Variable selection in functional linear concurrent regression

Rahul Ghosal, Arnab Maity, Timothy Clark and Stefano B. Longo
North Carolina State University, Raleigh, USA

[Received April 2019. Final revision February 2020]

Summary. We propose a novel method for variable selection in functional linear concurrent regression. Our research is motivated by a fisheries footprint study where the goal is to identify important time-varying sociostructural drivers influencing patterns of seafood consumption, and hence the fisheries footprint, over time, as well as estimating their dynamic effects. We develop a variable-selection method in functional linear concurrent regression extending the classically used scalar-on-scalar variable-selection methods like the lasso, smoothly clipped absolute deviation (SCAD) and minimax concave penalty (MCP). We show that in functional linear concurrent regression the variable-selection problem can be addressed as a group lasso, and their natural extension: the group SCAD or a group MCP problem. Through simulations, we illustrate that our method, particularly with the group SCAD or group MCP, can pick out the relevant variables with high accuracy and has minuscule false positive and false negative rate even when data are observed sparsely, are contaminated with noise and the error process is highly non-stationary. We also demonstrate two real data applications of our method in studies of dietary calcium absorption and fisheries footprint in the selection of influential time-varying covariates.

Keywords: Fisheries footprint; Functional linear concurrent regression; Variable selection

1. Introduction

Function-on-function regression is an active area of research in functional data with new statistical methods frequently emerging to address data where both the response variable and the covariates are functions over some continuous index such as time. The functional concurrent regression model is a special case of function-on-function regression where the predictor variables influence the response variable only through their value at the current time point (Kim et al., 2018). The commonly used functional linear concurrent regression model assumes a linear relationship between the response and the predictors, where the value of the response at a particular time point is modelled as a linear combination of the covariates at that specific time point, and the coefficients of the functional covariates are univariate smooth functions over time (Ramsay and Silverman, 2005). Multiple methods exist in the literature for estimation of these regression functions in functional linear concurrent regression and the closely related varying-coefficient model (Hastie and Tibshirani, 1993), using kernel–local polynomial smoothing (Wu et al., 1998; Hoover et al., 1998; Fan and Zhang, 1999; Kauermann and Tutz, 1999), polynomial splines (Huang et al., 2002, 2004) and smoothing splines (Hastie and Tibshirani, 1993; Hoover et al., 1998; Chiang et al., 2001; Eubank et al., 2004) among many others. Similarly to classical scalar regression, when there are a large number of covariates present, the primary
interest might be to select only the set of influential variables and to estimate their effects. Although doing significance testing and building confidence bands can help for assessing the individual effect of a predictor, they are computationally infeasible when the number of covariates is large. Thus arises the need to perform variable selection in functional linear concurrent regression.

Our research in this paper is motivated by a fisheries footprint study, where the fisheries footprint is defined as Global Footprint Network’s measure of the total marine area that is required to produce the amount of seafood products that a nation consumes. Our goal is to identify important time-varying sociostructural and economic drivers influencing fisheries footprints and to estimate their time-varying effects. Such work is essential to understand the interaction between marine and social systems (Longo and Clark, 2016). The primary dependent variable of interest in this study is the fisheries footprint. Fig. 1 displays the fisheries footprint (measured in hectares) of the nations over the study years on a logarithmic scale (natural logarithms). The fisheries footprint values of three representative countries (Afghanistan, Albania and the Bahamas) are displayed by using full, broken and dotted curves. Along with fisheries footprint data, we have data on 20 time-varying socio-economic drivers covering the broad spectrum of the economy, agriculture and population dynamics. More details on the variables and the study are provided in Section 4. Since we aim to capture the dynamic time-varying effect of the socio-economic drivers, the functional linear concurrent model presents itself as a natural choice.

Although various variable-selection methods have been developed for scalar-on-function regression (Gertheiss et al., 2013; Fan et al., 2015) and function-on-scalar regression (Chen et al., 2016), the literature for variable selection in functional linear concurrent regression is relatively sparse. Recently Goldsmith and Schwartz (2017) developed a variable-selection method for functional linear concurrent models by using a variational Bayes approach with sparsity being introduced through a spike-and-slab prior on the coefficients of the basis expansion of the regression functions. In this paper, we propose a variable-selection method in functional linear concurrent regression extending the classically used variable-selection methods like the lasso (Tibshirani, 1996), smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001) and the minimax concave penalty (MCP) (Zhang, 2010).

Our work is inspired by Gertheiss et al. (2013), who showed that the variable-selection problem in a scalar-on-function regression scenario can be reduced to a group lasso (Yuan and Lin, 2006) problem. We have shown in functional linear concurrent regression that the variable-selection problem can be addressed as a group lasso, and their natural extension group SCAD or group MCP problem. Chen et al. (2016) also used the group MCP for their variable selection in function-on-scalar regression. Our model is fundamentally different from theirs in the sense that the covariates that we consider are time-varying functions and possibly observed with measurement error. Our method is similar to Wang et al. (2008) in which they used a group SCAD penalty for variable selection in varying-coefficient models, but we propose a different penalty on the coefficient functions which simultaneously penalizes departure from sparsity as well as roughness of the coefficient functions, and our research shows that there is much to be gained by using the group MCP. We employ a prewhitening procedure similar to Chen et al. (2016) to take into account the possible temporal dependence within functions. We also consider that the covariates might be contaminated with measurement error and therefore we use functional principal component analysis (FPCA) to obtain denoised trajectories of the covariates, which improves the estimation accuracy of our approach. Through simulations, we illustrate that the method proposed, particularly with the group SCAD and group MCP, can pick out the relevant variables with very high accuracy and has minuscule false positive and false
negative rates even when data are observed sparsely and are contaminated with measurement error. We demonstrate two real data applications of our method in a study of dietary calcium absorption (Davis, 2002) and the fisheries footprint study in selection of the influential time-varying covariates.

The rest of the paper is organized as follows. In Section 2, we present our modelling framework and illustrate our variable-selection method. In Section 3, we conduct a simulation study to evaluate the performance of our method and summarize the simulation results. In Section 4, we go back to the two real data examples, the calcium absorption study and fisheries footprint study, apply our variable-selection method to find out the influential covariates and present our findings. We conclude in Section 5 with a discussion about some limitations and possible extensions of our work.
functions; however, other basis functions can be used as well. Then the minimization in the penalized residual sum of squares, \( Y_t = \sum_{j=1}^{p} Z_{ij}(t) \beta_j(t) + \epsilon_i(t), \) (1) where \( \beta_j(t) \) (\( j = 1, 2, \ldots, p \)) are smooth functions (with finite second derivative) representing the functional regression parameters. We assume that \( Z_{ij}(t) \) are independent and identically distributed (IID) copies of \( Z_j(t) \) (\( j = 1, 2, \ldots, p \)), where \( Z_j(t) \) are underlying smooth stochastic processes. We further assume that \( \epsilon_i(t) \) are IID copies of \( \epsilon(t) \), which is a mean 0 stochastic process. Model (1) in stacked form can be rewritten as \( Y(t) = Z(t) \beta(t) + \epsilon(t) \). Generally, in functional linear concurrent regression, estimation is done (Ramsay and Silverman, 2005) by minimizing the penalized residual sum of squares, \[ \text{SSE} = \int r(t)^T r(t) dt + \sum_{j=1}^{p} \lambda_j \int L_j \beta_j(t)^2 dt, \] where \( r(t) = Y(t) - Z(t) \beta(t) \). For example when \( L_j = I \), we minimize \( \int r(t)^T r(t) dt + \sum_{j=1}^{p} \lambda_j \int \beta_j(t)^2 dt \). Now suppose that \( \{ \theta_j(t), k = 1, 2, \ldots, k_j \} \) is a set of known basis functions for \( j = 1, 2, \ldots, p \). We model the unknown coefficient functions by using a basis function expansion as \( \beta_j(t) = \sum_{k=1}^{k_j} b_{kj} \theta_k(t) = \theta_j(t)^T b_j \), where \( \theta_j(t) = (\theta_{1j}(t), \theta_{2j}(t), \ldots, \theta_{kj}(t))^T \) and \( b_j = (b_{1j}, b_{2j}, \ldots, b_{kj})^T \) is a vector of unknown coefficients. In this paper, we use B-spline basis functions; however, other basis functions can be used as well. Then the minimization in the example mentioned above can be carried out by minimizing \[ \int (Y(t) - Z(t) \Theta(t)b)^T (Y(t) - Z(t) \Theta(t)b) dt + b^T R b. \] Here \( b, \Theta(t) \) and penalty matrix \( R \) are defined in stacked form as \( b = (b_1^T, b_2^T, \ldots, b_p^T)^T, \Theta(t) = \{ \theta_1(t)^T, \theta_2(t)^T, \ldots, \theta_p(t)^T \} \) and \( R = \text{diag}(R_1, R_2, \ldots, R_p) \), where \( R_j = \lambda_j b_j \int \theta_j(t) \theta_j(t)^T dt \). For our variable-selection method, we define the penalty on regression functions \( \beta_j(t) \) as \[ P_{\lambda, \psi} \{ \beta_j(t) \} = \lambda \left\{ \int \beta_j(t)^2 dt + \psi \int \beta_j''(t)^2 dt \right\}^{1/2} = \lambda (b_j^T R_j b_j + \psi b_j^T Q_j b_j)^{1/2} = \lambda (b_j^T K_{\psi,j} b_j)^{1/2}, \] where \( K_{\psi,j} = R_j + \psi Q_j \), \( R_j = \{ \int \theta_j(t) \theta_j(t)^T dt \} \) and \( Q_j = \{ \int \theta_j''(t) \theta_j''(t)^T dt \} \). This penalty was originally proposed by Meier et al. (2009) and later used by Gertheiss et al. (2013) for their
variable-selec tion method in scalar-on-function regression. The parameter $\psi \geq 0$ controls the amount of penalization on the roughness penalty. The proposed penalty simultaneously penalizes departure from sparsity and roughness of the coefficient functions ensuring that the resulting coefficient functions are smooth and small coefficient functions are shrunk to 0, introducing sparsity. Subsequently, we propose to minimize the following penalized mean sum of squares of the residuals for performing variable selection:

$$ L(b) = \frac{1}{n} \int (Y(t) - Z(t)\Theta(t)b)^T(Y(t) - Z(t)\Theta(t)b) \, dt + \lambda \sum_{j=1}^{p} (b_j^T \|\psi_j^T b_j\|)^{1/2}. \quad (2) $$

Since we assume that data are observed on a dense equispaced grid, the variable selection in practice is carried out by minimizing the following equivalent criterion:

$$ \sum_{i=1}^{n} \sum_{l=1}^{m} \left[ Y_i(t_l) - \sum_{j=1}^{p} Z_{ij}(t_l) \left\{ \sum_{k=1}^{k_j} b_{kj}(t_l) \right\} \right]^2 + \lambda mn \sum_{j=1}^{p} (b_j^T \|\psi_j^T b_j\|)^{1/2}. \quad (3) $$

The above formulation ignores the temporal dependence that might be present within the functions. In Section 2.2, we modify this criterion to take possible correlation in the error process into account. Now, using a Cholesky decomposition of $\|\psi_j^T b_j\| = \|\psi_j^T b_j\|^2$ and denoting $\gamma_j = \|\psi_j^T b_j\|$, the penalized sum of squares of residuals can be reformulated as

$$ R(\gamma) = \sum_{i=1}^{n} \sum_{l=1}^{m} \left\{ Y_i(t_l) - \sum_{j=1}^{p} Z_{ij}^*(t_l) b_j \right\}^2 + \lambda mn \sum_{j=1}^{p} (b_j^T \|\psi_j^T b_j\|)^{1/2} $$

where $Y_i = (Y_i(t_1), Y_i(t_2), \ldots, Y_i(t_m))^T$ and $\gamma = (\gamma_1, \gamma_2, \ldots, \gamma_p)^T$, and $Z_{ij}^*$ is defined as follows:

$$ Z_{ij}^* = \begin{pmatrix} Z_{i1}^*(t_1) & Z_{i2}^*(t_1) & \cdots & Z_{i1}^*(t_p) \\ Z_{i2}^*(t_1) & Z_{i2}^*(t_2) & \cdots & Z_{i2}^*(t_p) \\ \vdots & \vdots & \ddots & \vdots \\ Z_{im}^*(t_1) & Z_{im}^*(t_2) & \cdots & Z_{im}^*(t_p) \end{pmatrix}. $$

where $Z_{ij}^*$ refers to the $j$th block column in this matrix. We recognize this minimization problem as performing a group lasso (Yuan and Lin, 2006), where the grouping is introduced by covariates. In particular, we obtain estimates of $\gamma_j$ by minimizing similar penalized least squares as in the group lasso, namely
We extend this group lasso formulation to non-convex penalties, which are known (Breheny and Huang, 2015; Mazumder et al., 2011) to produce sparser solutions especially when there are large numbers of variables. In particular, we propose to use two non-convex penalties: SCAD (Fan and Li, 2001) and the MCP (Zhang, 2010). These two penalties overcome the high bias problem of the lasso as they relax the rate of penalization as the magnitude of the coefficient grows large. SCAD and the MCP have been shown to ensure selection consistency and estimation consistency under standard assumptions in the scalar regression case. They also enjoy the so-called oracle property in which they behave like the oracle maximum likelihood estimate asymptotically. Unlike the adaptive lasso, these methods do not require initial estimates of weights. These facts motivate us to use them in our functional variable selection context. Then the problem of variable selection reduces to a group SCAD or group MCP problem in our modelling set-up as follows.

2.1.1. Group smoothly clipped absolute deviation method

In the group SCAD method, we perform variable selection and obtain estimates of \( \gamma \) by using a penalized least squares criterion as in equation (4), where we now use a group SCAD penalty on the coefficients instead of a group lasso. In particular we estimate

\[
\hat{\gamma} = \arg\min_{\gamma_j, j=1,2,...,p} \sum_{i=1}^{n} \|Y_i - \sum_{j=1}^{p} Z_i^j \gamma_j \|_2^2 + \lambda mn \sum_{j=1}^{p} (\gamma_j^T \gamma_j)^{1/2} \\
= \arg\min_{\gamma_j, j=1,2,...,p} \sum_{i=1}^{n} \|Y_i - \sum_{j=1}^{p} Z_i^j \gamma_j \|_2^2 + \lambda mn \sum_{j=1}^{p} \|\gamma_j\|_2 \\
= \arg\min_{\gamma_j, j=1,2,...,p} \sum_{i=1}^{n} \|Y_i - \sum_{j=1}^{p} Z_i^j \gamma_j \|_2^2 + mn \sum_{j=1}^{p} P_{asso, \lambda}(\|\gamma_j\|_2),
\]

where \( P_{SCAD, \lambda, \phi}(\|\gamma_j\|_2) \) is defined in the following way:

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

2.1.2. Group minimax concave penalty method

For the group MCP method we use a group MCP on the coefficients and estimate \( \gamma \) as

\[
\hat{\gamma} = \arg\min_{\gamma_j, j=1,2,...,p} \sum_{i=1}^{n} \|Y_i - \sum_{j=1}^{p} Z_i^j \gamma_j \|_2^2 + mn \sum_{j=1}^{p} P_{MCP, \lambda, \phi}(\|\gamma_j\|_2),
\]

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

\[
P_{MCP, \lambda, \phi}(\|\gamma_j\|_2) = \begin{cases} 
\lambda \|\gamma_j\|_2 - \frac{\|\gamma_j\|_2^2}{2 \phi} & \text{if } \|\gamma_j\|_2 \leq \lambda \phi, \\
0.5 \lambda^2 \phi & \text{if } \|\gamma_j\|_2 > \lambda \phi.
\end{cases}
\]
2.2. Incorporating covariance structure into variable selection

The variable-selection method that was proposed in Section 2.1 does not account for possible correlation in the error process. In reality, however, temporal correlation is more likely to be present within functions. Although using an independent working correlation structure can yield consistent and unbiased estimates, incorporating the true covariance structure in the variable-selection criterion (4), (5) or (6) may give definite gains in terms of performance, as illustrated by Chen et al. (2016). We follow a similar prewhitening procedure to that employed by Chen et al. (2016) and Kim et al. (2018) to take into account the correct covariance structure. We assume that the error process $(\epsilon(t))$ follows a random walk $\epsilon(t) = V(t) + w_t$, where $V(t)$ is a smooth mean 0 stochastic process with covariance kernel $G(s, t)$ and $w_t$ is white noise with variance $\sigma^2$. The covariance function of the error process is then given by $\Sigma(s, t) = \text{cov}\{\epsilon(s), \epsilon(t)\} = G(s, t) + \sigma^2 I(s = t)$. For data observed on a dense and regular grid, the covariance matrix of the residual vector is then given by $\Sigma = \text{diag}(\Sigma_{m \times m}, \Sigma_{m \times m}, \ldots, \Sigma_{m \times m})$, where $\Sigma_{m \times m}$ denotes the covariance kernel $\Sigma(s, t)$ evaluated at $s = \{t_1, t_2, \ldots, t_m\}$. If $\Sigma_{m \times m}$ is known, $Y_i$ and $Z_{i}^{*j}$ as $Y_i = \{\Sigma_{m \times m}^{-1/2}\} Y_i$ and $Z_{i}^{*j} = \{\Sigma_{m \times m}\}^{-1/2}_j$, the same penalized criterion (4), (5) or (6) can be used to perform variable selection.

In reality $\Sigma$ is unknown, and we need an estimator $\hat{\Sigma}$. In the context of functional data, we want to estimate $\Sigma(\cdot, \cdot)$ non-parametrically. If we had the original residuals $\epsilon_{ij}$ available, we could use FPCA, e.g. Yao et al. (2005) or Zhang and Chen (2007), to estimate $\Sigma(s, t)$. If the covariance kernel $G(s, t)$ of the smooth part $V(t)$ is a Mercer kernel (Mercer, 1909), by Mercer’s theorem $G(s, t)$ must have a spectral decomposition

$$G(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t),$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$ are the ordered eigenvalues and $\phi_k(\cdot)$s are the corresponding eigenfunctions. Thus we have the decomposition $\Sigma(s, t) = \sum_{k=1}^{\infty} \lambda_k \phi_k(s) \phi_k(t) + \sigma^2 I(s = t)$. Given $\epsilon_{ij} = V(t_{ij}) + w_{ij}$, one could employ FPCA-based methods to obtain $\hat{\phi}_k(\cdot)$, $\hat{\lambda}_k$s and $\hat{\sigma}^2$. So an estimator of $\Sigma(s, t)$ can be formed as

$$\hat{\Sigma}(s, t) = \sum_{k=1}^{K} \hat{\lambda}_k \hat{\phi}_k(s) \hat{\phi}_k(t) + \hat{\sigma}^2 I(s = t),$$

where $K$ is sufficiently large for the convergence to hold and is typically chosen such that the percentage of variance explained, PVE, by the selected eigencomponents exceeds some prespecified value such as 99% or 95%. In reality, we do not have the original residuals $\epsilon_{ij}$ and use the full model (1) to obtain residuals $\hat{\epsilon}_{ij} = Y_i(t_{ij}) - \hat{Y}_i(t_{ij})$. Then, treating $\hat{\epsilon}_{ij}$ as our original residuals, we obtain $\hat{\Sigma}(s, t)$ by using FPCA.

Remark 1. We use a cubic B-spline basis with the same number of the basis functions to model the regression functions $\beta_j(t)$s, where the number of the basis is large so the basis is sufficiently rich. For selection of the tuning parameter $\psi$ (for smoothness) and the penalty parameter $\lambda$, we use the extended Bayesian information criteria (BIC) (Chen and Chen, 2008) corresponding to the equivalent linear model of criterion (4), (5) or (6) and this has shown good performance in our simulation study. Chen and Chen (2008) established consistency of the extended BIC under standard assumptions and illustrated its superiority over other methods like cross-validation, the Akaike information criterion AIC and BIC, which tend to overselect the variables. For tuning parameter $\phi$ we use the values 4 for SCAD and 3 for the MCP, as proposed by the original authors. For model fitting we use the gprpreg package (Breheny and Huang, 2015) in R (R Core Team, 2018).
Remark 2. In practice, we recommend standardizing the variables either by using the Euclidean norm (which is automatically performed in \texttt{grpreg}) or using FPCA-based methods \((X_j^* (t) = \{X_j(t) - \mu_j(t)\}/\sigma_j(t))\), which are especially useful for highly sparse data where some B-splines might not have observed data on its support. Standardizing is necessary to make sure that the variable-selection process does not depend on the variance of the predictors and is essential in applications where the variables have different units, e.g. the fisheries footprint study in this paper. Standardization can also help in faster convergence of the method proposed. By default, we use standardization using the Euclidean norm since we have implemented our procedure by using the \texttt{grpreg} package in R. However, we investigated both standardization methods in our simulation studies (the results are not shown) and obtained very similar results.

2.3. Extension to sparse data and noisy covariates

More generally we can consider the case where data are observed sparsely and covariates are observed with measurement error. This is most often the case for longitudinal data. Here the observed data are the response \(\{Y_i(t_{ij}), t_{ij}\}, j=1, 2, \ldots, m_i\}\) and the observed covariates \(\{(U_1(t_{ij}), t_{ij}), j=1, 2, \ldots, m_i\}, \{(U_2(t_{ij}), t_{ij}), j=1, 2, \ldots, m_i\}, \ldots\) \(\{(U_p(t_{ij}), t_{ij}), j=1, 2, \ldots, m_p\}\). Denote \(U_k(t_{ij})\)\(s (k=1, 2, 3, \ldots, p)\) by \(U_{ijk}\). Here \(U_{ijk}\)s represent the observed covariates with measurement error, i.e. we have \(U_{ijk} = Z_k(t_{ij}) + \epsilon_{ijk}\) for \(i=1, 2, \ldots, n, j=1, 2, \ldots, m_k\) and \(k=1, 2, \ldots, p\). The measurement errors \(\epsilon_{ijk}\) are assumed to be white noise with zero mean and variance \(\sigma_k^2\). In a sparse data set-up it is generally assumed (Kim et al., 2018) 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 \([0, T]\). Then we reconstruct the original curves from the observed sparse and noisy curves by using FPCA methods (Yao et al., 2005) by estimating the eigenvalues and eigenfunctions corresponding to the original curves. Li and Hsing (2010) proved uniform convergence of the mean, eigenvalues and eigenfunctions associated with the curves for both dense and in particular sparse designs under suitable regularity conditions. For prediction of the scores, we use the principal analysis by conditional expectation method as in Yao et al. (2005). Then these estimates are put together by using a Karhunen–Loève expansion to obtain estimates \(\hat{Z}_{ik}(\cdot)\) of the true curves \(Z_{ik}(\cdot)\) as \(\hat{Z}_{ik}(t) = \hat{\mu}_k(t) + \sum_{s=1}^S \hat{\zeta}_{ik} \hat{\psi}_s(t)\), where the number of eigenfunctions \(S\) to use is chosen by using the percentage of variance explained criterion PVE, which is the percentage of variance explained by the first few eigencomponents. Alternatively one can also use multivariate FPCA (Happ and Greven, 2018) instead of running FPCA on each predictor variable separately. Then, for sparse data observed on an irregular grid and observed with measurement error, we use \(\{Y_i(t_{ij}), \hat{Z}_{i1}(t_{ij}), \hat{Z}_{i2}(t_{ij}), \ldots, \hat{Z}_{ip}(t_{ij}), j=1, 2, \ldots, m_i\}_{i=1}^n\) as our original data for performing variable selection.

3. Simulation study

3.1. Simulation set-up

In this section, we evaluate the performance of our variable-selection method by using a simulation study. For this we generate data from the model

\[
Y_i(t) = \beta_0(t) + \frac{20}{j=1} Z_{ij}(t) \beta_j(t) + \epsilon_i(t), \quad i = 1, 2, \ldots, n, \quad t \in [0, 100].
\]

The regression functions are given by \(\beta_0(t) = 8 \sin(\pi t/50), \beta_1(t) = 5 \sin(\pi t/100), \beta_2(t) = 4 \sin(\pi t/50) + 4 \cos(\pi t/50), \beta_3(t) = 25 \exp(-t/20)\) and the rest of the \(\beta_j(t) = 0\) for \(j = 4, 5, 6, \ldots, 20\), i.e. the last 17 covariates are not relevant. The original covariates \(Z_{ij}(\cdot) \sim \text{IID} Z_j(\cdot)\), where \(Z_j(t) (j = 1, 2, \ldots, 20)\) are given by \(Z_j(t) = a_1 \sqrt{2} \sin(\pi j t/400) + b_1 \sqrt{2} \cos(\pi j t/400)\), where \(a_1 \sim \text{unif}(-1, 1)\) and \(b_1 \sim \text{unif}(-1, 1)\). The measurement errors \(\epsilon_i(t)\) are assumed to be white noise with zero mean and variance \(\sigma^2\). In a sparse data set-up it is generally assumed (Kim et al., 2018) 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 \([0, T]\). Then we reconstruct the original curves from the observed sparse and noisy curves by using FPCA methods (Yao et al., 2005) by estimating the eigenvalues and eigenfunctions corresponding to the original curves. Li and Hsing (2010) proved uniform convergence of the mean, eigenvalues and eigenfunctions associated with the curves for both dense and in particular sparse designs under suitable regularity conditions. For prediction of the scores, we use the principal analysis by conditional expectation method as in Yao et al. (2005). Then these estimates are put together by using a Karhunen–Loève expansion to obtain estimates \(\hat{Z}_{ik}(\cdot)\) of the true curves \(Z_{ik}(\cdot)\) as \(\hat{Z}_{ik}(t) = \hat{\mu}_k(t) + \sum_{s=1}^S \hat{\zeta}_{ik} \hat{\psi}_s(t)\), where the number of eigenfunctions \(S\) to use is chosen by using the percentage of variance explained criterion PVE, which is the percentage of variance explained by the first few eigencomponents. Alternatively one can also use multivariate FPCA (Happ and Greven, 2018) instead of running FPCA on each predictor variable separately. Then, for sparse data observed on an irregular grid and observed with measurement error, we use \(\{Y_i(t_{ij}), \hat{Z}_{i1}(t_{ij}), \hat{Z}_{i2}(t_{ij}), \ldots, \hat{Z}_{ip}(t_{ij}), j=1, 2, \ldots, m_i\}_{i=1}^n\) as our original data for performing variable selection.
\( \mathcal{N}(50, 2^2) \) and \( b_j \sim \mathcal{N}(50, 2^2) \). We moreover assume that \( Z_{ij}(t) \) are observed with measurement error, i.e. we observe \( U_{ij}(t) = Z_{ij}(t) + \delta_j, \) where \( \delta_j \sim \mathcal{N}(0, 0.6^2) \). The error process \( \epsilon_i(t) \) is generated as follows:

\[ \epsilon_i(t) = \xi_{i1} \cos(t) + \xi_{i2} \sin(t) + \mathcal{N}(0, 1), \]

where \( \xi_{i1} \sim \text{IID}\ \mathcal{N}(0, 0.5^2) \) and \( \xi_{i2} \sim \text{IID}\ \mathcal{N}(0, 0.75^2) \). The response \( Y_i(t) \) and noisy covariate \( U_{ij}(t) \) are observed sparsely for randomly chosen \( m_i \) points in \( S \), the set of \( m = 81 \) equidistant time points in \([0, 100]\) and \( m_i \sim \text{IID } \text{Unif}\{30, 31, \ldots, 41\} \). Three sample sizes \( n \in \{100, 200, 400\} \) are considered. For each sample size, we use 500 generated data sets for evaluation of our method.

### 3.2. Simulation results

Our primary interest is selection (identification) of the relevant covariates \( Z_1(\cdot), Z_2(\cdot) \) and \( Z_3(\cdot) \) and estimating their effects \( \beta_1(t), \beta_2(t) \) and \( \beta_3(t) \) accurately. As the covariates are observed sparsely and with measurement error, we apply FPCA as discussed in Section 2.3 with PVE = 99% and obtain the denoised curves \( \hat{Z}_{ij}(t) \) before applying our variable-selection method. We apply the proposed variable-selection method with and without the prewhitening procedure that was mentioned in Section 2.2. Table 1 and Table 2 display the selection percentage of each variable for each of the three selection methods that were discussed in Section 2 and for the three sample sizes \( n = 100, 200, 400 \), for the non-prewhitened and prewhitened case respectively. We call the functional lasso ‘FLASSO’, functional SCAD ‘FSCAD’ and functional MCP ‘FMCP’ for the proposed variable-selection methods for the functional linear concurrent model. We expect that the group lasso selection method will have a higher false positive rate and use this as a benchmark for comparison. It can be seen from Tables 1 and 2 that all the three methods (the group lasso, group SCAD and group MCP) pick out the three true covariates \( Z_1(\cdot), Z_2(\cdot) \) and \( Z_3(\cdot) \) 100% of the time. The group lasso method has a high false positive selection percentage as can be seen in both Table 1 and Table 2, with selection accuracy improving with increasing sample size. The group SCAD and group MCP method, in contrast, have a false selection percentage in the range of 0.2 – 1% for the non-prewhitened case and exactly 0% for the prewhitened case.

In other words, the group SCAD and group MCP methods can identify the true model by using the prewhitening procedure. In scalar regression, SCAD and the MCP are known to produce sparser solutions than does the lasso because of its concave nature and, here also in the context of variable selection in functional linear concurrent regression, we observe that these two methods (their group extension) outperform the lasso. The average model sizes for each scenario are also given in Table 1, and the group SCAD and group MCP method produce smaller and closer values to the true model size 3 (exactly 3 with the prewhitening procedure in Table 2). These results also illustrate the benefit of prewhitening and henceforth we use prewhitening as a preprocessing step to perform variable selection by using the methods proposed.

Next, as an assessment of the accuracy of the estimates \( \hat{\beta}_k(t) \) \((k = 1, 2, 3)\), we plot the true regression curves overlaid by their Monte Carlo (MC) mean estimate from the three methods. MC pointwise confidence intervals (95%) (corresponding to pointwise 2.5- and 97.5-percentiles of the estimated curves over 500 replicates) for each of the three curves are also displayed to assess variability of the estimates. Fig. 2 displays this plot for \( n = 200 \); the plots for \( n = 100 \) and \( n = 400 \) are similar with more accuracy and less variability for larger sample sizes. The group lasso estimates (the broken curve) have a larger bias which is again expected, as the lasso is known to have a relatively high bias when the magnitude of the regression coefficient is large. The group SCAD (the dotted curve) and group MCP (the chain curve) estimates have almost identical accuracy and variability as seen from Fig. 2; they have superimposed on each other and on the true curves represented by the full curves.
### Table 1. Comparison of selection percentages of different variables and average model size, without prewhitening

| Sample size n | Method   | Results (%) for the following variables: | Average model size |
|---------------|----------|------------------------------------------|--------------------|
|               |          | Var1 Var2 Var3 Var4 Var5 Var6 Var7 Var8 Var9 Var10 Var11 Var12 Var13 Var14 Var15 Var16 Var17 Var18 Var19 Var20 |                     |
| 100           | FLASSO   | 100 100 100 16.4 18.4 16.6 10 14.4 15 15.6 17.2 15.2 13 14.2 17.8 16.4 15.4 16 14.4 13 | 5.59               |
|               | FSCAD    | 100 100 100 0.6 0.4 0.2 1.0 1.2 0.4 0.8 0.4 0.2 0.4 0.8 0.2 0.2 0.2 0.6 0 1 | 3.088              |
|               | FMCP     | 100 100 100 0.6 0.2 0.2 1.0 1.0 0.4 0.4 0.6 0.4 0.2 0.2 0.4 0.2 0.2 0.6 0 1 | 3.076              |
| 200           | FLASSO   | 100 100 100 15.8 14.6 16.8 14 13.4 15.4 11.6 14.2 14.8 14.4 15 14.4 14 11.2 14.6 10.8 15 | 5.4                |
|               | FSCAD    | 100 100 100 0.2 0.6 1 0.6 0.4 0.8 0.8 1.2 0.2 0.2 0 0.4 0.2 0.4 0.8 0.4 0 1 | 3.092              |
|               | FMCP     | 100 100 100 0.2 0.6 0.8 0.4 0.4 0.8 0.6 1 0.2 0.2 0 0.4 0.2 0.4 0.8 0.2 0 0.8 | 3.08               |
| 400           | FLASSO   | 100 100 100 13.8 13.4 14.8 11.2 12.6 12.8 10.8 13.4 11 12.4 12.4 15.2 11.2 12.8 13.2 12 13.6 | 5.176              |
|               | FSCAD    | 100 100 100 0.4 0 0.2 0.4 0.4 0 0 0 0.4 0 0.2 0.2 0.2 0.2 0 0 0 | 3.026              |
|               | FMCP     | 100 100 100 0.4 0 0.2 0.4 0.2 0 0 0 0.4 0 0.2 0.2 0.2 0.2 0 0 0 | 3.024              |

### Table 2. Comparison of selection percentages of different variables and average model size, with prewhitening

| Sample size n | Method   | Results (%) for the following variables: | Average model size |
|---------------|----------|------------------------------------------|--------------------|
|               |          | Var1 Var2 Var3 Var4 Var5 Var6 Var7 Var8 Var9 Var10 Var11 Var12 Var13 Var14 Var15 Var16 Var17 Var18 Var19 Var20 |                     |
| 100           | FLASSO   | 100 100 100 6.2 7.8 7.6 6 7.8 6.4 6.6 7.4 5.8 5.4 6.4 6.4 7.4 4.8 6 6 6.2 4.112 |                     |
|               | FSCAD    | 100 100 100 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 |                     |
|               | FMCP     | 100 100 100 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 |                     |
| 200           | FLASSO   | 100 100 100 5.8 3 6 4.4 5.2 4.8 5.6 4.2 7.6 4.4 4.8 3.2 3.4 2.8 5.6 4.4 16 3.932 |                     |
|               | FSCAD    | 100 100 100 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 |                     |
|               | FMCP     | 100 100 100 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 |                     |
| 400           | FLASSO   | 100 100 100 4.2 2.6 4.6 3.8 3.6 3 2.6 3.4 5 5.2 5.4 3.6 3.4 3.4 4.6 4.8 29 3.922 |                     |
|               | FSCAD    | 100 100 100 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 |                     |
|               | FMCP     | 100 100 100 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 3 |                     |
To evaluate the performance of the estimates further we calculate the absolute bias and the MC mean-square error of the estimates averaged across 100 equally spaced points in [0, 100], for all the selection methods and the three sample sizes. These are displayed in Table 3. We again observe that the group SCAD and group MCP methods outperform the group lasso method, in terms of both absolute bias and mean-square error, the performance of the estimators improving with increasing sample size. We compared these mean-square errors of the estimates arising from the prewhitening procedure with the same from the non-prewhitening procedure and found that these were only marginally higher, which is expected because of the uncertainty that is associated with estimating the covariance matrix. The mean-square errors appear to be converging to 0 across all the three methods with an increase in sample size indicating consistency of the estimators.

The simulation results illustrate superior performance of the proposed group SCAD
Table 3. Comparison of MC absolute bias and mean-square error MSE

| Sample size n | Method     | $\hat{\beta}_1(t)$ Bias | MSE | $\hat{\beta}_2(t)$ Bias | MSE | $\hat{\beta}_3(t)$ Bias | MSE |
|---------------|------------|-------------------------|-----|-------------------------|-----|-------------------------|-----|
| 100           | FLASSO     | 0.083                   | 0.033 | 0.092                   | 0.048 | 0.112                   | 0.191 |
|               | FSCAD      | 0.011                   | 0.025 | 0.015                   | 0.038 | 0.022                   | 0.165 |
|               | FMCP       | 0.011                   | 0.025 | 0.015                   | 0.038 | 0.022                   | 0.165 |
| 200           | FLASSO     | 0.061                   | 0.017 | 0.069                   | 0.024 | 0.092                   | 0.109 |
|               | FSCAD      | 0.007                   | 0.013 | 0.008                   | 0.019 | 0.010                   | 0.091 |
|               | FMCP       | 0.007                   | 0.013 | 0.008                   | 0.019 | 0.010                   | 0.091 |
| 400           | FLASSO     | 0.047                   | 0.009 | 0.051                   | 0.013 | 0.070                   | 0.063 |
|               | FSCAD      | 0.004                   | 0.007 | 0.004                   | 0.010 | 0.010                   | 0.050 |
|               | FMCP       | 0.004                   | 0.007 | 0.004                   | 0.010 | 0.010                   | 0.050 |

(FSCAD) or group MCP (FMCP) selection method in the context of a functional linear concurrent model and are the recommended methods of this paper.

4. Real data applications

In this section, we demonstrate applications of our variable-selection method in the selection of influential time-varying predictors in two real data studies. For performing variable selection, we use only the FSCAD and FMCP methods along with the initial prewhitening procedure, as the group LASSO method yields a significantly higher false positive rate which is illustrated by our simulations. We first consider a small dietary calcium absorption data set (three time-varying covariates) with added pseudocovariates as an illustration of our method. The addition of pseudocovariates is a popular way (Wang et al., 2008; Wu et al., 2007; Miller, 2002) of assessing the false selection rate in real data sets. Pseudovariables can, therefore, be used effectively for tuning variable-selection procedures. We show that our proposed method can select the relevant predictors and discard the pseudovariables successfully. Finally, we apply our variable-selection method to the fisheries data set to find out relevant socio-economic drivers influencing the fisheries footprint of nations over time.

4.1. Study of dietary calcium absorption

We consider the study of dietary calcium absorption in Davis (2002). In this study, the subjects are a group of 188 patients. We have data on calcium absorption $Y(t)$, dietary calcium intake $Z_1(t)$, body mass index (BMI) $Z_2(t)$ and body surface area (BSA) $Z_3(t)$ of these patients, at irregular time points between 35 and 64 years of their ages. At the beginning of the study patients aged between 35 and 45 years and subsequent observations were taken approximately every 5 years. The number of repeated measurements for each patient varies from 1 to 4. Fig. 3 displays the individual curves of patients’ calcium absorption, calcium intake, BSA and BMI along their ages.

We are primarily interested in finding out which covariates influence the calcium absorption profile of the patients. Kim et al. (2018) also investigated the effect of calcium intake on calcium absorption by using an additive non-linear functional concurrent model and found that the effect was virtually linear while comparing it with a functional linear concurrent model. So we use functional linear concurrent regression to model the dependence of calcium absorption on calcium intake, BSA and BMI. As data are observed very sparsely and the original covariates might...
be observed with measurement error, we apply FPCA methods (PVE = 95%) as discussed in Section 2.3 and obtain the denoised trajectories $\hat{Z}_{j}(t)$ for $j = 1, 2, 3$. We expect that calcium intake among the three covariates will be associated with calcium absorption. We add 15 pseudocovariates by simulating from the following functional model to illustrate the selection performance and false positive rate of our variable-selection method. We generate $Z_{ij}(\cdot) \sim \text{IID} Z_{j}(\cdot)$ where

$$Z_{j}(t) \sim \mathcal{N}(0, 2^2) \quad \text{and} \quad b_{j} \sim \mathcal{N}(0, 2^2).$$

So in total, we have 18 covariates, where the first three are the denoised original covariates and the rest are simulated predictors. Then we apply our variable-selection method to $Y(t)$ and $\hat{Z}_{1}(t), \hat{Z}_{2}(t), \hat{Z}_{3}(t), Z_{4}(t), Z_{5}(t), \ldots, Z_{18}(t)$. We repeat this a large number of times and observe which variables are being selected in each iteration. We expect that our variable-selection method will pick out the truly influential predictors and ignore the randomly generated functional covariates the majority of the time. To illustrate the benefit of using our proposed variable-selection method in a functional regression model for these particular data we compare its performance with a backward selection method which uses a model selection criterion like the BIC or Mallows’s $C_p$, under a linear model approach (using an independent working correlation structure), and with a penalized generalized estimating
Table 4. Selection percentages of variables in the calcium absorption study

| Method                  | Results (%) for the following variables: |
|------------------------|------------------------------------------|
|                        | Var 1 (calcium intake) | Var 2 (BSA) | Var 3 (BMI) | Maximum Var(4–18) |
| FSCAD                  | 100                       | 0           | 0           | 0                |
| FMCP                   | 100                       | 0           | 0           | 0                |
| BIC (linear model)     | 100                       | 100         | 0           | 9                |
| C_p (linear model)     | 100                       | 100         | 2           | 33               |
| PGEE (independent)     | 100                       | 0           | 100         | 31               |
| PGEE (AR(1))           | 100                       | 0           | 100         | 31               |

The equation (PGEE) procedure (Wang et al., 2012) which was developed to analyse longitudinal data with a large number of covariates. We use the PGEE package in R (Inan et al., 2017) for implementing the PGEE procedure under two working correlation structures (independent and auto-regressive AR(1)).

Table 4 illustrates the selection percentage of each of the variables under the various methods. We note that both the proposed FSCAD and FMCP method identify calcium intake $Z_1(t)$ as a significant predictor 100% of the time. All other variables including all the pseudocovariates are ignored in 100% of the iterations. In contrast, the BIC, $C_p$ and the PGEE procedure exhibit a high false selection percentage for the pseudovariables. The case of selection of BSA or BMI appears to be overselection as their individual effects were not found to be statistically significant. This demonstrates that, when the underlying model is functional, the use of naive variable-selection methods using scalar regression techniques can lead to wrong inference as they do not account for the functional nature of the data.

As calcium intake is the only significant variable selected by both the methods proposed we want to estimate its effect and also to obtain a measure of uncertainty of our estimate. For this, we use a subject level bootstrap on our original data (no pseudocovariates added) while performing variable selection to come up with an estimated regression curve $\hat{\beta}_1(t)$ and a pointwise confidence interval for the effect of calcium intake. This is displayed in Fig. 4. We note that, as calcium intake increases, calcium absorption should decrease particularly until age 60 years, as $\hat{\beta}_1(t) < 0$ up to this age and the confidence interval strictly lies below zero, which might be due to dietary calcium saturation or to interaction with some other elements in the body, although the overall magnitude of the effect seems to decrease with age. Above age 60 years, the estimate appears to have high variability associated with it, which is primarily because we have very few observations (5.62%) above this mark (illustrated in Fig. 4). The uncertainty in estimating $\hat{\beta}_1$ by using FPCA and the uncertainty due to the bootstrap are reflected in its variability. Hence, some care should be taken in interpreting the estimated regression curve beyond 60 years because of such high uncertainty.

Remark 3. The pseudovariables that are considered in this section are generated independently of the true covariates. To investigate what happens when there is significant correlation between the truly relevant and the pseudovariables we considered the following scenario. The pseudovariables $Z_j(t)$ ($j = 4, 5, \ldots, 18$) were generated conditionally on $Z_1(t)$ (calcium intake) from a normal distribution with mean $\rho (Z_1(t) - a)/b$ and variance $1 - \rho^2$. We set $\rho = 0.9$ in our analysis. The constants $a$ and $b$ were taken to be the scalar mean and standard deviation of the samples $\{Z_{ij}(t_{ij}), j = 1, 2, \ldots, m_i\}_{i=1}^n$. We found that the proposed FMCP method performs
Fig. 4. (a) Bootstrap estimate and pointwise confidence interval of the effect of calcium intake (---, group SCAD; --, group MCP) and (b) percentage observation available in different age groups.
well, even in the presence of such a high correlation with negligible false positive rate (less than 0.07%) on the pseudovariables. In contrast, for the FSCAD method, the average false positive rate increases to 10.8%. This illustrates that the FMCP method might be better suited to handle correlated predictors.

4.2. **Study of fisheries footprint**

Fisheries production is a source of protein as well as an economic livelihood across the world. Along with the increasing global population, the importance of fish production and consumption has steadily increased through the modern era. The fisheries footprint is defined as the Global Footprint Network’s measure of total marine area required to sustain consumption levels of aquatic production of fish, crustaceans (e.g. shrimps), shellfish and seaweed from captures and aquaculture; so the fisheries footprint basically represents the coastal and marine area that is required to sustain the amount of seafood products that a nation consumes.

Over the last two decades, social scientists have accomplished much in advancing scholarly knowledge on the social drivers of ecological impact at a macroscale. Such work is essential, as ecological problems are becoming increasingly interlinked and severe at a global or planetary scale (Steffen et al., 2011). Over time, for example, economic development, population structuring (e.g. urbanization or age structure), trade relations and technological change have been shown to affect measures of environmental impact across nations (Clark and Longo, 2019; Jorgenson and Clark, 2010; York et al., 2003). This body of literature centres on the ecological effects of globalization and modernization, under the sociostructural parameters of a capitalist economy. There is still much debate over the effects of industrial and agricultural modernization on ecosystems. For example, the development and resource economics literature (World Bank, 2007) advocate for the utilization of innovation and technological developments to improve marine system sustainability and food security (Valderrama and Anderson, 2010), whereas, in contrast, environmental sociologists have demonstrated that such innovation, chiefly aquaculture, does not displace the deleterious ecological effects of capture fisheries (Longo et al., 2019). Nevertheless, despite such progress, the fisheries footprint remains an understudied metric, and its drivers are less understood in social research (Clark et al., 2018; Jorgenson et al., 2005).

The goal of this study, therefore, is to identify the relevant socio-economic drivers such as levels of economic development, population size and transformations in food system dynamics that influence the fisheries footprint of nations over time and also to capture their time-varying effects. Data for this study are collected from the World Bank, Fish StatJ of the United Nations Food and Agriculture Organization and the Ecological Footprint Network for the years between 1970 and 2009, across 136 nations.

The main dependent variable of interest in this study is the fisheries footprint (measured in hectares) which is log-transformed before analysis. To capture the trend of fisheries footprints over the years, we plot the mean fisheries footprint of the nations along with their pointwise 95% confidence interval. This is displayed in Fig. 5. We note an overall upward trend as well as heterogeneity across the years.

There are 20 independent time-varying covariates in the study, broadly covering various sectors of population dynamics (e.g. population density, urban population, total population and working age population percentage), agriculture (e.g. tractors and agriculture value added), food consumption and other fisheries variables (e.g. meat consumption and aquaculture production in tons), international relations (e.g. food export as a percentage of merchandize and foreign direct investment inflow) and economy (e.g. gross domestic product (GDP) per capita at constant US dollars and trade percentage of GDP). The full list of the variables is given in Table 5. The
predictors here are also time varying and can be expected to have dynamic effects on the fisheries footprint. Generally panel data methods like fixed effects or random-effects modelling are used (Torres-Reyna, 2007; Clark and Longo, 2019) for analysing such data where the effects of the covariates are taken to be constant, since we are interested in dynamic effects of the covariates,
the functional linear concurrent model can be seen as a generalization of this approach with time-varying effects of predictors. Therefore we use the functional linear concurrent regression model (1) discussed in this paper to model the dynamic effects of the socio-economic predictors on the fisheries footprint. The predictors in their original scale are also very large and therefore have been converted into a log-scale; the covariates observed as percentages are used without any conversion. Before applying our variable-selection method all the covariates are preprocessed by using FPCA methods (PVE = 95%) as discussed in Section 2.3.

We use the prewhitening procedure that was discussed in Section 2.2 and apply our proposed variable-selection method. Out of 20 covariates, the proposed FSCAD and FMCP method both identify GDP per capita and urban population as the two significant predictors. GDP is a measure of the market value of all the final goods and services that are produced in a specific time period, and GDP per capita is a measure of a country’s economic output adjusting for its
number of people. As the major economic indicator GDP per capita is associated with primary aspects of economic growth, consumer behaviour and trade and therefore is a key indicator of the fisheries footprint of nations over time. Furthermore, GDP per capita is a common metric in extant social science research to operationalize the extent to which a nation is successfully developing according to the standards of the world capitalist economy (Dietz and Jorgenson, 2013). Fig. 6(a) shows the estimated regression curve for GDP per capita obtained by applying the FSCAD selection method. The estimate from the FMCP method is similar. We observe the net effect of GDP per capita on the fisheries footprint to be positive and linear, although the magnitude of the effect has decreased over time.

As urban population is the key market, it also plays a crucial role in the total seafood consumption of nations and therefore influences the fisheries footprint. A change in urban population reflects urbanization and urbanization has important effects on food security and farming (Satterthwaite et al., 2010; Cohen and Garrett, 2010). Fig. 6(b) shows the estimated regression curve illustrating the effect of urban population on the fisheries footprint. Here also we note that the net effect of urban population on the fisheries footprint is positive, although the effect appears to be virtually constant with a very marginal decrease (on the log-scale).

Here, it is important to note that the fisheries footprint represents the metabolic potential of an ecosystem to reproduce itself ecologically. According to the Food and Agriculture Organization of the United Nations (Food and Agriculture Organization, 2016), about 58% of global fish stocks are currently fully exploited, and about 55% of ocean territory (conservatively) was subjected to industrial fishing in the past year (Kroodsma et al., 2018). Thus, there is declining metabolic potential for the expansion of capture fisheries, which probably helps to account for why variable effects were stronger in earlier, more ecologically productive decades.

Both these variables are important in the sense that they represent the primary indicators in economics, food consumption, population dynamics, trade, etc., which directly interact with a nation's need for seafood and therefore should influence the fisheries footprint. It is therefore not surprising that countries having high GDP per capita and/or high urban populations, e.g. the USA, Australia and Singapore, also have a high fisheries footprint. In Fig. 7 we display the fisheries footprint, GDP per capita and urban population profile of the three representative countries Afghanistan, Albania and the Bahamas. We note that the overall trend in the fisheries footprint profile can be described well by their GDP per capita and urban population profile, both of which were shown to have a positive effect on the fisheries footprint.

**Remark 4.** We have successfully applied our proposed variable-selection method to find the relevant time-varying predictors and their time-varying effects on the fisheries footprint. It is very plausible that there might be country- or region-specific effects on the fisheries footprint as revealed in the study by Clark and Longo (2019), and one might be interested in estimating these effects. The proposed variable-selection method for the functional linear concurrent model can be extended to handle such region-specific effects in its existing form.

**Remark 5.** We have considered concurrent effects of the predictors whereas some of the predictors might have lagged effects on the fisheries footprint. For example, in an economic crisis or recession, the predictors could very probably have reverberating effects on development for a few years. As invested capital takes time to flow through the economy, considering such lagged effects would be interesting. We applied our variable-selection method with lagged predictors present along with the original predictors (with lag windows 1 and 3 years). We found that, for lag 1 year, the proposed FMCP and FSCAD methods select almost identical models with the FSCAD method selecting ‘services value growth pct’ as an additional variable. Considering a lag window of 3 years, the FSCAD method selects an urban population lag instead of urban
Fig. 7. Profiles of the fisheries footprint, GDP per capita and urban population of the three representative countries (a) Afghanistan, (b) Albania and (c) the Bahamas

population whereas the FMCP method additionally selects ‘aquaculture production tons’ and ‘services value growth pct lag’ as influential covariates. These results indicate that some of the predictors could have reverberating effects and a more general framework like the historical functional regression model (Malfait and Ramsay, 2003) might be more suitable to model past effects of covariates on the response at the current time point.

5. Discussion

In this paper, we have proposed a variable-selection method in functional linear concurrent
regression extending the classically used penalized variable-selection methods like the lasso, SCAD and MCP. We have shown that the problem can be addressed as a group lasso and their natural extensions the group SCAD or group MCP problem. We have used a prewhitening procedure to take into account the temporal dependence within functions and, through numerical simulations, have illustrated that our proposed selection method with group SCAD or group MCP penalty can select the true underlying variables with high accuracy and has minuscule false positive and false negative rate even when data are observed sparsely, are contaminated with measurement error and the error process is highly non-stationary. We have illustrated the usefulness of the proposed method by applying it to two real data sets: the dietary calcium absorption study data and the fisheries footprint data for identification of the relevant time-varying covariates. In this paper we have used a resampling subject-based bootstrap method to measure uncertainty of the regression functions estimates; the theoretical properties corresponding to such a bootstrap method is something that we would like to explore more deeply in the future.

There are many interesting research directions that this work can head into. In real data, the dynamic effects of the predictors might always not be linear. In future, we would like to extend our variable-selection method to a non-parametric functional concurrent regression model (Maity, 2017), which is a more general and flexible model to capture complex relationships between the response and covariates. As mentioned earlier it would be also of interest to consider the lagged effects of covariates through a more general historical functional regression model (Malfait and Ramsay, 2003).

In developing our method we assumed that the covariates are independent and identically distributed. In many cases this might not be a reasonable assumption. For example, in the fisheries footprint data some countries could be very similar and form clusters; in contrast, they might not be even independent with the interplay of economies and other variables among nations. Even if the covariates are not independent over subjects, the variable-selection criterion that is proposed in this paper can still be used in practice as a penalized least squares method. The heterogeneity among the subjects can be addressed by using the interaction effect of covariates with regions, which can be clustered on the basis of the level of affluence. This can be done similarly to Clark and Longo (2019). Alternatively, one can also use subject-specific functional random effects for covariates, especially if one is interested in individual specific trajectories. A functional linear mixed model (Liu et al., 2017) might be an appropriate choice in such situations. Extending the proposed variable-selection method to such general functional regression models would be an extension of this work and remains an area for future research.

6. Software

All the methods that are discussed in this paper have been implemented by using the grpreg package (Breheny and Huang, 2015) in R. Illustrations of implementation of our method by using R are available from https://rss.onlinelibrary.wiley.com/hub/journal/14679879/series-c-datasets and at GitHub (https://github.com/rahulfrodo/FLCM_Selection).

Acknowledgements

We thank the Joint Editor, the Associate Editor and a referee for their valuable inputs and suggestions which have greatly helped in improving this paper.
References

Bretheny, P. and Huang, J. (2015) Group descent algorithms for nonconvex penalized linear and logistic regression models with grouped predictors. *Statist. Comput.*, 25, 173–187.

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

Chen, Y., Goldsmith, J. and Ogden, R. T. (2016) Variable selection in function-on-scalar regression. *Stat.*, 5, 88–101.

Chiang, C.-T., Rice, J. A. and Wu, C. O. (2001) Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *J. Am. Statist. Ass.*, 96, 605–619.

Clark, T. P. and Longo, S. B. (2019) Examining the effect of economic development, region, and time period on the fisheries footprints of nations (1961–2010). *Int. J. Compar. Sociol.*, 60, 225–248.

Clark, T. P., Longo, S. B., Clark, B. and Jorgenson, A. K. (2018) Socio-structural drivers, fisheries footprints, and seafood consumption: a comparative international study. 1961-2012. *J. Rurl Stud.*, 57, 140–146.

Cohen, M. J. and Garrett, J. L. (2010) The food price crisis and urban food (in)security. *Environ. Urbanizn.*, 22, 467–482.

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

Dietz, T. and Jorgenson, A. (2013) *Structural Human Ecology: New Essays in Risk, Energy, and Sustainability*. Pullman: Washington State University Press.

Eubank, R. L., Huang, C., Muñoz Maldonado, Y., Wang, N., Wang, S. and Buchanan, R. J. (2004) Smoothing spline estimation in varying-coefficient models. *J. R. Statist. Soc. B*, 66, 653–667.

Fan, J. and Li, R. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Statist. Ass.*, 96, 1348–1360.

Fan, J. and Zhang, W. (1999) Statistical estimation in varying coefficient models. *Ann. Statist.*, 27, 1491–1518.

Fan, Y., James, G. M. and Radchenko, P. (2015) Functional additive regression. *Ann. Statist.*, 43, 2296–2325.

Food and Agriculture Organization (2016) The state of the worlds fisheries and aquaculture. United Nations Food and Agriculture Organization, Rome.

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

Goldsmith, J. and Schwartz, J. E. (2017) Variable selection in the functional linear concurrent model. *Statist. Med.*, 36, 2237–2250.

Happ, C. and Greven, S. (2018) Multivariate functional principal component analysis for data observed on different (dimensional) domains. *J. Am. Statist. Ass.*, 113, 649–659.

Hastie, T. and Tibshirani, R. (1993) Varying-coefficient models with discussion). *J. R. Statist. Soc. B*, 55, 757–796.

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

Huang, J. Z., Wu, C. O. and Zhou, L. (2002) Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89, 111–128.

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

Inan, G., Zhou, J. and Wang, L. (2017) PGEE: penalized generalized estimating equations in high-dimension. *R Package Version 1.5*. (Available from https://CRAN.R-project.org/package=PGEE.)

Jorgenson, A. K. and Clark, B. (2010) Assessing the temporal stability of the population/environment relationship in comparative perspective: a cross-national panel study of carbon dioxide emissions, 1960–2005. *Popul Environ.*, 32, 27–41.

Jorgenson, A. K., Rice, J. and Crowe, J. (2005) Unpacking the ecological footprint of nations (1961–2010). *Int. J. Compar. Sociol.*, 60, 225–248.

Kauermann, G. and Tutz, G. (1999) On model diagnostics using varying coefficient models. *Biometrika*, 86, 119–128.

Kim, J., Maity, A. and Staicu, A.-M. (2018) Additive nonlinear functional concurrent model. *Statist. Interf.*, 11, 669–685.

Kroodsma, D. A., Mayorga, J., Hochberg, T., Miller, N. A., Boerker, K., Ferrett, F., Wilson, A., Bergman, B., White, T. D., Block, B. A., Woods, P., Sullivan, B., Costello, C. and Worm, B. (2018) Tracking the global footprint of fisheries. *Science*, 359, 904–908.

Li, Y. and Hsing, T. (2010) Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *Ann. Statist.*, 38, 3321–3351.

Liu, B., Wang, L. and Cao, J. (2017) Estimating functional linear mixed-effects regression models. *Computat Statist. Data Anal.*, 106, 153–164.

Longo, S. B. and Clark, B. (2016) An ocean of troubles: advancing marine sociology. *Socl Prob.*, 63, 463–479.

Longo, S. B., Clark, B., York, R. and Jorgenson, A. K. (2019) Aquaculture and the displacement of fisheries captures. *Conservn Biol.*, 33, 832–841.

Maity, A. (2017) Nonparametric functional concurrent regression models. *Wiley Interdisc. Rev. Computatnl Statist.*, 9, article e1394.
Malfait, N. and Ramsay, J. O. (2003) The historical functional linear model. "Can. J. Statist., 31, 115–128.

Mazumder, R., Friedman, J. H. and Hastie, T. (2011) Sparsenet: coordinate descent with nonconvex penalties. "J. Am. Statist. Ass., 106, 1125–1138.

Meier, L., Van de Geer, S. and Bühlmann, P. (2009) High-dimensional additive modeling. "Ann. Statist., 37, 3779–3821.

Mercer, J. (1909) Functions of positive and negative type, and their connection with the theory of integral equations. "Phil. Trans. R. Soc. A, 209, 415–446.

Miller, A. (2002) "Subset Selection in Regression." New York: Chapman and Hall.

Ramsay, J. and Silverman, B. (2005) "Functional Data Analysis." New York: Springer.

R Core Team (2018) "R: a Language and Environment for Statistical Computing." Vienna: R Foundation for Statistical Computing.

Satterthwaite, D., McGranahan, G. and Tacoli, C. (2010) Urbanization and its implications for food and farming. "Phil. Trans. R. Soc. B, 365, 2809–2820.

Steffen, W., Persson, A., Deutsch, L., Zalasiewicz, J., Williams, M., Richardson, K., Crutzen, P., Folke, C., Gordon, L., Molina, M., Ramanathan, V., Rockström, J., Scheffer, M., Schellnhuber, H. J. and Svedin, U. (2011) The anthropocene: from global change to planetary stewardship. "Ambio, 40, 739–761.

Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. "J. R. Statist. Soc. B, 58, 267–288.

Torres-Reyna, O. (2007) Panel data analysis fixed and random effects using stata (v. 4.2). Data and Statistical Services, Princeton University, Princeton.

Valderrama, D. and Anderson, J. L. (2010) Market interactions between aquaculture and common-property fisheries: recent evidence from the Bristol Bay sockeye salmon fishery in Alaska. "J. Environ. Econ. Mangmnt, 59, 115–128.

Wang, L., Li, H. and Huang, J. Z. (2008) Variable selection in nonparametric varying-coefficient models for analysis of repeated measurements. "J. Am. Statist. Ass., 103, 1556–1569.

Wang, L., Zhou, J. and Qu, A. (2012) Penalized generalized estimating equations for high-dimensional longitudinal data analysis. "Biometrics, 68, 353–360.

World Bank (2007) Changing the face of waters: the promise and challenge of sustainable aquaculture. World Bank Group, Washington DC.

Wu, C. O., Chiang, C.-T. and Hoover, D. R. (1998) Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. "J. Am. Statist. Ass., 93, 1388–1402.

Wu, Y., Boos, D. D. and Stefanski, L. A. (2007) Controlling variable selection by the addition of pseudovariables. "J. Am. Statist. Ass., 102, 235–243.

Yao, F., Müller, H.-G. and Wang, J.-L. (2005) Functional data analysis for sparse longitudinal data. "J. Am. Statist. Ass., 100, 577–590.

York, R., Rosa, E. A. and Dietz, T. (2003) Footprints on the earth: the environmental consequences of modernity. "Am. Sociol. Rev., 68, 279–300.

Yuan, M. and Lin, Y. (2006) Model selection and estimation in regression with grouped variables. "J. R. Statist. Soc. B, 68, 49–67.

Zhang, C.-H. (2010) Nearly unbiased variable selection under minimax concave penalty. "Ann. Statist., 38, 894–942.

Zhang, J.-T. and Chen, J. (2007) Statistical inferences for functional data. "Ann. Statist., 35, 1052–1079.