



HAL
open science

Nonparametric regression for functional responses with application to conditional functional principal components analysis

Hervé Cardot

► **To cite this version:**

Hervé Cardot. Nonparametric regression for functional responses with application to conditional functional principal components analysis. 2005. hal-00004472v1

HAL Id: hal-00004472

<https://hal.science/hal-00004472v1>

Preprint submitted on 15 Mar 2005 (v1), last revised 1 Mar 2006 (v2)

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.

Nonparametric regression for functional responses with application to conditional functional principal components analysis

Hervé Cardot

Unité Biométrie et Intelligence Artificielle, INRA Toulouse,
31326 Castanet-Tolosan Cedex, FRANCE

email: cardot@toulouse.inra.fr

&

LSP, UMR C5583, Université Paul Sabatier,
118, route de Narbonne, 31062 Toulouse Cedex, FRANCE

March 3, 2005

Abstract

This work proposes a nonparametric estimator of the regression function when the predictor is real and the dependent variable is a curve. The estimator is based on kernel smoothing. This approach is motivated by an extension of the functional principal components analysis, or Karhunen-Loève expansion, which can take into account nonparametrically the effects of an additional covariate. Such a model can also be interpreted as a nonparametric mixed effects model for functional data. Convergence rates are given for the regression function estimator and for the estimator of the conditional covariance operator. The good behaviour of the estimator for functional principal components analysis is illustrated on a simulation study.

Keywords : covariance function, functional mixed effects, Karhunen Loève expansion, weighted covariance operator, eigenelements, almost sure convergence.

1 Introduction

Since the pioneer work by Deville (1974) much attention has been given to the analysis of functional data in the statistical community (see *e.g.* Ramsay and Silverman, 1997, 2002 and references therein). Many studies are devoted

to the statistical description of a sample of curves (growth curves, temperature curves, spectrometric curves, ...) by means of the functional principal components analysis (Besse and Ramsay 1986, Castro *et al.* 1986, Rice and Silverman 1991, Kneip and Utikal 2001, ...). Performing the spectral decomposition of the empirical covariance operator, which is the analogous of a covariance matrix in a function space, allows to get a small dimension space which exhibits, in a optimal way according to a variance criterion, the main modes of variation of the data. Indeed, let us consider a random function $Y(t)$ where t varies in a compact interval T of \mathbb{R} , with mean $\mu(t) = \mathbb{E}(Y(t))$ and covariance function $\gamma(s, t) = \text{Cov}(Y(s), Y(t))$, $s \in T$. Under general conditions, the covariance function may be expressed as follows

$$\gamma(s, t) = \sum_{j \geq 1} \lambda_j v_j(s) v_j(t) , \quad (1)$$

where the λ_j are the ordered eigenvalues, $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, of the covariance operator and the functions v_j the associated orthonormal eigenfunctions. Then, the best linear approximation to Y in a function space with finite dimension q is given by projecting $Y - \mu$ onto the space generated by $\{v_1, \dots, v_q\}$. This expansion is obtained by means of the functional principal components analysis (FPCA) and is also known as the Karhunen-Loève expansion of Y truncated at order q . The reader is referred to Loève (1978) or Chiou *et al.* (2003b) for a comprehensive introduction to this topic.

This work aims at deriving a Karhunen-Loève expansion or FPCA which is able to take into account nonparametrically the effect of a quantitative covariate X in order to get a decomposition similar to (1) that incorporates this additional information. The introduction of an additional information in such a framework has not received much attention in the litterature whereas it can be very interesting in many situations. For instance, in medicine, when observing and describing electrocardiogram curves, the medic often knows the age of his patients or the concentration of some components in their blood. These quantitative factors may have a certain impact on the statistical characteristics of the electrocardiogram curves and consequently one could imagine that taking into account properly this additional information can lead to a better representation of this set of curves, adapting for instance the characteristics of the FPCA to the age of the individuals. This can also be a way to detect outliers taking into account the knowledge of some cofactors.

Silverman (1995) suggested a practical approach that could handle this kind of problem with parametric models. The estimation procedure is rather

heavy and parametric models are not always adapted when one does not know in advance what can be the relationship between the dependent functional observations and the covariates. More recently, Chiou *et al* (2003b) considered a general approach that incorporates a covariate effect through a semi-parametric model. The problem was to estimate the number of eggs laid per day by $n = 936$ female Mediterranean fruit flies (see Carey *et al.* 1998 for a description of the experiments and of the data) for a time period restricted to the first 50 days of eggs laying. The mean function, that is to say the number of laid eggs per day during the first 50 days of lifetime, and the Karhunen-Loève basis are estimated on the whole population but the coordinates of a functional observation (an eggs laying curve) in this basis are obtained thanks to a single index model which take into account the covariate effect.

This paper aims at proposing a very simple nonparametric approach that can be of real interest for such studies when the sample size is sufficiently large. Such large data sets of functional observations are not unusual nowadays. For instance the sample size is $n = 936$ for the Mediterranean egg laying curves with 50 equispaced discretization points whereas Cardot *et al.* (2004) look for estimators of the temporal responses of different crops when observing $n = 1209$ pixels at thirty different instants during a year with remote sensing.

Instead of incorporating directly the covariate effect in the Karhunen-Loève expansion, we consider a nonparametric estimator of the conditional expectation and the conditional covariance function. Once the estimators of the conditional expectation and the conditional covariance function are obtained, we can derive directly estimators of the conditional eigenvalues $\lambda_j(x)$ and conditional eigenfunctions $v_j(x, t)$ by means of a spectral decomposition. Thus the basis functions $v_j(x, s)$ are allowed to depend nonparametrically on the covariate effect x . That gives a small dimension functional space which can represent Y in an optimal way knowing that X takes the value x . The estimated conditional mean eggs laying curves are drawn in Figure (1) for the first quartile, the median and the third quartile of the total number of eggs. It can be seen that their shapes clearly depend on the total number of eggs and that is why Chiou *et al.* (2003a) proposed a model based on a multiplicative effect on the mean function which seems to be adapted to that problem. Nevertheless, if we display the first eigenfunctions v_j estimated by the approach described below, we clearly see that the covariance structure also varies when the total number of eggs varies. Thus a more general approach that can into account the effect of the covariate on the structure of

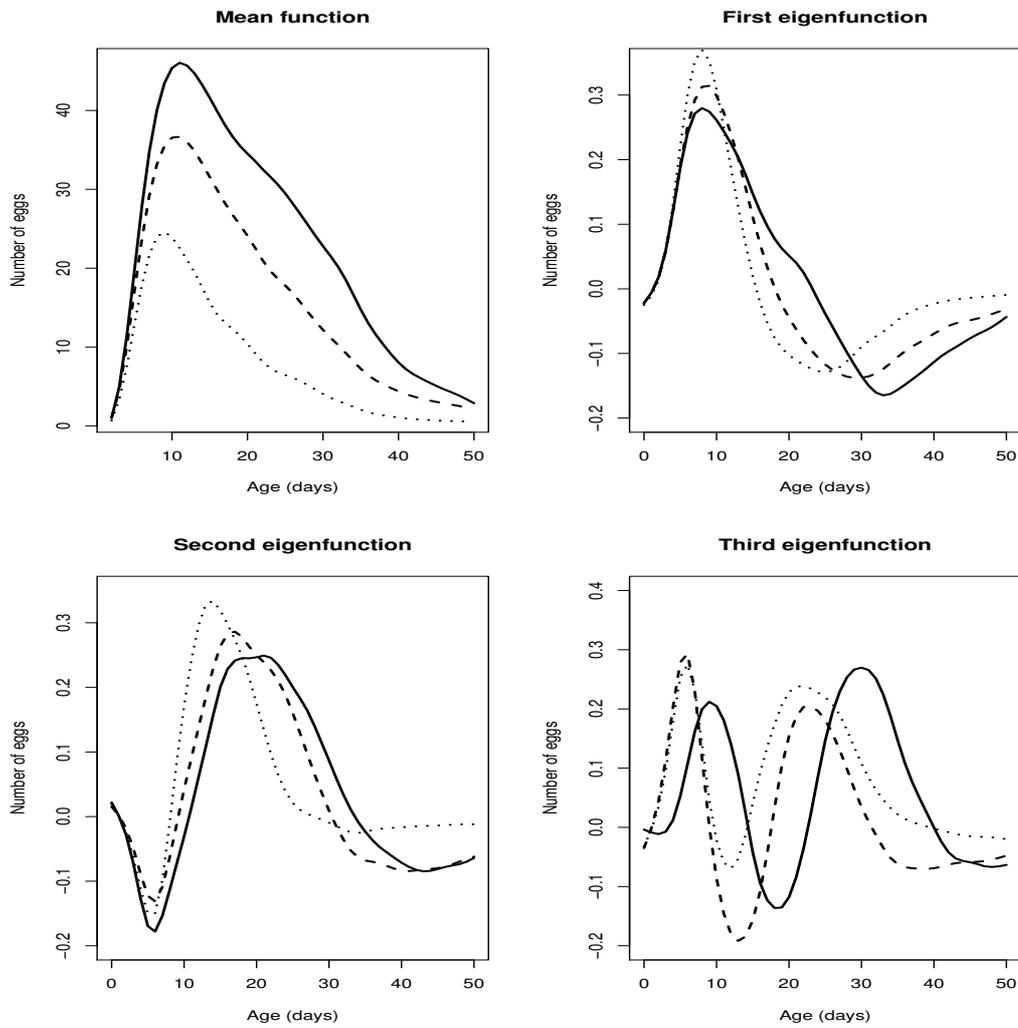


Figure 1: Mean egg laying curves and first, second and third eigenfunctions for the first (\cdots), second ($- -$) and third ($-$) quartiles of the total number of eggs. Bandwidths values are selected by minimizing a cross-validation criterion (see (7) and (17)).

the covariance function seems to be even more adapted by expanding the random functions in low dimension spaces depending on the value taken by the covariate.

We first present in section 2 a general estimation procedure of the regression function based on kernel smoothing when the response is functional and the covariate is real. In this framework, Lecoutre (1990) proposed an estimator based on statistically equivalent blocks methods but he did not give any rates of convergence. More recently, Ferré and Yao (2004) proposed an estimator of the covariance function of the conditional expectation, which is not what we are looking for in this paper, in the framework of sliced inverse regression for functional data. Convergence rates for the regression function estimator derived with our kernel smoothing approach are given and we obtain the usual nonparametric rates of convergence. In section 3, we show how this procedure can be adapted to the estimation of the conditional FPCA whereas Section 4 illustrates with a brief simulation study the good behaviour of the estimator. R programs are available on request to the author. In section 5 we propose a discussion about possible extensions. The proofs are gathered in section 6.

2 Nonparametric regression with functional response

Consider a sample $(X_i, Y_i), i = 1, \dots, n$ of *i.i.d.* realizations of (X, Y) where the X_i 's take values in a compact interval of \mathbb{R} and the Y_i 's belong to $H = L^2(T)$, the space of square integrable functions defined on the compact interval T . The inner product in H is denoted by $\langle \cdot, \cdot \rangle_H$ and the induced norm is denoted $\|\cdot\|_H$.

We want to estimate the regression function defined by

$$r(x, t) = \mathbb{E}(Y(t) | X = x) \quad t \in T. \quad (2)$$

For that purpose we consider the following kernel smoother :

$$\hat{r}(x, t) = \sum_{i=1}^n w_i(x, h) Y_i(t) \quad t \in T, \quad (3)$$

where

$$w_i(x, h) = \frac{K((X_i - x)/h)}{\sum_i K((X_i - x)/h)}, \quad (4)$$

the kernel K is a positive, symmetric around zero and bounded function with compact support and h is the bandwidth. Let us notice that when t

varies in T , $r(x, \cdot)$ and $\widehat{r}(x, \cdot)$ are functions taking values in H . For sake of clarity, the function of t , $r(x, \cdot)$ will be denoted by $r(x)$.

We assume that the conditional expectation satisfies some Lipschitz condition and that $\|Y\|_H$ is bounded. Conditions (H.3) and (H.4) are classical assumptions in nonparametric regression (see *e.g.* Sarda and Vieu 2000).

$$(H.1) \quad \|r(x) - r(z)\|_H \leq c_1 |x - z|^\beta \text{ for some } \beta > 0.$$

$$(H.2) \quad \|Y\|_H \leq c_2 < \infty .$$

(H.3) X has strictly positive density defined on a compact interval. Moreover, it also satisfies a Lipschitz condition with coefficient β .

(H.4) The kernel K is positive, symmetric around zero, with compact support and integrates to one.

The following theorem means that our estimator is consistent, for each fixed value x of the real covariate X .

Theorem 2.1 *Under assumptions (H.1) to (H.4), if $h \rightarrow 0$, $(\log n)/(nh) \rightarrow 0$ as n tends to infinity,*

$$\|r(x) - \widehat{r}(x)\|_H = O\left(\frac{\log n}{nh}\right)^{1/2} + O(h^\beta) \quad a.s.$$

Let us notice that assumptions (H.1) to (H.4) are not as general as they could be but they have the advantage to lead to a short and simple proof. Indeed, assumption H.1 could be replaced by a differentiability hypothesis of function r which would lead to better rates of convergence. Assumption H.2 could be replaced by an hypothesis on the existence of all the moments of Y and then truncation methods would lead to the same rate of convergence. Finally, let us remark that Theorem 2.1 remains true when H is a general separable Hilbert space such as a Sobolev space.

2.1 Discretized data

In practice one observes discretized data, that is to say a sample of n vectors

$$\mathbf{Y}_i = (Y_i(t_1), Y_i(t_2), \dots, Y_i(t_p)), \quad i = 1, \dots, n, \quad (5)$$

where the discretization points $t_1 < t_2 < \dots < t_p$ are supposed in the following to be the same for each curve which is a rather frequent situation. If it is not the case, one can consider for instance a B-splines expansion of the trajectories (Besse *et al.* 1997, James *et al.* 2000) and then deals with the coordinates instead of the observed data. Other approaches based on

kernel smoothing (Staniswalis and Lee, 1998) or local polynomials (Yao *et al.* 2005) can also be adapted to this situation. Assuming the discretization points are the same for each observed curve, it is easy to build estimators for the conditional mean $\mathbf{r}(x) = \mathbb{E}(\mathbf{Y}|X = x)$,

$$\widehat{\boldsymbol{\mu}}(x) = \sum_{i=1}^n w_i(x, h_1) \mathbf{Y}_i \in \mathbb{R}^p, \quad (6)$$

at discretization the points t_1, t_2, \dots, t_p .

Then one needs to choose a reasonable value for the smoothing parameter h_1 . A natural approach is to minimize a cross-validation criterion,

$$\text{CV}(h) = \frac{1}{np} \sum_{i=1}^n \sum_{\ell=1}^p \left(Y_i(t_\ell) - \widehat{\boldsymbol{\mu}}^{-i}(x_i)(t_\ell) \right)^2 \quad (7)$$

where $\widehat{\boldsymbol{\mu}}^{-i}(x_i)(t_\ell)$ is the estimator of $\mu(x_i)$ at $t = t_\ell$ obtained by leaving out the observation (x_i, \mathbf{Y}_i) from the initial sample.

Conditional eggs laying curves

This cross-validation criterion has been used for estimating the conditional mean function of the eggs laying curves with a Gaussian kernel. The covariate X which represents the total number of eggs has been normalized, without loss of generality, in order to belong to the interval $[0, 1]$. Cross-validated mean square errors are given in Table (1) and show that our functional approach performs well compared to those proposed by Chiou *et al.* (2003a, 2003b). Indeed their best predictions, according to the same criterion, have an error taking values around 315.

bandwidths	0.0625	0.0312	0.0156	0.0078	0.0039	0.0020	0.0010	0.0005
CV(h)	243.6	232.9	228.7	227.5	227.7	228.6	229.9	231.8

Table 1: Cross validation scores for estimating the conditional mean egg laying curves.

3 Conditional functional principal components analysis

This section shows how we can derive a conditional functional PCA or conditional Karhunen Loève expansion applying the estimator defined in (3) to

get estimators of the first and the second conditional moment of the variable Y .

3.1 The conditional Karhunen-Loève expansion

The Karhunen-Loève expansion (Castro *et al.* 1986, Cardot *et al.* 1999), also called empirical orthogonal functions in climatology (Preisendorfer and Mobley, 1988), is a direct generalization of the principal components analysis when the observations are realizations of a random function. It can be seen as an optimal linear decomposition, according to a variance criterion, of the random function Y in a finite dimension functional space. This finite dimension space is span by the eigenfunctions associated to the largest values of the covariance operator of Y .

Let us consider the conditional expectation $\mu(x, t) = E(Y(t)|X = x)$ and the conditional covariance function of Y , denoted by

$$\gamma(x, s, t) = \text{Cov}(Y(s), Y(t)|X = x), \quad (s, t) \in T \times T .$$

Then its conditional covariance operator is the integral operator satisfying

$$\Gamma^x f(t) = \int_T \gamma(x, s, t) f(s) ds \quad t \in T. \quad (8)$$

Let us denote by $v_j(x, t)$ the j th orthonormal eigenfunction of Γ^x , associated to the j th largest eigenvalue $\lambda_j(x)$. They satisfy

$$\Gamma^x v_j(x, t) = \lambda_j(x) v_j(x, t) \quad t \in T. \quad (9)$$

with $\lambda_1(x) \geq \lambda_2(x) \geq \dots \geq 0$, and $\langle v_j(x), v_{j'}(x) \rangle = 1$ if $j = j'$ and zero else. Then one can decompose the covariance function

$$\gamma(x, s, t) = \sum_{j \geq 1} \lambda_j(x) v_j(x, s) v_j(x, t), \quad (s, t) \in T \times T . \quad (10)$$

which is the analogous of the decomposition (1) of the covariance function incorporating conditional information. Since Y knowing that $X = x$ is a bounded function it is clear that $\sum_j \lambda_j(x) < \infty$ and the eigenvalues tend rapidly to zero as j goes to infinity. This means that a small dimensional space can generally represent well the main conditional modes of variation of Y .

Indeed, the best linear representation denoted by \tilde{Y}_q of Y knowing that $X = x$ in a q dimensional space is given by

$$\tilde{Y}_q(t | X = x) = \mu(x, t) + \sum_{j=1}^q \langle Y - \mu(x), v_j(x) \rangle v_j(x, t), \quad t \in T \quad (11)$$

meaning that the $v_j(x, t)$'s give an idea of the main conditional modes of variation of Y with associated variance

$$\mathbb{E} \left(\langle Y - \mu(x), v_j(x) \rangle^2 \mid X = x \right) = \lambda_j(x) .$$

This expansion can also be interpreted as a mixed functional effects model (James *et al.* 2000, Rice & Wu 2001), the trajectories of Y being expanded in a deterministic basis $(v_1(x), \dots, v_q(x))$ with random components $\langle Y - \mu(x), v_j(x) \rangle$, $j = 1, \dots, q$, which are also called principal components. The main innovation being here that we take into account nonparametrically the effect of the covariate X through the spectral decomposition of the conditional variance. This allows for instance to determine how the main modes of variation of Y vary according to X by comparing the shape of the functions $v_j(x, t)$ associated to the largest eigenvalues for different values of x as shown in Figure (1).

It is easy to build estimators of the conditional mean and conditional covariance function. The estimator of the conditional mean function is defined as before

$$\hat{\mu}(x, t) = \sum_{i=1}^n w_i(x, h_1) Y_i(t), \quad t \in T. \quad (12)$$

To define the estimator of the covariance function, let us introduce the tensor product notation. For two functions $(Z_i, Y_i) \in L^2(T) \times L^2(T)$, it is the bivariate function $Z_i \otimes Y_i$ belonging to the functional space $\mathcal{H} = L^2(T \times T)$ such that $Z_i \otimes Y_i(s, t) = Z_i(s)Y_i(t)$, for all $(s, t) \in T \times T$. Then, we get can build an estimator of the covariance function as follows

$$\hat{\gamma}(x, s, t) = \sum_{i=1}^n w_i(x, h_2) (Y_i - \hat{\mu}(x)) \otimes (Y_i - \hat{\mu}(x))(s, t), \quad (s, t) \in T \times (\mathbf{T3})$$

This being done, we can deduce estimators of the eigenfunctions and eigenvalues when $X = x$ by considering the conditional covariance operator $\hat{\Gamma}^x$,

$$\hat{\Gamma}^x f(s) = \int_T \hat{\gamma}(x, s, t) f(t) dt . \quad (14)$$

and performing its spectral decomposition (or eigen-analysis) :

$$\hat{\Gamma}^x \hat{v}_j(x, t) = \hat{\lambda}_j(x) \hat{v}_j(x, t) \quad (15)$$

with $\hat{\lambda}_1(x) \geq \hat{\lambda}_2(x) \geq \dots \geq 0$, and the orthonormality constraints $\langle \hat{v}_j(x), \hat{v}_{j'}(x) \rangle = 1$ if $j = j'$ and zero else.

3.2 Discretized data

When we have discrete observations, which is always the case for real data sets, the discretized mean function is estimated by (6) whereas the discretized covariance operator can be approximated as follows

$$\widehat{\mathbf{\Gamma}}^x = \sum_{i=1}^n w_i(x, h_2) (\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}^x) (\mathbf{Y}_i - \widehat{\boldsymbol{\mu}}^x)' . \quad (16)$$

We need to introduce a quadrature rule in order to estimate the eigenelements of the covariance operator with discrete data (Rice and Silverman, 1991, Staniswalis and Lee, 1998). Let us consider some quadrature weights m_1, \dots, m_p such that

$$\int_T f(t) dt \approx \sum_{\ell=1}^p m_\ell f(t_\ell) .$$

Define \mathbf{M} to be the diagonal matrix with diagonal elements $[\mathbf{M}]_{\ell,\ell} = m_\ell$. The discretized eigenfunctions $\widehat{\mathbf{v}}_j(x)$ are obtained by performing the eigenanalysis of the matrix $\widehat{\mathbf{\Gamma}}^x$ with metric \mathbf{M} :

$$\widehat{\mathbf{\Gamma}}^x \mathbf{M} \widehat{\mathbf{v}}_j(x) = \widehat{\lambda}_j(x) \widehat{\mathbf{v}}_j(x) ,$$

where $(\widehat{\mathbf{v}}_j(x))' \mathbf{M} \widehat{\mathbf{v}}_j(x) = 1$ if $j = j'$ and zero else.

Considering $\widehat{\mathbf{Z}}_i = \mathbf{Y}_i - \widehat{\boldsymbol{\mu}}(x_i)$, where the value of the bandwidth h_1 in $\widehat{\boldsymbol{\mu}}(x_i)$ has been chosen by cross-validation (see (7)), we can define a cross-validation criterion for the covariance function

$$CV(h) = \frac{1}{np^2} \sum_{i=1}^n \sum_{\ell,\ell'}^p \left\{ \mathbf{Z}_i(t_\ell) \mathbf{Z}_i(t_{\ell'}) - \widehat{\mathbf{\Gamma}}^{x_i, -i}(t_\ell, t_{\ell'}) \right\}^2 \quad (17)$$

to choose values for the bandwidth h_2 where $\widehat{\mathbf{\Gamma}}^{x_i, -i}$ is the estimated covariance matrix by leaving-out the observation (x_i, \mathbf{Y}_i) from the initial sample.

Covariance structure of the eggs laying curves

This cross-validation procedure was used to determine the value of h_2 in the estimation of the conditional covariance function of the eggs laying curves. The first, second and third eigenfunctions are drawn in Figure (1) for three different values (first quartile, median and third quartile) of the total number of eggs. They appear to be different and it seems that there is a kind of translation of the eigenfunctions to the right as x increases. We also notice

that the estimated conditional eigenvalues, that is to say a measure of the explained variance, also vary with x . The first eigenfunction explains 31 %, the second 17 % and the third 8 % of the total variation of Y when x is the first quartile of the total number of eggs. The first eigenfunction explains 35 %, the second 14 % and the third 7 % of the total variation when x is the median of the total number of eggs. The first eigenfunction explains 37 %, the second 12 % and the third 7 % of the total variation when x is the third quartile of the total number of eggs.

To end this section, let us also notice that adding a smoothing step, in the estimation procedure of the mean function and the eigenelements can lead to better estimates (Rice and Silverman 1991, Silverman 1996, Cardot 2000).

3.3 Some consistency properties

Let us consider the usual norm in $\mathcal{H} = L^2(T \times T)$ as a criterion error. It is defined by

$$\|\gamma(x)\|_{\mathcal{H}}^2 = \int_T \gamma(x, s, t)^2 ds dt \quad (18)$$

We assume now that the second order moment conditional function $r_2(x, s, t) = \mathbb{E}(Y(s)Y(t)|X = x)$ satisfies a Lipschitz condition

$$(H.5) \quad \|r_2(x) - r_2(z)\|_{\mathcal{H}} \leq C|x - z|^\alpha \text{ for some } \alpha > 0.$$

This assumptions means that a small variation of x implies a small variation of the covariance function and does not seem to be very restrictive.

The following proposition shows that we get a consistent estimator of the true conditional covariance operator.

Theorem 3.1 *Under assumptions (H.1) to (H.5), if $\sup(h_1, h_2) \rightarrow 0$, $(\log n)/(n \min(h_1, h_2)) \rightarrow 0$ as n tends to infinity,*

$$\|\mu(x) - \hat{\mu}(x)\|_H = O(h_1^\beta) + O\left(\frac{\log n}{nh_1}\right)^{1/2}, \quad a.s$$

and

$$\|\gamma(x) - \hat{\gamma}(x)\|_{\mathcal{H}} = O(h_1^\beta) + O(h_2^\alpha) + O\left(\frac{\log n}{n \min(h_1, h_2)}\right)^{1/2}, \quad a.s .$$

Thus, we can deduce that we obtain consistent estimators for the eigenelements of the conditional covariance operator. We need to assume that the

conditional eigenvalues are distincts and strictly positive for the eigenfunctions to be identifiable.

$$(H.6) \quad \lambda_1(x) > \lambda_2(x) > \dots > 0 .$$

Since the eigenfunctions are uniquely determined up to a sign change, we choose, without loss of generality, to consider $\widehat{v}_j(x)$ such that $\langle \widehat{v}_j(x), v_j(x) \rangle \geq 0$.

Corollary 3.1 *Under assumptions (H.1) to (H.6), if $\sup(h_1, h_2) \rightarrow 0$, $(\log n)/(n \min(h_1, h_2)) \rightarrow 0$ as n tends to infinity,*

$$\sup_j |\widehat{\lambda}_j(x) - \lambda_j(x)| = O(h_1^\beta) + O(h_2^\alpha) + O\left(\frac{\log n}{n \min(h_1, h_2)}\right)^{1/2}, \quad a.s.$$

and there exists a strictly positive constant C such that for each j ,

$$\|\widehat{v}_j(x) - v_j(x)\|_H \leq C\delta_j \left[h_1^\beta + h_2^\alpha + \left(\frac{\log n}{n \min(h_1, h_2)}\right)^{1/2} \right] \quad a.s.$$

where $\delta_1 = 2\sqrt{2}(\lambda_1(x) - \lambda_2(x))^{-1}$

and for $j \geq 2$, $\delta_j = 2\sqrt{2} \max [(\lambda_{j-1}(x) - \lambda_j(x))^{-1}, (\lambda_j(x) - \lambda_{j+1}(x))^{-1}]$.

Let us remarks that, for a fixed sample size n , the estimation of the eigenfunctions are getting poorer and poorer as j increases since δ_j is generally an increasing sequence.

4 A simulation study

We perform a small simulation study to evaluate the capacity of our estimators to get accurate estimations of the conditional mean and the conditional covariance function. This also allows us to see how the estimators are sensitive to the bandwidths values and the sample size.

We considered a real random variable X , drawn from an uniform distribution in $[0, 1]$ and a random function Y defined as follows :

$$Y(t) = X Z_1(t) + (1 - X) Z_2(t) ,$$

where $t \in T = [0, 1]$ and Z_1 and Z_2 are independent random functions such that

- Z_1 is a brownian motion with mean function $\mu_1(t) = \sin(4\pi t)$, $t \in [0, 1]$ and covariance function $\gamma_1(s, t) = \min(s, t)$.

- Z_2 is a Gaussian process with mean function $\mu_2(t) = \cos(4\pi t)$, $t \in [0, 1]$ and covariance function $\gamma_2(s, t) = \min(1 - s, 1 - t)$.

It is easy to see that, for all $(s, t) \in T \times T$,

$$\begin{cases} \mu(x, t) &= x \sin(4\pi t) + (1 - x) \cos(4\pi t) \\ \gamma(x, s, t) &= x^2 \gamma_1(s, t) + (1 - x)^2 \gamma_2(s, t) \end{cases}$$

We considered samples of n independent realizations of (X, Y) . The realizations of the random function Y are discretized in $p = 50$ equispaced sampling points in $[0, 1]$, denoted by t_1, t_2, \dots, t_p .

The following normalized quadratic criterions are used to measure the estimation error for the estimator (6) of the mean function

$$\text{MSE}(\hat{\mu}_{h_1}^x) = \frac{\sum_{\ell=1}^p (\mathbb{E}(Y(t_\ell)|X = x) - \hat{\mu}_{h_1}^x(t_\ell))^2}{\sum_{\ell=1}^p (\mathbb{E}(Y(t_\ell)|X = x))^2}$$

and for the estimator (16) of the covariance function

$$\text{MSE}(\hat{\gamma}_{h_2}^x) = \frac{\sum_{\ell, \ell'=1}^p (\text{Cov}(Y(t_\ell), Y(t_{\ell'})|X = x) - \hat{\gamma}_{h_2}^x(t_\ell, t_{\ell'}))^2}{\sum_{\ell, \ell'=1}^p (\text{Cov}(Y(t_\ell), Y(t_{\ell'})|X = x))^2}.$$

We performed a hundred of simulations for two different sample sizes, $n = 100$ and $n = 500$. The mean values of the MSE criterions are gathered in Tables 2, 3 and 4 for different values of the bandwidths h_1 and h_2 . The results are presented for X taking the value $x = 0.6$ but as a matter of fact they do not vary much provided x is not too close to the edges of the interval $[0, 1]$.

	bandwidth values for h_1						
sample size	1	0.5	0.25	0.5^4	0.5^5	0.5^6	0.5^7
100	0.045	0.036	0.026	0.019	0.019	0.025	0.035
500	0.032	0.024	0.013	0.006	0.004	0.005	0.006

Table 2: Estimation error for the mean function of Y conditional to $x = 0.6$ for different sample sizes and different bandwidth values.

Let us first notice that even for moderate sample size, *i.e.* when $n = 100$, the estimators perform well for the mean function (see Table 2) and the covariance function (see Table 3) provided that the bandwidth values are reasonable.

bandwidth values for h_2	bandwidth values for h_1						
	0.5^2	0.5^3	0.5^4	0.5^5	0.5^6	0.5^7	0.5^8
0.5	0.211	0.144	0.102	0.094	0.112	0.146	0.197
0.5^2	0.209	0.139	0.097	0.090	0.108	0.142	0.192
0.5^3	0.212	0.137	0.093	0.087	0.105	0.139	0.188
0.5^4	0.219	0.138	0.092	0.086	0.104	0.136	0.184
0.5^5	0.225	0.141	0.093	0.085	0.102	0.134	0.181
0.5^6	0.233	0.147	0.096	0.086	0.101	0.132	0.177
0.5^7	0.246	0.157	0.103	0.089	0.102	0.131	0.175
0.5^8	0.267	0.174	0.115	0.097	0.107	0.132	0.174
0.5^9	0.297	0.199	0.135	0.112	0.117	0.138	0.175

Table 3: Estimation error for the covariance function of Y conditional to $x = 0.6$ for a sample size $n = 100$ and different bandwidth values.

Another important conclusion of this simulation study is that one may choose different values for the bandwidth h_1 associated to the estimator of the mean function. A kind of diagonal structure appears in Table 3 meaning that the choice of h_1 has an impact on the best value for h_2 . The best values of h_1 for estimating the mean function seem to be also the best values for the mean when estimating the covariance function. They must have, in this example, the same order of magnitude.

One can also notice in Table 4 that when the sample size is large enough, the choice of a value for h_2 seems to be of second importance, particularly if h_1 is well chosen, since the values of the criterion error do not vary much according to h_2 .

Figure 2 shows that the estimators are very close to the true conditional mean function and not too far from the true conditional covariance function. As said before, a smoothing procedure of the eigenfunctions has been introduced in order to get better estimations.

5 Concluding remarks

We have proposed in this work a simple and powerful tool to analyse functional data when auxiliary information is available. This approach can be used in many practical studies since nowadays it is frequent to have a huge amount of data.

This short paper is first step to conditional FPCA and one can imag-

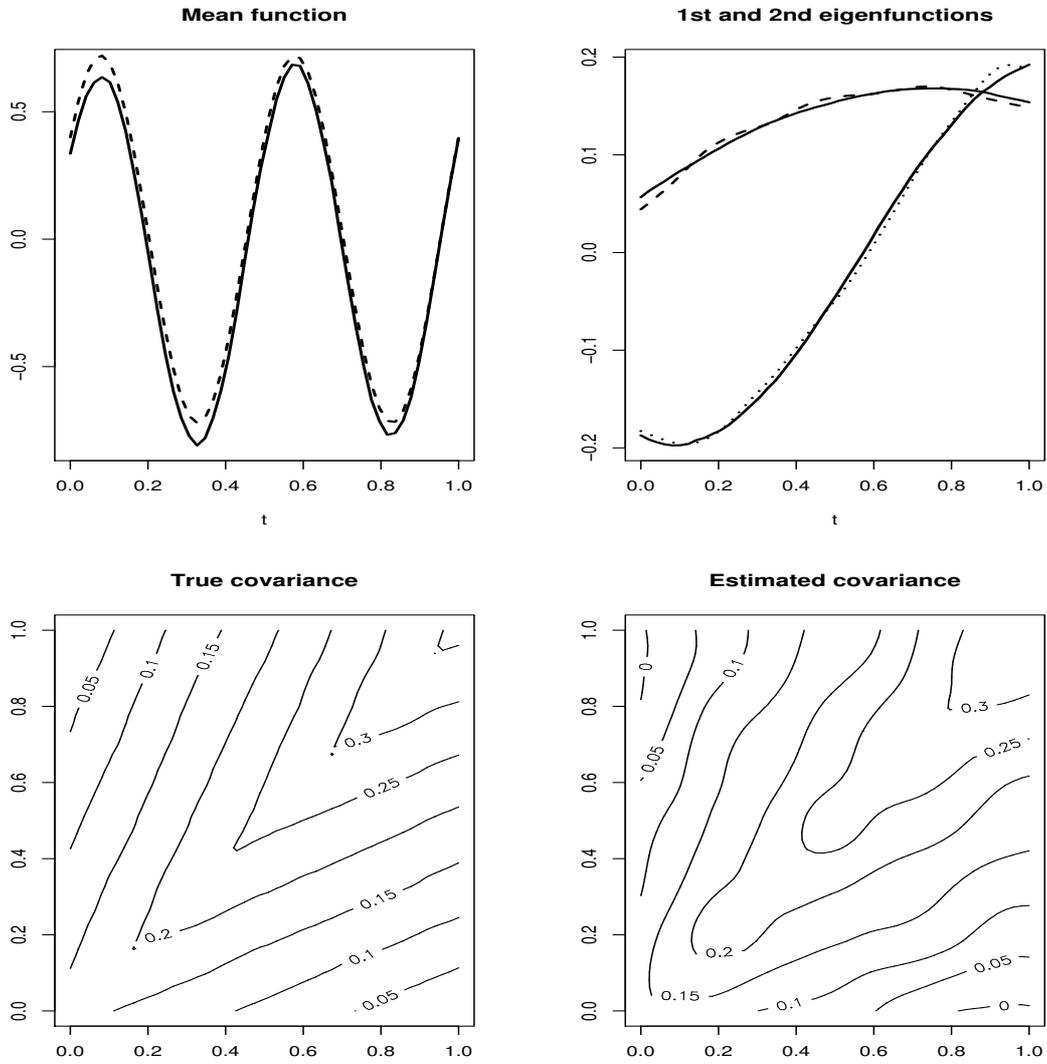


Figure 2: An example when the sample size $n = 500$, with $x = 0.6$ for $h_1 = 0.5^6$ and $h_2 = 0.5^7$. Estimation of the mean function is drawn in (a) with dotted lines and the first eigenvector and second eigenvector in (b). The true covariance function and its conditional estimation are drawn in (c) and (d).

bandwidth values for h_2	bandwidth values for h_1						
	0.5^2	0.5^3	0.5^4	0.5^5	0.5^6	0.5^7	0.5^8
0.5	0.166	0.104	0.056	0.032	0.027	0.030	0.037
0.5^2	0.164	0.096	0.048	0.027	0.023	0.027	0.034
0.5^3	0.167	0.094	0.045	0.025	0.022	0.026	0.033
0.5^4	0.172	0.095	0.044	0.024	0.021	0.025	0.033
0.5^5	0.174	0.096	0.044	0.024	0.021	0.025	0.033
0.5^6	0.175	0.097	0.045	0.024	0.021	0.025	0.033
0.5^7	0.177	0.098	0.045	0.024	0.021	0.025	0.033
0.5^8	0.179	0.100	0.046	0.025	0.021	0.025	0.033
0.5^9	0.183	0.102	0.048	0.026	0.022	0.026	0.034

Table 4: Estimation error for the covariance function of Y conditional to $x = 0.6$ for a sample size $n = 500$ and different bandwidth values.

ine many extensions. For instance it is immediate to extend this approach to multivariate conditional information and one can consider additive models to manage with the curse of dimensionality. One can also extend the conditional FPCA to functional auxiliary information by adapting the kernel methods proposed by Ferraty and Vieu (2002, 2004). Such an approach could also be used for prediction of autoregressive functional processes (Besse *et al.* 2000, Damon and Guillas 2002) allowing to incorporate at the same time a functional and real covariate effects.

It remains for the method to be effective in practice to define data driven criterions that allow to choose reasonable values for the bandwidths h_1 and h_2 . Cross-validation seems to be a good candidate but a detailed study of its theoretical performances is beyond the scope of this paper and deserves further attention.

6 Proofs

Proof of Theorem 2.1.

Let us define $\hat{g}(x) = \frac{1}{nh} \sum_i K((X_i - x)/h) Y_i$ and $\hat{f}(x) = \frac{1}{nh} \sum_i K((X_i - x)/h)$, then

$$\hat{r}(x) = \frac{\hat{g}(x)}{\hat{f}(x)}.$$

We can write

$$r(x) - \hat{r}(x) = \frac{g(x) - \hat{g}(x)}{\hat{f}(x)} + \left(\hat{f}(x) - f(x) \right) \frac{r(x)}{\hat{f}(x)}. \quad (19)$$

The beginning of the proof is rather classical and we get under previous assumptions that (see *e.g* Sarda and Vieu 2000) :

$$\widehat{f}(x) - f(x) = O(h^\beta) + O(\sqrt{\log n/(nh)}), \quad a.s \quad (20)$$

On the other hand, reminding that $g = rf$ also satisfies the Lipschitz condition in H and considering $z = (u - x)/h$ we get directly for the bias, under (H.1) and (H.4) :

$$\begin{aligned} \mathbb{E}\widehat{g}(x) - g(x) &= \frac{1}{h} \int (K((u-x)/h)Y - g(x)) f(u) du \\ &= \int_{\mathbb{R}} (g(x-zh) - g(x)) K(z) dz \\ &= O(h^\beta). \end{aligned} \quad (21)$$

The variance term $\widehat{g}(x) - \mathbb{E}\widehat{g}(x)$ is dealt with exponential inequalities for separable Hilbert space valued random variables. Defining

$$\Delta_i = \frac{1}{h} (K((X_i - x)/h) Y_i - \mathbb{E}K((X - x)/h) Y)$$

it is easy to see that $\|\Delta_i\| \leq C/h$ and $\mathbb{E}\|\Delta_i\|^2 \leq \frac{C}{h}$ under assumptions (H.1) to (H.3). Applying Yurinskii's Lemma (Yurinskii, 1976), we obtain

$$P[\|\widehat{g}(x) - \mathbb{E}\widehat{g}(x)\|_H > \epsilon] \leq 2 \exp\left(-\frac{n\epsilon^2 h}{4C}\right)$$

which is a convergent sequence if we take $\epsilon = \epsilon_0 (\log n/(nh))^{1/2}$, and thus by the Borel-Cantelli Lemma, we get that

$$\|\widehat{g}(x) - \mathbb{E}\widehat{g}(x)\|_H = O\left(\frac{\log n}{nh}\right)^{1/2} \quad a.s. \quad (22)$$

which concludes the proof combining (20), (21) and (22) in (19). \square

Proof of Theorem 3.1.

Under previous assumptions we get directly by Theorem 2.1

$$\|\widehat{\mu}(x) - \mu(x)\|_H = O(h_1^\beta) + O\left(\frac{\log n}{nh_1}\right)^{1/2}, \quad a.s. \quad (23)$$

Let us introduce now the empirical counterpart

$$\widehat{r}_2(x, s, t) = \sum_{i=1}^n w_i(x, h_2) Y_i(s) Y_i(t),$$

of the second order moment function $r_2(x, s, t) = \mathbb{E}(Y(s)Y(t) \mid X = x)$. Let us decompose the estimator of the covariance function as follows

$$\widehat{\gamma}(x) = \widehat{r}_2(x) - \widetilde{\mu}(x) \otimes \widehat{\mu}(x) - \widehat{\mu}(x) \otimes \widetilde{\mu}(x) + \widehat{\mu}(x) \otimes \widehat{\mu}(x) \quad (24)$$

where

$$\widetilde{\mu}^x = \sum_{i=1}^n w_i(x, h_2) Y_i .$$

Let us notice, that under assumptions on the bandwidth h_2 we have by Theorem 2.1,

$$\|\widetilde{\mu}(x) - \mu(x)\|_H = O(h_2^\beta) + O\left(\frac{\log n}{nh_2}\right)^{1/2}, \quad a.s. \quad (25)$$

Looking now at the difference

$$\begin{aligned} \gamma(x) - \widehat{\gamma}(x) &= r_2(x) - \mu(x) \otimes \mu(x) - \widehat{r}_2(x) + \widetilde{\mu}(x) \otimes \widehat{\mu}(x) + \widehat{\mu}(x) \otimes (\widetilde{\mu}(x) - \widehat{\mu}(x)) \\ &= r_2(x) - \widehat{r}_2(x) - \mu(x) \otimes \mu(x) + \widetilde{\mu}(x) \otimes \widehat{\mu}(x) + \widehat{\mu}(x) \otimes (\widetilde{\mu}(x) - \widehat{\mu}(x)) \end{aligned} \quad (26)$$

we get that

$$\begin{aligned} \|\gamma(x) - \widehat{\gamma}(x)\|_{\mathcal{H}} &\leq \|r_2(x) - \widehat{r}_2(x)\|_{\mathcal{H}} + \|\widetilde{\mu}(x) \otimes \widehat{\mu}(x) - \mu(x) \otimes \mu(x)\|_{\mathcal{H}} \\ &\quad + \|\widehat{\mu}^x \otimes (\widetilde{\mu}^x - \widehat{\mu}^x)\|_{\mathcal{H}} \end{aligned} \quad (27)$$

Applying Theorem 2.1 to the new functional observations $Z_i(s, t) = Y_i(s)Y_i(t)$ we get directly that

$$\|\widehat{r}_2(x) - r_2(x)\|_{\mathcal{H}} = O(h_2^\alpha) + O\left(\frac{\log n}{nh_2}\right)^{1/2} \quad a.s. \quad (28)$$

On the other hand, expanding

$$\widetilde{\mu}(x) \otimes \widehat{\mu}(x) - \mu(x) \otimes \mu(x) = (\widetilde{\mu}(x) - \mu(x)) \otimes \widehat{\mu}(x) - \mu(x) \otimes (\widehat{\mu}(x) - \mu(x))$$

we get by (23) and (25)

$$\begin{aligned} \|\widetilde{\mu}(x) \otimes \widehat{\mu}(x) - \mu(x) \otimes \mu(x)\|_{\mathcal{H}} &\leq \|\widetilde{\mu}(x) - \mu(x)\|_H \|\widehat{\mu}(x)\|_H + \|\mu(x)\|_H \|\widehat{\mu}(x) - \mu(x)\|_H \\ &= O(h_1^\beta) + O(h_2^\alpha) + O\left(\frac{\log n}{n \min(h_1, h_2)}\right)^{1/2}, \quad a.s. \end{aligned} \quad (29)$$

since $\|\widehat{\mu}(x)\|_H$ is bounded by a positive constant under assumption (H.2).

Noticing finally that $\widetilde{\mu}(x) - \widehat{\mu}(x) = \widetilde{\mu}(x) - \mu(x) + \mu(x) - \widehat{\mu}(x)$ we get directly that

$$\begin{aligned} \|\widehat{\mu}(x) \otimes (\widetilde{\mu}(x) - \widehat{\mu}(x))\|_{\mathcal{H}} &\leq \|\widehat{\mu}(x)\|_H (\|\widetilde{\mu}(x) - \mu(x)\|_H + \|\widehat{\mu}(x) - \mu(x)\|_H) \\ &= O(h_1^\beta) + O(h_2^\alpha) + O\left(\frac{\log n}{n \min(h_1, h_2)}\right)^{1/2}, \quad a.s. \end{aligned} \quad (30)$$

which concludes the proof combining (28), (29) and (30) in (27). □

Proof of Corollary 3.1.

The proof of the first part of the Corollary is an immediate consequence of classical properties of the eigenlements of covariance operators. The eigenvalues (see *e.g* Dauxois *et al.* 1982) satisfy:

$$|\widehat{\lambda}_j(x) - \lambda_j(x)| \leq \|\widehat{\Gamma}^x - \Gamma^x\|$$

where the norm $\|\cdot\|$ for operator is the Hilbert-Schmidt norm which is equivalent to the norm in \mathcal{H} for integral operators

$$\begin{aligned} \|\Gamma^x\|^2 &= \int_T \gamma(x, s, t)^2 ds dt \\ &= \|\gamma(x)\|_{\mathcal{H}}^2 . \end{aligned}$$

On the other hand, Lemma 4.3 by Bosq (2000) tells us that

$$\|\widehat{v}_j(x) - v_j(x)\|_H \leq C\delta_j \|\Gamma^x - \widehat{\Gamma}^x\| ,$$

which concludes the proof. □

Acknowledgements: I would like to thank all the participants at the STAPH working group on Functional Statistics in Toulouse for fruitful discussions. I would also like to thank Prof. Hans-Georg Müller who kindly allowed me to illustrate the method with the Mediterranean flies data.

Bibliography

- Besse, P., Cardot, H., and Ferraty, F. (1997). Simultaneous non-parametric regressions of unbalanced longitudinal data. *Computational Statistics & Data Analysis*, **24**, 255-270.
- Besse, P.C and Ramsay, J.O. (1986). Principal component analysis of sampled curves. *Psychometrika*, **51**, 285-311.
- Besse, P.C., Cardot, H. and Stephenson, D.B. (2000). Autoregressive Forecasting of Some Functional Climatic Variations. *Scandinavian Journal of Statistics*, **27**, 673-687.

- Bosq, D. (2000). *Linear Processes in Function Spaces*. Lecture Notes in Statistics, 149, Springer.
- Carey, J.R., Liedo, P., Müller, H.G., Wang, J.L., Chiou, J.M. (1998). Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproduction in a large cohort of Mediterranean fruit fly females. *J. of Gerontology Biological Sciences* **53**, 245-251.
- Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional Linear Model. *Statist. & Prob. Letters*, **45**, 11-22.
- Cardot, H. (2000). Nonparametric estimation of the smoothed principal components analysis of sampled noisy functions. *Journal of Nonparametric Statistics*, **12**, 503-538.
- Cardot, H., Faivre, R. and Maisongrande, P. (2004). Random Effects Varying Time Regression Models: Application to Remote Sensing. *Compstat 2004 proceedings*, ed. J. Antoch, Physica Verlag, 777-784.
- Castro, P., Lawton, W. and Sylvestre, E. (1986). Principal Modes of Variation for Processes with Continuous Sample Curves. *Technometrics*, **28**, 329-337.
- Chiou, J.M., Müller, H.G., Wang, J.L., Carey, J.R. (2003a). A functional multiplicative effects model for longitudinal data, with application to reproductive histories of female medflies. *Statistica Sinica* **13**, 1119-1133.
- Chiou, J-M., Müller, H.G. and Wang, J.L. (2003b). Functional quasi-likelihood regression models with smooth random effects. *Journal of the Royal Statistical Society, B*, **65**, 405-423.
- Damon, J. and Guillas, S. (2002). The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics*, **13**, 759-774.
- Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a random vector function: some applications to statistical inference. *Journal of Multivariate Analysis*, **12**, 136-154.
- Deville, J.C. (1974). Méthodes statistiques et numériques de l'analyse harmonique. *Ann. Insee*, **15**, 3-104.

- Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Comput. Statist.*, **17**, 545-564.
- Ferraty, F. and Vieu, P. (2004). Nonparametric models for functional data, with application in regression, time series prediction and classification. *Nonparametric Statistics*, **16**, 111-127.
- Ferré, L. and Yao, A.F. (2004). Smoothed Functional Inverse Regression. *Statistica Sinica*, to appear.
- James, G., Hastie, T., and Sugar, C. (2000). *Principal Component Models for Sparse Functional Data*. *Biometrika*, **87**, 587-602.
- Kneip, A. and Utikal, K.J. (2001). Inference for Density Families Using Functional Principal Component Analysis. *J. Am. Statist. Ass.*, **96**, 519-542
- Lecoutre, J.P. (1990). Uniform consistency of a class of regression function estimators for Banach-space valued random variable. *Statistics & Probability Letters*, **10**, 145-149.
- Loève, M. (1978). *Probability Theory*, Springer, New-York.
- Preisendorfer, R. and Mobley, C. (1988). *Principal Components Analysis in Meteorology and Oceanography*. Elsevier : Amsterdam.
- Ramsay, J. O. and Silverman, B.W. (1997). *Functional Data Analysis*. Springer-Verlag.
- Ramsay, J.O. and Silverman, B.W. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer-Verlag.
- Rice, J. and Silverman, B.W. (1991). Estimating the Mean and Covariance Structure Nonparametrically when the Data are Curves. *Journal of the Royal Statistical Society, B*, **53**, 233-243.
- Rice, J. and Wu, C. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics*, **57**, 253-259.
- Sarda, P. and Vieu, P. (2000). Kernel regression. in *Smoothing and Regression: Approaches, computation and application*, Ed M.G. Schimek, 43-70, Wiley Series in Probability and Statistics.
- Silverman, B.W. (1995). Incorporating parametric effects into functional principal components analysis. *Journal of the Royal Statistical Society, B*, **57**, 673-689.

- Silverman, B.W. (1996). Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics*, **24**, 1-24.
- Staniswalis, J.G., Lee, J.J. (1998). Nonparametric Regression Analysis of Longitudinal Data. *J.A.S.A.*, **93**, 1403-1418.
- Yao, F., Müller, H.G. and Wang, J.L. (2005). Functional Data Analysis for Sparse Longitudinal Data. *J.A.S.A.*, to appear.
- Yurinskii, V.V. (1976). Exponential Inequalities for Sums of Random Vectors. *Journal of Multivariate Analysis*, **6**, 473-499.