ABSTRACT

KIM, JANET S. Flexible Regression Models for Functional Responses. (Under the direction of Ana-Maria Staicu.)

Regression models where both the response and the covariate are curves have become increasingly common in many scientific fields including medicine, finance, and agriculture, to name a few. These problems are often called function-on-function regression, and their primary goal is to identify associations between the response and predictor functional variables. In the first part of the thesis, we introduce a flexible regression model to study the association between a functional response and a functional covariate that are observed on the same domain. Our modeling describes the relationship between the mean current response and the covariate by a smooth unknown bi-variate function that depends on both the current value of the covariate and the time point itself. We develop estimation methodology that accommodates realistic scenarios where the covariates are sampled with or without error on a sparse and/or irregular design, and prediction that accounts for unknown model correlation structure. In this framework we also discuss the problem of testing the null hypothesis that the covariate has no association with the response. The proposed methods are evaluated numerically through simulations and two real data applications.

In the second part of the thesis, we consider non-linear regression models for functional responses and functional predictors observed on possible different domains. We introduce flexible models where the mean response at a particular time point depends on the time point itself as well as the entire covariate trajectory. In this framework, we develop computationally efficient estimation methodology and discuss prediction of a new response trajectory. We propose an inference procedure that accounts for total variability in the predicted response curves, and construct point-wise prediction intervals. The proposed estimation/inferential procedure accommodates realistic scenarios such as correlated error structure as well as sparse and/or irregular design. We investigate our methodology in finite sample size through simulations and
two real data applications.

In the third part of the thesis, we propose a method for testing linearity in function-on-function regression models for functional responses and predictors that are observed on possible different domains. Specifically, we discuss the problem of testing the null hypothesis that the covariate has a linear association with the response against a complex non-linear dependence structure. The alternative hypothesis assumes a flexible dependence structure and models their relationship through smooth tri-variate function that depends on the current time point as well as the entire covariate trajectory, where estimation of this type of models are discussed in the previous chapter. The null hypothesis is tested by representing this general additive class of models using a linear mixed effects model representation and then by testing the variance component with a nuisance variance component under the null hypothesis. Its size and power properties are assessed through simulations under various realistic scenarios and two real data applications.
Flexible Regression Models for Functional Responses

by
Janet S. Kim

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

Statistics

Raleigh, North Carolina
2016

APPROVED BY:

__________________________  ________________________
Arnab Maity               David Dickey

__________________________  ________________________
Yichao Wu                  Ana-Maria Staicu
Chair of Advisory Committee
DEDICATION

For Tack, Mom, Dad, and Christine.
BIOGRAPHY

Janet S. Kim was born in Cleveland, Ohio on November 4, 1984 and grew up in Pohang, Republic of Korea. She received a Bachelor of Science with majors in Mathematics and Statistics in August of 2008 from Ewha Womans University, Seoul, Republic of Korea. She earned a Master of Science in Mathematics in August of 2010 from Ewha Womans University. She moved to Raleigh, North Carolina in August of 2010 to pursue graduate studies in Statistics at North Carolina State University. She earned a Master of Statistics degree in May of 2012 and her PhD in May of 2016.
ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my Advisor, Ana-Maria Staicu, for her support and guidance throughout my graduate career. This research would not have been possible without her dedication and endless encouragement. In particular, I appreciate her special commitment to advising me whenever I encountered difficulties in my personal life as well as in my professional development. I would like to sincerely thank Arnab Maity for collaborating and providing invaluable input necessary to complete the research contained in this dissertation. I would like to thank my committee members, Yichao Wu and David Dickey, for providing valuable insights into this research. I also thank Siamak Khorram for serving as the graduate school representative on my committee. To all of the faculty and the staff in Statistics department at North Carolina State University (NCSU), I extend my appreciation.

To my family: I thank you for your love, encouragement, patience, and support throughout this process. Most importantly, I have the deepest gratitude for mom. She has continually given guidance and support to find my own life and to overcome difficult situations. I am grateful to my dad for teaching me to do all my best at every moment. I thank my sister, Christine, for helping family members in family medical emergency situations during my PhD studies.

Finally, to my husband, Taek: Your support, humor, and exceeding love brought this success. I am so blessed to have this journey with you. Thanks for everything.
# TABLE OF CONTENTS

List of Tables ........................................................................... viii

List of Figures .......................................................................... x

## Chapter 1 Introduction ............................................................ 1

1.1 Modeling the Functional Data ........................................... 1

1.2 Smoothing ........................................................................... 4

1.2.1 Smoothing Based on Known Basis Functions Expansions .... 4

1.2.2 Smoothing through Roughness Penalties ....................... 6

1.2.3 Data-Driven Basis Expansions .................................. 7

1.3 Regression Models for Functional Responses and Functional Covariates ......................... 9

## Chapter 2 General Functional Concurrent Model ................ 12

2.1 Introduction ....................................................................... 12

2.2 General Functional Concurrent Model ............................... 14

2.2.1 Modeling Framework ........................................... 14

2.2.2 Estimation .......................................................... 16

2.2.3 Variance Estimation ........................................... 17

2.2.4 Prediction .......................................................... 18

2.3 Hypothesis Testing .......................................................... 20

2.4 Transformation of Functional Covariate ......................... 22

2.5 Extensions ........................................................................ 23

2.5.1 Data-Processing for Irregular and Sparse Design .......... 23

2.5.2 Model Estimation for Irregular and Sparse Design .......... 24

2.5.3 Model Estimation for Multiple Predictors .................. 25

2.6 Numerical Study ............................................................. 28

2.6.1 Details of Simulation Setup .................................. 28

2.6.2 Simulation Results ............................................. 31

2.6.3 Applications ........................................................ 37

## Chapter 3 General Additive Function-on-Function Regression 43

3.1 Introduction ...................................................................... 43

3.2 Methodology ..................................................................... 45

3.2.1 Statistical Framework and Modeling .......................... 45

3.2.2 Estimation and Prediction ..................................... 47

3.3 Out-of-Sample Prediction and Inference ......................... 49

3.3.1 Estimation of Error Covariance .............................. 50

3.3.2 Out-of-Sample Prediction Inference ......................... 51

3.4 Implementation and Extensions ..................................... 53

3.5 Simulation Study ............................................................ 55

3.5.1 Details of Simulation Setup .................................. 55

3.5.2 Competitive Methods ........................................... 58
Chapter 4 Testing for Linearity in General Additive Function-on-Function Regression

4.1 Introduction .................................................. 69
4.2 Restricted Likelihood Ratio Tests in Function-on-Function Regression ....... 72
  4.2.1 Testing Procedure ........................................... 72
  4.2.2 Review of Restricted Likelihood Ratio Tests in Scalar-on-Function Re-
       gression ...................................................... 75
4.3 Data Preprocessing ........................................... 76
  4.3.1 Transformation of Covariate ................................ 76
  4.3.2 Data Pre-processing for Sparse Sampling Design .............. 76
4.4 Simulation Study .............................................. 77
  4.4.1 Simulation Setup .......................................... 77
  4.4.2 Competitive Method ....................................... 79
  4.4.3 Simulation Results ........................................ 80
4.5 Applications ................................................ 86
  4.5.1 Capital Bike Share Study .................................. 86
  4.5.2 US Yield Curves Study .................................... 87

Chapter 5 Conclusion .............................................. 89

REFERENCES ....................................................... 92

APPENDICES ......................................................... 97

Appendix A Additional Details for Chapter 2 ..................................... 98
  A.1 Additional Simulation Results ................................... 98
    A.1.1 Further Investigation of Prediction Error .................. 98
    A.1.2 Further Investigation of Different Covariance Estimation Methods 102
    A.1.3 Further Investigation of Computation Time ................ 102
  A.2 Further Investigation of Gait Data Example ...................... 104
  A.3 Implementation Details ....................................... 109

Appendix B Additional Details for Chapter 3 .................................. 111
  B.1 Additional Simulation Results ................................ 111
    B.1.1 Additional Simulations for Irregular and Sparse Design ... 111
    B.1.2 Sensitivity Analysis ....................................... 112
  B.2 Further Investigation of Real Data Example .................... 112
    B.2.1 Histogram of Standardized Average Humidity .............. 112
    B.2.2 Validation of $L^2$-Norm Based Test ..................... 113

Appendix C Additional Details for Chapter 4 ................................. 116
  C.1 Proof of Proposition 1 ........................................ 116
C.2 Additional Simulations for Irregular and Sparse Design . . . . . . . . . . . . 117
<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1</td>
<td>Summary of RMSPE$^\text{in}$, RMSPE$^\text{out}$, ICP, IL, and R(SE) based on 1000 simulated data sets. The models fitted by our method and the linear FCM are indicated by GFCM and FCM, respectively.</td>
<td>32</td>
</tr>
<tr>
<td>Table 2.2</td>
<td>Rejection probabilities ($\times 100$) at $\alpha = 5%$ and $10%$ significance levels. The values in the parenthesis are the estimated standard errors ($\times 100$) of the rejection probabilities.</td>
<td>35</td>
</tr>
<tr>
<td>Table 2.3</td>
<td>Results from Gait data example. Displayed are the summaries of RMSPE$^\text{in}$, RMSPE$^\text{out}$, ICP, IL, and R(SE). The models fitted by our method and the linear FCM are indicated by GFCM and FCM, respectively.</td>
<td>38</td>
</tr>
<tr>
<td>Table 2.4</td>
<td>Results from the calcium absorption data example. Displayed are the summaries of RMSPE$^\text{in}$, RMSPE$^\text{out}$, ICP, IL, and R(SE). The models fitted by our method and the linear FCM are indicated by GFCM and FCM, respectively.</td>
<td>41</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>Summary of (1) RMSPE$^\text{in}$ and (2) RMSPE$^\text{out}$ based on 1000 simulated data sets. Results correspond to fitting the method of GAFFR and FLM.</td>
<td>59</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>Comparisons with FAM and SFF in terms of (1) RMSPE$^\text{in}$ and (2) RMSPE$^\text{out}$, and (3) computation time (in seconds) averaged over 1000 simulations. Results correspond to the cases with $n = 50$.</td>
<td>60</td>
</tr>
<tr>
<td>Table 3.3</td>
<td>Summary of average coverage probability (ACP) for the new response $Y_{0,i}'(t)</td>
<td>X_{0,i}'(\cdot)$ at nominal significance levels $1 - \alpha = 0.85$, 0.90, and 0.95. The results are based on 1000 simulated data sets with 100 bootstrap replications per data.</td>
</tr>
<tr>
<td>Table 3.4</td>
<td>Summary of (1) RMSPE$^\text{in}$ and (2) RMSPE$^\text{out}$ in bike share data analysis.</td>
<td>64</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>Type I error rates ($\times 100$) for dense sampling design ($m_W = 51$, $m_Y = 81$ for all $i$) based on 2000 Monte Carlo simulations. The values in the parenthesis are the estimated standard errors ($\times 100$) of the rejection probabilities. Results corresponding to the proposed test and the bootstrap-based algorithm are indicated by pRLRT and bootstrap, respectively.</td>
<td>82</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Type I error rates ($\times 100$) for sparse sampling design ($m_{W,i} \sim {37,\ldots,42}$, $m_{Y,i} \sim {59,\ldots,64}$) based on 2000 Monte Carlo simulations. The values in the parenthesis are the estimated standard errors ($\times 100$) of the rejection probabilities. Results corresponding to the proposed test and the bootstrap-based algorithm are indicated by pRLRT and bootstrap, respectively.</td>
<td>83</td>
</tr>
<tr>
<td>Table A.1</td>
<td>Summary of RMSPE$^\text{in}$, RMSPE$^\text{out}$, ICP, IL, and R(SE) for a moderately sparse design. The models fitted by our method and the linear FCM are indicated by GFCM and FCM, respectively.</td>
<td>99</td>
</tr>
<tr>
<td>Table A.2</td>
<td>Summary of RMSPE$^\text{in}$ and standard deviations (in parentheses) obtained by fitting the GFCM based on 1000 simulations. The simulation settings correspond to the case where $E_i = E_i^3$.</td>
<td>100</td>
</tr>
</tbody>
</table>
Table A.3 Summary of RMSPE\textsuperscript{in} and standard deviations (in parentheses) obtained by fitting the GFCM based on 1000 simulations. The simulation settings correspond to the case where $E_i = E_3^i$. Reduced variance is used to generate the covariate trajectories.

Table A.4 Summary of ICP, IL, and R(SE) for sample size 100 and $E_i = E_3^i$ based on 1000 simulations.

Table A.5 Summary of (1) RMSPE\textsuperscript{in}, (2) RMSPE\textsuperscript{out}, and (3) average computation time in seconds obtained by fitting the GFCM based on 1000 simulations. The left (right) table is obtained by using \texttt{gam} (\texttt{bam}) function of \texttt{mgcv} R package.

Table A.6 Summary of RMSPE\textsuperscript{in}, RMSPE\textsuperscript{out}, ICP, IL, and R(SE) obtained from the simulation studies of the gait data example. The models fitted by our method and the linear FCM are indicated by GFCM and FCM, respectively.

Table B.1 Summary of (1) RMSPE\textsuperscript{in} and (2) RMSPE\textsuperscript{out} based on 1000 simulated data sets.

Table B.2 Summary of (1) RMSPE\textsuperscript{in} and (2) RMSPE\textsuperscript{out} based on 1000 simulated data sets. Results are obtained by applying the GAFFR model.

Table B.3 Summary of ACP for the new response $Y_0(t)|X_0(\cdot)$, i.e., ACPp, at nominal significance levels $1 - \alpha = 0.85, 0.90, \text{ and } 0.95$. Results are based on 1000 simulated data sets with 100 bootstrap replications per data.

Table B.4 Summary of estimated rejection probabilities ($\times 100$). The values in the parenthesis are the estimated standard errors ($\times 100$) of the rejection probabilities.

Table C.1 Type I error rates ($\times 100$) for sparse sampling design ($m_{W,i} \sim \{27, \ldots, 32\}$, $m_{Y,i} \sim \{35, \ldots, 40\}$) based on 2000 Monte Carlo simulations. The values in the parenthesis are the estimated standard errors ($\times 100$) of the rejection probabilities. Results are based on the RLRTs.
LIST OF FIGURES

Figure 2.1 Comparison of 95% prediction bands constructed for three subject-level trajectories in the test data. The case corresponds to a dense design with 100 subjects and $E^3_i$ error structure. “+” are the response $Y_{0,i}(\cdot)$ in the test data, and dotted (“•”) lines are the true response without measurement errors. Solid and dashed lines are the prediction bands obtained by fitting the GFCM and the linear FCM, respectively. The top (bottom) panel corresponds to the case when the true function $F$ is non-linear (linear). ................................................................. 34

Figure 2.2 Powers ($\times 100$) of the tests at significance level $\alpha = 5\%$. The top (bottom) panel displays the result from the densely (sparsely) sampled. The error process in the left, middle and right panels is assumed to be $E^1_i$, $E^2_i$ and $E^3_i$, respectively. ................................................................. 36

Figure 2.3 Longitudinal measurements of the hip angle (left) and the knee angle (right) obtained from 39 children while they go through a single gait cycle. 37

Figure 2.4 Results from the gait data analysis. The top panel displays 95% prediction bands obtained by fitting the GFCM (grey solid lines) and the linear FCM (dashed lines) for three subject-level trajectories in the test data. “•” represent the knee angles in the test data. The bottom panel shows the heat map of $\hat{Y}_{0,i}(t)$ obtained from the test data set of the gait data example. ................................................................. 39

Figure 2.5 Longitudinal measurements of calcium intake (left) and calcium absorption (right) obtained from 188 patients. ................................................................. 40

Figure 2.6 Results from the calcium absorption data analysis. Displayed is 95% prediction bands obtained by fitting the GFCM (grey solid lines) and the linear FCM (dashed lines) for three subject-level trajectories in the test data. “•” indicate the calcium absorption measured. Thin solid lines are the smoothed curve of the calcium absorption. ................................................................. 42

Figure 3.1 True surface of $F_3(x, s, 0.05)$ (left), $F_3(x, s, 0.5)$ (middle), and $F_3(x, s, 1)$ (right). The red solid line represents the curve obtained by fixing $s$ as 0.6. 55

Figure 3.2 The number of casual bike users (left panel) and hourly temperatures (right panel) collected every Saturday. The measurements taken in three different days on January, April, and July in 2011 are indicated by solid, dashed, and dotted lines, respectively. ................................................................. 63

Figure 3.3 Changes in the US yield curves (left panel) and in the European yield curves (right panel) during the period from January 2, 2006 to December 30, 2011. The measurements taken in three different days are indicated by red, blue, and green lines. ................................................................. 67

Figure 4.1 True surface of $F_2(x, s, 0.05)$ corresponding to the model under the null hypothesis (left) and under the alternative hypothesis (right). The red solid line shows the curve when $s$ is fixed as 0.75. ................................................................. 79
Figure 4.2  Power \(\times 100\) of the tests at significance level \(\alpha = 5\%\) based on 1000 Monte Carlo simulations. Results are for the proposed RLRTs in the case \(F(x, s, t) = F_2(x, s, t)\) and \(E_i = E^3_i\). The left and right panel correspond to the dense and sparse sampling design, respectively. 84

Figure 4.3  Power \(\times 100\) of the tests at significance level \(\alpha = 5\%\) based on 1000 Monte Carlo simulations. The left most two panels correspond to the case where \(E^1_i\) and \(E^2_i\). The right most two panels correspond to \(E^3_i\) and \(E^4_i\), respectively. Results for \(F(x, s, t) = F_1(x, s, t)\) and \(F(x, s, t) = F_2(x, s, t)\) are displayed in the top and bottom row, respectively. 85

Figure A.1  Displayed plots are curves for sample size 100 simulated based on the gait data. The smoothed version of covariate functions \(\hat{X}_{\text{sim}}\) are presented in the left. The middle and the rightmost panel present response curves \(Y_i(t)\) generated based on \(F_{\text{L,gait}}(x, t)\) and \(F_{\text{NL,gait}}(x, t)\), respectively. The last two subjects are highlighted in different colors. 106

Figure B.1  A histogram of standardized average humidity (left) and the estimated coefficient function \(\hat{g}(t)\) (right). 114
Chapter 1

Introduction

1.1 Modeling the Functional Data

Functional data analysis (FDA) has become increasingly important due to the advent of computations and technology that allow measurements to be taken on very fine grids; for a comprehensive review of FDA, refer to the monograph by Ramsay and Silverman (2005); Ferraty and Vieu (2006); Ramsay, Hooker and Graves (2009), among many others. An example of functional data is chemometric data (Ferraty and Vieu, 2003; Ferraty and Vieu, 2006). The chemometric data consists of a 100 channel spectrum of light absorbances obtained from 240 samples of finely chopped meat. For each sample, the absorbances are recorded on the Tecator Infratec food and feed analyzer operating in the 100 different wavelengths between 850nm and 1050nm. As the light absorption occurs at the wavelength sequentially, we can consider the absorbances from each sample as a function of wavelength.

Although a function is defined over an infinite-dimensional space, in practice we observe only a finite number of realization of the underlying process. In FDA, the observed data can be expressed by \( \{(t_{ij}, Y_{ij}) : j = 1, \ldots, m_1\} \), where \( Y_{ij} \) are the measurements from the \( i \)th subject at the \( j \)th time point, and \( t_{ij} \in T_X \) for some compact interval \( T_X \). In reality, the data are observed
with noise. This feature can be expressed by the model

\[ Y_{ij} = X_i(t_{ij}) + \epsilon_{ij}, \]  

where \( X_i(t_{ij}) \) are the smooth latent curve from \( i \)th subject, and \( \epsilon_{ij} \) are the errors with zero-mean and unknown covariance. Specifically, it is assumed that \( X_i(t) \) are the square-integrable random process with \( \int_{T_X} E\{X_i(t)\}^2 dt < \infty \). The number of observations, \( m_i \), does not necessarily be the same for each subject. When the data are observed on a fine and equally spaced grid, i.e., \( m_i = m \) for all \( i \), this setting is commonly referred to as dense design. The chemometric data correspond to the densely observed functional data. When the number of observations \( m_i \) is small, we call this setting sparse design. In the case of sparsely observed data, it is typically assumed that the pooled time points across the subjects, \( \bigcup_{i,j} t_{ij} \), are dense in \( T_X \).

Many theoretical and methodological developments have been undertaken in the FDA. Major topics of interest are (i) to study important sources of variation in the curves and (ii) to predict subject specific trajectories based on the observed data. A key research direction is focused on studying the variability of curves. Functional principal component analysis (FPCA) is the methodology developed to deal with this type of problems; it determines main directions of variability within the curves. FPCA is first employed by Rao (1958) and developed for both dense (see e.g., Rice and Silverman, 1991; Hall and Hosseini-Nasab, 2006; Cardot, 2007; Zhang and Chen, 2007) and sparse design (see e.g., Staniswalis and Lee, 1998; James, Hastie and Sugar, 2000; Yao et al., 2005a). FPCA is coined differently, depending on whether the sampling design is dense or fixed. If the design is dense, the common approach is to smooth each individual curve to remove the noise and then to apply PCA to the smoothed curves; Section 1.2 discusses several ways to smooth the curves. When the design is sparse, the FPCA requires pooling the data together; this approach is also detailed in Section 1.2.3. The standard FPCA methods facilitate reconstruction of full trajectories based on few observation points by representing the underlying smooth curves using eigenbasis expansions.
Various extensions and techniques have been studied up to date. Relevant research includes Ramsay and Dalzell (1991) and Ramsay and Silverman (2005) for smoothing based on FPCA, and Di et al. (2009) for multilevel FPCA, among many others. For a complete review of FPCA, see Shang (2014).

Another main research direction is developing functional regression models, where one or both of response and covariates have functional characteristics in the sense described above. Scalar-on-function regression is a relevant area that has been developed to quantify a relationship between a scalar response and functional covariate/s. For example, functional linear model (see e.g., James, 2002; James and Silverman, 2005; Müller and Stadtmüller, 2005; James, Wang and Zhu, 2009; Goldsmith et al., 2011) assumes that the effect of functional covariate is linear and predicts the scalar responses using the entire covariate trajectory through a weighted integral. Sometimes, trends may not necessarily be captured linearly, and extensions of the functional linear model to non-linear case are addressed in McLean et al. (2014). Function-on-function regression framework also has been studied through various association models such as functional concurrent model, functional historical linear model, and functional linear model (for functional responses). The functional concurrent models (Ramsay and Silverman, 2005) assume that the response profile at the current time point is only affected by the covariate profile at the same time point. However, this type of models does not allow the past or future observations of the covariate to affect the current response. Those cases are considered by the functional historical models (see e.g., Malfait and Ramsay, 2003) and functional linear models (see e.g., Ramsay and Dalzell, 1991; Ramsay and Silverman, 2005; Yao, Müller and Wang, 2005b). The functional historical models predict the current response using the past observations of the covariate, whereas the functional linear models relates the current response using the entire trajectory of the predictor.

In this thesis, we continue the study of regression models for function-on-function regression to describe more general relationships between the response and the covariate. For the rest of this section, we introduce common smoothing techniques used in the FDA such as pre-specified
basis functions expansions with or without penalty and data-driven basis expansions in Section 1.2. We discuss function-on-function regression in more detail with potential applications in Section 1.3.

1.2 Smoothing

In FDA, smoothing the functional data is an essential procedure due to the representation in (1.1). There are two important reasons for smoothing the data. One reason is to remove the measurement errors $\epsilon_{ij}$ from the observed data and recover the smooth underlying process $X_i(t)$. The other reason is to predict $X_i(t)$ at any values of $t$ when only one or few observations are available for each curve.

Smoothing can be achieved by using basis functions expansions, and the number of basis functions used to represent the curve affects to the amount of smoothness. Several different types of basis functions have been used in the literature. One of which is a known set of basis functions such as splines, wavelets, and Fourier basis. When using the pre-specified set of basis functions, one can further incorporate roughness penalty to further impose smoothness on the result. The other choice is data-driven basis such as eigenbasis of a covariance function. The data-driven basis is unknown and can be estimated from the observed data using FPCA techniques. In the next subsequent sections, we introduce smoothing based on the known basis functions expansions (Section 1.2.1), smoothing in conjunction with a roughness penalty (Section 1.2.2), and the data-driven basis expansions (Section 1.2.3).

1.2.1 Smoothing Based on Known Basis Functions Expansions

The smooth latent process $X_i(t)$ in (1.1) can be represented as a linear combination of known basis functions. For illustration, we use B-spline basis functions of order 4 (cubic B-splines). Let \{b_k(t) : k \geq 1\} be a set of B-spline basis functions, and denote by $K$ the number of basis functions. Using a basis expansion, the function $X_i(t)$ can be approximated by $X_i(t) = \sum_{k=1}^{K} b_k(t) \beta_k$, where $\beta_k$ is the unknown basis coefficients associated with $b_k(t)$. A standard method for estimat-
ing the basis coefficients is minimizing the sum of squares criterion $\sum_{i=1}^{n} ||Y_i(\cdot) - \sum_{k=1}^{K} b_k(\cdot) \beta_k ||^2$ with respect to $\beta_k$, where $||\cdot||^2$ corresponds to the common $L^2$-norm induced by the inner product $<f,g> = \int fg$.

The choice of basis functions has a little effect on the estimates $\hat{X}_i(t)$. The usual convention is to select based on the characteristics of the data. Fourier basis $\{1, \sin(2\pi t), \cos(2\pi t), \sin(4\pi t), \cos(4\pi t), \ldots\}$ is often chosen when the data has periodic nature, and it is commonly used in signal processing. When the data has non-periodic trends, spline basis functions such as B-splines and truncated power functions can be used. When there are spikes in the curve, one would prefer wavelet bases. Since our study uses B-splines for the representation of the models, we provide further details on the B-splines.

B-splines are developed by de Boor (1978) and have many attractive properties. B-splines are constructed by polynomial pieces which connect smoothly at the joints and are useful to represent a smooth curve. For example, cubic B-splines consist of four cubic polynomial pieces, and their first and second derivatives are continuous at the joined points. We call the joined points knots. Selecting the location of knots might be a challenging issue. When the resolution of the curve is dense, we use equally-spaced knots. When the resolution of the curve is low, we place more knots over regions where the function has high curvature. B-splines have compact support and are locally non-zero. Due to this property, computing the expansion of B-splines is relatively fast compared to the other basis expansions. For a comprehensive review of B-splines, we refer the reader to Eilers and Marx (1996); Ruppert et al. (2003); Marx and Eilers (2005); Ramsay and Silverman (2005); Wood (2006a).

The amount of smoothing can be controlled by selecting an optimal number of basis functions. A large number of basis functions can accommodate the complexity of function, but the estimated curve can be wiggly. A small number of basis functions can lead to overly smooth fit, but the estimated curve might be less accurate.
1.2.2 Smoothing through Roughness Penalties

We use penalized basis functions expansions (see e.g., Wahba, 1990; Ruppert et al., 2003; Wood, 2006a; Li and Ruppert, 2008) to balance the trade-off between goodness of fit as well as smoothness of fit. The idea behind this approach is that we use a large number of basis functions to fully capture the variability of the function, and impose a penalty to control the smoothness of the resulting fit. In this approach, the penalty is often referred to as roughness penalty.

To illustrate the idea, consider model (1.1), and let $Y_i = [Y_{i1}, \ldots, Y_{im_i}]^T$. An estimate of the function $X_i(t)$ can be obtained by minimizing the penalized criterion $\sum_{i=1}^{\infty} ||Y_i - X_i(\cdot)||^2 + \lambda \text{Pen}(X)$, where the first term is the sum of squares criterion, and Pen$(X)$ is the penalty that measures variability of the function $X_i(\cdot)$. The parameter $\lambda$ is known as a smoothness parameter that controls the amount of penalization. The amount of penalization is associated with the magnitude of the derivative of the function, $X_i^{(r)}(t)$, where $r$ is the order of the derivative. A common choice is the second derivative $X_i^{''}(t)$, and the size of curvature can be measured through the integrated squared second derivative $\int_T X_i^{''}(t)dt$. Then, the penalized criterion to be minimized is given by $\sum_{i=1}^{\infty} ||Y_i - X_i(\cdot)||^2 + \lambda \int_T X_i^{''}(t)dt$.

If we represent the function $X_i(t)$ using a basis functions expansion, the penalty Pen$(X)$ can be defined using the curvature of the basis functions. Let the basis expansion of $X_i(t)$ be $X_i(t) \approx B(t)\theta$, where $B(t) = [b_1(t), \ldots, b_K(t)]$ and $\theta^T = [\beta_1, \ldots, \beta_k]$. Then, we estimate the basis coefficients by minimizing the penalized criterion $\sum_{i=1}^{\infty} ||Y_i - B(\cdot)\theta||^2 + \lambda \theta^T P \theta$, where $P$ is the $K$-dimensional penalty matrix with $(k, k')$th entry equal to $\int_T b_k''(t)b_{k'}''(t)dt$.

The smoothness parameter $\lambda$ controls the smoothness of the resulting fit. As $\lambda$ becomes larger, the estimated curve becomes smoother. On the other hand, as $\lambda$ gets smaller, the estimated curve becomes rougher. If the penalty is based on the integrated squared second derivative of $X(t)$, i.e., $\lambda \int_T X_i^{''}(t)dt$, and $\lambda = \infty$, the fit approaches a straight line since every departure from linearity is penalized. The optimal smoothness parameter can be chosen based on appropriate criteria such as generalized cross validation (GCV) (see e.g., Wahba, 1990; Ruppert et al., 2003; Wood, 2006a), maximum likelihood (ML), and restricted maximum likeli-
hood (REML) (see e.g., Ruppert et al., 2003; Wood, 2006a). The REML approach is commonly
used as a method for estimating model parameters in a linear mixed effects model where the
smoothness parameter corresponds to variance components in this framework. When interest
is to obtain better estimates of the variance components, the REML method is recommended
instead of the ML approach since the ML estimation of the variance components does not com-
penstate for the loss of degrees of freedom occurred by estimating the fixed effect parameters,
which in turn may lead to biased estimators of the variance components; see Chapter 6.2.5 of
Wood (2006a) for details.

1.2.3 Data-Driven Basis Expansions

Pre-specified basis functions introduced in Section 1.2.1 are not derived from the observed data.
In this section, we introduce how to represent a smooth function using data-driven basis.

Consider model (1.1), and let \( \mu_X(t) \) be the mean of \( X_i(t) \) and \( \Sigma(t, t') = \text{Cov}\{X_i(t), X_i(t')\} \)
be the covariance function of the same process, which admits spectral decomposition. Using
data-driven basis denoted by \( \{\phi_1(\cdot), \phi_2(\cdot), \ldots\} \), the observed data \( Y_{ij} \) can be represented as
\[
Y_{ij} = \sum_{k \geq 1} \xi_{ik} \phi_k(t_{ij}) + \epsilon_{ij},
\]
where \( \xi_{ik} \) are the loadings associated with \( \phi_k(t) \). The data-driven
basis is unknown and has to be estimated from the observed data, and this makes a key difference
between the data-driven basis and the pre-specified basis. The data-driven basis is orthogonal
and describes the main directions of variability in the observed data.

In order to estimate the data-driven basis, one begins by finding FPCs from the observed
data, and the method relies on Mercer’s theorem (Bosq, 2000). Mercer’s theorem represents
a symmetric positive-definite function in terms of orthogonal eigenfunctions \( \phi_k(\cdot) \) and non-
negative eigenvalues \( \lambda_1 \geq \lambda_2 \geq \ldots \geq 0 \). Given a continuous covariance function \( \Sigma(t, t') \), Mer-
cer’s theorem allows us to represent the covariance function by \( \Sigma(t, t') = \sum_{k \geq 1} \lambda_k \phi_k(t) \phi_k(t') \).
The eigenfunctions \( \phi_k(\cdot) \) are considered as data-driven basis functions, and we commonly
refer to as eigenbasis. Once the eigenfunctions and the eigenvalues are obtained, we apply
Karhunen-Loève (KL) expansion to represent the curve \( X_i(t) \):
\[
X_i(t) = \mu_X(t) + \sum_{k \geq 1} \xi_{ik} \phi_k(t),
\]
where $\xi_{ik} = \int_{TX} \{X_i(t) - \mu_X(t)\} \phi_k(t) dt$ are uncorrelated random variables with $E(\xi_{ik}) = 0$ and $\text{Var}(\xi_{ik}) = \lambda_k$. In FPCA, $\xi_{ik}$ are called FPC scores for $X_i(t)$. The $k$th eigenvalue $\lambda_k$ measures the average variance of the $k$th FPC, $\phi_k(\cdot)$; it can be seen from a result that $\lambda_k = \lambda_k \int_{TX} \phi_k^2(t) dt = \int_{TX} \text{Var}\{\xi_{ik} \phi_k(t)\} dt$. In practice, we approximate the curve using first few eigenfunctions $\{\phi_k(t)\}_{k=1}^K$ that explain more variation in the data. The finite truncation $K$ can be determined based on pre-specified percentage of variance explained (PVE); for example, $K$ can be chosen as the smallest integer that provides $\sum_{k=1}^K \lambda_k / \sum_{k \geq 1} \lambda_k$ greater than 0.95 or 0.99 (Di et al., 2009; Staicu et al., 2010).

Estimation of the eigenfunction/eigenvalue pairs $\{\phi_k(\cdot), \lambda_k\}_{k \geq 1}$ must be treated differently, depending on sampling designs of the functional data. For densely sampled curves, we first estimate the mean $\mu_X(t)$ and covariance function $\Sigma(t, t')$ either by the method described in Section 1.2.2 or local polynomial kernel smoothing (Zhang and Chen, 2007). The eigenfunction/eigenvalue pairs can be obtained from the spectral decomposition of the estimated covariance function $\hat{\Sigma}(t, t')$. A common approach for estimating the scores $\xi_{ik}$ is based on numerical integration. For example, when Riemann sum is used, the scores can be approximated as $\hat{\xi}_{ik} \approx \sum_j \{Y_{ij} - \hat{\mu}_X(t_{ij})\} \hat{\phi}_k(t_{ij})(t_{ij} - t_{i,j-1})$, where $\{\hat{\phi}_k(t)\}_k$ are the estimated eigenfunctions. In typical functional data setting, where the grid of points for each curve is sparse, numeric integration may not provide reasonable approximations for $\xi_{ik}$. To treat this situation, we first estimate the smooth mean and covariance functions of $X_i(t)$ using kernel-based local linear smoothing in Yao et al. (2005a). The spectral decomposition of the estimated covariance function $\hat{\Sigma}(t, t')$ yields the estimated pairs of eigenvalues/eigenfunctions $\{\hat{\phi}_k(\cdot), \hat{\lambda}_k\}_{k \geq 1}$. To predict the scores $\xi_{ik}$, Yao et al. (2005a) uses conditional expectation $\hat{\xi}_{ik} = \hat{E}[\xi_{ik}|Y_{i1}, \ldots, Y_{i,m_i}]$, assuming that $\xi_{ik}$ and $\epsilon_{ij}$ are jointly Gaussian. An analytical expression of $\hat{\xi}_{ik}$ can be found as $\hat{\xi}_{ik} = \hat{\lambda}_k \hat{\phi}_k^T \Sigma_{Y_i}^{-1}(Y_i - \hat{\mu}_{X_i})$, where $Y_i = \{Y_{i1}, \ldots, Y_{i,m_i}\}^T$, $\hat{\phi}_k = \{\hat{\phi}_k(t_{i1}), \ldots, \hat{\phi}_k(t_{i,m_i})\}^T$, and $\hat{\mu}_{X_i} = \{\hat{\mu}_X(t_{i1}), \ldots, \hat{\mu}_X(t_{i,m_i})\}^T$. Here, $\Sigma_{Y_i}$ is the $m_i \times m_i$-dimensional matrix with the $(j, j')$th entry given by $\hat{\text{Cov}}\{Y_{ij}, Y_{ij'}\}$. Asymptotic consistency and asymptotic distributions of the FPC scores are discussed in Yao et al. (2005a) as well.
In Chapter 2, 3 and 4, we employ FPCA techniques (Di et al., 2009; Zhang and Chen, 2007; Yao et al., 2005a; Yao et al., 2005b) to reconstruct underlying subject-trajectories when the grid of points for each curve is sparse as well as when the observed data is contaminated with errors. In Chapter 3 and 4, we borrow low dimensional features of the data-driven basis, which allows us to develop a parsimonious modeling framework in a complex regression setting, and the process of estimating the data-driven basis uses the FPCA methods. The idea of using the data-driven basis also can be found in many other papers; see for example, Aston et al. (2010); Jiang and Wang (2010); Pomann et al. (2013); Park and Staicu (2015).

1.3 Regression Models for Functional Responses and Functional Covariates

A key focus of this thesis is to develop flexible regression models for functional responses and functional covariates. This section details several existent models which are commonly used in real data applications and are recurrent throughout the next subsequent chapters.

As mentioned previously, functional regression models where both the response and the covariate are functional have been long researched in the literature. One of the commonly known models is the functional concurrent model (FCM), where the current response is modeled based on the current value of the covariate. Functional linear concurrent models assume a linear relationship between the response and the covariate; they can be thought of a series of linear regressions for each time point, with the constraint that the regression coefficient is a univariate smooth function of time. This type of models is firstly introduced as varying coefficient model by Hastie and Tibshirani (1993), and its estimation procedure has been discussed in Ramsay and Silverman (2005) and Sentürk and Nguyen (2011). The crucial dependence assumption for this type of models may be impractical in many real data situations. To bypass this limitation, one might use the functional historical models, where the current response is modeled using only the past observations of the covariate. Such models quantify the relation
between the response and the functional covariate/s using a linear relationship via an unknown bi-variate coefficient function. Several different versions of the historical functional linear models have been introduced in the literature. For example, Malfait and Ramsay (2003) proposed a model that relates the current response to covariate’s past information using a history index. Their model describes the backward causation by defining the coefficient function on triangular support. Sentürk and Müller (2011) also developed a similar model but with different representation, assuming that the current response depends on the recent history of the covariate. Another alternative is to relate the current response to past, current and future information of the covariate. Functional linear models (FLMs) rely on the assumption that the current response depends on full trajectory of the covariate; the dependence is modeled through a weighted integral of the full covariate trajectory through an unknown bi-variate coefficient surface as weight. Estimation procedures for this model have been discussed in Ramsay and Silverman (2005); Yao et al. (2005b); Wu et al. (2010), among many others.

This thesis comprises of three projects and considers generalization of current regression models for the data observed as a pair of continuous curves. Chapter 2 proposes a flexible functional model framework to study associations between a curve response and a curve predictor observed on the same domain, and this study extends the classical functional linear concurrent model to non-linear case. The data applications in Chapter 2 involve a gait data (Olshen et al., 1989; Ramsay and Silverman 2005) and a dietary calcium absorption data (Davis, 2002; Sentürk and Nguyen, 2011) analysis. In the gait data analysis, we relate the knee angles measured at specific time points to hip angles measured at the same time points as well as the time point itself, assuming both the knee angels and the hip angels are realizations from random stochastic process. The angles are measured on a fine and regular grid of points, hence this case corresponds to densely observed functional data. The dietary calcium absorption data consists of longitudinal measurements of calcium intake and calcium absorption, where the measurements are taken occasionally due to the patient’s missed visits; this example corresponds to functional data observed on a sparse design. A primary objective of this study is to investigate
an association between the calcium intake and calcium absorption measured at the same age under a more general dependence assumption.

Chapter 3 considers a flexible class of function-on-function regression models for functional responses and covariates observed on possible different domains. Therefore, the work extends the previous study by relaxing the main assumptions that the response and the predictor are observed on the same domain and the relationship is concurrent-wise. These models allow the relationship between the response and the covariate to be non-linear and reduce to the standard FLM as a special case. Methods developed in this chapter are applied to capital bike share data (Fanaee-T and Gama, 2013) and daily yield curves data (Ruppert and Matteson, 2015) as illustrative examples. The bike share data consists of the number of casual bikes rented and weather information such as temperature and humidity, which are measured on an hourly basis. In this application, we study patterns of causal bike rental in conjunction with the weather condition of the entire day. The yield curves data has daily observations of US and European interest-rate yield curves, and we focus on the changes in the yield curves at various maturities. Specifically, we study how the changes in US yield curves are associated with the changes in European yield curves.

Chapter 4 considers the same function-on-function regression framework used in the previous chapter, but we shift our focus to testing for linearity. A primary goal in this chapter is to formally assess whether the effect of a functional covariate is linear. A testing procedure developed in this chapter is also applied to capital bike share data and daily yield curves data. In the previous chapter, the capital bike share study and the yield curves study allowed in principle to study a complex dependence structure between the functional response and the functional covariate; whereas in this chapter we investigate whether such general modeling is really necessary by assessing whether the form of the relationship is, in fact, linear.
Chapter 2

General Functional Concurrent Model

2.1 Introduction

This chapter proposes a flexible class of functional concurrent models for functional responses and functional covariates observed on the same domain. Specifically, we propose a model where the value of the response variable at a certain time point depends on both the time point and the covariate value at that time point through a smooth bi-variate unknown function. Such formulation allows us to capture potential complex relationships between response and predictor functions, while it contains the standard functional linear concurrent model as a special case.

We will show through numerical study that when the true underlying relationship is linear (that is, the linear concurrent model is in fact optimal), fitting our proposed model maintains prediction accuracy. On the other hand, when the true relationship is non-linear, fitting the linear concurrent model results in high bias and loss of prediction accuracy.

We make two main contributions. First, we propose a general non-linear functional concurrent model to describe complex association between two functional variables measured on the same domain; the approach allows prediction of the full response trajectory given a new
functional covariate and its evaluation points. We model the relationship via an unknown bi-
variate function, which we represent using tensor products of B-spline basis functions and
estimate using a penalized least squares estimation procedure in conjunction with difference
penalties (Eilers and Marx, 1996; Li and Ruppert, 2008). We discuss prediction of the response
trajectory and develop point-wise prediction intervals that account for the correlated error
structure. Accounting for the non-trivial dependence of the residuals is key for constructing
valid inference in regression models with functional outcomes; see for example Guo (2002)
Reiss et al. (2010) proposed inference for the fixed effects parameters in function-on-scalar re-
gression by using estimates of the residual covariance obtained using an iterative procedure,
and Goldsmith et al. (2015) extended these ideas to generalized multilevel function-on-scalar
regression. Scheipl et al. (2015) considered function-on-function regression models with flexible
residual correlation structure, but did not investigate numerically the effect of different correla-
tion structures on estimation and inference. We also assume a flexible correlation structure for
the residuals and account for this non-trivial dependence in the proposed statistical inference.
Specifically, we estimate the residual covariance using a two-steps estimation procedure: 1) es-
timate the population level parameters using an independent error assumption; and 2) employ
standard FPCA based methods (see e.g., Yao et al., 2005b; Di et al., 2009; Goldsmith et al.,
2013) to the residuals. The proposed inference uses the resulted estimate of the residual covari-
ance.

Second, we develop a testing procedure for assessing the global effect of the functional co-
variate, that is, to test the null hypothesis that the functional covariate has no association with
the response variable. We consider an F-ratio type test statistic (see e.g., Shen and Faraway,
2004; Xu et al., 2011) and propose a resampling based algorithm to construct the null distribu-
tion of the test statistics. Our resampling procedure takes into account the correlated error
structure and thus maintains the correct nominal size.

Our model is inspired by the model proposed in McLean et al. (2014). In particular, the non-
linear relationship that describes the conditional mean of the current response given the current value of the covariate is reminiscent of the one used in McLean et al. (2014). The key differences come from: 1) the type of response considered and 2) the covariance model assumed for the residuals. Specifically, we consider functional responses in this chapter, whereas McLean et al. (2014) studied scalar responses. As well, we assume unknown complex dependence structure of the residuals, whereas they assume independent and identically distributed (iid) normal residuals, which is reasonable in their scalar-on-function regression setting. Accounting for the dependence within the error process is an important development in the proposed inference methodology. Additionally, the proposed estimation and prediction procedures can accommodate various sampling designs for both responses and covariates, such as densely and/or sparsely sampled predictors with or without error. In contrast, the methods of McLean et al. (2014) are presented only for densely sampled functional covariates.

The structure of this chapter is as follows. Section 2.2 introduces the proposed general functional concurrent model and its estimation and prediction methods. Section 2.3 describes our resampling-based testing procedure. In Section 2.4, we describe transformation of the functional covariates required by our estimation/prediction procedure as a preliminary step, and Section 2.5 provides various extensions. We study our methods numerically and apply the methods to gait data and dietary calcium absorption data in Section 2.6.

2.2 General Functional Concurrent Model

2.2.1 Modeling Framework

Suppose for \( i = 1, \ldots, n \), we observe \( \{(W_{ij}, t_{ij}) : j = 1, \ldots, m_{Wi}\} \) and \( \{(Y_{ik}, t_{ik}) : k = 1, \ldots, m_{Yi}\} \) where \( W_{ij} \)'s and \( Y_{ik} \)'s denote the covariate and response, respectively, observed at points \( t_{ij} \) and \( t_{ik} \). It is assumed that \( t_{ij}, t_{ik} \in \mathcal{T} \) for all \( i, j \) and \( k \), and \( W_{ij} = X_i(t_{ij}) + \delta_{ij} \) where \( X_i(\cdot) \) is a random square-integrable curve defined on the compact interval \( \mathcal{T} \); for convenience we take \( \mathcal{T} = [0,1] \). It is assumed that \( \delta_{ij} \) are the iid measurement errors with mean
zero and variance $\tau^2$. To illustrate ideas, we first consider $W_{ij} = X_i(t_{ij})$, which is equivalent to $\tau^2 = 0$ and furthermore that $t_{ij} = t_j$, $t_{ik} = t_k$ and furthermore that $m_{W,i} = m_{Y,i} = m$. We treat $Y_{ik} = Y_i(t_{ik})$ to emphasize the dependence on the time points $t_{ik}$. Adaptation of our methods to more realistic scenarios where $\tau^2 > 0$ and different sampling designs for $X_i$’s and $Y_i$’s are discussed in Section 2.5.

We introduce the following general functional concurrent model (GFCM)

$$Y_i(t) = F\{X_i(t), t\} + \epsilon_i(t), \quad (2.1)$$

where $F(\cdot, \cdot)$ is an unknown smooth function defined on $\mathbb{R} \times T$, and $\epsilon_i(\cdot)$ is an error process independent of the predictor $X_i(\cdot)$ and has mean zero and unknown autocovariance function $G(\cdot, \cdot)$. The standard linear functional concurrent model is a special case of model in (2.1) with $F(x, t) = \beta_0(t) + x\beta_1(t)$, where $\beta_0(\cdot)$ and $\beta_1(\cdot)$ are unknown parameter functions. We introduce two main innovations in (2.1). First, the general bi-variate function $F(\cdot, \cdot)$ allows us to model potentially complicated relationships between $Y_i(\cdot)$ and $X_i(\cdot)$, and extends the effect of the covariate beyond linearity. Second, the covariance structure for the residual process $\epsilon_i(s)$ is assumed unknown.

In the following we develop an estimation procedure for the model components $F(\cdot, \cdot)$ and $G(\cdot, \cdot)$, discuss prediction of a new trajectory, and develop a testing procedure to formally assess the association between the response $Y_i(\cdot)$ and the true latent predictor $X_i(\cdot)$ in this general framework. The inferential procedures account for the nontrivial dependence within the subject. Our model and methodology are presented for the setting involving a single functional covariate; nevertheless they can be extended straightforwardly to incorporate other vector-valued covariates via a linear or smooth effect without much complication (refer to Section 2.5.3).
2.2.2 Estimation

We model $F(\cdot, \cdot)$ using bi-variate basis expansion using tensor product of univariate B-spline basis functions (Marx and Eilers, 2005; Wood, 2006a; McLean et al., 2014). We write \( F(x,t) = \sum_{k=1}^{K_x} \sum_{l=1}^{K_t} B_{X,k}(x) B_{T,l}(t) \theta_{k,l} \), where \( \{B_{X,k}(x) : k = 1, \ldots, K_x\} \) and \( \{B_{T,l}(t) : l, \ldots, K_t\} \) are B-splines defined on \([0,1]\), and \( \theta_{k,l} \) are unknown parameters. Then, model (2.1) can be written as

\[
Y_i(t) = \sum_{k=1}^{K_x} \sum_{l=1}^{K_t} Z_{i,k,l}(t) \theta_{k,l} + \epsilon_i(t), \tag{2.2}
\]

where \( Z_{i,k,l}(t) = B_{X,k}\{X_i(t)\} B_{T,l}(t) \), and \( K_x \) and \( K_t \) are the number of basis functions used. A larger number of basis functions would result in a better but rougher fit, while a small number of basis functions results in overly smooth estimate. As is typical in the literature, we use rich bases to fully capture the complexity of the function, and penalize the coefficients to ensure smoothness of the resulting fit. Such strategies are discussed in Wood (2006a), Wood (2006b) and Goldsmith et al. (2011), among many others.

Define the \( K_x K_t \)-vector \( Z_i(t) = [Z_{i,k,l}(t)]_{k=1}^{K_x} \) and the \( K_x K_t \)-vector of unknown coefficients \( \Theta^T = [\theta_{k,l}]_{k=1}^{K_x} \). We can rewrite (2.2) as \( Y_i(t) = Z_i(t) \Theta + \epsilon_i(t) \). To prevent overfitting, we propose to estimate \( \Theta \) by minimizing a penalized criterion \( \sum_{i=1}^{n} ||Y_i(\cdot) - Z_i(\cdot)\Theta||^2 + \Theta^T P \Theta \), where \( || \cdot ||^2 \) is the usual \( L^2 \)-norm corresponding to the inner product \( \langle f, g \rangle = \int fg \), and \( P \) is a penalty matrix containing penalty parameters that regularize the trade-off between the goodness of fit and the smoothness of fit. In practice, we observe \( Y_i(\cdot) \) and \( X_i(\cdot) \) at fine grids of points \( t_1, \ldots, t_m \); thus, we approximate the \( L^2 \)-norm terms using numerical integration. The penalized sum of square fitting criterion becomes

\[
PENSS(\Theta) = \sum_{i=1}^{n} \sum_{j=1}^{m} \{Y_i(t_j) - Z_i(t_j)\}^2 / m + \Theta^T P \Theta. \tag{2.3}
\]

Here, \( P \) is the penalty matrix defined as \( P = \lambda_x D_x^T D_x \otimes I_{K_t} + \lambda_t I_{K_x} \otimes D_t^T D_t \), where the notation \( \otimes \) stands for the Kronecker product, \( I_K \) is the identity matrix with dimension \( K \), and \( D_x \) and \( D_t \) are matrices representing the row and column of second order difference penalties.
An explicit form of the estimator $\hat{\Theta}$ is readily available for fixed values of the penalty parameters. Define the $m$-dimensional vector of response $\mathbf{y}_i = [Y_i(t_1), \ldots, Y_i(t_m)]^T$. Similarly, define $\mathbf{x}_i$ and $\mathbf{e}_i$ as $m \times K_x K_t$ dimensional vectors of covariate $\mathbf{x}_i$ and errors $\mathbf{e}_i$. Define $Z_i$ as $m \times K_x K_t$ dimensional matrix with the $j$th row given by $Z_i(t_j)$. The estimator $\hat{\Theta}$ is calculated as

$$\hat{\Theta} = H\{\sum_{i=1}^n Z_i^T \mathbf{y}_i\},$$  

(2.4)

where $H = \{(\sum_{i=1}^n Z_i^T Z_i + P)^{-1}$. The penalty parameters $\lambda_x$ and $\lambda_t$ can be chosen based on some appropriate criteria such as GCV (Ruppert et al., 2003; Wood, 2006a) or REML (Ruppert et al., 2003; Wood, 2006a). Estimation under (2.3) can be fully implemented in R using functions of the mgcv package (Wood, 2015). Modification of the estimation procedure for the case where the grid of points is irregular and sparse is presented in Section 2.5.2.

### 2.2.3 Variance Estimation

The penalized criterion (2.3) is based on working independence assumption and thus does not account for the possibly correlated error process. For valid inference about $\Theta$, one needs to account for the dependence of the residuals when deriving the variance of $\hat{\Theta}$. The variance of the parameter estimate $\hat{\Theta}$ can be calculated as $\text{Var}(\hat{\Theta}) = H\{\sum_{i=1}^n Z_i^T G Z_i\}H^T$, where $G = \text{Cov}(\mathbf{e}_i) = [G(t_j, t_k)]_{1 \leq j,k \leq m}$ is the $m \times m$ covariance matrix evaluated corresponding to the observed time points. We model the non-trivial dependence of the errors process $\epsilon(t)$ assuming that the error process has the form $\epsilon(t) = \epsilon_S(t) + \epsilon_{WN}(t)$, where $\epsilon_S$ is a zero-mean smooth stochastic process, and $\epsilon_{WN}(t)$ is a zero-mean white noise measurement error with variance $\sigma^2$. Let $\Sigma(s, t)$ be the autocovariance function of $\epsilon_S$. It follows that the autocovariance of the random deviation $\epsilon(t)$, $G(s, t) = \Sigma(s, t) + \sigma^2 I(s = t)$ where $I(\cdot)$ is the indicator function, is unknown and needs to be estimated. To this end, we assume that $\Sigma$ admits a spectral decomposition
\[ \Sigma(s,t) = \sum_{k \geq 1} \phi_k(s)\phi_k(t)\lambda_k, \] where \( \{\phi_k(\cdot), \lambda_k\} \) are the pairs of eigenvalues/eigenfunctions. We first compute the residuals \( e_{ij} = Y_i(t_j) - Z_i(t_j)\hat{\Theta} \) from the model fit, and employ FPCA methods (e.g., Yao et al., 2005a; Di et al., 2009) to estimate \( \phi_k(\cdot), \lambda_k \), and \( \sigma^2 \). Specifically, we obtain an initial smooth estimate of the covariance \( \Sigma \), remove the negative eigenvalues, and obtain a final estimate of \( \hat{\Sigma}(s,t) = \sum_{k=1}^{K} \hat{\phi}_k(s)\hat{\phi}_k(t)\hat{\lambda}_k \), where \( \{\hat{\phi}_k(\cdot), \hat{\lambda}_k\} \) are the eigenfunctions/eigenvalues of the estimated covariance \( \hat{\Sigma}(s,t) \) with \( \hat{\lambda}_1 > \hat{\lambda}_2 > \ldots > \hat{\lambda}_K > 0 \), and \( K \) is the number of eigencomponents used in the estimation. Then, we estimate \( G \) by \( \hat{G}(s,t) \approx \sum_{k=1}^{K} \hat{\lambda}_k\hat{\phi}_k(s)\hat{\phi}_k(t) + \hat{\sigma}^2 I(s=t) \) for any \( s, t \in [0,1] \), where \( \hat{\sigma}^2 \) is the estimated error variance. The finite truncation \( K \) is typically chosen by setting the PVE by the first few eigencomponents to some pre-specified value, such as 90% or 95%.

### 2.2.4 Prediction

A main focus in this chapter is prediction of response trajectory when a new covariate and its evaluation points are given. For example, in fire management an important problem is that of prediction of fuel moisture content, defined as proportion of free and absorbed water in the fuel. Study of fuel moisture content remains important for understanding fire dynamics and adequately predicting fire danger in an area of interest (see e.g., Slijepcevic et al., 2013; Slijepcevic et al., 2015). However, dynamically measuring fuel moisture content on the spot over time is difficult, and a substantial amount of research has been directed to develop physical models for predicting moisture content profiles over time based on predictors that are easily available either from weather forecast (e.g., relative humidity and temperature) or predictable from seasons (e.g., solar radiation); see for example Slijepcevic et al. (2013) for a discussion on this topic. One viable alternative is to model the past years available data using the proposed function-on-function regression model and then predict the fuel moisture trajectory for a future day based on the day’s weather forecast, so that an informative decision about fire danger can be made apriori.

Suppose that we wish to predict new, unknown response \( Y_0(t_j) \) when new observations
\(X_0(t_j) \ (j = 1, \ldots, m)\) are given. We assume that the model \(Y_0(t) = F\{X_0(t), t\} + \epsilon_0(t)\) still holds for the new data, where the error process \(\epsilon_0(t)\) has the same distributional assumption as \(\epsilon_i(t)\) in (2.1) and is independent of the new covariate \(X_0(t)\). We predict the new response \(Y_0(t)\) by \(\hat{Y}_0(t) = \sum_{k=1}^{K_x} \sum_{l=1}^{K_t} Z_{0,k,l}(t) \hat{\theta}_{k,l} \), where \(Z_{0,k,l}(t) = B_{X,k}(X_0(t))B_{T,l}(t)\), and \(\hat{\theta}_{k,l}\) are estimated based on (2.4).

Uncertainty in the prediction depends on how small the difference is between the predicted response \(\hat{Y}_0(t)\) and the true response \(Y_0(t)\). We follow an approach similar to Ruppert et al. (2003) to estimate the prediction variance. Specifically, conditional on the new covariate, we have \(\text{Var}\{Y_0(t) - \hat{Y}_0(t)\} = \text{Var}\{\epsilon_0(t)\} + \text{Var}\{\hat{Y}_0(t)\}\). Note that \(\epsilon_0(t)\) is a realization of the same error process with zero-mean and covariance structure \(G(\cdot, \cdot)\).

Let \(Z_0(t) = [Z_{0,k,l}(t)]_{k=1}^{K_x} {l=1} \) be the \(K_xK_t\)-dimensional vector defined as earlier, and \(\bar{Y}_0\) be the \(m\)-dimensional vector of \(Y_0(t_j)\) and \(\hat{Y}_0(t_j)\) respectively. Then the prediction variance becomes \(\text{Var}\{\bar{Y}_0 - \hat{Y}_0\} = G + \sum_{i=1}^{n} Z_i^T G Z_i H^T Z_0^T\), where \(Z_0\) is \(m \times K_xK_t\)-dimensional matrix with the \(j\)th row given by \(Z_0(t_j)\). Then, the prediction variance can be estimated by plugging-in the sample estimate of \(G(\cdot, \cdot)\) in this formula. One can further define a \(100(1 - \alpha)\%\) point-wise prediction interval for the new observation \(Y_0(t)\) by \(C_{1-\alpha}(t) = \hat{Y}_0(t) \pm \Phi^{-1}(1 - \alpha/2)\sqrt{\text{Var}\{Y_0(t) - \hat{Y}_0(t)\}}\), where \(\Phi(\cdot)\) is the standard Gaussian cumulative distribution function (cdf). In Section 2.5, we provide details about performing prediction in the more general case when the new covariate \(X_0(t_j)\) is only observed on a sparsely sampled grid or with measurement error.

**Remark.** The proposed estimation and prediction requires some preliminary steps. To be specific, we propose to transform the covariate functions by subtracting the point-wise mean and dividing by point-wise standard deviation function before applying the estimation and prediction procedures. The transformation of covariate is important since the set of the covariate values \(\{X_i(t_j) : i, j\}\) may not be necessarily dense over the entire domain on which the B-spline functions are defined. Therefore, there might be some situations when there are no observed data on the support of some of the B-spline basis functions. Such transformation strategies are also discussed in McLean et al. (2014). Details about the preliminary transformation are given
2.3 Hypothesis Testing

In many situations, testing for association between the response and predictor variables is as important, if not more, as it is to estimate the model components. Often before performing any estimation, it is preferred to test for association first to determine whether there is association to begin with and then a more in-depth analysis is done to determine the form of the relationship. In this section we consider the problem of testing whether the functional predictor variable is associated with the response. Specifically, we want to test

\[ H_0 : E[Y(t)|X(t) = x] = F_0(t) \quad \forall x \quad \text{versus} \quad H_1 : E[Y(t)|X(t) = x] = F_1(x,t), \]

(2.5)

where \( F_0(t) \) is univariate function and \( F_1(x,t) \) is bi-variate function, both assumed unknown.

Our testing procedure is based on first modeling the null effect \( F_0(t) \) and the full model \( F_1(x,t) \) using basis function expressions in a manner that ensures that the null model is nested within the full model. Specifically, we propose to use \( \mathbb{B}_T = \{ B_{T,0}(t) = 1, B_{T,l}(t), l \geq 1 \} \) to model \( F_0(\cdot) \) under the null model, where \( B_{T,l}(t) \) (\( l \geq 1 \)) are the B-splines evaluated at time point \( t \). To model \( F_1(\cdot,\cdot) \) under the full model, we use the same set of basis functions defined over the domain \( T \), but \( \mathbb{B}_X = \{ B_{X,0}(x) = 1, B_{X,l}(x), l \geq 1 \} \) for \( x \), where \( B_{X,l}(x) \) (\( l \geq 1 \)) are the B-splines defined over \( X \) - the image of the process \( X \), where \( X_i \sim X \). Under the full model, we can write \( F_1(x,t) = F_0(t) + \sum_{k=1}^{\infty} \sum_{l=0}^{\infty} B_{X,k}(x)B_{T,l}(t)\theta_{k,l} \). We propose to use the following \( F \)-type test statistic

\[ T_n = \frac{(RSS_0 - RSS_1)/(df_1 - df_0)}{RSS_1/(N - df_1)}, \]

(2.6)

where \( RSS_0 \) and \( df_0 \) are the residual sums of squares and the effective degrees-of-freedom (Wood, 2006a) under the null model; \( RSS_1 \) and \( df_1 \) are defined similarly but corresponding to the full model. Here, \( N \) denotes the total number of observed data points. In this case, we have
$n$ subjects and $m$ observations per subject, and thus the total number of observed data becomes $N = nm$. This can be easily generalized when each subject has different number of observations. In general, it is difficult to derive the null distribution of the proposed test statistic $T_n$ (2.6) due to the smoothing techniques and the dependence in the data.

To bypass this complication, we propose to approximate the null distribution of the test statistic $T_n$ using bootstrap of the subjects. Specifically, we use the following steps:

(i) Fit the full model described by the alternative hypothesis in (2.5) using the estimation procedure of the GFCM. Calculate the residuals $e_i(t_j) = Y_i(t_j) - \hat{Y}_i(t_j)$, for all $i$ and $j$.
(ii) Fit the null model described by the null hypothesis in (2.5) using the estimation procedure of the GFCM and estimate $F_0(t)$, $\hat{F}_0(t)$.
(iii) Calculate the value of the test statistic in (2.6) based on the null and the full model fits; call this value $T_{n,\text{obs}}$.
(iv) Resample $B$ sets of bootstrap residuals $E^*_b(t) = \{e^*_b,i(t)\}^n_{i=1}$ ($b = 1, \ldots, B$) with replacement from the residuals $\{e_i(t)\}^n_{i=1}$ obtained in step (i).
(v) For $b = 1, \ldots, B$, generate response curves under the null model as $Y^*_b,i(t) = \hat{F}_0(t) + e^*_b,i(t)$.
   This provides $B$ bootstrap data sets $\{X_i(t), Y^*_1,i(t)\}^n_{i=1}, \ldots, \{X_i(t), Y^*_B,i(t)\}^n_{i=1}$.
(vi) Given the $B$ bootstrap data sets, fit the null and the full models and evaluate the test statistic in (2.6) for each data set, $\{T^*_b\}^B_{b=1}$. These can be viewed as realizations from the distribution of $T_n$ under the assumption that $H_0$ is true.
(vii) Compute the p-value by $\hat{p} = \sum_{b=1}^B I\{T^*_b \geq T_{n,\text{obs}}\} / B$.

Our proposed resampling algorithm has two advantages. First, the exact form of the null distribution of the test statistic $T_n$ is not required; the resampled version of the test statistic approximates the null distribution automatically. Second, our algorithm accounts for correlated error process $\epsilon(\cdot)$. This is done by sampling the entire residual vectors (“curve”) for each subject; and thus preserving the correlation structure within the residuals. We observed in our numerical study (results are not shown) that preserving such correlation structure is of particular importance, as ignoring the correlation results in severely inflated type I error.
2.4 Transformation of Functional Covariate

In this section we discuss the preprocessing of the functional covariates. In particular, one challenge of our estimation approach is that some B-splines might not have observed data on its support; this issue has been also emphasized in McLean et al. (2014). This problem typically arises when the realizations of the functional covariate $X_i(t_j)$ are not dense over $\mathbb{R}$. To bypass this difficulty, we propose to first apply point-wise center/scaling transformation of the covariates; it is worthwhile to note that McLean et al. (2014) used a different approach to address this problem.

We define point-wise center/scaling transformation of $X(t)$ by

$$X^*(t) = \{X(t) - \mu_X(t)\}/\sigma_X(t)$$

where $\mu_X(t)$ and $\sigma_X(t)$ are mean and standard deviation of $X(t)$. One can interpret the transformed covariate $X^*(t)$ as the amount of standard deviation $X(t)$ is away from the mean at time $t$. In practice, we estimate the mean and the standard deviation by the sample mean $\hat{\mu}_X(t)$ and the sample standard deviation $\hat{\sigma}_X(t)$, respectively, of the covariates. Thus for a fixed point $t_j$ we will obtain realizations of the transformed covariates $\{X^*_i(t_j)\}_{i=1}^n$ based on the sample mean $\hat{\mu}_X(t_j)$ and the sample standard deviation $\hat{\sigma}_X(t_j)$ at the same point.

Our GFCM based on the transformed covariate $X^*(t)$ can be written as

$$Y_i(t) = F^*\{X_i^*(t), t\} + \epsilon_i(t) = \sum_{k=1}^{K_x} \sum_{l=1}^{K_t} Z^*_{i,k,l}(t) \theta^*_{k,l} + \epsilon_i(t), \quad (2.7)$$

where $Z^*_{i,k,l}(t) = B_{X,k}\{X_i^*(t)\} B_{T,l}(t)$ and $\theta^*_{k,l}$ are the parameters to be estimated. Note that $Z^*_{i,k,l}(t)$ and $\theta^*_{k,l}$ are different from the previous quantities denoted by $Z_{i,k,l}(t)$ and $\theta_{k,l}$ in model (2) in (2.2). To emphasize the difference, we introduce a new smooth function $F^*$ in (2.7). Let $Z^*_i(t) = [Z^*_{i,k,l}(t)]_{k=1}^{K_x} \text{ and } \Theta^* = [\theta^*_{k,l}, \ldots, \theta^*_{K_x,K_t}]^T$ be the updated notations for the model components. Then the parameter estimates $\hat{\Theta}^*$ are obtained as $\hat{\Theta}^* = \{\sum_{i=1}^n Z^*_i Z^*_i +$
\( P^{-1}\{\sum_{i=1}^{n}Z_{i}^{*T}Y_{i}\} \) where the \( j \)th row of \( Z_{i}^{*} \) is \( Z_{i}^{*}(t_{j}) \).

The prediction procedure described in Section 2.2.4 proceeds as before with the understanding that one now uses the transformed version of the new covariates. The dependence structure between \( Y_{i}(t) \) and \( X_{i}(t) \) is, in general, not the same as the one between \( Y_{i}(t) \) and \( X_{i}^{*}(t) \). Hence, the model fit \( \hat{F}^{*}(x,t) \) of \( F^{*}(x,t) \) cannot be used directly to investigate the relationship between the response \( Y_{i}(t) \) and the covariate \( X_{i}(t) \). Instead, we compare the response \( Y_{i}(t) \) and its estimator \( \hat{Y}_{i}(t) \) in a sense that the resulting estimator will not be affected by the transformation of the covariates. If the true relationship between the conditional mean of \( Y_{i}(t) \) and \( X_{i}(t) \) is captured by \( F\{X_{i}(t),t\} \) for some unknown bi-variate function \( F \), it is expected that \( \hat{Y}_{i}(t) \) well approximates \( Y_{i}(t) \).

### 2.5 Extensions

#### 2.5.1 Data-Processing for Irregular and Sparse Design

This section discusses modifications of the methodology that are required by realistic situations. In particular, we consider the case when the functional covariate is observed densely with error, or sparsely with or without noise, as well as when the sparseness of the covariate is different from that of the response.

Assume first that functional covariate is observed on a fine and regular grid of points but with error; i.e., the observed predictors are \( W_{ij} \)'s with \( W_{ij} = X_{i}(t_{j}) + \delta_{ij} \), and the deviation \( \delta_{ij} \) has variance \( \tau^{2} > 0 \). Several approaches have been proposed to adjust for the measurement errors; Zhang and Chen (2007) proposed to first smooth each noisy trajectory using local polynomial kernel smoothing, and then estimate the mean and standard deviation of the covariate \( X_{i}(t_{ij}) \) by their sample estimators. The recovered trajectories, say \( \hat{X}_{i}(\cdot) \) will estimate the latent ones \( X_{i}(\cdot) \) with negligible error. The methodology described in Section 2.2.2 can be applied with \( \hat{X}_{i}(\cdot) \) in place of \( X_{i}(\cdot) \). Numerical investigation of this approach is included in the simulation section.
Then consider the case that the functional covariate is observed on a sparse and irregular grid of points with measurement error, i.e., \( W_{ij} = X_i(t_{ij}) + \delta_{ij} \). The common assumption made for this setting is that the number of observations \( m_{W,i} \) for each subject is small, but \( \bigcup_{i=1}^{n} \{ t_{ij} \}_{j=1}^{m_{W,i}} \) is dense in \( T = [0, 1] \). Reconstructing the latent trajectories \( X_i(\cdot) \) is based on employing FPCA for sparse design to the observed \( W_{ij} \)’s. Yao et al. (2005a) proposed to 1) use local linear smoothers to estimate the mean and covariance functions, 2) estimate eigenvalues/eigenfunctions from the spectral decomposition of the estimated covariance, and 3) predict FPC scores using conditional expectation. The latent trajectories are predicted using a finite KL truncation. This method may be further applied to the response variable when the sampling design of the response is sparse as well; i.e., \( Y_{ik} = Y_i(t_{ik}) \) for \( k = 1, \ldots, m_{Y,i} \). An alternative for the latter situation is to use the prediction of the covariates at the time points \( t_{ik} \) at which the response is observed, \( \hat{X}_i(t_{ik}) \) and then continue the estimation using the data \( \{ Y_{ik}, \hat{X}_i(t_{ik}) : k \}_{i=1}^{n} \). Preliminary investigation indicated that the latter method shows good performance in both estimation and testing evaluation; the former approach seems to yield slightly increased type I error rates.

The prediction procedure can also be extended to accommodate the more general case when the new covariate \( X_0(t_j) \) is only observed on a sparsely sampled grid. We first construct a smooth version of this new covariate using the FPCA. To this end, we compute the FPCA scores for the new covariate via the conditional expectation formula in Yao et al. (2005a) with the estimated eigenfunctions from the training data, implicitly assuming that the new covariate \( X_0(\cdot) \) and the originally observed covariate \( X_i(\cdot) \) are generated from the same distribution. Then the prediction procedure can be readily applied with the smooth version of this new covariate in place of \( X_0(t_j) \).

### 2.5.2 Model Estimation for Irregular and Sparse Design

The proposed method easily accommodates more realistic situations where the covariate and/or the response are observed on sparse sampling design. Here we discuss the modifications required
by the proposed estimation procedure to accommodate such sparseness. This approach was used in the simulation study as well as in the analysis of the dietary calcium absorption data in Section 2.6. Suppose for \( i = 1, \ldots, n \) we observe the functional covariate \( X_i(t_{ik}) \) for \( j = 1, \ldots, m_{W,i} \) and the functional response \( Y_i(t_{ik}) \) for \( k = 1, \ldots, m_{Y,i} \). We first smooth the covariate \( X_i(\cdot) \) to obtain the estimated smooth curve \( \hat{X}_i(\cdot) \) of \( X_i(\cdot) \), and we evaluate the smooth curve \( \hat{X}_i(\cdot) \) at the points \( t_{ik} \) - the points at which the response is observed. Then we approximate the penalized sum of squares by

\[
PENSS(\Theta) = \sum_{i=1}^{n} \left[ \sum_{k=1}^{m_{Y,i}} \frac{\{Y_i(t_{ik}) - Z_i(t_{ik})\Theta\}^2}{m_{Y,i}} \right] + \Theta^T P \Theta,
\]

where the penalty matrix \( P \) is defined equivalently as in Section 2.2.2. To estimate the unknown parameter \( \Theta \), we minimize the \( PENSS(\Theta) \). For simplicity of exposition, let \( Y_i = [Y_i(t_{i1}), \ldots, Y_i(t_{im_{Y,i}})]^T \) and \( X_i = [X_i(t_{i1}), \ldots, X_i(t_{im_{Y,i}})]^T \) be the \( m_{Y,i} \)-dimensional vectors of response and covariate for subject \( i \). Similarly, let \( E_i = [\epsilon_i(t_{i1}), \ldots, \epsilon_i(t_{im_{Y,i}})]^T \) be the \( m_{Y,i} \)-dimensional vector of the errors. Define \( m_{Y,i} \times K_x K_T \)-dimensional matrices \( Z_i \) such that the \( j \)th row of \( Z_i \) is \( Z_i(t_{ij}) \). Then the estimator \( \hat{\Theta} \) can be computed as \( \hat{\Theta} = H\{\sum_{i=1}^{n} Z_i^TY_i\} \) with \( H = (\sum_{i=1}^{n} Z_i^T Z_i + P)^{-1} \). The variance of the parameter estimator is calculated by following the same procedure as described in Section 2.2.3:

\[
\text{Var}(\hat{\Theta}) = H\{\sum_{i=1}^{n} Z_i^T G_i Z_i\} H^T,
\]

where \( G_i = \text{Cov}(E_i) \) with dimension \( m_{Y,i} \times m_{Y,i} \) for each \( i \).

### 2.5.3 Model Estimation for Multiple Predictors

An important advantage of the proposed method is that it can easily accommodate multiple functional and scalar predictors.

Consider first the case of multiple functional predictors. For convenience, we discuss extensions of our method to two functional covariates. Assume that \( X_{1,i}(\cdot) i.i.d. \sim X_1(\cdot) \) and \( X_{2,i}(\cdot) i.i.d. \sim X_2(\cdot) \) are square-integrable random functions over \( T \), where \( X_1(\cdot) \) and \( X_2(\cdot) \) are some underlying random processes. Assume the following conditional model for the functional
response $Y_i(t)$:

$$Y_i(t) = \beta_0(t) + F_1 \{X_{1,i}(t), t\} + F_2 \{X_{2,i}(t), t\} + \epsilon_i(t), \quad (2.8)$$

where $\beta_0(t)$ is unknown and smooth intercept function, and $F_1$ and $F_2$ are unknown bi-variate functions. For identifiability we assume that $E[F_1 \{X_1(t), t\}] = 0$ and $E[F_2 \{X_2(t), t\}] = 0$. Thus $\beta_0(t)$ is the marginal mean of the functional response. For simplicity we assume that the response has zero mean. The estimation is a simple extension of the method described in Section 2.2.2.

Specifically, we use tensor product of B-splines for modeling the unknown functions $F_q$ ($q = 1, 2$):

$$F_q \{X_{q,i}(t), t\} = \sum_{k=1}^{K_{xq}} \sum_{l=1}^{K_{tq}} B_{X_q,k} \{X_{q,i}(t)\} B_{T_{q,l}}(t) \theta_{q,k,l},$$

where $\{B_{X_q,k}(x) : k = 1, \ldots, K_{xq}\}$ and $\{B_{T_{q,l}}(t) : l, \ldots, K_{tq}\}$ are the B-spline basis functions, and $\theta_{q,k,l}$ are the unknown parameters. For simplicity, define the $K_{xq} \times K_{tq}$-dimensional vector $Z_{q,i}(t) = [B_{X_q,k} \{X_{q,i}(t)\} B_{T_{q,l}}(t)]^{l=1,\ldots,K_{tq}}_{k=1,\ldots,K_{xq}}$ and the $K_{xq} \times K_{tq}$-dimensional vector of unknown coefficients $\Theta^T_q = [\theta_{q,k,l}]_{k=1,\ldots,K_{xq}}^{l=1,\ldots,K_{tq}}$ for $q = 1, 2$.

Then we can rewrite model (2.8) as $Y_i(t) = Z_{1,i}(t) \Theta_1 + Z_{2,i}(t) \Theta_2 + \epsilon_i(t)$. We estimate the unknown parameters $\Theta_1$ and $\Theta_2$ by minimizing the penalized criterion

$$\sum_{i=1}^{n} \| Y_i(\cdot) - Z_{1,i}(\cdot) \Theta_1 - Z_{2,i}(\cdot) \Theta_2 \|^2 + \Theta_1^T P_1 \Theta_1 + \Theta_2^T P_2 \Theta_2,$$

where the last two terms are the penalty terms associated with smoothness of the functions $F_1$ and $F_2$. In each of the penalty term, the penalty matrix is defined by $P_q = \lambda_{xq} D_{xq}^T D_{xq} \otimes I_{K_{tq}} + \lambda_{tq} I_{K_{xq}} \otimes D_{tq}^T D_{tq}$ and depends on two penalty parameters $\lambda_{xq}$ and $\lambda_{tq}$ ($q = 1, 2$). Here $D_{xq}$ and $D_{tq}$ are matrices representing the row and column of second order difference penalties (Eilers and Marx, 1996; Marx and Eilers, 2005; McLean et al., 2014), and the parameters $\lambda_{xq}$ and $\lambda_{tq}$ control the smoothness of the function $F_q$ in directions $x_q$ and $t$, respectively. This estimation procedure requires the transformation technique described in Section 2.4 for both $X_{1,i}(t)$ and $X_{2,i}(t)$.

Next, we consider the case of a single functional covariate $X_{1i}(t)$ and a single scalar predictor $X_{3i}$. We discuss two approaches to account for the scalar covariate: (i) in a linear way (the approach presented next) and (ii) in a non-linear way (the approach presented following the
next). Both cases require the transformation technique only for $X_{1,i}(t)$.

(i) The linear effect of the scalar covariate can be modeled through

$$
Y_i(t) = F_1\{X_{1,i}(t), t\} + X_{3,i} \beta_1(t) + \epsilon_i(t),
$$

where as before we assume $F_1$ is an unknown smooth bi-variate function, and $\beta_1(t)$ is an unknown smooth function that quantifies the effect of the scalar covariate. It is assumed that $E[F_1\{X_1(t), t\}] = 0$, where $X_{1,i}(\cdot) \overset{iid}{\sim} X_1(\cdot)$ and that $Y_i(t)$ has zero mean function. In this framework we model the coefficient function $\beta_1(t)$ by expanding univariate B-spline basis functions such that $\beta_1(t) = \sum_{l=1}^{K_{t_3}} B_{T_3,l}(t) \theta_{3,l}$, where $\{B_{T_3,l}(t): l = 1, \ldots, K_{t_3}\}$ are the B-spline basis functions for $t$, and $\theta_{3,l}$ are the unknown basis coefficients. For simplicity, define the $K_{t_3}$-dimensional vector $Z_3(t) = [B_{T_3,l}(t)]_{l=1,\ldots,K_{t_3}}$ and the $K_{t_3}$-dimensional vector of unknown coefficients $\Theta_3 = [\theta_{3,l}]_{l=1,\ldots,K_{t_3}}$. Then the above model further can be written as

$$
Y_i(t) = Z_{1,i}(t) \Theta_1 + X_{3,i} Z_3(t) \Theta_3 + \epsilon_i(t).
$$

To estimate the unknown parameters $\Theta_1$ and $\Theta_3$, we use the penalized criterion $\sum_{i=1}^{n} ||Y_i(\cdot) - Z_{1,i}(\cdot) \Theta_1 - X_{3,i} Z_3(\cdot) \Theta_3||^2 + \Theta_1^T P_1 \Theta_1 + \Theta_3^T P_3 \Theta_3$. Here the penalty matrix $P_3$ is defined by $P_3 = \lambda_{t_3} D_{t_3}^T D_{t_3}$, where $D_{t_3}$ is the second order difference penalty, and the penalty parameter $\lambda_{t_3}$ controls the smoothness of the coefficient function $\beta_1(t)$.

Accommodating a linear effect of a vector covariate is straightforward.

(ii) The smooth non-linear effect of the scalar covariate can be modeled through

$$
Y_i(t) = F_1\{X_{1,i}(t), t\} + F_3(X_{3,i}, t) + \epsilon_i(t),
$$

where $X_{3,i} \overset{iid}{\sim} X_3$, and $F_1$ and $F_3$ are the unknown bi-variate functions; here too it is assumed that the response has zero mean function. For identifiability, it is also assumed that $E[F_1\{X_1(t), t\}] = 0$ and $E[F_3(X_3, t)] = 0$. Estimation of $F_1$ and $F_3$ is analogous to that from the two functional covariates case except that transformation of covariate is not needed in the term $F_3(X_{3,i}, t)$.

The computational complexity of the algorithm could be affected by the number and the
type of additional predictors as well as their assumed effect. Specifically, for fixed values of the smoothing parameters, the solution is ready available irrespective of the number of terms in the model. Selecting the optimal values of the smoothing parameters may be computationally demanding, and thus the computational time depends on the number of smoothing parameters. Our numerical investigation confirms that the additional computation expense is rather minimal compared to the computational time corresponding to single functional covariate (the results are provided in Appendix A.1.3).

2.6 Numerical Study

In this section, we investigate the finite sample performance of our proposed methodology. Prediction accuracy is studied in Section 2.6.2, testing performance is presented in Section 2.6.2. Finally in Section 2.6.3 we apply the proposed method to the gait study (Olshen et al., 1989; Ramsay and Silverman, 2005; Huang et al., 2015) and the dietary calcium absorption study (Davis, 2002; Sentürk and Nguyen, 2011).

2.6.1 Details of Simulation Setup

In the following we provide details about the simulation specifications and the evaluation criteria.

Simulation Design

We generate the true functional covariate $X_i(t)$ by a non-linear random process $a_{0i} + a_{1i} \sqrt{2} \sin(\pi t) + a_{2i} \sqrt{2} \cos(\pi t)$ where $a_{0i} \sim N(0, 1)$, $a_{1i} \sim N(0, 0.85^2)$ and $a_{2i} \sim N(0, 0.70^2)$ for $i = 1, \ldots, n$. Throughout the study, it is assumed that the covariate $X_i(t)$ are not observed directly. Instead we observe $W_{ij} = X_i(t) + WN(0, \tau^2)$, where $\tau^2 = 0.35$. The response $Y_i(\cdot)$ is generated based on model (2.1) for the all combinations of the following factors:

(1) True function $F(x, s, t)$:
(i) linear case: $F^L(x, t) = 1 + x + t$

(ii) non-linear case: $F^{NL}(x, t) = 1 + x + t + 2x^2t$

(2) Error process $\mathbb{E}_i = [\epsilon_i(t_{i1}), \ldots, \epsilon_i(t_{im,i})]^T$:

(i) $\mathbb{E}^1_i \sim N(0, \sigma^2 \epsilon I_{m_i})$

(ii) $\mathbb{E}^2_i \sim N(0, \sigma^2 \epsilon \Sigma) + N(0, \sigma^2 \epsilon I_{m_i})$ where $\Sigma$ has AR(1) structure

(iii) $\mathbb{E}^3_i \sim \xi_{i1} \sqrt{2} \cos(\pi t) + \xi_{i2} \sqrt{2} \sin(\pi t) + N(0, \sigma^2 \epsilon I_{m_i})$

We set $\sigma^2 = 0.8$ and $\rho = 0.2$. The random variables $\xi_{i1}$ and $\xi_{i2}$ are independently generated from $N(0,2)$ and $N(0, 0.75^2)$, respectively.

(3) Sampling design:

(i) dense: $m = 81$ equidistant time points in $[0,1]$ for all $i$

(ii) sparse: $m_{Y,i}, m_{W,i} \sim \text{Uniform}(20, 31)$ (observe $24.7\% \sim 38.3\%$ of the data per curve)

We do not consider the case where the sampling design of the covariate is sparse but the response curves are sampled densely, and vice versa.

(4) Number of subjects for training data: (i) $n = 50$, (ii) $n = 100$, and (iii) $n = 300$

Our aim is to investigate the prediction accuracy of our method for both within the sample and out of sample. To achieve this, we construct training and test data sets assuming both are independent. The test sets contain $n' = 100$ subjects and are obtained using the set of $m = 81$ equispaced points in $[0,1]$.

For each of these choices we fit both the standard linear functional concurrent model (FCM) and our proposed GFCM. The purpose of our simulation study is two-fold: (i) When the true model is non-linear, i.e. $F(x, t) = F^{NL}(x, t)$, we expect to see that the GFCM to perform better than FCM; and (ii) If the true model is in fact linear, i.e. $F(x, t) = F^L(x, t)$, we expect the GFCM still maintains prediction accuracy relatively to the FCM.
Evaluation Criteria

We perform $N = 1000$ Monte Carlo simulations. Our performance measures are the root mean squared prediction error (RMSPE), the integrated coverage probability (ICP) and the integrated length (IL) of prediction bands. Define the in-sample RMSPE by

$$\text{RMSPE}^{\text{in}} = \sum_{r=1}^{N} \left[ \sum_{i=1}^{n} \frac{1}{m_i} \sum_{j=1}^{m_i} \left( Y_{i}^{(r)}(t_{ij}) - \hat{Y}_{i}^{(r)}(t_{ij}) \right) \right]^2 / \left( n \cdot N \right),$$

where $Y_{i}^{(r)}(t_{ij})$ and its estimate $\hat{Y}_{i}^{(r)}(t_{ij})$ are from the $r$th Monte Carlo simulation. The out-of-sample RMSPE, denoted by $\text{RMSPE}^{\text{out}}$, is defined similarly.

We approximate $(1 - \alpha)$ level point-wise prediction intervals to observe coverage probabilities at the nominal level. The ICP at the $(1 - \alpha)$ level is given by

$$\text{ICP}(1 - \alpha) = \sum_{r=1}^{N} \sum_{i'=1}^{100} \int_{0}^{1} I\{ Y_{0,i'}(t) \in C_{1-\alpha,i'}^{(r)}(t) \} dt / (100 \cdot N),$$

where $C_{1-\alpha,i'}^{(r)}(t)$ are the point-wise prediction intervals from the $r$th Monte Carlo simulation and $I(\cdot)$ is the indicator function. We observe particular features of the prediction intervals by measuring their length. The IL of the $(1 - \alpha)$ level prediction intervals is defined by

$$\text{IL} = \sum_{r=1}^{N} \sum_{i'=1}^{100} \int_{0}^{1} 2\text{MOE}_{1-\alpha,i'}^{(r)}(t) dt / (100 \cdot N),$$

with $\text{MOE}_{1-\alpha,i'}^{(r)}(t) = \Phi^{-1}(1 - \alpha/2) \times \hat{sd}\{ Y_{0,i'}(t) - \hat{Y}_{0,i'}^{(r)}(t) \}$. Although the IL of the prediction intervals remains small, the band might fluctuate dramatically at some time point, and this will demonstrate a poor performance of variance estimation. Hence, we examine the range of the estimated standard errors (SE). For the prediction intervals, we define the minimum SE by

$$\min(\text{SE}) = \sum_{r=1}^{N} \sum_{i'=1}^{100} \min_{t} \{ 2\text{MOE}_{1-\alpha,i'}^{(r)}(t) \} / (100 \cdot N).$$

We define the maximum SE, $\max(\text{SE})$, similarly. Then, $\text{R(SE)} = [\min(\text{SE}), \max(\text{SE})]$ provides
the range of SE at the \((1 - \alpha)\) level. For the prediction intervals, the nominal significance level is set as 0.95, 0.90, and 0.85.

### 2.6.2 Simulation Results

#### Prediction Performance

We obtain both the GFCM and the linear FCM fits using \(K_x = K_t = 7\) cubic B-splines. For estimation of the residual covariance, the percentage of explained variance is set to 99%. Table 2.1 summarizes the predictive performance of the proposed procedure for different simulation scenarios based on 1000 Monte-Carlo simulations.

We first discuss the case when the true function is non-linear (the top two panels in Table 2.1). In this case the proposed GFCM outperforms the common linear FCM. We observe that \(\text{RMSPE}^\text{in}\) and \(\text{RMSPE}^\text{out}\) from fitting the GFCM are smaller than those from the linear FCM in all scenarios. The ICPS from the GFCM and the linear FCM are fairly close to the nominal levels of 0.95, 0.90, and 0.85. However, on an average the linear FCM produces larger intervals, indicated by larger IL values and wider range of the estimated standard error denoted by \(R(\text{SE})\) compared to the GFCM. Such patterns confirm that the variability in prediction is not properly captured by the linear FCM when the true model is non-linear. For less complicated error patterns such as \(E_i^1\) (independent error structure), both models produce smaller prediction errors; nevertheless GFCM still produces smaller errors compared to linear FCM. The results are valid for different sample sizes as well as for dense/sparse sampling designs.

The bottom two panels of Table 2.1 show the results when the underlying model is linear. In this case, it is expected that the linear FCM provides the optimal results. Nevertheless, GFCM continues to show very good performance; the results are almost identical to the ones yielded by the linear FCM. Thus, even when the true model is linear and one fits a more complicated model using the GFCM, the prediction accuracy and coverage performance remain impressive. The results are slightly affected by the number of subjects, sparseness of the sampling, and the error covariance structure.
Table 2.1: Summary of RMSPE\text{in}, RMSPE\text{out}, ICP, IL, and (R(SE) based on 1000 simulated data sets. The models fitted by our method and the linear FCM are indicated by GFCM and FCM, respectively.

<table>
<thead>
<tr>
<th>n</th>
<th>RMSPE\text{in}</th>
<th>RMSPE\text{out}</th>
<th>ICP</th>
<th>IL</th>
<th>R(SE)</th>
<th>ICP</th>
<th>IL</th>
<th>R(SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>True model is $F(x) = F(x, (F(x, \text{true}))$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.37 ± 0.07</td>
<td>1.34 ± 0.06</td>
<td>0.909 ± 0.012</td>
<td>0.920 ± 0.015</td>
<td>3.22 ± 0.25</td>
<td>3.25 ± 0.25</td>
<td>0.86 ± 0.02</td>
<td>0.89 ± 0.02</td>
</tr>
<tr>
<td>Densely Sampled Design</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>1.36 ± 0.07</td>
<td>1.34 ± 0.06</td>
<td>0.909 ± 0.012</td>
<td>0.920 ± 0.015</td>
<td>3.22 ± 0.25</td>
<td>3.25 ± 0.25</td>
<td>0.86 ± 0.02</td>
<td>0.89 ± 0.02</td>
</tr>
<tr>
<td>100</td>
<td>1.37 ± 0.07</td>
<td>1.34 ± 0.06</td>
<td>0.909 ± 0.012</td>
<td>0.920 ± 0.015</td>
<td>3.22 ± 0.25</td>
<td>3.25 ± 0.25</td>
<td>0.86 ± 0.02</td>
<td>0.89 ± 0.02</td>
</tr>
<tr>
<td>300</td>
<td>1.36 ± 0.07</td>
<td>1.34 ± 0.06</td>
<td>0.909 ± 0.012</td>
<td>0.920 ± 0.015</td>
<td>3.22 ± 0.25</td>
<td>3.25 ± 0.25</td>
<td>0.86 ± 0.02</td>
<td>0.89 ± 0.02</td>
</tr>
</tbody>
</table>

| True model is $F(x) = F(x, (F(x, \text{true}))$ |
| 50 | 1.37 ± 0.07 | 1.34 ± 0.06 | 0.909 ± 0.012 | 0.920 ± 0.015 | 3.22 ± 0.25 | 3.25 ± 0.25 | 0.86 ± 0.02 | 0.89 ± 0.02 |
| 100 | 1.37 ± 0.07 | 1.34 ± 0.06 | 0.909 ± 0.012 | 0.920 ± 0.015 | 3.22 ± 0.25 | 3.25 ± 0.25 | 0.86 ± 0.02 | 0.89 ± 0.02 |
| 300 | 1.36 ± 0.07 | 1.34 ± 0.06 | 0.909 ± 0.012 | 0.920 ± 0.015 | 3.22 ± 0.25 | 3.25 ± 0.25 | 0.86 ± 0.02 | 0.89 ± 0.02 |

| True model is $F(x) = F(x, (F(x, \text{true}))$ |
| 50 | 1.37 ± 0.07 | 1.34 ± 0.06 | 0.909 ± 0.012 | 0.920 ± 0.015 | 3.22 ± 0.25 | 3.25 ± 0.25 | 0.86 ± 0.02 | 0.89 ± 0.02 |
| 100 | 1.37 ± 0.07 | 1.34 ± 0.06 | 0.909 ± 0.012 | 0.920 ± 0.015 | 3.22 ± 0.25 | 3.25 ± 0.25 | 0.86 ± 0.02 | 0.89 ± 0.02 |
| 300 | 1.36 ± 0.07 | 1.34 ± 0.06 | 0.909 ± 0.012 | 0.920 ± 0.015 | 3.22 ± 0.25 | 3.25 ± 0.25 | 0.86 ± 0.02 | 0.89 ± 0.02 |
To aid understanding these results, Figure 2.1 displays prediction bands for three subject-level trajectories, when the true model is non-linear (top panel) and linear (bottom panel), and the covariates are observed densely. In the top panel, the prediction bands for the linear FCM (dashed line) are much wider than the bands from the GFCM (solid line). This indicates that variance estimation from the linear FCM is less accurate. In the bottom panel, the prediction bands of the GFCM (grey solid line) and the linear FCM (dashed line) are almost identical, indicating a similar performance in terms of the variance estimation when the true model is linear.

To summarize, the numerical investigation shows that GFCM may result in significant gain in prediction accuracy over the standard linear FCM, when the true model is non-linear; GFCM has similar prediction performance relative to the linear FCM, when the true model is linear. In Appendix A.1, we include additional simulation results. Appendix A.1.1 discusses additional simulation results corresponding to another level of sparseness. Appendix A.1.2 compares the results corresponding to two competitive approaches for covariance estimation: using local linear smoothing Yao et al. (2005a) which is implemented in Matlab using the functions of the PACE package and using global smoothing Xiao, Li and Ruppert (2013) which is implemented in R using fpca.face function of the refund package (Huang et al., 2015). Appendix A.1.3 provides simulation results based on two functional covariates, and discuss computational aspects of our method.

**Testing Performance**

Next, we assess the performance of the proposed testing procedure. Assume the data is generated corresponding to the true model by \( F_d(x, t) = 1 + 2t + t^2 + d(xt/8) \) and using the error covariance structures and different sampling design of the covariates as above. When \( d = 0 \), the true model is a univariate function of time point \( t \), while when \( d > 0 \), the true model depends on both \( x \) and \( t \). Thus, the parameter \( d \) indexes the departure from the null hypothesis given in (2.5). Type I error of the test is investigated by setting \( d = 0 \), and the power of the test is studied for
Figure 2.1: Comparison of 95% prediction bands constructed for three subject-level trajectories in the test data. The case corresponds to a dense design with 100 subjects and $E^3_i$ error structure. “+” are the response $Y_{0,i'}(\cdot)$ in the test data, and dotted (“•”) lines are the true response without measurement errors. Solid and dashed lines are the prediction bands obtained by fitting the GFCM and the linear FCM, respectively. The top (bottom) panel corresponds to the case when the true function $F$ is non-linear (linear).
Table 2.2: Rejection probabilities (×100) at \( \alpha = 5\% \) and \( 10\% \) significance levels. The values in the parenthesis are the estimated standard errors (×100) of the rejection probabilities.

<table>
<thead>
<tr>
<th>n</th>
<th>( \bar{E}^1_1 )</th>
<th>( \bar{E}^2_1 )</th>
<th>( \bar{E}^3_1 )</th>
<th>( \bar{E}^1_2 )</th>
<th>( \bar{E}^2_2 )</th>
<th>( \bar{E}^3_2 )</th>
<th>( \bar{E}^1_3 )</th>
<th>( \bar{E}^2_3 )</th>
<th>( \bar{E}^3_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>6.5(0.8)</td>
<td>7.6(0.8)</td>
<td>7.6(0.8)</td>
<td>13.8(1.1)</td>
<td>14.1(1.1)</td>
<td>13.8(1.1)</td>
<td>4.9(0.7)</td>
<td>7.1(0.8)</td>
<td>6.0(0.8)</td>
</tr>
<tr>
<td>100</td>
<td>5.3(0.7)</td>
<td>5.4(0.7)</td>
<td>6.6(0.8)</td>
<td>11.1(1.0)</td>
<td>10.6(1.0)</td>
<td>10.3(1.0)</td>
<td>5.5(0.7)</td>
<td>5.0(0.7)</td>
<td>4.6(0.7)</td>
</tr>
<tr>
<td>300</td>
<td>6.1(0.8)</td>
<td>5.0(0.7)</td>
<td>5.1(0.7)</td>
<td>10.6(1.0)</td>
<td>10.9(1.0)</td>
<td>9.4(0.9)</td>
<td>4.8(0.7)</td>
<td>4.2(0.6)</td>
<td>5.2(0.7)</td>
</tr>
</tbody>
</table>

positive values of \( d \). We report the performance of the algorithm based on 1000 Monte Carlo simulations and \( B = 200 \) bootstraps for each simulation.

To assess the size of the testing procedure, we fix the nominal levels to \( \alpha = 5\% \) and \( 10\% \). The rejection probabilities are presented in Table 2.2 for sample sizes ranging from 50 to 300. We observe that the estimated sizes are slightly larger than the respective nominal levels for \( n = 50 \), but they approach the nominal levels for larger sample sizes such as \( n = 100 \) and \( n = 300 \). The performance is similar across different covariance structures and sampling designs of the covariates (sparse and dense).

We evaluate the power performance of our proposed test, for fixed nominal level \( \alpha = 5\% \). Figure 2.2 displays the rejection probabilities for \( d = 0.1, 1, 2, 3, 4, 5, 6 \) for different error covariance and sampling designs. As expected, the power of the testing procedure increases with the sample size and decreases when the covariates are sparsely observed. Furthermore, the power is affected by the complexity of the error covariance: the power corresponding to non-stationary error covariance is much lower than the counterpart corresponding to AR(1) covariance error structure.

Finally, it is worthwhile noting that, when calculating the proposed test \( T_{n,obs} \) or \( T_{b*} \), the difference \( RSS_0 - RSS_1 \) occasionally comes out negative; this is true irrespective whether the sampling design is dense or sparse. In these cases, we set \( RSS_0 - RSS_1 = 0 \). Typically \( df_0 - df_1 \) is positive. However for very few cases this difference was returned negative; we excluded such cases from our study.
Figure 2.2: Powers ($\times 100$) of the tests at significance level $\alpha = 5\%$. The top (bottom) panel displays the result from the densely (sparsely) sampled. The error process in the left, middle and right panels is assumed to be $E_1$, $E_2$ and $E_3$, respectively.
2.6.3 Applications

Gait Data

We turn now to our data applications. We consider first the study of gait deficiency, where the objective is to understand how the joints in hip and knee interact during a gait cycle (Theologis, 2009). Typically, one represents the timing of events occurring during a gait cycle as a percentage of this cycle, where the initial contact of a foot is recorded as 0% and the second contact of the same foot as 100%. The data consist of longitudinal measurements of hip and knee angles taken on 39 children as they walk through a single gait cycle (Olshen et al., 1989; Ramsay and Silverman, 2005). The hip and knee angles are measured at 20 evaluation points \( \{t_j\}_{j=1}^{20} \) in \([0,1] \), which are translated from percent values of the cycle. Figure 2.3 displays the observed individual trajectories of the hip and knee angles.

We consider our proposed methodology to relate the hip and knee angles, and this is an example of densely observed functional covariates and response. Let \( Y_{ij} = Y_i(t_j) \) be the knee angle and \( W_{ij} = X_i(t_j) + \delta_{ij} \) be the hip angle corresponding to the \( i \)th child and percentage of gait cycle \( t_j \). Here \( \delta_{ij} \) are the measurement errors. We first employ our resampling based test to
Table 2.3: Results from Gait data example. Displayed are the summaries of $\text{RMSPE}^{\text{in}}$, $\text{RMSPE}^{\text{out}}$, ICP, IL, and R(SE). The models fitted by our method and the linear FCM are indicated by GFCM and FCM, respectively.

<table>
<thead>
<tr>
<th></th>
<th>$\text{RMSPE}^{\text{in}}$</th>
<th>$\text{RMSPE}^{\text{out}}$</th>
<th>ICP</th>
<th>IL</th>
<th>R(SE)</th>
<th>ICP</th>
<th>IL</th>
<th>R(SE)</th>
<th>ICP</th>
<th>IL</th>
<th>R(SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFCM</td>
<td>5.26</td>
<td>5.55</td>
<td>0.956</td>
<td>20.37</td>
<td>[15.78, 36.54]</td>
<td>0.900</td>
<td>17.10</td>
<td>[13.25, 30.66]</td>
<td>0.844</td>
<td>14.96</td>
<td>[11.59, 26.83]</td>
</tr>
<tr>
<td>FCM</td>
<td>5.37</td>
<td>5.59</td>
<td>0.956</td>
<td>20.65</td>
<td>[16.08, 36.44]</td>
<td>0.883</td>
<td>17.33</td>
<td>[13.50, 30.58]</td>
<td>0.828</td>
<td>15.17</td>
<td>[11.81, 26.76]</td>
</tr>
<tr>
<td>LME</td>
<td>18.93</td>
<td>19.05</td>
<td>0.972</td>
<td>76.55</td>
<td>[73.75, 82.26]</td>
<td>0.883</td>
<td>64.24</td>
<td>[61.90, 69.04]</td>
<td>0.844</td>
<td>56.22</td>
<td>[54.17, 60.42]</td>
</tr>
</tbody>
</table>

To assess how the hip and knee angles are related to each other, we fit our proposed GFCM as well as the linear FCM. We assess the predictive accuracy by splitting the data into training and test sets of size 30 and 9. Assuming that the hip angles are observed with measurement errors, we smooth the covariate by FPCA and then apply the center/scaling transformation. We compare prediction errors obtained by fitting both the GFCM and the linear FCM. Also, as a benchmark model we further fit a linear mixed effect (LME) model $Y_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i})X_{ij} + (\beta_2 + b_{2i})t_{ij} + \epsilon_{ij}$, where $(b_{0i}, b_{1i}, b_{2i})^T$ are the subject random coefficients from $N(0, R)$ with some $3 \times 3$ unknown covariance matrix $R$, $\epsilon_{ij}$ are the errors from $N(0, \sigma^2)$, and $(b_{0i}, b_{1i}, b_{2i})^T$ and $\epsilon_{ij}$ are assumed to be independent. For the GFCM and the linear FCM, we report the RMSPE, the ICP, the IL and the R(SE). For the LME model, we report similar measures, but we take average over the repeated measurements instead of integrating over the time domain.

The results are summarized in Table 2.3. We observe that the LME model provides a poor predictive performance compared to the others, implying that models in the framework of concurrent regression model are obviously better. GFCM yields slightly better predictive performances relative to the linear FCM (negligible difference). Specifically, the prediction errors from the GFCM are smaller than the linear FCM. The R(SE) obtained from the GFCM are
Figure 2.4: Results from the gait data analysis. The top panel displays 95% prediction bands obtained by fitting the GFCM (grey solid lines) and the linear FCM (dashed lines) for three subject-level trajectories in the test data. “•” represent the knee angles in the test data. The bottom panel shows the heat map of $\hat{Y}_{0,v}(t)$ obtained from the test data set of the gait data example.
Figure 2.5: Longitudinal measurements of calcium intake (left) and calcium absorption (right) obtained from 188 patients.

narrower at all significance levels. Figure 2.4, top panel, shows the prediction bands obtained for few subjects in the test data set. The two competitive models, GFCM (solid line) and the linear FCM (dashed line), show similar results. The bottom panel in Figure 2.4 displays a heat map plot of the predicted surface using the GFCM. The results corroborate that the relationship between the hip angles and the knee angles is linear. This finding is also confirmed by additional simulations results using a generating model that mimics the gait data, which are included in Appendix A.2.

Dietary Calcium Absorption Data

Next, we consider an application to dietary calcium absorption study (Davis, 2002). In a group of 188 patients dietary and bone measurement tests are conducted approximately every five years and calcium intake and calcium absorption are measured for each subject. Let $Y_{ij} = Y_i(t_{ij})$ be the calcium absorption and $W_{ij} = X_i(t_{ij}) + \delta_{ij}$ be the calcium intake corresponding to the $i$th subject and the $j$th time point. Here $\delta_{ij}$ are the noise. The patients are between 35 and 45 years old at the beginning of the study, with the overall ages ranging from 35 to 64 years old. The number of repeated measurements per subject varies between 1 to 4 times. This is an example
Table 2.4: Results from the calcium absorption data example. Displayed are the summaries of $RMSPE^{\text{in}}$, $RMSPE^{\text{out}}$, ICP, IL, and R(SE). The models fitted by our method and the linear FCM are indicated by GFCM and FCM, respectively.

<table>
<thead>
<tr>
<th></th>
<th>$RMSPE^{\text{in}}$</th>
<th>$RMSPE^{\text{out}}$</th>
<th>$1 - \alpha = 0.95$</th>
<th>$1 - \alpha = 0.90$</th>
<th>$1 - \alpha = 0.85$</th>
<th>$RMSPE^{\text{in}}$</th>
<th>$RMSPE^{\text{out}}$</th>
<th>$ICP$</th>
<th>$IL$</th>
<th>R(SE)</th>
<th>$ICP$</th>
<th>$IL$</th>
<th>R(SE)</th>
<th>$ICP$</th>
<th>$IL$</th>
<th>R(SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFCM</td>
<td>0.079</td>
<td>0.091</td>
<td>0.990 0.34 [0.31, 0.50]</td>
<td>0.974 0.29 [0.26, 0.42]</td>
<td>0.954 0.25 [0.23, 0.37]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FCM</td>
<td>0.080</td>
<td>0.092</td>
<td>0.986 0.34 [0.31, 0.49]</td>
<td>0.972 0.29 [0.26, 0.41]</td>
<td>0.954 0.25 [0.23, 0.36]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

of data where both the functional response and covariate are observed on a sparse design.

We investigate the relationship between the calcium intake and the calcium absorption measured using the proposed methodology. For completeness we also study the dependence assuming a linear FCM. In the analysis, ages are transformed into the values in $[0,1]$, and the results are considered as evaluation points of the functions. Figure 2.5 displays the observed individual trajectories of the calcium intake and the calcium absorption along the patient’s age at the visit.

As before, we begin by testing the null hypothesis of no association between the calcium intake and the calcium absorption. We select $K_x = K_t = 11$ cubic B-splines in directions $x$ and $t$ respectively to fit the GFCM and use $B = 250$ bootstrap replications. The p-value of our test is obtained to be less than 0.004, indicating a significant association between the current calcium intake measured and the current calcium absorption.

We next analyze the predictive performance of the GFCM by using a training set of 148 random patients and a test set formed by the remaining 40 patients. Shown are also the results obtained with the linear FCM. Several adjustments are required to accommodate the sparse sampling design of the covariates. The covariates in the training set are smoothed using the standard FPCA toolkit for sparse functional data. The resulted estimated model (using the training data) is later used to reconstruct the trajectories in the test set.

The prediction results using GFCM and the competitive linear FCM are presented in Table 2.4. Both the GFCM and the linear FCM show similar in-sample and out-of-sample performance, $RMSPE^{\text{in}}$ and $RMSPE^{\text{out}}$; this indicates that a simple linear association between
Figure 2.6: Results from the calcium absorption data analysis. Displayed is 95% prediction bands obtained by fitting the GFCM (grey solid lines) and the linear FCM (dashed lines) for three subject-level trajectories in the test data. “•” indicate the calcium absorption measured. Thin solid lines are the smoothed curve of the calcium absorption.

the calcium intake and absorption is more appropriate. Furthermore, Figure 2.6 displays the point-wise prediction intervals/bands for three randomly selected subjects from the test data set. The differences between the GFCM (grey solid lines) and the linear FCM (dashed lines) are rather negligible, further confirming a linearity dependence between the calcium intake and the calcium absorption.

Remark. We have implemented our proposed estimation and testing methodology using R software, and details about the implementation are provided in Appendix A.3.
Chapter 3

General Additive

Function-on-Function Regression

3.1 Introduction

There has been a great growth in the development of function-on-function regression modeling. One of the commonly known models is the FCM where the current response relates to the current values of the covariate/s; see for example, Ramsay and Silverman (2005); Sentürek and Nguyen (2011); Kim et al. (2016). When the current response depends on the past values of the covariate/s, the historical functional linear model (Malfait and Ramsay, 2003) is more appropriate. Both the FCM and the historical functional linear model rely on a crucial assumption that both the response and the covariate/s are defined on the same domain.

In this chapter, we consider functional regression models that relate the current response to the past as well as the future of the covariate/s for responses and covariates that are defined on possible different domains. FLMs (Ramsay and Silverman, 2005; Yao et al., 2005b; Wu et al., 2010) assume that the relationship is linear; the effect of the full covariate trajectory is modeled through a weighted integral using an unknown bivariate coefficient function as weight. The linearity assumption is extended by the functional additive models (Müller and Yao, 2008)
which model the effect of covariate by the sum of smooth functions of the functional principal components (FPCs) of the covariate. A limitation of this approach is that the estimated effects are not easily interpretable. This chapter considers flexible non-linear regression models that capture potential complex relationships between the response and the full covariate trajectory directly. In this framework, we develop estimation and prediction procedures that describe the unknown model dependence under various realistic settings, and derive a valid inference for predicted response curves.

We discuss general additive function-on-function regression where the mean of the current response is modeled by a smooth tri-variate function that depends on the current time and the entire covariate trajectory. The model reduces to the standard FLM as a special case, allowing in principle to study whether such general modeling is really necessary; such direction is out of the scope of this chapter and is left for Chapter 4. These models are related to McLean et al. (2014), where the response is scalar, and they have recently been introduced by Scheipl et al. (2015), although they were not investigated formally. Scheipl et al. (2015) described a spline-based methodology which uses tri-variate B-splines obtained from tensor product of univariate B-spline basis functions, essentially extending directly the estimation proposed by McLean et al. (2014) to functional response. As expected and observed in our numerical study, this method tends to get very computationally intensive.

There are three major contributions in this paper. First, we propose an estimation approach that combines B-splines (Marx and Eilers, 2005; Wood, 2006a; McLean et al., 2014) and functional principal component basis in a computationally efficient manner. Second, we develop valid inference for out-of-sample prediction of full trajectories. The proposed prediction inference procedure accounts for non-trivial correlation structures in the error process. Finally, our method accommodates realistic scenarios such as densely or sparsely observed functional responses and covariates, possibly corrupted by measurement errors. We show numerically that when the true relationship is non-linear, our model provides an improved prediction performance over the FLM. At the same time, when the true relationship is linear, fitting our model
still maintains prediction accuracy.

The rest of the chapter is organized as follows. Section 3.2 introduces the proposed modeling framework and estimation procedure. Section 3.3 discusses out-of-sample prediction inference, and Section 3.4 presents some data-processing steps. In Section 3.5, we investigate the empirical performance of our method through simulations. Section 3.6 presents applications to bike share study and yield curves study.

3.2 Methodology

3.2.1 Statistical Framework and Modeling

Suppose for $i = 1, \ldots, n$ we observe \{(W_{ik}, s_{ik}) : k = 1, \ldots, m_{W,i}\} and \{(Y_{ij}, t_{ij}) : j = 1, \ldots, m_{Y,i}\}, where $W_{ik}$ and $Y_{ij}$ are the noisy covariate and response observed at time points $s_{ik}$ and $t_{ij}$, respectively. We assume that $s_{ik} \in T_X$ for all $i$ and $k$ and $t_{ij} \in T_Y$ for all $i$ and $j$, where $T_X$ and $T_Y$ are compact intervals. It is also assumed that $W_{ik} = X_i(s_{ik}) + \delta_{ik}$, where $X_i(\cdot)$ is the square-integrable, true smooth signal defined on $T_X$, and $\delta_{ik}$ are the iid measurement errors with zero-mean and variance equal to $\tau^2$. It is further assumed that $Y_{ij} = Y_i(t_{ij})$, where $Y_i(\cdot)$ is defined on the compact interval $T_Y$. We assume that the response has zero-mean.

To illustrate the ideas, we assume that both the response and predictor are observed on a fine, regular, and common grid of points so that $s_{ik} = s_k$ with $k = 1, \ldots, m_W$ and $t_{ij} = t_j$ with $j = 1, \ldots, m_Y$ for all $i$. We also assume that the covariates are observed without error, which is equivalent to $\tau^2 = 0$. However, these assumptions are made for illustration purposes only; the proposed methodology is applicable to more general situations.

We consider the following general additive function-on-function regression (GAFFR) model introduced by Scheipl et al. (2015)

$$ Y_i(t) = \int_{T_X} F\{X_i(s), s, t\} ds + \epsilon_i(t), \quad (3.1) $$
where $F(\cdot, \cdot, \cdot)$ is an unknown smooth tri-variate function defined on $\mathbb{R} \times \mathcal{T}_X \times \mathcal{T}_Y$, and $\epsilon(\cdot)$ is an error process with zero-mean and unknown autocovariance function $R(t, t')$ and is independent of the covariate $X_i(s)$. The form $F(\cdot, \cdot, t)$ quantifies the unknown dependence between the current response $Y_i(t)$ and the full covariate trajectory $X_i(\cdot)$. For example, if $F(x, s, t) = \beta(s, t)x$, then model (3.1) reduces to the standard FLM. For convenience, we take $\mathcal{T}_X = \mathcal{T}_Y = [0, 1]$.

One possible approach for modeling $F$ is using a tensor product of univariate B-spline basis functions for $x$, $s$, and $t$. This approach was proposed by Scheipl et al. (2015) and implemented in the R package refund (Huang et al., 2015). However, this estimation approach has not been investigated numerically. As expected and also observed in our numerical experience, such an approach requires a heavy computational burden. For example, if $F$ is modeled using a tensor product of 10 basis functions per dimension, then there are $10^3 = 1000$ basis functions in total.

We propose an efficient estimation procedure that reduces computational cost substantially, while preserving high accuracy. Let $\{\phi_k(\cdot)\}_k$ be an orthogonal basis in $L^2[0, 1]$: $\int_0^1 \phi_k(t)\phi_{k'}(t)dt = 1$ if $k = k'$ and 0 otherwise. For fixed $x$ and $s$, we represent the function $F(x, s, t)$ using orthogonal basis functions $\{\phi_k(\cdot)\}_{k \geq 1}$ in $L^2[0, 1]$: $F(x, s, t) = \sum_{k=1}^{\infty} G_k(x, s)\phi_k(t)$.

Here, $G_k(x, s)$ are the unknown basis coefficients that vary smoothly over $x$ and $s$, and calculated as $G_k(x, s) = \int_0^1 F(x, s, t)\phi_k(t)dt$. We model the unknown coefficient functions $G_k(\cdot, \cdot)$ by tensor product of finite B-spline bases. Specifically, $G_k(\cdot, \cdot)$ can be represented as $G_k(x, s) = \sum_{l=1}^{K_x} \sum_{l'=1}^{K_s} B_{X,l}(x)B_{S,l'}(s)\theta_{l,l',k}$, where $\{B_{X,l}(x)\}_{l=1}^{K_x}$ and $\{B_{S,l'}(s)\}_{l'=1}^{K_s}$ are orthogonal B-spline basis functions for $x$ and $s$ of dimensions $K_x$ and $K_s$, respectively.

Combining all these expansions, the tri-variate function $F$ can be written as $F(x, s, t) = \sum_{k=1}^{\infty} \sum_{l=1}^{K_x} \sum_{l'=1}^{K_s} B_{X,l}(x)B_{S,l'}(s)\phi_k(t)\theta_{l,l',k}$, where $\theta_{l,l',k}$ are the unknown parameters. In practice, we approximate the infinite summation by a finite truncation $K$. This representation corresponds to using tri-variate basis functions obtained by the tensor product of univariate B-spline basis functions in directions $x$ and $s$ and $L^2[0, 1]$ orthogonal basis functions $\phi_k(\cdot)$. For exposition simplicity, define the $K_xK_s$-vector $Z(i) = [\int_0^1 B_{X,l}(X_i(s))B_{S,l'}(s)ds]_{l'=1}^{K_s}$ for each $k$. Using the vector
notations, model (3.1) can be approximated as

\[ Y_i(t) \approx \sum_{k=1}^{K} Z(i) \phi_k(t) \Theta_k + \epsilon_i(t). \] (3.2)

### 3.2.2 Estimation and Prediction

We estimate the unknown parameters \( \Theta_k \)'s in (3.2) using a penalized likelihood criterion. However, unlike the standard penalized likelihood approach (Ruppert et al. 2003; Wood 2006a) which penalizes the basis coefficients in all directions, we employ quadratic penalties for the directions \( x \) and \( s \), and control the roughness in the direction \( t \) by the number of orthogonal basis functions, \( K \). Specifically, the size of the curvature in the \( x \)-direction is measured through

\[
\int \int \int \left\{ \frac{\partial^2 F(x,s,t)}{\partial x^2} \right\}^2 dx ds dt = \sum_{k=1}^{K} \int \int \left\{ \frac{\partial^2 G_k(x,s)}{\partial x^2} \right\}^2 dx ds
\]

where the \( (l,r) \)-th entry of the marginal penalty matrix \( \mathbb{P}_x \) is

\[
\int \partial_{xx} \{ B_{X,l}(x) \} \partial_{xx} \{ B_{X,r}(x) \} \quad (l,r = 1, \ldots, K_x). \]

Similarly, the size of the curvature in the \( s \)-direction is

\[
\int \int \int \left\{ \frac{\partial^2 F(x,s,t)}{\partial s^2} \right\}^2 dx ds dt = \sum_{k=1}^{K} \int \int \left\{ \frac{\partial^2 G_k(x,s)}{\partial s^2} \right\}^2 dx ds
\]

where the \( (l',r') \)-th entry of the marginal penalty matrix \( \mathbb{P}_s \) is

\[
\int \partial_{ss} \{ B_{S,l'}(s) \} \partial_{ss} \{ B_{S,r'}(s) \} \quad (l',r' = 1, \ldots, K_s). \]

The notation \( \otimes \) stands for the Kronecker product, and \( I_K \) is the identity matrix of dimension \( K \). Then the penalized criterion to be minimized is

\[
\sum_{i=1}^{n} \| Y_i(\cdot) - \sum_{k=1}^{K} Z(i) \phi_k(\cdot) \Theta_k \|^2 + \sum_{k=1}^{K} \Theta_k^T \left( \lambda_x \mathbb{P}_x \otimes I_{K_s} + \lambda_s I_{K_x} \otimes \mathbb{P}_s \right) \Theta_k,
\] (3.3)

where \( \| \cdot \|^2 \) is common \( L^2 \)-norm corresponding to the inner product \( < f, g > = \int fg \), and \( \lambda_x \) and \( \lambda_s \) are smoothness parameters that control the tradeoff between the roughness of the function and the goodness of fit. The smoothness parameters \( \lambda_x \) and \( \lambda_s \), in fact, control the smoothness of the coefficient functions \( G_k(x,s) \) in directions \( x \) and \( s \), respectively.

One convenient way to calculate the first term in (3.3) is to expand \( Y_i(\cdot) \) using the same basis function \( \{ \phi_k(\cdot) \}_k \). Specifically, if \( \{ \phi_k(\cdot) \}_k \) is the eigenbasis of the marginal covariance of
Y_i(·), then KL expansion yields
\[ Y_i(t) = \sum_k \xi_{ik} \phi_k(t) + e_{it} \]
where \( e_{it} \) is a zero-mean measurement error and \( \xi_{ik} = \int_0^1 Y_i(t) \phi_k(t) dt \); recall that the marginal mean of \( Y_i(·) \) is assumed to be zero.

Then, criterion (3.3) can be equivalently written as
\[
\sum_{k=1}^K \left[ \sum_{i=1}^n \left( \xi_{ik} - Z(i) \Theta_k \right)^2 + \Theta_k^T (\lambda_x \mathbb{P}_x \bigotimes I_{K_x} + \lambda_s I_{K_s} \bigotimes \mathbb{P}_s) \Theta_k \right].
\] (3.4)

We set \( K_x \) and \( K_s \) to be sufficiently large to capture the potential complexity of the model and penalize the basis coefficients to balance the bias and the variance of resulting fits. The smoothness parameters \( \lambda_x \) and \( \lambda_s \) can be chosen based on appropriate criteria such as GCV (Ruppert et al., 2003; Wood, 2006a) or REML (Ruppert et al., 2003; Wood, 2006a). In our numerical studies, the smoothness parameters are selected using REML.

The penalized criterion (3.4) uses the true FPC scores. In practice, we use estimates of FPC scores obtained via FPCA; more details will be provided shortly. Using the eigenbasis of the marginal covariance of the response is important due to the resulting parsimonious representation of the response and has been often used in the literature; see for example, Aston et al. (2010); Jiang and Wang (2010); Park and Staicu (2015). In our framework, using the eigenbasis can be also viewed as positing the following regression model:
\[
E[Y_i(t) | X_i(·)] = \sum_{k=1}^K \phi_k(t) E[\xi_{ik} | X_i(·)] + \epsilon_i(t),
\] (3.5)

where \( E[\xi_{ik} | X_i(·)] = \sum_{l=1}^{K_x} \sum_{l' = 1}^{K_s} \sum_{k=1}^K \theta_{l,l',k} \int_0^1 B_{X,l} \{X_i(s)\} B_{S,l'}(s) ds \), and \( \xi_{ik} \) are the FPC scores of response. The representation (3.5) is novel and extends the ideas of Aston et al. (2010) and Pomann et al. (2013) to the case of functional covariate. Also, it relates to Müller and Yao (2008) with \( E[\xi_{ik} | X_i(·)] = \sum_{m=1}^M f_{mk}(\xi_{im}) \), where \( f_{mk}(·) \) are unknown smooth functions for \( m = 1, \ldots, M \) and \( k = 1, \ldots, K \), \( \xi_{im} \) are FPC scores of the covariate \( X_i(·) \), and \( M \) is a finite truncation of the number of principal components associated with the covariate.

Estimation and prediction of the response curves \( Y_i(·) \) follow a three-step procedure: 1) reconstruct the smooth trajectory of the response \( \tilde{Y}_i(·) \) by smoothing the data of each \( i \)
(Zhang and Chen, 2007) and de-mean it, 
\[ \tilde{Y}_c^e(\cdot) = \tilde{Y}_i(\cdot) - \bar{\mu}_Y(\cdot) \]
where \( \bar{\mu}_Y(\cdot) \) is the estimated mean function; 2) use FPCA to estimate eigenbasis \( \tilde{\phi}_k(\cdot) \) of the marginal covariance of \( \tilde{Y}_i(\cdot) \), and then obtain the FPC scores \( \tilde{\xi}_{ik} = \int_0^1 \tilde{Y}_c^e(t) \tilde{\phi}_k(t) dt \); and 3) Estimate the basis coefficients \( \Theta_k \) by minimizing the penalized criterion in (3.4) with respect to \( \Theta_k \), using \( \tilde{\xi}_{ik} \) in place of \( \xi_{ik} \).

The truncation \( K \) is determined through the pre-specified percent of variance explained. The solution \( \Theta_k \) has a closed form expression:

\[ \hat{\Theta}_k = H \{ \sum_{i=1}^n Z(i)^T \tilde{\xi}_{ik} \}, \]  

where \( H = \{ \sum_{i=1}^n Z(i)^T Z(i) + P \}^{-1} \) and \( P = \lambda_x P_x \otimes I_{K_x} + \lambda_s I_{K_s} \otimes P_s \). Once the basis coefficients are estimated, \( F(\cdot, \cdot, \cdot) \) can be estimated by \( \hat{F}(x, s, t) = \sum_{k=1}^K \sum_{l=1}^{K_x} \sum_{l'=1}^{K_s} B_{X,l}(x) B_{S,l'}(s) \hat{\theta}_{l,l',k} \). Furthermore, for any \( X(s) \), the response curve can be predicted by

\[ \hat{Y}(t) = \sum_{k=1}^K \tilde{\phi}_k(t) \left[ \sum_{l=1}^{K_x} \sum_{l'=1}^{K_s} \hat{\theta}_{l,l',k} \int_0^1 B_{X,l}\{X(s)\} B_{S,l'}(s) ds \right]. \]  

The notational convention we used above is ‘\( \hat{\cdot} \)’ for prediction based on the function-on-function regression model and ‘\( \tilde{\cdot} \)’ for estimation based on the marginal analysis of response \( Y_i(\cdot) \).

### 3.3 Out-of-Sample Prediction and Inference

In this section, we focus on out-of-sample prediction and its associated inference. For example, in the capital bike share study (Fanaee-T and Gama, 2013), the primary objective is to better understand how different weather conditions (e.g., temperature and humidity) affect bike rental patterns over time. Making a valid prediction of the future hourly demands has the potential to help the company avoid situations such as unnecessarily many bikes or too few bikes available for renting. We use the weather condition of the current day and the bike supply needed on the previous day to predict the supply required for the current day. In practice, one can use the
weather forecast for the current day.

Inference on predicted response curves is not straightforward due to two important sources of variability: (1) uncertainty produced by predicting response curves conditional on the particular estimate of the eigenbasis \{\phi_k(\cdot)\}_k and (2) uncertainty induced by estimating the eigenbasis \{\phi_k(\cdot)\}. Ignoring the second source of variability may deteriorate the accuracy of the estimated total variance. Inspired by the ideas of Goldsmith et al. (2013), we assess the total variability of the predicted response curves by combining both sources of variability. As the two sources are based on the estimated error covariance, we first describe the estimation of the error covariance in Section 3.3.1, and then discuss the out-of-sample prediction inference in Section 3.3.2.

Let \( \tilde{\xi}_{ik} = \int_0^1 \tilde{Y}_i'(t)\tilde{\phi}_k(t)dt \) be the projection of the de-meaned full response curve onto the direction \( \tilde{\phi}_k(t) \), where recall \{\tilde{\phi}_k(\cdot)\}_k are obtained from the spectral decomposition of the estimated marginal covariance of the response. Define \( \text{Var}(\xi_{ik}) = \sigma^2_k, \text{Var}(\tilde{\xi}_{ik}) = \nu_{kk}, \) and \( \text{Cov}(\tilde{\xi}_{ik}, \tilde{\xi}_{ik'}) = \nu_{kk'} (k \neq k') \). For notation simplicity, let \( \eta = [K, \{\sigma^2_k, \phi_k(\cdot)\}_{k=1}^K] \) be the set of all parameters that describe the marginal covariance of the response.

### 3.3.1 Estimation of Error Covariance

To account for the non-trivial dependence of the errors process, we use ideas similar to Kim et al. (2016). We model the error process as the sum between a zero-mean smooth stochastic process, \( \epsilon_s(t) \), and a white noise process, \( \epsilon_{WN}(t) \); i.e., \( \epsilon(t) = \epsilon_s(t) + \epsilon_{WN}(t) \). It is assumed that \( \Sigma(t, t') \) is the autocovariance function of \( \epsilon_s(t) \), and \( \sigma^2 \) is the variance of \( \epsilon_{WN}(t) \). It follows that the autocovariance function of \( \epsilon(t) \) is \( R(t, t') = \Sigma(t, t') + \sigma^2 I(t = t') \), where \( I(\cdot) \) is the indicator function. Then, the estimation of \( R(t, t') \) follows two steps: 1) fit the GAFFR model using working independence assumption and obtain residuals, \( e_{ij} = Y_{ij} - \hat{Y}_i(t_j) \) where \( \hat{Y}_i(t) = \sum_{k=1}^K Z(i) \tilde{\phi}_k(t)\hat{\Theta}_k \); and 2) employ standard FPCA based methods (see e.g., Yao et al., 2005a; Di et al., 2009) to the resulted residuals to estimate a finite rank approximation of \( R(t, t') \), based on the estimated eigencomponents and estimated error variance.
3.3.2 Out-of-Sample Prediction Inference

We now discuss the variability of the predicted response curves when new covariate profiles are observed. Let $X_0(\cdot)$ be the new functional covariate observed completely on $T_X$ and assume $Y_0(t) = \int_{T_X} F\{X_0(s), s, t\} ds + \epsilon_0(t)$. It is further assumed that the error process $\epsilon_0(t)$ has the same distributional assumptions as $\epsilon_i(t)$ in (3.1) and is independent of $X_0(s)$. We measure the uncertainty in the prediction by approximating the variance of the prediction error $\hat{Y}_0(t) - Y_0(t)$ (Ruppert et al., 2003), which is defined as $\text{Var}\{\hat{Y}_0(t) - Y_0(t)\} = \text{Var}\{\hat{Y}_0(t)\} + \text{Var}\{\epsilon_0(t)\}$. The variance of $\epsilon_0(t)$ can be estimated by $\hat{R}(t, t')$, as the error process $\epsilon_0(t)$ follows the same error process as $\epsilon_i(t)$. We approximate $\text{Var}\{\hat{Y}_0(t)\}$ by applying the iterated variance formula:

$$\text{Var}\{\hat{Y}_0(t)\} = \mathbb{E}_{\tilde{\eta}}[\text{Var}\{\hat{Y}_0(t)|\tilde{\eta}\}] + \text{Var}_{\tilde{\eta}}[\mathbb{E}\{\hat{Y}_0(t)|\tilde{\eta}\}], \quad (3.8)$$

where $\tilde{\eta}$ is the estimator of $\eta$.

We begin with deriving the model-based variance that specifically quantifies the variability of the conditional estimates, $\text{Var}\{\hat{Y}_0(t)|\tilde{\eta}\}$. From equation (3.7), the model-based variance can be written as

$$\text{Var}\{\hat{Y}_0(t)|\tilde{\eta}\} = \sum_{k=0}^{K} \tilde{\phi}_k(t) \theta_k Z_0 \phi_k(t) + \sum_{k \neq k'} \tilde{\phi}_k(t) Z_0 \Sigma_{kk'} Z_0 \phi_{k'}(t),$$

where $Z_0 = \int_{T_0} B_{X,t'} \{X_0(s)\} B_{S,t'} ds, \Sigma_{kk'} = \text{Var}(\hat{\theta}_k, \hat{\theta}_{k'})$. The variance of $\hat{\theta}_k$ can be calculated as

$$\text{Var}(\hat{\theta}_k) = \nu_{kk} H\{\sum_{i=1}^{n} \phi(i)^T \phi(i)\} H^T,$$

and its covariance is $\text{Cov}(\hat{\theta}_k, \hat{\theta}_{k'}) = \nu_{kk'} H\{\sum_{i=1}^{n} \phi(i)^T \phi(i)\} H^T$. Then, the model-based conditional variance of the predicted response is given by

$$\text{Var}\{\hat{Y}_0(t)|\tilde{\eta}\} = \sum_{k=1}^{K} \nu_{kk} \tilde{\phi}_k(t) \Omega_0 \phi_k(t) + \sum_{k \neq k'} \nu_{kk'} \tilde{\phi}_k(t) \Omega_0 \phi_{k'}(t), \quad (3.9)$$

where $\Omega_0 = Z_0 H\{\sum_{i=1}^{n} \phi(i)^T \phi(i)\} H^T Z_0^T$, implicitly assuming that this variance is conditioned by $X_0(s)$. We estimate $\text{Var}\{\hat{Y}_0(t)|\tilde{\eta}\}$ by plugging estimates of $\nu_{kk}$ and $\nu_{kk'}$ in (3.9). When the response curve is observed on a fine and regular grid of points, we estimate $\nu_{kk}$ by

$$\bar{\nu}_{kk} = \int \int \tilde{\Sigma}_Y(t, t') \phi_k(t) \phi_k(t') dt dt',$$

where $\tilde{\Sigma}_Y(\cdot, \cdot)$ is the estimated marginal covariance function.
of response, and $\bar{\nu}_{kk'} \approx 0$ for $k \neq k'$. When the response curve is observed on sparse and irregular grid of points, modification is needed to obtain $\bar{\nu}_{kk}$ and $\bar{\nu}_{kk'}$; we detail this case in Section 3.4.

To account for the second source of variability, we use bootstrapping of the subjects. We approximate the total variance of $\hat{Y}_0(t)$ using the iterated variance formula in (3.8); the first term, $E_\eta[\text{Var}\{\hat{Y}_0(t)|\hat{\eta}\}]$, can be estimated by averaging the model-based conditional variances across bootstrap samples. The second term, $\text{Var}_\hat{\eta}[E\{\hat{Y}_0(t)|\hat{\eta}\}]$, is estimated by the sample variance of the predicted responses obtained for each bootstrap sample. Algorithm 1 details how to compute the total variance of $\hat{Y}_0(t)$. Using this result, we can further construct a $100(1 - \alpha)\%$ point-wise prediction interval for the new response $Y_0(t)$ as $\hat{Y}_0(t) \pm Z_{\alpha/2} \text{SE}\{\hat{Y}_0(t) - Y_0(t)\}$, where $Z_{\alpha/2}$ is the critical value corresponding to $\alpha/2$ upper quantile of the standard normal distribution and $\text{SE}\{\hat{Y}_0(t) - Y_0(t)\} = \left[\text{Var}\{\hat{Y}_0(t) - Y_0(t)\}\right]^{1/2}$ is obtained by bootstrapping the subjects as detailed by Algorithm 1.

**Algorithm 1 Bootstrap of subjects**

1. **for** $b = 1$ **to** $B$ **do**
2. Bootstrap the subjects with replacement. Let $\{b_1, \ldots, b_n\}$ be the subject index of the bootstrap sample.
3. Define the covariate and the response curves in the $b$th bootstrap sample as $\{X^{(b)}_i(\cdot) = X_k(\cdot)\}_{i=1}^n$ and $\{Y^{(b)}_i(\cdot) = Y_k(\cdot)\}_{i=1}^n$, respectively. The bootstrap data for the $b$th subject is obtained by collecting the trajectories $\{X^{(b)}_i(s_k, s_k)\}_{k=1}^{m_y}$ and $\{Y^{(b)}_i(t_j, t_j)\}_{j=1}^{m_y}$.
4. Apply FPCA to $\{Y^{(b)}_i(\cdot)\}_{i=1}^n$ and obtain estimate of the eigenbasis $\{\phi^{(b)}_k(\cdot)\}_{k=1}^{K^{(b)}}$, where $K^{(b)}$ is the finite truncation estimated using PVE and the pre-specified threshold level.
5. For $l = 1, \ldots, K_x$, $l' = 1, \ldots, K_s$, and $k = 1, \ldots, K^{(b)}$, obtain parameter estimates $\hat{\theta}^{(b)}_{l,l',k}$ by fitting the GAFPR model based on $\{X^{(b)}_i(s_k, s_k)\}_{k=1}^{m_y}$ and $\{Y^{(b)}_i(t_j, t_j)\}_{j=1}^{m_y}$.
6. For a new covariate $X_0(s)$, obtain the predicted response by $\hat{Y}^{(b)}_0(t) = \sum_{k=1}^{K^{(b)}} \phi^{(b)}_k(t) \sum_{l=1}^{K_x} \sum_{l'=1}^{K_s} \hat{\theta}^{(b)}_{l,l',k} \int T X l B X l B B S l S (s) ds$
7. Compute $V^{(b)}(t) = \text{Var}\{\hat{Y}^{(b)}_0(t)|\hat{\nu}_b\}$ using the model-based formula in (3.9).
8. **end for**
9. Approximate the marginal variance of predicted response by

$$\text{Var}\{\hat{Y}_0(t)\} \approx \frac{1}{B} \sum_{b=1}^B V^{(b)}(t) + \frac{1}{B} \sum_{b=1}^B \left(\hat{Y}^{(b)}_0(t) - \bar{Y}_0(t)\right)^2,$$

where $\overline{Y}_0(t)$ is the sample mean of $\hat{Y}^{(b)}_0(t)$'s.
Our inferential procedure has two advantages. First, the procedure accommodates complex correlation structures within the subject. Second, the iterated expectation and variance formula combines the model-based prediction variance and the variance of $\tilde{\eta}$, and better captures the total variance of the predicted response curves; our numerical study confirms the standard error characteristics in finite samples. One possible alternative to estimate the error covariance is to use $B^{-1}\sum_{b=1}^{B}\tilde{R}^{b}(t,t')$ where $\tilde{R}^{b}(t,t')$ are estimated based on each of the bootstrap samples, and our numerical study is based on this approach. Our numerical experience is that using the latter estimate of the covariance yields similar results as using the estimated model covariance $\tilde{R}(t,t')$ derived in Section 3.3.1.

3.4 Implementation and Extensions

Implementation of our method requires transformation of covariate as a preliminary step since the realizations of the covariate functions $\{X_{i}(s_{k}) : i, k\}$ may not be dense over the entire domain of the B-spline basis functions for $x$. In this situation, some of the B-spline basis functions may not have observed data on its support. This problem is already addressed by McLean et al. (2014) and Kim et al. (2016) with different strategies. This chapter uses point-wise center/scaling transformation of the functional covariate proposed by Kim et al. (2016). Full details are presented in Section 2.4 as well.

Our methodology can be adjusted easily to accommodate other realistic settings as we describe next. Specifically, we consider the cases where the functional covariate is observed on (i) a fine and regular grid of points with error, or on (ii) a sparse and irregular gird of points with or without noise, as well as where (iii) the functional response is observed on a sparse grid of points. The method described in this section includes a review of 2.5.1.

(i) We first consider the case of functional covariate observed on a fine and regular grids of points but with error; i.e., $W_{ik} = X_{i}(s_{k}) + \delta_{ik}$, and the random deviation $\delta_{ik}$ has variance $\tau^{2} > 0$. In FDA, various smoothing techniques have been applied to remove the observational errors $\delta_{ik}$; see for example Ramsay and Silverman (2005) and the references therein. Zhang and Chen
(2007) proposed to approximate the true latent process $X_i(\cdot)$ by smoothing each noisy trajectory using local polynomial kernel smoothing, and proved that the estimated curves, say $\hat{X}_i(\cdot)$, are asymptotically identical to the true latent process. The mean and the covariance function of the covariate $X_i(s_k)$ are then estimated by their sample estimators. (ii) If the covariates are observed on a sparse and irregular grid of points with errors, i.e., $W_{ik} = X_i(s_{ik}) + \delta_{ik}$ and $\tau^2 > 0$, we estimate the underlying smooth curves, say $\hat{X}_i(\cdot)$, by employing FPCA techniques for sparsely sampled functional data (Yao et al., 2005a). Their method uses local linear smoothers to estimate the mean and covariance functions of the covariate $X_i(s_k)$, and then estimate the pairs of eigenvalues/eigenfunctions by representing the random curves in a mixed model framework; they predict FPC scores using conditional expectation, and predict the latent trajectories using a finite KL expansion. If the new covariate $X_0(s)$ is observed on a sparse design, we first estimate eigenfunctions from the original observed data assuming that $X_0(s)$ follows the same process as $X_i(s)$, and then predict the FPC scores for the new covariates via conditional expectation formula in Yao et al., 2005a. (iii) Finally, when the sampling design of the response is sparse, we follow the same approach used for the sparsely observed covariates, and then estimate the smooth underlying curves for the response.

When $Y_i(\cdot)$ are observed on a sparse and irregular grid of points, $\{t_{i1}, \ldots, t_{im_{y,i}}\}$, numerical integration may not provide an accurate approximation of the $\xi_{ik}$’s. In this case, we use best linear unbiased predictors (BLUPs) proposed by Yao et al. (2005a). For completeness, we review some of the results presented in Yao et al. (2005a). Let $Y_i = [Y_i(t_{i1}), \ldots, Y_i(t_{im_{y,i}})]^T$, and let $\Sigma_{Y,i} = [\text{Cov}\{Y_i(t_{ij}), Y_i(t_{ij'})\}]_{1 \leq j,j' \leq m_{Y,i}}$ be the covariance matrix evaluated corresponding to the observed time points $t_{ij}$ and $t_{ij'}$. For $i = 1, \ldots, n$, define the $m_{Y,i}$-dimensional vector $\tilde{\phi}_{ik} = [\tilde{\phi}_k(t_{i1}), \ldots, \tilde{\phi}_k(t_{im_{y,i}})]^T$, where $\tilde{\phi}_k(\cdot)$ is the estimated eigenfunctions obtained from the spectral decomposition of the estimated covariance matrix of response. The BLUPs of $\xi_{ik}$ can be found as $\tilde{\xi}_{ik} = \sigma_k^2 \tilde{\phi}_{ik} \Sigma_{Y,i}^{-1} Y_i$, where $\sigma_k^2$ and $\Sigma_{Y,i}$ are the eigenvalues and the covariance matrix of marginal response estimated using the entire data set. Then, it follows that $\tilde{\nu}_{kk} = \text{Var}(\tilde{\xi}_{ik}) = \sigma_k^2 \tilde{\phi}_{ik}^T \Sigma_{Y,i}^{-1} \tilde{\phi}_{ik} \sigma_k^2$ and $\tilde{\nu}_{kk'} = \text{Cov}(\tilde{\xi}_{ik}, \tilde{\xi}_{ik'}) = \sigma_k^2 \tilde{\phi}_{ik}^T \Sigma_{Y,i}^{-1} \tilde{\phi}_{ik'} \sigma_k^2$ ($k \neq k'$). The inference procedure
described in Section 3.3.2 is now available with $\tilde{\nu}_{kk}$ and $\tilde{\nu}_{kk'}$.

### 3.5 Simulation Study

We investigate the finite sample performance of our method through simulations. The performance is assessed in terms of in-sample and out-of-sample predictive accuracy, average computation time, and coverage probabilities of prediction intervals (PIs).

#### 3.5.1 Details of Simulation Setup

In this section, we describe the design of the numerical study and the evaluation criteria used in Section 3.5.3.

**Simulation Design**

We generate $N = 1000$ samples from model (3.1) with the true functional covariate given by $X(s) = a_1 + a_2 \sqrt{2} \sin(\pi s) + a_3 \sqrt{2} \cos(\pi s)$ where $a_p \sim N(0, 2^{(1-p)^2})$ for $p = 1, 2, 3$, and each covariate is observed with noise. We construct noisy trajectories $W_{ik}$ by $W_{ik} = X_i(s_{ik}) + \text{WN}(0, \tau^2)$ with a noise variance equal to $\tau = 0.5$. For simplicity, denote $\phi_1(t) = 1$, $\phi_2(t) = \sqrt{2} \sin(2\pi t)$,
\[ \phi_3(t) = \sqrt{2} \cos(2\pi t), \phi_4(t) = \sqrt{2} \sin(4\pi t), \text{ and } \phi_5(t) = \sqrt{2} \cos(4\pi t). \] Then, the response \( Y_i(t) \) is generated based on model (3.1) for the all combinations of the following factors:

(1) True function \( F(x, s, t) \):

(i) linear case: \( F_1(x, s, t) = [\sqrt{2} \sin(2\pi t)s + \sqrt{2} \cos(2\pi t)2 \cos(\pi s)]x \)

(ii) simple non-linear case: \( F_2(x, s, t) = \phi_1(t)xs + \phi_2(t)(4 - 2(\frac{t}{5})^2 - 10s) \)

(iii) complex non-linear case: \( F_3(x, s, t) = \phi_1(t)xs + \phi_2(t)(4 - 2(\frac{t}{5})^2 - 10s) + \phi_3(t) \sin(2 - x - 2s) + \phi_4(t)2x \cos(\pi s) + \phi_5(t) \exp\{-\frac{(s-0.5)^2}{0.3}\} - 0.5 \)

To aid understanding the complex dependence structure between the response and the covariate, we present the true surface of \( F_3(x, s, t) \) along with \( x \) and \( s \) at fixed points \( t = 0.1, 0.5 \) and \( 1 \) in Figure 3.1.

(2) Error process \( E_i = [\epsilon_i(t_{i1}), \ldots, \epsilon_i(t_{imY,i})]^T \):

(i) \( E_i^1 \sim N(0, \sigma^2 I_{mY,i}) \)

(ii) \( E_i^2 \sim N(0, \sigma^2 \Sigma) + N(0, \sigma^2 I_{mY,i}) \) where \( \Sigma \) has \( AR_\rho(1) \) structure

(iii) \( E_i^3 \sim \omega_{i1} + \omega_{i2} \sqrt{2} \sin(2\pi t) + N(0, \sigma^2 I_{mY,i}) \)

(iv) \( E_i^4 \sim \omega_{i1} \sqrt{2} \cos(2\pi t) + \omega_{i2} \sqrt{2} \sin(4\pi t) + N(0, \sigma^2 I_{mY,i}) \)

We set \( \sigma = 0.3 \) and \( \rho = 0.1 \). The random variables \( \omega_{i1} \) and \( \omega_{i2} \) are uncorrelated and are generated from \( N(0, 0.15^2) \) and \( N(0, 0.1^2) \), respectively.

(3) Sampling design:

(i) dense design: \( m_Y = 101 \) and \( m_W = 81 \) equidistant time points in \([0,1]\)

(ii) sparse design:

\( m_{Y,i} \overset{iid}{\sim} Uniform(35, 44) \) (preserve 34.7\%−43.6\% of the data per curve)

\( m_{W,i} \overset{iid}{\sim} Uniform(45, 54) \) (preserve 55.6\%−66.7\% of the data per curve)
We do not consider the case where the sampling design of the covariate is sparse but the response curves are sampled densely, and vice versa.

(4) Number of subjects in the training data: (i) $n = 50$, (ii) $n = 100$, and (iii) $n = 300$

We generate the covariate $X_{0,i'}(s)$ ($i' = 1, 2, \ldots, n'$) of test data set assuming that the test data follows the same distribution as the training data. We assess the predictive performance using the sample size setting $n' = 50$, $m_Y = 101$, and $m_X = 81$. When the coverage performance is assessed, we increase the sample size in the test data to $n' = 100$. Throughout the study, it is assumed that the covariate in the test data is observed with noise.

**Evaluation Criteria**

For each scenario, we perform Monte-Carlo simulations with $N = 1000$ replications. The performance measure for the in-sample and out-of-sample accuracy is the RMSPE denoted by RMSPE$^{\text{in}}$ and RMSPE$^{\text{out}}$, respectively. We define the in-sample RMSPE by

$$\text{RMSPE}^{\text{in}} = \sum_{r=1}^{N} \sum_{i=1}^{n} \sum_{j=1}^{m_Y,i} (Y_i^{(r)}(t_{ij}) - \hat{Y}_i^{(r)}(t_{ij}))^2 \frac{1}{n \cdot N},$$

where $Y_i^{(r)}(t_{ij})$ and its estimate $\hat{Y}_i^{(r)}(t_{ij})$ are from the $r$-th Monte Carlo simulation. We define the out-of-sample RMSPE by

$$\text{RMSPE}^{\text{out}} = \sum_{r=1}^{N} \sum_{i'=1}^{50} \sum_{j=1}^{101} (\tilde{Y}_{0,i'}^{(r)}(t_j) - \hat{Y}_{0,i'}^{(r)}(t_j))^2 \frac{1}{(50 \cdot 101 \cdot N)},$$

where $\tilde{Y}_{0,i'}^{(r)}(t)$ = $\int_{T_X} F\{X_{0,i'}^{(r)}(s), s, t\} ds$ and $\hat{Y}_{0,i'}^{(r)}(t_j)$ are from the $r$-the Monte Carlo simulation. The RMSPE$^{\text{out}}$ will capture prediction errors, and one would expect that the values of RMSPE$^{\text{out}}$ converge to zero for large sample sizes.

Finally, we approximate $(1 - \alpha)$ level point-wise PIs to observe coverage probabilities at the nominal level. We define the average coverage probability (ACP) of the PIs at the $(1 - \alpha)$ level
by
\[ \text{ACP}_p(1 - \alpha) = \sum_{r=1}^N \left( \sum_{i'=1}^{100} \sum_{j'=1}^{101} I\{Y_{0,i'}(t_j) \in P_{1-\alpha,i'}^{(r)}(t_j)\} \right)/(100 \cdot 101 \cdot N), \]

where \( P_{1-\alpha,i'}^{(r)}(t_j) \) are the point-wise PIs from the \( r \)-th Monte Carlo simulation, and \( I(\cdot) \) is the indicator function. Note that in this calculation the PIs are constructed using the same fixed test data set over the Monte Carlo replications.

### 3.5.2 Competitive Methods

We compare our method to three other competitive approaches: the FLM, the functional additive model (FAM) of Müller and Yao (2008), which we label FAM, and the B-spline based estimation of Scheipl et al. (2015), which we label SFF. Specifically, we assess the prediction performance of the proposed approach and three competitive alternatives. We also compare these methods via their computational cost. Due to the high computational cost of the estimation procedure in FAM and SFF, we restrict our comparison with them to the case where \( n = 50 \) and the error process (\( \mathbb{E}_i \)) is \( \mathbb{E}_i^2 \), \( \mathbb{E}_i^4 \). For the FLM, we consider a model defined by
\[ E[Y_i(t)|X_i(s)] = \int_{T_X} X_i(s) \beta(s,t) ds. \]

When true model is non-linear in \( x \), we expect to see that our modeling yields better prediction results compared to the FLM. When the true model is linear in \( x \), we expect our model and the FLM show similar prediction performance.

Next, we summarize implementation details of our method and other three approaches.

We fit the GAFFR model using \( K_x = K_s = 7 \) cubic B-splines for \( x \) and \( s \). The number of eigenbasis \( K \) is selected by a preset level PVE=95%. The smoothing parameters \( \lambda_x \) and \( \lambda_s \) are selected by REML, and the integration is approximated by Simpson’s rule. Also, the FPC decompositions are implemented using \texttt{fpca.sc} function in the \texttt{refund} \( R \) package. We fit the FLM using \texttt{ff} function in the same package; the bi-variate coefficient function, \( \beta(s,t) \), is modeled using 7 cubic B-splines for each direction. For the Bspline-based estimation (Scheipl et al., 2015), we fit model (3.1) using \texttt{sff} function in the \texttt{refund} \( R \) package with 6 cubic B-splines for direction \( x \) and 5 cubic B-splines for directions \( s \) and \( t \). The FAMs (Müller and Yao, 2008) are implemented using \texttt{FPCfam} function in \texttt{PACE} package written in
Matlab. When estimating the additive model components, Gaussian kernel is used, and the bandwidth is selected via GCV.

3.5.3 Simulation Results

Performance of Prediction

Comparisons with the FLM are summarized in Table 3.1. We first consider the case where the true model is linear in $x$; see Table 3.1, top post. As expected, the two competitive models, the GAFFR and the FLM, provide relatively similar in-sample and out-of-sample prediction performance in all scenarios. The number of subjects, the sampling design of the grid points, and the error structure slightly affect the numerical results. The results confirm that the GAFFR model has similar prediction performance to the FLM when the true relationship is linear.

Next, we discuss the case when the true function is non-linear in $x$; see Table 3.1, bottom two
Table 3.2: Comparisons with FAM and SFF in terms of (1) RMSPE\textsuperscript{in} and (2) RMSPE\textsuperscript{out}, and (3) computation time (in seconds) averaged over 1000 simulations. Results correspond to the cases with \( n = 50 \).

<table>
<thead>
<tr>
<th>( F(x, s, t) ) method</th>
<th>( E_i = E_i^2 )</th>
<th>( E_i = E_i^4 )</th>
<th>( E_i = E_i^1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) (2) (3)</td>
<td>(1) (2) (3)</td>
<td>(1) (2) (3)</td>
</tr>
<tr>
<td>( F(x, s, t) ) FAM</td>
<td>0.446 0.174 94.0</td>
<td>0.373 0.180 92.9</td>
<td>0.453 0.209 687.9</td>
</tr>
<tr>
<td>( F_1(x, s, t) ) SFF</td>
<td>0.443 0.149 99.4</td>
<td>0.362 0.213 82.3</td>
<td>0.449 0.162 43.0</td>
</tr>
<tr>
<td>( F_1(x, s, t) ) GAFFR</td>
<td>0.444 0.144 10.2</td>
<td>0.371 0.150 8.7</td>
<td>0.452 0.167 10.2</td>
</tr>
<tr>
<td>( F_2(x, s, t) ) SFF</td>
<td>0.430 0.115 93.9</td>
<td>0.355 0.134 93.4</td>
<td>0.435 0.166 564.2</td>
</tr>
<tr>
<td>( F_2(x, s, t) ) GAFFR</td>
<td>0.425 0.061 143.5</td>
<td>0.342 0.116 116.4</td>
<td>0.424 0.076 39.0</td>
</tr>
<tr>
<td>( F_3(x, s, t) ) SFF</td>
<td>0.470 0.281 94.4</td>
<td>0.410 0.289 92.7</td>
<td>0.488 0.324 687.2</td>
</tr>
<tr>
<td>( F_3(x, s, t) ) GAFFR</td>
<td>0.447 0.207 130.3</td>
<td>0.368 0.272 124.1</td>
<td>0.452 0.226 50.6</td>
</tr>
</tbody>
</table>

It is evident from the results that our method provides smaller in-sample and out-of-sample prediction errors than the FLM in all scenarios. As the true model gets complicated (e.g., \( F(x, s, t) = F_3(x, s, t) \)), such patterns become noticeable. Thus, our model better captures complex non-linear relationships.

Next, we compare our method to the B-spline based estimation. The results are presented in Table 3.2. Comparing the columns labeled (1) and (2) in the two panels, we observe that the two estimation approaches result in similar accuracy, with precision accuracy that varies slightly with the complexity of the relationship. Column labeled (3) shows the average computation time (in seconds), indicating an order of magnitude gain with our estimation. The measurements are taken on a 2.3GHz AMD Opteron Processor. In terms of computational aspects, using the B-spline based estimation is computationally more intensive than our approach; for example, when the sampling design is dense, using the B-spline based method is more than ten times expensive than fitting the GAFFR model.

Table 3.2 also summarizes results for comparisons with our approach and FAM. As the model complexity increases and the grid points are sparsely sampled, results based on FAM are more different from the results from the GAFFR model, providing larger values of RMSPE\textsuperscript{in} and RMSPE\textsuperscript{out}. Comparing the average computation times measured on a 2.3GHz AMD Opteron Processor, FAM takes more computation times than our approach especially when the gird points
Table 3.3: Summary of average coverage probability (ACP) for the new response $Y_{0,i}(t)|X_{0,i}(\cdot)$ at nominal significance levels $1-\alpha=0.85, 0.90, \text{and} 0.95$. The results are based on 1000 simulated data sets with 100 bootstrap replications per data.

<table>
<thead>
<tr>
<th>$F(x,s,t) = F_2(x,s,t)$, dense design</th>
<th>$F(x,s,t) = F_2(x,s,t)$, sparse design</th>
<th>$F(x,s,t) = F_3(x,s,t)$, dense design</th>
<th>$F(x,s,t) = F_3(x,s,t)$, sparse design</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n$</td>
<td>0.85</td>
<td>0.90</td>
<td>0.95</td>
</tr>
<tr>
<td>50</td>
<td>0.904</td>
<td>0.942</td>
<td>0.976</td>
</tr>
<tr>
<td>100</td>
<td>0.884</td>
<td>0.928</td>
<td>0.967</td>
</tr>
<tr>
<td>300</td>
<td>0.868</td>
<td>0.915</td>
<td>0.960</td>
</tr>
</tbody>
</table>

In summary, our numerical investigation confirms that our model better captures the complex non-linear relationships than the FLM when the true model is non-linear; when the true relationship is simply linear, our method still retains good prediction performance. The B-spline based estimation and our method have similar prediction performance, while our method is computationally more efficient. Compared with the FAM, our method better captures the true relationship in a more efficient manner.

Performance of Inference

Next, we assess coverage performance of the point-wise PIs. The point-wise PIs are approximated using the method described in Section 3.3 with 100 bootstrap samples per simulated
data. Table 3.3 reports the ACP for both the dense and sparse design, and nominal levels are 85%, 90%, and 95%. When the sample size is small (e.g., \( n = 50 \)), the results are in fact conservative, providing overestimated coverage probabilities. However, the coverage probabilities get closer to these respective nominal levels as the sample size increases. The complexity of true function \( F(x, s, t) \) slightly affects the coverage performance in all scenarios. If the true function is complex, \( F(x, s, t) = F_3(x, s, t) \), the coverage probability converges slower to the nominal levels compared to the case when the true function is simple, \( F(x, s, t) = F_2(x, s, t) \). The number of subjects, the sampling design of the grid points, and the error covariance structure slightly affect the coverage performance as well.

Remark. Appendix B.1.1 includes additional simulation results corresponding to another level of sparseness, and the results indicate that our approach still maintains prediction accuracy. Appendix B.1.2 shows how sensitive our method is depending on the choice of \( K \). Specifically, we assessed the performance of prediction accuracy and coverage estimation for the two choices of \( K \) (one is selected by 95% of variance explained, and the other one is by 99% of variance explained), and we could derive a conclusion that our method is not very sensitive to the choice of \( K \).

### 3.6 Applications

#### 3.6.1 Capital Bike Share Data

We now turn to the capital bike share study (Fanaee-T and Gama, 2013). The data is collected from Capital Bike Share system at Washington, D.C., which offers bike rental services on an hourly basis. In recent years, there have been increased demands for renting bicycles, as renting is a great alternative to purchasing own bicycles. In order to better understand how many bikes are needed on site, we wish to examine the patterns of bike rentals in conjunction with weather information. The data set has information on counts of casual bike rentals during the period from January 1, 2011 to December 31, 2012 - a total of 105 weeks. In addition to the
Figure 3.2: The number of casual bike users (left panel) and hourly temperatures (right panel) collected every Saturday. The measurements taken in three different days on January, April, and July in 2011 are indicated by solid, dashed, and dotted lines, respectively.

count of rentals, the data also has the date on which the rental services are made and the weather information such as temperature (°C) and humidity. We analyze the data focusing on the measurements taken every Saturday because the bike rentals on Saturday are particularly in high demand. The data are recorded on an hourly basis. Figure 3.2 presents counts of casual bike rentals (right panel) and hourly temperature (left panel) on Saturday, and each curve corresponds to a particular week. Also, the measurements taken in three different days are indicated by solid, dotted, and dashed lines, respectively. A simple visual inspection indicates that the higher temperature does not imply the increased demand for renting bikes.

**Predicting the Number of Casual Bike Users Based on Temperature**

To assess how the temperature and the number of casual bike users are related to each other, we fit the GAFFR model. Let $Y_{ij} = \tilde{Y}_i(t_{ij}) = \log(1 + \text{count}_{ij})$ be the log-transformed count of bike rentals, where count$_{ij}$ are the number of casual bikes rented at the time point $t_{ij}$ in the $i$th week, and $\{t_{ij}\}_{j=1}^{24}$ are the equidistant points in [0,1]. Let $W_{ik} = X_i(s_{ik}) + \delta_{ik}$ be the temperature measured at the time point $s_{ik}$ in the $i$th week, where $\{s_{ik}\}_{k=1}^{24}$ are the equidistant points
Table 3.4: Summary of (1) RMSPE\textsuperscript{in} and (2) RMSPE\textsuperscript{out} in bike share data analysis.

<table>
<thead>
<tr>
<th>Method</th>
<th>log-transformed data</th>
<th>Original data</th>
<th>log-transformed data</th>
<th>Original data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>FLM</td>
<td>0.755</td>
<td>0.626</td>
<td>57.835</td>
<td>41.940</td>
</tr>
<tr>
<td>GAFFR</td>
<td>0.669</td>
<td>0.540</td>
<td>39.716</td>
<td>31.764</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>log-transformed data</th>
<th>Original data</th>
<th>log-transformed data</th>
<th>Original data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>FLM</td>
<td>0.741</td>
<td>0.606</td>
<td>61.674</td>
<td>43.265</td>
</tr>
<tr>
<td>GAFFR</td>
<td>0.638</td>
<td>0.492</td>
<td>37.917</td>
<td>30.060</td>
</tr>
</tbody>
</table>

in [0,1], and $\delta_{ik}$ is the white noise random deviation. We consider two models, the proposed GAFFR model and the existing FLM. The temperature and the counts of bike rentals have lower level of missingness. Therefore, we follow our data processing steps by smoothing the covariate using FPCA method and then by applying the center/scaling transformation. We assess both in-sample and out-of-sample prediction accuracy by splitting the data into training and test sets of size 89 and 16. To fit the GAFFR model, we use $K_x = K_s = 7$ cubic B-splines for $x$ and $s$, and the number of eigenbasis $K$ is selected by 95% of variance explained by the first few eigencomponents. In the FLM, the bi-variate coefficient function, $\beta(s,t)$, is modeled using 7 cubic B-splines for each direction. The results are presented in Table 3.4 (see top left panel labeled “log-transformed data”). It seems that the prediction errors obtained from the FLM are inflated compared to the ones from our method. In Table 3.4 (see top right panel labeled ‘original data”), we also report prediction errors on the original scale, i.e., $\exp(\hat{Y}_{ij}) - 1$. It seems that our model still provides smaller in-sample and out-of-sample prediction errors compared to the FLM. The overall results clearly support that the relationship between the temperature and the counts of rental bikes is not simply linear.

Next, we explore coverage performance by constructing the point-wise PIs using 1000 bootstrap replications. For the nominal levels of 0.85, 0.90, and 0.95, the ACPs are approximated as 0.841 (0.034), 0.896 (0.029), and 0.953 (0.024), respectively, with the estimated standard errors in parenthesis. Thus, the results are fairly close to the nominal levels.
Accounting for Average Humidity

In this section we illustrate how we incorporate additional covariate/s in our method. Let $h_{um_i}$ denote the average humidity corresponding to week $i$ in the study. We assume that humidity is a time-varying effect and consider two alternative models, $E[Y_i(t)|X_i(s), h_{um_i}] = \int_{T_X} F\{X_i(s), s, t\} ds + h_{um_i} \gamma(t)$ for the GAFFR model and $E[Y_i(t)|X_i(s), h_{um_i}] = \int_{T_X} X_i(s) \beta(s, t) ds + h_{um_i} \gamma(t)$ for the FLM, where $\gamma(t)$ is the smooth coefficient function that determines the effect of average humidity on the response. One remark is that when additional covariate is combined, standardizing all variables is important since variables measured at different scales may not contribute equally to the response. When fitting the GAFFR model, we expand $\gamma(t)$ by the same basis function $\{\phi_k(\cdot)\}_k$; i.e., $\gamma(t) = \sum_{k=1}^{K} \phi_k(t) \zeta_k$, where $\zeta_k$ are the unknown basis coefficients. Such representation allows us to maintain the size of curvature in $t$-direction in the fitted coefficient function, $\hat{\gamma}(t)$. We assessed both in-sample and out-of-sample prediction performance by fitting the two competitive models, and results are summarized in Table 3.4 (see bottom table). If we compare the GAFFR model with the FLM, the GAFFR model provides smaller prediction errors. Furthermore, if we compare the results from the top table with the one from the bottom table, adding the average humidity improves the prediction accuracy. To aid understanding of the effect of average humidity, we provide a histogram of standardized average humidity and the curve estimate of $\gamma(t)$ obtained by fitting the GAFFR model in Appendix B.2.1.

We carried out additional analysis to see formally whether $h_{um_i}$ has a significant effect on the response, i.e., $H_0: \gamma(t) = 0$. To test the null hypothesis, we use a $L^2$-norm based test statistic (see e.g., Park et al. 2015) defined as $T_n = ||\hat{\gamma}(t)||^2$ and approximate its null distribution using bootstrap as we describe in Algorithm 2. In the analysis, Algorithm 2 is applied with $B = 1000$ bootstrap replications, and the p-value of our test is obtained to be 0.001, indicating that the average humidity is strongly associated with the log count of bike rentals. To confirm this result, we carried out additional numerical studies based on the simulated data sets that mimic the feature of the bike share data. Specifically, the data sets are generated using
E[Y_i(t)|X_i(s), hum_i] = \int_{T_X} \hat{F}\{X_i(s), s, t\} ds + hum_i \hat{\gamma}(t), \text{ where } \hat{F}(\cdot, \cdot, \cdot) \text{ and } \hat{\gamma}(t) \text{ are obtained by fitting the GAFFR model using the original observed data. We assessed the size and power performance of the } L^2 \text{-norm based test, and the results are presented in Appendix B.2.2. We obtained the size of test close to the nominal levels of 5\%, 10\% and 15\%, and the empirical power increased at a fast rate.}

Algorithm 2 $L^2$-norm based test

1: for $b = 1$ to $B$ do
2: Estimate the coefficient function $\gamma(t)$ using the original observed data, and compute the observed test statistic, $T_{n,\text{obs}} = ||\hat{\gamma}(t)||^2$.
3: Obtain $Y^*_i(t) = Y_i(t) - hum_i \hat{\gamma}(t)$, where $\hat{\gamma}(t)$ is obtained from step 2.
4: Bootstrap the subjects with replacement. Let $\{b_1, \ldots, b_n\}$ be the subject index of the bootstrap sample.
5: Generate bootstrap data: $\{X_i^{(b)}(s_k), s_k\}_{k=1}^{n}$, $\{hum_i^{(b)}\}_{i=1}^{n}$, and $\{Y_i^{(b)}(t_j), t_j\}_{j=1}^{n}$, where $X_i^{(b)}(s) = X_{b_i}(s)$, $hum_i^{(b)} = hum_{b_i}$, and $Y_i^{(b)}(t) = Y^*_b(t)$.
6: Fit the GAFFR model using the bootstrap sample.
7: Obtain $\{\hat{\gamma}_b(t)\}_b$, and calculate the bootstrap test statistic by $\{T^*_b\}_b = ||\hat{\gamma}_b(t)||^2$.
8: end for
9: Compute the p-value by $\hat{p} = B^{-1}\sum_{b=1}^{B} I\{T^*_b \geq T_{n,\text{obs}}\}$.

3.6.2 Yield Curves Data

Next, we consider an application to yield curves. The yield at maturity $T$ years is the average interest rate that is earned on a bond maturing in $T$ years (Ruppert and Matteson, 2015). The yield curve for a given type of bond is the plot of yield against maturity, and the shape of a yield curve changes each day. Maturities range from short-term (e.g., 1 year, 2 years) to long-term (e.g., 20 years, 30 years). The curves can take three primary shapes: positive slope (so-called “normal”), negative slope (so-called “inverted”), and flat curve. The slope of the yield curve is a powerful indication of the future economic growth (Plosser and Rouwenhorst, 1994). If the slopes are positive, one can expect a strong future economic growth and higher interest rates. The negative slopes indicate a slow future economic growth and rise in unemployment.

The changes in US yield curves are often discussed in relation to the changes of yield
Figure 3.3: Changes in the US yield curves (left panel) and in the European yield curves (right panel) during the period from January 2, 2006 to December 30, 2011. The measurements taken in three different days are indicated by red, blue, and green lines.

curves in other countries (see e.g., Plosser and Rouwenhorst, 1994; Mehl, 2009). In particular, Plosser and Rouwenhorst (1994) studied linkages between the real economic growth in the United States and interest-rates across other countries including Germany and the UK. Motivated by those studies, we consider the problem of predicting the changes in the US daily yield curves using the changes in the European yield curves on the same day. Our data consist of US and European daily yields during the period from January 2, 2006 to December 30, 2011 - a total of 1286 days. To visualize the data, Figure 3.3 presents the changes in the US yield curves (left panel) and European yield curves (right panel) over maturity (years), and each curve corresponds to a particular day. Simple visual inspection reveals that the changing patterns between the US and the European yield curves are very similar.

Let $Y_{ij} = Y_i(t_j)$ be the changes in the US yield curves corresponding to $i$th day ($i = 1, \ldots, 1286$) measured at maturity $t_j$ ($j = 1, \ldots, 30$), where the maturities $t_j$ range from 1 year to 30 years. Let $W_{ik} = X_i(s_k) + \delta_{ik}$ be the changes in the European yield curves corresponding to $i$th day measured at maturity $s_k$ ($k = 1, \ldots, 30$), and $\delta_{ik}$ is the white noise random deviation.

We consider two modeling approaches, the GAFFR model and the FLM, and assess both the in-sample and out-of-sample prediction accuracy based on RMSPE. For the GAFFR model,
the functional covariate is processed using the methods described in Section 3.4. The first 1000 days are taken as a training data set, and the last 286 days are taken as the test data. Thus, our models are estimated based on the past changes in the yield curves, while the model assessment is for future changes. To fit the GAFFR model, we used $K_x = K_s = 7$ cubic B-splines for $x$ and $s$, and the number of eigenbasis $K$ is selected by 95% of variance explained. In the FLM, the bi-variate coefficient function, $\beta(s, t)$, is modeled using 7 cubic B-splines for each direction. Both GAFFR model and the FLM provided the similar values for the in-sample and out-of-sample predication errors ($\text{RMSPE}^{\text{in}}=0.059$ and $\text{RMSPE}^{\text{out}}=0.064$), indicating that a simple linear association is more appropriate. The coverage performance is also assessed by constructing the point-wise PIs using 1000 bootstrap replications. For the nominal levels of 0.85, 0.90, and 0.95, the ACPs are approximated as 0.847, 0.885, 0.920, respectively.
Chapter 4

Testing for Linearity in General
Additive Function-on-Function
Regression

4.1 Introduction

FLMs are very popular tools for describing the relationship between functional response and functional predictors and have been studied extensively in the literature of FDA (see e.g., Ramsay and Silverman, 2005; Yao et al., 2005b; Wu et al., 2010). In FLMs, the conditional mean of the current response given the full true covariate profile is modeled as the inner product between a bi-variate function and the covariate profile. In this chapter, we consider the problem of testing whether the relationship between a functional response and predictor is linear. Specifically, if \( Y(\cdot) \) denotes the generic response profile and \( X(\cdot) \) denotes the generic true predictor profile, we are interested to test the null hypothesis

\[
H_0 : E[Y(\cdot)|X(s)] = \int_{\mathcal{T}_X} \beta(s, \cdot)X(s)ds
\]  

(4.1)
for some smooth bi-variate function \( \beta(\cdot, \cdot) \) versus the alternative that the model is not linear. We restrict the analysis to the class of models of the form considered in the previous chapter, where the conditional mean of the response is described through an unknown smooth tri-variate function. While the proposed general additive class of models has superior performance when the true model is indeed non-linear and has similar performance when the true model is linear, it involves more complex objects - a tri-variate smooth function.

To formulate the problem, denote the data for the \( i \)th subject by \( \{(W_{ik}, s_{ik}) : k = 1, \ldots, m_{W,i}\} \) and \( \{(Y_{ij}, t_{ij}) : j = 1, \ldots, m_{Y,i}\} \) for \( i = 1, \ldots, n \), where \( W_{ik} \)'s and \( Y_{ik} \)'s are the covariate and response profile observed at time points \( s_{ik} \in \mathcal{T}_X \) and \( t_{ij} \in \mathcal{T}_Y \), respectively, for some compact intervals \( \mathcal{T}_X \) and \( \mathcal{T}_Y \); for simplicity we take \( \mathcal{T}_X = \mathcal{T}_Y = [0, 1] \). It is assumed that \( W_{ik} = X_i(s_{ik}) + \delta_{ik} \), where \( X_i(\cdot) \) is a continuous, square-integrable random function over \( \mathcal{T}_X \), and \( \delta_{ik} \) are iid measurement errors with mean zero and variance equal to \( \tau^2 \). It is also assumed that \( Y_{ij} = Y_i(t_{ij}) \), where \( Y_i(\cdot) \) is a squared-integrable process defined over \( \mathcal{T}_Y \). Without loss of generality, we assume that the response \( Y_i(\cdot) \) has mean function equal to zero. For convenience, we describe the testing procedures for the case where (i) the response \( Y_i(\cdot) \) and the covariate \( X_i(s) \) are observed on a fine and regular grid of points such that \( t_{ij} = t_j \) and \( s_{ik} = s_k \) for all \( i \), \( j \), and \( k \), and \( m_{W,i} = m_W \) and \( m_{Y,i} = m_Y \) for all \( i \); and (ii) the covariate is observed without noise (\( \tau^2 = 0 \)). However, our method can be applied to more realistic situations as we discuss in Section 4.3. We assume that \( \mathbb{E}[Y_i(t)|X_i(s)] = \int_{\mathcal{T}_X} F\{X_i(s), s, t\}ds \), where \( F(\cdot, \cdot, \cdot) \) is an unknown smooth tri-variate function defined over \( \mathbb{R} \times \mathcal{T}_X \times \mathcal{T}_Y \) and accommodates a flexible non-linear dependence between the response \( Y_i(t) \) and the full covariate trajectory \( X_i(s) \). The standard FLM is a special case of this type of models where \( F(x, s, t) = \beta(s, t)x \), and the resulting model can be written as \( \mathbb{E}[Y_i(t)|X_i(s)] = \int_{\mathcal{T}_X} \beta(s, t)X_i(s)ds \).

One convenient way for testing the null hypothesis given in (4.1) is to use a F-ratio type test statistic (Shen and Faraway, 2004; Xu et al., 2011; Wang and Chen, 2012), which compares the residual sum of squares (RSS) under the null and the alternative hypothesis. In particular, Wang and Chen (2012) discussed F-tests for testing an unspecified function through mixed
model representations, but they did not scale the RSSs in test statistics using appropriate degrees of freedom; this type of test statistics can be defined as $T = (RSS_0 - RSS_1)/RSS_1$, where $RSS_0 = \sum_i ||Y_i(\cdot) - \hat{Y}_i^\text{null}(\cdot)||^2$ and $RSS_1 = \sum_i ||Y_i(\cdot) - \hat{Y}_i^\text{full}(\cdot)||^2$ are the RSSs under the null and the alternative hypothesis, respectively, with the estimated response curves $\hat{Y}_i^\text{null}(\cdot)$ and $\hat{Y}_i^\text{full}(\cdot)$. Although these test statistics are not sensitive to the choice of the degrees of freedom, deriving their null distributions is still not straightforward in the presence of variance components. They proposed to use spectral decomposition of the test statistic to derive the null distribution. Another plausible way might be using bootstrap approximation (see e.g., García-Portugués, 2014). However, approximating the null distribution of test statistics is rather computationally intensive because of the multiple fitting of the model. As an alternative, we develop a restricted likelihood ratio based testing procedure (McLean et al., 2015), and approximates its null distribution using simulated samples; such methods are called pseudo-restricted likelihood ratio tests and developed by Greven et al. (2008). The proposed testing procedure is computationally efficient and accommodates different sampling designs for the covariate and response as well as non-trivial correlation structures in the error process. We show through numerical studies that it has good size and power performance.

Testing in FDA has attracted considerable interest recently. Recent developments include Cardot et al. (2003); Swihart et al. (2014); Kong, Staicu and Maity (2015) which discussed testing for no association between the scalar response and functional covariate. In particular, Swihart et al. (2014) considered the restricted likelihood ratio test (RLRT) to assess the global effect of the functional covariate using a mixed effects framework for the FLM. McLean et al. (2015) developed testing for linearity in the class of functional generalized additive models by representing the mean structure of the response using a linear mixed effects framework and employing the RLRTs for zero variance components. RLRTs with a single variance component have non-trivial null distribution when iid assumption does not hold for the response because under the null hypothesis the parameter is on the boundary of the parameter space. For models with a single variance component, Crainiceanu and Ruppert (2004) derived finite sample dis-
tributions of the RLRT; Greven et al. (2008) extended these results for more general cases such as models with multiple variance components. Both methods are computationally inexpensive than the bootstrap-based algorithm and can be easily implemented using the R package RLRsim (Scheipl et al., 2008).

The rest of the chapter is organized as follows. Section 4.2 describes the method for testing linearity in function-on-function regression setting. Section 4.3 presents extensions to various realistic scenarios. Section 4.4 investigates the size and power performance of the test through simulations. 4.5 illustrates applications of our method to capital bike share study and US yield curves study.

4.2 Restricted Likelihood Ratio Tests in Function-on-Function Regression

As we proposed in Chapter 3, we model the tri-variate function $F(\cdot, \cdot, \cdot)$ in the alternative hypothesis using a mixture of pre-determined basis functions and unknown data-driven basis. Let $\phi_k(\cdot)$ be an orthogonal basis in $L^2[0,1]$, and let $\{B_{X,l}(x)\}_{l=1}^{K_x}$ and $\{B_{S,l'}(s)\}_{l'=1}^{K_{s'}}$ be the orthogonal B-splines for $x$ and $s$, respectively. We approximate the tri-variate function $F(x, s, t)$ by $F(x, s, t) \approx \sum_{k=1}^{K} \sum_{l=1}^{K_x} B_{X,l}(x) B_{S,l'}(s) \phi_k(t) \theta_{l,l',k}$, where $K$ is the number of orthogonal basis $\{\phi_k(\cdot)\}_{k \geq 1}$. In fact, as discussed in Chapter 3, we take $\phi_k(\cdot)$’s to be the eigenbasis of $Y_i(\cdot)$’s.

A consequence of the orthogonality of the functions $\phi_k(\cdot)$’s is that if $\xi_{ik} = \int_0^1 Y_i(t) \phi_k(t) dt$ where $Y_i(t)$ is the de-meaned full response curve, then it follows that $E[\xi_{ik} | X_i(\cdot)] = \int_0^1 G_k \{X_i(s), s\} ds$ for all $k$.

4.2.1 Testing Procedure

Intuitively, our testing procedure is based on the following result.

**Proposition** 1. Conditions (A1)-(A5) are equivalent.

(A1) $F(x, s, t) = \beta(s, t)x$ for bi-variate function $\beta(\cdot, \cdot)$. 

\( (A2) \frac{\partial^2 F(x,s,t)}{\partial x^2} = 0 \text{ for } \forall x. \)

\( (A3) \sum_{k \geq 1} \left| \frac{\partial^2 G_k(x,s)}{\partial x^2} \right|^2 = 0 \text{ for } \forall x. \)

\( (A4) \frac{\partial^2 G_k(x,s)}{\partial x^2} = 0 \text{ for all } k \geq 1 \text{ and for } \forall x. \)

\( (A5) G_k(x,s) \) is linear in \( x \) for all \( k \geq 1 \) and for \( \forall x. \)

The proof of Proposition 1 is given in Appendix C.1. Our modeling is based on the truncated basis expansions, where the leading \( K \) eigenbasis provides a parsimonious representation of \( F(x,s,t) \). Based on the truncated model representation and the result from Proposition 1, the null hypothesis given in (4.1) can be reduced to testing simultaneously the following \( K \) hypotheses:

\[
\begin{align*}
H_{0,1} : E[\xi_1 | X_i(\cdot)] &= \int_0^1 \beta_1(s)X_i(s)ds, \\
&\vdots \\
H_{0,K} : E[\xi_K | X_i(\cdot)] &= \int_0^1 \beta_K(s)X_i(s)ds,
\end{align*}
\]

(4.2)

for unknown smooth functions \( \beta_k(s) \) (\( k = 1, \ldots, K \)). Here, the truncation \( K \) can be determined by a preset PVE (Di et al., 2009; Staicu et al., 2010). We propose to test each null hypothesis separately using a level of significance adjusted by the Bonferroni criterion. McLean et al. (2015) studied testing for linearity in scalar-on-function regression in the case when \( \xi_{ik} \)'s are known, and \( X_i(\cdot) \) are fully observed without noise. We adopt their technique and use it for estimated \( \xi_{ik} \)'s. For completeness, we describe their technique in Section 4.2.2.

To test the null hypothesis in (4.2), we first obtain an underlying smooth trajectory of response \( \tilde{Y}_i(\cdot) \) by employing a smoothing technique such as local polynomial kernel smoothing (Zhang and Chen, 2007) or kernel-based local linear smoothing (Yao et al., 2005a), and de-mean the smooth trajectory, \( \tilde{Y}_i^c(t) = \tilde{Y}_i(t) - \tilde{\mu}_Y(t) \) where \( \tilde{\mu}_Y(t) \) is the estimated marginal mean of the response. Next, we estimate the eigenbasis \( \{\phi_k(t)\}_k \) based on the spectral decomposition of the estimated covariance function of the response \( Y_i(\cdot) \)'s, and denote by \( \{\tilde{\phi}_k(t)\}_k \) the estimated eigenbasis. Then, we follow the steps provided in Algorithm 3.
Algorithm 3 RLRT algorithm in function-on-function regression

1: for $k = 1$ to $K$ do
2: Estimate the loadings $\xi_{ik}$ by $\tilde{\xi}_{ik} = \int_0^1 \tilde{Y}_i(t)\tilde{\phi}_k(t)dt$.
3: Test linearity in scalar-on-function regression setting, $E[\xi_{ik}|X_i(\cdot)] = \int_0^1 G_k\{X_i(s), s\}ds$, using $\tilde{\xi}_{ik}$ in place of $\xi_{ik}$.
4: Calculate p-value, $p_k$.
5: end for
6: Reject $H_0$ given in (4.1) if $p_k < \frac{\alpha}{K}$ for some $k$ and for a nominal level $\alpha$.

In this algorithm, the finite truncation $K$ is chosen by the pre-specified PVE such as 95% or 99%. Step 3 is based on the pseudo RLRT using a mixed model representation of the scalar-on-function regression model (McLean et al., 2015).

Our algorithm has three advantages. First, it significantly improves the computational efficiency over bootstrap-based tests, and computation time increases minimally with the number of eigenbasis $K$; for example, in the case $K = 2$, it took less than 3 seconds on a 2.3GHz AMD Opteron Processor to follow the entire steps when data settings are $n = 300$, $m_W = 51$ and $m_Y = 81$. When $K = 4$, it took less than 5 seconds to proceed the algorithm with the same data settings. Second, our algorithm accommodates correlated error process; if we write $Y_i(t) = \sum_k \xi_{ik}\phi_k(t) + \epsilon_i(t)$, the orthogonal projection of response $Y_i(t)$ yields $e_{ik} = \int_0^1 \epsilon_i(t)\phi_k(t)dt$, where $e_{ik}$ are iid zero-mean measurement errors, and thereby the error correlation structure does not affect the testing procedure in step 3. Finally, the proposed test can be applied to more general situations such as curves with missing observations and measurement errors. However, it is worth noting that when the proportion of missingness is relatively large (e.g., more than 50% of the observations are missing for each subject), our test maintains the correct size only for sufficiently large sample sizes such as $n = 500$. Simulation results corresponding to this case are reported in Appendix C.2.
4.2.2 Review of Restricted Likelihood Ratio Tests in Scalar-on-Function Regression

We briefly review the method for testing linearity in scalar-on-function regression proposed by McLean et al. (2015); recall that in our testing procedure $\xi_{ik}$’s are unknown, true FPC scores, whereas $\xi_{ik}$’s are known scalar responses in their problem setting. The goal is to test whether the model $E[\xi_{ik}|X_i(\cdot)] = \int_{0}^{1} \beta_k(s)X_i(s)ds$ is true versus a general alternative model $E[\xi_{ik}|X_i(\cdot)] = \int_{0}^{1} G_k\{X_i(s), s\}ds$, assuming that $\xi_{ik}$ has mean zero. They test the null hypothesis through a mixed model representation by re-parameterizing the bi-variate function $G(\cdot, \cdot)$ as the sum of fixed effects and iid Gaussian random effects. To this end, they first represent the unknown smooth function $G(\cdot, \cdot)$ using tensor product of B-spline basis functions, and define a penalized log-likelihood criterion to estimate the basis coefficients. Then, the fixed and random components are formulated using eigendecomposition of two marginal penalties, both of which are used to control smoothness of the function in $x$ and $s$ directions. Specifically, they construct design matrices of the fixed and random effects by projecting the full model onto the null and range space of the marginal penalties, and similar ideas are also discussed in Wood (2006b), Lee and Durbán (2011) and Wood et al. (2013) among others. Finally, testing the null hypothesis that model is linear in $x$ reduces to testing the variance component in linear mixed models with a nuisance variance component under the null hypothesis, and one can apply the pseudo RLRTs proposed by Greven et al. (2008). With the iid assumption in the random effects, the mixed model components can be estimated using the standard mixed modeling software such as lme4 (Bates et al., 2015) in R package or PROC MIXED procedure in SAS (SAS Institute Inc, 2008). The RLRTs can be implemented using RLRsim (Scheipl et al., 2008).
4.3 Data Preprocessing

In this section, we describe several steps need to be taken to follow our testing procedure, as a part of data preprocessing.

4.3.1 Transformation of Covariate

Algorithm 3 requires the preprocessing of the functional covariate before fitting the models under the null and the alternative hypotheses. One challenge of using the B-spline basis functions for the representation of the models is that some B-splines for $x$ might not have observed data on its support. These problems are more likely to arise when the realizations of the functional covariate $X(s_k)$ are not dense over $\mathbb{R}$. Our numerical investigations are based on point-wise center/scaling transformation as described in Section 2.4, which transforms the covariate $X_i(s_k)$ at each point $s_k$ by subtracting the mean function and dividing by standard deviation function at the same point $s_k$. When the null hypothesis is true, the model with the transformed covariate preserves the linearity of the true function, and thus testing of linearity is still valid.

4.3.2 Data Pre-processing for Sparse Sampling Design

The testing methodology described in Section 4.2 is applicable to more general situations. If the covariate are measured at a fine and regular grid of points but with errors, i.e., $W_{ik} = X_i(s_k) + \delta_{ik}$ with $\tau^2 > 0$, we apply the approach of Zhang and Chen (2007) to recover the true latent process $X_i(\cdot)$ at the points $s_k$; this type of data collection is often called dense sampling design. The estimated curves, $\hat{X}_i(t)$, are asymptotically identical to the true latent process as discussed in Zhang and Chen (2007). Our procedure is available with the estimated smooth trajectory $\hat{X}_i(s_k)$ in place of the true one. In the case of functional covariate observed sparsely with noise, i.e., $W_{ik} = X_i(s_{ik}) + \delta_{ik}$ with $\tau^2 > 0$, we treat the data differently; the sparse sampling design refers to the case where the number of observations vary across subjects and the number of observations per subject is small. In this case, we employ the method of Yao et al. (2005a) to reconstruct the full underlying trajectories as follows: 1) estimate the mean and covariance
functions of the covariate $X_i(\cdot)$ based on local linear smoothing, 2) estimate the pairs of eigenvalues/eigenfunctions through a spectral decomposition of the estimated covariance function, and 3) predict the FPC scores via the conditional expectation formula. The latent trajectories can be approximated by the KL expansion using the estimated eigenvalues/eigenfunctions and the conditional estimates of the scores. Then, the testing procedure can be applied with the estimated smooth trajectory. The case of sparsely sampled response can be treated similarly.

### 4.4 Simulation Study

We conducted a simulation study to investigate the finite sample performance of the proposed test under various settings. Specifically, we compared the performance of the two approaches, the RLRT presented in Section 4.2 and the bootstrap-based test; details on the bootstrap algorithm will be provided shortly.

#### 4.4.1 Simulation Setup

We first describe the design of the numerical study. Each data set is generated from the model:

$$Y_i(t) = \int_0^1 [dF_L\{X_i(s), s, t\}] ds + (1 - d)F_{NL}\{X_i(s), s, t\} ds + \epsilon_i(t), \quad (4.3)$$

where $0 \leq d \leq 1$, $F_L(x, s, t)$ is a tri-variate function linear in $x$, and $F_{NL}\{X_i(s), s, t\}$ is non-linear in $x$. The parameter $d$ controls the departure from the null hypothesis given in (4.1) with $d = 1$ corresponding to $H_0$. When $d = 1$, the true model is linear in $x$.

We generate the covariate $X_i(t)$ as $a_{1i}\sqrt{2}\sin(\pi t) + a_{2i}\sqrt{2}\cos(\pi t)$ where $a_{1i} \sim N(0, 4^2)$ and $a_{2i} \sim N(0, 2^2)$ for $i = 1, \ldots, n$. Throughout the study, it is assumed that the covariate $X_i(t)$ are not observed directly. Instead we observe $W_{ij} = X_i(t) + WN(0, \tau^2)$, where $\tau = 0.5$. The response $Y_i(\cdot)$ is generated based on model (4.3) for the all combinations of the following factors:

1. **True function $F(x, s, t)$:**
   - $F_1(x, s, t)$ with $F_L(x, s, t) = 4\sqrt{2}\exp(2s)x\sin(2\pi t)$ and $F_{NL}(x, s, t) = \sqrt{2}x^2s\cos(2\pi t)$
(ii) $F_2(x,s,t)$ with $F_L(x,s,t) = 4\sqrt{2}\exp(2s)x\sin(2\pi t) + 12\sqrt{2}\cos(\pi s)x\sin(4\pi t)$ and $F_{NL}(x,s,t) = \sqrt{2}x^2s\cos(2\pi t) + \sqrt{2}\exp(-0.25x + 2s)\cos(4\pi t)$

(2) Error process $E_i = [\epsilon_i(t_{i1}), \ldots, \epsilon_i(t_{imY,i})]^T$:

(i) $E_{i1} \sim N(0, \sigma^2 I_{mY,i})$

(ii) $E_{i2} \sim N(0, \sigma^2 \Sigma) + N(0, \sigma^2 I_{mY,i})$ where $\Sigma$ has $AR(1)$ structure

(iii) $E_{i3} \sim \xi_{i1}\sqrt{2}\cos(2\pi t) + \xi_{i2}\sqrt{2}\sin(2\pi t) + N(0, \sigma^2 I_{mY,i})$

(iv) $E_{i4} \sim \xi_{i1}\sqrt{2}\cos(2\pi t) + \xi_{i2}\sqrt{2}\sin(4\pi t) + N(0, \sigma^2 I_{mY,i})$

We set $\sigma^2 = 1$ and $\rho = 0.2$. The random variables $\xi_{i1}$ and $\xi_{i2}$ are independently generated from $N(0, 0.2^2)$ and $N(0, 0.1^2)$, respectively.

(3) Sampling design:

(i) dense: $m_Y = 81$ and $m_W = 51$ equidistant points in $[0,1]$ for the response and covariate profile

(ii) sparse: $m_{Y,i}$ and $m_{W,i}$ points in $[0,1]$ for the response and covariate profile with $m_{Y,i} \overset{iid}{\sim} Uniform(59, 64)$ and $m_{W,i} \overset{iid}{\sim} Uniform(37, 42)$

(4) Number of subjects: (i) $n = 50$, (ii) $n = 100$, (iii) $n = 300$, and (iv) $n=500$

We implement the proposed test using $K_x = K_s = 7$ cubic B-splines for $x$ and $s$, and the number of true eigenbasis $K$ is estimated by the percentage of explained variance equal to 95%. We fit the mixed effects model in Algorithm 3 using the R package lme4, and the p-value from the RLRTs is obtained from the R package RLRsim. The null distribution of RLRT statistic is approximated using 10000 simulated samples. The size of test is obtained based on 2000 simulations, and the Bonferroni correction is used to maintain the desired level of the tests. The power performance is assessed based on 1000 simulations. Figure 4.1 shows the true surface of $F_2(x,s,t)$ under the null and the alternative hypothesis along with $x$ and $s$ for a fixed point $t = 0.55$. The red line in the left panel is obtained when $s$ is further fixed as 0.75 and shows linearity of the true surface; the true surface in the right panel is clearly non-linear in $x$. 

78
Figure 4.1: True surface of $F_2(x, s, 0.55)$ corresponding to the model under the null hypothesis (left) and under the alternative hypothesis (right). The red solid line shows the curve when $s$ is fixed as 0.75.

4.4.2 Competitive Method

As a competitive alternative, we consider testing based on F-ratio type test statistic with the null distribution approximated through bootstrap procedure. The test statistic is given as

$$T_n = \frac{RSS_0 - RSS_1}{RSS_1},$$

(4.4)

where $RSS_0$ and $RSS_1$ are the RSSs under the null and the alternative hypothesis, respectively, and are defined earlier in Section 4.1. In general, it is difficult to derive the null distribution of $T_n$. Here, we consider the approach based on bootstrap of the residuals. Specifically, we follow the steps provided in Algorithm 4.

Algorithm 4 is described for dense sampling design for both predictor and response. For the case of sparsely observed data, it proceeds with the smooth version of covariate and response, and the smooth trajectories can be obtained using the method described in Section 4.3. In simulations, we fit the full model using $K_x = K_s = 7$ cubic B-splines for $x$ and $s$. The bi-variate coefficient function $\beta(s, t)$ is estimated using 7 cubic B-splines for $s$ and $t$. Smoothing parameters are selected via REML criterion (Ruppert et al., 2003; Wood, 2006a), and the integration is
Algorithm 4 Bootstrap-based algorithm

1: Fit the full model, \( E[Y_i(t)|X(s)] = \int_0^1 F\{X_i(s), s, d\} ds \), using the estimation procedure of the GAFFR. Calculate the residuals \( e_i(t_j) = Y_i(t_j) - \hat{Y}_i(t_j) \) for all \( i \) and \( j \).
2: Fit the null model, \( E[Y_i(t)|X(s)] = \int_0^1 \beta(s, t)X_i(s)ds \), and obtain the estimated coefficient function \( \hat{\beta}(s, t) \).
3: Calculate the test statistic in (4.4) based on the null and the full model fits; call this value \( T_{n,\text{obs}} \).
4: Let \( \{b_1, \ldots, b_n\} \) be the subject index of bootstrap samples. Resample \( B \) sets of bootstrap residuals \( \{e_i^{(b)}(t_j) = e_{b_i}(t_j) : j\}^n_{i=1} (b = 1, \ldots, B) \) with replacement from the residuals \( \{e_i(t_j) : j\}^n_{i=1} \) obtained in step 1.
5: \textbf{for} \( b = 1 \) to \( B \) \textbf{do}
6: \hspace{1em} Generate response curves under the null model as \( Y_i^{(b)}(t) = \int_0^1 \hat{\beta}(s, t)X_i(s)ds + e_i^{(b)}(t) \) where \( \hat{\beta}(s, t) \) is from step 2, and construct a bootstrap set \( \{X_i(s_k), s_k\}^n_{i=1} \) and \( \{Y_i^{(b)}(t_j), t_j\}^n_{i=1} \).
7: \hspace{1em} Calculate the test statistic in (4.4) based on the model fits obtained from the bootstrap set; call this value \( T^*_b \).
8: \textbf{end for}
9: Compute the p-value by \( \hat{p} = \sum_{b=1}^B I\{T^*_b \geq T_{n,\text{obs}}\} / B \).

approximated by Simpson’s rule. The distribution of the test statistic \( T_n \) is approximated using \( B = 200 \) bootstrap samples for each simulation.

4.4.3 Simulation Results

For the simulation settings considered in Section 4.4.1, we assess size and power performance of the proposed test and the bootstrap alternative, in the next consecutive sections.

Size Performance

We report the estimated size of tests corresponding to the significance levels \( \alpha = 5\%, 10\% \) and \( 15\% \) for both approaches. Table 4.1 shows the rejection probabilities averaged over 2000 simulated data sets when the sampling design is dense. For sample sizes ranging from 50 to 500, the estimated type I error rates of the two tests are within 2 standard errors of the nominal values for all significance levels. For more complex models, \( F(x, s, t) = F_2(x, s, t) \), the two methods accurately estimate the type I error rates with only negligible differences. Theses results are not affected by the complexity of the error covariance as well. Nevertheless, in terms of computational cost, our testing algorithm offers fast computation than the bootstrap
approach; for example, in the case $F(x,s,t) = F_1(x,s,t)$, $E_i = E_i^1$ and $n = 50$, the proposed algorithm took roughly 0.71 seconds to estimate the p-value for a single data set, while it took 215.61 seconds on average using the bootstrap-based algorithm, and these are all measured on a 2.3GHz AMD Opteron Processor.

Next, the estimated type I error rates corresponding to the sparse sampling scenarios are presented in Table 4.2. Due to the high computational cost of the bootstrap-based algorithm, the case corresponding to $n = 500$ is omitted in our comparison. In the case $n = 50$, the proposed test yields slightly inflated type I errors, but the values are mostly within 2 standard errors of the nominal values. As the sample size increases, the estimated type I errors approach the nominal levels in all scenarios. On the other hand, the bootstrap algorithm results in very large type I errors in all cases, and the results are not improved with large sample sizes. Therefore, for sparse sampling designs, the approximate null distribution of the test is not reliable when the bootstrap algorithm is applied. Additional numerical studies also have shown that as the number of observations on each curve decreases (e.g., more than 50% of the data are missing for each curve), the RLRTs had a poor size performance. Nevertheless, for very large sample sizes such as $n = 500$, the proposed test showed improved type I errors in such situations (results are provided in Appendix C.2).

**Power Performance**

The power performance is assessed based on the significance level $\alpha = 5\%$, and the estimated rejection probabilities averaged over 1000 Monte Carlo simulations are presented in Figure 4.2 and 4.3. Figure 4.2 shows the power curves for testing $H_0 : E[Y(\cdot)|X(s)] = \int_{\mathbb{T}} X(s) ds$ based on the proposed test, and the left and right panel correspond to dense and sparse sampling scenario, respectively. Here, we only present the case when the true underlying model is $F(x,s,t) = F_2(x,s,t)$ and $E_i = E_i^3$. Results corresponding to other simulation settings were similar. In the plot, the departure from the null is stronger with the larger value of $1 - d$. As expected, the power of the testing procedure increases with the sample size. The type of
Table 4.1: Type I error rates (×100) for dense sampling design ($m_W = 51$, $m_Y = 81$ for all $i$) based on 2000 Monte Carlo simulations. The values in the parenthesis are the estimated standard errors (×100) of the rejection probabilities. Results corresponding to the proposed test and the bootstrap-based algorithm are indicated by pRLRT and bootstrap, respectively.

<table>
<thead>
<tr>
<th>$n$</th>
<th>method</th>
<th>$\alpha = 5%$</th>
<th>$\alpha = 10%$</th>
<th>$\alpha = 15%$</th>
<th>$\alpha = 5%$</th>
<th>$\alpha = 10%$</th>
<th>$\alpha = 15%$</th>
<th>$\alpha = 5%$</th>
<th>$\alpha = 10%$</th>
<th>$\alpha = 15%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>pRLRT</td>
<td>4.9(0.5)</td>
<td>10.4(0.7)</td>
<td>16.5(0.8)</td>
<td>5.3(0.5)</td>
<td>10.7(0.7)</td>
<td>16.4(0.8)</td>
<td>4.5(0.5)</td>
<td>10.4(0.7)</td>
<td>16.5(0.8)</td>
</tr>
<tr>
<td></td>
<td>bootstrap</td>
<td>6.6(0.6)</td>
<td>10.8(0.7)</td>
<td>15.9(0.8)</td>
<td>6.6(0.6)</td>
<td>11.2(0.7)</td>
<td>15.9(0.8)</td>
<td>5.9(0.5)</td>
<td>11.0(0.7)</td>
<td>15.6(0.8)</td>
</tr>
<tr>
<td>100</td>
<td>pRLRT</td>
<td>4.4(0.5)</td>
<td>9.8(0.7)</td>
<td>15.4(0.8)</td>
<td>4.2(0.4)</td>
<td>9.9(0.7)</td>
<td>16.2(0.8)</td>
<td>4.7(0.5)</td>
<td>10.2(0.7)</td>
<td>15.2(0.8)</td>
</tr>
<tr>
<td></td>
<td>bootstrap</td>
<td>5.7(0.5)</td>
<td>10.1(0.7)</td>
<td>14.8(0.8)</td>
<td>5.8(0.5)</td>
<td>10.3(0.7)</td>
<td>15.4(0.8)</td>
<td>5.7(0.5)</td>
<td>10.0(0.7)</td>
<td>15.0(0.8)</td>
</tr>
<tr>
<td>300</td>
<td>pRLRT</td>
<td>5.2(0.5)</td>
<td>10.5(0.7)</td>
<td>15.8(0.8)</td>
<td>5.6(0.5)</td>
<td>10.4(0.7)</td>
<td>16.2(0.8)</td>
<td>5.0(0.5)</td>
<td>10.4(0.7)</td>
<td>15.4(0.8)</td>
</tr>
<tr>
<td></td>
<td>bootstrap</td>
<td>5.7(0.5)</td>
<td>10.8(0.7)</td>
<td>16.2(0.8)</td>
<td>5.7(0.5)</td>
<td>11.1(0.7)</td>
<td>16.8(0.8)</td>
<td>5.6(0.5)</td>
<td>10.9(0.7)</td>
<td>16.2(0.8)</td>
</tr>
<tr>
<td>500</td>
<td>pRLRT</td>
<td>5.6(0.5)</td>
<td>10.8(0.7)</td>
<td>16.4(0.8)</td>
<td>5.8(0.5)</td>
<td>11.1(0.7)</td>
<td>16.2(0.8)</td>
<td>5.8(0.5)</td>
<td>10.5(0.7)</td>
<td>16.4(0.8)</td>
</tr>
<tr>
<td></td>
<td>bootstrap</td>
<td>4.9(0.5)</td>
<td>10.7(0.7)</td>
<td>16.4(0.8)</td>
<td>5.3(0.5)</td>
<td>10.6(0.7)</td>
<td>15.6(0.8)</td>
<td>5.0(0.5)</td>
<td>10.8(0.7)</td>
<td>16.1(0.8)</td>
</tr>
</tbody>
</table>

Complex model: $F(x, s, t) = F_2(x, s, t)$

<table>
<thead>
<tr>
<th>$n$</th>
<th>method</th>
<th>$\alpha = 5%$</th>
<th>$\alpha = 10%$</th>
<th>$\alpha = 15%$</th>
<th>$\alpha = 5%$</th>
<th>$\alpha = 10%$</th>
<th>$\alpha = 15%$</th>
<th>$\alpha = 5%$</th>
<th>$\alpha = 10%$</th>
<th>$\alpha = 15%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>pRLRT</td>
<td>4.5(0.5)</td>
<td>9.2(0.6)</td>
<td>13.8(0.8)</td>
<td>5.6(0.5)</td>
<td>9.4(0.7)</td>
<td>14.0(0.8)</td>
<td>4.5(0.5)</td>
<td>8.8(0.6)</td>
<td>14.1(0.8)</td>
</tr>
<tr>
<td></td>
<td>bootstrap</td>
<td>5.5(0.5)</td>
<td>10.5(0.7)</td>
<td>15.3(0.8)</td>
<td>6.4(0.5)</td>
<td>11.0(0.7)</td>
<td>16.1(0.8)</td>
<td>5.6(0.5)</td>
<td>10.0(0.7)</td>
<td>15.4(0.8)</td>
</tr>
<tr>
<td>100</td>
<td>pRLRT</td>
<td>4.5(0.5)</td>
<td>9.2(0.6)</td>
<td>13.6(0.8)</td>
<td>4.6(0.5)</td>
<td>9.3(0.6)</td>
<td>13.9(0.8)</td>
<td>4.6(0.5)</td>
<td>9.4(0.7)</td>
<td>13.6(0.8)</td>
</tr>
<tr>
<td></td>
<td>bootstrap</td>
<td>5.5(0.5)</td>
<td>9.8(0.7)</td>
<td>14.3(0.8)</td>
<td>5.4(0.5)</td>
<td>9.4(0.7)</td>
<td>14.2(0.8)</td>
<td>5.5(0.5)</td>
<td>10.2(0.7)</td>
<td>14.8(0.8)</td>
</tr>
<tr>
<td>300</td>
<td>pRLRT</td>
<td>4.9(0.5)</td>
<td>9.3(0.6)</td>
<td>14.2(0.8)</td>
<td>5.1(0.5)</td>
<td>9.5(0.7)</td>
<td>14.3(0.8)</td>
<td>4.5(0.5)</td>
<td>9.4(0.7)</td>
<td>14.0(0.8)</td>
</tr>
<tr>
<td></td>
<td>bootstrap</td>
<td>6.0(0.5)</td>
<td>10.7(0.7)</td>
<td>16.1(0.8)</td>
<td>6.2(0.5)</td>
<td>11.5(0.7)</td>
<td>15.6(0.8)</td>
<td>5.8(0.5)</td>
<td>10.9(0.7)</td>
<td>16.3(0.8)</td>
</tr>
<tr>
<td>500</td>
<td>pRLRT</td>
<td>5.7(0.5)</td>
<td>10.4(0.7)</td>
<td>15.2(0.8)</td>
<td>4.9(0.5)</td>
<td>10.4(0.7)</td>
<td>15.8(0.8)</td>
<td>5.8(0.5)</td>
<td>10.7(0.7)</td>
<td>15.2(0.8)</td>
</tr>
<tr>
<td></td>
<td>bootstrap</td>
<td>5.1(0.5)</td>
<td>10.4(0.7)</td>
<td>16.4(0.8)</td>
<td>5.2(0.5)</td>
<td>10.6(0.7)</td>
<td>15.7(0.8)</td>
<td>4.9(0.5)</td>
<td>11.1(0.7)</td>
<td>15.8(0.8)</td>
</tr>
</tbody>
</table>
Table 4.2: Type I error rates (×100) for sparse sampling design ($m_{W,i} \sim \{37, \ldots, 42\}$, $m_{Y,i} \sim \{59, \ldots, 64\}$) based on 2000 Monte Carlo simulations. The values in the parenthesis are the estimated standard errors (×100) of the rejection probabilities. Results corresponding to the proposed test and the bootstrap-based algorithm are indicated by \textit{pRLRT} and \textit{bootstrap}, respectively.

<table>
<thead>
<tr>
<th>$n$</th>
<th>method</th>
<th>$\alpha = 5%$</th>
<th>$\alpha = 10%$</th>
<th>$\alpha = 15%$</th>
<th>$\alpha = 5%$</th>
<th>$\alpha = 10%$</th>
<th>$\alpha = 15%$</th>
<th>$\alpha = 5%$</th>
<th>$\alpha = 10%$</th>
<th>$\alpha = 15%$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>pRLRT</td>
<td>6.0(0.5)</td>
<td>11.8(0.7)</td>
<td>18.2(0.9)</td>
<td>6.2(0.5)</td>
<td>12.1(0.7)</td>
<td>18.1(0.9)</td>
<td>6.3(0.5)</td>
<td>11.9(0.7)</td>
<td>17.8(0.9)</td>
</tr>
<tr>
<td></td>
<td>bootstrap</td>
<td>16.5(0.8)</td>
<td>24.0(1.0)</td>
<td>30.3(1.0)</td>
<td>14.4(0.8)</td>
<td>22.1(1.0)</td>
<td>27.8(1.0)</td>
<td>16.2(0.8)</td>
<td>22.0(1.0)</td>
<td>28.0(1.0)</td>
</tr>
<tr>
<td>100</td>
<td>pRLRT</td>
<td>5.5(0.5)</td>
<td>10.9(0.7)</td>
<td>16.3(0.8)</td>
<td>5.1(0.5)</td>
<td>10.6(0.7)</td>
<td>16.3(0.8)</td>
<td>5.6(0.5)</td>
<td>10.6(0.7)</td>
<td>15.2(0.8)</td>
</tr>
<tr>
<td></td>
<td>bootstrap</td>
<td>13.9(0.8)</td>
<td>20.8(1.0)</td>
<td>27.1(1.0)</td>
<td>13.0(0.8)</td>
<td>19.3(1.0)</td>
<td>24.4(1.0)</td>
<td>14.0(0.8)</td>
<td>20.7(1.0)</td>
<td>27.4(1.0)</td>
</tr>
<tr>
<td>300</td>
<td>pRLRT</td>
<td>5.2(0.5)</td>
<td>10.5(0.7)</td>
<td>15.7(0.8)</td>
<td>5.5(0.5)</td>
<td>11.0(0.7)</td>
<td>17.0(0.8)</td>
<td>5.5(0.5)</td>
<td>10.7(0.7)</td>
<td>16.1(0.8)</td>
</tr>
<tr>
<td></td>
<td>bootstrap</td>
<td>12.3(0.7)</td>
<td>18.7(0.9)</td>
<td>25.0(1.0)</td>
<td>11.7(0.7)</td>
<td>18.5(0.9)</td>
<td>23.3(0.9)</td>
<td>11.8(0.7)</td>
<td>18.2(0.9)</td>
<td>23.8(1.0)</td>
</tr>
</tbody>
</table>

Simple model: $F(x,s,t) = F_1(x,s,t)$

Complex model: $F(x,s,t) = F_2(x,s,t)$
Figure 4.2: Power ($\times 100$) of the tests at significance level $\alpha = 5\%$ based on 1000 Monte Carlo simulations. Results are for the proposed RLRTs in the case $F(x, s, t) = F_2(x, s, t)$ and $E_i = E_i^3$. The left and right panel correspond to the dense and sparse sampling design, respectively.

dense sampling design has little effect on the results.

The power performance of the RLRTs is also compared with its counterpart, the bootstrap algorithm. Since the bootstrap-based test does not exhibit a good size performance in the sparsely sampled settings, we only compare the power performance of the two methods for the dense sampling design. Results are based on the sample size $n = 100$, and the top and bottom panel correspond to the case where $F(x, s, t) = F_1(x, s, t)$ and $F(x, s, t) = F_2(x, s, t)$, respectively. The power comparisons in Figure 4.3 shows that the proposed test is much more powerful than the bootstrap-based tests in all simulation settings, and the results remain unchanged with the error covariance structures. Furthermore, the difference in power between the two approaches becomes noticeable as the model complexity increases.

To summarize, in simulations the proposed test maintained the correct size for all simulation settings considered and showed excellent power performance. At the same time, our method was not only computationally efficient but also more powerful than the bootstrap-based test, and gave a type I error fairly close to the nominal level even when applied to the settings where the curves are observed on a relatively sparse grid of points.
Figure 4.3: Power (×100) of the tests at significance level $\alpha = 5\%$ based on 1000 Monte Carlo simulations. The left most two panels correspond to the case where $E_1^i$ and $E_2^i$. The right most two panels correspond to $E_3^i$ and $E_4^i$, respectively. Results for $F(x,s,t) = F_1(x,s,t)$ and $F(x,s,t) = F_2(x,s,t)$ are displayed in the top and bottom row, respectively.
4.5 Applications

We turn now to our data applications. We apply the proposed RLRT and the bootstrap alternative to the bike share data (Fanaee-T and Gama, 2013) and the yield curves data (Ruppert and Matteson, 2015); both data are also used to investigate the modeling of GAFFR in Section 3.6; we compared the model fits from the FLM and the model class, general additive function-on-function regression.

4.5.1 Capital Bike Share Study

We first consider the bike share study, where the objective is to understand the relationship between the casual bike rental trajectories and the daily temperature. As in the previous analysis from Section 3.6.1, we analyze the data focusing on the measurements taken every Saturday.

Recall that $Y_{ij} = Y_i(t_{ij}) = \log(1 + \text{count}_{ij})$ is the log-transformed count of bike rentals, where $\text{count}_{ij}$ are the number of casual bike rented at the time point $t_{ij}$ in the $i$th week ($i = 1, \ldots, 105$). Also, recall that $W_{ik}$ is the temperature measured at the time point $s_{ik}$ in the $i$th week. Here, $\{t_{ij}\}_{j=1}^{24}$ and $\{s_{ik}\}_{k=1}^{24}$ are the equidistant points in $[0,1]$. Both the response and covariate trajectories have some missingness, but the proportion of missingness is very low. It is also assumed that the data are measured with noise, and thus the preprocessing steps described in Section 4.3 is necessary; we smooth the covariate and response using the FPCA method and then apply the center/scaling transformation presented in Section 4.3.1. For the implementation of the RLRT, we used $K_x = K_s = 10$ cubic B-splines for $x$ and $s$, respectively, and we approximated the null distribution of the RLRT statistic using 10000 simulated samples. By setting the PVE equal to 95%, three eigenfunctions are estimated (i.e., $K = 3$), and thus we have multiple tests for the null hypotheses $H_{0,1}, H_{0,2}$ and $H_{0,3}$. The proposed test provided p-values for each of the null hypothesis, and the corresponding p-values were equal to $p_1 \approx 0$, $p_2 \approx 0.0002$ and $p_3 \approx 1$. This result clearly provides evidence of a non-linear association between the response and the covariate. For comparison the bootstrap-based test is applied using the same implementation option described in Section 4.4.2, and $B = 1000$ bootstrap replications are used in the analysis.
The bootstrap p-value was nearly 0. This result further supports that the effect of temperature is not linear.

In summary, our findings based on the two tests are consistent with the analysis result presented in Section 3.6.1; recall that fitting the GAFFR model improved prediction accuracy over fitting the FLM.

4.5.2 US Yield Curves Study

Next, we consider the US yield curves study, where the goal of the study is to investigate an association between the changes in US and European daily yield curves, measured at different maturities ranging from 1 year to 30 years. In Chapter 3.6.2, there was an evidence to suggest linear association is true. To assess this formally, we consider the proposed RLRT and the resampling-based algorithm.

Recall that our data consists of US and European daily yields during the period from January 2, 2006 to December 30, 2011 - a total of 1286 days. Let \( Y_{ij} = Y_i(t_j) \) be the changes in the US yield curves corresponding to \( i \)th day (\( i = 1, \ldots, 1286 \)) measured at maturity \( t_j \) (\( j = 1, \ldots, 30 \)), where the maturities \( t_j \) range from 1 year to 30 years. Also, let \( W_{ik} = X_i(s_k) + \delta_{ik} \) be the changes in the European yield curves corresponding to \( i \)th day measured at maturity \( s_k \) (\( k = 1, \ldots, 30 \)), and \( \delta_{ik} \) is the white noise random deviation. As a preliminary step, we smooth the noisy covariate \( W_{ik} \) and apply the center/scaling transformation technique. The RLRT is carried out with \( K_x = K_s = 10 \) cubic B-splines for \( x \) and \( s \), respectively, and we approximated the null distribution of RLRT statistic using 10000 simulated samples. By setting the PVE equal to 95%, two eigenfunctions are estimated (i.e., \( K = 2 \)), and the respective p-values were estimated as \( p_1 \approx 1 \) and \( p_2 \approx 0.2763 \). This result provides evidence that the changes in the European yield curves are linearly associated with the changes in the US yield curves. As a competitor, we also applied the bootstrap-based test using the same implementation option described in Section 4.4.2, and \( B = 1000 \) bootstrap replications are used in the analysis. The bootstrap p-value was estimated as 0.401, indicating a strong linear relationship between the changes in US
and European yields.
Chapter 5

Conclusion

Chapter 2 proposes a wider class of function-on-function regression models, the general functional concurrent model, for functional response and functional covariate observed on the same domain, and discusses significance testing of no association. Our proposed hypothesis testing can formally assess whether the effect of a functional covariate is significant, under the assumption that the relationship between the response and the predictor is general; the linear dependence is a special case of the proposed general model, as described by our proposed modeling. In contrast, the existing literature assumes a linear dependence between the response and the covariate/s. Thus, similar significance tests are only valid when the linearity dependence assumption between the response and the covariate is true. For the two applications - the gait data and the dietary calcium absorption data - our testing procedure found significant association between the response and covariate, under a more general dependence assumption. Furthermore, using the proposed methods, we found evidence that the relationship between the response and covariate is indeed linear; in contrast, a functional linear concurrent model assumes a linear dependence is valid. Thus, our proposed procedure allows one to approach the problem from a more general point of view. Software in the form of R code, together with a sample input data set and complete documentation is available at http://www4.stat.ncsu.edu/~maity/software/GFCM-R_code.zip.
Chapter 3 considers a general class of function-on-function regression models for functional response and functional covariate observed on possible different domains. These models are a generalization of the standard FLMs, and allow the relationship between the response and the covariate to be non-linear. We propose a novel estimation technique that is computationally very fast. We develop prediction inference for a future covariate function. Our methodology can be extended to accommodate multiple functional covariates. We show through numerical study that when the true model is linear, the proposed prediction performance is very close to that from FLM. At the same time, if the true model is non-linear, the GAFFR and the proposed methodology yield considerably improved prediction performance. In the capital bike share study, we found evidence that our method provides optimal model fits compared to the FLM. This result indicates that the effect of daily temperature is not simply linear. In the daily yield curves study, our method and the FLM provided similar prediction performances, indicating a simple linear association is more appropriate. The R code used in the simulation is available at the link: http://www4.stat.ncsu.edu/~staicu/Code/GAFFRcode.zip.

Chapter 4 develops a restricted likelihood ratio test procedure for testing linearity in general additive function-on-function regression. The methodology may be viewed as an extension of McLean et al. (2015) for scalar-on-function regression to function-on-function regression models. In numerical studies, our method showed good size and power performance in various realistic simulation scenarios. Nevertheless, the performance was somewhat affected by sparsity of the design of both response and covariate, and improving the size performance is still important when the curves are measured at only a small number of grid points; such direction is left for future research. In simulations, we found that the proposed test outperforms the bootstrap-based test, and furthermore it requires little computational cost. The bike share data is again used to test whether the true relationship between the covariate and response is linear, and the two competitive tests provided strong evidence that the relationship is not simply linear. In the daily yields curve study, the proposed method and the resampling-based algorithm exhibited strong evidence of linearity.
Finally, in regards to further improvements, an interesting problem would be extending the modeling of GFCM and GAFFR to non-Gaussian response case and developing testing procedures. In the GAFFR framework, testing procedure for sparsely sampled response and covariate can be further studied.
REFERENCES


URL: http://CRAN.R-project.org/package=refund


Rice, J. and Silverman, B. (1991). Estimating the mean and covariance structure nonparametri-


2nd edn, Springer.

New York: Cambridge University Press.


effect variance or polynomial regression in additive and linear mixed models, *Computational 


Sinica* 14: 1239–1257.


to predict daily fine fuel moisture content in eucalypt forest, *Forest Ecology and Management* 335: 261–269.


*Journal of the American Statistical Association* 93(1403-1418).

Swihart, B. J., Goldsmith, J. and Crainiceanu, C. M. (2014). Restricted likelihood ratio tests 


URL: [http://CRAN.R-project.org/package=mgcv](http://CRAN.R-project.org/package=mgcv)


APPENDICES
Appendix A

Additional Details for Chapter 2

This section comprises of three sections. Additional simulation results are reported in Appendix A.1. Additional investigation for the gait data application is included in Appendix A.2. Appendix A.3 provides the implementation details.

A.1 Additional Simulation Results

This section provides additional simulation results. Specifically in Appendix A.1.1 we show further investigation of the prediction error, in Appendix A.1.2 we compare the prediction results obtained by using different covariance estimation methods. In Appendix A.1.3 we investigate the computational complexity of the proposed algorithm and its increase for two functional covariates.

A.1.1 Further Investigation of Prediction Error

Additional Simulations for Irregular and Sparse Design

We first provide additional supporting results corresponding to another level of sparseness. We now consider the case with moderate sparseness relative to the one presented in Section 2.6.1, and compare the prediction results with those in Section 2.6.2. Here, we generate $m_i$ time
Table A.1: Summary of RMSPE\textsuperscript{in}, RMSPE\textsuperscript{out}, ICP, IL, and R(SE) for a moderately sparse design. The models fitted by our method and the linear FCM are indicated by GFCM and FCM, respectively.

(a) True model is \( F(x, t) = F\text{NL}(x, t) \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( E )</th>
<th>RMSPE\textsuperscript{in}</th>
<th>RMSPE\textsuperscript{out}</th>
<th>ICP</th>
<th>IL</th>
<th>R(SE)</th>
<th>ICP</th>
<th>IL</th>
<th>ICP</th>
<th>IL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GFCM</td>
<td>FCM</td>
<td>GFCM</td>
<td>FCM</td>
<td>GFCM</td>
<td>FCM</td>
<td>GFCM</td>
<td>FCM</td>
<td>GFCM</td>
</tr>
<tr>
<td>1</td>
<td>( E\text{\textsuperscript{1}} )</td>
<td>1.09</td>
<td>3.38</td>
<td>1.31</td>
<td>3.86</td>
<td>0.952</td>
<td>0.924</td>
<td>4.35</td>
<td>12.12</td>
<td>[3.66, 6.85]</td>
</tr>
<tr>
<td>50</td>
<td>( E\text{\textsuperscript{2}} )</td>
<td>1.41</td>
<td>3.50</td>
<td>1.56</td>
<td>3.96</td>
<td>0.950</td>
<td>0.929</td>
<td>5.61</td>
<td>12.75</td>
<td>[5.03, 7.72]</td>
</tr>
<tr>
<td>100</td>
<td>( E\text{\textsuperscript{3}} )</td>
<td>1.86</td>
<td>3.74</td>
<td>2.25</td>
<td>4.20</td>
<td>0.936</td>
<td>0.933</td>
<td>7.63</td>
<td>14.06</td>
<td>[5.61, 11.48]</td>
</tr>
<tr>
<td>300</td>
<td>( E\text{\textsuperscript{4}} )</td>
<td>1.09</td>
<td>3.55</td>
<td>1.10</td>
<td>3.76</td>
<td>0.962</td>
<td>0.932</td>
<td>4.26</td>
<td>12.62</td>
<td>[3.62, 6.53]</td>
</tr>
</tbody>
</table>

(b) True model is \( F(x, t) = F\text{L}(x, t) \)

<table>
<thead>
<tr>
<th>( n )</th>
<th>( E )</th>
<th>RMSPE\textsuperscript{in}</th>
<th>RMSPE\textsuperscript{out}</th>
<th>ICP</th>
<th>IL</th>
<th>R(SE)</th>
<th>ICP</th>
<th>IL</th>
<th>ICP</th>
<th>IL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>GFCM</td>
<td>FCM</td>
<td>GFCM</td>
<td>FCM</td>
<td>GFCM</td>
<td>FCM</td>
<td>GFCM</td>
<td>FCM</td>
<td>GFCM</td>
</tr>
<tr>
<td>1</td>
<td>( E\text{\textsuperscript{1}} )</td>
<td>0.91</td>
<td>0.91</td>
<td>0.90</td>
<td>0.90</td>
<td>0.953</td>
<td>0.953</td>
<td>3.59</td>
<td>3.59</td>
<td>[3.54, 3.76]</td>
</tr>
<tr>
<td>50</td>
<td>( E\text{\textsuperscript{2}} )</td>
<td>1.28</td>
<td>1.28</td>
<td>1.27</td>
<td>1.27</td>
<td>0.951</td>
<td>0.951</td>
<td>5.02</td>
<td>5.02</td>
<td>[4.94, 5.26]</td>
</tr>
<tr>
<td>100</td>
<td>( E\text{\textsuperscript{3}} )</td>
<td>1.77</td>
<td>1.82</td>
<td>1.98</td>
<td>1.87</td>
<td>0.935</td>
<td>0.942</td>
<td>7.14</td>
<td>7.13</td>
<td>[5.29, 9.88]</td>
</tr>
<tr>
<td>300</td>
<td>( E\text{\textsuperscript{4}} )</td>
<td>1.08</td>
<td>3.64</td>
<td>0.96</td>
<td>3.68</td>
<td>0.969</td>
<td>0.942</td>
<td>4.20</td>
<td>12.91</td>
<td>[3.61, 6.36]</td>
</tr>
</tbody>
</table>

99
Table A.2: Summary of RMSPE and standard deviations (in parentheses) obtained by fitting the GFCM based on 1000 simulations. The simulation settings correspond to the case where \( E_i = E^3_i \).

<table>
<thead>
<tr>
<th>( n )</th>
<th>Dense ( m_i = m = 81 )</th>
<th>Moderately Sparse ( m_i \sim \text{Unif}(20,31) )</th>
<th>Sparse ( m_i \sim \text{Unif}(29,41) )</th>
<th>Dense ( m_i = m = 81 )</th>
<th>Moderately Sparse ( m_i \sim \text{Unif}(20,31) )</th>
<th>Sparse ( m_i \sim \text{Unif}(29,41) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>1.79 (0.11)</td>
<td>1.86 (0.12)</td>
<td>1.91 (0.13)</td>
<td>1.76 (0.11)</td>
<td>1.77 (0.11)</td>
<td>1.77 (0.11)</td>
</tr>
<tr>
<td>100</td>
<td>1.84 (0.08)</td>
<td>1.90 (0.09)</td>
<td>1.94 (0.09)</td>
<td>1.80 (0.08)</td>
<td>1.81 (0.08)</td>
<td>1.81 (0.08)</td>
</tr>
<tr>
<td>300</td>
<td>1.87 (0.05)</td>
<td>1.92 (0.05)</td>
<td>1.96 (0.05)</td>
<td>1.83 (0.05)</td>
<td>1.83 (0.05)</td>
<td>1.84 (0.05)</td>
</tr>
</tbody>
</table>

points in \([0, 1]\) by setting \( m_i \sim \text{Uniform}(29, 41)\) for each \( i \). For convenience we call this setting \textit{moderately sparse design}, whereas the sampling design used in Section 2.6.2 is referred to as \textit{sparse design}. The results presented in Table A.1 indicate that our approach still maintains in-sample and out-of-sample prediction accuracy.

**Additional Simulations for Complicated Error Structure**

Next, we further discuss prediction accuracy investigated in Section 2.6.2. In Table 2.1, the value of RMSPE slightly increases with a larger sample size when the error structure is complicated (see results corresponding to \( E^3_i \)). Indeed, there are two major sources that cause such phenomena: (i) sampling design of \( t \) and (ii) variance of the latent functional covariate. To aid understanding these results, one needs to compare the values of RMSPE with the true variance of the error process. In our simulation experiment, the true variance (averaged over \( t \)) of each error process can be computed as:

1. \( \text{Var}(E^1_i) = \sqrt{0.8} \approx 0.89. \)
2. \( \text{Var}(E^2_i) = \sqrt{0.8 + 0.8} \approx 1.26. \)
3. \( \text{Var}(E^3_i) = \sqrt{2 + 0.75^2 + 0.8} \approx 1.83. \)

It is expected that the values of RMSPE get closer to the above variance typically with a large sample size.

(i) We first analyze the sensitivity of prediction accuracy with regards to sampling design of \( t \). Consider the results corresponding to densely sampled design presented in Table 2.1. For a
Table A.3: Summary of RMSPE\textsuperscript{in} and standard deviations (in parentheses) obtained by fitting the GFCM based on 1000 simulations. The simulation settings correspond to the case where $E_i = E_i^3$. Reduced variance is used to generate the covariate trajectories.

<table>
<thead>
<tr>
<th></th>
<th>(a) True model is $F(x, t) = F_{NL}(x, t)$</th>
<th></th>
<th>(b) True model is $F(x, t) = F_{L}(x, t)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Dense</td>
<td>Moderately Sparse</td>
<td>Sparse</td>
</tr>
<tr>
<td>$n$</td>
<td>$m_i$</td>
<td>$m$</td>
<td>$m_i \sim \text{Unif}(29, 41)$</td>
</tr>
<tr>
<td>50</td>
<td>1.75 (0.11)</td>
<td>1.75 (0.11)</td>
<td>1.75 (0.11)</td>
</tr>
<tr>
<td>100</td>
<td>1.80 (0.08)</td>
<td>1.80 (0.08)</td>
<td>1.80 (0.08)</td>
</tr>
<tr>
<td>300</td>
<td>1.83 (0.05)</td>
<td>1.82 (0.05)</td>
<td>1.82 (0.05)</td>
</tr>
</tbody>
</table>

complicated error pattern such as $E_i^3$, the RMSPE\textsuperscript{in}'s are smaller than the above error variance when the sample size is small (e.g., $n = 50$). With larger sample sizes, the values get closer to the true error variance. On the other hand, when the sampling design is sparse and $E_i = E_i^3$, it seems that the RMSPE\textsuperscript{in}'s are slightly inflated. This is because the sampling design of $t$ slightly affects the prediction results. However, with the moderately sparse design (more observed data per curve), the values of RMSPE\textsuperscript{in}'s are less inflated compared to the results from sparsely sampled design. To clarify this point, simulation results corresponding to the above settings are summarized in Table A.2 with the estimated standard deviations in the parentheses. It is evident from the results that the targeted values (true error variances) are within two standard deviations of the estimated values, and thus our estimation method still preserves prediction accuracy even when the error patterns are complicated.

(ii) The variance of the covariate profiles is another factor that affects the prediction accuracy, as we investigate through additional simulation studies; here, we consider the case with a complicated error structure ($E_i = E_i^2$) as well. We generate $N = 1000$ samples with the true functional covariate given by $X_i(t) = a_0i + a_{1i}\sqrt{2}\sin(\pi t) + a_{2i}\sqrt{2}\cos(\pi t)$ where $a_{0i} \sim N(0, 0.5^2)$, $a_{1i} \sim N(0, 0.35^2)$ and $a_{2i} \sim N(0, 0.15^2)$ for $i = 1, \ldots, n$, and the noisy trajectories are obtained from $W_{ij} = X_i(t) + \text{WN}(0, 0.2^2)$. Thus the true variance of the covariate is reduced compared to the one used in Section 2.6.1. Table A.3 shows the in-sample prediction results under this setting. If we compare the results from Table A.3 with the ones from Table A.2, the prediction using the reduced variance of the covariate improves the prediction accuracy.
A.1.2 Further Investigation of Different Covariance Estimation Methods

In the literature, there are various approaches to estimate the covariance of the residual process: using local polynomial smoothing and using global smoothing via B-spline basis functions. So far the latter approach is implemented in \texttt{R} where as the former one is implemented in \texttt{Matlab}. The proposed method in Section 2.2.3 can be implemented by \texttt{fpca.sc} function in the \texttt{refund} \texttt{R} package. We now consider two alternative approaches: (i) Yao et al. (2005a) which is implemented in \texttt{Matlab} using the functions of the \texttt{PACE} package and (ii) Xiao et al. (2013) implemented in \texttt{R} using \texttt{fpca.face} function of the \texttt{refund} package. We carried out additional simulation studies to compare the performance of variance estimation for two cases - when the functional covariates are observed densely or sparsely and with measurement error. Table A.4 presents the results corresponding to the setting where \(n = 100\) and \(E_i = E_3^3\). As in the table, the results are quite robust to the methods. However, one remark is that \texttt{fpca.face} is originally not developed for sparsely sampled data, and one might encounter a problem if sparseness of the sampling is significantly increased; for example, we only observe 5\%\textasciitilde10\% of the data for each curve.

A.1.3 Further Investigation of Computation Time

This section investigates the computational complexity of the proposed estimation algorithm with respect to the number of additional covariates. Specifically, we compared computational time taken to fit the following two models. We first obtained model fits when the true model is

\[
Y_i(t) = F_1\{X_{1,i}(t), t\} + F_2\{X_{2,i}(t), t\} + \epsilon_i(t),
\]

(A.1)

where \(F_1(x, t) = 1 + x + t + 2x^2t\) and \(F_2(x, t) = 0.75\exp(xt)\). For comparisons, we also obtained model fits when the true model is

\[
Y_i(t) = F_1\{X_{1,i}(t), t\} + \epsilon_i(t).
\]

(A.2)
Table A.4: Summary of ICP, IL, and R(SE) for sample size 100 and $E_i = E_i^3$ based on 1000 simulations.

<table>
<thead>
<tr>
<th>Method</th>
<th>$fpca.sc$</th>
<th>$fpca.face$</th>
<th>PACE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong></td>
<td><strong>1</strong></td>
<td><strong>1</strong></td>
<td><strong>1</strong></td>
</tr>
<tr>
<td><strong>α</strong></td>
<td><strong>α</strong></td>
<td><strong>α</strong></td>
<td><strong>α</strong></td>
</tr>
<tr>
<td><strong>95</strong></td>
<td><strong>95</strong></td>
<td><strong>95</strong></td>
<td><strong>95</strong></td>
</tr>
<tr>
<td><strong>90</strong></td>
<td><strong>90</strong></td>
<td><strong>90</strong></td>
<td><strong>90</strong></td>
</tr>
<tr>
<td><strong>85</strong></td>
<td><strong>85</strong></td>
<td><strong>85</strong></td>
<td><strong>85</strong></td>
</tr>
<tr>
<td><strong>Densely Sampled Design:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_i = m = 81$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$fpca.sc$</td>
<td>0.946 0.944 7.34 14.56 [5.53, 9.84] [8.37, 23.27] 0.895 0.915 6.16 12.22 0.844 0.884 5.39 10.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$fpca.face$</td>
<td>0.940 0.950 7.22 14.51 [5.49, 9.82] [8.51, 23.61] 0.886 0.927 6.06 12.18 0.835 0.894 5.30 10.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACE</td>
<td>0.943 0.960 7.29 10.77 [5.43, 9.73] [8.41, 23.45] 0.889 0.929 6.12 12.31 0.839 0.897 5.35 10.77</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Moderately Sparse Design:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_i \sim Unif(29, 41)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$fpca.sc$</td>
<td>0.948 0.943 7.55 14.52 [5.69, 10.67] [8.18, 24.34] 0.899 0.912 6.34 12.19 0.849 0.881 5.55 10.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$fpca.face$</td>
<td>0.943 0.958 7.37 14.45 [5.64, 10.24] [8.37, 22.60] 0.890 0.926 6.18 12.13 0.841 0.893 5.41 10.62</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACE</td>
<td>0.948 0.959 7.62 14.73 [5.65, 11.12] [8.23, 25.50] 0.897 0.928 6.40 12.36 0.849 0.896 5.60 10.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sparsely Sampled Design:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_i \sim Unif(20, 31)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$fpca.sc$</td>
<td>0.949 0.942 7.69 14.51 [5.82, 11.09] [8.13, 24.67] 0.900 0.911 6.46 12.17 0.852 0.879 5.65 10.65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$fpca.face$</td>
<td>0.939 0.960 7.38 14.19 [5.59, 11.38] [8.97, 24.28] 0.885 0.929 6.19 11.91 0.834 0.895 5.42 10.42</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACE</td>
<td>0.950 0.960 7.85 14.84 [5.84, 11.78] [8.26, 26.51] 0.902 0.929 6.59 12.46 0.855 0.897 5.77 10.90</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) True model is $F(x,t) = F_N^L(x,t)$

<table>
<thead>
<tr>
<th>Method</th>
<th>$fpca.sc$</th>
<th>$fpca.face$</th>
<th>PACE</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>1</strong></td>
<td><strong>1</strong></td>
<td><strong>1</strong></td>
<td><strong>1</strong></td>
</tr>
<tr>
<td><strong>α</strong></td>
<td><strong>α</strong></td>
<td><strong>α</strong></td>
<td><strong>α</strong></td>
</tr>
<tr>
<td><strong>95</strong></td>
<td><strong>95</strong></td>
<td><strong>95</strong></td>
<td><strong>95</strong></td>
</tr>
<tr>
<td><strong>90</strong></td>
<td><strong>90</strong></td>
<td><strong>90</strong></td>
<td><strong>90</strong></td>
</tr>
<tr>
<td><strong>85</strong></td>
<td><strong>85</strong></td>
<td><strong>85</strong></td>
<td><strong>85</strong></td>
</tr>
<tr>
<td><strong>Densely Sampled Design:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_i = m = 81$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$fpca.sc$</td>
<td>0.943 0.947 7.16 7.13 [5.41, 9.10] [5.40, 8.81] 0.891 0.896 6.01 5.98 0.839 0.846 5.26 5.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$fpca.face$</td>
<td>0.936 0.939 7.06 7.08 [5.37, 9.37] [5.36, 9.40] 0.880 0.884 5.92 5.94 0.828 0.833 5.18 5.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACE</td>
<td>0.939 0.943 7.13 7.16 [5.31, 9.08] [5.33, 8.96] 0.884 0.889 5.98 6.01 0.833 0.839 5.24 5.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Moderately Sparse Design:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_i \sim Unif(29, 41)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$fpca.sc$</td>
<td>0.943 0.947 7.14 7.15 [5.37, 9.36] [5.37, 9.27] 0.891 0.896 5.99 6.00 0.840 0.846 5.24 5.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$fpca.face$</td>
<td>0.935 0.938 7.03 7.06 [5.37, 9.24] [5.38, 9.30] 0.879 0.884 5.90 5.93 0.828 0.832 5.16 5.19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACE</td>
<td>0.941 0.944 7.21 7.25 [5.31, 9.76] [5.33, 9.75] 0.887 0.892 6.05 6.09 0.837 0.842 5.29 5.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Sparsely Sampled Design:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$m_i \sim Unif(20, 31)$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$fpca.sc$</td>
<td>0.943 0.947 7.14 7.17 [5.38, 9.56] [5.37, 9.54] 0.891 0.896 6.00 6.01 0.840 0.846 5.25 5.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$fpca.face$</td>
<td>0.925 0.928 6.87 6.92 [4.94, 10.19] [4.90, 10.29] 0.866 0.870 5.76 5.81 0.812 0.818 5.04 5.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PACE</td>
<td>0.943 0.946 7.27 7.32 [5.35, 10.20] [5.36, 10.23] 0.890 0.895 6.10 6.14 0.840 0.845 5.34 5.38</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
In this experiment, the covariate functions $X_{q,i}(t)$ are generated by a non-linear random process $a_{0i,q} + a_{1i,q}\sqrt{2}\sin(\pi t) + a_{2i,q}\sqrt{2}\cos(\pi t)$, where $a_{0i,q} \sim N(0, \{2^{-0.5(q-1)}\}^2)$, $a_{1i,q} \sim N(0, \{0.85 \times 2^{-0.5(q-1)}\}^2)$ and $a_{2i,q} \sim N(0, \{0.70 \times 2^{-0.5(q-1)}\}^2)$ for $i = 1, \ldots, n$ and for $q = 1, 2$. For the rest of the simulation settings, we use the same setup described in Section 2.6.1. There are two available implementation tools to obtain predictions: \texttt{gam} and \texttt{bam} function of \texttt{mgcv} (Wood, 2015) \texttt{R} package. Both functions provide almost identical model fits, but computational advantage can be gained when using the \texttt{bam} function with very large data sets. Details about the functions can be found in Wood (2015).

We obtained the GFCM fits using $K_{x_1} = K_{t_1} = 7$ and $K_{x_2} = K_{t_2} = 6$ cubic B-splines for both $F_1\{X_{1,i}(t), t\}$ and $F_2\{X_{2,i}(t), t\}$, respectively. We obtained smooth versions of covariates using the FPCA techniques with the percent variance set to 99%. Table A.5 displays RMSPE\textsubscript{in}, RMSPE\textsubscript{out} and average computation time in seconds averaged over 1000 Monte Carlo replications. The computation time is measured on a 2.3GHz AMD Opteron Processor. The results obtained by using \texttt{gam} and \texttt{bam} functions are presented in the left and right post, respectively. The top and bottom post display the results corresponding to model (A.1) and model (A.2), respectively. We found that adding additional covariate does not result in loss of prediction accuracy. Both \texttt{gam} and \texttt{bam} slightly increase the average computation time as the sample size increases. Nevertheless, the additional computational expense of running model (A.1) is still reasonable. The results also show that computations can be further sped up if \texttt{bam} is used. We recommend using \texttt{bam} for large sample sizes.

### A.2 Further Investigation of Gait Data Example

The curves in the gait data have different characteristics compared to the curves generated for simulation studies. Furthermore, the sample size ($n = 39$) in the gait data is smaller than the settings ($n = 50, 100, 300$) considered in the experiment. In this section, we investigate the prediction performance of our method as well as the small sample properties for data sets which imitate principal features of the gait data. The goal of the study is to confirm all the results
Table A.5: Summary of (1) \text{RMSPE}^{in}, (2) \text{RMSPE}^{out}, and (3) average computation time in seconds obtained by fitting the GFCM based on 1000 simulations. The left (right) table is obtained by using \texttt{gam} (\texttt{bam}) function of \texttt{mgcv} R package.

(a) True model is $Y_i(t) = F_1 \{X_{1,i}(t), t\} + F_2 \{X_{2,i}(t), t\} + \epsilon_i(t)$

<table>
<thead>
<tr>
<th>n</th>
<th>$E_i$</th>
<th>\text{Dense} $m_i = m = 81$</th>
<th>\text{gam}</th>
<th>$m_i \sim \text{Unif}(20, 31)$</th>
<th>\text{Sparse}</th>
<th>$m_i = m = 81$</th>
<th>\text{bam}</th>
<th>$m_i \sim \text{Unif}(20, 31)$</th>
<th>\text{Sparse}</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>$E_i^1$</td>
<td>1.30 1.31 11.36</td>
<td>1.43 1.52 2.68</td>
<td>1.30 1.30 2.70</td>
<td>1.43 1.52 3.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>$E_i^1$</td>
<td>1.73 2.11 11.43</td>
<td>1.84 2.18 2.65</td>
<td>1.73 2.12 3.28</td>
<td>1.84 2.19 3.29</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>$E_i^1$</td>
<td>0.96 0.94 24.74</td>
<td>1.13 1.07 4.20</td>
<td>0.96 0.93 4.44</td>
<td>1.13 1.07 3.20</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(b) True model is $Y_i(t) = F_1 \{X_{1,i}(t), t\} + \epsilon_i(t)$

<table>
<thead>
<tr>
<th>n</th>
<th>$E_i$</th>
<th>\text{Dense} $m_i = m = 81$</th>
<th>\text{gam}</th>
<th>$m_i \sim \text{Unif}(20, 31)$</th>
<th>\text{Sparse}</th>
<th>$m_i = m = 81$</th>
<th>\text{bam}</th>
<th>$m_i \sim \text{Unif}(20, 31)$</th>
<th>\text{Sparse}</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>$E_i^1$</td>
<td>1.31 1.28 24.37</td>
<td>1.44 1.38 5.12</td>
<td>1.31 1.28 4.17</td>
<td>1.44 1.38 2.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>$E_i^1$</td>
<td>1.80 1.96 24.15</td>
<td>1.91 2.01 4.86</td>
<td>1.80 1.96 3.37</td>
<td>1.91 2.01 3.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>300</td>
<td>$E_i^1$</td>
<td>0.97 0.92 57.74</td>
<td>1.13 0.97 18.14</td>
<td>0.97 0.92 4.61</td>
<td>1.13 0.97 4.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

105
Figure A.1: Displayed plots are curves for sample size 100 simulated based on the gait data. The smoothed version of covariate functions $\hat{X}_i^{\text{sim}}$ are presented in the left. The middle and the rightmost panel present response curves $Y_i(t)$ generated based on $F^{L,\text{gait}}(x,t)$ and $F^{NL,\text{gait}}(x,t)$, respectively. The last two subjects are highlighted in different colors.

obtained from the gait data analysis in Section 2.6.3.

The new simulation study generates the covariates $X_i(t)$ from a process with the mean and covariance functions that equal their estimated counterparts from the data. For this purpose, we first apply the FPCA to the observed hip angles using the entire 39 subjects, and obtain a smoothed version of $n$ curves by computing $\hat{X}_i^{\text{sim}}(t_j) = \hat{\mu}_X(t) + \sum_{k=1}^{K} \xi_{ik} \hat{\phi}_k(t)$, where $\xi_{ik}$ ($k = 1, \ldots, K$) are normally distributed with zero-mean and variance $\hat{\lambda}_k$ ($i = 1, \ldots, n$). The estimates $\hat{\mu}_X(t), \hat{\phi}_k(t)$, and $\hat{\lambda}_k$ are obtained from the observed data, and the finite truncation $K$ is chosen by a preset level PVE=99%. We generate noisy covariate trajectories from $W_{ij}^{\text{sim}} = \hat{X}_i^{\text{sim}}(t_{ij}) + \delta_{ij}$. The noise $\delta_{ij}$ are normally distributed with zero-mean, and the noise variance is estimated from the original data. To assess both the in-sample and out-of-sample prediction capabilities we consider $n = 30, 100$ and $300$ subjects for the training data, and the test data contains 9 subjects. For evaluation points $t$, we use the same time points used in the data analysis.

The response $Y_i(\cdot)$ is generated from the model $Y_i(t) = F\{X_i(t), t\} + \epsilon_i(t)$ with the following true functions for $F(\cdot, \cdot)$: a linear version $F^{L,\text{gait}}(x,t)$ and a non-linear version $F^{NL,\text{gait}}(x,t)$. We
define the linear version by \( F^{L,gait}(x,t) = \beta_0(t) + \beta_1(t)x \), where \( \beta_0(t) = \hat{\beta}_0^*(t) - \hat{\beta}_1^*(t)\hat{\mu}_X(t)/\hat{\sigma}_X(t) \) and \( \beta_1(t) = \hat{\beta}_1^*(t)/\hat{\sigma}_X(t) \). Here, \( \hat{\beta}_0^*(t) \) and \( \hat{\beta}_1^*(t) \) are the intercept and the slope estimated from the gait data, and \( \hat{\mu}_X(t) \) and \( \hat{\sigma}_X(t) \) are the sample mean and the sample standard deviation at the point \( t \) obtained from the original data as well. It can be easily shown that \( F^{L,gait}\{X_i(t),t\} = \hat{\beta}_0^*(t) + \hat{\beta}_1^*(t)X_i(t) - \hat{\mu}_X(t)\hat{\sigma}_X(t) \), and thus the response curves generated from \( F^{L,gait}(x,t) \), in fact, are defined similarly to the response curves fitted by the linear FCM based on the original gait data. Next, we define the non-linear version by \( F^{NL,gait}(x,t) = \exp(xt/12) - x \). To generate the random errors \( \epsilon_i(t) \), we first obtain residuals from the fitted linear model of the gait data, and employ the FPCA methods to estimate the variance of the random errors. Figure A.1 displays the simulated covariates \( \tilde{X}_i^{sim}(t) \) \((i = 1,\ldots,100)\) evaluated at the points \( t_j \) (leftmost panel) as well as the response curves \( Y_i(t) \) obtained from the models \( F^{L,gait}(x,t) \) (middle panel) and \( F^{NL,gait}(x,t) \) (rightmost panel). In the plot, when the true function is \( F^{L,gait}(x,t) \), patterns in the response curves are very similar to the ones from the original gait data.

Finally, we assess the prediction performance of the proposed method using \( N = 1000 \) samples, and compare its performance with the linear FCM. We fit the GFCM using \( K_x = K_t = 11 \) cubic B-splines for \( x \) and \( t \). When the true function is \( F^{L,gait}(x,t) \) (the top three panels in Table A.6), the overall predictive performance of the GFCM and the linear FCM is relatively similar. These results indicate that the underlying relationship between the covariate and the response is linear, as we investigated through the gait data analysis. When the true function is \( F^{NL,gait}(x,t) \) (the bottom three panels in Table A.6), the GFCM better captures the complex non-linear relationships than the linear FCM in all scenarios. Other measures also indicate that the prediction variance obtained from the linear FCM is less accurate. Overall, the GFCM and our estimation method successfully capture the underlying relationship even with the small sample sizes, and the prediction results are only little affected by the complexity of true model.
Table A.6: Summary of RMSPE\textsuperscript{in}, RMSPE\textsuperscript{out}, ICP, IL, and R(SE) obtained from the simulation studies of the gait data example. The models fitted by our method and the linear FCM are indicated by GFCM and FCM, respectively.

| n | RMSPE\textsuperscript{in} | RMSPE\textsuperscript{out} | $1 - \alpha = 0.95$ | ICP | IL | R(SE) | ICP | IL | $1 - \alpha = 0.90$ | ICP | IL | $1 - \alpha = 0.85$ | ICP | IL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 30 | 4.83 | 5.00 | 5.41 | 5.26 | 0.920 | 0.933 | 18.98 | 19.36 | [12.96, 33.92] | [13.56, 34.61] | 0.862 | 0.878 | 15.93 | 16.25 | 0.808 | 0.827 | 13.84 | 14.22 |
| 100 | 5.08 | 5.12 | 5.20 | 5.16 | 0.942 | 0.945 | 19.48 | 19.58 | [14.47, 35.24] | [14.57, 35.46] | 0.890 | 0.894 | 16.35 | 16.43 | 0.840 | 0.844 | 14.31 | 14.38 |
| 300 | 5.13 | 5.14 | 5.14 | 5.13 | 0.947 | 0.948 | 19.55 | 19.59 | [14.78, 35.44] | [14.81, 35.50] | 0.897 | 0.898 | 16.41 | 16.44 | 0.847 | 0.848 | 14.36 | 14.39 |

(a) True model is $F(x, t) = F\text{\textsuperscript{L-gait}}(x, t)$

(b) True model is $F(x, t) = F\text{\textsuperscript{NL-gait}}(x, t)$
A.3 Implementation Details

We implemented our proposed estimation and testing methodology using R software. The model components of the GFCM can be estimated using the \texttt{gam/bam} functions of \texttt{mgcv} package (Wood, 2015). The smoothing parameter choice is automatic in \texttt{gam/bam}; we use REML criteria to select the smoothing parameters. For the sparsely sampled design, we employ the R package \texttt{refund} to carry out FPCA. In the following, we illustrate how the R software codes can be used to implement our procedures.

We first pool all observed data. Let \( \text{trep} \) be the \( N \)-dimensional vector of evaluation points pooled from all subjects. Let \( y.\text{vec} \) and \( x.\text{vec} \) be the \( N \)-dimensional vectors of response and transformed covariate, respectively, where the evaluation points of the functional correspond to the vector \( \text{trep} \). The transformed covariate indicates that the point-wise center/scaling transformation is applied. Then a simple command

\[
\text{fit} <- \text{gam}(y.\text{vec} \sim \text{te}(x.\text{vec}, \text{trep}, \text{bs}='ps', k=c(Kx, Kt)), \text{method}='\text{REML}')
\]

performs our estimation procedure. The function \texttt{te()} specifies the tensor product of basis functions. The \texttt{bs} argument selects the type of penalized splines. In our case, we set \texttt{bs}='ps' to incorporate B-splines with the second order difference penalties. The \texttt{k} argument specifies the number of basis functions, which is \( Kx \) for \( x.\text{vec} \) and \( Kt \) for \( \text{trep} \). Smoothing for the penalized splines is indicated by \texttt{method}='REML'. The \texttt{gam()} function will offer the \( KxKt \)-dimensional parameter estimates \( \hat{\Theta} \) as well as the estimated response \( y.\text{vec}\text{Est} \) at the points \( \text{trep} \). For the large data sets, one can use \texttt{bam()} in place of \texttt{gam()}.

To estimate \( G(\cdot, \cdot) \) at specific time points, we use \texttt{fpca.sc} function of \texttt{refund} package in R. The model residuals are computed from \( \text{res.vec}=y.\text{vec}-y.\text{vec}\text{Est} \); FPCA is then applied to the residuals using the function \texttt{fpca.sc} of R package \texttt{refund} as

\[
\text{fpc} <- \text{fpca.sc(matrix(res.vec, nrow=n, ncol=m, byrow=TRUE), pve, var=TRUE}).
\]
Note that the residuals must be transformed into a matrix format in \texttt{fpca.sc}. The \texttt{pve} argument specifies the percent of variance explained by the first few eigencomponents such as \texttt{pve=0.9} or \texttt{pve=0.95}. We set \texttt{var=TRUE} to estimate the variance of measurement errors $\sigma^2$. This procedure offers the estimate of eigencomponents $\{\phi_k(\cdot), \lambda_k\}$ and the estimate of $\sigma^2$, which will be used to reconstruct $G = \text{Cov}(E_i)$. For the case where the data are observed on a sparse grid of points, the $n \times m$-dimensional matrix of the residuals contains NAs as components, considered as missing values. Nevertheless, the \texttt{fpca.sc()} function can still estimate the underlying smooth curves and the eigencomponents. One can use \texttt{fpca.face()} in place of \texttt{fpca.sc()} for the densely sampled design.
Appendix B

Additional Details for Chapter 3

This section comprises of two sections. Additional simulation results are reported in Appendix B.1. Additional investigation for the bike share data application is included in Appendix B.2.

B.1 Additional Simulation Results

B.1.1 Additional Simulations for Irregular and Sparse Design

We further investigate predictive performance for sparse design with a different level of sparseness. In this experiment, we generate the data using increased number of time points for $s$ and $t$; for convenience, we call this setting moderately sparse design. On the other hand, the simulations in Section 3.5 used smaller number of time points per curve; we continue to call this setting sparse design. For the response trajectories, $m_{Y,i} = 55 \sim 66$ points are randomly selected from $[0, 1]$ interval for each curve, while the simulations in Section 3.5 use only $m_{Y,i} = 35 \sim 44$ points for $t$. For covariate trajectories, we randomly select $m_{W,i} = 50 \sim 62$ points from $[0, 1]$ interval for each curve, while the settings in Section 3.5 use $m_{W,i} = 45 \sim 54$ points for $s$. 

111
Table B.1: Summary of (1) RMSPE in and (2) RMSPE out based on 1000 simulated data sets.

<table>
<thead>
<tr>
<th>n</th>
<th>method</th>
<th>F(x, s, t) = F_2(x, s, t), moderately sparse design</th>
<th>F(x, s, t) = F_3(x, s, t), moderately sparse design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>E_i = E_i^1</td>
<td>E_i = E_i^2</td>
</tr>
<tr>
<td>50</td>
<td>FLM</td>
<td>0.328 0.149</td>
<td>0.445 0.150</td>
</tr>
<tr>
<td></td>
<td>GAFFR</td>
<td>0.310 0.100</td>
<td>0.429 0.099</td>
</tr>
<tr>
<td>100</td>
<td>FLM</td>
<td>0.330 0.143</td>
<td>0.446 0.144</td>
</tr>
<tr>
<td></td>
<td>GAFFR</td>
<td>0.309 0.081</td>
<td>0.429 0.077</td>
</tr>
<tr>
<td>300</td>
<td>FLM</td>
<td>0.331 0.138</td>
<td>0.447 0.138</td>
</tr>
<tr>
<td></td>
<td>GAFFR</td>
<td>0.306 0.058</td>
<td>0.428 0.057</td>
</tr>
</tbody>
</table>

B.1.2 Sensitivity Analysis

We investigate the finite sample performance of our method using a different choice of $K$. Both prediction and inference performances are assessed, considering the cases where $F(x, s, t) = F_3(x, s, t)$ and $E_i = E^3$. For comparisons, two choices are considered for the selection of $K$. One is selected by a preset level $PVE=95\%$, and the other one is by $PVE=99\%$. Simulation results are presented in Table B.2 and Table B.3. Table B.2 summarizes the in-sample and out-of-sample RMSPE’s. It seems that the choice of $K$ slightly affects both the in-sample and out-of-sample prediction accuracy, but the overall predictive performance does not change significantly. Table B.3 summarizes the ACP of the point-wise PIs. The coverage performance is less affected by the choice of $K$ as well.

B.2 Further Investigation of Real Data Example

B.2.1 Histogram of Standardized Average Humidity

This section provides additional figures for the bike share data application illustrated in Section 3.6.1. Figure B.1 presents a histogram of standardized average humidity (left) and the estimated coefficient function $\hat{\gamma}(t)$ (right) obtained by fitting the GAFFR model.
Table B.2: Summary of (1) RMSPE_{in} and (2) RMSPE_{out} based on 1000 simulated data sets. Results are obtained by applying the GAFFR model.

<table>
<thead>
<tr>
<th>n</th>
<th>dense design</th>
<th>sparse design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PVE=95% (1)</td>
<td>PVE=99% (2)</td>
</tr>
<tr>
<td>50</td>
<td>0.374 0.201 0.369 0.194</td>
<td>0.387 0.229 0.382 0.221</td>
</tr>
<tr>
<td>100</td>
<td>0.372 0.161 0.366 0.150</td>
<td>0.386 0.186 0.380 0.176</td>
</tr>
<tr>
<td>300</td>
<td>0.370 0.131 0.364 0.116</td>
<td>0.380 0.143 0.375 0.129</td>
</tr>
</tbody>
</table>

Table B.3: Summary of ACP for the new response $Y_0(t) | X_0(.)$, i.e., ACP_p, at nominal significance levels $1 - \alpha = 0.85, 0.90, \text{and } 0.95$. Results are based on 1000 simulated data sets with 100 bootstrap replications per data.

<table>
<thead>
<tr>
<th>n</th>
<th>dense design</th>
<th>sparse design</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PVE=95%</td>
<td>PVE=99%</td>
</tr>
<tr>
<td>50</td>
<td>0.912 0.947 0.978</td>
<td>0.910 0.946 0.977</td>
</tr>
<tr>
<td>100</td>
<td>0.895 0.935 0.972</td>
<td>0.893 0.934 0.971</td>
</tr>
<tr>
<td>300</td>
<td>0.870 0.916 0.962</td>
<td>0.868 0.914 0.960</td>
</tr>
</tbody>
</table>

### B.2.2 Validation of $L^2$-Norm Based Test

In Section 3.6.1, a $L^2$-norm based testing procedure is performed to test the null hypothesis that the average humidity is a significant effect, i.e., $H_0: \gamma(t) = 0$; refer to Algorithm 2 in Section 3.6.1 for the details about the procedure. We carried out additional numerical study to validate the size and power performance of the test using simulated data sets which mimic the main feature of the bike share data.

The new simulation study generates the covariates $X_i(s)$ from a process in which the mean and the covariance functions equal their estimated counterparts from the data. We first apply the FPCA to the observed temperature, and obtain a smoothed version of 105 curves by computing $X_i^{\text{sim}}(s) = \tilde{\mu}_X(s) + \sum_{k=1}^{K} \xi ik \tilde{\phi}_{X,k}(s)$, where $\xi ik (i = 1, \ldots, 105)$ are normally distributed with zero-mean and variance $\tilde{\lambda}_{X,k}$. Here, $\tilde{\mu}_X(s)$ are the estimated mean of covariate, and $\{\tilde{\lambda}_{X,k}, \tilde{\phi}_{X,k}(s)\}_k$ are the estimated pairs of eigenvalues/eigenfunctions obtained from the spectral decomposition of the estimated covariance function of the covariate. The estimates $\tilde{\mu}_X(s)$ and $\{\tilde{\lambda}_{X,k}, \tilde{\phi}_{X,k}(s)\}_k$ are obtained from the observed data, and the truncation $K$ are determined by setting the PVE to 99%. Assuming that the covariate trajectories are observed
with noise, we generate noisy covariates $W^\text{sim}_{ik}$ by $W^\text{sim}_{ik} = X^\text{sim}_i(s) + \delta_{ik}$, where the noise $\delta_{ik}$ are normally distributed with zero-mean, and the noise variance is estimated from the original data.

To generate the response curves $Y_i(t)$, we first fit the GAFFR model using the original observed data, and obtain the estimated function $\hat{F}(x,s,t)$ and the estimated coefficient function $\hat{\gamma}(t)$ as described in Section 3.6.1. Then, we simulate the response curves based on the following model:

$$Y^\text{sim}_{i,d}(t) = \int_{\mathcal{T}_X} \hat{F}\{X^\text{sim}_i(s), s, t\} ds + d\{hum_i\hat{\gamma}(t)\} + \epsilon_i(t),$$

where $hum_i$ is the standardized average humidity obtained from the original data. To generate the random errors $\epsilon_i(t)$, we obtain residuals from the fitted model of the bike share data, and employ the FPCA methods to estimate the variance of random errors. When $d = 0$, the effect of average humidity does not exist, while when $0 < d \leq 1$, the model depends on the average humidity gradually. The size performance is investigated based on 3000 Monte Carlo simulations by setting $d = 0$. The power of the test is studied for $d = 0.25, 0.5, 0.75, 1$ based on 1000 Monte Carlo simulations. Simulations are based on $B = 200$ bootstrap replications.

The size of the test is assessed by computing rejection probabilities at 5%, 10%, and 15%
Table B.4: Summary of estimated rejection probabilities (×100). The values in the parenthesis are the estimated standard errors (×100) of the rejection probabilities.

<table>
<thead>
<tr>
<th>100α%</th>
<th>d = 0</th>
<th>d = 0.25</th>
<th>d = 0.5</th>
<th>d = 0.75</th>
<th>d = 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>5%</td>
<td>5.1(0.4)</td>
<td>8.9(0.5)</td>
<td>14.2(0.6)</td>
<td>24.5</td>
<td>71.4</td>
</tr>
</tbody>
</table>

nominal levels. We evaluated the power performance of the test for fixed nominal level \( \alpha = 5\% \); results are summarized in Table B.4. We observe that the estimated sizes are very close to the nominal levels. The power of the testing procedure increases as the values of \( d \) increase, indicating excellent power properties.
Appendix C

Additional Details for Chapter 4

C.1 Proof of Proposition 1

This section includes the proof of Proposition 1.

Proof. The proof of Proposition 1 proceeds by showing equivalences of all paired conditions.

In this setting, the tri-variate function $F(x, s, t)$ is twice differentiable on its domain $\mathbb{R} \times T_X \times T_Y$ for any $s \in T_X$ and for any $t \in T_Y$, and by the definition of partial derivative of $F(x, s, t)$ with respect to $x$ the equivalence of Condition (A1) and (A2) is automatic. The equivalence of Condition (A2) and (A3) is immediately by the property of $L^2$-norm.

Equalities in Condition (A3) follow from the result that

$$\left\| \frac{\partial^2 F(x, s, t)}{\partial x^2} \right\|^2 = \left\| \sum_k \phi_k \frac{\partial^2 G_k(x, s)}{\partial x^2} \right\|^2 = \sum_{k, k'} \left\{ \int \phi_k(t) \phi_k'(t) dt \right\} \left\{ \int \int \frac{\partial^2 G_k(x, s)}{\partial x^2} \frac{\partial^2 G_{k'}(x, s)}{\partial x^2} dxds \right\} = \sum_k \left\| \frac{\partial^2 G_k(x, s)}{\partial x^2} \right\|^2 = 0$$

for $\forall x$ and $\int \phi_k(t) \phi_{k'}(t) dt = I(k = k')$, where $I(\cdot)$ is the indicator function. Therefore,
\[ \left\| \frac{\partial^2 G_k(x,s)}{\partial x^2} \right\|^2 = 0 \] holds for all \( k \), and again by the property of \( L^2 \)-norm the equivalence of Condition (A3) and (A4) can be seen.

Finally, the equivalence of Condition (A4) and (A5) follows by the definition of partial derivative of \( G_k(x,s) \) with respect to \( x \), which ends the proof.

**C.2 Additional Simulations for Irregular and Sparse Design**

We report additional simulation results for the sparse sampling design with a smaller number of observations per curve. Table C.1 illustrates the size performance of the RLRTs for the setting where \( m_{Y,i} = 35 \sim 40 \) and \( m_{W,i} = 27 \sim 32 \), and these are randomly selected from [0,1] interval. To assess the size of the testing procedure, we fixed the nominal levels to \( \alpha = 5\% \) and \( 10\% \).

Table C.1: Type I error rates \((\times100)\) for sparse sampling design \((m_{W,i} \sim \{27, \ldots, 32\}, m_{Y,i} \sim \{35, \ldots, 40\})\) based on 2000 Monte Carlo simulations. The values in the parenthesis are the estimated standard errors \((\times100)\) of the rejection probabilities. Results are based on the RLRTs.

<table>
<thead>
<tr>
<th></th>
<th>Simple model: ( F(x,s,t) = F_1(x,s,t) )</th>
<th>Complex model: ( F(x,s,t) = F_2(x,s,t) )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( E_i = E_i^1 )</td>
<td>( E_i = E_i^2 )</td>
</tr>
<tr>
<td>( n ) ( \alpha = 5% ) ( \alpha = 10% ) ( \alpha = 5% ) ( \alpha = 10% ) ( \alpha = 5% ) ( \alpha = 10% ) ( \alpha = 5% ) ( \alpha = 10% ) ( \alpha = 5% ) ( \alpha = 10% )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>9.9(0.7) 16.8(0.8)</td>
<td>10.4(0.7) 17.9(0.9)</td>
</tr>
<tr>
<td>100</td>
<td>7.1(0.6) 13.0(0.8)</td>
<td>6.3(0.5) 11.8(0.7)</td>
</tr>
<tr>
<td>300</td>
<td>5.8(0.5) 12.1(0.7)</td>
<td>5.6(0.6) 11.5(0.7)</td>
</tr>
<tr>
<td>500</td>
<td>6.0(0.5) 10.6(0.7)</td>
<td>6.9(0.6) 12.8(0.7)</td>
</tr>
</tbody>
</table>