HHS Public Access

Author manuscript

J Comput Graph Stat. Author manuscript; available in PMC 2018 May 17.

Published in final edited form as:

J Comput Graph Stat. 2018 ; 27(1): 234–244. doi:10.1080/10618600.2017.1356730.

Additive Function-on-Function Regression

Janet S. Kim[*], **Ana-Maria Staicu**[†], **Arnab Maity**[‡], **Raymond J. Carroll**[§], and **David Ruppert**[¶]

[*]Department of Statistics, North Carolina State University

[†]Department of Statistics, North Carolina State University

[‡]Department of Statistics, North Carolina State University

[§]Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143 USA and School of Mathematical and Physical Sciences, University of Technology Sydney, Broadway NSW 2007, Australia

[¶]School of Operations Research and Information Engineering and Department of Statistical Science, Cornell University

Abstract

We study additive function-on-function regression where the mean response at a particular time point depends on the time point itself, as well as the entire covariate trajectory. We develop a computationally efficient estimation methodology based on a novel combination of spline bases with an eigenbasis to represent the trivariate kernel function. We discuss prediction of a new response trajectory, propose an inference procedure that accounts for total variability in the predicted response curves, and construct pointwise prediction intervals. The estimation/inferential procedure accommodates realistic scenarios, such as correlated error structure as well as sparse and/or irregular designs. We investigate our methodology in finite sample size through simulations and two real data applications. Supplementary Material for this article is available online.

Keywords

Functional data analysis; Eigenbasis; Nonlinear models; Orthogonal projection; Penalized B-splines; Prediction

1 Introduction

Regression models where both the response and the covariate are curves have become common in many scientific fields such as medicine, finance, and agriculture. These models are often called function-on-function regression. One of the commonly known models is the functional concurrent model where the current response relates to the current values of the covariate/s; see for example, Ramsay and Silverman (2005); Sentürk and Nguyen (2011);

Supplementary Material

Supplementary Material: Additional descriptions for methodology extensions, simulation setup, and simulation results are provided. (pdf file)

R code: The R code developed for the simulation. (zip file)

Kim et al. (2016). When the current response depends on the past values of the covariate/s, the historical functional linear model (Malfait and Ramsay, 2003) is more appropriate.

We consider functional regression models that relate the current response to the full trajectory of the covariate. The functional linear model (Ramsay and Silverman, 2005; Yao et al., 2005b; Wu et al., 2010) assumes that the relationship is linear: the effect of the full covariate trajectory is modeled through a weighted integral using an unknown bivariate coefficient function as the weights. The linearity assumption was extended to the functional additive model (FAM) of Müller and Yao (2008), which models the effect of the covariate by a sum of smooth functions of the functional principal component scores of the covariate. A limitation of this approach is that the estimated effects are not easily interpretable. This paper considers flexible nonlinear regression models that can capture complex relationships between the response and the full covariate trajectory more directly.

Additive models have enjoyed great popularity since they were introduced by Friedman and Stuetzle (1981) for a scalar response and scalar predictors. Their model replaces the linear model $Y_i = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p} + \varepsilon_i$, where each of Y_i, $X_{i,1}$, …, $X_{i,p}$, $i = 1, \ldots, n$, is scalar by, $Y_i = \beta_0 + f_1(X_{i,1}) + \cdots f_p(X_{i,p}) + \varepsilon_i$. Here f_1, …, f_p are smooth functions. Additive models allow nonparametric modeling of the relationship between the response and the predictors while avoiding the so-called curse of dimensionality and being easily interpreted. Additive and generalized additive models for a scalar response and functional predictors were introduced by McLean et al. (2014) and Müller et al. (2013).

Additive function-on-function regression models where the current mean response depends on the time point itself as well as the full covariate trajectory were introduced by Scheipl et al. (2015), but the present paper is the first to investigate them fully. We develop a novel estimation procedure that is an order of magnitude faster than the existing algorithm and discuss inference for the predicted response curves. The methodology is applicable for realistic scenarios involving densely and/or sparsely observed functional responses and predictors, as well as various residual dependence structures.

There are three major contributions in this paper. First, we combine B-spline bases for the covariate function, $X(s)$, and for its argument, s, (Marx and Eilers, 2005; Wood, 2006; McLean et al., 2014) with a functional principal component basis for the argument, t, of the response function; see (2) below. The replacement of the spline basis used by Scheipl et al. (2015) for the argument of the response function with an eigenbasis is a key step to creating a computationally efficient algorithm. Second, we develop inferential methods for out-of-sample prediction of the full response trajectories, accounting for correlation in the error process. Finally, our method accommodates realistic scenarios such as densely or sparsely observed functional responses and covariates, possibly corrupted by measurement error. We show numerically that when the true relationship is nonlinear, our model provides improved predictions over the functional linear model. At the same time, when the true relationship is linear, our model maintains prediction accuracy and its fit is comparable to that of a functional linear model.

Section 2 introduces our modeling framework and a novel estimation procedure, that we call AFF-PC (additive function-on-function regression with a principal component basis). Additionally it discusses implementation details and extensions. Section 3 discusses out-of-sample prediction inference. In Section 4, we investigate the performance of AFF-PC through simulations. Section 5 presents an application of AFF-PC to a bike share study. A second application, to yield curves, is in the Supplementary Material. Section 6 provides a brief discussion.

2 Methodology

2.1 Statistical Framework and Modeling

Suppose for $i = 1, \ldots, n$ we observe $\{(X_{ik}, s_{ik}) : k = 1, \ldots, m_i\}$ and $\{(Y_{ij}, t_{ij}) : j = 1, \ldots, m_{Y,i}\}$, where X_{ik} and Y_{ij} are the covariate and response observed at time points s_{ik} and t_{ij}, respectively. We assume that $s_{ik} \in \mathcal{T}_X$ for all i and k and $t_{ij} \in \mathcal{T}_Y$ for all i and j, where $\mathcal{T}_X$ and $\mathcal{T}_Y$ are compact intervals. It is assumed that $X_{ik} = X_i(s_{ik})$, where $X_i(\cdot)$ is a square-integrable, true smooth signal defined on $\mathcal{T}_X$. It is further assumed that $Y_{ij} = Y_i(t_{ij})$, where $Y_i(\cdot)$ is defined on $\mathcal{T}_Y$. For convenience, we assume that the response has zero-mean. In practice, this is achieved by de-meaning, that is, by subtracting the response sample mean from the individual response curves.

To illustrate our ideas, we assume that both the response and the predictor are observed on a fine, regular, and common grid of points so that $s_{ik} = s_k$ with $k = 1, \ldots, m$ and $t_{ij} = t_j$ with $j = 1, \ldots, m_Y$ for all i. This assumption, as well as the assumption that the functional covariate is observed on a fine grid and without noise are made for illustration only; our methodology can accommodate more general situations, as we show in Section 4.

We consider a general additive function-on-function regression model

$$Y_i(t) = \int_{\mathcal{T}_X} F\{X_i(s), s, t\} ds + \varepsilon_i(t), \quad (1)$$

where $F(\cdot, \cdot, t)$ is an unknown smooth trivariate function defined on $\mathbb{R} \times \mathcal{T}_X \times \mathcal{T}_Y$, and $\varepsilon_i(\cdot)$ is an error process with mean zero and unknown autocovariance function $R(t, t')$ and is independent of the covariate $X_i(s)$. Model (1) was introduced by Scheipl et al. (2015). The form $F(\cdot, \cdot, t)$ quantifies the unknown dependence between the current response $Y_i(t)$ and the full covariate trajectory $X_i(\cdot)$. If $F(x, s, t) = \beta(s, t)x$, then model (1) reduces to the standard functional linear model. In principle, this allows us to study whether an additive rather than a linear functional model is necessary, but this topic is left for future research.

One possible approach for modeling F is using a tensor product of univariate B-spline basis functions for x, s, and t. This approach was proposed by Scheipl et al. (2015) and implemented in the R package `refund` (Goldsmith et al., 2016), although the accuracy of their estimation approach has been investigated neither numerically nor theoretically. As

expected, and also as observed in our numerical results, using a trivariate spline basis imposes a heavy computational burden. The main reason for the high computational cost is that the trivariate spline basis requires a large number of basis functions. For example, if F is modeled using a tensor product of 10 basis functions per dimension, then there are $10^3 = 1000$ basis functions in total. Secondly, this estimation methodology requires selection of three smoothing parameters, one for each spline basis, which is computationally very expensive. Thirdly, the associated penalized criterion uses the response data directly, rather than the projection of the data onto a lower dimension basis. In this paper we consider an alternative approach that uses a low-rank representation of the response data and, since we have only two spline bases, fewer smoothing parameters. The low-dimensional representation of the response curves, especially, leads to computationally efficient estimation. We will refer to the Scheipl et al. (2015) algorithm as AFF-S, where "S" refers to the spline basis for t in $F(x, s, t)$. Our algorithm uses a principal component basis for t and so is called AFF-PC.

For some insight, consider a smooth function $\phi(\cdot) \in L^2(\mathcal{T}_Y)$ and let $y_{i,\phi} = \int_{\mathcal{T}_Y} Y_i(t)\phi(t)dt$ be the projection of Y_i onto $\phi(\cdot)$. Model (1) implies that $y_{i,\phi} = \int_{\mathcal{T}_Y}\int_{\mathcal{T}_X} F\{X_i(s), s, t\}\phi(t)dsdt + e_{i,\phi} = \int_{\mathcal{T}_X} G_\phi\{X_i(s), s\}ds + e_{i,\phi}$, where $G_\phi\{X_i(s), s\} = \int_{\mathcal{T}_Y} F\{X_i(s), s, t\}\phi(t)dt$ and $e_{i,\phi} = \int_{\mathcal{T}_Y} \varepsilon_i(t)\phi(t)dt$, assuming these integrals exist. The implied final model is exactly the one proposed by McLean et al. (2014); thus the unknown bivariate function $G_\phi(\cdot, \cdot)$ can be estimated by modeling it using a tensor product of two univariate known bases functions and controlling its smoothness through two tuning parameters.

Inspired by this result, let $\{\phi_k(\cdot)\}_k$ be an orthogonal basis in $L^2(\mathcal{T}_Y)$: $\int_{\mathcal{T}_Y} \phi_k(t)\phi_{k'}(t)dt = 1$ if $k = k'$ and 0 otherwise. We represent the function $F(x, s, t)$ as $F(x, s, t) = \sum_{k=1}^{\infty} G_k(x, s)\phi_k(t)$. Here, $G_k(x, s) = \int_{\mathcal{T}_Y} F(x, s, t)\phi_k(t)dt$, $k = 1, \ldots$, are unknown basis coefficients that vary smoothly over x and s. We model $G_k(\cdot, \cdot)$ as a tensor product of spline bases, $G_k(x, s) = \sum_{l=1}^{K_x}\sum_{l'=1}^{K_s} B_{X,l}(x)B_{S,l'}(s)\theta_{l,l',k}$, where $\{B_{X,l}(x)\}_{l=1}^{K_x}$ and $\{B_{S,l'}(s)\}_{l'=1}^{K_s}$ are orthogonalized B-spline bases (Redd, 2012) of dimensions K_x and K_s, respectively. Combining these expansions, the trivariate "kernel" function F can be written as

$$F(x, s, t) = \sum_{k=1}^{\infty}\sum_{l=1}^{K_x}\sum_{l'=1}^{K_s} B_{X,l}(x)B_{S,l'}(s)\phi_k(t)\theta_{l,l',k}, \quad (2)$$

where $\theta_{l,l',k}$ are the unknown parameters. In practice, we truncate the summation in k at some finite K. Thus, this representation uses trivariate basis functions obtained by the tensor

product of univariate B-spline bases functions in directions x and s and $L^2(\mathcal{T}_Y)$ orthogonal basis functions, $\phi_k(\cdot)$. Let $\mathbb{Z}_i$ be the K_xK_s-column vector of $\int_{\mathcal{T}_Y} B_{X,l}\{X_i(s)\}B_{S,l'}(s)ds$ and let Θ_k be the K_xK_s-column vector of unknown coefficients $\theta_{l,l',k}$, where $l = 1, \ldots, K_x$, $l' = 1, \ldots, K_s$. Then, model (1) can be approximated as

$$Y_i(t) \approx \sum_{k=1}^{K} \mathbb{Z}_i^T \Theta_k \phi_k(t) + \varepsilon_i(t). \quad (3)$$

2.2 Estimation and Prediction

2.2.1 Estimation—We estimate the unknown Θ_k's parameters in (3) by penalized least squares. However, unlike the standard penalized likelihood approaches (Ruppert et al. 2003; Wood 2006), which penalize the basis coefficients in all directions, we use quadratic penalties for the directions x and s, and control the roughness in the direction t by the number of orthogonal basis functions, K. Here $\otimes$ is the Kronecker product, and I_K is the identity matrix of dimension K. Specifically, the curvature in the x-direction is measured through $\int\int\int \{\partial^2 F(x,s,t)/\partial x^2\}^2 dxdsdt = \sum_{k=1}^{K} \int\int \{\partial^2 G_k(x,s)/\partial x^2\}^2 dxds$, where $\mathbb{P}_x$ is $= \sum_{k=1}^{K} \Theta_k^T (\mathbb{P}_x \otimes I_{K_s}) \Theta_k$

the $K_x \times K_x$ penalty matrix with the (l, r) entry equal to $\int \{\partial_{xx} B_{X,l}(x)\}\{\partial_{xx} B_{X,r}(x)\}dx$ $l, r = 1, \ldots, K_x$. Using the orthogonality of $\{\phi_k, k = 1, \ldots, K\}$, the curvature in the s-direction is

$$\int\int\int \left\{\frac{\partial^2 F(x,s,t)}{\partial s^2}\right\}^2 dxdsdt = \sum_{k=1}^{K} \int\int \left\{\frac{\partial^2 G_k(x,s)}{\partial s^2}\right\}^2 dxds = \sum_{k=1}^{K} \Theta_k^T (I_{K_x} \otimes \mathbb{P}_s) \Theta_k,$$

and $\mathbb{P}_s$ is the $K_s \times K_s$ penalty matrix with the (l, r) entry equal to $\int \{\partial_{ss} B_{S,l'}(s)\}\{\partial_{ss} B_{S,r'}(s)\}ds$ $(l', r' = 1, \ldots, K_s)$. The penalized criterion to be minimized is

$$\sum_{i=1}^{n} \|Y_i(\cdot) - \sum_{k=1}^{K} \mathbb{Z}_i^T \Theta_k \phi_k(\cdot)\|^2 + \sum_{k=1}^{K} \Theta_k^T (\lambda_x \mathbb{P}_x \otimes I_{K_s} + \lambda_s I_{K_x} \otimes \mathbb{P}_s) \Theta_k, \quad (4)$$

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

One convenient way to calculate the first term in (4) is to expand $Y_i(\cdot)$ using the same basis functions $\{\phi_k(\cdot)\}_k$. Specifically, if $\{\phi_k(\cdot)\}_k$ is the eigenbasis of the marginal covariance of $Y_i(\cdot)$, then Karhunen-Loève (KL) expansion yields $Y_i(t) = \sum_k \xi_{ik}\phi_k(t) + e_{it}$ where e_{it} is a zero-mean error and $\xi_{ik} = \int_{\mathcal{T}_Y} Y_i(t)\phi_k(t)dt$; recall that the marginal mean of $Y_i(\cdot)$ is assumed

to be zero. Here “marginal” means marginalized over the covariate function. Criterion (4) can be equivalently written as

$$\sum_{k=1}^{K}\left[\sum_{i=1}^{n}\{\xi_{ik}-\mathbb{Z}_i^T\Theta_k\}^2+\Theta_k^T(\lambda_x\mathbb{P}_x\otimes I_{K_s}+\lambda_s I_{K_x}\otimes\mathbb{P}_s)\Theta_k\right]. \quad (5)$$

Using the eigenbasis of the response covariance allows a low-dimensional representation of (4) that improves computation time and yet preserves model complexity. Our numerical results show that AFF-PC is orders of magnitude faster than its closest competitor, AFF-S; see Table 2.

We set K_x and K_s to be sufficiently large to capture the complexity of the model and penalize the basis coefficients to balance the bias and the variance. The smoothness parameters λ_x and λ_s can be chosen based on appropriate criteria, such as generalized cross validation (GCV) (see e.g., Ruppert et al., 2003; Wood, 2006) or restricted maximum likelihood (REML) (see e.g., Ruppert et al., 2003; Wood, 2006). In our numerical studies, we let K_x and K_s be as large as possible, while ensuring that $K_xK_s < n$, and select the smoothness parameters using REML.

The penalized criterion (5) uses the true functional principal component (FPC) scores. In practice, we use estimates of FPC scores from functional principal component analysis (FPCA), as we show next. Using the eigenbasis of the marginal covariance of the response, rather than a spline basis, is appealing because of the resulting parsimonious representation of the response and has been often used in the literature; see for example, Aston et al. (2010); Jiang and Wang (2010); Park and Staicu (2015). This choice of orthogonal basis also allows us to formulate the mean model for the conditional response profile, given scalar/vector covariates, based on mean models for the conditional FPC scores given the covariates: $\mathrm{E}[Y_i(t)|X_i(\cdot)]=\sum_{k=1}^{K}\phi_k(t)\mathrm{E}[\xi_{ik}|X_i(\cdot)]$, where $\mathrm{E}[\xi_{ik}|X_i(\cdot)]=\int_{\mathcal{T}_X}G_k\{X_i(s),s\}ds$, $G_k(\cdot,\cdot)$ are unknown bivariate functions and ξ_{ik} are the FPC scores of response. The representation is novel and extends ideas of Aston et al. (2010) and Pomann et al. (2015) to the case of a functional covariate. Also, it is related to Müller and Yao (2008) for $\mathrm{E}[\xi_{ik}|X_i(\cdot)]=\sum_{m=1}^{M}f_{km}(\xi_{im}^X)$, where $f_{km}(\cdot)$ are unknown smooth functions for $m=1,\ldots,M$ and $k=1,\ldots,K$, $\{\xi_{i1}^X,\ldots,\xi_{iM}^X\}$ are the FPC scores of the functional covariate $X_i(\cdot)$, and M is a finite truncation.

2.2.2 Prediction—We use the following notation: ‘^’ for prediction based on the function-on-function regression model and ‘~’ for estimation based on the marginal analysis of response $Y_i(\cdot)$. Estimation and prediction of the response curves $Y_i(\cdot)$ follows a three-step procedure: 1) reconstruct the smooth trajectory of the response $\widetilde{Y}_i(\cdot)$ by smoothing the data for each I (Zhang and Chen, 2007) and de-mean it, $\widetilde{Y}_i^c(\cdot)=\widetilde{Y}_i(\cdot)-\widetilde{\mu}_Y(\cdot)$ where $\widetilde{\mu}_Y(\cdot)$ is the estimated mean function; 2) use functional principal components analysis (PCA) to estimate the eigenbasis $\widetilde{\phi}_k(\cdot)$ of the (marginal) covariance of $\widetilde{Y}_i(\cdot)$, and then obtain the functional

PCA scores $\tilde{\xi}_{ik} = \int_{\mathcal{T}_Y} \tilde{Y}_i^c(t)\tilde{\phi}_k(t)dt$; and 3) Obtain estimates $\hat{\Theta}_k$, $k = 1, \ldots, K$, of the basis coefficients by minimizing the penalized criterion (5) with respect to Θ_k's, and using $\tilde{\xi}_{ik}$ in place of ξ_{ik}. The truncation point K is determined through a pre-specified percent of variance explained; in our numerical work we use 95%. For fixed smoothness parameters, the minimizer of (5) has a closed form:

$$\widehat{\Theta}_k = H_\lambda\left(\sum_{i=1}^n \mathbb{Z}_i \tilde{\xi}_{ik}\right), \quad (6)$$

where $H_\lambda = \left(\sum_{i=1}^n \mathbb{Z}_i \mathbb{Z}_i^T + P_\lambda\right)^{-1}$, $P_\lambda = \lambda_x \mathbb{P}_x \otimes I_{K_s} + \lambda_s I_{K_x} \otimes \mathbb{P}_s$, and $\lambda = (\lambda_x, \lambda_s)^T$. Once the basis coefficients are estimated, $F(\cdot, \cdot, \cdot)$ can be estimated by

$$\hat{F}(x, s, t) = \sum_{k=1}^K \sum_{l=1}^{K_x} \sum_{l'=1}^{K_s} B_{X,l}(x) B_{S,l'}(s) \phi_k(t) \hat{\theta}_{l,l',k}.$$

Furthermore, for any $X(s)$, the response curve can be predicted by the estimated $E[Y|X(\cdot)]$,

$$\hat{Y}(t) = \sum_{k=1}^K \tilde{\phi}_k(t) \left[\sum_{l=1}^{K_x} \sum_{l'=1}^{K_s} \hat{\theta}_{l,l',k} \int_{\mathcal{T}_X} B_{X,l}\{X(s)\} B_{S,l'}(s) ds\right], \quad (7)$$

which is obtained by plugging in the expression of $\hat{F}(x, s, t)$ into the integral term of (1).

2.3 Implementation and Extensions

Implementation of our method requires transformation of the covariate as a preliminary step since the realizations of the covariate functions $\{X_i(s_k): k, i\}$ may not be dense over the entire domain of the B-spline basis functions for x. In this situation, some of the B-spline basis functions may not have observed data on their support. This problem has been addressed by McLean et al. (2014) and Kim et al. (2016) with different strategies. This paper uses pointwise center/scaling transformation of the functional covariate proposed by Kim et al. (2016). For completeness, we present the full details in Section B of the Supplementary Material.

We have presented our methodology for the case where, for each subject, the functional covariate is observed on a fine grid and without measurement error. The approach can be easily modified to accommodate a variety of other realistic settings such as noisy functional covariates observed on either a dense or sparse grid of points for each subject, or a functional response observed on a sparse grid of points for each subject. Details on the necessary modifications are provided in the Supplementary Material, Section A. Our numerical investigation, to be discussed in Section 4, considers settings where the functional covariates are observed at dense or moderately sparse grids of points and the measurements are corrupted with noise.

3 Out-of-Sample Prediction

In this section, we focus on out-of-sample prediction and its associated inference. For example, in the capital bike share study (Fanaee-T and Gama, 2013), an important research objective is to understand better how the hourly temperature profile for a weekend day affects the bike rental patterns for that day. The idea is that nowadays with reasonably accurate weather forecasts, AFF-PC could be applied to the next day's weather forecast to predict bike rental demand that day; this could help the company avoid deploying too many or too few bikes for rental.

Inference for predicted response curves is not straightforward due to two important sources of variability: (1) uncertainty produced by predicting response curves *conditional* on the particular estimate of the orthogonal basis $\{\phi_k(\cdot)\}_k$ and (2) uncertainty induced by estimating the eigenbasis $\{\phi_k(\cdot)\}_k$. Ignoring the second source of variability could cause underestimation of total variance. Inspired by the ideas of Goldsmith et al. (2013), we assess the total variability of the predicted response curves by combining the two sources of variability. As the two sources are assessed based on the estimated error covariance, we first describe the estimation of the error covariance in Section 3.1, and then discuss the out-of-sample prediction inference in Section 3.2.

Let $\tilde{\xi}_{ik} = \int_{\mathcal{T}_Y} \tilde{Y}_i^c(t)\tilde{\phi}_k(t)dt$ be the projection of the de-meaned full response curve onto $\tilde{\phi}_k(t)$; recall that the $\{\tilde{\phi}_k(\cdot)\}_k$ are obtained from the spectral decomposition of the estimated marginal covariance of the response. Define $\text{var}(\xi_{ik}) = \sigma_k^2$, $\text{var}(\tilde{\xi}_{ik}) = \nu_{kk}$, and $\text{cov}(\tilde{\xi}_{ik}, \tilde{\xi}_{ik'}) = \nu_{kk'}(k \neq k')$. For notational simplicity, let $\eta = \left[K, \{\sigma_k^2, \phi_k(\cdot)\}_{k=1}^K\right]$ be the set of all parameters that describe the marginal covariance of the response.

3.1 Estimation of Error Covariance

For inference about the model parameters, we account for dependence of the errors using ideas similar to those of Kim et al. (2016). We assume that the covariance function of $\varepsilon(t)$, denoted by $R(t, t')$, can be decomposed as $R(t, t') = \Sigma(t, t') + \sigma^2 I(t = t')$, where $\Sigma(t, t')$ is a continuous covariance function, $\sigma^2 > 0$, and $I(\cdot)$ is the indicator function. Estimation of $R(t, t')$ follows two steps: 1) fit the additive function-on-function model using a working independence assumption and obtain residuals, $e_{ij} = Y_{ij} - \hat{Y}_i(t_j)$ where $\hat{Y}_i(t) = \sum_{k=1}^K \mathbb{Z}_i^T \hat{\Theta}_k \tilde{\phi}_k(t)$; and 2) apply standard functional PCA based methods (see e.g., Yao et al., 2005a; Di et al., 2009) to the residual curves and estimate a finite rank approximation of $\hat{\Sigma}(t, t')$; this yields estimated eigencomponents and estimated error variance, $\hat{\sigma}^2$.

3.2 Inference

We now discuss the variability of the predicted response curves when new covariate profiles are observed. Let $X_0(\cdot)$ be the new functional covariate and assume

$Y_0(t) = \int_{\mathcal{T}_X} F\{X_0(s), s, t\}ds + \varepsilon_0(t)$ as in (1). Let $\hat{Y}_0(t)$ be the right-hand side of (7) with $X = X_0$. We measure the uncertainty in the prediction by the prediction error $\hat{Y}_0(t) - Y_0(t)$ (Ruppert et al., 2003), which is defined as

$$\text{var}\{\hat{Y}_0(t) - Y_0(t)\} = \text{var}\{\hat{Y}_0(t)\} + \text{var}\{\varepsilon_0(t)\}. \quad (8)$$

Assume that the error process $\varepsilon_0(t)$ has the same distribution as $\varepsilon_i(t)$ in (1) and is independent of $X_0(s)$. Then, the variance of $\varepsilon_0(t)$ can be estimated by $\hat{R}(t, t')$; here $\hat{R}(t, t')$ is obtained as in the previous section. We approximate $\{\hat{Y}_0(t)\}$ using the iterated variance formula:

$$\text{var}\{\hat{Y}_0(t)\} = \text{E}_{\tilde{\eta}}[\text{var}\{\hat{Y}_0(t)|\tilde{\eta}\}] + \text{var}_{\tilde{\eta}}[\text{E}\{\hat{Y}_0(t)|\tilde{\eta}\}], \quad (9)$$

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

We begin by deriving a model-based variance estimate of $\text{var}\{\hat{Y}_0(t)|\tilde{\eta}\}$. From (7),

$$\text{Var}\{\hat{Y}_0(t)|\tilde{\eta}\} = \sum_{k=1}^{K} \tilde{\phi}_k(t)\mathbb{Z}_0^T \text{var}(\hat{\Theta}_k)\mathbb{Z}_0\tilde{\phi}_k(t) + \sum_{k \neq k'} \tilde{\phi}_k(t)\mathbb{Z}_0^T \text{cov}(\hat{\Theta}_k, \hat{\Theta}_{k'})\mathbb{Z}_0\tilde{\phi}_{k'}(t),$$

where $\mathbb{Z}_0$ is the $K_x K_s$-column vector of $\int_0^1 B_{X,l}\{X_0(s)\}B_{S,l'}(s)ds$ for $l = 1, \ldots, K_x$, $l' = 1, \ldots, K_x$. Next, $\text{var}(\hat{\Theta}_k) = \nu_{kk} H_\lambda\{\sum_{i=1}^{n} \mathbb{Z}_i\mathbb{Z}_i^T\}H_\lambda^T$ and $\text{cov}(\hat{\Theta}_k, \hat{\Theta}_{k'}) = \nu_{kk'} H_\lambda\{\sum_{i=1}^{n} \mathbb{Z}_i\mathbb{Z}_i^T\}H_\lambda^T$. The conditional variance of $\hat{Y}_0(t)$ is

$$\text{var}\{\hat{Y}_0(t)|\tilde{\eta}\} = \sum_{k=1}^{K} \nu_{kk}\tilde{\phi}_k(t)\Omega_0\tilde{\phi}_k(t) + \sum_{k \neq k'} \nu_{kk'}\tilde{\phi}_k(t)\Omega_0\tilde{\phi}_{k'}(t), \quad (10)$$

where $\Omega_0 = \mathbb{Z}_0^T H_\lambda\{\sum_{i=1}^{n} \mathbb{Z}_i\mathbb{Z}_i^T\}H_\lambda^T\mathbb{Z}_0$ and where implicitly this variance is conditioned on $X_0(s)$. We estimate $\text{var}\{\hat{Y}_0(t)|\tilde{\eta}\}$ by plugging estimates of ν_{kk} and $\nu_{kk'}$ into (10). When the response curve is observed on a fine and regular grid of points, we estimate ν_{kk} by $\tilde{\nu}_{kk} = \int\int \tilde{\Sigma}_Y(t, t')\tilde{\phi}_k(t)\tilde{\phi}_k(t')dtdt'$, where $\tilde{\Sigma}_Y(\cdot, \cdot)$ is the estimated marginal covariance function of response, and $\tilde{\nu}_{kk'}$ is approximately 0 for $k \neq k'$, since $\{\tilde{\phi}(\cdot)\}k$ is the eigenbasis of $\tilde{\Sigma}_Y(\cdot, \cdot)$ and therefore orthogonal. When the response curve is observed at sparse and irregular grid of points, modification is needed to obtain $\tilde{\nu}_{kk}$ and $\tilde{\nu}_{kk'}$; see Section A of the Supplementary Material.

To account for the second source of variability, we use bootstrapping of subjects. We approximate the total variance of $\hat{Y}_0(t)$ using the iterated variance formula in (9); the first term, $\mathrm{E}_{\tilde{\eta}}[\mathrm{var}\{\hat{Y}_0(t)|\tilde{\eta}\}]$, can be estimated by averaging the model-based conditional variances across bootstrap samples. The second term, $\mathrm{var}_{\tilde{\eta}}[\mathrm{E}\{\hat{Y}_0(t)|\tilde{\eta}\}]$, is estimated by the sample variance of the predicted responses obtained from the bootstrap samples. Algorithm 1, displayed below, computes the total variance of $\hat{Y}_0(t)$. Using this result, we can construct a $100(1-\alpha)\%$ pointwise prediction interval for the new response $Y_0(t)$ as $\hat{Y}_0(t) \pm z_{\alpha/2}\widehat{\mathrm{SE}}\{\hat{Y}_0(t) - Y_0(t)\}$, where $z_{\alpha/2}$ is the $\alpha/2$ upper quantile of the standard normal distribution and $\mathrm{SE}\{\hat{Y}_0(t) - Y_0(t)\} = [\widehat{\mathrm{var}}\{\hat{Y}_0(t) - Y_0(t)\}]^{1/2}$ is obtained by bootstrapping the subjects using Algorithm 1.

Our inferential procedure has two advantages. First, the procedure accommodates complex correlation structure within the subject. Second, the iterated expectation and variance formula combines the model-based prediction variance and the variance of $\tilde{\eta}$, and better captures the total variance of the predicted response curves; our numerical study confirms the standard error characteristics in finite samples. One possible alternative for estimating the error covariance function $R(t, t')$ is to use $B^{-1}\sum_{b=1}^{B}\hat{R}^b(t,t')$ where $\hat{R}^b(t,t')$ is estimated using the b^{th} bootstrap sample, and our numerical study is based on this approach. Our numerical experience is that using the latter estimate of the covariance yields similar results as using the estimated model covariance $\hat{R}(t,t')$ derived in Section 3.1.

Algorithm 1

Bootstrap of subjects

1: for $b = 1$ to B do

2: Resample the subjects with replacement. Let $\{b_1, \ldots, b_n\}$ be the subject index of the bootstrap resample.

3: Define the covariate and the response curves in the b^{th} bootstrap sample as $\{X_i^{(b)}(\cdot) = X_{b_i}(\cdot)\}_{i=1}^{n}$ and $\{Y_i^{(b)}(\cdot) = Y_{b_i}(\cdot)\}_{i=1}^{n}$, respectively. The bootstrap data for the i^{th} subject is obtained by collecting the trajectories $\{X_i^{(b)}(s_k), s_k\}_{k=1}^{m}$ and $\{Y_i^{(b)}(t_j), t_j\}_{j=1}^{m_Y}$.

4: Apply FPCA to $\{Y_i^{(b)}(\cdot)\}_{i=1}^{n}$ and obtain an estimate of the eigenbasis $\{\phi_k^{(b)}(\cdot)\}_{k=1}^{K^{(b)}}$, where $K^{(b)}$ is the finite truncation that explains a pre-specified percent of variance.

5: For $l = 1, \ldots, K_x$, $l' = 1, \ldots, K_s$, and $k = 1, \ldots, K^{(b)}$, obtain parameter estimates $\hat{\theta}^{(b)}_{l,l',k}$ by applying AFF-PC to $\{X_i^{(b)}(s_k), s_k\}_{k=1}^{m}$ and $\{Y_i^{(b)}(t_j), t_j\}_{j=1}^{m_Y}$.

6: For a new covariate $X_0(s)$, obtain the predicted response by
$\hat{Y}_0^{(b)}(t) = \sum_{k=1}^{K^{(b)}} \tilde{\phi}_k^{(b)}(t) \sum_{l=1}^{K_x} \sum_{l'=1}^{K_s} \hat{\theta}_{l,l',k}^{(b)} \int_{\mathcal{T}_X} B_{X,l}\{X_0(s)\} B_{S,l'}(s) ds.$

7: Compute $V^{(b)}(t) = \widehat{\text{var}}\{\hat{Y}_0^{(b)}(t) | \tilde{\eta}_b\}$ using the model-based formula in (10).

8: **end for**

9: Approximate the marginal variance of predicted response by
$\widehat{\text{var}}\{\hat{Y}_0(t)\} \approx B^{-1} \sum_{b=1}^{B} V^{(b)}(t) + B^{-1} \sum_{b=1}^{B} \{\hat{Y}_0^{(b)}(t) - \bar{Y}_0(t)\}^2,$

where $\bar{Y}_0(t)$ is the sample mean of $\hat{Y}_0^{(b)}(t)$.

As the Associate Editor pointed out, the proposed approach to construct prediction bands relies on the validity of the involved bootstrap approximations. We use resampling of the subjects (see also Benko et al. (2009); Park et al. (2017)) to approximate both the unconditional model-based variance component and the variance of the predicted trajectories. But a rigorous study of the bootstrap techniques is somewhat limited in the functional data analysis. In particular, there is no consistency result about the bootstrap procedure involved here. While our numerical investigation, based on the coverage of the prediction bands (see Table 3), confirms that the methodology has desired property in the settings considered here, there is no guarantee that it is generally valid and the approach is for illustration.

4 Simulation Study

We investigate the finite sample performance of our method through simulations. We generate $N = 1000$ samples from model (1) with the true functional covariate given by $X(s) = a_1 + a_2\sqrt{2}\sin(\pi s) + a_3\sqrt{2}\cos(\pi s)$ where a_1, a_2, and a_3 vary independently across subjects, specifically, $a_p \sim \text{Normal}(0, 2^{(1-p)2})$ for $p = 1, 2, 3$. Also, the covariate is observed with noise, $W_{ik} = X_i(s_{ik}) + \delta_{ik}$ where the δ_{ik} are independent Normal(0, 0.5). For each sample we generate training sets of size $n = 50$, 100, and 300 and a test set of size 50; also $\mathcal{T}_X = \mathcal{T}_Y = [0, 1]$. The training sets include two different scenarios for the sampling of s and t. (i) *Dense design* - the grids of points $\{s_{ik}: k = 1, \ldots, m_i\}$ and $\{t_{ik} : k = 1, \ldots, m_{Y,i}\}$ are the same across i, $m_i = m$ and $m_{Y,i} = m_Y$, and are defined as the set of 81 and 101 equidistant points in [0, 1] respectively. (ii) *Sparse design* - for each i the number of observation points $m_i \sim \text{Uniform}(45, 54)$ and $m_{Y,i} \sim \text{Uniform}(35, 44)$; the time-points $\{s_{ik} : k = 1, \ldots, m_i\}$ and $\{t_{ik} : k = 1, \ldots, m_{Y,i}\}$ are randomly sampled without replacement from a set of 81 and 101 equidistant points in [0, 1] respectively.

The test set is generated using the set of 81 and 101 equispaced points in [0, 1] for s and t, respectively. We denote realizations of the error process by $\mathbb{E}_i = [\epsilon_i(t_{i1}), \ldots, \epsilon_i(t_{im_{Y,i}})]^T$ and generate them using four different covariance structures; these cases are denoted by $\mathbb{E}_i^1$, $\mathbb{E}_i^2$,

$\mathbb{E}_i^3$, $\mathbb{E}_i^4$, where $\mathbb{E}_i^1$ assumes a simple independent error structure, and other cases have correlated structure with increasing complexity and are described in Section C of the Supplementary Material. We consider three forms of true function F: linear function $F_1(x, s, t)$, simple nonlinear function $F_2(x, s, t)$, and complex nonlinear function $F_3(x, s, t)$; they too are defined in the Supplementary Material, Section C. Figure 1 shows the true surface of $F_3(x, s, t)$ along with x and s at fixed points $t = 0.05$, 0.5, and 1. The thick solid line is F_3 evaluated at fixed values for t and s so that only x varies; the nonlinearity of this curve indicates a departure from a functional linear model where $F(x, s, t)$ would be linear in x. The remaining details about the simulation are in the Supplementary Material, Section C. The R packages `Matrix` (Bates and Maechler, 2017) and `MASS` (Venables and Ripley, 2002) were used to generate data.

The performance of AFF-PC was assessed in terms of in-sample and out-of-sample predictive accuracy, as measured by the root mean squared prediction error (RMSPE), average computation time, and coverage probabilities of prediction intervals. The in-sample and out-of-sample root mean squared prediction error (RMSPE) are denoted by RMSPE^{in} and $\text{RMSPE}^{\text{out}}$, respectively. We define the in-sample RMSPE by

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

where $Y_i^{(r)}(t_{ij})$ and its estimate $\hat{Y}_i^{(r)}(t_{ij})$ are from the r^{th} Monte Carlo simulation. The out-of-sample RMSPE, denoted by $\text{RMSPE}^{\text{out}}$, is defined similarly. For each prediction we calculate the average coverage probability of the pointwise prediction intervals.

4.1 Competitive Methods

We compare our method to three other approaches: the functional linear model, the functional additive model of Müller and Yao (2008), which we label `FAM`, and the B-spline based estimation of Scheipl et al. (2015), AFF-S. Our approach is implemented using the R packages `refund` (Goldsmith et al., 2016) and `mgcv` (Wood, 2011, 2004, 2003). Details about the selection of the tuning parameters for our approach and the competitive approaches are in Section C of the Supplementary Material. We assess the prediction accuracy of the proposed approach and three competitive alternatives and compare their computational efficiency. Due to the high computational cost of the functional additive model and AFF-S, we restrict our comparisons with these methods to the case where $n = 50$ and the error process $(\mathbb{E}_i)$ is either $\mathbb{E}_i^2$ or $\mathbb{E}_i^4$ as described in Section C of the Supplementary Material. For the functional linear model, we consider a model defined by $E[Y_i(t)|X_i] = \int_{\mathcal{T}_X} X_i(s)\beta(s,t)ds$. Implementation details of our method and the three other approaches are summarized in the Supplementary Material, Section C.3.

4.2 Simulation Results

4.2.1 Prediction Performance—The comparison with the functional linear model is summarized in Table 1. For in-sample prediction accuracy, we report the relative percent gain in prediction with respect to functional linear model by computing $100 \times (1 - \text{RMSPE}^{\text{in}}_{\text{AFF}-\text{PC}}/\text{RMSPE}^{\text{in}}_{\text{FLM}})$, where $\text{RMSPE}^{\text{in}}_{\text{AFF}-\text{PC}}$ and $\text{RMSPE}^{\text{in}}_{\text{FLM}}$ are the in-sample prediction errors obtained by fitting the AFF-PC and functional linear model, respectively. Relative improvement for out-of-sample prediction is measured similarly. Thus, values closer to 0 indicate similar prediction performance between the two models, while larger positive values are indicative of AFF-PC having greater prediction accuracy than the functional linear model. The top part of Table 1 contains the case when the underlying true model is linear in x; the true relationship is described by F_1. Both AFF-PC and the functional linear model, provide relatively similar in-sample and out-of-sample prediction performance in all scenarios. The number of subjects, the sampling design of the grid points, and the error structure slightly affect the numerical results. The results confirm that when the true relationship is linear, then AFF-PC has similar prediction performance to the functional linear model, although there are few cases, especially for sparse designs and smaller sample sizes, where AFF-PC is slightly worse with respect to out-of-sample prediction.

The prediction results for the case where the true model in nonlinear are shown in the middle and bottom parts of the table: the true relationship is described by F_2 (simple nonlinear, middle) and F_3 (complex nonlinear, bottom). The results confirm that if the true model is nonlinear, then AFF-PC shows a dramatic improvement in prediction accuracy over the functional linear model. Depending on the complexity of the mean model, AFF-PC improves prediction accuracy compared to the functional linear model by over 50%. This improvement increases as the sample size gets larger.

Next, we compare AFF-PC to the AFF-S estimator (Scheipl et al., 2015), which uses B-splines rather than an eigenbasis to represent the trajectories. The results are presented in Table 2. Comparing the columns labeled (1) and (2) in the two panels, we observe that the two estimators have similar accuracy, with accuracy varying slightly with the complexity of the relationship. Column labeled (3) shows the average computation time (in seconds), indicating an order of magnitude improvement by AFF-PC over AFF-S. The models were run on a 2.3GHz AMD Opteron Processor.

Table 2 also compares AFF-PC and FAM. As the model complexity increases, the out-of-sample prediction accuracy of AFF-PC increases compared to FAM. Also, FAM takes much more computation time than AFF-PC, especially when the grid points are sparsely sampled. Computation time is less affected by the error covariance structure than is prediction accuracy.

In summary, AFF-PC better captures complex nonlinear relationships than the functional linear model, and yet AFF-PC performs as well as the functional linear model when the latter is true. The B-spline based estimator, AFF-S, and AFF-PC have similar prediction performance, while AFF-S and FAM are much slower than AFF-PC.

4.2.2 Performance of the Prediction Intervals—Next, we assess coverage accuracy of the pointwise prediction intervals. These intervals are approximated using the method described in Section 3 with 100 bootstrap samples per simulated data set. Table 3 reports the average coverage probability for both the dense and sparse design at nominal levels of 85%, 90%, and 95%. When the sample size is small (e.g., $n = 50$), the prediction intervals are conservative, providing greater coverage probabilities than the nominal values. However, the coverage probabilities approach the nominal levels as the sample size increases. The complexity of the true function $F(x, s, t)$ affects the coverage performance slightly. If the true function is complex, e.g., $F(x, s, t) = F_3(x, s, t)$, the coverage probability converges more slowly to the nominal levels as n increases compared to when the true function is simple, e.g., $F(x, s, t) = F_2(x, s, t)$. The number of subjects, the sampling design of the grid points, and the error covariance structure also affect the coverage performance slightly.

Remark: Section D.1 of the Supplementary Material includes additional simulation results corresponding to another level of sparseness, and the results indicate that our approach still maintains prediction accuracy. Section D.2 of the Supplementary Material illustrates numerically that our method is not sensitive to the choice of K.

5 Capital Bike Share Data

We now turn to the capital bike share study (Fanaee-T and Gama, 2013). The data were collected from the Capital Bike Share system in Washington, D.C., which offers bike rental services on an hourly basis. In recent years, there has been an increased demand for bicycle rentals; renting is viewed as an attractive alternative to owing bicycles. Ensuring a sufficient bike supply represents an important factor for a successful business in this area. In this paper we try to gain a better understanding of the customers' rental behavior during a weekend day in relation to the weather condition for that day. We are interested in casual rentals, which are rentals to cyclists without membership in the Capital Bike Share program. The counts of casual bike rentals are recorded at every hour of the day, during the period from January 1, 2011 to December 31, 2012, for a total of 105 weeks. Also collected are weather information such as temperature (°C) and humidity on an hourly basis.

Bike rentals have different dynamics on weekends compared to weekends. We restrict our study to Saturday rentals, when there is a particularly high demand for casual bike rentals. Our focus is on how Saturday rentals relate to the temperature, while accounting for humidity. Understanding the nature of this association could help one predict the casual rental demand based on the weather forecast available on the previous day. Figure 2 shows the counts of casual bike rentals (left panel) and hourly temperature (right panel) on Saturdays; each curve corresponds to a particular week. The solid, dotted, and dashed lines are the observations for three different Saturdays. On weekends, many renters can be flexible about when during the day to rent, so it is assumed that the entire temperature curve affects the number of casual bikes rental at any time on Saturday. To remove skewness, we log-transform the response data, $x \rightarrow \log(x + 1)$, before we proceed with our analysis.

Let $CB_i(t)$ be the number of casual bikes rented recorded, on the log-scale, for the i^{th} Saturday at the t^{th} hour of the day; also let $Temp_i(t)$ denote the true temperature for the i^{th}

Saturday at the t^{th} hour of the day and let $AHum_i$ be the average humidity for the corresponding Saturday. We consider the general additive function-on-function regression AFF-PC model

$$\mathrm{E}[CB_i(t)|Temp_i(\cdot), AHum_i] = \alpha(t) + \int_0^{24} F\{Temp_i(s), s, t\}ds + AHum_i\gamma(t), \quad (11)$$

where $\alpha(\cdot)$ is the marginal mean of the response, $F(\cdot, \cdot, \cdot)$ is an unknown trivariate function capturing the effect of the daily temperature and $\gamma(\cdot)$ is a smooth univariate function that quantifies the time-varying effect of the average humidity.

The temperature and the counts of bike rentals have a small amount of missingness. Therefore, we smoothed the temperature profiles using functional principal component analysis before applying the center/scaling transformation. We assessed both in-sample and out-of-sample prediction accuracy by splitting the data into training and test sets of size 89 and 16, respectively. To model the function F, we used $K_x = K_s = 7$ cubic B-splines for the x- and s-directions and selected K, the number of eigenfunctions $\{\phi_k(\cdot)\}_{k=1}^K$ for modeling F in the t direction, by fixing the percentage of explained variance to 95%; this resulted in $K = 3$. These choices for the tuning parameters are supported by additional sensitivity analysis included in Section E.2 of the Supplementary Material. We also used $\{\phi_k(\cdot)\}_{k=1}^K$ to model the marginal mean function $\alpha(\cdot)$ and the smooth effect of average humidity γ(t), $\alpha(t) = \sum_{k=1}^K \phi_k(t)\beta_k$ and $\gamma(t) = \sum_{k=1}^K \phi_k(t)\zeta_k$, where β_k and ζ_k are the unknown basis coefficients. Such a representation allows us to use K also to control the smoothness of the fitted coefficient function, $\hat{\gamma}(t)$. Parameter estimation was done as described in Section 2.2 with minor modifications due to the additional covariate, average humidity. Briefly, to estimate the unknown parameters, β_k, ζ_k and $\theta_{l,l',k}$, we constructed

$$\mathbb{Z}(i) = \left[1, AHum_i, \{\int_0^1 B_{X,l}\{Temp_i(s)\}B_{S,l'}(s)ds\}_{l,l'}\right], \Theta_k = [\beta_k, \zeta_k, \{\theta_{l,l}\}_{l,l'}], \widetilde{\mathbb{P}}_x = \mathrm{diag}(0, 0, \mathbb{P}_x)$$

and $\widetilde{\mathbb{P}}_s = \mathrm{diag}(0, 0, \mathbb{P}_s)$ and then minimized the penalized criterion (5) using $\widetilde{\mathbb{P}}x$ and $\widetilde{\mathbb{P}}s$ in place of $\mathbb{P}x$ and $\mathbb{P}s$, respectively.

Figure 3 shows the estimated parameter functions: the top two plots illustrate the estimated intercept function $\hat{\alpha}(\cdot)$ and $\hat{\gamma}(\cdot)$. On average the number of casual bike rentals decreases until 5AM ($t = 5$ on the plot) and then increases steadily peaking at about 3:00PM ($t = 15$). As expected, humidity is negatively associated with the bike rentals; the effect seems to be largest at 3:00PM. The bottom panels show the contour plots of the function $\hat{F}(x, s, t)$ for three values of t, $t = 0$ (midnight), $t = 12$ (noon) and $t = 20$ (evening, 8PM); the values of x have a standardized interpretation. For example, $x = 1$ is interpreted as one standard deviation away from the mean temperature profile. The plots were produced using the `R` packages `gridExtra` (Auguie, 2016) and `lattice` (Sarkar, 2008).

As mentioned in Section 2.1, the functional linear model is the special case of (1) where $F(x, s, t) = \beta(s, t)x$, which implies that $\partial F(x, s, t)/\partial x$ does not depend on x. The nonlinearity of $\hat{F}$ in x can be noted in all these plots but in particular in the middle bottom panel ($t = 12$). Consider the case when $s = 10$; simple calculations yield that the partial derivative of F with respect to x at $x = -1$ is different from the one for $x = 1$, and thus that $\hat{F}$ is not linear in x.

Table 4 compares AFF-PC, AFF-S, and the functional linear model in terms of prediction accuracy. AFF-PC results in better prediction performance than functional linear model for both in-sample and out-of sample. As expected, AFF-PC and AFF-S have similar accuracy but AFF-PC is much faster than AFF-S.

Furthermore, we can construct bootstrap-based prediction intervals for the predicted trajectories in the test set, by slightly modifying the bootstrap procedure included in Section 3.2. For completeness, the algorithm is provided in the Supplementary Material, Section E.3. Figure 4 illustrates the 95% prediction bands constructed for three different Saturdays in the test set. Finally, we assessed the coverage probability of the prediction intervals. AFF-PC tended to produce conservative prediction intervals. For example, using 1000 bootstrap replications, the actual coverage probability of the 95% prediction intervals was 0.988 with a standard error of 0.003.

6 Discussion

This article considered additive regression models for functional responses and functional covariates. These models are a generalization of the functional linear model and allow for a nonlinear relationship between the response and the covariate. We proposed a novel estimation technique, AFF-PC, that is computationally very fast. We developed prediction inference for a future functional outcome when the functional covariate is known. As illustrated by the bike share study, AFF-PC can accommodate additional scalar or vector covariates. Furthermore, AFF-PC can easily be extended to accommodate multiple functional covariates. We showed through numerical study that when the true model is linear, AFF-PC's performance is very close to that of the functional linear model, but if the true model is nonlinear, AFF-PC can yield considerably improved prediction performance. The capital bike share data is available at: https://archive.ics.uci.edu/ml/datasets/Bike+Sharing+Dataset. The `R` code used in the simulation is available at: http://www4.stat.ncsu.edu/~staicu/Code/affpccode.zip.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Staicu's research was funded by National Science Foundation grant DMS 1454942 and National Institute of Health grants R01 NS085211 and R01 MH086633. Maity's research was partially funded by National Institutes of Health award R00 ES017744 and a North Carolina State University Faculty Research and Professional Development award. Carroll's research was supported by a grant from the National Cancer Institute (U01-CA057030). The research of Ruppert was partially supported by National Science Foundation grant AST 1312903 and by a grant from the National Cancer Institute (U01-CA057030).

References

Aston JAD, Chiou JM, Evans JP. Linguistic pitch analysis using functional principal component mixed effect models. Journal of the Royal Statistical Society, Series C. 2010; 59:297–317.

Auguie, B. gridExtra: Miscellaneous Functions for ”Grid” Graphics. 2016. R package version 2.2.1

Bates, D., Maechler, M. Matrix: Sparse and Dense Matrix Classes and Methods. 2017. R package version 1.2.8

Benko M, Härdle W, Kneip A, et al. Common functional principal components. The Annals of Statistics. 2009; 37(1):1–34.

Di CZ, Crainiceanu CM, Caffo B, Punjabi NM. Multilevel functional principal component analysis. Annals of Applied Statistics. 2009; 3:458–488. [PubMed: 20221415]

Fanaee-T H, Gama J. Event labeling combining ensemble detectors and background knowledge. Progress in Artificial Intelligence. 2013:1–15.

Friedman JH, Stuetzle W. Projection pursuit regression. Journal of the American statistical Association. 1981; 76(376):817–823.

Goldsmith J, Greven S, Crainiceanu C. Corrected confidence bands for functional data using principal components. Biometrics. 2013; 69:41–51. [PubMed: 23003003]

Goldsmith, J., Scheipl, F., Huang, L., Wrobel, J., Gellar, J., Harezlak, J., McLean, MW., Swihart, B., Xiao, L., Crainiceanu, C., Reiss, PT. refund: Regression with Functional Data. 2016. R package version 0.1.16

Jiang CR, Wang J-L. Covariate adjusted functional principal components analysis for longitudinal data. Annals of Statistics. 2010; 38:1194–1226.

Kim J, Maity A, Staicu AM. General functional concurrent model. 2016 Unpublished manuscript (under review).

Malfait N, Ramsay JO. The historical functional linear model. The Canadian Journal of Statistics. 2003; 31:115–128.

Marx BD, Eilers PHC. Multivariate penalized signal regression. Technometrics. 2005; 47:13–22.

McLean MW, Hooker G, Staicu AM, Scheipl F, Ruppert D. Functional generalized additive models. Journal of Computational and Graphical Statistics. 2014; 23:249–269. [PubMed: 24729671]

Müller HG, Wu Y, Yao F. Continuously additive models for nonlinear functional regression. Biometrika. 2013; 103:607–622.

Müller HG, Yao F. Functional additive models. Journal of the American Statistical Association. 2008; 103:1534–1544.

Park SY, Staicu AM. Longitudinal functional data analysis. Stat. 2015; 4:212–226. [PubMed: 26594358]

Park SY, Staicu A-M, Xiao L, Crainiceanu CM. Inference on fixed effects in complex functional mixed models. 2017

Pomann GM, Staicu A-M, Ghosh S. Two sample hypothesis testing for functional data. 2015 Unpublished manuscript (submitted).

Ramsay, JO., Silverman, BW. Functional Data Analysis. 2nd. New York: Springer; 2005.

Redd A. A comment on the orthogonalization of b-spline basis functions and their derivatives. Statistics and Computing. 2012; 22:251–257.

Ruppert, D., Wand, MP., Carroll, RJ. Semiparametric Regression. Cambridge, New York: Cambridge University Press; 2003.

Sarkar, D. Lattice: Multivariate Data Visualization with R. New York: Springer; 2008.

Scheipl F, Staicu A-M, Greven S. Functional additive mixed models. Journal of Computational and Graphical Statistics. 2015; 24:477–501. [PubMed: 26347592]

Sentürk D, Nguyen DV. Varying coefficient models for sparse noise-contaminated longitudinal data. Statistica Sinica. 2011; 21:1831–1856. [PubMed: 25589822]

Venables, WN., Ripley, BD. Modern Applied Statistics with S. Fourth. New York: Springer; 2002.

Wood SN. Thin-plate regression splines. Journal of the Royal Statistical Society (B). 2003; 65(1):95–114.

Wood SN. Stable and efficient multiple smoothing parameter estimation for generalized additive models. Journal of the American Statistical Association. 2004; 99(467):673–686.

Wood, SN. Generalized Additive Models: An Introduction with R. Boca Raton, Florida: Chapman and Hall/CRC; 2006.

Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society (B). 2011; 73(1):3–36.

Wu Y, Fan J, Müller H-G. Varying-coefficient functional linear regression. Bernoulli. 2010; 16:730–758.

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

Yao F, Müller H-G, Wang J-L. Functional linear regression analysis for longitudinal data. Annals of Statistics. 2005b; 33:2873–2903.

Zhang JT, Chen J. Statistical inference for functional data. Annals of Statistics. 2007; 35:1052–1079.

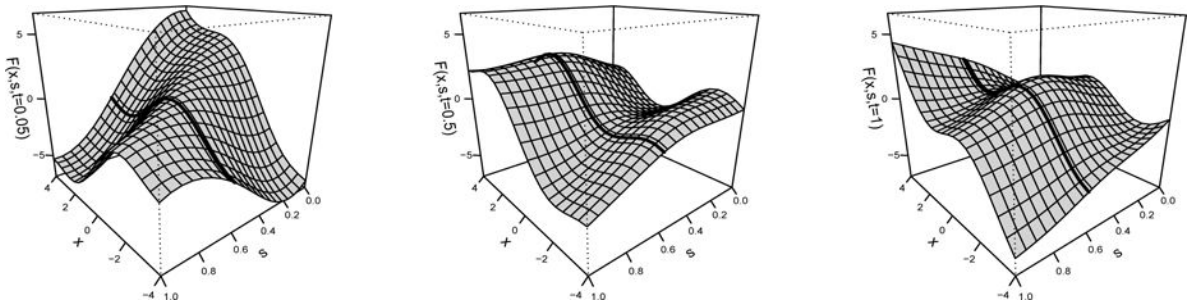

Figure 1.

The three panels show the complex nonlinear function $F_3(\cdot)$. Plotted are $F_3(x, s, 0.05)$ (left), $F_3(x, s, 0.5)$ (middle), and $F_3(x, s, 1)$ (right). The thick solid line represents the curve obtained by fixing s at 0.6. Notice its nonlinearity as a function of x.

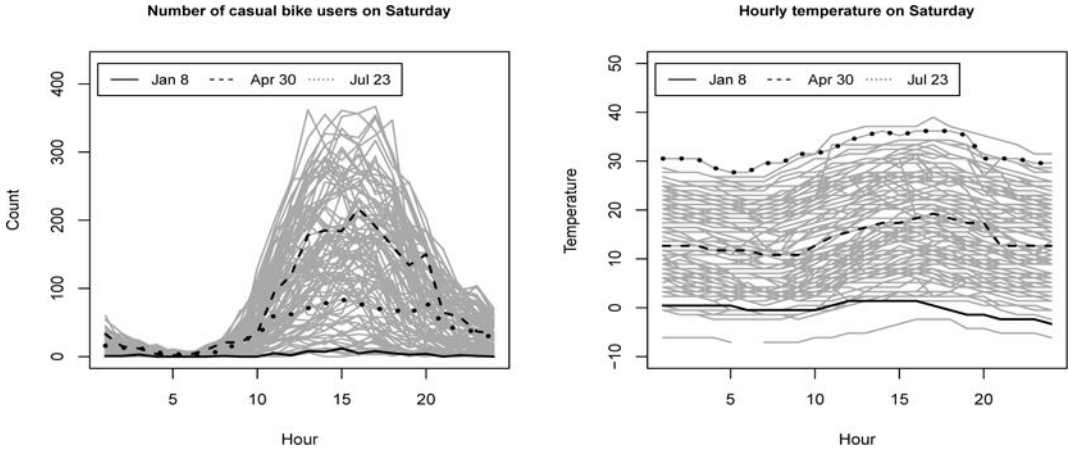

Figure 2.

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

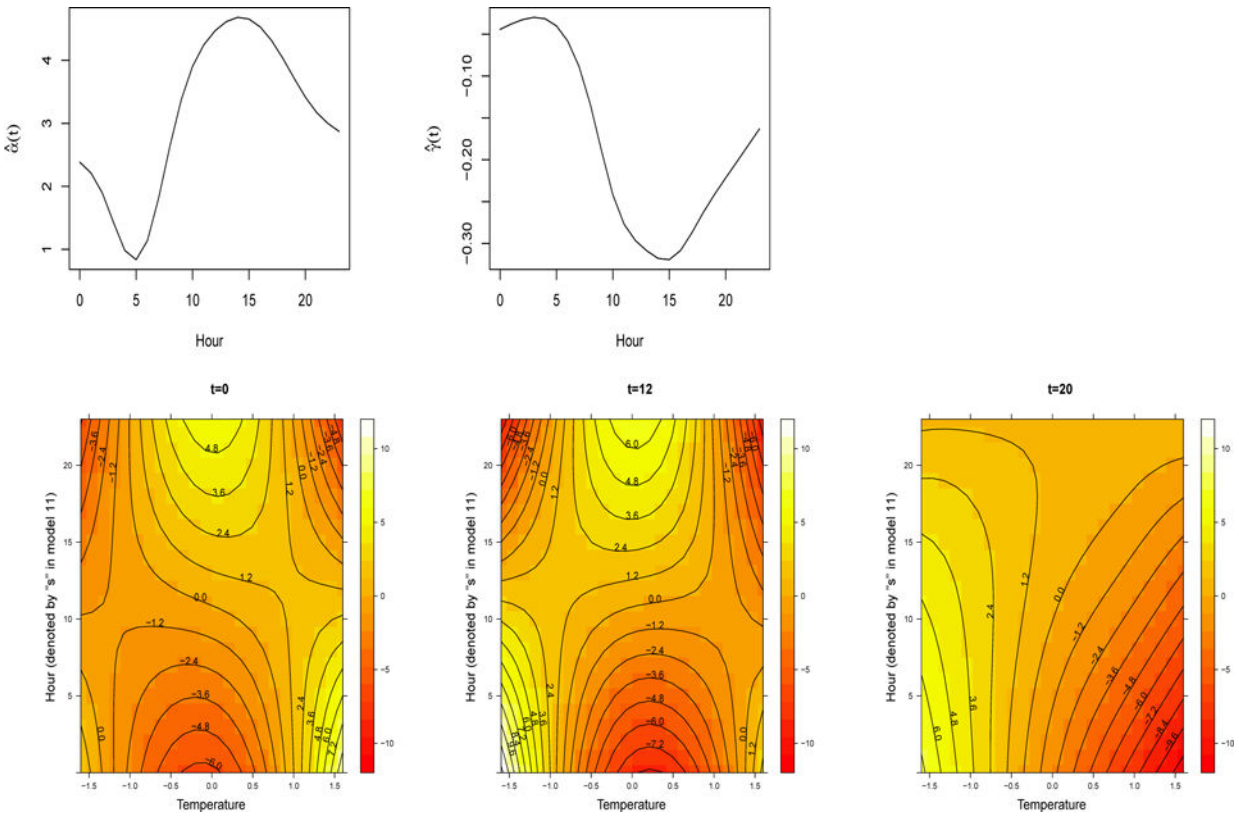

Figure 3.

Displayed are the estimated parameter functions obtained by regressing $\log(1+\text{count}_{ij})$ on the transformed temperature (°C) and average humidity. Top panels: marginal mean, $\hat{\alpha}(t)$ and the effect of average humidity, $\hat{\gamma}(t)$. Bottom panels: contour plots of the estimated surface, $\hat{F}(x, s, 0)$ (left), $\hat{F}(x, s, 12)$ (middle) and $\hat{F}(x, s, 20)$ (right).

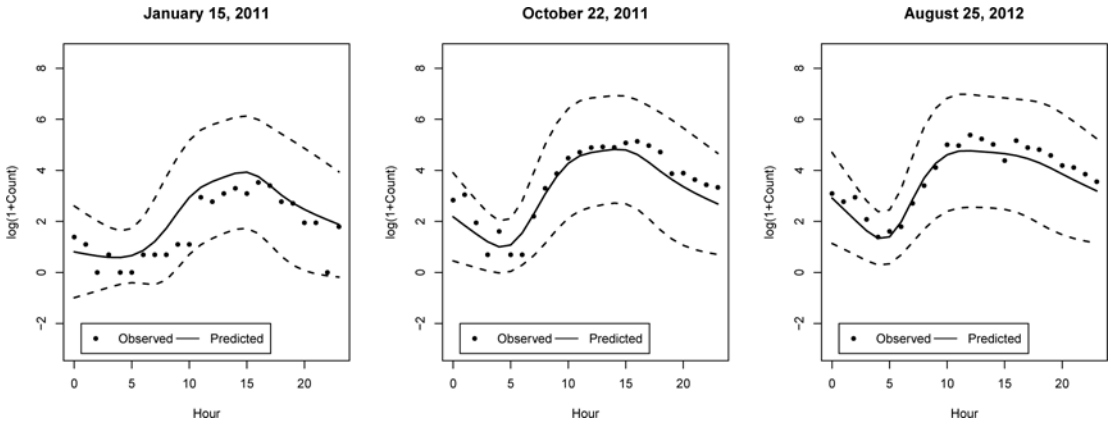

Figure 4.

95% prediction bands constructed for three subject-level trajectories in the bike data. "●" are the observed response trajectories, solid lines are predicted response. Dashed lines are the prediction bands obtained by applying the method of AFF-PC.

Table 1

Relative percent gain in prediction accuracy of the AFF-PC compared to the functional linear model. The percent improvements are measured for (1) in-sample and (2) out-of-sample. The functions are a linear function $F_1(x, s, t)$, a simple nonlinear function $F_2(x, s, t)$, and a complex nonlinear function $F_3(x, s, t)$ and $\mathbb{E}_i^1 - \mathbb{E}_i^4$ are four correlation structures, with $\mathbb{E}_i^1$ being independent.

	$\mathbb{E}_i = \mathbb{E}_i^1$		$\mathbb{E}_i = \mathbb{E}_i^2$		$\mathbb{E}_i = \mathbb{E}_i^3$		$\mathbb{E}_i = \mathbb{E}_i^4$		$\mathbb{E}_i = \mathbb{E}_i^1$		$\mathbb{E}_i = \mathbb{E}_i^2$		$\mathbb{E}_i = \mathbb{E}_i^3$		$\mathbb{E}_i = \mathbb{E}_i^4$	
	(1)	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**	**(1)**	**(2)**
n	$F(x, s, t) = F_1(x, s, t)$, dense design								$F(x, s, t) = F_1(x, s, t)$, sparse design							
50	0.00	−0.71	0.22	−1.41	0.27	−0.67	0.27	−2.04	−0.29	−5.81	0.00	−6.37	0.00	−4.37	0.00	−5.66
100	0.31	−0.88	0.00	−0.87	0.00	−0.84	0.00	−0.85	−0.30	−3.88	0.00	−4.58	−0.27	−3.79	−0.27	−3.79
300	0.00	0.00	0.00	−1.06	0.00	0.00	0.00	0.00	0.00	−1.98	−0.23	−2.94	0.00	−1.96	−0.27	−1.96
n	$F(x, s, t) = F_2(x, s, t)$, **dense design**								$F(x, s, t) = F_2(x, s, t)$, **sparse design**							
50	5.20	31.97	3.38	38.51	6.74	44.81	5.93	35.95	5.47	29.80	3.15	26.80	5.91	32.91	4.84	22.29
100	6.36	41.84	4.04	50.00	6.95	56.55	6.67	50.34	6.65	41.67	3.59	37.93	6.13	45.58	5.59	36.05
300	6.97	55.80	4.26	63.04	6.91	66.19	6.65	64.75	7.53	55.40	4.03	52.52	6.63	60.00	6.10	53.57
n	$F(x, s, t) = F_3(x, s, t)$, **dense design**								$F(x, s, t) = F_3(x, s, t)$, **sparse design**							
50	34.38	58.77	24.32	58.21	30.61	56.21	30.61	55.34	31.71	52.37	22.69	51.18	28.60	50.86	28.60	50.21
100	36.36	65.30	25.59	64.61	31.99	63.33	31.99	62.64	34.16	59.68	24.38	58.43	30.07	58.20	30.07	57.75
300	37.57	70.19	26.38	69.95	32.60	69.32	32.60	68.85	35.88	67.29	25.33	66.59	31.41	66.67	31.23	66.43

Table 2

Comparison of FAM and AFF-S by root means squared prediction errors (1) RMSPE^{in} and (2) RMSPE^{out}, and (3) computation time (in seconds) averaged over 1000 simulations. Results correspond to $n = 50$.

		dense design						sparse design					
		$\mathbb{E}_i = \mathbb{E}_i^2$			$\mathbb{E}_i = \mathbb{E}_i^4$			$\mathbb{E}_i = \mathbb{E}_i^2$			$\mathbb{E}_i = \mathbb{E}_i^4$		
***F*(*x*, *s*, *t*)**	**method**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**	**(1)**	**(2)**	**(3)**
	FAM	0.45	0.17	94.0	0.37	0.18	92.9	0.45	0.21	687.9	0.38	0.21	920.5
$F_1(x, s, t)$	AFF-S	0.44	0.15	99.4	0.36	0.21	82.3	0.45	0.16	43.0	0.37	0.19	38.7
	AFF-PC	0.44	0.14	10.2	0.37	0.15	8.7	0.45	0.17	10.2	0.38	0.17	8.9
	FAM	0.43	0.12	93.9	0.36	0.13	93.4	0.44	0.17	564.2	0.36	0.17	697.7
$F_2(x, s, t)$	AFF-S	0.43	0.06	143.5	0.34	0.12	116.4	0.42	0.08	39.0	0.34	0.10	37.5
	AFF-PC	0.43	0.09	6.1	0.35	0.10	7.8	0.43	0.11	7.1	0.35	0.12	9.9
	FAM	0.48	0.28	94.4	0.41	0.29	92.7	0.49	0.32	687.2	0.42	0.33	656.3
$F_3(x, s, t)$	AFF-S	0.45	0.21	130.3	0.37	0.27	124.1	0.45	0.23	50.6	0.38	0.26	31.0
	AFF-PC	0.45	0.19	10.5	0.37	0.21	10.1	0.46	0.23	9.7	0.39	0.23	9.9

Table 3

Summary of average coverage probabilities for predicting a new response $Y_0(t)|X_0(\cdot)$ at nominal significance levels $1 - \alpha = 0.85$, 0.90, and 0.95. Results are based on 1000 simulated data sets with 100 bootstrap replications per data.

	$\mathbb{E}_i = \mathbb{E}_i^1$			$\mathbb{E}_i = \mathbb{E}_i^2$			$\mathbb{E}_i = \mathbb{E}_i^3$			$\mathbb{E}_i = \mathbb{E}_i^4$		
n	**0.85**	**0.90**	**0.95**	**0.85**	**0.90**	**0.95**	**0.85**	**0.90**	**0.95**	**0.85**	**0.90**	**0.95**
***F*(*x*, *s*, *t*) = *F*$_2$(*x*, *s*, *t*), dense design**												
50	0.904	0.942	0.976	0.883	0.926	0.966	0.880	0.926	0.967	0.883	0.925	0.965
100	0.884	0.928	0.967	0.869	0.916	0.960	0.866	0.916	0.961	0.869	0.915	0.959
300	0.868	0.915	0.960	0.859	0.908	0.955	0.856	0.908	0.957	0.858	0.908	0.953
F(*x*, *s*, *t*) = *F*$_2$(*x*, *s*, *t*), **sparse design**												
50	0.910	0.946	0.977	0.882	0.926	0.965	0.888	0.930	0.969	0.887	0.928	0.967
100	0.887	0.930	0.969	0.870	0.916	0.960	0.870	0.918	0.963	0.873	0.918	0.961
300	0.867	0.915	0.960	0.860	0.909	0.956	0.858	0.910	0.958	0.862	0.910	0.955
F(*x*, *s*, *t*) = *F*$_3$(*x*, *s*, *t*), **dense design**												
50	0.936	0.963	0.986	0.914	0.948	0.978	0.912	0.947	0.978	0.911	0.946	0.977
100	0.913	0.949	0.979	0.895	0.935	0.970	0.895	0.935	0.972	0.893	0.933	0.970
300	0.880	0.924	0.966	0.871	0.917	0.961	0.870	0.916	0.962	0.869	0.914	0.959
F(*x*, *s*, *t*) = *F*$_3$(*x*, *s*, *t*), **sparse design**												
50	0.949	0.971	0.989	0.913	0.947	0.977	0.931	0.958	0.982	0.932	0.959	0.983
100	0.923	0.954	0.982	0.895	0.936	0.971	0.903	0.941	0.975	0.903	0.940	0.974
300	0.889	0.931	0.970	0.877	0.922	0.964	0.879	0.923	0.966	0.878	0.921	0.963

Table 4

Results from the Capital Bike Share study described in Section 5. Displayed are the summaries of (1) RMSPE^{in}, (2) RMSPE^{out} and (3) computation time (in seconds) obtained by regressing $\log(1+\text{count}_{ij})$ on temperature and average humidity.

		log-transformed data		original data		
Method	(K_x, K_s, K_t)	**(1)**	**(2)**	**(1)**	**(2)**	**(3)**
FLM	(NA, 7,7)	0.740	0.606	62.079	43.603	2.12
AFF-S	(7,7,7)	0.637	0.494	37.275	28.826	25.36
AFF-PC	(7,7, K b = 3)	0.635	0.493	38.184	31.715	1.97