Statistics and Its Interface Volume 11 (2018) 669–685

Additive nonlinear functional concurrent model

Janet S. Kim, Arnab Maity*, and Ana-Maria Staicu

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 the current response to current values of the covariates by a sum of smooth unknown bivariate 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 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.

AMS 2000 subject classifications: Primary 62G05, 62G10.
Keywords and phrases: Functional concurrent models, F-test, Nonlinear models, Penalized B-splines, Prediction.

1. INTRODUCTION

Functional regression models where both the response and the covariate are functional have been long researched in the literature. Such models typically assume that the current response depends on the full trajectory of the covariate. The dependence is modeled through a weighted integral of the full covariate trajectory through an unknown bivariate coefficient surface as weight. Estimation procedures for this model have been discussed in [27], [43] and [40], among others. The crucial dependence assumption for this type of model may be impractical in many real data situations. To bypass this limitation, one might use the functional historical models (see, e.g., [19]), 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 bivariate 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., [27]; [26]; [31]) assume a linear relationship; they can be thought of as a series of linear regressions for each time point, with the constraint that the regression coefficient is a univariate smooth function of time. These types of models, commonly known as varying coefficient models [11] are useful in many situations. There is a substantial amount of literature on such models for both scalar and longitudinal/functional response variables. In the latter case, several methods to estimate the model components have been developed, such as local polynomial kernel smoothing [13, 39, 5, 1], regression and smoothing spline methods [14, 4], and penalized spline and quadratic inference function based methods [25]. There have been many subsequent developments in this area such as an extension to spatial imaging [45], ridge regression [12]. We refer the readers to [6] and [22], and references therein, for detailed reviews on varying coefficient models and functional concurrent models. 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 nonlinear. 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 bivariate 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 correct), fitting our proposed model maintains prediction accuracy. On the other hand, when the true relationship is nonlinear, fitting the linear concurrent model results in high bias and loss of prediction accuracy.

We make three main contributions. First, we propose a general nonlinear functional concurrent model to describe the complex association between two or more functional variables measured on the same domain. We model the relationship via unknown bivariate 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 ([3]; [18]; [21]). We discuss prediction of the response trajectory and develop point-wise prediction intervals that account for the

*Corresponding author.

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 [10] and [23] who considered inference in a functional mixed model framework. [28] 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 [9] extended these ideas to generalized multilevel function-on-scalar regression. [30], extending the work of [16], considered function-on-function regression models with flexible residual correlation structure, but did not numerically investigate 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., [43]; [46]; [8]) to the residuals. The proposed inference uses the resulting estimate of the residual covariance.

Second, our model allows us to incorporate multiple functional and scalar covariates, assuming that the effects of the covariates are additive. Specifically, we represent the effects of the covariates by a sum of smooth bivariate 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. Another way to incorporate multiple functional covariates in a functional regression is proposed by [17], where they use a single index model by first forming a linear effect of the functional covariates and then modeling the combined effect via a nonparametric function. Nevertheless, the models proposed by [17] are not our special case, and neither is our model a special case of theirs. This is because our methodology models the individual effect of each functional covariate nonlinearly and develops an additive model. We also employ a bivariate tensor product of B-spline based estimation procedure. Recently, [7] proposed a function-on-function nonlinear additive model with multiple functional and scalar covariates. They propose to summarize each functional covariate by taking an integral of the product of the covariate and an unknown two-dimensional index function, and then model the resulting single indices using a nonlinear additive model. As such, this model uses the entirety of each functional covariate to model the response function at any given time point. In contrast, we consider a concurrent relationship.

Third, we develop a testing procedure for assessing whether any of the functional covariates are related to the response variable. To the best of our knowledge, no other work considers testing for nonlinear concurrent models in functional setting and provides numerical results. 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., [32]; [42]) 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 [21]. In particular, the nonlinear relationship that describes the conditional mean of the current response given the current value of the covariate is reminiscent of the one used in [21]. The key differences come from 1) the type of response considered, 2) ability to accommodate various sampling designs and 3) the covariance model assumed for the residuals. Specifically, we consider functional responses in this paper, whereas [21] studied scalar responses. Also, the proposed estimation and prediction procedures can accommodate various sampling designs for both responses and covariates, such as densely or sparsely sampled predictors with or without error. In contrast, the methods of [21] are presented only for densely sampled functional covariates. Additionally, we assume unknown complex dependence structure of the residuals, whereas [21] 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.

2. ADDITIVE NONLINEAR 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. For simplicity of presentation, our model and methodology are outlined for the setting involving two functional covariates; nevertheless, our methodology is general and can accommodate more than two scalar or functional covariates via linear or smooth effects.

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 $X_{q,i}(\cdot)$ are realizations from random square-integrable processes, $X_q(\cdot)$, 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 our 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 common regularly spaced grid of points $t_1, \ldots, t_m$. 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, e.g., sparsely observed data or when the covariates and response are not observed on the same grid of points, are discussed in Section 4 and Section A of the Supplementary Material.

We introduce the following additive nonlinear functional concurrent model (ANFCM)

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

where $\mu_Y(t)$ is an unknown and smooth intercept function, F_1 and F_2 are smooth unknown bivariate functions defined on $\mathbb{R} \times \mathcal{T}$, and $\epsilon_i(\cdot)$ is a Gaussian 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 $\mathrm{E}\left[F_q\{X_q(t), t\}\right] = 0$ for any $t \in \mathcal{T}$, $q = 1, 2$, and thus $\mu_Y(t)$ is the marginal mean function of the response. We introduce two main innovations in (1). First, the general bivariate 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, we estimate and incorporate the unknown covariance structure for the residual process $\epsilon_i(t)$ in our inference, e.g., in prediction of a new response curve and hypothesis testing.

An important advantage of our framework is that it can easily accommodate multiple functional covariates with various types of effects; a few examples are shown in Table 1. We assume that models in Table 1 have flexible correlation structure for the residuals. The standard linear functional concurrent model with only one functional covariate is a special case of the model in (a) with $F_1(x, t) = \beta_0(t) + x\beta_1(t)$ and $F_2(x, t) = 0$, where $\beta_0(\cdot)$ and $\beta_1(\cdot)$ are unknown parameter functions. All the above models can easily accommodate scalar (time invariant) covariate effects of the form $X\beta(t)$ or $F(X, t)$ as well. While we present our methodology using only two functional covariates, the proposed framework can easily include more. This comes at the cost of computational time due to increasing number of model components and penalty parameters; although an efficient implementation (e.g., using the `bam` function instead of the default `gam` function in R package `mgcv`) helps to reduce computational time substantially. We have conducted a numerical study with up to five functional covariates (results shown in Supplementary Material, Section D.5), observed that our proposed methods perform reasonably well in all the cases considered, with increasing computation time.

2.2 Estimation

We focus on model (1) and describe estimation of the marginal mean function $\mu_Y(t)$ and the bivariate 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 $\mu_Y(t) = \sum_{d=1}^{K_\mu} B_{\mu,d}(t)\theta_{\mu,d} = \mathbf{B}_\mu^T(t)\mathbf{\Theta}_\mu$, where $\mathbf{B}_\mu(t)$ is the K_μ-dimensional vector of $B_{\mu,d}(t)$'s, and $\mathbf{\Theta}_\mu$ is the vector of unknown parameters $\theta_{\mu,d}$'s. We also model $F_1(\cdot, \cdot)$ and $F_2(\cdot, \cdot)$ using bivariate basis expansion using tensor product of univariate B-spline basis functions ([20]; [36]; [21]; [30]). 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 $F_1\{X_{1,i}(t), t\} = \sum_{k=1}^{K_{x1}} \sum_{l=1}^{K_{t1}} B_{X1,k}\{X_{1,i}(t)\}B_{T1,l}(t)\theta_{1,k,l} = \mathbf{Z}_{1,i}^T(t)\mathbf{\Theta}_1$, where $\mathbf{Z}_{1,i}(t)$ is the $K_{x1}K_{t1}$-dimensional vector of $B_{X1,k}\{X_{1,i}(t)\}B_{T1,l}(t)$'s, and $\mathbf{\Theta}_1$ is the vector of unknown parameters, $\theta_{1,k,l}$'s. Similarly, we write $F_2\{X_{2,i}(t), t\} = \sum_{k=1}^{K_{x2}} \sum_{l=1}^{K_{t2}} B_{X2,k}\{X_{2,i}(t)\}B_{T2,l}(t)\theta_{2,k,l} = \mathbf{Z}_{2,i}^T(t)\mathbf{\Theta}_2$, where $\mathbf{Z}_{2,i}(t)$ is the $K_{x2}K_{t2}$-dimensional vector of $B_{X2,k}\{X_{2,i}(t)\}B_{T2,l}(t)$'s, and $\mathbf{\Theta}_2$ is the vector of unknown parameters, $\theta_{2,k,l}$'s. Based on the above expansions, model (1) can be written as

$$(2) \quad Y_i(t) = \mathbf{B}_\mu^T(t)\mathbf{\Theta}_\mu + \mathbf{Z}_{1,i}^T(t)\mathbf{\Theta}_1 + \mathbf{Z}_{2,i}^T(t)\mathbf{\Theta}_2 + \epsilon_i(t).$$

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 a large number of basis functions to fully capture the complexity of the function, and penalize the coefficients to ensure smoothness of the resulting fit.

We propose to estimate $\mathbf{\Theta}_\mu$, $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ by minimizing a penalized criterion $\sum_{i=1}^n ||Y_i(\cdot) - \mathbf{B}_\mu^T(\cdot)\mathbf{\Theta}_\mu - \mathbf{Z}_{1,i}^T(\cdot)\mathbf{\Theta}_1 - \mathbf{Z}_{2,i}^T(\cdot)\mathbf{\Theta}_2||^2 + \mathbf{\Theta}_\mu^T\mathbb{P}_\mu\mathbf{\Theta}_\mu + \mathbf{\Theta}_1^T\mathbb{P}_1\mathbf{\Theta}_1 + \mathbf{\Theta}_2^T\mathbb{P}_2\mathbf{\Theta}_2$, where $\mathbb{P}_\mu$, $\mathbb{P}_1$ and $\mathbb{P}_2$ 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 $||\cdot||^2$

Table 1. Examples of general additive functional concurrent models

Model	Form
(a) General model	$\mathrm{E}\{Y(t)\|X_1(t), X_2(t)\} = \mu_Y(t) + F_1\{X_1(t), t\} + F_2\{X_2(t), t\}$
(b) Linear concurrent model	$\mathrm{E}\{Y(t)\|X_1(t), X_2(t)\} = \mu_Y(t) + X_1(t)\beta_1(t) + X_2(t)\beta_2(t)$
(c) Partially linear concurrent model	$\mathrm{E}\{Y(t)\|X_1(t), X_2(t)\} = \mu_Y(t) + F_1\{X_1(t), t\} + X_2(t)\beta_2(t)$

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

$$\sum_{i=1}^{n}\sum_{j=1}^{m}\{Y_i(t_j) - \mathbf{B}_\mu^T(t_j)\boldsymbol{\Theta}_\mu - \mathbf{Z}_{1,i}^T(t_j)\boldsymbol{\Theta}_1 - \mathbf{Z}_{2,i}^T(t_j)\boldsymbol{\Theta}_2\}^2/m + \boldsymbol{\Theta}_\mu^T\mathbb{P}_\mu\boldsymbol{\Theta}_\mu + \boldsymbol{\Theta}_1^T\mathbb{P}_1\boldsymbol{\Theta}_1 + \boldsymbol{\Theta}_2^T\mathbb{P}_2\boldsymbol{\Theta}_2. \tag{3}$$

Here, $\mathbb{P}_\mu$ is given by $\mathbb{P}_\mu = \lambda_\mu\mathbb{D}_\mu^T\mathbb{D}_\mu$, where $\mathbb{D}_\mu$ represents the second order difference penalty ([3]; [20]; [21]), and λ_μ is its penalty parameter. Also, $\mathbb{P}_q$ $(q = 1, 2)$ is given by $\mathbb{P}_q = \lambda_{xq}\mathbb{D}_{xq}^T\mathbb{D}_{xq} \bigotimes \mathbb{I}_{K_{tq}} + \lambda_{tq}\mathbb{I}_{K_{xq}} \bigotimes \mathbb{D}_{tq}^T\mathbb{D}_{tq}$, where the notation $\bigotimes$ stands for the Kronecker product, $\mathbb{I}_K$ is the identity matrix with dimension K, and $\mathbb{D}_{xq}$ and $\mathbb{D}_{tq}$ 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 $\widehat{\boldsymbol{\Theta}}_\mu$, $\widehat{\boldsymbol{\Theta}}_1$ and $\widehat{\boldsymbol{\Theta}}_2$ is readily available for fixed values of the penalty parameters. Define the m-dimensional vector of response $\mathbf{Y}_i = [Y_i(t_1), \ldots, Y_i(t_m)]^T$ and the m-dimensional vector of errors $\mathbf{E}_i = [\epsilon_i(t_1), \ldots, \epsilon_i(t_m)]^T$. Also, we define $\mathbb{B}_\mu$ as $m \times K_\mu$-dimensional matrix with the j-th row given by $\mathbf{B}_\mu^T(t_j)$ and $\mathbb{Z}_{q,i}$ as $m \times K_{xq}K_{tq}$-dimensional matrix with the j-th row given by $\mathbf{Z}_{q,i}^T(t_j)$ $(q = 1, 2)$. For simplicity, denote $\mathbb{Z}_i = [\mathbb{B}_\mu|\mathbb{Z}_{1,i}|\mathbb{Z}_{2,i}]$, $\boldsymbol{\Theta}^T = [\boldsymbol{\Theta}_\mu^T, \boldsymbol{\Theta}_1^T, \boldsymbol{\Theta}_2^T]^T$ and $\mathbb{P} = \mathrm{diag}(\mathbb{P}_\mu, \mathbb{P}_1, \mathbb{P}_2)$. Then the solution of $\boldsymbol{\Theta}$ is calculated as

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

where $\mathbb{H} = \{\sum_{i=1}^{n}\mathbb{Z}_i^T\mathbb{Z}_i + \mathbb{P}\}^{-1}$. The penalty parameters can be chosen based on some appropriate criteria such as generalized cross validation (GCV) ([29]; [36]) or restricted maximum likelihood (REML) ([29]; [36]). Estimation under (3) can be fully implemented in `R` using functions of the `mgcv` package [37]. Then, given $X_{1,i}(t)$ and $X_{2,i}(t)$, the response curve can be estimated by $\widehat{Y}_i(t) = \mathbf{B}_\mu^T(t)\widehat{\boldsymbol{\Theta}}_\mu + \mathbf{Z}_{1,i}^T(t)\widehat{\boldsymbol{\Theta}}_1 + \mathbf{Z}_{2,i}^T(t)\widehat{\boldsymbol{\Theta}}_2$. Furthermore, one can estimate the marginal mean of response by $\widehat{\mu}_Y(t) = \mathbf{B}_\mu^T(t)\widehat{\boldsymbol{\Theta}}_\mu$. Modification of the estimation procedure for the case where the grid of points is irregular and sparse is presented in the Supplementary Material, Section A.

2.3 Variance estimation

The penalized criterion (3) 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 $\widehat{\boldsymbol{\Theta}}$. The variance of the parameter estimate $\widehat{\boldsymbol{\Theta}}$ can be calculated as $\mathrm{var}(\widehat{\boldsymbol{\Theta}}) = \mathbb{H}\{\sum_{i=1}^{n}\mathbb{Z}_i^T\mathbb{G}\mathbb{Z}_i\}\mathbb{H}^T$, where $\mathbb{G} = \mathrm{cov}(\mathbf{E}_i) = [G(t_j, t_k)]_{1\le j,k\le m}$ is the $m \times m$ covariance matrix evaluated corresponding to the observed time points. We model the 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 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 $\Sigma(s, t) = \sum_{k\ge1}\phi_k(s)\phi_k(t)\lambda_k$, where $\{\phi_k(\cdot), \lambda_k\}$ are the pairs of eigenfunctions/eigenvalues. 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., [44]; [46]) 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 the 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., [33]; [34]). 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 [33] for a discussion on this topic. One viable alternative is to model the past year's 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 a priori.

However, while the research was motivated by the fire management data, we have been unable to gain access to the data at the time of preparing this article. Instead, we will analyze the dietary calcium absorption data [2] and the gait data [27]. For the dietary calcium absorption data, our interest is to predict calcium absorption (as a function of age) of an individual based on their calcium intake over their age. In the context of gait data, the problem is to predict

knee angle of an individual as a function of percent of the gait cycle based on the hip angle. We will analyze both the data sets in Section 6.

Suppose that we wish to predict new, unknown response $Y_{\text{new}}(t_j)$ when new observations $X_{1,\text{new}}(t_j)$ and $X_{2,\text{new}}(t_j)$ $(j = 1, \ldots, m)$ are given. We assume that the model $Y_{\text{new}}(t) = \mu_Y(t) + F_1\{X_{1,\text{new}}(t), t\} + F_2\{X_{2,\text{new}}(t), t\} + \epsilon_{\text{new}}(t)$ still holds for the new data, where the error process $\epsilon_{\text{new}}(t)$ has the same distributional assumption as $\epsilon_i(t)$ in (1) and is independent of the new covariates $X_{1,\text{new}}(t)$ and $X_{2,\text{new}}(t)$. We predict the new response by

$$\begin{aligned}\widehat{Y}_{\text{new}}(t) &= \textstyle\sum_{d=1}^{K_\mu} B_{\mu,d}(t)\widehat{\theta}_{\mu,d} \\ &+ \textstyle\sum_{k=1}^{K_{x1}} \sum_{l=1}^{K_{t1}} B_{X1,k}\{X_{1,\text{new}}(t)\} B_{T1,l}(t)\widehat{\theta}_{1,k,l} \\ &+ \textstyle\sum_{k=1}^{K_{x2}} \sum_{l=1}^{K_{t2}} B_{X2,k}\{X_{2,\text{new}}(t)\} B_{T2,l}(t)\widehat{\theta}_{2,k,l},\end{aligned}$$

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 $\widehat{Y}_{\text{new}}(t)$ and the true response $Y_{\text{new}}(t)$. We follow an approach similar to [29] to estimate the prediction variance. Specifically, conditional on the new covariates $X_{1,\text{new}}(t)$ and $X_{2,\text{new}}(t)$, we have $\text{var}\{Y_{\text{new}}(t) - \widehat{Y}_{\text{new}}(t)\} = \text{var}\{\epsilon_{\text{new}}(t)\} + \text{var}\{\widehat{Y}_{\text{new}}(t)\}$. Note that $\epsilon_{\text{new}}(t)$ is a realization of the same error process with zero-mean and covariance structure $G(\cdot, \cdot)$. Let $\mathbf{Z}_{1,\text{new}}(t)$ be the $K_{x1}K_{t1}$-dimensional vector of $B_{X1,k}\{X_{1,\text{new}}(t)\}B_{T1,l}(t)$'s, and $\mathbf{Z}_{2,\text{new}}(t)$ be the $K_{x2}K_{t2}$-dimensional vectors of $B_{X2,k}\{X_{2,\text{new}}(t)\}B_{T2,l}(t)$'s as defined earlier. Also, let $\mathbf{Y}_{\text{new}}$ and $\widehat{\mathbf{Y}}_{\text{new}}$ be the m-dimensional vector of $Y_{\text{new}}(t_j)$'s and $\widehat{Y}_{\text{new}}(t_j)$'s respectively. Then the prediction variance becomes $\text{var}\{\mathbf{Y}_{\text{new}} - \widehat{\mathbf{Y}}_{\text{new}}\} = \mathbb{G} + \mathbb{Z}_{\text{new}}\mathbb{H}\{\sum_{i=1}^n \mathbb{Z}_i^T \mathbb{G}\mathbb{Z}_i\}\mathbb{H}^T\mathbb{Z}_{\text{new}}^T$, where $\mathbb{Z}_{\text{new}} = [\mathbb{B}_\mu|\mathbb{Z}_{1,\text{new}}|\mathbb{Z}_{2,\text{new}}]$ is $m \times (K_\mu + K_{x1}K_{t1} + K_{x2}K_{t2})$-dimensional matrix with the jth row given by $[\mathbf{B}_\mu^T(t_j)|\mathbf{Z}_{1,\text{new}}^T(t_j)|\mathbf{Z}_{2,\text{new}}^T(t_j)]$. 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 $Y_{\text{new}}(t)$ by $C_{1-\alpha}(t) = \widehat{Y}_{\text{new}}(t) \pm \Phi^{-1}(1 - \alpha/2)\left[\widehat{\text{var}}\{Y_{\text{new}}(t) - \widehat{Y}_{\text{new}}(t)\}\right]^{1/2}$ 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 $X_{1,\text{new}}(\cdot)$ and $X_{2,\text{new}}(\cdot)$ 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 the covariates 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 [21]. Details about the preliminary transformation are given in the Supplementary Material, Section B.

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. However, to the best of our knowledge, there are no currently available methods that allow testing the effect of functional covariates in nonlinear concurrent models. In this section, we consider the problem of testing whether the functional predictor variable is associated with the response. To illustrate our 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

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

where $F_0(t)$ is univariate function and $F_1(x_1, t)$ is bivariate function, both assumed unknown; see Section 5.2.2 for examples. Our testing procedure is based on first modeling the null effect $F_0(t)$ and the full model $F_1(x_1, t)$ using basis function expressions in a manner that ensures that the null model is nested within the full model. Specifically, we propose to use $\{B_{T,l}(t), l \geq 1\}$ to model $F_0(\cdot)$ under the null model, where $B_{T,l}(t)$ 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 $\{B_{X,0}(x_1) = 1, B_{X,k}(x_1), k \geq 1\}$ for x_1, where $B_{X,k}(x_1)$ are the B-splines defined over $\mathcal{X}_1$ - the state space of the process X_1, where $X_{1,i}(\cdot) \sim X_1(\cdot)$. Under the full model, we can write $F_1(x_1, t) = F_0(t) + \sum_{k=1}^\infty \sum_{l=1}^\infty B_{X,k}(x_1)B_{T,l}(t)\theta_{k,l}$. We propose to use the following F-type test statistic

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

where RSS_0 and df_0 are the residual sums of squares and the effective degrees-of-freedom [36] 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 a 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 follow the steps provided in Algorithm 1.

Algorithm 1 Bootstrap algorithm for testing global effect

1: Fit the full model described by the alternative hypothesis in (5) using the estimation procedure of the ANFCM. 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 ANFCM 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.

for $b = 1$ **to** B

5: Generate response curves under the null model as $Y_{b,i}^*(t) = \widehat{F}_0(t) + e_{b,i}^*(t)$.
6: 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^*.

end for

7: Compute the p-value by $\widehat{p} = [1 + \sum_{b=1}^B I\{T_b^* \geq T_{n,\text{obs}}\}]/(B+1)$.

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

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

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 bivariate 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 $\{B_{X1,0}(x_1) = 1, B_{X1,k}(x_1) : k \geq 1\}$ and $\{B_{T,l}(t) : l \geq 1\}$ for x_1 and t to model $F_0(x_1, t)$ under the null hypothesis; here, $B_{X1,k}(x_1)$ are the B-splines defined over the state space of the process X_1, and $B_{T,l}(t)$ 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 $\{B_{X1,0}(x_1) = 1, B_{X1,k}(x_1), B_{X2,k}(x_2) : k \geq 1\}$ for x_1 and x_2, where $B_{X2,k}(x_2)$ are defined over the state space of the process X_2. Therefore, under the alternative, we can write $E[Y(t)|X_1(t) = x_1, X_2(t) = x_2] = F_0(x_1, t) + \sum_{k=1}^{K_{x2}} \sum_{l=1}^{K_{t2}} B_{X2,k}(x_2) B_{T,l}(t) \theta_{2,k,l}$.

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 ANFCM. 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 ANFCM 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.

for $b = 1$ **to** B

5: Generate response curves under the null model as $Y_{b,i}^*(t) = \widehat{F}_0(x_1, t) + e_{b,i}^*(t)$.
6: 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^*.

end for

7: Compute the p-value by $\widehat{p} = [1 + \sum_{b=1}^B I\{T_b^* \geq T_{n,\text{obs}}\}]/(B+1)$.

Our proposed resampling algorithms described in Sections 3.1 and 3.2 account for the 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 not shown) that ignoring the correlation results in severely inflated type I error.

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; [46] proposed to first smooth each noisy trajectory using local

polynomial kernel smoothing, and then estimate the mean and standard deviation of the covariate $X_i(t_j)$ 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 ([44]) to the observed W_{ij}'s. [44] 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 by combining the newly estimated FPC scores and the estimated eigenfunctions. 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 $X_{\text{new}}(t_j)$ 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 [44] with the estimated eigenfunctions from the training data, implicitly assuming that the new covariate $X_{\text{new}}(\cdot)$ 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 $X_{\text{new}}(t_j)$.

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 6 we apply the proposed method to the gait study [24, 27, 15] and the dietary calcium absorption study [2, 31].

5.1 Simulation design

We generate 1,000 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 $\mathbf{E}_i = [\epsilon_i(t_{i1}), \ldots, \epsilon_i(t_{im_{Y,i}})]^T$: (i) $\mathbf{E}_i^1 \sim N(0, 0.9^2\mathbb{I}_{m_i})$; (ii) $\mathbf{E}_i^2 \sim N(0, 0.9^2\Sigma) + N(0, 0.9^2\mathbb{I}_{m_i})$ where Σ has $AR_{0.2}(1)$ structure; and (iii) $\mathbf{E}_i^3 \sim \xi_{i1}\sqrt{2}\cos(\pi t) + \xi_{i2}\sqrt{2}\sin(\pi t) + N(0, 0.9^2\mathbb{I}_{m_i})$, 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: $m_i \overset{iid}{\sim} Unif(20, 31)$ 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 width (IW) of prediction bands. The definition of these measures is given in the Supplementary Material, Section C.

Result from Scenario A and B: For these two scenarios, we fit both the standard linear functional concurrent model (FCM) and our proposed ANFCM for each of these choices. Our objective is two-fold: (i) When the true model is nonlinear, i.e. $F_1(x_1, t) = 1 + x_1 + t + 2x_1^2 t$, we expect to see that the ANFCM performs better than the 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 ANFCM still maintains prediction accuracy relative to the FCM. We obtained both the ANFCM 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 2 summarizes the predictive performance of our method for these two scenarios. We first discuss the case when the true function is nonlinear (the top two panels in Table 2). In this case, we observe that RMSPE^{in} and $\text{RMSPE}^{\text{out}}$ from fitting the ANFCM are smaller than those

Table 2. Summary of RMSPE^{in}, RMSPE^{out}, ICP, IW, and R(SE) based on 1000 simulated data sets. The models fitted by our method and the linear FCM are indicated by `ANFCM` and `FCM`, respectively

						$1-\alpha=0.95$						$1-\alpha=0.90$				$1-\alpha=0.85$			
		RMSPE$^{\text{in}}$		RMSPE$^{\text{out}}$		ICP		IW		R(SE)		ICP		IW		ICP		IW	
n	$\mathbf{E}_i$	ANFCM	FCM	ANFCM	FCM	ANFCM	FCM	ANFCM	FCM	ANFCM	FCM	ANFCM	FCM	ANFCM	FCM	ANFCM	FCM	ANFCM	FCM
Scenario B (true relationship is nonlinear), $m = 81$																			
100	$\mathbf{E}_i^1$	**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
	$\mathbf{E}_i^2$	**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
	$\mathbf{E}_i^3$	**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	$\mathbf{E}_i^1$	**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
	$\mathbf{E}_i^2$	**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
	$\mathbf{E}_i^3$	**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 nonlinear), $m_i \overset{iid}{\sim} Unif(20, 31)$																			
100	$\mathbf{E}_i^1$	**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
	$\mathbf{E}_i^2$	**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
	$\mathbf{E}_i^3$	**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	$\mathbf{E}_i^1$	**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
	$\mathbf{E}_i^2$	**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
	$\mathbf{E}_i^3$	**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	$\mathbf{E}_i^1$	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
	$\mathbf{E}_i^2$	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
	$\mathbf{E}_i^3$	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	$\mathbf{E}_i^1$	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
	$\mathbf{E}_i^2$	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
	$\mathbf{E}_i^3$	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	$\mathbf{E}_i^1$	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
	$\mathbf{E}_i^2$	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
	$\mathbf{E}_i^3$	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	$\mathbf{E}_i^1$	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
	$\mathbf{E}_i^2$	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
	$\mathbf{E}_i^3$	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

from the linear FCM in all cases, indicating that the proposed ANFCM outperforms the linear FCM. The ICPs from the ANFCM and the linear FCM are fairly close to the nominal levels of 0.95, 0.90, and 0.85. However, on average, the linear FCM produces larger intervals, indicated by larger IW values and wider range of the estimated standard error denoted by R(SE) compared to the ANFCM. Such patterns confirm that the variability in prediction is not properly captured by the linear FCM when the true model is nonlinear. For less complicated error patterns such as $\mathbf{E}_i^1$ (independent error structure), both models produce smaller prediction errors; nevertheless, ANFCM 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 show the results when the underlying model is linear. In this case, the ANFCM 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 nonlinear (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 ANFCM (solid line). This indicates that variance estimation from the linear FCM is less accurate. In the bottom panel, the prediction bands of the ANFCM (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.

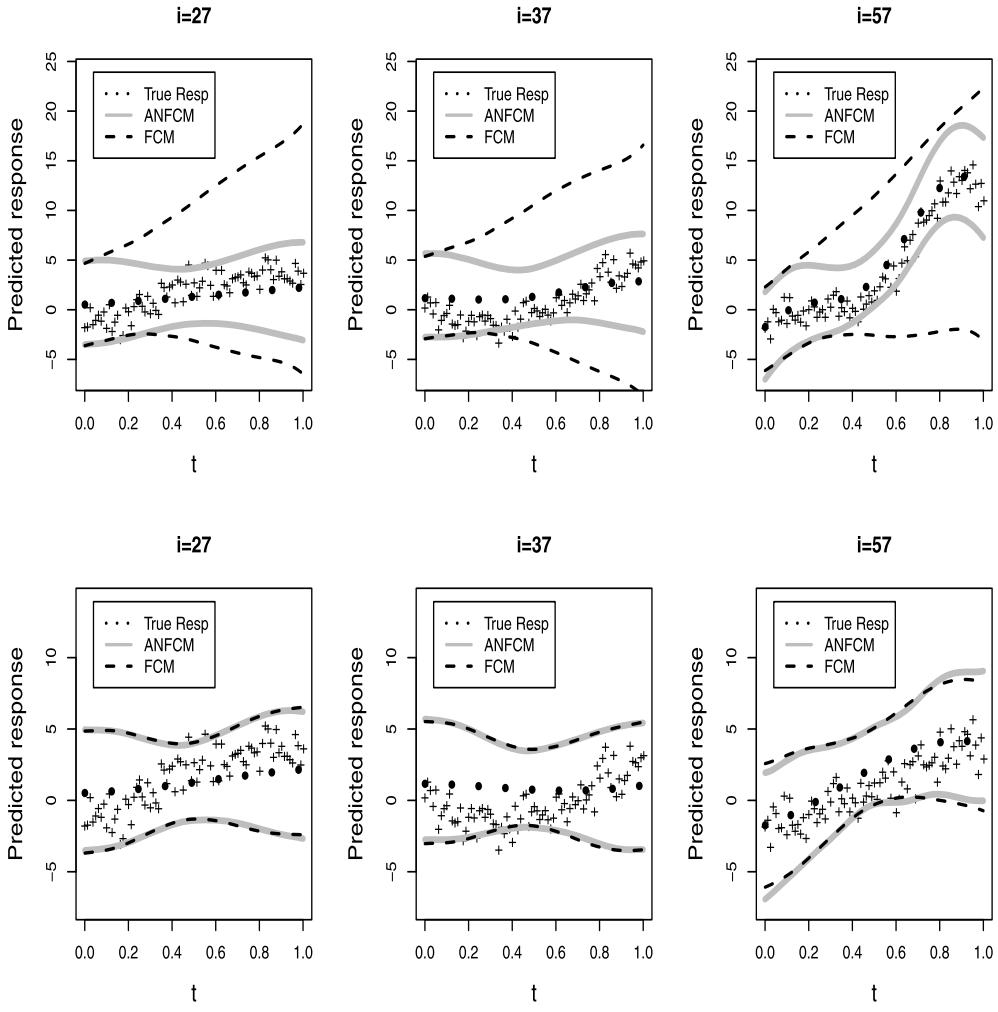

Figure 1. 95% prediction bands constructed for three subject-level trajectories in the test data. The results are from scenario A and B (involving a single functional covariate) for the setting where sampling design is dense, $n = 100$, and $\mathbf{E}_i = \mathbf{E}_i^3$. The top (bottom) panel corresponds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

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` ([37]) `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 (the model with two functional covariates).

We obtained the ANFCM 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 3 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 1,000 Monte Carlo replications. The top and bottom panels 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

Table 3. Summary of RMSPE^{in}, RMSPE^{out}, ICP and average computation time (in seconds) based on 1000 simulated data sets. The models are fitted by the estimation procedure of ANFCM. The average computation times are obtained using `gam` and `bam` functions in `mgcv` `R` package

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

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 ANFCM may result in a significant gain in prediction accuracy over the standard linear FCM, when the true model is nonlinear; ANFCM 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.

In the Supplementary Material, we include additional simulation results. Section D.1 reports simulation results corresponding to another level of sparseness. Section D.2 compares the results corresponding to two competitive approaches for covariance estimation: using local linear smoothing ([44]) which is implemented in `Matlab` using the functions of the `PACE` toolbox and using a fast covariance smoothing method ([41]) which is implemented using `fpca.face` function of the `refund R` package ([15]). Section D.3 presents additional results for larger measurement error variance ($\tau^2 = 1$ and 2) in functional covariates as well as a smaller sample size $n = 40$ (as in the Gait data example) in addition to $n = 100$ and 300. We also conducted an additional simulation study to investigate the effect of different choices of the number of basis functions. The results are displayed in Section D.4 of the Supplementary Material. We found that the prediction errors and the coverage at different nominal levels did not vary too much across different numbers of basis functions. Section D.5 investigates the model performance with up to five functional covariates through a numerical study.

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) $E[Y_d(t)|X_1(t) = x_1, X_2(t) = x_2] = 2t + t^2 + x_1 \sin(\pi t)/4 + d\{2\cos(x_2 t)\}$, where d is a constant. 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, we are interested in testing the null hypothesis in (5). Here, $F_1\{X(t), t\} = F_0(t) + d\{X(t)\beta_1(t)\}$, where $F_0(t) = 1 + 2t + t^2$ and $\beta_1(t) = t/8$. This scenario is designed to capture the standard linear concurrent effect of the functional covariate. When $d = 0$, the true model is a univariate function of time point t, and $X(\cdot)$ has no association with Y; when $d > 0$, the true model depends on both $X(t)$ and t in a linear concurrent relationship. Thus the parameter d indexes the departure from the null hypothesis given in (5) towards stronger linear concurrent relationship. This setting is similar to the real data analysis (in Section 6), where the functional covariate is found to have significant association with the response but performance of the proposed nonlinear model is similar to that of a linear concurrent model.

In Scenario B, we are interested in testing the null hypothesis given in (7). Here, $F_1\{X_1(t), t\} = X_1(t)\sin(\pi t)/4$ and $F_2\{X_2(t), t)\} = d\{2\cos(X_2(t)t)\}$, where the parameter $d \geq 0$ controls the departure from the null hypothesis given in (7). When $d = 0$, the response depends on X_1 using a linear concurrent relationship, but does not depend on X_2. However, when $d > 0$, Y has a linear concurrent relationship with X_1 and a nonlinear relationship with X_2. Thus Scenario B is designed to detect such a nonlinear relationship in the case with two covariates.

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 2,000 samples to assess the type I error rate, and 1,000 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 the top panel of Table 4. 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 $d = 0.1, 1, 2, \ldots, 7$ 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 the densely sampled scenario are provided in the Supplementary material, Section D.6. 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 4 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

Table 4. Type I error probabilities (×100) based on 2000 simulations. The values in the parenthesis are the estimated standard errors (×100) of the rejection probabilities

Scenario	Sampling	n	$\alpha = 5\%$ $\mathbf{E}_i^1$	$\mathbf{E}_i^2$	$\mathbf{E}_i^3$	$\alpha = 10\%$ $\mathbf{E}_i^1$	$\mathbf{E}_i^2$	$\mathbf{E}_i^3$
A	$m = 81$	100	5.3(0.5)	5.4(0.5)	6.6(0.6)	11.1(0.7)	10.6(0.7)	10.3(0.7)
		300	6.1(0.5)	5.0(0.5)	5.1(0.5)	10.6(0.7)	10.9(0.7)	9.4(0.7)
	$m_i \overset{iid}{\sim} Unif(20, 31)$	100	5.5(0.5)	5.0(0.5)	4.6(0.5)	10.0(0.7)	10.8(0.7)	9.8(0.7)
		300	4.8(0.5)	4.2(0.4)	5.2(0.5)	9.5(0.7)	10.4(0.7)	10.7(0.7)
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	6.4(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)

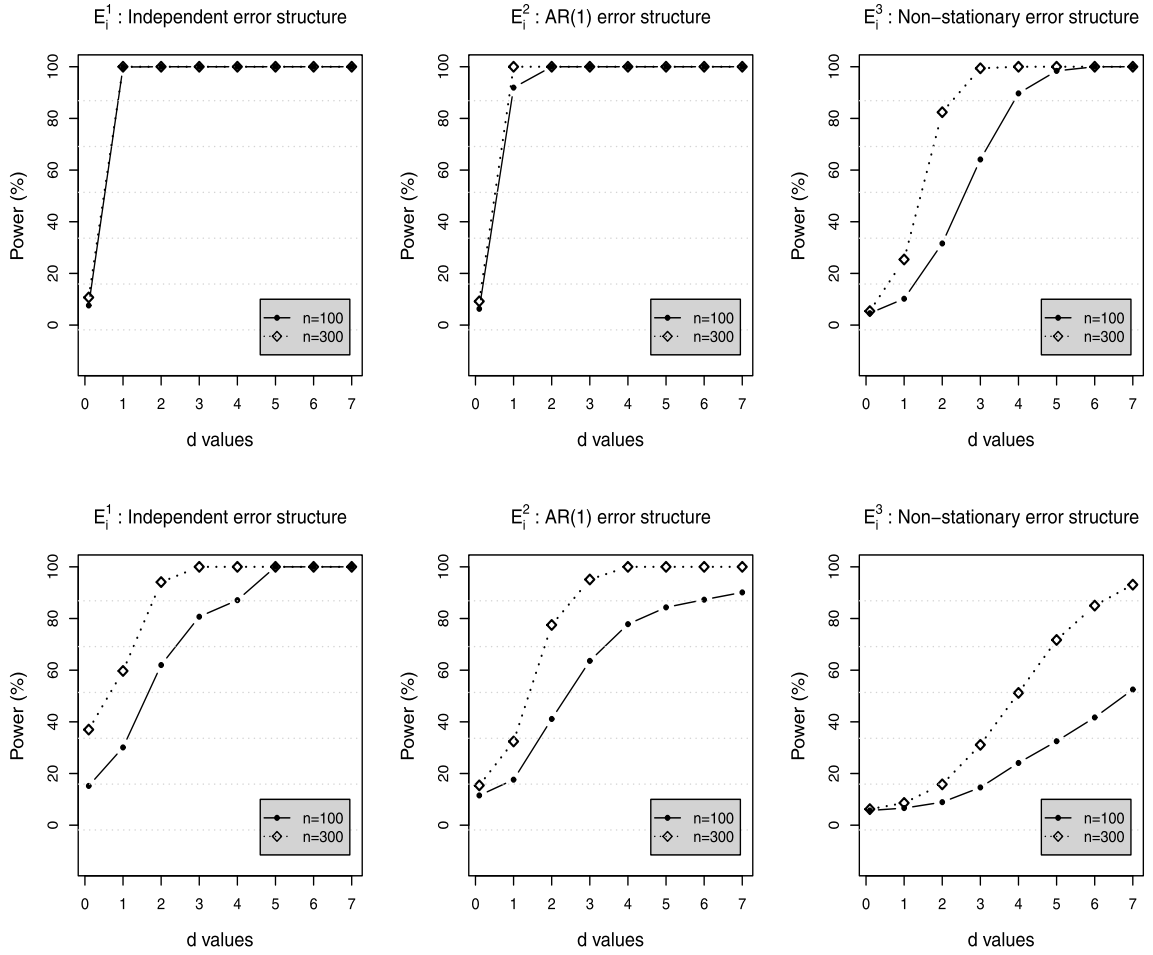

Figure 2. Powers (×100) of the tests at significance level $\alpha = 5\%$. The top (bottom) panel displays the results from scenario A (scenario B) for the setting where sampling design is sparse. The error process in the left, middle and right panels is assumed to be $\mathbf{E}_i^1$, $\mathbf{E}_i^2$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

displays the rejection probabilities at the $\alpha = 0.05$ level for $d = 0.1, 1, 2, \ldots, 7$ 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 patterns such as $\mathbf{E}_i^3$ (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 Material, Section D.6.

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.

6. APPLICATIONS

6.1 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 [35]. 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 [24, 27]. 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. In the Supplementary Material, Figure 2 of Section 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, which 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, were δ_{ij} are the measurement errors. The measurement error variance in the Hip angle curves are estimated (via FPCA) to be 1.714. 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 ANFCM, $\mathrm{E}[Y_i(t)|X_i(t)] = F\{X_i(t), t\}$, 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 ANFCM 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 ANFCM 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 ANFCM and the linear FCM, we report the in-sample and the out-of-sample RMSPE, the ICP, the IW, and the R(SE). For the LME model, we report similar measures, but take the average over the repeated measurements instead of integrating over the time domain.

The results are summarized in Table 5 (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 models are obviously better. ANFCM yields similar predictive performance relative to the linear FCM. Specifically, the prediction errors from the ANFCM are similar to the linear FCM. The R(SE) obtained from the ANFCM are only slightly narrower compared to FCM 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, ANFCM (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 ANFCM. The results corroborate that the relationship between the hip angles and the knee angles is linear. This finding is also confirmed by additional simulation results using a generating model that mimics the gait data, which are included in the Supplementary Material, Section E.2.

6.2 Dietary calcium absorption data

Next, we consider an application to dietary calcium absorption study [2]. 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. In the Supplementary Material, Figure 2 of Section E.1 displays the observed individual trajectories of calcium intake and absorption along the patient's age at the visit.

Table 5. Results from data examples as described in Section 6. Displayed are the summaries of RMSPE^{in}, RMSPE^{out}, ICP, IW, and R(SE). The models fitted by our method and the linear FCM are indicated by `ANFCM` and `FCM`, respectively

Gait data

	RMSPE$^{\text{in}}$	RMSPE$^{\text{out}}$	$1-\alpha=0.95$			$1-\alpha=0.90$			$1-\alpha=0.85$		
			ICP	IW	R(SE)	ICP	IW	R(SE)	ICP	IW	R(SE)
`ANFCM`	5.47	5.99	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

	RMSPE$^{\text{in}}$	RMSPE$^{\text{out}}$	$1-\alpha=0.95$			$1-\alpha=0.90$			$1-\alpha=0.85$		
			ICP	IW	R(SE)	ICP	IW	R(SE)	ICP	IW	R(SE)
`ANFCM`	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]

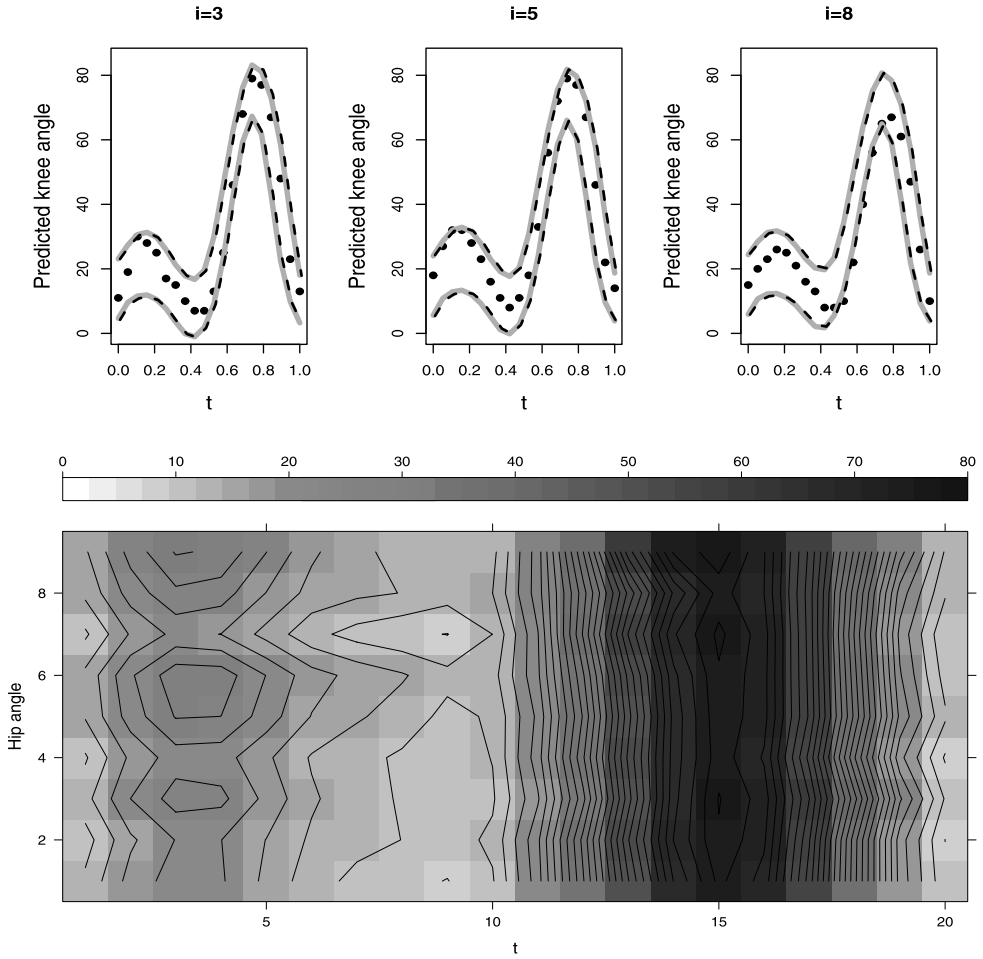

Figure 3. Results from gait data analysis as described in Section 6. The top panel displays 95% prediction bands obtained by fitting the ANFCM (grey solid lines) and the linear FCM (dashed lines) for three subject-level trajectories in the test data. "•" represent the knee angles in the test data. The bottom panel shows the heat map of $\widehat{Y}_{\text{new,i}}(t)$ obtained from the test data set of the gait data example.

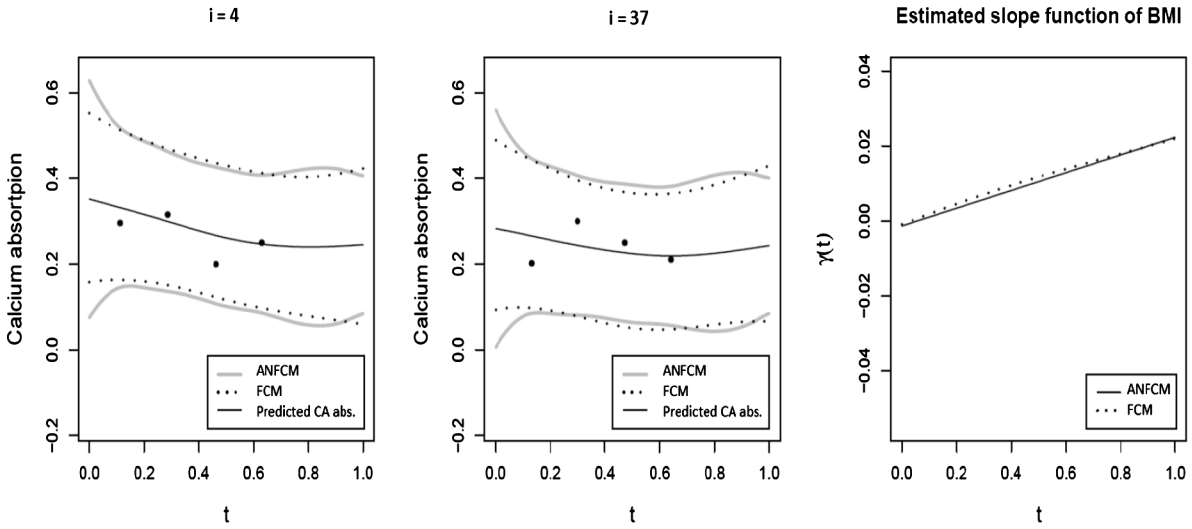

Figure 4. Results from calcium data analysis as described in Section 6. The leftmost two panels display 95% prediction bands obtained by fitting the ANFCM (grey solid lines) and the linear FCM (black dotted lines) for two subject-level trajectories in the test data. "•" indicate the calcium absorption measured, and thin solid lines are the predicted calcium absorption from the ANFCM estimation method. The rightmost panel displays the estimated slope function of BMI, $\widehat{\gamma}(t)$, obtained from the ANFCM (black solid line) and the linear FCM (black dotted line).

In the analysis, we fit a ANFCM with a linear BMI effect, $\mathrm{E}[Y_i(t)|X_{1,i}(t), X_{2,i}] = F\{X_{1,i}(t), t\} + \gamma(t)X_{2,i}$, 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 ANFCM, 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 absorption.

We next analyze the predictive performance of the ANFCM 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 [44]. The resulting estimated model (using the training data) is later used to reconstruct the trajectories in the test set. The prediction results using the ANFCM and the competitive linear FCM are presented in Table 5 (bottom panel). Both the ANFCM 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 ANFCM (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 ANFCM and the linear FCM. On 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.

In the Supplementary Material (Section E.3), we have also investigated the effect of choosing different number of basis functions, K_x and K_t, for the calcium absorption data. The results for three different choices $(K_x, K_t) = (7, 7), (9, 9)$ and $(11, 11)$ are similar. We also investigated another ANFCM for the calcium absorption data set with a bivariate effect of BMI and age, $\mathrm{E}[Y_i(t)|X_{1,i}(t), X_{2,i}] = F_1\{X_{1,i}(t), t\} + F_2(X_{2,i}, t)$, and obtained similar results.

7. DISCUSSION

We propose a wide class of function-on-function regression models, the additive nonlinear 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 Section F of the Supplementary Material.

Models similar to (1) are also considered by [30]; software implementing their approach is available in the `refund` package in R. However, the main difference is that [30] assume that the error process is white noise, that is, iid Gaussian errors; although functional random effects are allowed. In the specific context of nonlinear concurrent modeling, no numerical investigation has been done by [30] regarding estimation or prediction. In contrast, we present methodology and numerical results for prediction, estimation of prediction uncertainty as well as algorithms for two types of hypothesis testing.

We note that the model parameters are estimated by minimizing a penalized least squares criterion (3), and the smoothing parameters are chosen using REML or GCV. While minimizing (3) is reasonable regardless whether the errors are dependent or not, it can be viewed as a penalized likelihood criterion only when errors are assumed to be independent. As a result, we propose methods for variance estimation that takes into account the error covariance. One potential issue in this procedure is that the smoothing parameters are still chosen based on (3). While our numerical results show that the proposed method still has good performance in terms of prediction error and coverage, more research is needed to address this issue. Also, the variance estimation procedure does not take into account the variability due to the estimation of the smoothing parameters. It might be possible to develop methodology to overcome this issue using recently published works such as that of [38] (see, e.g., their Section 4). Finally, in our variance estimation procedure, we have essentially adopted a frequentist approach; however, it might be possible to implement a Bayesian variance estimation procedure (see, e.g., Chapter 4.8 of [36]) in our framework. All these issues are highly interesting research topics and are reserved for future research.

SUPPLEMENTAL MATERIAL

Online supplemental material is available in the submitted file http://intlpress.com/site/pub/pages/journals/items/sii/content/vols/0011/0004/s003.

ACKNOWLEDGEMENTS

Maity's research was supported by National Institute of Health grant R00 ES017744 and an NCSU Faculty Research and Professional Development grant. Staicu's research was supported by National Institute of Health grants R01 NS085211 and R01 MH086633 and National Science Foundation grant DMS 1454942. The authors are extremely grateful to the three anonymous referees whose suggestions and comments greatly improved the results and presentation of the article. Majority of this work was done while the first author was a PhD student at North Carolina State University.

Received 31 August 2016

REFERENCES

[1] Chiang, C. T., Rice, J. A. and Wu, C. O. (2001). Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *Journal of the American Statistical Association* **96** 605–619. MR1946428

[2] Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements*. New York, NY: Springer. MR1883764

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

[4] Eubank, R. L., Huang, C., Muñoz Maldonado, Y., Wang, N., Wang, S. and Buchanan, R. J. (2004). Smoothing spline estimation in varying-coefficient models. *Journal of the Royal Statistical Society, Ser. B* **66** 653–667.

[5] Fan, J. and Zhang, J. (2000). Two-Step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society, Ser. B* **62** 303–322. MR1749541

[6] Fan, J. and Zhang, W. (2008). Statistical methods with varying coefficient models. *Statistics and Its Interface* **1** 179–195.

[7] Fan, Y., Foutz, N., James, G. M. and Jank, W. (2014). Functional Response additive model estimation with online virtual stock markets. *Annual Review of Statistics and Its Application* **8** 2435–2460. MR3292504

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

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

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

[11] Hastie, T. and Tibshirani, R. (1993). Varying coefficient models. *Journal of the Royal Statistical Society, Ser. B* **55** 757–796.

[12] Hilgert, N., Manrique Chuquillanqui, T., Crambes, C. and Mas, A. (2015). Ridge regression for the functional concurrent model. *Presented at 60. World Statistics Congress ISI2015, Rio de Janeiro, BRA (2015-07-26 - 2015-07-31). http://prodinra.inra.fr/record/310110.*

[13] Hoover, D. R., Rice, J. A., Wu, C. O. and Yang, L. P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* **85** 809–822.

[14] Huang, J. Z., Wu, C. O. and Zhou, L. (2004). Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica* **14** 763–788.

[15] Huang, L., Scheipl, F., Goldsmith, J., Gellar, J., Harezlak, J., McLean, M. W., Swihart, B., Xiao, L., Crainiceanu, C. and Reiss, P. (2015). refund: Regression with Functional Data R package version 0.1-13.
[16] Ivanescu, A. E., Staicu, A.-M., Scheipl, F. and Greven, S. (2015). Penalized function on function regression. *Computational Statistics* **30** 539–568. MR3357075
[17] Jiang, C. R. and Wang, J. L. (2011). Functional single index models for longitudinal data. *The Annals of Statistics* **39** 362–388.
[18] Li, Y. and Ruppert, D. (2008). On the asymptotics of penalized splines. *Biometrika* **95** 415–436.
[19] Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *The Canadian Journal of statistics* **31** 115–128.
[20] Marx, B. D. and Eilers, P. H. C. (2005). Multivariate penalized signal regression. *Technometrics* **47** 13–22.
[21] McLean, M. W., Hooker, G., Staicu, A. M., Scheipl, F. and Ruppert, D. (2014). Functional generalized additive models. *Journal of Computational and Graphical Statistics* **23** 249–269.
[22] Morris, J. S. (2015). Functional Regression. *Annual Review of Statistics and Its Application* **2** 321–359.
[23] Morris, J. S. and Carroll, R. J. (2006). Wavelet-based functional mixed models. *Journal of the Royal Statistical Society: Series B* **68** 179–199.
[24] Olshen, R. A., Biden, E. N., Wyatt, M. P. and Sutherland, D. H. (1989). Gait analysis and the bootstrap. *The Annals of Statistics* **17** 1419–1440.
[25] Qu, A. and Li, R. (2006). Quadratic inference functions for varying coefficient models with longitudinal data. *Biometrics* **62** 379–391.
[26] Ramsay, J. O., Hooker, G. and Graves, S. (2009). *Functional data analysis in R and Matlab*. New York, NY: Springer.
[27] Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis*, 2nd ed. New York; Berlin: Springer.
[28] Reiss, P. T., Huang, L. and Mennes, M. (2010). Fast function-on-scalar regression with penalized basis expansions. *International Journal of Biostatistics* **6**.
[29] Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric regression*. Cambridge, New York: Cambridge University Press.
[30] Scheipl, F., Staicu, A. M. and Greven, S. (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics* **24** 477–501.
[31] Sentürk, D. and Nguyen, D. V. (2011). Varying coefficient models for sparse noise-contaminated longitudinal data. *Statistica Sinica* **21** 1831–1856.
[32] Shen, Q. and Faraway, J. (2004). An F test for linear models with functional responses. *Statistica Sinica* **14** 1239–1257.
[33] Slijepcevic, A., Anderson, W. R. and Matthews, S. (2013). Testing existing models for predicting hourly variation in fine fuel moisture in eucalypt forests. *Forest Ecology and Management* **306** 202–215.
[34] Slijepcevic, A., Anderson, W. R., Matthews, S. and Anderson, D. H. (2015). Evaluating models to predict daily fine fuel moisture content in eucalypt forest. *Forest Ecology and Management* **335** 261–269.
[35] Theologis, T. N. (2009). *Children's Orthopaedics and Fractures (Chapter 6)*. London: Springer.
[36] Wood, S. N. (2006). *Generalized Additive Models: An Introduction with R*. Boca Raton, Florida: Chapman and Hall/CRC.
[37] Wood, S. N. (2015). Mixed GAM computation vehicle with GCV/AIC/REML smoothness estimation R package version 1.8-7.
[38] Wood, S. N., Pya, N. and Safken, B. (2017). Smoothing parameter and model selection for general smooth models. *Journal of the American Statistical Association* **111** 1548–1563.
[39] Wu, C. O., Chiang, C. T. and Hoover, D. R. (1998). Asymptotic confidence regions for kernel smoothing of the varying coefficients model. *Journal of the American Statistical Association* **93** 1388–1402.
[40] Wu, Y., Fan, J. and Müller, H. G. (2010). Varying-coefficient functional linear regression. *Bernoulli* **16** 730–758.
[41] Xiao, L., Zipunnikov, V., Ruppert, D. and Crainiceanu, C. (2016). Fast covariance function estimation for high-dimensional functional data. *Statistics and Computing* **26** 409–421.
[42] Xu, H., Shen, Q., Yang, X. and Shoptaw, S. (2011). A quasi F-test for functional linear models with functional covariates and its application to longitudinal data. *Statistics in Medicine* **30** 2842–2853.
[43] Yao, F., Müller, H. G. and Wang, J. L. (2005). Functional linear regression analysis for longitudinal data. *The Annals of Statistics* **33** 2873–2903. MR2253106
[44] Yao, F., Müller, H. G. and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100** 577–590.
[45] Zhang, J., Clayton, M. K. and Townsend, P. A. (2011). Functional Concurrent Linear Regression Model for Spatial Images. *Journal of Agricultural, Biological, and Environmental Statistics* **16** 105–130.
[46] Zhang, J. T. and Chen, J. (2007). Statistical inference for functional data. *The Annals of Statistics* **35** 1052–1079.

Janet S. Kim
Astellas Pharma Global Development, Inc.
Northbrook, Illinois 60062
USA
E-mail address: janet.kim@astellas.com

Arnab Maity
Department of Statistics
North Carolina State University
Raleigh, North Carolina 27695
USA
E-mail address: amaity@ncsu.edu

Ana-Maria Staicu
Department of Statistics
North Carolina State University
Raleigh, North Carolina 27695
USA
E-mail address: ana-maria_staicu@ncsu.edu