



HAL
open science

PCR and PLS for Clusterwise Regression on Functional Data

Cristian Preda, Gilbert Saporta

► **To cite this version:**

Cristian Preda, Gilbert Saporta. PCR and PLS for Clusterwise Regression on Functional Data. Brito P., Bertrand P., Cucumel G., De Carvalho F. (Eds.). Selected Contributions in Data Analysis and Classification, Springer, pp.589-598, 2007, 978-3-540-73558-8. hal-01125361

HAL Id: hal-01125361

<https://hal.science/hal-01125361v1>

Submitted on 24 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PCR and PLS for Clusterwise Regression on Functional Data

Cristian Preda¹ and Gilbert Saporta²

¹ Faculté de Médecine, Université de Lille 2

CERIM - Département de Statistique

1, Place de Verdun, 59045 Lille Cedex, France, cpreda@univ-lille2.fr

² Chaire de Statistique Appliquée, CEDRIC, CNAM

292, Rue Saint Martin, 75141 Paris Cedex 03, France, saporta@cnam.fr

Abstract. Clusterwise regression is applied to functional data, using PCR and PLS as regularization methods for the functional linear regression model. We compare these two approaches on simulated data as well as on stock-exchange data.

1 Introduction

Clusterwise linear regression method provides classification of data such that each cluster is generated by some linear regression model. More precisely, if $\{Y, X_1, \dots, X_q\}$, $q \geq 1$, are real-valued random variables, the homogeneity of subjects within a cluster is given not only by similarities of the observed values of these variables but mainly by the proximity of subjects with respect to some linear model. One can consider that data is generated by a mixture of several regression models (DeSarbo and Cron (1988)), Hennig (1999),(2000)), that is, there exists a latent categorical random variable \mathcal{G} , $\mathcal{G} \in \{1, \dots, K\}$, $K \geq 2$, defining the clusters such that for $\forall k \in 1, \dots, K$, $\mathbb{P}(\mathcal{G} = k) \neq 0$ and

$$\mathbb{E}(Y | X_1 = x_1, \dots, X_q = x_q) = \beta_0^k + \beta_1^k x_1 + \dots + \beta_q^k x_q,$$

where $\{\beta_i^k\}_{i=0, \dots, q}$ are the regression coefficients for the cluster defined by $\{\mathcal{G} = k\}$.

The estimation aspects in clusterwise linear regression was addressed firstly by the pioneering works of Bock (1969) and Diday (1976) who propose a piecewise linear regression algorithm as a special case of k -means clustering with a criterion based on the minimization of the squared residuals instead of the classical within-class dispersion. The problem of multicollinearity and overfit under the least squares criterion is the subject of works of Charles (1977) which establish properties and conditions for convergence of the alternating algorithm proposed by Diday(1976) and introduce the ridge regression as a regularization method for the clusterwise procedure. One can also mention the works of Spaeth(1979) which propose an estimation procedure of clusterwise regression models by an exchange algorithm.

These clusterwise algorithms are largely used nowadays but few significant modifications have been done since then (Plaia (2004)). Recent contributions in this area are due mainly to the development of techniques for estimating the linear models within clusters subject to different inconsistency issues : multicollinearity of predictors, number of observations within a cluster smaller than the number of predictors, etc.

In this paper we are interested in clusterwise linear regression when the set of explanatory variables (predictors) are of functional type, i.e., data are functions or curves of some continuous parameter t (usually time or wavelength). A well accepted model for this kind of data is to consider them as paths of a stochastic process $X = \{X_t\}_{t \in T}$ taking values in a Hilbert space H of functions on some set T . For example, a second order stochastic process $X = \{X_t\}_{t \in [0,1]}$, L_2 -continuous with sample paths in $L_2([0,1])$ can be used as model for index stock-exchange evolution during a time period or for the knee flexion angle measure over a complete gait cycle.

There is a rich and recent literature devoted to functional data, the last contributions being reported by the monographs of Ferraty and Vieu (2006), Ramsay and Silverman (1997, 2002). As an alternative to the work of Abraham et al. (2002) on unsupervised classification of functional data, Preda and Saporta (2005b) proposed the PLS approach for clusterwise regression on functional data.

We propose a comparative study of the partial least squares (PLS) and the regression on principal components (PCR) approaches for estimating coefficient regression functions within clusters in the context of clusterwise linear regression with predictors of functional type. The paper is divided into three parts. After a brief introduction to PCR and PLS regularization methods for functional data, we describe the clusterwise linear model using the estimations given by PCR and PLS. In the last section we present a simulation study as well as an application on stock exchange data.

2 PCR and PLS for functional data

Let us consider the functional data as sample paths of a stochastic process $\mathbf{X} = \{X_t\}_{t \in [0,T]}$ with continuous time, and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)$, $p \geq 1$, a random vector defined on the same probability space as \mathbf{X} , (Ω, \mathcal{A}, P) . We assume that $\{X_t\}_{t \in [0,T]}$ and \mathbf{Y} are of second order, $\{X_t\}_{t \in [0,T]}$ is L_2 -continuous and for any $\omega \in \Omega$, $t \mapsto X_t(\omega)$ is an element of $L_2([0,T])$. Without loss of generality we assume also that $E(X_t) = 0$, $\forall t \in [0,T]$ and $E(Y_i) = 0$, $\forall i = 1, \dots, p$.

The functional linear regression model assumes that

$$\mathbf{Y} = \int_0^T \beta(t) X(t) dt + \varepsilon, \quad (1)$$

where β is a \mathbb{R}^p -valued function on $[0, T]$ and ε is the random error term.

It is well known that the approximation of \mathbf{Y} obtained by the classical linear regression on $\mathbf{X} = \{X_t\}_{t \in [0, T]}$, i.e., $\hat{\mathbf{Y}} = \int_0^T \beta(t) X_t dt$, is such that β is in general a distribution rather than a function of $L_2([0, T])$ (Saporta (1981)). This difficulty appears also in practice because one has generally more predictors than the number of observations, the least squares criterion providing inconsistent estimators (infinite number of solutions). Regression on principal components (PCR) of \mathbf{X} (Deville (1978)) and PLS approach (Preda and Saporta (2005a)) give satisfactory solutions to this problem.

2.1 Linear regression on principal components (PCR)

The principal components of the stochastic process $\mathbf{X} = \{X_t\}_{t \in [0, T]}$ are linear combinations of X_t , $t \in [0, T]$, given by the eigenfunctions of the covariance operator of \mathbf{X} :

$$\xi_i = \int_0^T f_i(t) X_t dt,$$

where $\{f_i\}_{i \geq 1}$ are solution of the eigenvalue equation

$$\int_0^T C(t, s) f_i(s) ds = \lambda_i f_i(t),$$

and $C(t, s) = \text{cov}(X_t, X_s)$, $\forall t, s \in [0, T]$.

Observe that the principal components $\{\xi_i\}_{i \geq 1}$ are eigenvectors of the Escoufier operator, \mathbf{W}^X , defined by

$$\mathbf{W}^X Z = \int_0^T E(X_t Z) X_t dt,$$

for any real-random variable Z in $L_2(\Omega)$ (Escoufier (1970)).

As in the classical setting, the process $\{X_t\}_{t \in [0, T]}$ and the set of its principal components, $\{\xi_k\}_{k \geq 1}$, span the same linear space. Thus, the regression of \mathbf{Y} on \mathbf{X} is equivalent to the regression on $\{\xi_k\}_{k \geq 1}$ and we have

$$\hat{\mathbf{Y}} = \sum_{k \geq 1} \frac{E(\mathbf{Y} \xi_k)}{\lambda_k} \xi_k.$$

In practice one has to choose an approximation of order q , $q \geq 1$:

$$\hat{\mathbf{Y}}_{PCR(q)} = \sum_{k=1}^q \frac{E(\mathbf{Y} \xi_k)}{\lambda_k} \xi_k = \int_0^T \hat{\beta}_{PCR(q)}(t) X_t dt, \quad (2)$$

where

$$\hat{\beta}_{PCR(q)} = \sum_{k=1}^q \frac{E(\mathbf{Y} \xi_k)}{\lambda_k} f_k(t)$$

is the estimator of the coefficient regression function β obtained with the first q principal components.

Using the first q principal components raises a problem since they are computed independently of the response. Principal components with a great power of explanation yield generally stable models but could be uncorrelated with the response, whereas the principal components highly correlated with the response could be less explanatory for \mathbf{X} . Moreover, for functional data, the number of principal components could be infinite. Thus, the choice of principal components is a trade-off between stability of the linear model and its predictive power (see also Escabias et al. (2004)). A solution to this problem is the PLS approach.

2.2 Partial least squares regression (PLS)

The PLS approach offers a good alternative to the PCR method by replacing the least squares criterion with that of maximal covariance between $\{X_t\}_{t \in [0, T]}$ and \mathbf{Y} (Preda and Saporta (2005a)).

One obtains a set of PLS components $\{t_i\}_{i \geq 1}$ using an iterative procedure. At each step, the PLS component being defined as the linear combination of X_t variables that attains maximum covariance with the response or between residuals :

Let $X_{0,t} = X_t$, $\forall t \in [0, T]$ and $\mathbf{Y}_0 = \mathbf{Y}$. At step q , $q \geq 1$, of the PLS regression of \mathbf{Y} on $\{X_t\}_{t \in [0, T]}$, we define the q^{th} PLS component, t_q , by the eigenvector associated to the largest eigenvalue of the operator $\mathbf{W}_{q-1}^X \mathbf{W}_{q-1}^Y$, where \mathbf{W}_{q-1}^X , respectively \mathbf{W}_{q-1}^Y , are the Escoufier's operators associated to $\{X_{q-1,t}\}_{t \in [0, T]}$, respectively to \mathbf{Y}_{q-1} . The PLS step is completed by the ordinary linear regression of $X_{q-1,t}$ and \mathbf{Y}_{q-1} on t_q . Let $X_{q,t}$, $t \in [0, T]$ and \mathbf{Y}_q be the random variables which represent the error of these regressions : $X_{q,t} = X_{q-1,t} - p_q(t)t_q$ and $\mathbf{Y}_q = \mathbf{Y}_{q-1} - \mathbf{c}_q t_q$.

Then, for each $q \geq 1$, $\{t_q\}_{q \geq 1}$ forms an orthogonal system in $L_2(X)$ and the following decomposition formulas hold :

$$\begin{aligned} \mathbf{Y} &= \mathbf{c}_1 t_1 + \mathbf{c}_2 t_2 + \dots + \mathbf{c}_q t_q + \mathbf{Y}_q, \\ X_t &= p_1(t)t_1 + p_2(t)t_2 + \dots + p_q(t)t_q + X_{q,t}, \quad t \in [0, T]. \end{aligned}$$

The PLS approximation of \mathbf{Y} by $\{X_t\}_{t \in [0, T]}$ at step q , $q \geq 1$, is given by :

$$\hat{\mathbf{Y}}_{PLS(q)} = \mathbf{c}_1 t_1 + \dots + \mathbf{c}_q t_q = \int_0^T \hat{\beta}_{PLS(q)}(t) X_t dt. \quad (3)$$

Notice that de Jong (1993) and Phatak (2001) show that for a fixed q , the PLS regression fits closer than PCR, in that sense

$$R^2(\mathbf{Y}, \hat{\mathbf{Y}}_{PCR(q)}) \leq R^2(\mathbf{Y}, \hat{\mathbf{Y}}_{PLS(q)}),$$

where R is the multiple correlation coefficient.

The number of PLS components used for regression is generally determined by cross-validation.

3 Clusterwise regression model and functional data

Let us suppose that the response Y is one dimensional ($p = 1$). The clusterwise linear model assumes that there exists a random variable \mathcal{G} , $\mathcal{G} \in \{1, 2, \dots, K\}$, $K \geq 2$, such that for each cluster defined by $\{\mathcal{G} = i\}$ one has

$$\begin{aligned} \mathbb{E}(Y | \mathbf{X} = x, \mathcal{G} = i) &= \alpha^i + \int_0^T \beta^i(t)x(t)dt, \\ V(Y | \mathbf{X} = x, \mathcal{G} = i) &= \sigma_i^2 > 0, \quad x \in L_2([0, T]), \forall i = 1, \dots, K. \end{aligned} \quad (4)$$

i.e.,

$$Y_{|\mathcal{G}=i} = \alpha^i + \int_0^T \beta^i(t)X(t)dt + \varepsilon_i, \quad \forall i = 1, \dots, K.$$

Let us assume that K is known and the homoscedasticity hypothesis holds, i.e. the variance of the random error term ε_i within each cluster are equals, $\sigma_i^2 = \sigma^2$, $\forall i = 1, \dots, K$.

In such a model, the parameters that have to be estimated are the regression coefficient functions for each cluster $\{(\alpha^i, \beta^i)\}_{i=1, \dots, K}$ and σ^2 . Charles (1997) and Bock (1969) use the following criterion for estimating the linear models within clusters, $\{\alpha^i, \beta^i\}_{i=1}^K$:

$$\min_{\{\alpha^i, \beta^i\}_{i=1}^K, \mathcal{L}(\mathcal{G})} \left\{ V(Y - \hat{Y}^L) \right\}, \quad (5)$$

where $\hat{Y}^L = \sum_{i=1}^K \hat{Y}^i \mathbf{1}_{\mathcal{G}=i}$ and $\hat{Y}^i = \alpha^i + \langle \hat{\beta}^i, \mathbf{X} \rangle$ is the approximation of Y given by the linear regression of Y on \mathbf{X} within the cluster i , $i = 1, \dots, K$.

If n data points $\{x_i, y_i\}_{i=1}^n$ have been collected, the cluster linear regression algorithm finds simultaneously an optimal partition of the n points, $\hat{\mathcal{G}}$ (as estimation of the distribution of \mathcal{G} , $\mathcal{L}(\mathcal{G})$), and the regression models for each cluster (element of partition) $(\hat{\alpha}, \hat{\beta}) = \{\hat{\alpha}^i, \hat{\beta}^i\}_{i=1}^K$, which minimize the criterion :

$$\mathcal{V}(K, \hat{\mathcal{G}}, \hat{\alpha}, \hat{\beta}) = \sum_{i=1}^K \sum_{\hat{\mathcal{G}}(j)=i} (y_j - (\hat{\alpha}^i + \langle \hat{\beta}^i, x_j \rangle))^2. \quad (6)$$

In order to minimize (6), the clusterwise linear regression algorithms iterates the following two steps :

- i) For given $\hat{\mathcal{G}}$, $\mathcal{V}(K, \hat{\mathcal{G}}, \hat{\alpha}, \hat{\beta})$ is minimized by the LS-estimator $(\hat{\alpha}^i, \hat{\beta}^i)$ from the points (x_j, y_j) with $\hat{\mathcal{G}}(j) = i$.
- ii) For given $\{\hat{\alpha}^i, \hat{\beta}^i\}_{i=1}^K$, $\mathcal{V}(K, \hat{\mathcal{G}}, \hat{\alpha}, \hat{\beta})$ is minimized according to

$$\hat{\mathcal{G}}(j) = \arg \min_{i \in \{1, \dots, K\}} (y_j - (\hat{\alpha}^i + \langle \hat{\beta}^i, x_j \rangle))^2. \quad (7)$$

That is, $\mathcal{V}(K, \hat{\mathcal{G}}, \hat{\alpha}, \hat{\beta})$ is monotonely decreasing if the steps *i*) and *ii*) are carried out alternately :

$$\underbrace{\hat{\mathcal{G}}_0 \Rightarrow (\hat{\alpha}_0, \hat{\beta}_0)}_{\mathcal{V}_0} \geq \underbrace{\hat{\mathcal{G}}_1 \Rightarrow (\hat{\alpha}_1, \hat{\beta}_1)}_{\mathcal{V}_1} \geq \dots \geq \underbrace{\hat{\mathcal{G}}_l \Rightarrow (\hat{\alpha}_l, \hat{\beta}_l)}_{\mathcal{V}_l} \geq \dots \quad (8)$$

where the index of each term denotes the iteration step, $\hat{\mathcal{G}}_0$ being an initial partition of the n data points.

When the predictor is of functional type, the classical linear regression is not adequate to provide estimators for the linear models within clusters, $\{\alpha^i, \beta^i\}_{i=1}^K$. We propose to adapt the PLS and PCR regression approaches for the clusterwise algorithm in order to overcome this problem. The convergence of the clusterwise algorithm using these regularization methods is discussed in Preda and Saporta (2005b).

Let us denote by $\{\hat{\alpha}_{PLS}^i, \hat{\beta}_{PLS}^i\}_{i=1}^K$, respectively by $\{\hat{\alpha}_{PCR}^i, \hat{\beta}_{PCR}^i\}_{i=1}^K$ the estimators for the coefficient regression functions within clusters.

As a quality measure of the fit in clusterwise regression one can use the square of the correlation coefficient between the response (Y) and the predictor (\mathbf{X}) within each cluster. If a clusterwise linear model underlies data, it is interesting to compare each cluster regression quality with that obtained by the linear model without clusters. For comparison of several techniques for estimating the clusterwise model (for example, PLS and PCR) the criterion given in (7) is a natural choice.

4 Numerical results

In this section we compare the clusterwise PLS and PCR approaches in the context of functional data both on simulated and real data.

Firstly we consider simulated data with two clusters each having its own linear structure with respect to a one dimensional response Y and a set of curves $\{X_t, t \in [0, T]\}$ drawn from the one-dimensional Brownian motion. The aim is to check the capability of the clusterwise regression to identify these two clusters. The second application concerns stock exchange data and the aim is to "predict" the last five minutes of the evolution of a particular share, considered on a certain interval of time.

We quote by CW-PLS(K) and CW-PCR(K) the clusterwise PLS, respectively PCR, regressions with K clusters, by PCR and PLS, the global linear regression models obtained with the principal components, respectively on the first PLS components. The number of components considered for regression (PLS and PCR) is determined by cross-validation (leave-one-out) using a significance level of 95%.

4.1 Simulation study

Let us consider that the stochastic process underlying the functional data is the standard Brownian motion on the interval $[0, 1]$, $\mathbf{X} = \{X_t\}_{t \in [0, 1]}$, $\mathbb{E}(X_t) = 0$, $\mathbb{E}(X_t X_s) = \inf\{t, s\}$, $\forall t, s \in [0, 1]$. The response variable Y is defined with respect to a group variable, \mathcal{G} , with two modalities in the following way :

$$\text{Class 1 : } Y = \int_0^1 t X_t dt + \varepsilon_1$$

$$\text{Class 2 : } Y = \int_0^1 (1 - t) X_t dt + \varepsilon_2$$

where ε_1 and ε_2 are Gaussian noises such that $\sigma_1^2 = \sigma_2^2 = \sigma^2$. We consider two situations $\sigma^2 = 0.01$ and $\sigma^2 = 0.02$ which correspond to the following ratios $\sigma^2/\mathbb{V}(Y)$:

	$\sigma^2 = 0.01$	$\sigma^2 = 0.02$
Class 1	0.069	0.137
Class 2	0.167	0.285

Table 1. Noise to response ratio, $\sigma^2/\mathbb{V}(Y)$

Our simulation is based on the following conditions:

- the trajectories of X are discretized in 101 equidistant points.
- values of Y as well as the principal and PLS components are computed using integration by trapezoidal interpolation.
- the training sample sizes are identical for both groups, $n = 500$.
- 100 simulations.

Table 2 presents the performance values of PLS and PCR models in terms of response variance explained by the predictor, i.e. the multiple correlation coefficient, R^2 . For clusterwise models we present also the error classification rate (ECR). Both measures are averaged over 100 samples.

Figure 1 plots $\hat{\beta}_{PLS}^i$, $i = 1, 2$., the two regression coefficient functions obtained with the PLS approach attaining the best model with respect to the criterion given by (7).

The results obtained on this example show that PLS fits slightly better than PCR especially when the noise to response ratio is increasing. This is mainly due to the fact that the PLS takes into account, for computing PLS components, the correlation between the response and predictor, whereas that is not the case for PCR. Notice that these results are in agreement with those obtained by Barker and Rayens (2003) and Preda et al. (2007) on the capability of PLS models for classification purpose.

Model	$\sigma^2 = 0.01$			$\sigma^2 = 0.02$		
R^2 -PCR	0.718			0.597		
R^2 -PLS	0.724			0.612		
	cluster 1	cluster 2	ECR	cluster 1	cluster 2	ECR
CW-PCR(2)	0.882	0.794	0.112	0.752	0.625	0.322
CW-PLS(2)	0.908	0.812	0.103	0.826	0.674	0.260

Table 2. Model quality : R^2 and error classification rate (ECR) averaged over 100 simulations

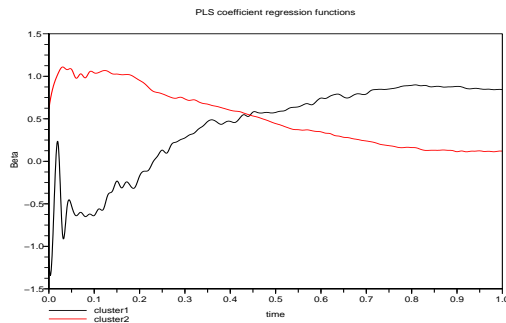


Fig. 1. Cluster-specific regression coefficient functions for PLS approach ($\sigma^2 = 0.01$).

4.2 Application on stock exchange data

We have 84 shares quoted at the Paris stock exchange, for which we know the whole behavior of the growth index during one hour (between 10^{00} and 11^{00}). Notice that a share is likely to change every second. We also know the evolution of the growth index of a new share (indexed 85) between 10^{00} and 10^{55} .

Linear models for this data set were fitted with PLS and PCR regression techniques in order to predict the way in which the new share will behave between 10^{55} and 11^{00} (Preda and Saporta (2005a)). We have shown (Preda and Saporta (2005b)) that this prediction is improved when the clusterwise approach is considered.

Since the curves are completely known, we use the time average approximation developed in Preda (2000) by taking an equidistant discretization of the interval $[0, 3600]$ (time expressed in seconds) in 60 subintervals. The forecasts obtained will then match the average level of the growth index of share 85 considered on each interval $[60 \cdot (i - 1), 60 \cdot i]$, $i = 56, \dots, 60$.

The results of the best models are presented in the Table 3.

	$\hat{m}_{56}(85)$	$\hat{m}_{57}(85)$	$\hat{m}_{58}(85)$	$\hat{m}_{59}(85)$	$\hat{m}_{60}(85)$	SSE
Observed	0.700	0.678	0.659	0.516	-0.233	-
PLS	0.312	0.355	0.377	0.456	0.534	0.911
PCR	0.613	0.638	0.669	0.825	0.963	1.511
CW-PLS(3)	0.643	0.667	0.675	0.482	0.235	0.215
CW-PLS(4)	0.653	0.723	0.554	0.652	-0.324	0.044

Table 3. Forecasts for share 85

Using the sum of squared errors (SSE) as measure of fit, let us observe that the clusterwise models give better results than the global ones. The clusterwise models predict better the crash of the share 85 for the last 5 minutes, whereas the global models do not. For the PLS model with 4 clusters, the size of each cluster is given by the distribution $(\frac{17}{84}, \frac{32}{84}, \frac{10}{84}, \frac{25}{84})$. Following the K-NN procedure proposed by Charles (1977), the share 85 belongs to the first cluster.

5 Conclusion

PLS and PCR approaches are regularization techniques for linear regression used with success when the least squares criterion produces inconsistent estimators, in particular, when multicollinearity and sample size problems occur. This is the case for functional data (multicollinearity) and the clusterwise algorithm (cluster size less than the number of predictors). We show by a simulation study and an application on stock-exchange data the efficiency of these two methods and point out the accuracy of PLS with respect to PCR.

References

- ABRAHAM C., CORNILLON P., MATZNER-LÖBER E. and MOLINARI N. (2003): Unsupervised curve clustering using B-splines, *Scand. J. Statist.*, Vol. 30, p. 581–595.
- BARKER M. and RAYENS W. (2003): Partial least squares for discrimination. *Journal of Chemometrics*, 17, 166–173.
- BOCK H. H. (1989): The equivalence of two extremal problems and its application to the iterative classification of multivariate data. *Lecture note, Mathematisches Forschungsinstitut Oberwolfach*.
- CHARLES C. (1977): Régression typologique et reconnaissance des formes, *Thèse de doctorat, Université Paris IX*.
- DE JONG S. (1993): PLS fits closer than PCR, *Journal of Chemometrics*, 7, 551–557.
- DESARBO W. S. and CRON W. L. (1988): A Maximum Likelihood Methodology for Clusterwise Linear Regression, *Journal of Classification*, 5, 249–282.
- DEVILLE J.C. (1978): Analyse et prévision des séries chronologiques multiples non stationnaires, *Statistique et Analyse des Données*, 3, 19–29.

- DIDAY E. (1976): Classification et sélection de paramètres sous contraintes, *Rapport de Recherche IRIA-LABORIA*, no. 188.
- ESCABIAS M., AGUILERA A.M., and VALEDERRAMA M.J. (2004): Principal component estimation of functional logistic regression : discussion of two different approaches. *Journal of Nonparametric Statistics*, 3-4, 365-385.
- ESCOUFIER Y. (1970) Echantillonnage dans une population de variables aléatoires réelles, *Publications de l'Institut de Statistique de l'Université de Paris*, 19, Fasc. 4, 1-47.
- FERRATY F. and VIEU P. (2006): *Nonparametric Functional Data Analysis. Theory and Practice*, Springer Series in Statistics.
- HENNIG C. (1999): Models and Methods for Clusterwise Linear Regression, *Classification in the Information Age*, Springer, Berlin, 179-187.
- HENNIG C. (2000): Identifiability of Models for Clusterwise Linear Regression, *Journal of Classification*, 17, 273-296.
- PLAIA A. (2004): Constrained clusterwise linear regression, In: M. Vichi, P. Monari, S. Mignani, A. Montanari (Eds): *New Developments in Classification and Data Analysis*, Springer, 78-86.
- PHATAK A. and DE HOOG F. (2001): PLSR, Lanczos, and Conjugate Gradients, *CSIRO Mathematical & Information Sciences, Report No. CMIS 01/122, Canberra*.
- PREDA C. and SAPORTA G. (2005a): PLS regression on a stochastic process, *Computational Statistics and Data Analysis*, 48, 149-158.
- PREDA C. and SAPORTA G. (2005b): Clusterwise PLS regression on a stochastic process, *Computational Statistics and Data Analysis*, 49, 99-108.
- PREDA C., SAPORTA G. and LÉVÉDER C. (2007): PLS classification of functional data, *Computational Statistics*, In Press, doi : 10.1007/s00180-007-0041-4.
- RAMSAY J.O. and SILVERMAN B.W. (1997): *Functional Data Analysis*, Springer Series in Statistics, Springer-Verlag, New York.
- RAMSAY J.O. and SILVERMAN B.W. (2002): *Applied Functional Data Analysis : Methods and Case Studies*, Springer, Berlin-Heidelberg.
- SAPORTA G. (1981): Méthodes exploratoires d'analyse de données temporelles, *Cahiers du B.U.R.O., No. 37-38, Université Pierre et Marie Curie, Paris*.
- SPAETH H. (1979): Clusterwise linear regression, *Computing* 22, 367-373.