



**T.R.**  
**MIMAR SINAN FINE ARTS UNIVERSITY**  
**INSTITUTE OF SCIENCE**

**A FUNCTIONAL DATA ANALYSIS APPROACH  
TO REMOTE SENSING DATA**

**Ph.D. THESIS**

**Nihan ACAR DENİZLİ**

**Statistics Department**

**Statistics Programme**

**Thesis Advisor: Prof. Gülay BAŞARIR**

**Thesis Co-Advisor: Assoc. Prof. Pedro DELICADO**

**NOVEMBER 2016**

**T.R.**  
**MIMAR SINAN FINE ARTS UNIVERSITY**  
**INSTITUTE OF SCIENCE**

**A FUNCTIONAL DATA ANALYSIS APPROACH  
TO REMOTE SENSING DATA**

**Ph.D. THESIS**

**Nihan ACAR DENİZLİ**  
**(20117229)**

**Statistics Department**

**Statistics Programme**

**Thesis Advisor: Prof. Gülay BAŞARIR**

**Thesis Co-Advisor: Assoc. Prof. Pedro DELICADO**

**NOVEMBER 2016**

**T.C.  
MİMAR SİNAN GÜZEL SANATLAR ÜNİVERSİTESİ  
FEN BİLİMLERİ ENSTİTÜSÜ**

**UZAKTAN ALGILAMA VERİLERİNE  
FONKSİYONEL VERİ ANALİZİ YAKLAŞIMININ UYGULANMASI**

**DOKTORA TEZİ**

**Nihan ACAR DENİZLİ  
(20117229)**

**İstatistik Anabilim Dalı**

**İstatistik Programı**

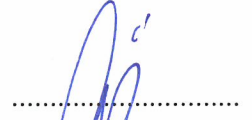
**Tez Danışmanı: Prof. Dr. Gülay BAŞARIR**

**Tez Eş Danışmanı: Doç. Dr. Pedro DELICADO**

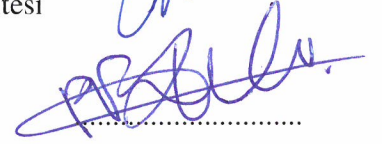
**KASIM 2016**

Nihan ACAR DENIZLI, a Ph.D. student of Mimar Sinan Fine Arts University Institute of Science 20117229 successfully defended the thesis entitled “A FUNCTIONAL DATA ANALYSIS APPROACH TO REMOTE SENSING DATA”, which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

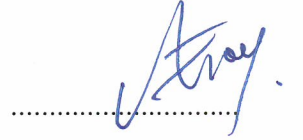
Thesis Advisor : Prof. Gülay BAŞARIR  
Mimar Sinan Güzel Sanatlar Üniversitesi



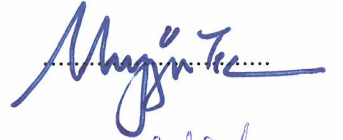
Co-advisor : Assoc. Prof. Pedro DELICADO  
Universitat Politècnica de Catalunya



Jury Members : Prof. Aydın ERAR  
Mimar Sinan Güzel Sanatlar Üniversitesi



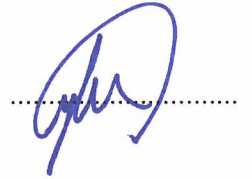
Prof. Müjgan TEZ  
Marmara Üniversitesi



Prof. Ahmet Mete ÇİLİNGİRTÜRK  
Marmara Üniversitesi



Assoc. Prof. Eylem Deniz HOWE  
Mimar Sinan Güzel Sanatlar Üniversitesi



## FOREWORD

I would like to express my sincere thanks to my supervisor Gülay Başarır who encouraged me to study this subject and helped me improve during this process with her support, to my co-supervisor and mastermind of this work Pedro Delicado who not only has answered my endless questions patiently so that I have learned a lot from him but also supported me and helped me meet new people who also study on this area, to Isabel Caballero who gently shared her data and knowledge with me and turned a collaboration into a friendship.

I appreciate my dear professors Aydın Erar and Mujgan Tez who helped me form this study with their precious comments, Gülay Ilona Telsiz Kayaoğlu who has offered me references for the theoretical background of the study and helped me organize appendices, my dearest friend Elif Ertürk and Ebru Payaslıoğlu Kılıç who helped me with English translations for their contributions.

I would like to express my gratitude to my dear professor Nalan Cinemre for her trust, support and understanding during my journey, to all the professors, PhD students and members of UPC who support me during my stay in Barcelona, to my friends from Argentina, Mexico, Spain and other countries who cheered me up and helped me a lot during this process, to Lilliana Tolomei and Daris Fernandez who are in particular just a phone away from me, to Manuel Ramos-Casals and Antoni Sisó-Almirall for their moral support, to my professors and friends in my department for their understanding and to all friends for being there with me in tough times.

I also would like to thank my mother who could bear not seeing me for a long time and is looking forward to the end of this process, my father who encouraged me for my academic career and now watching me smiling from somewhere up above, my lovely niece Ela who awaits for me to play with her, my brother who has always been my side, my little nephew and all my family for their trust, moral support and belief in me.

And lastly I would like to thank my pillar of strength, my fellow traveller, my husband Belchin Kostov (Belçin Denizli) with whom I met by means of this project and who did his best to help me and be there for me during this process.

November 2016

Nihan ACAR DENİZLİ

## TABLE OF CONTENTS

	<u>Page</u>
<b>FOREWORD.....</b>	<b>iv</b>
<b>TABLE OF CONTENTS.....</b>	<b>v</b>
<b>ABBREVIATIONS .....</b>	<b>viii</b>
<b>LIST OF TABLES .....</b>	<b>ix</b>
<b>LIST OF FIGURES .....</b>	<b>x</b>
<b>LIST OF SYMBOLS .....</b>	<b>xi</b>
<b>SUMMARY .....</b>	<b>xiii</b>
<b>ÖZET .....</b>	<b>xv</b>
<b>1. INTRODUCTION .....</b>	<b>1</b>
<b>2. GENERAL FRAMEWORK OF FUNCTIONAL DATA .....</b>	<b>6</b>
2.1 The Structure of Functional Data .....	6
2.2 Functional Exploratory Data Analysis .....	7
2.2.1 Mean, Variance and Covariance Functions .....	7
2.2.2 Depth Measures .....	10
2.3 Representing Functional Data .....	12
2.3.1 Basis Representation .....	13
2.3.1.1 Fourier basis.....	14
2.3.1.2 B-spline basis.....	15
2.3.1.3 Bias - Variance Trade Off .....	16
2.3.2 Kernel Smoothing.....	17
2.3.3 Functional Principal Components (FPC) Basis.....	19
2.3.4 Functional Partial Least Squares (FPLS) Basis.....	19
2.4 Smoothing Functional Data.....	20
2.4.1 Fitting by Least Squares Criterion.....	21
2.4.1.1 Ordinary (Unweighted) Least Squares Fit.....	21
2.4.1.2 Weighted Least Squares Fit .....	22
2.4.2 Fitting by Localized Least Squares Criterion.....	23
2.4.2.1 Localized basis functions smoothing.....	23
2.4.2.2 Local polynomial smoothing .....	24
2.4.3 Fitting by Roughness Penalty Approach .....	24
2.4.3.1 The Choice of Smoothing Parameter.....	26
<b>3. FUNCTIONAL PRINCIPAL COMPONENTS ANALYSIS .....</b>	<b>28</b>
3.1 The Approximation of Principal Components Analysis .....	29
3.2 Principal Component Analysis for Functional Data.....	32

<b>4. FUNCTIONAL LINEAR MODELS.....</b>	<b>40</b>
4.1 Functional Linear Regression Models (FLRM) for Scalar Responses.....	41
4.1.1 Functional Linear Regression Models (FLRM) with Basis Representation.....	42
4.1.2 Roughness Penalty Approach.....	44
4.1.3 Functional Principal Component Regression (FPCR).....	46
4.1.4 Functional Partial Least Squares Regression (FPLSR).....	51
<b>5. APPLICATION.....</b>	<b>57</b>
5.1 Material and Methods.....	58
5.1.1 Remotely Sensed Data.....	58
5.1.2 In-situ data.....	59
5.1.3 Validation Procedure.....	61
5.1.4 Statistical Methods.....	61
5.1.4.1 Generalized Additive Models (GAM).....	62
5.1.4.2 Least Absolute Shrinkage and Selection Operator (LASSO) Model.....	63
5.2 Results.....	64
5.2.1 Exponential Regression Models.....	64
5.2.2 GAM and LASSO Models.....	68
5.2.3 Functional Linear Regression Models.....	70
5.3 Concluding Remarks.....	76
<b>6. SIMULATION STUDY.....</b>	<b>77</b>
6.1 The Choice of Functional Predictor.....	77
6.2 Generating the Response.....	78
6.3 The Comparison of the Prediction Models.....	80
6.4 Simulation Results.....	81
<b>7. CONCLUSION.....</b>	<b>85</b>
<b>REFERENCES.....</b>	<b>87</b>
<b>APPENDICES.....</b>	<b>96</b>
Some Definitions On Functional Spaces.....	98
1.1 Metric Spaces and Its Properties.....	98
1.2 Inner Product Spaces and Its Properties.....	98
1.3 Normed Spaces and Its Properties.....	99
1.4 Hilbert Spaces.....	100
Operators and Some Usefull Theorems for Functional Data.....	101
2.1 Linear Operator, Integral Operator and Hilbert-Schmidt Operator.....	101
2.2 Covariance Operator and Its Properties.....	101
2.3 Singular Value Decomposition of a Linear Operator.....	102
2.4 Spectral Decomposition of a Hilbert-Schmidt Operator.....	102
2.5 Parseval's Equality.....	103
2.6 Riesz Representation Theorem.....	103
2.7 Mercer's Theorem.....	103
2.8 Karhunen-Loève Expansion.....	103
Simulation Function Implemented In R.....	104
<b>CURRICULUM VITAE.....</b>	<b>109</b>





## ABBREVIATIONS

<b>FDA</b>	: Functional Data Analysis
<b>MRI</b>	: Magnetic Resonance Imaging
<b>HSD</b>	: Half-Space Depth
<b>FMD</b>	: Fraiman-Muniz Depth
<b>FD</b>	: Functional Depth
<b>SD</b>	: Simplicial Depth
<b>MD</b>	: Modal Depth
<b>RTD</b>	: Random Tukey Depth
<b>RPD</b>	: Random Projection Depth
<b>MSE</b>	: Mean Square Error
<b>KNN</b>	: $K$ Nearest Neighbourhood
<b>FPC</b>	: Functional Principal Components
<b>LS</b>	: Least Squares
<b>PLS</b>	: Partial Least Squares
<b>FPLS</b>	: Functional Partial Least Squares
<b>SSE</b>	: Sum of Square Errors
<b>OLS</b>	: Ordinary Least Squares
<b>WLS</b>	: Weighted Least Squares
<b>LLS</b>	: Localized Least Squares
<b>SSE<sub>LLS</sub></b>	: Sum of Square Errors for Localized Least Squares
<b>LPLS</b>	: Localized Polynomial Least Squares
<b>SSE<sub>LPS</sub></b>	: Sum of Square Errors for Localized Polynomial Least Squares
<b>SSE<sub>PEN</sub></b>	: Penalized Sum of Square Errors
<b>PEN</b>	: Penalization
<b>CV</b>	: Cross Validation
<b>GCV</b>	: Generalized Cross Validation
<b>Df</b>	: Degrees of freedom
<b>FANOVA</b>	: Functional Analysis of Variance
<b>FLM</b>	: Functional Linear Models
<b>FLRM</b>	: Functional Linear Regression Models
<b>PCR</b>	: Principal Components Regression
<b>PLSR</b>	: Partial Least Squares Regression
<b>FPCR</b>	: Functional Principal Components Regression
<b>FPLSR</b>	: Functional Partial Least Squares Regression
<b>MSC</b>	: Model Selection Criteria
<b>SIC</b>	: Schwarz Information Criterion
<b>SICc</b>	: Corrected Schwarz Information Criterion
<b>AIC</b>	: Akaike Information Criterion
<b>AICc</b>	: Corrected Akaike Information Criterion
<b>TSS</b>	: Total Suspended Solids
<b>Rrs</b>	: Reflectance of Remote Sensing

## LIST OF TABLES

1.1	Linear and Nonparametric Models for Multivariate and Functional Data .....	4
3.1	The comparison of PCA and FPCA methods .....	28
4.1	Functional Linear Regression Models .....	40
5.1	L2 flags.....	59
5.2	MEP Values for Exponential Regression Models .....	67
5.3	MEP Values for PC and PLS Regression Models.....	68
5.4	AMEP Values for GAM.....	69
5.5	AMEP Values for FLRM .....	72
5.6	Parameter Estimates of FPCR CV model .....	73
5.7	Parameter Estimates of FPLSR CV model .....	74
6.1	Simulation Results for Sample Size 25.....	82
6.2	Simulation Results for Sample Size 50.....	82
6.3	Simulation Results for Sample Size 100.....	83

## LIST OF FIGURES

5.1	The study area and ROI. ....	60
5.2	The correlation plot between variables .....	65
5.3	The scatter plot between in situ TSS and Rrs values at wavelengths 665 nm and 681 nm.....	66
5.4	The Residual Plots of Used Exponential Models .....	67
5.5	Mean Squared Error versus a range of values of lambda. ....	69
5.6	LASSO Model. ....	70
5.7	Functional Explanatory Variables.....	71
5.8	The summary graphs for the model FPCR CV between log TSS and Rrs .....	73
5.9	The summary graphs for the model FPLSR CV between log TSS and Rrs .....	74
5.10	The functional components of cross validated FPCR and FPLSR Models.....	75
5.11	The parameter estimate of cross validated FPCR and FPLSR Models	76
6.1	Images obtained from satellite data for different wavelengths. ....	78
6.2	Functional observations used for simulation. ....	78
6.3	The Simulation Design .....	81

## LIST OF SYMBOLS

$X$	: Random variable defined on a vector space
$\chi$	: Functional random variable defined on a functional space
$\mathbb{R}$	: Real line
$F$	: Functional space
$L^2(T)$	: Functional space of square integrable functions defined on T
$t_j$	: Argument values of a functional random variable
$n$	: Number of argument values of a functional random variable
$T$	: The interval on which the functional variable is defined
$N$	: The number of functional observations
$\mathbb{E}$	: The expected value
$\mu_\chi(t)$	: The mean function of the population
$\sigma_\chi^2(t)$	: The variance function of the population
$c_\chi(t, s)$	: The covariance function of the population
$r_\chi(t, s)$	: The correlation function of the population
$\Gamma_\chi$	: The covariance operator of the population
$\hat{\mu}_\chi(t)$	: The mean function of the sample
$\hat{\sigma}_\chi^2(t)$	: The variance function of the sample
$\hat{c}_\chi(t, s)$	: The covariance function of the sample
$\hat{r}_\chi(t, s)$	: The correlation function of the sample
$\hat{\Gamma}_\chi$	: The covariance operator of the sample
$\hat{c}_{\chi, \gamma}(t, s)$	: The cross-covariance function
$\hat{r}_{\chi, \gamma}(t, s)$	: The cross-correlation function
$F_N$	: Cumulative distribution function
$y_i$	: Observations of a random variable $y$
$\phi_k$	: Basis function
$c_k$	: The vector of coefficient
$K$	: The number of basis functions
$D$	: The derivative operator
$\tau_l$	: The breakpoints of a spline
$K(\cdot)$	: Kernel function
$h$	: The bandwidth in kernel smoothing
$w_i(t)$	: Weight function
$S_j(t)$	: Linear smoother
$f_{ij}$	: Principal component scores
$\xi_j$	: Eigenfunctions
$\varphi_l$	: FPLS components
$v_{il}$	: FPLS scores
$\mathbf{S}$	: Smoothing matrix
$\mathbf{R}$	: Penalty matrix

$\lambda$	: Smoothing parameter in spline smoothing
$\omega_j$	: Weight value of the argument $t_j$
$\Xi(v)$	: Penalty function
$x_{ij}$	: Centered observation vectors
$\mathbf{u}_j$	: The weight vector of the $j$ 'th component
$f_{ik}$	: The weight vector of the $k$ 'th component $i$ 'th observation
$\chi_i^c(t)$	: Centered observation curves
$\hat{\xi}_j(t)$	: The empirical weight function of the $j$ 'th principal component
$Y_i$	: The scalar response variable
$\gamma_i(t)$	: The functional response variable
$\varepsilon_i(t)$	: The functional error term
$\boldsymbol{\varepsilon}_i$	: The vectorel error term
$\mathbf{Z}$	: $N \times p$ dimension design matrix
$\boldsymbol{\beta}$	: $p \times 1$ dimension parameter vector
$\beta(s)$	: The parameter function
$\langle \cdot, \cdot \rangle$	: Inner product
$\hat{Y}_i$	: The estimate of the scalar response
$\hat{\beta}(s)$	: The estimate of the parameter function
$\hat{\beta}_{FPC}(s)$	: The FPCR estimate of the parameter function
$\hat{\beta}_{FPLS}(s)$	: The FPLSR estimate of the parameter function
$\beta^*(s)$	: The basis expansion of the functional parameter estimate
$\theta_l$	: The basis function used to expand parameter function
$b_l$	: The basis expansion coefficients of the parameter function
$\hat{b}_l$	: The basis coefficient vector of the function $\hat{\beta}$
$\mathbf{C}$	: The vector of basis coefficients of function $\chi_i(s)$
$\mathbf{J}_{\phi\theta}$	: $K \times L$ dimension matrix that is composed of the inner product of $\phi$ and $\theta$ basis functions
$\boldsymbol{\zeta}$	: $L + 1$ length parameter vector
$\hat{\boldsymbol{\zeta}}$	: $L + 1$ length parameter vector estimate
$\mathbf{H}$	: The hat matrix
$h_{ii}$	: The diagonal elements of the hat matrix
$p_N$	: A parameter used for generalization of information criteria
$\tilde{Y}_{exp}$	: The response vector generated from exponential regression model
$\tilde{Y}_{basis}$	: The response vector generated from FLRM with B-spline Basis expansion
$\tilde{Y}_{FPC}$	: The response vector generated from FPCR model with all components
$\tilde{Y}_{FPCv}$	: The response vector generated from FPCR model with the components chosen by CV criterion

# **A FUNCTIONAL DATA ANALYSIS APPROACH TO REMOTE SENSING DATA**

## **SUMMARY**

Functional Data Analysis (FDA) is a statistical field which has gained importance due to the progress in modern science, mainly in the ability to measure in continuous time results of an experiment and the possibility to record them. Many methods such as discriminant analysis, principal components analysis and regression analysis that are used on vector spaces for classification, dimension reduction and modelling have been adapted to the functional case. FDA is concerned on variables that are defined on a continuum or that have continuous structure. Therefore, FDA has an important role in the analysis of spectral data sets and images that are mostly recorded in the fields of chemometry, medicine and ecology. Especially in ecology, the analysis of images that are recorded in satellite sensors inform us in a fast and economical way about the use of land, the crop production in land, the water pollution and the amount of minerals include the water.

The aim of this study is to propose the use of FDA approach and to predict the amount of Total Suspended Solids (TSS) in the estuary of Guadalquivir river in Cadiz on remote sensing data by using different Functional Linear Regression Models (FLRM). Besides, it is purposed to compare the results obtained from various FLRMs and classical statistical methods practically, to design a simulation study in order to support findings and to determine the best prediction model.

In accordance with this purpose, the following chapter reviews the studies on this area and the literature on FDA.

In the second chapter, the general framework of FDA is explained, descriptive statistics and exploratory data analysis for functional variables are handled and the methods which can be used to convert a discrete variable to a functional variable are presented.

In the third chapter, the theoretical background of the extension of principal component analysis from multivariate case to the functional case is explained.

In the fourth chapter, FLRM for the case of scalar response is explained comprehensively.

In the fifth chapter, the acquisition of the satellite images and in-situ data are explained in detail. In order to estimate the amount of TSS on satellite data, some classical statistical models and functional models are used and their results are compared.

In the sixth chapter, a simulation study is designed to support findings and to measure the performance of the FLRMs and later its results are presented.

As to final chapter, the results obtained are discussed and compared to the results of the studies in the literature and prospective suggestions are made.

Key Words: Functional data analysis, functional linear regression models, functional principal components regression, functional partial least squares regression, remote sensing data.





## UZAKTAN ALGILAMA VERİLERİNE FONKSİYONEL VERİ ANALİZİ YAKLAŞIMININ UYGULANMASI

### ÖZET

Son zamanlarda modern bilimin gelişmesi ve özellikle de verilerin sürekli zamanda ölçülme eğiliminin artmasıyla Fonksiyonel Veri Analizi (FVA) istatistikte önem kazanmıştır. Klasik istatistikte vektör uzayları üzerinde tanımlanan diskriminant analizi, temel bileşenler analizi, regresyon analizi gibi sınıflama, boyut indirgeme ve modelleme amaçlarıyla kullanılan pek çok yöntem fonksiyonel duruma uyarlanmıştır. FVA, bir süreklilik üzerinde tanımlanan ya da sürekli yapıya sahip olan değişkenlerin analizi ile ilgilenir. Bu nedenle, FVA kemometri, tıp ve çevrebilim gibi alanlarda bir spektrum üzerinde tanımlanan ya da görüntü olarak kaydedilen verilerin analizinde önemli yere sahiptir. Özellikle çevrebilimde, uydulardan elde edilen görüntü verilerinin analiz edilmesi karada toprak kullanımı, sudaki kirlilik ya da mineral oranı gibi pek çok konuda daha ucuz ve hızlı bir şekilde bilgi sahibi olmamızı sağlar.

Bu çalışmanın amacı, FVA yaklaşımını önermek, Fonksiyonel Doğrusal Regresyon Modellerini (FDRM) uzaktan algılama verileri üzerinde İspanya'nın Cadiz bölgesinde bulunan Guadalquivir nehir ağzında biriken katı madde oranının tahmin edilmesi için uygulamak ve çeşitli FDRM ile klasik istatistiksel modellerden elde edilen sonuçları uygulamalı olarak karşılaştırmaktır. Ayrıca, uygulama sonuçlarının desteklenmesi amacıyla bir simülasyon çalışması tasarlamak ve bu şekilde en iyi tahmin performansını gösteren modelleri belirlemek amaçlanmaktadır.

Bu amaç doğrultusunda çalışmanın izleyen bölümünde, fonksiyonel veri analizi konusunda literatür taraması yapılarak bu alandaki çalışmalardan bahsedilmiştir.

İkinci bölümde, fonksiyonel veri analizi genel çerçevesi ile açıklanmış, fonksiyonel değişkenler için betimsel istatistiklerin hesabı ve keşifsel veri analizi konularına değinilerek kesikli bir değişkeni fonksiyonel değişkene dönüştürmek için kullanılacak yöntemler tanıtılmıştır.

Üçüncü bölümde, önemli bir boyut indirgeme yöntemi olan temel bileşenler analizinin çok değişkenli analizden fonksiyonel veri analizine uyarlanması ve teorik alt yapısı ele alınmıştır.

Dördüncü bölümde, yanıt değişkenin skaler olduğu durumda kullanılacak FDRM kapsamlı olarak açıklanmıştır.

Beşinci bölümde, uygulamada kullanılan uydu verilerinin ve skaler yanıt vektörünü oluşturan katı madde oranı değerlerinin nasıl elde edildiği anlatılmış, katı madde oranının uydu verilerinden tahmin edilmesi için bazı klasik istatistiksel yöntemler ve fonksiyonel yöntemler kullanılarak elde edilen sonuçlar karşılaştırılmıştır.

Altıncı bölümde, uygulamadan elde edilen sonuçların desteklenmesi ve kullanılan fonksiyonel modellerin tahmin performansının karşılaştırılabilmesi için tasarlanan simülasyon çalışması anlatılmış ve sonuçları verilmiştir.

Çalışmanın son bölümünde ise elde edilen sonuçlar genel olarak yorumlanmış, literatürde yapılan çalışmalarla karşılaştırılmış ve ileriye yönelik önerilerde bulunulmuştur.

Anahtar kelimeler: Fonksiyonel veri analizi, fonksiyonel doğrusal regresyon modelleri, fonksiyonel temel bileşenler regresyon, fonksiyonel kısmi en küçük kareler regresyon, uzaktan algılama verileri.



# 1 INTRODUCTION

Due to the progress of modern science, recently it is more common to work on large data bases in many fields such as medicine, economy and environmental studies. The analysis of large data sets require further developments of existing methods. Functional Data Analysis (FDA) is an extension of multivariate statistical analysis methods which concerns with the analysis of curves instead of vectors. FDA is an important approach in analyzing the data sets measured over time, space or any other continuum. Therefore, it has gained importance in the fields of biomedical science and bioinformatics (Wu and Müller, 2010; Escabias et al., 2012). According to Ullah and Finch (2013), 21% of the published articles between the years 2005-2010 that use FDA approach were related to the biomedical science. Several applications of FDA are found in the literature including biomechanics, magnetic resonance imaging (MRI) and gene expression profiles (Ullah and Finch, 2013; Ramsay, 2000; Viviani et al., 2005; Müller et al., 2008).

Differently from multivariate data analysis, functional data analysis consider a sequence of individual observations as a whole single entity under the assumption that the process generating the data is smooth. For a function to be smooth means that it has differentiability until a certain degree.

The functional data analysis methods help to reveal additional information contained in the functions or their derivatives, which is not available through traditional methods (Levitin et al., 2007). The main goals of functional data analysis are to represent the data in ways that aid further analysis, to display the data so as to highlight various characteristics, to study the pattern and variation among the data, to explain variation in an outcome or dependent variable by using input or independent variable information, to compare two or more sets of data where it is possible to contain different sets of replicates of the same function or different functions for a common set of replicates. In this sense, some of the methods that are adapted from multivariate

statistical analysis to functional data analysis are functional descriptive statistics, functional principal components analysis, functional linear models and functional canonical correlation analysis (Ramsay and Silverman, 2005).

Because of the data structure that consists of repeated measurements over time per observation, functional data analysis is involved with longitudinal data analysis which is a multivariate method. Although the data structure that two methods deal with seem similar, in fact they are quite different from each other.

Longitudinal data analysis (LDA) concerns with small number of repeated measurements which are collected at different time points per subject, whereas functional data analysis deals with large number of repeated measurements which can be collected at different time points, frequencies or any other continuum. Besides, longitudinal data is often viewed as a random vector in a parametric model while functional data is handled as a smooth process observed at discrete time points with few model structural assumptions which yields nonparametric or semi parametric approaches (Davidian et al., 2004). A good comparison of the methodology of FDA and LDA can be found in the studies of Rice (2004) and Hall et al. (2006).

The importance of functions and the difference between the classical and modern concepts of mathematical analysis was explained by Dieudonné (1960) :

"The idea that a function  $f$  is a single object, which may itself vary and is in general to be thought of as a point in a large functional space; indeed it may be said that one of the main differences between the classical and modern concepts of Analysis is that, in classical mathematics, when one writes  $f(x)$ ,  $f$  is visualized as "fixed" and  $x$  as variables, whereas nowadays both  $f$  and  $x$  are considered as variables (and sometimes it is  $x$  which is fixed, and  $f$  becomes the varying object)."

Although the importance of functional analysis had been mentioned in 60's, the first articles about the context of using functions as random observations were published in 80's (Dauxois et al., 1982; Ramsay, 1982; Besse and Ramsay, 1986). Ramsay (1982) gives an functional analytic view of the data and explains how to extend the statistical concepts to functional data where Besse and Ramsay (1986) offers a new technique for principal components analysis when the data consists of sampled functions. The approach for estimating the mean function of a sample of curves which are assumed

as independent observations of a random function coming from a stochastic process is defined in 90's by Rice and Silverman (1991). Following this, the notion of functional data analysis is first proposed in the publication of Ramsay and Dalzell (1991) in which classical multivariate statistical methods such as linear modeling and principal components analysis were applied into the infinite dimensional space. Later Silverman (1996) proposed a new approach with a smoothing parameter. As mentioned in Manteiga and Vieu (2007) and Shang (2014) principal component analysis and linear regression modeling were the first classical multivariate statistical methods that were adapted to functional data. Functional principal components analysis (FPCA) is an important method that since it can be used with different aims such as exploring, modelling and clustering. Therefore FPCA has found wide field of application from modeling fMRI data (Viviani et al., 2005) to modeling aircraft trajectories (Nicol, 2013). A wide literature review of FPCA can be found in Shang (2014).

Functional linear regression models (FLRM) were proposed for explaining the relationship between two variables where at least one of them has functional structure. FLRM was firstly developed for the case where both the response and the predictor are functions (Ramsay and Dalzell, 1991). Then, they were extended to the case where the covariates are functional but the response is scalar and to the case where the response is functional but the covariates are scalar (Cardot et al., 1999; James, 2002). Following that, functional logistic regression models were developed to predict a binary response variable from functional covariates (Escabias et al., 2004; Müller and Stadtmüller, 2005). In order to avoid problems arise from multicollinearity and high dimensionality, the jointly use of functional logistic regression and principal component analysis was suggested by Escabias et al. (2005). Different applications of logistic regression on functional data can be found in Ratcliffe et al. (2002); Müller and Stadtmüller (2005); Aguilera (2008). Functional logistic regression models were later extended to generalized functional linear models (James, 2002; Müller and Stadtmüller, 2005). Cuevas et al. (2004) proposed a an ANOVA test for functional data. Various applications of functional data analysis approach was adapted to the areas such as the analysis of pinch force of human fingers (Ramsay et al., 1995), the analysis of lip motion (Ramsay et al., 1996), the analysis of handwritings (Ramsay, 2000), the analysis of growth curves, climatic variation, criminology, the nondurable good index

**Table 1.1:** Linear and Nonparametric Models for Multivariate and Functional Data

Data Type	Linear Models	Nonparametric Models
Multivariate	$X \in \mathbb{R}^p$ $Y = a_0 + \sum_{j=1}^p a_j X_j + \varepsilon$ $C = \{r \text{ linear}\}$	$X \in \mathbb{R}^p$ $Y = r(X_1, \dots, X_p) + \varepsilon$ $C = \{r \text{ continuous}\}$
Functional	$\chi \in \mathbb{F} = L^2(T)$ $Y = \int_T \rho(t) \chi(t) dt + \varepsilon$ $C = \{\chi \mapsto \int_T \rho(t) \chi(t) dt \in \mathbb{R}\}$	$\chi \in \mathbb{F} = L^2(T)$ $Y = r(\chi) + \varepsilon$ $C = \{r \text{ continuous}\}$

and the movement of hip and knee ankles (Ramsay and Silverman, 2002, 2005), the analysis of musical performance (Almansa and Delicado, 2009) and the analysis of aircraft trajectories (Nicol, 2013).

Functional models are divided in parametric and nonparametric models as in the multivariate case. But the structure of functional models are quite different from multivariate models in terms of the space that the data is defined. The new approach that combines nonparametric analysis and the functional data was proposed by Ferraty and Vieu (2003) where the relationship between a curve and a categorical response is investigated by means of nonparametric Kernel methods. Nonparametric functional approach was extended to supervised and unsupervised curve classification methods. Different applications of nonparametric methods were done related to the areas like chemometry, speech recognition and electricity consumption (Ferraty and Vieu, 2006).

Considering the general form of regression models,

$$Y = r(X) + \varepsilon$$

the distinction between parametric and nonparametric models for multivariate and functional data can be summarized as in the Table (1.1) where  $X$  denotes the multivariate variables defined on a vector space  $\mathbb{R}$  and  $\chi = \chi(t)$ ,  $t \in T = (t_1, t_n)$  denotes the functional variables defined on a bounded functional space  $L^2(T)$ .

Another important tool in functional data analysis is to determine outliers. Different outlier detection methods for functional data based on depth measures were offered (Febrero et al., 2007, 2008). Following that, robust outlier detection methods for functional data were also studied (Sawant et al., 2012).

Recently, different literature reviews in FDA have been published (Cuevas, 2014; Wang et al., 2016; Morris, 2014; Reiss et al., 2016). Among these Cuevas (2014) explained the theory of statistics for functional data and focused on the recent advances on regression, classification and dimension reduction methods, Wang et al. (2016) gave a summary of the developments in FDA by time, Morris (2014) focused on functional regression and Reiss et al. (2016) gave an extensive literature review of functional regression with a scalar response. Due to the advances in FDA and the popularity of the subject, many journals devoted special issues to this topic such as *Statistica Sinica*, issue 14, 3 (2004), *Computational Statistics*, 22, 3 (2007), *Computational Statistics & Data Analysis*, 51, 10 (2007) and *Journal of Multivariate Analysis*, 101, 2 (2010) with the introductions of Davidian et al. (2004); Manteiga and Vieu (2007); Valderrama (2007) and Ferraty (2010). Recently, functional data analysis gain importance in national literature (Keser, 2010, 2007; Gündüz, 2012; Sözen, 2014; Özçomak and Gündüz, 2014). Keser (2010) and Sözen (2014) analysed the rainfall data in the Aegean and Black Sea regions through functional principal components analysis. In the study of Gündüz (2012), functional canonical correlation analysis is used as a tool to estimate the Istanbul Stock Exchange index 30. The national studies are mostly focused on Functional Principal Components and Functional Canonical Correlation Analysis. Lately, there aren't any studies that focus on other areas of functional data in the national literature.

## 2 GENERAL FRAMEWORK OF FUNCTIONAL DATA

Classical statistical analysis methods are concerned with the observations that are points. However, functional data analysis is concerned with observations in the form of real functions. The functional data sample consists of  $N$  functions  $\chi_1, \chi_2, \dots, \chi_N$  defined on some set of  $T$ . In functional data analysis the sample space is an infinite dimensional functional space rather than a vector space. Due to this infinite-dimensional nature of the sample space FDA is regarded as a new branch of the statistical theory (Cuevas, 2014).

### 2.1 The Structure of Functional Data

As it can be seen in Table (1.1), the most important distinction between the multivariate and functional variable is the space where the variable is defined. A functional variable  $\chi$  is a random variable taking values in a functional space  $F$  which can be either a metric or a semi-metric space in general. A functional data set consists of a sample of  $N$  functional variables  $\chi_1, \chi_2, \dots, \chi_N$  (can be also denoted by  $\chi_1(t), \chi_2(t), \dots, \chi_N(t)$ ) that depend on argument values  $t_j, (j = 1, \dots, n)$  which are identically distributed as  $\chi$  (Febrero-Bande and Oviedo de la Fuente, 2012; Ferraty and Vieu, 2006). The argument values  $t_j$  can define same measurement points for all observations or can vary from record to record (Ramsay and Silverman, 2005).

Considering the definition of functional data set, any process  $\{\chi(t) : t \geq 0\}$  that is observed in discrete grid points  $t_1, t_2, \dots, t_n$  can also be regarded as a functional data. But this type of discrete processes are not always functional. According to Cuevas (2014):

"There are at least two reasons that could lead us to consider this as a functional data: first the probability (at least theoretical) of observing the phenomenon in a much finer



grid and, in the limit, to observe  $x(t)$  at any fixed instant  $t$ . Second, the choice of a functional model to approximately represent it."

Some properties of a functional data set can be listed as follows (Ramsay and Silverman, 2002):

- Although the functional data can be observed at discrete points, conceptually it is defined as continuous.
- The individual datum of a functional data set is a whole function rather than a single observed point. Although it is assumed that the observed functional datum is independent from one another, there is no assumption about the independence of observations that consist the same functional datum.
- The functional data consists of functions of time. Time is not considered as a variable as in the multivariate context.
- Even though the functional data is not necessarily smoothly observed, smoothness is required for modeling the functional data.

Similarly to classical data analysis to estimate means and standard deviations and outlier detections are important steps in functional data analysis. However, because of the functional structure of the data the estimation of mean and standard deviation is a bit harder than in the multivariate case.

## 2.2 Functional Exploratory Data Analysis

Mean, standard deviation, variance and other summary statistics of a functional data such as depth are calculated in analogy to the multivariate case. The main difference is that these statistics are calculated by using point-wise averages across replications and the result is now a function rather than a scalar.

### 2.2.1 Mean, Variance and Covariance Functions

Consider a square integrable random function  $\chi = \chi(t), t \in T = [t_1, t_n]$  in  $L^2(T)$  space. Square integrability requires that

$$\mathbb{E}\|\chi\|^2 = \mathbb{E} \int_T \chi^2(t) dt < \infty. \quad (2.1)$$

The main two parameters of the population that consists of square integrable functions are the mean function and the variance function which are respectively given by equations (2.2) and (2.3)

$$\mu_\chi(t) = \mathbb{E}[\chi(t)], \quad (2.2)$$

$$\text{Var}_\chi(t) = \sigma_\chi^2(t) = \mathbb{E}[(\chi(t) - \mu(t))^2]. \quad (2.3)$$

The covariance function of a functional data set gives an information about the dependency of the records across different argument values. The covariance function of two argument values  $t, s \in T$  is found from the equation (2.4)

$$\text{Cov}_\chi(t, s) = c_\chi(t, s) = \mathbb{E}[(\chi(t) - \mu_\chi(t))(\chi(s) - \mu_\chi(s))]. \quad (2.4)$$

Hence, the correlation function is calculated through the variance and covariance functions given by the equations (2.3) and (2.4):

$$\text{Cor}_\chi(t, s) = r_\chi(t, s) = \frac{\text{Cov}_\chi(t, s)}{\sqrt{\text{Var}_\chi(t)\text{Var}_\chi(s)}}. \quad (2.5)$$

Another important parameter for the functional case is the covariance operator which is defined by the equation (2.6)

$$\Gamma_\chi(\xi) = \mathbb{E}[\langle (\chi - \mu_\chi), \xi \rangle (\chi - \mu_\chi)]. \quad (2.6)$$

for all  $t \in T$  and  $\xi(t) \in L^2(T)$ .

Covariance operator takes the role of variance-covariance matrix in functional data analysis. Therefore, it is especially important in computing functional principal components which will be explained detailedly in Chapter 3.

Assume that a functional data set consists of independent identically distributed, square integrable observations  $\chi_1(t), \chi_2(t), \dots, \chi_N(t)$  in  $L_2(T)$  which have the same

distribution as  $\chi(t)$ . Then the mean, the variance and the covariance functions of the functional sample can be estimated from the sample counterparts of the equations (2.2), (2.3) and (2.4). The sample mean function is calculated as in the equation (2.7):

$$\hat{\mu}_\chi(t) = \bar{\chi}(t) = \frac{1}{N} \sum_{i=1}^N \chi_i(t). \quad (2.7)$$

Similarly, the sample variance function is obtained from the equation (2.8)

$$\hat{\sigma}_\chi^2(t) = \frac{1}{N-1} \sum_{i=1}^N [\chi_i(t) - \bar{\chi}(t)]^2. \quad (2.8)$$

The standard deviation function is the point-wise square root of the variance function and it is denoted by  $\hat{\sigma}_\chi(t)$ .

The empirical versions of the covariance function, the correlation function and the covariance operator are given by following equations:

$$\hat{c}_\chi(t, s) = \frac{1}{N-1} \sum_{i=1}^N [\chi_i(t) - \bar{\chi}(t)][\chi_i(s) - \bar{\chi}(s)], \quad (2.9)$$

$$\hat{r}_\chi(t, s) = \frac{\hat{c}_\chi(t, s)}{\sqrt{\hat{\sigma}_\chi^2(t)\hat{\sigma}_\chi^2(s)}}, \quad (2.10)$$

$$\hat{\Gamma}_\chi(\xi) = (N-1)^{-1} \sum_{i=1}^N [(\chi_i - \bar{\chi}), \xi(t)] (\chi_i - \bar{\chi}), \quad \xi(t) \in L^2(T). \quad (2.11)$$

The mean of a functional sample is an estimator of the center of the functional distribution, whereas the covariance of a functional sample is an estimator of the scale and the correlation structure of the functional distribution (Febrero et al., 2007).

When there are pairs of observed functions such as  $(\chi_i, \gamma_i)$ , to investigate if they depend on each other, cross-covariance or cross-correlation functions which are respectively given in the equations (2.12) and (2.13) are used.

$$\text{Cov}_{\chi, \gamma}(t, s) = \hat{c}_{\chi, \gamma}(t, s) = \frac{1}{N-1} \sum_{i=1}^n [\chi_i(t) - \bar{\chi}(t)][\gamma_i(s) - \bar{\gamma}(s)], \quad (2.12)$$

$$\text{Cor}_{\chi, \gamma}(t, s) = \hat{r}_{\chi, \gamma}(t, s) = \text{Cov}_{\chi, \gamma}(t, s) / \text{Var}_\chi(t) \text{Var}_\gamma(s) \quad (2.13)$$

To estimate center and scale properties of a data set, some different estimators can also be used rather than ordinary mean and standard deviation functions. Median and trimmed mean are some of robust location measures where median absolute deviation and trimmed standard deviation are robust scale measures. Functional  $\alpha$ -trimmed mean gives the mean of the most central curves where  $\alpha$  is  $0 \leq \alpha \leq (n-1)/n$ . The depth measure can also be used to estimate the most central curve.

### 2.2.2 Depth Measures

The depth measures the centrality of an observation within a given data cloud. The most popular depth measures for scalar covariates are the half-space depth, the simplicial depth and the Fraiman-Muniz depth measure.

In the univariate context, the depth of a point is calculated based on the cumulative distribution function  $F_N$  of the sample consists of scalar observations  $X_1, X_2, \dots, X_N$ .

$$F_N(x) = \frac{1}{N} \sum_{k=1}^N \mathbf{I}\{X_k \leq x\}. \quad (2.14)$$

The halfspace depth (HSD) of the observation  $X_i$  is defined by means of (2.14) as

$$\text{HSD}(X_i) = \min(F(X_i), 1 - F(X_i)) \quad (2.15)$$

In case of  $X_i$  is the median,  $F(X_i)$  is equal to  $1/2$ , and so  $\text{HSD}(X_i)$  is equal to  $1/2$  which gives the largest possible depth. If  $X_i$  is the largest point, then  $F(X_i) = 0$  and  $\text{HSD}(X_i) = 0$  and it gives the least possible depth (Horváth and Kokoszka, 2012).

The simplicial depth (SD) is calculated as

$$\text{SD}(X_i) = 2F(X_i)(1 - F(X_i)) \quad (2.16)$$

where the largest and the smallest possible depths take the values of  $1/2$  and  $0$  respectively.

The depth measure offered by Fraiman and Muniz (2001) takes values between  $1/2$  and  $1$  and defined as

$$\text{FMD}(X_i) = 1 - \left| \frac{1}{2} - F(X_i) \right| \quad (2.17)$$

For a sample of functions  $\{\chi_i(t), t \in T, i = 1, 2, \dots, N\}$ , the empirical cumulative distribution function at point  $t \in [t_{\min}, t_{\max}]$  is described as

$$F_{N,t}(\chi_i(t)) = \frac{1}{N} \sum_{k=1}^N \mathbf{I}\{\chi_k(t) \leq \chi_i(t)\} \quad (2.18)$$

where  $\mathbf{I}(\cdot)$  is an indicator function (Febrero et al., 2007).

If the univariate depth of the data  $X_i$  at point  $t$  is defined as  $D$  on  $\mathbb{R}$ , then the functional depth (FD) of the curves  $\chi_1, \chi_2, \dots, \chi_N$  can be found from the integration of the univariate depths.

$$\text{FD}(\chi_i) = \int_{t_{\min}}^{t_{\max}} D(\chi_i(t)) dt, \quad i = 1, \dots, N. \quad (2.19)$$

For instance, the functional Fraiman-Muniz depth (FMD) of curve  $\chi_i(t)$  is defined as

$$\text{FMD}(\chi_i) = \int_{t_{\min}}^{t_{\max}} \left[ 1 - \left| \frac{1}{2} - F_{N,t}(\chi_i(t)) \right| \right] dt. \quad (2.20)$$

Likewise in the univariate context, the curves which attain the maximum and the minimum values are the deepest and the least deepest curves, respectively (Febrero et al., 2007).

There are also other types of functional depths such as modal depth, random projection depth (RPD) and band-depth. Detailed information about these techniques can be found in Febrero et al. (2008) and López-Pintado and Romo (2009).

The depth measure is an important indicator of outlying. Because, the curves that have significantly low depth values are more close to be outliers. In functional context, a functional outlier is a curve which comes from a different distribution in a function space than the rest of the curves, which are assumed to be identically distributed.

The algorithm of identifying outliers by using depth measures is as follows (Horváth and Kokoszka, 2012):

1. Calculate  $F(\chi_1), F(\chi_2), \dots, F(\chi_N)$ .

2. Determine a threshold value of  $C$  and remove the curves with depth smaller than  $C$ . Classify these curves as outliers. If there are no such curves, the procedure ends here.
3. Return to step 1 and apply it to the rest of the sample obtained by removing curves in step 2.

Febrero-Bande and Oviedo de la Fuente (2012) proposed two different outlier detection procedures based on bootstrap samples. One of them uses trimming and the other one uses weighting principle. These methods are implemented respectively as functions "outliers.depth.trim()" and "outliers.depth.pond()" in the R package "fda.usc" (Febrero-Bande and Oviedo de la Fuente, 2012; Oviedo de la Fuente, 2011). In the first method, the bootstrap samples are chosen after removing  $\% \alpha$  of data curves. In the second procedure, each observation (each curve) is weighted depending on its depth value. The last steps of both procedures are similar. The algorithm of for "outliers.depth.pond()" function can be summarized as follows:

1. A measure of functional depth such as SD, FMD or RPD is chosen and the functional depths  $D(\chi_1), \dots, D(\chi_n)$  are obtained for the data set  $\chi_1(t), \dots, \chi_n(t)$ .
2. B number of bootstrap samples  $\chi_i^b$  are taken from the functional data set in a way that each curve is sampled with a probability proportional to its depth.
3. Considering that  $Z_i^b(t_1), \dots, Z_i^b(t_N)$  is normally distributed with mean zero and the covariance matrix  $\gamma \Sigma_\chi$  where  $\Sigma_\chi$  is the covariance matrix of  $\chi(t_1), \dots, \chi(t_N)$  and  $\gamma$  is a smoothing parameter, the bootstrap samples are smoothed such that  $Y_i^b = \chi_i^b + Z_i^b$ .
4. For each bootstrap sample  $b = 1, \dots, B$ , a cut off value  $C^b$  is chosen such that  $P(D_n(\chi_i) \leq C^b) = c$ ,  $i = 1, \dots, N$  where by default it is equal to the 1th percentile ( $c = 0.01$ ) of the distribution of the depths  $D(Y_i^b)$ .
5. The median of the values  $C^b$  are taken and denoted by  $C$ .

Then the second and third steps of the previous algorithm by Horváth and Kokoszka (2012) are applied. For detailed information see Oviedo de la Fuente (2011).

### 2.3 Representing Functional Data

The first step in a functional data analysis is to convert raw data to functional objects (Levitin et al., 2007). This can be done in two different ways. If it is assumed that

the discrete values are errorless, interpolation is used. But if it is assumed that an observational error exists, then the conversion is made by smoothing (Ramsay and Silverman, 2005). Therefore, smoothing or interpolation is the most important part of functional data analysis. It serves to convert relatively enough discrete points of a functional datum to a smooth function.

The smoothness implies that two adjacent points that compose a function  $\chi$  are linked together to some extent and unlikely to be too different from each other (Ramsay and Silverman, 2005). If the data is defined on a  $L^2$  or a Hilbert space smoothing can be done based on basis functions representation. Otherwise kernel methods can be used in smoothing (Febrero-Bande and Oviedo de la Fuente, 2012).

### 2.3.1 Basis Representation

A basis is a system of functions which are independent of each other. Basis function procedure represents a function  $\chi$  as a linear combination of  $K$  known basis functions  $\phi_k$ :

$$\chi(t) = \sum_{k=1}^K c_k \phi_k(t). \quad (2.21)$$

This expression can be written in matrix terms:

$$\chi(t) = \mathbf{c}'\boldsymbol{\phi} = \boldsymbol{\phi}'\mathbf{c}, \quad (2.22)$$

where  $\mathbf{c}$  denotes the vector of length  $K$  of the coefficients  $c_k$  and  $\boldsymbol{\phi}$  is a  $n \times K$  matrix contains the basis functions  $\phi_k(t_j)$  (Ramsay and Silverman, 2005). The coefficients  $c_k$  define the weight related to each basis function in constructing a curve.

The dimension of expansion is equal to number of basis functions  $K$ . The basis expansion approach enables to work with infinite dimensional functions by representing them in finite dimension. The choice of number  $K$  of functions and the type of basis are the most important steps in smoothing. The smaller  $K$  provides convenience in computation.

The type of the basis expansion is determined according to the type of the data. Mostly used basis functions are Fourier basis and B-spline basis functions. Fourier

basis is mostly preferred for periodic data whereas B-spline basis are mostly used for non-periodic data.

Except these two methods, there are also other types of basis functions such as monomials, wavelet bases, polynomial bases, exponential and power bases, the polygonal bases and constant basis. Wavelet bases are mostly chosen where the derivatives are not required. Constant basis is the simplest of all. It is mostly used to define a scalar observation as a functional datum. Therefore, it has particularly importance in determining the intercept function in functional linear models. Polynomial basis can be used for simpler functional problems.

### 2.3.1.1 Fourier basis

The best known basis expansion is the Fourier series expansion which can be given in the form of

$$\hat{\chi}(t) = c_0 + c_1 \sin \omega t + c_2 \cos \omega t + c_3 \sin 2\omega t + c_4 \cos 2\omega t + \dots \quad (2.23)$$

where the coefficients are  $\phi_0(t) = 1$ ,  $\phi_{2r-1}(t) = \sin r\omega t$ ,  $\phi_{2r}(t) = \cos r\omega t$ . It is periodic with the period  $2\pi/\omega$ .

The basis expansion of a derivative estimate can be written in the form:

$$D\hat{\chi}(t) = \sum_{i=1}^K \hat{c}_i D\phi_i(t) = \hat{\mathbf{c}}' D\boldsymbol{\phi}(t). \quad (2.24)$$

A basis system that gives good representation of function estimates is not always appropriate for derivative estimates. Therefore, it is suggested to control if the approximation of one or more derivatives behave reasonably or not (Ramsay and Silverman, 2005).

The Fourier expansion coefficients for derivatives of a function can be easily computed as well from the following equations:

$$\begin{aligned} D\sin r\omega t &= r\omega \cos r\omega t \\ D\cos r\omega t &= -r\omega \sin r\omega t \end{aligned} \quad (2.25)$$



The Fourier expansions of higher derivatives can be also easily found by multiplying coefficients by suitable powers of  $r\omega$ . For example, the coefficient vector of a second order derivative  $D^2\chi$  of a function can be found as

$$(0, -\omega^2 c_1, -\omega^2 c_2, -4\omega^2 c_3, -4\omega^2 c_4, \dots).$$

Fourier basis has good computational facilities for equally spaced and periodic data. They are especially appropriate for stable functions where the curvature tends to be of the same order everywhere. Fourier series expansions are very popular in mathematics, statistics and engineering applications. But they are not offered for smoothing of non periodic and not equally spaced data. In this case, it would be better to use other types of basis systems such as B-splines.

As mentioned in the book of Ramsay and Silverman (2005):

"A Fourier series is like margarine: It's cheap and you can spread it on practically anything, but don't expect that the result will be exciting eating."

### **2.3.1.2 B-spline basis**

Spline functions are the most common choice of approximation system for non-periodic functional data or parameters (Ramsay and Silverman, 2005). The advantage of splines are that they are fast to compute and they achieve a good approximation of the data with a relatively small number of  $K$ 's (Horváth and Kokoszka, 2012).

The first step to construct a a spline is to divide the interval on which the function is defined, into  $L$  subintervals. The points that separates one interval from an other are called breakpoints and they are denoted by  $\tau_l$  where  $l = 1, \dots, L - 1$ . Splines gain more flexibility as the number of breakpoints increases. The sequence of values at breakpoints of a spline basis are called knots. The number of knots related to a break point can be more than one. The knots and breakpoints define the same thing if and only if the knots are distinct.

A spline function is a polynomial of a fixed degree or an order that is defined on each interval. The degree of a polynomial refers to the highest power of the polynomial. The order of a polynomial is the number of constants required to define it which is

one more than the value of its degree (Ramsay et al., 2009). It is represented by  $m$ . If  $m = 1$ , it is a step function of degree zero. For  $m = 2$ , it is piecewise linear. Order 3 and 4 splines define a quadratic and a cubic polynomial, respectively. In applications mostly order 4 splines are used because they seem quite smooth and their first and second degree of derivatives exist. If there is the need of smooth derivatives, Ramsay et al. (2009) suggests that the order of the spline basis should be at least two higher than the highest order derivative to be used.

For the case that the number of knots are equal to the number of breakpoints, the number  $K$  of basis functions of a spline system is equal to the sum of the order  $m$  of the polynomial and the number of interior knots (Ramsay and Silverman, 2005).

$$K = m + L - 1. \quad (2.26)$$

Some basic properties of splines can be summarized as follows (Ramsay and Silverman, 2005):

1. Each basis function  $\phi_k(t)$  is an order  $m$  spline function with a knot sequence  $\tau$ .
2. The sums and differences of splines are splines. The multiple and the linear combination of spline functions still define a spline function.
3. Any spline function defined by order  $m$  and the knot sequence  $\tau$  can be expressed as a linear combination of these basis functions.

### 2.3.1.3 Bias - Variance Trade Off

The determination of number of basis is an important part of smoothing. While determining the number  $K$  of basis functions, the bias and the variance of the estimation should be considered to find out the best approximation. Using high number of basis can give a better representation of the data with low bias

$$\text{Bias}[\hat{\chi}(t)] = \chi(t) - \mathbb{E}[\hat{\chi}(t)] \quad (2.27)$$

but it increases the variance of the estimate  $\hat{\chi}(t)$ .

$$\text{Var}[\hat{\chi}(t)] = \mathbb{E}[\hat{\chi}(t) - \mathbb{E}[\hat{\chi}(t)]^2] \quad (2.28)$$

Similarly, using small number of basis functions produces high bias but decreases the sampling variance (Levitin et al., 2007). This contradiction between the number of basis and smoothness is called "bias-variance trade-off".

Therefore, another criterion called Mean Squared Error (MSE) is used to decide to the number of basis. MSE criterion considers both the bias and the variance of the estimate. MSE is calculated as in the equation (2.29)

$$\text{MSE}[\hat{\chi}(t)] = \mathbb{E}[\hat{\chi}(t) - \chi(t)]^2 \quad (2.29)$$

which can be equivalently given in terms of bias and variance as

$$\text{MSE}[\hat{\chi}(t)] = \text{Bias}^2[\hat{\chi}(t)] + \text{Var}[\hat{\chi}(t)] \quad (2.30)$$

This criteria gives permission to tolerate a little bias if there is a big reduction in the sampling variance (Ramsay and Silverman, 2005).

Another method offered to determine the number of basis is stepwise variable selection method which is based on adding the basis functions one after another and testing at each step if the fit is improved or not (Ramsay and Silverman, 2005).

### 2.3.2 Kernel Smoothing

Functional data can also be represented by using nonparametric methods that are based on localized weighting principle. Kernel smoothing method is one of the most popular nonparametric smoothing methods. In kernel smoothing the weight values depend on a Kernel function given as

$$w_i(t) = K\left(\frac{t_i - t}{h}\right) \quad (2.31)$$

where  $h$  is called the bandwidth parameter and determines the weight value  $w_i(t)$  according to the distance to related point  $t$ . Small values of  $h$  states that only observations close to  $t$  receive any weight while the reverse means that the observations

considerable away from  $t$  take weights, too. Here, kernel function  $K(\cdot)$  can be chosen as one of well known kernel functions such as Gaussian, Uniform, Quadratic, Epanechnikov, Triweight or Cosine Kernel.

The bandwidth parameter  $h$  is a measure of concentration. Small values of  $h$  imply that only observations close to  $t$  receive weight where as the higher values of  $h$  means that even the observations distant from  $t$  have relevant weight (Ramsay and Silverman, 2005).

In kernel smoothing the estimate at a given point can be written as a linear combination of observed responses

$$\hat{\chi}(t_j) = \sum_{j=1}^n S_j(t) y_j \quad (2.32)$$

where  $S_j(t)$  is a measure of weight that is based on local weights and it is called as a linear smoother. Kernel estimator is the simplest and most classic case of an estimator that uses local weights (Ramsay and Silverman, 2005).

The most popular Kernel estimator is the Nadaraya-Watson estimator. The Nadaraya-Watson kernel estimator is defined by.

$$S_j(t) = \frac{K \left[ \frac{t_j - t}{h} \right]}{\sum_{l=1} K \left[ \frac{(t_l - t)}{h} \right]} \quad (2.33)$$

Other possible estimators that can be used in kernel smoothing are  $k$  nearest neighbors (KNN) estimator, local linear regression estimator and local polynomial estimator (Wasserman, 2006; Febrero-Bande and Oviedo de la Fuente, 2012).

In kernel smoothing, the derivatives can not be easily calculated by taking the derivative of the kernel estimator. Because some kernels such as uniform and quadratic kernels are not differentiable. For estimating derivatives, it is suggested that the bandwidth value  $h$  to be larger than the one used in estimating the function (Ramsay and Silverman, 2005).

### 2.3.3 Functional Principal Components (FPC) Basis

In functional data analysis, principal components analysis is used as a dimension reduction technique as in the multivariate case. The most important difference is that in functional context the components are functions rather than vectors (Levitin et al., 2007).

Functional Principal Components Analysis (FPCA) helps to represent a functional data in terms of a combination of orthonormal variables that are obtained by maximizing the variance of the component scores which is equal to  $N^{-1} \sum_{j=1}^K f_{ij}^2$ . This is essential to reveal the most important modes of variation in the variables (Ramsay and Silverman, 2005).

The main algorithm of functional principal components can be summarized as follows:

1. The determination of principal component weight function  $\xi_1(t)$  of norm 1 for which the principal component scores  $f_{i1} = \int_T \xi_1(t) \chi_i(t) dt$  maximize  $\sum f_{i1}^2$  subject to  $\|\xi_1\|^2 = 1$  in order to minimize sum of square errors.
2. The computation of following weight function  $\xi_2(t)$  and the principal component scores  $f_{i2} = \int \xi_2(t) \chi_i(t) dt$  that maximize  $\sum f_{i2}^2$  subject to the constraint  $\|\xi_2\|^2 = 1$  considering the constraint of orthogonality of two components as given  $\int \xi_1(t) \xi_2(t) dt = 0$ .
3. The repetition of the steps as much as the number of principal components.

The functional variables can be written in terms of principal component scores  $f_{ij}$  and the eigenfunctions  $\xi_j$  in a finite orthonormal basis

$$\hat{\chi}_i(t) = \sum_{j=1}^K f_{ij} \xi_j(t), \quad i = 1, \dots, N. \quad (2.34)$$

Detailed information about this methodology will be given in Chapter 3.

### 2.3.4 Functional Partial Least Squares (FPLS) Basis

Partial Least squares (PLS) is a dimension reduction method alternative to principal components analysis (PCA). PLS approach and regression are generally used together.

Therefore, it is denominated as Partial Least Squares Regression (PLSR). PLSR is extended to the case that the explanatory variables are in the form of a stochastic process by Preda and Saporta (2005). In functional context, it is used with functional linear regression analysis with a scalar response and called Functional Partial Least Squares (FPLS) Regression. This approach will has been explained detailedly in Chapter 4. FPLS is based on the determination of FPLS components by considering the correlation between the scalar response variable and the functional predictors. Different applications of FPLS regression especially on chemometrics, can be found in the literature (Saeys et al., 2008; Aguilera et al., 2010; Preda and Schiltz, 2011).

The main idea of FPLS is to represent the functional variable  $\chi_i(t)$  in terms of FPLS components  $\varphi_l$  and FPLS scores  $v_{il}$ ,

$$\hat{\chi}_i(t) = \sum_{l=1}^K v_{il} \varphi_l(t), \quad i = 1, \dots, N. \quad (2.35)$$

The main algorithm to obtain FPLS components has been given in the next sections. The detailed information about this technique can be found in Preda and Saporta (2005) and Aguilera et al. (2010). The algorithm of FPLS regression also takes part in `fda.usc` package of R software Febrero-Bande and Oviedo de la Fuente (2012).

## 2.4 Smoothing Functional Data

The vital part of smoothing functional data is to determine the model that fits the raw data. Different fitting algorithms can be chosen according to the smoothing method that is used to represent the data. Three main criteria used in fitting functional data are Least Squares Criteria - which can be separated into two as Ordinary (Unweighted) Least Squares (OLS) and Weighted Least Squares(WLS), Localized Least Squares (LLS) and Roughness Penalty Approach.

Least Squares Criteria is generally used for basis representation to determine basis coefficients. In the case of autocorrelation, weighted least squares criterion is used rather than OLS criterion. LLS is mainly used for kernel smoothing where the weight functions are determined by a kernel function. Roughness penalty approach is another method which takes into account an extra smoothing parameter and uses penalized least squares criterion in fitting the data to a smooth function.

## 2.4.1 Fitting by Least Squares Criterion

Fitting the raw data by using a basis system can be seen as a regression problem in the form of

$$y_j = \chi(t_j) + \varepsilon_j \quad (2.36)$$

where  $y_j$  states discretely observed raw data vector  $(y_1, y_2, \dots, y_n)$ ,  $\chi(t_j)$  is the basis function expansion of given in the form of (2.21) where  $\mathbf{t}$  is the vector consisting of the observation points  $t_j$ ,  $j = 1, \dots, n$  and  $\varepsilon_j$  the  $n$  length vector contains the error terms.

### 2.4.1.1 Ordinary (Unweighted) Least Squares Fit

The coefficients  $c_k$  in the equation (2.21) should be determined in a way to obtain an optimal fit to data. Therefore, they are chosen in order to minimize the error sum of squares by the least squares estimate

$$\text{SSE} = \sum_{j=1}^n \left[ y_j - \sum_{k=1}^K c_k \phi_k(t_j) \right]^2 = \sum_{j=1}^n [y_j - \boldsymbol{\phi}(t_j)' \mathbf{c}]^2 \quad (2.37)$$

which can also be defined in matrix terms as

$$(\mathbf{y} - \boldsymbol{\phi} \mathbf{c})' (\mathbf{y} - \boldsymbol{\phi} \mathbf{c}) \quad (2.38)$$

(Ramsay et al., 2009; Hooker, 2010).

In functional notation, this can be shown as a norm  $\| \mathbf{y} - \boldsymbol{\phi} \mathbf{c} \|^2$ .

Taking the derivative of the expression in (2.38) respect to  $\mathbf{c}$

$$2 \boldsymbol{\phi} \boldsymbol{\phi}' \mathbf{c} - 2 \boldsymbol{\phi}' \mathbf{y} = 0, \quad (2.39)$$

and solving it for  $\mathbf{c}$  leads us to the ordinary least squares estimate  $\hat{\mathbf{c}}$ .

$$\hat{\mathbf{c}} = (\boldsymbol{\phi}' \boldsymbol{\phi})^{-1} \boldsymbol{\phi}' \mathbf{y} \quad (2.40)$$

By using equation (2.40), the estimated curve values  $\hat{x}(t)$  are obtained from

$$\hat{\chi}(t) = \phi(\phi'\phi)^{-1}\phi'y \quad (2.41)$$

The smoothed values  $\hat{\chi}(t)$  can be written as a linear transformation of the observed discrete values  $y_j$  by using the smoothing matrix. In this case, the smoothing operation is called a linear smoother.

$$\hat{\chi}(t) = \sum_{j=1}^n S_j(t) y_j \quad (2.42)$$

Here,  $S_j(t)$  is a measure of weight that is based on local weights. It weights the  $j$ th discrete data value.

In matrix terms, it can be defined as (2.43)

$$\hat{\chi}(\mathbf{t}) = \mathbf{S}\mathbf{y} \quad (2.43)$$

Here,  $\hat{\chi}(\mathbf{t})$  is a column vector that consists of the values of the estimate of the function  $\chi(\mathbf{t})$  at each sampling point  $t_j$  and  $\mathbf{S}$  is the smoothing matrix that converts the dependent variable vector into the fitted values likewise in linear regression. In the least squares context,  $\mathbf{S}$  is a projection matrix with the property of idempotency (Ramsay and Silverman, 2005).

The smoothing matrix of the ordinary or unweighted least squares is in the form of (2.44)

$$\mathbf{S} = \phi(\phi'\phi)^{-1}\phi' \quad (2.44)$$

#### 2.4.1.2 Weighted Least Squares Fit

This simple least squares approximation is appropriate in the case that the residuals  $\epsilon_j$  are independently and identically distributed with mean zero and constant variance  $\sigma^2$ . When the residuals are autocorrelated or the variance vary with the observation time, the use of weighted least squares method is offered to smooth the data (Ramsay and Silverman, 2005).



The least squares criterion of WLS is an extension of OLS with an additional weight matrix  $\mathbf{W}$  which is a symmetric and positive definite matrix.

$$(\mathbf{y} - \boldsymbol{\phi}\mathbf{c})' \mathbf{W} (\mathbf{y} - \boldsymbol{\phi}\mathbf{c}) \quad (2.45)$$

As in the multivariate context the weight matrix  $\mathbf{W}$  can simply be obtained from the variance-covariance matrix of residuals as

$$\mathbf{W} = \boldsymbol{\Sigma}_e^{-1} \quad (2.46)$$

and in this case the weighted least squares estimate  $\hat{\mathbf{c}}$  is expressed as

$$\hat{\mathbf{c}} = (\boldsymbol{\phi}' \mathbf{W} \boldsymbol{\phi})^{-1} \boldsymbol{\phi}' \mathbf{W} \mathbf{y} \quad (2.47)$$

Hence, the smoothing matrix for weighted least squares estimation is written in the form

$$\mathbf{S} = \boldsymbol{\phi} (\boldsymbol{\phi}' \mathbf{W} \boldsymbol{\phi})^{-1} \boldsymbol{\phi}' \mathbf{W}. \quad (2.48)$$

## 2.4.2 Fitting by Localized Least Squares Criterion

The values of the function estimate at a point  $t_j$  is affected by the observations nearby. Considering this, the function estimates at point  $t_j$  can be written as

$$\chi(t_j) = \sum_j^n w_j y_j \quad (2.49)$$

by using local weight functions  $w_j$ . The localized weights here are obtained from Kernel functions in the way that is shown in the equation (2.31).

### 2.4.2.1 Localized basis functions smoothing

The least squares criterion can be extended in a way to give a local measure of error. This method combines the ideas of basis function estimators and the kernel estimators.

The localized basis function estimators of this method are obtained by minimizing the sum of square errors which is given as

$$\text{SSE}_{\text{LLS}} = \sum_{j=1}^n w_j(t) \left[ y_j - \sum_{k=1}^K c_k \phi_k(t_j) \right]^2 \quad (2.50)$$

where  $w_j(t)$  denotes the weights obtained from a kernel function in the form of (2.31).

In matrix terms it can be written as

$$\text{SSE}_{\text{LLS}} = (\mathbf{y} - \boldsymbol{\phi}\mathbf{c})' \mathbf{W} (\mathbf{y} - \boldsymbol{\phi}\mathbf{c}). \quad (2.51)$$

The localized least squares estimates are found from

$$\hat{\mathbf{c}} = (\boldsymbol{\phi}' \mathbf{W} \boldsymbol{\phi})^{-1} \boldsymbol{\phi}' \mathbf{W} \mathbf{y}. \quad (2.52)$$

#### 2.4.2.2 Local polynomial smoothing

Another technique that can be used in smoothing is local polynomial smoothing. In this method, the estimated curve values  $\hat{\chi}_i(t)$  are obtained from the minimization of the formula

$$\text{SSE}_{\text{LPS}} = \sum_{j=1}^n \text{Kern}_h(t_j, t) \left[ y_j - \sum_{l=0}^L c_l (t - t_j)^l \right]^2 \quad (2.53)$$

In general, the value of  $L$  is taken at least one. In the situation that the derivatives are used, it is suggested to take  $l$  two higher than the highest order derivative required (Ramsay and Silverman, 2005).

#### 2.4.3 Fitting by Roughness Penalty Approach

Another popular smoothing method is the roughness penalty approach. This approach enables to use large number of basis functions while imposing smoothness by penalizing some measure of function complexity (Ramsay et al., 2009). A popular

measure of a roughness of a function  $\chi$  is the integrated squared second derivative which can also be called curvature.

$$\text{PEN}_2(\chi) = \int [D^2\chi(t)]^2 dt. \quad (2.54)$$

The choice of the roughness penalty is dependent on the type of the data. If the aim is to smooth second derivatives then the penalization of the curvature of the second derivative can give good results.

$$\text{PEN}_4(\chi) = \int [D^4\chi(t)]^2 dt \quad (2.55)$$

Accordingly, when dealing with periodic data the penalization of the harmonic acceleration operator would be used as a measure of roughness

$$\text{PEN}_L(\chi) = \int [(L\chi)^2](t)dt = \|L\chi\|^2 \quad (2.56)$$

where

$$L\chi = D^3\chi + \omega^2 d\chi \quad (2.57)$$

The general form of roughness penalty can be defined in matrix form

$$\begin{aligned} \text{PEN}_m(\chi) &= \int [D^m\chi(t)]^2 dt \\ &= \int [D^m\mathbf{c}\boldsymbol{\phi}(t)]^2 dt \\ &= \int \mathbf{c}' D^m\boldsymbol{\phi}(t) D^m\boldsymbol{\phi}(t)' \mathbf{c} dt \\ &= \mathbf{c}' \left[ \int D^m\boldsymbol{\phi}(t) D^m\boldsymbol{\phi}'(t) dt \right] \mathbf{c} \\ &= \mathbf{c}' \mathbf{R} \mathbf{c} \end{aligned} \quad (2.58)$$

where

$$\mathbf{R} = \int D^m\boldsymbol{\phi}(t) D^m\boldsymbol{\phi}(t)' dt. \quad (2.59)$$

The penalized residual sum of squares approach is obtained from the addition of the weighted sum of squares error criteria and the multiple of  $\text{PEN}_m(\lambda)$  by a smoothing parameter  $\lambda$

$$\text{SSE}_{\text{PEN}} = (\mathbf{y} - \boldsymbol{\phi}\mathbf{c})'\mathbf{W}(\mathbf{y} - \boldsymbol{\phi}\mathbf{c}) + \lambda\mathbf{c}'\mathbf{R}\mathbf{c}. \quad (2.60)$$

When the derivative of (2.60) is taken, it is seen that

$$-2\boldsymbol{\phi}'\mathbf{W}\mathbf{y} + \boldsymbol{\phi}'\mathbf{W}\boldsymbol{\phi}\mathbf{c} + \lambda\mathbf{R}\mathbf{c} = 0. \quad (2.61)$$

From this equation, the estimated coefficient vector is obtained.

$$\hat{\mathbf{c}} = (\boldsymbol{\phi}'\mathbf{W}\boldsymbol{\phi} + \lambda\mathbf{R})^{-1}\boldsymbol{\phi}'\mathbf{W}\mathbf{y}. \quad (2.62)$$

Hence, the smoothing matrix takes the form of

$$\mathbf{S}_{\boldsymbol{\phi},\lambda} = \boldsymbol{\phi}(\boldsymbol{\phi}'\mathbf{W}\boldsymbol{\phi} + \lambda\mathbf{R})^{-1}\boldsymbol{\phi}'\mathbf{W}. \quad (2.63)$$

When  $\lambda = 0$ , it is equal to the smoothing matrix of LSS which was given in the expression (2.48).

The degrees of freedom (df) of a least squares smoother is equal to the trace of the smoothing matrix which is equal to the number of parameters  $K$  that the coefficient vector consists (Ramsay and Silverman, 2005).

$$df = \text{trace}(\mathbf{S}) = K. \quad (2.64)$$

#### 2.4.3.1 The Choice of Smoothing Parameter

The most popular methods that are used to decide the value of the smoothing parameter are Cross-Validation (CV) and Generalized Cross Validation criteria.

Cross Validation (CV) criteria can be written as

$$\text{CV}(\mathbf{v}) = \frac{1}{n} \sum_j^n (y_j - \hat{y}_{j(-j)})^2 \omega_j \quad (2.65)$$

where  $\hat{y}_{j(-j)}$  denotes the estimator when leaving out the  $j$ th pair  $(t_j, y_j)$  and  $\omega_j$  is the weight at point  $t_j$ . Although this method is widely used, it has some problems. This method is computationally intensive and it can lead to under smoothing the data (Ramsay and Silverman, 2005). Therefore, generalized cross validation (GCV) method that has less tendency to undersmooth data is offered. It can be calculated as in (2.66)

$$\text{GCV}(\mathbf{v}) = \frac{1}{n} \sum_j^n (y_j - \hat{y}_{j(-j)})^2 \omega_j \Xi(\mathbf{v}) \quad (2.66)$$

where  $\Xi(\mathbf{v})$  denotes the type of penalizing function. Here  $\mathbf{v}$  indicates a parameter that depends both on the number  $K$  of basis and on the smoothing parameter  $\lambda$ .

For GCV criteria  $\Xi(\mathbf{v})$  is equal to

$$\Xi(\mathbf{v}) = (1 - \text{Tr}(\mathbf{S}) n^{-1})^{-2} \quad (2.67)$$

According to the change in this function, different criteria can be used such as Akaike Information Criterion (AIC), Finite Prediction Error (FPE), Rice Bandwidth Selector and Shibata's Model Selector criterion. The  $\Xi(\mathbf{v})$  functions according to mentioned criteria have been explained detailedly in Febrero-Bande and Oviedo de la Fuente (2012).

### 3 FUNCTIONAL PRINCIPAL COMPONENTS ANALYSIS

Principal components analysis method is an important technique in multivariate analysis which is used to reduce dimension and helps to reveal uncorrelated orthogonal components as a linear combination of existing random variables considering the correlation or covariances between them.

Functional principal components analysis (FPCA) is an extension of multivariate principal components analysis (PCA). The main difference between two methods is that FPCA deals with curves instead of vectors. Therefore, the covariance matrix and eigenvectors are replaced with the linear operator and eigenfunctions, respectively. The principal components are elements of  $L^2$  space rather than vector space  $\mathbb{R}^p$ . The methodology of FPCA is based on Karhunen-Loève decomposition which is used to define random variables that come from a stochastic process. Detailed information about this expansion can be found in the Appendix. The main differences between the methods of PCA and FPCA can be summarized as in Table (3.1).

**Table 3.1:** The comparison of PCA and FPCA methods

	PCA	FPCA
<b>Data</b>	Vectors $\in \mathbb{R}^p$	Functions $\in L^2(T)$ , $T = [t_1, t_n]$
<b>Variables</b>	$X = [X_1, X_2, \dots, X_p]$ $X_j = [x_{1j}, \dots, x_{nj}]$ , $j = 1, \dots, p$	$\chi(t) = [\chi_1(t), \dots, \chi_N(t)]$ $t \in T = [t_1, t_n]$
<b>Covariance</b>	Matrix $\mathbf{V}$ $\mathbf{V} = \text{Cov}(X) \in \mathbb{R}$	Operator $\Gamma_\chi$ bounded between $t_1$ and $t_n$ $\Gamma_\chi : L^2(T) \rightarrow L^2(T)$
<b>Eigenstructure</b>	Vector $u_j \in \mathbb{R}^p$ $\mathbf{V}u_j = \lambda_j u_j$	Function $\xi_j \in L^2(T)$ , $t \in T$ $\int_T c(s, t) \xi_j(s) ds = \lambda_j \xi_j(t)$
<b>Components</b>	Random variables in $\mathbb{R}^p$	Random functions in $L^2(T)$

Despite these differences, the main idea of both analysis is the same which is to obtain significant factors that explain the variability of the data most.

### 3.1 The Approximation of Principal Components Analysis

In multivariate case, the principal components or the scores are linear combinations of the centered observations  $x_{ij}$ 's of the  $j$ th variable with the loadings  $\mathbf{u}_j = (u_{1j}, u_{2j}, \dots, u_{pj})'$  which are chosen in a way to reveal the strongest mode of variation in the data. This can be written as

$$f_{ik} = \sum_{j=1}^p \mathbf{u}_{kj} x_{ij} = \mathbf{u}'_k \mathbf{x}_i, \quad i = 1, \dots, N, \quad j = 1, \dots, p, \quad k = 1, \dots, p \quad (3.1)$$

where  $\mathbf{x}_i$  denotes the vector  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})'$ ,  $\mathbf{u}_j$  denotes the weight vector of the  $j$ th component and  $k$  is the number of components considering that there are as many components as the number of variables  $p$ .

The weight vector of the first component  $\mathbf{u}_1 = (u_{11}, \dots, u_{p1})'$  is obtained from the maximization of the variance of the first component  $f_{i1}$

$$f_{i1} = \sum_{j=1}^p \mathbf{u}_{j1} x_{ij} = \mathbf{u}'_1 \mathbf{x}_i \quad (3.2)$$

subject to the constraint

$$\|\mathbf{u}_1\|^2 = \mathbf{u}'_1 \mathbf{u}_1 = 1. \quad (3.3)$$

The weight vector of the second component  $\mathbf{u}_2$  is obtained from the maximization of the variance of the second component  $f_{i2}$  subject to the constraint  $\|\mathbf{u}_2\|^2 = \mathbf{u}'_2 \mathbf{u}_2 = 1$  and additionally subject to the constraint

$$\mathbf{u}'_2 \mathbf{u}_1 = 0 \quad (3.4)$$

which indicates the orthogonality of the components.

The weight vector related to the  $m$ th component is obtained from the repetition of the mentioned steps so as to let  $f_{im}$  have maximum variance subject to the constraint

$$\|\mathbf{u}_m\|^2 = \mathbf{u}'_m \mathbf{u}_m = 1 \quad (3.5)$$

and subject to  $m - 1$  additional constraints

$$\mathbf{u}'_l \mathbf{u}_m = 0, l < m \quad (3.6)$$

In this manner, the most important modes of variation in the data are determined and the amount of variation of the components decrease in each step.

Considering that  $\mathbf{X}$  is a  $N \times p$  matrix consists of the centered observations  $x_{ij}$ ,  $\mathbf{u}$  is a  $p$  length weight vector and  $f_{ik}$  are the principal component scores which are computed from the maximization criteria which is given in matrix form as

$$\max \quad N^{-1} \mathbf{u}' \mathbf{X}' \mathbf{X} \mathbf{u} \quad \text{subject to} \quad \mathbf{u}' \mathbf{u} = 1. \quad (3.7)$$

Since the variance-covariance matrix  $\mathbf{V}$  is defined as

$$\mathbf{V} = N^{-1} \mathbf{X}' \mathbf{X}, \quad (3.8)$$

the criterion (3.7) takes the form

$$\max \quad \mathbf{u}' \mathbf{V} \mathbf{u} \quad \text{subject to} \quad \mathbf{u}' \mathbf{u} = 1. \quad (3.9)$$

The solution of this maximization problem is obtained by solving the eigenequation problem given as

$$\mathbf{V} \mathbf{u}_j = \lambda_j \mathbf{u}_j \quad (3.10)$$

where  $\lambda_j$  indicates the eigenvalues and  $\mathbf{u}_j$  indicates the eigenvectors of the variance covariance matrix. For each  $j$ , eigenvectors satisfy the conditions given in the equations (3.5) and (3.6) respectively.

Due to the effect of centering, the rank of the centered data matrix is equal to  $N - 1$  and the  $p \times p$  symmetric matrix  $\mathbf{V}$  has at most  $\min(p, N - 1)$  non-zero eigenvalues  $\lambda_j$ . Different eigenvalue-eigenvector pairs  $(\lambda_j, \mathbf{u}_j)$  can be found satisfying the condition (3.10).



Considering that  $\mathbf{V}$  is a  $p \times p$  symmetric matrix, the solution of any maximization problem in the form of

$$\max \mathbf{x}'\mathbf{V}\mathbf{x} \quad \text{subject to} \quad \mathbf{x}'\mathbf{x} = 1 \quad (3.11)$$

can be calculated by using quadratic forms.

Let  $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$  be the positive eigenvalues of  $\mathbf{V}$  given in decreasing order and  $\mathbf{u}_j$  be corresponding eigenvectors of  $\mathbf{V}$  that satisfy the condition of orthonormality. Assume that  $\mathbf{D}$  is a diagonal matrix with diagonal elements  $\lambda_j$ , ( $j = 1, \dots, p$ ) and  $\mathbf{U}$  is an orthonormal matrix whose columns are the eigenvectors of  $\mathbf{V}$  given as  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p]$ . The elements  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p$  of the orthonormal matrix  $\mathbf{U}$  compose an orthonormal basis in  $\mathbb{R}^p$ . According to the singular value decomposition theorem,  $\mathbf{V}$  can be written as

$$\mathbf{V} = \mathbf{U}\mathbf{D}\mathbf{U}' \quad (3.12)$$

where due to the orthonormality  $\mathbf{U}' = \mathbf{U}^{-1}$  and

$$\mathbf{U}\mathbf{U}' = \mathbf{U}'\mathbf{U} = \mathbf{I} \quad (3.13)$$

In order to obtain a unit length vector  $\mathbf{x}$ , a vector  $\mathbf{y} = \mathbf{U}'\mathbf{x}$  is taken, so that  $\mathbf{x} = \mathbf{U}\mathbf{y}$ . Since  $\|\mathbf{u}_j\| = 1$ , the expressions  $\mathbf{x}'\mathbf{x}$  and  $\mathbf{y}'\mathbf{y}$  are equivalent (Horváth and Kokoszka, 2012).

$$\mathbf{x}'\mathbf{x} = \mathbf{y}'\mathbf{U}'\mathbf{U}\mathbf{y} = \mathbf{y}'\mathbf{y} \quad (3.14)$$

If the singular value decomposition of the variance covariance matrix  $\mathbf{V}$  in the equation (3.12) is used and  $\mathbf{x} = \mathbf{U}\mathbf{y}$  is set in the equation (3.11), the maximization problem takes the form of

$$\max \mathbf{y}'\mathbf{D}\mathbf{y} \quad \text{subject to} \quad \mathbf{y}'\mathbf{y} = 1. \quad (3.15)$$

This problem can be solved by setting  $\mathbf{y}$  to a unit vector  $\mathbf{y} = [1, 0, \dots, 0]'$ . Then,  $\mathbf{x}$  is obtained as the first column of  $\mathbf{U}$ , as  $\mathbf{x} = \mathbf{u}_1$  with the maximum value of eigenvalue  $\lambda_1$ .

These steps can be extended to find all eigenvectors of  $\mathbf{V}$  with the additional constraints  $\mathbf{x}'\mathbf{u}_1 = \mathbf{x}'\mathbf{u}_2 = \dots = \mathbf{x}'\mathbf{u}_{m-1} = 0$ .

In the  $m$ th step,  $\mathbf{x}$  is found as  $\mathbf{x} = \mathbf{u}_m$  where  $\mathbf{x}'\mathbf{V}\mathbf{x} = \lambda_m \mathbf{u}_m' \mathbf{u}_m = \lambda_m$ , is equal to the  $m$ th eigenvalue. So that, it is proved that the weight vectors  $\mathbf{u}_j$  in the equation (3.9) are equivalent to the eigenvectors of the variance covariance matrix related to the corresponding eigenvalue.

### 3.2 Principal Component Analysis for Functional Data

In functional principal component analysis (FPCA), the data consists of continuous functions. The weight vector  $\mathbf{u}_j$  and the centered observation vector  $\mathbf{x}_i$  in the multivariate context now become functions and are denoted by  $\xi_j(t)$  and  $\chi^c(t)$ , respectively. Therefore, the summation, which is used to take a linear combination of vectors, is replaced by an integral. In this case, the functional principal component scores are given as the inner product of empirical weight functions  $\hat{\xi}_j(t)$  and centered observation curves  $\chi^c(t)$  in  $q$  dimensional functional space.

$$f_{ij} = \int_T \hat{\xi}_j(t) \chi_i^c(t) dt, \quad i = 1, \dots, N, \quad j = 1, \dots, q \quad (3.16)$$

where  $\chi_i^c(t)$  are centered curves obtained from subtracting the mean function from each curve.

The steps of functional principal component analysis are similar to principal component analysis with a difference that now the operations are on functionals. As a first step, the weight function of the first functional principal component  $\xi_1(t)$  is found by maximizing the variance of the principal component scores

$$f_{i1} = \int_T \hat{\xi}_1(t) \chi_i^c(t) dt \quad (3.17)$$

under the restriction

$$\int_T \hat{\xi}_1^2(t) dt = \|\hat{\xi}_1\|^2 = 1. \quad (3.18)$$

In the next steps, other weight functions are computed for which the variance of functional principal component scores are maximized subject to the constraint  $\|\hat{\xi}_m\|^2 = 1$  and subject to  $m - 1$  additional constraints of orthogonality

$$\int_T \hat{\xi}_l(t) \hat{\xi}_m(t) dt = 0, \quad l < m. \quad (3.19)$$

In functional context, the variance covariance matrix  $\mathbf{V}$  leaves its place to the empirical covariance operator  $\hat{\Gamma}_\chi$ . Assume that the centered random curves  $\chi_i^c(t)$  are elements of the Hilbert space  $L^2[T]$  and it is provided that  $\mathbb{E}\|\chi\|^2 = \mathbb{E}[\int \chi(t)^2 dt] < \infty$  for every  $t \in T = [t_1, t_n]$ .

Then the maximization problem of the quadratic forms for the functional case is defined by

$$\max \langle \hat{\Gamma}_\chi \hat{\xi}, \hat{\xi} \rangle \quad \text{subject to} \quad \|\hat{\xi}\| = 1 \quad (3.20)$$

where  $\hat{\xi} \in H$ .

By using Riesz representation theorem mentioned in 2.6, this maximization problem can be rewritten as:

$$\langle \hat{\Gamma}_\chi \hat{\xi}, \hat{\xi} \rangle = \langle \mathbb{E}[\langle \chi_i^c(t), \hat{\xi}(s) \rangle \chi_i^c(s)], \hat{\xi}(t) \rangle = \mathbb{E} \left[ \langle \chi_i^c, \hat{\xi} \rangle^2 \right], \quad \hat{\xi} \in L^2[T]. \quad (3.21)$$

On the other hand, the goal of functional principal components analysis is to reduce dimension and to represent the data in  $q$  dimensional space for ease of calculation. Let  $\hat{\xi}_1, \hat{\xi}_2, \dots, \hat{\xi}_q$  be an orthonormal basis that consists of orthonormal functions. Once such an orthonormal basis is found so as to minimize the sum of square errors (SSE),

$$\text{SSE} = \sum_{i=1}^N \|\chi_i^c - \hat{\chi}_i^c\|^2, \quad (3.22)$$

each curve  $\hat{\chi}_i^c$  can be replaced by an expansion,

$$\hat{\chi}_i^c = \sum_{j=1}^q \langle \chi_i^c, \hat{\xi}_j \rangle \hat{\xi}_j, \quad (3.23)$$

such that  $q < N$ . The inner product  $\langle \chi_i^c, \hat{\xi}_j \rangle$  is the  $j$ th score of  $\chi_i$  which is defined by (3.16).

Under the condition of orthonormality  $\|\hat{\xi}_j\|^2 = 1$ , the equation (3.22) can be reexpressed as:

$$\begin{aligned} \sum_{i=1}^N \|\chi_i^c - \langle \chi_i^c, \hat{\xi}_j \rangle \hat{\xi}_j\|^2 &= \sum_{i=1}^N \|\chi_i^c\|^2 - 2 \sum_{i=1}^N \sum_{j=1}^q \langle \chi_i^c, \hat{\xi}_j \rangle^2 + \sum_{i=1}^N \sum_{j=1}^q \langle \chi_i^c, \hat{\xi}_j \rangle^2 \|\hat{\xi}_j\|^2 \\ &= \sum_{i=1}^N \|\chi_i^c\|^2 - \sum_{i=1}^N \sum_{j=1}^q \langle \chi_i^c, \hat{\xi}_j \rangle^2. \end{aligned} \quad (3.24)$$

So, minimizing SSE is equivalent to maximize the term  $\sum_{i=1}^N \sum_{j=1}^q \langle \chi_i^c, \hat{\xi}_j \rangle^2$  which is equal to maximize  $\langle \hat{\Gamma}_\chi \hat{\xi}, \hat{\xi} \rangle$ . The solutions of this maximization problem is obtained from the theorem below.

**Theorem** (Horváth and Kokoszka, 2012): Suppose that  $\Psi$  is a symmetric, positive definite Hilbert-Schmidt operator with eigenfunctions  $v_j$  and let  $\lambda_1 > \lambda_2 > \dots > \lambda_p > \lambda_{p+1}$  be the eigenvalues. Then,

$$\sup \{ \langle \Psi \xi, \xi \rangle : \|\xi\| = 1, \langle \xi, v_j \rangle = 0, \quad 1 \leq j \leq i-1, i < p \} = \lambda_i \quad (3.25)$$

is realized and the supremum is reached if  $\xi = v_i$ .

Assume that  $\hat{v}_j$ 's are the orthonormal eigenfunctions of the sample covariance operator  $\hat{\Gamma}_\chi$ . Then by Spectral Decomposition Theorem given in the Appendix 2.4, the maximization criterion can be expanded as follows,

$$\begin{aligned} \sum_{i=1}^N \sum_{j=1}^q \langle \chi_i^c, \hat{\xi}_j \rangle^2 &= \sum_{j=1}^q \langle \hat{\Gamma}_\chi \hat{\xi}_j, \hat{\xi}_j \rangle \\ &= \sum_{j=1}^{\infty} \hat{\lambda}_j \langle \hat{\xi}_1, \hat{v}_j \rangle^2 + \sum_{j=1}^{\infty} \hat{\lambda}_j \langle \hat{\xi}_2, \hat{v}_j \rangle^2 + \dots + \sum_{j=1}^{\infty} \hat{\lambda}_j \langle \hat{\xi}_q, \hat{v}_j \rangle^2. \end{aligned} \quad (3.26)$$

where  $\hat{\lambda}_j$  denote the empirical eigenvalues in decreasing order  $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots$

From the theorem mentioned above, it is obviously seen that  $\hat{\xi}_1 = \hat{v}_1, \hat{\xi}_2 = \hat{v}_2, \dots, \hat{\xi}_q = \hat{v}_q$ . Consequently, the weight functions  $\hat{\xi}_j$ 's are in fact orthonormal eigenfunctions of the sample covariance operator and they are called the empirical functional principal

component (EFPC). The  $j$ th score of  $\chi_i^c$  is interpreted as the weight of the contribution of the functional principal component  $\hat{\xi}_j$  to the curve  $\chi_i^c$ .

The empirical eigenfunctions of the covariance operator  $\hat{\xi}_j$  are solutions of the eigenequation defined by (3.27):

$$\hat{\Gamma}_\chi \hat{\xi}_j = \hat{\lambda}_j \hat{\xi}_j, \quad j \geq 1. \quad (3.27)$$

Considering that the empirical covariance operator can be defined as an integral transform with the kernel  $\hat{c}(s, t)$ , it can be reexpressed by

$$\int_T \hat{c}(t, s) \hat{\xi}_j(s) ds = \hat{\lambda}_j \hat{\xi}_j(t), \quad j \geq 1. \quad (3.28)$$

There are different ways to compute functional principal components. Mostly used method is the basis representation.

Considering that each observed function has a basis expansion such as mentioned in (2.21) and (2.22) in the previous chapter, the eigenfunctions can be defined in terms of basis functions as

$$\hat{\xi}(t) = \sum_{k=1}^K b_k \phi_k(t), \quad (3.29)$$

or equivalently in matrix notation

$$\hat{\xi}(t) = \boldsymbol{\phi}(t)' \mathbf{b}. \quad (3.30)$$

Using the basis expansion of functional data in the matrix form,

$$\hat{\boldsymbol{\chi}} = \mathbf{C} \boldsymbol{\phi} \quad (3.31)$$

the empirical variance covariance function can be written in matrix notation,

$$\hat{c}(t, s) = \frac{1}{N-1} \boldsymbol{\phi}(t)' \mathbf{C}' \mathbf{C} \boldsymbol{\phi}(s). \quad (3.32)$$

If an order  $K$  symmetric matrix  $\mathbf{W}$  is defined as

$$\mathbf{W} = \int \boldsymbol{\phi}(s)\boldsymbol{\phi}(s)'ds \quad (3.33)$$

then the eigenequation is equivalent to

$$\begin{aligned} \int_T \hat{c}(t,s)\hat{\xi}(s)ds &= \int \frac{1}{N-1}\boldsymbol{\phi}(t)'\mathbf{C}'\mathbf{C}\boldsymbol{\phi}(s)\boldsymbol{\phi}(s)'\mathbf{b}ds \\ &= \frac{1}{N-1}\boldsymbol{\phi}(t)'\mathbf{C}'\mathbf{C}\mathbf{W}\mathbf{b}. \end{aligned} \quad (3.34)$$

Since the weight functions can be written in terms of basis functions given as in the equation (3.30), the eigenequation problem takes the form of

$$\frac{1}{N-1}\boldsymbol{\phi}(t)'\mathbf{C}'\mathbf{C}\mathbf{W}\mathbf{b} = \hat{\lambda}\boldsymbol{\phi}(t)'\mathbf{b}. \quad (3.35)$$

This equation is realized for all  $t \in T$ , then  $\boldsymbol{\phi}(t)'$  can be removed and the equation can be redefined by

$$\frac{1}{N-1}\mathbf{C}'\mathbf{C}\mathbf{W}\mathbf{b} = \hat{\lambda}\mathbf{b}. \quad (3.36)$$

Here, the constraint  $\|\hat{\xi}\| = 1$  implies that  $\mathbf{b}\mathbf{W}\mathbf{b} = 1$  and the additional constraint of orthogonality of two functions  $\hat{\xi}_1$  and  $\hat{\xi}_2$  is denoted by  $\mathbf{b}_1\mathbf{W}\mathbf{b}_2 = 0$ .

To obtain principal components, the eigenvector  $\mathbf{u} = \mathbf{W}^{1/2}\mathbf{b}$  is defined and set into the equation (3.36). So it becomes,

$$\frac{1}{N-1}\mathbf{W}^{1/2}\mathbf{C}'\mathbf{C}\mathbf{W}^{1/2}\mathbf{u} = \hat{\lambda}\mathbf{u}. \quad (3.37)$$

and  $\mathbf{b} = \mathbf{W}^{-1/2}\mathbf{u}$  is computed for each eigenvector.

For orthonormal basis expansion, the equivalence  $\mathbf{W} = \mathbf{I}$  is realized which denotes that the weight matrix is equal to the identity matrix. In this case, functional principal component analysis is reduced to the multivariate principal component analysis with the following eigenequation problem

$$\frac{1}{N-1}\mathbf{C}'\mathbf{C}\mathbf{u} = \hat{\lambda}\mathbf{u}. \quad (3.38)$$

The second way to compute functional principal components is to use data driven basis instead of basis expansion (Ramsay and Silverman, 2005). In this case, the integral involved in the left side of the eigenequation problem should be approximated to a sum of discrete values. For a function  $f$ , this approximation can be shown as

$$\int f(t)dt \approx \sum_{j=1}^n w_j f(t_j) \quad (3.39)$$

where  $t_j$  indicates the argument values or in other words quadrature points,  $n$  is the number of argument values and  $w_j$ 's are the weights belonged to each function value in the sum which can also be called quadrature weights.

If an approximation such as given in the equation (3.39) is applied to the covariance operator, it takes the form of

$$\hat{\Gamma}_X \hat{\xi} \approx \mathbf{V} \mathbf{W} \tilde{\xi} \quad (3.40)$$

where the matrix  $\mathbf{V}$  composes of the covariance values  $\hat{c}(t_j, t_k)$  at the argument values,  $\tilde{\xi}$  is an order  $n$  vector that contains the values of  $\hat{\xi}(t_j)$  and  $\mathbf{W}$  is a diagonal matrix with diagonal elements equal to the quadrature weights  $w_j$ .

The approximately equivalent matrix eigenanalysis problem can be written in the form

$$\mathbf{V} \mathbf{W} \tilde{\xi} = \lambda \tilde{\xi}. \quad (3.41)$$

In this case, the constraints of orthonormality and orthogonality are defined by using weight matrix as in the equation (3.42).

$$\tilde{\xi}_m' \mathbf{W} \tilde{\xi}_m = 1 \quad \text{and} \quad \tilde{\xi}_m' \mathbf{W} \tilde{\xi}_k = 0, \quad k < m. \quad (3.42)$$

Similar to the rewritten eigenequation problem given in (3.37), the equation (3.41) can be rewritten by using the equality  $\mathbf{u} = \mathbf{W}^{1/2} \tilde{\xi}$  in terms of orthonormal vector  $\mathbf{u}$ .

$$\mathbf{W}^{1/2} \mathbf{V} \mathbf{W}^{1/2} \mathbf{u} = \hat{\lambda} \mathbf{u}. \quad (3.43)$$

The solution of this problem requires numerical iterations and consists of four steps:

- 1) The determination of the values of  $n, w_j$ 's and  $t_j$ 's
- 2) The computation of eigenvalues  $\hat{\lambda}_j$ 's and eigenvectors  $\mathbf{u}_j$ 's of the matrix  $\mathbf{W}^{1/2}\mathbf{V}\mathbf{W}^{1/2}$
- 3) The computation of

$$\tilde{\xi}_m = \mathbf{W}^{1/2}\mathbf{u}_m. \quad (3.44)$$

- 4) Using an interpolation technique to convert discretized values of the vector  $\tilde{\xi}_m$  to a function  $\hat{\xi}_m$ .

From this procedure, the number of obtained eigenfunctions are as much as the number of argument values  $n$ .

There are several methods proposed to determine the number  $q$  of the functional principal components is an important step. The most popular of these methods is the scree plot which plots the eigenvalues  $\hat{\lambda}_j$  against  $j$ . The point where the decrease of the eigenvalues appear is chosen as the selected value of  $q$  (Horváth and Kokoszka, 2012).

Another method to determine the number of components is the Cumulative Percentage of Variance (CPV). CPV criteria is defined as follows.

$$\text{CPV}(q) = \frac{\sum_{k=1}^q \hat{\lambda}_k}{\sum_{k=1}^N \hat{\lambda}_k}. \quad (3.45)$$

Horváth and Kokoszka (2012) recommend that the  $p$  is chosen for which CPV criteria exceeds a level of 85%. In addition to these methods, pseudo-AIC or cross-validation methods can also be used to determine the number of components.

One of the important tool in functional principal components analysis is to visualise the results. This step is important to interpret the components. To visualise results the overall mean function of the functional data is computed and a suitable multiple of each principal component curve is added to and subtracted from the functional mean. The mean function and the effects of adding and subtracting the multiple of each curve is plotted for each component. The choice of the multiple of the principal component function is important. It is a constant and can be defined by  $C$ . It is generally chosen



subjectively. This constant  $C$  can differ from function to function but for ease of comparison it is offered to be the same for all the principal components functions (Ramsay and Silverman, 2005). In the study of Ramsay and Silverman (2005)  $C$  is chosen as 0.2 to give easily interpretable results.



## 4 FUNCTIONAL LINEAR MODELS

Linear models has an important place in classical statistical analysis to reveal the linear relationship between variables and to make predictions in classical statistical analysis. As in the case of principal component analysis, linear models can also be adapted to functional data with the name functional linear models.

A linear model is called a functional linear model (FLM) due to the functional structure of the variables that construct the linear model. FLMs are divided into different groups according to the structure of the response and independent variables. A FLM with a functional response and a categorical independent variable is called a functional analysis of variance model whereas a model with scalar response and functional variables or a model with a functional response is called a functional multiple regression model.

Functional linear regression models (FLRM) can be summarized basically as in Table (4.1).

**Table 4.1:** Functional Linear Regression Models

Model	Response Variable	Regressors	Functional Linear Model
The Functional Response Model	Functional	Scalar	$\gamma_i(t) = \beta(t)X_i + \varepsilon_i(t)$
The Fully Functional Model	Functional	Functional	$\gamma_i(t) = \int \beta(t,s)\chi_i(s)ds + \varepsilon_i(t)$
The Scalar Response Model	Scalar	Functional	$Y_i = \int \beta(s)\chi_i(s)ds + \varepsilon_i$

In this chapter, we will focus on functional linear regression models for scalar responses which are used to model the relationship between a real random variable and a functional random variable defined on a interval of length  $T$ . Several studies can be found in the literature that aims to explain a scalar response form a functional

variable (Ramsay and Silverman, 2005; Cardot et al., 1999; Aguilera et al., 2010; Preda and Schiltz, 2011). Specifically, these models are mostly used in chemometrics where the explanatory variables are functions of wavelenghts such as in the studies of Ferraty and Vieu (2006) and Aguilera et al. (2010).

#### 4.1 Functional Linear Regression Models (FLRM) for Scalar Responses

A general linear regression model with scalar response is in the form of

$$\mathbf{Y} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (4.1)$$

where  $\mathbf{Y}$  denotes the  $N \times 1$  response vector,  $\mathbf{Z}$  is the  $N \times p$  design matrix,  $\boldsymbol{\beta}$  is the  $p \times 1$  parameter vector and  $\boldsymbol{\varepsilon}$  is the  $N \times 1$  error term.

The main difference between a linear regression model and a functional linear regression model is that some elements are now square integrable functions rather than vectors. In this manner, the functional scalar response model takes the form

$$\mathbf{Y} = \langle \boldsymbol{\chi}, \boldsymbol{\beta} \rangle + \boldsymbol{\varepsilon} = \int_T \boldsymbol{\chi}(s)\boldsymbol{\beta}(s)ds + \boldsymbol{\varepsilon}. \quad (4.2)$$

Here the predictor is a square integrable function taking values at point  $t \in T$  and satisfy the conditions  $E[\|\boldsymbol{\chi}\|^2] \leq \infty$  and  $\mu_{\boldsymbol{\chi}}(t) = E[\boldsymbol{\chi}(t)]$ , the functional slope of the model  $\boldsymbol{\beta}(s)$  is a square integrable function in  $L^2(T)$ , the error term  $\boldsymbol{\varepsilon}$  is a real random variable with mean 0 and variance  $\sigma_{\boldsymbol{\varepsilon}}^2$  (Febrero-Bande et al., 2015).

The functional linear regression model given in (4.2) has infinite number of solutions that give perfect prediction of the response (Ramsay and Silverman, 2005). Therefore, different techniques based on basis functions and nonparametric smoothing are proposed to assess an interpretable estimate of the parameter function  $\boldsymbol{\beta}(s)$  (Ramsay and Silverman, 2005; Ferraty and Vieu, 2006; Aguilera et al., 2010). Here, mostly preferred three methods will be explained. The first method is to use classical basis representation. The other methods are based on dimension reduction methods principal components analysis and partial least squares analysis. Therefore, they are named as Functional Principal Components Regression (FPCR) and Functional Partial Least Squares Regression (FPLS). FPCR and FPLSR models can be constructed based on

two approaches. The first one is the basis expansion approach. The other approach was offered by Febrero-Bande and Oviedo de la Fuente (2012) and it is based on the functional components obtained from the eigenstructure of the data. In this study, we will focus on the second approach. More information about constructing FPCR and FPLSR models in terms of basis functions can be found in the studies of Ocaña et al. (2007); Aguilera et al. (2010); Aguilera Morillo (2013).

#### 4.1.1 Functional Linear Regression Models (FLRM) with Basis Representation

The functional linear model in the equation (4.2) can be estimated from the following expression:

$$\hat{Y}_i = \int_T \chi_i(s) \hat{\beta}(s) ds. \quad (4.3)$$

This expression defines the inner product  $\langle \chi_i, \hat{\beta} \rangle$  and can be approximated by using basis expansion.

An estimate of the parameter function  $\beta(s)$  is the basis expansion in terms of basis functions  $\theta_l$ 's

$$\beta^*(s) = \sum_{l=1}^L b_l \theta_l(s). \quad (4.4)$$

Here the coefficients  $b_1, b_2, \dots, b_L$  are estimated from the minimization of the criteria sum of square errors

$$\sum_{i=1}^N \left[ Y_i - \int_T \chi_i(s) \beta^*(s) ds \right]^2 = \sum_{i=1}^N \left[ Y_i - \sum_{l=1}^L \langle \chi_i, \theta_l \rangle b_l \right]^2. \quad (4.5)$$

So the estimation of the parameter function is defined by

$$\hat{\beta}(s) = \sum_{k=1}^L \hat{b}_k \theta_k(s) = \boldsymbol{\theta}(s)' \hat{\mathbf{b}}. \quad (4.6)$$

On the other hand, the functional predictors  $\chi_i(s)$  can also be expanded in terms of basis functions  $\phi_1(s), \phi_2(s), \dots, \phi_K(s)$  such as mentioned in Chapter 2,

$$\chi_i(s) = \sum_{k=1}^K c_k \phi_k(s) = \mathbf{C}\boldsymbol{\phi}(s). \quad (4.7)$$

By substituting the expansions given in the equations (4.6) and (4.7) into the model (4.3), the functional linear model can be redefined by

$$\hat{Y}_i = \int \mathbf{C}\boldsymbol{\phi}(s)\boldsymbol{\theta}(s)'\hat{\mathbf{b}}ds, \quad (4.8)$$

such that  $\mathbf{C}$  and  $\hat{\mathbf{b}}$  indicate the basis coefficient vectors of the functions  $\chi_i(s)$  and  $\hat{\beta}(s)$  respectively.

Let  $\mathbf{J}_{\phi\theta}$  define a  $K \times L$  dimension matrix of inner product between the basis functions,

$$\mathbf{J}_{\phi\theta} = \int \boldsymbol{\phi}(s)\boldsymbol{\theta}'(s)ds. \quad (4.9)$$

Then the equation (4.8) can be rewritten in matrix form as

$$\hat{Y}_i = \mathbf{C}\mathbf{J}_{\phi\theta}\hat{\mathbf{b}}. \quad (4.10)$$

Considering the intercept term  $b_0$ , a  $L + 1$  length parameter vector  $\boldsymbol{\zeta} = (b_0, b_1, \dots, b_K)$  and a  $N \times K$  dimension design matrix  $\mathbf{Z} = [1 \ \mathbf{C}\mathbf{J}_{\phi\theta}]$  can be defined. Hence, the functional linear regression problem given in the expression (4.10) is induced to an ordinary linear regression problem with the estimated parameter vector  $\hat{\boldsymbol{\zeta}}$  such that,

$$\hat{\mathbf{Y}} = \mathbf{Z}\hat{\boldsymbol{\zeta}}. \quad (4.11)$$

The sum of square errors (SSE) of the functional linear regression model is defined as

$$\text{SSE} = \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^N (Y_i - \langle \chi_i, \hat{\beta} \rangle)^2. \quad (4.12)$$

By using equations (4.11) and (4.12) SSE can be rewritten in matrix form as

$$\text{SSE} = \|\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\zeta}}\|^2. \quad (4.13)$$

This problem can be seen as an ordinary linear regression problem and the least squares estimate of  $\boldsymbol{\zeta}$  is estimated by minimizing (4.13), yielding to

$$\mathbf{Z}'\mathbf{Z}\hat{\boldsymbol{\zeta}} - \mathbf{Z}'\mathbf{Y} = 0. \quad (4.14)$$

Hence,  $\hat{\boldsymbol{\zeta}}$  is obtained as

$$\hat{\boldsymbol{\zeta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}. \quad (4.15)$$

Substituting (4.15) into the equation (4.11) gives the predicted values of the response  $\hat{\mathbf{Y}}$

$$\hat{\mathbf{Y}} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}, \quad (4.16)$$

where  $\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$  defines the hat matrix such as in ordinary multiple linear regression with the degrees of freedom  $df = \text{trace}(\mathbf{H})$ .

#### 4.1.2 Roughness Penalty Approach

A functional linear regression model in the form of (4.3) can also be fitted by a roughness penalty approach to obtain a smooth estimate of the parameter function. Then the penalized sum of squares problem in functional context is defined by

$$\sum_{i=1}^N [Y_i - \hat{Y}_i]^2 + \lambda \int [D^m \beta(s)]^2 ds. \quad (4.17)$$

where  $\lambda$  is a smoothing parameter and  $D^m$  is the linear differential operator which denotes the  $m$ th derivative of the parameter function  $\beta(s)$ . The last term of (4.17) refers the roughness penalty and it is chosen depending on the structure of the parameter function  $\beta(s)$ . Usually the second derivative of the parameter function is chosen as the roughness penalty (Ramsay and Silverman, 2005).

Assume that  $R$  is a penalization matrix as defined in the equation (2.59). Then the roughness penalty can be written in matrix form. Considering (4.10), the penalized residual sum of squares problem in matrix form is expressed by

$$\text{PENSSE}_\lambda(\hat{b}_0, \hat{\mathbf{b}}) = \|\mathbf{Y} - \hat{b}_0 - \mathbf{C}\mathbf{J}_{\phi, \theta} \hat{\mathbf{b}}\|^2 + \lambda \hat{\mathbf{b}}' \mathbf{R} \hat{\mathbf{b}}. \quad (4.18)$$

The expression (4.18) can be rewritten in a simpler form in terms of augmented parameter vector  $\hat{\boldsymbol{\zeta}} = (\hat{b}_0, \hat{\mathbf{b}}')'$  and the design matrix  $\mathbf{Z}$

$$\text{PENSSE}_\lambda(\hat{\boldsymbol{\zeta}}) = \|\mathbf{Y} - \mathbf{Z}\hat{\boldsymbol{\zeta}}\|^2 + \lambda \hat{\boldsymbol{\zeta}}' \mathbf{R}_0 \hat{\boldsymbol{\zeta}}. \quad (4.19)$$

where  $\mathbf{R}_0$  is a augmented penalty matrix obtained by attaching a column and row of  $K$  zeros.

The solution of this problem is obtained by solving the equation

$$(\mathbf{Z}'\mathbf{Z} + \lambda \mathbf{R}_0)\hat{\boldsymbol{\zeta}} = \mathbf{Z}'\mathbf{Y}. \quad (4.20)$$

Hence, the parameter vector  $\boldsymbol{\zeta}$  is estimated as

$$\hat{\boldsymbol{\zeta}} = (\mathbf{Z}'\mathbf{Z} + \lambda \mathbf{R}_0)^{-1} \mathbf{Z}'\mathbf{Y} \quad (4.21)$$

with the variance

$$\text{Var}[\hat{\boldsymbol{\zeta}}] = \sigma_\varepsilon^2 (\mathbf{Z}'\mathbf{Z} + \lambda \mathbf{R}_0)^{-1} \mathbf{Z}'\mathbf{Z} (\mathbf{Z}'\mathbf{Z} + \lambda \mathbf{R}_0)^{-1}. \quad (4.22)$$

The smoothing parameter  $\lambda$  is chosen subjectively or depending on one of the methods mentioned in Chapter 2 such as cross validation.

The cross validation algorithm can be adapted to this approach as follows:

$$\text{CV}_\lambda = \sum_{i=1}^N [Y_i - \hat{b}_0^{(-i)} - \int z_i(s) \hat{\beta}_\lambda^{(-i)}(s) ds]^2 ds \quad (4.23)$$

where  $b_0^{(-i)}$  and  $\beta_\lambda^{(-i)}(s)$  are respectively the estimates of the intercept and  $\beta$  parameter obtained by minimizing the penalized residual sum of squares based on all the data except the  $i$ th data pair  $(z_i, y_i)$ . The  $\lambda$  for which the CV criteria is minimum gives the optimum lambda value (Ramsay and Silverman, 2005).

Another definition of  $\text{CV}_\lambda$  is based on the the hat matrix  $\mathbf{H}$ . In the penalized least squares regression the hat matrix  $\mathbf{H}$  is defined by the equation

$$\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z} + \lambda \mathbf{R})^{-1} \mathbf{Z}' \quad (4.24)$$

Let  $h_{ii}$  denote the diagonal elements of the hat matrix, then  $CV_\lambda$  can be computed from

$$CV_\lambda = \sum_{i=1}^N \left( \frac{Y_i - \hat{Y}_i}{1 - h_{ii}} \right)^2. \quad (4.25)$$

Due to its computational simplicity, mostly basis expansion approach is preferred in solving functional linear regression problems. But in the case of multicollinearity, this method alone would not be enough. In this case, functional principal components regression and functional partial least squares regression methods are offered to avoid multicollinearity and to reduce dimension. The latter depends on the relationship between the scalar response variable and the functional explanatory variable(s) and is mostly used in the case of functional scalar response models.

### 4.1.3 Functional Principal Component Regression (FPCR)

The idea of FPCR is to predict scalar response vector  $\mathbf{Y}$  based on the functional principal component scores. Considering that the eigenfunctions  $\xi_1, \xi_2, \dots$  of the covariance operator  $\Gamma_\chi$  form an orthonormal basis in  $L^2(T)$ , the stochastic process  $\chi(t)$ ,  $t \in T$  with mean  $\mathbb{E}[\chi] = \mu_\chi$  can be represented in terms of functional principal components according to the Karhunen-Loève Theorem given in (B.12),

$$\hat{\chi}(t) = \mu_\chi + \sum_{j=1}^{\infty} f_j \xi_j(t), \quad i = 1, \dots, N \quad (4.26)$$

where  $f_j$ ,  $j \geq 1$  are uncorrelated zero mean random variables that compose of principal component scores.

Assume that  $\mathbf{Y}$  is the response vector with mean  $\mathbb{E}[\mathbf{Y}] = \mu_{\mathbf{Y}}$  and the variance  $\sigma_{\mathbf{Y}}^2 = \mathbb{E}[(\mathbf{Y} - \mu_{\mathbf{Y}})^2]$ . So the functional linear model with the scalar response  $\mathbf{Y}$  can be expressed by

$$\mathbf{Y} = \mu_{\mathbf{Y}} + \langle \chi - \mu_\chi, \beta \rangle + \varepsilon = \mu_{\mathbf{Y}} + \int_T (\chi(t) - \mu_\chi(t)) \beta(t) dt + \varepsilon. \quad (4.27)$$

$\chi(t)$  and the set of functional principal component scores  $f_j$  span the same linear space. Thus, the functional linear regression of the scalar response on  $\chi(t)$  is equivalent to



the linear regression of the response on the principal component scores  $f_j$  (Preda and Saporta, 2005). This can easily be seen when the expansion (4.26) is set into the equation (4.27).

$$\begin{aligned}
\mathbf{Y} &= \mu_{\mathbf{Y}} + \int_T \chi(t)\beta(t)dt - \int_T \mu_{\chi}(t)\beta(t)dt + \varepsilon \\
&= \mu_{\mathbf{Y}} + \int_T \mu_{\chi}(t)\beta(t)dt + \int_T \sum_{j=1}^{\infty} f_j \xi_j(t)\beta(t)dt - \int_T \mu_{\chi}(t)\beta(t)dt + \varepsilon \\
&= \mu_{\mathbf{Y}} + \int_T \sum_{j=1}^{\infty} f_j \xi_j(t)\beta(t)dt + \varepsilon.
\end{aligned} \tag{4.28}$$

Equation (4.28) defines an ordinary simple linear regression problem in the form of

$$\mathbf{Y} = \mu_{\mathbf{Y}} + \sum_{j=1}^{\infty} f_j b_j + \varepsilon, \tag{4.29}$$

where  $f_j = \langle \chi^c, \xi_j \rangle$  is the explanatory variable and  $b_j = \int_T \xi_j(t)\beta(t)dt = \langle \beta, \xi_j \rangle$  is the coefficient vector.

Hence, the functional slope  $\beta(t)$  can be written in terms of orthonormal principal component functions such that,

$$\beta(t) = \sum_{j=1}^{\infty} b_j \xi_j(t). \tag{4.30}$$

If the term  $f_j$  in the model (4.29) is expanded, the intercept term of the functional linear regression model is found as,

$$b_0 = \mu_{\mathbf{Y}} - \int_T \mu_{\chi}(t)\beta(t)dt. \tag{4.31}$$

Due to the fact that the parameter function  $\beta(t)$  is a square integrable function, it is satisfied that  $\sum_{j=1}^{\infty} b_j^2 < \infty$ . So the regression coefficients  $b_j$  of the model (4.29) can be computed from (4.32) such as in an ordinary simple linear regression problem,

$$b_j = \frac{\text{Cov}(\mathbf{Y}, f_j)}{\text{Var}(f_j)}. \tag{4.32}$$

Then, considering that the variance of the principal components scores is equal to the related eigenvalue of the covariance operator obtained from the equation  $\Gamma_{\chi} \xi_j = \lambda_j \xi_j$ , the functional slope  $\beta(t)$  can be rewritten as

$$\beta(t) = \sum_{j=1}^{\infty} \frac{\text{Cov}(\mathbf{Y}, f_j)}{\text{Var}(f_j)} \xi_j(t) = \sum_{j=1}^{\infty} \frac{c_{\mathbf{Y}, f_j}}{\lambda_j} \xi_j(t). \quad (4.33)$$

This result can also be given by a Lemma.

**Lemma** (Horváth and Kokoszka, 2012): Suppose that  $\chi(t)$  is a centered process,  $Y$  and  $\varepsilon$  are zero mean random variables. The following linear model holds

$$Y = \int_T \beta(t) \chi(t) dt + \varepsilon, \quad (4.34)$$

under the condition that  $\beta$  is a square integrable function with  $\int \int \beta^2(t) dt < \infty$ .

Let  $\xi_j(t)$  be the functional principal components and  $f_j$  be the scores defined by the inner product  $\langle \chi, \xi_j \rangle$ . Then it is realized that

$$\beta(t) = \sum_{j=1}^{\infty} \frac{\mathbb{E}[f_j Y]}{\mathbb{E}[f_j^2]} \xi_j(t). \quad (4.35)$$

The goodness of fit of the functional linear model is measured by the coefficient of the determination that is denoted by  $R^2$ .  $R^2$  is the proportion of the scalar response variance explained by the functional predictor  $\chi$  (Febrero-Bande et al., 2015). In functional case it is computed from

$$R^2 = \frac{\text{Var}(\mathbb{E}[\mathbf{Y}|\chi])}{\text{Var}[\mathbf{Y}]} \quad (4.36)$$

The numerator of the formula (4.36) is equal to

$$\text{Var}(\mathbb{E}[\mathbf{Y}|\chi]) = \sum_{j=1}^{\infty} \lambda_j b_j^2 = \sum_{j=1}^{\infty} \frac{c_{\mathbf{Y}, f_j}^2}{\lambda_j}. \quad (4.37)$$

Therefore the coefficient of determination can be rewritten as

$$R^2 = \frac{1}{\text{Var}[\mathbf{Y}]} \sum_{j=1}^{\infty} \frac{c_{\mathbf{Y}, f_j}^2}{\lambda_j} = \sum_{j=1}^{\infty} r_{\mathbf{Y}, f_j}^2 \quad (4.38)$$

where  $r_{\mathbf{Y},f_j}^2$  denotes the correlation between the scalar response and FPC scores. Hence, it can be said that the goodness of fit of FPCR depends on the relationship between the response and the FPC scores.

Consider a sample of independent identically distributed observations  $\chi_1, \chi_2, \dots, \chi_N$  that have the same distribution as  $\chi(t)$ . Then the functional linear regression model for the sample drawn from  $(\chi, \mathbf{Y})$  is written in the form of

$$\hat{Y}_i = \bar{Y} + \int_T (\chi_i(t) - \bar{\chi}(t)) \hat{\beta}_j(t) dt, \quad i = 1, 2, \dots, N, \quad j = 1, \dots, K. \quad (4.39)$$

Let  $\hat{\xi}_j$  and  $\hat{\lambda}_j$  denote the empirical eigenfunctions and eigenvalues of the sample covariance operator  $\hat{\Gamma}_\chi$  and  $\hat{f}_j = (\hat{f}_{1j}, \dots, \hat{f}_{Nj})^T$  be the sample functional principal component scores where  $\hat{f}_{ij} = \langle \chi_i - \bar{\chi}, \hat{\xi}_j \rangle$  for  $i = 1, \dots, N$  and  $j = 1, \dots, K$ . Then the slope of the functional linear regression model is estimated from the projection of the regressors onto the  $K$  empirical functional components  $\hat{\xi}_j, j = 1, \dots, K$  related to the largest eigenvalues  $\hat{\lambda}_j$

$$\hat{\beta}_j(t) = \sum_{j=1}^K \hat{b}_j \hat{\xi}_j(t). \quad (4.40)$$

To compute  $\hat{\beta}_j(t)$ , first the values of  $\hat{b}_j$  which minimize the criteria (4.41) should be estimated,

$$\sum_{i=1}^N [Y_i - \sum_{j=1}^K \langle \chi_i^c(t), \hat{\xi}_j \rangle \hat{b}_j]^2. \quad (4.41)$$

From Lemma (Horváth and Kokoszka, 2012) mentioned above, it is known that the estimated regression coefficients  $\hat{b}_j$  is computed from the ratio of covariance between the scalar response and the sample principal component scores to the variance of the FPC scores  $f_{ij}$  that is equal to the associated eigenvalue  $\hat{\lambda}_j$ ,

$$\hat{b}_j = \frac{\text{Cov}(\mathbf{Y}, f_j)}{\text{Var}(f_j)} = \frac{\hat{c}_{\mathbf{Y},f_j}}{\hat{\lambda}_j}, \quad j = 1, \dots, K \quad (4.42)$$

So, the estimation of the parameter function  $\beta(t)$  can be rewritten in terms of sample counterparts of the covariance function, eigenfunctions and the eigenvalues as in the equation (4.43)

$$\hat{\beta}_{FPC}(t) = \sum_{j=1}^K \frac{\hat{c}_{\mathbf{Y},f_j}}{\hat{\lambda}_j} \hat{\xi}_j(t). \quad (4.43)$$

The coefficient vector  $\hat{b}_j$  estimated from the first  $K$  principal components is expressed by

$$\hat{b}_j = \left( \frac{f_{.1}^T \mathbf{Y}}{N \hat{\lambda}_1}, \frac{f_{.2}^T \mathbf{Y}}{N \hat{\lambda}_2}, \dots, \frac{f_{.K}^T \mathbf{Y}}{N \hat{\lambda}_K} \right). \quad (4.44)$$

The predicted values of the response are obtained from the equality  $\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}$  where  $\mathbf{H}$  is a hat matrix with degrees of freedom  $df = \text{trace}(\mathbf{H}) = K$ .  $\mathbf{H}$  consists of the elements

$$\mathbf{H} = \left( \frac{f_{.1} f_{.1}^T \mathbf{Y}}{N \hat{\lambda}_1}, \frac{f_{.2} f_{.2}^T \mathbf{Y}}{N \hat{\lambda}_2}, \dots, \frac{f_{.K} f_{.K}^T \mathbf{Y}}{N \hat{\lambda}_K} \right). \quad (4.45)$$

Hence, the vector of residuals is equal to

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{H})\mathbf{Y}. \quad (4.46)$$

with the variance

$$\hat{\sigma}_{\boldsymbol{\varepsilon}}^2 = \frac{\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}_i^T}{N - K - 1}. \quad (4.47)$$

To measure the goodness of fit of an estimated functional linear model, the sample coefficient of determination is used which is defined by

$$\hat{R}^2 = \sum_{j=1}^K \hat{\lambda}_j \hat{b}_j^2 = \sum_{j=1}^K \hat{r}_{\mathbf{Y},f_j}^2 \quad (4.48)$$

The most important problem in FPCR is to choose the optimum number of principal components. Cross-validation criteria (CV) or different model selection criteria can be used to select the number  $K$  of principal components that best estimate the response (Febrero-Bande and Oviedo de la Fuente, 2012).

The CV criteria is given with the formula

$$\text{CV}(K) = \frac{1}{N} \sum_{i=1}^N (Y_i - \langle \chi_i, \hat{\beta}_{(-i,K)} \rangle)^2. \quad (4.49)$$

The general formula for the model selection criteria (MSC) can be written as

$$\text{MSC}(K) = \log \left[ \frac{1}{N} \sum_{i=1}^N \left( Y_i - \langle \chi_i, \hat{\beta}_{(i,K)} \rangle \right)^2 \right] + p_N \frac{K}{N}. \quad (4.50)$$

The last term  $p_n$  in this formula differs according to the type of the information criterion. The mostly used criteria are Schwarz Information Criteria (SIC), Corrected SIC (SICc), Akaike-Information Criteria (AIC) and Corrected AIC (AICc). The value of  $K$  which minimize the criteria MSC is taken as the number of projections.

The values of  $p_n$  for the criterias SIC, SICc, AIC and AICc are given respectively as below (Febrero-Bande and Oviedo de la Fuente, 2012).

$$\text{SIC}(p_N) = \frac{\log(N)}{N}. \quad (4.51)$$

$$\text{SICc}(p_N) = \frac{\log(N)}{N - K - 2}. \quad (4.52)$$

$$\text{AIC}(p_N) = 2. \quad (4.53)$$

$$\text{AICc}(p_N) = \frac{2N}{N - K - 2}. \quad (4.54)$$

#### 4.1.4 Functional Partial Least Squares Regression (FPLSR)

Partial Least Squares Regression (PLSR) is a popular estimation method alternative to PCR which takes into account the correlation between the response and the predictors (Delaigle et al., 2012). This method is mostly preferred in chemometrics (Aguilera et al., 2010). PLS approach is adapted to the functional context by Preda and Saporta (2005). Wang et al. (2009) proposed a new PLS approach to solve the multicollinearity problem that is encountered in the multiple linear regression of functional data. There are several studies of FPLSR in the literature for the case of scalar response (Preda and Saporta, 2005; Reiss and Ogden, 2007; Aguilera et al., 2010; Febrero-Bande et al., 2015). Recently, Preda and Schiltz (2011) extended this approach to the case where both the response and the predictors are functionals.

Functional partial least squares (FPLS) method is based on the maximization of the covariance between the functional independent variable  $\chi(t)$  and the scalar response  $Y$ . Due to the fact that FPLS components are more related to the variability of the response, they are more relevant to predict the outcome (Oviedo de la Fuente, 2011; Reiss and Ogden, 2007). The maximization criterion, which is known as Tucker's criterion in the literature, can be defined by

$$\max \text{Cov}^2\left(\int_T \chi(t)\varphi(t)dt, \mathbf{Y}\right) \quad (4.55)$$

where  $\varphi(t)$  is a functional element of  $L^2(T)$  with the norm  $\|\varphi(t)\|_{L^2(T)} = 1$  (Aguilera et al., 2010).

FPLS components are computed from an iterative algorithm. The steps of this algorithm can be summarized as follows (Febrero-Bande et al., 2015):

1. The response vector and the functional predictors are centered and defined as  $Y^c = Y - \mu_Y$  and  $\chi^c = \chi - \mu_\chi$  respectively. For the first step, let  $l = 0$ . The initial values of the response  $Y_l$  and predictor  $\chi_l$  are taken as  $Y_0 = Y^c$  and  $\chi_0 = \chi^c$ .
2. Let  $\varphi_{l+1} \in L^2(T)$  be the associated weight function defined by (4.56),

$$\varphi_{l+1} = \frac{\text{Cov}(Y_l, \chi_l)}{\|\text{Cov}(Y_l, \chi_l)\|} = \frac{E[Y_l \chi_l(t)]}{\int_T E[Y_l \chi_l(t)] dt}, \quad t \in [0, T], \quad (4.56)$$

and let  $v_{l+1} = \langle \chi_l, \varphi_{l+1} \rangle$  denote FPLS component score which is chosen so as to maximize the square of  $\text{Cov}^2[Y_l, v_{l+1}] = c_{Y_l, v_{l+1}}^2$  as given in Tucker's criterion.

3. Regress  $Y_l$  and  $\chi_l$  on the partial least squares components to obtain the regression coefficient estimates  $\psi_{l+1} \in \mathbb{R}$  and  $\delta_{l+1} \in L^2[T]$  which are obtained from the regression models (4.57) and (4.58):

$$Y_l = \psi_{l+1} v_{l+1} + \varepsilon_l(t), \quad (4.57)$$

$$\chi_l = \delta_{l+1} v_{l+1} + \eta_l(t), \quad (4.58)$$

as in given in the equations (4.59) and (4.60):

$$\psi_{l+1} = \frac{\text{Cov}[Y_l, v_{l+1}]}{\text{Var}[v_{l+1}]}, \quad (4.59)$$

$$\delta_{l+1} = \frac{\text{Cov}[\chi_l, v_{l+1}]}{\text{Var}[v_{l+1}]}. \quad (4.60)$$

4. Set  $Y_{l+1}$  and  $\chi_{l+1}$  as error terms of the models (4.57) and (4.58) such as given in the equations (4.61) and (4.62) respectively:

$$Y_{l+1} = Y_l - \psi_{l+1} v_{l+1}, \quad (4.61)$$

$$\chi_{l+1} = \chi_l - \delta_{l+1} v_{l+1}. \quad (4.62)$$

5. Take  $l = l + 1$  and repeat the algorithm from second step on.

The FPLS components  $v_l$  are defined by the inner product of  $\{\chi(t) : t \in [0, T]\}$  and weight functions  $\varphi_l(t)$  such as given in (4.63):

$$v_l = \langle \chi_{l-1}, \varphi_l \rangle = \int_T \chi_{l-1}(t) \varphi_l(t) dt. \quad (4.63)$$

As in the case of FPCR, FPLSR is based on regressing the response vector on the FPLS components. The functional linear model (4.29) can be represented as a linear combination of PLS components (Oviedo de la Fuente, 2011). The parameter function  $\beta(t)$  can be found from expansion of this equation:

$$\mathbf{Y} = \mu_{\mathbf{Y}} + \sum_{l=1}^{\infty} c_l v_l + \varepsilon \quad (4.64)$$

For beginning, assume that just one FPLS component is used. For  $l = 1$ , the equation (4.64) is equal to

$$\mathbf{Y} = \mu_{\mathbf{Y}} + c_1 v_1 + \varepsilon. \quad (4.65)$$

From the definition of FPLS components as given in (4.63),  $v_1$  can be written as

$$v_1 = \int_T \chi_0(t) \varphi_1(t) dt. \quad (4.66)$$

By using the initial value of  $\chi_0 = \chi^c$  and setting  $\varphi_1(t)$  as  $\phi_1(t)$  the equation takes the form

$$v_1 = \int_T \chi^c(t) \phi_1(t) dt. \quad (4.67)$$

Then (4.65) can be rewritten as

$$\mathbf{Y} = \mu_{\mathbf{Y}} + c_1 \int_T \chi^c(t) \phi_1(t) dt + \varepsilon. \quad (4.68)$$

Hence it can be easily seen that for  $l = 1$  the  $\beta(t)$  functional parameter is equal to

$$\beta(t) = c_1 \phi_1(t). \quad (4.69)$$

Let us take two FPLS components. For  $l = 2$ , the scalar response variable can be written as a linear combination of 2 FPLS components.

$$\begin{aligned} \mathbf{Y} &= \mu_{\mathbf{Y}} + \sum_{l=1}^2 c_l v_l + \varepsilon \\ &= \mu_{\mathbf{Y}} + c_1 v_1 + c_2 v_2 + \varepsilon. \end{aligned} \quad (4.70)$$

The expansion of  $v_1$  was given in (4.67). By using the equation (4.62),  $\chi_1(t)$  is equal to

$$\chi_1(t) = \chi_0(t) - \delta_1 v_1 = \chi^c(t) - \delta_1 v_1. \quad (4.71)$$

Then the expansion of  $v_2$  can be written as

$$\begin{aligned} v_2 &= \langle \chi_1, \varphi_2 \rangle = \int_T (\chi^c - \delta_1 v_1) \varphi_2(t) dt \\ &= \int_T (\chi^c \varphi_2(t) - \langle \delta_1, \varphi_2 \rangle v_1) dt \\ &= \int_T \chi^c \varphi_2(t) dt - \int_T \chi^c(t) \langle \delta_1, \varphi_2 \rangle \varphi_1(t) dt \\ &= \int_T \chi^c(t) [\varphi_2(t) - \langle \delta_1, \varphi_2 \rangle \varphi_1(t)] dt \end{aligned} \quad (4.72)$$

By setting (4.72) into the equation (4.70) the model takes the form of

$$Y = \mu_Y + c_1 \int_T \chi^c(t) \phi_1(t) dt + \int_T \chi^c(t) [\varphi_2(t) - \langle \delta_1, \varphi_2 \rangle \varphi_1(t)] dt. \quad (4.73)$$



Hence the approximation of  $\beta(t)$  parameter with two FPLS components is found as

$$\beta(t) = c_1\phi_1 + c_2[\varphi_2(t) - \langle \delta_1, \varphi_2 \rangle \phi_1(t)] = c_1\phi_1(t) + c_2\phi_2(t). \quad (4.74)$$

Then it is found that  $\phi_2(t) = \varphi_2(t) - \langle \delta_1, \varphi_2 \rangle \phi_1(t)$ .

For  $l = 3$ , on top of  $v_1$  and  $v_2$ , the expansion of  $v_3$  should be set into the equation

$$y = \mu_Y + \sum_{i=1}^3 c_i v_i = c_1 v_1 + c_2 v_2 + c_3 v_3. \quad (4.75)$$

By using the definition (4.63) and the equation (4.62), the expansion the third FPLS component is found as

$$\begin{aligned} v_3 &= \langle \chi_2, \varphi_3 \rangle = \int_T (\chi^c - \delta_1 v_1 - \delta_2 v_2) \varphi_3(t) dt \\ &= \int_T (\chi^c \varphi_3(t) - \langle \delta_1, \varphi_3 \rangle v_1) dt - \langle \delta_2, \varphi_3 \rangle v_2 dt \\ &= \int_T \chi^c \varphi_3(t) dt - \int_T \chi^c(t) \langle \delta_1, \varphi_3 \rangle \phi_1 - \int_T \chi^c(t) \langle \delta_2, \varphi_3 \rangle \phi_2 \\ &= \int_T \chi^c(t) [\varphi_3(t) - \langle \delta_1, \varphi_3 \rangle \phi_1 - \langle \delta_2, \varphi_3 \rangle \phi_2] dt \end{aligned} \quad (4.76)$$

Hence,  $\phi_3(t) = \varphi_3(t) - \langle \delta_1, \varphi_3 \rangle \phi_1 - \langle \delta_2, \varphi_3 \rangle \phi_2$  is found and the parameter function  $\beta(t)$  as an approximation of 3 FPLS components is given by

$$\begin{aligned} \beta(t) &= c_1\phi_1(t) + c_2\phi_2(t) + c_3[\varphi_3(t) - \langle \delta_1, \varphi_3 \rangle \phi_1 - \langle \delta_2, \varphi_3 \rangle \phi_2] \\ &= c_1\phi_1(t) + c_2\phi_2(t) + c_3\phi_3(t). \end{aligned} \quad (4.77)$$

By calculating the further steps, it can be seen that  $\phi_l(t)$  functions shows the same pattern. So the definition of  $\phi_l(t)$  can be generalized by following equations.

$$\phi_1 = \varphi_1, \quad \text{for } l = 1 \quad (4.78)$$

$$\phi_l = \varphi_l - \langle \delta_1, \varphi_l \rangle \phi_1 - \dots - \langle \delta_{l-1}, \varphi_l \rangle \phi_{l-1}, \quad \text{for } l \geq 2. \quad (4.79)$$

Hence, the regression coefficient estimator of the functional linear model can be written as in the equation (4.80) for the infinite case,

$$\beta = \sum_{l=1}^{\infty} c_l \phi_l, \quad (4.80)$$

where  $\phi_l$  satisfies the conditions given by (4.78) and (4.79).

In case of  $q$  number of FPLS components, the slope of the functional linear model can be estimated from

$$\hat{\beta}_{FPLS} = \sum_{i=1}^q \hat{c}_i \hat{\phi}_i. \quad (4.81)$$

The number of FPLS components can be determined by the same methods given for their counterpart FPC's (Oviedo de la Fuente, 2011).

The goodness of fit of the FPLSR model is determined by the coefficient of determination which is a measure of squared correlations between the scalar response  $\mathbf{Y}$  and the FPLS components:

$$R^2 = \sum_l r_{Y, v_l}^2. \quad (4.82)$$

## 5 APPLICATION

FDA assumes that the functional data set is measured on a continuity such as a dense time interval or a spectrum that consists of different frequency channels or wavelengths. FDA techniques are mostly popular in analyzing spectral data sets that are measured on a spectrum. There are many applications of FDA techniques on spectral data, especially in chemometrics (Saeys et al., 2008; Aguilera et al., 2013). The data gathered from remote sensing sensors via transmission of electromagnetic energy is also a kind of spectral data. As mentioned by Pidwirny (2006) remote sensing has many applications in the fields of land-use mapping, agriculture, forestry and oceanography (Faivre and Fischer, 1997; Caballero et al., 2014a,b; Nezlin and DiGiacomo, 2005). In oceanography, remote sensing data are used to estimate ocean characteristic parameters such as Sea Surface Temperature (SST), Chlorophyll-a content (Chl-a) and Total Suspended Solids (TSS) (Devi et al., 2015; Caballero et al., 2014a,b; Clarke et al., 2006; Schwarz et al., 2008). Recently, FDA gain importance in analyzing remote sensing sensor data sets. Cardot et al. (2003) and Besse et al. (2005) applied functional data analysis approaches to predict land use from remote sensing data obtained from the Vegetation sensor of the SPOT4 Satellite. Liu et al. (2012) offered a new rotation approach for functional factor analysis with an application on periodic remote sensing data. Gong et al. (2015) used FPCA to model high-dimensional temperature curves and temperature surfaces of Lake Victoria. Lately, Ferraty et al. (2016) study on predicting the chlorophyll content from a hyperspectrum data between wavelengths 400 - 2500 nm by using nonparametric functional models. Although, there are many applications of multivariate analysis techniques on remote sensing satellite data in oceanography (Clarke et al., 2006; Caballero et al., 2014a,b; Nezlin and DiGiacomo, 2005), there are few studies that use FDA approach (Gong et al., 2015; Ferraty et al., 2016).

In this study, it is aimed to apply functional data analysis techniques on the Remote Sensing (RS) data to estimate the amount of TSS in the coastal zone adjacent to the Guadalquivir estuary and to compare the performance of functional data analysis approach with alternative multivariate analysis techniques.

## **5.1 Material and Methods**

The data set consists of two parts: RS data and in-situ data. The satellite data is obtained from 300 m full spatial resolution (FRS) MEdium Resolution Imaging Spectrometer (MERIS) on board the Envisat multispectral platform where in-situ data is comprised of the samples collected from ocean by different campaigns.

### **5.1.1 Remotely Sensed Data**

The study area corresponds to the coastal region of the Gulf of Cadiz in the southwest coast of the Iberian Peninsula ( $35.5^{\circ} - 37.5^{\circ}$  N latitude and  $1^{\circ} - 10^{\circ}$  W longitude). Specifically, we will study the Guadalquivir estuary, one of the largest and most productive estuarine systems of the west Europe. The satellite data belong to the Region Of Interest (ROI) was downloaded from the Ocean Colour Website (<http://oceancolor.gsfc.nasa.gov>) in hdf format. It consists of Level-2 Remote Sensing Reflectance (Rrs) ( $\text{sr}^{-1}$ ) at eight different wavelengths (413 nm, 443 nm, 490 nm, 510 nm, 560 nm, 620 nm, 665 nm, 681 nm) with 300 m full spatial resolution between the years 2002-2011. A Level-2 data product is the result of the sensor calibration and atmospheric correction, consisting of derived geophysical variables generated from the corresponding radiometrically corrected Level-1A product by using the standard NASA processing methodologies. The MERIS overpass time for central Europe is between 9:30 and 11:00 UTC, with a global coverage every 3 days.

SeaDAS image analysis software (SeaWifs Data Analysis System, version 6, <http://seadas.gsfc.nasa.gov/>) and the interface VMware Workstation 12 Player (<https://www.vmware.com/>) were used to convert data from hdf format to ascii format.

The data was passed through a quality control process corresponding to the L2 flags given in the Table 5.1 to remove the suspicious and low-quality data points. This

**Table 5.1:** L2 flags

<b>L2 flags</b>	<b>Description</b>
LAND	pixel is over land
CLOUD	cloud contamination
ATMFAIL	atmospheric corection failure
HIGLINT	high sun glit
HILT	total radiance above knee
HISATZEN	large satellite zenith
CLDICE	clouds and/or ice
COCCOLITH	coccolithophores detected
HISOLZEN	large solar zenith
LOWLW	very low water-leaving radiance
CHLFAIL	chlorophyll algorithm failure
NAVWARN	questionable navigation
MAXAERITER	maximum iterations
CHLWARN	chlorophyl out of range
ATMWARN	atmospheric correction is suspect

filtering process is done by using MATLAB 7.12.0-R2011a software and the filtered data was saved in mat format for all wavelengths and all years.

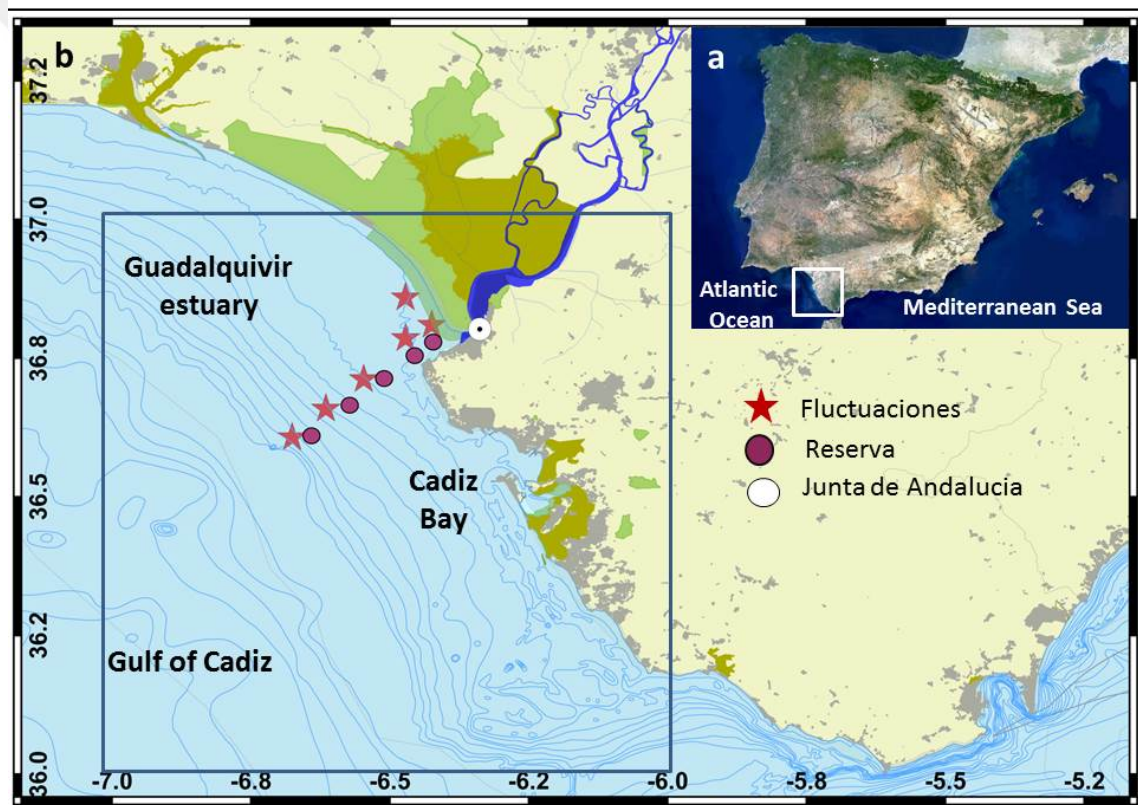
The difference between the latitudes and longitudes of the region of interest is 2° and 9°, respectively. Considering that there is 111.12 km distance between each degree, the total distance between latitudes and longitudes are approximately 222 km and 999 km. Considering that the resolution of images is 300 m it is seen that there are 740 pixels for the latitude and 3330 pixels for the longitude. So, the data set consists of 740 × 3330 pixel images. The download process gave us 2464200 pixel vectors for each wavelength. These generated data files were edited and analyzed using MATLAB7.12.0 software (m\_map toolbox), and the study area was then subset from the images to a Region Of Interest (ROI) with geographic extents of 36.01 – 37° N latitude and 7 – 6.01° W longitude.

### **5.1.2 In-situ data**

The in-situ data consist of the records of TSS values which are obtained from the samples collected by the station of Junta de Andalucia and by the cruises of Reserva and Fluctuaciones in the Guadalquivir estuary. The surface samples taken into analysis

were collected with a rosette sampler (5 m below water surface) with a distance from coast from 1km to 25 km offshore.

The samples were collected during different time periods. The samples collected by Junta de Andalucía covers the period between April 2008 - May 2011 where the samples of Reserva and Fluctuaciones were collected between the periods July 2002 - September 2004 and May 2005 - May 2007 respectively. Each sample is collected by one of the campaigns from a determined coordinate. The coordinate of the station Junta de Andalucía was fixed with the latitude  $36.78^{\circ}$  N and longitude  $6.37^{\circ}$  W where the coordinates of the stations Reserva and Fluctuaciones were chosen according to the campaign planning. The coordinates of each station can be seen in the Figure 5.1.



**Figure 5.1:** The study area and ROI. a) The study area. b) Map of Guadalquivir estuary and the Gulf of Cadiz coastal area showing the ROI. Pink stars and circles indicate the stations of Fluctuaciones and Reserva, respectively. The white round circle is the station of Junta de Andalucía.

The amount of TSS concentration in a sample is measured according to the protocols mentioned in Caballero et al. (2014b).

### 5.1.3 Validation Procedure

In the validation process, the filtered satellite data were matched up with in-situ data collected from the stations Junta de Andalucía, Reserva and Fluctuaciones considering the coordinate and the time that the sample is collected. In this step, the coordinates of in-situ and the satellite data were matched up by choosing the nearest point in the  $2 \times 2$  pixel box area. All the matched-ups were done in MATLAB 7.12.0-R2011a software.

After matching, totally 71 observations were obtained. Setting the condition that the time difference between in-situ samples and the satellite data should be at most one and a half hour, in order to reduce differences because the large spatio-temporal variability of these estuarine waters, the number of observations decrease from 71 to 31. Four observations were removed from the data set due to the fact that they were not collected from the water surface, one observation was removed due to the measurement error during filtering process and one observation was removed due to the missing Rrs values at the wavelengths 413 nm and 665 nm. Finally, the analysis were conducted on 25 observations left.

### 5.1.4 Statistical Methods

Several multivariate and functional methods were used to investigate the relationship between TSS and Rrs values and their performances are compared. It is assumed that FLRM gives better estimation than other models with two reasons: FLRM uses all the information obtained from the satellite and the spectral structure of the data allows us to take the observations as curves rather than points.

Due to the spectral structure of the data, functional linear regression models for scalar response were used in analyzing the relationship between TSS and Rrs where the scalar response vector consists of TSS values and the functional explanatory variable consists of the Rrs values recorded at eight different wavelengths.

Usually wavelength Rrs 665 is used to predict TSS concentration (Binding et al., 2003, 2005; Caballero et al., 2014b). For MERIS data Nechad et al. (2010) offered to use the bands 665 nm and 681 nm to model TSS. In study of Caballero et al. (2014b) a simple exponential regression model is found reasonable to analyse the relationship

between TSS and Rrs 665. Since, the highest correlation in our data set is found in the wavelengths 665 nm and 681 nm, two simple exponential regression models were constructed between TSS values and the Rrs values at the wavelengths 665 nm and 681 nm respectively. Following that, a stepwise exponential regression model is fitted to choose the optimum number of wavelengths that explain the response.

Alternative to FLRM and exponential regression models, Least Absolute Shrinkage and Selection Operator (LASSO) model and Generalized Additive Models (GAM) with different link functions and parameters were used to model TSS. The performance of the models were compared by using an adjusted Mean Error of Prediction (MEP) computed from Leave One Out Cross Validation (LOOCV) results.

For 2-fold cross validation MEP is defined by,

$$\text{MEP} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n}{\text{Var}(y)} = \frac{\text{SSE}}{\text{Var}(y)}, \quad (5.1)$$

where  $y$  indicates the response vector of the set that will be predicted (Febrero-Bande and Oviedo de la Fuente, 2012).

In the case of LOOCV, in every step there is just one observation. Since, the variance can not be calculated just for one observation an Adjusted MEP (AMEP) criteria is used to compare the models. AMEP is defined by the equation 5.2,

$$\text{AMEP} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2 / n}{\sum_{i=1}^n (y_i - \bar{y}_{-i})^2} = \frac{\text{SSE}}{\sum_{i=1}^n (y_i - \bar{y}_{-i})^2}, \quad (5.2)$$

where the term  $\sum_{i=1}^n (y_i - \bar{y}_{-i})^2$  is a constant that is used to scale MEP values of different models.

All the statistical analysis were done in R v3.2.3 for Windows statistical software package.

#### 5.1.4.1 Generalized Additive Models (GAM)

Generalized additive models (GAM) were first introduced by Hastie and Tibshirani (1986). GAM are nonparametric extensions of generalized linear models (GLM) which allows to model some function of the expected value of the response by a sum of



nonparametric or parametric functions of the predictors (Hothorn and Everitt, 2014). The model is defined by the equation 5.3,

$$g(\mathbb{E}[Y|X_1, X_2, \dots, X_p]) = \alpha + f(X_1) + f(X_2) + \dots + f(X_p), \quad (5.3)$$

where the response may follow one of the member distributions of exponential distribution family (Wood, 2006). The link function  $g$  can be identity, logit, probit or log link according to the distribution of the response. Identity link function is used mostly for Gaussian response where log link function is used for Poisson count data. For modelling binomial probabilities, logit or probit link functions are preferred (Friedman et al., 2001; Hastie et al., 2011).

GAM are halfway between generalized nonparametric multiple regression model and generalized linear model. In the case of multiple regressors, the multiple nonparametric regression models may face a problem called the curse of dimensionality which means that the neighbourhood of any point  $t$  may contain no observational data (Delicado, 2015). In the case of high dimensional data, GAM are alternative to nonparametric multiple regression models.

If the functions on the right side of the equation (5.3) are all linear, then generalized additive model is equivalent to generalized linear model. If some of the functions in the model have a linear effect on the response while some others do not, then it is called a semi-parametric model. Semi-parametric models are more flexible comparing to nonparametric models.

#### **5.1.4.2 Least Absolute Shrinkage and Selection Operator (LASSO) Model**

In a multiple linear regression problem if the number of predictors  $p$  are greater than the number of observations  $n$ , different shrinkage methods are offered to penalize the estimator to avoid the problems that can occur from high dimensionality. LASSO is one of these shrinkage methods which is used with the aim of variable selection. The LASSO estimate is found by minimizing penalized sum of squares criteria with norm  $L_1$  as the penalty term,

$$\sum_{i=1}^N \left( Y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j|, \quad (5.4)$$

subject to  $\sum_{j=1}^p |\beta_j| \leq t$  (Hastie et al., 2011). Here  $t$  is called the tuning parameter. For ease of interpretation, the coefficients are drawn versus a range of standardized tuning parameters  $s = t / \sum_{j=1}^p |\beta_j|$ . The optimum value of  $s$  is determined by using cross validation. The coefficients which take value for the optimum  $s$  are taken into the analysis.

## 5.2 Results

The data set consists of 25 observations collected from 3 different stations: 14 observations from "Junta de Andalucía" station, 7 observations from "Reserva" cruise stations and 4 observations from "Fluctuaciones" cruise stations.

The correlation plot in the Figure 5.2 implies that the wavelengths are highly correlated and there is an exponential relationship between in situ TSS values and Rrs values at each wavelength.

### 5.2.1 Exponential Regression Models

In order to decide which wavelength to be used in the exponential regression, the correlations between the logarithm of TSS and wavelengths were analyzed. The correlations between TSS and the wavelengths 413 nm, 443 nm, 490 nm, 510 nm, 560 nm, 620 nm, 665 nm, 681 nm were found as 0.723 ( $p < 0.001$ ), 0.671 ( $p < 0.001$ ), 0.596 ( $p=0.002$ ), 0.567 ( $p=0.003$ ), 0.533 ( $p=0.005$ ), 0.681 ( $p < 0.001$ ), 0.729 ( $p < 0.001$ ), 0.734 ( $p < 0.001$ ), respectively. Since, the highest correlations are found for the bands 665 nm and 681 nm, two exponential regression models were constructed between TSS and these wavelengths.

In situ TSS measurements ranged between 3 - 327 mg/L while Rrs values at the wavelength 681 nm ranged between 0.000 - 0.0275  $\text{sr}^{-1}$  and Rrs values at the wavelength 665 nm ranged between 0.000 - 0.028  $\text{sr}^{-1}$ . As it is seen from the scatter plot of observations in Figure 5.3, the dispersion of Rrs values for both bands are similar.

The fitted exponential regression model for the band 681 nm is,



**Figure 5.2:** The correlation plot between variables

$$\text{TSS} = 24.14 * \exp(74.38 * \text{Rrs } 681), \quad (5.5)$$

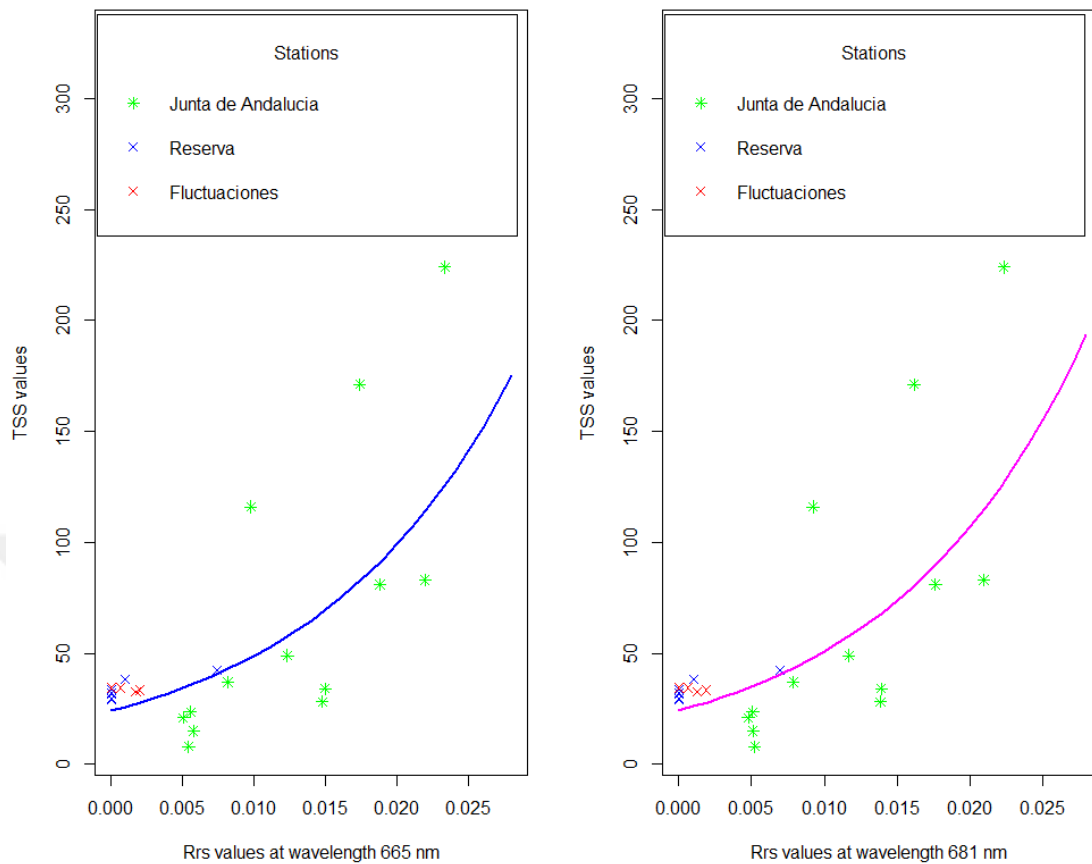
with significant parameters ( $p < 0.001$ ), 0.54 explained variance and 0.57 standard error.

The second highly correlated wavelength was Rrs 665. Exponential model between TSS and Rrs 665 satisfies the following equation.

$$\text{TSS} = 24.04 * \exp(70.83 * \text{Rrs } 665), \quad (5.6)$$

where R square is 0.53 and the standard error is 0.57.

Following these models, a stepwise linear regression model is constructed between the logarithm of TSS and the Rrs values. As a result, the covariates Rrs 413 ( $p=0.121$ ), Rrs 560 ( $p=0.003$ ) and Rrs 620 ( $p=0.003$ ) were chosen according to AIC criteria and this



**Figure 5.3:** The scatter plot between in situ TSS and Rrs values at wavelengths 665 nm and 681 nm. Blue curve indicates the relationship between TSS and Rrs 665 nm while pink curve indicates the relationship between TSS and Rrs 681 nm.

four parameter model was found significant ( $p < 0.001$ ) with R square 0.69, adjusted R square 0.65 and the standard error 0.49.

The estimated model is

$$\log \text{TSS} = 3.42 + 82.90 * \text{Rrs } 413 - 100.46 * \text{Rrs } 560 + 142.92 * \text{Rrs } 620. \quad (5.7)$$

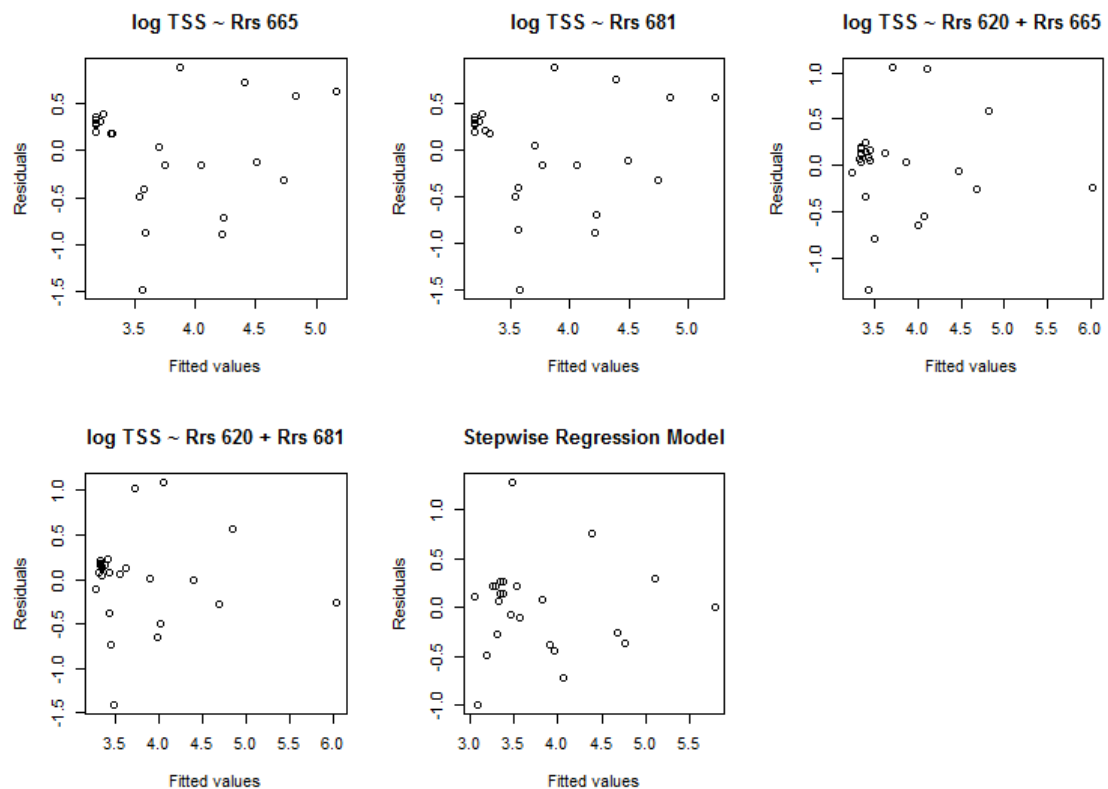
As well as these models, exponential regression models with two explanatory variables were also fitted to data. The exponential model with the parameters Rrs 620 nm and Rrs 681 nm and the exponential model with the parameters Rrs 620 nm and Rrs 665 nm were found significant. Though, the adjusted R square of these models were lower than the stepwise regression model.

The standard error, R square, Adjusted R square and AMEP values of the mentioned models are given in Table 5.2.

**Table 5.2:** MEP Values for Exponential Regression Models

Model	Std Error	Adj R <sup>2</sup>	AMEP
$\log \text{TSS} = b_0 + b_1 \text{ Rrs } 681$	0.57	0.52	<b>0.46</b>
$\log \text{TSS} = b_0 + b_1 \text{ Rrs } 665$	0.58	0.51	<b>0.48</b>
$\log \text{TSS} = b_0 + b_1 \text{ Rrs } 620 + b_2 \text{ Rrs } 681$	0.53	0.58	5.07
$\log \text{TSS} = b_0 + b_1 \text{ Rrs } 620 + b_2 \text{ Rrs } 665$	0.53	0.59	2.98
$\log \text{TSS} = b_0 + b_1 \text{ Rrs } 413 + b_2 \text{ Rrs } 560 + b_3 \text{ Rrs } 620$	0.50	0.65	1.37

According to AMEP values, the exponential regression models with one explanatory variable (Rrs 665 or Rrs 681 as the coefficient) seem reasonable. Although the standard error and the explained variance of these models are lower than the rest, they are the best models to predict TSS. First of all, they are not effected by multicollinearity. Secondly, as it can be seen from the residual plots given in Figure 5.4, other exponential models violates the assumption of the homogeneity of variance.



**Figure 5.4:** The Residual Plots of Used Exponential Models

In addition to stepwise exponential regression model, Principal Components Regression (PCR) and Partial Least Squares Regression (PLSR) methods were used to avoid correlation between the variables and to reduce dimension. The optimum number of components for both models were determined by cross validation. The AMEP values of the ordinary PC and PLS regression models including all components and Cross Validated(CV) PCR and PLSR models including just chosen components are as given in the Table 5.3.

**Table 5.3:** MEP Values for PC and PLS Regression Models

<b>Model</b>	<b>MSE</b>	<b>R<sup>2</sup></b>	<b>AMEP</b>
<b>PCR</b>	0.19	0.70	3.91
<b>PLSR</b>	0.19	0.70	3.40
<b>PCR CV with 2 components</b>	0.29	0.55	<b>0.86</b>
<b>PLSR CV with 2 components</b>	0.23	0.64	0.90

According to the table regression models with the components chosen by CV criterion gave better results than regression models that include all the components.

### 5.2.2 GAM and LASSO Models

In this study, mainly three types of GAM were used to model TSS values. Firstly, GAM with Gaussian distribution and identity link is used to model the logarithm of TSS by means of Rrs 665 and Rrs 681. This way it is aimed to obtain a more flexible model than the exponential regression models. Secondly, the dependent variable is taken as TSS and Gamma distribution with inverse link is used in modelling. Besides, alternative to one term models, two term models were also tested to find out the best GAM model that explains TSS values.

The adjusted R square, the percentage of explained deviance (Dev. Exp.) and AMEP values of the related GAM models are summarized in Table 5.4:

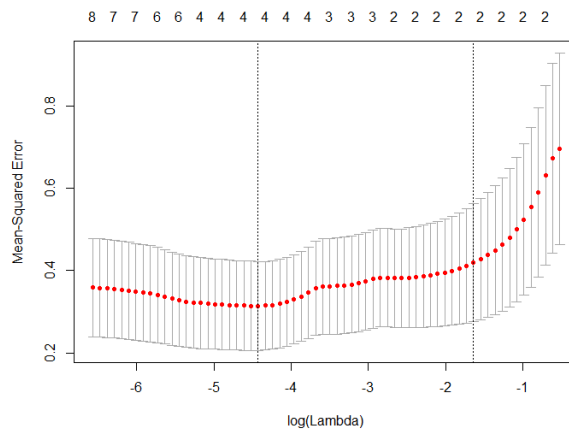
The LASSO model is constructed by taking logarithm of TSS as dependent variable. The optimum number of coefficients that should be taken into the analysis were determined by considering the optimum value of lambda that is found from 10-fold cross validation. Optimum value of lambda can be determined in two ways: The lambda value that gives the mean cross-validated error or the lambda value which

**Table 5.4:** AMEP Values for GAM

Model	Family	Link	Adj R <sup>2</sup>	Dev. Exp.	AMEP
log TSS ~ $s(\text{Rrs } 665)$	Gaussian	identity	0.89	93.2	966.01
log TSS ~ $s(\text{Rrs } 681)$	Gaussian	identity	0.87	92.0	1866.72
TSS ~ $s(\text{Rrs } 620) + s(\text{Rrs } 665)$	Gaussian	log	0.88	89.0	<b>1.53</b>
TSS ~ $s(\text{Rrs } 620) + s(\text{Rrs } 681)$	Gaussian	log	0.85	90.6	<b>1.53</b>
TSS ~ $s(\text{Rrs } 665)$	Gamma	inverse	0.73	69.0	1.63
TSS ~ $s(\text{Rrs } 681)$	Gamma	inverse	0.74	69.0	1.63

gives the maximum value such that error is within 1 standard error of the minimum is chosen as optimum lambda.

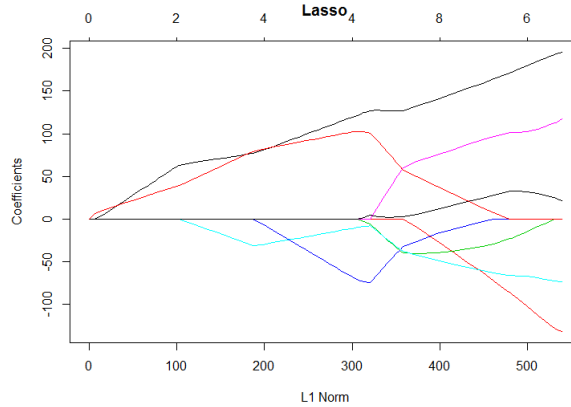
As it can be seen in the Figure 5.5, the lambda value based on the mean cross validated error is found as 1.01 and the number of variables that should be taken into the analysis is determined as 4 where the lambda value based on the standard error is found as 1.21 with the number of variables equal to 2. On the upper side of the Figure, for different values of lambda the number of variables that should be taken into the analysis are given.



**Figure 5.5:** Mean Squared Error versus a range of values of lambda. The dashed line on the left side of the plot shows the lambda value based on cross-validated error where as the dashed line on the right side of the plot shows the lambda based on standard error.

The coefficients that entered to the analysis step by step according to the value of  $L_1$  norm can be seen from Figure 5.6.

the optimum number of parameters of LASSO model were chosen as minimum 2, maximum 4 by CV criterion. The AMEP value was found equal to 33.71. This value



**Figure 5.6:** LASSO Model. The coefficients that are in the analysis for the range of  $L_1$  norm values. On x-axis the value of norm is seen. Each curve defines one coefficient.

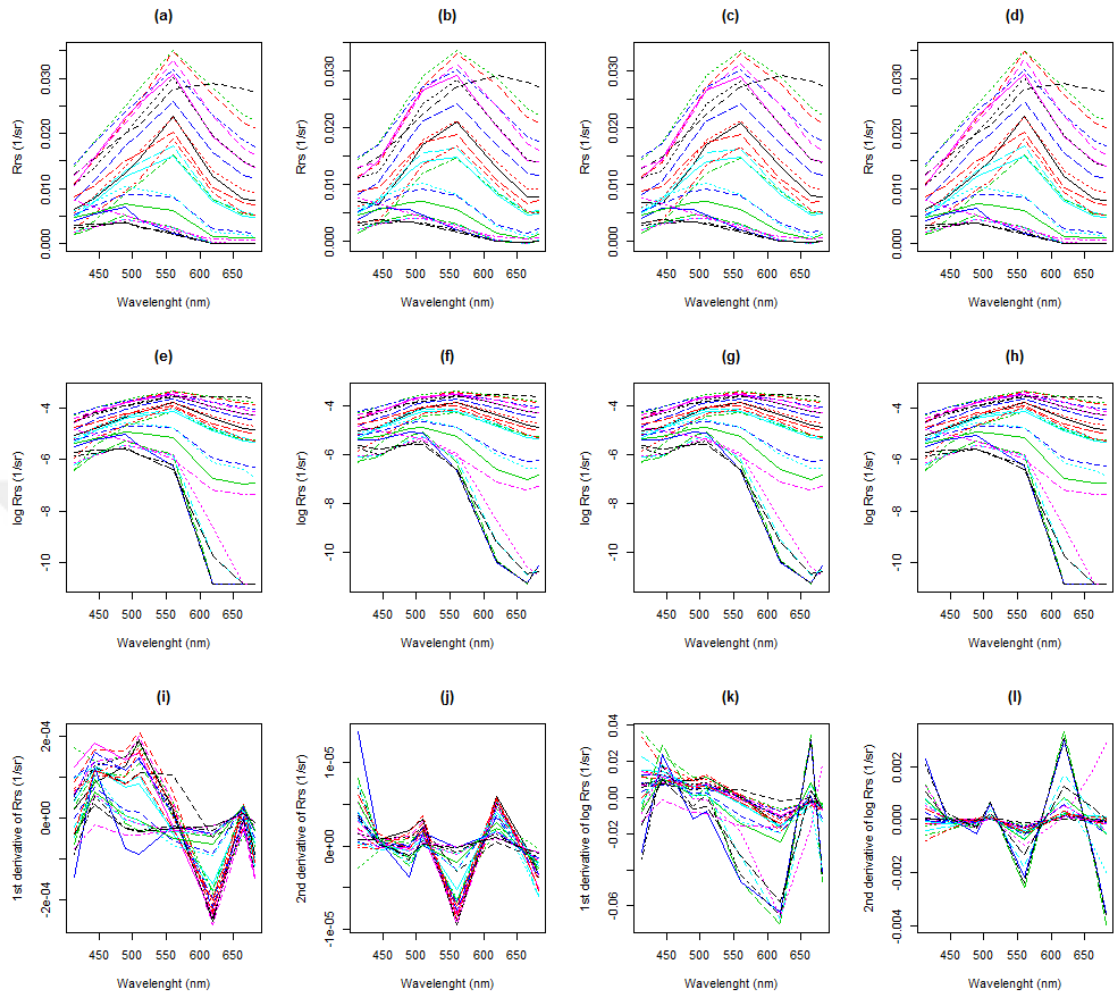
is quite higher than the AMEP values of previous models. Therefore, it has been seen that LASSO model is not a good option for modelling TSS values.

### 5.2.3 Functional Linear Regression Models

The models that are used in modelling the scalar response are functional linear regression by using B-spline basis, FPCR including all the components, FPCR with Cross Validation (FPCR CV), FPLSR including all the components, FPLSR with Cross Validation (FPLSR CV) and the FLRM with the first and second derivatives of Rrs and log Rrs curves as explanatory variables. All the models were compared in terms of AMEP values given by the equation (5.2). Initially, the scalar response was taken as raw TSS values and the functional linear models were constructed between TSS and the Rrs curves. Then the logarithm of TSS values were taken as response and they were regressed respectively on Rrs curves, on the logarithm of Rrs curves and on the first and second derivatives of them. In order to do functional modelling, first the curves were smoothed by using several methods: B-spline basis expansion with optimum number of basis chosen by CV criterion, Penalized B-spline method with optimum number of basis and optimum value of lambda chosen by CV, and Nadaraya-Watson kernel method. The graphs of these methods didn't show a big difference. Therefore, for ease of computation B-spline smoothing method with 8 number of basis is preferred in smoothing the data. For the first and the second derivatives of Rrs and log Rrs curves also B-spline smoothing with 8 number of basis is used. The smoothed curves of



functional explanatory variables obtained from different methods are shown in Figure 5.7.



**Figure 5.7:** Functional Explanatory Variables. (a) Rrs curves. (b) Rrs curves smoothed by B-spline method. (c) Rrs curves smoothed by penalized B-spline method. (d) Rrs curves smoothed by Nadaraya-Watson method. (e) log Rrs curves. (f) log Rrs curves smoothed by B-spline method. (g) log Rrs curves smoothed by penalized B-spline method. (h) log Rrs curves smoothed by Nadaraya-Watson method. (i) 1st derivative of Rrs curves. (j) 2nd derivative of Rrs curves. (k) 1st derivative of log Rrs curves. (l) 2nd derivative of log Rrs curves.

For functional outlier detection one of the procedures offered by Febrero-Bande and Oviedo de la Fuente (2012) was used which is based on weighting and bootstrap. This procedure is repeated for four types of depth measures: Fraiman-Muniz Depth (FMD), Modal Depth (MD), Random Tukey Depth (RTD) and Random Projection Depth (RPD). The number of bootstrap samples and the quantile to determine the cut off value obtained from Bootstrap sample were taken 200 and 0.5 since they are offered by Febrero-Bande and Oviedo de la Fuente (2012) as default

values. Although, there are suspicious observations in the data set, none of them were recognized as outliers. The analysis were done on 25 observations.

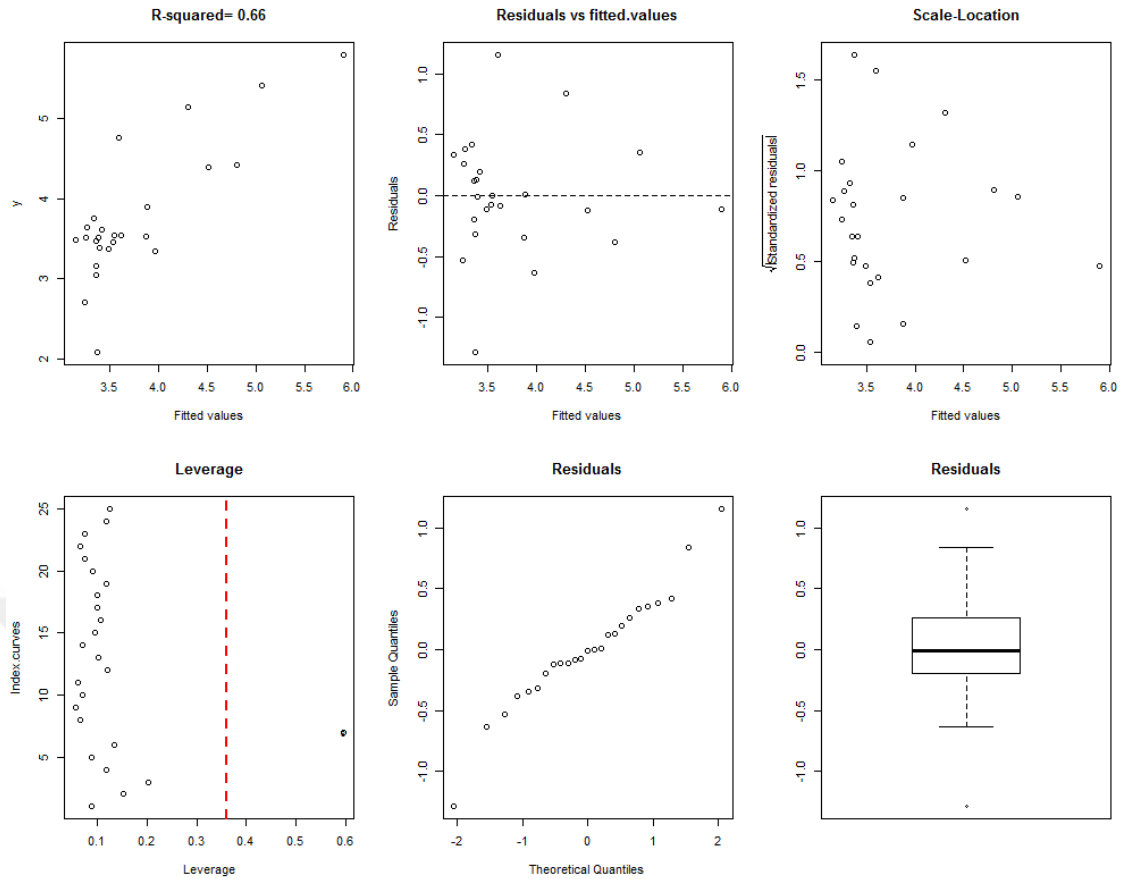
The AMEP values for the related FLRMs are as given in the Table 5.5.

**Table 5.5:** AMEP Values for FLRM

MODEL	TSS ~ Rrs	log TSS ~ Rrs	log TSS ~ log Rrs
<b>B-Spline Basis</b>	0.34	0.43	0.77
<b>FPCR</b>	0.64	0.36	0.85
<b>FPCR CV</b>	0.34	<b>0.22</b>	0.75
<b>FPLSR</b>	0.64	0.36	0.85
<b>FPLSR CV</b>	0.33	<b>0.24</b>	0.79
<b>Der1</b>	0.57	0.64	0.96
<b>Der1 Pen</b>	0.54	0.63	0.97
<b>Der2</b>	0.57	0.39	1.02
<b>Der2 Pen</b>	0.51	0.33	0.92

All the models given in Table 5.5 were found significant with  $p < 0.05$ . The highest AMEP values belong to the FLRMs between logarithm of TSS and logarithm of Rrs curve. Therefore, it is seen that predicting TSS from the logarithm of Rrs curves is not convenient. On the other hand, the lowest AMEP values are found for FLRMs between the logarithm of TSS and Rrs curves except for the methods B-spline basis, FLR on the first derivative and penalized FLR on the first derivative. These methods give better predictions when TSS is modelled on RsS curves. Generally, cross validated FPCR and FPLSR methods give lower AMEP values. The best predictions are obtained from the methods FPCR CV and FPLSR CV for the case of "log TSS ~ Rrs" with AMEP values respectively 0.22 and 0.24. Here CV criterion chooses two number of components for both methods. For the case of FPCR the fifth and the second components are chosen while for the case of FPLSR first two components are chosen.

The coefficient of determination ( $R^2$ ) of FPCR CV model is found as 0.66 with the standart deviation 0.50. Top middle graph shows that there exists heterogeneity of variance. This can arise from the less number of observations. As far as it is seen from the Leverage graph given at the the bottom left graph of Figure 5.8, the 8th curve is a possible influence curve. According to the Q-Q plot and boxplot of residuals given at the bottom of Figure 5.8, it can be said that the residuals approximately follow a skew normal distribution with two extreme values.



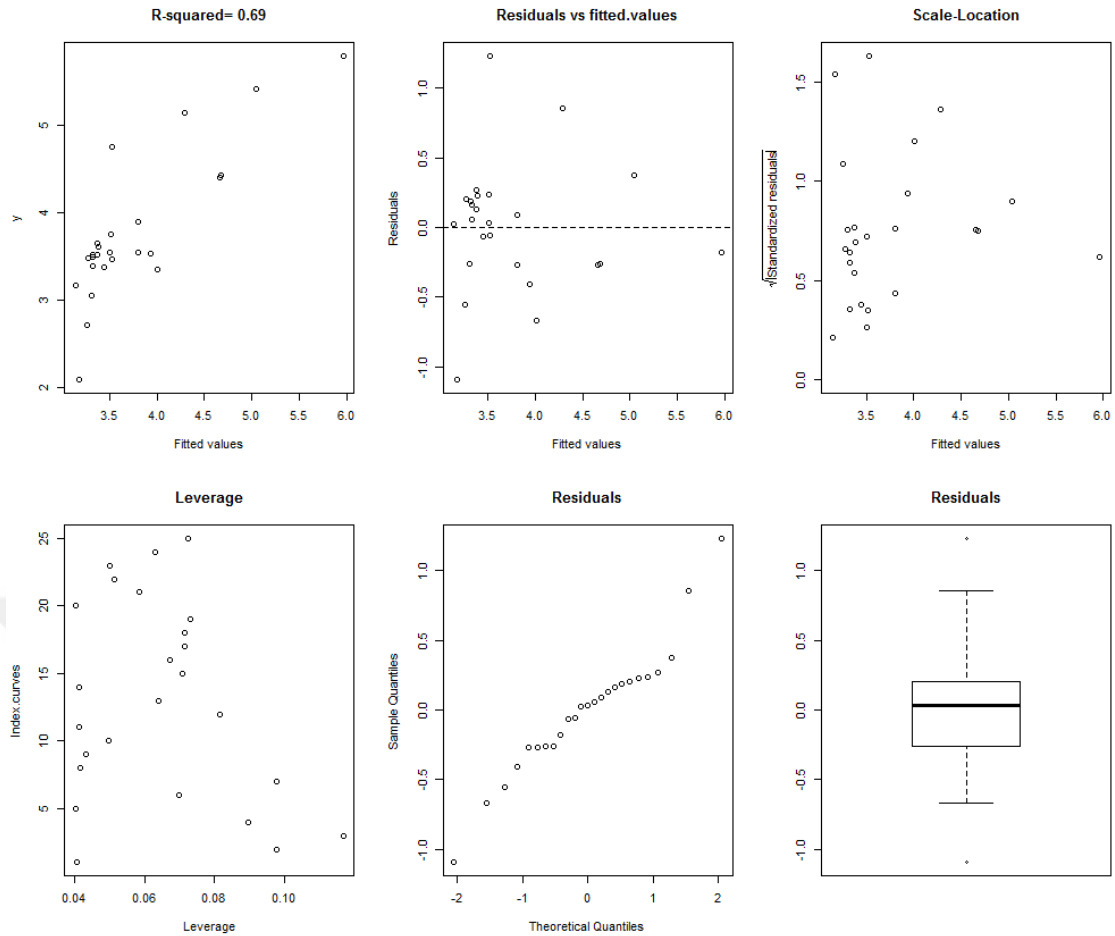
**Figure 5.8:** The summary graphs for the model FPCR CV between log TSS and Rrs

The fifth and the second PCs explain respectively 4.3 and 6.8 percentage of the variability. The parameter estimates of the PCs with the standard error, t statistic values and p values are given in the Table 5.6.

**Table 5.6:** Parameter Estimates of FPCR CV model

Parameter	Estimate	Std. Error	t value	p value
Intercept	3.76	0.10	37.4	< 0.001
FPC 5	-24.05	3.92	-6.13	< 0.001
FPC 2	-8.35	3.12	-2.68	0.013

On the other hand, for FPLSR CV model the coefficient of determination ( $R^2$ ) of is found as as 0.69 which is slightly higher than FPCR CV model with the standard deviation 0.23. As it is seen from Leverage graph given at the bottom left of Figure 5.9, there weren't observed any atypical nor influence curves. The residual plots indicate that the distribution of residuals seem similar to the case of FPCR CV model.



**Figure 5.9:** The summary graphs for the model FPLSR CV between log TSS and Rrs

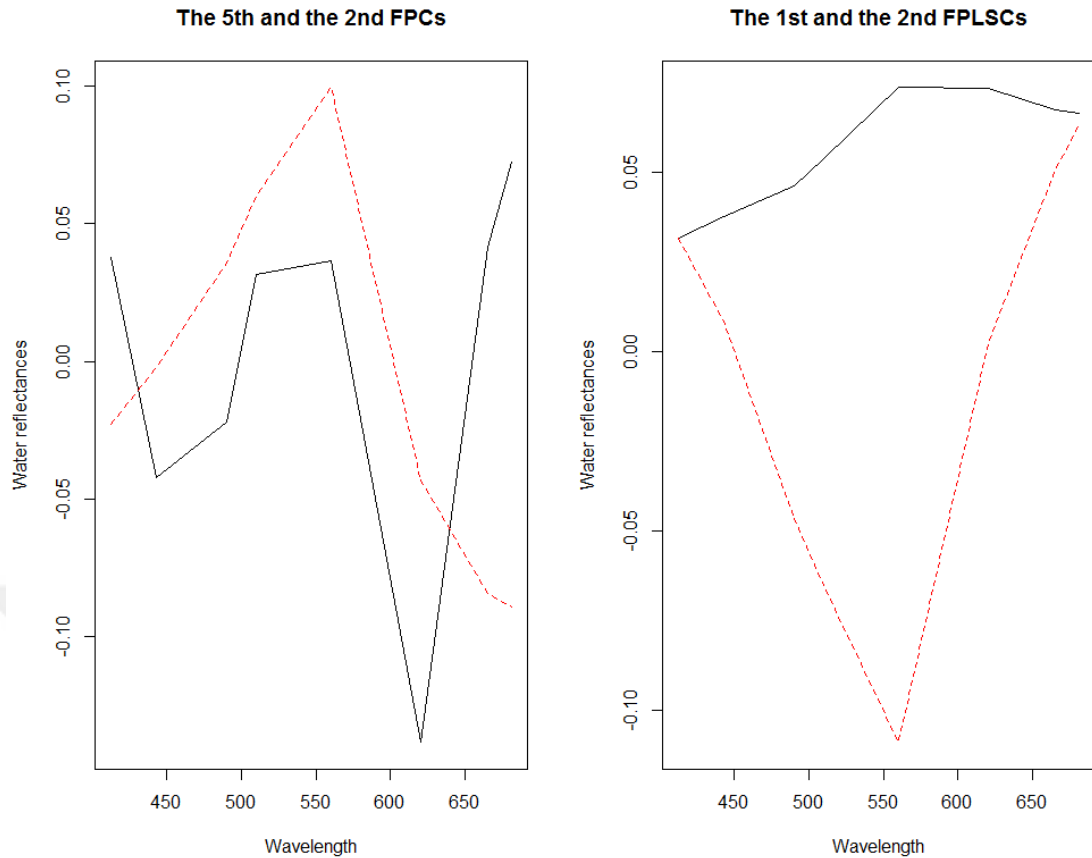
The parameter estimates and the standard errors of the first and the second component of FPLSR CV model can be seen in the Table 5.7.

**Table 5.7:** Parameter Estimates of FPLSR CV model

Parameter	Estimate	Std. Error	t value	p value
Intercept	3.76	0.09	38.89	< 0.001
FPLSC 1	0.12	0.005	23.89	< 0.001
FPLSC 2	0.20	0.001	178.77	< 0.001

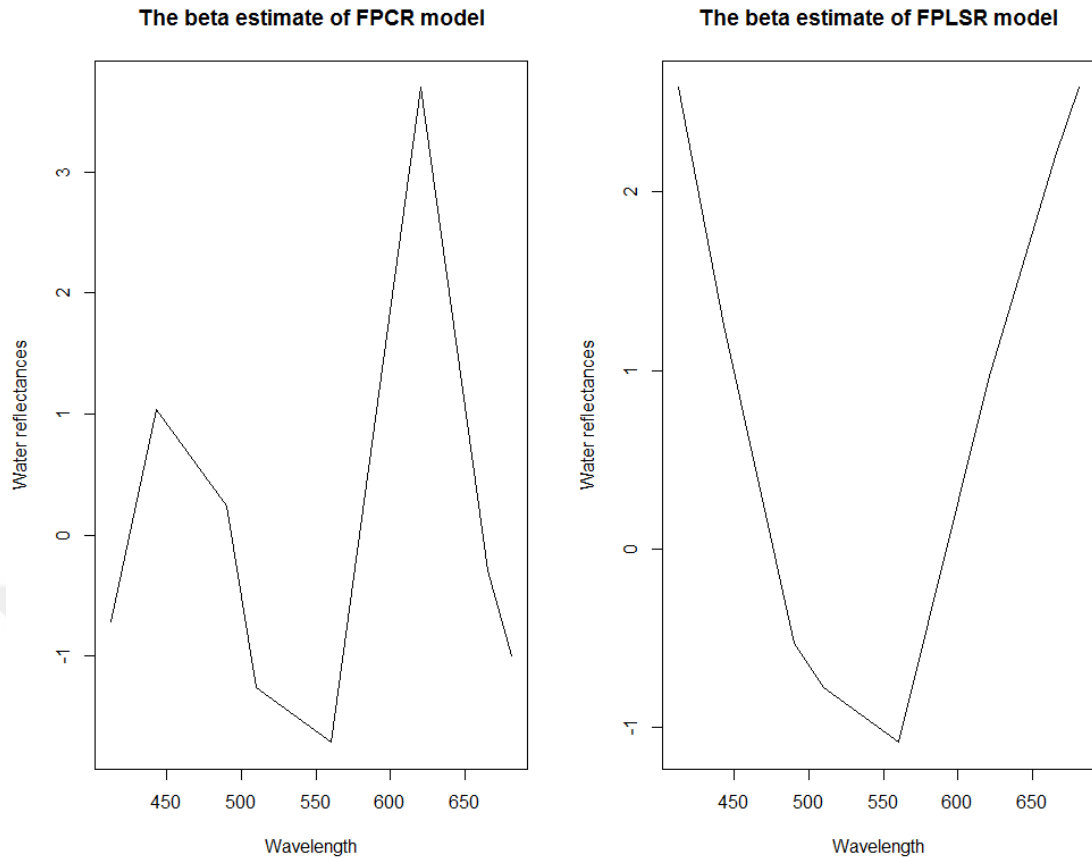
The components chosen by CV criterion both for FPCR and FPLSR models were found significant. The chosen FPCR and FPLS components can be drawn as in the Figure 5.10.

As far as seen from the left panel of Figure 5.10, the second FPC (red line) gives more weight to the band value 550 nm while it gives lower weight to the highest and lowest



**Figure 5.10:** The functional components of cross validated FPCR and FPLSR Models. band values. On the contrary, the fifth FPC (black line) is harder to interpret while it is curly. For the case of FPLSR, the first FPLS component (black line) is a more stable function that gives higher values for the higher bands while the second FPLS component gives lower weight to the middle band and higher weight to the lower and higher band values. In fact, the second FPLSC can be interpreted as the inverse of the second FPC.

The estimated parameter function of cross validated FPCR and FPLSR models are shown in the Figure 5.11. As it is seen from the left panel, the higher band values are more efficient in estimating the parameter for both cases. The main difference between the two parameter functions is that the lower band values are more important for FPLSR parameter comparing to FPCR parameter.



**Figure 5.11:** The parameter estimate of cross validated FPCR and FPLSR Models

### 5.3 Concluding Remarks

FLRM generally have lower AMEP values comparing to other statistical methods. Among classical statistical models (exponential regression model, GAM and LASSO), the simple exponential regression models with the wavelength 681 nm and the wavelength 665 nm predict TSS better than its counterparts. These are the mostly used models in the literature to predict TSS values. When the scalar response is taken as TSS, FLRM with B-Spline basis expansion, FPCR CV and FPLSR CV models are better. All the FLRM between log TSS and Rrs are better in predicting response comparing to exponential regression models and other statistical methods except the ones in which the first derivative of the Rrs curves are used as explanatory variable. Among all the models FPCR CV and FPLSR CV models gave the best predictions.

As a result, FLRM are good alternative to classical statistical models in predicting TSS value from Rrs values. Especially, in the future studies with more number of wavelengths FLRM would gain much more importance in modelling Rrs data.

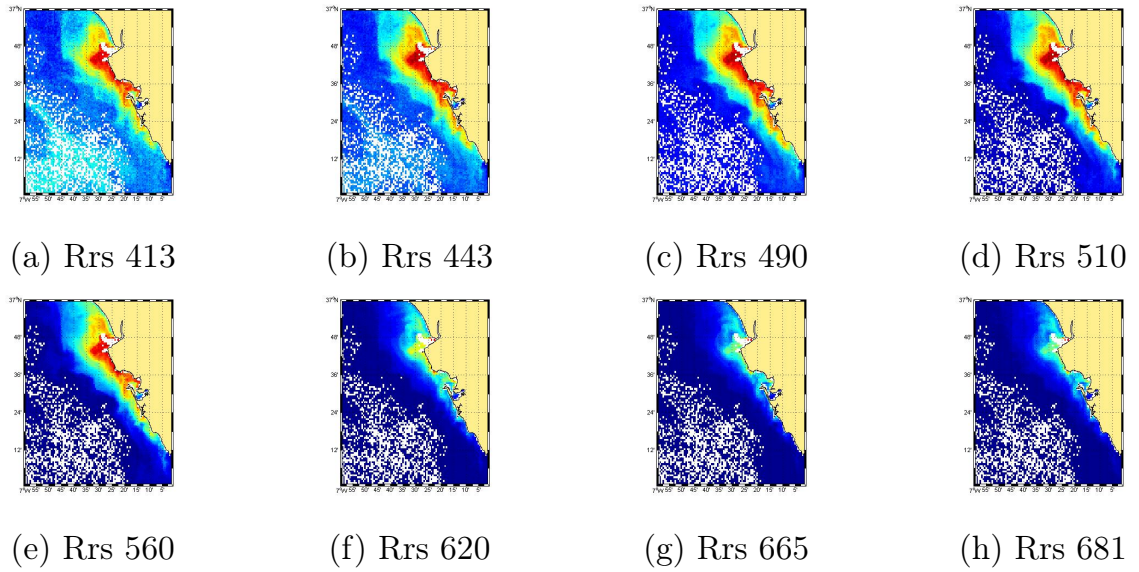
## 6 SIMULATION STUDY

In order to compare predictive performance of the proposed functional models and to support findings, a simulation study is designed. The simulation study has four main steps. In the first step, day images obtained from satellite were taken as functional predictor in order to generate a scalar response vector. In the second step, the response vector  $\tilde{Y}$  was generated based on the satellite image that was chosen in the previous step. Four different models and three different standard deviations were used in producing the response. As the third step, the generated responses were predicted based on six different models. In this sense, 2-fold CV method based on three different sample sizes were used and the models used in prediction were compared in terms of their MEP values. These steps will be detailed in the following sections.

### 6.1 The Choice of Functional Predictor

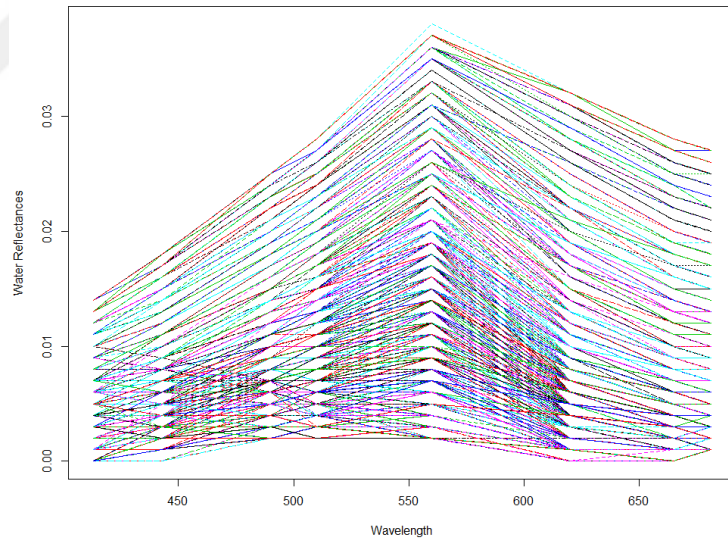
The day images recorded at eight different bands 413 nm, 443 nm, 490 nm, 510 nm, 560 nm, 620 nm, 665 nm, 681 nm were used to create the functional predictor. The images show the area bounded by  $36.01 - 37^\circ$  N latitude and  $7 - 6.01^\circ$  W longitude. Therefore, each image consist of 136900 pixels. Each pixel corresponds to the information of water reflectances. In order to use as much information as possible, the pixels belong to the land were removed and only the pixels that have value greater than 0 were considered in choosing functional predictor. As the wavelength increases, the quantity of information in pixels decreases (see Figure 6.1). Therefore, the choice of the pixels that compose the functional predictor was done based on the highest wavelength 681 nm.

The image that contains more information was determined as one of the days belonging to the year 2009 with 9843 pixels. This image was used as functional predictor  $\tilde{\chi}(t)$ . Then, the RRS values at eight wavelengths for that day were merged and the functional data object is created. Totally 66 observations were removed due to the missing values



**Figure 6.1:** Images obtained from satellite data for different wavelengths.

at different wavelengths. Finally, 9777 functional observations were left ( $N=9777$ ). The curves for each pixel were drawn in Figure 6.2.



**Figure 6.2:** Functional observations used for simulation.

## 6.2 Generating the Response

The scalar response  $\tilde{Y}$  was generated based on the parameters of the models between the logarithm of TSS and Rrs values, since the best predictions in real life data were obtained from those models. Four models ( $M = 4$ ) were used in generating the logarithm of TSS values: simple exponential regression with 665 nm wavelength,



FLRM with B-spline Basis approach, FPCR with all components, FPCR with the components chosen by CV criterion. FPLSR could not be used since this method requires that the covariance between the response and the predictor is known. Here, the covariance can not be computed because the response is unknown due to the fact that it is generated based on the chosen predictor.

Initially the response vector ( $\tilde{Y}_{exp}$ ) was generated based on the classical simple exponential regression model given in the equation (5.6). As explanatory variable Rrs values recorded at the band 665 nm for the chosen day was used.

Then, FLRMs which can be expressed in the general form,

$$\tilde{Y}_i = \int \tilde{\chi}_i^*(t)\beta(t)dt + \varepsilon, \quad (6.1)$$

were used to simulate the response. Here,  $\tilde{\chi}_i^*(t)$  denotes the functional predictor composed of Rrs values at 8 different wavelengths related to the day chosen in the previous step and  $\beta(t)$  is the functional parameter estimate that was taken respectively from the models FLRM B-spline basis expansion, FPCR and FPCR with CV method used in the application. The responses of functional linear models will be denoted respectively by  $\tilde{Y}_{basis}$ ,  $\tilde{Y}_{FPC}$  and  $\tilde{Y}_{FPCV}$ .

The response ( $\tilde{Y}_{exp}$ ) for FLRM with B-spline basis approach is generated from the equation (6.2),

$$\tilde{Y}_{basis} = \int \tilde{\chi}^*(t)\beta(t)dt = \mathbf{c}_i^* \mathbf{J}_{\phi\theta} \mathbf{b}^*, \quad (6.2)$$

where  $\mathbf{J}_{\phi\theta}$  is the matrix computed from the inner product of basis functions  $\boldsymbol{\phi}(t) = [\phi_1(t), \dots, \phi_8(t)]'$  and  $\boldsymbol{\theta}(t) = [\theta_1(t), \dots, \theta_5(t)]'$  that are used to extend  $\tilde{\chi}_i^*(t) = \sum_{k=1}^8 c_{ik}^* \phi_k(t)$  and  $\beta(t) = \sum_{l=1}^5 b_l^* \theta_l(t)$ , respectively. Here,  $\mathbf{c}_i^* = [c_{i1}^*, \dots, c_{i8}^*]'$  and  $\mathbf{b}^* = [b_1^*, \dots, b_5^*]'$  indicate the coefficient vectors that are used in the extension of the functional predictor and the parameter function.  $\mathbf{b}^*$  and  $\mathbf{J}_{\phi\theta}$  parameters in the model (6.2) were taken directly from the related FLRM in the application where  $\mathbf{c}_i^*$  is obtained from the smoothing of the chosen functional predictor  $\tilde{\chi}_i^*(t)$  on 8 number of B-spline basis functions.

In the case of FPCR, the model given in the equation (6.1) is induced to,

$$\tilde{Y}_{FPC} = \mu_Y + \int \tilde{\chi}^*(t)\beta(t)dt = \mu_Y + \mathbf{F}_{pc}^* \mathbf{b}^* \quad (6.3)$$

where  $\mu_Y$  is the mean of the real response vector,  $\mathbf{F}_{pc}^*$  denotes the score matrix which is computed from  $\mathbf{F}_{pc}^* = \langle \chi, \hat{\xi} \rangle$  and  $\mathbf{b}^*$  is the parameter vector obtained from the extension  $\beta_{FPC}(t) = \sum_{j=1}^J b_j^* \hat{\xi}_j$  computed in the application.  $J$  indicates the number of components that are used in the extension of the data which is equal to 8 for FPCR with all the components and is equal to 2 for FPCR with CV method.

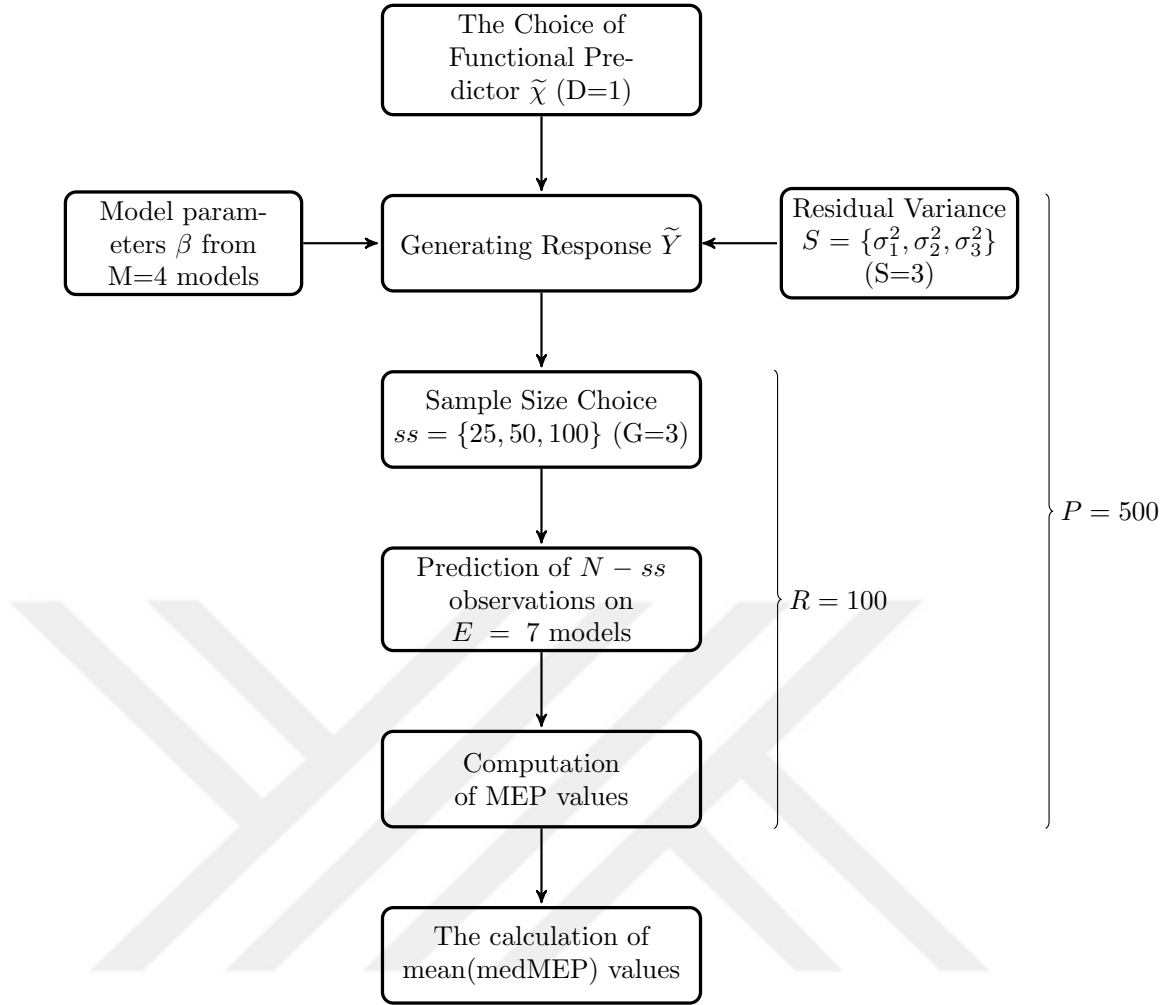
The error term  $\varepsilon$  is a normally distributed vector with zero mean and the variance equal to the raw residual variance of the model from which the functional parameter was taken. The raw residual variances of the models in the application were quite high due to the less number of observations. So that three different values ( $S = 3$ ) for the model variance were used in the simulation: the raw residual variances of the models in the application ( $\sigma_1^2 = \hat{\sigma}^2$ ), one fifth of the model residual variance ( $\sigma_2^2 = \hat{\sigma}^2/5$ ), one tenth of the model residual variance ( $\sigma_3^2 = \hat{\sigma}^2/10$ ). For each value of the residual variance and each type of model, the response was generated five hundred ( $P = 500$ ) times.

### 6.3 The Comparison of the Prediction Models

Seven ( $E=7$ ) different models were used in predicting the response: exponential regression with band 665 nm, FLRM with B-spline basis expansion, FPCR, FPCR CV, FPLSR, FPLSR CV models and FLRM with second derivative as explanatory variable (FLRM Der2). In order to measure the predictive performance of the models their MEP values were computed from 2-fold cross validation based on the equation (5.1).

The data was splitted into two considering three ( $G = 3$ ) different sample sizes  $ss = \{25, 50, 100\}$ . The maximum value of the sample size was taken as 100, while in real life data the predictions are always done on limited number of observations.  $ss$  out of 9777 pixels were randomly chosen from the image and the response was predicted on  $ss$  observations for rest of the ( $N - ss$ ) pixels.

This procedure was repeated  $R = 100$  times in order to use as much different pixels as possible for each sample size. The MEP values for each response and each prediction



**Figure 6.3:** The Simulation Design

model were computed. Then, the median of MEP values (medMEP) were taken instead of mean due to the skewness of the distribution.

Finally, the mean of medMEP values were taken over  $P = 500$  simulations. The models were compared in terms of their mean(medMEP) values.

The simulation study can be summarized as in the Figure 6.3. The simulation design has  $M \times S \times G$ , i.e. 36, possible scenarios. For each scenario, the results obtained from  $E = 7$  different models are compared.

#### 6.4 Simulation Results

The results of simulation study regarding to the sample sizes were given respectively in Table 6.1, Table 6.2 and Table 6.3.

**Table 6.1:** Simulation Results for Sample Size 25

Sample size 25 (ss=25)								
Type of Response	Residual Variance	Exp. Reg.	FLRM Basis	FPCR	FPCR CV	FPLSR	FPLSR CV	FLRM Der2
$\tilde{Y}_{exp}$	$\sigma_1^2$	<b>0.99</b>	1.26	1.57	1.25	1.57	<b>0.98</b>	1.04
	$\sigma_2^2$	<b>0.14</b>	0.25	0.36	0.29	0.36	0.31	0.67
	$\sigma_3^2$	<b>0.04</b>	0.07	0.09	0.09	0.09	0.09	0.65
$\tilde{Y}_{bas}$	$\sigma_1^2$	1.10	1.11	1.32	1.15	1.32	<b>1.09</b>	1.10
	$\sigma_2^2$	1.19	<b>0.13</b>	0.20	0.17	0.20	0.16	1.03
	$\sigma_3^2$	1.20	<b>0.03</b>	0.05	0.05	0.05	0.06	1.02
$\tilde{Y}_{FPC}$	$\sigma_1^2$	1.06	1.22	1.27	1.16	1.27	1.13	<b>1.01</b>
	$\sigma_2^2$	1.08	0.76	<b>0.17</b>	0.21	<b>0.17</b>	0.29	0.85
	$\sigma_3^2$	1.07	0.75	<b>0.04</b>	0.05	<b>0.04</b>	0.05	0.83
$\tilde{Y}_{FPCcv}$	$\sigma_1^2$	0.88	0.78	0.91	0.82	0.91	<b>0.74</b>	0.78
	$\sigma_2^2$	0.52	0.08	0.09	0.08	0.09	<b>0.07</b>	0.15
	$\sigma_3^2$	0.49	0.03	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	0.11

**Table 6.2:** Simulation Results for Sample Size 50

Sample Size 50 (ss=50)								
Type of Response	Residual Variance	Exp. Reg.	FLRM Basis	FPCR	FPCR CV	FPLSR	FPLSR CV	FLRM Der2
$\tilde{Y}_{exp}$	$\sigma_1^2$	0.90	1.00	1.08	0.98	1.08	<b>0.89</b>	0.97
	$\sigma_2^2$	<b>0.11</b>	0.14	0.16	0.15	0.16	0.20	0.60
	$\sigma_3^2$	<b>0.03</b>	0.04	0.04	0.04	0.04	0.04	0.57
$\tilde{Y}_{bas}$	$\sigma_1^2$	1.02	<b>0.84</b>	0.91	0.89	0.91	0.86	1.04
	$\sigma_2^2$	0.94	<b>0.08</b>	0.09	0.09	0.09	0.10	0.95
	$\sigma_3^2$	0.94	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	0.03	0.94
$\tilde{Y}_{FPC}$	$\sigma_1^2$	0.97	0.90	<b>0.85</b>	0.88	<b>0.85</b>	0.88	0.95
	$\sigma_2^2$	0.83	0.40	<b>0.08</b>	<b>0.08</b>	<b>0.08</b>	0.09	0.76
	$\sigma_3^2$	0.82	0.36	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	0.75
$\tilde{Y}_{FPCcv}$	$\sigma_1^2$	0.89	<b>0.71</b>	0.75	0.73	0.75	<b>0.71</b>	0.73
	$\sigma_2^2$	0.54	<b>0.07</b>	<b>0.07</b>	<b>0.07</b>	<b>0.07</b>	<b>0.07</b>	0.14
	$\sigma_3^2$	0.52	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	0.09

**Table 6.3:** Simulation Results for Sample Size 100

Sample Size 100 (ss=100)								
Type of Response	Residual Variance	Exp. Reg.	FLRM Basis	FPCR	FPCR CV	FPLSR	FPLSR CV	FLRM Der2
$\tilde{Y}_{exp}$	$\sigma_1^2$	<b>0.84</b>	0.88	0.91	0.88	0.91	0.86	0.95
	$\sigma_2^2$	<b>0.10</b>	0.11	0.12	0.12	0.12	0.12	0.56
	$\sigma_3^2$	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	<b>0.03</b>	0.53
$\tilde{Y}_{bas}$	$\sigma_1^2$	0.96	<b>0.72</b>	0.75	0.75	0.75	0.74	1.01
	$\sigma_2^2$	0.77	<b>0.07</b>	<b>0.07</b>	<b>0.07</b>	<b>0.07</b>	<b>0.07</b>	0.91
	$\sigma_3^2$	0.76	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	0.91
$\tilde{Y}_{FPC}$	$\sigma_1^2$	0.91	0.75	<b>0.69</b>	0.72	<b>0.69</b>	0.72	0.92
	$\sigma_2^2$	0.69	0.27	<b>0.06</b>	<b>0.06</b>	<b>0.06</b>	<b>0.06</b>	0.72
	$\sigma_3^2$	0.67	0.24	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	<b>0.01</b>	0.71
$\tilde{Y}_{FPCcv}$	$\sigma_1^2$	0.90	<b>0.69</b>	0.70	<b>0.69</b>	0.70	<b>0.69</b>	0.71
	$\sigma_2^2$	0.59	0.07	<b>0.06</b>	<b>0.06</b>	<b>0.06</b>	<b>0.06</b>	0.13
	$\sigma_3^2$	0.57	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	<b>0.02</b>	0.09

As seen from Table 6.1, for the sample size 25, as the error variance decreases the mean medMEP values of the models except the exponential regression model and FLRM Der2 decrease for the response types  $\tilde{Y}_{basis}$  and  $\tilde{Y}_{FPC}$ . For the case of high error variance, the response types  $\tilde{Y}_{exp}$  ve  $\tilde{Y}_{basis}$  have been predicted better by the models FPLSR CV and FLRM Der2 where the response  $\tilde{Y}_{FPCcv}$  has been predicted better by FPLSR CV model. As the residual variance decreases, usually the best predicitions are obtained from the models from which the response was generated from. As an exception, FPLSR CV model predicted the response  $\tilde{Y}_{FPCcv}$  better than FPCR CV model. FPCR CV and FPLSR CV models that include all the components give always the same results.

When the sample size is 50 and the error variance is high ( $\sigma_1^2$ ),  $\tilde{Y}_{exp}$  is predicted better by FPLSR CV method and the response  $\tilde{Y}_{FPCcv}$  is predicted better by FLRM Basis and FPLSR CV methods. Except these, the best predictions are obtained from the models that the response comes from. As the sample size increases and the error variance decreases it is seen that the mean medMEP values decrease. However, this decrease is higher in FLRMs with Rrs curves as explanatory variables compared to other models.

As it is seen in Table 6.3, for lower values of variance the response all the models except FLRM Der2 gave better predictions for the response  $\tilde{Y}_{exp}$ . Among the responses generated from functional models, for the high value of variance the response  $\tilde{Y}_{basis}$  has been predicted better by FLRM Basis method but when the error variance decreases the predictive performance of all functional regression models gets closer. The best predictions for the responses  $\tilde{Y}_{FPC}$  and  $\tilde{Y}_{FPCcv}$  have been obtained from FLRMs based on dimension reduction methods from which they are generated from.

Generally as the sample size increases and the residual variance decreases, the medMEP values of the models decrease for all type of responses. Particularly, the mean of medMEP values of functional linear models get closer. Among functional linear models, FLRM with the second derivative as explanatory variable (FLRM Der2) make difference with regards to high MEP values.

When the response comes from exponential regression model, there is not found a big difference between MEP values of exponential regression model and FLRMs. But when the response comes from a functional model, the difference between MEP values of exponential regression model and FLRMs increase, generally FLRM predict better than exponential regression model.

Considering all types of response, generally FLRMs make better predictions than exponential regression model that is mostly used in the literature. Particularly, FLRM based on dimension reduction methods (FPCR, FPLSR, FPCR CV and FPLSR CV) and FLRM basis make better predictions with lower values of MEP.

As a result, the simulation study support the findings obtained from the application. FLRM can be used alternative to classical statistical models for remote sensing data applications even though when there the data is recorded at limited number of wavelengths.

## 7 CONCLUSION

In this study, FDA approach has been proposed to make predictions from remote sensing data in oceanography as an alternative to classical statistical methods. In this sense various statistical models have been applied to predict the amount of TSS in the coastal region of Guadalquivir estuary in Cadiz from satellite data. While in oceanography mostly classical exponential regression models have been used to model TSS, in this study many different models have been used and their results have been compared. For FLRMs different responses and explanatory variables were taken. Among these, FLRMs between the logarithm of TSS and Rrs values that are based on dimension reduction methods gave better predictions compared to GAM, LASSO and classical exponential regression models.

The importance of this study is that FLRMs have been used for the first time to predict the amount of TSS from satellite data in oceanography. In previous studies that use the remote sensing data obtained from MERIS, TSS values were mostly predicted from regression models with single band. Binding et al. (2003) have predicted the amount of minerals from the quadratic regression model that use the reflectance values at the band 665 nm. In the study of Nechad et al. (2010) exponential regression model with band 681 nm have been used to model TSS while in the study of Caballero et al. (2016) the highest amount of correlation were found in band 665 nm. There are studies that use the RS data obtained from other satellite sensors such as MODIS or DEIMOS (Caballero et al., 2014a,b). However due to its high amount of resolution MERIS is preferable. Nechad et al. (2010) has mentioned that it would be better to use multiple bands rather than single band to model particularly low values of TSS. In this study, as well as FLRMs many other single and multiple band models that had never been used before have been applied and compared. The most suitable models have been found to be FLRMs.

The limitation of this study is the low number of observations and wavelengths. The number of observations were increased in the designed simulation study. Unfortunately, because of the sensors that were used in that period, it is not possible to increase the number of wavelengths. Recently, the sensors that record data in more number of wavelengths are in use. However, there haven't been enough number of collected samples that match with that period. In this branch, Ferraty et al. (2016) study on a new project in which some parameters related to the environment are predicted from hyperspectral remote sensing data by using nonparametric FLRMs. Naturally, the increase in the number of wavelengths increases the size of the data set.

The results obtained in the application have been supported by a simulation study. According to the results of the simulation as the sample size increases and the error variance decreases, the predictive performance of FLRMs generally gets better than exponential regression models. In analogy to the application, FLRMs based on dimension reduction methods have given better predictions than other functional models. Due to the limited time, the simulation study has been designed with restricted number of parameters. In the future, the simulation study can be improved by changing some parameters. For instance, the sample size used for prediction can be taken as 200 rather than 100, the pixels that compose the functional variable can be chosen based on a determined route instead of using randomly chosen pixels, the images recorded at different days can be considered in order to use more number of images while creating the functional variable.

In matching process in the application the nearest pixels in the  $2 \times 2$  pixel area were taken in order to have more number of observations. It can be analysed how the results are effected when the exact pixels of the data are used. Considering the 3.5 TB size of the remote sensing data set, it can be said that this study also includes a big data problem. It is foreseen that in the future FLRMs will gain more importance and will provide more convenience in analysis compared to other statistical models.

It is thought that this study is the first most detailed study in our country that has been done in this area and it would be an usefull resource for further researchs on this area.



## REFERENCES

- Aguilera, A. M., Escabias, M., Preda, C., and Saporta, G. (2010). Using basis expansions for estimating functional pls regression: Applications with chemometric data. *Chemometrics and Intelligent Laboratory Systems*, 104(2):289–305.
- Aguilera, A. M., Escabias, M., Valderrama, M. J., and Aguilera-Morillo, M. C. (2013). Functional analysis of chemometric data. *Open Journal of Statistics*, 3(5):333–343.
- Aguilera, Ana M ve Escabias, M. v. V. M. J. (2008). Discussion of different logistic models with functional data. application to systemic lupus erythematosus. *Computational Statistics & Data Analysis*, 53(1):151–163.
- Aguilera Morillo, M. d. C. (2013). *Penalized Estimation Methods in Functional Data Analysis*. PhD thesis, University of Granada.
- Almansa, J. and Delicado, P. (2009). Analysing musical performance through functional data analysis: rhythmic structure in schumann’s träumerei. *Connection Science*, 21(2-3):207–225.
- Besse, P. and Ramsay, J. O. (1986). Principal components analysis of sampled functions. *Psychometrika*, 51(2):285–311.
- Besse, P. C., Cardot, H., Faivre, R., and Goulard, M. (2005). Statistical modelling of functional data. *Applied Stochastic Models in Business and Industry*, 21(2):165–173.
- Binding, C., Bowers, D., and Mitchelson-Jacob, E. (2003). An algorithm for the retrieval of suspended sediment concentrations in the irish sea from seawifs ocean colour satellite imagery. *International Journal of Remote Sensing*, 24(19):3791–3806.

- Binding, C., Bowers, D., and Mitchelson-Jacob, E. (2005). Estimating suspended sediment concentrations from ocean colour measurements in moderately turbid waters; the impact of variable particle scattering properties. *Remote sensing of Environment*, 94(3):373–383.
- Caballero, I., Morris, E. P., Ruiz, J., and Navarro, G. (2014a). Assessment of suspended solids in the Guadalquivir estuary using new Deimos-1 medium spatial resolution imagery. *Remote Sensing of Environment*, 146:148–158.
- Caballero, I., Morris, E. P., Ruiz, J., and Navarro, G. (2014b). The influence of the Guadalquivir river on the spatio-temporal variability of suspended solids and chlorophyll in the eastern Gulf of Cadiz. *Mediterranean Marine Science*, 15(4):721–738.
- Caballero, I., Ruiz, J., and Navarro, G. (2016). Dynamics of the turbidity plume in the Guadalquivir estuary (SW Spain): A remote sensing approach. ESA Living Planet Symposium, Prague, Czech Republic.
- Cardot, H., Faivre, R., and Goulard, M. (2003). Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. *Journal of Applied Statistics*, 30(10):1185–1199.
- Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters*, 45(1):11–22.
- Clarke, E., Speirs, D., Heath, M., Wood, S., Gurney, W., and Holmes, S. (2006). Calibrating remotely sensed chlorophyll-a data by using penalized regression splines. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(3):331–353.
- Cuevas, A. (2014). A partial overview of the theory of statistics with functional data. *Journal of Statistical Planning and Inference*, 147:1–23.
- Cuevas, A., Febrero, M., and Fraiman, R. (2004). An ANOVA test for functional data. *Computational Statistics & Data Analysis*, 47(1):111–122.

- Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *Journal of multivariate analysis*, 12(1):136–154.
- Davidian, M., Lin, X., and Wang, J. (2004). Introduction: Emerging issues in longitudinal and functional data analysis. *Statistica Sinica*, 14(3):613–614.
- Delaigle, A., Hall, P., et al. (2012). Methodology and theory for partial least squares applied to functional data. *The Annals of Statistics*, 40(1):322–352.
- Delicado, P. (2015). Advanced statistical modelling. Lecture Notes.
- Devi, G. K., Ganasri, B., and Dwarakish, G. (2015). Applications of remote sensing in satellite oceanography: A review. *Aquatic Procedia*, 4:579–584.
- Dieudonné, J. (1960). *Foundations of modern analysis*, volume 286. Academic press New York.
- Escabias, M., Aguilera, A. M., and Valderrama, M. J. (2004). Principal component estimation of functional logistic regression: discussion of two different approaches. *Journal of Nonparametric Statistics*, 16(3-4):365–384.
- Escabias, M., Aguilera, A. M., and Valderrama, M. J. (2005). Modeling environmental data by functional principal component logistic regression. *Environmetrics*, 16(1):95–107.
- Escabias, M., Valderrama, M. J., and Aguilera-Morillo, M. C. (2012). Functional data analysis in biometrics and biostatistics. *J Biom Biostat*, 3(8):e120.
- Faivre, R. and Fischer, A. (1997). Predicting crop reflectances using satellite data observing mixed pixels. *Journal of Agricultural, Biological, and Environmental Statistics*, 2(1):87–107.
- Febrero, M., Galeano, P., and Manteiga-Gonzales, W. (2007). A functional analysis of nox levels: location and scale estimation and outlier detection. *Computational Statistics*, 22:411–427.
- Febrero, M., Galeano, P., and Manteiga-Gonzales, W. (2008). Outlier detection in functional data by depth measures, with application to identify abnormal nox levels. *Environmetrics*, 19:331–345.

- Febrero-Bande, M., Galeano, P., and González-Manteiga, W. (2015). Functional principal component regression and functional partial least-squares regression: An overview and a comparative study. *International Statistical Review*, 0(0):1–23.
- Febrero-Bande, M. and Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: the r package fda.usc. *Journal of Statistical Software*, 51(4):1–28.
- Ferraty, F. (2010). High-dimensional data: a fascinating statistical challenge. *Journal of Multivariate Analysis*, 101(2):305–306.
- Ferraty, F., Fauvel, M., and Zullo, A. (2016). Nonparametric prediction based on contaminated functional predictor and application to hyperspectral data. 2016 CRoNoS Summer Course on Functional Data Analysis, Oviedo, Spain.
- Ferraty, F. and Vieu, P. (2003). Curves discrimination: a nonparametric functional approach. *Computational Statistics & Data Analysis*, 44(1):161–173.
- Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media.
- Fraiman, R. and Muniz, G. (2001). Trimmed means for functional data. *Test*, 10(2):419–440.
- Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin.
- Gündüz, M. (2012). *The Survey of The Relationship Between Closing Price of the Companies Trading on Istanbul Stock Exchange and Their Trading Volumes Through Functional Canonical Correlation Analysis*. PhD thesis, Atatürk Üniversitesi.
- Gong, M., Miller, C., and Scott, E. (2015). Functional pca for remotely sensed lake surface water temperature data. *Procedia Environmental Sciences*, (26):127–130.
- Hall, P., Müller, H.-G., and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *The annals of statistics*, 34(3):1493–1517.

- Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical science*, 1(3):297–310.
- Hastie, T. J., Tibshirani, R. J., and Friedman, J. H. (2011). *The elements of statistical learning: data mining, inference, and prediction*. Springer.
- Hooker, G. (2010). Functional data analysis: A short course. International Workshop on Statistical Modeling.
- Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*, volume 200. Springer Science & Business Media.
- Hothorn, T. and Everitt, B. S. (2014). *A handbook of statistical analyses using R*. CRC press.
- James, G. M. (2002). Generalized linear models with functional predictors. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):411–432.
- Keser, I. K. (2007). *A study on regularized functional principal component analysis as a statistical dimension reduction technique*. PhD thesis, Dokuz Eylül Üniversitesi.
- Keser, I. K. (2010). Ege bölgesi yağış verilerinin fonksiyonel veri analizi İle İncelenmesi. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 25(1):41–67.
- Kreyszig, E. (1978). *Introductory Functional Analysis with Applications*. John Wiley & Sons Inc., New York.
- Levitin, D., Nuzzo, R., Vines, B., and Ramsay, J. (2007). Introduction to functional data analysis. *Canadian Psychology*, 48:135–155.
- Liu, C., Ray, S., Hooker, G., and Friedl, M. (2012). Functional factor analysis for periodic remote sensing data. *The Annals of Applied Statistics*, 6(2):601–624.
- López-Pintado, S. and Romo, J. (2009). On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718–734.
- Manteiga, W. G. and Vieu, P. (2007). Statistics for functional data. *Computational Statistics & Data Analysis*, 51(10):4788–4792.

- Morris, J. S. (2014). Functional regression. <https://arxiv.org/abs/1406.4068>.
- Müller, H.-G., Chiou, J.-M., and Leng, X. (2008). Inferring gene expression dynamics via functional regression analysis. *BMC bioinformatics*, 9(1):60.
- Müller, H.-G. and Stadtmüller, U. (2005). Generalized functional linear models. *Annals of Statistics*, 33(2):774–805.
- Nechad, B., Ruddick, K., and Park, Y. (2010). Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. *Remote Sensing of Environment*, 114(4):854–866.
- Nezlin, N. P. and DiGiacomo, P. M. (2005). Satellite ocean color observations of stormwater runoff plumes along the san pedro shelf (southern california) during 1997–2003. *Continental Shelf Research*, 25(14):1692–1711.
- Nicol, F. (2013). Functional principal component analysis of aircraft trajectories. In *ISIATM 2013, 2nd International Conference on Interdisciplinary Science for Innovative Air Traffic Management*. Toulouse, France.
- Ocaña, F. A., Aguilera, A. M., and Escabias, M. (2007). Computational considerations in functional principal component analysis. *Computational Statistics*, 22(3):449–465.
- Oviedo de la Fuente, M. (2011). Utilities for statistical computing in functional data analysis: The r package 'fda.usc'. Master's thesis, Universidad de Santiago de Compostela.
- Pidwirny, M. (2006). Introduction to geographic information systems.
- Preda, C. and Saporta, G. (2005). Pls regression on a stochastic process. *Computational Statistics & Data Analysis*, 48(1):149–158.
- Preda, C. and Schiltz, J. (2011). Functional pls regression with functional response: the basis expansion approach. In *14th Applied Stochastic Models and Data Analysis Conference*, pages 1126–1133. Università di Roma La Sapienza.
- Ramsay, J. (1982). When the data are functions. *Psychometrika*, 47(4):379–396.

- Ramsay, J. (2000). Functional components of variation in handwriting. *Journal of the American Statistical Association*, 95(449):9–15.
- Ramsay, J. and Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 53(3):539–572.
- Ramsay, J., Munhall, K. G., Gracco, V. L., and Ostry, D. J. (1996). Functional data analyses of lip motion. *The Journal of the Acoustical Society of America*, 99(6):3718–3727.
- Ramsay, J. and Silverman, B. (2002). *Applied Functional Data Analysis*. Springer, USA.
- Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer, USA.
- Ramsay, J., Wang, X., and Flanagan, R. (1995). A functional data analysis of the pinch force of human fingers. *Applied Statistics*, 44(1):17–30.
- Ramsay, J. O., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer.
- Ratcliffe, S. J., Heller, G. Z., and Leader, L. R. (2002). Functional data analysis with application to periodically stimulated foetal heart rate data. ii: Functional logistic regression. *Statistics in medicine*, 21(8):1115–1127.
- Reiss, P. T., Goldsmith, J., Shang, H. L., and Ogden, R. T. (2016). Methods for scalar-on-function regression. *International Statistical Review*, 0(0):1–22.
- Reiss, P. T. and Ogden, R. T. (2007). Functional principal component regression and functional partial least squares. *Journal of the American Statistical Association*, 102(479):984–996.
- Rice, J. and Silverman, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J.R. Statist. Soc. B*, 53(1):233–243.
- Rice, J. A. (2004). Functional and longitudinal data analysis: perspectives on smoothing. *Statistica Sinica*, 14(3):631–648.
- Saeyns, W., De Ketelaere, B., and Darius, P. (2008). Potential applications of functional data analysis in chemometrics. *Journal of chemometrics*, 22(5):335–344.

- Sawant, P., Billor, N., and Shin, H. (2012). Functional outlier detection with robust functional principal component analysis. *Computational Statistics*, 27(1):83–102.
- Schwarz, J., Pinkerton, M., Wood, S., and Zeldis, J. (2008). Remote sensing of river plumes in the canterbury bight: Stage i report. *NIWA Client Report CHC2008-169*, 68.
- Shang, H. L. (2014). A survey of functional principal component analysis. *AStA Advances in Statistical Analysis*, 98(2):121–142.
- Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics*, 24(1):1–24.
- Sözen, C. (2014). *Ayrık Noktalarda Gözlenen Verilerin Analizi ve Bir Uygulama*. PhD thesis, On Dokuz Mayıs Üniversitesi.
- Ullah, S. and Finch, C. F. (2013). Applications of functional data analysis: A systematic review. *BMC medical research methodology*, 13(1):43.
- Valderrama, M. J. (2007). An overview to modelling functional data. *Computational Statistics*, 22(3):331–334.
- Viviani, R., Grön, G., and Spitzer, M. (2005). Functional principal component analysis of fmri data. *Human brain mapping*, 24(2):109–129.
- Wang, J.-L., Chiou, J.-M., and Mueller, H.-G. (2016). Functional data analysis. *The Annual Review of Statistics and Its Application*, 3(2):257–295.
- Wang, S., Wang, J., Wang, H., and Saporta, G. (2009). An Approach for PLS Regression Modeling of Functional Data. In *PLS'09, 6th Int. Conf. on Partial Least Squares and Related Methods, Pekin*, pages 28–33.
- Wasserman, L. (2006). *All of nonparametric statistics*. Springer Science & Business Media.
- Wood, S. (2006). *Generalized additive models: an introduction with R*. CRC press.
- Wu, P.-S. and Müller, H.-G. (2010). Functional embedding for the classification of gene expression profiles. *Bioinformatics*, 26(4):509–517.



Özçomak, M. S. and Gündüz, M. (2014). Borsa İstanbul'da İşlem gören Şirketlerin kapanış fiyatları ile İşlem miktarları arasındaki İlişkinin fonksiyonel kanonik korelasyon ile analizi. *International Journal of Economic and Administrative Studies*, 7(13):233–253.



## **APPENDICES**

**APPENDIX A** : Some Definitions on Functional Spaces

**APPENDIX B** : Operators and Some Usefull Theorems for Functional Data

**APPENDIX C** : Simulation Function Implemented in R





## A Some Definitions On Functional Spaces

In functional data analysis we work on functional spaces rather than finite dimensional vector spaces. Functional spaces have almost the same properties as vector spaces with the difference that vectors are replaced by functions and matrices are replaced by linear operators. A functional data is always defined on a metric or a semi-metric space. The primary step of functional data analysis is to decide the space on which the functional data is defined.

In this part some definitions and theorems have been given which are useful while dealing with functional data. Some of the references that are used are Kreyszig (1978); Ramsay and Silverman (2005); Horváth and Kokoszka (2012) and Cuevas (2014).

### 1.1 Metric Spaces and Its Properties

A *metric space* is a pair  $(X, d)$  where  $X$  is a set and  $d : X \times X \rightarrow R$  is a distance function (metric) on  $X$  which satisfies following conditions for all  $x, y, z \in X$  :

1.  $d$  is real valued, non-negative and finite. (Positivity)
2.  $d(x, y) = 0$  if and only if  $x = y$ . (Identity)
3.  $d(x, y) = d(y, x)$  (Symmetry)
4.  $d(x, y) \leq d(x, z) + d(z, y)$  (Triangle inequality)

The real line  $\mathbb{R}$  with the usual metric  $d(x, y) = |x - y|$  and the plane  $\mathbb{R}^2$  with the Euclidean metric  $d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$  are examples of metric spaces.

A *semi-metric space* satisfies the same properties as a metric space except the second property mentioned below which means that in semi-metric spaces the distance  $d(x, y)$  between two elements  $x, y \in X$  can be equal to zero even when  $x$  and  $y$  are distinct points.

An example of a semi metric is the set of absolute integrable functions defined on the interval  $[0, 1]$  which can be given as

$$d(\chi, \gamma) = \int_0^1 |\chi(t) - \gamma(t)| dt$$

### 1.2 Inner Product Spaces and Its Properties

An *inner product* is an operator with properties positivity, symmetry and bilinearity. The inner product of  $x$  and  $y$  is notated by  $\langle x, y \rangle$  and have following properties:

1.  $\langle x, x \rangle \geq 0$  for all  $x$ .  $\langle x, x \rangle = 0$  if and only if  $x = 0$ . (Positivity)
2.  $\langle x, y \rangle = \langle y, x \rangle$  for all  $x$  and  $y$ . (Symmetry)
3. For all  $a, b \in \mathbb{R}$  and all vectors  $x, y$  and  $z$ ,  
 $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$ . (Bilinearity)

A vector space on which an inner product can be defined is called an *inner product space*.

For Euclidean space the inner product of two vectors  $x = (x_1, x_2, \dots, x_n)$  and  $y = (y_1, y_2, \dots, y_n)$  is given as dot product.

$$\langle (x_1, x_2, \dots, x_n), (y_1, y_2, \dots, y_n) \rangle = x_1y_1 + x_2y_2 + \dots + x_ny_n = \sum_i x_iy_i$$

In the vector space of real functions with inner product, the sum is replaced by integral. For the functions  $\chi(t)$  and  $\gamma(t)$ , where  $t$  takes values in the interval  $[0, T]$ , the inner product is defined as

$$\langle \chi, \gamma \rangle = \int_0^T \chi(t)\gamma(t)dt$$

### 1.3 Normed Spaces and Its Properties

In a vector space the inner product of an element  $x$  is a measure of its size. The positive square root of this size is called the *norm* and the norm of  $x$  is shown as

$$\|x\|^2 = \langle x, x \rangle$$

where  $\|x\| \geq 0$  according to the positivity axiom of inner product.

The norm of an element of  $n$ -dimensional Euclidean space is equal to the length of the vector. The norm of a function  $\chi$  is called its  $\mathcal{L}^2$  norm and shown as

$$\|\chi\| = \left( \int \chi(t)^2 dt \right)^{1/2}$$

The properties of norm are similar to the properties of inner product and can be summarized as follows:

1.  $\|x\| \geq 0$  for all  $x$ .  $\|x\| = 0$  if and only if  $x = 0$ . (Positivity)
2.  $\|x + y\| = \|y + x\|$  for all  $x$  and  $y$ . (Symmetry)
3. For all real numbers  $a$ ,  
 $\|ax\| = |a|\|x\|$ .

Let  $X$  be a vector space and a function  $\|\cdot\| : X \rightarrow \mathbb{R}$  is given. If for all elements  $x, y \in X$  and a scalar  $\lambda \in \mathbb{R}$  the properties mentioned below are satisfied, then  $(X, \|\cdot\|)$  is called a *normed space*.

*All normed spaces are metric spaces but all metric spaces are not normed spaces. Similarly, all inner product spaces are normed spaces but the reciprocal is not true.*

## 1.4 Hilbert Spaces

If a  $H$  is a complete inner product space, then  $H$  is called a *Hilbert space*.

The space  $L^2 = L^2([0, T])$  is the set of measurable real valued functions defined on  $[0, T]$  satisfying  $\int_0^T \chi^2(t) dt < \infty$ . It is called as *the space of square integrable functions*. The space  $L^2$  is a separable Hilbert space with an inner product of  $\chi$  and  $\gamma$  is given by

$$\langle \chi, \gamma \rangle = \int \chi(t)\gamma(t)dt.$$



## B Operators and Some Useful Theorems for Functional Data

### 2.1 Linear Operator, Integral Operator and Hilbert-Schmidt Operator

A *linear operator* is a mapping between two vector spaces.

A linear operator  $\Psi(x)$  is said to be symmetric and positive definite if it satisfies the properties:

1.  $\langle \Psi(x), x \rangle = \langle x, \Psi(x) \rangle$ . (Symmetry)
2.  $\langle \Psi(x), x \rangle \geq 0$ . (Positive-Definiteness)

An *integral operator*  $\Psi$  on  $L^2(T)$ ,  $\Psi : \xi \rightarrow \Psi\xi$ ,  $\xi \in L^2(T)$  with the kernel function  $\psi : T \times T \rightarrow \mathbb{R}$  is defined by

$$\Psi(\xi)(t) = \int_T \psi(t, s) \xi(s) ds, \quad \xi \in L_2(T). \quad (\text{B.1})$$

If the kernel function  $\psi$  satisfies the condition  $\int \int \psi^2(t, s) dt ds < \infty$ , then this kernel is called a *Hilbert-Schmidt kernel*. A linear and bounded integral operator with a Hilbert-Schmidt kernel is called a *Hilbert-Schmidt operator*.

If the related kernel of an integral operator is symmetric and positive-definite, then the integral operator is also symmetric and positive definite.

### 2.2 Covariance Operator and Its Properties

The covariance operator is an integral operator where the covariance function  $c(t, s) = \mathbb{E}[\chi(t)\chi(s)]$  is its kernel and it is expressed by

$$\Gamma_\chi \xi(t) = \int c(t, s) \xi(s) ds, \quad \xi \in L^2(T). \quad (\text{B.2})$$

The covariance function  $c(s, t)$  satisfies the properties of symmetry and positive definiteness.

$$c(s, t) = c(t, s) \quad (\text{B.3})$$

$$\int \int c(t,s)\xi(t)\xi(s)dtds = \int \int E[\chi^c(t)\chi^c(s)\xi(t)\xi(s)] = E[(\int \chi^c(t)\xi(t)dt)^2] \geq 0. \quad (\text{B.4})$$

Therefore the covariance operator  $\Gamma_\chi$  is also symmetric and positive-definite.

Considering the eigenequation  $\Gamma_\chi \xi_j = \lambda_j \xi_j$  it can be shown that the covariance operator has non-negative eigenvalues:

$$\lambda_j = \langle \Gamma_\chi \xi_j, \xi_j \rangle = \langle \mathbb{E}[\langle \chi, \xi_j \rangle \chi], \xi_j \rangle = \mathbb{E}[\langle \chi, \xi_j \rangle^2]. \quad (\text{B.5})$$

If infinite sum is taken at the both sides of the equation (B.5) it is seen that

$$\sum_{j=1}^{\infty} \lambda_j = \sum_{j=1}^{\infty} \mathbb{E}[\langle \chi, \xi_j \rangle^2] = \mathbb{E}\|\chi\|^2 < \infty, \quad (\text{B.6})$$

so the eigenvalues of the covariance operator satisfy the condition  $\sum_{j=1}^{\infty} \lambda_j < \infty$ .

An operator is a covariance operator if and only if it is symmetric, positive-definite and bounded.

### 2.3 Singular Value Decomposition of a Linear Operator

Consider a separable Hilbert space  $H$  with inner product  $\langle \cdot, \cdot \rangle$  that generates the norm  $\|\cdot\|$ . Let  $\mathcal{L}$  denote the space of bounded linear operators on  $H$  with the norm

$$\|\Psi\|_{\mathcal{L}} = \sup\{\|\Psi(x)\| : \|x\| = 1\}.$$

Singular value decomposition establishes that a compact operator  $\Psi \in \mathcal{L}$  can be represented by

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \xi_j \rangle \eta_j, \quad (\text{B.7})$$

where  $\eta_j$  and  $\xi_j$  are orthonormal bases and  $\lambda_j$  is a real sequence converging to zero.

### 2.4 Spectral Decomposition of a Hilbert-Schmidt Operator

A symmetric and positive definite Hilbert-Schmidt operator ensures the spectral decomposition

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, \xi_j \rangle \xi_j, \quad x \in H \quad (\text{B.8})$$

where  $\xi_j$ 's are the orthonormal eigenfunctions of  $\Psi$  which are assumed to form an orthonormal basis.



## 2.5 Parseval's Equality

Let  $\xi_j$  be an orthonormal basis in a separable Hilbert space  $H$ . Then Parseval's equality states that

$$\sum_j \langle \chi, \xi_j \rangle = \|\chi\|^2 = \langle \chi, \chi \rangle. \quad (\text{B.9})$$

## 2.6 Riesz Representation Theorem

Riesz Representation theorem establishes that the covariance operator  $\Gamma_\chi : L^2(T) \rightarrow L^2(T)$  can be written as

$$\Gamma_\chi \xi = \mathbb{E}[\langle \chi, \xi \rangle \chi], \quad \xi \in L^2(T). \quad (\text{B.10})$$

## 2.7 Mercer's Theorem

The continuous, symmetric, positive-definite kernel  $\psi(t, s)$  of a Hilbert Schmidt operator can be represented by

$$\psi(t, s) = \sum_{j=1}^{\infty} \lambda_j \xi_j(t) \xi_j(s) \text{ in } L^2(T \times T) \quad (\text{B.11})$$

where  $\xi_j$  are orthonormal functions in  $L^2(T)$  and  $\lambda_j$  is a decreasing sequence of positive numbers. Mercer's Lemma states that the kernel of a symmetric, positive definite linear operator can be written in terms of its eigenvalues and eigenfunctions.

## 2.8 Karhunen-Loève Expansion

Let be a centered stochastic process with  $\mathbb{E}[\chi(t)] = 0$  and  $\mathbb{E}[\chi^2(t)] \leq \infty$  for all values of  $t$ . Then  $\chi(t)$  can be represented by

$$\hat{\chi}(t) = \sum_{j=1}^{\infty} f_j \xi_j(t) \quad (\text{B.12})$$

where  $\{\xi_j\}_{j \in \mathbb{N}}$  is an orthonormal basis given by the eigenfunctions of the covariance operator  $\Gamma_\chi$  associated with the corresponding eigenvalue  $\lambda_j$  and  $f_j$  is a sequence of orthogonal random variables with  $E[f_j] = 0$ ,  $E[f_i f_j] = 0$  for  $i \neq j$  and  $E[f_j^2] = \lambda_j$  (Cuevas (2014) has been followed here).

## C Simulation Function Implemented In R

```
library(fda)
library(fda.usc)

#simulation<-function(Xt,rep=100,P=500){
rep=100
P=500
for(p in 1:P){
print(p)
print(date())
mep<-function(newx,newy,mod,exp=FALSE){
pred<-predict(mod,newx)
if(exp) MEP <- ((1 / length(newy)) * sum((newy - exp(pred))^2))
else MEP <- ((1 / length(newy)) * sum((newy - pred)^2))
return(MEP)
}

simexp<-function(data_665s,samp=samp,rep=rep){
amep.665s<-numeric()
for (r in 1:rep){
data_665s.sam<-data_665s[samp[[r]],]
aux<-lm(lY~Xt_665,data=data_665s.sam)
mep.reg665s<-mep(newx=data_665s[-samp[[r]],2,drop=FALSE],newy=exp(data_665s[-
samp[[r]],1]),mod=aux,exp=TRUE)
scalvar<-var(exp(data_665s[-samp[[r]],1]))
amep.665s[r]<-mep.reg665s/scalvar
}
return(median(amep.665s))
}

simfunc<-
function(chi,lgY,mod="bas",samp=samp,rep=rep,bas1=basis1,bas2=basis2,bas2d2=basis2d2){
amep.sim<-numeric()
for (r in 1:rep){
chi.aux<-chi[samp[[r]],]
lgY.aux<-lgY[samp[[r]],]
if(mod=="bas") aux<-fregre.basis(chi.aux,lgY.aux,bas1,bas2)
if(mod=="fpc") aux<-fregre.pc(chi.aux,lgY.aux,l=1:8)
if(mod=="fpcv"){
aux0<-fregre.pc.cv(chi.aux,lgY.aux,8)
```

```

aux<-fregre.pc(chi.aux,lgY.aux,aux0$pc.opt)
}
if(mod=="fpls") aux<-fregre.pls(chi.aux,lgY.aux,l=1:8)
if(mod=="fplscv"){
aux0<-fregre.pls.cv(chi.aux,lgY.aux,8)
aux<-fregre.pls(chi.aux,lgY.aux,aux0$pls.opt)
}
if(mod=="fd2") aux<-fregre.basis(chi.aux,lgY.aux,bas1,bas2d2)

mep.sim<-mep(chi[-samp[[r]],,drop=FALSE],exp(lgY[-samp[[r]]]),mod=aux,exp=TRUE)
scalvar<-var(exp(lgY[-samp[[r]]]))
amep.sim[r]<-mep.sim/scalvar
}
return(median(amep.sim))
}
MEPlist_exp<-vector(mode="list",length=P)
MEPlist_bas<-vector(mode="list",length=P)
MEPlist_pc<-vector(mode="list",length=P)
MEPlist_pcv<-vector(mode="list",length=P)
MEPlist_d2<-vector(mode="list",length=P)
MEPS<-vector(mode="list",length=P)
lys<-vector(mode="list",length=5)
for(i in 1:5) lys[[i]]<-vector(mode="list",length=3)
###Exponential Regression Model with Rrs 665nm
#a vector composed of rrs at 665 nm
Xt_665<-Xt[,7]
b<-reg.665$coefficients

###generating IY.exp
lys[[1]][[1]]<-IYexp.sd1<-b[1]+Xt_665*b[2]+rnorm(length(Xt_665),sd=sd(reg.665$residuals))
lys[[1]][[2]]<-IYexp.sd2<-b[1]+Xt_665*b[2]+rnorm(length(Xt_665),sd=sd(reg.665$residuals)/5)
lys[[1]][[3]]<-IYexp.sd3<-b[1]+Xt_665*b[2]+rnorm(length(Xt_665),sd=sd(reg.665$residuals)/10)

###generating IY.bas
listXt<-list()
wl<-Xt[,c(1:8)]
listXt$rres<-wl[,]
rres.nm<-as.numeric(substring(names(listXt$rres),5))
Xtfddata<-fdata(listXt$rres,argvals=rres.nm,names=list(main='Water reflectances of box
data',xlab='Wavelength',ylab='Water Reflectances'))
ts<-Xtfddata[["argvals"]]
bsXt<-create.bspline.basis(rangeval=range(ts),nbasis=8)
fdXt <- Data2fd(argvals=ts,y=t(Xtfddata$data),basisobj=bsXt)
Cnew<-fdXt$coefs
Xtfd<-fdata2fd(Xtfddata,nbasis = 8)
J<-flrly$J
B<-flrly$beta.est$coefs
a<-flrly$a.est
lys[[2]][[1]]<-IYbas.sd1<-a+t(Cnew)%*%J%*%B+rnorm(n=nrow(t(Cnew)),sd=sd(flrly$residuals))
lys[[2]][[2]]<-IYbas.sd2<-a+t(Cnew)%*%J%*%B+rnorm(n=nrow(t(Cnew)),sd=sd(flrly$residuals)/5)
lys[[2]][[3]]<-IYbas.sd3<-a+t(Cnew)%*%J%*%B+rnorm(n=nrow(t(Cnew)),sd=sd(flrly$residuals)/10)

```

```

##generating IY.pc
Xtcen<-fdata.cen(Xtfdata) Xtfdatacen<-Xtcen$Xcen
Fpc<-inprod.fdata(Xtfdatacen,fpcly$fdata.comp$rotation)
bpc<-fpcly$coefficients[2:9]
sdpc<-sd(fpcly$residuals)
lys[[3]][[1]]<-IYpc.sd1<-mean(fpcly$y)+Fpc%%as.matrix(bpc)+rnorm(n=9777,0,sd=sdpc)
lys[[3]][[2]]<-IYpc.sd2<-mean(fpcly$y)+Fpc%%as.matrix(bpc)+rnorm(n=9777,0,sd=sdpc/5)
lys[[3]][[3]]<-IYpc.sd3<-mean(fpcly$y)+Fpc%%as.matrix(bpc)+rnorm(n=9777,0,sd=sdpc/10)

##generating IY.pcv
Fpcv<-inprod.fdata(Xtfdatacen,fpclv$fregre.pc$fdata.comp$rotation)
bpcv<-fpclv$fregre.pc$coefficients[2:3]
sdpcv<-sd(fpclv$fregre.pc$residuals)
lys[[4]][[1]]<-IYpcv.sd1<-
mean(fpclv$fregre.pc$y)+Fpcv[,c(2,5)]%%as.matrix(bpcv)+rnorm(n=9777,0,sd=sdpcv)
lys[[4]][[2]]<-IYpcv.sd2<-
mean(fpclv$fregre.pc$y)+Fpcv[,c(2,5)]%%as.matrix(bpcv)+rnorm(n=9777,0,sd=sdpcv/5)
lys[[4]][[3]]<-IYpcv.sd3<-
mean(fpclv$fregre.pc$y)+Fpcv[,c(2,5)]%%as.matrix(bpcv)+rnorm(n=9777,0,sd=sdpcv/10)

##generating IY.der2
der2Xt<-fdata.deriv(Xtfdata, nbasis=8, nderiv=2)
listder2Xt<-list()
wld<-der2Xt[,c(1:8)]
listder2Xt$rres<-wld[,]
rss.nm<-as.numeric(substring(names(listder2Xt$rres),5))
ts<-Xtfdata[["argvals"]]

bsXt<-create.bspline.basis(rangeval=range(ts),nbasis=8)
der2Xtfd <- Data2fd(argvals=ts,y=t(der2Xt$data),basisobj=bsXt)
Cd2<-der2Xtfd$coefs
Jd2<-regd2ly$J
Bd2<-regd2ly$coefficients[2:5]
ad2<-regd2ly$coefficients[1]
lys[[5]][[1]]<-IYd2.sd1<-ad2+t(Cd2)%*%Jd2)%*%Bd2+rnorm(n=nrow(t(Cd2)),sd=sd(regd2ly$residuals))
lys[[5]][[2]]<-IYd2.sd2<-
ad2+t(Cd2)%*%Jd2)%*%Bd2+rnorm(n=nrow(t(Cd2)),sd=sd(regd2ly$residuals)/5)
lys[[5]][[3]]<-IYd2.sd3<-
ad2+t(Cd2)%*%Jd2)%*%Bd2+rnorm(n=nrow(t(Cd2)),sd=sd(regd2ly$residuals)/10)
samp25<-samp50<-samp100<-list()
for(r in 1:rep){
  samp25[[r]]<-sample(1:length(data_665s[,1]),25)
  samp50[[r]]<-sample(1:length(data_665s[,1]),50)
  samp100[[r]]<-sample(1:length(data_665s[,1]),100)
}
cont<-TRUE
for(z in 1:5){
  if(cont){
    MEPaux<-as.data.frame(matrix(nrow=3,ncol=21,NA))
    for(i in 1:3){
      data_665s$Y<-lys[[z]][[i]]
    }
  }
}

```

```

for(j in 1:3){
  if(j==1) sampl<-samp25
  if(j==2) sampl<-samp50
  if(j==3) sampl<-samp100
  if(cont){
    MEPaux[i,((j-1)*7)+1]<-try(simexp(data_665s,samp=sampl,rep=rep),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+1])
  }
  if(cont){
    MEPaux[i,((j-1)*7)+2]<-
    try(simfunc(chi=Xtfddata,lgY=as.matrix(lys[[z]][[i]]),samp=sampl,rep=rep,mod="bas"),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+2])
  }
  if(cont){
    MEPaux[i,((j-1)*7)+3]<-
    try(simfunc(Xtfddata,as.matrix(lys[[z]][[i]]),samp=sampl,rep=rep,mod="fpc"),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+3])
  }
  if(cont){
    MEPaux[i,((j-1)*7)+4]<-
    try(simfunc(Xtfddata,as.matrix(lys[[z]][[i]]),samp=sampl,rep=rep,mod="fpcv"),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+4])
  }
  if(cont){
    MEPaux[i,((j-1)*7)+5]<-
    try(simfunc(Xtfddata,as.matrix(lys[[z]][[i]]),samp=sampl,rep=rep,mod="fpls"),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+5])
  }
  if(cont){
    MEPaux[i,((j-1)*7)+6]<-
    try(simfunc(Xtfddata,as.matrix(lys[[z]][[i]]),samp=sampl,rep=rep,mod="fplscv"),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+6])
  }
  if(cont){
    MEPaux[i,((j-1)*7)+7]<-
    try(simfunc(der2Xt,as.matrix(lys[[z]][[i]]),samp=sampl,rep=rep,mod="fd2"),TRUE)
    cont<-is.numeric(MEPaux[i,((j-1)*7)+7])
  }
}
}
}
colnames(MEPaux)<-
paste(rep(c('exp','bas','fpc','fpcv','fpls','fplscv','der2'),3),'_',rep(c(25,50,100),each=7),sep="")
if(z==1){
  rownames(MEPaux)<-c('IYexp.sd1','IYexp.sd2','IYexp.sd3')
  MEPlist_exp[[p]]<-MEPaux
}
if(z==2){
  rownames(MEPaux)<-c('IYbas.sd1','IYbas.sd2','IYbas.sd3')
  MEPlist_bas[[p]]<-MEPaux
}
if(z==3){
  rownames(MEPaux)<-c('IYpc.sd1','IYpc.sd2','IYpc.sd3')
}
}
}

```

```

MEPlist_pc[[p]]<-MEPaux
}
if(z==4){
rownames(MEPaux)<-c('IYpcv.sd1','IYpcv.sd2','IYpcv.sd3')
MEPlist_pcv[[p]]<-MEPaux
}
if(z==5){
rownames(MEPaux)<-c('IYd2.sd1','IYd2.sd2','IYd2.sd3')
MEPlist_d2[[p]]<-MEPaux
}
if(z==1) MEPSaux<-MEPaux
else MEPSaux<-rbind(MEPSaux,MEPaux)
}
}
MEPS[[p]]<-MEPSaux
if(cont){
arch<-paste("iter_NAD_Y_",p,".csv",sep="")
write.csv2(MEPSaux,arch)
}
}
#}

for (p in 1:P){
if (p==1) sumMEP=as.matrix(MEPS[[1]])
else sumMEP=sumMEP+as.matrix(MEPS[[p]])
}
meanMEP<-sumMEP/P
write.csv2(sumMEP,"sumMEP_500.csv")
write.csv2(meanMEP,"meanMEP_500.csv")

```

## CURRICULUM VITAE

**Name:** Nihan ACAR DENIZLI

**Birth Date and Place:** 07/03/1985 BURSA

**E-Mail:** nihan.acar@msgsu.edu.tr

**Bachelor:** Marmara University Department of Mathematics

**Master:** Mimar Sinan Fine Arts University Statistics Programme

### CONFERENCE PAPERS

- **Acar-Denizli N.**, Başarır G., Caballero I., Delicado P., 2016: The Application Of Functional Linear Models On Remote Sensing Data In Oceanography. *2016 CRoNoS Summer Course On Functional Data Analysis*, August 26-28, 2016, Oviedo, SPAIN.
- **Acar-Denizli N.**, Delicado P., Başarır G., Caballero I., 2016: A simulation based comparative study of functional linear regression models on remote sensing data. *ERCIM 2016*, December 9-11, 2016, Sevilla, SPAIN. (invited to the session "EO117: Functional data analysis" organised by Alicia Nieto-Reyes.)
- **Acar-Denizli N.**, Delicado P., Başarır G., Caballero I., 2017: The Evaluation of the Performance of Functional Linear Regression Models on a Remote Sensing Application. *61st World Statistics Congress, ISI 2017*, July 16-21, 2016, Marrakech, MOROCCO. (accepted to the session "IPS112: Young researchers session on Functional Data Analysis" organized by Kristina Lurz and Nihan Acar-Denizli.)