Statistical models for spatially indexed functional data

by

Daniel Clayton Fortin

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Program of Study Committee:

Petrutza Caragea, Co-major Professor

Zhengyuan Zhu, Co-major Professor

Heike Hofmann

Dan Nordman

Stephen Vardeman

Iowa State University

Ames, Iowa

2013

TABLE OF CONTENTS

LIST OF TABLES iv

LIST OF FIGURES v

CHAPTER 1. INTRODUCTION 1

CHAPTER 2. NONPARAMETRIC COVARIANCE FUNCTION AND PRINCIPAL COMPONENT FUNCTION ESTIMATION FOR FUNCTIONAL DATA 4

2.1 Abstract 4

2.2 Introduction 4

2.3 Methodology 7

2.4 Covariance function estimation 9

2.5 Estimation of functional principal components 12

2.6 Simulations 14

2.7 Discussion 15

CHAPTER 3. ORDINARY KRIGING FOR SPATIALLY INDEXED FUNCTIONAL DATA 18

3.1 Introduction 18

3.2 Statistical model for spatially indexed functional data 19

3.3 Functional Kriging predictor 20

3.4 Numerical Experiments 20

3.5 Future Work 22

3.5.1 Simulation framework useful for studying spatial dependence between curves . . . 22
3.5.2 Adjustments to the covariance function estimation which account for spatial dependence . . . 24

CHAPTER 4. FUNCTIONAL DATA ANALYSIS OF SATELLITE MEASUREMENTS OF PHENOLOGICAL PROCESSES . . . 26

4.1 Introduction . . . 26
4.2 Initial Exploratory Work . . . 26
4.3 Future Work . . . 34

CHAPTER 5. SUMMARY OF COMPLETED WORK AND FUTURE WORK . . . 35

5.1 Completed Work . . . 35
5.2 Future Work . . . 36

APPENDIX A. Theoretical background on smoothing splines and reproducing kernel Hilbert space . . . 37

A.1 Historical note on the use of reproducing kernel Hilbert spaces in statistics 37
A.2 Going beyond ordinary least squares: Penalized least squares and the cubic smoothing spline . . . 38
A.2.1 Least Squares Estimation . . . 38
A.2.2 Penalized Least Squares . . . 39
A.3 Reproducing kernel Hilbert spaces (RKHS) . . . 41
A.3.1 Notation and basic definitions . . . 41
A.3.2 Motivation for the use of RKHS in penalized regression . . . 42

BIBLIOGRAPHY . . . 45

LIST OF TABLES

2.1 Tensor product space decomposition with corresponding reproducing kernels. 10

2.2 Average integrated squared error of the principal component functions $\left\|PC - \widehat{PC}\right\|_{L_2}^2$. The standard errors are reported in parentheses next to the estimates. 15

LIST OF FIGURES

2.1 (a) 50 simulated curves from the process $X(t)$ in (2.18) (b) data set of curves evaluated at five random locations with noise, $\sigma_0 = 0.392$ (c) Scatterplot of values based on the data used to estimate covariance surface (d) Estimated covariance function. 15

2.2 Estimated covariance functions based on the 50 curves simulated from the process $X(t)$ in (2.18) using $\alpha = 2$ and observation standard deviation equal to $\sigma_0 = 0.329$. (a) five observations per curve (b) ten observations per curve (c) forty observations per curve (d) true covariance functions 16

3.1 Locations of simulated curves. Solid dots are prediction locations. 21

3.2 Predicted curves at unobserved locations. 21

3.3 One hundred curves simulated independently from the process $X(t)$ in (3.6). 23

3.4 1st eigenfunction estimated from 100 simulated curves, where each curves is observed at m random locations with noise ($\sigma = 0.0368$). The number of observations per curve is indicated in the plot label. 23

3.5 2nd eigenfunction estimated from 100 simulated curves, where each curves is observed at m random locations with noise ($\sigma = 0.0368$). The number of observations per curve is indicated in the plot label. 24

4.1 Diagram illustrating typical phenology patterns for agricultural land and natural vegetation. 27
4.2 Map of southern India with pixels colored according to dominant vegetation type. The white dots indicate a 100-pixel subregion used for exploratory analysis. 28
4.3 One year of MTCI data for the subset of locations on a 10×10 grid. 29
4.4 Nonparametric estimate of the covariance function using one year of MTCI data for the subset of locations on a 10×10 grid. . . . 29
4.5 Basis functions consisting of a constant function and the first five eigenfunctions of the estimated covariance function. The curves were not centered, so the first eigenfunction shown in (b) represents the mean. 30
4.6 Figure (b) is the first eigenfunction of the uncentered curve and represents the mean. Figures (c) - (f) show the mean curve and the curves that result when you add (+) and subtract (-) a multiple of the respective eigenfunctions. Figure (c) indicates that the second eigenfunction represents the type of concavity curves exhibit at the beginning of the year. 31
4.7 Observed values for the 10×10 grid of locations and fitted curves. 32
4.8 Boxplots of the coefficients for the fitted curves using the eigenfunction basis. 32
4.9 Histograms of coefficients corresponding to the third basis function for tropical evergreen and coastal vegetation. 33
4.10 Observed values for the 10×10 grid of locations and fitted curves. 33

CHAPTER 1. INTRODUCTION

In many instances, it is useful to view data as a collection of curves. The study of human growth is a common example used to motivate this view, where data arising from the study can be viewed as collection of individual growth curves. The second derivative of the growth curves are used to study growth acceleration and identify common growth spurts. In studies like this, where the questions motivating the analysis relate to properties of curves, it makes sense to view the fundamental datum as a curve and refer to the collection of curves as *functional data*.

The publication of the book *Functional Data Analysis*, the seminal work by Ramsay and Silverman (2005), inspired substantial interest in developing statistical models for functional data. Though their book describes the extension of linear models to the functional setting, it stops short of allowing for complex dependence structures by assuming independence between observed curves. For curves observed in space, this assumption is often violated. Examples of functional data which have a spatial index are becoming more common in scientific studies: data collected by weather stations, oceanology measurements taken by seals equipped with sensor devices, peak electron density measurements collected by ionosonde stations in the northern hemisphere, and satellite measurements of environmental variables. Therefore, research on functional models for spatial data is a very important and actual topic.

When data are random curves rather than scalars or vectors, it is necessary to have a convenient way to represent the curves. A standard approach is to express the curves in terms of a known basis set (e.g. b-spline, wavelets, Fourier). An alternative, and

what we believe to be a more attractive choice for spatial functional data, is an empirical basis consisting of functional principal components (FPC). When curves are represented by a linear combination of basis functions, the variation among curves can be investigated through the variation among the coefficients of the basis functions. Thus, spatial similarity between curves can be modeled through the correlation structure associated with the basis function coefficients viewed as scalar random fields. Using a finite basis representation of the curves, each curve is identified by its vector of coefficients, allowing the use of multivariate methods; moreover, it is well known that principal components achieve maximum efficiency in terms of representing variation in the data for a fixed dimension.

This approach introduces the non-trivial problem of estimating the principal component functions from observed data. Estimation of principal component functions can be accomplished by estimating the covariance function through bivariate smoothing and then computing the eigenfunctions of the estimated covariance surface (some alternative methods are described in Ramsay and Silverman (2005) Ch. 8 and Ch. 9). Because of this, much effort has been devoted to methods for nonparametric covariance estimation and eigenfunction estimation (Yao et al. (2005), Li et al. (2007), Cai and Yuan (2010)). Of the literature on nonparametric covariance function estimation, Cai and Yuan (2010) provides the most general framework by assuming curves belong to a reproducing kernel Hilbert space (RKHS) and utilizing theoretical results developed for spline smoothing methods. This framework allows for closed form estimation of the eigenfunctions and their derivatives, so one does not need to discretize the covariance function and approximate eigenfunctions with eigenvectors. This is particularly useful because we use eigenfunctions as a basis set for further model building and not just as an exploratory tool.

In this dissertation we work in a reproducing kernel Hilbert space framework, representing curves with a principal component function basis, and model spatial dependence

between curves through FPC scores.

- In Chapter 2 we describe an estimator for principal component functions by extending the methods developed in Cai and Yuan (2010). We have developed an R implementation of this method. In our implementation one can easily produce list object containing the estimated principal component functions. We have also developed an empirical basis specification within the fda package framework, so that the estimated principal component functions can be used as a basis object.

- In Chapter 3 we describe how this framework can be used for ordinary kriging prediction for functional data. We also describe our current work on improving the nonparametric covariance estimator by incorporating spatial dependence in the estimation of the covariance function.

- In Chapter 4 we discuss future directions in model development motivated by spatial functional data arising from a phenological study in India. Some of the scientific questions related to these data involve identifying biological life-cycles, which can be cast in terms of derivatives of the curves (e.g., global mixima, inflection points). This fact further motivates the functional data approach, as it is a natural way in which one can estimate and represent derivatives.

- In Appendix A I have included some of the theoretical background necessary to understand the developments in Chapter 2. There is no original work in this chapter and it only serves to make this document a little more self-contained.

Plans for future work to complete the dissertation are described in the last section of the respective chapters.

CHAPTER 2. NONPARAMETRIC COVARIANCE FUNCTION AND PRINCIPAL COMPONENT FUNCTION ESTIMATION FOR FUNCTIONAL DATA

2.1 Abstract

Functional principal components are often used in exploratory analysis and as a modeling tool. As in multivariate methods, functional principal components are derived from the covariance function. By using a reproducing kernel Hilbert space framework, it has been shown that efficient covariance function estimation is achieved through a regularization approach on the tensor product space. We adopt this framework and show how this method can be extended to allow for unpenalized subspaces which arise naturally when the smoothing penalty is based on a high order derivatives. Closed form estimators of the principal component functions are also derived without discretizing the covariance function, making this method ideally suited for use as an empirical basis set for functional data analysis.

2.2 Introduction

Functional principal components are often used in exploratory analysis and as a modeling tool. In multivariate methods, principal components are derived through an eigen-decomposition of the covariance; however, with functional data the covariance is not a matrix, but a continuous bivariate function. Cai and Yuan (2010) propose a

framework for nonparametric covariance function estimation for a second order stochastic process, $X(\cdot)$, with support on a compact domain $\mathcal{T} \subset \mathbb{R}$ and $X(\cdot) \in \mathcal{H}$, where $\mathcal{H}$ is a reproducing kernel Hilbert space. By examining the form of the covariance function within this framework, it can be shown that the covariance function resides in the tensor product reproducing kernel Hilbert space $\mathcal{H} \otimes \mathcal{H}$. Based on this fact, it is natural to consider a regularization procedure for covariance function estimation that utilizes the tensor product norm. Cai and Yuan (2010) proposed the following

$$\hat{C}_\lambda = \underset{C \in \mathcal{H}\otimes\mathcal{H}}{\operatorname{argmin}} \{l_n(C) + \lambda \|C\|^2_{\mathcal{H}\otimes\mathcal{H}}\}, \tag{2.1}$$

where

$$l_n(C) = \sum_{i=1}^{n} \sum_{1 \le j_1 \ne j_2 \le m} ([Y_{ij_1} - \mu_0(t_{ij})][Y_{ij_2} - \mu_0(t_{ij_2})] - C(t_{ij_1}, t_{ij_2}))^2 \tag{2.2}$$

and $\lambda \geq 0$ is a tuning parameter that balances the fidelity to the data measured by l_n and smoothness of the estimate measured by the squared RKHS norm.

The convergence rate of this estimator has been shown to be superior to the optimal rate for general bivariate smoothing on $[0,1] \times [0,1]$. This implies that the tensor product RKHS is to a certain degree smaller than the typical Sobelev space used for bivariate smoothing, thus reducing the effect of "the curse of dimensionality".

By establishing a representer theorem (see Wahba (1990)), the covariance function estimator has a finite dimensional representation. Further, using this representation, closed form expressions for the eigenfunctions can be derived. These developments provide a strong argument for the use of reproducing kernel Hilbert space framework when interest is in nonparametric estimation of the covariance function and principal component functions.

This approach assumes that the reproducing kernel is known and that the tensor product norm corresponding to the reproducing kernel penalizes smoothness in an appropriate way. It seems important, from a practical point of view, to have a clear understanding of how the penalty functional is operating. To accomplish this we pro-

pose an approach that begins with defining how univariate functions are penalized on the marginal domain (typically done by defining an appropriate high order derivative), then use the implied penalty on the tensor product domain to estimate the covariance function. This approach is more attractive from a practitioner's point of view, because it is more natural to define a differential operator to penalize smoothness, and in this case the functions annihilated by the differential operator form a non-null subspace. Our approach allows for a decomposition of the function space into penalized and unpenalized subspaces. To illustrate this, recall the general development for the smoothing spline in the univariate case (Wahba (1990)): The solution of

$$\widehat{f}_\lambda = \underset{f\in\mathcal{H}}{\operatorname{argmin}} \left\{ \frac{1}{n}\sum_{i=1}^{n}(y_i - f(t_i))^2 + \lambda \|f\|^2_{\mathcal{H}} \right\} \tag{2.3}$$

has the form

$$\hat{f}(t) = \sum_{i=1}^{N} c_i R(t, t_i), \tag{2.4}$$

while the solution of

$$\widehat{f}_\lambda = \underset{f\in\mathcal{H}}{\operatorname{argmin}} \left\{ \frac{1}{n}\sum_{i=1}^{n}(y_i - f(t_i))^2 + \lambda \|P_1(f)\|^2_{\mathcal{H}_1} \right\} \tag{2.5}$$

has the form

$$\hat{f}(t) = \sum_{j=1}^{M} d_j \phi_j(t) + \sum_{i=1}^{N} c_i R_1(t, t_i), \tag{2.6}$$

where $\{\phi_j(\cdot)\}_{j=1}^{M}$ are a basis for the space of unpenalized functions $\mathcal{H}_0$ and $P_1(f) = f_1$ is the projection of f onto the space $\mathcal{H}_1$. For details of the derivation of (2.6) from (2.5) see section **??**. In Section 2.4 we generalize regularization procedure for covariance function in Cai and Yuan (2010) to account for this type of decomposition. In Section 2.5 we derive closed form estimators of the principal component functions based on the generalized covariance estimator.

2.3 Methodology

In the following we assume $\mathcal{T}$ to be the interval $[0, 1]$. Let $X(\cdot)$ be a second order stochastic process with covariance function

$$C_0(s,t) = E([X(s) - E(X(s))][X(t) - E(X(t))]),\ \forall s,t \in \mathcal{T}.$$

Further, assume $X(\cdot)$ takes values in a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$ with corresponding reproducing kernel $R(s,t)$. The reproducing kernel $R(s,t)$ has the property $\langle f(s), R_t(s)\rangle_{\mathcal{H}} = f(t)$ for all $f \in \mathcal{H}$, where $R_t(s)$ is notation for $R(s,t)$ holding the second coordinate fixed. The function $R_t(s)$ belongs to $\mathcal{H}$; as an immediate consequence of the reproducing property $\langle R_{t_1}(s), R_{t_2}(s)\rangle = R(t_1,t_2)$. This is a key property, since it shows that inner products involving the reproducing kernel function are equivalent to function evaluation.

Let $\{X_1, X_2, \ldots, X_N\}$ be a collection of independent realizations of X, and we consider the following observation model

$$Y_{ij} = X_i(t_{ij}) + \epsilon_{ij},\ j = 1, \ldots, m;\ i = 1, \ldots, N,$$

where the sampling locations are independently drawn from a common distribution on $\mathcal{T}$, and ϵ_{ij} are independently and identically distributed measurement errors with mean zero and finite variance σ_0^2. It is further assumed that the random functions X, sampling locations t_{ij}, and measurement errors ϵ are mutually independent.

The development in this section relies on the fact that a closed subspace of a Hilbert space induces a natural partition of the space into a direct sum of the closed subspace and its orthogonal complement (for a concise overview of the relevant Hilbert space theory, see Gu (2002)). The subspace consisting unpenalized functions is a closed subspace and it is convenient to express $\mathcal{H}$ as a direct sum decomposition $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, where on $\mathcal{H}_1$ the penalty functional is a full squared norm and $\mathcal{H}_0$ consists of functions in $\mathcal{H}$ which will not be penalized. With this orthogonal decomposition, any function $f \in \mathcal{H}$ has

the representation $f = f_0 + f_1$, where $f_0 \in \mathcal{H}_0$ and $f_1 \in \mathcal{H}_1$. A well known result (see Theorem A.3.1), is that the reproducing kernel R on $\mathcal{H}$ can be expressed as $R = R_0 + R_1$, where R_0 is the reproducing kernel on $\mathcal{H}_0$ and R_1 is the reproducing kernel on $\mathcal{H}_1$.

Covariance function estimation takes place on the product domain $[0,1] \times [0,1] = \mathcal{T}_1 \times \mathcal{T}_2$. Denote by $\mathcal{H}_{<1>}$ and $\mathcal{H}_{<2>}$ the reproducing kernel Hilbert space on $\mathcal{T}_1$ and $\mathcal{T}_2$, respectively. Let $\mathcal{H}_{<1>} = \mathcal{H}_{0<1>} \oplus \mathcal{H}_{1<1>}$ and $\mathcal{H}_{<2>} = \mathcal{H}_{0<2>} \oplus \mathcal{H}_{1<2>}$ be the direct sum decomposition of spaces on the marginal domains into their unpenalized and penalized subspaces. Using these decompositions, the tensor product space $\mathcal{H}_{<1>} \otimes \mathcal{H}_{<2>}$ has the representation

$$\mathcal{H}_{<1>} \otimes \mathcal{H}_{<2>} = (\mathcal{H}_{0<1>} \oplus \mathcal{H}_{1<1>}) \otimes (\mathcal{H}_{0<2>} \oplus \mathcal{H}_{1<2>}) \tag{2.7}$$

$$= (\mathcal{H}_{0<1>} \otimes \mathcal{H}_{0<2>}) \oplus (\mathcal{H}_{0<1>} \otimes \mathcal{H}_{1<2>}) \oplus (\mathcal{H}_{1<1>} \otimes \mathcal{H}_{0<2>}) \oplus (\mathcal{H}_{1<1>} \otimes \mathcal{H}_{1<2>}) \tag{2.8}$$

Equation (2.8) is a direct sum of tensor product reproducing kernel Hilbert spaces, where the subspace corresponding to $\mathcal{H}_{0<1>} \otimes \mathcal{H}_{0<2>}$ consists only of unpenalized functions. As in the univariate case, the solution to the penalized smoothing problem when unpenalized subspaces are allowed will involve linear combination of the basis functions for the unpenalized subspace.

The main goal in what follows is to describe the form of the covariance function estimator when the tensor product space has a decomposition as in (2.8). This setting is more attractive from a practitioner's point of view, because it is more natural to define a differential operator to penalize smoothness, and in this case the functions annihilated by the differential operator form a non-null subspace. We describe the form of the estimator in this case as well as derive the principle component functions. Since our focus is on practical implementation of this method, we focus on the most common case where the smoothing penalty on the marginal domain is $\int (f'')^2$, i.e. the total curvature.

2.4 Covariance function estimation

Consider the space $\mathcal{H} = \{f : f, f' \text{ absolutely continuous}, f'' \in L_2[0,1]\}$. On this space the term $\int_0^1 f''g''dx$ is a semi-inner-product which can be extended to a full inner product by defining an inner product on the subspace $\mathcal{H}_0 = \{f : f'' = 0\}$. Common choices for inner products in $\mathcal{H}_0$ are $\langle f, g\rangle_0 = f(0)g(0) + f'(0)g'(0)$ or $\langle f, g\rangle_0 = (\int_0^1 fdx)(\int_0^1 gdx) + (\int_0^1 f'dx)(\int_0^1 g'dx)$. We work with the latter because it corresponds to a mathematically convenient expression for the reproducing kernel.

The space $\mathcal{H} = \{f : f, f' \text{ absolutely continuous}, f'' \in L_2[0,1]\}$ with inner product

$$\begin{aligned}\langle f, g\rangle &= \langle f, g\rangle_0 + \langle f, g\rangle_1 \\ &= \left(\int_0^1 fdx\right)\left(\int_0^1 gdx\right) + \left(\int_0^1 f'dx\right)\left(\int_0^1 g'dx\right) + \int_0^1 f''g''dx\end{aligned}$$

is a RKHS with a reproducing kernel that can be conveniently expressed in terms of the functions

$$k_r(x) = -\left(\sum_{\mu=-\infty}^{-1} + \sum_{\mu=1}^{\infty}\right)\frac{\exp(2\pi i\mu x)}{(2\pi i\mu)^r}, r = 1, 2, \ldots. \tag{2.9}$$

The functions $k_r(x)$ in (2.9) are scaled Bernoulli polynomials, $k_r(x) = \frac{B(r)}{r!}$. The space $\mathcal{H}$ has an orthogonal decomposition $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ with corresponding reproducing kernel $R(x, y) = R_0(x, y) + R_1(x, y)$, where

$$R_0(x, y) = 1 + k_1(x)k_1(y) \tag{2.10}$$

$$R_1(x, y) = k_2(x)k_2(y) - k_4(x - y). \tag{2.11}$$

Note that the functions $k_r(x)$ in (2.9) have a rather simple form

$$\begin{aligned}k_1(x) &= x - 0.5 \\ k_2(x) &= \frac{1}{2}(k_1^2(x) - \frac{1}{12}) \\ k_4(x) &= \frac{1}{24}\left(k_1^4(x) - \frac{k_1^2(x)}{2} + \frac{7}{240}\right)\end{aligned}$$

for $x \in [0, 1]$.

The unpenalized space $\mathcal{H}_0$ can be decomposed further as $\mathcal{H}_0 = \mathcal{H}_{00} \oplus \mathcal{H}_{01}$ with reproducing kernels

$$R_{00}(x, y) = 1$$
$$R_{01}(x, y) = k_1(x)k_1(y).$$

This formulation provides a decomposition of the unpenalized space into functions spanned by a constant and functions spanned by a linear term. This construction results in the overall decomposition $\mathcal{H} = \mathcal{H}_{00} \oplus \mathcal{H}_{01} \oplus \mathcal{H}_1$. Using this decomposition of $\mathcal{H}$ on both marginal domains of $[0, 1] \times [0, 1]$ results in a natural decomposition of the tensor product space

$$\mathcal{H}_{<1>} \otimes \mathcal{H}_{<2>} = (\mathcal{H}_{00<1>} \oplus \mathcal{H}_{01<1>} \oplus \mathcal{H}_{1<1>}) \otimes (\mathcal{H}_{00<2>} \oplus \mathcal{H}_{01<2>} \oplus \mathcal{H}_{1<2>})$$

into a sum of nine subspaces of the tensor product space. These nine subspaces and their corresponding reproducing kernels are shown in Table 2.1. It is straight forward to derive the reproducing kernels in Table **??** using the fact that the reproducing kernel on the tensor product space is the product of the reproducing kernels on the marginal spaces, i.e. $R((x_{<1>}, x_{<2>}), (y_{<1>}, y_{<2>})) = R_{<1>}(x_{<1>}, y_{<1>}) \times R_{<2>}(x_{<2>}, y_{<2>})$.

Table 2.1: Tensor product space decomposition with corresponding reproducing kernels.

Subspace	Reproducing Kernel
$H_{00<1>} \otimes H_{00<2>}$	1
$H_{00<1>} \otimes H_{01<2>}$	$k_1(x_{<2>})k_1(y_{<2>})$
$H_{00<1>} \otimes H_{1<2>}$	$k_2(x_{<2>})k_2(y_{<2>}) - k_4(x_{<2>} - y_{<2>})$
$H_{01<1>} \otimes H_{00<2>}$	$k_1(x_{<1>})k_1(y_{<1>})$
$H_{01<1>} \otimes H_{01<2>}$	$k_1(x_{<1>})k_1(y_{<1>})k_1(x_{<2>})k_1(y_{<2>})$
$H_{01<1>} \otimes H_{1<2>}$	$k_1(x_{<1>})k_1(y_{<1>})[k_2(x_{<2>})k_2(y_{<2>}) - k_4(x_{<2>} - y_{<2>})]$
$H_{1<1>} \otimes H_{00<2>}$	$k_2(x_{<1>})k_2(y_{<1>}) - k_4(x_{<1>} - y_{<1>})$
$H_{1<1>} \otimes H_{01<2>}$	$[k_2(x_{<1>})k_2(y_{<1>}) - k_4(x_{<1>} - y_{<1>})]k_1(x_{<1>})k_1(y_{<1>})$
$H_{1<1>} \otimes H_{1<2>}$	$[k_2(x_{<1>})k_2(y_{<1>}) - k_4(x_{<1>} - y_{<1>})][k_2(x_{<2>})k_2(y_{<2>}) - k_4(x_{<2>} - y_{<2>})]$

Of the tensor product subspaces listed in Table 2.1, only five have $\mathcal{H}_1$ as one of their marginal domain spaces. If we denote $\mathcal{H}_{\nu,\mu} = \mathcal{H}_{\nu<1>} \otimes \mathcal{H}_{\mu<2>}$ and $R_{\nu,\mu} = R_{\nu<1>}R_{\mu<2>}$,

then $\breve{R} = R_{1,00} + R_{1,01} + R_{00,1} + R_{01,1} + R_{1,1}$ is the reproducing kernel on $\breve{\mathcal{H}} = \mathcal{H}_{1,00} \oplus \mathcal{H}_{1,01} \oplus \mathcal{H}_{00,1} \oplus \mathcal{H}_{01,1} \oplus \mathcal{H}_{1,1}$. The space $\breve{\mathcal{H}}$ contain all functions we wish to penalize.

Let $\mathbf{b}^{(i)} = [(y_{ij} - \mu(t_{ij}))(y_{ij'} - \mu(t_{ij'}))]_{1 \le j \ne j' \le m}$, $i = 1, \ldots, n$. Let

$$\mathbf{b} = (\mathbf{b}^{(1)T}, \mathbf{b}^{(2)T}, \ldots, \mathbf{b}^{(n)T})^T,$$

then the vectors $\mathbf{b}^{(i)}$ contain all pairwise products of observations on the ith curve. We propose the following estimator

$$\widehat{C}_\lambda = \underset{C \in \mathcal{H} \otimes \mathcal{H}}{\operatorname{argmin}} \left\{ (\mathbf{b} - \mathbf{C})^T (\mathbf{b} - \mathbf{C}) + \lambda \left\| C \right\|_{\breve{\mathcal{H}}}^2 \right\},$$

where

$$\mathbf{C} = [C(t_{i,j}, t_{i'j'})].$$

Using the representer theorem in Wahba (1990) theorem 1.3.1, it can be shown that the estimated covariance function has the form

$$\hat{C}(s,t) = \sum_{\nu,\mu=00,01} d_{\nu,\mu}\phi_{\nu,\mu}(s,t) + \sum_{i,j} c_{i,j}\breve{R}((t_i,t_j),(s,t)) \tag{2.12}$$

The four basis functions for the unpenalized space are $\phi_{\nu,\mu}$ are $\{1, k_1(s), k_1(t), k_1(s)k_1(t)\}$, so the solution can be written more explicitly as

$$\hat{C}(s,t) = d_{00,00} + d_{01,00}k_1(t) + d_{00,01}k_1(s) + d_{01,01}k_1(s)k_1(t) + \sum_{i,j} c_{i,j}\breve{R}((t_i,t_j),(s,t)). \tag{2.13}$$

Efficient estimation of (2.13) including smoothing parameter selection can be accomplished using methods described in Gu (2002).

In practice the number of observed pairs (t_i, t_j) will be large and we may want to specify the knot locations instead of using the observed locations. Clarke et al. (2009) recommend using knot locations forming a regular grid within the convex hull of the observed points. Let $t_i \in [0,1]; i = 1, \ldots, K$ be the chosen knot locations on the univariate space, making $(t_i, t_j)_{1 \le i,j \le K}$ a regular grid of knot locations on $[0,1] \times [0,1]$. Kim and Gu (2004) investigate this with a simulation study and give a general recommendation of $10n^{2/9}$, where n is the sample size on the tensor product domain.

2.5 Estimation of functional principal components

Functional principal components are related to the well-known Karhunen-Loeve representation theorem. Here is a brief review of the relevant mathematical results.

For a square-integrable stochastic process $X(t)$ defined on a closed interval $[a, b]$, with continuous covariance $C(s, t)$, there corresponds a linear operator $[T_C f](s) = \int_a^b C(s, t)f(t)dt$. Since $C(s, t)$ is symmetric and non-negative definite, it has the following representation (see Mercer's theorem)

$$C(s, t) = \sum_{i=1}^{\infty} \lambda_i \psi_i(s)\psi_i(t),$$

where $\{\psi_m(t)\}_{m=1,2,...}$ are a sequence of orthonormal eigenfunctions which form a complete basis, and $\{\lambda_m\}_{m=1,2,...}$ are nonnegative and nondecreasing eigenvalues. In this context, an eigenfunction-eigenvalue pair $\{\lambda_j, \psi_j(t)\}$ satisfy $\int_a^b C(s, t)\psi_j(t)dt = \lambda_j\psi_j(t)$. The Karhunen-Loeve theorem states that the process $X(t)$ admits the representation

$$X(t) = \sum_{m=1}^{\infty} \alpha_m \psi_m(t), \text{ where } \alpha_m = \int_a^b X(t)\psi_m(t)dt,$$

and the random variables $\{\alpha_m\}_{m=1,2,...}$ are uncorrelated and satisfy $E(\alpha_m) = 0$ and $\text{Var}(\alpha_m) = \lambda_m$, $\sum_m \lambda_m < \infty$. The eigenfunctions $\{\psi_m(t)\}_{m=1,2,...}$ corresponding to $C(s, t)$ are called the principal component functions and the coefficients $\{\alpha_m\}$ are the functional principal component scores of $X(t)$.

We seek functions $\hat{\psi}(s)$ that satisfy satisfy

$$\int \hat{C}(s, t)\hat{\psi}(t)dt = \theta\hat{\psi}(s). \tag{2.14}$$

Methods for deriving principal component functions were developed in Ramsay and Silverman (2005) for functions known to have a finite basis representation, i.e. $X(t) = \mathbf{b}'\mathbf{g}(t)$, where $\mathbf{g}(t)$ is a vector of basis functions. Using the finite dimensional covariance function representation in (2.13) we adapt these results by considering the vector of functions

$$\mathbf{g}(\cdot) = (1, k_1(\cdot), R_1(\cdot, t_1), R_1(\cdot, t_2), \ldots, R_1(\cdot, t_K))'. \tag{2.15}$$

Using $\mathbf{g}$ in (2.15) the covariance function estimator in (2.13) has the representation $\hat{C}(s,t) = \mathbf{g}(s)'A\mathbf{g}(t)$ where

$$A = \left(\begin{array}{cc:cccc} d_{00,00} & d_{01,00} & c_{1.} & c_{2.} & \dots & c_{K.} \\ d_{00,01} & d_{01,01} & \sum_j c_{1j}k_1(t_j) & \sum_j c_{2j}k_1(t_j) & \dots & \sum_j c_{Kj}k_1(t_j) \\ \hdashline c_{.1} & \sum_j c_{j1}k_1(t_j) & c_{11} & c_{12} & \dots & c_{1K} \\ c_{.2} & \sum_j c_{j2}k_1(t_j) & c_{21} & c_{22} & \dots & c_{2K} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{.K} & \sum_j c_{jK}k_1(t_j) & c_{K1} & c_{K2} & \dots & c_{KK} \end{array}\right)$$

Define the matrix Q to be

$$Q_{ij} = \int_0^1 \mathbf{g_i}(t)\mathbf{g}_j(t)dt, \tag{2.16}$$

then the following result states that the eigenfunctions can be expressed as a linear combination of the elements of $\mathbf{g}$.

Lemma 2.5.1. *The eigenfunctions of $\hat{C}(s,t)$ can be expressed as*

$$\hat{\psi}_k(\cdot) = b_k'\mathbf{g}(\cdot), \tag{2.17}$$

where b_k is the k-th column of $B = Q^{-1/2}U$ and U is the eigenvectors of $Q^{1/2}AQ^{1/2}$, and

$$\mathbf{g}(\cdot) = (1, k_1(\cdot), R_1(\cdot,t_1), R_1(\cdot,t_2), \dots, R_1(\cdot,t_K))'.$$

Proof. Let θ_k be the eigenvalues of $Q^{1/2}AQ^{1/2}$, then

$$\begin{aligned} \sum \theta_k\hat{\psi}_k(s)\hat{\psi}_k(t) &= \sum \theta_k b_k'\mathbf{g}(s)b_k'\mathbf{g}(t) \\ &= \sum \theta_k \mathbf{g}'(s)b_k b_k'\mathbf{g}(t) \\ &= \mathbf{g}'(s)\left(B\begin{bmatrix}\theta_1 & & \\ & \ddots & \\ & & \theta_k\end{bmatrix}B'\right)\mathbf{g}(t). \end{aligned}$$

To complete the proof, we show that $B \begin{bmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_k \end{bmatrix} B' = A$,

$$
\begin{aligned}
B \begin{bmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_k \end{bmatrix} B' &= Q^{-1/2} U \begin{bmatrix} \theta_1 & & \\ & \ddots & \\ & & \theta_k \end{bmatrix} U' Q^{-1/2} \\
&= Q^{-1/2} [\theta_1 \mathbf{u}_1 | \dots | \theta_k \mathbf{u}_k] U' Q^{-1/2} \\
&= Q^{-1/2} [Q^{1/2} A Q^{1/2} \mathbf{u}_1 | \dots | Q^{1/2} A Q^{1/2} \mathbf{u}_k] U' Q^{-1/2} \\
&= Q^{-1/2} Q^{1/2} A Q^{1/2} U U' Q^{-1/2} \\
&= A.
\end{aligned}
$$

Thus, $\sum \theta_k \hat{\psi}_k(s) \hat{\psi}_k(t) = g'(s) A g(t) = \hat{C}(s,t)$. □

2.6 Simulations

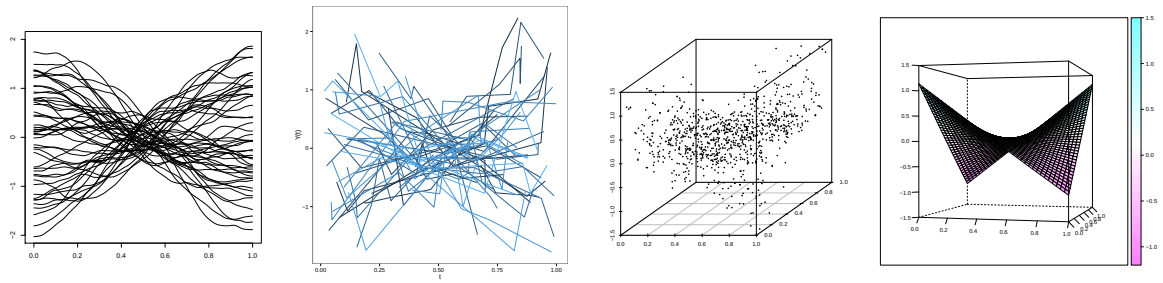
In this section, the finite sample performance of the proposed estimator is investigated by simulating random curves. Random curves are simulated independently as

$$X(t) = \sum_{k=1}^{50} \zeta_k U_k \cos(k\pi t), \quad t \in [0,1], \tag{2.18}$$

where U_k were independently sampled from a Unif$(-\sqrt{3}, \sqrt{3})$ distribution and $\zeta = (-1)^{k+1} k^{-\alpha}$. The covariance function for this process can be shown to be

$$C(s,t) = \sum_{k=1}^{50} k^{-2\alpha} \cos(k\pi s) \cos(k\pi t).$$

The parameter α controls the degree of smoothness. We simulated 50 curves using $\alpha = 2$. Figure 2.1 shows the 50 simulated curves, a data set consisting of 5 noisy observations per curve, and the fitted covariance function. Figure 2.2 show the estimated covariance functions when the number of observations per curve is 5, 10, and 40.



(a) (b) (c) (d)

Figure 2.1: (a) 50 simulated curves from the process $X(t)$ in (2.18) (b) data set of curves evaluated at five random locations with noise, $\sigma_0 = 0.392$ (c) Scatterplot of values based on the data used to estimate covariance surface (d) Estimated covariance function.

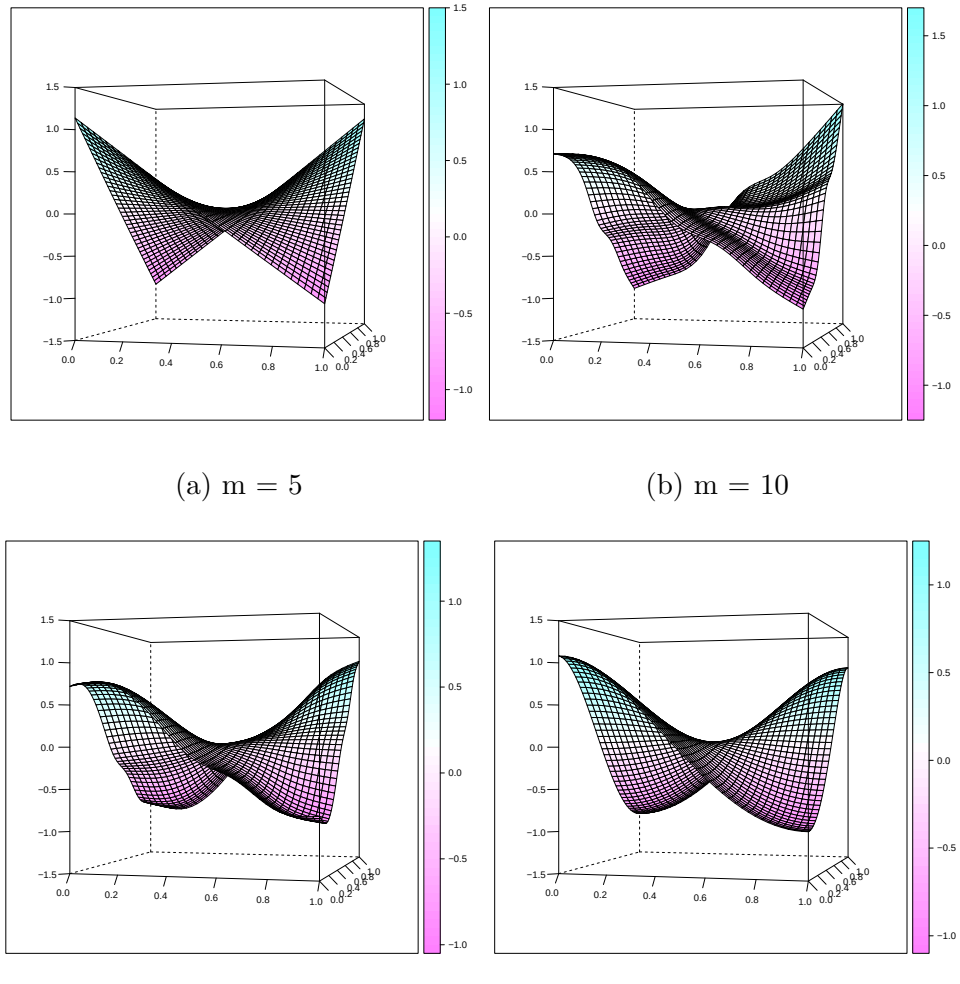
To understand the general behavior of the FPC estimator, we repeated this process 30 times using 50 curves with $\sigma_0 = 0.369$, and considered a small sample size per curve (m = 10) and moderate sample size per curve (m = 50). Talbe 2.2 summarizes the results.

Table 2.2: Average integrated squared error of the principal component functions $\left\| PC - \widehat{PC} \right\|_{L_2}^2$. The standard errors are reported in parentheses next to the estimates.

	PC 1		PC 2	
	m =10	m = 50	m = 10	m = 50
Ind	0.0452 (0.0360)	0.0119 (0.0085)	0.6302 (0.6113)	0.0893 (0.0777)
Dep	0.0373 (0.0258)	0.0128 (0.0101)	0.7438 (0.5778)	0.0893 (0.0833)

2.7 Discussion

We have shown how the reproducing kernel Hilbert space framework for covariance estimation can be extended to include unpenalized subspaces. We have also derived the form of the principal component functions. We believe that this is a more attractive framework because it allows smoothing penalties to be defined for larger function spaces. Simulation results indicate the estimator performs well even with sparsely observed curves. Even though development here was for a specific penalty, the method is



(a) m = 5 (b) m = 10

(c) m = 40 (d) truth

Figure 2.2: Estimated covariance functions based on the 50 curves simulated from the process $X(t)$ in (2.18) using $\alpha = 2$ and observation standard deviation equal to $\sigma_0 = 0.329$. (a) five observations per curve (b) ten observations per curve (c) forty observations per curve (d) true covariance functions

very general and could easily be applied to other penalties, though the form of the reproducing kernel and the basis for the null space will depend on this choice. We have also created an R package implementation of this method with user-friendly functions for estimating the covariance function and principal component functions for functional data, making it convenient to use empirical basis representation for functional data analyses. Documentation for the R package is included in appendix.

CHAPTER 3. ORDINARY KRIGING FOR SPATIALLY INDEXED FUNCTIONAL DATA

3.1 Introduction

Data that arise from environmental processes often exhibit similarities at small spatial scales. Geostatistical models incorporate this phenomenon by modeling correlations as a function of distance. Often a primary goal of the analysis is prediction of a environmental variable at an unobserved location. In this setting, geostatistical models are particularly useful because the spatial process is considered to be a continuous stochastic process. The extension of geostatistical models for functional data were first considered in Goulard and Voltz (1992) where two approaches are proposed: one approach involves cokriging by reducing the functional response to a multivariate response, while the other approach utilizes a functional version of the variogram. The functional variogram approach has been further developed by Giraldo et al. (2010) and Nerini et al. (2010). We pursue the first option, but do not make the assumption of a parametric model for the curves. We allow the curves to be represented nonparametrically using a principal component function basis. Typically, very few principal component functions are needed to represent the major modes of variation in the curves, thus making a multivariate geostatistical approach feasible.

3.2 Statistical model for spatially indexed functional data

In order to extend the methodology for spatial random fields to the functional setting we define a *spatial functional process* as

$$\left\{\boldsymbol{\chi}(s;t) : s \in D \subseteq \mathbb{R}^2, t \in \mathcal{T}\right\},$$

where, for fixed location s, $\boldsymbol{\chi}(s;\cdot)$ is a random function on a compact set $\mathcal{T} \subset \mathbb{R}$. We assume $\boldsymbol{\chi}(s;\cdot)$ takes values in a reproducing kernel Hilbert space of functions $\mathcal{H}$.

The trajectories $\chi(s;t)$ admit the following representation

$$\chi(s;t) = \mu(t) + \epsilon(s;t), \tag{3.1}$$

where $\mu(t)$ represents large-scale structure which does not depend on spatial location and $\epsilon(s;t)$ is a mean zero spatially correlated random effect. The methodology we propose assumes $\epsilon(s;t)$ can be represented by the Karhunen-Loeve expansion $\epsilon(s;t) = \sum_{k=1}^{\infty} \alpha_k(s)\psi_k(t)$. The functions $\psi_k(\cdot)$ are eigenfuncitons of the covariance function and form a basis for $\mathcal{L}_2(\mathcal{T})$. For each integer k, $\alpha_k(s) = \langle \epsilon(s;t), \psi_k(t) \rangle$ is assumed to be a second-order stationary isotropic random field. Spatial random fields connected to different eigenfunctions are assumed to be uncorrelated, that is

$$\text{Cov}(\alpha_j(s), \alpha_l(s')) = 0 \qquad \text{for } j \neq l. \tag{3.2}$$

Under assumption (3.2), the covariance between trajectories at locations s_j and s_l are given by

$$\text{Cov}(\chi(s_j,t), \chi(s_l,t')) = \sum_{k=1}^{\infty} \text{Cov}(\alpha_k(s_j), \alpha_k(s_l))\psi_k(t)\psi_k(t') \tag{3.3}$$

$$= \sum_{k=1}^{\infty} h_k(\|s_j - s_l\|)\psi_k(t)\psi_k(t'). \tag{3.4}$$

In practice we work with the truncated expansion $\epsilon(s;t) = \sum_{k=1}^{q} \alpha_k(s)\psi_k(t)$, where q is chosen to preserve most of the (interesting/low frequency) variation. Estimation of the eigenfunctions is done using a method described in Chapter 2.

3.3 Functional Kriging predictor

Steps for prediction at unobserved location s_0:

1. compute the least-squares projection of sample curves onto eigenfunctions: $\hat{\psi}^{(1)}, \hat{\psi}^{(2)}, \ldots, \hat{\psi}^{(q)}$
2. compute kriging estimates of the coefficients: $\hat{\alpha}_{s_0}^{(1)}, \hat{\alpha}_{s_0}^{(2)}, \cdots, \hat{\alpha}_{s_0}^{(q)}$
3. $\widehat{\chi_{s_0}(t)}_{ok} = \sum_{k=1}^{q} \hat{\alpha}_{s_0}^{(k)} \hat{\psi}^{(k)}(t)$

3.4 Numerical Experiments

In this section we investigate the performance of the estimator by simulating 110 random curves with spatial dependence using the process

$$X(t) = \sum_{k=1}^{3} \zeta_k Z_k \cos(k\pi t), \quad t \in [0,1], \tag{3.5}$$

where Z_k were sampled from a Gaussian random field and $\zeta = (-1)^{k+1} k^{-2}$. Ten of the simulated curves were used as a hold-out set and the remaining 100 curves we used for estimation. Figure 3.1 shows the locations of the curves. Figure 3.2 shows the predictions for the unobserved curves in the hold-out set.

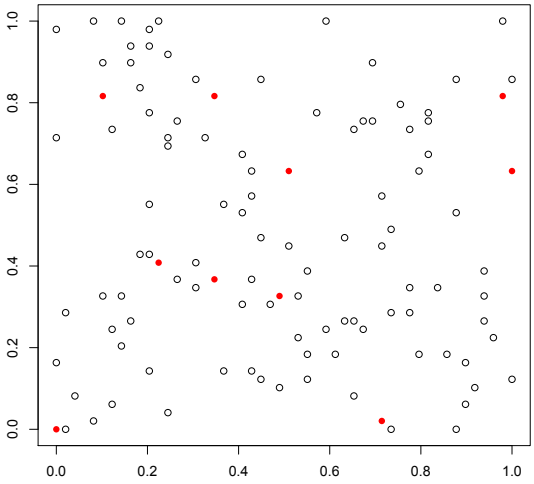


Figure 3.1: Locations of simulated curves. Solid dots are prediction locations.

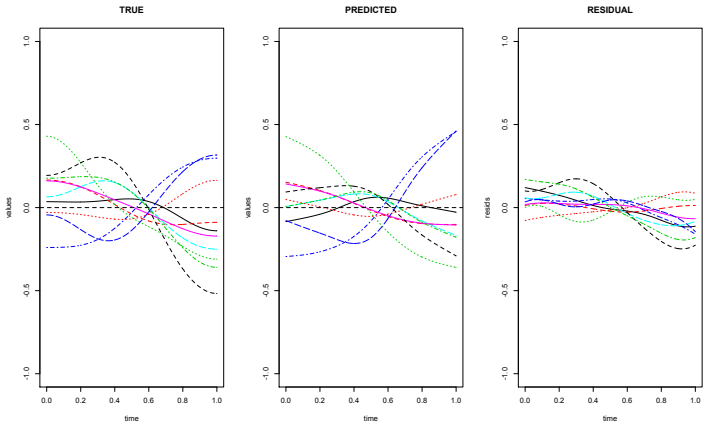


Figure 3.2: Predicted curves at unobserved locations.

3.5 Future Work

The future work we would like to accomplish for this chapter consists of gaining a better understanding of how the covariance function estimator performs when the independence assumption is violated, and to explore possible modifications to both the covariance function estimator and smoothing parameter selection which account for spatial dependence.

3.5.1 Simulation framework useful for studying spatial dependence between curves

By using the simulation framework in Chapter 2, but using a spatially correlated Gaussian random variables instead of uniform random variables we can investigate the effects of spatial dependence on estimation of functional principal components. Already included here is a simulated example under independence.

Random curves are simulated independently as

$$X(t) = \sum_{k=1}^{3} \zeta_k Z_k \cos(k\pi t), \quad t \in [0,1], \tag{3.6}$$

where Z_k were independently sampled from a $N(0,1)$ distribution and $\zeta = (-1)^{k+1}k^{-\alpha}$. The covariance function for this process can be easily show to be

$$C(s,t) = \sum_{k=1}^{3} k^{-2\alpha} \cos(k\pi s)\cos(k\pi t). \tag{3.7}$$

Figure 3.3 shows 100 curves simulated independently from the process in (2.18) with $\alpha = 2$.

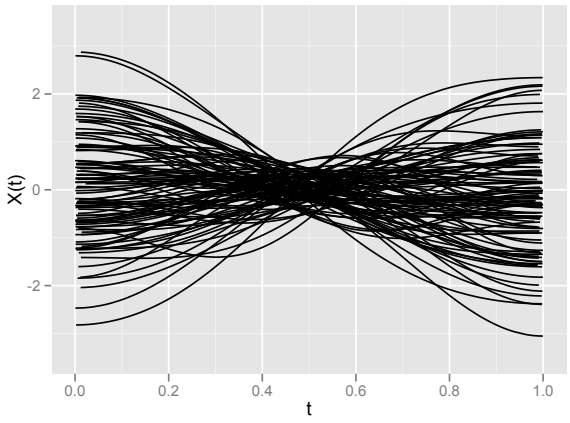


Figure 3.3: One hundred curves simulated independently from the process $X(t)$ in (3.6).

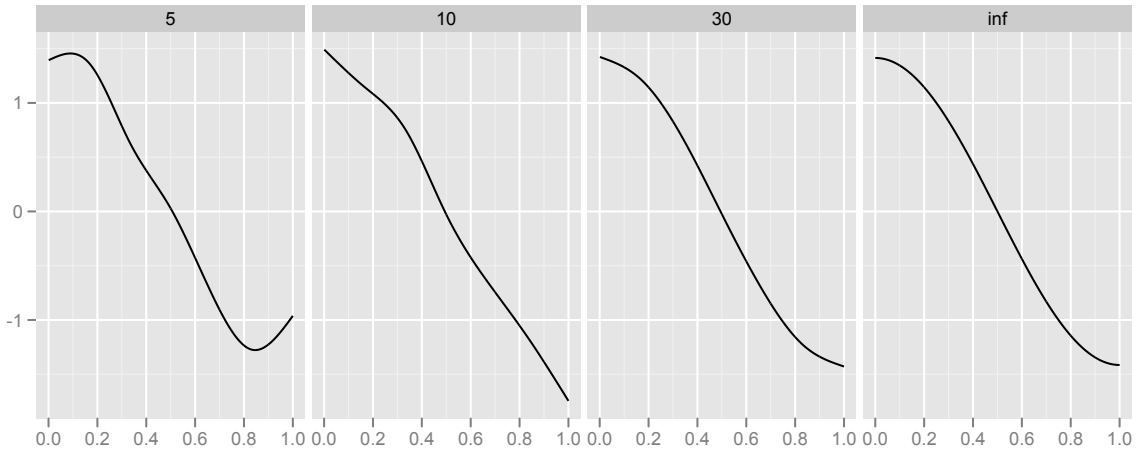


Figure 3.4: 1st eigenfunction estimated from 100 simulated curves, where each curves is observed at m random locations with noise ($\sigma = 0.0368$). The number of observations per curve is indicated in the plot label.

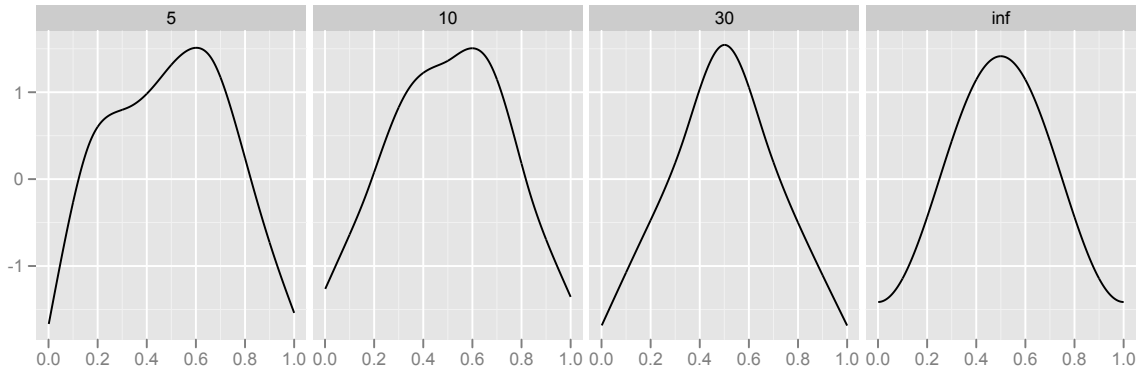


Figure 3.5: 2nd eigenfunction estimated from 100 simulated curves, where each curves is observed at m random locations with noise ($\sigma = 0.0368$). The number of observations per curve is indicated in the plot label.

3.5.2 Adjustments to the covariance function estimation which account for spatial dependence

Let $\mathbf{b}^{(i)} = [(y_{ij} - \mu(t_{ij}))(y_{ij'} - \mu(t_{ij'}))]_{1 \le j \ne j' \le m}$, $i = 1, \ldots, n$. Let

$$\mathbf{b} = (\mathbf{b}^{(1)T}, \mathbf{b}^{(2)T}, \ldots, \mathbf{b}^{(n)T})^T,$$

then the vectors $\mathbf{b}^{(i)}$ contain values related to the sample covariance from curve i, and the vector $\mathbf{b}$ contains all sample covariance terms. The elements of $\mathbf{b}$ have non-trivial covariances due to spatial correlation among curves. However, we show that the elements of Cov($\mathbf{b}$) can be compute using only covariances of the form (3.4). Let

$$\mathbf{W} = \text{Cov}(\mathbf{b}),$$

then elements of $\mathbf{W}$ can be computed as follows,

$$\begin{aligned}\text{Cov}(\epsilon(s_i;t_j)\epsilon(s_i;t_{j'}), \epsilon(s_{i'};t_l)\epsilon(s_{i'};t_{l'})) = &\text{Cov}(\epsilon(s_i;t_j), \epsilon(s_{i'};t_l))\text{Cov}(\epsilon(s_i;t_{j'}), \epsilon(s_{i'};t_{l'})) \\ &+ \text{Cov}(\epsilon(s_i;t_j), \epsilon(s_{i'};t_{l'}))\text{Cov}(\epsilon(s_i;t_{j'}), \epsilon(s_{i'};t_l))\end{aligned} \tag{3.8}$$

The right hand side of (3.8) holds under the assumption of Gaussian distributions (see Bohrnstedt and Goldberger (2010)).

We propose the following estimator

$$\widehat{C}_\lambda = \underset{C \in \mathcal{H}\otimes\mathcal{H}}{\operatorname{argmin}} \left\{ l_n(C) + \lambda \left\| C \right\|^2_{\breve{\mathcal{H}}} \right\},$$

where

$$l_n(C) = (\mathbf{b} - \mathbf{C})^T \mathbf{W}^{-1} (\mathbf{b} - \mathbf{C}) \tag{3.9}$$

and

$$\mathbf{C} = [C(t_{i,j}, t_{i'j'})]$$

- Though computation of $\mathbf{W}^{-1}$ is theoretically possible, the dimension of $\mathbf{W}$ is computationally prohibitive. To complete this chapter we intend to investigate the possibility of including weights in the loss function used in covariance function estimation (this would correspond to a diagonal $\mathbf{W}$ matrix in (3.9)). If we think of the locations of the curves as a point process, then we can use the intensity function to derive weights for the curves. A possible way to define appropriate weights would be the inverse intensity raised to some power. We intend on investigating appropriate choices through a simulation study, with the hope of being able to offer general guidelines for choosing a weight function. We also intend on including weight function in the smoothing parameter selection.

CHAPTER 4. FUNCTIONAL DATA ANALYSIS OF SATELLITE MEASUREMENTS OF PHENOLOGICAL PROCESSES

4.1 Introduction

Phenology is the study of annual life-cycles of terrestrial vegetation and how they are effected by climate change or other environmental variables (e.g. elevation). Understanding vegetation phenology and its spatio-temporal variation is required to reveal and predict ongoing changes in Earth system dynamics Jeganathan et al. (2010). The goal of this work is to develop a flexible model for vegetation life-cycles. A single life-cycle is determined by two variables: onset of greenness (OG) and end of senescence (ES). Figure 4.1 illustrates how these values are defined. We use Multi-temporal Medium Resolution Imaging Spectrometer (MERIS) Terrestrial Chlorophyll Index (MTCI) data to derive onset of greenness and end of senescence for major tropical vegetation types.

4.2 Initial Exploratory Work

We have only recently started working with this data set. As an initial step we have chosen a subset of the data corresponding to the 10 grid shown in Figure 4.2. The first year of data for these locations are shown in Figure 4.3, with colors indicating the four different land cover types in the subset. Using these data the nonparametric covariance estimator described in Chapter 2 was used to estimate the covariance function

and corresponding eigenfunctions. The first five eigenfunctions shown in Figure 4.5 were used as basis functions producing the fitted curves shown in Figure 4.10.

In order to understand the type of variation represented by an eigenfunction Ramsay and Silverman (2005) recommend adding and subtracting the eigenfunction from the mean function. Plots of this type are shown in Figure 4.6. Notice how the second eigenfunction explains variation in the concavity of the function at the beginning of the season. The difference between tropical evergreen vegetation and coastal vegetation can largely be explained by the values of the coefficients for the second eigenfunction (see Figure 4.9). The total variation in coefficient values for all basis functions are shown in Figure 4.8 illustrating how the majority of the variation is explained with only the first few eigenfunctions.

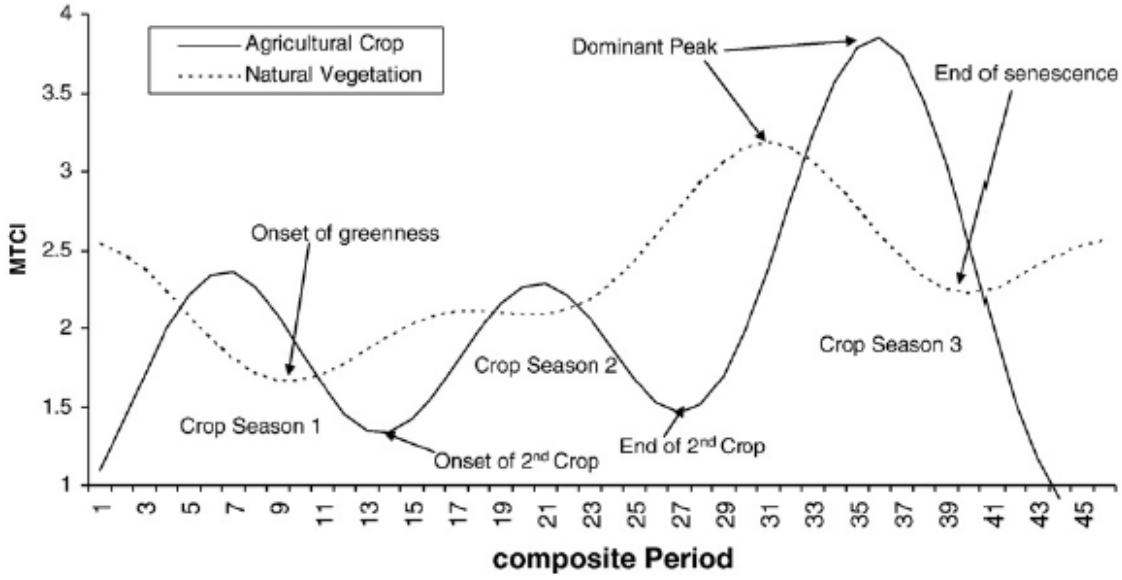


Figure 4.1: Diagram illustrating typical phenology patterns for agricultural land and natural vegetation.

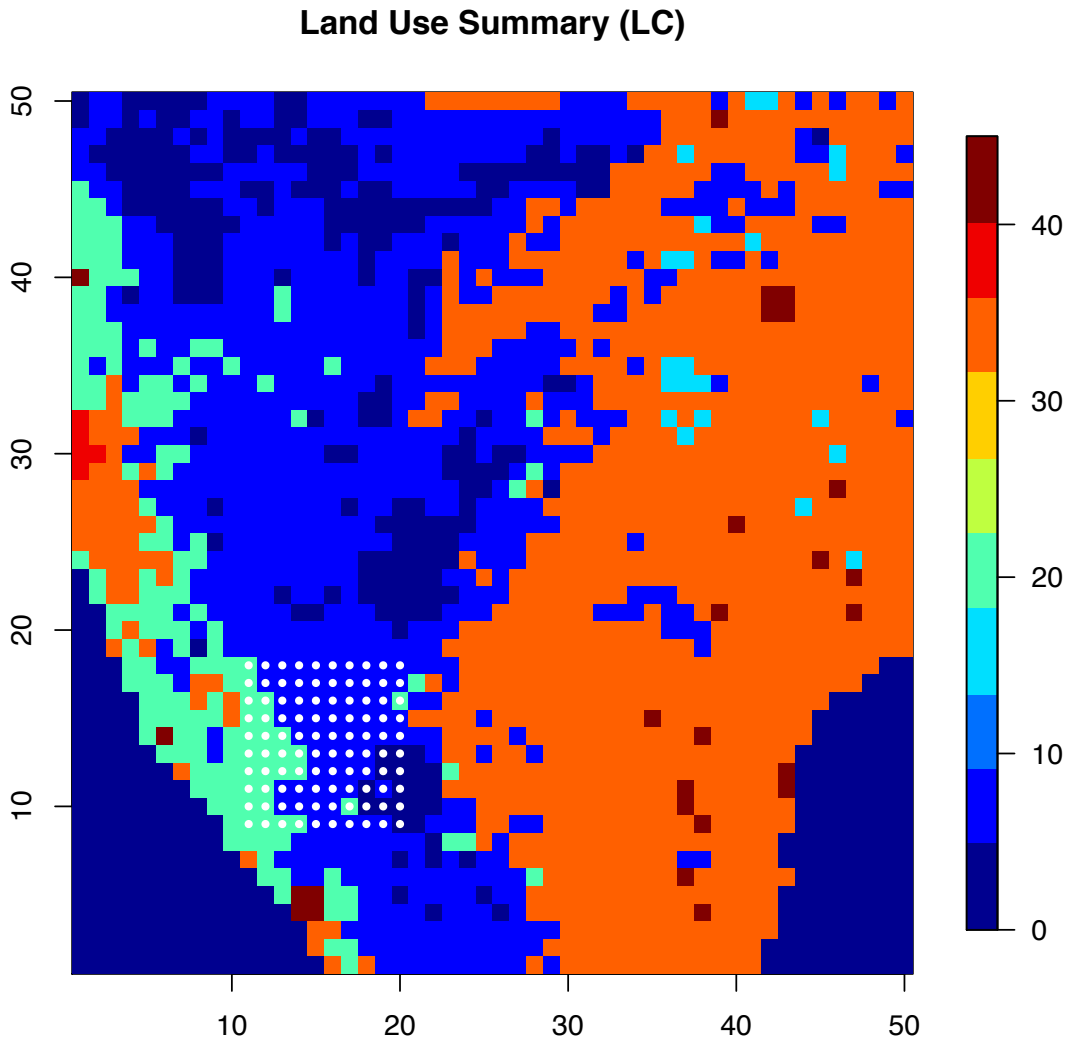


Figure 4.2: Map of southern India with pixels colored according to dominant vegetation type. The white dots indicate a 100-pixel subregion used for exploratory analysis.

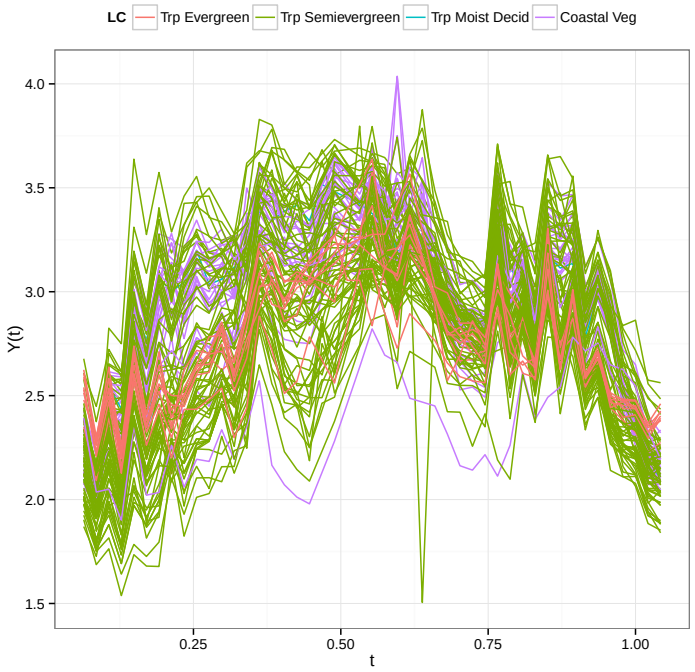


Figure 4.3: One year of MTCI data for the subset of locations on a 10×10 grid.

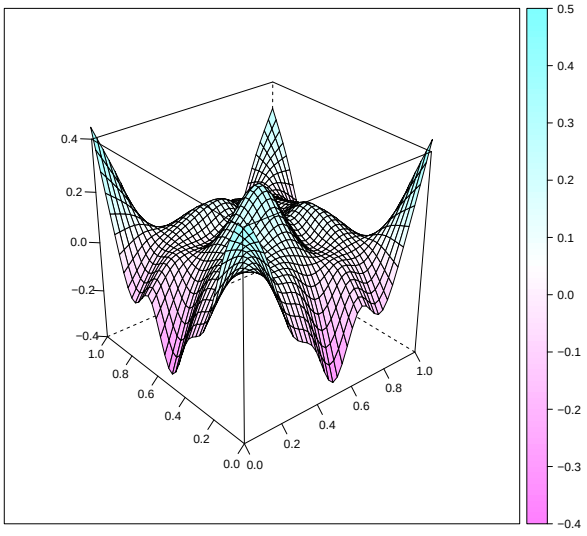
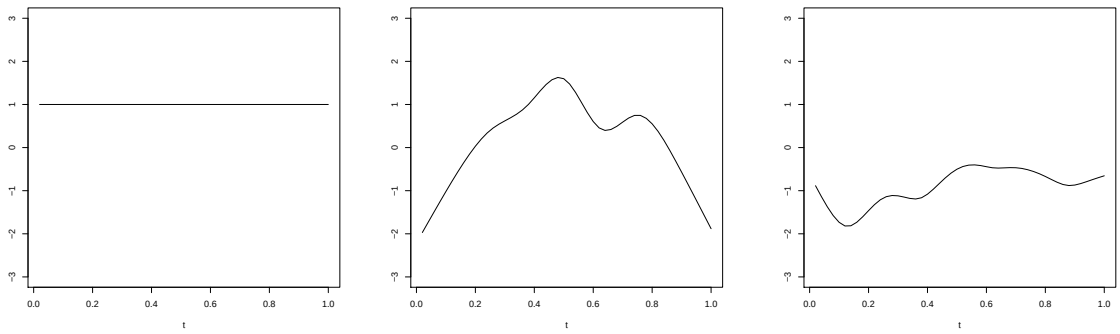
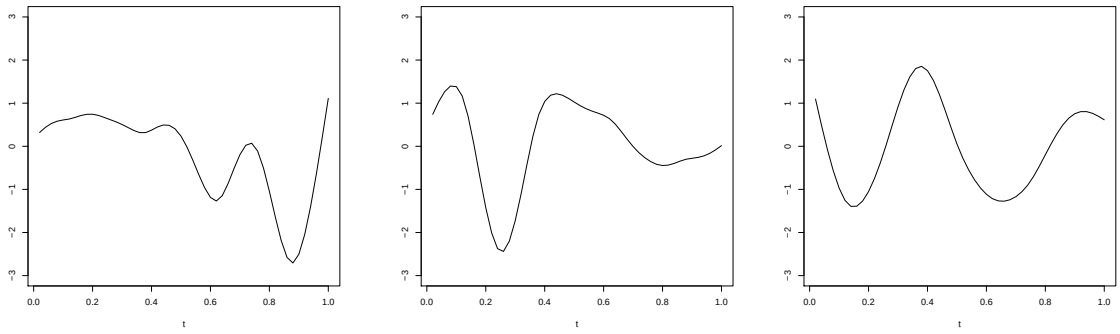


Figure 4.4: Nonparametric estimate of the covariance function using one year of MTCI data for the subset of locations on a 10×10 grid.



(a) (b) (c)



(d) (e) (f)

Figure 4.5: Basis functions consisting of a constant function and the first five eigenfunctions of the estimated covariance function. The curves were not centered, so the first eigenfunction shown in (b) represents the mean.

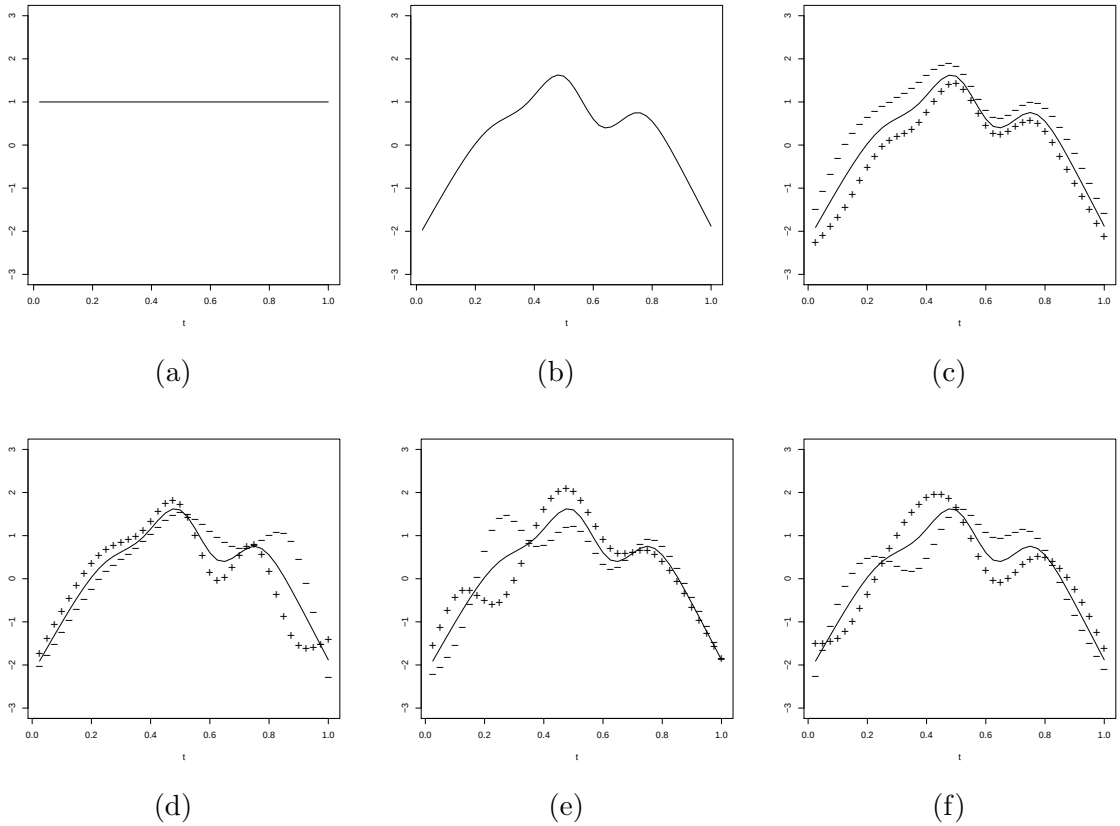


Figure 4.6: Figure (b) is the first eigenfunction of the uncentered curve and represents the mean. Figures (c) - (f) show the mean curve and the curves that result when you add (+) and subtract (-) a multiple of the respective eigenfunctions. Figure (c) indicates that the second eigenfunction represents the type of concavity curves exhibit at the beginning of the year.

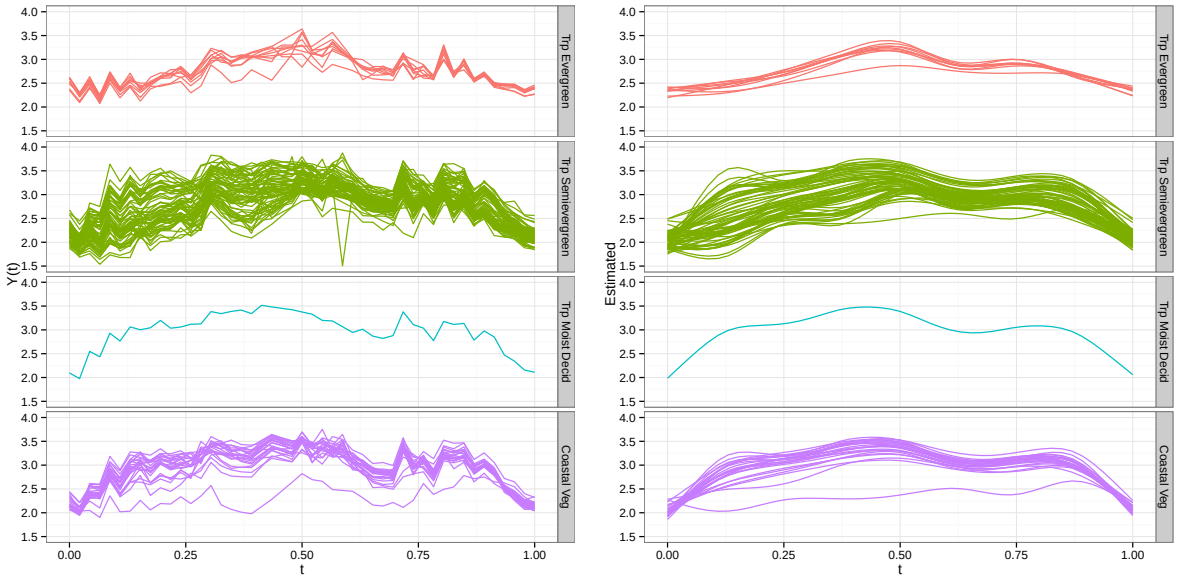


Figure 4.7: Observed values for the 10×10 grid of locations and fitted curves.

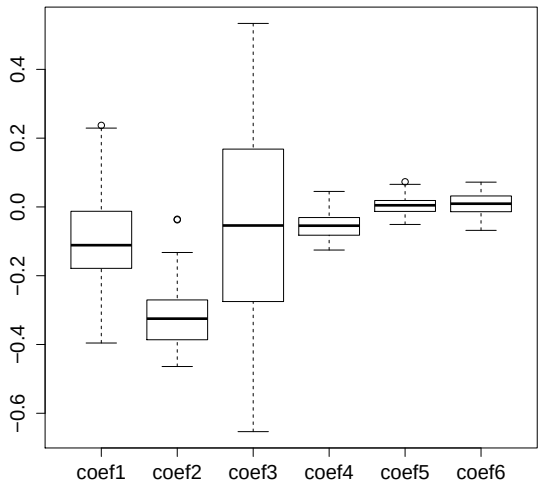


Figure 4.8: Boxplots of the coefficients for the fitted curves using the eigenfunction basis.

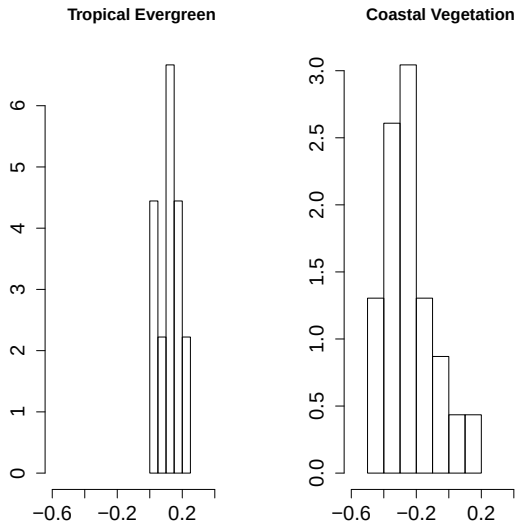


Figure 4.9: Histograms of coefficients corresponding to the third basis function for tropical evergreen and coastal vegetation.

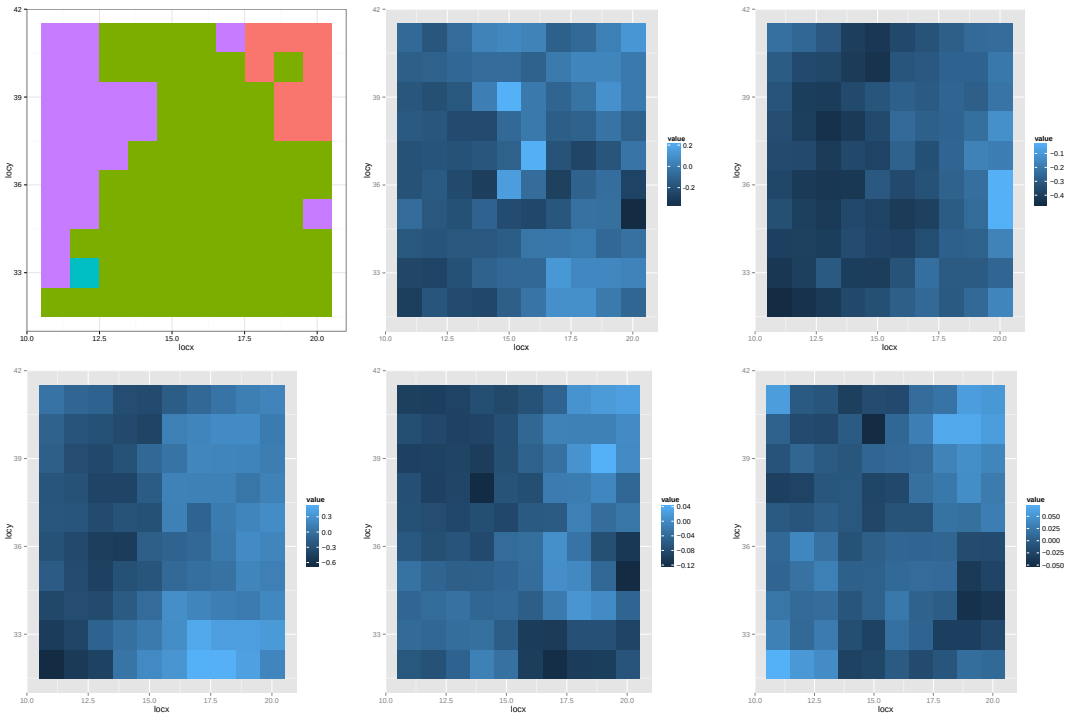


Figure 4.10: Observed values for the 10×10 grid of locations and fitted curves.

4.3 Future Work

1. Use model to estimate primary scientific variables of interest—onset of greeness and end of senescence—for each of the major vegetation types.

2. Explore spatial structure in these data and investigate how spatial dependence can be incorporated into the model framework.

3. Investigate how to incorporate other available covariates (e.g. elevation, slope, and aspect)

4. Investigate if this model can be used for classification of land-cover types.

CHAPTER 5. SUMMARY OF COMPLETED WORK AND FUTURE WORK

5.1 Completed Work

- In Chapter 2 we developed a nonparametric estimator for principal component functions using a reproducing kernel Hilbert space framework. An R package implementation of this method with user-friendly functions for estimating the covariance function and principal component functions for functional data, making it convenient to use empirical basis representation for functional data analyses.

- In Chapter 3 we introduced a kriging predictor for functional data based on an empirical basis representation of curves, where spatial dependence is modeled through basis function coefficients. We also describe our current work on improving the nonparametric covariance estimator by incorporating spatial dependence in the estimation of the covariance function.

- In Chapter 4 we discussed an application motivated by a phenology study in India. We successfully applied the nonparametric covariance function estimator to a subset of the data to produce and empirical basis for these data. We have also shown how to interpret the type of variation represented by the basis functions and how these are associated with biological patterns exhibited in specific vegetation types.

5.2 Future Work

- Chapter 3

    1. Investigate how the covariance estimator performs when curves are not independent. At the end of Chapter 3 we describe a simulation framework to investigate these effects. The focus of the simulation study will be on how the covariance estimator is affected by (1) the strength of spatial dependence, and (2) the spacing between curves (e.g. regular lattice, completely spatially random, spatially clustered).
    2. Investigate the possibility of including weights in the loss function used in covariance function estimation. If we think of the locations of the curves as a point process, then we can use the intensity function to derive weights for the curves. A possible way to define appropriate weights would be the inverse intensity raised to some power. We intend on investigating appropriate choices through a simulation study, with the hope of being able to offer general guidelines for choosing a weight function. We also intend on including weight function in the smoothing parameter selection.

- Chapter 4

    1. Use model to estimate primary scientific variables of interest—onset of greeness and end of senescence—for each of the major vegetation types.
    2. Explore spatial structure in these data and investigate how spatial dependence can be incorporated into the model framework.
    3. Investigate how to incorporate other available covariates (e.g. elevation, slope, and aspect)
    4. Investigate if this model can be used for classification of land-cover types.

APPENDIX A. Theoretical background on smoothing splines and reproducing kernel Hilbert space

A.1 Historical note on the use of reproducing kernel Hilbert spaces in statistics

My first encounter with reproducing kernel Hilbert spaces was through a paper describing a method of nonparametric covariance function estimation. At that point in time I had very little knowledge about nonparametric curve estimation, though I was certainly aware of smoothing splines and their basic properties. It was not at all obvious to me how this type of estimation could be related to something as mathematically exotic as a reproducing kernel Hilbert space. I felt like I needed some historical context to understand the motivation for the mathematical framework. In the following I would like to describe a little bit about how these special Hilbert spaces became mainstream in statistics and mention some of the key players.

The origin of reproducing kernel Hilbert space methods in statistics is due to Emanuel Parzen. In an interview in a 2002 issue of Statistical Science, Parzen describes how he was introduced to the topic. He states that when he was a graduate student at Berkely, students did not take the usual Ph.D qualifying exam; instead, each student was tasked with giving a lecture to their committee on a specific topic — Parzen's topic: reproducing kernels. When he began studying the topic, he only found one source in the literature and it was in French. The paper he found was work by Aronszajn, who had developed the abstract theory of reproducing kernels to generalize kernel methods popular in applied

mathematics.

The use of Hilbert space methods to model continuous-parameter time series appears to be the Parzen's initial application of reproducing kernels in Statistics. Though Parzen was working on time series analysis, it seems that the reproducing kernel framework is used more widely in the smoothing spline methods as a regularization approach for nonparametric curve estimation due to the fact that minimization problems over a RKHS have a unique solution with a finite dimensional representation (the so-caledl representer theorem). It was Parzen's graduate student Grace Wahba who championed regularization via RKHS in her book, *Spline Models for Observational Data*. Most of the current work using a RKHS framework for nonparametric estimation cite Wahbas work, and it is telling that a lot of the current researchers on smoothing spline methods are themselves former students of Grace Wahba (e.g. Chong Gu, Yuedong Wang, Ming Yuan, Doug Nychka).

In the introduction to her book, *Spline Models for Observational Data*, Wahba offers some encouragement, which I have included below in quotations. I took this advice to heart when I first embarked on this topic — I have to say that she has been true to her word.

> "I would like to assure the reader that the effort to master the basic properties of reproducing kernel Hilbert spaces, will be worth the effort."
> -Grace Wahba

A.2 Going beyond ordinary least squares: Penalized least squares and the cubic smoothing spline

A.2.1 Least Squares Estimation

Consider the regression problem where observations $(x_i, y_i), i = 1, \ldots, n$, are modeled as $y_i = f(x_i) + \epsilon_i, \epsilon_i \sim N(0, \sigma^2)$. This model assumes that we observe an underlying signal,

f, corrupted by noise. The desire is to recover f from the observations $(x_i, y_i), i = 1, \ldots, n$. The noise terms $\{\epsilon_i\}$ are random variables and exact recovery of f is not possible, thus our goal becomes to find a good approximation to f. The least squares estimator of f is given by the minimizer $\hat{f}$ of the sum of the squared deviations over all functions $f \in \mathcal{A}$

$$\hat{f} = \underset{f \in \mathcal{A}}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2. \tag{A.1}$$

If the underlying signal f does not lie in $\mathcal{A}$, then the estimator is guaranteed to be biased. In order to avoid misspecification of $\mathcal{A}$, a wide class of functions should be considered. As a concrete example, in simple linear regression we have $\mathcal{A} = \{f : f(x_i) = \beta_0 + \beta_1 x_i; i = 1, \ldots, n\}$. In this setting, $\mathcal{A}$ is defined as the two dimensional subspace of $\mathbb{R}^n$ spanned by a column of 1s and the column vector $(x_1, \ldots, x_n)'$. The matrix consisting of these columns constitutes the *model matrix.* Any function $f \in \mathcal{A}$ can be identified uniquely by the parameters β_0 and β_1 in its representation as a linear combination of the columns of the model matrix. Thus, the solution to the least squares minimization problem has the representation $\widehat{f} = X\widehat{\boldsymbol{\beta}}$ satisfying

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^2}{\operatorname{argmin}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) \tag{A.2}$$

where $\mathbf{Y} = (y_1, \ldots, y_n)'$, and $\mathbf{X}$ is the model matrix. Finding $\widehat{\boldsymbol{\beta}}$ requires simple matrix calculus manipulations.

A.2.2 Penalized Least Squares

The purpose of presenting the simple linear regression problem in the previous section was to emphasize how the choice of a parametric model can be framed as a choice of the structure of the function space $\mathcal{A}$. The simple linear regression model is equivalent to specifying $\mathcal{A}$ to be a known 2-dimensional subspace of $\mathbb{R}^n$ – the column space of $\mathbf{X}$. In practice, many relationships of interest are approximately linear making this model suitable in such cases. However, in most applications it is not the *truth* of the

model we believe in, but its adequacy as a mathematical representation of the underlying relationship. When a parametric model is adequate the parameters often give us insight into the structure of the relationship and scientific questions of interest can often be stated in using contrasts of the parameters. However, there is usually very little known about the functional form of f, and hence there is very little justification for restricting the estimator to belong to a particular class of parametric functions.

Therefore, it seems desirable to consider a larger class of functions. To decide on how large this class should be, we remark that any reasonably large class of functions will contain functions that interpolate the data. Since we believe the observational process is corrupted by noise, we do not wish to model the idiosyncrasies of the data; that is, we believe the estimator $\hat{f}$ must be smoother than an interpolating function. Penalized least squares is one method where this problem is addressed by including a term that penalizes the roughness of the function. The penalized least squares estimator of f is formulated to balance two competing functionals

$$\widehat{f}_\lambda = \underset{f\in\mathcal{A}}{\operatorname{argmin}} \left\{ \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \int_{\mathcal{X}} (f''(x))^2 dx \right\}, \tag{A.3}$$

where the smoothing parameter λ governs the trade-off between fitting the data and smoothness of the function. In (A.3) the smoothness of the function is measured by the integral of its squared second derivative, thus implicitly assuming the second derivative exists and that it is square integrable. A natural functional space to consider is the space $\mathcal{A} = \{f : f, f' \text{are absolutely continuous and } \int_0^1 f''(x)^2 dx < \infty\}$. It is important to note that linear functions are not penalized in (A.3) and as $\lambda \to \infty$ the estimator $\widehat{f}_\lambda$ converges to the least squares estimator. Two important questions to consider are:

1. Is there a solution to the minimization problem in (A.3)?

2. Is the solution unique?

The answer is affirmative to both questions; furthermore, the solution is a piecewise cubic polynomial. To better understand how to solve minimization problems like (A.3)

we consider the Hilbert space structure of $\mathcal{A}$, and show how the existence of a reproducing kernel is key.

A.3 Reproducing kernel Hilbert spaces (RKHS)

A.3.1 Notation and basic definitions

In this section we will cover a brief introduction to reproducing kernel Hilbert spaces. Two excellent introductions to this material are given in Heckman (1997) and Storlie et al. (2011), while more rigorous presentations are given by Wahba (1990) and Aronszajn (1950). Notice that in (A.3) the sum of squares term involves evaluation of f. A reproducing kernel Hilbert space is a Hilbert space where function evaluation is a continuous linear operator. That is a Hilbert space $\mathcal{H}$ on a domain $\mathcal{T}$ is a reproducing kernel Hilbert space if $L_t(f) = f(t)$ is continuous in $\mathcal{H}$, $\forall t \in \mathcal{T}$. The domain $\mathcal{T}$ often represents a continuous time domain and will typically be the close interval $[0, 1]$. It is well known that continuity of a linear functional is equivalent to boundedness of the functional, where boundedness means that there exists an M_t such that

$$|L_t(f)| = |f(t)| \leq M_t \|f\|_{\mathcal{H}}, \text{ for all } f \text{ in the RKHS.} \tag{A.4}$$

The Riesz representation theorem states that any continuous linear operator on $\mathcal{H}$ has a representation in terms of the inner product on $\mathcal{H}$

$$L_t(f) = \langle R_t(\cdot), f(\cdot) \rangle = f(t), \forall f \in \mathcal{H}, \tag{A.5}$$

where R_t is a function in $\mathcal{H}$. If we denote $R_t(s)$ by $R(s, t)$, then the reproducing property in (A.5) implies that $\langle R_s(\cdot), R_t(\cdot) \rangle = R(s, t)$. Note the $R(s, t)$ is a symmetric bivariate function defined on $\mathcal{T} \times \mathcal{T}$.

At this point, we note that this reproducing property may appear to be similar to the use of the Dirac delta function δ, where $f(x) = \int \delta(x - y) f(y) dy$. The key distinction is that δ is not an element of $\mathcal{H}$, in fact, δ is not even a function in the conventional sense. As

a concrete example the space $\mathcal{L}^2[0,1]$ is not a RKHS. To see this, consider the sequence of polynomials $f_n = x^n$ which converge to the null function, since $\|f_n - 0\|_2 = (2n+1)^{-1/2}$, yet $f_n(1) = 1$ for all n; i.e. $L_1(t)$ is not a continuous functional. It is easy to show that in RKHS, norm convergence implies pointwise convergence.

The following result (Theoreom 2.5 in Gu (2002)) will be useful in solving penalized least squares problems, where it is natural to consider a direct sum decomposition of the null space of the penalty term and its orthogonal complement.

Theorem A.3.1. *If the reproducing kernel R of a space $\mathcal{H}$ on domain $\mathcal{T}$ can be decomposed into $R = R_0 + R_1$, where R_0 and R_1 are both non-negative definite, $R_0(x,\cdot), R_1(x,\cdot) \in \mathcal{H}, \forall t \in \mathcal{T}$, and $\langle R_0(s,\cdot), R_1(t,\cdot)\rangle = 0, \forall s,t \in \mathcal{T}$, then the spaces $\mathcal{H}_0$ and $\mathcal{H}_1$ corresponding respectively to R_0 and R_1 form a tensor sum decomposition of $\mathcal{H}$. Conversely, if R_0 and R_1 are both non-negative definite and $\mathcal{H}_0 \cap \mathcal{H}_1 = 0$, then $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$ has a reproducing kernel $R = R_0 + R_1$.*

A.3.2 Motivation for the use of RKHS in penalized regression

In Section **??** the space $\mathcal{A} = \{f : f, f' \text{are absolutely continuous and } \int_0^1 f''(x)^2 dx < \infty\}$ was chosen as a solution space for the optimization problem in (A.3). This space with the inner product $\langle f, g\rangle = f(0)g(0) + f'(0)g'(0) + \int_0^1 f''(x)g''(x)dx$ is a Hilbert space with the reproducing property. In this section we derive the solution to (A.3) highlighting the use of the reproducing property.

Let $\mathcal{H}$ be a the Hilbert space $\{f : f, f' \text{are absolutely continuous and } \int_0^1 f''(x)^2 dx < \infty\}$ with inner product $\langle f, g\rangle = f(0)g(0) + f'(0)g'(0) + \int_0^1 f''(x)g''(x)dx$. Notice that the penalty functional in (A.3), $\int_0^1 f''(x)^2 dx$, corresponds to a squared semi-norm with null space $\mathcal{N} = \{f : f = \alpha_0 + \alpha_1 x\}$. This is a closed linear subspace of $\mathcal{H}$ which means $\mathcal{H}$ has a direct sum decomposition $\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$, where $\mathcal{H}_0 = \mathcal{N}$ and $\mathcal{H}_1 = \mathcal{N}^\perp = \{f : f(0) = f'(0) = 0 \text{ and } \int_0^1 (f''(x))^2 dx < \infty\}$. Any $f \in \mathcal{H}$ has the representation $f = f_0 + f_1$, where $f_0 \in \mathcal{H}_0$ and $f_1 \in \mathcal{H}_1$. On $\mathcal{H}_1$ the penalty term $\int_0^1 f''(x)^2 dx$ corresponds to a full

squared norm, thus the penalty term can be written as $\|P_1(f)\|^2_{\mathcal{H}_1}$ where $P_1(f) = f_1$ is the projection of f onto $\mathcal{H}_1$. This allows (A.3) to be put in the more general form

$$\widehat{f}_\lambda = \arg\min_{f\in\mathcal{H}} \left\{ \frac{1}{n}\sum_{i=1}^{n}(y_i - f(x_i))^2 + \lambda \|P_1(f)\|^2_{\mathcal{H}_1} \right\}. \tag{A.6}$$

Recall that in the discussion of the least squares estimation problem in Section **??** the minimization problem was reduced to a matrix calculus problem by restricting the solution space to a finite dimensional space spanned by the columns of the model matrix. In the current problem the minimizer lies in a finite dimensional subspace—this is a direct consequence of the reproducing property. To see this, let us further decompose $\mathcal{H}_1$ into two orthogonal subspaces $\mathcal{H}_1 = \text{span}\{R_1(\cdot, x_i); i = 1, \ldots, n\}$ and $\mathcal{H}_1 \ominus \text{span}\{R_1(\cdot, x_i); i = 1, \ldots, n\}$. Any function $f \in \mathcal{H}$ has the representation

$$f(x) = \alpha_0 + \alpha_1 x + \sum_{i=1}^{n} \beta_i R_1(x, x_i) + \eta(x) = f_0 + f_1 + \eta. \tag{A.7}$$

Substituting this representation into (A.6) we have

$$\widehat{f}_\lambda = \arg\min_{f\in\mathcal{H}} \left\{ \frac{1}{n}\sum_{i=1}^{n}(y_i - \alpha_0 - \alpha_1 x_i - \sum_{i=1}^{n}\beta_i R_1(\cdot, x_i) - \eta(x_i))^2 + \lambda \left\| \sum_{i=1}^{n}\beta_i R_1(\cdot, x_i) + \eta \right\|^2_{\mathcal{H}_1} \right\} \tag{A.8}$$

$$= \arg\min_{f\in\mathcal{H}} \left\{ \frac{1}{n}\sum_{i=1}^{n}(y_i - \alpha_0 - \alpha_1 x_i - \sum_{i=1}^{n}\beta_i R_1(\cdot, x_i) - \eta(x_i))^2 + \lambda \left\| \sum_{i=1}^{n}\beta_i R_1(\cdot, x_i) \right\|^2_{\mathcal{H}_1} + \lambda \|\eta\|^2_{\mathcal{H}_1} \right\}. \tag{A.9}$$

The reproducing property allows us to express the function evaluation terms $\eta(x_i); i = 1, \ldots, n$ as inner products with the reproducing kernel on $\mathcal{H}_1$

$$\eta(x_i) = \langle R_1(x_i, \cdot), \eta \rangle; \quad i = 1, \ldots, n, \tag{A.10}$$

which is equal to zero for $i = 1, \ldots, n$ due to orthogonality. Thus, with this construction, η only contributes to the minimization term through its norm, which is clearly minimized for η equal to the zero function.

We now see that for a Hilbert space $\mathcal{H}$ where function evaluation is a continuous linear operator, the minimizer of (A.9) has the form

$$\widehat{f}_\lambda = \alpha_0 + \alpha_1 x_1 + \sum_{i=1}^{n} \beta_i R_1(\cdot, x_i). \tag{A.11}$$

In summary, we started with a minimization problem involving a smoothing penalty over a Hilbert space $\mathcal{H}$ and approached the problem by decomposing the space into orthogonal subspaces. The space $\mathcal{H}_0$ is the null space of the penalty term and defines the parametric contrasts; while the space $\mathcal{H}_1$ defines the non-parametric contrasts. If $\mathcal{H}$ is a RKHS, then the problem is simplified by only having to consider a finite dimensional subspace of $\mathcal{H}_1$. This simplification came from recognizing that point evaluation in the least squares functional could be re-expressed in terms of an inner product involving the reproducing kernel on $\mathcal{H}_1$, thus by choosing a subspace spanned by slices of the reproducing kernel on $\mathcal{H}_1$, the function η no longer contributes to the least squares functional. The geometry of the Hilbert space, namely orthogonality, plays a critical role in this type of result and it is precisely continuity of the evaluation functional which allows point evaluation to be cast in terms of the geometry of the space.

BIBLIOGRAPHY

Aronszajn, N. (1950). Theory of reproducing kernels. *Trans. Amer. Math. Soc*, 68(3):337–404.

Bohrnstedt, G. W. and Goldberger, A. S. (2010). On the Exact Covariance of Products of Random Variables. *Journal of the American Statistical Association, Vol. , No. Dec., , pp.*, pages 1–5.

Cai, T. and Yuan, M. (2010). Nonparametric covariance function estimation for functional and longitudinal data. *Technical Report*.

Clarke, B., Fokoue, E., and Zhang, H. H. (2009). *Principles and Theory for Data Mining and Machine Learning*. Springer.

Giraldo, R., Delicado, P., and Mateu, J. (2010). Ordinary kriging for function-valued spatial data. *Environmental and Ecological Statistics*, 18(3):411–426.

Goulard, M. and Voltz, M. (1992). Geostatistical Interpolation of Curves: A Case Study in Soil Science. *Geostatistics Troia*, 2(Soares A (ed.)):805–816.

Gu, C. (2002). *Smoothing Spline ANOVA Models*. Springer.

Heckman, N. (1997). The theory and application of penalized least squares methods or reproducing kernel hilbert spaces made easy. *Technical Report*.

Jeganathan, C., Dash, J., and Atkinson, P. (2010). Characterising the spatial pattern of phenology for the tropical vegetation of india using multi-temporal meris chlorophyll data. *Landscape Ecology*, 25(7):1125–1141.

Kim, Y. J. and Gu, C. (2004). Smoothing spline Gaussian regression: More scalable computation via efficient approximation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 66(2):337–356.

Li, Y., Wang, N., Hong, M., Turner, N. D., Lupton, J. R., and Carroll, R. J. (2007). Nonparametric estimation of correlation functions in longitudinal and spatial data, with application to colon carcinogenesis experiments. *The Annals of Statistics*, 35(4):1608–1643.

Nerini, D., Monestiez, P., and Manté, C. (2010). Cokriging for spatial functional data. *Journal of Multivariate Analysis*, 101(2):409–418.

Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis*. Springer.

Storlie, C., Nosedal-Sanchez, A., and Lee, T. (2011). Reproducing Kernel Hilbert Spaces for Penalized Regression: A tutorial. pages 1–33.

Wahba, G. (1990). Spline models for observational data. *Society for Industrial and Applied Mathematics*.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional Data Analysis for Sparse Longitudinal Data. *Journal of the American Statistical Association*, 100(470):577–590.