particle sizes. Therefore, the preprocessing suggested by Marx and Eilers [53] is performed by differencing the columns of data matrix to eliminate sudden shifts, Figure 5.6 (a), which results in a data set of NIR spectra in 600 dimension. Observation 23 is known to be an outlier in most analyses, so it is suggested to exclude this observation. To show robustness of RFPCR, this analysis is used on the data matrix with all 40 observations. Figure 5.7 shows a boxplot of the percentage of water (y-data).

RFPCR (Section 5.2) is applied on the preprocessed data and robust diagnostic plots are obtained (Figure 5.8). The orthogonal, score and residual distances are obtained. The orthogonal-score diagnostic plot based on robust FPCA (Section 3.3) is displayed in Figure 5.8 (a). Observations 7, 20, 21, 23 and 24 are detected as outliers. Residual diagnostic plot obtained from RFPCR indicates observations 7, 21 and 23 as outliers. Figure 5.6 (b) shows preprocessed Biscuit-dough data with outliers. Figure 5.9 displays the estimated beta functions for Biscuit-dough data by classical and robust method. There is a slight difference

between the estimated beta functions by both the methods, which clearly shows that the classical method is being affected by presence of outliers.



(a) (b)

Figure 5.6: (a)Sample curves (X-data) of the preprocessed Biscuit-dough data;(b)X-data with outliers.

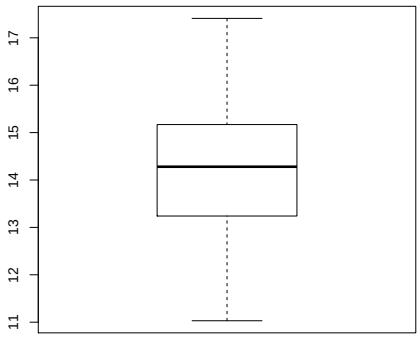
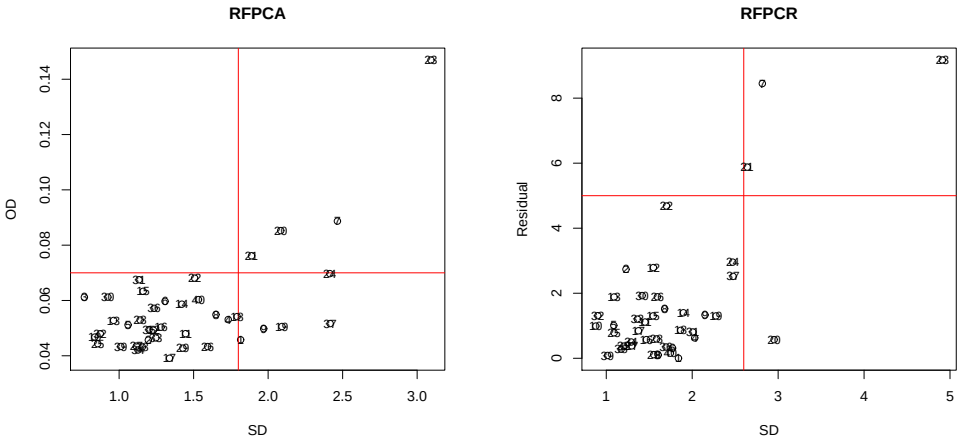


Figure 5.7: Boxplot of the percentage of water.



(a) Orthogonal-Score Plot (b) Residual-Score Plot

Figure 5.8: Diagnostic plots for the preprocessed Biscuit-dough data.

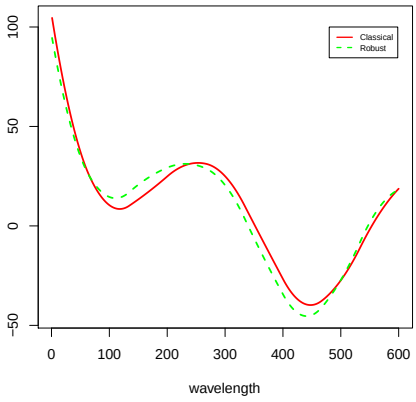
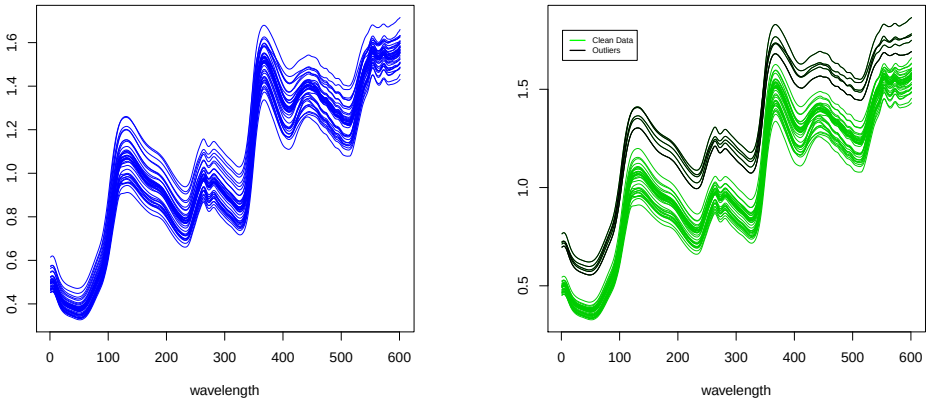


Figure 5.9: The estimated beta functions for the preprocessed Biscuit-dough data from classical and robust method .

Original Biscuit-Dough Data

In this section original Biscuit-dough data (Figure 5.10 (a)) is used to demonstrate the robustness of proposed method. The data set consists of n=40 biscuit dough samples with q = 601 wavelengths. The original data has four response variables, but for this analysis only one of them (the percentage of water) is used here (Figure 5.7). The results of analysis of preprocessed Biscuit-dough data indicates that this dataset contains outliers (observations 7, 20, 21, 23 and 24). To assess the performance of the proposed method these outliers are further contaminated by using Model-3 (Section 5.3). Figure 5.10 (b) shows original Biscuit-dough data with outliers (7, 20, 21, 23 and 24) obtained after using Model-3.

Figure 5.12 displays the diagnostic plots obtained by robust FPCA (Section 3.3) and RFPCR (Section 5.2). The orthogonal-score plot and residual-score plot indicate observations 7, 20, 21, 23 and 24 as outliers. Diagnostic plots based on robust methods detected all outliers effectively showing that the proposed approach is resistant to outlying observations. Figure 5.11 displays the estimated beta functions for Biscuit-dough data by classical and robust method. There is a difference between the estimated beta functions by both the methods clearly indicating that the classical method is being affected by the presence of outliers.



(a) (b)

Figure 5.10: (a)Sample curves (X-data) of the Biscuit-dough data;(b)X-data with outliers.

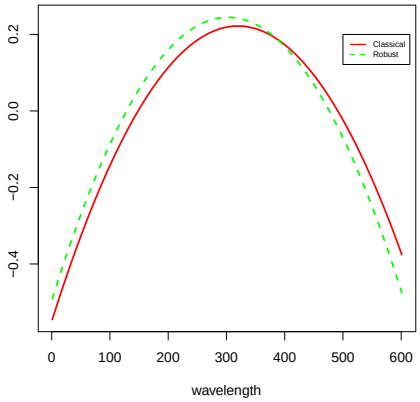
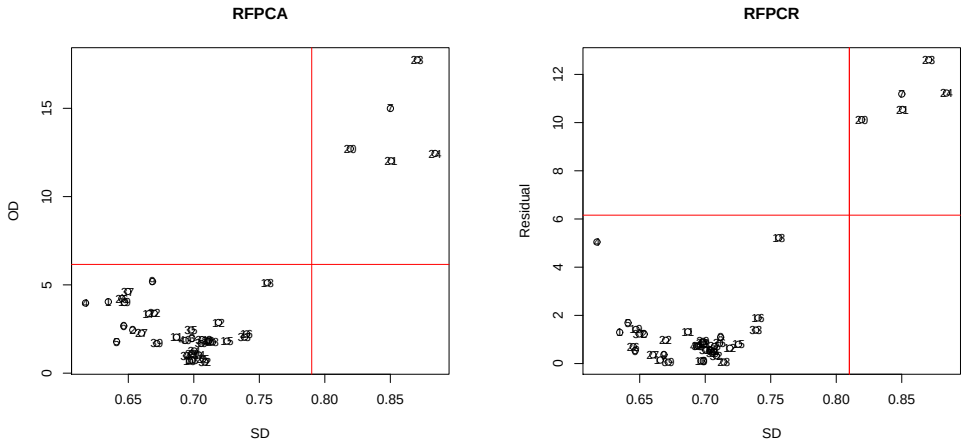


Figure 5.11: The estimated beta functions for Biscuit-dough data by classical and robust method.



(a) Orthogonal-Score Plot (b) Residual-Score Plot

Figure 5.12: Diagnostic plots for the Biscuit-dough data.

Chapter 6

Conclusions and Future Work

In this dissertation, different aspects of functional data analysis have been studied. In this chapter, final conclusions on all results obtained throughout dissertation are summarized. We also discuss some possibilities for future research.

In Chapter 1 and 2, the main concepts of FD, MFD and FPCA are introduced and reviewed. Two procedures of FPCA, for univariate and multivariate functional data; are described. To take into account the functional nature of the data, basis expansion is reviewed.

In Chapter 3, two existing robust techniques of FPCA for univariate functional data are reviewed. First method reduces the functional object to the multivariate data by basis expansion and then uses robust multivariate PCA techniques. In the second method robust principal functions are estimated by the penalized robust regression with a smoothness inducing penalty. Since abnormally observed measurements in a single curve are separately down weighed, it makes maximal use of the information. Second method is fast in computation even for long and dense functional data.

In Chapter 4, two robust MFPCA (RMFPCA1 and RMFPCA2) for dimension reduction of multivariate functional data under presence of outlying trajectories are proposed. The first method RMFPCA1 is a basis expansion approach that uses robust multivariate techniques (ROBPCA) to find robust principal functions. The second method RMFPCA2 uses Karhunen-Loeve expansion for multivariate functional data and estimates principal functions by M-estimation with roughness penalty. Tuning parameter lambda is obtained from robust leave-out-one-column cross validation. Instead of deleting a single curve it removes abnormal observations at t_j grid for assessing prediction of data. It is computationally efficient and takes less time than the existing methods. Both the methods can also be used to detect

functional outliers. An extensive simulation study is conducted and a real dataset is used to asses the performance of the RMFPCA1 and RMFPCA2. From the simulation study for RMFPCA1 based on different contamination configurations, and when contamination is around 15%, we concluded that robust PCA based multivariate functional data analysis yields better results than CPCA based multivariate functional data analysis. RMFPCA2 is tested using simulated data sets under various scenarios and real data set. The performance is compared with existing classical MFPCA. Results show that the RMFPCA2 is resistant towards many types of contamination, whilst their performance is also good for uncontaminated data sets.

In Chapter 5, the effect of outliers on one of the existing functional regression method is investigated and a new robust functional regression algorithm (RFPCR) for estimating functional coefficient is proposed. The first step consists of regressing scalar outcome on a space spanned by eigenfunctions of functional predictor. Then a robust iteratively reweighted least squares method, which gives low weights to outlying points, is proposed to estimate the regression coefficient function. Initial weights as robust distances are obtained from outlier detection methods, ROBPCA, to down-weight outlying points in predictor space. Reweighted FPCR is performed iteratively to obtain robust coefficient function. It is shown that the proposed method is very effective for uncontaminated data and it yields better results when data contain outliers.

In this dissertation, we have shown promising results for RMFPCA1, RMFPCA2 and RFPCR. There is, of course, more research to be done. We would like to extend RFPCR to multivariate case and use RMFPCA1 and RMFPCA2 in different fields of applications for instance fMRI, genomics, etc. Proposed robust techniques can be further explored in a Sobolev space. Proposed RFPCR fails to estimate coefficient function with spikes therefore, further investigation is needed. Theoretical and robustness properties of the proposed estimators also need to be studied.

Bibliography

[1] Ash, R. B. and Gardner, M. F., *Topics in stochastic processes*. New York: Academic Press, 1975.

[2] Bali, J. L., Boente, G., Tyler, D. E. and Wang, J.-L., "Robust functional principal components: A projection-pursuit approach", *The Annals of Statistics*, 39, 2852–2882, 2011.

[3] Berrendero, J., Justel, A. and Svarc, M., "Principal components for multivariate functional data". *Computational Statistics and Data Analysis*, 55, 2619-2634, 2011.

[4] Billor, Nedret, Hadi A. S., and Velleman P. F., "BACON: blocked adaptive computationally efficient outlier nominators", *Computatuional Statistics and Data Analysis*, 34, 279–298, 2000.

[5] Billor, N., Kiral, G. and Turkmen, A., "Outlier Detection Using Principal Components", *Twelfth International Conference on Statistics, Combinatorics, Mathematics and Applications*, Auburn, 2005. (Unpublished Manuscript).

[6] Billor, N., Chatterjee, S., and Hadi, A. S., "Iteratively Re-weighted Least Squares Method for Outlier Detection in Linear Regression, *American Journal for Mathematical and Management Science*, 26, 3, 229–252, 2006.

[7] Boente, G. and Fraiman, R., Discussion of "Robust principal components for functional data", by Locantore *et al.*, Test, 8, 28-35, 1999.

[8] Cardot, H., "Nonparametric estimation of smoothed principal components analysis of sampled noisy functions", *J Nonparam. Statist.*, 12, 503–538, 2000.

[9] Cardot, H., Ferraty, F. and Sarda, P., "Functional linear model", *Statist. Probab. Lett.*, 45, 11-22, 1991.

[10] Cardot, H., Ferraty, F. and Sarda, P., "Spline Estimators for the Functional Linear Model", *Statistica Sinica*, 13, 571–591, 2003.

[11] Cardot, H. and Sarda, P., "Estimation in Generalized Linear Model for Functional Data via Penalized Likelihood", *Journal of Multivariate Analysis*, 92, 24–41, 2005.

[12] Chiou, J. M. and Müller, H. G., "Diagnostic for functional regression via residual processes", *Computational Statistics and Data Analysis*, 51, 4849–4863, 2006.

[13] Cook, R. D., "Detection of influential observations in linear regression", *Technometrics*, 19, 15-18, 1977.

[14] Crambes, C., Delsol, L. and Laksaci, A., "Robust nonparametric estimation for functional data", *Functional and Operatorial Statistics*, Dabo-Niang, S. and Ferraty, F. (eds.), Springer, Heidelberg, 109-116, 2008.

[15] Craven, P. and Wahba G., "Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation", *Numerische Mathematik*, 31, 377-403, 1979.

[16] Cuevas, A., Febrero, M., Fraiman R., "Linear Functional regression: the case of fixed design and functional response", *Canad. J. Statist*, 30, 285-300, 2002.

[17] Dauxois, J., Pousse, A. and Romain, Y., "Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference", *Journal of Multivariate Analysis*, 12, 136-154, 1982.

[18] Farawayy, J. J., "Regression analysis for a functional response", *Technometrics*, 39, 254-261, 1977.

[19] Febrero, M., Galeano, P. and Gonzales-Mantegia, W., "A functional analysis of NOx levels: location and scale estimation and outlier detection", *Computational Statistics*, 22, 411-427, 2007.

[20] Febrero, M., Galeano, P. and Gonzales-Mantegia, W., "Measures of influence for the functional linear model with scalar response", *Journal of Multivariate Analysis*, 101, 327–339, 2010.

[21] Fraiman, R. and Muniz, G., "Trimmed means for functional data", *Test*, 10, 419-440, 2001.

[22] Febrero, M., Galeano, P. and Gonzales-Mantegia, W., "Influence in the Functional Linear Model with Scalar Response", In *Functional and Operatorial Statistics*, edited by Sophie Dabo-Niang and Frdric Ferraty. Physica-Verlag Heidelberg, 1 edition, 165-171, 2008.

[23] Febrero, M., Galeano, P. and Gonzales-Mantegia, W., "Measures of influence for the functional linear model with scalar response", *Journal of Multivariate Analysis*, In Press, Corrected Proof, Available online 25 December 2008.

[24] Ferraty, F. and Vieu, P., "Nonparametric Functional Data Analysis: Theory and Practice", *Springer*, New York, 2006.

[25] Gentle, J. E., "Matrix Algebra: Theory, Computations, and Applications in Statistics", Springer, 2007.

[26] Gervini, D., "Robust functional estimation using the median and spherical principal components", *Biometrika*, 95, 587-600, 2008.

[27] Gervini, D., "Detecting and handling outlying trajectories in irregularly sampled functional datasets", *The Annals of Applied Statistics*, 3, 1758-1775, 2009.

[28] Gervini, D., "Functional robust regression for longitudinal data", arXiv:1211.7332[stat.ME], 2012.

[29] Green, P. J. and Silverman, B. W., "Nonparametric Regression and Generalized Linear Models: A roughness penalty approach", Chapman & Hall, 1994.

[30] Goldsmith, J., Feder, J., Crainiceanu, C. M., Caffo, B., and Reich, D., "Penalized Functional Regression", *Journal of Computational and Graphical Statistics*, 20(4), 830–851, 2011.

[31] Hall, P. and Hosseini-Nasab, M., "On properties of functional principal components analysis", *Journal of Royal Statistical Society*, Series B, 68, 109-126, 2006.

[32] Hall, P., Müller, H.G. and Wang, J.L., "Properties of principal component methods for functional and longitudinal data analysis", *The Annals of Statistics*, 34, 1493-1517, 2006.

[33] Hastie, T. and Mallows, C., "A statistical view of some chemometrics regression tools", *Technometrics*, 35, 140-143, 1993.

[34] Huang, J. Z., Shen, H. and Buja, A., "Functional principal components analysis via penalized rank one approximation", *Electronic Journal of Statistics*, 2, 678-695, 2008.

[35] Hubert, M., and Verboven, S., "A robust PCR method for high-dimensional regressors", *Journal of Chemometrics*, 17, 438-452, 2003.

[36] Hubert, M., and Engelen, S., "Robust PCA and classification in biosciences", *Bioinformatics*, 20, 1728-1736, 2004.

[37] Hubert, M., Rousseeuw, P. J. and Branden, K. V., "ROBPCA: A new approach to Robust Principal Component analysis", *Technometrics*, 47, no. 1, 64-79, 2005.

[38] Hyndman, R. J. and Ullah, S., "Robust forecasting of mortality and fertility rates: A functional data approach", *Computational Statistics and Data Analysis*, 51, 4942-4956, 2007.

[39] Iglewicz, B. and Hoaglin D. C., "How to Detect and Handle Outliers", Milwaukee: American Society for Quality Control, WI, 1993.

[40] James, G. M., Hastie, T. J. and Sugar, C. A., "Principal component models for sparse functional data", *Biometrika*, 87, 587-602, 2000.

[41] James, G., "Generalized linear models with functional predictors", *Journal Of The Royal Statistical Society Series B*, 64, 3, 411432, 2002.

[42] Jacques, J. and Preda, C., "Model-based clustering for multivariate functional data", *Computational Statistics and Data Analysis*", doi:10.1016/j.csda.2012.12.004, 2013.

[43] Joliffe, I. T. *Principal Component Analysis*. New York: Springer-Verlag, 1986.

[44] Kayano, M. and Sadanori, S., "Functional principal component analysis via regularized Gaussian basis expansions and its application to unbalanced data", *Journal of Statistical Planning and Inference*, 139, 2388-2398, 2009.

[45] Kolmogorov, A. N., and Fomin, S. V., *Introductory real analysis* [English translation], Englewood Cliffs, NJ: Prentice-Hall, 1968.

[46] Lee, S., Shin, H. and Billor, N., "M-type Smoothing Spline Estimators for Principal Functions", *Computational Statistics and Data Analysis*, 66, 89–100, 2013.

[47] Levitin, D. J., Nuzzo, R. L., Vines, B. W. and Ramsay, J. O., "Introduction to Functional Data Analysis", *Canadian Psychology*, 48(3), 135-155, 2007.

[48] Li, Y. and Ruppert, D., "On the asymptotics of penalized splines", *Biometrika*, 60, 95, 415-436, 2008.

[49] Li, P.L. and Chiou, J.M., "Identifying cluster number for subspace projected functional data clustering", *Computational Statistics & Data Analysis* 55, 2090-2103, 2011.

[50] Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T. and Cohen, K. L., "Robust principal component analysis for functional data", *TEST*, 8(1), 1-73, 1999.

[51] Lopez-Pintado, S. and Romo J., "Depth based inference for functional data", *Computational Statistics and Data Analysis*, 51(10), 4957-4968, 2007.

[52] Marx, B. D. and Eilers P. H., "Generalized linear regression on sampled signals with penalized likelihood", *Statistical Modelling*, 1996.

[53] Marx, B.D. and Eilers, P.H., "Generalized Linear Regression on Sampled Signals and Curves: A P-spline Approach. *Technometrics*, 41, 1-13, 1999.

[54] Maronna, R., "Principal components and orthogonal regression based on robust scales", *Technometrics*, 47, 264-273, 2005.

[55] Maronna, R. A., Martin, D. and Yohai, V. J., "Robust Statistics: Theory and Methods", *Wiley*, 2006.

[56] Maronna, R. A. and Yohai, J. V., "Robust functional linear regression based on splines", (Unpublished Manuscript), 2012.

[57] Matsui, H. and Konishi, S. "Variable selection for functional regression models via the L_1 regularization", *Computational Statistics and Data Analysis*, 55, 3304-3310, 2011.

[58] Müller, H.G., "Functional modelling and classification of longitudinal data", *Scandinavian Journal of Statistics* 23, 223-240, 2005.

[59] Müller, H. and Stadtmüller, U., "Generalized functional linear models", *Annals of Statistics*, 33, 2, 774-805, 2005.

[60] Osborne, B.G., Fearn, T. , Miller, A.R. and Douglas, S., "Application of Near In- frared Reflectance Spectroscopy to the Compositional Analysis of Biscuits and Biscuit Dough, *Journal of Scientific Food Agriculture*, 35, 99-105, 1984.

[61] Peña, D. "A new statistic for influence in linear regression", *Technometrics*, 47, 1-12, 2005.

[62] Ramsay, J. O. and Dalzell, C. J., "Some tools for functional data analysis", *Journal of Royal Statistical Society*, 53(3), 539-572, 1991.

[63] Ramsay, J. O. and Silverman, B. W., "Functional Data Analysis Software", MATLAB edition.
Online at http://www.psych.mcgill.ca/faculty/ramsay/ software.html, 2001.

[64] Ramsay, J. O. and Silverman, B. W., *Functional Data Analysis*. Second Edition. New York: Springer-Verlag, 2005.

[65] Reiss, P. and Ogden, R., "Functional principal component regression and functional partial least squares", *Journal of the American Statistical Association*, 102, 479, 984-996, 2007.

[66] Rice, J. A. and Silverman, B. W., "Estimating the mean and covariance structure non-parametrically when the data are curves", *Journal of Royal Statistical Society*, 53, 233-243, 1991.

[67] Rousseeuw, P. J. "Least median of squares regression", *Journal of the American Statistical Association*, 79, 871-880, 1984.

[68] Rousseeuw, P. J. "Multivariate estimation with high break-down point", In *Mathematical Statistics and Applications*, edited by Grossmann, W. *et al.*. Reidel, Dordrecht, Vol. B., 283-297, 1985.

[69] Ruppert, D., "Selecting the number of knots for penalized splines", *Journal of Computational and Graphical Statistics*, 11(4), 737–757, 2002.

[70] Saporta, G., "Exploratory analysis methods temporal data", PhD dissertation, 1981.

[71] Sawant, P., Billor, N., and Shin, H., "Robust Functional Principal Component Analysis: Functional Outlier Detection", *Computational Statistics*, 27(1), 83-102, 2012.

[72] Shen, Q. and Xu, H., "Diagnostic for linear models with functional resposes", *Technometrics*, 49, 26-33, 2007.

[73] Silverman, B. W., "Smoothed functional principal components analysis by choice of norm", *The Annals of Statistics*", 24, 1-24, 1996.

[74] Turkmen, A. S., "Robust Partial Least Squares For Regression and Classification", PhD dissertation, Auburn University, 2008.

[75] Yao, F. and Lee, T. C. M., "Penalized spline models for functional principal component analysis", *Journal of the Royal Statistical Society*, B, 68, 3–25, 2006.

[76] Yao, F., M*ü*ller, H.G. and Wang, J.L., "Functional data analysis for sparse longitudinal data", *Journal of American Statistical Association* 100, 577-590, 2005.

[77] Zhang, J. T., "Smoothed Functional Data Analysis", PhD dissertation, The University of North Carolina at Chapel Hill, 1999.

[78] Zhang, J. T. and Chen, J., "Statistical inferences for functional data", *The Annals of Statistics*", 35, 1052-1079, 2007.