General Functional Concurrent Model

Janet S. Kim* Arnab Maity† Ana-Maria Staicu‡

June 17, 2016

Abstract

We propose a flexible regression model to study the association between a functional response and multiple functional covariates that are observed on the same domain. Specifically, we relate the mean of current response to current values of the covariates by sum of smooth unknown bi-variate functions, where each of the functions depends on the current value of the covariate and the time point itself. In this framework, 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. 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.

Keywords: Functional concurrent models; F-test; Non-linear models; Penalized B-splines; Prediction.

*PhD Student, Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695 (Email: jskim3@ncsu.edu)

†Associate Professor, Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695 (Email: amaity@ncsu.edu)

‡Associate Professor, Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695 (Email: ana-maria_staicu@ncsu.edu)

1 Introduction

Functional regression models where both the response and the covariate are functional have been long researched in the literature. Functional linear models 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) and Wu et al. (2010), among others. 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 (see e.g., Malfait and Ramsay 2003), 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. Another alternative is to assume a concurrent relationship, where the current response is modeled based on only the current value of the covariate function. Functional linear concurrent models (see e.g., Ramsay and Silverman 2005; Ramsay et al. 2009; Sentürk and Nguyen 2011) assume a linear relationship; 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. While the linear approach provides easy interpretation for the estimated coefficient function, it may not capture all the variability of the response in practical situations where the underlying relationship is complex.

In this article, we consider the functional concurrent model where we allow the relationship between the response and the covariate functions to be non-linear. 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 values of the covariates at that time point through smooth unknown bi-variate functions. Such formulation allows us to capture potential complex relationships between response and predictor functions as well as better capture out-of-sample prediction performance, as we will observe in our numerical investigation. Our model contains the standard 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 three main contributions. First, we propose a general non-linear functional concurrent model to describe complex association between two or more functional variables measured on the same domain. We model the relationship via unknown bi-variate functions, 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; McLean et al. 2014). 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) and Morris and Carroll (2006) who considered inference in functional mixed model framework. Reiss et al. (2010) proposed inference for the fixed effects parameters in function-on-scalar regression 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 correlation 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-step estimation procedure: 1) estimate the population level parameters using an independent error assumption; and 2) employ standard functional principal component analysis (FPCA) based methods (see e.g., Yao et al. 2005b; Zhang and Chen 2007; Goldsmith et al. 2013) to the residuals. The proposed inference uses the resulted estimate of the residual covariance.

Second, our model allows to incorporate multiple functional and/or scalar covariates, assuming that the effects of the covariates are additive. Specifically, we represent the effects of the covariates by sum of smooth bi-variate functions, each of which separately quantifies the dependence between the response and one of the covariates. The model involving a single functional covariate is a particular case of these models; functional single index models (FSIMs) by Jiang and Wang (2011) may also cover this type of

models, but they did not investigate these cases formally. They used an unknown bivariate link function to relate a functional/longitudinal response to multiple predictors but considered different estimation methodology. Furthermore, while it is possible to incorporate multiple functional covariates in their model framework, our model and the FSIM have different interpretation for the effects of the covariates; the link function in the FSIM accounts for an average effect of the covariates over time.

Third, we develop a testing procedure for assessing whether any of the functional covariates are related to the response variable. We discuss two particular testing problems: 1) testing the global effect of the functional covariate against a model involving a single functional covariate; and 2) testing the null hypothesis of no association between the response and a particular covariate against a model involving two functional covariates. 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 distribution 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 paper, 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.

2 General Functional Concurrent Model

In this section, we introduce our model framework for a functional response and functional covariates, develop an estimation procedure for unknown model components, and discuss prediction of a new trajectory. Our model and methodology are outlined for the setting involving two functional covariates; nevertheless, they can be extended straightforwardly to incorporate other vector-valued and/or functional covariates via a linear or smooth effect without much complication.

2.1 Modeling Framework

Suppose for $i = 1, \ldots, n$, we observe $\{(W_{1ij}, t_{1ij}) : j = 1, \ldots, m_{1i}\}$, $\{(W_{2ij}, t_{2ij}) : j = 1, \ldots, m_{2i}\}$ and $\{(Y_{ik}, t_{ik}) : k = 1, \ldots, m_{Y,i}\}$, where W_{1ij}'s and W_{2ij}'s denote the covariates observed at points t_{1ij} and t_{2ij}, respectively, and Y_{ik}'s are the response observed at t_{ik}. It is assumed that $t_{1ij}, t_{2ij}, t_{ik} \in \mathcal{T}$ for all i, j and k and $W_{qij} = X_{q,i}(t_{qij}) + \delta_{qij}$ for $q = 1, 2$ where $X_{q,i}(\cdot)$ 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 δ_{1ij} and δ_{2ij} are the iid measurement errors with mean zero and variances equal to τ_1^2 and τ_2^2, respectively. To illustrate ideas, we first consider $W_{1ij} = X_{1,i}(t_{1ij})$ and $W_{2ij} = X_{2,i}(t_{2ij})$, which is equivalent to $\tau_1^2 = \tau_2^2 = 0$. Furthermore, we assume that the response and the covariates are observed on a fine and regular grid of points, or equivalently $t_{1ij} = t_{1j}$, $t_{2ij} = t_{2j}$, $t_{ik} = t_k$ and $m_{1i} = m_{2i} = m_{Y,i} = m$. 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 $\tau_q^2 > 0$ and different sampling designs for $X_{q,i}$'s and Y_i's are discussed in Section 4.

Assume that $X_{1,i}(\cdot) \overset{iid}{\sim} X_1(\cdot)$ and $X_{2,i}(\cdot) \overset{iid}{\sim} X_2(\cdot)$, where $X_1(\cdot)$ and $X_2(\cdot)$ are some underlying random processes. We introduce the following general functional concurrent model (GFCM)

$$Y_i(t) = \mu_Y(t) + F_1\{X_{1,i}(t), t\} + F_2\{X_{2,i}(t), t\} + \epsilon_i(t), \tag{1}$$

where $\mu_Y(t)$ is an unknown and smooth intercept function, F_1 and F_2 are smooth unknown bi-variate functions defined on $\mathbb{R} \times \mathcal{T}$, and $\epsilon_i(\cdot)$ is an error process independent of the predictors $X_{1,i}(\cdot)$ and $X_{2,i}(\cdot)$. The error process $\epsilon_i(\cdot)$ is assumed to have mean

zero and unknown autocovariance function $G(\cdot,\cdot)$. For identifiability of the two functions F_1 and F_2, we assume that $\mathrm{E}\left[F_1\{X_1(t),t\}|X_1(t)\right]=0$ and $\mathrm{E}\left[F_2\{X_2(t),t\}|X_2(t)\right]=0$ for any $t\in\mathcal{T}$, and thus $\mu_Y(t)$ is the marginal mean function of the response. We introduce two main innovations in (1). First, the general bi-variate functions $F_q(\cdot,\cdot)$ allow us to model potentially complicated relationships between $Y(\cdot)$ and $X_q(\cdot)$ for $q=1,2$, and extends the effect of the covariate beyond linearity. Second, the covariance structure for the residual process $\epsilon_i(s)$ is assumed unknown.

An important advantage of our framework is that it can easily accommodate multiple functional and scalar predictors, and we introduce several examples. When the setting involves a single functional covariate $X_{1,i}(t)$, the form of GFCM is $Y_i(t)=F_1\{X_{1,i}(t),t\}+\epsilon_i(t)$. The standard linear functional concurrent model is a special case of this model with $F_1(x,t)=\beta_0(t)+x\beta_1(t)$, where $\beta_0(\cdot)$ and $\beta_1(\cdot)$ are unknown parameter functions. When the setting involves a single functional covariate $X_{1,i}(t)\overset{iid}{\sim}X_1(t)$ and a scalar predictor $X_{2,i}\overset{iid}{\sim}X_2$, two approaches are possible to incorporate the scalar predictor: in a linear way and in a non-linear way. The linear effect of the scalar covariate $X_{2,i}$ can be modeled through $Y_i(t)=\mu_Y(t)+F_1\{X_{1,i}(t),t\}+X_{2,i}\beta(t)+\epsilon_i(t)$, where $\beta(t)$ is an unknown coefficient function that quantifies the effect of the scalar covariate. For identifiability we require $\mathrm{E}\left[F_1\{X_1(t),t\}|X_1(t)\right]=0$ in this model. The smooth non-linear effect of the scalar covariate $X_{2,i}$ can be modeled through $Y_i(t)=\mu_Y(t)+F_1\{X_{1,i}(t),t\}+F_2(X_{2,i},t)+\epsilon_i(t)$, where as before F_1 and F_2 are unknown and smooth bi-variate functions; here too it is assumed that $E[F_1\{X_1(t),t\}|X_1(t)]=0$ and $E[F_2(X_2,t)|X_2]=0$. Accommodating linear or non-linear effects of vector-valued covariates are similar.

2.2 Estimation

We focus on model (1) and describe estimation of the marginal mean function $\mu_Y(t)$ and the bi-variate surfaces $F_1(\cdot,\cdot)$ and $F_2(\cdot,\cdot)$. We model the mean function $\mu_Y(t)$ by spline-based estimation methodology which represents the smooth effect by a linear combination of univariate B-spline basis functions. Let $\{B_{\mu,d}(t)\}_{d=1}^{K_\mu}$ be a set of univariate B-spline basis functions defined on [0,1], where K_μ is the basis dimension. Using the basis functions, we can write $\mu_Y(t)=\sum_{d=1}^{K_\mu}B_{\mu,d}(t)\theta_{\mu,d}=\mathbb{B}_\mu(t)\Theta_\mu$, where $\mathbb{B}_\mu(t)$ is the K_μ-dimensional vector of $B_{\mu,d}(t)$'s, and Θ_μ is the vector of unknown parameters

$\theta_{\mu,d}$'s. We also model $F_1(\cdot,\cdot)$ and $F_2(\cdot,\cdot)$ using bi-variate basis expansion using tensor product of univariate B-spline basis functions (Marx and Eilers 2005; Wood 2006; McLean et al. 2014; Scheipl et al. 2015). For $q = 1, 2$ let $\{B_{Xq,k}(x)\}_{k=1}^{K_{xq}}$ and $\{B_{Tq,l}(t)\}_{l=1}^{K_{tq}}$ be the B-spline basis functions for x and t, respectively, used to model $F_q(x,t)$. Then, $F_1(\cdot,\cdot)$ can be represented as $F_1\{X_{1,i}(t),t\} = \sum_{k=1}^{K_{x1}}\sum_{l=1}^{K_{t1}} B_{X1,k}\{X_{1,i}(t)\}B_{T1,l}(t)\theta_{1,k,l} = \mathbb{Z}_{1,i}(t)\Theta_1$, where $\mathbb{Z}_{1,i}(t)$ is the $K_{x1}K_{t1}$-dimensional vector of $B_{X1,k}\{X_{1,i}(t)\}B_{T1,l}(t)$'s, and Θ_1 is the vector of unknown parameters, $\theta_{1,k,l}$'s. Similarly, we write $F_2\{X_{2,i}(t),t\} = \sum_{k=1}^{K_{x2}}\sum_{l=1}^{K_{t2}} B_{X2,k}\{X_{2,i}(t)\}B_{T2,l}(t)\theta_{2,k,l} = \mathbb{Z}_{2,i}(t)\Theta_2$, where $\mathbb{Z}_{2,i}(t)$ is the $K_{x2}K_{t2}$-dimensional vector of $B_{X2,k}\{X_{2,i}(t)\}B_{T2,l}(t)$'s, and Θ_2 is the vector of unknown parameters, $\theta_{2,k,l}$'s. Based on the above expansions, model (1) can be written as

$$Y_i(t) = \mathbb{B}_\mu(t)\Theta_\mu + \mathbb{Z}_{1,i}(t)\Theta_1 + \mathbb{Z}_{2,i}(t)\Theta_2 + \epsilon_i(t), \tag{2}$$

In this representation, 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.

To prevent overfitting, we propose to estimate Θ_μ, Θ_1 and Θ_2 by minimizing a penalized criterion $\sum_{i=1}^n ||Y_i(\cdot) - \mathbb{B}_\mu(\cdot)\Theta_\mu - \mathbb{Z}_{1,i}(\cdot)\Theta_1 - \mathbb{Z}_{2,i}(\cdot)\Theta_2||^2 + \Theta_\mu^T P_\mu \Theta_\mu + \Theta_1^T P_1 \Theta_1 + \Theta_2^T P_2 \Theta_2$, where P_μ, P_1 and P_2 are the penalty matrices for smoothness of $\mu_Y(t)$, $F_1(x,t)$ and $F_2(x,t)$, respectively, and contain penalty parameters that regularize the trade-off between the goodness of fit and the smoothness of fit. The notation $||\cdot||^2$ is the usual L^2-norm corresponding to the inner product $< f, g >= \int fg$. In practice, we observe $Y_i(\cdot)$, $X_{1,i}(\cdot)$ and $X_{2,i}(\cdot)$ 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

$$\sum_{i=1}^n \sum_{j=1}^m \{Y_i(t_j) - \mathbb{B}_\mu(t_j)\Theta_\mu - \mathbb{Z}_{1,i}(t_j)\Theta_1 - \mathbb{Z}_{2,i}(t_j)\Theta_2\}^2/m + \Theta_\mu^T P_\mu \Theta_\mu + \Theta_1^T P_1 \Theta_1 + \Theta_2^T P_2 \Theta_2. \tag{3}$$

Here, P_μ is given by $P_\mu = \lambda_\mu D_\mu^T D_\mu$, where D_μ represents the second order difference penalty (Eilers and Marx 1996; Marx and Eilers 2005; McLean et al. 2014), and λ_μ is its penalty parameter. Also, P_q $(q = 1, 2)$ is given by $P_q = \lambda_{xq} D_{xq}^T D_{xq} \bigotimes I_{K_{tq}} +$

$\lambda_{tq} I_{K_{xq}} \bigotimes D_{tq}^T D_{tq}$, where the notation $\bigotimes$ stands for the Kronecker product, I_K is the identity matrix with dimension K, and D_{xq} and D_{tq} are matrices representing the row and column of second order difference penalties. The penalty parameters λ_{xq} and λ_{tq} control the roughness of the function in directions x and t, respectively.

An explicit form of the estimators $\widehat{\Theta}_\mu$, $\widehat{\Theta}_1$ and $\widehat{\Theta}_2$ is readily available for fixed values of the penalty parameters. Define the m-dimensional vector of response $\mathbb{Y}_i = [Y_i(t_1), \ldots, Y_i(t_m)]^T$ and the m-dimensional vector of errors $\mathbb{E}_i = [\epsilon_i(t_1), \ldots, \epsilon_i(t_m)]^T$. Also, we define $\widetilde{\mathbb{B}}_\mu$ as $m \times K_\mu$-dimensional matrix with the j-th row given by $\mathbb{B}_\mu(t_j)$ and $\mathcal{Z}_{q,i}$ as $m \times K_{xq}K_{tq}$-dimensional matrix with the j-th row given by $\mathbb{Z}_{q,i}(t_j)$ $(q = 1, 2)$. For simplicity, denote $\mathcal{Z}_i = [\widetilde{\mathbb{B}}_\mu | \mathcal{Z}_{1,i} | \mathcal{Z}_{2,i}]$, $\Theta^T = [\Theta_\mu, \Theta_1, \Theta_2]^T$ and $\mathbb{P} = \mathrm{diag}(P_\mu, P_1, P_2)$. Then the solution of Θ is calculated as

$$\widehat{\Theta} = H\{\textstyle\sum_{i=1}^n \mathcal{Z}_i^T \mathbb{Y}_i\}, \tag{4}$$

where $H = \{\sum_{i=1}^n \mathcal{Z}_i^T \mathcal{Z}_i + \mathbb{P}\}^{-1}$. The penalty parameters can be chosen based on some appropriate criteria such as generalized cross validation (GCV) (Ruppert et al. 2003; Wood 2006) or restricted maximum likelihood (REML) (Ruppert et al. 2003; Wood 2006). Estimation under (3) can be fully implemented in `R` using functions of the `mgcv` package (Wood 2015). Then, given $X_{1,i}(t)$ and $X_{2,i}(t)$, the response curve can be estimated by $\widehat{Y}_i(t) = \mathbb{B}_\mu(t)\widehat{\Theta}_\mu + \mathbb{Z}_{1,i}(t)\widehat{\Theta}_1 + \mathbb{Z}_{2,i}(t)\widehat{\Theta}_2$. Furthermore, one can estimate the marginal mean of response by $\widehat{\mu}_Y(t) = \mathbb{B}_\mu(t)\widehat{\Theta}_\mu$. Modification of the estimation procedure for the case where the grid of points is irregular and sparse is presented in Appendix A of the Supplementary Materials.

2.3 Variance Estimation

The penalized criterion (3) is based on working independence assumption and thus does not account for the possibly correlated error process. For valid inference about Θ, one needs to account for the dependence of the residuals when deriving the variance of $\widehat{\Theta}$. The variance of the parameter estimate $\widehat{\Theta}$ can be calculated as $\mathrm{var}(\widehat{\Theta}) = H\{\sum_{i=1}^n \mathcal{Z}_i^T \mathbb{G} \mathcal{Z}_i\} H^T$, where $\mathbb{G} = \mathrm{cov}(\mathbb{E}_i) = [G(t_j, t_k)]_{1 \le j,k \le m}$ is the $m \times m$ covariance matrix evaluated corresponding to the observed time points. We model the nontrivial dependence of the errors

process $\epsilon(t)$ assuming that the error process has the form $\epsilon(t) = \epsilon_S(t) + \epsilon_{WN}(t)$, where ϵ_S is a zero-mean smooth stochastic process, and $\epsilon_{WN}(t)$ is a zero-mean white noise measurement error with variance σ^2. Let $\Sigma(s,t)$ be the autocovariance function of ϵ_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 Σ 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 $e_{ij} = Y_i(t_j) - \widehat{Y}_i(t_j)$ from the model fit, and employ FPCA methods (e.g., Yao et al. 2005a; Zhang and Chen 2007) to estimate $\phi_k(\cdot)$, λ_k, and σ^2. Specifically, we obtain an initial smooth estimate of the covariance Σ, remove the negative eigenvalues, and obtain a final estimate of Σ, $\widehat{\Sigma}(s,t) = \sum_{k=1}^{K} \widehat{\phi}_k(s)\widehat{\phi}_k(t)\widehat{\lambda}_k$, where $\{\widehat{\phi}_k(\cdot), \widehat{\lambda}_k\}$ are the eigenfunctions/eigenvalues of the estimated covariance $\widehat{\Sigma}(s,t)$ with $\widehat{\lambda}_1 > \widehat{\lambda}_2 > \ldots > \widehat{\lambda}_K > 0$, and K is the number of eigencomponents used in the estimation. Then, we estimate G by $\widehat{G}(s,t) \approx \sum_{k=1}^{K} \widehat{\lambda}_k \widehat{\phi}_k(s)\widehat{\phi}_k(t) + \widehat{\sigma}^2 I(s = t)$ for any $s, t \in [0,1]$, where $\widehat{\sigma}^2$ is the estimated error variance. The finite truncation K is typically chosen by setting the percent of variance explained (PVE) by the first few eigencomponents to some pre-specified value, such as 90% or 95%.

2.4 Prediction

A main focus in this paper 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 $X_{10}(t_j)$ and $X_{20}(t_j)$ $(j = 1, \ldots, m)$ are given. We assume that the model $Y_0(t) = \mu_Y(t) + F_1\{X_{10}(t), t\} + F_2\{X_{20}(t), t\} + \epsilon_0(t)$ still holds for the new data, where the error process $\epsilon_0(t)$ has the same distributional assumption as $\epsilon_i(t)$ in (1) and is independent of the new covariates $X_{10}(t)$ and $X_{20}(t)$. We predict the new response by $\widehat{Y}_0(t) = \sum_{d=1}^{K_\mu} B_{\mu,d}(t)\widehat{\theta}_{\mu,d} + \sum_{k=1}^{K_{x1}}\sum_{l=1}^{K_{t1}} B_{X1,k}\{X_{10}(t)\}B_{T1,l}(t)\widehat{\theta}_{1,k,l} + \sum_{k=1}^{K_{x2}}\sum_{l=1}^{K_{t2}} B_{X2,k}\{X_{20}(t)\}B_{T2,l}(t)\widehat{\theta}_{2,k,l}$, where $\widehat{\theta}_{\mu,d}$, $\widehat{\theta}_{1,k,l}$ and $\widehat{\theta}_{2,k,l}$ are estimated based on (4).

Uncertainty in the prediction depends on how small the difference is between the predicted response $\widehat{Y}_0(t)$ 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 covariates $X_{10}(t)$ and $X_{20}(t)$, we have $\text{var}\{Y_0(t) - \widehat{Y}_0(t)\} = \text{var}\{\epsilon_0(t)\} + \text{var}\{\widehat{Y}_0(t)\}$. Note that $\epsilon_0(t)$ is a realization of the same error process with zero-mean and covariance structure $G(\cdot, \cdot)$. Let $\mathbb{Z}_{10}(t)$ be the $K_{x1}K_{t1}$-dimensional vector of $B_{X1,k}\{X_{10}(t)\}B_{T1,l}(t)$'s, and $\mathbb{Z}_{20}(t)$ be the $K_{x2}K_{t2}$-dimensional vectors of $B_{X2,k}\{X_{20}(t)\}B_{T2,l}(t)$'s as defined earlier. Also, let $\mathbb{Y}_0$ and $\hat{\mathbb{Y}}_0$ be the m-dimensional vector of $Y_0(t_j)$'s and $\widehat{Y}_0(t_j)$'s respectively. Then the prediction variance becomes $\text{var}\{\mathbb{Y}_0 - \widehat{\mathbb{Y}}_0\} = \mathbb{G} + \mathcal{Z}_0 H\{\sum_{i=1}^n \mathcal{Z}_i^T \mathbb{G} \mathcal{Z}_i\} H^T \mathcal{Z}_0^T$, where $\mathcal{Z}_0 = [\widetilde{\mathbb{B}}_\mu | \mathcal{Z}_{10} | \mathcal{Z}_{20}]$ is $m \times (K_\mu + K_{x1}K_{t1} + K_{x2}K_{t2})$-dimensional matrix with the jth row given by $[\mathbb{B}_\mu(t_j) | \mathbb{Z}_{10}(t_j) | \mathbb{Z}_{20}(t_j)]$. 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 $C_{1-\alpha}(t) = \widehat{Y}_0(t) \pm \Phi^{-1}(1-\alpha/2)\left[\widehat{\text{var}}\{Y_0(t) - \widehat{Y}_0(t)\}\right]^{1/2}$ where $\Phi(\cdot)$ is the standard Gaussian cumulative distribution function. In Section 4, we provide details about performing prediction in the more general case when the new covariates $X_{10}(\cdot)$ and $X_{20}(\cdot)$ are 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 in Appendix B of the Supplementary Materials.

3 Hypothesis Testing

In many situations, testing for association between the response and the 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. To illustrate ideas, two particular cases of the proposed framework are considered: 1) testing the no effect of the covariate against a model involving a single functional covariate (see Section 3.1); and 2) testing the significance of a particular covariate against a model involving two functional covariates (see Section 3.2).

3.1 Testing of Global effect

Our focus in this section is testing the global effect of the functional covariate. Specifically, we want to test

$$H_0 : E[Y(t)|X_1(t) = x_1] = F_0(t) \ \ \forall x_1 \ \ \text{versus} \ \ H_1 : E[Y(t)|X_1(t) = x_1] = F_1(x_1, t), \quad (5)$$

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 $\mathcal{T}$, but $\mathbb{B}_X = \{B_{X,0}(x) =$

$1, B_{X,l}(x), l \geq 1\}$ for x, where $B_{X1,k}(x_1)$ $(k \geq 1)$ are the B-splines defined over $\mathcal{X}$ - the image of the process X_1, where $X_{1,i}(\cdot) \sim X_1(\cdot)$. Under the full model, we can write $F_1(x_1, t) = F_0(t) + \sum_{k=1}^{\infty} \sum_{l=0}^{\infty} B_{X1,k}(x_1) B_{T,l}(t) \theta_{k,l}$. We propose to use the following F-type test statistic

$$T_n = \frac{(RSS_0 - RSS_1)/(df_1 - df_0)}{RSS_1/(N - df_1)}, \tag{6}$$

where RSS_0 and df_0 are the residual sums of squares and the effective degrees-of-freedom (Wood 2006) 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 (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 algorithm:

Algorithm 1 Bootstrap algorithm for testing the global effect

1: Fit the full model described by the alternative hypothesis in (5) using the estimation procedure of the GFCM. Calculate the residuals $e_i(t_j) = Y_i(t_j) - \widehat{Y}_i(t_j)$ for all i and j.
2: Fit the null model described by the null hypothesis in (5) using the estimation procedure of the GFCM and estimate $F_0(t)$, $\widehat{F}_0(t)$.
3: Calculate the value of the test statistic in (6) based on the null and the full model fits; call this value $T_{n,\text{obs}}$.
4: Resample B sets of bootstrap residuals $\mathbb{E}_b^*(t) = \{e_{b,i}^*(t)\}_{i=1}^n$ $(b = 1, \ldots, B)$ with replacement from the residuals $\{e_i(t)\}_{i=1}^n$ obtained in step 1.
5: **for** $b = 1$ **to** B **do**
6: Generate response curves under the null model as $Y_{b,i}^* = \widehat{F}_0(t) + e_{b,i}^*(t)$.
7: Given the bootstrap data set $\{X_{1,i}(t), Y_{b,i}^*(t)\}_{i=1}^n$, fit the null and the full models and evaluate the test statistic in (6), T_b^*.
8: **end for**
9: Compute the p-value by $\widehat{p} = \sum_{b=1}^{B} I\{T_b^* \geq T_{n,\text{obs}}\}/B$.

In Algorithm 1, the test statistics $\{T_b^*\}_{b=1}^B$ obtained from each of the bootstrap samples can be viewed as realizations from the distribution of T_n under the assumption that H_0 is true.

3.2 Testing of Inclusion

In the context of multiple predictors, one might be interested to know which of the predictors are related to the response variable. Suppose we want to test

$$H_0 : E[Y(t)|X_1(t) = x_1, X_2(t) = x_2] = \mu_Y(t) + F_1(x_1, t) \quad \forall x_1, x_2 \tag{7}$$

versus the alternative $H_1 : E[Y(t)|X_1(t) = x_1, X_2(t) = x_2] = \mu_Y(t) + F_1(x_1, t) + F_2(x_2, t)$, where $F_1(x_1, t)$ and $F_2(x_2, t)$ are bi-variate functions assumed unknown. For simplicity, denote by $F_0(x_1, t) = \mu_Y(t) + F_1(x_1, t)$. To test the null hypothesis in (7), we represent the null and the full models using B-spline basis functions, and follow a similar logic used in the previous section. Specifically, we propose to use $\mathbb{B}_{X1} = \{1, B_{X1,k}(x_1) : k \geq 1\}$ and $\mathbb{B}_T = \{B_{T,l}(t) : l \geq 1\}$ for x_1 and t to model $F_0(x_1, t)$ under the null hypothesis; here, $B_{X1,k}(x_1)$ $(k \geq 1)$ are the B-splines defined over the image of the process X_1, and $B_{T,l}(t)$ $(l \geq 1)$ are the B-splines defined over $\mathcal{T}$. To formulate the full model, we use the same set of basis functions defined over the domain $\mathcal{T}$, but $\mathbb{B}_{X12} = \{1, B_{X1,k}(x_1), B_{X2,k}(x_2) : k \geq 1\}$ for x_1 and x_2, where $B_{X2,k}(x_2)$ $(k \geq 1)$ are defined over the image of the process X_2. Therefore, under the alternative, we can write $E[Y(t)|X_1(t) = x_1, X_2(t) = x_2] = F_0(x_1, t) + \sum_{k=1}^{\infty} \sum_{l=0}^{\infty} B_{X2,k}(x_2) B_{T,l}(t) \theta_{2,k,l}$.

The null hypothesis in (7) can be tested using Algorithm 2 with the test statistic in (6).

Algorithm 2 Bootstrap algorithm for testing significance

1: Fit the full model described by the alternative hypothesis using the estimation procedure of the GFCM. Calculate the residuals $e_i(t_j) = Y_i(t_j) - \widehat{Y}_i(t_j)$ for all i and j.
2: Fit the null model described by the null hypothesis in (7) using the estimation procedure of the GFCM and estimate $F_0(x_1, t)$, $\widehat{F}_0(x_1, t)$.
3: Calculate the value of the test statistic in (6) based on the null and the full model fits; call this value $T_{n,\text{obs}}$.
4: Resample B sets of bootstrap residuals $\mathbb{E}_b^*(t) = \{e_{b,i}^*(t)\}_{i=1}^n$ $(b = 1, \ldots, B)$ with replacement from the residuals $\{e_i(t)\}_{i=1}^n$ obtained in step 1.
5: **for** $b = 1$ **to** B **do**
6: Generate response curves under the null model as $Y_{b,i}^* = \widehat{F}_0(x_1, t) + e_{b,i}^*(t)$.
7: Given the bootstrap data set $\{X_{1,i}(t), X_{2,i}(t), Y_{b,i}^*(t)\}_{i=1}^n$, fit the null and the full models and evaluate the test statistic in (6), T_b^*.
8: **end for**
9: Compute the p-value by $\widehat{p} = \sum_{b=1}^{B} I\{T_b^* \geq T_{n,\text{obs}}\}/B$.

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. Testing the significance of $F_1(x_1, t)$ against the full model can also be performed similarly.

4 Extensions

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 δ_{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 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_i for each subject is small, but $\bigcup_{i=1}^{n} \{t_{ij}\}_{j=1}^{m_i}$ is dense in [0,1]. Reconstructing the latent trajectories $X_i(\cdot)$ is based on employing FPCA for sparse design (Yao et al. 2005a) to the observed W_{ij}'s. Yao et al. (2005a) proposed to 1) estimate the mean and covariance functions using local linear smoothers; 2) estimate eigenvalues/eigenfunctions from spectral decomposition of

the estimated covariance; and 3) predict FPC scores via conditional expectation. Then, the latent trajectories are predicted using a finite Karhunen-Loève (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 $X_0(t_j)$.

5 Numerical Study

In this section, we investigate the finite sample performance of our proposed methodology. Prediction accuracy is studied in Section 5.2.1, and testing performance is presented in Section 5.2.2. Finally in Section 5.3 we apply the proposed method to the gait study (Huang et al. 2015; Olshen et al. 1989; Ramsay and Silverman 2005) and the dietary calcium absorption study (Davis 2002; Sentürk and Nguyen 2011).

5.1 Simulation Design

We generate 1000 samples for the three simulation scenarios: (A) $Y_i(t) = F_1\{X_{1,i}(t), t\} + \epsilon_i(t)$ where $F_1(x_1, t) = 1 + x_1 + t$; (B) $Y_i(t) = F_1\{X_{1,i}(t), t\} + \epsilon_i(t)$ where $F_1(x_1, t) = 1 + x_1 + t + 2x_1^2 t$; and (C) $Y_i(t) = F_1\{X_{1,i}(t), t\} + F_2\{X_{2,i}(t), t\} + \epsilon_i(t)$ where $F_1(x_1, t) = 1 + x_1 + t + 2x_1^2 t$ and $F_2(x_2, t) = 0.75 \exp(x_2 t)$. In all experiments, the true covariates are given

by $X_q(t) = a_{q0} + a_{q1}\sqrt{2}\sin(\pi t) + a_{q2}\sqrt{2}\cos(\pi t)$, where $a_{q0} \sim N(0, \{2^{-0.5(q-1)}\}^2)$, $a_{q1} \sim N(0, \{0.85 \times 2^{-0.5(q-1)}\}^2)$ and $a_{q2} \sim N(0, \{0.70 \times 2^{-0.5(q-1)}\}^2)$ for $q = 1, 2$. Throughout the study, it is assumed that the covariates $X_{1,i}(t)$ and $X_{2,i}(t)$ are not observed directly. Instead we observe $W_{1i} = X_{1,i}(t) + \text{WN}(0, 0.6^2)$ and $W_{2i} = X_{2,i}(t) + \text{WN}(0, 0.6^2)$. For each of the above scenarios, the response $Y_i(\cdot)$ is generated under all possible combinations of the following factors:

- Error process $\mathbb{E}_i = [\epsilon_i(t_{i1}), \ldots, \epsilon_i(t_{im_{Y,i}})]^T$: (i) $\mathbb{E}_i^1 \sim N(0, 0.9^2 I_{m_i})$; (ii) $\mathbb{E}_i^2 \sim N(0, 0.9^2\Sigma) + N(0, 0.9^2 I_{m_i})$ where Σ has $AR_{0.2}(1)$ structure; and (iii) $\mathbb{E}_i^3 \sim \xi_{i1}\sqrt{2}\cos(\pi t) + \xi_{i2}\sqrt{2}\sin(\pi t) + N(0, 0.9^2 I_{m_i})$, where $\xi_{i1} \overset{iid}{\sim} N(0, 2)$ and $\xi_{i2} \overset{iid}{\sim} N(0, 0.75^2)$
- Sampling design: (i) dense: $m = 81$ equidistant time points in [0,1] for all i; and (ii) sparse: $m_i \overset{iid}{\sim} Uniform(20, 31)$ points for response and covariates
- $n \in 100, 300$

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

5.2 Simulation Results

5.2.1 Prediction Performance

Our performance measure includes in-sample or out-of-sample root mean squared prediction error (RMSPE^{in} and $\text{RMSPE}^{\text{out}}$), integrated coverage probability (ICP) and integrated length (IL) of prediction bands. The definition of these measures is given in the Supplementary Materials, Appendix C.

Result from Scenario A and B: For these two scenarios, we fit both the standard linear functional concurrent model (FCM) and our proposed GFCM for each of these choices. Our objective is two-fold: (i) When the true model is non-linear, i.e. $F_1(x_1, t) = 1 + x_1 + t + 2x_1^2 t$, we expect to see that the GFCM to perform better than FCM; and (ii) If the true model is in fact linear, i.e. $F_1(x_1, t) = 1 + x_1 + t$, we expect the GFCM still maintains prediction accuracy relatively to the FCM. We obtained both the GFCM and the linear FCM fits using 7 cubic B-splines for x_1 and t. For estimation of the residual covariance, the PVE is set equal to 99%.

Table 1 summarizes the predictive performance of our method for these two scenarios. We first discuss the case when the true function is non-linear (the top two panels in Table 1). In this case, we observe that RMSPE^{in} and $\text{RMSPE}^{\text{out}}$ from fitting the GFCM are smaller than those from the linear FCM in all cases, indicating that the proposed GFCM outperforms the linear FCM. 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(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 $\mathbb{E}_i^1$ (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 1 show the results when the underlying model is linear. In this case, the GFCM continues to show very good prediction performance; the results are almost identical to the ones yielded by the linear FCM irrespective of the number of subjects, sparseness of the sampling, and the error covariance structure.

To aid understanding these results, Figure 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.

Result from Scenario C: Now consider the setting involving two functional covariates. In this experiment, we investigate both in-sample and out-of-sample prediction performance as well as computational complexity of our algorithm with respect to the number of additional covariates. It is worth noting that the computational complexity can 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

readily 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. There are two available implementation tools to obtain predictions: `gam` and `bam` functions of `mgcv` (Wood 2015) `R` package. Both functions provide almost identical model fits, but computational advantage can be gained when using the `bam` function with very large data sets. For completeness, we report simulation results for Scenario B (the model involving a single functional covariate) and C.

We obtained the GFCM fits using 7 cubic B-splines for x_1, x_2 and t to model the smooth functions $F_1(x_1, t)$ and $F_2(x_2, t)$. For estimation of the residual covariance, we preset PVE = 99%. Table 2 summarizes RMSPE^{in}, $\text{RMSPE}^{\text{out}}$, ICP's at the nominal levels of 0.95, 0.90, and 0.85, and average computation time (in seconds). The computation time is measured on a 2.3GHz AMD Opteron Processor and averaged over 1000 Monte Carlo replications. The top and bottom post display the results corresponding to Scenario B and C, respectively. We observe from the results that both in-sample and out-of-sample predictive accuracy are still maintained irrespective of the number of predictive variables. We also observe that the coverage (indicated by ICP) is fairly close to the nominal levels in all cases. In terms of computational cost, both `gam` and `bam` slightly increase the average computation time with the increased sample size and model complexity. Nevertheless, the additional computation expense is rather minimal compared to the computational time corresponding to single functional covariate. The results also show that computations can be further sped up if `bam` is used.

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 nonlinear; GFCM has similar prediction performance relative to the linear FCM, when the true model is linear. Furthermore, in the presence of two functional covariates, our method still preserves high predictive accuracy, and the additional computation cost is not substantial compared to the case of single functional covariate. Finally, although results are not shown, we found that the above prediction performance is not sensitive to the number of basis functions selected.

In the Supplementary Materials, we include additional simulation results. Appendix

D.1 reports simulation results corresponding to another level of sparseness. Appendix D.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 a fast covariance smoothing method (Xiao et al. 2016) which is implemented using `fpca.face` function of the `refund` `R` package (Huang et al. 2015).

5.2.2 Testing Performance

Now we assess the performance of the proposed testing procedure for the following two scenarios: (A) $E[Y_d(t)|X_1(t) = x_1] = 1 + 2t + t^2 + d(x_1 t/8)$; and (B) $E[Y_d(t)|X_1(t) = x_1, X_2(t) = x_2] = 2t + t^2 + x_1 \sin(\pi t)/4 + d\{2\cos(x_2 t)\}$. We generate the data corresponding to the above true models using the error covariance structure and sampling design defined in Section 5.1. In Scenario (A), 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_1 and t. Thus the parameter d indexes the departure from the null hypothesis given in (5). In Scenario (B), the parameter $d \geq 0$ controls the departure from the null hypothesis given in (7). For each of the scenarios, type I error of the test is investigated by setting $d = 0$, and the power of the test is studied for positive values of d. We generated 2000 samples to assess the type I error rate, and 1000 samples to assess the power. The distribution of the test statistic in (6) is approximated using $B = 200$ bootstrap samples for each simulation.

Result from Scenario A: We first examine the size and power performance of the global test presented in Algorithm 1. The empirical type I error rates are evaluated at nominal levels of $\alpha = 5\%$ and 10% for sample sizes 100 and 300, and the estimated rejection probabilities are presented in top panel of Table 3. We observe that estimated type I error rates are mostly within 2 standard errors of the nominal values, and the larger sample size ($n = 300$) improves the size performance. The results also indicate that the performance is similar across different covariance structures and sampling designs. The power performance of our proposed test is evaluated for a fixed nominal level $\alpha = 5\%$. The top panel in Figure 2 displays the rejection probabilities for $d = 0.1 \sim 7$ and for different error covariance structure; in the interest of space, we only present the case of

sparse design. As expected, the power of the testing procedure increases with the sample size, while the results are 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. Power curves for densely sampled scenario are provided in the Supplementary materials, Appendix D.3. We found that the results for densely sampled data are very similar to the one obtained for sparsely sampled data.

Result from Scenario B: We now assess the size and power properties of the significance testing presented in Algorithm 2. The bottom panel of Table 3 shows the empirical type I error rates corresponding to nominal levels $\alpha = 5\%$ and 10% for sample sizes 100 and 300. The results show that the estimated sizes are mostly within 2 standard errors of the nominal values for all significance levels, and the results are less affected by different error covariance structure and sampling designs. The bottom panel in Figure 2 displays the rejection probabilities at the $\alpha = 0.05$ level for $d = 0.1 \sim 7$ and for different error covariance structure; again, we only present the case corresponding to sparse design. For the moderate sample size ($n = 100$), the power of the testing procedure increases with the value of d. For the larger sample size ($n = 300$) the empirical rejection probabilities converge to 1 at a fast rate as the value of d increases. As expected, there is some loss of power for complicated error pattern such as $\mathbb{E}_i^3$ (non-stationary error covariance structure). However, the power performance for this case improves with the larger sample size ($n = 300$). We also detected that the power of the test is affected by how the data is sampled, but with only a negligible difference; the power curves corresponding to the dense design are provided in the Supplementary materials, Appendix D.3.

Finally, it is worthwhile noting that, when calculating the size of proposed test $T_{n,\text{obs}}$ or T_b^*, the difference $RSS_0 - RSS_1$ occasionally comes out negative. We detected such cases 2.6% of the time on an average, and 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 (less than 0.1% of the time) this difference was returned negative; we excluded such cases from our study.

5.3 Applications

5.3.1 Gait Data

We turn our attention to data applications. We first consider 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 1 of Appendix E.1 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 ith child and the percentage of gait cycle t_j. Here δ_{ij} are the measurement errors. We first employ our resampling based test to investigate whether the hip angles are associated with the knee angles. We select 7 cubic B-splines for x and t to fit the GFCM, $\mathrm{E}[Y_i(t)|X_i(t)] = F\{X_i(t), t\}$, and $B = 250$ bootstrap replications are used. The bootstrap p-value is computed to be less than 0.004, thus we conclude that the hip angle measured at a specific time point has a strong effect on the knee angle at the same time point.

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 3×3 unknown covariance matrix R, ϵ_{ij} are the errors from $N(0, \sigma_\epsilon^2)$, and $(b_{0i}, b_{1i}, b_{2i})^T$ and ϵ_{ij} are assumed to be independent. For the GFCM and the linear FCM, we report the in-sample and the out-of-

sample 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 4 (top panel). 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 narrower at all significance levels. Figure 3, 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 3 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 E.2 of the Supplementary Materials.

5.3.2 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 absorption are measured for each subject. 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 of data where both the functional response and covariate are observed on a sparse design. Our objective is to examine the pattern of calcium absorption over the ages based on calcium intake as well as body mass index (BMI) of the patients averaged over their ages. To assess their relationship, let $Y_{ij} = Y_i(t_{ij})$ be the calcium absorption and $W_{1ij} = X_{1,i}(t_{1ij}) + \delta_{ij}$ be the calcium intake corresponding to ith subject and the jth time point, where δ_{ij} are the noise. Let $X_{2,i}$ denote the average BMI for the ith subject. Figure 2 of Appendix E.1 displays the observed individual trajectories of calcium intake and absorption along the patient's age

at the visit.

In the analysis, we fit the GFCM, $\mathrm{E}[Y_i(t)|X_{1,i}(t), X_{2,i}] = F\{X_{1,i}(t), t\} + \gamma(t)X_{2,i}$, where $\gamma(t)$ denotes the unknown slope function of the average BMI. As an alternative model, we also study the dependence assuming a linear FCM, $\mathrm{E}[Y_i(t)|X_{1,i}(t), X_{2,i}] = \beta_0(t) + \beta_1(t)X_{1,i}(t) + \gamma(t)X_{2,i}$. In the analysis, ages are transformed into the values in [0,1], and the results are considered as evaluation points of the functions. As before, we begin by testing the null hypothesis of no association between the calcium intake and the absorbtion. Specifically, we test the null hypothesis $H_0 : \mathrm{E}[Y(t)|X_1(t) = x_1, X_2 = x_2] = \beta_0(t) + x_2\gamma(t)$ versus the alternative $H_1 : \mathrm{E}[Y(t)|X_1(t) = x_1, X_2 = x_2] = F(x_1, t) + x_2\gamma(t)$ using Algorithm 2. We select 7 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 0.003, indicating a significant association between the current calcium intake measured and the current 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, $X_{1,i}(t)$. The covariates in the training set are smoothed using the standard FPCA toolkit for sparse functional data Yao et al. (2005a). The resulted estimated model (using the training data) is later used to reconstruct the trajectories in the test set. The prediction results using the GFCM and the competitive linear FCM are presented in Table 4 (bottom panel). Both the GFCM and the linear FCM show similar in-sample and out-of-sample performance, $\mathrm{RMSPE}^{\mathrm{in}}$ and $\mathrm{RMSPE}^{\mathrm{out}}$; this indicates that a simple linear association between the calcium intake and absorption is more appropriate.

Furthermore, in Figure 4, the leftmost two panels present the point-wise prediction intervals/bands for two selected subjects from the test data. 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 absorption. The rightmost panel in Figure 4 displays the estimated slope function $\widehat{\gamma}(t)$ obtained by fitting the GFCM and the linear FCM. On an average, both methods indicate a positive effect of the average BMI on the absorption but with a slight difference in the estimated effects.

Additional analysis has shown that the difference is due to the estimation errors.

6 Discussion

We propose a wider class of function-on-function regression models, the general functional concurrent model, and discuss significance testing of no association. In particular, 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, the linear FCM assumes a linear dependence is valid. Thus, our proposed procedure allows one to approach the problem from a more general point of view. We have implemented our proposed estimation and testing methodology using `R` software, and details about the implementation are provided in Appendix F of the Supplementary Materials.

References

Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements*. New York, NY: Springer.

Eilers, P. H. C. and B. D. Marx (1996). Flexible smoothing with b-splines and penalties. *Statistical Science 11*, 89–121.

Goldsmith, J., S. Greven, and C. Crainiceanu (2013). Corrected confidence bands for functional data using principal components. *Biometrics 69*, 41–51.

Goldsmith, J., V. Zipunnikov, and J. Schrack (2015). Generalized multilevel function-on-scalar regression and principal component analysis. *Biometrics 71*, 344–353.

Guo, W. (2002). Functional mixed effects models. *Biometrics 58*, 121–128.

Huang, L., F. Scheipl, J. Goldsmith, J. Gellar, J. Harezlak, M. W. McLean, B. Swihart, L. Xiao, C. Crainiceanu, and P. Reiss (2015). *refund: Regression with Functional Data*. R package version 0.1-13.

Jiang, C.-R. and J.-L. Wang (2011). Functional single index models for longitudinal data. *The Annals of Statistics 39*, 362–388.

Li, Y. and D. Ruppert (2008). On the asymptotics of penalized splines. *Biometrika 95*, 415–436.

Malfait, N. and J. O. Ramsay (2003). The historical functional linear model. *The Canadian Journal of statistics 31*, 115–128.

Marx, B. D. and P. H. C. Eilers (2005). Multivariate penalized signal regression. *Technometrics 47*, 13–22.

McLean, M. W., G. Hooker, A. M. Staicu, F. Scheipl, and D. Ruppert (2014). Functional generalized additive models. *Journal of Computational and Graphical Statistics 23*, 249–269.

Morris, J. S. and R. J. Carroll (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B 68*, 179–199.

Olshen, R. A., E. N. Biden, M. P. Wyatt, and D. H. Sutherland (1989). Gait analysis and the bootstrap. *The Annals of Statistics 17*, 1419–1440.

Ramsay, J. O., G. Hooker, and S. Graves (2009). *Functional data analysis in R and Matlab*. New York, NY: Springer.

Ramsay, J. O. and B. W. Silverman (2005). *Functional data analysis* (2nd ed.). New York ; Berlin : Springer.

Reiss, P. T., L. Huang, and M. Mennes (2010). Fast function-on-scalar regression with penalized basis expansions. *International Journal of Biostatistics 6*(28).

Ruppert, D., M. P. Wand, and R. J. Carroll (2003). *Semiparametric regression*. Cambridge, New York: Cambridge University Press.

Scheipl, F., A. M. Staicu, and S. Greven (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics 24*, 477–501.

Sentürk, D. and D. V. Nguyen (2011). Varying coefficient models for sparse noise-contaminated longitudinal data. *Statistica Sinica 21*, 1831–1856.

Shen, Q. and J. Faraway (2004). An f test for linear models with functional responses. *Statistica Sinica 14*, 1239–1257.

Slijepcevic, A., W. R. Anderson, and S. Matthews (2013). Testing existing models for predicting hourly variation in fine fuel moisture in eucalypt forests. *Forest Ecology and Management 306*, 202–215.

Slijepcevic, A., W. R. Anderson, S. Matthews, and D. H. Anderson (2015). Evaluating models to predict daily fine fuel moisture content in eucalypt forest. *Forest Ecology and Management 335*, 261–269.

Theologis, T. N. (2009). *Children's Orthopaedics and Fractures (Chapter 6)*. London: Springer.

Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Boca Raton, Florida: Chapman and Hall/CRC.

Wood, S. N. (2015). *Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation*. R package version 1.8-7.

Wu, Y., J. Fan, and H.-G. Müller (2010). Varying-coefficient functional linear regression. *Bernoulli 16*, 730–758.

Xiao, L., V. Zipunnikov, D. Ruppert, and C. Crainiceanu (2016). Fast covariance function estimation for high-dimensional functional data. *Statistics and Computing 26*, 409–421.

Xu, H., Q. Shen, X. Yang, and S. Shoptaw (2011). A quasi f-test for functional linear models with functional covariates and its application to longitudinal data. *Statistics in Medicine 30*, 2842–2853.

Yao, F., H.-G. Müller, and J.-L. Wang (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association 100*, 577–590.

Yao, F., H.-G. Müller, and J.-L. Wang (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics 33*, 2873–2903.

Zhang, J.-T. and J. Chen (2007). Statistical inference for functional data. *The Annals of Statistics 35*, 1052–1079.

Table 1: Summary of RMSPE^{in}, $\text{RMSPE}^{\text{out}}$, 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.

						$1-\alpha=0.95$						$1-\alpha=0.90$				$1-\alpha=0.85$			
		RMSPE^{in}		$\text{RMSPE}^{\text{out}}$		ICP		IL		R(SE)		ICP		IL		ICP		IL	
n	$\mathbb{E}_i$	`GFCM`	`FCM`	`GFCM`	`FCM`	`GFCM`	`FCM`	`GFCM`	`FCM`	`GFCM`	`FCM`	`GFCM`	`FCM`	`GFCM`	`FCM`	`GFCM`	`FCM`	`GFCM`	`FCM`
Scenario B (true relationship is non-linear), $m = 81$																			
100	$\mathbb{E}_i^1$	0.97	3.55	0.96	3.75	0.959	0.943	3.83	12.67	[3.54, 5.08]	[3.65, 21.90]	0.916	0.919	3.21	10.64	0.871	0.896	2.81	9.31
	$\mathbb{E}_i^2$	1.32	3.67	1.31	3.86	0.955	0.944	5.20	13.31	[4.96, 6.26]	[5.03, 22.24]	0.909	0.917	4.37	11.17	0.861	0.891	3.82	9.77
	$\mathbb{E}_i^3$	1.84	3.90	1.98	4.10	0.946	0.944	7.34	14.56	[5.53, 9.84]	[8.37, 23.29]	0.895	0.915	6.16	12.22	0.844	0.884	5.39	10.69
300	$\mathbb{E}_i^1$	0.98	3.65	0.92	3.68	0.963	0.948	3.83	12.94	[3.54, 5.11]	[3.63, 22.07]	0.921	0.925	3.21	10.86	0.877	0.902	2.81	9.50
	$\mathbb{E}_i^2$	1.33	3.75	1.28	3.79	0.958	0.948	5.20	13.56	[4.96, 6.30]	[5.02, 22.39]	0.912	0.923	4.37	11.38	0.866	0.898	3.82	9.96
	$\mathbb{E}_i^3$	1.87	3.99	1.87	4.02	0.952	0.949	7.30	14.79	[5.58, 9.48]	[8.44, 23.52]	0.903	0.921	6.12	12.41	0.854	0.892	5.36	10.87
Scenario B (true relationship is non-linear), $m_i \overset{iid}{\sim} Unif(20, 31)$																			
100	$\mathbb{E}_i^1$	1.16	3.55	1.18	3.76	0.964	0.928	4.52	12.60	[3.71, 7.16]	[3.32, 23.46]	0.929	0.903	3.79	10.57	0.891	0.880	3.32	9.25
	$\mathbb{E}_i^2$	1.46	3.66	1.47	3.87	0.959	0.932	5.75	13.17	[5.09, 8.04]	[4.69, 23.72]	0.918	0.904	4.82	11.06	0.875	0.877	4.22	9.68
	$\mathbb{E}_i^3$	1.94	3.89	2.06	4.10	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
300	$\mathbb{E}_i^1$	1.15	3.65	1.00	3.68	0.973	0.935	4.42	12.90	[3.67, 6.90]	[3.23, 23.25]	0.941	0.912	3.71	10.83	0.904	0.889	3.25	9.48
	$\mathbb{E}_i^2$	1.46	3.76	1.33	3.79	0.966	0.941	5.66	13.50	[5.06, 7.79]	[4.70, 23.50]	0.927	0.915	4.75	11.33	0.885	0.889	4.16	9.91
	$\mathbb{E}_i^3$	1.96	3.99	1.91	4.03	0.958	0.948	7.64	14.80	[5.87, 10.64]	[8.28, 24.56]	0.913	0.919	6.41	12.42	0.866	0.890	5.61	10.87
Scenario A (true relationship is linear), $m = 81$																			
100	$\mathbb{E}_i^1$	0.90	0.90	0.90	0.90	0.951	0.952	3.53	3.53	[3.52, 3.59]	[3.52, 3.59]	0.902	0.902	2.97	2.97	0.853	0.853	2.60	2.60
	$\mathbb{E}_i^2$	1.27	1.27	1.27	1.27	0.951	0.951	4.98	4.93	[4.93, 5.14]	[4.93, 5.15]	0.901	0.902	4.18	4.18	0.852	0.852	3.66	3.66
	$\mathbb{E}_i^3$	1.80	1.83	1.93	1.85	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
300	$\mathbb{E}_i^1$	0.90	0.90	0.89	0.89	0.952	0.952	3.53	3.53	[3.52, 3.58]	[3.52, 3.58]	0.902	0.902	2.97	2.97	0.853	0.853	2.59	2.60
	$\mathbb{E}_i^2$	1.27	1.27	1.27	1.27	0.951	0.951	4.98	4.98	[4.93, 5.15]	[4.93, 5.15]	0.902	0.902	4.18	4.18	0.852	0.852	3.66	3.66
	$\mathbb{E}_i^3$	1.83	1.84	1.86	1.84	0.948	0.949	7.13	7.13	[5.44, 8.77]	[5.44, 8.72]	0.897	0.899	5.98	5.98	0.846	0.849	5.24	5.24
Scenario A (true relationship is linear), $m_i \overset{iid}{\sim} Unif(20, 31)$																			
100	$\mathbb{E}_i^1$	0.92	0.92	0.90	0.90	0.955	0.955	3.61	3.61	[3.55, 3.77]	[3.55, 3.77]	0.907	0.908	3.03	3.03	0.859	0.859	2.65	2.65
	$\mathbb{E}_i^2$	1.28	1.28	1.27	1.27	0.952	0.952	5.04	5.04	[4.96, 5.26]	[4.96, 5.26]	0.904	0.904	4.23	4.23	0.854	0.855	3.70	3.70
	$\mathbb{E}_i^3$	1.81	1.84	1.89	1.86	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
300	$\mathbb{E}_i^1$	0.92	0.92	0.90	0.90	0.955	0.955	3.60	3.60	[3.55, 3.72]	[3.55, 3.73]	0.908	0.908	3.02	3.02	0.859	0.859	2.64	2.64
	$\mathbb{E}_i^2$	1.28	1.28	1.27	1.27	0.953	0.953	5.03	5.03	[4.97, 5.19]	[4.97, 5.19]	0.904	0.904	4.22	4.22	0.855	0.855	3.69	3.69
	$\mathbb{E}_i^3$	1.84	1.85	1.85	1.84	0.949	0.950	7.16	7.17	[5.45, 9.19]	[5.45, 9.19]	0.899	0.900	6.01	6.02	0.848	0.850	5.26	5.27

Table 2: Summary of RMSPE^{in}, $\text{RMSPE}^{\text{out}}$, ICP and average computation time (in seconds) based on 1000 simulated data sets. The models fitted by the estimation procedure of GFCM. The average computation times are obtained using `gam` and `bam` functions in `mgcv R` package.

n	$\mathbb{E}_i$	RMSPE^{in}	$\text{RMSPE}^{\text{out}}$	ICP at $1-\alpha=0.95$	ICP at $1-\alpha=0.90$	ICP at $1-\alpha=0.85$	Computation Time `gam`	Computation Time `bam`
Scenario B (model with single functional covariate), $m=81$								
100	$\mathbb{E}_i^1$	0.97	0.96	0.959	0.916	0.871	4.05	2.02
	$\mathbb{E}_i^2$	1.32	1.31	0.955	0.909	0.861	3.75	2.02
	$\mathbb{E}_i^3$	1.84	1.98	0.946	0.895	0.844	3.55	3.02
300	$\mathbb{E}_i^1$	0.98	0.92	0.963	0.921	0.877	22.09	4.07
	$\mathbb{E}_i^2$	1.33	1.28	0.958	0.912	0.866	15.60	3.90
	$\mathbb{E}_i^3$	1.87	1.87	0.952	0.903	0.854	12.03	3.06
Scenario B (model with single functional covariate), $m_i \overset{iid}{\sim} Unif(20,31)$								
100	$\mathbb{E}_i^1$	1.16	0.98	0.964	0.929	0.891	1.04	2.76
	$\mathbb{E}_i^2$	1.46	1.47	0.959	0.918	0.875	0.98	2.88
	$\mathbb{E}_i^3$	1.94	2.06	0.949	0.900	0.852	1.02	2.53
300	$\mathbb{E}_i^1$	1.15	1.00	0.973	0.941	0.904	3.76	2.64
	$\mathbb{E}_i^2$	1.46	1.33	0.966	0.927	0.885	3.63	2.97
	$\mathbb{E}_i^3$	1.96	1.91	0.958	0.913	0.866	3.33	2.53
Scenario C (model with two functional covariates), $m=81$								
100	$\mathbb{E}_i^1$	0.96	0.94	0.955	0.908	0.860	24.74	4.44
	$\mathbb{E}_i^2$	1.31	1.28	0.957	0.912	0.863	24.37	4.17
	$\mathbb{E}_i^3$	1.80	1.96	0.937	0.882	0.831	24.15	3.37
300	$\mathbb{E}_i^1$	0.97	0.92	0.958	0.912	0.866	57.74	4.61
	$\mathbb{E}_i^2$	1.32	1.27	0.959	0.915	0.867	56.52	4.85
	$\mathbb{E}_i^3$	1.85	1.90	0.946	0.892	0.842	60.43	3.16
Scenario C (model with two functional covariate), $m_i \overset{iid}{\sim} Unif(20,31)$								
100	$\mathbb{E}_i^1$	1.13	1.07	0.961	0.921	0.880	4.20	3.20
	$\mathbb{E}_i^2$	1.44	1.38	0.960	0.919	0.874	5.12	2.97
	$\mathbb{E}_i^3$	1.91	2.01	0.942	0.891	0.840	4.86	3.13
300	$\mathbb{E}_i^1$	1.13	0.97	0.970	0.934	0.895	18.14	4.05
	$\mathbb{E}_i^2$	1.44	1.30	0.967	0.928	0.885	14.15	4.33
	$\mathbb{E}_i^3$	1.94	1.91	0.953	0.903	0.856	11.15	3.96

Table 3: Rejection probabilities ($\times 100$) based on 2000 simulations. The values in the parenthesis are the estimated standard errors ($\times 100$) of the rejection probabilities.

Scenario	Sampling	n	$\alpha=5\%$ $\mathbb{E}_i^1$	$\alpha=5\%$ $\mathbb{E}_i^2$	$\alpha=5\%$ $\mathbb{E}_i^3$	$\alpha=10\%$ $\mathbb{E}_i^1$	$\alpha=10\%$ $\mathbb{E}_i^2$	$\alpha=10\%$ $\mathbb{E}_i^3$
A	$m=81$	100	5.3(0.7)	5.4(0.7)	6.6(0.8)	11.1(1.0)	10.6(1.0)	10.3(1.0)
		300	6.1(0.8)	5.0(0.7)	5.1(0.7)	10.6(1.0)	10.9(1.0)	9.4(0.9)
	$m_i \overset{iid}{\sim} Unif(20,31)$	100	5.5(0.7)	5.0(0.7)	4.6(0.7)	10.0(0.9)	10.8(1.0)	9.8(0.9)
		300	4.8(0.7)	4.2(0.6)	5.2(0.7)	9.5(0.9)	10.4(1.0)	10.7(1.0)
B	$m=81$	100	6.2(0.5)	4.5 (0.5)	5.1(0.5)	11.5(0.7)	9.7(0.7)	10.0(0.7)
		300	6.0(0.5)	4.8 (0.5)	5.0(0.5)	12.9(0.7)	10.7(0.7)	11.7(0.7)
	$m_i \overset{iid}{\sim} Unif(20,31)$	100	5.0(0.5)	4.4(0.5)	6.2(0.5)	11.2(0.7)	9.8(0.7)	10.8(0.7)
		300	6.2(0.5)	5.3(0.5)	6.5(0.5)	12.8(0.7)	10.7(0.7)	11.7(0.7)

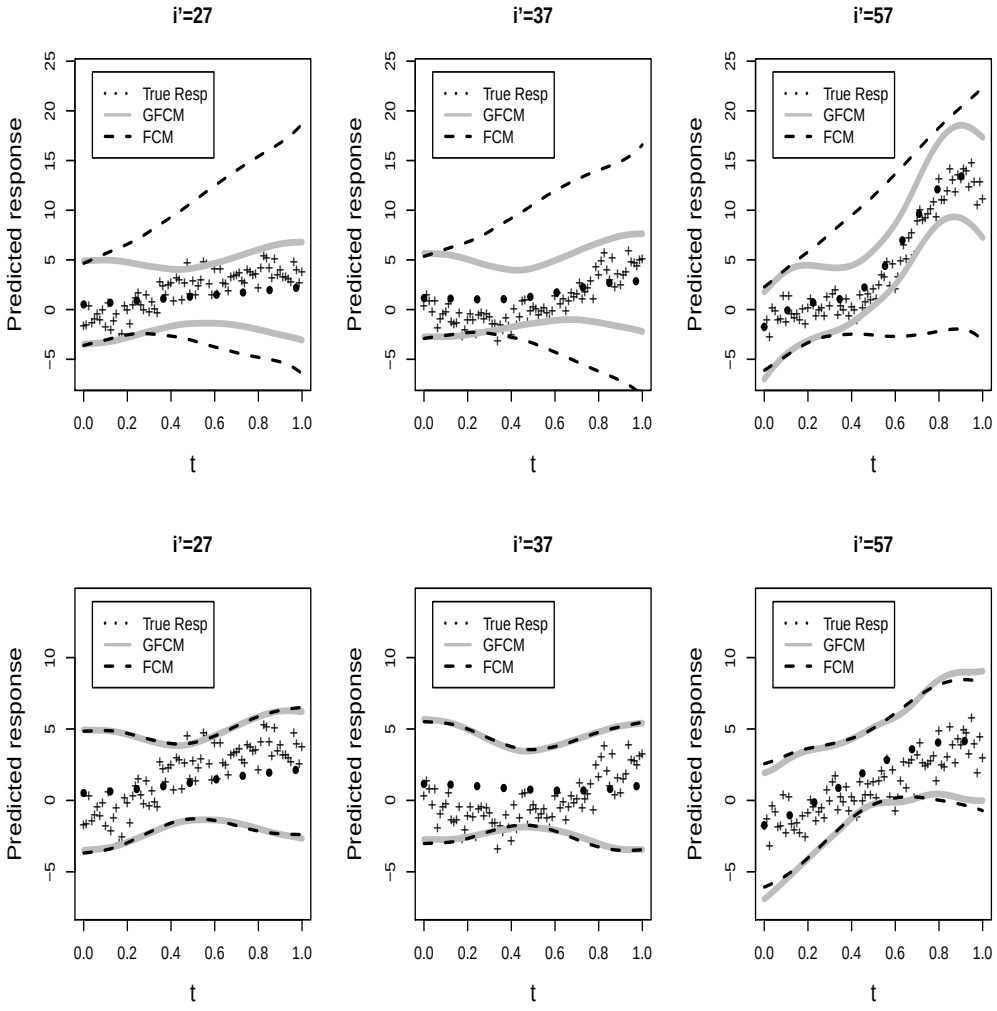

Figure 1: 95% prediction bands constructed for three subject-level trajectories in the test data. The results are from scenario A and B (involving a single functional covariate) for the setting where sampling design is dense, $n = 100$ and $\mathbb{E}_i = \mathbb{E}_i^3$. The top (bottom) panel corresponds to the case where the true function $F_1(x_1, t)$ is non-linear (linear) in x_1. "+" 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.

Table 4: Results from data examples as described in Section 5.3. Displayed are the summaries of RMSPE^{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.

Gait data

			$1-\alpha = 0.95$			$1-\alpha = 0.90$			$1-\alpha = 0.85$		
	RMSPE^{in}	$\text{RMSPE}^{\text{out}}$	ICP	IL	R(SE)	ICP	IL	R(SE)	ICP	IL	R(SE)
`GFCM`	5.50	5.72	0.939	20.43	[14.86, 35.75]	0.867	17.15	[12.47, 30.00]	0.828	15.01	[10.91, 26.26]
`FCM`	5.59	5.69	0.943	20.60	[14.91, 36.12]	0.861	17.29	[12.51, 30.32]	0.822	15.13	[10.95, 26.53]
`LME`	18.93	19.05	0.972	76.55	[73.75, 82.26]	0.883	64.24	[61.90, 69.04]	0.844	56.22	[54.17, 60.42]

Calcium absorption data

			$1-\alpha = 0.95$			$1-\alpha = 0.90$			$1-\alpha = 0.85$		
	RMSPE^{in}	$\text{RMSPE}^{\text{out}}$	ICP	IL	R(SE)	ICP	IL	R(SE)	ICP	IL	R(SE)
`GFCM`	0.079	0.112	0.950	0.35	[0.32, 0.56]	0.935	0.30	[0.27, 0.47]	0.919	0.26	[0.23, 0.41]
`FCM`	0.081	0.114	0.951	0.33	[0.31, 0.41]	0.935	0.28	[0.26, 0.34]	0.921	0.24	[0.23, 0.30]

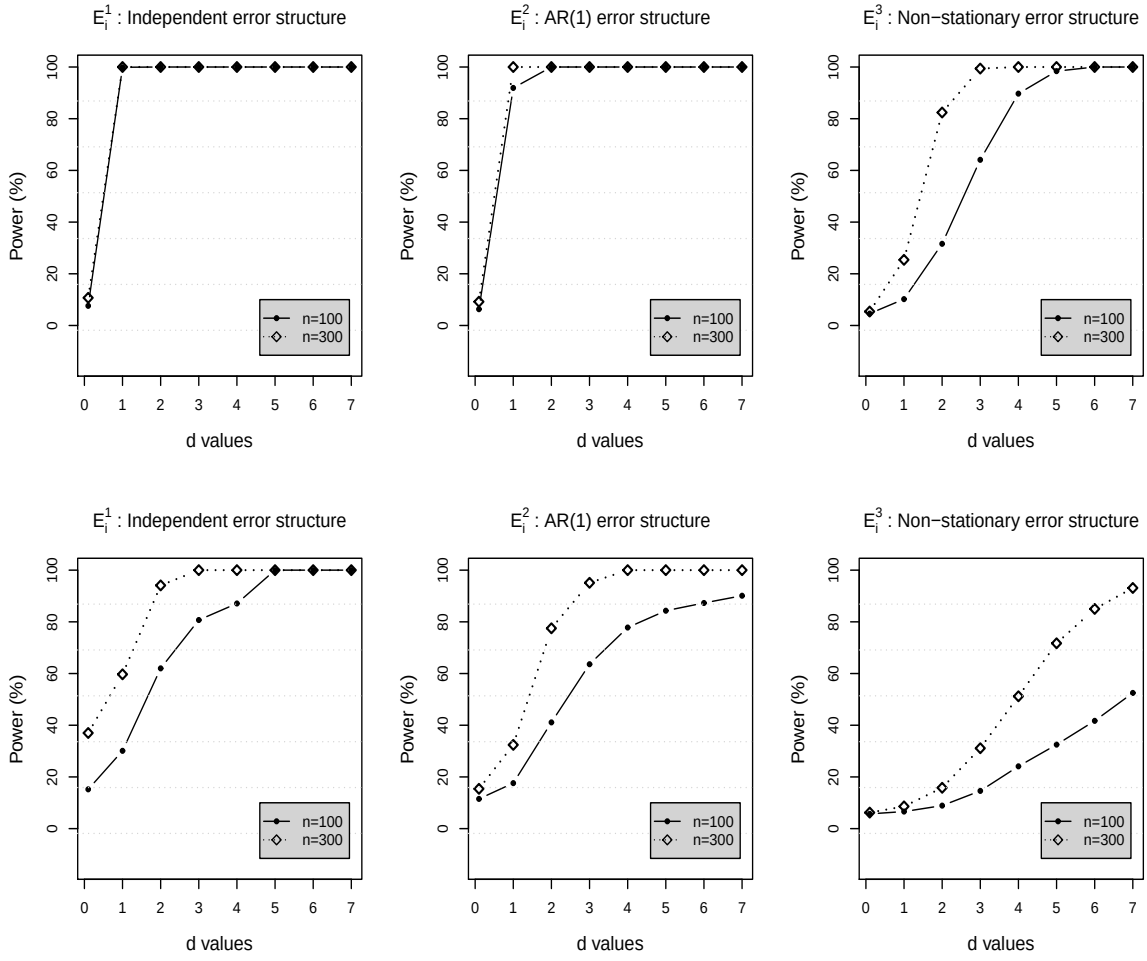

Figure 2: Powers ($\times 100$) of the tests at significance level $\alpha = 5\%$. The top (bottom) panel displays the results from scenario A (scenario B) for the setting where sampling design is sparse. The error process in the left, middle and right panels is assumed to be $\mathbb{E}_i^1$, $\mathbb{E}_i^2$ and $\mathbb{E}_i^3$, respectively.

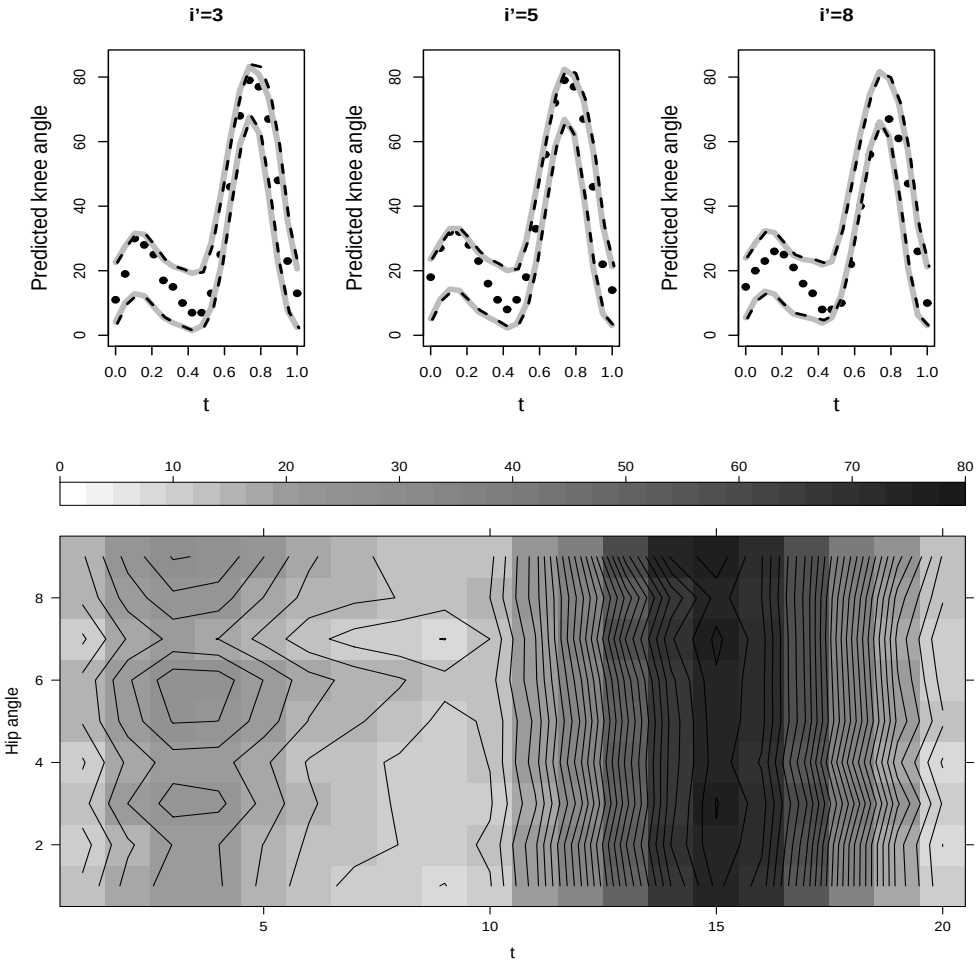

Figure 3: Results from gait data analysis as described in Section 5.3. The top panel displays 95% prediction bands obtained by fitting the GFCM (grey solid lines) and the linear FCM (dashed lines) for three subject-level trajectories in the test data. "•" represent the knee angles in the test data. The bottom panel shows the heat map of $\widehat{Y}_{0,i'}(t)$ obtained from the test data set of the gait data example.

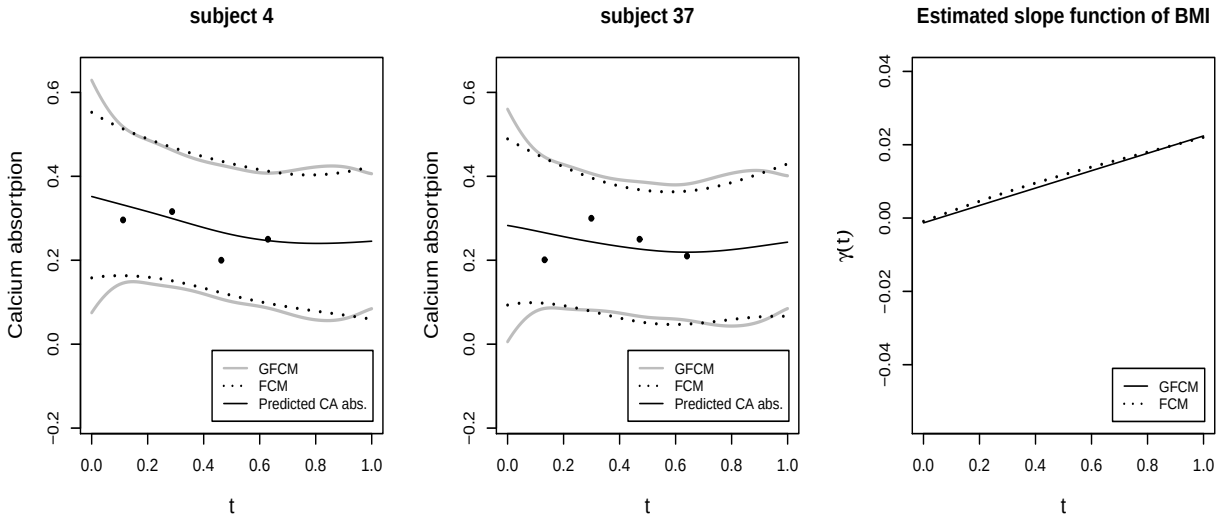

Figure 4: Results from calcium data analysis as described in Section 5.3. The leftmost two panels display 95% prediction bands obtained by fitting the GFCM (grey solid lines) and the linear FCM (black dotted lines) for two subject-level trajectories in the test data. "•" indicate the calcium absorption measured, and thin solid lines are the predicted calcium absorption from the GFCM estimation method. The rightmost panel displays the estimated slope function of BMI, $\widehat{\gamma}(t)$, obtained from the GFCM (black solid line) and the linear FCM (black dotted line).