




On function-on-function regression: partial least squares approach

Ufuk Beyaztas¹ · Han Lin Shang² 

Received: 1 June 2019 / Revised: 8 December 2019 / Published online: 7 January 2020
© Springer Science+Business Media, LLC, part of Springer Nature 2020

Abstract

Functional data analysis tools, such as function-on-function regression models, have received considerable attention in various scientific fields because of their observed high-dimensional and complex data structures. Several statistical procedures, including least squares, maximum likelihood, and maximum penalized likelihood, have been proposed to estimate such function-on-function regression models. However, these estimation techniques produce unstable estimates in the case of degenerate functional data or are computationally intensive. To overcome these issues, we proposed a partial least squares approach to estimate the model parameters in the function-on-function regression model. In the proposed method, the B -spline basis functions are utilized to convert discretely observed data into their functional forms. Generalized cross-validation is used to control the degrees of roughness. The finite-sample performance of the proposed method was evaluated using several Monte-Carlo simulations and an empirical data analysis. The results reveal that the proposed method competes favorably with existing estimation techniques and some other available function-on-function regression models, with significantly shorter computational time.

Keywords Basis function · Functional data · NIPALS · Nonparametric smoothing · SIMPLS

Handling Editor: Bryan F. J. Manly

✉ Han Lin Shang
hanlin.shang@anu.edu.au

¹ Department of Statistics, Bartın University, Bartın, Turkey

² Research School of Finance, Actuarial Studies, and Statistics, Australian National University, Level 4, Building 26C, Kingsley Street, Acton, ACT 2601, Australia

1 Introduction

Recent advances in computer storage and data collection have enabled researchers in diverse branches of science such as, for instance, chemometrics, meteorology, medicine, and finance, recording data of characteristics varying over a continuum (time, space, depth, wavelength, etc.). Given the complex nature of such data collection tools, the availability of functional data, in which observations are sampled over a fine grid, has progressively increased. Consequently, the interest in functional data analysis (FDA) tools is significantly increasing over the years. Ramsay and Silverman (2002, 2006), Ferraty and Vieu (2006), Horvath and Kokoszka (2012) and Cuevas (2014) provide excellent overviews of the research on theoretical developments and case studies of FDA tools.

Functional regression models in which both the response and predictors consist of curves known as, function-on-function regression, have received considerable attention in the literature. The main goal of these regression models is to explore the associations between the functional response and the functional predictors observed on the same or potentially different domains as the response function. In this context, two key models have been considered: the *varying-coefficient* model and the *function-on-function regression model* (FFRM). The varying-coefficient model assumes that the functional response $\mathcal{Y}(t)$ and functional predictors $\mathcal{X}(t)$ are observed in the same domain. Its estimation and test procedures have been studied by numerous authors, including Fan and Zhang (1999), Hoover et al. (1998), Brumback and Rice (1998), Wu and Chiang (2000), Huang et al. (2002, 2004), Şentürk and Müller (2005), Cardot and Sarda (2008), Wu et al. (2010) and Zhu et al. (2014) among many others. In contrast, the FFRM considers cases in which the functional response $\mathcal{Y}(t)$ for a given continuum t depends on the full trajectory of the predictors $\mathcal{X}(s)$. Compared with the varying-coefficient model, the FFRM is more natural; therefore, we restrict our attention to the FFRM for this study.

The FFRM was first proposed by Ramsay and Dalzell (1991), who extended the traditional multivariate regression model to the infinite-dimensional case. In the FFRM, the association between the functional response and the functional predictors is expressed by integrating the full functional predictor weighted by an unknown bivariate coefficient function. More precisely, if $\mathcal{Y}_i(t)$ ($i = 1, \dots, N$) and $\mathcal{X}_{im}(s)$ ($m = 1, \dots, M$), respectively, denote a set of functional responses and M sets of functional predictors with $s \in [0, S]$ and $t \in [0, \mathcal{T}]$, where S and \mathcal{T} are closed and bounded intervals on the real line, then the FFRM for $\mathcal{Y}_i(t)$ and $\mathcal{X}_{im}(s)$ is constructed as follows:

$$\mathcal{Y}_i(t) = \beta_0(t) + \sum_{m=1}^M \int_S \mathcal{X}_{im}(s) \beta_m(s, t) ds + \epsilon_i(t), \quad (1.1)$$

where $\beta_0(t)$ is the mean response function, $\beta_m(s, t)$ is the bivariate coefficient function, and $\epsilon_i(t)$ denotes an independent random error function having a normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix Σ_ϵ , i.e., $\epsilon_i(t) \sim N(\mathbf{0}, \Sigma_\epsilon)$. The main purpose of model (1.1) is to estimate the bivariate coefficient function $\beta_m(s, t)$.

In this context, Yamanishi and Tanaka (2003) proposed a geographically weighted regression model to explore the functional relationship between the variables; Ramsay and Silverman (2006) proposed a least squares (LS) method to estimate $\beta_m(s, t)$ by minimizing the integrated sum of squares; Yao et al. (2005) extended the FFRM to the analysis of sparse longitudinal data and discussed the estimation procedures; Müller and Yao (2008) proposed a functional additive regression model where regression parameters are estimated using regularization; Matsui et al. (2009) suggested a maximum penalized likelihood (MPL) approach to estimate the coefficient function in the FFRM; Wang (2014) proposed a linear mixed regression model and estimated the model parameters via the expectation/conditional maximization either algorithm; Ivanescu et al. (2015) developed several penalized spline approaches to estimate the FFRM parameters using the mixed model representation of penalized regression; and Chiou et al. (2016) proposed a multivariate functional regression model to analyze multivariate functional data.

Most investigations on parameter estimation in FFRM have generally focused on the LS, maximum likelihood (ML), and MPL approaches. While these approaches work well in certain circumstances, they are characterized by several drawbacks. For instance, the ML and LS methods produce unstable estimates when functional data have degenerate structures (see Matsui et al. 2009). They also encounter a singular matrix problem when a large number of functional predictors are included in the FFRM. Alternatively, the singular matrix problem can also occur when a large number of basis functions are used to approximate those functions. In such cases, the LS and ML methods typically fail to provide an estimate for $\beta_m(s, t)$. Although the MPL method can overcome such difficulties and produce consistent estimates, it is computationally time-consuming for a computer with standard memory. It may not be possible to obtain the MPL estimates where a large number of basis functions are used to approximate the functional data. In this paper, we propose a partial least squares (PLS) approach to estimate the parameter of the FFRM to overcome these vexing issues.

The functional counterparts of the PLS method, when the functional data consist of scalar response and functional predictors, were proposed by Preda and Saporta (2005), Reiss and Ogden (2007), Krämer et al. (2008), and Aguilera et al. (2010). Febrero-Bande et al. (2017) compared these methods and discussed their advantages and disadvantages. Hyndman and Shang (2009) proposed a weighted functional partial least squares regression method for forecasting functional time series. Their method is based on a lagged functional predictor and a functional response. In this paper, we proposed an extended version of the functional partial least squares regression (FPLSR) of Preda and Schiltz (2011). The proposed method differs from the previous FPLSR in two respects. First, while the FPLSR considers only one functional predictor in the model, our approach allows for more than one functional predictor. Second, the FPLSR uses a fixed smoothing parameter when converting the discretely observed data to functional form. However, our approach uses a grid search to determine the optimal smoothing parameter.

In summary, our proposed method works as follows. First, the B -spline basis function expansion is used to express discretely observed data as smooth functions. The number of basis functions is determined using the penalized LS, and the smoothing

parameter that controls the roughness of the expansion is specified by the generalized cross-validation (GCV). The discretized version of the smooth coefficient function obtained by the basis function expansion is solved for a matrix [(say \mathbf{B} in (2.6)] using a PLS algorithm. In this study, we used the two fundamental PLS algorithms found in the literature to estimate \mathbf{B} —nonlinear iterative partial least squares (NIPALS) (Wold 1974) and simple partial least squares (SIMPLS) (de Jong 1993). Finally, the estimate of the coefficient function $\beta_m(s, t)$ was obtained by applying the smoothing step. The main advantage of the proposed method is that it bypasses the singular matrix problem. Further, the proposed method increases the predicting accuracy of the FFRM and is more efficient compared with some other available estimation methods.

The remainder of this paper proceeds as follows. Section 2 is dedicated to the methodology of the proposed method. Section 3 evaluates the finite-sample performance of the proposed method using several Monte-Carlo experiments. Section 4 applies the proposed method to a dataset on solar radiation prediction. Section 5 concludes the paper and provides several future research directions.

2 Methodology

For the FFRM provided by (1.1), the functional random variables are assumed to be an element of \mathcal{L}_2 , which expresses square-integrable and real-valued functions. They are further assumed to be second-order stochastic processes with finite second-order moments. The association between these functional variables is characterized by the surface $\beta_m(s, t) \in \mathcal{L}_2$, where \mathcal{L}_2 denotes a square-integrable functional space. Without loss of generality, the mean response function $\beta_0(t)$ is eliminated from the model (1.1) by centering the functional response and functional predictor variables.

If $\mathcal{Y}_i^*(t) = \mathcal{Y}_i(t) - \bar{\mathcal{Y}}(t)$, $\mathcal{X}_{im}^*(s) = \mathcal{X}_{im}(s) - \bar{\mathcal{X}}_m(s)$ and $\epsilon_i^*(t) = \epsilon_i(t) - \bar{\epsilon}(t)$ are used to denote the centered versions of the functional variables and the error function defined in (1.1), the model (1.1) can be re-expressed as follows:

$$\mathcal{Y}_i^*(t) = \sum_{m=1}^M \int_S \mathcal{X}_{im}^*(s) \beta_m(s, t) ds + \epsilon_i^*(t). \quad (2.1)$$

By custom, we expressed the functional variables and the bivariate coefficient function as basis function expansions before fitting the FFRM.

Initially, let $x(t)$ denote a function finely sampled on a grid $t \in [0, T]$. Based on a pre-determined basis and a sufficiently large number of basis functions K , it can be approximated as $x(t) \approx \sum_{k=1}^K c_k \phi_k(t)$, where $\phi_k(t)$ and c_k , for $k = 1, \dots, K$, represent the k th basis function and its associated coefficient vector, respectively. In this study, the functions were approximated using B -spline basis and the number of basis functions were determined according to GCV. Similarly, the (centered) functional variables and the bivariate coefficient function in (2.1) can be written as basis function expansions as follows:

$$\mathcal{Y}_i^*(t) = \sum_{k=1}^{K_{\mathcal{Y}}} c_{ik} \phi_k(t) = \mathbf{c}_i \Phi(t) \quad \forall t \in \mathcal{T}, \tag{2.2}$$

$$\mathcal{X}_{im}^*(s) = \sum_{j=1}^{K_{m,\mathcal{X}}} d_{imj} \psi_{mj}(s) = \mathbf{d}_{im} \Psi(s) \quad \forall s \in S, \tag{2.3}$$

$$\beta_m(s, t) = \sum_{j,k} \psi_{mj}(s) b_{mjk} \phi_k(t) = \Psi_m(s) \mathbf{B}_m \Phi(t) \quad \forall t \in \mathcal{T}, \forall s \in S, \tag{2.4}$$

where $\Phi(t)$ and $\Psi(s)$ are the vectors of the basis functions with dimensions $K_{\mathcal{Y}}$ and $K_{m,\mathcal{X}}$, respectively, \mathbf{c}_i and \mathbf{d}_{im} , respectively, are the $K_{\mathcal{Y}}$ and $K_{m,\mathcal{X}}$ dimensional coefficient vectors, and \mathbf{B}_m is a $K_{m,\mathcal{X}} \times K_{\mathcal{Y}}$ dimensional coefficient matrix. Replacing (2.2) to (2.4) with (2.1) yields:

$$\begin{aligned} \mathbf{c}_i \Phi(t) &= \sum_{m=1}^M \mathbf{d}_{im} \zeta_{\psi_m} \mathbf{B}_m \Phi(t) + \epsilon_i^*(t), \\ &= z_i \mathbf{B} \Phi(t) + \epsilon_i^*(t), \end{aligned} \tag{2.5}$$

where $\zeta_{\psi_m} = \int_S \psi_m(s) \psi_m^\top(s) ds$ is a $K_{m,\mathcal{X}} \times K_{m,\mathcal{X}}$ cross-product matrix, $z_i = (\mathbf{d}_{i1}^\top \zeta_{\psi_1}, \dots, \mathbf{d}_{iM}^\top \zeta_{\psi_M})^\top$ is a vector of dimension $\sum_{m=1}^M K_{m,\mathcal{X}}$, and $\mathbf{B} = (\mathbf{B}_1, \dots, \mathbf{B}_M)^\top$ is the coefficient matrix with dimensions $\sum_{m=1}^M K_{m,\mathcal{X}} \times K_{\mathcal{Y}}$. Let $\mathbf{C} = (\mathbf{c}_1, \dots, \mathbf{c}_N)^\top$, $\mathbf{Z} = (\mathbf{z}_1, \dots, \mathbf{z}_N)^\top$ and $\boldsymbol{\epsilon}(t) = (\epsilon_1^*(t), \dots, \epsilon_N^*(t))^\top$, the model (2.5) can then be rewritten as follows:

$$\mathbf{C} \Phi(t) = \mathbf{Z} \mathbf{B} \Phi(t) + \boldsymbol{\epsilon}(t). \tag{2.6}$$

Assuming that the error function $\boldsymbol{\epsilon}(t)$ in (2.6) can also be represented as a basis function expansion, then $\boldsymbol{\epsilon}(t) = \mathbf{e} \Phi(t)$ with $\mathbf{e} = (\mathbf{e}_1, \dots, \mathbf{e}_N)^\top$, where each \mathbf{e}_i consists of independently and identically distributed (iid) random variables $\mathbf{e}_i = (e_{i1}, \dots, e_{iK})^\top$ having a normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\boldsymbol{\Sigma}$. Replacing $\boldsymbol{\epsilon}(t)$ with $\mathbf{e} \Phi(t)$ in (2.6), and multiplying the whole equation by $\Phi^\top(t)$ from the right and integrating with respect to \mathcal{T} , yields:

$$\mathbf{C} = \mathbf{Z} \mathbf{B} + \mathbf{e}.$$

Estimating \mathbf{B} is an ill-posed problem. The dimension of \mathbf{B} increases exponentially when a large number of basis functions are used to approximate the functions or when a large number of predictors are used in the model. In such cases, traditional estimation methods such as LS and ML fail to provide an estimate for \mathbf{B} . However, the MPL method can produce a stable estimate for \mathbf{B} as long as the functional data are approximated by a small number of basis functions. Because it is computationally intensive, obtaining an MPL estimate of \mathbf{B} may not be possible. This is the case when a relatively large number of basis functions are used to convert discretely observed data

into the functional form. In this paper, we propose using the PLS approach to obtain a stable estimate for \mathbf{B} . Compared with MPL, PLS has several important advantages, including flexibility, straightforward interpretation, and fast computation ability in high-dimensional settings. Note that our proposal is based on an extended version of the FPLSR suggested by Preda and Schiltz (2011).

2.1 PLS for the function-on-function regression model

Let $\mathcal{X}^*(s) = (\mathcal{X}_1^*(s), \dots, \mathcal{X}_M^*(s))$ with $\mathcal{X}_m^*(s) = (\mathcal{X}_{m1}^*, \dots, \mathcal{X}_{mN}^*)$ ($m = 1, \dots, M$) and $\mathcal{Y}^*(t) = (\mathcal{Y}_1^*(t), \dots, \mathcal{Y}_N^*(t))$ denote a matrix of M sets of centered functional predictors of size $(M \times N) \times J_x$ and a matrix of a set of centered functional response of size $N \times J_y$, respectively. Herein, the terms J_x and J_y denote the lengths of time spans where the predictors and response functions observed. Let us now denote the FFRM of $\mathcal{Y}^*(t)$ on $\mathcal{X}^*(s)$ as follows:

$$\mathcal{Y}^*(t) = \int_S \mathcal{X}^*(s) \boldsymbol{\beta}(s, t) ds + \boldsymbol{\epsilon}^*(t), \quad (2.7)$$

where $\boldsymbol{\beta}(s, t)$ and $\boldsymbol{\epsilon}^*(t)$ denote the M sets of bivariate coefficient functions and error functions, respectively. The PLS components of the FFRM (2.7) may be obtained as solutions of Tucker's criterion extended to functional variables as follows:

$$\max_{\substack{\kappa \in \mathcal{L}_2, \|\kappa\|_{\mathcal{L}_2}=1 \\ \zeta \in \mathcal{L}_2, \|\zeta\|_{\mathcal{L}_2}=1}} \text{Cov}^2 \left(\int_S \mathcal{X}^*(s) \kappa(s) ds, \int_T \mathcal{Y}^*(t) \zeta(t) dt \right).$$

The functional PLS components also correspond to the eigenvectors of Escoufier's operators (Preda and Saporta 2005). Let $Z \in \mathcal{L}_2$ denote a random variable. Then, the Escoufier's operators of the centered functional response, $W^{\mathcal{Y}^*}$, and the matrix of M sets of centered functional predictors, $W^{\mathcal{X}^*}$, are given as follows:

$$W^{\mathcal{Y}^*} = \int_T \mathbb{E}[\mathcal{Y}^*(t)Z] \mathcal{Y}^*(t) dt$$

$$W^{\mathcal{X}^*} = \int_S \mathbb{E}[\mathcal{X}^*(s)Z] \mathcal{X}^*(s) ds.$$

The first PLS component of the FFRM (2.7), η_1 , is then equal to the eigenvector of the largest eigenvalue of the product of Escoufier's operators, λ :

$$W^{\mathcal{X}^*} W^{\mathcal{Y}^*} \eta_1 = \lambda \eta_1.$$

The first PLS component is defined as follows:

$$\eta_1 = \int_S \kappa_1(s) \mathcal{X}^*(s) ds,$$

where the weight function $\kappa_1(s)$ is as follows:

$$\kappa_1(s) = \frac{\int_{\mathcal{T}} \mathbb{E} [\mathcal{Y}^*(t)\mathcal{X}^*(s)] dt}{\sqrt{\int_S (\int_{\mathcal{T}} \mathbb{E} [\mathcal{Y}^*(t)\mathcal{X}^*(s)] dt)^2 ds}}$$

The PLS approach is an iterative method, which maximizes the squared covariance between the response and predictor variables as a solution to Tucker’s criterion in each iteration. Let $h = 1, 2, \dots$ denote the iteration number. At each step h , the PLS components are determined by the residuals of the regression models constructed at the previous step as follows:

$$\begin{aligned} \mathcal{X}_h^*(s) &= \mathcal{X}_{h-1}^*(s) - p_h(s)\eta_h, \\ \mathcal{Y}_h^*(t) &= \mathcal{Y}_{h-1}^*(t)\zeta_h(t)\eta_h, \end{aligned}$$

where $\mathcal{X}_0^*(s) = \mathcal{X}^*(s)$, $\mathcal{Y}_0^*(t) = \mathcal{Y}^*(t)$, $p_h(s) = \frac{\mathbb{E}[\mathcal{X}_{h-1}^*(s)\eta_h]}{\mathbb{E}[\eta_h^2]}$, and $\zeta_h(t) = \frac{\mathbb{E}[\mathcal{Y}_h^*(t)\eta_h]}{\mathbb{E}[\eta_h^2]}$. Then, the h th PLS component, η_h corresponds to the eigenvector of the largest eigenvalue of the product of Escoufier’s operators computed at step $h - 1$ as follows:

$$W_{h-1}^{\mathcal{X}^*} W_{h-1}^{\mathcal{Y}^*} \eta_h = \lambda \eta_h.$$

Similarly to the first PLS component, the h th PLS component is obtained as follows:

$$\eta_h = \int_S \kappa_h(s)\mathcal{X}_{h-1}^*(s)ds,$$

where the weight function η_h is given by:

$$\kappa_h(s) = \frac{\int_{\mathcal{T}} \mathbb{E} [\mathcal{Y}_{h-1}^*(t)\mathcal{X}_{h-1}^*(s)] dt}{\sqrt{\int_S (\int_{\mathcal{T}} \mathbb{E} [\mathcal{Y}_{h-1}^*(t)\mathcal{X}_{h-1}^*(s)] dt)^2 ds}}$$

Finally, the ordinary linear regressions of $\mathcal{X}_{h-1}^*(s)$ and $\mathcal{Y}_{h-1}^*(t)$ on η_h are conducted to complete the PLS regression.

The observations of the functional response and functional predictors are intrinsically infinite-dimensional. However, in practice, they are observed in the sets of discrete time points. In this case, the direct estimation of a functional PLS regression becomes an ill-posed problem since the Escoufier’s operators are needed to be estimated using the discretely observed observations. To overcome this problem, we consider the basis function expansions of the functional variables.

Let us now consider the basis expansions of $\mathcal{Y}^*(t)$ and $\mathcal{X}^*(s)$ as follows:

$$\mathcal{Y}^*(t) = \sum_{k=1}^{K_{\mathcal{Y}}} C_k \phi_k(t) = \mathbf{C}\Phi(t)$$

$$\mathcal{X}^*(s) = \sum_{j=1}^{K_{\mathcal{X}}} \mathbf{C}_j \psi_j(s) = \mathbf{D}\Psi(s).$$

Denote by $\Phi = \int_{\mathcal{T}} \Phi(t)\Phi^{\top}(t)dt$ and $\Psi = \int_{\mathcal{S}} \Psi(s)\Psi^{\top}(s)ds$ the $K_{\mathcal{Y}} \times K_{\mathcal{Y}}$ and $K_{\mathcal{X}} \times K_{\mathcal{X}}$ dimensional symmetric matrices of the inner products of the basis functions, respectively. Also, let $\Phi^{1/2}$ and $\Psi^{1/2}$ denote the square roots of Φ and Ψ , respectively. Then, we consider the PLS regression of $\mathbf{C}\Phi^{1/2}$ on $\mathbf{D}\Psi^{1/2}$ to approximate the PLS regression of $\mathcal{Y}^*(t)$ on $\mathcal{X}^*(s)$ as follows:

$$\mathbf{C}\Phi^{1/2} = \mathbf{D}\Psi^{1/2}\Xi + \delta,$$

where Ξ and δ denote the regression coefficients and the residuals, respectively. Now let $\widehat{\Xi}^h$ denote the estimate of Ξ using the PLS regression at step h . Then we have,

$$\mathbf{C}\Phi^{1/2} = \mathbf{D}\Psi^{1/2}\widehat{\Xi}^h,$$

$$\widehat{\mathcal{Y}}^*(t) = \int_{\mathcal{S}} \mathcal{X}^*(s)\Theta^h(s, t)ds,$$

where

$$\Theta^h(s, t) = \sum_{k=1}^{K_{\mathcal{Y}}} \sum_{j=1}^{K_{\mathcal{X}}} \left((\Psi^{1/2})^{-1} \widehat{\Xi}^h (\Phi^{1/2})^{-1} \right) \psi_j(s)\phi_k(t).$$

Herein, the term $\Theta^h(s, t)$ denotes the PLS approximation of the coefficient function $\beta(s, t)$ given in (2.7).

Throughout this paper, two main PLS algorithms were used to obtain the model parameters: NIPALS and SIMPLS. While the NIPALS algorithm iteratively deflates the functional predictor and functional response, the SIMPLS algorithm iteratively deflates the covariance operator. In our numerical analyses, the functions `plsreg2` and `pls.regression` of the R packages `plsdepot` (Sanchez 2012) and `plsgenomics` (Boulesteix et al. 2018) were used to perform NIPALS and SIMPLS algorithms, respectively.

3 Simulation studies

Various Monte-Carlo experiments were conducted under different scenarios to investigate the finite-sample performances of the proposed PLS-based methods. Throughout these experiments, $MC = 1000$ Monte-Carlo simulations were performed, and the results were compared with LS, ML, MPL, and two available FFRM models: (1)

penalized flexible functional regression (PFFR) from Ivanescu et al. (2015) [refer to the R package “refund” from Goldsmith et al. (2018), for details] and (2) the functional regression with functional response (FREG) from Ramsay and Silverman (2006) (refer to the R package “fda.usc” from Febrero-Bande and Oviedo de la Fuente (2012), for details].

Throughout the experiments, the following simple FFRM was considered:

$$\mathcal{Y}_i(t) = \beta_0(t) + \int_{\mathcal{T}} \mathcal{X}_i(s)\beta_1(s, t)ds + \epsilon_i(t),$$

where $s \in S, t \in \mathcal{T}$, and $N = 100$ and 200 individuals were considered. A comparison was made using the average mean squared error (AMSE). For each experiment, the generated data were divided into two parts: (1) The first half of the data were used to build the FFRM, and the following AMSE was calculated:

$$AMSE = (N/2)^{-1} \sum_{i=1}^{N/2} [\mathcal{Y}_i(t) - \widehat{\mathcal{Y}}_i(t)]^2,$$

where $\widehat{\mathcal{Y}}_i(t)$ denotes the fitted function for i th individual. (2) The second part of the data was used to evaluate the prediction performances of the methods based on the constructed FFRMs using the first-half of the data:

$$AMSE_p = (N/2)^{-1} \sum_{i=N/2+1}^N [\mathcal{Y}_i(t) - \widehat{\mathcal{Y}}_i^*(t)]^2,$$

where $\widehat{\mathcal{Y}}_i^*(t)$ denotes the predicted response function for i th individual. Also, we applied the model confidence set (MCS) procedure proposed by Hansen et al. (2011) [refer to the R package “MCS” from Barnardi and Catania (2018), for details] on the prediction errors obtained by the FFRM procedures to determine superior method(s). The MCS procedure was performed using 5000 bootstrap replications at a 95% confidence level. Computations were performed using R Core Team (2019) on an Intel Core i7 6700HQ 2.6 GHz PC.

The following process was used to generate functional variables:

- Generate the observations of the predictor variable \mathcal{X} at discrete time points s_j as follows:

$$\mathcal{X}_{ij} = \kappa_i(s_j) + \epsilon_{ij},$$

where $j = 1, \dots, 50, \epsilon_{ij} \sim N(0, 1), s_j \sim U(-1, 1)$, and $\kappa_i(s)$ is generated as:

$$\kappa_i(s) = \cos[\exp(a_{1i}s)] + a_{2i}s,$$

where $a_{1i} \sim N(2, 0.02^2)$ and $a_{2i} \sim N(-3, 0.04^2)$.

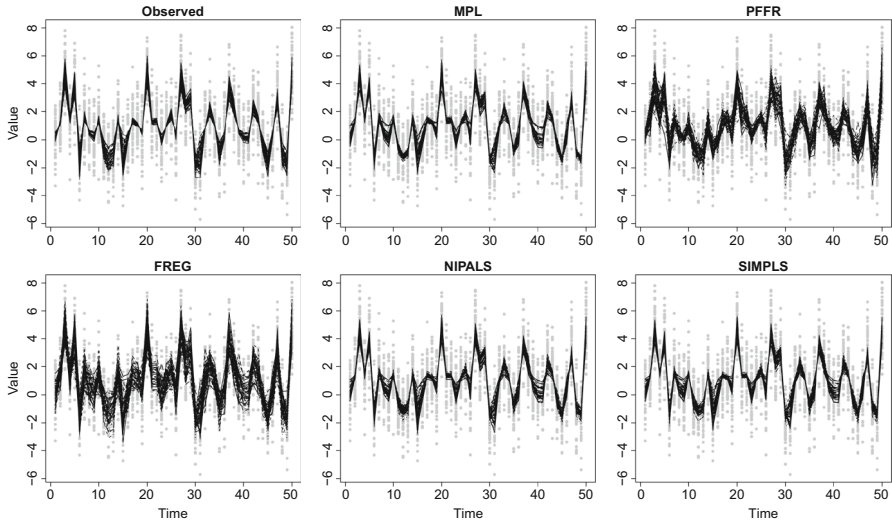


Fig. 1 Plots of the generated N sets of discrete data (gray points) and fitted smooth functions (black lines) when $\rho = 1$ and $K = 20$ numbers of basis functions were used in B -spline basis. The MPL, PFFR, FREG, NIPALS, and SIMPLS were used to obtain fitted response functions

- Similarly, generate the data points of the response variable \mathcal{Y} at time points t_j using the following process:

$$\mathcal{Y}_{ij} = \eta_i(t_j) + \epsilon_{ij},$$

where $t_j \sim U(-1, 1)$ and $\eta_i(t)$ is generated as:

$$\eta_i(t) = \vartheta_i(t) + \varepsilon_i(t),$$

where $\vartheta_i(t) = \sin[\exp(a_{1_i}t)] + a_{2_i}t + 2t^2$, $\varepsilon_i(t) = \mathbf{e}_i^\top \Phi(t)$, \mathbf{e}_i s are iid multivariate Gaussian random errors with mean $\mathbf{0}$ and variance-covariance matrix $\Sigma = [(0.5)^{|k-l|}\rho]_{k,l}$, and $\Phi(t)$ is the B -spline basis function. Throughout the simulations, four different variance parameters were considered: $\rho = [0.5, 1, 2, 4]$.

The data generated at discrete time points were first converted into functions using the B -spline basis with $K = [10, 20, 30, 40]$ numbers of basis functions. An example of the observed data with noise and the fitted smooth functions for the generated response variable is presented in Fig. 1.

Before presenting our findings, we note that the results do not vary considerably with different choices of N ; therefore, to save space, we only report the results for $N = 100$. The LS and ML methods failed to provide an estimate for \mathbf{B} because of the singular matrix problem and degenerate structure of the generated data; thus, we only report on the comparative studies with MPL, PFFR, FREG, NIPALS, and SIMPLS. Our results obtained from the fitted and predicted models are presented in Figs. 2 and 3, respectively.

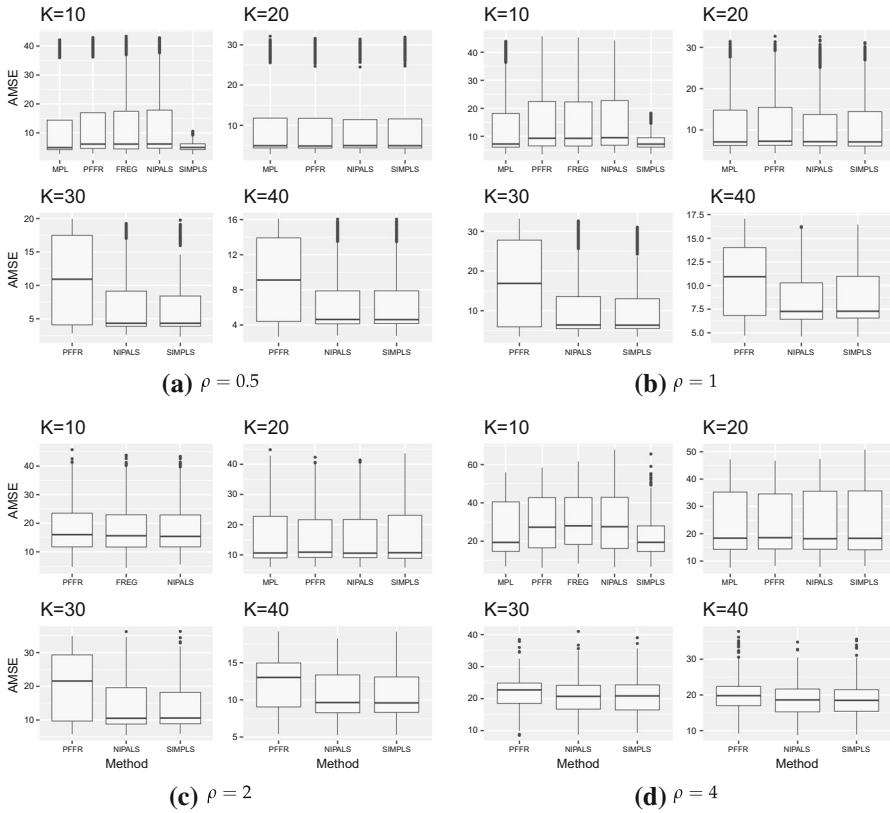


Fig. 2 Fitted model performances: Computed AMSE values of the MPL, PFFR, FREG, NIPALS, and SIMPLS methods. The data were generated based on the variance parameter $\rho = [0.5, 1, 2, 4]$ and $K = [10, 20, 30, 40]$ numbers of basis functions were used to convert the data to smooth functions

They illustrate that, when $K = 10$, the proposed SIMPLS algorithm performed considerably better than the other methods in terms of AMSE, $AMSE_p$, and their associated standard errors. Also, the NIPALS algorithm showed competitive performance to other methods. We observed that the FREG and MPL failed to provide an estimate for the model parameter when $K \geq 20$ and $K \geq 30$, respectively. For a small to moderate variance parameter, the proposed NIPALS and SIMPLS performed better than the PFFR, while all three estimation methods tended to have similar performances when $\rho = 4$.

The results for the MCS analysis are presented in Table 1. The values in this table correspond to the percentages of the superiorities of the methods from 1000 Monte-Carlo simulations. Our findings demonstrate that the proposed NIPALS and SIMPLS algorithms produced significantly better prediction performances compared with their competitors except when $K = 20$.

Furthermore, we examined the computing performances of the methods considered in this study. Figure 4 represents the elapsed computational times for a different number of basis functions obtained by a single Monte-Carlo experiment. This figure illustrates

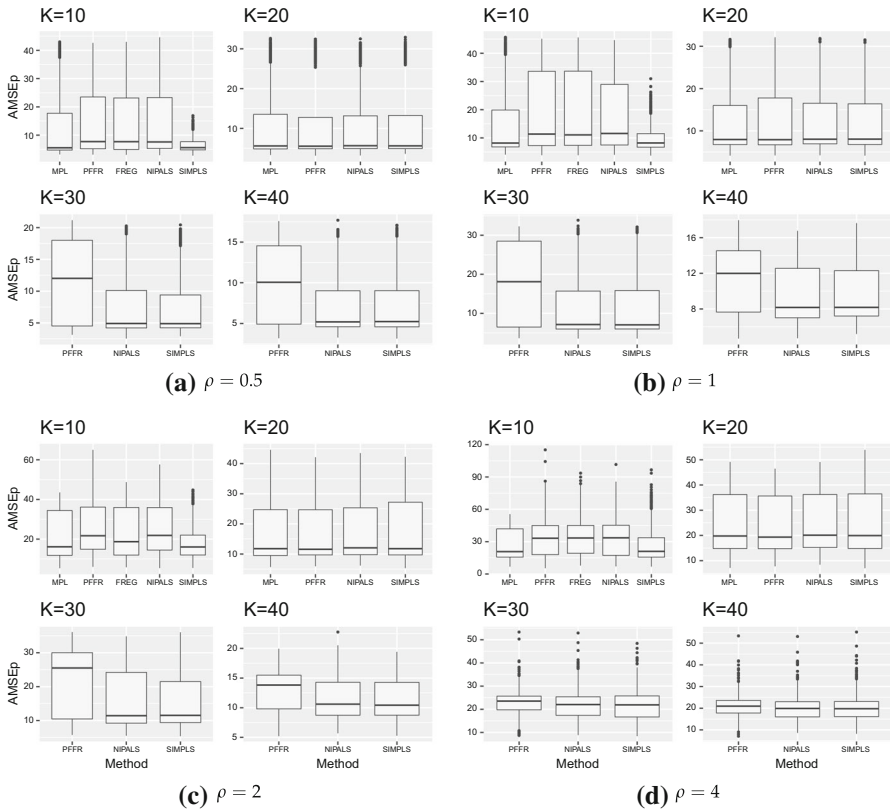


Fig. 3 Predicted model performances: Computed $AMSE_p$ values of the MPL, PFFR, FREG, NIPALS, and SIMPLS methods. Data were generated based on the variance parameter $\rho = [0.5, 1, 2, 4]$ and $K = [10, 20, 30, 40]$ numbers of basis functions were used to convert the data to smooth functions

that both the NIPALS and SIMPLS algorithms had considerably shorter computational times than other methods. The computational time of MPL increased exponentially with increasing K ; therefore, we do not recommend its use when a large number of basis functions are used in the FFRM.

4 Data analyses

In this section, we evaluate the performances of the proposed PLS-based methods using an empirical data example: daily North Dakota weather data. The daily dataset was collected from 70 stations across North Dakota (see Table 2), from January 2010 to December 2018 (dataset are available from the North Dakota Agricultural Weather Network Center: <https://ndawn.ndsu.nodak.edu>). The dataset has three meteorological variables: average temperature ($^{\circ}C$), average wind speed (m/s), and total solar radiation (MJ/m^2).

Table 1 MCS analysis results

ρ	Method	$K = 10$ (%)	$K = 20$	$K = 30$	$K = 40$
0.5	MPL	0.090	0.532%	–	–
	PFFR	0.000	0.000%	0.000%	0.000%
	FREG	0.000	–	–	–
	NIPALS	0.610	0.299%	0.547%	0.588%
	SIMPLS	0.318	0.197%	0.467%	0.424%
1	MPL	0.183	0.603%	–	–
	PFFR	0.000	0.000%	0.000%	0.000%
	FREG	0.000	–	–	–
	NIPALS	0.529	0.257%	0.485%	0.655%
	SIMPLS	0.301	0.161%	0.528%	0.352%
2	MPL	0.001	0.517%	–	–
	PFFR	0.000	0.000%	0.000%	0.004%
	FREG	0.000	–	–	–
	NIPALS	0.321	0.143%	0.478%	0.533%
	SIMPLS	0.682	0.354%	0.526%	0.447%
4	MPL	0.644	0.803%	–	–
	PFFR	0.000	0.000%	0.147%	0.163%
	FREG	0.000	–	–	–
	NIPALS	0.212	0.192%	0.410%	0.494%
	SIMPLS	0.150	0.008%	0.450%	0.357%

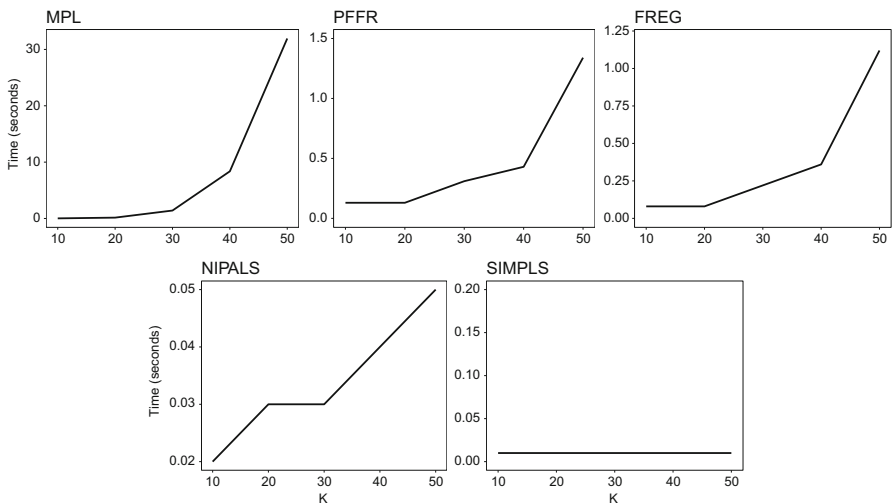
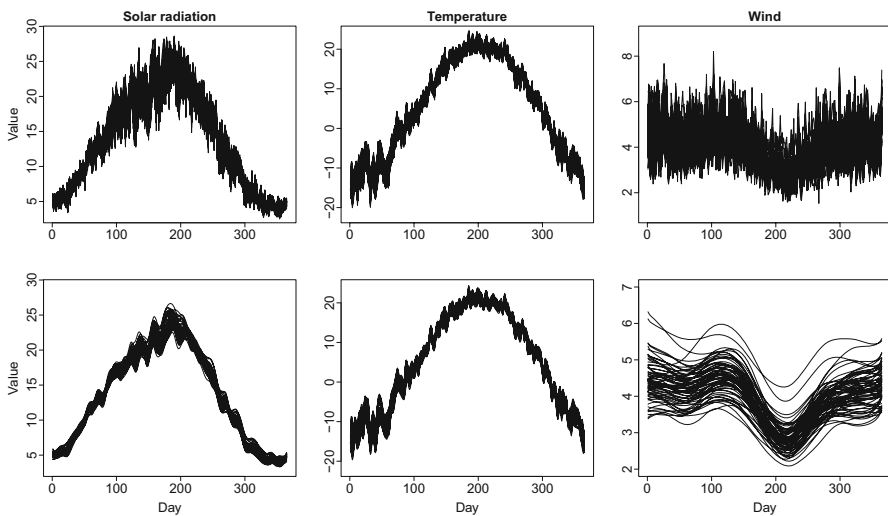


Fig. 4 Estimated computational times in second for the MPL, PFFR, FREG, NIPALS, and SIMPLS estimators

Table 2 Station names for the North Dakota weather data

Station	Station	Station	Station	Station	Station	Station
Ada	Cavalier	Fingal	Hofflund	Marion	Perley	Sidney
Baker	Crary	Forest River	Humboldt	Mavie	Pillsbury	Stephen
Beach	Crosby	Galesburg	Inkster	Mayville	Plaza	Streeter
Berthold	Dazey	Grafton	Jamestown	McHenry	Prosper	Thomas
Bottineau	Dickinson	Grand Forks	Karlsruhe	Michigan	Robinson	Tappen
Bowbells	Dunn	Greenbush	Langdon	Minot	Rolla	Turtle Lake
Bowman	Edgeley	Harvey	Leonard	Mohall	Roseau	Wahpeton
Brorson	Ekre	Hazen	Linton	Mooreton	Ross	Warren
Cando	Eldred	Hettinger	Lisbon	Mott	Rugby	Watford City
Carrington	Fargo	Hillsboro	Mandan	Oakes	Sabin	Williston

**Fig. 5** Plots of discrete data (first row) and their calculated smooth functions (second row) for daily weather data

The data were averaged over the entire time, and B -spline basis function expansion was used to convert the discretely observed data to functional forms. Using the GCV criterion, the estimated numbers of basis functions of the temperature, wind speed, and solar radiation variables were [147, 62, 150]. The plots of the observed dataset and its computed functions are presented in Fig. 5.

For the dataset, we predicted total solar radiation using temperature and wind speed variables. For this purpose, the dataset was divided into the following two parts: FFRMs were constructed based on the variables of the first 50 stations to predict the total solar radiation functions of the remaining 20 stations. However, FREG and PFFR do not allow for more than one functional predictor in the FFRM. Therefore, to compare all the methods considered in this study, we first constructed the FFRM using only one

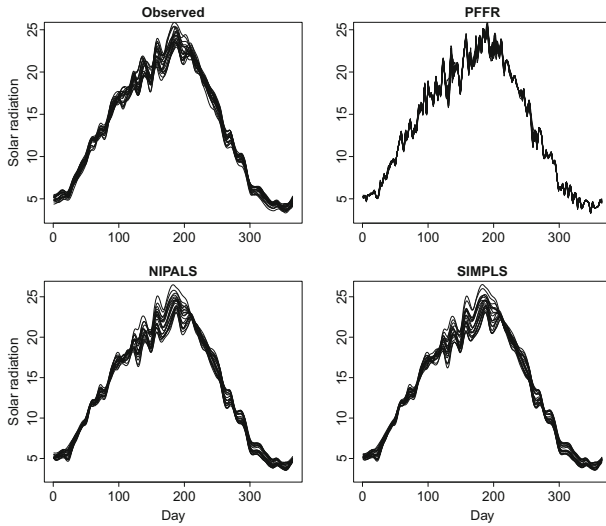


Fig. 6 Plots of observed and predicted solar radiation functions for the test stations. The FFRM was constructed using one predictor (wind speed). The PFFR, NIPALS, and SIMPLS methods were used to estimate the model parameter $\beta_1(s, t)$

functional predictor as follows:

$$\mathcal{Y}_i(t) = \int_S \mathcal{X}_{i1}(s)\beta_1(s, t) + \epsilon_i(t) \quad i = 1, \dots, 50, \tag{4.1}$$

where $\mathcal{Y}_i(t)$ and $\mathcal{X}_{i1}(s)$ denote the i th function of the solar radiation and wind speed. Then, we calculated the $AMSE_p$ as follows:

$$AMSE_p = \frac{1}{20} \sum_{i=51}^{70} [\mathcal{Y}_i(t) - \widehat{\mathcal{Y}}_i^*(t)]^2,$$

where $\widehat{\mathcal{Y}}_i^*(t)$ denotes the predicted response function for i th station. The MPL and FREG failed to provide an estimate for the regression parameter because of the singular matrix problem. The calculated $AMSE_p$ for the PFFR, NIPALS, and SIMPLS were [275.0590, 100.4674, 100.2813]. The results show that, of all methods, the proposed PLS-based methods were most effective. The observed and predicted total solar radiation functions of the test stations using model (4.1) are presented in Fig. 6.

Next, we constructed the FFRM using more than one functional predictor as follows:

$$\mathcal{Y}_i(t) = \int_S \mathcal{X}_{i1}(s)\beta_1(s, t) + \int_S \mathcal{X}_{i2}(s)\beta_2(s, t) + \epsilon_i(t), \quad i = 1, \dots, 50, \tag{4.2}$$

where $\mathcal{Y}_i(t)$, $\mathcal{X}_{i1}(s)$, and $\mathcal{X}_{i2}(s)$ denote the i th function of the solar radiation, wind speed, and temperature, respectively. In this case, we only compare the MPL, NIPALS,

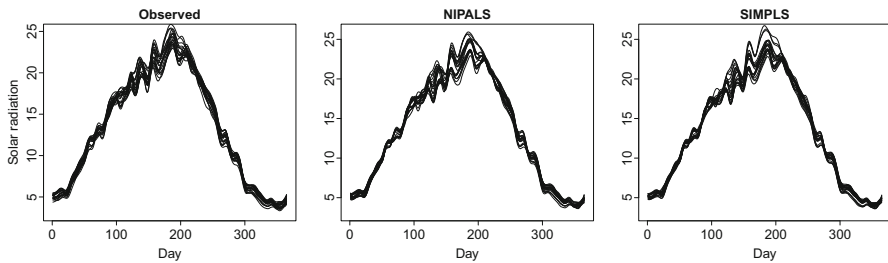


Fig. 7 Plots of observed and predicted solar radiation functions for the test stations: daily weather data. The FFRM was constructed using two predictors (wind speed and temperature). The NIPALS, and SIMPLS methods were used to estimate the model parameters $\beta_1(s, t)$ and $\beta_2(s, t)$

and SIMPLS methods because the FREG and PFFR do not allow more than one functional predictor in the model. For the data, the MPL failed to provide an estimate for the regression parameter because of the singular matrix problem; therefore, we only compared the proposed NIPALS and SIMPLS methods. The calculated $AMSE_p$ values for the NIPALS and SIMPLS, respectively, were [56.93, 57.70]. The results show that the NIPALS performed better than the SIMPLS. The observed and predicted total solar radiation functions for model (4.2) are provided in Fig. 7.

In summary, our proposed PLS methods tend to produce superior performances than existing estimation methods and other available FFRMs. Additionally, the proposed methods avoided common computing problems. Computational issues observed when using the MPL and FREG are presented as follows:

```
> Error: cannot allocate vector of size 7.3 Gb
> Error in chol.default(Asym): the leading minor of order 7264 is not positive definite
> Error in solve.default(Sigma): system is computationally singular: +
> reciprocal condition number = 1.64815e-20
```

These errors were attributable to the relatively large number of basis functions estimated by the GCV. A possible solution for overcoming these problems is to use a high-performance computer or a smaller number of basis functions in the modeling phase. However, the proposed PLS-based methods can successfully provide estimates for the model parameters in a few seconds without producing any errors listed above (an example R code for the analysis of daily North Dakota weather data is available at <https://github.com/hanshang/FPLSR>).

5 Conclusion

Analysis of the association between functional response and functional predictors has received considerable attention in many research fields. For this purpose, several FFRMs have been proposed, with their primary objective being to estimate the model parameters accurately. Existing estimation methods work well when a small number of predictors are used in the model. Existing estimation methods also work well when a finite number basis functions are used to convert discretely observed data to smooth functions. However, when the opposite occurs, estimation methods suffer from two

key problems. First, they fail to provide estimates for the model parameters because of the singular matrix problem. Second, they are computationally time-consuming.

In the present study, we integrated the PLS approach with an FFRM and used two principal algorithms, NIPALS and SIMPLS, to estimate the parameter matrix. The finite-sample performances of the proposed approaches were evaluated using Monte-Carlo experiments and empirical data analysis. We compared our results with some other estimation methods within an FFRM. Our findings illustrate that the proposed approaches perform better than several existing estimation methods. They avoid the singular matrix problem by decomposing the response and predictor variables into orthogonal matrices. Additionally, they are computationally more efficient compared with available estimation methods.

For the proposed methods, two points need to clarify: (1) Throughout this study, we assume that the functional predictor variables are observed on the same domain [(see model (1.1)]. However, there may be some cases where the dataset includes multiple predictors observed on different domains (see, e.g., Happ and Greven 2018). In such a case, the following FFRM can be considered:

$$\mathcal{Y}_i(t) = \beta_0(t) + \sum_{m=1}^M \int_{S_m} \mathcal{X}_{im}(s_m) \beta_m(s_m, t) + \epsilon_i(t), \quad (5.1)$$

where S_m denotes the domain of m th functional predictor. All the functional predictor matrices \mathcal{X}_m , for $m = 1, \dots, M$, have the same row lengths, and they can be stacked into a vector \mathcal{X} . Our proposed method can also be used to estimate the variable-domain FFRM given in (5.1). (2) In this study, we use the same finite-dimensional basis functions method (B -spline) to convert the discretely observed data points of predictor variables into their functional forms. However, using different basis functions methods for different predictors may be more useful in some cases; for example, B -spline and Fourier bases can be used to approximate the functional variables having non-periodic and periodic structures, respectively. In such a case, the basis coefficient matrices produced by different basis expansion methods will have the same row lengths; and thus, using different basis functions for different predictors does not interfere with the use of our proposed method.

The present research can be extended in three directions: (1) We only considered two fundamental algorithms, NIPALS and SIMPLS, to estimate the FFRM. However, numerous algorithms, such as improved kernel PLS (Dayal and MacGregor 1997), Bidiag2 (Golub and Kahan 1965), and non-orthogonalized scores (Martens and Naes 1989), are available in the PLS literature; and could be included for performance comparison. (2) In the presence of outliers, it may be advantageous to consider a robust PLS algorithm, such as the robust iteratively reweighted SIMPLS in Alin and Agostinelli (2017). (3) In our numerical analyses, the finite sample performance of the proposed method is evaluated using a fixed $h = 5$ number of PLS components. However, its performance may depend on different choices of the number of PLS components. Thus, a cross-validation approach of Yao and Tong (1998), Racine (2000), and Antoniadis et al. (2009) may be proposed to determine the optimum number of PLS components.

Acknowledgements The authors acknowledge R code assistance from Dr. Fabian Scheipl at the Ludwig-Maximilians University of Munich and Ms. Julia Wrobel at Columbia University. The authors thank two anonymous referees for their careful reading of our manuscript and valuable suggestions and comments, which have helped us produce a much-improved paper. The second author also acknowledges the financial support from a research grant at the College of Business and Economics at the Australian National University.

References

- Aguilera AM, Escabias M, Preda C, Saporta G (2010) Using basis expansions for estimating functional PLS regression: applications with chemometric data. *Chemometr Intell Lab Syst* 104(2):289–305
- Alin A, Agostinelli C (2017) Robust iteratively reweighted SIMPLS. *J Chemometr* 31(3):e2881
- Antoniadis A, Paparoditis E, Sapatinas T (2009) Bandwidth selection for functional time series prediction. *Stat Prob Lett* 79:733–740
- Barnardi M, Catania L (2018) The model confidence set package for R. *Int J Comput Econ Econometr* 8(2):144–158
- Boulesteix A-L, Durif G, Lambert-Lacroix S, Peyre J, Strimmer K (2018) *plsgenomics: PLS Analyses for Genomics*. R package version 1.5–2. <https://CRAN.R-project.org/package=plsgenomics>. Accessed 30 July 2019
- Brumback BA, Rice JA (1998) Smoothing spline models for the analysis of nested and crossed samples of curves. *J Am Stat Assoc* 93(443):961–976
- Cardot H, Sarda P (2008) Varying-coefficient functional linear regression models. *Commun Stat Theory Methods* 37(20):3168–3203
- Chiou J-M, Yang Y-F, Chen Y-T (2016) Multivariate functional linear regression and prediction. *J Multivar Anal* 146:301–312
- Cuevas A (2014) A partial overview of the theory of statistics with functional data. *J Stat Plan Inference* 147:1–23
- Dayal BS, MacGregor JF (1997) Improved PLS algorithms. *J Chemometr* 11(1):73–85
- de Jong S (1993) SIMPLS: an alternative approach to partial least squares regression. *Chemometr Intell Lab Syst* 18(3):251–263
- Fan J, Zhang W (1999) Statistical estimation in varying coefficient models. *Ann Stat* 27(5):1491–1518
- Febrero-Bande M, Oviedo de la Fuente M (2012) Statistical computing in functional data analysis: the R package *fda.usc*. *J Stat Softw* 51(4):1–28
- Febrero-Bande M, Galeano P, Gozalez-Manteiga W (2017) Functional principal component regression and functional partial least-squares regression: an overview and a comparative study. *Int Stat Rev* 85(1):61–83
- Ferraty F, Vieu P (2006) *Nonparametric functional data analysis*. Springer, New York
- Goldsmith J, Scheipl F, Huang L, Wrobel J, Gellar J, Harezlak J, McLean MW, Swihart B, Xiao L, Crainiceanu C, Reiss PT (2018) *Refund: regression with functional data*. R package version 0.1–17. <https://CRAN.R-project.org/package=refund>. Accessed 30 July 2019
- Golub G, Kahan W (1965) Calculating the singular values and pseudo-inverse of a matrix. *SIAM J Numer Anal* 2(2):205–224
- Hansen PR, Lunde A, Nason JM (2011) The model confidence set. *Econometrica* 79(2):453–497
- Happ C, Greven S (2018) Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J Am Stat Assoc Theory Methods* 113(522):649–659
- Hoover DR, Rice JA, Wu CO, Yang L-P (1998) Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 85(4):809–822
- Horvath L, Kokoszka P (2012) *Inference for functional data with applications*. Springer, New York
- Huang JZ, Wu CO, Zhou L (2002) Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika* 89(1):111–128
- Huang JZ, Wu CO, Zhou L (2004) Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Stat Sinica* 14(3):763–788
- Hyndman RJ, Shang HL (2009) Forecasting functional time series. *J Korean Stat Soc* 38(3):199–211
- Ivanescu AE, Staicu A-M, Scheipl F, Greven S (2015) Penalized function-on-function regression. *Comput Stat* 30(2):539–568

- Krämer N, Boulesteix A-L, Tutz G (2008) Penalized partial least squares with applications to B-spline transformations and functional data. *Chemometr Intell Lab Syst* 94(1):60–69
- Martens H, Naes T (1989) *Multivariate calibration*. Wiley, New York
- Matsui H, Kawano S, Konishi S (2009) Regularized functional regression modeling for functional response and predictors. *J Math-for-Ind* 1(A3):17–25
- Müller H-G, Yao F (2008) Functional additive models. *J Am Stat Assoc* 103(484):1534–1544
- Preda C, Saporta G (2005) PLS regression on a stochastic process. *Comput Stat Data Anal* 48(1):149–158
- Preda C, Schiltz J (2011) Functional PLS regression with functional response: the basis expansion approach. In: *Proceedings of the 14th applied stochastic models and data analysis conference*. Universita di Roma La Spienza. p. 1126–33
- R Core Team (2019) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>. Accessed 30 July 2019
- Racine J (2000) Consistent cross-validatory model-selection for dependent data: *h_v*-block cross-validation. *J Econometr* 99:39–61
- Ramsay JO, Dalzell CJ (1991) Some tools for functional data analysis. *J R Stat Soc Ser B* 53(3):539–572
- Ramsay JO, Silverman BW (2002) *Applied functional data analysis*. Springer, New York
- Ramsay JO, Silverman BW (2006) *Functional data analysis*. Springer, New York
- Reiss PT, Ogden TR (2007) Functional principal component regression and functional partial least squares. *J Am Stat Assoc* 102(479):984–996
- Sanchez G (2012) *plsdepot: Partial Least Squares (PLS) data analysis methods*. R package version 0.1.17. <https://CRAN.R-project.org/package=plsdepot>. Accessed 30 July 2019
- Şentürk D, Müller H-G (2005) Covariate adjusted correlation analysis via varying coefficient models. *Scand J Stat* 32(3):365–383
- Wang W (2014) Linear mixed function-on-function regression models. *Biometrics* 70(4):794–801
- Wold H (1974) Causal flows with latent variables: partings of the ways in the light of NIPALS modelling. *Eur Econ Rev* 5(1):67–86
- Wu CO, Chiang C-T (2000) Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Stat Sinica* 10(2):443–456
- Wu Y, Fand J, Müller H-G (2010) Varying-coefficient functional linear regression. *Bernoulli* 16(3):730–758
- Yamanishi Y, Tanaka Y (2003) Geographically weighted functional multiple regression analysis: a numerical investigation. *J Jpn Soc Comput Stat* 15(2):307–317
- Yao Q, Tong H (1998) Cross-validatory bandwidth selections for regression estimation based on dependent data. *J Stat Plan Inference* 68:387–415
- Yao F, Müller H-G, Wang J-L (2005) Functional linear regression analysis for longitudinal data. *Ann Stat* 33(6):2873–2903
- Zhu H, Fan J, Kong L (2014) Spatially varying coefficient model for neuroimaging data with jump discontinuities. *J Am Stat Assoc* 109(507):1084–1098

Ufuk Beyaztas is an assistant professor in the Department of Statistics at Bartın University, Turkey. Ufuk received his PhD in Statistics from Dokuz Eylül University. His research interests are regression analysis, resampling methods, time series analysis and forecasting, and functional data analysis.

Han Lin Shang is currently an Associate Professor of Statistics at the Research School of Finance, Actuarial Studies and Statistics at Australian National University in Canberra, Australia. He obtained a First Class Honours degree in Statistics in 2016 from the Department of Mathematics and Statistics from La Trobe University and earned his Ph.D. in Statistics with Mollie Holman medal from Department of Econometrics and Business Statistics from Monash University in 2010. He has research interests in functional time-series analysis, empirical finance, actuarial studies and demographic forecasting. His research publications include articles in *Statistics: Journal of Computational and Graphical Statistics*, *Annals of Applied Statistics*, *Statistics and Computing*, *Computational Statistics and Data Analysis*, *Journal of Time Series Analysis*, *AStA Advances in Statistical Analysis*, *Journal of Nonparametric Statistics*, *Journal of Multivariate Analysis*, *Journal of Royal Statistical Society: Series A*; *Journal of the American Statistical Association: Theory & Methods*. His research publications include articles in *Demography: Journal of Population Research*, *Demographic Research*, *Population Studies*, *Demography*. His research publications include articles in *Actuarial Science: Insurance: Mathematics and Economics*, *ASTIN Bulletin*. His research publications include articles in *Finance: Journal of Banking and Finance*, *European Financial*

Management, Annals of Operations Research. He is currently serving as an associate editor for Journal of Computational and Graphical Statistics, Journal of Royal Statistical Society: Series A, Computational Statistics and an editor for Australian & New Zealand Journal of Statistics.