Geostatistical Analysis of Functional Data

Ramón Giraldo Henao

Academic Dissertation for the Degree of PhD Advised by

Dr. Pedro Delicado
Universitat Politècnica de Catalunya

and

Dr. Jorge Mateu
Universitat Jaume I

PhD Thesis Submitted to Universitat Politècnica de Catalunya

Barcelona, July 2009
Acknowledgements

I am especially grateful to my advisors Pedro Delicado and Jorge Mateu for their constant support and guidance. During these years they have shared with me their expertise and research insight. All time they help me to find solutions to theoretical and practical problems in order to improve the papers and the thesis.

I would like to thank to the National University of Colombia and to the FUNDACION CAROLINA for financial support through a PhD scholarship. I also received financial support from the research groups Complex Data Analysis and Statistical Modeling of Environmental Problems, which are supported by the Spanish Ministry of Education and Science through grants MTM2006-09920 and MTM2007-62923.

I thank to my wife Astrid and my sons Alejandro and Isabella. They have lost a lot due to my research abroad. Without their encouragement and understanding it would have been impossible for me to finish this work. My special gratitude is due to my mother Loly and my brother Emilio.
Abstract

Functional data analysis concerns with statistical modeling of random variables taking values in a space of functions (functional variables). Several standard statistical techniques such as regression, ANOVA or principal components, among others, have been considered from a functional point of view. In general, these methodologies are focussed on independent and identically distributed functional variables. However, in several disciplines of applied sciences there exists an increasing interest in modeling correlated functional data. In particular in most of them the modeling of spatially correlated functional data is of interest. This is the topic treated here. Specifically this work concerns with spatial prediction of curves when we dispose of a sample of curves collected at sites of a region with spatial continuity.

Four methods for doing spatial prediction of functional data are developed. Initially, we propose a predictor having the same form as the classical kriging predictor, but considering curves instead of one-dimensional data. The other predictors arise from adaptations of functional linear models for functional response to the case of spatially correlated functional data. One the one hand, we define a predictor which is a combination of kriging and the functional linear point-wise (concurrent) model. On the other hand, we use the functional linear total model for extending two classical multivariable geostatistical methods to the functional context. The first predictor is defined in terms of scalar parameters. In the remaining cases the predictors involves functional parameters. We adapt an optimization criterion used in multivariable spatial prediction in order to estimate scalar and functional parameters involved in the predictors proposed. In all cases a non-parametric approach based on expansion in terms of basis functions is used for getting curves from discrete data.
The number of basis functions is chosen by cross-validation. The methodologies proposed are illustrated by analyzing three real data sets corresponding to curves of penetration resistance and temperature which are functions of depth and time, respectively.

*Keywords*: Basis functions; Cokriging; Cross-validation; Functional linear model; Kriging; Multivariable spatial prediction.
# Contents

1 Introduction .......................... 1  
   1.1 Data sets .......................... 6  
      1.1.1 Penetration resistance curves .......................... 6  
      1.1.2 Canadian temperatures .......................... 8  
      1.1.3 Canada’s Maritime provinces temperatures .......................... 8  

2 Geostatistics and functional data analysis .......................... 11  
   2.1 Geostatistics .......................... 11  
      2.1.1 Univariate geostatistics .......................... 12  
      2.1.2 Multivariable geostatistics .......................... 15  
      2.1.3 Space-time geostatistics .......................... 18  
   2.2 Functional data analysis .......................... 20  
      2.2.1 Notation .......................... 21  
      2.2.2 From data to functions by using basis functions .......................... 23  
      2.2.3 Some techniques for modeling functional data .......................... 32  

3 Ordinary kriging for functional data .......................... 38  
   3.1 Predictor and estimation of parameters .......................... 39  
   3.2 Estimating the trace-variogram .......................... 43  
   3.3 A non-parametric approach .......................... 45  
   3.4 Prediction of penetration resistance curves .......................... 47  
      3.4.1 Evaluation criterion of a kriging predictor .......................... 47
3.4.2 Analysis of data using B-splines ........................................ 49
3.4.3 Analysis of data using wavelets ........................................ 54

4 Continuos time-varying kriging ............................................. 58
  4.1 Predictor and minimization problem .................................... 58
  4.2 A solution based on basis functions .................................... 60
  4.3 Prediction of temperature in the Maritimes ............................ 63
  4.4 Application: Canadian temperature curves ............................ 70

5 From multivariable to functional geostatistics .......................... 79
  5.1 Cokriging based on functional data ..................................... 80
  5.2 Functional Kriging: Total model ....................................... 86
  5.3 Application: Canada’s Maritime provinces temperatures ............ 93
  5.4 A comparison with OKFD and CTKFD .................................. 98

6 Conclusions and further research ......................................... 101

Bibliography ......................................................................... 103

Appendix ............................................................................. 112

A R functions ....................................................................... 112
  A.1 Code for modeling temperature data by OKFD ...................... 112
  A.2 Code for modeling temperature data by CTKFD .................... 117
  A.3 Code for modeling the temperature data by FKTM ............... 122
List of Figures

1.1 Observed penetration resistance curves (right panel) on 32 sampling points (left panel). Data collected during 2004 at Marengo Experimental Station (National University of Colombia). ........................................... 7

1.2 Averages (over 30 years) of mean daily temperature curves (right panel) observed at 35 Canadian weather stations (left panel). Red dot corresponds to Slave Lake, an unsampled station .......................................................... 9

1.3 Averages (over 30 years) of daily temperature curves (right panel) observed at 35 Canadian Maritime weather stations (left panel). Red dot corresponds to Moncton, an unsampled location .......................................................... 9

2.1 B-Splines basis with L=10 ............................................................. 25

2.2 Fourier basis with K=5 and period 365 ........................................... 26

2.3 Wavelets families (a) Haar (b) Daubechies 4 (c) Coiflet1 (d) Symlet2 (e) Meyer (f) Morlet (g) Mexican hat. Source: www.dtic.upf.edu/ xserra/cursos/ 29

2.4 Simulated example of a Haar wavelet transform. Left panel: Signal. Right panel: Haar wavelet 1-level transform of the signal. ................................. 31

3.1 Set of 32 penetration resistances observed curves (left) and smoothed curves by B-splines (right) ......................................................... 47

3.2 Contour plot for the function $FCV(L, \eta)$, using a logarithmic scale for $\eta$ (eta in the graphic). ......................................................... 51
3.3 Left panel: Kriging weights $\lambda_i$ as a function of the Euclidean distance between sites $s_i$ and $s_0$. Right panel: Kriging prediction at an unsampled location (black thick line) over the set of smoothed curves. 52

3.4 Left panel: Kriging predictions for the 32 original sites based on functional cross-validation (data smoothed by B-splines). Right panel: Functional cross-validation residuals (grey lines), residual mean (black thick line) and residual standard deviation (dashed line). 53

3.5 Left panel: Inverse Haar wavelet transform of penetration resistance data. Right panel: Kriging prediction at an unsampled location (black thick line) over the set of inverse transform wavelet curves. 55

3.6 Left panel: Kriging predictions for the 32 original sites based on functional cross-validation (data convert to functions by a Haar wavelet transform). Right panel: Functional cross-validation residuals (grey lines), residual mean (black thick line) and residual standard deviation (dashed line). 56

4.1 Sum of squared errors as a function of $K$ obtained by non-parametric cross-validation for Canada’s Maritime provinces data set. 64

4.2 Estimated functional parameters (left, dark lines), sum of functional estimated parameters (left, dotted line), smoothed temperature curves (right, light lines), temperature prediction function at Moncton (right, dark line) and real temperature values at Moncton (right, circles). 66

4.3 Left panel: Point-wise kriging predictions based on cross-validation. Right panel: Cross-validation residuals (light lines), residual mean (dark line) and residual standard deviation (dashed line) for Canada’s Maritime provinces data set. 67

4.4 Temperature curves (observed, smoothed and predicted by point-wise kriging) at four weather stations in Canada’s Maritime provinces. 69

4.5 Point-wise kriging prediction variances of cross-validation analysis. Circle diameter is proportional to the prediction variance. 70
4.6 Estimated functional parameters $\hat{\alpha}(t)$ (top), $\hat{\beta}_1(t)$ (centre) and $\hat{\beta}_2(t)$ (bottom) for the estimation of temperature (functional response) from geographical coordinates (longitude, latitude) with the Canadian data set. Variables are standardized. The model includes a constant function. Dashed lines are 95% point-wise confidence limits for the functional parameters.

4.7 Left panel: Smoothed temperatures by using a Fourier basis function with 65 functions. Right panel: Estimations of temperature obtained by a functional multiple regression model with functional response (smoothed temperature) and two covariates (longitude and latitude coordinates).

4.8 Residuals of the functional regression model of temperature versus the coordinates (longitude and latitude) with the Canadian data set.

4.9 Estimated functional parameters for predicting a residual curve at Slave Lake (left, dark lines), sum of estimated functional parameters (left, dotted line), smoothed temperature curves (right, clear lines), temperature prediction function at Slave Lake (right, dark line) and real temperature values at Slave Lake (right, circles). Temperature prediction was obtained by adding the trend plus the residual prediction.

4.10 Right panel: Point-wise kriging predictions based on cross-validation. Left panel: Cross-validation residuals (clear lines), residual mean (dark line) and residual standard deviation (dashed line) for the Canadian temperature data set.

4.11 Temperature curves (observed, smoothed and predicted by continuous time kriging) at four Canadian weather stations. Temperature predictions obtained by adding the trend plus the residual prediction achieved by continuous time kriging.

5.1 Left panel: Cross-validation predictions obtained by FKTW. Right panel: Predicted temperature at Moncton station obtained by FKTW (dark line), smoothed temperature curves (gray color), and real data recorded at Moncton (circles).
5.2 Left panel: Variance of smoothed and predicted curves. Right panel: Cross-validation residuals obtained by FKTM (*clear lines*), residual mean (*dark line*) and residual standard deviation (*dashed line*).  

95

5.3 Top: Estimated functional parameters for Bouctouche (NB) and Nappan (NB). Bottom: Estimated functional parameter for Aroostook (NB) and functional parameters for all stations when we do prediction at Moncton on 9th April (day 100).  

96

5.4 Functional parameters for doing prediction at Moncton on July 18th (day 200) by means of CBFD (left panel) and FKTM (right panel).  

97
Chapter 1

Introduction

In an increasing number of problems in a wide range of applied sciences, the collected data are curves. Such curve data may be generated by densely space-time repeated measurements, or by automatic recordings of a quantity of interest. Functional Data Analysis (FDA) (Ramsay and Silverman, 2005) is used, since beginning of the nineties, in order to model these type of information. Agronomy, meteorology, ecology and other sciences where the geostatistical analysis (Cressie, 1993) is often used to describe spatial distribution, are not exceptions. However, we have found only a few references that use FDA techniques in the spatial context (Goulard and Voltz, 1993; Yamanishi and Tanaka, 2003; Illian et al., 2006; Baladandayuthapani et al., 2008; Nerini and Monestiez, 2008). Consequently the motivation of this work is to offer a solution to the problem of predicting curves on unsampled locations of a region, based on joint theories of these two statistical branches. A background about geostatistics and functional data analysis is given in Chapter 2.

The modeling of measured variables in different sites of a region with spatial continuity (Cressie, 1993) has had since the sixties, with the development of the geostatistical analysis (Matheron, 1962), an increasing use in several scientific disciplines like mining (Journel and Huijbregts, 1978), geology (Samper and Carrera, 1993), ecology (Robertson, 1987) or environmental sciences (Cressie and Majure, 1995; Diggle et al., 1995; Paez et al., 2005). Conventional geostatistical analysis is a two-step procedure (Isaaks and Srivastava, 1987;
First, the spatial structure of the variable is examined with the variogram analysis (Samper and Carrera, 1993). Once a spatial structure has been identified and accurately described by a suitable model (spherical, exponential, Gaussian or Matérn, for instance), some kriging procedure (simple, ordinary, universal) is used for interpolating the variable at unvisited sites (Isaaks and Srivastava, 1987).

Both the possibility that computational methods can be implemented and the necessity of including statistical assumptions not considered before, has made that new geostatistical theories have been developed. Diggle et al. (1998) and Diggle and Ribeiro (2000), propose a model-based geostatistics theory, where by assuming that the spatial random process follows some distribution of the exponential family, it is possible to estimate the autocorrelation parameters by maximum likelihood (ML) or restricted maximum likelihood (REML). In addition, if priors are assumed as a purpose of taking into account uncertainty about the parameters then Bayesian analysis could be employed. Le and Zidek (1992) and Handcock and Stein (1993) showed alternatives to the ordinary kriging assuming a Gaussian random field and unknown covariance matrix. Posterior predictive distribution is obtained based on both the likelihood and the priors for parameters of a correlation Matérn model (Matérn, 1980). De Oliveira et al. (1997) showed an alternative to the trans-Gaussian kriging called Bayesian transformed Gaussian model. They studied uncertainty proper of the transformation used to get normality. All Bayesian approximations mentioned above are based on Markov Chain Monte Carlo (MCMC) methods (Robert and Casella, 1999) for solving integrals of both the posteriori and predictive distributions.

In the multivariable geostatistical context $m$ spatial random processes are considered simultaneously (Ver Hoef and Cressie, 1993; Pebesma, 2004). It is necessary to find the spatial covariance model for all variables included in the analysis. We then use the multivariable information for predicting each one on unsampled points. A particular case of multivariable geostatistics is the bivariate cokriging method (Cressie, 1993). In multivariable geostatistics a linear model of coregionalization (LMC) (Journel and Huijbregts, 1978) is fitted to the spatial cross-covariance matrix as a first step. Wackernagel (1995, 1998) show the theory for carrying out spatial principal components analysis based on a LMC. Borgault and Marcotte (1991) and Borgault et al. (1992) also use LMC to fit the
multivariable covariogram in order to make cluster analysis of spatial data.

Another important branch of the geostatistical analysis is the modeling of space-time data. Geostatistical spatio-temporal models provide a probabilistic framework for data analysis and predictions based on temporal dependence between observations (Kyriakidis and Journel, 1999). In opposition to multivariable geostatistics, where often there are a few variables in each location, these spatio-temporal data sets can be very large. For instance, contamination measurements are frequently observed every day in many locations. First approaches to model spatio-temporal problems were based on separability in the space-time covariance matrix, indicating that spatial and temporal covariance models can be combined in sum (Rouhani and Hall, 1989) or product forms (De Cesare et al., 1997; De Iaco et al., 2002) to give spatio-temporal valid models. Evaluation of assumptions (stationarity, separability, symmetry) and properties (positive definite) about the space-time covariance matrix has been the most challenging problem considered recently (Cressie and Huang, 1999; Christakos, 2000; Gneiting, 2002; Gneiting et al., 2005). Links between space-time and multivariable geostatistics are discussed in (Kyriakidis and Journel, 1999).

In summary the theory about spatial random processes has been studied from many faces and proposed models for many types of data (including binary, discrete, continuous). However, it is possible to consider other experimental situations which instead of univariate, space-time or multivariable random fields, the sample is a set of curves spatially correlated. These class of data will be studied in this research.

Since the pioneer work by Deville (1974) and more recently with the work by Ramsay and Dalzell (1991) the statistical community has been interested in develop models for functional data. Functional versions for many branches of statistics have been given. Examples of such methods include exploratory and descriptive data analysis (Ramsay, 1995, 2000), linear models (Cardot et al., 1999; Malfait et al., 2000), generalized linear models (Escabias et al., 2004), analysis of variance (Cuevas et al., 2004; Delicado, 2007), non-parametric methods (Ferraty and Vieu, 2006), longitudinal data (Yao et al., 2005) or multivariable techniques like principal components (Pezulli and Silverman, 1993; Silverman, 1995), canonical correlation (Leurgans et al., 1993), discriminant analysis (Ferraty and Vieu, 2003) or cluster analysis (Clarkson et al., 2005). These procedures are
summarized in Ramsay and Silverman (1997, 2005). New methodologies which combine non-parametric modeling and functional data are proposed in Ferraty and Vieu (2006). Computational applications are shown in Clarkson et al. (2005). An overview of statistical methods for analyzing functional data is shown in Ramsay and Silverman (2005) and recent developments in this field are given in special issues of several journals (González-Manteiga and View, 2007; Valderrama, 2007). Many functional data analysis start fitting curves to observations obtained in a discrete set of points (Ramsay, 1998). Often this procedure is carried out by smoothing and non-parametric regression (Eubank, 1998; Simonoff, 1996).

The standard statistical techniques for modeling functional data are focused on independent functions. However, in several disciplines of applied sciences there exists an increasing interest in modeling correlated functional data: this is the case when samples of functions are observed over a discrete set of time points (temporally correlated functional data) or when these functions are observed in different sites of a region (spatially correlated functional data). In these cases the above-mentioned methodologies may not be appropriate as they do not incorporate dependence among functions into the analysis. For this reason, some statistical methods for modeling correlated variables, such as time series (Box and Jenkins, 1976) or geostatistical analysis (Cressie, 1993), have been adapted to the functional context.

An example of modeling temporally correlated functional data is shown in Ruiz-Medina et al. (2007). These authors consider an autoregressive Hilbertian model of order one to represent the dynamic of a sequence of functional data. For spatially correlated functional data, Yamanishi and Tanaka (2003) develop a regression model that enables the relation among variables over time and space to be studied, combining both geographically weighted regression (Brunsdon et al., 1998) and functional multiple regression (Ramsay and Silverman, 2005). Baladandayuthapani et al. (2008) show an alternative for analyzing an experimental design with a spatially correlated functional response. They both use a hierarchical model and a Bayesian approach. Contributions of Yamanishi and Tanaka (2003) and Baladandayuthapani et al. (2008) make it possible to include spatial dependence among curves into the standard functional analysis, such as functional multiple regression and functional analysis of variance.
In the same way that standard statistical methods have been generalized to be used in FDA, it is possible to think that geostatistical methods can be adapted to these types of data in order to do spatial prediction of curves. This topic has been considered from several points of view. Goulard and Voltz (1993) is a pioneer work in the spatial prediction of curves setting. They propose three geostatistical approaches to predict curves: a curve kriging approach and two multivariate approaches based on cokriging on either discrete data or coefficients of the parametric models that have been fitted to the observed curves. In Chapters 3, 4 and 5, we develop some alternative methodologies in this matter.

In Chapter 3 it is given a non-parametric approach to solve the first approach considered by Goulard and Voltz (1993). The predictor in the first proposal of Goulard and Voltz (1993) as well as that considered in Chapter 3 has the same form as the classical ordinary kriging predictor (Cressie, 1993) but considering curves instead of one-dimensional data, that is, each curve is weighted by an scalar parameter.

In Chapter 4, we solve the problem of spatial prediction of functional data by weighting each observed curve by a functional parameter. This approach is a combination of ordinary kriging and the functional linear concurrent (point-wise) model such as shown in Chapter 14 of Ramsay and Silverman (2005). This approach was mentioned in Goulard and Voltz (1993) but was not developed there.

In Chapter 5, we propose a cokriging predictor for doing univariate prediction (as in the cokriging multivariable sense), but considering as auxiliary information samples of curves instead of observations of random vectors. Likewise, we extend the multivariable kriging from random vectors to the functional context by defining a functional kriging predictor which allows to do prediction of a whole curve at an unvisited site by using as information the curves sampled in nearby sites to the prediction site. The solution considered in Chapter 5 is also studied by Nerini and Monestiez (2008). They propose

\footnote{The main results of Chapters 3, 4 and 5 have generated three papers: respectively, Giraldo et al. (2007) (in third revision in Environmental and Ecological Statistics), Giraldo et al. (2008) (accepted for publication in Journal of Agricultural, Biological and Environmental Statistics), Giraldo et al. (2009) (submitted). Additionally, Delicado et al. (2009) (accepted for publication in Environmetrics) offer a review on spatial statistics for functional data.}
a solution based on orthonormal basis functions. In our approach orthogonality is not a required condition.

As an illustration, the methodologies proposed in Chapters 3 to 5 are applied to three real data sets. In next Section we describe them.

1.1 Data sets

In applied sciences, it is common that data have both spatial and functional components. In agronomy, for instance, previous to the crop, measures of penetration resistance are taken in a sampling grid of the study area Chan et al. (2006). In this case, and though penetration resistance is measured only in some depths, it is possible to consider it as a functional variable after a smoothing or interpolation process have been applied. Other examples in environmental sciences are given when water samples are taken in different sites of an aquatic ecosystem and curves of conductivity are observed or when daily cycles of temperature or oxygen are measured in different points of a study zone (Mancera and Vidal, 1994). Likewise in meteorology when curves of temperature or precipitation are obtained in several weather stations of a country (Ramsay and Silverman, 2005) or when solar radiation is monitored in both space and time over a region, and smoothing methods are used to fit each time series (Bodas-Salcedo et al., 2003). For applying the methods proposed in the thesis we choose three real data sets having spatial and functional components. Now we show a detailed description of these data sets.

1.1.1 Penetration resistance curves

In agronomy, it is usual to measure penetration resistance in a region before sowing (Chan et al., 2006). Penetration resistance is an empirical measure of soil strength that can rapidly identify areas where soil depth or soil compaction may be limiting yields (Castrignano et al., 2002). The soil mechanical resistance to penetration shows high influence on vegetal development since the growth of the roots and the crops productivity change in the inversely proportional form with its value (Freddi et al., 2006). Knowing its spa-
Figure 1.1: Observed penetration resistance curves (right panel) on 32 sampling points (left panel). Data collected during 2004 at Marengo Experimental Station (National University of Colombia).

tial variability provides possibilities for site-specific soil treatments that can increase profitability and sustainability of crop production. Determination of soil resistance at different depths is also necessary to establish proper management strategies. In other words the goal of measuring penetration resistance is assessing soil strength for carrying out precision agriculture. In this work we use some soil penetration resistance profiles (Figure 1.1, right panel) obtained on a regular georreferenced grid with 32 sampling points over an agricultural farm at the National University of Colombia (Figure 1.1, left panel). The study area is situated at 4°, 42N, 74°, 12W. For each sampling point, 334 observations of penetration resistance (MPa) were obtained on depths varying from 0 to 45 cm. The positions of the locations were measured by GPS. The data set was obtained by Galindo (2004) as part of a research project focused in precision agriculture (Leiva, 2003). This data set is studied in Chapter 3. The results obtained suggest a low level of spatial dependence. For this reason it was not not consider subsequently and we work with other data sets in Chapters 4 and 5.
1.1.2 Canadian temperatures

Spatial prediction of meteorological data is an important input for many types of models including hydrological or those of regeneration, growth, and mortality of forest ecosystems. In particular, the modeling of spatially correlated temperature data is of interest for predicting microclimate conditions in mountainous terrain, resource management, calibration of satellite sensors or for studying the “greenhouse effect”, among others. Many methods have been developed and used for doing spatial prediction of temperatures, although to the best of our knowledge all of these ignore its functional character. Here we use a well-known meteorological data set in FDA consisting of daily mean temperature measurements recorded at 35 weather stations of Canada (Ramsay and Dalzell, 1991; Ramsay and Silverman, 2005). In particular we analyze information of mean daily temperature averaged over the years 1960 to 1994 (February 29th combined with February 28th) (Figure 1.2, right panel). The data for each station were obtained from Ramsay and Silverman’s home page (http://www.functionaldata.org/). The geographical coordinates of weather stations (Figure 1.2, left panel) were obtained from the Meteorological Service of Canada (http://www.climate.weatheroffice.ec.gc.ca/climateData/). This data are analyzed in Chapter 4. In Chapter 5 some summary statistics of cross-validation residuals obtained with this information are also considered for doing comparison among the approaches given in the thesis.

1.1.3 Canada’s Maritime provinces temperatures

The Maritime Provinces cover a region of Canada consisting of three provinces: Nova Scotia (NS), New Brunswick (NB), and Prince Edward Island (PEI). They occupy just over 1% of Canada’s land surface. We provide an applied context for our proposal by using a data set consisting of temperature measurements recorded at 35 weather stations located in these provinces (Figure 1.3, left panel). Given their locations on Canada’s Atlantic coast, the Maritime Provinces have a cool, temperate climate: cold continental air masses from the northwest alternate with warmer, humid maritime air from the southwest Stanley (2002). Their temperature is characterized by cool summers and mild winters, with a much
Figure 1.2: Averages (over 30 years) of mean daily temperature curves (right panel) observed at 35 Canadian weather stations (left panel). Red dot corresponds to Slave Lake, an unsampled station.

Figure 1.3: Averages (over 30 years) of daily temperature curves (right panel) observed at 35 Canadian Maritime weather stations (left panel). Red dot corresponds to Moncton, an unsampled location.
smaller annual temperature range than that recorded in other Canadian regions. As in the
data set described in Section 1.1.2, this data set contains information of daily temperatures
averaged over the years 1960 to 1994 (February 29th combined with February 28th) (Figure
1.3, right panel). The data for each station were obtained from the Meteorological Service
of Canada (http://www.climate.weatheroffice.ec.gc.ca/climateData/). The geographical
coordinates in decimal degrees of the weather stations (Figure 1.3, left panel) were obtained
from the database of geographic coordinate information (http://www.tageo.com).

This data set has the same structure as the data set introduced in Section 1.1.2, but
it covers a much smaller and more homogeneous area. Thus the effects of longitude,
latitude, coastal and inland climates are negligible. We use this information for showing
applications of the methods proposed in Chapters 4 and 5. Summary statistics of cross-
validation analysis obtained with this data set are used to make a comparison among the
approaches given in this research.
Chapter 2

Geostatistics and functional data analysis

2.1 Geostatistics

Geostatistics solves the problem of spatial prediction in a region with spatial continuity. In recent years there has been a tremendous growth of theoretical geostatistical models including univariate, multivariable and space-time geostatistics. In this section we will discuss the basic theory about each one. The theory about model-based geostatistics shown at the end of Section 2.1.1 and the space-time geostatistics given in Section 2.1.3 is not used subsequently. However we include these topics in the thesis for completeness in the presentation of geostatistical methods. Geostatistics is usually concerned with spatial prediction, but there are other important areas such as, for instance, simulation and sampling network design. However, taking into account the goals of this work, we will show only the basics of spatial prediction methods. Specifically, we will describe the methodology known as kriging.
2.1.1 Univariate geostatistics

Univariate geostatistics deals with a real-valued stochastic process \( \{Z(s) : s \in D \subset \mathbb{R}^d\} \). Usually spatial index \( s \) is two or three dimensional, i.e., \( d = 1, 2, 3 \). \( \{Z(s) : s \in D \subset \mathbb{R}^d\} \) is also called random process or random field. Given \( s_1, \ldots, s_n \) the random vector \( Z(s) = (Z(s_1), \ldots, Z(s_n))^t \) is defined by its joint cumulative distribution \( F(z_1, \ldots, z_n) = P(Z(s_1) \leq z_1, \ldots, Z(s_n) \leq z_n) \). The process is stationary if \( Z(s) = (Z(s_1), \ldots, Z(s_n))^t \) and \( Z(s + h) = (Z(s_1 + h), \ldots, Z(s_n + h))^t \) have the same joint cumulative distribution, for all \( h \) (Stein, 1999). Using moments of the random variables, second-order stationarity is given by

1. \( E(Z(s)) = \mu \) for all \( s \in D \subset \mathbb{R}^d \),

2. \( \text{Cov}(Z(s_i), Z(s_j)) = E[(Z(s_i) - \mu)(Z(s_j) - \mu)] = C(h) \) for all \( s_i, s_j \in D \subset \mathbb{R}^d \) and \( h = \|s_i - s_j\| \).

The process is called intrinsically stationary if \( (Z(s) - Z(s + h)) \) is stationary, that is, if the process satisfy

1. \( E(Z(s_i) - Z(s_j)) = 0 \) for all \( s_i, s_j \in D \subset \mathbb{R}^d \),

2. \( V(Z(s_i) - Z(s_j)) = E[Z(s_i) - Z(s_j)]^2 = 2\gamma(h) \) for all \( s_i, s_j \in D \subset \mathbb{R}^d \) and \( h = \|s_i - s_j\| \).

Assuming second order stationarity \( \gamma(h) = C(0) - C(h) \). \( C(h) \) is called (auto)covariance function and it is a definite positive function. \( 2\gamma(h) \) and \( \gamma(h) \) are called (auto)variogram and (auto)semivariogram, respectively (Wackernagel, 1995). \( \gamma(h) \) is a conditionally negative definite function in the sense that \( \gamma(\cdot) \) applied to a distance matrix gives us a matrix that is negative definite (Cressie, 1993). In addition to stationarity another important property of random fields is isotropy. The random field \( \{Z(s) : s \in D \subset \mathbb{R}^d\} \) is isotropic if \( h \) affects \( C(h) \) only through its length \( |h| \), and not through its direction (Stein, 1998). Estimation of the variogram is mostly done by the method of moments estimator (Ploner...
and Dutter, 2000)

$$2 \hat{\gamma}(h) = \frac{1}{|N(h)|} \sum_{N(h)} (Z(s_i) - Z(s_j))^2,$$  \hspace{1cm} (2.1)

where the average (2.1) is taken over $N(h) = \{(s_i, s_j) : s_i - s_j = h\}$ and $|N(h)|$ is the number of distinct elements in $N(h)$. For irregularly spaced data there are generally not enough observations separated by exactly $h$ then $N(h)$ is usually modified to $\{(s_i, s_j) : s_i - s_j \in T(h)\}$, where $T(h)$ is a tolerance region of $R^d$ surrounding $h$ (Cressie, 1989).

Due to the conditional negative definiteness of the variogram it is not possible to use the experimental variogram with some smoothing interpolation to describe the spatial autocorrelation (Ploner and Dutter, 2000). A classical approach for modeling spatial dependence structure is to fit parametric models (spherical, Gaussian, exponential, for instance) (Journel and Huijbregts, 1978; Isaaks and Srivastava, 1987) to the estimator (2.1). This procedure is usually carried out by weighted least squares (Cressie, 1989). Having a fitted model, the next step is to predict the variable over unsampled points $s_0$ based on estimates weights $\lambda_i$ in the kriging predictor given by

$$\hat{Z}(s_0) = \mu(s_0) + \sum_{i=1}^{n} \lambda_i (Z(s_i) - \mu(s_0)).$$  \hspace{1cm} (2.2)

Optimal weights $\lambda_i$s in (2.2) are those that produce unbiased predictions and minimize $E(Z(s_0) - \hat{Z}(s_0))^2$, that is, the mean squared prediction error (MSPE). Let $\mu(s) = \mu$, $\forall s \subset D \in R^d$. Then (2.2) is called simple kriging if $\mu$ is known or ordinary kriging if $\mu$ is estimated. If $\mu(s) = \sum_{j=0}^{p} \beta_j X_j(s)$ with $\beta_j$ constants and $X_j, j > 0$, the explanatory variables, then (2.2) is called universal kriging. Parameters in ordinary kriging are obtained by solving a set of simultaneous equations $\Gamma \lambda = \gamma$

\[
\begin{pmatrix}
\gamma(s_1 - s_1) & \gamma(s_1 - s_2) & \cdots & \gamma(s_1 - s_n) & 1 \\
\gamma(s_2 - s_1) & \gamma(s_2 - s_2) & \cdots & \gamma(s_2 - s_n) & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
\gamma(s_n - s_1) & \gamma(s_n - s_2) & \cdots & \gamma(s_n - s_n) & 1 \\
1 & 1 & \cdots & 1 & 0
\end{pmatrix}
\begin{pmatrix}
\lambda_1 \\
\lambda_2 \\
\vdots \\
\lambda_n \\
m
\end{pmatrix}
= \begin{pmatrix}
\gamma(s_0 - s_1) \\
\gamma(s_0 - s_2) \\
\vdots \\
\gamma(s_0 - s_n) \\
1
\end{pmatrix},
\]  \hspace{1cm} (2.3)
2.1. GEOSTATISTICS

where \( \gamma(s_i - s_j) \) is the semivariance function between \( Z(s_i) \) and \( Z(s_j) \), \( m \) is the Lagrange multiplier used to take into account the unbiasedness restriction of the predictor (2.2) given by \( \sum_{i=1}^{n} \lambda_i = 1 \). In simple kriging there is not restriction about \( \lambda \)'s and then last row in the system (2.3) is not included. In universal kriging the number of restrictions to get unbiasedness depend of the variables used to fit the mean. In this case the system (2.3) includes additional rows corresponding to each restriction. In practice plug-in methods are used to solve (2.3) and consequently \( \gamma(s_i - s_j) \) is estimated by the fitted model to the experimental variogram (2.1), where \( h \) is the Euclidean distance between \( s_i \) and \( s_j \). The minimized squared prediction error for ordinary kriging is given by

\[
\sigma_k^2 = \sum_{i=1}^{n} \lambda_i \gamma(s_0 - s_i) + m. \tag{2.4}
\]

In simple kriging last expression does not include \( m \) and in universal kriging \( m \) is replaced by \( \sum_{j=0}^{p} m_j X_j(s_0) \).

In classical geostatistical analysis often no explicit underlying stochastic model is declared. On the other hand, model-based approach (Diggle et al., 1998; Diggle and Ribeiro, 2000) starts with an explicit stochastic model and derives methods of parameter estimation and interpolation by the application of general statistical principles. The authors, assuming a stationary Gaussian multivariate process \( Z(s) \equiv (Z(s_1), \cdots, Z(s_n)) \), show that \( Z(s_i) \) can be written as \( Z(s_i) = \mu + S(s_i) + W(s_i) \), where \( \mu \) is the constant mean required by stationarity, \( S(s) \sim N(MV(0, \sigma^2 \rho)) \), \( \text{Cov}(S(s_i), S(s_j)) = \sigma^2 \rho(s_i - s_j) \), and \( W(s) \sim N(MV(0, \tau^2 I)) \). Conditionally on \( S(s_i) \), \( Z(s_i) \) are mutually independent with \( Z(s_i) \mid S(s_i) \sim N(\mu + S(s_i), \tau^2) \). \( Z(s) \) is the observable process, \( S(s) \) is called signal and \( W(s) \) is a white-noise. Diggle and Ribeiro (2000) show that \( (S(s_0), Z(s))_{(n+1) \times 1} \sim N(MV(\mu, V)) \). If we define \( r = (\rho(s_0 - s_1), \cdots, \rho(s_0 - s_1))^t \), matrix \( V \) is given by:

\[
V = \begin{pmatrix}
\sigma^2 & (\sigma^2 r)^t_{1 \times n} \\
(\sigma^2 r)_{n \times 1} & (\sigma^2 \rho + \tau^2 I)_{n \times n}
\end{pmatrix}.
\tag{2.4}
\]

Using basic theory of multivariate Gaussian distribution we obtain that the conditional distribution is

\[
(S(s_0) \mid Z(s))_{n \times 1} \sim N_1 (E(S(s_0) \mid Z(s)), V(S(s_0) \mid Z(s))),
\]
where

\[
E(S(s_0)|Z(s)) = \mu + \sigma^2 r^t(\sigma^2 \rho + \tau^2 \mathbf{I})^{-1}(Z(s) - \mu),
\]

(2.5)

and

\[
V(S(s_0)|Z(s)) = \sigma^2 - \sigma^2 r^t(\sigma^2 \rho + \tau^2 \mathbf{I})^{-1}\sigma^2 r.
\]

Assuming square loss, the optimal predictor is the conditional expectation. Consequently the best predictor \( \hat{Z}(s_0) \) is given by \( E(S(s_0)|Z(s)) \) in (2.5). In (2.2) \( \lambda = (\lambda_1, \cdots, \lambda_n) \) is estimated solving \( \lambda = \gamma \Gamma^{-1} \) in (2.3). Using \( \lambda = \sigma^2 r^t(\sigma^2 \rho + \tau^2 \mathbf{I})^{-1} \) in (2.5) both equations are equivalent. However it is better to assume a parametric model because in addition to weight least squares estimation other methods, as maximum likelihood, can be used (Stein, 1999). Generalization to other distributions are given in Diggle et al. (1998).

2.1.2 Multivariable geostatistics

A generalization of univariate geostatistics is obtained when instead of one stochastic process we consider \( m \) distinct random fields. In this section we show in particular the basics of cokriging (Myers, 1982; Bogaert, 1996) and multivariable spatial prediction (Ver Hoef and Cressie, 1993; Ver Hoef and Barry, 1998). Let \( \{Z(s) : s \in D\} \) be a multivariable spatial vector of \( m \) random process \( Z_1(s), \cdots, Z_m(s) \) on a spatial region \( D \subset \mathbb{R}^d \). We consider the model

\[
Z(s) = \mu(s) + \epsilon(s),
\]

(2.6)

where \( \mu(s) \) is a mean vector and \( \epsilon(s) \) a random vector with \( E(\epsilon(s)) = 0 \). We assume that the \( m \) processes are stationary, that is, the mean vector is assumed constant for all \( s \in D \), and the covariance and variogram functions depends only on the separation vector \( h \) and not on the location \( s \). We use the following notation:

- \( 2\gamma_{lq}(s_i, s_j) = V(Z_l(s_i) - Z_q(s_j)) \)
- \( \gamma_{lq}^T = (\gamma_{lq}(s_1, s_0), \cdots, \gamma_{lq}(s_n, s_0)) \)
2.1 GEOSTATISTICS

- \( \mathbf{\Gamma}_{lk} = \begin{pmatrix}
  \gamma_{lq}(s_1, s_1) & \cdots & \gamma_{lq}(s_1, s_n) \\
  \vdots & \ddots & \vdots \\
  \gamma_{lq}(s_n, s_1) & \cdots & \gamma_{lq}(s_n, s_n)
\end{pmatrix} \)

The cokriging predictor of the \( k \)-th variable, \( k = 1, \ldots, m \), in the location \( s_0 \) is given by

\[
\hat{Z}_k(s_0) = \sum_{j=1}^{m} \lambda_{1j} Z_j(s_1) + \cdots + \sum_{j=1}^{m} \lambda_{nj} Z_j(s_n) \\
= \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_{ij} Z_j(s_i). \tag{2.7}
\]

The predictor (2.7) is unbiased if \( \sum_{i=1}^{n} \lambda_{ik} = 1 \) and \( \sum_{i=1}^{n} \lambda_{ij} = 0 \) for all \( j \neq k, j = 1, \ldots, m \). A variant of cokriging uses only the condition \( \sum_{i=1}^{n} \sum_{j=1}^{m} \lambda_{ij} = 1 \) (Isaaks and Srivastava, 1987). Using the method of Lagrange multipliers to minimize mean squared prediction error subject to the unbiasedness constraints gives the cokriging system of equations expressed in matrix notation by

\[
\mathbf{C} \mathbf{\lambda} = \mathbf{c}, \tag{2.8}
\]

with

\[
\mathbf{C} = \begin{pmatrix}
  \mathbf{\Gamma}_{11} & \cdots & \mathbf{\Gamma}_{1k} & \cdots & \mathbf{\Gamma}_{1m} & 1 & \cdots & 0 & \cdots & 0 \\
  \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  \mathbf{\Gamma}_{m1} & \cdots & \mathbf{\Gamma}_{mk} & \cdots & \mathbf{\Gamma}_{mm} & 0 & \cdots & 0 & \cdots & 1 \\
  \mathbf{1}^T & \cdots & \mathbf{0}^T & \cdots & \mathbf{0}^T & 0 & \cdots & 0 & \cdots & 0 \\
  \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  \mathbf{0}^T & \cdots & \mathbf{1}^T & \cdots & \mathbf{0}^T & 0 & \cdots & 0 & \cdots & 1 \\
  \vdots & \ddots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\
  \mathbf{0}^T & \cdots & \mathbf{0}^T & \cdots & \mathbf{1}^T & 0 & \cdots & 0 & \cdots & 0 \\
\end{pmatrix} = \begin{pmatrix}
  \mathbf{\Gamma} \\
  \mathbf{X} \\
  \mathbf{X}^T \\
  \mathbf{0}
\end{pmatrix}, \tag{2.9}
\]
\[ \lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_k \\ \vdots \\ \lambda_m \\ \delta_1 \\ \vdots \\ \delta_k \\ \vdots \\ \delta_m \end{pmatrix}, \quad c = \begin{pmatrix} \gamma_{1k} \\ \vdots \\ \gamma_{kk} \\ \vdots \\ \gamma_{mk} \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \]

where \( \lambda_j = (\lambda_{1j}, \ldots, \lambda_{nj}) \).

In multivariable spatial prediction (Ver Hoef and Cressie, 1993; Ver Hoef and Barry, 1998) all the \( m \) variables are predicted simultaneously at \( s_0 \). In this case the kriging predictor is given by

\[
\begin{pmatrix}
\hat{Z}_1(s_0) \\
\vdots \\
\hat{Z}_m(s_0)
\end{pmatrix} = \begin{pmatrix}
\lambda_{11} & \cdots & \lambda_{1m} & \cdots & \lambda_{n1} & \cdots & \lambda_{nm}
\end{pmatrix} \begin{pmatrix}
Z_1(s_1) \\
\vdots \\
Z_m(s_n)
\end{pmatrix},
\tag{2.10}
\]

and the matrix of parameters is obtained by solving the system (Ver Hoef and Barry, 1998)

\[
\begin{pmatrix}
\Gamma & X \\
X^T & 0
\end{pmatrix} \begin{pmatrix}
\Lambda \\
\Delta
\end{pmatrix} = \begin{pmatrix}
G \\
I
\end{pmatrix},
\tag{2.11}
\]
where $\Gamma$ and $X$ are defined as in (2.9), $\Lambda$ is the matrix of parameters, $\Delta$ is a diagonal matrix of Lagrange multipliers, $I$ is an identity matrix and

$$G = \begin{pmatrix}
\gamma_{11} & \gamma_{12} & \cdots & \gamma_{1m} \\
\gamma_{21} & \gamma_{22} & \cdots & \gamma_{21} \\
\vdots & \ddots & \ddots & \vdots \\
\gamma_{m1} & \gamma_{m2} & \cdots & \gamma_{mm}
\end{pmatrix}.$$  

Cokriging could be used for predicting simultaneously all $m$ variables by cokriging each variable, one at a time. The cokriging prediction for one variable at a time is identical to the prediction of that same variable obtained by multivariable spatial prediction (Ver Hoef and Cressie, 1993). The difference between these approaches is given by their predictions variances. With cokriging analysis we obtain a prediction variance at each time. In multivariable spatial prediction, in addition to the prediction variances, it is possible to estimate a multidimensional prediction region with its long axis oriented towards regions where the predicted variables tend to covary (Ver Hoef and Cressie, 1993).

### 2.1.3 Space-time geostatistics

Spatio-temporal geostatistics provide a probabilistic framework for data analysis and predictions that builds on the joint spatial and temporal dependence between observations (Kyriakidis and Journel, 1999). The analysis can be focused on spatial interpolation over specific time instants. In this case the objective might be to compare the various maps over time. Moreover the analysis can be focused on modeling multiple time series where each spatial location is associated with a distinct time series. Now theory focused on geostatistical prediction including additional time dimension is shown. Let $\{Z(s, t) : S \subset D \subset R^d, t \in [0, \infty)\}$ denote a spatio-temporal stochastic process observed at $N$ space-time coordinates $(s_1, t_1), \cdots, (s_N, t_N)$. Optimal prediction (in space and time) of the unobserved parts of the process, based on $Z \equiv (Z(s_1, t_1), \cdots, Z(s_N, t_N))^t$, is the goal Cressie and Huang (1999). Analogous to (2.2) and (2.7) spatio-temporal kriging is given by
An example of a separable model is
\[ \hat{Z}(s_0, t_0) = \mu(s_0, t_0) + \sum_{i=1}^{n} \lambda_i(Z(s_i, t_i) - \mu(s_i, t_i)), \]
(2.12)
or in matrix notation
\[ \hat{Z}(s_0, t_0) = \mu(s_0, t_0) + cC^{-1}(Z - \mu), \]
(2.13)
where \( C = \text{Cov}(Z), c = \text{Cov}(Z(s_0, t_0), Z), \) and \( \mu = E(Z). \)

The process is stationary if \( \text{Cov}(Z(s_i, t_k), Z(s_j, t_l)) = C(s_i - s_j, t_k - t_l) = C(h, u), \)
hence \( C(h, u) = C(h + s_i - s_j, u + t_k - t_l) = C(h, u), \) and \( \mu = E(Z). \) The process is said to be isotropic if \( \text{Cov}(h, u) = c(||h||, |u|), \) that is, the covariance function depends upon the separation vectors only through their lengths. Predictor (2.13) will be valid if the function \( C \) satisfy a positive definiteness condition (Cressie and Huang, 1999), that is, for any \( (r_1, q_1), \cdots, (r_m, q_m), \) any real \( a_1, \cdots, a_m \) and any integer \( m, \sum_{i=1}^{m} \sum_{j=1}^{m} a_i a_j C(r_i - r_j, q_i - q_j) \geq 0. \) The random field \( Z \) is said to have separable covariance if there exist purely spatial and purely temporal covariance functions \( C_S(Z(s_i), Z(s_j)) \) and \( C_T(Z(t_k), Z(t_l)) \), respectively, such that (Gneiting et al., 2005)
\[
\text{Cov}(Z(s_i, t_k), Z(s_j, t_l)) = C(h, u) = C_S(Z(s_i), Z(s_j)) \cdot C_T(Z(t_k), Z(t_l)) = C_S(h) \cdot C_T(u)
\]
An example of a separable model is \( C_S(h) = \exp(-\theta_1||h||); \theta_1 > 0 \) and \( C_T(u) = \exp(-\theta_2|u|); \theta_2 > 0, \) and hence \( C(h, u) = \exp(-\theta_1||h|| - \theta_2|u|) \) Cressie and Huang (1999).

Another type of separability involves adding spatial and temporal covariances \( C(h, u) = C_S(h) + C_T(u) \) (Rouhani and Hall, 1989). For this model \( C(h, u) \) can be singular (Rouhani and Myers, 1990). The process \( Z \) has fully symmetric covariance if \( \text{Cov}(Z(s_i, t_k), Z(s_j, t_l)) = \text{Cov}(Z(s_i, t_l), Z(s_j, t_k)) \) for all \( (s_i, t_k), (s_j, t_l) \in D \times T. \) Covariance structures that are not fully symmetric are nonseparable (Gneiting et al., 2005).

There has been an increasing development of methods to build nonseparable covariance functions. An example is \( C(h, u) = (1 + u)^{-1} \exp(h/(1 + u)^{3/2}) \) (Gneiting, 2002). Other important references about separability, stationarity and symmetry are De Cesare et al.
Once established the covariance structure (separable or nonseparable) it is fitted a continuous two-dimensional model to the experimental spatio-temporal model given by

$$
\hat{\gamma}(h, u) = \frac{1}{2|N(h, u)|} \sum_{N(h,u)} (Z(s_i, t_k) - Z(s_j, t_l))^2,
$$

(2.14)

where $h = \|s_i - s_j\|$, $u = |t_k - t_l|$, $N(h,u) = \{(s_i, t_k), (s_j, t_l) : h = \|s_i - s_j\|, u = |t_k - t_l|\}$ and $|N(h,u)|$ is the number of distinct elements in $N(h,u)$. We can carry out prediction by finding $\lambda_i$ in (2.12) by solving the system $C\lambda = c$, where

$$
C = \begin{pmatrix}
\hat{C}(s_1 - s_1, t_1 - t_1) & \cdots & \hat{C}(s_1 - s_n, t_1 - t_n) \\
\vdots & \ddots & \vdots \\
\hat{C}(s_n - s_1, t_n - t_1) & \cdots & \hat{C}(s_n - s_n, t_n - t_n)
\end{pmatrix},
$$

$$\gamma = \begin{pmatrix}
\lambda_1 \\
\vdots \\
\lambda_n
\end{pmatrix},
$$

$$
c = \begin{pmatrix}
\hat{C}(s_0 - s_1, t_0 - t_1) \\
\vdots \\
\hat{C}(s_0 - s_n, t_0 - t_n)
\end{pmatrix},
$$

and $\hat{C}(s_i - s_j, t_k - t_l) = \hat{C}(h, u) = \hat{\gamma}(0,0) - \gamma(h,u)$.

### 2.2 Functional data analysis

Most statistical analysis involve one or more observations taken from a number of individuals in a sample, with the aim of making inferences about the general population from which the sample is drawn. In an increasing number of fields, these observations are curves or images. Curves and images are examples of functions, since an observed intensity is available at each point on a line segment, a portion of a plane or a volume.
FDA is a general way of thinking, where the basic unit of information is an entire observed function rather than a string of numbers. The \( i \)-th observation is a real function, \( x_i(t), i = 1, \ldots, n, t \in T \), where \( T \) is a real interval (\( T \subseteq \mathbb{R}^2 \) for images), and therefore each \( x_i \) is a point in some function space \( H \) (Ramsay and Dalzell, 1991). In the real context an observation can be expressed by the random family \( \{X(t_j)\}_{j=1,\ldots,J} \). On the other hand in functional analysis we can consider that the data are an observation of the continuous family \( \chi = \{X(t); t \in (t_{\text{min}}, t_{\text{max}})\} \) (Ferraty and Vieu, 2006). Two important definitions for establishing differences between real and functional context are (Ferraty and Vieu, 2006):

**Definition 2.1.** A random variable \( \chi \) is called functional variable (f.v.) if it takes values in an infinite dimensional space (or functional space). An observation \( \chi \) of \( \chi \) is called a functional data.

**Definition 2.2.** A functional data set \( \chi_1, \ldots, \chi_n \) is the observation of \( n \) functional variables \( \chi_1, \ldots, \chi_n \) identically distributed as \( \chi \).

In FDA the choice of the function space is crucial. Throughout the thesis we suppose that \( f(t) \) are valued in the separable Hilbert space of square integrable functions

\[
L_2(T) = \{f : T \to \mathbb{R}, \text{ such that } \int_T f(t)^2 dt < \infty\}.
\]

Note that \( L_2(T) \) with the inner product \( \langle f, g \rangle = \int_T f(t)g(t) dt \) defines a Euclidean space. Also note that we are implicitly working with a quotient space given by the equivalence relation \( f \sim g \iff \int_T (f(t) - g(t))^2 dt = 0 \).

### 2.2.1 Notation

In the rest of the thesis we use indistinctly \( \chi, \chi(t), X(t), Y(t) \) and \( Z(t) \) for denoting functional variables and \( x, \chi(t), x(t), y(t) \) and \( z(t) \) for observations of functional variables. In addition \( \mathbf{X}(t), \mathbf{Y}(t), \mathbf{B}(t), \mathbf{\theta}(t), \mathbf{\psi}(t) \) and \( \mathbf{\eta}(t) \) are used for denoting vectors whose components are functional variables, observations of functional variables or known functions.
(for instance functions of a basis). The context should make clear when the symbol refers
to each one of these.

Preliminary definitions about classical measures of size and distance between vectors
in \( \mathbb{R}^p \) are given in the functional context. Let us assume that \( X(t), Y(t), Z(t), t \in T \), are
functions in some function space, then expressions for inner product and norm are:

**Inner product**

\[
\langle X(t), Y(t) \rangle = \int_T X(t)Y(t)dt.
\]  (2.15)

**Properties**

1. Symmetry \( \langle X(t), Y(t) \rangle = \langle Y(t), X(t) \rangle \).
2. Positivity \( \langle X(t), X(t) \rangle \geq 0 \).
3. Bilinearity \( \langle aX(t) + bY(t), Z(t) \rangle = a\langle X(t), Z(t) \rangle + b\langle Y(t), Z(t) \rangle \), for all real numbers \( a, b \).

**Norm**

\[
\| X(t) \| = \sqrt{\| X(t) \|^2},
\]  (2.16)

where

\[
\| X(t) \|^2 = \langle X(t), X(t) \rangle = \int_T X(t)X(t)dt.
\]

**Properties**

1. \( \| X(t) \| \geq 0 \).
2. \( \| aX(t) \| = |a| \| X(t) \| \), for all real numbers \( a \).
3. \( \| X(t) + Y(t) \| \leq \| X(t) \| + \| Y(t) \| \).
4. \( |\langle X(t), Y(t) \rangle| \leq \| X(t) \| \| Y(t) \| = \sqrt{\langle X(t), X(t) \rangle \langle Y(t), Y(t) \rangle} \).
5. $-1 \leq \langle X(t), Y(t) \rangle / (\|X(t)\| \|Y(t)\|) \leq 1$.

A functional data $\chi_i(t), t \in T$, is usually represented as a finite set of pairs $(t_j, y_{ij}), t_j \in T, j = 1, \ldots, M$ and $y_{ij} = \chi_i(t_j)$ (if there is no observational noise) or $y_{ij} = \chi_i(t_j) + \varepsilon_j$ (if noise is present), $\varepsilon_j$ having zero mean. The set of points $\{t_j\}_{j=1}^M \subset T$ can be considered the same for all the functions in a functional data set, and usually they form a fine evenly spaced grid in $T$. Nowadays the number $M$ of observed values for functions $\chi_i$ is usually in the order of several hundreds or thousands in real applications. Interpolation methods (if there is no observation noise) or non-parametric smoothing methods (in the opposite case) are commonly used to represent the discrete sets $(t_j, y_{ij}), j = 1, \ldots, M$, as a real function $\chi_i$. In this sense we can say that functional data analysis inherits methodology from non-parametric functional estimation. Note that, actually, the curves are not observed, instead only points on the curves are observed. However, when the number of points in a curve is dense for simplicity we talk about “observed or measured curves”.

### 2.2.2 From data to functions by using basis functions

A basis function system is a set of known functions $B_K$ that are mathematically independent of each other and that have the property that we can approximate arbitrarily well any function by a linear combination of a sufficiently large number $K$ of these functions (Ramsay and Silverman, 2005). Basis functions procedures approximate a function $\chi$ by using a fixed truncated basis expansion

$$\chi(t) = \sum_{l=1}^{K} c_l B_l(t) = c^T B(t) \quad (2.17)$$

in terms of $K$ known basis functions (Ramsay and Silverman, 2005). In this research we perform smoothing using B-splines, Fourier and wavelets basis functions, but any other basis functions or non-parametric regression method could also be used. Good references for these methods can be found in Chapter 5 of Wasserman (2006), Chapters 3 and 5 of (Ramsay and Silverman, 2005), the ones we follow here, or in the book of Green and Silverman (1997). Once the representation by basis functions is adopted, three types of inquires need to be answered for computational issues (Lee, 2004).
2.2. FUNCTIONAL DATA ANALYSIS

- Which basis functions are appropriate.
- How many basis functions are selected to describe the sample paths.
- How the coefficients $c$ are determined based on partially observed functions.

In Chapters 3 to 5 of Ramsay and Silverman (2005) are given guidelines for solving these inquires. In the particular case of spatially correlated functional data some additional considerations can be made. In Chapter 3 we define a criterium useful for choosing the number of basis functions. Least squares is the standard method for estimating the vector of coefficients $c$, however a roughness penalty can be included in the minimization problem. Now, we show the basics of B-splines, Fourier and wavelets basis functions.

**B-splines basis functions**

Given a set of $L$ interior points of $T = [a, b]$, say $a < \tau_1 < \ldots < \tau_L < b$ (we also define $\tau_0 = a$ and $\tau_{L+1} = b$), a cubic spline $S$ is a function defined on $T$ such that $S$ is a cubic polynomial in $[\tau_{l-1}, \tau_l]$, $l = 1, \ldots, L + 1$ and $S$ has continuous second derivative in $T$ (in particular, $S$, $S'$ and $S''$ are continuous at all $\tau_l$). The points $\tau_l$, $l = 0, \ldots, L + 1$, are called knots. It can be proved that the set of cubic splines with knots $\tau_l$, $l = 0, \ldots, L + 1$, is a vector space with dimension $L + 4$. Cubic splines are very flexible and are commonly used to approximate unknown functions. In fact any set of points $(\tau_l, f_l)$, $l = 0, \ldots, L + 1$, can be interpolated using a cubic spline. Flexibility is gained when the number of knots is increased.

A useful system of basis functions for the set of cubic splines with knots $\tau_l$, $l = 0, \ldots, L + 1$ is the set of cubic B-splines $B_k(t)$, $k = 1, \ldots, L + 4$. They have an interesting property from the computational point of view: they are nonzero in no more that 4 inter-knots intervals. See Section 3.5 of Ramsay and Silverman (2005) for details on the shape of the cubic B-splines and other properties. Any cubic spline $S$ with knots $\tau_l$, $l = 0, \ldots, L + 1$ can be expressed as

$$S(t) = \sum_{k=1}^{L+4} c_k B_k(t) = c^T B(t),$$
where \( \mathbf{c} \) is the vector of coefficients \( c_k \) and \( \mathbf{B}(t) \) is a \((L + 4)\)-dimensional function with components \( B_k(t) \). Therefore working with cubic splines is equivalent to working with the \((L + 4)\)-dimensional vectors \( \mathbf{c} \) of coefficients. In Figure 2.1 we show a B-splines basis with \( L = 10 \). This is posteriorly used in Section 3.4.2 for smoothing the penetration resistance data shown is Section 1.1. Let us assume that a function \( \chi(t) \) defined on \( T \) has been observed at points \( t_1, \ldots, t_M \in T \), possibly with an observational error: the values \( y_j = \chi(t_j) + \epsilon_j \), where \( \epsilon_j \) are independent observations of a random variable with zero mean. The way we approximate function \( \chi(t) \) by a cubic spline is by solving the problem

\[
\min_{\mathbf{c} \in \mathbb{R}^{L+4}} \sum_{j=1}^{M} (y_j - S(t_j))^2 + \eta \int_T (S''(t))^2 \, dt.
\] (2.18)

Minimizing only the first term (fixing \( \eta = 0 \)) leads to splines \( S \) that fit well the observed data, but producing too much variable fitting splines. In the extreme case of choosing \( t_1, \ldots, t_M \in T \) as interior knots, the resulting cubic spline would interpolate the observations \( (t_j, y_j) \), and this has no sense because observational noise could be present. This is
the reason why a second term appears in (2.18): the integral of the square of the second derivative of \( S \) is a good measure of its total curvature, so it is a roughness penalty to too much wiggly splines candidates.

The parameter \( \eta \) is a smoothing parameter that controls the trade-off between the fit to the observed data and the smoothness of the approximating cubic spline. If \( \eta \) goes to infinity, the spline solution of (2.18) would approach the least square regression line. \( \eta \) is also known as roughness penalty parameter. In our approach there are additional parameters that affect the smoothing properties of the cubic spline approximation: the number and location of the interior knots. In the present work we are always using evenly spaced interior knots in \( T \). Therefore we work with two smoothing parameters: \( \eta \) and \( L \). The degree of smoothing is an increasing function of \( \eta \) and a decreasing function of \( L \).

The cubic spline found as the solution of (2.18) is one example of non-parametric regression estimators of the unknown function \( \chi(t) \). The definition of all these estimators involves one (or more) smoothing parameter that has to be fixed to obtain the actual

---

Figure 2.2: Fourier basis with K=5 and period 365
estimations of $\chi(t)$. The smoothing parameter choice is the most tricky step in non-parametric regression. When the objective of the estimation is to predict the value of the function $\chi(t)$ at a non-sampled value $t \in T$, non-parametric cross-validation (NPCV) is a useful way for choosing the smoothing parameter. In the case of cubic splines with $L$ interior knots, NPCV works as follows. For $j = 1, \ldots, M$, let $S_{L,\eta}^{(j)}$ be the solution of the problem (2.18) when the observation $(t_j, y_j)$ has been left out of the sample and the parameter values $L$ and $\eta$ are used to fit the rest of the data. We define

$$N_{PCV}(L, \eta) = \sum_{j=1}^{M} (y_j - S_{L,\eta}^{(j)}(t))^2.$$  

(2.19)

Then the smoothing parameters $(L, \eta)$ are chosen to minimize $N_{PCV}(L, \eta)$.

With huge quantity data the computation of the NPCV is quite long. An alternative is consider generalized cross-validation (GCV) instead of CV. Given the values of the basis functions at the $M$ sampling points and by using generalized least squares the coefficients in equation (2.18) are estimated by

$$\hat{c} = (\Phi^T G \Phi + \eta R)^{-1} \Phi^T G^T y,$$  

(2.20)

with $R = \int_T S^{''}(t) S^{''T}(t) dt$. The vector of estimated values is given by

$$\hat{y} = \Phi (\Phi^T G \Phi + \eta R)^{-1} \Phi^T G^T y$$  

$$= Hy.$$  

(2.21)

The non-parametric generalized cross-validation criterium (NPGCV) is expressed as

$$N_{PGCV}(\eta) = \frac{M^{-1} \sum_{j=1}^{M} (y_j - S(t_j))^2}{M^{-1} \text{trace}(I-H)}.$$  

(2.22)

Fourier basis functions

An function could be represented by a series of sines and cosines by means of

$$\chi(t) = c_0 + c_1 \sin \omega t + c_2 \cos \omega t + c_2 \sin 2\omega t + c_4 \cos 2\omega t + \cdots, t \in T$$  

(2.23)
which is defined by the basis $B_0(t) = 1, B_{2r-1}(t) = \sin r\omega t$, and $B_{2r}(t) = \cos r\omega t$. This basis is periodic, and the parameter $\omega$ determines the period $2\pi/\omega$. If the period is equal to the length of interval $T$, then the basis is orthogonal, that is, the matrix $W = \int_T B(t)B^T(t)dt$ is diagonal (Ramsay and Silverman, 2005). $W$ is orthonormal if we divide the basis functions by constants $\sqrt{n}$ for $l = 0$ and $\sqrt{n}/2$ for all other $j$. In Figure 2.2 some elements of a Fourier basis with period 365 and $K = 65$ are shown. In Chapters 4 and 5 we use a basis with equal period and $K = 65$ for smoothing the temperature data sets described in Section 1.1. Coefficients in 2.23 can be estimated by using the same methodology described for B-splines. However in this case would be more natural to use the harmonic acceleration operator as roughness penalty (Ramsay and Silverman, 2005), that is, we can obtain the coefficients by solving

$$\min_{c \in \mathbb{R}^K} \sum_{j=1}^{M} (y_j - \chi(t_j))^2 + \eta \int_T (\chi'''(t) + \omega^2 \chi'(t))^2 dt.$$  

(2.24)

In the case of Fourier basis with $K$ functions, CV works as follows. For $j = 1, \ldots, M$, let $\chi_{K,\eta}^{(j)}$ be the solution of the problem (2.24) when the observation $(t_j, y_j)$ has been leaved out of the sample and the parameter values $K$ and $\eta$ are used to fit the rest of the data. We define

$$NPCV(K, \eta) = \sum_{j=1}^{M} (y_j - \chi_{K,\eta}^{(j)}(t_j))^2.$$  

(2.25)

Then the smoothing parameters $(K, \eta)$ are chosen to minimize $NPCV(K, \eta)$.

If we have $i = 1, \ldots, n$ samples of curves the criteria (2.19) and (2.25) can be generalized as follows: Let $\tilde{\chi}_{s_i}^{(j)}(t_j)$ be the estimated function at $t_j$ by means of equation (2.17) when the datum $\chi_{s_i}(t_j)$ has been temporarily suppressed from the sample. Then the cross-validation sum of squared errors is calculated by

$$NPCV = \sum_{i=1}^{n} \sum_{j=1}^{M} (\chi_{s_i}(t_j) - \tilde{\chi}_{s_i}^{(j)}(t_j))^2.$$  

(2.26)

The strategy is choosing the parameters $(L, \eta)$ (for B-splines) or $(K, \eta)$ (for Fourier basis) that minimize $NPCV$. As mentioned in Section 2.2.2 an alternative for estimating the parameters by cross-validation when the sample size is large is the $NP_{GCV}$. 
The wavelet transform replaces the Fourier transform’s sinusoidal waves by a family generated by translations and dilations of a window called a wavelet. In contrast to Fourier series, wavelets expansions cope well with discontinuities or rapid changes in behavior; only those basis functions whose support includes the region of discontinuity or other bad behavior are reflected (Ramsay and Silverman, 2005). Wavelets are families of orthonormal basis functions that can be used to represent other functions parsimoniously. In Figure 2.3 are shown some wavelets families. The most known family of wavelets is the family Daubechies. Another mentioned wavelet is the simplest one, the Haar wavelet, which use a box function as the scaling function. In $L^2$ an orthogonal wavelet basis is obtained by
dilating and translating a mother wavelet $\psi$ as

$$\psi_{jk}(t) = 2^{j/2} \psi(2^j t - k), \quad (2.27)$$

with $j$ and $k$ integers that scale and dilate the mother function. Wavelets are constructed so that $\psi_{jk}(t)$ describes a particular level of detail in the signal. A square integrable function $x(t)$ can then be represented by the wavelet series

$$x(t) = \sum_{j,k} d_{jk} \psi_{jk}(t), \quad (2.28)$$

with $d_{jk}$ the wavelet coefficients (Morris and Carroll, 2006). Given the row vector $\mathbf{x} = (x(t_1), \ldots, x(t_N))$ containing values of a function $x(t)$ at $N$ points, a fast algorithm for decomposing $\mathbf{x}$ into a set of $N$ wavelet and scaling coefficients is the discrete wavelet transform (DWT). This is an implementation of the wavelet transform using a discrete set of wavelet scales and translations obeying some defined rules. This transform decomposes the signal into mutually orthogonal set of wavelets. There are several types of implementation of DWT. The oldest and most known is the pyramidal algorithm (Mallat, 1998). In this work we use the Haar wavelet transform. Next paragraphs are devoted to describe this wavelet transform. The orthogonal set of Haar wavelets $h_i(t)$ is a group of square waves with magnitude +1 or -1 in some intervals and zeros elsewhere (Hariharan et al., 2009). The Haar transform is defined by the scaling function $\psi_0(t) = 1, 0 \leq t < 1$ and the the Haar wavelet mother $\psi(t)$ defined as

$$\psi(t) = I_{[0,1/2)}(t) - I_{[1/2,1)}(t).$$

In order to perform wavelet transform, Haar wavelet uses translations and dilations by means of $\psi_1(t) = \psi(2^j t - k), \ j \geq 0, \ 0 \leq k < 2^j, \ j, k \in \mathbb{Z}$. $\psi_1(t) = \psi(t)$ is the fundamental square wave which spam the whole interval and $\psi_i(t), i > 1$, are generated from $\psi_1(t)$ via two operations: translations and dilations (Hsiao and Wang, 1999). Any square integrable function $x(t)$ can be approximated in a Haar series with $N = 2^j$ terms by (Hariharan et al., 2009)

$$x(t) \approx \sum_{i=0}^{N-1} c_i \psi_1(t) = \mathbf{c}^T \mathbf{\psi}(t). \quad (2.29)$$
Given the vector $\mathbf{x}^T = (x(t_1), \ldots, x(t_N))$, the vector of coefficients $\mathbf{c}$ in equation (2.29) is calculated by

$$\mathbf{c}^T = (x(t_1), \ldots, x(t_N)) \begin{pmatrix} \psi_0(t_1) & \psi_0(t_2) & \cdots & \psi_0(t_N) \\ \psi_1(t_1) & \psi_1(t_2) & \cdots & \psi_1(t_N) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{N-1}(t_1) & \psi_{N-1}(t_2) & \cdots & \psi_{N-1}(t_N) \end{pmatrix}^{-1}$$

As an example in Figure 2.4 we show simulated data from the function $x(t) = 40t^2(1 - t)^6 \sin(10\pi t), t \in [0, 1]$, and the Haar wavelet 1-level transform. We follow a similar example shown in Walker (1999). In right panel of Figure (2.4), the trend subsignal is graphed on the left half (over the interval $[0, 0.5]$) and the fluctuation subsignal is graphed on the right half (over the interval $[0.5, 1]$). The trend look like the original signal and the fluctuations area close to 0. This feature on the fluctuations are generally true in Haar transform (Walker, 1999).
2.2.3 Some techniques for modeling functional data

In this section we describe some techniques useful for modeling functional data (FD). Initially we show some basis of descriptive analysis. After that we resume the theory about functional linear models. References used for doing the resume are Malfait and Ramsay (2003) and Chapter 12 to 16 of Ramsay and Silverman (2005).

Descriptive analysis of FD

Classical summary statistics for univariate data applies equally to functional data. Assuming that a functional datum for replication \( i \) arrives as a set of discrete measured values \( X_{i1}, \ldots, X_{in} \), the first task is convert these values, using non parametric methods, to a function \( X_i(t) \) computable for any desired argument value \( t \). If the discrete values are assumed to be errorless, then the process is the interpolation, but if they have some observational error that needs removing, then the conversion from discrete to functions may involve smoothing. Mean, variance, covariance and correlation for samples of functional data are given for the next expressions (Ramsay and Silverman, 1997)

- **Mean**: \( \bar{X}(t) = N^{-1} \sum_{i=1}^{N} X_i(t) \).

- **Variance**: \( \text{Var}(X(t)) = (N - 1)^{-1} \sum_{i=1}^{N} (X_i(t) - \bar{X}(t))^2 \).

- **Covariance**: \( \text{Cov}(X(t_1), X(t_2)) = (N - 1)^{-1} \sum_{i=1}^{N} (X_i(t_1) - \bar{x}(t_1))(X_i(t_2) - \bar{x}(t_2)) \).

- **Correlation**: \( \text{Corr}(X(t_1), X(t_2)) = \frac{\text{Cov}(X(t_1), X(t_2))}{\sqrt{\text{Var}(X(t_1))\text{Var}(X(t_2))}} \).

- **Cross-covariance**: \( \text{Cov}(X(t_1), Y(t_2)) = (N - 1)^{-1} \sum_{i=1}^{N} (X_i(t_1) - \bar{X}(t_1))(Y_i(t_2) - \bar{Y}(t_2)) \).

- **Cross-correlation**: \( \text{Corr}(X(t_1), Y(t_2)) = \frac{\text{Cov}(X(t_1), Y(t_2))}{\sqrt{\text{Var}(X(t_1))\text{Var}(Y(t_2))}} \).
**Functional linear model: scalar response**

We can start recalling that given a response random variable $Y$ and $p$ fixed explanatory variables $X_1, \ldots, X_p$, the classical multiple linear regression is based on the estimation of the parameters $\alpha, \beta_1, \ldots, \beta_p$ in the model:

$$ Y_i = \alpha + \sum_j \beta_j X_{ij} + \epsilon_i, \quad (2.30) $$

where $\epsilon_i$ is a random error (traditionally assumed white noise), the parameters can be estimated by least squares minimizing

$$ SSE = \sum_{i=1}^{n} (Y_i - \alpha - \sum_j \beta_j X_{ij})^2. \quad (2.31) $$

Assuming we have an explanatory functional variable $X(t), t \in T$, where $T$ is a real interval, a functional linear model is a natural extension of multiple regression by replacing the summation in (2.30) by an integral. The functional linear model is then given by the expression

$$ Y_i = \alpha + \int_T X_i(t) \beta(t) dt + \epsilon_i. \quad (2.32) $$

The functional parameter $\beta(t)$ given in (2.32) cannot be estimated by least squares with a analogous expression to (2.31) because there are infinitely many sets of solutions. Estimation of this functional parameter can be obtained by using some regularization method (Ramsay and Silverman, 2005). It is usually assumed that the sampling trajectories and $\beta(t)$ are square integrable functions in a Hilbert space and they are generated by function basis $\psi(t) = (\psi_1(t), \ldots, \psi_{K_z}(t))^T$, $\theta(t) = (\theta_1(t), \ldots, \theta_{K_{\beta}}(t))^T$. Then trajectories can be expressed by

$$ X_i(t) = \sum_k^{K_z} a_{ik} \psi_k(t) = A^T_i \psi(t), $$

and the parameter function by
\[ \beta(t) = \sum_{k} b_k \theta_k(t) = \theta^T(t) b. \]

The model (2.32) can be expressed as

\[ \hat{Y}_i = \int_{T} X_i(t) \beta(t) dt = \int_{T} A_i^T \psi(t) \theta(t)^T b dt = A_i^T J_{\psi\theta} b, \]

where

\[ J_{\psi\theta} = \int_{T} \psi(t) \theta^T(t) dt. \]

Defining \( \zeta = (\alpha, b_1, \ldots, b_K)^T \) and \( Z = [1 \ A J_{\psi\theta}] \), the model (2.32) becomes simply

\[ \hat{Y} = Z \hat{\zeta} \]

and least squares estimate of \( \zeta \) is given by \( Z^T Z \hat{\zeta} = Z^T y. \)

Another alternative in functional linear model is to consider the response as a functional variable. In this case, the functional covariates can influence the functional response in two different ways: we might only use covariates at the same \( t \) or these could be extended over the whole \( T \). These models are called concurrent (point-wise) and total model, respectively (Ramsay and Silverman, 2005, chapter 12). The estimation is again carried out using basis function expansion and some regularization method.

**Functional linear model: concurrent model**

The model considered is given by

\[ Y_i(t) = \beta_0(t) + \beta_1(t) X_{i1}(t) + \beta_2(t) X_{i2}(t) + \cdots + \beta_q(t) X_{iq}(t) + \epsilon_i(t). \]

or in matrix notation

\[ Y(t) = X(t) \beta(t) + \epsilon(t). \] (2.33)
In equation (2.33) $\mathbf{X}(t)$ is a design matrix. A basis function expansion is estimated for each regression function $\beta_j(t)$ with roughness penalties to control the smoothness of the estimates. Roughness penalty for $\beta_j(t)$ is given by

$$\text{PEN}_j(\beta_j) = \lambda_j \int_T [L_j \beta_j(t)]^2 dt,$$

where $L_j$ is a linear differential operator. For example $L = D^2$, where $D$ is the derivative operator. The weighted regularized fitting criterion is

$$\text{SSE} = \int_T r(t)^T r(t) dt + \sum_j \lambda_j \int_T [L_j \beta_j(t)]^2 dt,$$

where

$$r(t) = \mathbf{Y}(t) - \mathbf{X}(t) \beta(t).$$

We define

$$\beta_j(t) = \sum_k b_{kj} \theta_{kj}(t) = \mathbf{\theta}_j(t)^T \mathbf{b}_j, \quad \mathbf{b} = (\mathbf{b}_1^T, \ldots, \mathbf{b}_q^T)^T$$

and

$$\mathbf{\Theta} = \begin{pmatrix}
\mathbf{\theta}_1^T & 0 & \cdots & 0 \\
0 & \mathbf{\theta}_2^T & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \mathbf{\theta}_q^T
\end{pmatrix}.$$ 

The model (2.33) is now expressed as

$$\mathbf{Y}(t) = \mathbf{X}(t) \mathbf{\Theta}(t) \mathbf{b} + \epsilon(t),$$

$$\mathbf{Y}(t) = \mathbf{X}^*(t) \mathbf{b} + \epsilon(t), \quad \mathbf{X}^*(t) = \mathbf{X}(t) \mathbf{\Theta}(t).$$

Then, the penalized least squares solution is

$$\left( \int_T \mathbf{X}^{*T}(t) \mathbf{X}^*(t) dt + \mathbf{R} \right) \mathbf{b} = \int_T \mathbf{X}^{*T}(t) \mathbf{y}(t) dt,$$

where $\mathbf{R}$ is the symmetric block diagonal matrix of roughness penalties.
2.2. FUNCTIONAL DATA ANALYSIS

Functional linear model: total model

The model is defined as

\[ Y_i(t) = \beta_0(t) + \int_T X_i(v) \beta(v, t) dv + \epsilon_i(t) \]  \hspace{1cm} (2.35)

From the normal equations

\[ \beta_0(t) = \mu_Y(t) - \int \mu_X(v) \beta(v, t) dv, \]

and consequently the model 2.35 can be expressed as

\[ Y_i(t) - \mu_Y(t) = \int_T (X_i(v) - \mu_X(v)) \beta(v, t) dv + \epsilon_i(t) \]
\[ Y_i^*(t) = \int_T X_i^*(v) \beta(v, t) dv + \epsilon_i(t), \]  \hspace{1cm} (2.36)

or in matrix notation

\[ Y^*(t) = \int_T X^*(v) \beta(v, t) dv + \epsilon(t). \]

Now the parameters are expressed with a double expansion in terms of \( K_1 \) basis functions \( \eta_k \) and \( K_2 \) basis functions \( \theta_l \) to give

\[ \beta(v, t) = \sum_{k=1}^{K_1} \sum_{l=1}^{K_2} b_{kl} \eta_k(v) \theta_l(t) = \eta(v)^T B \theta(t). \]  \hspace{1cm} (2.37)

Using (2.37), the model (2.35) is

\[ Y^*(t) = \int_T X^*(v) \theta^T(v) B \eta(t) dv + \epsilon(t) = X B \eta(t) + \epsilon(t), \]

where

\[ X = \int_T X^*(v) \theta^T(v) dv. \]
Then, without regularization, the normal equations are

\[ X^T X B \int_T \eta(t) \eta^T(t) dt = X^T \int_T Y^*(t) \eta^T(t) dt. \]  

(2.38)

Now, if we include regularization we need to define two roughness penalties: one for \( \beta \)'s variation with respect to \( v \), and another for its variation with respect to \( t \). The penalty respect to \( v \) is given by

\[
PEN_v(\beta) = \int \int (L_v \beta(v,t))^2 \, dv \, dt
= \int \int (L_v \theta^T(v) B \eta(t))(L_v \theta^T(v) B \eta(t))^T \, dv \, dt
= \int trace(B \eta(t) \eta^T(t) B^T R)
= trace(B^T R J_{\eta\eta}),
\]

where \( J_{\eta\eta} = \int_T \eta(t) \eta^T(t) dt \) and \( R = \int (L_v \theta(v))(L_v \theta^T(v)) \, dv \). Analogously the penalization of \( \beta \) respecto to \( t \) is given by \( trace(B^T J_{\theta\theta} S B) \), where \( J_{\theta\theta} = \int_T \theta(t) \theta^T(t) dt \) and \( S = \int (L_v \eta(t))(L_v \eta^T(t)) \, dt \). Then if we add these two penalties to equation 2.38 we obtain the system

\[ X^T X B J_{\eta\eta} + \lambda_v R J_{\eta\eta} + \lambda_t J_{\theta\theta} S = X^T \int_T Y^*(t) \eta^T(t) dt. \]  

(2.39)
Chapter 3

Ordinary kriging for function-value spatial data

In this chapter we consider a first approach to the problem of spatial prediction of functional data. We propose a functional kriging procedure where the predicted curve is a linear combination of observed curves where the coefficients are real numbers.

The problem considered in this Chapter was initially studied by Goulard and Voltz (1993). In the framework considered by Goulard and Voltz (1993) the functions were only known by a finite set of their points, and a parametric model was fitted to them for reconstructing the whole curve. In this context the parametric model was assumed to be known and both, the number of known points for each function and the number of parameters in the parametric model, were implicitly assumed to be small. We recover the contributions of Goulard and Voltz (1993) overcoming the restrictive assumptions on parametric modeling and small number of observed points per function. In particular, we propose to apply a non-parametric fitting pre-process to the observed functions (in this Chapter we use B-spline smoothing, but other approaches are possible) where the smoothing parameter is chosen by what we call functional cross-validation. We believe that our approach is in complete agreement with present trends in FDA, and in particular, with non-parametric functional estimation methodology. Our proposal for doing kriging-based spatial prediction of random curves formally coincides with the functional kriging intro-
duced in Goulard and Voltz (1993), but our non-parametric approach involves noticeable differences (for instance, data representation) and the additional problem of choosing the smoothing parameters (the keystone of non-parametric methods). Our predictor is based on the basic philosophy of functional data, that is, curves are single entities, rather than a sequence of individual observations (Ramsay and Silverman, 2005).

3.1 Predictor and estimation of parameters

Let us consider a functional random process \( \chi_s : s \in D \subseteq \mathbb{R}^d \), usually \( d = 2 \), such that \( \chi_s \) is a functional variable for any \( s \in D \). Let \( s_1, \ldots, s_n \) be arbitrary points in \( D \), and assume that we can observe a realization of the functional random process \( \chi_s \) at these \( n \) sites, \( \chi_{s_1}, \ldots, \chi_{s_n} \).

We assume that we have a second-order stationary and isotropic random process, that is, the mean and variance functions are constant and the covariance depends only on the distance between sampling points (however, the methodology could also be developed without assuming these conditions). Formally we assume that:

- \( E(\chi_s(t)) = m(t) \), for all \( t \in T, s \in D \).
- \( V(\chi_s(t)) = \sigma^2(t) \), for all \( t \in T, s \in D \).
- \( \text{Cov}(\chi_{s_i}(t), \chi_{s_j}(t)) = C(h; t) = C_{s_is_j}(t) \), for all \( s_i, s_j \in D, t \in T \), where \( h = \|s_i - s_j\| \).
- \( \frac{1}{2}V(\chi_{s_i}(t) - \chi_{s_j}(t)) = \gamma(h; t) = \gamma_{s_is_j}(t) \), for all \( s_i, s_j \in D, t \in T \), where \( h = \|s_i - s_j\| \).

The function \( \gamma(h; t) \), as a function of \( h \), is called semivariogram of \( \chi(t) \). We consider the family of linear predictors for \( \chi_{s_0} \):

\[
\hat{\chi}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i \chi_{s_i}(t), \quad \lambda_1, \ldots, \lambda_n \in \mathbb{R}.
\tag{3.1}
\]
The predictor (3.1) has the same expression as an ordinary kriging, but considering curves instead of variables; that is, the predicted curve is a linear combination of observed curves. This predictor was initially considered by Goulard and Voltz (1993). In Section 3.3 we comment on the differences between our approach and the solutions given by these authors. We are assuming that each measured curve is a complete datum. This approach treats the whole curve as a single entity and do not consider correlations between the repeated measurements. The $\lambda$'s in predictor (3.1) give the influence of the curves surrounding the unsampled location where we want to carry out prediction. Curves from locations closer to the prediction point will have greater influence than other more separated. This a first natural step in modeling of spatial functional data. In chapters 4 and 5 we consider other more flexible predictors which take into account correlations into the functional index.

To find the best linear unbiased predictor (BLUP), we consider first the unbiasedness. From the constant mean condition above, we require that $\sum_{i=1}^{n} \lambda_i = 1$. In a classical univariate geostatistical setting we assume that the observations are realizations of a random field $\{Z(s) : s \in D, D \in \mathbb{R}^d\}$. The kriging predictor is defined as $\sum_{i=1}^{n} \lambda_i Z(s_i)$, and the best linear unbiased predictor (BLUP) is obtained by minimizing

$$\sigma_{s_0}^2 = V(\hat{Z}(s_i) - Z(s_i)),$$

subject to $\sum_{i=1}^{n} \lambda_i = 1$. On the other hand in multivariable geostatistics (Myers, 1982; Ver Hoef and Cressie, 1993; Wackernagel, 1995) the data consist of $\{Z(s_1), \cdots, Z(s_n)\}$, that is, we have observations of a spatial vector-valued process $\{Z(s) : s \in D\}$, where $Z(s) \in \mathbb{R}^m$ and $D \in \mathbb{R}^d$. In this context $V(\hat{Z}(s_0) - Z(s_0))$ is a matrix, and the BLUP of $m$ variables on an unsampled location $s_0$ can be obtained by minimizing

$$\sigma_{s_0}^2 = \sum_{i=1}^{m} V\left(\hat{Z}_i(s_0) - Z_i(s_0)\right),$$

subject to constraints that guarantee unbiasedness conditions, that is, minimizing the trace of the mean-squared prediction error matrix subject to some restrictions given by the unbiasedness condition (Myers, 1982). Extending the criterium given in Myers (1982) to the functional context by replacing the summation by an integral, the $n$ parameters
in the kriging predictor of $\chi_{s_0}$ considered in this Section are given by the solution of the following optimization problem:

$$\min_{\lambda_1, \ldots, \lambda_n} \int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))dt, \text{ s.t. } \sum_{i=1}^n \lambda_i = 1. \hspace{1cm} (3.2)$$

where $\sum_{i=1}^n \lambda_i = 1$ is the unbiasedness constraint. Observe that unbiasedness and Fubini’s Theorem\(^1\) imply that

$$\int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))dt = \int_T E[(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))^2]dt = E\left[\int_T (\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))^2dt\right].$$

Consequently we need to minimize

$$\int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))dt + 2 \mu (\sum_{i=1}^n \lambda_i - 1), \hspace{1cm} (3.3)$$

where $\hat{\chi}_{s_0}(t) = \sum_{i=1}^n \lambda_i \chi_{s_i}(t)$ and $\mu$ is the Lagrange multiplier used to take into account the unbiasedness restriction. The integral in equation (3.3) can be written as

$$\sigma_{s_0}^2 = \int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))dt = \int_T V(\hat{\chi}_{s_0}(t))dt + \int_T V(\chi_{s_0}(t))dt - 2 \int_T C(\hat{\chi}_{s_0}(t), \chi_{s_0}(t))dt$$

$$= \int_T V(\sum_{i=1}^n \lambda_i \chi_{s_i}(t))dt + \int_T \sigma^2(t)dt - 2 \int_T C(\sum_{i=1}^n \lambda_i \chi_{s_i}(t), \chi_{s_0}(t))dt$$

$$= \int_T \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(\chi_{s_i}(t), \chi_{s_j}(t))dt + \int_T \sigma^2(t)dt$$

$$- 2 \int_T \sum_{i=1}^n \lambda_i C(\chi_{s_i}(t), \chi_{s_0}(t))dt$$

$$= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \int_T C_{ij}(t)dt + \int_T \sigma^2(t)dt - 2 \sum_{i=1}^n \lambda_i \int_T C_{i0}(t)dt. \hspace{1cm} (3.4)$$

Then, the objective function can be written as

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \int_T C_{ij}(t)dt + \int_T \sigma^2(t)dt - 2 \sum_{i=1}^n \int_T C_{i0}(t)dt + 2 \mu (\sum_{i=1}^n \lambda_i - 1). \hspace{1cm} (3.5)$$

\(^1\)Here and thereafter is sections 3.2, 4.1, and 5.2, respectively, we apply Fubini’s Theorem assuming that regularity conditions hold for square integrability of realization of curves.
Minimizing (3.5) with respect to $\lambda_1, \cdots, \lambda_n$ and $\mu$, we obtain the following set of $(n + 1)$ equations

\begin{align*}
\sum_{j=1}^{n} \lambda_j \int_T C_{1j}(t)dt + \mu &= \int_T C_{10}(t)dt \\
\sum_{j=1}^{n} \lambda_j \int_T C_{2j}(t)dt + \mu &= \int_T C_{20}(t)dt \\
&\vdots \\
\sum_{j=1}^{n} \lambda_j \int_T C_{nj}(t)dt + \mu &= \int_T C_{n0}(t)dt \\
\sum_{j=1}^{n} \lambda_j &= 1
\end{align*}

which can be expressed in matrix notation by

\[
\begin{pmatrix}
\int_T C_{s_1,s_1}(t)dt & \cdots & \int_T C_{s_1,s_n}(t)dt & 1 \\
\vdots & \ddots & \vdots & \vdots \\
\int_T C_{s_n,s_1}(t)dt & \cdots & \int_T C_{s_n,s_n}(t)dt & 1 \\
1 & \cdots & 1 & 0
\end{pmatrix}
\begin{pmatrix}
\lambda_1 \\
\vdots \\
\lambda_n \\
\mu
\end{pmatrix}
= 
\begin{pmatrix}
\int_T C_{s_1,s_0}(t)dt \\
\vdots \\
\int_T C_{s_n,s_0}(t)dt \\
1
\end{pmatrix}
\]  \hspace{1cm} (3.7)

We also can obtain the estimations based on trace-variogram. From stationarity assumptions, we have that

\[
\gamma_{s_is_j}(t) = \gamma(x_{s_i}(t), x_{s_j}(t)) = V(x_{s_i}(t) - x_{s_j}(t)) = \frac{1}{2}E(x_{s_i}(t) - x_{s_j}(t))^2 = \sigma^2(t) - C_{ij}(t).
\]

Then

\[
\int_T C_{ij}(t)dt = \int_T \sigma^2(t)dt - \int_T \gamma_{s_is_j}(t)dt.
\]  \hspace{1cm} (3.8)
By replacing equation (3.8) in the system (3.7), we obtain the system

\[
\begin{pmatrix}
\int T \gamma_{s_1s_1}(t) dt & \cdots & \int T \gamma_{s_1s_n}(t) dt & 1 \\
\vdots & \ddots & \vdots & \vdots \\
\int T \gamma_{s_ns_1}(t) dt & \cdots & \int T \gamma_{s郑s_n}(t) dt & 1 \\
1 & \cdots & 1 & 0
\end{pmatrix}
\begin{pmatrix}
\lambda_1 \\
\vdots \\
\lambda_n \\
-\mu
\end{pmatrix}
= 
\begin{pmatrix}
\int T \gamma_{s_0s_1}(t) dt \\
\vdots \\
\int T \gamma_{s_0s_n}(t) dt \\
1 \\
-\mu
\end{pmatrix}.
\tag{3.9}
\]

From the first \( n \) equations in system (3.6), we have the relation

\[
\sum_{i=1}^{n} \sum_{i=1}^{n} \lambda_i \lambda_j \int T C_{ij}(t) dt = \sum_{i=1}^{n} \lambda_i \int T C_{i0}(t) dt - \sum_{i=1}^{n} \lambda_i \mu.
\tag{3.10}
\]

Replacing equation (3.10) into equation (3.4) we obtain

\[
\sigma^2 = \int T \sigma^2(t) dt - \sum_{i=1}^{n} \lambda_i \int T C_{i0}(t) dt - \mu.
\]

If, in addition, we consider the relation (3.8), the prediction trace-variance of the functional ordinary kriging based on the trace-variogram is given by

\[
\sigma^2_{s_0} = \int T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t)) dt = \sum_{i=1}^{n} \lambda_i \int T \gamma_{s_is_0}(t) dt - \mu = \sum_{i=1}^{n} \lambda_i \gamma(h) - \mu.
\tag{3.11}
\]

The parameter defined in equation (3.11) should be considered as a global uncertainty measure, in the sense that it is an integrated version of the classical point-wise prediction variance of ordinary kriging. Under a specified trace-variogram model, we can use estimations of this parameter to identify those zones presenting greater uncertainty on the predictions. If we fix \( t \in T \) and replace the trace-variogram \( \gamma(h) \) in equation (3.11) for the variogram of a random variable \( \chi(t), t \in T \) we would obtain the prediction variance of classical ordinary kriging (Isaaks and Srivastava, 1987). This result could be used for calculating point-wise confidence intervals.

### 3.2 Estimating the trace-variogram

In order to solve the system in expression (3.9), an estimator of the trace-variogram is needed. Given that we are assuming that \( \chi_s(t) \) has a constant mean function \( m \) over \( D \),
Then an adaptation of the classical method-of-moments (MoM) for this quantity gives the following estimator

\[ \hat{\gamma}(h) = \frac{1}{2} \frac{1}{|N(h)|} \sum_{i,j \in N(h)} \int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt, \]  

(3.12)

where \( N(h) = \{(s_i, s_j) : \|s_i - s_j\| = h\} \), and \(|N(h)|\) is the number of distinct elements in \( N(h) \). For irregularly spaced data there are generally not enough observations separated by exactly \( h \). Then \( N(h) \) is modified to \( \{(s_i, s_j) : \|s_i - s_j\| \in (h - \epsilon, h + \epsilon)\} \), with \( \epsilon > 0 \) being a small value.

Once we have estimated the trace-variogram for a sequence of \( K \) values \( h_k \), we propose to fit a parametric model \( \gamma_\alpha(h) \) (any of the classical and widely used models such as spherical, Gaussian, exponential or Matérn could well be used) to the points \((h_k, \hat{\gamma}(h_k))\), \( k = 1, \ldots, K \), as if they were obtained under the classical geostatistical setting. Usually, this type of fitting is done by ordinary least squares (OLS) or weighted least squares (WLS) (see, for instance, Cressie, 1993). Note that the fitted parametric trace-variogram \( \gamma_\alpha(h) \) is always a valid variogram because its properties are those of a parametric variogram fitted from a univariate geostatistical data set. A different procedure, alternative to the parametric fitting, consists of applying smoothing techniques (splines or local linear regression, see Wasserman, 2006, and references therein) to the set of data \((h_k, \hat{\gamma}(h_k))\), \( k = 1, \ldots, K \), in order to be able to approximately evaluate \( \hat{\gamma}(h) \) for any value of \( h \in \mathbb{R}^+ \). However in this case, if \( \tilde{\gamma}(h) \) denotes the smoothed version of \( \hat{\gamma}(h) \), the question of conditionally definite-negativeness of \( \tilde{\gamma}(h) \) deserves more attention.

If \( \gamma_\alpha(h) \) denotes the parametric estimated trace-variogram, this functional form is used both to obtain the kriging coefficients \( \lambda_i \) in equation (3.9), and to estimate the prediction trace-variance through equation (3.11).
CHAPTER 3. ORDINARY KRIGING FOR FUNCTIONAL DATA

3.3 A non-parametric approach

Goulard and Voltz (1993) deal with the same problem of geostatistical interpolation of curves and specifically they consider the predictor 3.1. They consider that curves are only known by a finite set of their points: \( \chi_{s_i}(t_j), j = 1, \ldots, M, i = 1, \ldots, n \), and show three approaches for predicting curves at unvisited sites:

- **Multivariate approach 1: Cokrige first, Fit later (CFP).** The vector of observed values \( (\chi_{s_i}(t_1), \ldots, \chi_{s_i}(t_M)) \) is considered as the observation of a \( M \)-dimensional random variable at site \( s_i \). Cokriging is then applied to predict the values of this random vector at the unvisited site \( s_0 \): \( (\hat{\chi}_{s_0}(t_1), \ldots, \hat{\chi}_{s_0}(t_M)) \). Therefore, a parametric model \( \chi(\cdot; \theta), \theta \in \mathbb{R}^p \) is fitted to the values \( (\hat{\chi}_{s_0}(t_1), \ldots, \hat{\chi}_{s_0}(t_M)) \) for reconstructing a whole function at \( s_0 \): \( \chi(\cdot; \hat{\theta}_{s_0}) \).

- **Multivariate approach 2: Fit first, Cokrige later (FCP).** First, the parametric model is fitted to the observed curves: \( \chi(\cdot; \hat{\theta}_{s_i}), i = 1, \ldots, n \). The \( p \)-dimensional parameter values \( \hat{\theta}_{s_1}, \ldots, \hat{\theta}_{s_n} \) are considered as observations of a multivariate random variable. Then cokriging is applied to predict the value of the parameter \( \theta \) at site \( s_0 \), say \( \hat{\theta}_{s_0}^* \), and \( \chi(\cdot; \hat{\theta}_{s_0}^*) \) is the resulting predicted function at \( s_0 \).

- **A curve kriging approach (CKP).** Given that the functions \( \chi_{s_i} \) are known only for \( M \) values, Goulard and Voltz (1993) propose to fit a parametric model \( \chi(\cdot; \theta), \theta \in \mathbb{R}^p \), to these data to obtain \( \chi(\cdot; \hat{\theta}_{s_i}) \) as an approximation of the whole function \( \chi_{s_i} \). Therefore, equation (3.1) can be rewritten as

\[
\hat{\chi}_{s_0} = \sum_{i=1}^{n} \lambda_i \chi(\cdot; \hat{\theta}_{s_i}),
\]

and the integrals on \( T \) involved in estimating the coefficients \( \lambda_i \) are calculated using \( \chi(\cdot; \hat{\theta}_{s_i}) \) instead of \( \chi_{s_i} \) (see Section 3.2).

Now, let us consider the common case where a large number \( M \) of values are observed for each sampled function and there is no parametric model fitting them adequately. In this context the second proposal of Goulard and Voltz (1993), CFP (Cokrige first, Fit
later), becomes extremely expensive in terms of computing resources. The reason is that CFP involves a first cokriging step where the dimension of the multivariate observation is equal to \( M \). The computational cost of this step is reasonable if \( M \) is in tens, but it is unacceptable if \( M \) is in the order of several hundreds or thousands, as it usually happens nowadays.

When a parametric model is adequate for representing the observed functions, the alternative FCP (Fit first, Cokrige later) of Goulard and Voltz (1993) is feasible, because in this case the cokriging step involves \( p \)-dimensional vectors, where \( p \) is the number of parameters in the parametric model. Here the problem appears when no parametric model is considered to be adequate to fit the observed values. Then a non-parametric fit could be done, but this process is essentially equivalent to a parametric fit with a number of parameters \( p_M \) growing with the number of observed values \( M \). Moreover, the extra flexibility provided by a non-parametric model is usually obtained at the expense of allowing for large values of \( p_M \) (even if always \( p_M \leq M \)). Therefore we end up that the computational cost of the FCP approach could also be prohibitive.

So we limit ourselves to use the curve kriging predictor (CKP) proposed by Goulard and Voltz (1993). From equation (3.9) and Section 3.2, it is clear that the crucial step to estimate the trace-variogram \( \gamma(h) \) and to compute the kriging coefficients \( \lambda_i \) in (3.1), is the computation of integrals over \( T \) having the generic form \( \int_T (\chi_{s_i}(t) - \chi_{s_j}(t))^2 dt \). When a parametric model fits well the available values of the observed functions, Goulard and Voltz (1993) propose to replace this integral by \( \int_T (\chi(t; \hat{\theta}_{s_i}) - \chi(t; \hat{\theta}_{s_j}))^2 dt \), that usually has a closed analytical expression. Here we address the question of what can be done if there is no parametric model acceptably fitting the observed functions. Our proposal is to replace the parametric fitting step by its non-parametric counterpart.
3.4 Application: Spatial prediction of penetration resistance curves

An application of the proposed methodology is carried out by analyzing the data set described in Section 1.1.1. Figure 1.1 suggests that it is not easy to propose a parametric model for these curves and that there is a certain degree of observational noise. We smooth the observed functions using B-splines basis functions. Initially we give some guidelines for achieve this stage in the context of spatially correlated functional data when the objective is carry out spatial prediction and posteriorly we do the analysis of data by using the predictor proposed.

3.4.1 Evaluation criterion of a kriging predictor

In any prediction problem, the ideal model evaluation consists of splitting the whole data set into two parts: a training (estimation) sample for model fitting and a test (validation)
sample for model evaluation. This approach, however, is not efficient unless the sample size is large. The idea behind cross-validation proposals is to recycle data by switching the roles of training and test samples.

In the context of spatially correlated functional data, where the goal is to predict a whole function $\chi_{s_0}(t)$ at an unvisited site $s_0$, leave-one-out cross-validation works as follows: each data location is removed from the data set and a smoothed function is predicted at this location using a functional kriging predictor. We call this procedure functional cross-validation (FCV). We calculate the SSE of FCV by

$$SSE_{FCV} = \sum_{i=1}^{n} SSE_{FCV}(i) = \sum_{i=1}^{n} \sum_{j=1}^{M} (\chi_{s_i}(t_j) - \hat{\chi}^{(i)}_{s_i}(t_j))^2,$$

where $\hat{\chi}^{(i)}_{s_i}(t_j)$ is the prediction on $s_i$ evaluated at $t_j$, $j = 1, \cdots, M$, by leaving the site $s_i$ temporarily out of the sample.

In the particular case of using ordinary kriging for function-value data as prediction method and B-splines as smoothing method the functional cross-validation works as follows

- Minimize in $L \in [L_{\text{min}}, L_{\text{max}}]$ and $\eta \in [\eta_{\text{min}}, \eta_{\text{max}}]$ the function $FCV(L, \eta)$ that is computed by the following steps for each fixed value $(L, \eta)$:

1. For $i = 1, \ldots, n$, repeat:
   
   (a) Fit a cubic spline to $\chi_{s_i}$ using equation (2.18) and smoothing parameters $(L, \eta)$.
   
   (b) Leave the site $s_i$ out of the sample.
   
   (c) Use equation (3.12) to estimate the empirical trace-variogram from the data set $\tilde{\chi}_{s_i'}, i' = 1, \ldots, n, i' \neq i$, and then fit a parametric model for the trace-variogram, as it is described in Section 3.2.
   
   (d) Solve the system (3.9) with the trace-variogram estimated in the previous step and the data set $\tilde{\chi}_{s_i'}, i' = 1, \ldots, n, i' \neq i$, to predict the function at site $s_0 = s_i$. Let $\tilde{\chi}^{(i)}_{s_i}$ be the resulting function.
   
   (e) Compute a measure of distance between $\chi_{s_i}$ and $\tilde{\chi}^{(i)}_{s_i}$ at the observed values $t_1, \ldots, t_M$: $SSE_{FCV}(i) = \sum_{j=1}^{M} (\chi_{s_i}(t_j) - \tilde{\chi}^{(i)}_{s_i}(t_j))^2$. 
2. Define $FCV(L, \eta) = \sum_{i=1}^{n} SSE_{FCV}(i)$.

- Use the optimal values $(L^*, \eta^*)$ to smooth the whole sample. Then fit a parametric model for the trace-variogram and use it to predict functions at unvisited sites using the kriging equations (3.9).

Some comments are in order. When $t_j, j = 1, \ldots, M$, are equally spaced the quantity $SSE_{FCV}(i)$ is (up to a multiplicative constant) an approximation to the integral $\int_T (\chi_{s_i}(t) - \tilde{\chi}_{s_i}(t))^2 dt$.

The estimation of the empirical trace-variogram using equation (3.12) involves the computation of integrals that, in the case of fitting cubic splines with a common basis of B-splines, can be simplified to give

$$
\int_T (\tilde{\chi}_{s_i}(t) - \tilde{\chi}_{s_j}(t))^2 dt = \int_T (c_i^T B(t) - c_j^T B(t))^2 dt = \int_T ((c_i - c_j)^T B(t))^2 dt = (c_i - c_j)^T \left( \int_T B(t)B(t)^T dt \right) (c_i - c_j) = (c_i - c_j)^T W(c_i - c_j)^T.
$$

The matrix $W$ depends only on the knots, so it is common for all the sites $s_i$. Similar reasoning allows us to write the penalty term in equation (2.18) as $c_iDc_i^T$, where the term $(l, k)$ in matrix $D$ is $\int_T B_l''(t)B_k''(t)dt$, only depending on the knots.

The former considerations are powerful arguments to use a common value of parameter $L$ for all the functions $\chi_{s_i}$. The parameter $\eta$ could vary from site to site, but this implies a standard cross-validation step (CV) when fitting any $\tilde{\chi}_{s_i}$. This would increase the computational cost and therefore we opt for choosing a common value for all sites.

### 3.4.2 Analysis of data using B-splines

Penetration values must be non-negative, but the kriging prediction formula (3.1) does not guarantee that the predicted curve values are always non-negative, even if the observed curves are (this is because some $\lambda_i$ in the solution to the system of equations (3.9) could be negative). Therefore we have worked with the square root of penetration functions in the prediction process. Once a curve has been predicted, then the square of this function is
taken as prediction of the penetration resistance function. This way the non-negativeness of predicted penetration resistances is guaranteed. The square root transformation is used only as a method for obtaining positive predictions. Another Box-Cox transformation could be used. Further attention must be given to the bias effect induced for this procedure. The functional cross-validation process above described has been used to find adequate smoothing parameters: $L$, the number of interior knots, and $\eta$, the roughness penalty parameter. The sets of possible values considered for $L$ and $\eta$ have been, respectively, \{6, \ldots, 11\} and \{0, 1, 10, 10^2, 10^3\}. A previous exploration leads us to consider these ranges for smoothing parameters. Figure 3.2 shows the contour plot of the function $FCV(L, \eta)$, using a logarithmic scale for $\eta$. The minimum is achieved at $(L^* = 10, \eta^* = 0)$. It was mentioned in Section 2.2.2 that the degree of smoothing is an increasing function of $\eta$ and a decreasing function of $L$. In this case the estimated parameters $L^*$ and $\eta^*$ indicate that we can get a good level of smoothness only considering a low number of interior knots. The set of smoothed functions using these smoothing parameter values is shown in Figure 3.1, right panel. In step 1(c) of the algorithm that implements the FCV process, a spherical model was used to fit the empirical trace-variogram values calculated for several spatial lags (the fitting procedure was WLS). The spherical model has been widely used in the literature, and was chosen in terms of its flexibility, simplicity and interpretability of its parameters. Indeed, this model has three free parameters: nugget, partial sill ($\sigma^2$) and range ($\phi$). The estimated spherical model for the trace-variogram, when using the optimal smoothing parameters $(L^*, \eta^*)$, was as follows: the nugget was equal to 1.01, the partial sill equal to $\sigma^2 = 2.25$, and the range was $\phi = 39.10$. So the estimated trace-variogram was $\hat{\gamma}(h) = 1.01 + 2.25(1.5h/39.10 - 0.5(h/39.10)^3)$ for $h \leq 39.10$, and $\hat{\gamma}(h) = 3.26$ for $h > 39.10$.

Taking into account that in the data set both the maximum distance between two adjacent sites in any direction is 35.36 meters and the minimum distance between non-adjacent sites is 50.08 meters, and considering that the estimated range is 39.10 meters, we can say that two sampling sites are correlated if and only if they are adjacent. This result suggests, from a empirical point of view, that the data have have a weak spatial dependence.
As an example of the proposed methodology, kriging prediction on an unsampled location with UTM coordinates 11179 (eastings) and 9750 (northings) (see Figure 1.1) was performed. The kriging coefficients $\lambda_i$ were obtained by solving the system of equations (3.9) with $\gamma(h)$ estimated by the spherical variogram $\hat{\gamma}(h)$ as described above. Figure 3.3 (left panel) shows the weights $\lambda_i$ as a function of the Euclidean distance between sampled sites $s_i$ and the unvisited site $s_0$. It is clearly highlighted that the highest weights $\lambda_i$ correspond to the four sites surrounding $s_0$. The predicted curve (Figure 3.3, thick line in the right panel) indicates that in this location there is a good soil compaction level, because the predicted penetration resistance is less than 2 MPa, which is considered the critical limit for root growth (Chan et al., 2006). The functional cross-validation method (FCV) used to choose the smoothing parameters, can also be considered a useful tool to compare observed and predicted curves, as it defines a measure of distance between these two curves. Indeed, the $SSE(i)$ can be considered an approximation to the mean integrated squared error (MISE) (Myers, 1991). Note that when using cross-validation the idea is that the predicted values should be “close” to the observed values (using the

Figure 3.2: Contour plot for the function $FCV(L, \eta)$, using a logarithmic scale for $\eta$ (eta in the graphic).
3.4. PREDICTION OF PENETRATION RESISTANCE CURVES

Figure 3.3: *Left panel:* Kriging weights $\lambda_i$ as a function of the Euclidean distance between sites $s_i$ and $s_0$. *Right panel:* Kriging prediction at an unsampled location (black thick line) over the set of smoothed curves.

leave-one-out technique). However, there is not a single best way to quantify “closeness”. In the case of numerical valued random functions, there are at least six different statistics (Myers, 1991). In the thesis, we have chosen FCV. Figure 3.4 shows a graphical comparison between observed and predicted curves (using FCV). It shows that predicted curves are smoother than observed ones, as well as that the predicted data set has less variance. This was not surprising since kriging is itself a smoothing method (thus the variance decreases), and also because there is a significant high variability amongst penetration resistance values for some particular depth levels (see right panel of Figure 3.3). In our case, though noting that several senses of smoothness could be considered, our curves satisfy both situations: (a) a particular predicted curve is smoother than the corresponding observed curve, and (b) the variability among the predicted curves is less than the variability among the observed curves.

A detailed analysis of functional cross-validation residuals indicates that there was no evidence of biased predictions (see mean function in Figure 3.4). We also note that the
Figure 3.4: *Left panel:* Kriging predictions for the 32 original sites based on functional cross-validation (data smoothed by B-splines). *Right panel:* Functional cross-validation residuals (grey lines), residual mean (black thick line) and residual standard deviation (dashed line).
uncertainty on predictions is approximately constant for depths greater than 15 cm. (see residual standard deviation in Figure 3.4).

### 3.4.3 Analysis of data using wavelets

There are many applications of functional data analysis where the data are not smooth curves. The local adaptivity of wavelet-based curve estimation may yield favorable results when the curves have irregular and complex structures as shown in Figure 1.1. For that reason in this section we consider the use of wavelets and particularly the Haar wavelet transform for analyzing the penetration resistance data set. We do, as in section 3.4.2, prediction on an unvisited site and subsequently we carry out a cross-validation analysis. We use the cross-validation residuals for comparing the results obtained by wavelets with those obtained by B-splines in previous Section. Again to guarantee positive predictions we work with square root of penetration functions and subsequently the square of predictions are taken as penetration resistance predictions.

We fit a Haar wavelet 3-level transform to the square root of penetration values. Square values of the inverse transform are shown in left panel of Figure 3.5. A spherical model was fitted to the trace-variogram calculated with coefficients of the Haar wavelet transform. The estimated model was \( \hat{\gamma}(h) = 2.01 + 0.73(1.5h/68 - 0.5(h/68)^3) \) for \( h \leq 68 \), and \( \hat{\gamma}(h) = 2.74 \) for \( h > 68 \). According to the range (68 meters) of the trace-variogram model, the functions obtained by the Haar transform show a greater spatial correlation than curves smoothed by B-splines (the range in that case was 39.10 meters).

If we compare the prediction on an unvisited site (Figure 3.5), the cross-validation predictions (left panel in Figure 3.6), and the cross-validation residuals (right panel in Figure 3.6) achieved with the functions obtained by the Haar wavelet transform with those achieved with functions smoothed by B-splines (Figures 3.3, 3.4(left panel) and 3.4(right panel), respectively) we can conclude that the results are very similar. The difference in reflected only in terms of the smooth level achieved with of the B-spline basis considered in Section 3.4.2.

From Table 3.1 we can conclude that both methods (OKFD based on B-splines and
Figure 3.5: *Left panel:* Inverse Haar wavelet transform of penetration resistance data.  
*Right panel:* Kriging prediction at an unsampled location (black thick line) over the set of inverse transform wavelet curves.
Figure 3.6: *Left panel:* Kriging predictions for the 32 original sites based on functional cross-validation (data converted to functions by a Haar wavelet transform). *Right panel:* Functional cross-validation residuals (grey lines), residual mean (black thick line) and residual standard deviation (dashed line).
Table 3.1: Summary of cross-validation residuals of OKFD when data (penetration resistance) are smoothed by B-splines and transformed by a Haar Wavelet.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>B-splines</th>
<th>Wavelets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>4.4</td>
<td>4.8</td>
</tr>
<tr>
<td>Median</td>
<td>24.9</td>
<td>26.8</td>
</tr>
<tr>
<td>Mean</td>
<td>34.9</td>
<td>43.4</td>
</tr>
<tr>
<td>Maximum</td>
<td>175.8</td>
<td>176.1</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>34.3</td>
<td>34.0</td>
</tr>
<tr>
<td>Sum</td>
<td>1118.2</td>
<td>1146.5</td>
</tr>
</tbody>
</table>

Based on Haar wavelet transform) have similar performances (in terms of cross-validation residuals) with the Penetrations resistance data set. The smoothness achieved by B-splines is reflected in slightly better predictions.
Chapter 4

Continuous time-varying kriging for spatial prediction of functional data

In this Chapter we consider the problem of spatial prediction of functional data by weighting each observed curve by a functional parameter. Our modeling approach is a combination of ordinary kriging and the functional linear concurrent (point-wise) model such as shown in Ramsay and Silverman (2005). We propose a solution based on basis functions. Both the curves and the functional parameters are expanded in terms of a set of basis functions. Thus, the problem becomes one of estimating the coefficients of these basis functions for each functional parameter. In order to provide a solution, we use a linear model of coregionalization for estimating the covariances among the coefficients of each curve.

4.1 Predictor and minimization problem

In this Chapter we propose to use a family of point-wise linear predictors for \( \chi_{s_0}(t), t \in T \), given by

\[
\hat{\chi}_{s_0}(t) = \sum_{i=1}^{n} \lambda_i(t) \chi_{s_i}(t), \quad \lambda_1(t), \ldots, \lambda_n(t) : T \to \mathbb{R}.
\]  

(4.1)
CHAPTER 4. CONTINUOUS TIME-VARYING KRIGING

This predictor was mentioned in Goulard and Voltz (1993) without further development. We assume the same assumptions given in Chapter 3. For each \( t \in T \), the predictor (4.1) has the same expression as an ordinary kriging predictor. In the rest of the thesis this predictor is called continuous time-varying kriging for functional data (CTKFD). This modeling approach is coherent with the functional linear concurrent model (Hastie and Tibshirani, 1993; Ramsay and Silverman, 2005) as shown in equation (2.33), in which the influence of each covariate on the response is simultaneous or point-wise. In this model the response \( Y_i(t) \) and each covariate \( X_{ij}(t) \), \( j = 1, \cdots, q \), are functions of the same argument and \( X_j(t) \) only influences \( Y(t) \) through its value at time \( t \) (Ramsay and Silverman, 2005). Estimation of functional parameters \( \alpha(t), \beta_j(t), j = 1, \cdots, q \), is carried out by solving (Ramsay and Silverman, 2005)

\[
\min_{\alpha(\cdot), \beta_1(\cdot), \ldots, \beta_q(\cdot)} E\|\hat{\mathbf{Y}}(t) - \mathbf{Y}(t)\|^2.
\]

In our context, the covariates are the observed curves at \( n \) sites of a region and the functional response is an unobserved function on an unsampled location. Consequently, our optimization problem is

\[
\min_{\lambda_1(\cdot), \ldots, \lambda_n(\cdot)} E\|\hat{\mathbf{X}}_{s_0}(t) - \mathbf{X}_{s_0}(t)\|^2,
\]

or equivalently, by using Fubini’s Theorem,

\[
\min_{\lambda_1(\cdot), \ldots, \lambda_n(\cdot)} \int_T E(\hat{\mathbf{X}}_{s_0}(t) - \mathbf{X}_{s_0}(t))^2 dt.
\]

If we consider the stationarity assumptions, the problem becomes

\[
\min_{\lambda_1(\cdot), \ldots, \lambda_n(\cdot)} \int_T V(\hat{\mathbf{X}}_{s_0}(t) - \mathbf{X}_{s_0}(t)) dt. \tag{4.2}
\]

As in Chapter 3 the minimization problem in equation (4.2) is an extension of the minimization criterion given by Myers (1982) to the functional context, by replacing the summation by an integral and the random vectors \([\mathbf{Z}_1(s_0), \cdots, \mathbf{Z}_m(s_0)]\) and \([\hat{\mathbf{Z}}_1(s_0), \cdots, \hat{\mathbf{Z}}_m(s_0)]\) by the functional variables \( \mathbf{X}(t) \) and \( \hat{\mathbf{X}}(t) \), respectively, with \( t \in T \).
The predictor (4.1) is unbiased if 
\[ E(\hat{\chi}_0(t)) = m(t), \] for all \( t \in T \), that is, if \( \sum_{i=1}^{n} \lambda_i(t) = 1 \) for all \( t \in T \). Consequently, in order to find the BLUP, the \( n \) functional parameters in the predictor proposed are given by the solution of the following optimization problem

\[
\min_{\lambda_1(\cdot), \ldots, \lambda_n(\cdot)} \int_T V(\hat{\chi}_0(t) - \chi_0(t)) \, dt, \text{ s.t. } \sum_{i=1}^{n} \lambda_i(t) = 1, \text{ for all } t \in T. \tag{4.3}
\]

4.2 A solution based on basis functions

We assume that each observed function can be expressed in terms of \( K \) basis functions by

\[
\chi_i(t) = \sum_{l=1}^{K} a_{il} B_l(t) = a_i^T B(t), \quad i = 1, \ldots, n. \tag{4.4}
\]

By taking into account that \( \chi_i(t), i = 1, \ldots, n \), are random functions with spatial dependence, we assume that the matrix

\[
A = \begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1K} \\
a_{21} & a_{22} & \cdots & a_{2K} \\
\vdots & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nK}
\end{pmatrix} = (\alpha_1, \ldots, \alpha_K)_{(n \times K)}
\]

forms a \( K \) multivariable random field with \( E(\alpha_i) = v_{i(1 \times 1)} \) and covariance matrix

\[
\Sigma = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1K} \\
\Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2K} \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma_{K1} & \Sigma_{K2} & \cdots & \Sigma_{KK}
\end{pmatrix}_{(K \times n) \times (K \times n)} \tag{4.5}
\]

where \( \Sigma_{ij} = C(\alpha_i, \alpha_j)_{(n \times n)} \). The coefficients \( a_{ij} \) are assumed a realization of the spatial random field \( \alpha_j, j = 1, \ldots, K \). We propose to use multivariable geostatistics (Wackernagel, 1995) and specifically a linear model of coregionalization (LMC) for estimating the matrix (4.5). In order to establish the unbiasedness condition and for carrying out the
parameter estimation in (4.1) we further expand each functional parameter \( \lambda_i(t) \) by

\[
\lambda_i(t) = \sum_{i=1}^{K} b_{il} B_l(t) = b_i^T B(t).
\] (4.6)

Therefore, using (4.4) and (4.6) the predictor in equation (4.1) is given by

\[
\hat{\chi}_{s_0}(t) = \sum_{i=1}^{n} b_i^T B(t) a_i^T B(t)
\] (4.7)

\[
= \sum_{i=1}^{n} b_i^T B(t) B^T(t) a_i.
\]

Using (4.6) and by expanding the constant function 1 by means of \( \sum_{l=1}^{K} c_l B_l(t) = c^T B(t) = 1 \), the unbiasedness constraint can be expressed as

\[
\sum_{i=1}^{n} b_i^T 1 = c^T B(t), \forall t, \iff \sum_{i=1}^{n} b_i = c,
\]

or more specifically by

\[
\sum_{i=1}^{n} b_{i1} = c_1, \ldots, \sum_{i=1}^{n} b_{iK} = c_K.
\] (4.8)

Developing the variance in the objective function (4.3) we have

\[
V \left( \hat{\chi}_{s_0}(t) - \chi_{s_0}(t) \right) = V \left( \hat{\chi}_{s_0}(t) \right) + V \left( \chi_{s_0}(t) \right) - 2C(\hat{\chi}_{s_0}(t), \chi_{s_0}(t))
\]

\[
= V \left( \sum_{i=1}^{n} b_i^T B(t) B^T(t) a_i \right) + B^T(t) V(a_0) B(t)
\]

\[
- 2 \sum_{i=1}^{n} b_i^T B(t) B^T(t) C(a_i, a_0) B(t)
\]

\[
= \sum_{i=1}^{n} b_i^T B(t) B^T(t) V(a_i) B(t) B^T(t) b_i
\]

\[
+ 2 \sum_{i<j}^{n} b_i^T B(t) B^T(t) C(a_i, a_j) B(t) B^T(t) b_j
\]

\[
+ B^T(t) V(a_0) B(t)
\]

\[
- 2 \sum_{i=1}^{n} b_i^T B(t) B^T(t) C(a_0, a_i) B(t).
\] (4.9)
In equation (4.9), for \(i < j, i, j = 0, 1, \cdots, n\), we have

\[
V(a_i) = \begin{pmatrix} \text{var}(a_i) & \text{cov}(a_i, a_{i1}) & \cdots & \text{cov}(a_i, a_{iK}) \\ \text{cov}(a_{i1}, a_i) & \text{var}(a_{i1}) & \cdots & \text{cov}(a_{i1}, a_{iK}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(a_{iK}, a_i) & \text{cov}(a_{iK}, a_{i1}) & \cdots & \text{var}(a_{iK}) \end{pmatrix}_{(K \times K)}
\]

and

\[
C(a_i, a_j) = \begin{pmatrix} \text{cov}(a_{i1}, a_{j1}) & \text{cov}(a_{i1}, a_{j2}) & \cdots & \text{cov}(a_{i1}, a_{jK}) \\ \text{cov}(a_{i2}, a_{j1}) & \text{cov}(a_{i2}, a_{j2}) & \cdots & \text{cov}(a_{i2}, a_{jK}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{cov}(a_{iK}, a_{j1}) & \text{cov}(a_{iK}, a_{j2}) & \cdots & \text{cov}(a_{iK}, a_{jK}) \end{pmatrix}_{(K \times K)}
\]

If we define

\[
Q_i = \int_T (B(t)B^T(t)V(a_i)B(t)B^T(t)) \, dt,
\]

\[
Q_{ij} = \int_T (B(t)B^T(t)C(a_i, a_j)B(t)B^T(t)) \, dt,
\]

\[
D = \int_T B^T(t)V(a_0)B(t) \, dt,
\]

and

\[
J_i = \int_T (B(t)B^T(t)C(a_0, a_i)B(t)) \, dt,
\]

and by considering \(K\) Lagrange multipliers \(m^T = (m_1, \cdots, m_K)\), the objective function (4.3) can be expressed as

\[
\min_{b_1, \ldots, b_n, m} \sum_{i=1}^n b_i^T Q_i b_i + 2 \sum_{i<j} b_i^T Q_{ij} b_j + D - 2 \sum_{i=1}^n b_i^T J_i + 2m^T \left( \sum_{i=1}^n b_i - c \right) \quad (4.10)
\]

By taking \(\beta = (b_1^T, \cdots, b_n^T, m^T)_{(K(n+1) \times 1)}^T\), the expression (4.10) is given by

\[
\min_{\beta} \beta^T Q \beta + D - 2\beta^T J \quad (4.11)
\]
where
\[
Q = \begin{pmatrix}
Q_1 & Q_{12} & \cdots & Q_{1n} & I \\
Q_{21} & Q_2 & \cdots & Q_{2n} & I \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
Q_{n1} & Q_{n2} & \cdots & Q_n & I \\
I & I & \cdots & I & 0
\end{pmatrix},
\]
\[
J = \begin{pmatrix}
J_1 \\
J_2 \\
\vdots \\
J_n \\
c
\end{pmatrix}.
\]
(4.12)

The identity matrix in (4.12) is of order \(K\). Minimizing equation (4.11) with respect to \(\beta\) we obtain
\[
2Q\beta - 2J = 0 \Rightarrow Q\beta = J \Rightarrow \hat{\beta} = Q^{-1}J.
\]
(4.13)

In practice, we start estimating both a LMC for the multivariable random field \(A = (\alpha_1, \cdots, \alpha_K)\) and the matrix in equation (4.5). Subsequently, we can calculate the matrices \(Q\) and \(J\) in equation (4.12). Replacing these matrices in equation (4.13) we can estimate \(b_i, i = 1, \cdots, n\) and consequently the functional parameters given in (4.6).

On the other hand, a plug-in estimation of the integrated prediction variance \(\hat{\sigma}^2_{s_0} = \int_T V(\hat{\chi}_{s_0}(t) - \chi_{s_0}(t))\,dt\) is given by
\[
\hat{\sigma}^2_{s_0} = \hat{\beta}^T Q \hat{\beta} + D - 2\hat{\beta}^T J,
\]
(4.14)

where the matrix \(D\) is calculated by using \(\hat{V}(a_0)\), which is obtained from the fitted LMC. The integrated prediction variance \(\hat{\sigma}^2_{s_0}\) is a measure of the uncertainty in the prediction of a whole curve. Based on the estimated parameters and using equation (4.9), a point-wise prediction variance function can also be estimated.

### 4.3 Application: Spatial prediction of temperature curves in Canada’s Maritime provinces

In this section we illustrate our approach by using the temperature data set described in Section 1.1.3. We initially select an appropriate number of basis functions. In a second stage we perform prediction at an unvisited site using the proposed predictor and describe the results from a practical point of view. In order to avoid a large computational work we
Figure 4.1: Sum of squared errors as a function of $K$ obtained by non-parametric cross-validation for Canada’s Maritime provinces data set.

choose the number of basis functions by using directly the non-parametric cross-validation criterion described in equation (2.26) and posteriorly we do functional cross-validation with the $K$ chosen. We do not consider roughness penalty in this case, that is, we assume $\eta = 0$ in equation (2.24). When data are periodic, Fourier basis with an even number of basis functions is the most appropriate choice (Ramsay and Silverman, 2005). A Fourier basis with 65 basis functions is used to expand the Canadian temperature data (Ramsay and Silverman, 2005). Although we can expand in terms of a Fourier basis with an infinite number of sinusoids, we take 365 as the limit because this is the number of discrete data for each site in our data set. Frequencies greater than 365 in this case will distort the signal. This is known as the problem of aliasing (Lleachor and Jervis, 1993). Figure 4.1 shows the relation between $K$ and $NPCV(K)$ obtained by non-parametric cross-validation. We can observe that the SSE values decrease significantly until the number of basis functions is around 60. Then the rate of decrease is small. In particular, $NPCV(K)$ decreases by 44% when $K$ is between 5 and 65, whereas this percentage is 55% when $K$ varies between 5 and
145 (where the minimum of $NPCV(K)$ is attained). Thus SSE values indicate that there is no advantage in using a value of $K$ much larger than 65. As mentioned before, Ramsay and Silverman (2005) also use 65 Fourier basis functions for smoothing the Canadian temperature data set. Our results suggest that this number of basis functions is also appropriate for our data set. In consequence, a pragmatic choice for $K$ is 65. Therefore, in the following we assume that the data to be analyzed correspond to the temperatures curves obtained after smoothing each discrete data set by means of a Fourier basis with 65 functions.

CTKFD is used to predict a temperature curve at an unvisited site with coordinates $-64.69$ (easting) and $45.10$ (northing). This site corresponds to the Moncton station (Figure 1.3). Moncton is situated in southeastern New Brunswick, and at the geographic center of the Maritime Provinces (Figure 1.3). The climate of this station is more continental than maritime during the summer and winter seasons, whereas maritime influences tend to temper the transitional seasons of spring and autumn. It should be noted that the latitude does not create large climatic variations in this region (Stanley, 2002) and consequently it plays no role in the modeling.

In the first stage of the analysis, and using the library {gstat} of R language (Pebesma, 2004), a LMC was fitted to the multivariable random field $A = (\alpha_1, \cdots, \alpha_{65})$ consisting of the coefficients of the Fourier basis used for smoothing each observed sample. We assume stationarity for each process $\alpha_j$, $j = 1, \cdots, 65$. All single (direct) variograms and cross-variograms are modeled as a linear combination of nugget and exponential models. Based on the fitted LMC, the matrices $Q$ and $J$ given in (4.12) are estimated and used to solve the system (4.13) finally to find $b_i$ and the functional parameters $\lambda_i(t)$, $i = 1, \cdots, n$. Figure 4.2 (left panel) shows a plot of the estimated functional parameters. We note that an estimated functional parameter is much greater than the others (functional parameter with values around 0.5). This corresponds to Bouctouche (NB), the closest station to Moncton (Figure 1.3). Other stations near Moncton, and therefore with influence on the prediction, are Nappan (NS), and Alma (NB) with weights around 0.2 in Figure 4.2. The curves corresponding to the sites furthest from Moncton receive a weight of almost zero (Figure 4.2, left panel). This result is consistent with the kriging philosophy, that is, sites
closer to the prediction location have greater influence than others further apart. The sum of the estimated functional parameters is equal to 1 for all \( t \) (Figure 4.2). With this result we verify graphically that the system (4.8) guarantees the unbiasedness constraint.

A plot of the temperature prediction at Moncton is also shown in Figure 4.2 (right panel). The predicted curve is obtained essentially as a weighted sum of smoothed temperature curves corresponding to Bouctouche, Nappan and Alma stations. The predicted curve shows a seasonal behavior similar to the smoothed curves. In addition, predicted values are consistent with real values recorded for this weather station (Figure 4.2). Prediction errors (difference between observed and predicted values in Figure 4.2) vary from -1.98 °C to 2.28 °C. The mean of the errors is -0.0053 °C. These results provide evidence that the prediction using CTKFD is quite close to the observed data. Temperature values for the Moncton station are recorded from http://www.climate.weatheroffice.ec.gc.ca/climateData/.

To verify the goodness-of-fit of the proposed predictor, we use the functional cross-
validation results obtained with 65 Fourier basis functions. Each individual smoothed curve $\chi_{s_i}(t), i = 1, \cdots, 35,$ is temporarily removed, and further predicted from the remaining ones by means of CTKFD. A comparison between predicted and smoothed curves (Figures 4.2 and 4.3) shows that the predictions have the same temporal behavior as the smoothed curves. Note also that the latter curves have less variance, in particular in wintertime (where the Canadian weather is more variable, Figure 4.3). This is not surprising, given that kriging is itself a smoothing method.

Figure 4.3 (right panel) shows cross-validation residuals. The predictions are plausible in a high proportion of sites (those having residuals around zero). There are few stations with large positive or negative residuals. These are particularly obtained in wintertime (residuals higher than 4 °C and lower than -4 °C at the beginning and at the end of the year). This is due to the fact that the temperature curves show more variability in this season. The greatest residuals are obtained in Bertrand (NB) and Bathurst (NB) (Figure 4.4). The prediction at Bertrand is most influenced by the curve at Bathurst,
and vice versa because the distance between these stations is smaller than the distance to the remaining ones (Figure 1.3). The temperature curves at Bertrand and Bathurst have a very similar behavior throughout the time considered (Figure 4.4). However, for some days (19, 49 to 57, 353, 354 and 356) the difference between the temperature values are greater than 4 °C. For this reason we obtain high residuals at these stations.

Continuing with the cross-validation residual analysis, in Figure 4.3 (right panel) we note that, although there are outliers, the residual mean indicates that the predictions are unbiased (mean around zero). We can also observe that the variation on the prediction is lower in summer (days 100 and 300) than in winter (Figure 4.3, right panel) as a consequence of the reasons given above. Cross-validation prediction variances are estimated by using equation (4.14). As in multivariable kriging, this statistic depends only on estimations of simple (direct) and cross-covariances, that is, it depends on the distance between the prediction site and the sampling locations, and does not take into account the observed values, that is, the uncertainty on the prediction is directly related to the sampling configuration. The further the prediction site, the greater the prediction variance. This result is clearly highlighted in the map of prediction variances (Figure 4.5), which shows that the most widely separated weather stations in our data set, such as those located in the Nova Scotia’s north coast (Sydney, Baddeck, Cheticamp and Ingonish beach), Aroostook (NB) and Bertrand (NB), have greater variances. This result should be interpreted carefully because at a weather station with a high prediction variance we may find low prediction errors. This is the case of Baddeck (NS) and Cheticamp (NS). There we have a relatively high prediction variance (Figure 4.5) with a good prediction (Figure 4.4). In general, the cross-validation results show that the predictions by CTKFD are close to the smoothed curves, and therefore this method can be considered a valid technique for performing spatial prediction of functional data.
Figure 4.4: Temperature curves (observed, smoothed and predicted by point-wise kriging) at four weather stations in Canada’s Maritime provinces.
4.4 Application: Spatial prediction of Canadian temperature curves

In this section we illustrate our approach by using the Canadian temperature data set described in Section 1.1.2. As in Section 4.3 an analysis of non-parametric cross-validation suggest that a Fourier basis with 65 functions can be appropriate for smoothing the data set. In this Section we follow the same methodology of Section 4.3, that is, we do prediction at an unvisited location and after that we carry out a cross-validation analysis.

Taking into account that in the Canadian data set the mean function is not constant, the smoothed data are initially detrended by using a functional regression model with functional response (smoothed temperature curves) and two scalar covariates (longitude and latitude coordinates in decimal degress), that is, we consider the functional regression model

$$\chi_{s_0}(t) = \hat{\alpha}(t) + \hat{\beta}_1(t)\text{Longitude} + \hat{\beta}_2(t)\text{Latitude} + e_{s_0}(t).$$

(4.15)

Subsequently, the residual $e_{s_0}(t)$ on an unvisited site $s_0$ is predicted by CTKFD and finally the temperature prediction at $s_0$ is obtained by adding the kriging prediction to the trend given by the model 4.15. The estimated parameters for the fitted model are shown in Figure (4.6). The variables latitude and longitude in model 4.15 were previously standardized. The estimated parameters show clearly that the temperature is much more
influenced by the coordinates in the winter than in the summer.

In Figure (4.7) are shown the smoothed temperature curves and the curves estimated by the functional regression model given in equation 4.15. In Figure (4.8) are shown the corresponding residual curves of this model. An alternative model for detrending the mean in the Canadian data set is to consider a functional Anova model taking as factor the regions (Atlantic, Continental, Pacific and Artic) such as in shown in Chapter 13 of Ramsay and Silverman (2005). We choose the functional regression model given in equation 4.15 instead of the functional Anova model because with the functional Anova model assigning stations to regions could be ambiguous in some cases. From Figures (4.7) and (4.8) we can conclude that the fitted regression model has good performance. High residuals in Figure (4.8) are due to bad estimations at stations with extreme winters as Victoria, Vancouver, Prince Rupert (with warmer winters than most stations) or Resolute (with an extremely cold winter).

CTKFD was used to predict a residual curve at Slave Lake station (Figure 1.2). Figure 4.9 (left panel) shows a plot of the corresponding estimated functional parameters. The largest estimated functional parameter corresponds to Edmonton, the closest station to Slave Lake (Figure 1.2). Other stations near to Slave Lake, and therefore with influence on the prediction, are Yellowknife (weights around 0.2, Figure 4.9), Uranium City and Pr George (values around 0.1, Figure 4.9). As in Figure 4.2 the sum of the estimated functional parameters is equal to 1 for all $t$ (Figure 4.9) which indicates that the unbiasedness constraint is fulfill. The prediction at Slave Lake Figure 4.9 (right panel) is consistent with real values recorded for this weather station (http://www.climate.weatheroffice.ec.gc.ca/climateData/). A comparison between cross-validation prediction and smoothed curves (Figures 4.9 and 4.10) shows that the predictions have the same temporal behavior that the smoothed curves. As in the analysis of Canada’s Maritime provinces data set, the predictions has less variability specially in wintertime.

From figure 4.10 we can observe that there are some stations with large positive or negative residual curves. This is due to the fact that the temperature curves at Resolute, Inuvik, Iqaluit, Dawson, Churchill, Prince Rupert and St Johns are not well predicted.
Figure 4.6: Estimated functional parameters $\hat{\alpha}(t)$ (top), $\hat{\beta}_1(t)$ (centre) and $\hat{\beta}_2(t)$ (bottom) for the estimation of temperature (functional response) from geographical coordinates (longitude, latitude) with the Canadian data set. Variables are standardized. The model includes a constant function. Dashed lines are 95% point-wise confidence limits for the functional parameters.
Figure 4.7: *Left panel:* Smoothed temperatures by using a Fourier basis function with 65 functions. *Right panel:* Estimations of temperature obtained by a functional multiple regression model with functional response (smoothed temperature) and two covariates (longitude and latitude coordinates).
Figure 4.8: Residuals of the functional regression model of temperature versus the coordinates (longitude and latitude) with the Canadian data set.
Figure 4.9: Estimated functional parameters for predicting a residual curve at Slave Lake (left, dark lines), sum of estimated functional parameters (left, dotted line), smoothed temperature curves (right, clear lines), temperature prediction function at Slave Lake (right, dark line) and real temperature values at Slave Lake (right, circles). Temperature prediction was obtained by adding the trend plus the residual prediction.
because they have extreme temperature values, and are spatially very separated from the remaining ones (Figure 1.2). As an example of this phenomenon, we can compare predictions for Bagottville, Edmonton, Resolute and Prince Rupert stations (Figure 4.11). We observe a good fit for Bagottville and Edmonton which are close to other weather stations (Figure 1.2), whereas for Resolute, the farthest station considered in our data set, and for Prince Rupert with an oceanic climate, the difference between smoothed and predicted curves is greater than 10 degrees Celsius several days or months of the year, specially in wintertime.

The results obtained by the cross-validation analysis and the bad predictions for some far apart stations suggest that when considering temperatures over a large spatial area (through all Canada in this case) surely an anisotropic model and the effect of costal/inland must be included into the analysis. For example, if our unobserved location has 4 observed sites each 50 miles away, respectively directly north, south, east and west, we would expect that the ones east and west would be much better predictors than the ones north.
Figure 4.11: Temperature curves (observed, smoothed and predicted by continuous time kriging) at four Canadian weather stations. Temperature predictions obtained by adding the trend plus the residual prediction achieved by continuous time kriging.
and south. An exception to this is when sites are near the coast, due to the substantial
difference between coastal/island and continental climates. If in this case the one to the
west was on the coast then you would expect the site to the east to be by far the best
predictor. Similarly, if the unobserved site has two observed sites on the same latitude,
one 50 miles away on the coast and the other 150 miles away inland, we might still expect
the one 150 miles away to be a better predictor. In consequence Resolute is predicted
extremely badly (Figure 4.11), not just because it is a long way away from anywhere else
but also because it is much further north. Similarly Pr Rupert is predicted as having lower
winter temperatures (Figure 4.11) than it really does because it is coastal but the nearest
site, Pr George (Figure 1.2), is inland. These results suggest that including anisotropy
and the effect coastal/inland, if it is the case, could be need when we do prediction of
functional data based on a data set recorded over a large spatial scale. Comparing cross-
validation residuals in Figures 4.3 and 4.10 we can conclude that the effect of consider a
more homogeneous region is reflected in better predictions.
Chapter 5

From multivariable to functional geostatistics

In spatial statistics and specifically in geostatistics, both cokriging analysis (Myers, 1982; Bogaert, 1996) and multivariable kriging (Ver Hoef and Cressie, 1993; Ver Hoef and Barry, 1998) are used for modeling observations of vector-valued random fields. In this work we adapt these methodologies to functional context. We propose a cokriging predictor for doing univariate prediction (as in the cokriging multivariable sense), but considering as auxiliary information samples of functions instead of observations of random vectors. Likewise, we extend the multivariable kriging from random vectors to the functional context by defining a functional kriging predictor which allows to do prediction of a whole curve at an unvisited site by using as information the curves sampled in nearby sites to the prediction site. In both cases (cokriging based on functions and functional kriging) we give a non-parametric solution based on basis functions, and we prove that both proposals coincide when using such an approach. The problem of functional kriging prediction considered in this work is also studied by Nerini and Monestiez (2008). They propose a solution based on orthonormal basis functions. In our approach this is not a necessary condition. The proposed methodologies in this Chapter are in the same way of the solution given in Chapter 4, that is, each observed curve is weighted by a functional parameter. However here the flexibility increases because double indexed functional parameters are
COKRIGING BASED ON FUNCTIONAL DATA

estimated. Now, each curve is weighted by a functional parameter for carrying out the prediction at each time. This modeling approach follows the basic philosophy of the functional linear model for functional response (total model) given in equation (2.35) where a bivariate regression coefficient function must be estimated (Malfait and Ramsay, 2003). The results of applying these methodologies to the temperature data described in Chapter 1 are graphically similar to the obtained in Chapter 4. Consequently in order to avoid repetitions we only show the results obtained with one of these data sets. However, at the end of this Chapter, we use summary statistics of cross-validation residual of both data sets for doing comparison between the methods proposed in the thesis.

5.1 Cokriging based on functional data

Let \( \{\chi_s(t), t \in T, s \in D \subset \mathbb{R}^d\} \) be a random function defined on some compact set \( T \) of \( \mathbb{R} \). Suppose we observe a sample of curves \( \chi_{s_1}(t), \cdots, \chi_{s_n}(t) \) defined for \( t \in T, s_i \in D, i = 1, \cdots, n \). We assume that these curves belong to a separable Hilbert space \( H \) of square integrable functions defined on \( T \). We consider that for each \( t \in T \) we have a second-order stationary and isotropic random process, that is, the mean and variance functions are constant and the covariance depends only on the distance among sampling points. We want to predict a single variable at a single location from a sample of spatially correlated functional data \( \chi_{s_1}(t), \cdots, \chi_{s_n}(t) \). Let \( \chi_{s_0}(v) \) be the random variable to be predicted at an unsampled location \( s_0 \) at a time \( v \in T \). For carrying out this task we can extend the cokriging predictor shown in Section 2.1.2, replacing in (2.7) both the \( n \times m \) parameters \( \lambda_{ij}^k \) by \( n \) functional parameters \( \lambda_i^v(t) \) and the \( n \times m \) random variables \( Z_j(s_i) \) by \( n \) functional variables \( \chi_{s_i}(t), i = 1, \cdots, n, j = 1, \cdots, m \). We describe these replacements in schemes (5.1) and (5.2), respectively.
 Parameters

\[
\begin{align*}
\lambda_{11} \cdots \lambda_{1m} & \Rightarrow \lambda_1^v(t), \ t \in T \\
\lambda_{21} \cdots \lambda_{2m} & \Rightarrow \lambda_2^v(t), \ t \in T \\
\vdots & \vdots \\
\lambda_{n1} \cdots \lambda_{nm} & \Rightarrow \lambda_n^v(t), \ t \in T
\end{align*}
\] (5.1)

 Variables

\[
\begin{align*}
Z_1(s_1) \cdots Z_m(s_1) & \Rightarrow \chi_{s_1}(t), \ t \in T \\
Z_1(s_2) \cdots Z_m(s_2) & \Rightarrow \chi_{s_2}(t), \ t \in T \\
\vdots & \vdots \\
Z_1(s_n) \cdots Z_m(s_n) & \Rightarrow \chi_{s_n}(t), \ t \in T
\end{align*}
\] (5.2)

Thus, the cokriging predictor of \( \chi_{s_0}(v) \) based on functional data (CBFD) is given by

\[
\hat{\chi}_{s_0}(v) = \sum_{i=1}^{n} \int_T \lambda_i^v(t) \chi_{s_i}(t) dt.
\] (5.3)

For each specific \( v \in T \), the functional parameters \( \lambda_i^v(t), i = 1, \cdots, n \) in (5.3) are estimated by taking into account classical geostatistical constraints; that is, unbiasedness and minimum prediction variance. We solve this problem using an approach based on basis functions. We expand functional variables by using the expression (4.4) and functional parameters by

\[
\lambda_i^v(t) = \sum_{l=1}^{K} b_{ilv} B_l(t) = b_{iv}^T B(t).
\] (5.4)

Therefore using (4.4) and (5.4), the predictor (5.3) is expressed as

\[
\hat{\chi}_{s_0}(v) = \sum_{i=1}^{n} \int_T b_{iv}^T B(t) B^T(t) a_i dt \\
= \sum_{i=1}^{n} b_{iv}^T W a_i,
\] (5.5)
where

$$W = \int_T B(t)B^T(t)dt. \quad (5.6)$$

For any orthonormal basis such as the Fourier basis, the Gram matrix $W$ is the identity matrix. For other basis functions such as B-Splines basis $W$ must be calculated by using numerical integration. Assuming stationarity hypothesis of the random functions the matrix

$$A = \begin{pmatrix}
a_{11} & a_{12} & \cdots & a_{1K} \\
a_{21} & a_{22} & \cdots & a_{2K} \\
 & \vdots & \ddots & \vdots \\
a_{n1} & a_{n2} & \cdots & a_{nK}
\end{pmatrix} = (\alpha_1, \cdots, \alpha_K)_{(n \times K)}$$

forms a K multivariable random field with

$$E(\alpha_j) = \begin{pmatrix}
E(a_{1j}) \\
E(a_{2j}) \\
\vdots \\
E(a_{nj})
\end{pmatrix} = \begin{pmatrix}
\vartheta_j \\
\vartheta_j \\
\vdots \\
\vartheta_j
\end{pmatrix}, \quad j = 1, \cdots, k, \quad (5.7)$$

and covariance matrix

$$\Sigma = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} & \cdots & \Sigma_{1K} \\
\Sigma_{21} & \Sigma_{22} & \cdots & \Sigma_{2K} \\
& \vdots & \ddots & \vdots \\
\Sigma_{K1} & \Sigma_{K2} & \cdots & \Sigma_{KK}
\end{pmatrix}. \quad (5.8)$$

where $\Sigma_{ij} = C(\alpha_i, \alpha_j)_{(n \times n)}$. From equation (5.7) we have

$$E(\alpha_i) = \begin{pmatrix}
E(a_{i1}) \\
E(a_{i2}) \\
\vdots \\
E(a_{iK})
\end{pmatrix} = \begin{pmatrix}
\vartheta_1 \\
\vartheta_2 \\
\vdots \\
\vartheta_K
\end{pmatrix} = \vartheta. \quad (5.9)$$
Thus, the mean of the predictor (5.5) is given by

\[
E(\hat{\chi}_{s_0}(v)) = \sum_{i=1}^{n} b_{iv}^T W E(a_i) = \sum_{i=1}^{n} b_{iv}^T W \vartheta.
\]

On the other hand, the mean of the unobserved function on site \(s_0\) at a time \(v\) is

\[
E(\chi_{s_0}(v)) = \sum_{l=1}^{K} E(a_{0l}) B_l(v) = B^T(v) E(a_0) = B^T(v) \vartheta.
\]

Consequently the proposed predictor is unbiased if

\[
\sum_{i=1}^{n} b_{iv}^T = B^T(v) W^{-1}. \tag{5.10}
\]

Although a Gram matrix is in general positive semidefinite, the Gram matrix \(W\) in (5.6) is positive definite because the functions \(B_l(t), l = 1, \cdots, K\), are linearly independent. Consequently the equation (5.10) is well defined. In order to find the best linear unbiased predictor (BLUP), the \(n\) functional parameters in the proposed predictor are given by the solution of the following optimization problem:

\[
\text{Min}_{\lambda_1(v), \ldots, \lambda_K(v)} V(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)), \text{ s.t. } E(\hat{\chi}_{s_0}(v)) = E(\chi_{s_0}(v)). \tag{5.11}
\]

By developing the variance in the objective function (5.11), we find the next expression:
\[ V(\tilde{\chi}_s(v) - \chi_s(v)) = V(\tilde{\chi}_s(v)) + V(\chi_s(v)) - 2C(\tilde{\chi}_s(v), \chi_s(v)) \]
\[ = \sum_{i=1}^{n} b_{iv}^T WV(a_i)W^T b_{iv} \]
\[ + 2 \sum_{i<j} b_{iv}^T WC(a_i, a_j)W^T b_{jv} \]
\[ + B^T(v)V(a_0)B(v) \]
\[ - 2 \sum_{i=1}^{n} b_{iv}^T WC(a_0, a_i)B(v). \] (5.12)

\(V(a_i), C(a_i, a_j), \) and \(C(a_0, a_i)\) in equation (5.12) can be calculated if \(\Sigma\) in (5.8) has been previously estimated. We can use multivariable geostatistics (Wackernagel, 1995) and specifically a linear model of coregionalization (LMC) in order to estimate these matrices. If we define

\[ M_i = (WV(a_i)W^T)_{(K \times K)}, \]
\[ M_{ij} = (WC(a_i, a_j)W^T)_{(K \times K)}, \]
\[ N_i(v) = (WC(a_0, a_i)B(v))_{(K \times 1)} \] (5.13)

and

\[ D(v) = B^T(v)V(a_0)B(v), \] (5.14)

the equation (5.12) is given by

\[ V(\tilde{\chi}_s(v) - \chi_s(v)) = \sum_{i=1}^{n} b_{iv}^T M_i b_{iv} + 2 \sum_{i<j} b_{iv}^T M_{ij} b_{jv} \]
\[ + D(v) - 2 \sum_{i=1}^{n} b_{iv}^T N_i(v). \] (5.15)
From (5.15) and considering \( K \) Lagrange multipliers \( \mathbf{m}_v^T = (m_{1v}, \ldots, m_{Kv}) \), the optimization problem (5.11) can be expressed as

\[
\min_{b_{1v}, \ldots, b_{nv}, m_v} \sum_{i=1}^{n} b_{iv}^T M_i b_{iv} + 2 \sum_{i<j} b_{iv}^T M_{ij} b_{jv} + D(v)
\]

\[
- 2 \sum_{i=1}^{n} b_{iv}^T N_i(v) + 2 m_v^T \left( \sum_{i=1}^{n} b_{iv} - W^{-1} B(v) \right).
\]

(5.16)

Taking \( \beta_v = (b_{1v}^T, \ldots, b_{nv}^T, m_v^T)^T \), the equation (5.16) is given by

\[
\min_{\beta_v} \beta_v^T M \beta_v + D(v) - 2 \beta_v^T N(v)
\]

(5.17)

where

\[
M = \begin{pmatrix}
M_1 & M_{12} & \cdots & M_{1n} & I \\
M_{21} & M_2 & \cdots & M_{2n} & I \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
M_{n1} & M_{n2} & \cdots & M_n & I \\
I & I & \cdots & I & 0
\end{pmatrix}_{[K(n+1) \times [K(n+1)]}
\]

and

\[
N(v) = \begin{pmatrix}
N_1(v) \\
N_2(v) \\
\vdots \\
N_n(v) \\
W^{-1} B(v)
\end{pmatrix}_{(K(n+1) \times 1)}
\]

(5.18)

Finally, minimizing the equation (5.17) respect to \( \beta_v \), we obtain

\[
2M \beta_v - 2N(v) = 0 \Rightarrow M \beta_v = N(v) \Rightarrow \hat{\beta}_v = M^{-1} N(v).
\]

(5.19)

A plug-in estimation of the prediction variance \( \sigma_{s_0}^2(v) = V(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) \) is given by

\[
\hat{\sigma}_{s_0}^2(v) = \hat{\beta}_v^T M \hat{\beta}_v + D(v) - 2 \hat{\beta}_v^T N(v),
\]

(5.20)

where matrix \( D(v) \) defined as in equation (5.16) is calculated using an estimate of \( \hat{V}(a_0) \) obtained by means of the fitted LMC.
5.2 Functional Kriging: Total model

To define the functional kriging (total model) predictor (FKTM), we assume the same stationarity and isotropy assumptions given in the first paragraph of Section 5.1. The cokriging predictor given in (5.3) is defined for a specific \( v \in T \). If we want to predict the whole curve at \( s_0 \), the functional parameter \( \lambda_i(v) \) in (5.3) is replaced by a double indexed functional parameter \( \lambda_i(t, v) \). Thus, the predictor of the whole curve is given by

\[
\hat{\chi}_{s_0}(v) = \sum_{i=1}^{n} \int_T \lambda_i(t, v) \chi_{s_i}(t) dt, \quad v \in T,
\]

(5.21)

such that \( \lambda_1(t, v), \ldots, \lambda_n(t, v) : T \times T \rightarrow \mathbb{R} \). Note that, according to the replacements given in the schemes (5.1) and (5.2), the predictor given in (2.10) is extended to the functional context by replacing the vector \( \hat{Z}_1(s_0), \ldots, \hat{Z}_m(s_0) \) by \( \hat{\chi}_{s_0}(v), v \in T \), and the vectors of parameters \( (\lambda_1(t) \cdots \lambda_n(t)), \ldots, (\lambda_1^m(t) \cdots \lambda_n^m(t)) \) by \( (\lambda_1^1(t) \cdots \lambda_1^m(t)), \ldots, (\lambda_n^1(t) \cdots \lambda_n^m(t)) \), \( t \in T \), respectively. In addition, taking into account that \( v \) varies continuously in \( T \), the discrete set of functional parameters \( (\lambda_1^1(t) \cdots \lambda_1^m(t)), i = 1, \ldots, n \), is replaced by a double indexed functional parameter \( \lambda_i(t, v) \), \( t, v \in T \). It is clear that for a fixed \( v \in T \) the expression of the FKTM predictor in equation (5.21) is equal to the CBFD predictor given in (5.3). The functional parameter \( \lambda_i(t, v) \) in (5.21) determines the impact of the \( i \)-th observed function at time \( t \) on an unobserved function at time \( v \). This modeling approach is coherent with the functional linear model for functional responses (total model) shown in equation (2.35) In that framework the estimation of functional parameters is carried out by solving (Ramsay and Silverman, 2005)

\[
\min_{\alpha(\cdot), \beta_1(\cdot), \ldots, \beta_m(\cdot)} E \| \hat{Y}(v) - Y(v) \|^2.
\]

In our context the covariates are the observed curves in \( n \) sites of a region and the functional response is an unobserved function at an unvisited location. Consequently our objective function is

\[
E \| \hat{\chi}_{s_0}(v) - \chi_{s_0}(v) \|^2;
\]

depending on \( \lambda_1(\cdot, \cdot), \ldots, \lambda_n(\cdot, \cdot) \), or by using Fubini’s Theorem

\[
\int_T E \left( \hat{\chi}_{s_0}(v) - \chi_{s_0}(v) \right)^2 dv.
\]
Considering stationarity the objective function becomes

$$\int_T V (\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) \, dv.$$ 

Again the functional parameters $\lambda_i(t, v)$ in (5.21) are estimated taking into account constraints of unbiasedness and minimum prediction variance. Thus the optimization problem is

$$\min_{\lambda_1(\cdot, \cdot), \ldots, \lambda_n(\cdot, \cdot)} \int_T V (\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) \, dv \text{ s.t. } E(\hat{\chi}_{s_0}(v)) = E(\chi_{s_0}(v)), \forall v \in T. \quad (5.22)$$

We solve this problem using an approach based on basis functions. We expand functional variables as in the equation (4.4) and the bivariate functional parameters by

$$\lambda_i(t, v) = \sum_{j=1}^K \sum_{l=1}^K c_{jl} B_j(t) B_l(v) = B^T(t) C_i B(v), \quad (5.23)$$

where

$$C_i = \begin{pmatrix}
    c_{11} & c_{12} & \cdots & c_{1K} \\
    c_{21} & c_{22} & \cdots & c_{2K} \\
    \vdots & \vdots & \ddots & \vdots \\
    c_{K1} & c_{K2} & \cdots & c_{KK}
\end{pmatrix}_{(K \times K)}.$$

From equations (4.4) and (5.23) the predictor (5.21) can be expressed as

$$\hat{\chi}_{s_0}(v) = \sum_{i=1}^n \int_T a_i^T B(t) B^T(t) C_i B(v) \, dt$$

$$= \sum_{i=1}^n a_i^T W C_i B(v) = B^T(v) \sum_{i=1}^n C_i^T W a_i$$

$$= B^T(v) \sum_{i=1}^n C_i^T W a_i = B^T(v) \hat{a}_0, \quad (5.24)$$

where the matrix of inner products $W$ is defined as in equation (5.6). The predictor (5.21) is also considered by Nerini and Monestiez (2008). In the work of Nerini and Monestiez (2008) it is assumed that $W$ is a identity matrix because they consider a solution based on
orthonormal basis expansions. In our solution this is not a necessary condition. Now we consider the unbiasedness and minimum variance properties of the proposed predictor. As in Section 5.1, we assume that the coefficients \( a_i \) in the equation (5.24) follow a stationary multivariable random field. Consequently from equation (5.9) the expected value of the the curve on an unvisited site \( s_0 \) is given by

\[
E(\chi_{s_0}(v)) = E \left( \sum_{j=1}^{K} a_{0j} B_j(v) \right) = E(B^T(v)a_0) = B^T(v)E(a_0) = B^T(v)\vartheta \text{ for all } v \in T. \tag{5.25}
\]

On the other hand taking expected value in (5.24) we have

\[
E(\hat{\chi}_{s_0}(v)) = B^T(v) \sum_{i=1}^{n} C_i^T W E(a_i) = B^T(v) \sum_{i=1}^{n} C_i^T W \vartheta \text{ for all } v \in T. \tag{5.26}
\]

Consequently from equations (5.25) and (5.26) we note that the predictor (5.21) is unbiased if and only if

\[
B^T(v) \sum_{i=1}^{n} C_i^T W \vartheta = B^T(v)\vartheta \text{ for all } v \in T, \tag{5.27}
\]

that is, if and only if,

\[
\sum_{i=1}^{n} C_i^T W \vartheta = \vartheta.
\]

This condition is equivalent to

\[
\sum_{i=1}^{n} C_i = W^{-1}. \tag{5.28}
\]

The \( n \) functional parameters in the predictor (5.21) are given by the solution of the following optimization problem

\[
\min_{C_1, \ldots, C_n} \int_T V(B^T(v)a_0 - B^T(v)a_0)dv \text{ s.t. } \sum_{i=1}^{n} C_i = W^{-1}. \tag{5.29}
\]
Developing the integral in the objective function (5.29) we obtain
\[
\int_T B^T(v)(a_0 - a_0)B(v)dv = \int_T Tr\left( B^T(V(\hat{a}_0 - a_0)B(v)) \right) dv = Tr\left( V(\hat{a}_0 - a_0) \int_T B(v)B^T(v)dv \right) = Tr(V(\hat{a}_0 - a_0)W). \tag{5.30}
\]

The variance in (5.30) is
\[
V(\hat{a}_0 - a_0) = \sum_{i=1}^{n} C_i^T W V(a_i) W C_i + 2 \sum_{i<j}^{n} C_i^T W C(a_i, a_j) W C_j + V(a_0) - 2 \sum_{i=1}^{n} C_i^T N_i W. \tag{5.31}
\]

From (5.31) and defining the following matrices
\[
Q_{ii} = (WV(a_i)W), Q_{ij} = (WC(a_i, a_j)W), N_i = WC(a_i, a_0)
\]
the optimization problem (5.29) can be expressed as
\[
\min_{C_1, \ldots, C_n, m} \sum_{i=1}^{n} Tr\left( C_i^T Q_{ii} C_i W \right) + 2 \sum_{i<j}^{n} Tr\left( C_i^T Q_{ij} C_j W \right) + Tr(V(a_0)W) - 2 \sum_{i=1}^{n} Tr\left( C_i^T N_i W \right) - 2m \left( \sum_{i=1}^{n} C_i - W^{-1} \right). \tag{5.32}
\]

Derivatives with respect \( C_i, i = 1, \ldots, n \), and \( m \) in (5.32) are given by
\[
\frac{\partial}{\partial C_i} = 2 \sum_{j=1}^{n} Q_{ij} C_j W - 2N_i W + 2m, \quad \frac{\partial}{\partial m} = \sum_{i=1}^{n} C_i - W^{-1}. \tag{5.33}
\]
The solution of the problem given in (5.32) is achieved by setting these derivatives equal to zero. This solution can be represented in matrix notation as

\[
\begin{pmatrix}
Q_{11} & Q_{12} & \cdots & Q_{1n} & I \\
Q_{21} & Q_{22} & \cdots & Q_{2n} & I \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
Q_{n1} & Q_{n2} & \cdots & Q_{nn} & I \\
I & I & \cdots & I & 0
\end{pmatrix}
\begin{pmatrix}
C_1 \\
C_2 \\
\vdots \\
C_n \\
m^* \\
W^{-1}
\end{pmatrix} = \begin{pmatrix}
N_1 \\
N_2 \\
\vdots \\
N_n \\
0
\end{pmatrix},
\]

(5.34)

where \( m^* = mW^{-1} \). From equations (5.30) and (5.31), an estimation of the integrated prediction variance

\[
\sigma^2_{int} = \int_T \sigma^2_{s_0}(v)dv = \int_T V(\hat{\chi}_{s_0}(v) - \chi_{s_0}(v)) dv
\]

is given by

\[
\sigma^2_{int} = \sum_{i=1}^n Tr \left(C_i^T C_i W \right) + 2 \sum_{i<j}^n Tr \left(C_i^T C_{ij} C_j W \right) \\
+ Tr(V(a_0)W) - 2 \sum_{i=1}^n Tr \left(C_i N_i W \right),
\]

(5.35)

where the matrices \( \hat{C}_1, \ldots, \hat{C}_n \) are obtained by solving the system (5.34).

In the remainder of this section, we discuss about the connection between the two methods introduced in the thesis. The relationship between CBFD and FKTM is analogous to that of cokriging analysis and multivariable spatial prediction, in the sense that the prediction obtained by CBFD at a time is identical to the prediction achieved by FKTM at the same time. The expressions of these predictors, the unbiasedness constraints and the respective objective functions, are equivalent at each fixed time \( v \). In fact, the next Proposition establishes the equivalence between both alternatives when, as we propose, basis function expansion is used.

**Proposition 1.** Assume that representation (4.4) is correct for \( \chi_{s_i}, i = 0, 1, \ldots, n \). Let \( \hat{\chi}^C_{s_0}(v), v \in T, \) be the set of predictors for \( \chi_{s_0}(v), v \in T, \) derived from equations (5.3) and (5.4), where the coefficients \( b_{iv}, i = 0, 1, \ldots, n, \) are the solution of problem (5.17) for each
v ∈ T. Let \( \hat{\chi}^F_{s_0}(v), v ∈ T, \) be the predicted function given by equations (5.21) and (5.23), where coefficient matrices \( C_i, i = 0, 1, \ldots, n, \) are the solution of problem (5.32). Then

\[
\hat{\chi}^C_{s_0}(v) = \hat{\chi}^F_{s_0}(v)
\]

for all \( v ∈ T \) and all \( s_0 ∈ D. \)

**Proof.**

Let us define the following sets of functions: \( \mathcal{F} = \{λ : T × T → R\}, \mathcal{F}^n = \mathcal{F} × \cdots × \mathcal{F}, \)

\[
\mathcal{L}_C = \{λ ∈ \mathcal{F}^n : λ_i(t, v) = b_i^T(t)B_i(v), b_i(v) ∈ R^K, v ∈ T, t ∈ T, i ∈ \{1 \ldots n\}\},
\]

\[
\mathcal{L}_F = \{λ ∈ \mathcal{F}^n : λ_i(t, v) = B_i^T(v)C_iB_i(t), C_i ∈ R^{K×K}, v ∈ T, t ∈ T, i ∈ \{1 \ldots n\}\}.
\]

Observe that \( \mathcal{L}_F ⊈ \mathcal{L}_C. \)

Let \( \mathcal{F}_1 \) be the set of functions from \( T \) to \( R \) and let \( g \) be a generic function defined from \( \mathcal{F}_1^n × T \) to \( R. \) Let \( \mathcal{R} \) be a generic subset of \( \mathcal{F}^n \) such that \( \mathcal{L}_F ∩ \mathcal{R} ≠ ∅. \) For \( \mathcal{L} ∈ \{\mathcal{L}_C, \mathcal{L}_F\} \) we define the functionals

\[
Ψ^{(1)}_{\mathcal{L}}(g) = \int_T \min_{λ ∈ \mathcal{L} ∩ \mathcal{R}} g(λ_1(·, v), \ldots, λ_n(·, v), v) dv,
\]

\[
Ψ^{(2)}_{\mathcal{L}}(g) = \min_{λ ∈ \mathcal{L} ∩ \mathcal{R}} \int_T g(λ_1(·, v), \ldots, λ_n(·, v), v) dv.
\]

Observe that for all function \( g \) and for all \( \lambda^F ∈ \mathcal{L}_F ∩ \mathcal{R}, \)

\[
Ψ^{(1)}_{\mathcal{L}_C}(g) ≤ Ψ^{(1)}_{\mathcal{L}_F}(g) ≤ Ψ^{(2)}_{\mathcal{L}_F}(g) ≤ \int_T g(λ_1^F(·, v), \ldots, λ_n^F(·, v), v) dv,
\]

and

\[
Ψ^{(1)}_{\mathcal{L}_C}(g) ≤ Ψ^{(2)}_{\mathcal{L}_C}(g) ≤ Ψ^{(2)}_{\mathcal{L}_F}(g) ≤ \int_T g(λ_1^F(·, v), \ldots, λ_n^C(·, v), v) dv.
\]

Let \( \lambda^{C∗} \) be the optimal element in \( \mathcal{L}_C ∩ \mathcal{R} \) such that

\[
\min_{λ ∈ \mathcal{L}_C ∩ \mathcal{R}} g(λ_1(·, v), \ldots, λ_n(·, v), v) = g(λ_1^{C∗}(·, v), \ldots, λ_n^{C∗}(·, v), v),
\]

or equivalently

\[
\int_T g(λ_1^{C∗}(·, v), \ldots, λ_n^{C∗}(·, v), v) dv = Ψ^{(1)}_{\mathcal{L}_C}(g).
\]
If it were possible to prove that in fact $\lambda^{C*}$ is an element of $\mathcal{L}_F \cap \mathcal{R}$, then it would follow that

$$\int_T g(\lambda_1^{C*}(. , v), \ldots, \lambda_n^{C*}(. , v), v)dv = \Psi_{\mathcal{L}_F}^{(1)}(g) \leq \Psi_{\mathcal{L}_F}^{(2)}(g) \leq \int_T g(\lambda_1^{C*}(. , v), \ldots, \lambda_n^{C*}(. , v), v)dv$$

and then the inequalities would in fact be equalities. We would therefore conclude that the minima defining $\Psi_{\mathcal{L}_F}^{(1)}(g)$ and $\Psi_{\mathcal{L}_F}^{(2)}(g)$ are achieved at the same array of functions $\lambda^{C*}$.

We apply this result to the function $g$ defined as

$$g(\lambda_1(. , v), \ldots, \lambda_n(. , v), v) = V\left(\sum_{i=1}^{n} \int_T \lambda_i(t, v)\chi_{s_i}(t)dt - \chi_{s_0}(v)\right)$$

and the subset $\mathcal{R}$ of unbiased predictors,

$$\mathcal{R} = \left\{ \lambda \in \mathcal{F}^n : E\left(\sum_{i=1}^{n} \int_T \lambda_i(t, v)\chi_{s_i}(t)dt\right) = E(\chi_{s_0}(v)), \text{ for all } v \in T \right\}.$$ 

Then we only need to prove that the optimal coefficients $\lambda_i(v, t) = \lambda^*_i(t)$ defined in equation (5.4) as $b_{iw}^T B(t)$, are in fact of the form $B^T(v)\hat{C}_i B(t)$. Thus, we need to prove that the optimal $b_{iw}$ are

$$\hat{b}_{iw} = \hat{C}_i^T B(v)$$

for some matrix $\hat{C}_i \in \mathbb{R}^{K \times K}$. However, this is true from equations (5.13), (5.18) and (5.19), which establish that

$$\left(b_{1w}^T, \ldots, b_{nw}^T, m_c^T\right)^T = M^{-1} N(v) = M^{-1} \left(\begin{array}{c} WC(a_0, a_1) \\ \vdots \\ WC(a_0, a_n) \\ W^{-1} \end{array}\right) B(v),$$

and the proof is concluded. □

The difference between solving an infinite number of point-wise predictions by CBFD and doing only one prediction of a whole curve by FKTM is due to the use of a double expansion in terms of basis functions for the functional parameters of FKTM. A distinctive feature between these methodologies is given in terms of their prediction variances. The
estimated prediction variance of FKTM in equation (5.35) can be used as a global measure of uncertainty on the prediction of a whole curve whereas the estimated prediction variance of CBFD given in equation (5.20) can be used in a classical sense, that is, we can, for instance, calculate confidence intervals for the prediction.

5.3 Application: Spatial prediction of temperature curves in Canada’s Maritime provinces

We use the Canada’s Maritime provinces data set from Section 1.1.3 for giving an application of the methodologies developed in this Chapter. The another temperature data set (Section 1.1.2) is used in next section for doing a comparison among the methodologies proposed in the thesis. In Chapter 4 was concluded that a basis with 65 functions can be appropriate for smoothing these data sets. We use in this Section the same number of basis functions. From Proposition (1) we know that predictions obtained by CBFD coincide with those given by FKTM. So we only show the results obtained by FKTM. For each case, an LMC was estimated and used for calculating the matrices $Q_{ij}, N_i, i, j = 1, \ldots, 35$, in equation (5.34) and consequently for estimating the parameters given in that system. For fitting the LMC all single (direct) variograms and cross-variograms were modeled as linear combination of nugget and exponential models. We use the library gstat of R language (Pebesma, 2004) for fulfill this task. Figure 5.1 displays the cross-validation predictions, the smoothed curves (by using a 65 Fourier basis functions) and the prediction at Moncton (an unvisited weather station, Figure 1.2). A graphical comparison between smoothed and predicted curves shows the good performance of the predictions (Figure 5.1).

Figure 5.2 (right panel) shows cross-validation residuals. We can see in this figure that there are good predictions in a high proportion of sites (those having residuals around zero). Large positive or negative residuals are obtained in only a small number of stations. In wintertime the residuals are bigger than other seasons (residuals higher than 4 °C and lower than -4 °C at the beginning and at the end of the year); that is, we have greater uncertainty in the prediction during this period. This can be due to the fact that the
observed temperature curves have more variability in this time of year (Figure 5.2, left panel). The residual standard deviation is lower in the summer (Figure 5.2, right panel) where the smoothed and the predicted curves have less variation (Figure 5.2, left panel). The residual mean varies around zero which indicates that the predictions are unbiased (Figure 5.2, right panel).

Bertrand and Bathurst in the north eastern area of New Brunswick (Figure 1.3) are the stations with the greatest cross-validation residuals. Though the average daily temperature at these stations has a very similar behavior throughout the time, the difference between its values on some days (19, 49 to 57, 353, 354 and 356) was greater than 4 °C. This generate high residuals in both cases taking into account that in the cross-validation analysis the prediction at Bertrand is very influenced by the curve of Bathurst and vice-versa because these sites are relatively close (Figure 1.3). With respect to the prediction at Moncton, we observe in Figure 5.1 (right panel) that the predicted curve both shows a seasonal behavior similar to the smoothed curves and is consistent with real values recorded for this weather
station. Three of the estimated parameters for doing this prediction are shown in Figure 5.3. These correspond to Bouctouche, Nappan and Aroostook. In Figure 5.3 we also show the estimated functional parameters for all the stations when we do prediction at Moncton on 9th April (day 100). Two relevant aspects can be highlighted from Figure 5.3. First, Bouctouche and Nappan stations have greater influence on the prediction than Aroostook (whose functional parameter is almost null). Secondly, the estimated functional parameters reveal that there is a short temporal effect on the prediction. Both results are expected. On the one hand, the first result is coherent with the geostatistical philosophy, that is, sites closer to the prediction location have greater influence than others further apart. Bouctouche and Nappan, which are 40 km and 56 km from Moncton, respectively, are the nearest stations to this site in the data set (Figure 1.2) whereas Aroostook located in western New Brunswick is one of the most distant stations (approximately 310 kilometers from Moncton). As in the case of Aroostock, other stations very separated from Moncton have a low influence on the prediction. On the other hand, an analysis
Figure 5.3: Top: Estimated functional parameters for Bouctouche (NB) and Nappan (NB). Bottom: Estimated functional parameter for Aroostook (NB) and functional parameters for all stations when we do prediction at Moncton on 9th April (day 100).
of time series do not included in the thesis indicates that the temperature time series considered (after differences) have low autoregressive orders. This feature is reflected in Figure 5.3 (bottom, right panel) where the estimated parameters reveal a temporal effect on the prediction even lower than 10 days.

In Figure 5.4 we show the estimated functional parameters for doing prediction at Moncton on 18th July by means of CBFD (left panel) and FKTM (right panel). We see in this figure that the estimations coincide, that is, for $v = 200$ it is graphically verified the relationship $\hat{\lambda}_i^{200}(t) = \hat{\lambda}_i(t, 200), \, i = 1, \ldots, 35$, where $\hat{\lambda}_i^{200}(t)$ and $\hat{\lambda}_i(t, 200)$ are estimated functional parameters obtained by CBFD and FKTM, respectively. This result is in concordance with the remarks given at the end of Section 5.2. In both panels of Figure 5.4 the greatest estimated parameter (with values near to 0.09 on 18th July) corresponds to Bouctouche (the nearest station to Moncton). From Figures 5.3 (bottom, right) and 5.4 we see that the effect of changing $v$ is only reflected at the time when the estimated parameters get their larger values but not in the shape of their curves.
5.4 A comparison with ordinary kriging for function-value data and continuous time-varying kriging

We use the data sets described in Sections 1.1.2 and 1.1.3 and the summary statistics of $SSE_{FCV}(i), i = 1, \ldots, 35$, in equation (3.13) for comparing the methodologies proposed in the thesis. We evaluate on $j = 1, \ldots, 365$ the cross-validation predictions obtained by OKFD, CTKFD, and FKTM. For including the CBFD predictor into this analysis we calculate $\hat{\chi}_0(v)$ in equation (5.3) on $v = 1, \ldots, 365$. Summary statistics of $SSE_{FCV}$ values obtained by CBFD and FKTM coincide because predictions by CBFD are calculated at the same arguments where are evaluated the curves predicted by FKTM. For the Canadian data set, we initially detrended the data by using a functional regression model (see Section 4.4). We apply the spatial prediction methods proposed to residuals of regression. The temperature predictions for these data set are obtained adding the residual predictions to the trend.

Table 5.1: Summary statistics of sum of squared errors of functional cross-validation for the Canada’s Maritime provinces data set. OKFD: Ordinary kriging for function-valued spatial data; CTVFD: Continuous time-varying kriging for functional data; FKTM: Functional kriging (total model), which equals CBFD CBFD (Cokriging based on functional data).

<table>
<thead>
<tr>
<th>Statistic</th>
<th>OKFD</th>
<th>CTKFD</th>
<th>FKTM (or CBFD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>103.7</td>
<td>104.4</td>
<td>104.3</td>
</tr>
<tr>
<td>Median</td>
<td>253.4</td>
<td>252.7</td>
<td>252.8</td>
</tr>
<tr>
<td>Mean</td>
<td>299.5</td>
<td>299.2</td>
<td>298.9</td>
</tr>
<tr>
<td>Maximum</td>
<td>890.8</td>
<td>902.1</td>
<td>899.8</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>178.4</td>
<td>175.4</td>
<td>176.1</td>
</tr>
<tr>
<td>Sum</td>
<td>10483</td>
<td>10471</td>
<td>10461</td>
</tr>
</tbody>
</table>

From Table 5.1 we conclude that the all methods have similar performances with the Canada’s Maritime provinces data set. On the other hand respect to the Canadian data set we note in Table 5.2 that there are small differences among the methods in terms of
minimum or median values. This indicates that all of them have a similar behavior at well-predicted stations as Toronto, Quebec, Edmonton or Bagottville. In 23 stations (from 35) the $SSE_{FCV}$ are very similar and lower than 1000 by using either of three methods. High $SSE_{FCV}$ values (greater than 3000) are due to predictions at stations located on the coast as Pr Rupert or very separated from the remaining ones as Resolute, Iqaluit, Dawson, St Johns or Inuvik. The predictors have a similar behavior when the curves are relatively homogeneous and the differences among them with the Canadian data set are essentially due to their performance in stations with extreme winters (warm or cold) as before mentioned. The functional regression model fitted in section 4.4 do not estimate well the temperature at these stations (Figure 4.7). This result is also reflected in the cross-validation predictions.

Table 5.2: Summary statistics of sum of squared errors of functional cross-validation for the Canadian temperature data set. Smoothed data were detrended by using a functional regression model and the prediction methods were applied to residuals of regression. OKFD: Ordinary kriging for function-valued spatial data; CTVFD: Continuous time-varying kriging for functional data; FKT: Functional kriging (total model), which equals CBFD (Cokriging based on functional data).

<table>
<thead>
<tr>
<th>Statistic</th>
<th>OKFD</th>
<th>CTKFD</th>
<th>FKT (or CBFD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>177.2</td>
<td>170.0</td>
<td>170.9</td>
</tr>
<tr>
<td>Median</td>
<td>531.8</td>
<td>511.5</td>
<td>515.9</td>
</tr>
<tr>
<td>Mean</td>
<td>1575.0</td>
<td>1680.0</td>
<td>1675.0</td>
</tr>
<tr>
<td>Maximum</td>
<td>7363.0</td>
<td>10990.0</td>
<td>11056.0</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1902.3</td>
<td>2410.9</td>
<td>2419.3</td>
</tr>
<tr>
<td>Sum</td>
<td>55129</td>
<td>58796</td>
<td>58622</td>
</tr>
</tbody>
</table>

With the Canada’s Maritime provinces data set the sum of $SSE_{FCV}$ values (Tables 5.1) indicate that FKT (CBFD) has slightly better performance than other predictors. The results shown in Tables 5.1 indicate that the inclusion of double indexed functional parameters (temporal effects in the data sets studied) in the analysis is reflected in better
predictions. The differences between the predictors are smaller in the Canada’s Maritime provinces data set than in the Canadian data set. This result indicates that when the curves are homogeneous we could use any of three approaches. However if the observed data have high variability (as in the Canadian data set) OKFD could be a much better option.

In summary we can conclude that FKTM is the best option for carry out spatial prediction of functional data with the Canada’s Maritime provinces data set. CBFD can be used complementary if we want to do univariate inference. OKFD is a better option for the Canada data set. In both cases (two data sets) the differences among the methods are very small in a high proportion of sites. Consequently all of them are good alternatives for doing spatial prediction of functional data. Taking into account that OKFD is simpler than others from a practical and computational point of view in applications with large data sets this method can be preferable.
Chapter 6

Conclusions and further research

In this research we propose four methods for doing spatial prediction of functional data. We have enhanced the curve kriging predictor proposed by Goulard and Voltz (1993) by introducing a non-parametric smoothing step. We have introduced functional cross-validation to automatically choose the smoothing parameters, and we prove that both proposals coincide when using such an approach. A minimization criterion given in multivariable geostatistics has been adapted to the functional context. We develop the continuous time-varying kriging predictor which is based on the functional linear concurrent model. We also have shown how two predictors used in multivariable geostatistics can be extended to the functional context. The first of them called cokriging based on curves can be used with the same objective of classical multivariable cokriging (univariate prediction) but considering as auxiliary information curves instead of vectors. The second one functional kriging: total model allows to predict a whole curve at an unvisited site by considering a double indexed functional parameter. We propose the use of basis functions in order to estimate functional parameters. In our approaches the use of a orthonormal basis is computationally advantageous. However this is not a necessary condition for applying our method.

Our approaches were applied to agronomic and climatological data. The cross-validations results show a good performance of the proposed predictors, and indicate from a descriptive point of view that they can be used as methods for modeling spatially correlated
functional data. FKTM show better performance than OKFD and CTVKFD for data showing high variability (Canadian data set). For more homogeneous data (Canadian maritime data set) all predictors considered show practically the same performance.

There is still a long way of research necessary for spatial prediction of functional data. We have used B-splines and Fourier basis functions. However the use of other basis system to get functional data from discrete observations can be considered. Alternative methods of estimating the empirical trace-variogram, for instance, by using robust estimators (Cressie, 1993) or kernel-based estimation methods (Yu et al., 2005) could be used. The short temporal effect detected by the estimated parameter in our applications suggest to consider only parameters defined over the domain \( v - \delta \leq t \leq v + \delta \), where \( \delta \) is a time lag over which we use information for carry out the prediction. Further research is needed for establishing the performance of the predictors proposed under different levels of temporal cross-correlation. We have assumed stationarity and isotropy for developing the methods. A generalization to the case where the mean function change inside the study area is required. Models for carrying out spatial prediction based on information of several functional variables, that is, two or more functional variables observed at each sampling location, could also be studied.
Bibliography


Leiva, F. (2003). Manejo sostenible del suelo bajo la concepción de agricultura de precisión. Research project partially supported by COLCIENCIAS, Colombia.


Appendix A

R functions

The Appendix contains a brief description of R codes (including libraries and functions) used in the thesis. For simplicity we only show the code used for modeling the Canadian Maritimes data set described in Chapter 1. All code used in the thesis can be obtained from http://www.docentes.unal.edu.co/rgiralddoh/docs/

A.1 Code for modeling the Canada’s Maritime provinces data set by OKFD

#################################################################
# Code R ("main.okfd.maritimes.R") for carrying out prediction by
# OKFD. The function "okfd.R" enables to do prediction by OKFD
# and with the function "okfd.cv", we can do functional
# cross-validation. Data are smoothed by using Fourier basis
# functions. A spherical model is used for modeling the
# trace-variogram.
#################################################################
# Libraries and functions

```r
rmlist(ls())
library(fda)
library(geoR)
source("okfd.R")
source("okfd.cv.R")
```

# Data and coordinates

```r
# Coordinates and distances

coord <- matrix(scan("Coordenadas.txt", 0, dec="."), 35, 2, byrow=TRUE)
dista <- dist(coord, method="euclidean", diag=TRUE, upper=TRUE)
d <- as.matrix(dista)

# Distances to Moncton

dista.cero <- read.table("coordcero.txt", header=FALSE)
d.cero <- as.matrix(dista.cero)

# Data

temp <- matrix(scan("dailtemp.txt", 0, dec=","), 365, 35, byrow=TRUE)
day <- matrix(scan("day.txt", 0), 365, 35, byrow=TRUE)
matplot(temp, type="l", lty=1, col=1, ylab="Temperature (degrees C)", xlab="Day")

abline(h=0, lty=2)

"Pugwash", "Shearwater", "Charlottetown", "Summerside",
"Saint John", "Baddeck ", "Cheticamp", "Bridgewater",
"Middle Musquodoboit")
dimnames(temp) <- list(NULL,place)

#################################################################
# Prediction at Moncton by using the function "okfd"
#################################################################
coord.cero <- matrix(c(-64.69, 46.10),nrow=1,ncol=2)
n<-dim(temp)[1]
nbasis<-65
argvals<-seq(1,n, by=1)
okfd.res<-okfd(coord=coord,data=temp,argvals=argvals,nbasis=nbasis,
new.coord=coord.cero)
names(okfd.res)
plot(okfd.res$datafd,lty=1,col=8, xlab="Day", ylab="Temperature
(Degres C)"
lines(okfd.res$krig.new.data,col=1,lwd=2)
moncton<-matrix(scan("PromediosMoncton.txt",0, dec=",") , 365, 1,
byrow=TRUE)
lines(moncton, type="p", pch=20,cex=0.6)
legend(0, 20, legend = c("Smoothed", "Prediction", "Real data"),
  lty = c(1,1,-1), pch=c(-1,-1,20),lwd = c(1,1,1),
col=c(8,1,1),pt.cex =0.5)

#################################################################
# Prediction variance
#################################################################
okfd.res$pred.var

#################################################################
# Sum of Squared Errors (SSE) analysis

residual <- moncton-okfd.res$krig.new.data
residual2 <- residual * residual
summary(residual2)
sd(residual2)
sum(residual2)

# Cross-validation analysis using the function "okfd.cv"

n <- dim(temp)[1]
nbasis <- 65
argvals <- seq(1, n, by = 1)
array.nbasis <- seq(65, 65, by = 10)
okfd.cv.res <- okfd.cv(coord = coord, data = temp, argvals = argvals,
                        array.nbasis = array.nbasis,
                        max.dist.variogram = NULL, nugget.fix = NULL)
write.table(okfd.cv.res$MSE.CV, file = "MSE.CV.txt",
            row.names = FALSE, quote = FALSE, append = TRUE)
MSE.CV <- matrix(scan("MSE.CV.txt"), 365, 1, byrow = TRUE)

# Results and plots of cross-validation analysis

okfd.cv.res$k.opt
resultsMSE.CV <- cbind(array.nbasis, okfd.cv.res$MSE.CV)
plot(array.nbasis, okfd.cv.res$MSE.CV, type = "l",
     xlab = "Number of basis functions", ylab = "MSE of cross-validation")
matplot(okfd.cv.res$krig.CV[,], type = "l", col = 1,
        lty = 1, ylim = c(-15, 20), ylab = "Temperature (Degrees C)", xlab = "Day")
abline(h = 0, lty = 2)
residuals <- okfd.cv.res$krig.CV[1,,]-temp
SSE.stations <- apply(residuals*residuals,2,sum)
SSE.stations <- as.matrix(SSE.stations)
SSE.stations
write.table(SSE.stations,"SSEOKFD65.txt",quote=FALSE)
summary(SSE.stations)
sd(SSE.stations)
sum(SSE.stations)

# Plot of cross-validation predictions

k <- 65
n<-dim(temp)[1]
evalarg<-seq(1,n, by=1)
range <- c(1,n)
period <- n
nbasis <- k
basis <- create.fourier.basis(range, nbasis, period)
datafd <- data2fd(temp,argvals=evalarg,basis)
ajustes<-eval.fd(evalarg,datafd,Lfdobj=0)
predcv<-okfd.cv.res$krig.CV[1,,]
plot(temp[,27], main="Aroostook", type="p",lty=1, ylim=c(-15,20),
     lwd=1, xlab="Day", ylab="Temperature (degrees C)",cex=.5)
lines(ajustes[,27], type="l", lty=1, lwd=2, col=1)
lines(predcv[,27], type="l", lty=1, lwd=1, col=1)
legend(0, 20, legend = c("Observed", "Smoothed", "Predicted"),
       lty = c(-1,1,1), pch=c(1,-1,-1),lwd = c(1,2,1),
       col=c(1,1,1),pt.cex =0.5)
abline(h=0,lty=2)
residual<-temp[,27]-predcv[,27]
summary(residual)
summary(residual*residual)
plot(residual*residual,type="l", ylim=c(0,20))

A.2 Code for modeling the Canada’s Maritime provinces data set by CTKFD

# Code R ("main.pwkfd.maritimes.R") for carrying out prediction
# by point-wise kriging (continuous time varying kriging).
# This code enables to do prediction at an unvisited site (Moncton)
# by means of the function "pwkfd.R" and cross-validation analysis
# by means of the function "pwkfd.cv.R".
# Data are smoothed by using Fourier basis functions.
# All single (direct) variograms and cross-variograms of coefficients of Fourier basis are modeled as a linear combination of
# nugget and exponential models.

# Libraries and functions

rm(list=ls())
memory.limit(size=4000)
library(fda)
library(gstat)
source("chooseK.R")
source("fit.lmc.R")
source("parameters.lmc.R")
source("covariances.R")
source("product.fourier.basis.R")
source("integral.R")
source("matrixQ.pwkfd.R")
source("covariances.pred.R")
source("matrixJ.pwkfd.R")
source("functional.parameters.pwkfd.R")
source("prediction.pwkfd.R")
source("prediction.variance.R")
source("pwkfd.R")
source("pwkfd.cv.R")

################################################################################
# Data and coordinates
################################################################################
# Coordinates and distances
coord <- matrix(scan("Coordenadas.txt",0, dec="."), 35, 2, byrow=TRUE)
dista <- dist(coord, method="euclidean", diag=TRUE, upper=TRUE)
d <- as.matrix(dista)
# Distances to the predictions site
coord.cero<-read.table("coordcero.txt", header=FALSE)
d.cero <- as.matrix(coord.cero)
# Data
temp <- matrix(scan("dailtemp.txt",0, dec=""), 365, 35,byrow=TRUE)
day  <- matrix(scan("day.txt",0), 365, 35, byrow=TRUE)
matplot(temp, type="l", lty=1,col=1, ylab="Temperature (degrees C)",
       xlab="Day")
abline(h=0,lty=2)
place <- c(
"Fredericton", "Halifax", "Sydney ", "Miramichi", "Kentville",


"Bathurst", "Bertrand", "Bouctouche", "Sussex", "Gagetown",
"Liverpool", "Truro", "Greenwood", "Ingonish", "Parrsboro",
"Pugwash", "Shearwater", "Charlottetown", "Summerside",
"Saint John", "Baddeck", "Cheticamp", "Bridgewater",
"Middle Musquodoboit")
dimnames(temp) <- list(NULL, place)

# Choosing K by non-parametric cross-validation
# Cross-validation analysis by estimating the linear model of coregionalization for each iteration. We use the function "pwkfd.cv"

sec.k <- seq(5, 325, by=10)
n <- dim(temp)[1]
evalarg <- seq(1, n, by=1)
k.optimo <- escogencia.k(sec.k=sec.k, data=temp, evalarg)
SSE <- k.optimo$SSE
nbasis <- as.matrix(k.optimo$nbasis)
SSE <- cbind(nbasis, SSE)
plot(SSE, type="l", xlab="Number of basis functions", main="Non-parametric CV", ylab="SSE")
write.table(SSE, "SSE.txt", quote=FALSE)

n <- dim(temp)[1]
period <- n
evalarg <- seq(1, n, by=1)
pwkfd.cv.res <- pwkfd.cv(data=temp, evalarg, period=period, d, array.k)
plot(pwkfd.cv.res$MSE.CV,type="l", xlab="Number of basis functions", main="Functional CV", ylab="SSE")
predcv<-pwkfd.cv.res$krig.CV
predcv<-matrix(predcv,nrow=35,ncol=365)
predcv<-t(predcv)
write.table(predcv, file="valcruz65.txt", quote=FALSE, row.names=FALSE)
residuals<-temp-predcv
SSE.stations<-apply(residuals*residuals,2,sum)
SSE.stations<-as.matrix(SSE.stations)
SSE.stations
write.table(SSE.stations,"SSEPWFkFd.text",quote=FALSE)
summary(SSE.stations)
sum(SSE.stations)
sd(SSE.stations)
varpred<-pwkfd.cv.res$varpred.CV
write.table(varpred,"varpred65.txt",quote=FALSE)
dimnames(varpred) <- list(place, NULL)
varpred
summary(varpred)

# Point-wise kriging with K=65 by using the function "pwkfd"

k <- pwkfd.cv.res$k.opt
n<-dim(temp)[1]
evalarg<-seq(1,n, by=1)
range <- c(1,n)
period <- n
nbasis <- k
basis <- create.fourier.basis(range, nbasis, period)
datafd <- data2fd(temp,argvals=evalarg,basis)
\begin{verbatim}
g.fit<- fitlmc(coord,datafd,k)
pwkfd.res<-pwkfd(g.fit,basis,evalarg,k,datafd,d,d.cero,equispaced=TRUE)

# Plots of prediction and estimated functional parameters

plot(datafd, lty=1,col=8, ylab="Temperature (degrees C)", xlab="Day")
lines(pwkfd.res$prediction,type="l",col=1,lty=1, lwd=2)
moncton<-matrix(scan("PromediosMoncton.txt",0, dec=""), 365, 1,
byrow=TRUE)
lines(moncton, type="p", pch=20,cex=0.6)
legend(0, 20, legend = c("Smoothed", "Prediction", "Real data"),
   lty = c(1,1,-1), pch=c(-1,-1,20),lwd = c(1,1,1),
   col=c(8,1,1),pt.cex =0.5)
matplot(pwkfd.res$funpar,type="l",col=1,lty=1,ylim=c(-0.025,1),
   ylab="Parameters", xlab="Day")
lines(apply(pwkfd.res$funpar,1,sum),lty=2)
dimnames(pwkfd.res$funpar)<-list(NULL,place)
summary(pwkfd.res$funpar)

# Prediction variance

varpred<-pwkfd.res$pred.var

# Sum of Squared Errors (SSE) analysis

moncton<-matrix(scan("PromediosMoncton.txt",0, dec=""), 365, 1,
byrow=TRUE)
residual<-moncton-pwkfd.res$prediction
summary(residual)
residual2<-residual*residual
\end{verbatim}
A.3 Code for modeling the Canada’s Maritime provinces data set by FKTM

# Code R ("main.fktm.maritimes.R" for carrying out prediction by functional kriging (total model).
# The code enables to do prediction at Moncton by means of the function "fktm.R" and cross-validation analysis by means of the function "fktm.cv.R".
# Data are smoothed by using Fourier basis functions.
# All single (direct) variograms and cross-variograms of coefficients of Fourier basis are modeled as a linear combination of nugget and exponential models.

# Libraries and functions

rm(list=ls())
library(fda)
library(gstat)
source("chooseK.R")
source("fit.lmc.R")
source("parameters.lmc.R")
source("covariances.R")
source("product.fourier.basis.R")
source("integral.R")
source("matrixQ.fktm.R")
source("covariances.pred.R")
source("matrixN.fktm.R")
source("unbiased.fktm.R")
source("prediction.fktm.R")
source("functional.parameters.fktm.R")
source("fktm.R")
source("matrixC.R")
source("fktm.vc.R")

#################################################################
# Data set (temp, day) and coordinates
#################################################################

# Coordinates and distances
coord <- matrix(scan("Coordenadas.txt", 0, dec="."), 35, 2, byrow=TRUE)
dista <- dist(coord, method="euclidean", diag=TRUE, upper=TRUE)
d <- as.matrix(dista)

# Distances to Moncton
coord.cero <- read.table("coordcero.txt", header=FALSE)
d.cero <- as.matrix(coord.cero)

# Data
temp <- matrix(scan("dailtemp.txt", 0, dec=","), 365, 35, byrow=TRUE)
day <- matrix(scan("day.txt", 0), 365, 35, byrow=TRUE)
matplot(temp, type="l", lty=1, col=1, ylab="Temperature (degrees C)", xlab="Day")
abline(h=0, lty=2)

place <- c(
  "Fredericton", "Halifax", "Sydney ", "Miramichi", "Kentville",
  "Bathurst", "Bertrand", "Bouctouche", "Sussex", "Gagetown",
  "Liverpool", "Truro", "Greenwood", "Ingonish", "Parrsboro",
)
"Pugwash", "Shearwater", "Charlottetown", "Summerside",  
"Saint John", "Baddeck ", "Cheticamp", "Bridgewater",  
"Middle Musquodoboit")

dimnames(temp) <- list(NULL,place)

# Prediction at Moncton with K=65 by using the function "fktm"

k <- 65  
n<-dim(temp)[1]  
evalarg<-seq(1,n, by=1)  
range <- c(1,n)  
period <- n  
 nbasis <- k  
basis <- create.fourier.basis(range, nbasis, period)  
datafd <- data2fd(temp, argvals=evalarg,basis)  
g.fit<- fitlmc(coord,datafd,k)  
fktm.salida<-fktm(g.fit,basis,evalarg,k,datafd,d,d.cero)  
plot(datafd, lty=1,col=8, ylab="Temperature (degrees C)", xlab="Day",  
 ylim=c(-15,20))  
lines(fktm.salida$pred,type="l",col=1,lty=1, lwd=2)  
slave<-matrix(scan("promediosMoncton.txt",0, dec=",")  
 lines(slave, type="p", pch=20,cex=0.6)  
 legend(0, 20, legend = c("Smoothed", "Prediction", "Real data"),  
 lty = c(1,1,-1), pch=c(-1,-1,20),lwd = c(1,1,1),  
 col=c(8,1,1),pt.cex =0.5)  
x<-seq(1,n,1)  
y<-x  
persp(x,y,fktm.salida$funpar[,,7], ticktype="detailed",col=8,  
 border=NA,main="Nappan",ltheta=-120,shade=0.5,lty=4,
APPENDIX A. R FUNCTIONS

zlim=c(-0.20,0.20), theta=30, phi=30, expand=0.5, xlab="v", ylab="t", zlab="Paramater")
persp(x,y,fktm.salida$funpar[,13], zlim=c(-0.2,0.2),main="Bouctouche",
ticktype="detailed",col=8, border=NA, ltheta=-120, shade=0.5, lty=4, theta=30, phi=30, expand=0.5, xlab="v", ylab="t", zlab="Paramater")

# Cross-validation analysis with K=65 by using the function # "fktm.cv"

array.k<-65
valcruz<-fktm.cv(temp, evalarg, period, d, array.k)
write.table(t(valcruz$krig.CV[1,,]), file="valcruz65.txt", quote=FALSE, row.names=FALSE)
predcv65<-read.table("valcruz65.txt", header=TRUE)
matplot(predcv65,type="l", lty=1, col=1, xlab="Day", ylab="Temperature (degrees C)", ylim=c(-35,25))
abline(h=0, lty=2)
residuals65<-temp-predcv65
matplot(residuals65,type="l", col=8, lty=1, xlab="Day", ylab="Residuals (degrees C)")
lines(apply(residuals65,1,mean), lty=1, lwd=2)
lines(apply(residuals65,1,sd), lty=2, lwd=2)
SSE.stations<-apply(residuals65*residuals65,2,sum)
as.matrix(SSE.stations)
summary(SSE.stations)
sd(SSE.stations)
sum(SSE.stations)
smoothed65<-eval.fd(evalarg, datafd, Lfdobj=0)
resmoothed65<-smoothed65-predcv65
matplot(resmoothed65,type="l", col=8, lty=1, xlab="Day",
ylab="Residuals (degrees C)"
lines(apply(resmoothed65,1,mean),lty=1,lwd=2)
lines(apply(resmoothed65,1,sd),lty=2,lwd=2)