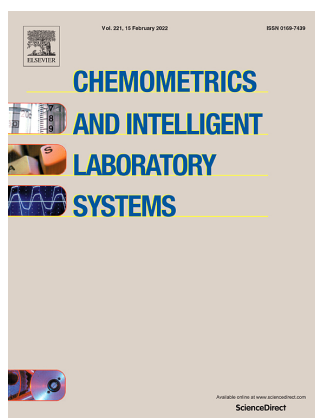
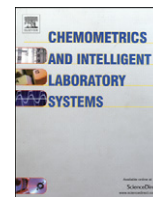


Penalized versions of functional PLS regression

- Ana M. Aguilera; M. Carmen Aguilera-Morillo; C. Preda
- Penalized versions of functional PLS regression
- *Chemometrics and Intelligent Laboratory Systems* 154 (2016) 80–92
- DOI: <https://doi.org/10.1016/j.chemolab.2016.03.013>





Penalized versions of functional PLS regression



A.M. Aguilera^{a,*}, M.C. Aguilera-Morillo^b, C. Preda^c

^aDepartment of Statistics and O.R. and IEMath-GR, University of Granada, Spain

^bDepartment of Statistics, University Carlos III of Madrid, Spain

^cDepartment of Statistics, Université des Sciences et Technologies de Lille 1, France and ISMMA Romanian Academy, Romania

ARTICLE INFO

Article history:

Received 24 August 2015

Received in revised form 22 January 2016

Accepted 11 March 2016

Available online 22 March 2016

2010 MSC:

62H99

60G12

Keywords:

PLS

Functional data

Penalized splines

Basis representation

ABSTRACT

Least squares estimation of the functional linear regression model with scalar response is an ill-posed problem due to the infinite dimension of the functional predictor. Dimension reduction approaches as principal component regression or partial least squares regression are proposed and widely used in applications. In both cases the interpretation of the model could be difficult because of the roughness of the coefficient regression function. In this paper, two penalized estimations of this model based on modifying the partial least squares criterion with roughness penalties for the weight functions are proposed. One introduces the penalty in the definition of the norm in the functional space, and the other one in the cross-covariance operator. A simulation study and several applications on real data show the efficiency of the penalized approaches with respect to the non-penalized ones.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The linear regression model was the first regression model considered in the framework of functional data analysis (FDA). Called the functional linear regression model (FLRM), it was intensively studied in the literature, depending on the type of the response and the predictor random variables. The first theoretical contributions to the study of FLRM are due to [1] for scalar response and functional predictor. The case of both predictor and response of functional type is considered in [2] and [3]. These linear models were first introduced in [4,5] to forecast a continuous time stochastic process from its recent past. Linear models with scalar or finite multidimensional predictor and functional response have been studied in [6]. Finally, the functional analysis of variance as a model for the mean of a functional response was introduced in [7].

To fit such linear models, the general methodology is to combine dimension reduction techniques and to modify the least squares criterion by introducing penalties. Principal component regression (PCR) [8,9] and partial least squares regression (PLSR) [10,11] are

two of the main dimension reduction approaches. Several smoothing estimation approaches based on penalizing the least squares criterion (roughness penalty) in terms of a B-spline expansion of the functional parameter and smoothed principal component regression were considered by [12].

In this paper we propose the estimation of the FLRM by modifying the PLS criterion with roughness penalties for the coefficient regression function. Our approach is motivated by the following considerations. It is well known that PLSR fits closer than PCR [13]. In [14] functional PCR and functional PLSR with basis expansion of the functional predictor were compared with their multivariate versions by an extensive simulation study. From this study, the authors concluded that, although discrete and functional models have similar prediction ability, the functional models provide a more accurate estimation of the functional parameter, PLSR outperforming PCR. However, in both cases (PLSR and PCR) even if the fit is very good in terms of prediction, the interpretation of the model could be difficult because of the complexity (roughness) of the coefficient regression function. A similar idea has been developed in [15] and [16] for smoothing principal components and in [12] in the regression framework. In [17] the authors introduce two discrete roughness penalty estimation approaches for PCR and PLSR based on B-spline basis expansion of the functional parameter. The main difference between these two approaches is in the way of introducing the

* Corresponding author.

E-mail address: aaguiler@ugr.es (A. Aguilera).

penalty: in the likelihood function of the model or in the construction of the PLS or principal components. These penalized estimation approaches do not assume a basis representation for the sample paths and are based on multivariate linear regression of the response in terms of the matrix of discrete-time observations of the sample curves. A penalized version of multivariate PLSR was also applied for the estimation of additive functional models in terms of B-spline expansions [18].

A roughness penalty for PLS regression based on redefining the inner product in the functional space data belong, is introduced in Section 3. A simulation study compares the performance of this penalized approach with the corresponding non-penalized versions in Section 4. Two applications with chemometric data (Section 5) close the paper. Finally, the methodological and practical results related with an alternative approach that introduces the penalty in the covariance have been presented as supplementary material because this type of penalization consistently performed worse than the other in all simulations and applications developed in this paper.

2. Functional PLS regression

Let Y be a real random variable (response) and X be a real-valued second order stochastic process $\{X(t) : t \in T\}$ (predictor), where T is some interval on the real line. We assume that the sample paths of X belong to the space $L_2(T)$ of the square integrable functions endowed with the usual inner product $\langle \cdot, \cdot \rangle$. Without loss of generality, we consider that $E[Y] = 0$ and $E[X(t)] = 0, \forall t \in T$.

The functional linear regression model with response Y and predictor X assumes that

$$Y = \int_T X(t)\beta(t)dt + \epsilon, \tag{1}$$

where β is the functional parameter and ϵ is the residual. It is well known [1,9] that the use of the least squares criterion for estimating the model in Eq. (1) yields to an ill-posed problem because of the covariance operator of X which, in general, is not invertible. Indeed, the estimation of the regression coefficient function β under the least squares criterion yields to the integral Wiener–Hopf equation

$$\mathbb{E}(YX(t)) = \int_0^T \mathbb{E}(X(t)X(s))\beta(s)ds. \tag{2}$$

The PLS approach is an efficient solution to the inverse problem encountered in Eq. (2). In the framework of functional data, that method was widely investigated by [10], [14] and [19]. It consists of building PLS components as linear functionals of X ,

$$t = \int_T X(t)w(t)dt,$$

obtained by maximizing the squared covariance

$$\max_{w, \|w\|^2=1} \text{Cov}^2 \left(\int_T X(t)w(t)dt, Y \right). \tag{3}$$

In the multivariate context this maximization problem is known as Tucker’s criterion [see 20, for a detailed study]. The PLS algorithm for multivariate linear regression was first introduced by [21]. The fact that the PLS components are solutions to the Tucker’s criterion was subsequently discovered by [22].

In what follows it is useful to detail how the PLS methodology works for fitting the model in Eq. (1).

2.1. The PLS algorithm

Let C_{YX} be the cross-covariance operator of X and Y and C_{XY} be its adjoint defined by

$$\begin{aligned} C_{YX} : L_2(T) &\rightarrow \mathbb{R} \\ f &\mapsto x = \int_T \text{Cov}(X(t), Y)f(t)dt, \\ C_{XY} : \mathbb{R} &\rightarrow L_2(T), \\ x &\mapsto f(t) = x \cdot \text{Cov}(X(t), Y) \quad \forall t \in [0, T]. \end{aligned}$$

The optimization problem (3) can be rewritten as

$$\max_{w \in L_2(T)} \frac{\langle \mathcal{U}w, w \rangle}{\langle w, w \rangle},$$

where $\mathcal{U} = C_{XY} \circ C_{YX}$.

Therefore, the solution to Eq. (3) is the eigenfunction of the operator \mathcal{U} associated to its largest eigenvalue λ_1 ,

$$\mathcal{U}w_1 = \lambda_1 w_1,$$

and the first PLS component is defined as

$$t_1 = \int_T X(t)w_1(t)dt.$$

The PLS algorithm is an iterative procedure.

Let $X_0 = X$ and $Y_0 = Y$. For any positive integer h , let X_h and Y_h be the residuals of the linear regressions of X_{h-1} , respectively of Y_{h-1} , with predictor the h -th PLS component t_h , i.e

$$\begin{aligned} X_h(t) &= X_{h-1}(t) - p_h(t)t_h, \quad t \in T, \\ Y_h &= Y_{h-1} - c_h t_h, \end{aligned}$$

where $p_h(t) = (\mathbb{E}(X_{h-1}(t)t_h)/\mathbb{E}(t_h^2))$ and $c_h = (\mathbb{E}(Y_{h-1}t_h)/\mathbb{E}(t_h^2))$.

Then, at the step h , the h -th PLS component is defined as the random variable maximizing the Tucker criterion (3) using the residuals X_{h-1} and Y_{h-1} ,

$$t_h = \int_T X_{h-1}(t)w_h(t)dt, \tag{4}$$

where $w_h(t)$ is the solution of

$$w_h = \arg \max_{w, \|w\|^2=1} \text{Cov}^2 \left(\int_T X_{h-1}(t)w(t)dt, Y_{h-1} \right), \tag{5}$$

given by the largest eigenvalue of $\mathcal{U}_{h-1} = C_{XY}^{h-1} \circ C_{YX}^{h-1}$. That is,

$$\mathcal{U}_{h-1}(w_h) = \lambda_h w_h, \tag{6}$$

with C_{XY}^{h-1} and C_{YX}^{h-1} being the cross-covariance operators of $X_{h-1}(t)$ and Y_{h-1} , respectively.

The properties of the PLS components are summarized by the following proposition [10].

Proposition 1. For any $h \geq 1$

- a) $\{t_h\}_{h \geq 1}$ forms an orthogonal system in $L(X)$, the linear space spanned by $\{X(t) : t \in T\}$,
- b) $Y = c_1 t_1 + c_2 t_2 + \dots + c_h t_h + Y_h$,
- c) $X(t) = p_1(t)t_1 + p_2(t)t_2 + \dots + p_h(t)t_h + X_h(t), \forall t \in T$,

- d) $\mathbb{E}(Y_h t_j) = 0, \quad \forall j = 1, \dots, h,$
 e) $\mathbb{E}(X_h(t) t_j) = 0, \quad \forall t \in T, \forall j = 1, \dots, h.$

The PLS approximation of Y at the step h is then given by

$$\hat{Y}^h = c_1 t_1 + c_2 t_2 + \dots + c_h t_h. \quad (7)$$

Notice that the expression of the PLS components defined by Eq. (4) can be rewritten as elements of the linear space spanned by $\{X(t) : t \in T\}$, i.e.

$$t_i = \int_T v_i(t) X(t) dt,$$

with $v_i \in \text{span}\{w_1, \dots, w_i\}$, $i = 1, \dots, h$. Thus, the PLS linear approximation at the step h becomes

$$\hat{Y}^h = c_1 \int_T v_1(t) X(t) dt + \dots + c_h \int_T v_h(t) X(t) dt = \int_T \beta^{PLS,h}(t) X(t) dt, \quad (8)$$

where $\beta^{PLS,h}$ is the approximation of the slope parameter function β in Eq. (1) provided by the functional PLS approach with h components.

2.2. Basis expansion estimation

In practice, the functional predictor X is often observed in a finite set of time points $\{t_k : k = 1, \dots, m\}$ and data is represented by a finite vector $\{X(t_k) : k = 1, \dots, m\}$. Notice that the points t_k and m could be random. Thus, first step in FDA is often the reconstruction of the functional form of data from discrete observations. The most common approach to this problem is to consider that sample paths belong to a finite dimension space spanned by a basis of functions [see, for example, 23].

Let $\{\phi_1, \dots, \phi_p\}$, $p \geq 1$, be a p -dimensional basis in $L_2(T)$ and assume that the functional predictor admits the basis expansion

$$X(t) = \sum_{k=1}^p \alpha_k \phi_k(t), \quad (9)$$

with the α_k 's as random coefficients.

Let $\Phi = (\phi_{jk})$ be the $p \times p$ symmetric matrix with entries the inner products between basis functions $(\phi_{jk}) = \langle \phi_j, \phi_k \rangle = \int_T \phi_j(t) \phi_k(t) dt$, and let denote by $\Phi^{1/2}$ a square root of the matrix Φ such that $\Phi = (\Phi^{1/2})(\Phi^{1/2})'$. (This square root is computed in practice by using the SVD of the matrix Φ). Let us look closely to the weights functions w_i defining the PLS components. From Eq. (6), the w_h 's have the basis representation

$$w_h(t) = \sum_{k=1}^p w_{hk} \phi_k(t), \quad h \geq 1,$$

where $w_h = (w_{h1}, \dots, w_{hp})'$ is the vector of the expansion coefficients. Therefore, the parameter function β can be expressed in the same basis by

$$\beta(t) = \sum_{k=1}^p \beta_k \phi_k(t).$$

Let us denote by $\alpha = (\alpha_1, \dots, \alpha_p)'$ and $\beta = (\beta_1, \dots, \beta_p)'$ the vectors of expansion coefficients of X and β , respectively.

PLS regression is then equivalent to the finite multivariate PLS regression using as predictor the expansion basis coefficients in the

sense that, at each step of the PLS iteration, one obtains the same prediction [14]. The main results are in the following proposition.

Proposition 2. *The functional PLS regression of Y on $X = \{X(t) : t \in T\}$ is equivalent to the PLS regression of Y on the finite random vector $(\Phi^{1/2})\alpha$ with the usual metric in \mathbb{R}^p . In this sense, at each step h of the PLS algorithm, $1 \leq h \leq p$, we have the same PLS component, and so, the same PLS approximation.*

Then, at each step of the FPLS algorithm, the weight functions w_h defining the FPLS components are obtained, in terms of expansion coefficients, as $w_h = (\Phi^{-1/2})\tilde{w}_h$ where \tilde{w}_h are the eigenvectors associated to the largest eigenvalues of the eigen-problem

$$(\Phi^{1/2})C_{h-1}(\Phi^{1/2})\tilde{w} = \lambda \tilde{w} \quad \tilde{w} \in \mathbb{R}^p, \langle \tilde{w}, \tilde{w} \rangle_{\mathbb{R}^p} = 1,$$

where C_{h-1} is the $p \times p$ matrix with entries

$$C_{h-1}(j, k) = \text{Cov}(\alpha_{h-1,j} Y_{h-1}) \text{Cov}(\alpha_{h-1,k} Y_{h-1}), \quad j, k \in 1 \dots p$$

and $\alpha_{h-1,j}$'s are the expansion coefficients of the residuals X_{h-1} . In other words, the FPLSR is reduced to the finite multivariate PLSR with predictor α , and metric Φ . The weight functions w_h are then given by the coefficients w associated to the largest eigenvalue of the eigen-problem

$$C_{h-1}\Phi w = \lambda w, \quad w \in \mathbb{R}^p, \langle w, \Phi w \rangle_{\mathbb{R}^p} = 1.$$

3. Penalized functional PLSR

The main goal of the present work is to provide smoothed and more accurate estimates of β within the PLS algorithm. For this purpose, one way to do this is to smooth the weights w_i defining the PLS components. Therefore, by Eq. (8), one obtains smoothed estimates of β .

Two different versions of penalized FPLSR were considered at the time of developing this work. The first one makes use of the methodology exposed in [15] for regularized FPCA that introduces the penalty in the norm of the PCA weights. The second one introduces the penalty in the covariance by following the penalized FPCA proposed in [16]. Taking into account that the second penalized approach (FPLS by penalizing the covariance) consistently performed worse than the first in all simulations and applications developed in this paper, the methodological and practical results related with it were omitted from the main article and presented as supplementary material.

Next, a review about the penalty function which will be used in the proposed penalized versions of FPLSR is given.

3.1. Roughness penalty function

The approach developed in this work to obtain smoothed estimates of β , is based on the roughness penalty. In order to quantify the "roughness" of a function w , the derivatives of some order d , $D^d(w) = w^{(d)}$, $d \geq 1$, are used. The first reference about penalties based on such a roughness is introduced in [24] who proposed

$$\text{Pen}_d(w) = \int_T [D^d(w)(t)]^2 dt,$$

with $d = 2$, as a measure of roughness of w . Thus, the flexibility of the fitted curve is restricted and the over-fitting is prevented.

By considering the basis function expansion of $w(t)$ given by

$$w(t) = \sum_{k=1}^p w_k \phi_k(t), \tag{10}$$

the roughness penalty function is given by

$$Pen_d(w) = \int_T [D^d w(t)]^2 dt = w' P_d w, \tag{11}$$

where $w = (w_1, \dots, w_p)'$ is the vector of basis coefficients of $w(t)$ and P_d the matrix of the cross inner products of the d -order derivatives of basis functions.

In [25] the approximation of O'Sullivan was generalized, such that it could be applied in any context where regression on B-splines was useful. They proposed to work with a relatively large number of knots and a penalty based on d -order differences between coefficients of adjacent B-splines. This kind of penalty is known as P-spline. In that paper, the relationship between the two penalties was shown. More recently, [26] concluded from a simulation study that P-splines provide the lowest approximation errors, have less numerical complexity making easier its computational implementation and are quite insensitive to the choice of knots so that it is sufficient to choose a relatively large number of equally spaced basis knots. In this

context, $P_d = (\Delta^d)' \Delta^d$, with Δ^d the matrix of d -order differences between the adjacent basis coefficients.

In the following, the notation P_d is used for both, continuous and discrete penalty matrices.

3.2. FPLS by penalizing the norm

In order to smooth the PLS weight functions, the smoothing is incorporated in the definition of the norm with respect to an inner product which takes into account the roughness.

The FPLS regression by penalizing the norm is defined as a generalization of the Tucker criterion (3) by introducing the following roughness penalty for the weights defining the PLS components:

$$\max_{w \in L_2(T)} \frac{\text{Cov}^2(\int_T X(t)w(t)dt, Y)}{\langle w, w \rangle + \lambda Pen_d(w)}, \tag{12}$$

with Pen_d being the penalty defined in Eq. (11) for some given positive integer d .

Notice that, for any of the two considered penalties (continuous or discrete), this is equivalent to consider a new inner product in the functional space so that the h -th PLS component is given by

$$t_h = \int_T X_{h-1}(t)w_h(t)dt,$$

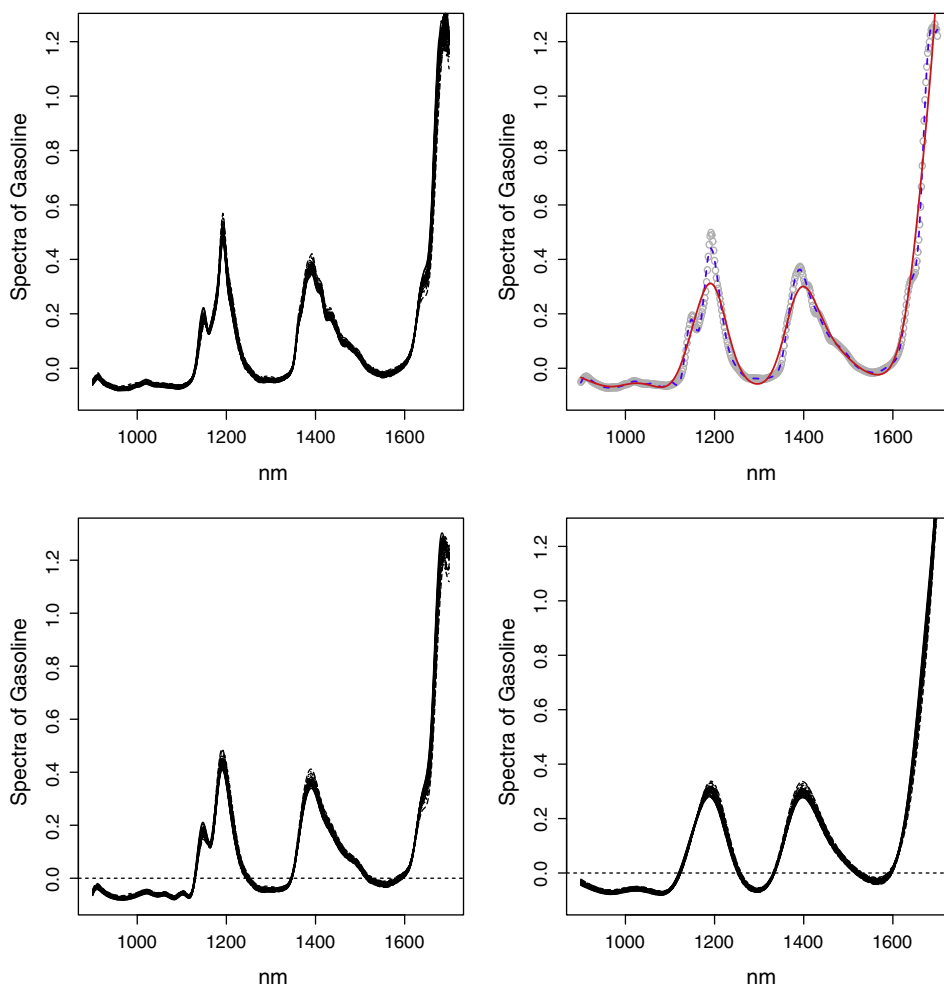


Fig. 1. Simulation study. Spectrometric raw curves of 60 gasoline samples measured in 2-nm intervals from 900 nm to 1700 nm (top left). Discrete observations (circles), regression spline (blue dashed line) and P-spline (red solid line) for one of the sample paths (top right). At the bottom, the regression splines (left) and P-splines (right) for all the sample paths.

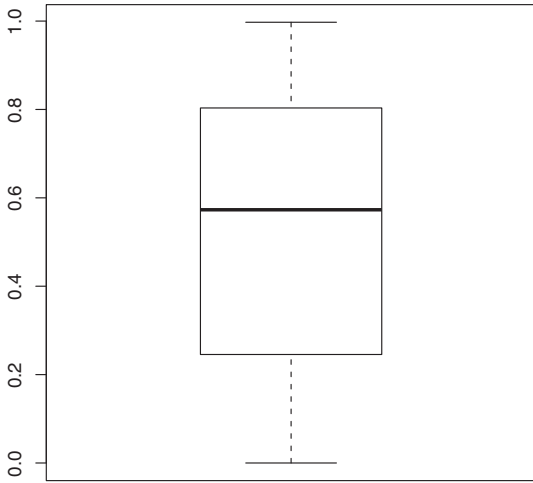


Fig. 2. Simulation study. Box plot of the distribution of correlations between columns of matrix $A\Phi$.

where $w_h(t)$ is obtained by solving the following problem

$$w_h = \arg \max_{w, \|w\|_\lambda^2 = 1} \text{Cov}^2 \left(\int_T X_{h-1}(t)w(t)dt, Y_{h-1} \right), \quad (13)$$

and $\|\cdot\|_\lambda$ is the norm associated to a new inner product defined by

$$\langle f, g \rangle_\lambda = \langle f, g \rangle + \lambda \langle f', P_d g \rangle,$$

with $f = (f_1, \dots, f_p)'$ and $g = (g_1, \dots, g_p)'$ being the vectors of basis coefficients of $f(t)$ and $g(t)$, respectively.

The main result of this section is the following proposition.

Proposition 3. *If X admits the expansion in (9), then, for any $\lambda > 0$, the FPLS regression by penalizing the norm is equivalent with the finite multivariate PLS regression of Y on the random vector of basis coefficients α using the metric*

$$M = (L^{-1}\Phi)'L^{-1}\Phi,$$

where L is such that $LL' = \Phi + \lambda P_d$. Then, at each step h of the penalized FPLS algorithm, the weights functions w_h defining the penalized FPLS

components are obtained, in terms of expansion coefficients, as $w_h = (L^{-1})'L^{-1}\Phi u_h$, where u_h are the eigenvectors associated to the largest eigenvalues of the eigen-problem

$$C_{h-1}Mu = \lambda u, \quad u \in \mathbb{R}^p, \langle u, Mu \rangle_{\mathbb{R}^p} = 1.$$

Proof. Let us consider now the basis expansions of $X(t)$ and $w(t)$ given by Eqs. (9) and (10), respectively. Then, $\langle w, w \rangle_\lambda = w'\Phi w + \lambda w'P_d w$, with P_d being the penalty matrix, and $w = (w_1, \dots, w_p)'$ the vector of basis coefficients of $w(t)$. It is easy to convert the initial problem (12) into a new problem given by

$$\max_w \frac{\langle \mathcal{L}w, w \rangle}{w'\Phi w + \lambda w'P_d w} = \max_w \frac{w'\Phi C\Phi'w}{w'(\Phi + \lambda P_d)w}, \quad (14)$$

where C is the $p \times p$ matrix with entries

$$C(j, k) = \text{Cov}(\alpha_j Y) \text{Cov}(\alpha_k Y), \quad j, k \in 1 \dots p.$$

By assuming the decomposition $LL' = \Phi + \lambda P_d (L = (\Phi + \lambda P_d)^{1/2})$, the maximization problem (14) is equivalent to

$$\max_w \frac{w'\Phi C\Phi'w}{w'(LL')w}.$$

Defining $\tilde{w} = L'w (w = (L^{-1})'\tilde{w})$ the problem is reduced to

$$\max_w \frac{\tilde{w}'(L^{-1})\Phi C\Phi'(L^{-1})'\tilde{w}}{\tilde{w}'\tilde{w}}.$$

By analogy with the non-penalized FPLS, the associated eigen-problem is

$$(L^{-1})\Phi C\Phi'(L^{-1})'\tilde{w} = \lambda \tilde{w},$$

so that the weight function associated with the first penalized PLS component t_1 is given by $w_1 = (L^{-1})'\tilde{w}_1$, with \tilde{w}_1 being the eigenvector associated with the largest eigenvalue of that problem.

The first PLS step is completed by ordinary linear regression of $X_0(t) = X(t)$ and $Y_0 = Y$ on t_1 , and denoting by $X_1(t)$ and Y_1 the corresponding residuals.

In general, the weight function that defines the h -th PLS component ($h > 1$) is obtained by solving the maximization problem (13). Reasoning as above, it is shown that $w_h = (L^{-1})'\tilde{w}_h$ with \tilde{w}_h being the

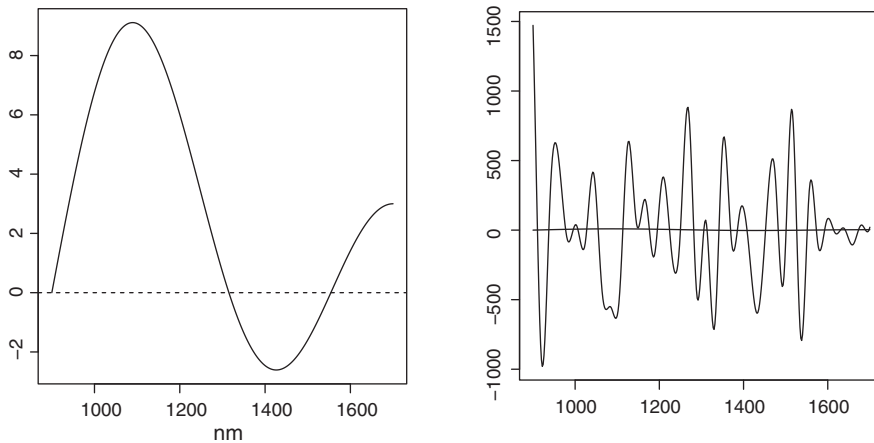


Fig. 3. Simulation study. Parameter function used for simulating the response variable of the functional linear model (left) and the estimation provided by the functional linear model in terms of regression spline approximation of the spectrometric curves (right).

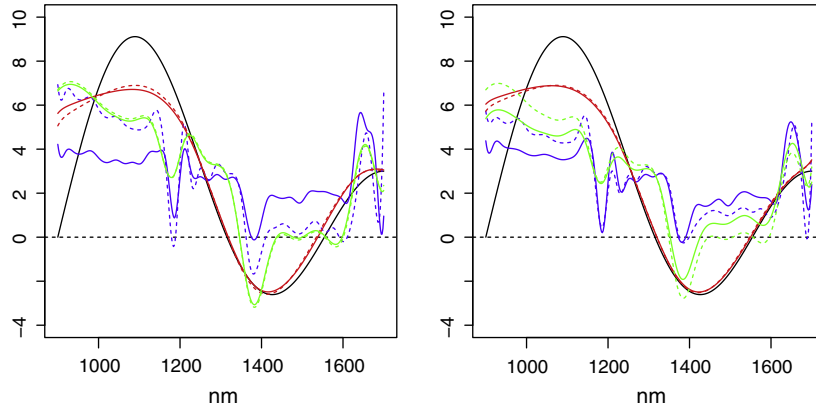


Fig. 4. Simulation study. Case I (left) and Case II (right). Simulated parameter function (black), mean of the 100 parameter functions estimated by Methods I (blue dotted line), II (green dashed line) and III (red solid line), using Criterion 1 (solid line) and Criterion 2 (dashed line) to select the number of PLS components.

eigenfunction associated with the largest eigenvalue in the following eigen-problem:

$$(L^{-1})\Phi C_{h-1}\Phi'(L^{-1})\tilde{w} = \lambda\tilde{w}. \tag{15}$$

The h -th PLS step is concluded with the linear regression of $X_{h-1}(t)$ and Y_{h-1} on t_h , and the corresponding residuals $X_h(t)$ and Y_h .

From Eq. (15) it can be concluded that FPLS by penalizing the norm is equivalent to multivariate PLS of Y on the random vector $L^{-1}\Phi\alpha$, with the PLS components given by

$$t_h = \tilde{w}'L^{-1}\Phi\alpha = w'\Phi\alpha.$$

Defining $u = (L^{-1}\Phi)^{-1}\tilde{w}$, ($\tilde{w} = L^{-1}\Phi u$) the eigen-problem (15) is equivalent to

$$C_{h-1}\Phi'(L^{-1})\Phi u = \lambda u \quad u \in \mathbb{R}^p, u'\Phi'(L^{-1})\Phi u = 1.$$

This means that the FPLSR by penalizing the norm is equivalent to the finite multivariate PLS regression of Y on the random vector $L^{-1}\Phi\alpha$ with the usual metric in \mathbb{R}^p . Then, the weights functions w_h defining the penalized FPLS components are obtained, in terms of expansion coefficients, as $w_h = (L^{-1})\tilde{w}_h$ where \tilde{w}_h are the eigenvectors associated to the largest eigenvalues of the eigen-problem

$$(L^{-1})\Phi C_{h-1}\Phi'(L^{-1})\tilde{w} = \lambda\tilde{w} \quad \tilde{w} \in \mathbb{R}^p, (\tilde{w}, \tilde{w})_{\mathbb{R}^p} = 1.$$

Remarks.

- Proposition 3 shows that by penalizing the norm, the functional framework is reduced to a classical multivariate PLS one. Algorithms and software devoted to the finite multivariate framework can then be employed to estimate the functional model (for example, the pls package in the R software).
- Proposition 3 assumes that the value of the smoothing parameter λ is constant over all the steps of the PLS algorithm. As the number of steps of the PLS algorithm, the optimal value of λ can be obtained by cross validation.
- The penalty introduced in the estimation process is given by the matrix P_d . In Section 4, we provide a simulation study using $d = 2$ and P_d defined by P-spline approaches in terms of cubic B-spline bases. The advantage of the smoothed approaches based on P-spline penalties with respect to the ones based on penalizing the integrated squared d -order derivatives is that they are mathematically simpler because the difference matrix is easier to compute than the matrix of integrals of products of

d -order derivatives between B-spline basis functions [see 27, for a detailed study with functional PCA].

3.3. Sample estimation

The estimation of the different PLS functional regression models – without and with penalty – is developed in this section.

Let $\{x_i(t) : t \in T, i = 1, \dots, n\}$ be a random sample of curves drawn from the distribution of the functional variable X . Assume that x_i is observed in the knots $\{t_{ik}, k = 1, \dots, m_i\}$, and denote by x_i the vector of the observed values for the sample curve $x_i(t)$,

$$x_i = (x_i(t_{i1}), \dots, x_i(t_{im_i}))', \quad i = 1, \dots, n.$$

Let $\{y_1, y_2, \dots, y_n\}$ be a random sample of Y associated with the sample $\{x_i, i = 1, \dots, n\}$ (i.e. the pairs (x_i, y_i) are drawn from the joint distribution of (X, Y) , for all $i = 1, \dots, n$).

The functional linear model assumes

$$y_i = \beta_0 + \int_T x_i(t)\beta(t)dt + \varepsilon_i,$$

where $\{\varepsilon_i : i = 1, \dots, n\}$ are independent and zero-mean random errors.

Let $\{\phi_1, \dots, \phi_p\}$, $p \geq 1$, be a p -dimensional basis in $L_2(T)$ and assume that the functional predictor X admits an expansion in that basis. Then, the estimation procedure of the parameter function $\beta(t)$, based on Propositions 2–3 consists of the following steps:

1. Compute the basis expansion approximation of $\{x_i(t), i = 1, \dots, n\}$. The vector of basis coefficients of the i -th sample path is estimated by the least squares criterion in terms of a B-spline basis, so that $\hat{\alpha}_i = (\Psi_i'\Psi_i)^{-1}\Psi_i'x_i$, with $\Psi_i = (\phi_j(t_{ik}))_{m_i \times p}$. Let us denote with A the matrix of size $n \times p$ with rows the expansion coefficients of $x_i, i = 1, \dots, n$.
2. Compute the first $h(h \geq 1)$ PLS components. The matrix of PLS components T^h for each method is given by
 - Non-penalized FPLS: $T^h = A\Phi^{1/2}V^h$
 - FPLS by penalizing the norm: $T^h = A\Phi(L^{-1})V^h$ with $LL' = \Phi + \lambda P_d$.

where V^h is the matrix comprising the columns of the first h eigenvectors $\tilde{w}_1, \dots, \tilde{w}_h$ associated with the t_1, \dots, t_h PLS components of each considered method (see Propositions 2–3).

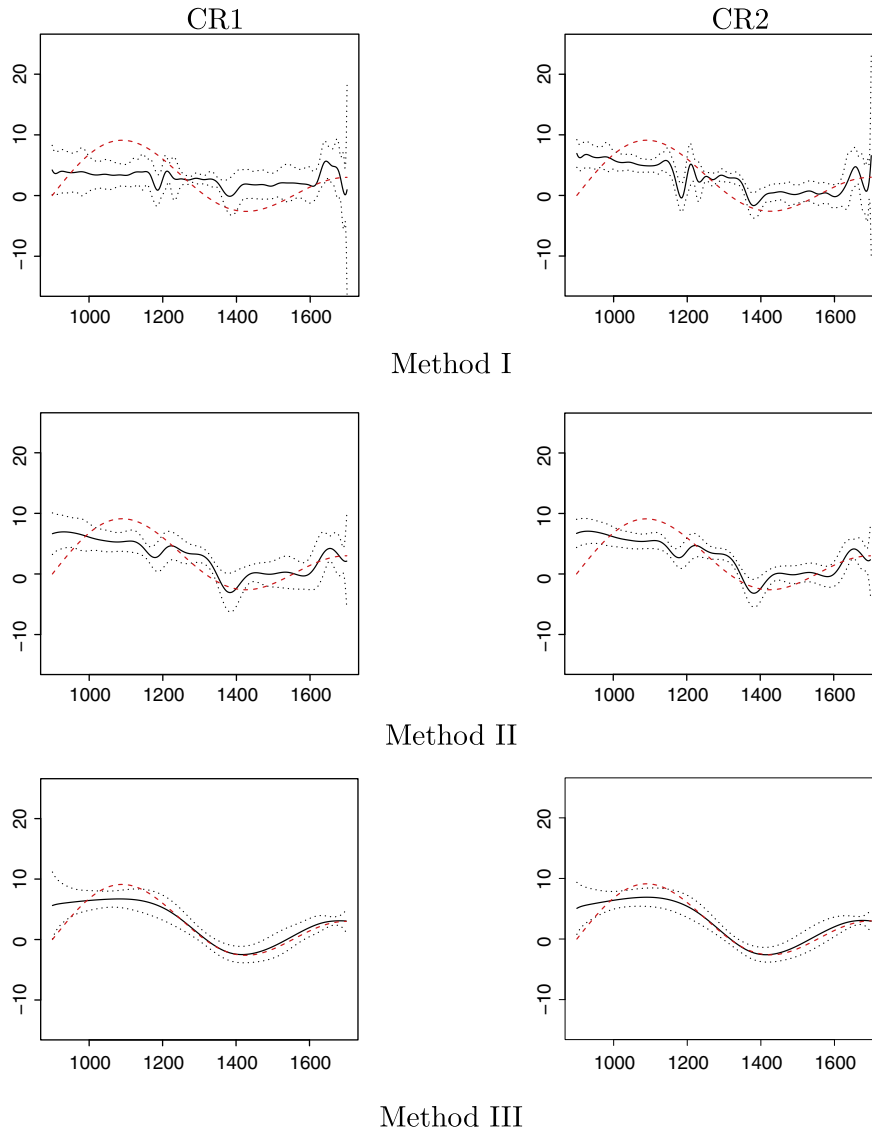


Fig. 5. Simulation study. Case I. CR1 (left) and CR2 (right). Simulated parameter function (red dashed line) and the mean of the 100 parameter functions estimated by Methods I, II and III (black solid line) next to confidence bands (black dashed line) computed as ± 2 times the standard deviation at each time. The number of PLS components was selected by Criterion 1 (left) and Criterion 2 (right).

3. The estimated functional linear model of Y in terms of the first h PLS components is given by

$$\hat{Y}^h = \mathbf{1}\hat{\gamma}_0^h + T^h\hat{\gamma}^h = \mathbf{1}\hat{\gamma}_0^h + A\Phi\hat{\beta}^h,$$

where T^h is the matrix whose columns are the first h PLS components, $\hat{\gamma}^h$ is the vector of the regression coefficients of Y on T^h , and $\hat{\beta}^h$ the vector of basis coefficients of the estimated parameter function

$$\hat{\beta}^h(t) = \sum_{j=1}^p \hat{\beta}_j^h \phi_j(t).$$

Then, the vector $\hat{\beta}^h$ of coefficients estimated by each method is given by

- Non-penalized FPLS: $\hat{\beta}^h = (\Phi^{-1/2})^T V^h \hat{\gamma}^h$

- FPLS by penalizing the norm: $\hat{\beta}^h = (L^{-1})^T V^h \hat{\gamma}^h$.

3.4. Model selection

In order to select the optimum number h of PLS components and the smoothing parameter λ , two different criteria have been considered.

- Criterion 1: choosing simultaneously the values of h and λ by the K -fold cross-validation criterion. The sample data are divided into K segments of equal size, $S_k (k = 1, \dots, K)$, using each one as the validation data for testing the model and the remaining $K - 1$ segments as training sample. The K -fold cross-validation error is given by

$$K - FCVE(h, \lambda) = \frac{1}{n} \sum_{k=1}^K E_k(h, \lambda),$$

where $E_k(h, \lambda) = \sum_{i \in S_k} (y_i - \hat{y}_i(h, \lambda))^2$, with $\hat{y}_i(h, \lambda)$ being the predictions on the k -th segment S_k , from a model fitted with the first h PLS components and the smoothing parameter λ . In practice we have considered $K = 10$.

- Criterion 2: for each fixed number of PLS components h , choosing the value of λ that minimizes the $K-FCVE(h, \lambda)$. Then, the number of PLS components h is selected by minimizing the integrated mean squared error of the parameter function

$$IMSE\beta(h) = \left(\frac{1}{T} \int_T (\beta(t) - \hat{\beta}^h(t))^2 dt \right)^{1/2},$$

where $\hat{\beta}^h(t)$ is the parameter function estimated with h PLS components and the smoothing parameter λ previously selected by K -fold cross-validation. This criterion can be computed only in simulations where the parameter function is known.

For non-penalized FPLS both criteria are reduced to select only the number of PLS components.

4. Simulation study

The ability of the proposed penalized FPLS approaches to predict the response and to provide an accurate estimation of the functional parameter is tested and compared in this section with simulated data.

4.1. Description

The simulation study developed in this paper is based on the spectroscopic data set of gasoline described by [28]. The gasoline data set consists of the NIR spectra of 60 gasoline samples measured in 2-nm intervals from 900 nm to 1700 nm (400 discrete observations for each sample curve). The NIR spectra of these gasoline samples are shown in Fig. 1 (left).

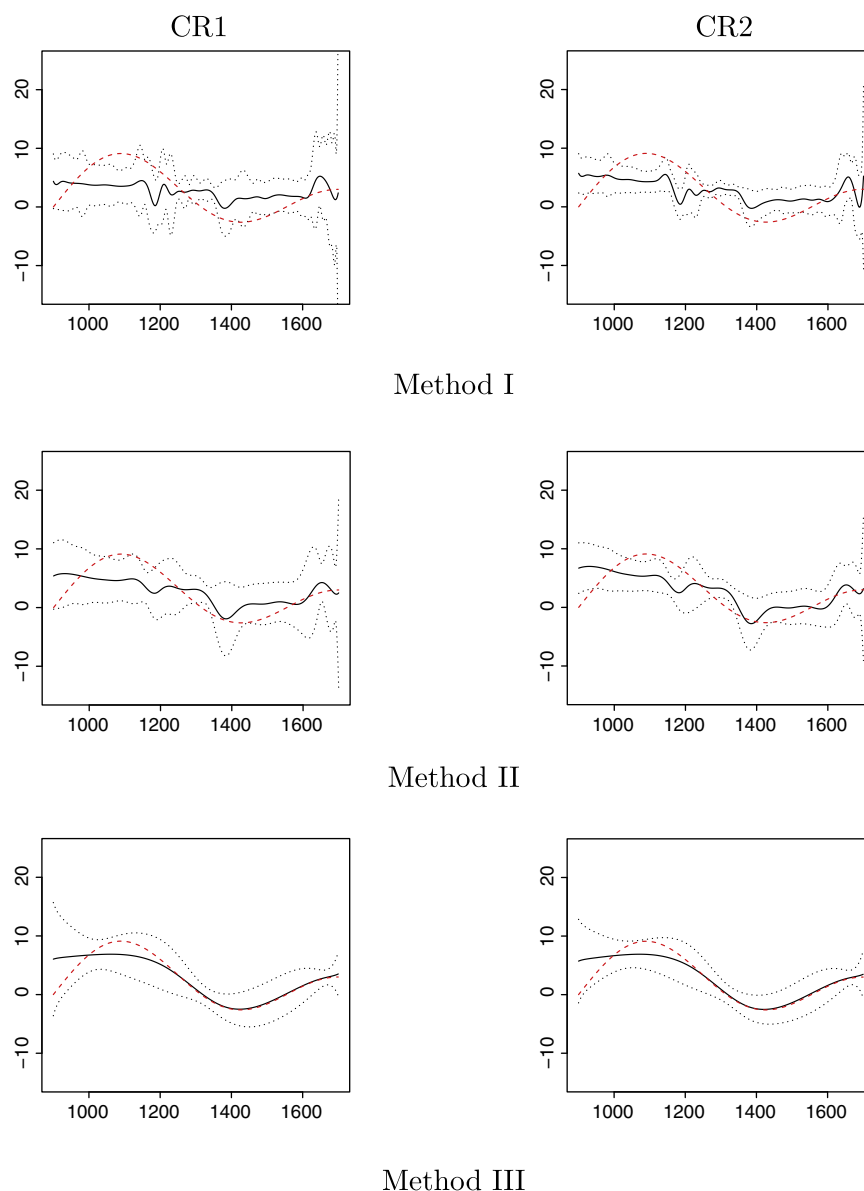


Fig. 6. Simulation study. Case II. CR1 (left) and CR2 (right). Simulated parameter function (red dashed line) and the mean of the 100 parameter functions estimated by Methods I, II and III (black solid line) next to confidence bands (black dashed line) computed as ± 2 times the standard deviation at each time. The number of PLS components was selected by Criterion 1 (left) and Criterion 2 (right).

The parameter function $\beta(t)$ used for simulating the response variable Y of the functional linear model is a relatively smooth function

$$\beta(t) = 2 \sin(0.5\pi t) + 4 \sin(1.5\pi t) + 5 \sin(2.5\pi t), \quad t \in [0, 1],$$

used in [12] and [17] by transforming its domain to the spectra domain (see Fig. 3 (left)).

After least squares approximation of the spectrometric curves and the functional parameter in terms of the cubic B-splines defined

on 40 equally spaced knots in the interval [900, 1700], the response values simulated in this work are given by

$$y_i = \int_{900}^{1700} x_i(t)\beta(t)dt + \varepsilon_i,$$

where ε_i ($i = 1, \dots, n$) are simulated independent random errors with normal distribution. The standard deviation of the errors, σ_ε , is chosen so that the squared multiple correlation coefficient of the true model equals 0.9 (Case I) and 0.7 (Case II). An example of regression spline and P-spline for one of the sample paths is shown in Fig. 1 (top right).

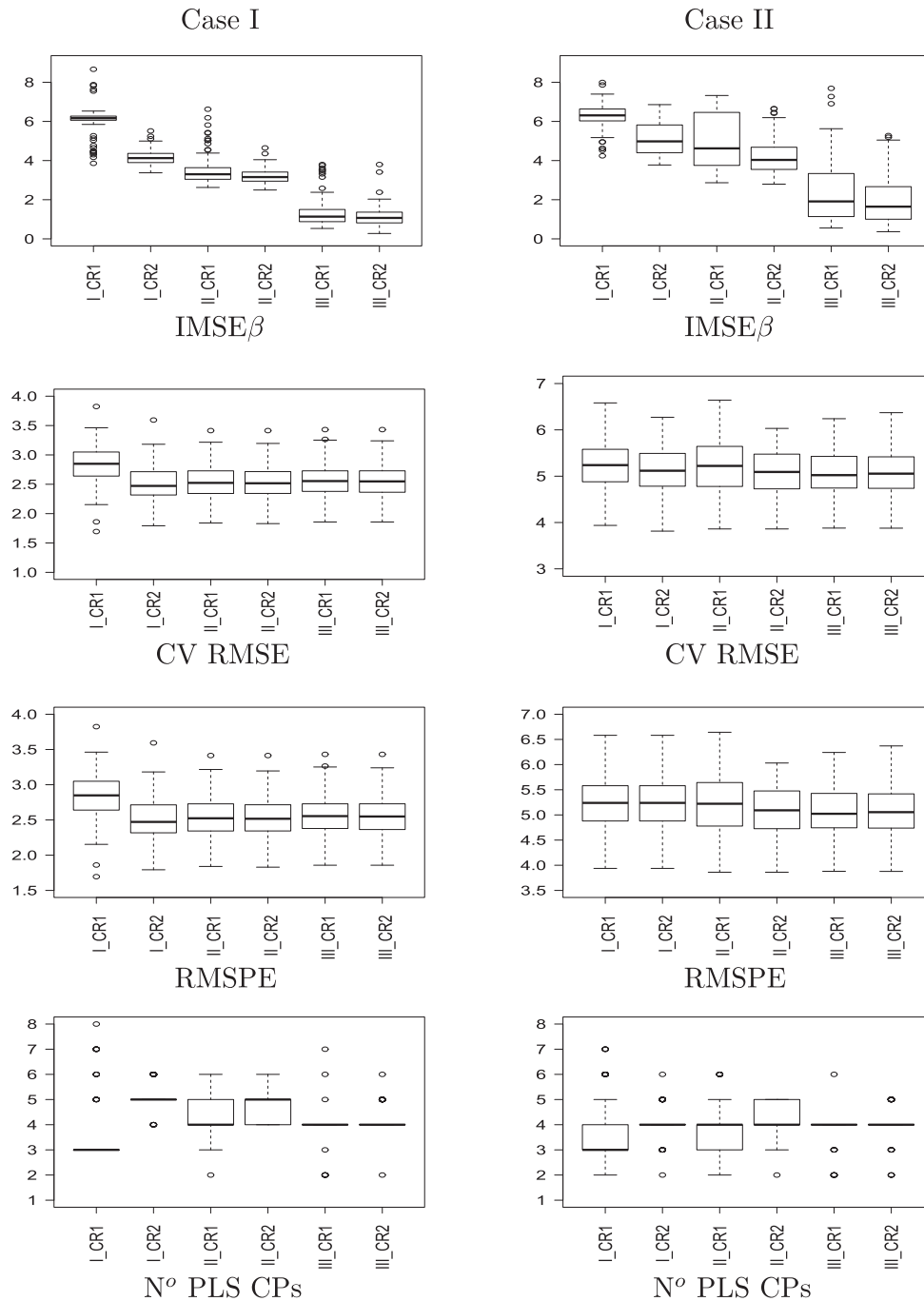


Fig. 7. Simulation study. Case I (left) and Case II (right). Box plots related to the distribution of $IMSE\beta$, CV RMSE, RMSPE number of PLS components for the FPLS regression models estimated by Methods I, II and III, with the number of predictors selected by CR1 and CR2 on 100 simulations.

The simulated response variable in matrix form is given by

$$Y = A\Phi\beta + \varepsilon, \quad (16)$$

with A being the matrix of basis coefficients of the spectrometric curves, β the vector of basis coefficients of the parameter function and Φ the matrix of inner products between the B-spline basis functions. It is well known that an important problem related with the estimation of this linear model is multicollinearity (high correlations between columns of its design matrix). This could produce inaccurate estimates of the functional parameter. The distribution of the correlations between the columns of the design matrix can be seen in Fig. 2. As a consequence, the estimation of the functional parameter from the equivalent linear model (16) is really poor (see Fig. 3 (right) for case I). This inaccurate estimation makes it very difficult to interpret the relationship between the functional predictor and the response variable.

This problem is solved in this paper by applying different dimension reduction approaches based on taking an optimum set of functional PLS components as predictor variables.

The problem of lack of smoothness of the parameter function estimated by non-penalized FPLS regression (Method I) is solved by considering a penalized approach.

First, two non-penalized FPLS regression approaches based on non-penalized FPLSR on regression spline approximation of the sample paths (Method I) and the other is based on non-penalized FPLSR on P-spline smoothing of the sample curves (Method II). In both cases, a cubic B-spline basis with 40 equally spaced knots in the interval [900, 1700] is considered for computing the basis expansion of the spectral curves. A detailed comparative study between the behavior of regression splines, smoothing splines and P-splines in the functional data context can be seen in [26].

Second, a penalized estimation of FPLSR (by penalizing the norm) is introduced in this paper and it will be called Method III. The smoothing parameters associated with the penalized FPLS version are chosen by 10-fold cross-validation (10-FCV). On the other hand, the optimal number of FPLS components in all compared methods was chosen by two different criteria: minimizing the 10-FCV error (Criterion 1 denoted by CR1) and minimizing the integrated mean squared error with respect to the functional parameter (IMSE β) (Criterion 2 denoted by CR2). In order to corroborate the good performance of the penalized FPLS estimations proposed in this paper, 100 repetitions of each simulation scheme (Case I and Case II) were carried out.

4.2. Discussion of results

The response values were simulated by fixing $R^2 = 0.9$ (Case I) and $R^2 = 0.7$ (Case II) for the simulation of the random errors associated with the functional linear model.

The means of the estimated parameter functions over 100 simulations provided by Methods I, II and III, with criteria CR1 and CR2 used for model selection, are presented in Fig. 4. The mean and pointwise confidence bands computed as the sample mean ± 2 times the standard deviation at each time point are shown in Fig. 5 (Case I. CR1 and CR2) and Fig. 6 (Case II. CR1 and CR2).

The box plots for the distribution of the integrated mean squared error with respect to the original parameter function (IMSE β) are drawn in Fig. 7 for the four considered methods and the two model selection criteria. Let us observe that in Case II the variance of errors is higher and then the estimation of the parameter function is a bit less accurate and with higher variability than in Case I. This fact is reflected in the confidence bands for the mean functions that are wider for Case II. Independently of the criterion to be used for model

selection, the penalized Method III provides lower IMSE β than the non-penalized ones (Methods I and II).

According to the results related to the estimation of the parameter function, it can be seen that Method I does not provide a fairly accurate estimation by showing a great lack of smoothness that makes very difficult its interpretation. The parameter functions estimated by Method II are smoother than the ones provided by Method I, but they are noisy and far from the original function. On the other hand, looking at the plots about the mean of the parameter functions, the penalized FPLS version proposed in this paper (Method III) achieves the best parameter functions in the sense of smoothness and variability.

In order to check the forecasting ability of the three considered methods, box plots for the distribution of the CV RMSE and RMSPE (squared root of the mean squared prediction errors) on the 100 simulations are presented in Fig. 7. In general, in Case II, both the CV RMSE and RMSPE are higher than in Case I. And just for the more realistic case (Case II, where $R^2 = 0.7$) the box plots show that Method III provides the minimum RMSPE (in median).

To compare the degree of dimension reduction produced by the two model selection criteria, box plots for the distribution of the selected number of PLS components are drawn in Fig. 7 for the two cases and the three FPLSR approaches considered in this paper. Let us observe that the penalized version of FPLS regression (Method III) selects a balanced number of PLS components, with some outliers which do not exceed six components. Method I presents more variability in the selected number of PLS components. On the other hand, the number of PLS components selected by the Criteria 1 and 2 with Method III is different to the non-penalized procedures. Let us take into account that in real applications the parameter function is unknown and CR2 criterion based on minimizing the IMSE β can not be used. Based on the previous comments, it can be concluded that using K-fold cross-validation criterion for model selection is a good option to predict the response and to estimate the parameter function with the penalized approach.

In order to see numerical differences among the results given by Methods I, II and III, the model selection criteria CR1 and CR2, and the

Table 1

Simulation study. Cases I and II. Sample mean and standard deviation related to the distribution of IMSE β , CV RMSE, RMSPE and number of PLS components for the optimum FPLS regression models estimated by Methods I, II and III, with the number of predictors selected by CR1 and CR2 criteria on 100 simulations.

		$R^2 = 0.9$					
		Criterion 1			Criterion 2		
	Method	I	II	III	I	II	III
IMSE β	Mean	6.08	3.52	1.36	4.16	3.20	1.14
	sd	0.70	0.77	0.76	0.40	0.35	0.53
CV RMSE	Mean	2.82	2.55	2.58	2.53	2.54	2.57
	sd	0.34	0.28	0.29	0.30	0.27	0.28
RMSPE	Mean	2.82	2.55	2.58	2.53	2.54	2.57
	sd	0.34	0.28	0.29	0.30	0.27	0.28
N $^\circ$ CPs	Mean	3.56	4.42	3.95	5.16	4.67	4.23
	sd	1.27	0.74	0.59	0.55	0.49	0.51
		$R^2 = 0.7$					
		Criterion 1			Criterion 2		
	Method	I	II	III	I	II	III
IMSE β	Mean	7.29	5.55	2.49	5.13	4.16	2.00
	sd	4.04	3.30	2.18	0.80	0.82	1.22
CV RMSE	Mean	5.23	5.20	5.09	5.12	5.09	5.09
	sd	0.54	0.56	0.49	0.50	0.50	0.50
RMSPE	Mean	5.23	5.20	5.09	5.23	5.09	5.09
	sd	0.54	0.56	0.49	0.54	0.50	0.50
N $^\circ$ CPs	Mean	3.54	3.74	3.83	4.10	4.24	4.07
	sd	1.32	1.10	0.59	0.59	0.64	0.61

two different R^2 used in the simulation study, Table 1 summarizes the sample mean and the standard deviation of the errors IMSE β , CV RMSE and RMSPE, and the number of PLS components for each of the eight possible combinations. The results in this table corroborate the previous ones given by Figs. 4–7. Summarizing, it can be said that independently of the model selection criterion and the simulation scheme ($R^2 = 0.9$ or $R^2 = 0.7$), we observe that the estimation of the coefficient functions is significantly improved by Method III. With respect to the prediction errors, Methods I, II and III give similar results with Method I-CR1 providing the worse predictions for $R^2 = 0.9$ and Method III-CR1 being the most accurate for $R^2 = 0.7$.

As expected, the significant differences between the non-penalized and penalized estimations of FPLSR are not in their forecasting ability but in their capacity to provide an accurate estimation of the functional parameter.

5. Real data applications

In order to test the good performance of the proposed methods, two applications to different real data sets are considered in the following subsections.

5.1. Gasoline spectra

In this section, the spectroscopic data set of gasoline described by [28] is considered again. Let $\{y_i : i = 1, \dots, 60\}$ be a sample of a scalar response variable related to the octane number of each gasoline sample. The aim is to forecast the octane number from the NIR spectra of 60 gasoline samples, measured in 2-nm intervals from 900 nm to 1700 nm. In order to test the results, the sample was divided into five sets of 12 gasoline samples, as in [17]. Then, each of these subsets is taken as a test sample and the remaining 48 observations as a training sample to fit the model.

First, least squares cubic B-spline smoothing with 40 equally spaced knots of the spectral curves was carried out. After that, the three proposed methods were applied. In this case, the parameter function is unknown, and then the number of PLS components for each method was chosen by the leave-one-out cross-validation criterion. The means of the parameter functions estimated by the three proposed methods are overlaid in Fig. 8.

The non-penalized FPLS coefficient functions are quite difficult to interpret because of their high local variation. The coefficient function provided by Method III is more useful not only for prediction but also for interpretation. Let us observe that the weights are positive and significantly different from zero for wavelengths in the interval (1300,1450) and negative in the interval (1150,1300), approximately, with the highest positive weight around 1350–1375 and the smallest negative weight around 1200–1225. This means that high values of the spectrum in the interval (1150,1300) are associated with gasoline of less octane number while high values in the interval (1300,1450) are related to high level of octane number. In other words, a gasoline with high octane must have less value of the spectrum in the wavelength range (1150,1300) and higher in the range (1300,1450).

In order to check the forecasting performance of the different proposed methods, the squared root of the mean squared prediction errors (RMSPE) was computed on the five test samples. The sample mean and the standard deviation of the RMSPEs are summarized in Table 2. The mean and the standard deviation of the squared multiple correlation coefficients associated with the fitted models for each method appear also in this table. Let us observe that (as in simulation studies) the estimation of the functional parameter given by Method III is smoother and more interpretable with similar prediction errors than Method II and better than Method I.

Let us observe that an accurate estimation of the parameter function could be useful not only for getting nice interpretations and predictions but also for designing efficient selection methods

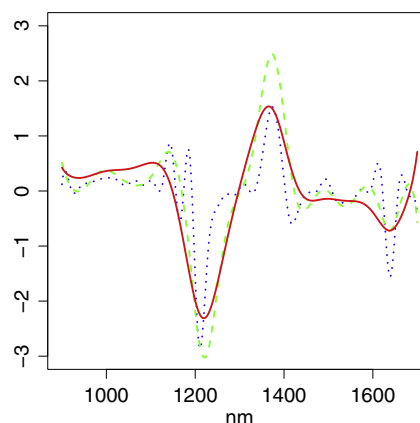


Fig. 8. Application to gasoline data. Mean of the parameter functions estimated by Methods I (blue dotted line), II (green dashed line) and III (red solid line), for 5 training samples.

of explanatory variables (see, for example, [29]). A comparison of selection methods of explanatory variables in multiple PLS regression with application to manufacturing process data was developed in [30]. Some of these methods are based on eliminating those explanatory variables presenting the smallest non-significant PLS regression coefficients, and then recalculating a new PLS model on the remaining covariates. Following this idea, an algorithm to select the significant wavelength domain to predict the response could be developed for the FPLS regression model. In fact, in the application with gasoline data, if we eliminate the nonsignificant wavelet intervals (900,1150) and (1450,1700) and estimate the FPLSR model taking as explanatory functional variable only the spectra in the interval (1150,1450) then we obtain a much less mean (RMSPE) given by 0.1491 with standard deviation 0.0214.

5.2. Beer spectra

In this section, let us consider an original data set related to the spectra of different types of beers. These data were proposed in the competition “Challenge 2010” during the “CHIMOMETRIE 2010 Symposium” organized by the French Chemometrics Society. Data can be downloaded from <http://www.chimometrie.fr/challenge2010.html>. Originally, this data was presented as a classification problem, so that the best approaches proposed by the participants and organizers were summarized in [31].

In order to apply FPLR a scalar response variable is required. So, given the type of beer associated with each spectrum, the chemical composition of alcohol (%) of each sample has been observed. Let $\{y_i : i = 1, \dots, 200\}$ be a sample of a scalar response variable related to the alcohol content of each beer sample (ranging from 5 to 11.3%). The aim is to forecast the degree of alcohol from the NIR spectra of 200 beer samples, measured in 2-nm intervals from 400 nm to 2498 nm. The 200 spectra were divided into training sample (120 spectra) to fit the model and a test sample of 80 spectra. A cubic B-spline basis defined on 40 equally spaced knots was considered to approximate the functional data. As in the other application, the

Table 2
Application to gasoline data. Mean and standard deviation of the RMSPE on test samples and the R^2 coefficients estimated by Methods I, II and III.

	Mean (RMSPE)	sd (RMSPE)	Mean (R^2)	sd (R^2)
Method I	0.2680	0.0758	0.9780	0.0028
Method II	0.2224	0.0355	0.9868	0.0012
Method III	0.2200	0.1432	0.9055	0.0638

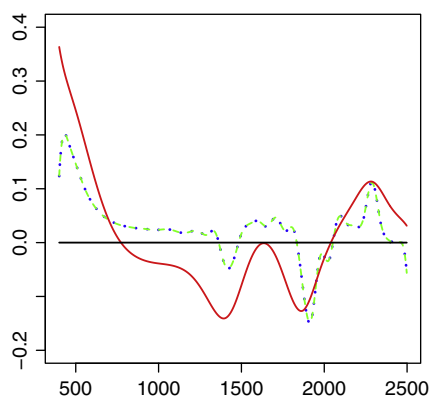


Fig. 9. Application to beer data. Parameter functions estimated by Methods I (blue dotted line), II (green dashed line) and III (red solid line).

number of PLS components was selected by means of leave-one-out cross-validation criterion.

In Fig. 9 the coefficient functions estimated by the three proposed methods are shown. There are no differences between the two non-penalized methods (Methods I and II), meanwhile Method III provides a function easier to interpret in the sense that clearly identifies periods in which the function is positive (from 400 nm to 772 nm – period A; and from 2040 nm to 2498 nm – period C) and negative (from 774 nm to 2038 nm – period B). This means that high values of the spectrum in the period B are associated with beers of less alcohol content, meanwhile high values of the spectrum in the periods A and C are related to beers with high level of alcohol.

In Table 3 the squared root of the mean prediction error on the test sample (RMSPE) and the determination coefficient (R^2) were also provided. In this application the differences between non-penalized and penalized methods are more clear, being Method III which achieves the smallest prediction error and lightly increases the proportion of variability of the response explained by the model.

6. Conclusions

The aim of this paper is to improve the estimation of the functional parameter associated with the functional linear model for a scalar response.

In order to solve the problems of high dimension and multicollinearity in the estimation of the functional linear model, and also to provide an accurate estimation of the functional parameter by controlling its degree of smoothness, two different penalized approaches based on functional partial least squares regression (FPLS) are developed. The first approach introduces the penalty in the definition of the norm of the PLS component weight functions (Method III). The second one considers a penalized estimation of the covariance between the response and the PLS components (Method IV that is presented only in the supplementary material because it does not improve the results provided by Method III). Discrete and continuous penalties can be used in terms of basis expansions of the sample curves.

Table 3
Application to beer data. RMSPE on test sample and the R^2 coefficients estimated for Methods I, II and III.

	RMSPE	R^2
Method I	1.2862	0.6574
Method II	1.2867	0.6569
Method III	1.0443	0.6986

Two different criteria based on minimizing the K-fold cross-validation errors and the integrated mean squared error of the parameter function (Criteria 1 and 2, respectively) were adapted to select the different parameters (the smoothing parameter and the number of PLS components) associated with the considered approaches.

The performance of the penalized FPLSR approach was tested and compared with non-penalized FPLSR on non-penalized and penalized least squares approximation of the sample curves with cubic B-spline basis (Methods I and II, respectively). A simulation study and two applications with chemometric functional data, measuring the NIR spectra of gasoline and beer samples, were developed. In the simulation study two different schemes were considered so that $R^2 = 0.9$ and $R^2 = 0.7$.

From the simulation study, it can be concluded that independently of the model selection criterion and the simulation scheme ($R^2 = 0.9$ or $R^2 = 0.7$), the most accurate estimation of the functional parameter is given by Method III (penalizing the norm).

With respect to the forecasting performance, the penalized and non-penalized approaches provided similar quality for the prediction errors with the penalized approach (Method III) giving slightly smaller errors (in median) for the more realistic case of $R^2 = 0.7$ and the selection model criteria valid in practice (CR1), and Method I-CR1 providing the worse predictions for $R^2 = 0.9$.

Then, the significant differences between the non-penalized and penalized estimations of FPLS regression are mainly in their capacity to provide an accurate estimation of the functional parameter, which is useful not only for obtaining accurate predictions but also for interpreting the relationship between the functional predictor and the scalar response. On the other hand, after comparing with criterion based on minimizing the IMSE with respect to the true parameter function, we can say that K-fold cross-validation is a good criterion for model selection for a prediction and estimation point of view.

In the applications with the spectroscopic data sets of gasoline and beer, the aim was to forecast a related response variable from the NIR spectra, and to get an accurate estimation of the functional parameter that explains the relationship between the response and the functional predictor. The results of this applications corroborate that the estimation of the coefficient function is significantly improved by Method III. On the other hand, the forecasting performance of the penalized and non-penalized approaches is quite similar with Method III giving slightly smaller prediction error in the second application. Based on this accurate estimation, an interpretation of the coefficient regression functions was provided.

Summarizing, in both, simulation and application studies, the penalized FPLSR approach introduced in this paper presents the best performance. Finally, it is important to highlight that an accurate estimation of the parameter function could be useful not only for getting nice interpretations and predictions but also for designing efficient selection methods of explanatory variables. Based on this idea, an algorithm to select the significant wavelength domain to predict the response could be developed for the FPLS regression model.

Conflict of interests

There is no conflict of interest.

Acknowledgments

This research has been funded by project P11-FQM-8068 from Consejería de Innovación, Ciencia y Empresa. Junta de Andalucía, Spain and the projects MTM2013-47929-P and MTM 2011-28285-C02-C2 from Secretaría de Estado Investigación, Desarrollo e Innovación, Ministerio de Economía y Competitividad, Spain, and by Fondo Europeo de Desarrollo Regional (FEDER).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.10.1016/j.chemolab.2016.03.013>.

References

- [1] H. Cardot, F. Ferraty, P. Sarda, Functional linear model, *Stat. Probab. Lett.* 45 (1) (1999) 11–22.
- [2] F. Yao, H.G. Müller, J.L. Wang, Functional linear regression analysis for longitudinal data, *Ann. Stat.* 33 (6) (2005) 2873–2903.
- [3] A.M. Aguilera, M. Escabias, F.A. Ocaña, M.J. Valderrama, Functional wavelet-based modelling of dependence between lupus and stress, *Methodol. Comput. Appl. Probab.* 17 (2015) 1015–1028.
- [4] A.M. Aguilera, F.A. Ocaña, M.J. Valderrama, An approximated principal component prediction model for continuous-time stochastic processes, *Appl. Stoch. Model. Data Anal.* 13 (2) (1997) 61–72.
- [5] A.M. Aguilera, F.A. Ocaña, M.J. Valderrama, Forecasting with unequally spaced data by a functional principal component approach, *Test* 8 (1) (1999) 233–254.
- [6] J.M. Chiou, H.G. Müller, J.L. Wang, Functional response models, *Stat. Sin.* 14 (3) (2004) 659–677.
- [7] A. Cuevas, M. Febrero, R. Fraiman, Linear functional regression: The case of fixed design and functional response, *Can. J. Stat.* 30 (2) (2002) 285–300.
- [8] J.C. Deville, Analyse et prevision des series chronologique multiples non stationnaires, *Statistique et analyse des donnees* 3 (3) (1978) 19–29.
- [9] G. Saporta, Méthodes exploratoires d'analyse de données temporelles, *Cahiers du B.U.R.O., Université Pierre et Marie Curie* 1981, 37–38.
- [10] C. Preda, G. Saporta, PLS Regression on a stochastic process, *Comput. Stat. Data Anal.* 48 (1) (2005) 149–158.
- [11] C. Preda, G. Saporta, C. Lévêder, PLS Classification for functional data, *Comput. Stat.* 22 (2) (2007) 223–235.
- [12] H. Cardot, F. Ferraty, P. Sarda, Spline estimators for the functional linear model, *Stat. Sin.* 13 (3) (2003) 571–591.
- [13] S. de Jong, PLS Fits closer than PCR, *Chemometrics* 7 (6) (1993) 551–557.
- [14] A.M. Aguilera, M. Escabias, C. Preda, G. Saporta, Using basis expansion for estimating functional PLS regression. Applications with chemometric data, *Chemom. Intell. Lab. Syst.* 104 (2) (2010) 289–305.
- [15] B.W. Silverman, Smoothed functional principal component analysis by choice of norm, *Ann. Stat.* 24 (1) (1996) 1–24.
- [16] J.A. Rice, B.W. Silverman, Estimating the mean and covariance structure nonparametrically when the data are curves, *J. R. Stat. Soc. B* 53 (1) (1991) 233–527.
- [17] P.T. Reiss, R.T. Ogden, Functional principal component regression and functional partial least squares, *J. Am. Stat. Assoc.* 102 (479) (2007) 984–996.
- [18] N. Krämer, A.L. Boulesteix, G. Tutz, Penalized partial least squares with applications to B-spline transformations and functional data, *Chemom. Intell. Lab. Syst.* 94 (1) (2008) 60–69.
- [19] A. Delaigle, P. Hall, Methodology and theory for partial least squares applied to functional data, *Ann. of Stat.* 40 (1) (2012) 322–352.
- [20] M. Tenenhaus, *La Régression PLS: Théorie et Pratique*, Editions Technip, 1998.
- [21] H. Wold, Nonlinear Iterative Partial Least Squares (NIPALS) Modeling: Some Current Developments, in: P.R. Krishnaiah (Ed.), *Multivariate Analysis II, Proceedings of an International Symposium on Multivariate Analysis held at Wright State University, Dayton, Ohio, June 19–24, 1972, 1973*, pp. 383–407.
- [22] M. Stone, Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression, *J. R. Stat. Soc. B* 52 (2) (1990) 237–269.
- [23] J.O. Ramsay, B.W. Silverman, *Functional Data Analysis*, Second edition ed., Springer-Verlag 2005.
- [24] F. O'Sullivan, A stastical perspective on ill-posed inverse problems, *Stat. Sci.* 1 (1986) 505–527.
- [25] P.H.C. Eilers, B.D. Marx, Flexible smoothing with B-splines and penalties, *Stat. Sci.* 11 (2) (1996) 89–121.
- [26] A.M. Aguilera, M.C. Aguilera-Morillo, Comparative study of different B-spline approaches for functional data, *Math. Comput. Model.* 58 (7–8) (2013) 1568–1579.
- [27] A.M. Aguilera, M.C. Aguilera-Morillo, Penalized PCA approaches for B-spline expansions of smooth functional data, *Appl. Math. Comput.* 219 (14) (2013) 7805–7819.
- [28] J.H. Kalivas, Two data sets of near-infrared spectra, *Chemom. Intell. Lab. Syst.* 37 (2) (1997) 255–259.
- [29] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.* 58 (2001) 109–130.
- [30] J.P. Gauchi, P. Chagnon, Comparison of selection methods of explanatory variables in PLS regression with application to manufacturing process data, *Chemom. Intell. Lab. Syst.* 58 (2001) 171–193.
- [31] J.A.F. Pierna, L. Duponchel, C. Ruckebusch, D. Bertrand, V. Baeten, P. Dardenne, Trappist beer identification by vibrational spectroscopy: a chemometric challenge posed at the “Chimiométrie 2010” congress, *Chemom. Intell. Lab. Syst.* 113 (2012) 2–9.