Variable selection in nonparametric functional concurrent regression

Rahul GHOSAL1* and Arnab MAITY2

1Department of Biostatistics, Johns Hopkins University, Baltimore, MD, U.S.A.
2Department of Statistics, North Carolina State University, Raleigh, NC, U.S.A.

Key words and phrases: Functional data analysis; nonparametric functional concurrent regression; variable selection.

MSC 2020: Primary 62G08; secondary 65D10.

Abstract: We develop a new method for variable selection in nonparametric functional concurrent regression. The commonly used functional linear concurrent model (FLCM) is far too restrictive in assuming linearity of the covariate effects, which is not necessarily true in many real-world applications. The nonparametric functional concurrent model (NPFCM), on the other hand, is much more flexible and can capture complex dynamic relationships present between the response and the covariates. We extend the classically used variable selection methods, e.g., group LASSO, group SCAD and group MCP, to perform variable selection in NPFCM. We show via numerical simulations that the proposed variable selection method with the non-convex penalties can identify the true functional predictors with minimal false-positive rate and negligible false-negative rate. The proposed method also provides better out-of-sample prediction accuracy compared to the FLCM in the presence of nonlinear effects of the functional predictors. The proposed method’s application is demonstrated by identifying the influential predictor variables in two real data studies: a dietary calcium absorption study, and some bike-sharing data. The Canadian Journal of Statistics 50: 142–161; 2022 © 2021 The Authors. The Canadian Journal of Statistics/La revue canadienne de statistique published by Wiley Periodicals LLC on behalf of Statistical Society of Canada.

Résumé: Cet article propose une nouvelle méthode de sélection de variables dans un contexte de modèles concurrents fonctionnels non paramétriques. Comme l’hypothèse de linéarité des effets des covariables qu’impose un modèle concurrent fonctionnel linéaire usuel soit trop restrictive pour que ce dernier soit valide et utile dans des applications réelles, les auteurs de ce travail ont opté pour des modèles fonctionnels concurrents non paramétriques (NPFCM). Ces modèles sont bien plus souples et capables de capturer les relations dynamiques complexes liant une variable réponse et ses covariables. Ils ont, donc, repris des méthodes de sélection de variables classiques, telles: LASSO, SCAD et MCP de groupes pour les étendre aux modèles de type NPFCM. Des simulations numériques montrent clairement que la méthode proposée avec les pénalités non convexes peut identifier les véritables prédicteurs fonctionnels avec un taux de faux positifs minimal et un taux de faux négatifs négligeable. Aussi, en présence d’effets non linéaires des prédicteurs fonctionnels, elle fournit une meilleure précision de prédiction hors échantillon comparativement à celle d’un modèle concurrent fonctionnel linéaire. En guise d’illustration pratique, la méthode proposée a été appliquée à deux jeux de données, l’un portant sur l’absorption de calcium alimentaire et l’autre sur un système de partage de vélos. La revue canadienne de statistique 50: 142–161; 2022 © 2021 Les auteurs. La revue canadienne de statistique/The Canadian Journal of Statistics, publiée par Wiley Periodicals LLC au nom de la Société statistique du Canada.

* Corresponding author: rghosal@ncsu.edu

© 2021 The Authors. The Canadian Journal of Statistics/La revue canadienne de statistique published by Wiley Periodicals LLC on behalf of Statistical Society of Canada. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
1. INTRODUCTION

Function-on-function regression refers to the class of regression models with both the response variable and the covariates being functions over a continuous domain, such as time. The functional concurrent regression model is a particular class of function-on-function regression, where the value of the response at the current time point depends only on the values of the predictors at that specific time point. The most commonly used concurrent regression model is the functional linear concurrent model (FLCM) (Ramsay & Silverman, 2005), which assumes that the relationship between the response and the covariates is linear and modelled using smooth univariate regression functions. These regression functions can capture the time-varying effect of the covariates on the response. Estimation and inference in the functional linear concurrent regression and the closely related varying coefficient model (Hastie & Tibshirani, 1993) have been widely studied in the literature.

Although FLCM is useful and finds applications in longitudinal data analysis, it is restrictive in assuming a linear relationship between the response and predictors. The nonparametric functional concurrent model (NPFCM) overcomes this limitation by specifying a nonparametric relationship that is more general and flexible for modelling purposes. Multiple methods exist in the literature for estimation and inference in NPFCM using smoothing splines (Kim, Maity & Staicu, 2018a), Gaussian process regression (Shi, Murray-Smith & Titterington, 2005) and local kernel smoothing techniques (Jiang & Wang, 2011); see Maity (2017) and the references therein for a review of various methods regarding NPFCM. With the advent of modern high-throughput technologies, functional data structures have recently become complex and high-dimensional in nature, and this has motivated researchers to develop variable selection methods to identify the true underlying set of influential predictors. Several variable selection methods exist in the functional data analysis literature for scalar-on-function (Gertheiss, Maity & Staicu, 2013; Fan, James & Radchenko, 2015) and function-on-scalar (Chen, Goldsmith & Ogden, 2016; Parodi & Reimherr, 2018) regression models. However, the literature on variable selection in the functional concurrent regression model is relatively sparse. Goldsmith & Schwartz (2017) proposed a variable selection method for the FLCM using the variational Bayes approach. Recently, Ghosal et al. (2020) developed a variable selection method for FLCM regression by extending the classically used penalized variable selection methods. However, these methods are for the FLCM, which might not capture the actual relationship between the response and the predictors.

In this article, we consider an additive nonparametric functional concurrent regression model (Kim et al., 2018b), which is an extension of the functional generalized additive model (FGAM) (McLean et al., 2014) for functional responses. Scheipl et al. (2016) extended this class of models to non-Gaussian functional responses, allowing for nested or crossed functional random effects. We develop a variable selection method for the NPFCM, extending the classically used penalized variable selection methods like LASSO (Tibshirani, 1996), SCAD (Fan & Li, 2001) and MCP (Zhang, 2010). To the best of our knowledge, ours is the first work to develop a variable selection method for NPFCM. We follow the approach of Ghosal et al. (2020) and show that for nonparametric functional concurrent regression, the variable selection problem reduces to a group LASSO (Yuan & Lin, 2006) and its natural extension, namely a group SCAD or a group MCP problem. Through simulations, we illustrate that the proposed variable selection method can identify the true underlying predictors with minimal false-positive rate (FPR) and negligible false-negative rate (FNR). The proposed method is flexible and shown to produce satisfactory performance even when the functional covariates are sparsely observed and are contaminated with measurement error. Comparisons with the variable selection method in FLCM are provided to illustrate potential gain in terms of selection accuracy and predictive performance using the variable selection method for NPFCM. We demonstrate the application of
the proposed method on a dietary calcium absorption study data (Davis, 2002) and bikeshare data (Fanaee-T & Gama, 2014) for modelling the number of daily bike rentals in Washington, D.C., while taking into account the diurnal effect of continuous meteorological covariates, e.g., temperature, humidity and wind speed. The proposed variable selection method is employed to identify the relevant meteorological factors while simultaneously estimating their nonlinear dynamic effects.

The rest of this article is organized in the following way. We present our modelling framework, set up the nonparametric functional concurrent regression model and illustrate the proposed variable selection method in Section 2. In Section 3, we present simulation studies to evaluate the empirical performance of the proposed method. In Section 4, the real data applications are presented. In Section 5, we deliberate over the contribution and some technical aspects of our method and also propose some future extensions based on the current research.

2. METHODOLOGY

2.1. Modelling Framework

We assume that the observed data for the ith subject is \( \{Y_i(t), X_{i1}(t), X_{i2}(t), \ldots, X_{ip}(t)\} \) \( i = 1, 2, \ldots, n \), where \( Y(\cdot) \) represents the functional response and \( X_{ij}(\cdot) \) are the corresponding functional covariates. In practice, we observe both the response and the predictors only on finitely many time points over some closed and bounded interval. In developing our method, we assume that the covariates and the response are observed on a dense and regular grid of points \( S_0 = \{t_1, t_2, \ldots, t_m\} \subset S = [a, b] \) for some \( a, b \in \mathbb{R} \) and that the covariates are observed without any measurement error. We later illustrate how the proposed method can be easily extended to accommodate more general scenarios where the covariates are contaminated with measurement error and observed on an irregularly spaced grid. We consider the following nonparametric functional concurrent regression model:

\[
Y_i(t) = \mu_Y(t) + \sum_{j=1}^{p} F_j \{X_{ij}(t), t\} + \epsilon_i(t), \quad \text{for } i = 1, 2, \ldots, n.
\]

Here, \( \mu_Y(t) \) is the functional intercept, and \( F_j \{X_{ij}(t), t\} \) are smooth bivariate functions on \( \mathbb{R} \times S \) capturing the concurrent effect of covariate \( X_{ij}(\cdot) \). Estimation and prediction in the NPF CM has been studied by Kim, Maity & Staicu (2018a). If \( E[F_j \{X_{ij}(t), t\}] = 0 \), the functions \( F_j(\cdot, \cdot) \) are identifiable and \( \mu_Y(\cdot) \) represents the marginal mean of the response curves. The covariates \( X_{ij}(\cdot) \) are assumed to be independent and identically distributed (i.i.d.) copies of \( X_{j}(\cdot) \) \( j = 1, 2, \ldots, p \), where \( X_{j}(\cdot) \) is a smooth stochastic process. We further assume that \( \epsilon_i(\cdot) \) are i.i.d. copies of \( \epsilon(\cdot), \) which is a mean-zero stochastic process with unknown and nontrivial covariance structure. We model the intercept function \( \mu_Y(t) \) using the truncated univariate basis expansion \( \mu_Y(t) = \sum_{l=1}^{p_0} \theta_{\mu,l} B_{\mu,l}(t) = B_{\mu}(t)^T \theta_{\mu} \), where \( B_{\mu}(t) = [B_{\mu,1}(t), B_{\mu,2}(t), \ldots, B_{\mu,p_0}(t)]^T \) and \( \theta_{\mu} = (\theta_{\mu,1}, \theta_{\mu,2}, \ldots, \theta_{\mu,p_0})^T \) is a vector of unknown coefficients. We use cubic B-spline basis functions in this article; however, other basis functions can also be used. For modelling the bivariate functions \( F_j(\cdot, \cdot) \), we express them in terms of bivariate basis expansion using the tensor product of B-spline basis functions (McLean et al., 2014; Kim, Maity & Staicu, 2018a). Suppose for \( j = 1, \ldots, p \), \( \{B_{X_j,\ell}(x)\}_{\ell=1}^{K_j} \) and \( \{B_{T_j,\ell}(t)\}_{\ell=1}^{L_j} \) are sets of known B-spline basis functions over \( x \) (the image of the process \( X_j(\cdot) \)) and \( t \), respectively. Let us define \( B_{T_j,0}(t) = 1 \) for \( j = 1, \ldots, p \) and include them in the basis set over \( t \) as \( \{B_{T_j,\ell}(t)\}_{\ell=1}^{L_j} \). We model \( F_j(\cdot, \cdot) \) using a tensor product of these two sets of basis functions as

\[
F_j \{X_{ij}(t), t\} = \sum_{k=1}^{K_j} \sum_{\ell=0}^{L_j} \theta_{j,k,\ell} B_{X_{j,k}} \{X_{ij}(t)\} B_{T_j,\ell}(t) = Z_{ij}(t)^T \theta_j,
\]

where we denote the stacked
vector \( \{B_{X_j,k}(X_{ij}(t))B_{T_j,k}(t)\}_{k=1}^{K_j,L_j} \) of length \( K_j(L_j + 1) \) as \( Z_{ij}(t)^T \) and \( \theta_j \) is the vector of unknown basis coefficients \( \theta_{j,k} \), i.e., \( \theta_j = (\theta_{j,1,0}, \theta_{j,1,1}, \ldots, \theta_{j,K_j,L_j})^T \). For the purpose of variable selection, we do not include 1 as a basis in \( \{B_{X_j,k}(x)\}_{k=1}^{K_j} \). Hence, we are able to estimate the joint effect of \( x, t \) on response \( Y(t) \) adjusted for the main effect of \( t \), which gets included in the functional intercept \( \mu_Y(t) \). Using the basis expansions of \( \mu_Y(\cdot) \) and \( F_j(\cdot, \cdot) \), NPFCM (1) can be reformulated as

\[
Y_i(t) = B_{\mu}(t)^T \theta_{\mu} + \sum_{j=1}^{p} Z_{ij}(t)^T \theta_j + \epsilon_i(t). \tag{2}
\]

2.2. Variable Selection Method for NPFCM

Since we assume data are observed on a dense and regular grid of points \( S = \{t_1, t_2, \ldots, t_m\} \), the linear model corresponding to the functional model (2) can be expressed as

\[
Y_i = B_{\mu} \theta_{\mu} + \sum_{j=1}^{p} Z_{ij} \theta_j + \epsilon_i, \tag{3}
\]

where \( Y_i = (Y_i(t_1), Y_i(t_2), \ldots, Y_i(t_m))^T \), \( Z_{ij} = (Z_{ij}(t_1), Z_{ij}(t_2), \ldots, Z_{ij}(t_m))^T \) and \( \epsilon_i = (\epsilon_i(t_1), \epsilon_i(t_2), \ldots, \epsilon_i(t_m))^T \). Note that a functional predictor \( X_j(t) \) taking value \( x \) influences the response \( Y(t) \) unless \( F_j(x, t) = a_j(t) \ \forall x \), which, in terms of the basis representation used, is equivalent to the condition \( \theta_j = 0 \). Therefore, we propose to minimize the following penalized residual sum-of-squares criterion to perform variable selection:

\[
R(\theta) = \sum_{i=1}^{n} \left\| Y_i - B_{\mu} \theta_{\mu} - \sum_{j=1}^{p} Z_{ij} \theta_j \right\|_2^2 + \lambda mn \sum_{j=1}^{p} |\theta_j|^{1/2}
\]

\[
= \sum_{i=1}^{n} \left\| Y_i - B_{\mu} \theta_{\mu} - \sum_{j=1}^{p} Z_{ij} \theta_j \right\|_2^2 + \lambda mn \sum_{j=1}^{p} |\theta_j|_2
\]

\[
= \sum_{i=1}^{n} \left\| Y_i - B_{\mu} \theta_{\mu} - \sum_{j=1}^{p} Z_{ij} \theta_j \right\|_2^2 + mn \sum_{j=1}^{p} P_{\text{LASSO}, \lambda}(|\theta_j|_2), \tag{4}
\]

where \( \theta = (\theta_{\mu}^T, \theta_1^T, \ldots, \theta_p^T)^T \). This minimization problem is identical to performing a group LASSO (Yuan & Lin, 2006), where the grouping is introduced by the covariates \( Z_j \), and depending on whether \( \theta_j = 0 \), \( X_j(\cdot) \) will be either absent or present in the model. The tuning parameters for this minimization are the penalty parameter \( \lambda \) and the number of B-spline basis functions \( K_j \) and \( L_j \), which control the smoothness of the function \( F_j(\cdot, \cdot) \) in \( x \) and \( t \) directions, respectively. Instead of directly using a roughness penalty on the functions, we follow a truncated basis approach (Ramsay & Silverman, 2005; Fan, James & Radchenko, 2015) and restrict the number of basis functions to be small, which ensures that the resulting functions are smooth in both the variables. Next, we extend this group LASSO formulation to two non-convex penalties. It was illustrated in Ghosal et al. (2020) that group LASSO can have a very high FPR for variable selection in FLCM and produce biased estimates, which in turn can affect the prediction accuracy of the model. The non-convex penalties SCAD (Fan & Li, 2001) and MCP (Zhang, 2010), on the other hand, are known (Mazumder, Friedman & Hastie, 2011; Breheny & Huang, 2015) to produce sparser solutions in classical regression literature and also to overcome the bias problem of LASSO due to their relaxed rate of penalization. Unlike adaptive LASSO (Zou, 2006), these
methods do not require initial estimates of weights. As pointed out by Fan, Samworth & Wu (2009), one drawback of adaptive LASSO is that the final solution is always sparser than the initial estimate owing to zero being an absorbing state. On the contrary, SCAD and MCP do not face this problem.

Subsequently, we propose a group SCAD- and a group MCP-based criterion, extending the group LASSO formulation (4) for performing variable selection in the NPFCM. For the group SCAD method, we perform variable selection and obtain an estimate of the parameters and estimate \( \arg \min_{\boldsymbol{\theta}, \mu} \sum_{i=1}^{n} \left| Y_i - \mathbb{B}_\mu \theta_\mu - \sum_{j=1}^{p} Z_{ij} \theta_j \right|^2 + mn \sum_{j=1}^{p} P_{SCAD,\lambda,\phi}(||\theta_j||_2), \) (5)

where the SCAD penalty \( P_{SCAD,\lambda,\phi}(||\theta_j||_2) \) is defined as

\[
P_{SCAD,\lambda,\phi}(||\theta_j||_2) = \begin{cases} 
\lambda ||\theta_j||_2 & \text{if } ||\theta_j||_2 \leq \lambda. \\
\frac{\lambda \phi ||\theta_j||_2 - 0.5(\frac{||\theta_j||_2^2}{\phi} + \lambda^2)}{\phi - 1} & \text{if } \lambda < ||\theta_j||_2 \leq \lambda \phi. \\
0.5 \lambda^2 (\phi + 1) & \text{if } ||\theta_j||_2 > \lambda \phi.
\end{cases}
\]

Analogously, for the group MCP method, we use the group minimax concave penalty (MCP) on the parameters and estimate \( \arg \min_{\boldsymbol{\theta}, \mu} \sum_{i=1}^{n} \left| Y_i - \mathbb{B}_\mu \theta_\mu - \sum_{j=1}^{p} Z_{ij} \theta_j \right|^2 + mn \sum_{j=1}^{p} P_{MCP,\lambda,\phi}(||\theta_j||_2), \) (6)

where the penalty \( P_{MCP,\lambda,\phi}(||\theta_j||_2) \) is defined as

\[
P_{MCP,\lambda,\phi}(||\theta_j||_2) = \begin{cases} 
\lambda ||\theta_j||_2 - \frac{||\theta_j||_2^2}{2\phi} & \text{if } ||\theta_j||_2 \leq \lambda \phi. \\
0.5 \lambda^2 \phi & \text{if } ||\theta_j||_2 > \lambda \phi.
\end{cases}
\]

Remark 1. For the construction of \( B_{X_j}(x) \) (B-spline basis), we use equally spaced knots in the range of observed data \( \{X_{ij}(t)\}_{i=1}^{n} \). The tuning parameters \( K_j \) and \( L_j \) (the number of basis functions in each direction) and the penalty parameter \( \lambda \) are selected using the extended BIC (EBIC) (Chen & Chen, 2008) of a corresponding Gaussian likelihood. Although we make no distributional assumption, this has shown satisfactory performance, as illustrated in our empirical studies. Some other information criterion or data-driven method can also be used, depending on the problem. We use the same number of basis functions \( K_j = K \) for \( j = 1, \ldots, p \), which is a standard assumption for computational tractability and assumes that all the nonparametric functional effects have more or less the same level of smoothness in both the variables. For the tuning parameter \( \phi \), we use the values 4 for SCAD and 3 for MCP, as recommended by the original authors of these methods. We use the \texttt{grpreg} package (Breheny & Huang, 2015) in R (R Core Team, 2018) for implementation of the group SCAD and group MCP method.

2.3. Extension to Sparse and Noisy Data

As a general extension, we consider the case where data are observed on an irregular and sparse grid and covariates are possibly contaminated with measurement error. This situation arises frequently for longitudinal studies where there are few subject-specific visits. Here we have response \( \{Y_i(t_{ij}), t_{ij}, j = 1, 2, \ldots, m_i\} \), and the observed covariates are
given by \( \{ (U_1(t_{1ij}), t_{1ij}), j = 1, 2, \ldots, m_1 \}, \{ (U_2(t_{2ij}), t_{2ij}), j = 1, 2, \ldots, m_2 \}, \ldots, \{ (U_p(t_{p_{ij}}), t_{p_{ij}}), j = 1, 2, \ldots, m_p \} \). We denote \( U_k(t_{k_{ij}})s, (k = 1, 2, 3, \ldots, p) \) by \( U_{ijk} \), which represent the observed covariate with measurement error, i.e., we have \( U_{ijk} = X_k(t_{k_{ij}}) + e_{ijk} \) for \( i = 1, 2, \ldots, n, j = 1, 2, \ldots, m_k \) and \( k = 1, 2, \ldots, p \), where the measurement errors \( e_{ijk} \) are assumed to be white noises with zero mean and variance \( \sigma_k^2 \). We assume that, although the individual number of observations \( m_i \) is small, \( \bigcup_{i=1}^n \bigcup_{j=1}^{m_i} t_{ij} \) is dense in \([0, T]\) (Kim, Maity & Staicu, 2018a; Ghosal et al., 2020). Under such a set-up, the eigenvalues and eigenfunctions corresponding to the original curves can be estimated using functional principal component analysis (FPCA) (Yao, Müller & Wang, 2005). The scores can be estimated through conditional expectation, to the original curves can be estimated using functional principal component analysis (FPCA) (Ghosal et al., 2020). Under such a set-up, the eigenvalues and eigenfunctions corresponding to the observed curves \( \{ Y_i(t_{i_{ij}}) \} \) can be estimated using functional principal component analysis (FPCA) (Ghosal et al., 2020). Under such a set-up, the eigenvalues and eigenfunctions corresponding to the original curves can be estimated using functional principal component analysis (FPCA) (Ghosal et al., 2020). Under such a set-up, the eigenvalues and eigenfunctions corresponding to the original curves can be estimated using functional principal component analysis (FPCA) (Ghosal et al., 2020).

2.4. Centring and Scaling of Covariates

One of the problems that is likely to arise for irregularly observed sparse functional data is that some of the B-splines \( B_{X_{ij,k}}(x) \) might not have any observed data in their support, particularly if the image of the process \( X_j(t) \) is not dense over \( \mathbb{R} \). To overcome this limitation, we employ point-wise centring and scaling of functional covariates \( X_j(t) \), as recommended in Kim, Maity & Staicu (2018a). In particular, we transform the functional covariate \( X_j(t) \) as \( X^*_j(t) = \frac{X_j(t) - \mu_j(t)}{\sigma_j(t)} \), where \( \mu_j(t) \) and \( \sigma_j(t) \) are the mean and standard deviation of \( X_j(t) \). The transformed covariate can be interpreted as how much the amount of standard deviation \( X_j(t) \) is away from the mean \( \mu_j(t) \) at time \( t \). This normalization, first of all, brings all the covariates to a comparable scale. Second, and more importantly, since \( X^*_j(t) \) is bounded (at least stochastically), it facilitates construction of the B-splines \( B_{X^*_j,k}(x) \). In practice, we estimate \( \hat{\mu}_j(t) \) and \( \hat{\sigma}_j(t) \) from the FPCA step and use the transformed denoised covariates \( \hat{X}^*_j(t) = \left\{ \frac{\hat{X}_j(t) - \hat{\mu}_j(t)}{\hat{\sigma}_j(t)} \right\} \) for performing variable selection.

Our empirical analysis shows good performance in terms of selection accuracy and out-of-sample prediction performance using the above strategy; henceforth, throughout this article, we pre-process (denoise and standardize) the observed functional covariates before applying the proposed variable selection method. The overall procedure for performing variable selection in the NPFCM is presented as an algorithm in Algorithm 1.

**Algorithm 1** Variable selection method for NPFCM

1. Pre-process functional covariate \( X_{ij}(t) \) using FPCA and obtain the denoised and standardized predictor \( \hat{X}^*_j(t) = \frac{X^*_j(t) - \hat{\mu}_j(t)}{\hat{\sigma}_j(t)} \) (using “fpca.sc” function in **Refund** package in R).
2. Form basis matrix \( B_{\mu} \) and \( Z_{ij} \) as in Equation (3) with transformed covariates \( \hat{X}^*_j(t) \).
3. Use \( Y_i, B_{\mu} \) and \( Z_{ij} \) in the penalized selection criterion (4), (5), (6) and implement group LASSO, group SCAD, or group MCP using the “grpreg” function within **grpreg** package in R.
4. Use EBIC for choosing the tuning parameters \( K_j, L_j \), and the penalty parameter \( \lambda \).
3. SIMULATION STUDY

3.1. Simulation Setup

In this section, we investigate the performance of the proposed variable selection method using numerical simulations. We evaluate the proposed method in terms of selection accuracy and out-of-sample prediction performance. We also compare the proposed variable selection method for NPFCM with the variable selection method in FLCM (Ghosal et al., 2020). We consider the following simulation designs:

**Scenario A:** We generate data from the model

\[ Y_i(t) = \mu_Y(t) + \sum_{j=1}^{20} F_j(X_{ij}(t), t) + \epsilon_i(t), \quad \text{for } i = 1, \ldots, n, \quad t \in [0, 100], \quad (7) \]

where we have \( \mu_Y(t) = 1 + 2(t/100) + (t/100)^2 \), and the bivariate functions are given by \( F_1(x, t) = 2x(t/100), F_2(x, t) = 5x \sin(\pi t/100), F_3(x, t) = 8 \sin(xt/100) \) and \( F_j(x, t) = 0 \) for \( j = 4, 5, 6, \ldots, 20 \), i.e., the last 17 covariates do not have any effect on the response. The original covariates \( X_{ij}(-) \overset{i.i.d.}{\sim} X_j(\cdot) \), where \( X_j(t) \) (\( j = 1, 2, \ldots, 20 \)) are generated by \( X_j(t) = a_j \sqrt{2} \sin(\pi j t/400) + b_j \sqrt{2} \cos(\pi j t/400) \), where \( a_j \sim \mathcal{N}(0, (2)^2), b_j \sim \mathcal{N}(0, (2)^2) \). Moreover, following the discussion in Section 2.3, we assume that the covariates \( X_{ij}(t) \) are observed with measurement error, i.e., we observe \( U_{ij}(t) = X_{ij}(t) + \delta_j \), where \( \delta_j \sim \mathcal{N}(0, 0.6^2) \). The error process \( \epsilon_i(t) \) is generated as \( \epsilon_i(t) = \xi_{i1} \cos(t) + \xi_{i2} \sin(t) + N(0, 1^2 I_{m_i}) \), where \( \xi_{i1} \overset{i.i.d.}{\sim} \mathcal{N}(0, 0.5^2) \) and \( \xi_{i2} \overset{i.i.d.}{\sim} \mathcal{N}(0, 0.75^2) \). The response \( Y_i(t) \) and the noisy covariates \( U_{ij}(t) \) are observed sparsely in random \( m_i \) points in \( S \), where \( S \) is the set of \( m = 81 \) equidistant time points in \( [0, 100] \) and \( m_i \overset{i.i.d.}{\sim} \text{Uniform} \{30, 31, \ldots, 41\} \). Three sets of sample size \( n \in \{200, 400, 600\} \) are considered. For each set, 80\% of data is used for model fitting and the rest is kept aside for validation.

**Scenario B:** We follow a similar data generation set-up as in (7) with \( \mu_Y(t) = 1 + 2(t/100) + (t/100)^2 \), \( F_1(x, t) = 2x^3 \exp(t/100), F_2(x, t) = 5x^2 \sin(\pi t/100), F_3(x, t) = 16 \sin(xt/100) \) and \( F_j(x, t) = 0 \) for \( j = 4, 5, 6, \ldots, 20 \). In this scenario, the concurrent effects of the functional predictors are highly nonlinear unlike in scenario A, and this scenario is specifically used to compare the performance of the proposed variable selection method for NPFCM with that of the variable selection method in FLCM (Ghosal et al., 2020). We consider sample size \( n = 200 \) for this simulation scenario.

**Scenario C:** We consider an additional high-dimensional scenario with \( p = 50 \) functional predictors and sample size \( n = 40 \). We set \( \mu_Y(\cdot) \) and \( F_1(\cdot, \cdot), F_2(\cdot, \cdot) \) and \( F_3(\cdot, \cdot) \) as in scenario A, and \( F_j(X, t) = 0 \) for \( j = 4, 5, 6, \ldots, 50 \). The rest of the simulation design is kept exactly the same as in scenario A.

For all the simulation scenarios, we generate 500 replicated datasets from the specified models to assess the proposed variable selection method’s performance.

3.2. Simulation Results

**Scenario A:** We first assess the selection accuracy of the proposed variable selection method. As the covariates are observed sparsely and with measurement error, we apply the FPCA method (PVE = 99\%) and pre-process the curves as discussed in Section 2.3 to obtain denoised and standardized predictors \( \tilde{X}_{ij}^*(t) \) before applying our variable selection method. We then use the proposed variable selection method for NPFCM and note which predictors are selected in the model for each of the three penalized selection criteria proposed in Section 2.2. We repeat this across 500 Monte Carlo replications and obtain the selection percentage of each predictor,
reflecting the true-positive rate (TPR) and FPR of the proposed variable selection method. The selection percentages are reported in Table 1. We observe that all three penalization methods (group LASSO, group SCAD and group MCP) pick out the nonparametric concurrent effect of \( X_1(\cdot), X_2(\cdot), X_3(\cdot) \) in 100\% of the replications, illustrating satisfactory TPR of the proposed method. The group LASSO method exhibits a small false selection percentage around 0.2\% for some of the variables. The group SCAD and the group MCP method, on the other hand, show zero FPRs across all the sample sizes in our simulation for all the predictors. In other words, they are able to identify the true model in 100\% of the replications. The average model sizes for each scenario are also provided in Table 1. The group LASSO method has an average model size in the range of 3.002–3.006 for the three different sample sizes. The group SCAD and the group MCP method produce an average model size 3, successfully identifying the true model each time.

Next, we compare the proposed selection methods (with different penalties) in terms of their out-of-sample prediction performance. For this purpose, we define a criterion similar to \( R^2 \)-squared (\( R^2 \)), based on the sum-of-squared deviations of the predicted curves from the actual curves. For normalization purpose, we use the sum-of-squared deviations of the actual curve from their scalar mean \( \bar{Y} \). The measure is defined as

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} \sum_{j=1}^{m_i} (Y_i(t_{ij}) - \hat{Y}_i(t_{ij}))^2}{\sum_{i=1}^{n} \sum_{j=1}^{m_i} (Y_i(t_{ij}) - \bar{Y})^2}.
\]

We report the quartiles and 10th and 90th percentiles of the out-of-sample \( R^2 \) criterion based on all the Monte Carlo replications in Table 2. We notice that for all three penalization methods, the median of the out-of-sample \( R^2 \) is around 96\%–97\% and increasing with the training set sample size. This indicates a satisfactory out-of-sample prediction performance of the proposed variable selection method. The empirical results, in particular, illustrate that the proposed group SCAD- and group MCP-based selection method for the NPFCM can not only identify the underlying model successfully but also provide high out-of-sample prediction accuracy.

**Scenario B:** In this scenario, the primary goal is to illustrate the benefit of the proposed variable selection method for NPFCM by comparing it with the variable selection method for FLCM (Ghosal et al., 2020). In particular, we want to demonstrate what can be gained using the proposed method in terms of selection accuracy and prediction performance in the presence of nonlinear effects of the predictors. As in scenario A, we first obtain the denoised curves \( \hat{X}_j(t) \) before applying our variable selection method. All the predictors in this scenario have a nonlinear effect on the response. We obtain the selection percentage of each variable for each of the three penalized selection criteria discussed in this article and compare them with the same for the variable selection method in FLCM. The results are displayed in Table 3. We observe that for the variable selection method in FLCM, the effect of \( X_2(\cdot) \) (\( X_3(\cdot) \)) is captured in 69\% (73.2\%) of the replications by the group LASSO penalty, in 58.4\% (64.6\%) of the replications by the group SCAD penalty, and in 50.8\% (60\%) of the replications by the group MCP. The results illustrate quite a large (30\%–50\%) FNR for the variable selection method in FLCM. The average FPR is also considerably high (29\%, 21\%, 16\%) using the variable selection method for FLCM.

In comparison, the variable selection method for NPFCM captures the nonlinear effect of \( X_2(\cdot) \) in 100\% of the replications for all the penalties. The effect of \( X_3(\cdot) \) is also captured with a higher TPR (87\%–99\%). The proposed group SCAD- and group MCP-based selection methods for the NPFCM also exhibit a negligible FPR (<1\%). The results illustrate that the proposed variable selection method for NPFCM is more suitable in the presence of nonlinear effects of the functional predictors, as the FNR remains minimal along with the FPR.

We provide a similar comparison as in scenario A, in terms of out-of-sample prediction performance of the selection methods. FLCM intuitively should be able to capture only the linear
Table 1: Comparison of selection percentages (%) of variables and average model size across 500 Monte Carlo replications: scenario A.

| Sample size | Method  | Var 1 | Var 2 | Var 3 | Var 4 | Var 5 | Var 6 | Var 7 | Var 8 | Var 9 | Var 10 | Var 11 | Var 12 | Var 13 | Var 14 | Var 15 | Var 16 | Var 17 | Var 18 | Var 19 | Var 20 | Avg. model size |
|-------------|---------|-------|-------|-------|-------|-------|-------|-------|-------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|----------------|
| n = 160     | LASSO   | 100   | 100   | 100   | 0.2   | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 3.002 |
|             | SCAD    | 100   | 100   | 100   | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 3     |
|             | MCP     | 100   | 100   | 100   | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 3     |
| n = 320     | LASSO   | 100   | 100   | 100   | 0     | 0     | 0     | 0     | 0.2   | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 3.002 |
|             | SCAD    | 100   | 100   | 100   | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 3     |
|             | MCP     | 100   | 100   | 100   | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 3     |
| n = 480     | LASSO   | 100   | 100   | 100   | 0.2   | 0.2   | 0.2   | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 3.006 |
|             | SCAD    | 100   | 100   | 100   | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 3     |
|             | MCP     | 100   | 100   | 100   | 0     | 0     | 0     | 0     | 0     | 0     | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 0      | 3     |
TABLE 2: Quantiles of out-of-sample $R^2$ based on MC simulation: scenario A.

| Sample size (train) | $R^2$(test) method | 10%  | 25%  | 50%  | 75%  | 90%  |
|---------------------|--------------------|------|------|------|------|------|
| $n = 160$           | LASSO              | 0.797| 0.937| 0.963| 0.972| 0.977|
|                     | SCAD               | 0.769| 0.937| 0.965| 0.974| 0.979|
|                     | MCP                | 0.769| 0.937| 0.965| 0.974| 0.979|
| $n = 320$           | LASSO              | 0.859| 0.944| 0.964| 0.973| 0.977|
|                     | SCAD               | 0.842| 0.943| 0.966| 0.974| 0.978|
|                     | MCP                | 0.838| 0.943| 0.966| 0.974| 0.978|
| $n = 480$           | LASSO              | 0.885| 0.945| 0.963| 0.971| 0.976|
|                     | SCAD               | 0.876| 0.945| 0.964| 0.972| 0.977|
|                     | MCP                | 0.876| 0.945| 0.964| 0.972| 0.977|

TABLE 3: Comparison of selection percentages (%) of variables, scenario B.

| Method   | Penalty | Var 1 | Var 2 | Var 3 | Mean (Var 4–Var 20) | Avg. model size |
|----------|---------|-------|-------|-------|---------------------|-----------------|
| FLCM     | LASSO   | 100   | 69    | 73.2  | 29.46               | 7.43            |
|          | SCAD    | 100   | 58.4  | 64.6  | 21.07               | 5.812           |
|          | MCP     | 100   | 50.8  | 60    | 15.66               | 4.77            |
| NPFCM    | LASSO   | 100   | 100   | 98.8  | 2.59                | 3.428           |
|          | SCAD    | 100   | 100   | 87.8  | 0.962               | 3.032           |
|          | MCP     | 100   | 100   | 94.4  | 0.847               | 3.088           |

TABLE 4: Quantiles of out-of-sample $R^2$ based on Monte Carlo simulation, scenario B.

| Method   | $R^2$(test) penalty | 10%  | 25%  | 50%  | 75%  | 90%  |
|----------|---------------------|------|------|------|------|------|
| FLCM     | LASSO               | 0.396| 0.515| 0.598| 0.665| 0.708|
|          | SCAD                | 0.383| 0.515| 0.600| 0.663| 0.712|
|          | MCP                 | 0.383| 0.515| 0.600| 0.662| 0.712|
| NPFCM    | LASSO               | 0.974| 0.983| 0.988| 0.991| 0.994|
|          | SCAD                | 0.972| 0.982| 0.987| 0.991| 0.994|
|          | MCP                 | 0.971| 0.982| 0.987| 0.991| 0.994|

component of the concurrent effect and thus may not be suitable for prediction in the presence of nonlinear effects. The results from both the methods are displayed in Table 4. We notice that the median out-of-sample $R^2$ is approximately 60% for the variable selection method in FLCM, whereas this is around 99% for the variable selection method in NPFCM. The variable selection
method in NPFCM produces a superior performance on account of successfully identifying the nonlinear effects of the functional predictors.

**Scenario C:** This scenario illustrates the performance of the proposed variable selection method in the high-dimensional case of \( p > n \). The TPR over the 500 replicated datasets for all three penalization methods is found to be 100%. The average FPR for the group LASSO, the group SCAD, and the group MCP method are found to be 0.39\%, 0\% and 0.008\%, respectively. The results illustrate the proposed variable selection method’s satisfactory performance even in the high-dimensional scenario of \( p > n \).

The simulation results in this section illustrate that the proposed variable selection method for NPFCM can identify the underlying functional predictors having nonlinear concurrent effects with high accuracy and also gives satisfactory out-of-sample prediction performance. The group SCAD- and group MCP-based selection methods, in particular, are shown to have minuscule FPR and FNR, and therefore are the recommended methods for variable selection in NPFCM in this article.

### 4. REAL DATA APPLICATIONS

In this section, we demonstrate the application of the proposed variable selection method in NPFCM for the selection of the influential functional covariates in two real data studies. We first consider a dietary calcium absorption study, with added pseudo-variables as an illustration of the proposed method. Pseudo-variables have been used before (Miller, 2002; Wu, Boos & Stefanski, 2007) to assess the false selection rate in real datasets and therefore to tune variable selection procedures. Next, we apply the proposed method for identifying the influential meteorological variables in the capital bikeshare data and estimate their concurrent effects on the number of casual bike rentals.

#### 4.1. Study of Dietary Calcium Absorption

We consider the study of dietary calcium absorption given in Davis (2002). In this article, we have data on calcium absorption \( Y(t) \), dietary calcium intake \( X_1(t) \), body mass index (BMI) \( X_2(t) \) and body surface area (BSA) \( X_3(t) \) of 188 patients at irregular time points between 35 and 64 years of age \( (t) \). For each patient, the number of repeated measurements varies between 1 and 4. Figure 1 displays the individual curves of the patients’ calcium absorption, calcium intake, BSA and BMI as a function of their ages. We are primarily interested in finding out which of the three predictors influence calcium absorption \( Y(t) \). We intuitively expect calcium intake to be associated with calcium absorption based on prior literature (Kim, Maity & Staicu, 2018a; Ghosal et al., 2020). As the data are sparse and the original covariates might be observed with measurement error, we apply FPCA-based pre-processing (PVE = 95\%) as discussed in Section 2.3 and get the denoised and standardized trajectories \( \hat{X}_j^*(t) \) for \( j = 1, 2, 3 \), before applying our variable selection method.

To further illustrate the selection performance and FPR of the proposed variable selection method, we add 15 functional pseudo-variables by simulating \( X_{ij}(\cdot) \overset{i.i.d.}{\sim} X_j(\cdot) \), where \( X_j(t) \) \((j = 4, 5, \ldots, 18)\) are given by \( X_j(t) = a_j \sqrt{2} \sin(\pi(j - 3)/200) + b_j \sqrt{2} \cos(\pi(j - 3)/200) \), where \( a_j \overset{i.i.d.}{\sim} \mathcal{N}(0,(2)^2) \), \( b_j \overset{i.i.d.}{\sim} \mathcal{N}(0,(2)^2) \). Finally, we apply the variable selection method developed in this article for NPFCM to the functional response \( Y(t) \) and the covariates \( \hat{X}_1^*(t), \hat{X}_2^*(t), \hat{X}_3^*(t), \hat{X}_4^*(t), \hat{X}_5^*(t), \ldots, \hat{X}_{18}^*(t) \) (the pseudo-variables are standardized). We repeat the whole procedure a large number of times \((B = 100)\). Table 5 displays the selection percentage of each of the variables. We notice that both the group SCAD- and group MCP-based methods identify calcium intake \( X_1(t) \) as a significant predictor in 100\% of the iterations. All other variables including the randomly generated predictors are completely ignored in all the iterations. So the proposed variable selection method for NPFCM does a good job in identifying the true predictor set and discarding the pseudo-variables. As both the
penalization methods select calcium intake, we want to estimate its concurrent effect. For this purpose, we use the proposed method for performing variable selection in NPFCM on the original data without adding any pseudo-variables. The estimated nonparametric effect of calcium intake (adjusted for the main effect of $t$) is displayed in Figure 2. The estimated linear effect of calcium intake, obtained using the variable selection method for FLCM (Ghosal et al., 2020), is also provided for comparison. We notice that the estimated nonparametric effect of calcium intake (standardized) is more or less linear, which matches with prior findings of Kim, Maity & Staicu (2018a). For fixed and moderate values of $t$ (age), as $x$ (calcium intake) increases (or is higher than the mean at time $t$), the functional effect

**FIGURE 1:** Calcium absorption and covariate profiles of patients by age. Profiles of three randomly chosen subjects are illustrated with the solid, dashed and dotted lines.

**TABLE 5:** Selection percentages (%) of variables in the calcium absorption study.

| Method | Var 1 (calcium intake) | Var 2 (BSA) | Var 3 (BMI) | Max. Var (4–18) |
|--------|------------------------|-------------|-------------|-----------------|
| SCAD   | 100                    | 0           | 0           | 0               |
| MCP    | 100                    | 0           | 0           | 0               |

DOI: 10.1002/cjs.11654

The Canadian Journal of Statistics / La revue canadienne de statistique
FIGURE 2: Estimated nonparametric effect of calcium intake on calcium absorption from the proposed variable selection method in NPFCM (left panel) and linear concurrent effect estimated using FLCM (right panel).

$F(x, t)$ is decreasing in most of the regions, indicating that, as calcium intake increases, dietary calcium absorption should decrease at most ages. A similar observation was made in Ghosal et al. (2020), although one needs to be careful in interpreting the results near boundaries due to lack of observations. Our analysis using the variable selection method for NPFCM provides a more comprehensive understanding of the underlying association between calcium absorption and dietary calcium intake.

Remark 2. To investigate what happens in the presence of a significant correlation between the truly relevant predictors and the pseudo-variables, we considered an alternative scenario where the pseudo-variables $Z_j(t)$ ($j = 4, 5, \ldots, 18$) were generated conditionally on $Z_1(t)$ (calcium intake) from a normal distribution with mean $\rho(Z_1(t) - a)$ and variance $(1 - \rho^2)$. We set $\rho = 0.5$ in our analysis to introduce a moderate correlation among the variables. The constants $a, b$ were taken to be the scalar mean and standard deviation of the samples $\{Z_{i1}(t_{ij}), j = 1, 2, \ldots, m_i\}_{i=1}^n$. We applied the proposed variable selection method for NPFCM and obtained identical results as reported in Table 5 in terms of variable selection. This result demonstrates that the proposed variable selection method can distinguish between the truly influential predictors and irrelevant ones with high accuracy even when there is a significant correlation among them.

4.2. Study of Bike-Sharing Data

We consider application of the proposed variable selection method on the Capital bikeshare study in Fanaee-T & Gama (2014). The data were initially obtained from the Capital bikeshare system in Washington, D.C., and contain information on casual bike rentals along with other meteorological variables such as temperature (temp), feels-like temperature (atemp), humidity (hum) and wind speed (windspeed) on an hourly basis. Casual bike rentals are defined as rentals to cyclists without membership in the Capital bikeshare program. Since bike rentals can have different dynamics on weekends compared to weekdays, following the analysis of Kim et al. (2018b), we restrict our attention to casual bike rentals on Saturdays. The counts of casual bike rentals (hourly) are available during the period from 1 January 2011 to 31 December 2012, on a total of 105 Saturdays barring some exceptions (8 missing). Figure 3 displays the diurnal curves of casual bike rentals and other meteorological variables. It is plausible that the meteorological variables have real-time dynamic effects on the number of bike rentals,
and we are primarily interested in selecting the influential ones as well as estimating their concurrent effects. The hourly bike rentals $Y(t)$ are discrete and their distribution is skewed, therefore we apply a log transformation prior to our analysis ($Y(\cdot) = \log\{Y(\cdot) + 1\}$). We add 16 functional pseudo-variables to the existing 4 meteorological variables by generating from the covariate model $X_j(t) = a_j \sqrt{2} \sin(2\pi(j - 4)t) + b_j \sqrt{2} \cos(2\pi(j - 4)t)$, where $a_j \sim \mathcal{N}(0, (2)^2)$, $b_j \sim \mathcal{N}(0, (2)^2)$, for $j = 5, 6, \ldots, 20$. This is again done in order to assess the false selection performance of the proposed method. We pre-process the meteorological variables $X_1(t), X_2(t), X_3(t), X_4(t)$ using FPCA methods as discussed in Section 2.3, and then apply the proposed variable selection method to $Y(t)$ and $\hat{X}_1^*(t), \hat{X}_2^*(t), \hat{X}_3^*(t), \hat{X}_4^*(t), \ldots, \hat{X}_{20}^*(t)$ (the pseudo-variables are standardized). The whole procedure is repeated several times ($B = 100$). Table 6 displays the selection percentage of each of the predictors for the group SCAD- and group MCP-based methods for NPFCM. Also reported, for comparison, are the selection percentages of the FMCP method (Ghosal et al., 2020) with and without pre-whitening (PW and NPW, respectively).

We observe that the group SCAD method selects feels-like temperature $X_2(t)$, humidity $X_3(t)$ and wind speed $X_4(t)$ 100% of the time, whereas the group MCP method selects temperature $X_1(t)$, humidity $X_3(t)$ and wind speed $X_4(t)$ 100% of the time. The covariates temperature $X_1(t)$ and feels-like temperature $X_2(t)$ should be highly correlated, and it is ideal that the proposed variable selection method is picking only one of them consistently. In terms of the FPR, we see that both group SCAD and group MCP methods discard the pseudo-variables in all the iterations.
TABLE 6: Comparison of selection percentages (%) of variables in the bike-sharing study.

| Method          | \( X_1 \) (temp) | \( X_2 \) (atemp) | \( X_3 \) (hum) | \( X_4 \) (windspeed) | Mean \( X_5, \ldots, X_{20} \) |
|-----------------|------------------|-------------------|----------------|------------------------|-----------------------------|
| SCAD (NPFCM)    | 0                | 100               | 100            | 100                    | 0                           |
| MCP (NPFCM)     | 100              | 0                 | 100            | 100                    | 0                           |
| FMCP (PW)       | 0                | 100               | 0              | 0                      | 4.85                        |
| FMCP (NPW)      | 100              | 100               | 100            | 100                    | 4.85                        |

For the FMCP method, which uses an FLCM, we observe that feels-like temperature (atemp) is selected using pre-whitening, and all four predictors are selected without pre-whitening. In this case, the average number of false positives is calculated to be 4.85. In the presence of nonlinear effects, pre-whitening may not be a good idea, as the FLCM used to estimate the covariance matrix becomes mis-specified. Hence, both results (with and without pre-whitening) are presented to facilitate comparison. Nevertheless, the proposed variable selection method for NPFCM performs better than FMCP irrespective of pre-whitening in terms of selection accuracy.

Next, we estimate the concurrent effects of the meteorological factors on hourly bike rentals by applying the proposed group MCP method for NPFCM on the actual covariates (no pseudo-variables added). The estimated concurrent effects of the selected meteorological variables (in standardized scale) are displayed in Figure 4. The estimated effects from the group SCAD method are similar and therefore not reported. We observe that the effect of temperature on bike rentals is maximum when the temperature is in the mid-range (near the mean) and during midday, as this should be an ideal condition for biking. For too cold or too warm temperatures, the effect on the number of bike rentals decreases, which is natural. For mid-range to higher temperatures, the impact of temperature on casual bike rentals is more or less uniform throughout the day, which is expected on weekends and suitable weather conditions in Washington, D.C.

There is a negative association for each fixed time point for humidity, as its effect on the number of bike rentals decreases with increasing humidity. A similar observation was made in Jim et al. (2018b). For lower humidity (lower than the mean at a particular time), the effect increases during midday and is highly positive. For higher humidity (higher than the mean), we see a drop in the effect, indicating that on humid days there should be a fall in bike rentals, which is again what one would naturally expect in these weather conditions. For wind speed, the effect is more or less uniform across the domain, except in high wind speed regions where bikers might face additional difficulties. In such high wind speed regions, during midday (plausibly the hottest time of the day), there is a noticeable dip, which can be attributed to the difficulty one might face while cycling amidst high-velocity winds. Our findings match with those of Gebhart & Noland (2014), who made similar observations using a multiple linear regression model on the Capital bikeshare data. Our analysis using NPFCM provides a more comprehensive understanding of the effect of meteorological factors on daily bike rentals.

It is evident from Figure 4 that the concurrent effects of the meteorological predictors are nonlinear. To illustrate the efficiency gain in predictive performance using the variable selection method for NPFCM, we compare it with the FMCP method (Ghosal et al., 2020) for FLCM in terms of their out-of-sample prediction performance. For the FMCP method, we do not use pre-whitening. In the presence of nonlinear effects, the estimate of the covariance matrix obtained using the FLCM can be highly biased and lead to inferior performance. As a comparison metric, we use the out-of-sample mean-square prediction error of both methods.
FIGURE 4: Estimated concurrent effect of the meteorological variables. Displayed are the effects of temperature (top left), humidity (top right) and wind speed (bottom row).

on test data, which is defined as $\text{MSPE} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} (Y_i(t_{ij}) - \hat{Y}_i(t_{ij}))^2$. We perform repeated training (80%) and testing (20%) using both methods for a large number of iterations ($B = 100$). Boxplots illustrating the distribution of out-of-sample prediction error corresponding to both methods are displayed in Figure 5. We notice that the variable selection method for NPFCM with MCP produces much smaller out-of-sample MSPE than the FMCP method, demonstrating the advantage of the proposed method, particularly in the presence of nonlinear effects of functional predictors.

Our data analysis in this section illustrates that, with the variable selection method for NPFCM, it is possible to identify the underlying functional predictors driving the functional response more accurately and capture their concurrent nonlinear effects, providing a better understanding of the dynamics between them.

Remark 3. The pre-whitened FMCP, as reported in Table 6, was unable to capture the concurrent effect of humidity and wind speed and produced inferior prediction performance compared to the non-pre-whitened version of the variable selection method in FLMC. Hence, for comparison of out-of-sample prediction performance, the non-pre-whitened version of FMCP was used.
5. DISCUSSION

In this article, we proposed a variable selection method in the nonparametric functional concurrent regression, extending the classically used penalized variable selection methods like LASSO, SCAD and MCP. The proposed method provides a unified framework for variable selection in additive function-on-function concurrent regression models. We have shown through numerical simulations that the proposed variable selection method for NPFCM, particularly with the group SCAD penalty or the group MCP, identifies the true underlying variables with minimal FPR and minuscule FNR while also providing satisfactory out-of-sample prediction accuracy. In the presence of nonlinear concurrent effects of the functional predictors, the proposed variable selection method for NPFCM with non-convex (SCAD or MCP) penalties was illustrated to be more accurate compared to the variable selection method for FLCM and therefore is the recommended method for variable selection in this article. In classical penalized regression, the non-convex penalties (SCAD and MCP) are known to have the oracle property (Fan & Li, 2001; Zhang, 2010) under standard regularity conditions. The linear model representation of NPFCM in Equation (3) could be used to establish the oracle properties of the regularized estimators in NPFCM, under the assumption that the actual nonparametric functions $F_j(\cdot, \cdot)$ are in the space spanned by the tensor product splines. This could be done in the same way as in Chen, Goldsmith & Ogden (2016), where the oracle properties of the group MCP regularized estimate were established in the context of function-on-scalar regression.

The proposed method can also be readily applied for variable selection in function-on-scalar regression, where the covariates are not time-varying. So far, almost the entire existing literature for variable selection in function-on-scalar regression (Chen, Goldsmith & Ogden, 2016; Fan & Reimherr, 2017; Parodi & Reimherr, 2018) has focused on the linear effect of the covariates. The proposed method can be used to identify the nonlinear effects of the predictors. We have demonstrated our method on a dietary calcium absorption study and bike-sharing data to select the influential functional covariates and estimate their underlying concurrent effects.
One of the proposed method’s limitations is forecasting when a new functional predictor is observed outside the range of the training data. The B-splines (De Boor, 1978) used for modelling in the training data will not be defined on this new observation. Note that this aspect of prediction is not unique to our proposed variable selection method or NPFCM. In general, for any spline-based method for nonparametric regression, extrapolation beyond the range of observed data is troublesome and has to be done with additional care (e.g., Eilers & Marx, 2010). We have followed a centring and scaling approach as suggested in Kim, Maity & Staicu (2018a) for NPFCM in this article. Alternatively, based on the suggestions for the FGAM (McLean et al., 2014), the transformations \( \hat{G}_t(x) = n^{-1} \sum_{i=1}^{n} I(X_i(t) < x) \) and \( \hat{H}_t(x) = n^{-1} \sum_{i=1}^{n} \Phi \left( \frac{x - X_i(t)}{h_t} \right) \) (where \( \Phi \) is the cdf of the standard normal and \( h_t \) is a user-chosen bandwidth) can be used on \( X(t) \) so that the transformed predictors cover \([0, 1]\) uniformly. Another working strategy, often employed by practitioners, can be to include additional “ghost” knots beyond the range of the observed data with the aim of extrapolating (Eilers & Marx, 2010; Kosinka, Sabin & Dodgson, 2014; Corlay, 2016) or supplementing the basis with functions of infinite support following the proposal in Schumaker (2007).

We have used an independent working correlation structure in our proposed variable selection method for NPFCM, which has shown satisfactory performance in our empirical analysis in terms of selection accuracy, which is the primary objective of this article. It is plausible to consider a pre-whitening method to take into account temporal dependence within functions. This can be done similarly to Chen, Goldsmith & Ogden (2016) and Ghosal et al. (2020). However, if the model is mis-specified, the use of pre-whitening can be counterproductive, as observed in our analysis. An additional challenge would be to handle the computational complexity arising from the initial pre-whitening step.

There are plenty of research problems that remain to be explored based on this work. It is of particular interest to study small-sample properties of the estimated nonparametric functions and get a measure of their uncertainty. The perturbation-based bootstrap (Das, Gregory & Lahiri, 2019) procedure for penalized regression methods would be interesting to explore in such situations. For distribution-free predictive inference, split conformal prediction bands (Lei et al., 2018) can also be explored by extending the conformal prediction approach to functional response (Lei, Rinaldo & Wasserman, 2015). The proposed method can be generalized to other function-on-function regression scenarios, e.g., the historical functional regression model or the additive function on function regression model (Kim et al., 2018b), where one might consider nonlinear lagged effects of functional predictors on the response at the current time point. Extending the proposed method to these general classes of functional regression models can have diverse applications and remain areas for future research.

ACKNOWLEDGEMENTS

We would like to thank the editor, the associate editor and two anonymous referees for their valuable input and suggestions, which have greatly helped to improve this article.

DATA AVAILABILITY STATEMENT

Illustrations of the implementation of our method using R, along with the datasets analyzed, are available on Github at https://github.com/rahulfrodo/NPFCM_Selection.

REFERENCES

Breheny, P. & Huang, J. (2015). Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statistics and Computing*, 25, 173–187.

Chen, J. & Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model spaces. *Biometrika*, 95, 759–771.

DOI: 10.1002/cjs.11654
Chen, Y., Goldsmith, J., & Ogden, R. T. (2016). Variable selection in function-on-scalar regression. *Stat*, 5, 88–101.

Corlay, S. (2016). B-spline techniques for volatility modeling. *Journal of Computational Finance*, 19, 97–135.

Das, D., Gregory, K., & Lahiri, S. (2019). Perturbation bootstrap in adaptive lasso. *The Annals of Statistics*, 47, 2080–2116.

Davis, C. S. (2002). *Statistical Methods for the Analysis of Repeated Measurements*. Springer Science & Business Media, New York.

De Boor, C. (1978). *A Practical Guide to Splines*, Vol. 27. Springer-Verlag, New York.

Eilers, P. H. & Marx, B. D. (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2, 637–653.

Fan, J. & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96, 1348–1360.

Fan, J., Samworth, R., & Wu, Y. (2009). Ultrahigh dimensional feature selection: Beyond the linear model. *Journal of Machine Learning Research*, 10, 2013–2038.

Fan, Y., James, G. M., & Radchenko, P. (2015). Functional additive regression. *The Annals of Statistics*, 43, 2296–2325.

Fan, Z. & Reimherr, M. (2017). High-dimensional adaptive function-on-scalar regression. *Econometrics and Statistics*, 1, 167–183.

Fanee-T, H. & Gama, J. (2014). Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence*, 2, 113–127.

Gebhart, K. & Noland, R. B. (2014). The impact of weather conditions on bikeshare trips in Washington, DC. *Transportation*, 41, 1205–1225.

Gertheiss, J., Maity, A., & Staicu, A.-M. (2013). Variable selection in generalized functional linear models. *Stat*, 2, 86–101.

Ghosal, R., Maity, A., Clark, T., & Longo, S. B. (2020). Variable selection in functional linear concurrent regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 69, 565–587.

Goldsmith, J. & Schwartz, J. E. (2017). Variable selection in the functional linear concurrent model. *Statistics in Medicine*, 36, 2237–2250.

Hastie, T. & Tibshirani, R. (1993). Varying-coefficient models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55, 757–796.

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

Kim, J., Maity, A., & Staicu, A.-M. (2018a). Additive nonlinear functional concurrent model. *Statistics and its Interface*, 11, 669–685.

Kim, J. S., Staicu, A.-M., Maity, A., Carroll, R. J., & Ruppert, D. (2018b). Additive function-on-function regression. *Journal of Computational and Graphical Statistics*, 27, 234–244.

Kosinka, J., Sabin, M., & Dodgson, N. (2014). Creases and boundary conditions for subdivision curves. *Graphical Models*, 76, 240–251.

Lei, J., G’Sell, M., Rinaldo, A., Tibshirani, R. J., & Wasserman, L. (2018). Distribution-free predictive inference for regression. *Journal of the American Statistical Association*, 113, 1094–1111.

Lei, J., Rinaldo, A., & Wasserman, L. (2015). A conformal prediction approach to explore functional data. *Annals of Mathematics and Artificial Intelligence*, 74, 29–43.

Maity, A. (2017). Nonparametric functional concurrent regression models. *Wiley Interdisciplinary Reviews: Computational Statistics*, 9, e1394.

Mazumder, R., Friedman, J. H., & Hastie, T. (2011). SparseNet: Coordinate descent with nonconvex penalties. *Journal of the American Statistical Association*, 106, 1125–1138.

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

Miller, A. (2002). *Subset Selection in Regression*. Chapman and Hall/CRC, New York.

Parodi, A. & Reimherr, M. (2018). Simultaneous variable selection and smoothing for high-dimensional function-on-scalar regression. *Electronic Journal of Statistics*, 12, 4602–4639.

R Core Team. (2018). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ramsay, J. & Silverman, B. (2005). *Functional Data Analysis*. Springer-Verlag, New York.
Scheipl, F., Gertheiss, J., & Greven, S. (2016). Generalized functional additive mixed models. *Electronic Journal of Statistics*, 10, 1455–1492.

Schumaker, L. (2007). *Spline Functions: Basic Theory*. Cambridge University Press, Cambridge.

Shi, J. Q., Murray-Smith, R., & Titterington, D. (2005). Hierarchical Gaussian process mixtures for regression. *Statistics and Computing*, 15, 31–41.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58, 267–288.

Wu, Y., Boos, D. D., & Stefanski, L. A. (2007). Controlling variable selection by the addition of pseudovariables. *Journal of the American Statistical Association*, 102, 235–243.

Yao, F., Müller, H.-G., & Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 68, 267–288.

Yuan, M. & Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68, 49–67.

Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38, 894–942.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, 101, 1418–1429.

Received 9 June 2020
Accepted 30 April 2021