APPLIED STOCHASTIC MODELS IN BUSINESS AND INDUSTRY
Appl. Stochastic Models Bus. Ind., 2005; **21**:165–173
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asmb.539

Statistical modelling of functional data

Philippe C. Besse[1,*,†], Hervé Cardot[1,2,‡], Robert Faivre[2,§] and Michel Goulard[2,¶]

[1] *Laboratoire de Statistique et Probabilités, UMR CNRS C5583, Université Paul Sabatier, 31062 Toulouse Cedex, France*
[2] *INRA Toulouse, Biométrie et Intelligence Artificielle, 31326 Castanet-Tolosan Cedex, France*

SUMMARY

This paper presents some statistical models and estimation procedures when the explanatory variables are functions. We put the stress on the fact that regularization techniques are needed in order to get stable and reliable estimations. Then, an application in remote sensing in presented. It shows the potential of these kind of models to handle real life problems. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: functional principal components analysis; ill-posed problems; Karhunen–Loeve expansion; longitudinal data; multilogit model; penalty; regularization; smoothing; SPOT4/Végétation sensor

1. INTRODUCTION

This paper addresses a specific case of statistical learning: the predictive modelling when the training data are discretizcd curves. In the illustrative example, these curves are the evolution along time of the reflectance of coarse pixels. The images are obtained from the Végétation sensor of the SPOT4 satellite. The aim of this application is to predict the land use, i.e. the proportion of each type of vegetation or land cover inside each mixed pixel. Of course discretized curves can be considered as random vectors in a finite dimensional vectorial space and then statistical or training methods can be performed as usual. Nevertheless, this kind of data rises some particular questions and then requires some special care to adapt classical methods. The main question is how the functional properties as the regularity or the smoothness of the curves could be taken into account. Automatic measurements easily generate highly dimensional data. Both this dimensionality and regularity drive us to take care of severe multicollinearity and overfitting problems.

This approach is also a particular case of functional data analysis. A survey on this topic is given in Reference [1]. The statistical analysis of data with variables taking values in a function

*Correspondence to: Philippe C. Besse, Laboratoire de Statistique et Probabilités, UMR CNRS C5583, Université Paul Sabatier, 31062 Toulouse Cedex, France.
[†] E-mail: philippe.besse@math.ups-tlse.fr
[‡] E-mail: cardot@toulouse.inra.fr
[§] E-mail: faivre@toulouse.inra.fr
[¶] E-mail: goulard@toulouse.inra.fr

space has been firstly introduced for exploratory methods such as principal component analysis (PCA) and many articles were published treating examples from chemometrics, economy, climatology, etc., adapting multivariable tools for functional data analysis.

Our aim is slightly different. We will mainly focus on statistical estimation, or learning, in models with functional covariates. Nevertheless, the mathematical background is quite similar and a reduction dimension method such as PCA is one way to prevent from multicollinearity. Suppose we observe a sample of curves $X_1(t), \ldots, X_n(t)$, defined for t belonging to a time interval T. It is usually assumed that these curves are drawn from the same distribution and belong to a separable Hilbert space H, equipped with the inner product $\langle ., . \rangle$ and the norm $\|.\|$. We can take for instance $H = L^2[T]$, the Hilbert space of squared integrable functions defined on T. We also suppose that $E(\|X\|^2)$ is finite so that we can define the expectation $\mu(t) = E(X(t))$ and the covariance operator Γ, mapping H to H,

$$\Gamma f(s) = \int_T \gamma(s,t) f(t)\,\mathrm{d}t, \quad f \in H \tag{1}$$

$\gamma(s,t)$ being the covariance between the two random variables $X(s)$ and $X(t)$. This operator is known to be non-negative and nuclear, that is to say the sum of its eigenvalues is finite.

Many studies have dealt with the statistical description of this kind of data and the main idea was to extend PCA to a functional framework.

This was initially done by Deville [2] and Dauxois and Pousse [3] by expanding the curves in the basis of eigenfunctions of the empirical covariance operator in order to obtain a small dimension space which represents as well as possible the main variations of the data. This expansion is also known as the Karhunen–Loeve expansion. Then, Besse and Ramsay [4], Rice and Silverman [5], Besse *et al.* [6]... added interpolating and smoothing procedures in order to take into account both the discretization of the curves and to give a smooth representation of the data. Pezzulli and Silverman [7], Silverman [8] and Cardot [9] showed that incorporating a smoothing step could also improve the quality of the estimators.

We present, in Section 2, extensions of classical statistical models such as the linear regression model, the AR processes and the generalized linear models for functional data. In Section 3, we explain why, in this context, estimation is an ill-posed problem and thus regularization is needed, either by a dimension reduction approach or by adding a penalty in the loss criterion. Finally, in Section 4, an application of generalized linear models for remote sensing data is presented.

2. SOME STATISTICAL MODELS FOR FUNCTIONAL DATA

Suppose now we also have observations $Y_1, \ldots, Y_n$, supposed to be real or functional again, linked to the curves $X_1, \ldots, X_n$. We may be interested in modelling and/or predicting Y as a function of X. Almost all existing statistical models can be derived for functional data but we will present the most studied ones which are the linear model, the generalized linear model and the functional autoregressive process. The linear model [10–12] can be expressed as follows,

$$Y_i = \mu + \int_T \beta(s) X_i(s)\,\mathrm{d}s + \varepsilon_i \tag{2}$$

and we are willing to estimate the functional parameter β. Generalized linear models are built by adding a link function that allows to describe distributions that are discrete, positive, etc. These types of models have been implemented and studied by Marx and Eilers [13] or Cardot and Sarda [14]. The remote sensing application presented in Section 4 is a particular case of generalized linear model with a multilogit link function. Another model studied in the literature is the functional autoregressive process of order one. It was introduced by Bosq (15) and is defined as follows,

$$X_i(t) = \mu(t) + \int_T \rho(s,t)(X_{i-1}(s) - \mu(s))\,\mathrm{d}s + \varepsilon_i(t) \tag{3}$$

where ε_i are i.i.d. second order random variables taking values in H. More generally, this model may be written $X_i = \mu + \rho(X_{i-1} - \mu) + \varepsilon_i$ and estimating operator ρ and mean function μ allows then to make a prediction. Note that fully non-parametric models have also been investigated recently [16].

3. ESTIMATION PROCEDURES

The main problem is that covariance matrices are compact operators in an infinite dimension space and so estimation is an *ill-posed* problem. For instance in the linear model, it is easy to check that the solution of the unconstrained least squares criterion satisfies the score equation:

$$\frac{1}{n}\sum_{i=1}^{n} \langle X_i, \hat{\beta}\rangle X_i = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i \tag{4}$$

and we can find as many $\hat{\beta}$ as we want satisfying (4).

A first idea consists in reducing the dimension of the space in which we seek for a solution. Denoting by $v_1, \ldots, v_q$ the orthonormal eigenfunctions of the empirical covariance operator associated to the eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_q$, we get a solution in the q-dimensional function space V_q span by $v_1, \ldots, v_q$,

$$\hat{\beta}_q = \sum_{j=1}^{q} \frac{\langle \Delta_n, v_j\rangle}{\lambda_j} v_j \tag{5}$$

where $\Delta_n = 1/n \sum_{i=1}^{n} Y_i X_i$. Smoothing steps can also be combined with this approach and they generally give better results on real data [17, 18]. Asymptotic properties of such estimators have been derived [12, 15, 19] and show that q must not tend too rapidly to infinity as the sample size increases. Other strategies to reduce the dimension of the space in which estimators are built exist. For instance, Preda and Saporta [20] extended PLS to the functional framework.

A second idea consists in adding a penalty in order to get a stable and a unique solution [13, 21, 22] since estimation can be seen in this context as a kind of ill-posed inverse problem [23]. Consider a penalty operator $J(.)$ which maps a subset of H to R; it can be for instance the norm of the second derivative $J(\beta) = \|\beta^{(2)}\|^2$. Assuming that $J(.)$ is differentiable, we are seeking, in the linear model, the function $\hat{\beta}_\ell$ satisfying

$$\frac{1}{n}\sum_{i=1}^{n} \langle X_i, \hat{\beta}\rangle X_i + \ell \nabla J(\hat{\beta}) = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i \tag{6}$$

ℓ being a tuning parameter that controls the trade off between the fidelity to the data and the 'regularity' of the solution. Note that generally the solution is expanded in a basis of B-splines or Fourier series both for computational reasons and approximation properties of such functions. This also allows to deal with the discretization problem since curves are never observed continuously along time. Indeed, let us consider a basis of functions $B_1, \dots, B_k$ and expand the functional parameter β in this basis

$$\beta(t) \approx \sum_{j=1}^{k} \beta_j \, B_j(t), \quad t \in T$$

Then, $\langle X_i, \beta \rangle \approx \sum_{j=1}^{k} \beta_j \langle X_i, B_j \rangle$ can be evaluated with a quadrature rule and the estimation of β is performed through the estimation of the vector of parameters $(\beta_1, \dots, \beta_k)$ with classical algorithms.

More generally, we can consider a loss function $\mathscr{L}$ such as a least squares criterion or the opposite of the log-likelihood. Then, the regularization approach consists in minimizing

$$\min_{\beta \in S} \mathscr{L}(Y_1, \dots, Y_n, X_1, \dots, X_n, \beta) + \ell J(\beta) \tag{7}$$

in a space $S \subset H$. The dimension reduction approach leads to find the optimum of

$$\min_{\beta \in \hat{S}_q} \mathscr{L}(Y_1, \dots, Y_n, \hat{\Pi}_q X_1, \dots, \hat{\Pi}_q X_n, \beta) \tag{8}$$

where $\hat{S}_q$ is a q-dimensional function space and $\hat{\Pi}_q$ is a projector onto $\hat{S}_q$. Cardot *et al.* [24] and Cardot and Sarda [14] showed, in the context of linear and generalized linear models, for quadratic penalties based on the L_2 norm of given order derivatives, that the parameter ℓ should not tend too rapidly to zero to get consistent estimators. The values of the tuning parameters q and ℓ are generally obtained in practical situations with criterions based on cross-validation.

4. A REMOTE SENSING APPLICATION

On board SPOT4, a satellite launched in March 1998, the Végétation sensor gives, at a high temporal resolution, daily images of Europe at a coarse spatial resolution, each pixel corresponding to a ground area of 1 km^2. The information given by this sensor are the reflectances, i.e. the proportion of reflected radiation, in the four spectral bands blue (B), red (R), near infra-red (NIR) and short wave infra-red (SWIR). We also considered two vegetation indices, that are frequently used in bioclimatology and remote sensing [25], the normalized difference vegetation index (NDVI), NDVI = (NIR−R)/(NIR+R), and the perpendicular vegetation index (PVI), $\mathrm{PVI} = (\mathrm{NIR} - 1.2\mathrm{R})/(\sqrt{1 + (1.2)^2})$, which are functions of the reflectances in the red (R) and NIR channels. This information allows to characterize the developpement of vegetation and crops at the scale of a small country [25]. Because in Europe, and particularly in France, the size of plots is much less than 1 km^2, the observed reflectances are a mixture of different informations since they contain different agricultural plots (maize, wheat, etc.), forests or urban areas.

We aim at estimating the land use, i.e. the proportion of each types of culture or land cover inside each mixed pixel. This is the first step in predicting regional crop productions. Despite the medium spatial resolution, we take advantage of the high temporal resolution of such a sensor to derive estimations of the land use. We consider two approaches to achieve that. A direct approach that is based on the generalized linear models for functional data. The proportions are assumed to be drawn from a multinomial distribution whose parameters depend on the temporal evolution of the reflectance. The other approach, which is an inverse approach, assumes that the observed reflectance in a mixed pixel is a weighted combination of the pure reflectances of each theme.

Let us denote by $\pi_{ij},\ j = 1, \ldots, p$, the proportion of land use of crop j in pixel i of 1 km^2. In our application the observed area is about 40 km × 40 km so that $i = 1, \ldots, n = 1554$. Ten $(p = 10)$ different classes of crops were present. The curves of reflectance for each pixel i, in each channel and index, are denoted by $\boldsymbol{X}_i = [X_i(t_i), \ldots, X_i(t_K)]^{\mathrm{T}}$ where $t_1 < \cdots < t_k < \cdots < t_K$ are the instants of measure. The images in which the clouds were too important were removed to finally get $K = 39$ different images from March to August 1998. We assume that the land use is fixed during the observation period.

4.1. The multilogit model for functional data

We suppose now that the proportions π_{ij} given the temporal evolution of the reflectance $\{X_i(t),\ t \in T\}$ can be modelled as resulting from a multinomial distribution whose parameters satisfy

$$\mathbb{E}(\pi_{ij}|X_i) = \frac{\exp\left(\delta_j + \int_T \alpha_j(t) X_i(t)\,\mathrm{d}t\right)}{\sum_{\ell=1}^{p} \exp\left(\delta_l + \int_T \alpha_l(t) X_i(t)\,\mathrm{d}t\right)} \tag{9}$$

For identifiability reasons we took $\alpha_p = 0$ and $\delta_p = 0$.

We can give a sketch of interpretation on this approach. Indeed, each coarse resolution pixel is assumed to be composed of numerous small agricultural plots of similar area. Each of these plots of a pixel i is, with a probability $E(\pi_{ij}|X_i)$, of the theme j of the land use. So that the number of plots of the themes follows exactly a multinomial distribution and we observe the proportions. Each functional coefficient α_j may have an interpretation by comparison to the reference function $\alpha_p = 0$. For instance, if α_j is a positive function, then the ratio of the proportion will be higher than the mean value and thus the class j will be more important in the pixel i, if the centred reflectance curve is positive. Nevertheless, if the estimations have oscillating features, giving an interpretation may be a hazardous task.

We aim at estimating the vector $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{p-1})^{\mathrm{T}}$ and the functional coefficients $\alpha_j(t),\ j = 1, \ldots, p-1$. The estimations are obtained by means of the maximum likelihood criterion.

For computational purposes, we preferred the dimension reduction approach based on a functional principal components analysis. The number of covariates (the principal components) still may be large and we decided to select the most significative parameters by means of the likelihood ratio test with an ascendant procedure. More details may be found in Cardot *et al.* [26].

4.2. Results

Different models can be built to deal with the problem of land use estimation. The *most natural* approach consists in assuming that the reflectance $X_i(t)$ of a pixel i at date t is a linear combination of the pure reflectance curves or characteristic curves, say $\rho_j(t)$, of each theme j

weighted by the associated proportions π_{ij}

$$X_i(t) = \sum_{j=1}^{p} \pi_{ij}\rho_j(t) + \varepsilon_i(t) \tag{10}$$

where $\varepsilon_i(t)$ is an error term. This model was proposed in Reference [27] incorporating moreover random effects.

Model (10) is a *varying-time regression model* and assuming the land use is known, we can obtain global non-parametric estimations of the characteristic curves ρ_j, $j = 1, \ldots, p$, by means of penalized splines as in Reference [28]. Then, using the estimated curves, we are able to get estimators of the vector of proportions in a new pixel i' by considering a constrained least squares estimator ($\pi_{i'j} \geqslant 0$ and $\sum_j \pi_{i'j} = 1$). More details may be found in Reference [26].

The initial sample was split into a learning sample composed of 1055 pixels used to perform the estimations and select the best models and a test sample composed of 499 pixels used to evaluate and compare the two different approaches. The following criterion was used to compare on the test sample the skill of the two approaches:

$$R_{ij} = \frac{|\pi_{ij} - \hat{\pi}_{ij}|}{1/n \sum_{i=1}^{n} \pi_{ij}}$$

where $\hat{\pi}_{ij}$ is the predicted proportion of theme j in pixel i. We also considered the most simple model, named M_0, as a benchmark to indicate if it is worth building sophisticated statistical models. It consists in predicting the land use of one crop by its empirical mean in the learning sample. This is a particular case of the multilogit model with no covariates.

The estimators for functions α_j in the multilogit model are shown in Figure 1. The reference curve is taken for the theme 'Urban' since we expect that it varies less along time. We recognize a biological cycle for the curves associated to crops such as 'Winter crops' or 'Peas' and a rather flat coefficient for themes such as 'Forest' or 'Permanent crops'.

We noticed (see Tables I and II) that the functional multilogit model, even if it can appear to be less natural since it has no direct physical interpretation, gave generally better predictions than the mixture of curves approach which appeared to be unstable. Indeed, it seems that these poor results (see e.g. the predictions for 'rapeseed', 'winter crops' or 'permanent crops' in Table I) are mainly due to identifiability problems of the parameters. A bayesian approach would certainly help us to improve the prediction.

Nevertheless, the effectiveness of the multilogit model may be moderated. Indeed, it relies on a precise estimation of the location parameters δ_j which represent a kind of mean value of each crop j. If we have to make predictions in an area whose repartition of the crops is very different then we might get into trouble. The best predictions seem to be obtained when using the PVI index. For instance, the errors were reduced of about 60% compared to the reference model M_0 which consists in predicting in the test sample by the mean value in the training sample.

Thus combinations of the original wavelengths may be more appropriate to predict the land use and our future work will deal with finding optimal combinations of the available channels.

Other methods could also have been used and regression trees or neural networks are potential candidates (see Reference [29]). These methods are known to be better if threshold effects or strong non-linear effects are suspected to exist. A preliminary work with random forests, not presented in this paper, gave good results but this point needs further investigation.

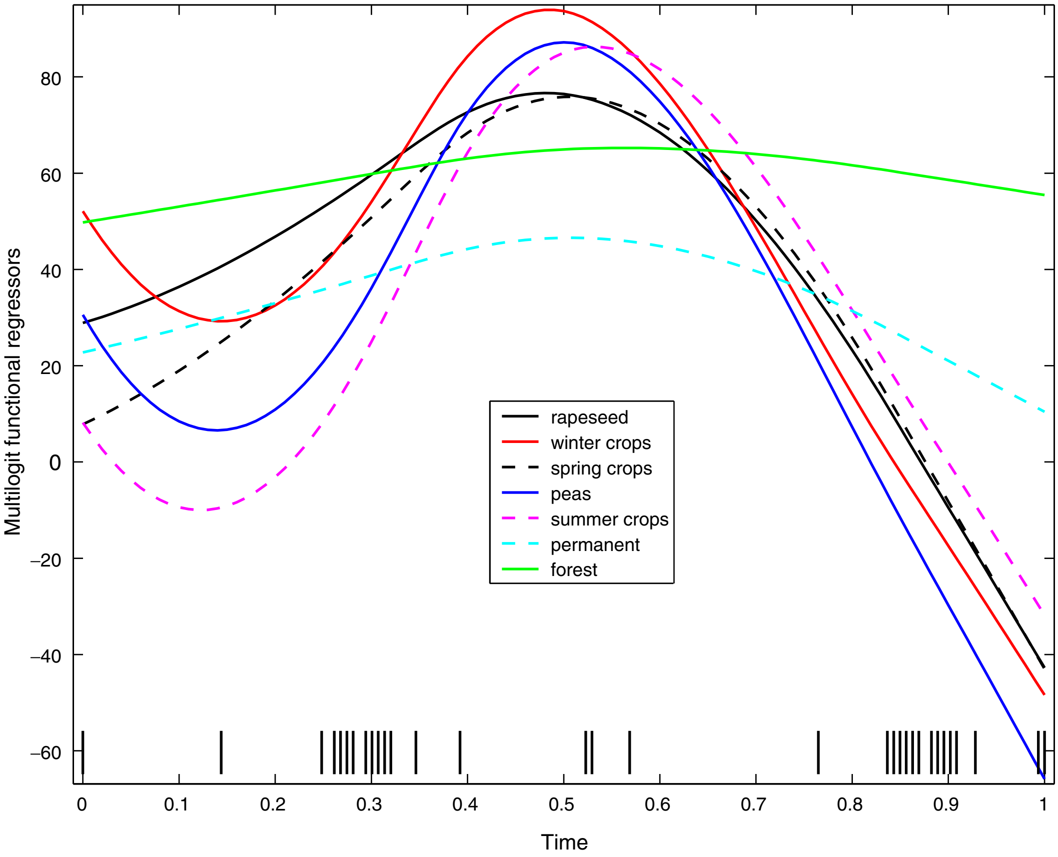

Figure 1. Estimated functional parameters for the NDVI index.

Table I. Median value of the criterion error when predicting land use in the test sample with the mixture of characteristic curves approach.

Themes	NDVI	PVI	Blue	Red	NIR	SWIR	M_0
Urban	0.19	0.18	0.18	**0.13**	0.48	0.18	0.86
Water	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	**0.00**	1.30
Rapeseed	0.90	1.15	**0.82**	1.19	1.01	1.26	0.59
Winter crops	**0.35**	0.49	0.36	0.51	0.41	0.57	0.30
Spring crops	0.49	0.85	0.49	0.43	0.67	**0.43**	0.69
Peas	**0.77**	0.94	0.81	0.98	0.84	0.86	0.63
Summer crops	0.21	0.23	3.41	0.34	**0.19**	0.21	0.88
Permanent crops	0.79	**0.72**	1.01	0.78	0.91	0.95	0.61
Forest	0.36	**0.33**	0.48	0.54	0.37	0.34	0.98
Potatoes	0.37	0.24	**0.23**	0.30	0.72	0.30	1.30

Bold face numbers correspond to the best predictions. Model M_0 is used as a benchmark.

Table II. Median value of the criterion error when predicting land use in the test sample with the GLM approach.

Themes	NDVI	PVI	Blue	Red	NIR	SWIR	M_0
Urban	0.49	**0.36**	0.47	0.54	0.41	0.51	0.86
Water	0.43	**0.29**	0.78	0.62	0.61	0.31	1.30
Rapeseed	0.48	0.46	**0.45**	0.50	0.47	0.47	0.59
Winter crops	0.20	0.21	**0.19**	0.20	0.22	**0.19**	0.30
Spring crops	0.58	**0.56**	0.60	0.61	0.65	0.61	0.69
Peas	0.50	**0.43**	0.45	**0.43**	0.48	0.46	0.63
Summer crops	0.61	0.68	0.61	0.60	0.76	**0.53**	0.88
Permanent crops	0.47	**0.46**	0.52	0.49	**0.46**	0.50	0.61
Forest	0.34	0.36	0.34	**0.31**	0.45	0.35	0.98
Potatoes	0.90	0.93	0.94	0.90	1.06	**0.85**	1.31

Bold face numbers correspond to the best predictions. Model M_0 is used as a benchmark.

5. CONCLUDING REMARKS

A large number of publications, in various fields of science, deal now with functional data since this approach has proved to be a promising alternative to more classical statistical models when the covariates can be considered as discretized curves. It takes account of the functional nature of the data and allows to deal with numerous discretization points and irregular sampling designs. Nevertheless a huge amount of work is still to be done both from a theoretical and practical point of view. More theory would help, for instance, the statisticians to test hypotheses or build confidence intervals and the development of efficient algorithms and packages in statistical softwares would certainly permit these approaches to be used by scientists from other communities.

ACKNOWLEDGEMENTS

The satellite data have been obtained, thanks to the IUC-Végétation program. We would like to thank the participants to the working group ‘staph’ in Toulouse for stimulating discussions.

REFERENCES

1. Ramsay JO, Silverman BW. *Functional Data Analysis*. Springer: Berlin, 1997.
2. Deville JC. Méthodes statistiques et numériques de l’analyse harmonique. *Annales de l’INSEE* 1974; **15**:3–101.
3. Dauxois J, Pousse R. Les analyses factorielles en calcul dei probabilités ét en statistique : essai d’étude synthétique. *Ph.D. Thesis*, Thèse d’État, Université Toulouse III, 1976.
4. Besse PC, Ramsay JO. Principal component analysis of sampled curves. *Psychometrika* 1986; **51**:285–311.
5. Rice JA, Silverman BW. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B* 1991; **53**:233–243.
6. Besse PC, Cardot H, Ferraty F. Simultaneous nonparametric regressions of unbalanced longitudinal data. *Computational Statistics & Data Analysis* 1997; **24**:255–270.
7. Pezzulli S, Silverman BW. On smoothed principal components analysis. *Computational Statistics* 1993; **8**:1–16.
8. Silverman BW. Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics* 1996; **24**:1–24.
9. Cardot H. Nonparametric estimation of the smoothed principal components analysis of sampled noisy functions. *Journal of Nonparametric Statistics* 2000; **12**:503–538.

10. Hastie TJ, Mallows C. A discussion of 'A statistical view of some chemometrics regression tools' by I.E. Fránk and J.H. Friedman. *Technometrics* 1993; **35**:140–143.
11. Ramsay JO, Dalzell C. Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B* 1991; **53**:539–572 (with discussion).
12. Cardot H, Ferraty F, Sarda P. Functional linear model. *Statistics and Probability Letters* 1999; **45**:11–22.
13. Marx BD, Eilers PH. Generalized linear regression on sampled signals and curves: a *P*-spline approach. *Technometrics* 1999; **41**:1–13.
14. Cardot H, Sarda P. Estimation in generalized linear models for functional data via penalized likelihood. *Journal of Multivariate Analysis* 2005; **92**(1):24–41. DOI: 10.1016/j.jmva.2003.08.008.
15. Bosq D. Modelization, non-parametric estimation and prediction continuous time processes. In *Nonparametric Functional Estimation and Related Topics*, Roussas G (ed.). Nato Asi series, 1991; 509–529.
16. Ferraty F, Vieu P. The functional nonparametric model and application to spectrometric data. *Computational Statistics* 2002; **17**:545–564.
17. Besse PC, Cardot H. Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canadian Journal of Statistics* 1996; **24**:467–487.
18. Besse PC, Cardot H, Stephenson D. Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics* 2000; **27**:673–688.
19. Bosq D. *Linear Processes in Function Spaces*. Springer: Berlin, 2000.
20. Preda C, Saporta G. Régression PLS sur un processus stochastique. *Revue de Statistique Appliquée* 2002; **50**:27–45.
21. Leurgans S, Moyeed R, Silverman BW. Canonical correlation analysis when the data are curves. *Journal of the Royal Statistical Society, Series B* 1993; **55**:725–740.
22. Do KA, Kirk K. Discriminant analysis of event-related potential curves using smoothed principal components. *Biometrics* 1999; **55**:174–181.
23. Mas A. Estimation d'opérateurs de corrélation de processus fonctionnels: lois limites, tests, déviations modérées. *Ph.D. Thesis*, Université Paris 6, 1999.
24. Cardot H, Ferraty F, Sarda P. Spline estimators for the functional linear model. *Statistica Sinica* 2003; **13**:571–591.
25. Tucker CJ. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment* 1979; **8**:127–150.
26. Cardot H, Faivre R, Goulard M. Functional approaches for predicting land use with the temporal evolution of coarse resolution remote sensing data. *Journal of Applied Statistics* 2003; **30**:1185–1199.
27. Faivre R, Fischer A. Predicting crop reflectances using satellite data observing mixed pixels. *Journal of Agricultural, Biological and Environmental Statistics* 1997; **2**:87–107.
28. Hoover DR, Rice JA, Wu CO, Yang LP. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika* 1998; **85**:809–822.
29. Hastie TJ, Tibshirani RJ, Friedman J. *Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer: New York, 2001.