Conditional functional principal components analysis

Hervé CARDOT
CESAER, UMR INRA-ENESAD.

March 27, 2006

Abstract

This work proposes an extension of the functional principal components analysis, or Karhunen-Loève expansion, which can take into account non-parametrically the effects of an additional covariate. Such models can also be interpreted as non-parametric mixed effects models for functional data. We propose estimators based on kernel smoothers and a data-driven selection procedure of the smoothing parameters based on a two-steps cross-validation criterion. The conditional functional principal components analysis is illustrated with the analysis of a data set consisting of egg laying curves for female fruit flies. Convergence rates are given for estimators of the conditional mean function and the conditional covariance operator when the entire curves are collected. Almost sure convergence is also proven when one only observes discretized noisy sample paths. A simulation study allows us to check the good behavior of the estimators.

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

1 Introduction

Since the pioneer work by Deville (1974) much attention has been given to functional data analysis in the statistical community (see *e.g* Ramsay & Silverman, 2002, 2005 and references therein). Many publications are devoted to the statistical description of a sample of curves (*e.g.* growth curves, temperature curves, spectrometric curves) by means of the functional principal components analysis (see *e.g.* Besse & Ramsay 1986; Castro *et al.*, 1986; Kirkpatrick & Heckman, 1989; Rice & Silverman 1991, Kneip & Utikal

2001). Performing the spectral decomposition of the empirical covariance operator, which is the analogue of the covariance matrix in a function space, allows us to get a low dimensional space which exhibits, in an 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 index t varies in a compact interval T of $\mathbb{R}$, with mean $\mu(t) = \mathbb{E}(Y(t))$ and covariance function $\gamma(s,t) = \mathrm{Cov}(Y(s), Y(t)),\ s \in T$. Under general conditions (see *e.g.* Loève, 1978), the covariance function may be expressed as follows

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

where the λ_j are the ordered eigenvalues, $\lambda_1 \geq \lambda_2 \geq ... \geq 0$, of the covariance operator and the functions v_j the associated orthonormal eigenfunctions. Then, the best linear approximation $\widetilde{Y}^q$ to Y in a function space with finite dimension q is given by projecting the centered random function $Y - \mu$ onto the space generated by $\{v_1, \ldots, v_q\}$,

$$\widetilde{Y}^q(t) \quad = \quad \mu(t) + \sum_{j=1}^{q} c_j \ v_j(t) \ , \tag{2}$$

where the random coordinates $c_j = \int_T (Y(t) - \mu(t)) v_j(t) dt$, also called principal components (Dauxois *et al.*, 1982), are centered with variance $\mathrm{var}(c_j) = \lambda_j$. This expansion is also known as the Karhunen-Loève expansion of Y truncated at order q. The reader is referred to Loève (1978), Kirkpatrick & Heckman (1989) or Chiou *et al.* (2003b) for a comprehensive introduction on this topic.

This work aims at deriving a Karhunen-Loève expansion or functional principal components analysis (FPCA) which is able to take into account non-parametrically the effect of a quantitative covariate X on Y in order to get a decomposition similar to (2) that incorporates this additional information. Conditional on $X = x$, we would like to get the following optimal decomposition

$$\widetilde{Y}^q(x,t) \quad = \quad \mu(x,t) + \sum_{j=1}^{q} c_j(x) \ v_j(x,t) \ , \tag{3}$$

allowing the mean function and the basis functions $v_j(x,t)$ to depend non-parametrically on the covariate effect x.

The introduction of an additional information in such a framework has not received much attention in the literature whereas it can be very interesting in many situations. For instance, in medicine, when observing and describing electrocardiogram curves, physicians often know the age of their 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 adequate 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 egg laying, conditional on the covariate X which is the total number of eggs laid during the period. 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, *i.e.* the principal components, of an egg laying curve in this basis are obtained thanks to a single index model which take into account the covariate effect. A sample of 80 egg laying curves is drawn in Figure (1,(a)), showing a large variability in their shapes. This example will serve as an illustration of the proposed methodology.

Figure 1 is around here

This paper aims at proposing a simple non-parametric 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: there are 936 observations in the egg laying curves data set, sampled at 50 equispaced design points whereas Cardot *et al.* (2004) deal with $n = 1209$ coarse resolution pixels observed at thirty different instants during a year with remote sensing data. Instead of incorporating directly the covariate effect in the Karhunen-Loève expansion, we consider non-parametric estimators of the conditional expectation and the conditional covariance function. Then, we can derive estimators of the conditional eigenvalues $\lambda_j(x)$ and conditional eigenfunctions $v_j(x, t)$ by means of a spectral decomposition. The estimated conditional mean egg laying curves are drawn in Figure

(1, (b)) for the first quartile, the median and the third quartile of the total number of laid 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 on the mean function based on a multiplicative effect which seems to be adapted to that problem. Nevertheless, if we display the first eigenfunctions v_j estimated by the approach described below, we also see that the covariance structure varies when the total number of eggs laid by each fly varies. Thus a more general approach that can take into account the effect of the covariate on the structure of the covariance function seems to be even more adapted by expanding the random functions in low dimension spaces whose basis functions depend on the values taken by the covariate.

We first present in section 2 a general description of the conditional principal components analysis and give estimators of the conditional mean function and conditional covariance operators based on kernel smoothers. When dealing with real data, one can not assume anymore the curves are observed on the whole interval but at a finite number of design points and may be corrupted by noise. We propose to first approximate the discrete trajectories with non parametric smoothers in order to get curves. This is a common practice (see *e.g.* Ramsay & Silverman, 2005) when the design points are not too sparse, which is often the case with real data. The conditional moments estimators depend on smoothing parameters and we propose a two steps cross-validation criterion in order to select good values. In section 3, convergence rates are given and we obtain the usual non-parametric rates of convergence when the curves are observed entirely without noise. In a similar framework of Hilbert-valued dependent variable, Lecoutre (1990) proposed an estimator of the regression based on statistically equivalent blocks methods but he did not give any rates of convergence. For noisy and sampled data, we also show that the estimators remain consistent under classical hypotheses. In Section 4, a simulation study confirms the good behavior of the estimators and allows us to evaluate the impact of the different smoothing parameters on their accuracy. R programs are available on request to the author. In section 5, we discuss possible improvements and extensions. The proofs are gathered in the appendix.

2 Conditional functional principal components analysis

Consider a sample $(X_i, Y_i), i = 1, \ldots, n$ of *i.i.d.* realizations of (X, Y) where X is a real random variable and Y is a random variable taking values in

$H = L^2(T)$, the space of square integrable functions defined on the compact interval T. The inner product in H is denoted by $< .,. >_H$ and the induced norm is denoted by $\|.\|_H$. We first suppose that the trajectories are observed entirely , *i.e.* for every t in T, and are not corrupted by noise.

2.1 The conditional Karhunen-Loève expansion

The Karhunen-Loève expansion (Loève, 1978; Castro *et al.*, 1986), also called empirical orthogonal functions in climatology (Preisendorfer & Mobley, 1988), is a direct extension 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 spanned by the eigenfunctions associated to the largest values of the covariance operator of Y.

Assuming $E(\|Y\|^2 \; |X = x) < \infty$, we can define the conditional expectation $\mu(x,t) = E(Y(t)|X = x)$ and the conditional covariance operator Γ^x,

$$\Gamma^x f(t) \;\; = \;\; \int_T \gamma(x,s,t) f(s) \; ds \quad t \in T, \tag{4}$$

where

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

is the conditional covariance function of Y. Let us denote by $v_j(x,t)$ the jth orthonormal eigenfunction of Γ^x, associated to the jth largest eigenvalue $\lambda_j(x)$. The eigenelements satisfy

$$\Gamma^x \; v_j(x,t) \;\; = \;\; \lambda_j(x) \; v_j(x,t) \quad t \in T. \tag{5}$$

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

$$\gamma(x,s,t) \;\; = \;\; \sum_{j \geq 1} \lambda_j(x) \; v_j(x,s) v_j(x,t), \quad (s,t) \in T \times T \ . \tag{6}$$

which is the analogue of the decomposition (1) of the covariance function incorporating conditional information. Since $E(\|Y\|^2 \; |X = x) < \infty$, it is clear that $\sum_j \lambda_j(x) < \infty$ and the eigenvalues tend rapidly to zero as j goes to infinity.

Then, the best linear representation denoted by $\widetilde{Y}_q$ of Y conditional on $X = x$ in a q dimensional space is given by

$$\widetilde{Y}_q(x,t) \;\; = \;\; \mu(x,t) + \sum_{j=1}^{q} < Y - \mu(x), v_j(x) > v_j(x,t), \quad t \in T \tag{7}$$

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

$$\mathbb{E}\left(< Y - \mu(x), v_j(x) >^2 |X = x\right) = \lambda_j(x) \ .$$

Furthermore, it can be checked easily that, for every dimension q,

$$\mathbb{E}\left(\|Y - \widetilde{Y}_q(x)\|^2 |X = x\right) \ \leq \ \mathbb{E}\left(\|Y - \widetilde{Y}_q\|^2\right), \tag{8}$$

where the overall approximation $\widetilde{Y}_q$ is defined in (2). 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 $(v_1(x), \ldots, v_q(x))$ with random coordinates $< Y - \mu(x), v_j(x) >, \ j = 1, \ldots, q$, which are also called principal components. The main innovation proposed here is that we take into account non-parametrically the effect of a covariate X through the conditional mean function and the spectral decomposition of the conditional variance. This allows us 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 (2, (c) and (d)).

2.2 Non-parametric estimators of the conditional moments

The conditional mean $\mu(x,t)$ and the conditional covariance function are estimated with kernel smoothers. Let us consider a kernel K which is positive, bounded, symmetric around zero and has compact support. A natural estimator of $\mu(x,t)$ is given by

$$\widetilde{\mu}(x,t) \ = \ \sum_{i=1}^n w_i(x,h_1) Y_i(t), \quad t \in T, \tag{9}$$

where the weights, which depend on a bandwidth h, are defined by

$$w_i(x,h) \ = \ \frac{K\left((X_i - x)/h\right)}{\sum_i K\left((X_i - x)/h\right)} \ . \tag{10}$$

For sake of clarity, the function of t, $\mu(x,.)$ will be denoted by $\mu(x) \in L^2(T)$.

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

$$\widetilde{\gamma}(x,s,t) \ = \ \sum_{i=1}^n w_i(x,h_2) \left(Y_i - \widetilde{\mu}(x)\right) \otimes \left(Y_i - \widetilde{\mu}(x)\right)(s,t), \quad (s,t) \in T \times T \tag{11}$$

Estimators of the eigenfunctions and eigenvalues when $X = x$ are obtained by considering the conditional covariance operator $\widetilde{\Gamma}^x$,

$$\widetilde{\Gamma}^x f(s) \quad = \quad \int_T \widetilde{\gamma}(x,s,t) f(t) \; dt \; , \tag{12}$$

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

$$\widetilde{\Gamma}^x \; \widetilde{v}_j(x,t) \quad = \quad \widetilde{\lambda}_j(x) \; , \widetilde{v}_j(x,t) \tag{13}$$

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

2.3 Discretized curves

With real data we do not observe the whole curves but discretized trajectories, generally supposed to be noisy,

$$y_{i\ell} \quad = \quad Y_i(t_{i\ell}) + \epsilon_{i\ell}, \quad \ell = 1, \ldots, p_i \tag{14}$$

at design points $t_{i1} < t_{i2} < \cdots < t_{ip_i}$ which may vary from one trajectory to another and where $\epsilon_{i,\ell}$ is a white noise, $\mathbb{E}(\epsilon_{i\ell}) = 0$ and $\mathbb{E}(\epsilon_{i\ell}^2) = \sigma_i^2$. In this general context, one can consider for instance a B-splines expansion of the trajectories (Besse *et al.*, 1997, Cardot, 2000; James *et al.*, 2000) and then deal with the coordinates instead of the observed data. Other approaches based on kernel smoothing (Staniswalis & Lee, 1998) or local polynomials (Yao *et al.* 2005) can also be adapted to this situation.

Assuming that the number of design points p_i is large enough, we can get smooth approximations to the discretized observed curves by applying classical non-parametric estimators based on local polynomials, kernel smoothers, wavelets, smoothing splines or regression splines. Let us denote by $\widehat{Y}_i(t)$ the non-parametric approximation to $Y_i(t)$ obtained by regressing the noisy discretized curves $(t_{i\ell}, y_{i\ell})$, $j = 1, \ldots, p_i$.

Then, we can build estimators of the conditional mean and covariance functions as follows,

$$\widehat{\mu}(x,t) = \sum_{i=1}^n w_i(x,h) \widehat{Y}_i(t)$$

and

$$\widehat{\gamma}(x,s,t) \quad = \quad \sum_{i=1}^n w_i(x,h_2) \left(\widehat{Y}_i - \widehat{\mu}(x) \right) \otimes \left(\widehat{Y}_i - \widehat{\mu}(x) \right) (s,t), \quad (s,t) \in T \times T \tag{15}$$

This pre-smoothing step allows us to sample all curves at the same design points and then to use quadrature rules in order to approximate integrals by summations (see *e.g.* Rice & Silverman, 1991).

2.4 Selecting smoothing parameter values

Then one needs to choose reasonable values for the smoothing parameters h_1 and h_2. A natural approach consists in looking for the bandwidth values that minimize a prediction error. We consider a two-steps cross-validation criterion similar to the one proposed by Chiou *et al.* (2003b).

We first look for the best bandwidth value for the conditional mean by minimizing, according to h_1,

$$\mathrm{CV}_\mu(h_1) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}\sum_{\ell=1}^{p_i}\left(y_i(t_{i\ell}) - \widehat{\mu}^{-i}(x_i, t_{i\ell})\right)^2 \tag{16}$$

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

$$\widehat{\mu}^{-i}(x_i, t_{i\ell}) = \sum_{k\neq i}\frac{w_k(x_i, h_1)}{\sum_{k'\neq i} w_{k'}(x_i, h_1)}\widehat{Y}_k(t_{i\ell})$$

The estimator of the conditional mean associated to the optimal bandwidth is denoted by $\widehat{\mu}_{CV}$.

In a second step, we minimize the prediction error of $y_i(t_{i\ell})$ in a q dimensional space, q being fixed in advance, with the cross-validation criterion

$$\mathrm{CV}_\gamma(h_2) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{p_i}\sum_{\ell=1}^{p_i}\left(y_i(t_{i\ell}) - \widehat{Y}_i^{q,-i}(t_{i\ell}))\right)^2 \tag{17}$$

where

$$\widehat{Y}_i^{q,-i}(t_{i\ell}) = \widehat{\mu}_{CV}(x_i, t_{i\ell}) + \sum_{j=1}^{q}\widehat{c}_{ij}\ \widehat{v}_j^{-i}(x_i, t_{i\ell}),$$

$\widehat{c}_{ij} = \int(\widehat{Y}_i(t) - \widehat{\mu}_{CV}(x_i, t))\widehat{v}_j^{-i}(x_i, t)dt$ and $\widehat{v}_j^{-i}(x_i, t_{i\ell})$ is the jth eigenfunction of

$$\widehat{\Gamma}_{x_i}^{-i} = \sum_{k\neq i}^{n}\frac{w_k(x_i, h_2)}{\sum_{k'\neq i}^{n} w_{k'}(x_i, h_2)}\left(\widehat{Y}_k - \widehat{\mu}_{CV}(x_i)\right)\otimes\left(\widehat{Y}_k - \widehat{\mu}_{CV}(x_i)\right)$$

evaluated at $X = x_i$ and $t = t_{i\ell}$.

2.5 Conditional functional principal components analysis of the egg laying curves

The data consist of $n = 936$ egg laying curves of Mediterranean fruit flies observed daily during the first 50 days of egg laying. The issue of determining how reproductive patterns are associated with overall reproductive success, measured by the total number of laid eggs during the period, is of real interest (Chiou *et al.*, 2003b).

The original curves are rather rough and a pre-smoothing step was performed by kernel regression using a Gaussian kernel. As noticed by Kneip & Utikal (2001) or Chiou *et al.* (2003b) under-smoothing seems to lead to better estimation in this framework and we consider, for each curve, four smoothed functional approximations based on individual smoothing parameters $h_{i,cv}$ chosen by minimizing a classical cross-validation criterion as well as under-smoothed approximations taking the bandwidths $h_{i,cv}/2$, $h_{i,cv}/3$ and $h_{i,cv}/6$.

The covariate X which represents the total number of eggs has been normalized, without loss of generality, in order to take values in the interval $[0,1]$. Before normalization, the mean number of eggs laid was 801, the first quartile was 386, the median was 793 and the third quartile 1170. The minimum value was 2 and the maximum was 2349.

Table 1 is around here

Cross-validated mean square errors using the predictive mean squared error criterion (16) are given in Table (1). We first notice that even if there are no large differences between pre-smoothing approaches, functional approximations to the discretized curves obtained with small bandwidth values lead to better predictions and no smoothing should be preferred to usual smoothing procedures based on cross-validation. It appears that the most important tuning parameter is the parameter h_1 which controls the dependence of the mean function on the covariate X. Taking h_1 around 0.004 leads to a prediction error less than 228 for "under-smoothed" curves. If we consider the unconditional mean function, the leave out one curve criterion gives a prediction error around 352. Let us also remark that our functional approach performs well compared to those proposed by Chiou *et al.* (2003a, 2003b) whose best prediction error is around 315, according to the same criterion.

Table 2 is around here

The cross-validation procedure (17) was used to determine the value of h_2 and cross-validation scores are given in Table (2). We also remark that a pre-smoothing step performed with small bandwidth values lead to better prediction but, as before, the most important tuning parameter seems to be h_2 which controls the effect of the covariate X on the covariance function. The first and second conditional eigenfunctions are drawn in Figure (2, (c) and (d)) for three different values (first quartile, median and third quartile) of the total number of eggs and individual pre-smoothing steps performed with $h_{i,cv}/3$. A comparison with the overall eigenfunctions clearly indicates a kind of "lag" in the dimension, meaning that the first eigenfunction try to capture the vertical shift information brought by the covariate which is already included in the conditional mean function. On the other hand, the second overall eigenfunction has roughly the same shape as the first conditional eigenfunctions. If we compare now the conditional eigenfunctions, it clearly appear that their shapes are different, and they take larger values, for fixed ages greater than 30 days, as the covariate increases. This means that larger variations occur at the end of the time interval when the number of laid eggs is large.

Figure 2 is around here

We also noticed that the estimated conditional eigenvalues, which give a measure of the conditional explained variance, also vary with x. The first eigenfunction explains 48 %, the second 22 % and the third 8 % of the total variation of Y when x takes values around the first quartile of the total number of eggs. The first eigenfunction explains 52 %, the second 19 % and the third 8 % of the total variation for x around the median value of the total number of eggs. The first eigenfunction explains 56 %, the second 15 % and the third 7 % of the total variation for x close to third quartile of the total number of eggs.

3 Some consistency properties

We assume that the conditional expectation $\mu(x,t)$ satisfies some Lipschitz condition and that $\|Y\|_H$ is bounded. Conditions (H.3) and (H.4) are classical assumptions in non-parametric regression.

(H.1) $\|\mu(x) - \mu(z)\|_H \leq c_1 |x - z|^\beta$ for some $\beta > 0$.

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

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

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

We also assume 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) $\|r_2(x) - r_2(z)\|_{\mathcal{H}} \leq C|x - z|^\alpha$ 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.

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

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

Next proposition shows that we get consistent estimators of the conditional mean and covariance operator, for each fixed value x of the real covariate X.

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

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

and

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

In order to obtain consistent estimators for the eigenelements of the conditional covariance operator, we need to assume that the conditional eigenvalues are distinct and strictly positive for the eigenfunctions to be identifiable.

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

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

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

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

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

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

where $\delta_1 = 2\sqrt{2}(\lambda_1(x) - \lambda_2(x))^{-1}$
and for $j \geq 2$, $\delta_j = 2\sqrt{2} \max\left[(\lambda_{j-1}(x) - \lambda_j(x))^{-1}, (\lambda_j(x) - \lambda_{j+1}(x))^{-1} \right]$.

Let us remark 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.

Next proposition shows that even when the curves are corrupted by noise at the design points, one can still get consistent approximations to the conditional functional principal components analysis. Nevertheless, convergence rates are harder to obtain and are not given here since they will depend on many ingredients such as the regularity of the trajectories, the design of the discretization points and the way the conditional mean and covariance function are estimated. We only suppose here that the following hypothesis is fulfilled

(H.7) $\max_i \left\| Y_i - \widehat{Y}_i \right\|_H \to 0 \quad a.s.$

Such an assumption is satisfied under general conditions for classical non-parametric smoothers such as kernels, local polynomials or smoothing splines. It assumes implicitly that the grid of the design points gets finer and finer and the trajectories are regular enough.

Theorem 2 *Under assumptions (H.1) to (H.7), if* $\sup(h_1, h_2) \to 0$, $(\log n)/(n \min(h_1, h_2)) \to 0$ *as n tends to infinity,*

$$\|\widehat{\mu}(x) - \mu(x)\|_H \to 0 \quad a.s$$

and

$$\|\widehat{\gamma}(x) - \gamma(x)\|_{\mathcal{H}} \to 0 \quad a.s \ .$$

Thus, for each j

$$\left|\widehat{\lambda}_j - \lambda_j\right| \to 0 \quad a.s$$

and

$$\|\widehat{v}_j(x) - v_j(x)\|_H \to 0 \quad a.s \ .$$

4 A simulation study

We propose now to perform a simulation study in order to evaluate the ability of our estimators to get accurate estimations of the conditional mean and the conditional covariance function. This allows us to see how the estimators are sensitive to the bandwidths values and the sample size. We also generate discretized noisy sample paths in order to evaluate what amount of smoothing should be preferred when performing non-parametric regression of the trajectories.

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

$$Y(t) \quad = \quad X\ Z_1(t) + (1-X)\ Z_2(t)\ , \tag{19}$$

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 $(s,t) \in T \times T$,

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

We make 100 replications of model (19) considering two different sample sizes, $n = 100$ and $n = 500$. In the implementation, the realizations of the random function Y are discretized at $p = 50$ equispaced design points in $[0,1]$, denoted by $t_1, t_2, \ldots, t_p$ and we use a quadrature rule to approximate integrals by summations. The following normalized quadratic criterions are used to measure the estimation error for the estimator (9) of the mean function

$$\mathrm{MSE}(\widehat{\mu}^x_{h_1}) \quad = \quad \frac{\sum_{\ell=1}^p \left(\mathbb{E}(Y(t_\ell)|X=x) - \widehat{\mu}^x_{h_1}(t_\ell)\right)^2}{\sum_{\ell=1}^p \left(\mathbb{E}(Y(t_\ell)|X=x)\right)^2} \tag{20}$$

and for the estimator (11) of the covariance function

$$\mathrm{MSE}(\widehat{\gamma}^x_{h_2}) \quad = \quad \frac{\sum_{\ell,\ell'=1}^p \left(\mathrm{Cov}(Y(t_\ell), Y(t_{\ell'})|X=x) - \widehat{\gamma}^x_{h_2}(t_\ell, t_{\ell'})\right)^2}{\sum_{\ell,\ell'=1}^p \left(\mathrm{Cov}(Y(t_\ell), Y(t_{\ell'})|X=x)\right)^2}\ . \tag{21}$$

The mean values of the MSE criterions are gathered in Tables 3, 4 and 5 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 $[0,1]$.

Table 3 is around here

Table 4 is around here

Let us first note that even for moderate sample sizes, *i.e.* when $n = 100$, the estimators perform well for the mean function (see Table 3) and the covariance function (see Table 4) provided that the bandwidth values are reasonable.

Another important remark 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 4 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.

Table 5 is around here

One can also notice in Table 5 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 criterion error does not vary much according to h_2.

Figure 3 shows that the estimators are very close to the true conditional mean function and not too far from the true conditional covariance function. A smoothing procedure of the eigenfunctions described in Rice & Silverman (1991) has also been introduced in order to get better estimations. Even if not presented here, we have remarked that the conditional eigenvalues are also well estimated.

Figure 3 is around here

4.1 Discretized noisy data

A second aim of this simulation study is to evaluate the impact of pre-smoothing noisy sampled trajectories on the accuracy of estimators. We now generate 100 replications of n sampled curves each corrupted by noise at p equispaced design points in $[0, 1]$,

$$y_i(t_\ell) = Y_i(t_\ell) + \epsilon_{i\ell}, \quad \ell = 1, \ldots, p, \ i = 1, \ldots n,$$

where $\epsilon_{i\ell}$ are *i.i.d.* realizations of a centered gaussian variable with variance $\sigma^2 = 0.05$.

We compare the effect of different pre-smoothing steps of the noisy curves on the estimation of the conditional mean as well as the approximation error of Y at the design points in a $q = 3$ dimensional space. The estimators obtained when observing $\mathbf{Y}_i$ without noise serve as a benchmark. For each noisy discrete trajectory $\mathbf{y}_i$, pre-smoothing is done with kernel smoothers and we consider three different individual bandwidth values, $h_{i,cv}$ those obtained by minimizing a classical cross-validation criterion, $h_{i,cv}/2$ and $h_{i,cv}/6$ which lead to "under-smoothing".

The following criterion error

$$MSE(\widehat{\mu}) = \frac{1}{np}\sum_{i=1}^{n}\sum_{\ell=1}^{p}\left(\mu(x_i,t_\ell)-\widehat{\mu}(x_i,t_\ell)\right)^2 \tag{22}$$

allows us to measure a global estimation error of the conditional mean function.

The mean squared errors according to criterion (22) when the bandwidth value h_1 is chosen by cross-validation are given in Table (6). We compare estimators built with the discrete non noisy trajectories Y, the noisy trajectories y and the pre-smoothed trajectories.

Table 6 is around here

It first appears that even if the data are noisy, estimation errors are close to those obtained with estimators built with non noisy data. We can also remark that under-smoothing in the pre-smoothing step, *i.e.* by choosing the individual bandwidths $h_{i,cv}/6$, can lead to a slight improvement of the estimators compared to no smoothing at all, that is to say by performing the estimation using directly the noisy data.

We also compute the following criterion

$$MSE(\widehat{Y}^q) = \frac{1}{np}\sum_{i=1}^{n}\sum_{\ell=1}^{p}\left(Y_i(t_\ell)-\widehat{Y}_i^q(x_i,t_\ell))\right)^2 \tag{23}$$

to measure the approximation error of the true discrete trajectories in a q dimension space. Approximation errors according to criterion (23) for $q=3$ when the bandwidths h_1 and h_2 are selected by the cross-validation criterions (16) and (17) are gathered in Table (7). We first notice that now the pre-smoothing steps can lead to a moderate improvement of the estimation, specially when the sample size is small, meaning that we get better estimations of the conditional eigenfunctions by incorporating a smoothing procedure. When the sample size is large ($n=500$), all the estimation errors have similar mean values and are close to those obtained with non noisy data.

Table 7 is around here

It appears in this simulation study that even if there is no real gain in performing a pre-smoothing step when one is interested by the estimation of the conditional mean function, there can be a non negligible gain by performing a pre-smoothing step when one is interested in estimating the conditional eigenfunctions. Furthermore, when the design points vary from one curve to another, pre-smoothing allows us to use quadrature rules to estimate the eigenfunctions by sampling all the curves at the same design points.

At last, even if no smoothing generally gives rather good approximations to the true signal, one should be aware that it can lead to overestimated eigenvalues since it incorporates in the conditional covariance function a "diagonal" term due to the noise variance. Then, adding a pre-smoothing step must always be preferred when one is interesting in measuring the conditional variance associated to the eigenfunctions.

5 Concluding remarks

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

This short paper is first step to conditional FPCA and one can imagine many extensions. For instance it is immediate to extend this approach to multivariate conditional information and one can consider additive models to handle 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 employed to improve the prediction skill of autoregressive functional processes (Besse *et al.*, 2000; Damon and Guillas, 2002) by incorporating simultaneously functional and real covariates effects.

Another interesting issue is to determine if there exists a dependence between Y and the covariate X. Test statistics based on permutation approaches which break down artificially the correlation by resampling can be used successfully in this functional context (Cardot *et al.*, 2006).

We also remarked in the simulation study that the smoothness of the approximation to noisy sampled curves has no real impact on the estimation of the conditional mean whereas it can induce a certain gain when considering the covariance function. The consistency proof proposed in Section 6 relies on rough inequalities and should be dealt with a more precise and careful asymptotic study on the effect of pre-smoothing borrowing for instance ideas from a recent work by Benko *et al.* (2006). Indeed, estimation of the covariance function could certainly be improved by considering a modified estimator which takes into account the fact that a bias term due to the noise variance may appear. More precisely, suppose the pre-smoothing step is performed by linear smoothers, such as kernels, local polynomials or regression splines. Then,

$$\widehat{Y}_i(t) = \sum_{\ell=1}^{p_i} s_i(t, t_\ell, h_i) y_{i\ell},$$

where the weights $s_i(t, t_\ell, h_i)$ depends on some smoothing parameter h_i, and one can consider the following modified estimator of the covariance function,

$$\widehat{\gamma}(x, s, t) = \sum_{i=1}^{n} w_i(x, h_2) \sum_{\ell=1}^{p_i} \sum_{k \neq \ell} s_i(t, t_\ell, h_i) s_i(s, t_k, h_i) y_{i\ell} y_{ik} - \widehat{\mu}(x, t)\widehat{\mu}(x, s) \ ,$$

in order to eliminate the variance terms due to the noise. I believe that this direct approach can provide consistent estimates to the true covariance function even when the number of sampling points is finite, provided their location is random. This issue certainly deserves further attention but is beyond the scope of this paper.

Acknowledgments: I would like to thank all the participants, and specially Luboš Prchal and Prof. Pascal Sarda, 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. I wish to thank the Editor, Associate Editor and two anonymous reviewers for helpful comments that led to several improvements in the paper.

References

Benko, M., Härdle, W. and Kneip, A. (2006). Common functional principal components. *SFB649 Economic Risk Discussion Paper 2006-10*, Humboldt University, Berlin.

Besse, P., Cardot, H., and Ferraty, F. (1997). Simultaneous non-parametric regressions of unbalanced longitudinal data. *Comput. Statist. Data Anal.*, **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. *Scand. J. Statist.*, **27**, 673-687.

Bosq, D. (2000). *Linear Processes in Function Spaces.* Lecture Notes in Statistics, 149, Springer.

Cardot, H. (2000). Nonparametric estimation of the smoothed principal components analysis of sampled noisy functions. *J. Nonparametr. Stat.*, **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.

Cardot, H., Prchal, L. and Sarda, P. (2006). No effect and lack-of-fit permutation tests for functional regression. *Preprint.*

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.

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. *Statist. Sinica* **13**, 1119-1133.

Chiou, J-M., Müller, H.G. and Wang, J.L. (2003b). Functional quasi-likelihood regression models with smooth random effects. *J. Roy. Statist. Soc. Ser. 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. *J. Multivariate Anal.*, **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. *J. Nonparametr. Stat.*, **16**, 111-127.

James, G., Hastie, T., and Sugar, C. (2000). Principal Component Models for Sparse Functional Data. *Biometrika*, **87** , 587-602.

Kirkpatrick, M. and Heckman, N. (1989). A quantitative genetic model for growth, shape, reaction norms and other infinite dimensional characters. *J. Math. Biol.*, **27**, 429-450

Kneip, A. and Utikal, K.J. (2001). Inference for Density Families Using Functional Principal Component Analysis. *J. Amer. Statist. Assoc.*, **96**, 519-542.

Lecoutre, J.P. (1990). Uniform consistency of a class of regression function estimators for Banach-space valued random variable. *Statist. Probab. Lett.*, **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. (2002). *Applied Functional Data Analysis: Methods and Case Studies.* Springer-Verlag.

Ramsay, J. O. and Silverman, B.W. (2005). *Functional Data Analysis.* Springer-Verlag, 2nd ed.

Rice, J. and Silverman, B.W. (1991). Estimating the Mean and Covariance Structure Nonparametrically when the Data are Curves. *J. Roy. Statist. Soc. Ser. 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 Statsitics.

Silverman, B.W. (1995). Incorporating parametric effects into functional principal components analysis. *J. Roy. Statist. Soc. Ser. B*, **57**, 673-689.

Staniswalis, J.G., Lee, J.J. (1998). Nonparametric Regression Analysis of Longitudinal Data. *J. Amer. Statist. Assoc.*, **93**, 1403-1418.

Yao, F., Müller, H.G. and Wang, J.L. (2005). Functional Data Analysis for Sparse Longitudinal Data. *J. Amer. Statist. Assoc.*, **100**, 577-590.

Yurinskiĭ, V.V. (1976). Exponential Inequalities for Sums of Random Vectors. *J. Multivariate Anal.*, **6**, 473-499.

Hervé Cardot, CESAER, UMR INRA-ENESAD, 26, Bd Docteur Petitjean, BP 87999, 21079 Dijon Cedex - France.
email: cardot@enesad.inra.fr

Appendix: Proofs

Proof of Theorem 1.

We first prove the consistency of the conditional mean function estimator. Defining $\widetilde{g}(x) = \frac{1}{nh}\sum_i K\left((X_i - x)/h\right) Y_i$ and $\widetilde{f}(x) = \frac{1}{nh}\sum_i K\left((X_i - x)/h\right)$, we have

$$\widetilde{\mu}(x) = \frac{\widetilde{g}(x)}{\widetilde{f}(x)} \ ,$$

and we can write

$$\mu(x) - \widetilde{\mu}(x) \quad = \quad \frac{g(x) - \widetilde{g}(x)}{\widetilde{f}(x)} + \left(\widetilde{f}(x) - f(x)\right) \frac{r(x)}{\widetilde{f}(x)} \ . \tag{24}$$

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

$$\widetilde{f}(x) - f(x) \quad = \quad O(h^{\beta}) + O(\sqrt{\log n/(nh)}), \quad a.s. \tag{25}$$

On the other hand, since $g = \mu f$ also satisfies the Lipschitz condition in H and considering $z = (u - x)/h$, we get under (H.1) and (H.4),

$$\begin{aligned} \mathbb{E}\widetilde{g}(x) - g(x) \quad &= \quad \frac{1}{h}\int \left(K\left((u-x)/h\right) Y - g(x)\right) \ f(u) \ du \\ &= \quad \int_{\mathbb{R}} \left(g(x - zh) - g(x)\right) K(z) dz \\ &= \quad O(h^{\beta}) \ . \end{aligned} \tag{26}$$

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

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

it is easy to check 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

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

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

$$\|\widetilde{g}(x) - \mathbb{E}\widetilde{g}(x)\|_H \quad = \quad O\left(\frac{\log n}{nh}\right)^{1/2} \quad a.s. \tag{27}$$

Combining (25), (26) and (27) in (24), we get under previous assumptions

$$\|\widetilde{\mu}(x) - \mu(x)\|_H \quad = \quad O(h_1^\beta) + O\left(\frac{\log n}{nh_1}\right)^{1/2}, \quad a.s. \qquad (28)$$

Let us introduce now the empirical counterpart

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

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

$$\widetilde{\gamma}(x) \quad = \quad \widetilde{r}_2(x) - \breve{\mu}(x) \otimes \widetilde{\mu}(x) - \widetilde{\mu}(x) \otimes \breve{\mu}(x) + \widetilde{\mu}(x) \otimes \widetilde{\mu}(x) \qquad (29)$$

where $\breve{\mu}^x = \sum_{i=1}^n w_i(x,h_2) Y_i$. Considering the same decomposition as in (24), we can show under assumptions on the bandwidth h_2 that

$$\|\breve{\mu}(x) - \mu(x)\|_H \quad = \quad O(h_2^\beta) + O\left(\frac{\log n}{nh_2}\right)^{1/2}, \quad a.s. \qquad (30)$$

Looking now at the difference

$$\begin{aligned} \gamma(x) - \widetilde{\gamma}(x) &= r_2(x) - \mu(x) \otimes \mu(x) - \widetilde{r}_2(x) + \breve{\mu}(x) \otimes \widetilde{\mu}(x) + \widetilde{\mu}(x) \otimes (\breve{\mu}(x) - \widetilde{\mu}(x)) \\ &= r_2(x) - \widetilde{r}_2(x) - \mu(x) \otimes \mu(x) + \breve{\mu}(x) \otimes \widetilde{\mu}(x) + \widetilde{\mu}(x) \otimes (\breve{\mu}(x) - \widetilde{\mu}(x)) \qquad (31) \end{aligned}$$

we get that

$$\begin{aligned} \|\gamma(x) - \widetilde{\gamma}(x)\|_{\mathcal{H}} \quad \leq \quad & \|r_2(x) - \widetilde{r}_2(x)\|_{\mathcal{H}} + \|\breve{\mu}(x) \otimes \widetilde{\mu}(x) - \mu(x) \otimes \mu(x)\|_{\mathcal{H}} \\ & + \|\widetilde{\mu}^x \otimes (\breve{\mu}^x - \widetilde{\mu}^x)\|_{\mathcal{H}} \ . \qquad (32) \end{aligned}$$

Considering again the decomposition (24) for the functional observations $Z_i(s,t) = Y_i(s)Y_i(t)$ we get directly that

$$\|\widetilde{r}_2(x) - r_2(x)\|_{\mathcal{H}} \quad = \quad O(h_2^\alpha) + O\left(\frac{\log n}{nh_2}\right)^{1/2} \quad a.s. \qquad (33)$$

On the other hand, expanding

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

we get by (28) and (30)

$$\begin{aligned} \|\breve{\mu}(x) \otimes \widetilde{\mu}(x) - \mu(x) \otimes \mu(x)\|_{\mathcal{H}} \quad \leq \quad & \|\breve{\mu}(x) - \mu(x)\|_H \|\widetilde{\mu}(x)\|_H + \|\mu(x)\|_H \|\widetilde{\mu}(x) - \mu(x)\|_H \\ = \quad & O(h_1^\beta) + O(h_2^\alpha) + O\left(\frac{\log n}{n \min(h_1,h_2)}\right)^{1/2} \quad a.s. \ (34) \end{aligned}$$

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

Remarking now that $\breve{\mu}(x) - \widetilde{\mu}(x) = \breve{\mu}(x) - \mu(x) + \mu(x) - \widetilde{\mu}(x)$ we get directly that

$$\begin{aligned} \|\widetilde{\mu}(x) \otimes (\breve{\mu}(x) - \widetilde{\mu}(x))\|_{\mathcal{H}} &\leq \|\widetilde{\mu}(x)\|_H \left(\|\breve{\mu}(x) - \mu(x)\|_H + \|\widetilde{\mu}(x) - \mu(x)\|_H\right) \\ &= O(h_1^\beta) + O(h_2^\alpha) + O\left(\frac{\log n}{n \min(h_1, h_2)}\right)^{1/2} \end{aligned} \quad (35)$$

which concludes the proof combining (33), (34) and (35) in (32).

Proof of Corollary 1.

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

$$|\widetilde{\lambda}_j(x) - \lambda_j(x)| \leq \left\|\widetilde{\Gamma}^x - \Gamma^x\right\| , \quad (36)$$

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

$$\|\Gamma^x\|^2 = \int_T \gamma(x,s,t)^2 \; ds \; dt = \|\gamma(x)\|_{\mathcal{H}}^2 . \quad (37)$$

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

$$\|\widetilde{v}_j(x) - v_j(x)\|_H \leq C\delta_j \left\|\Gamma^x - \widetilde{\Gamma}^x\right\| , \quad (38)$$

which concludes the proof.

Proof of Theorem 2.

By assumption (H.7), we can bound

$$\begin{aligned} \|\widetilde{\mu}(x) - \widehat{\mu}(x)\|_H &\leq \sum_{i=1}^n w_i(x,h_1) \left\|Y_i - \widehat{Y}_i\right\|_H & (39) \\ &\leq \max_i \left\|Y_i - \widehat{Y}_i\right\|_H \sum_{i=1}^n w_i(x,h_1) \quad \to 0 \;\; a.s. & (40) \end{aligned}$$

Dealing now with the covariance operator, let us study

$$\widetilde{\Gamma} - \widehat{\Gamma} = \sum_{i=1}^n w_i(x,h_2) \left\{ (Y_i - \widetilde{\mu}(x)) \otimes (Y_i - \widetilde{\mu}(x)) - \left(\widehat{Y}_i - \widehat{\mu}(x)\right) \otimes \left(\widehat{Y}_i - \widehat{\mu}(x)\right) \right\} .$$

We have

$$\begin{aligned} &(Y_i - \widetilde{\mu}(x)) \otimes (Y_i - \widetilde{\mu}(x)) - \left(\widehat{Y}_i - \widehat{\mu}(x)\right) \otimes \left(\widehat{Y}_i - \widehat{\mu}(x)\right) \\ &= \left(Y_i - \widehat{Y}_i + \widehat{\mu}(x) - \widetilde{\mu}(x)\right) \otimes (Y_i - \widetilde{\mu}(x)) - \left(\widehat{Y}_i - \widehat{\mu}(x)\right) \otimes \left(\widehat{Y}_i - Y_i + \widetilde{\mu}(x) - \widehat{\mu}(x)\right) \end{aligned}$$

and we get with (H.7),

$$\left\|Y_i - \widehat{Y}_i + \widehat{\mu}(x) - \widetilde{\mu}(x)\right\|_H \leq \left\|Y_i - \widehat{Y}_i\right\|_H + \|\widehat{\mu}(x) - \widetilde{\mu}(x)\|_H \to 0 \ a.s. \tag{41}$$

Then

$$\left\|\Gamma^x - \widehat{\Gamma}^x\right\| \leq \left\|\Gamma^x - \widetilde{\Gamma}^x\right\| + \left\|\widetilde{\Gamma}^x - \widehat{\Gamma}^x\right\| \to 0 \quad a.s. \tag{42}$$

Applying again (36) and (38) the proof is complete.

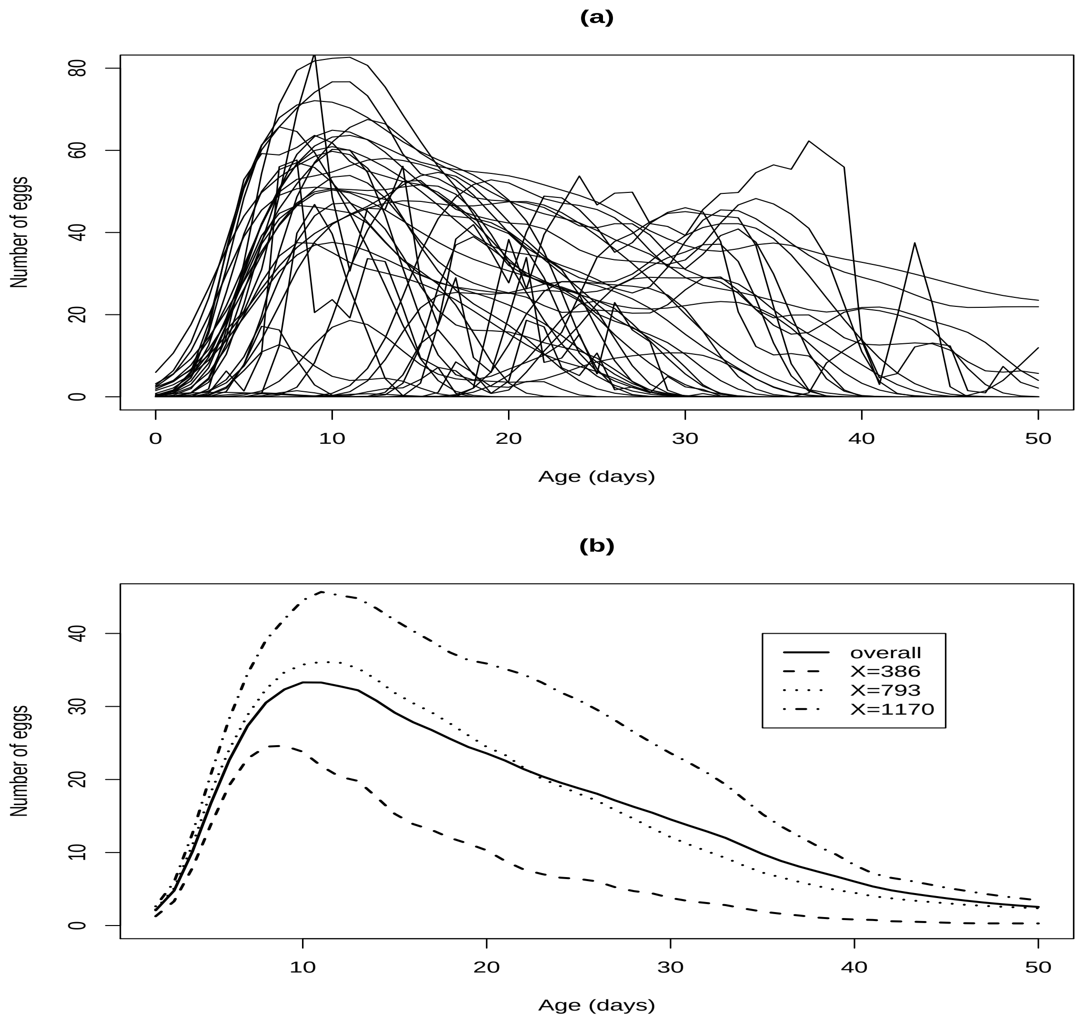

Figure 1: (a): a sample of 80 smoothed egg laying curves. (b) : A comparison of the overall mean egg laying curve with conditional mean egg laying curves estimated for the first ($X = 386$), second ($X = 793$) and third ($X = 1170$) quartiles of the total number of eggs. Bandwidths values are selected by minimizing the cross-validation criterion (16).

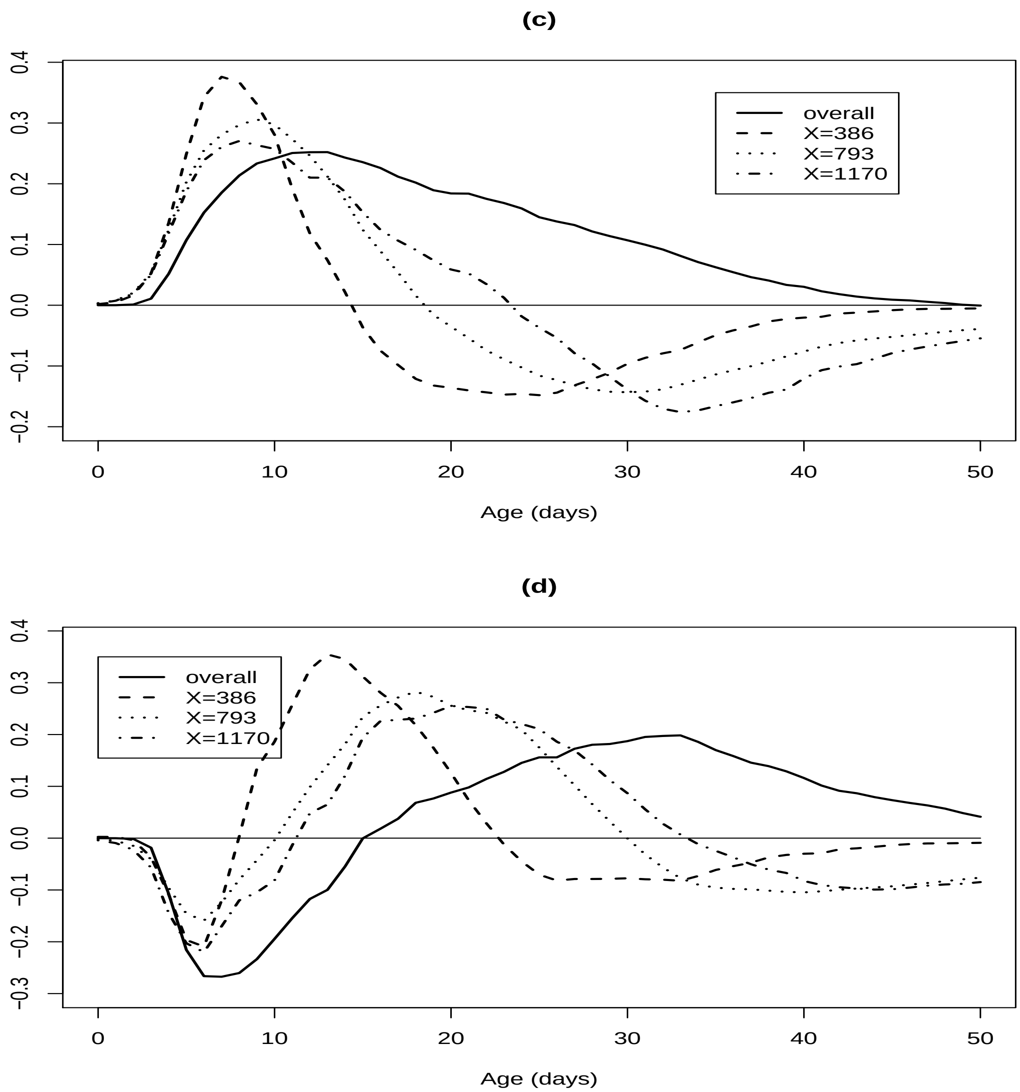

Figure 2: A comparison of the overall eigenfunctions and the conditional ones estimated for the first ($X = 386$), second ($X = 793$) and third ($X = 1170$) quartiles of the total number of eggs. (c) : First overall and conditional eigenfunctions. (d): Second overall and conditional eigenfunctions. Bandwidths values are selected by minimizing the cross-validation criterions (16) and (17).

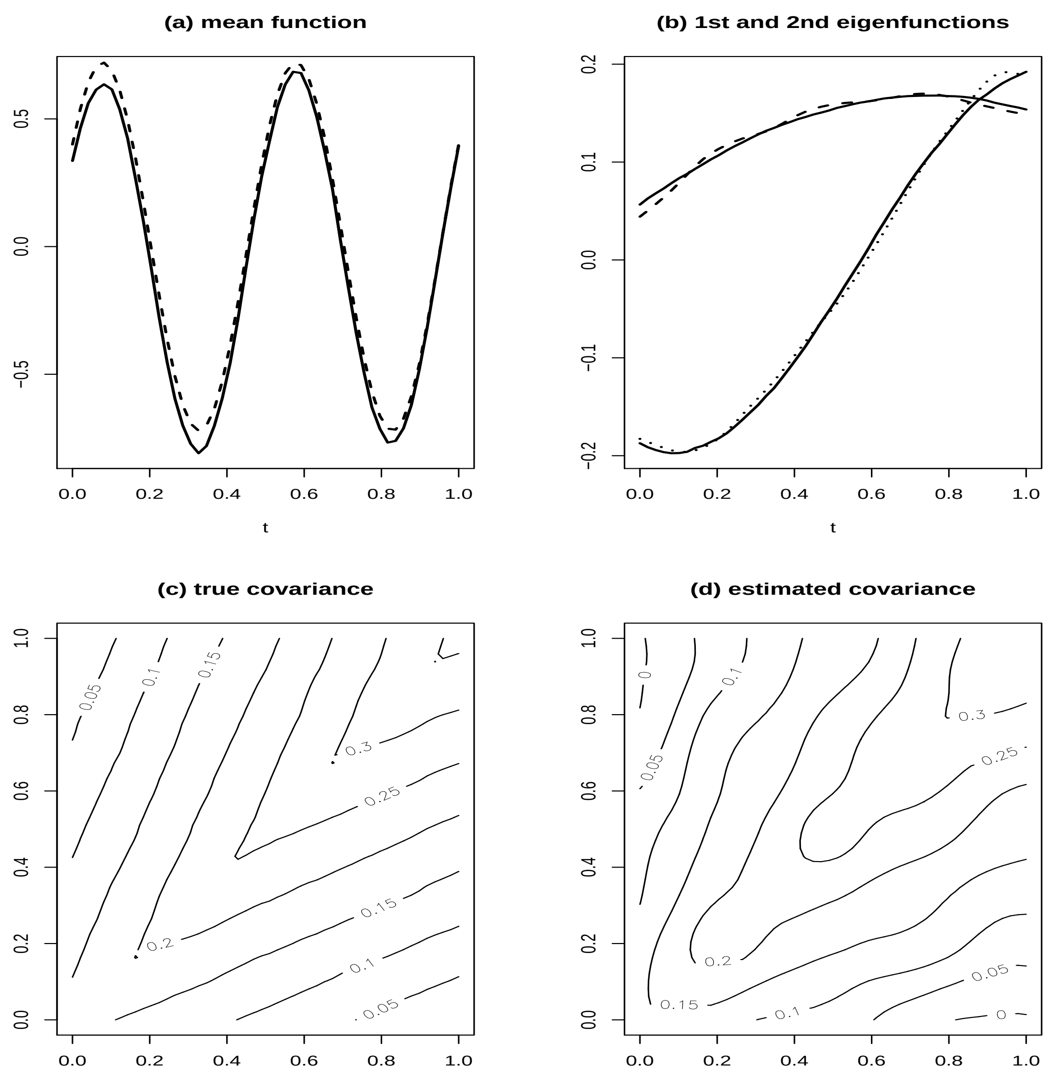

Figure 3: An example when the sample size is $n = 500$, with $x = 0.6$ for $h_1 = 0.5^6$ and $h_2 = 0.5^7$. Estimation of the mean function is drawn in (a) with dotted lines and the first eigenfunction and the second eigenfunction in (b). The true conditional covariance function and its estimate are drawn in (c) and (d).

bandwidth h_1	0.0312	0.0156	0.0078	0.0039	0.0020	0.0010	0.0005
$h_{i,cv}$	234.6	230.1	228.7	228.6	229.0	229.9	230.9
$h_{i,cv}/2$	233.6	229.3	227.9	227.8	228.4	229.3	230.4
$h_{i,cv}/3$	233.2	228.9	227.6	227.5	228.2	229.1	230.4
$h_{i,cv}/6$	233.1	228.8	227.5	227.5	228.1	229.1	230.5
no smoothing	232.9	228.7	227.5	227.7	228.6	229.9	231.8

Table 1: Leaving one curve out prediction error of the conditional mean egg laying curves for different pre-smoothing steps and bandwidth values.

bandwidth h_2	0.0156	0.0078	0.0039	0.0020	0.0010
$h_{i,cv}$	120.2	119.2	119.2	119.7	120.6
$h_{i,cv}/2$	118.1	117.4	118.2	119.2	120.4
$h_{i,cv}/3$	117.1	116.5	116.8	117.6	118.7
$h_{i,cv}/6$	117.4	116.9	117.2	118.2	119.4
no smoothing	117.5	116.9	117.2	118.5	120.2

Table 2: Leaving one curve out prediction error of the approximated egg laying curves in a $q = 2$ dimensional space for different pre-smoothing steps and bandwidth values.

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

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 4: Mean estimation error for the covariance function of Y conditional on $x = 0.6$ for a sample size $n = 100$ and different bandwidth values.

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 5: Mean estimation error for the covariance function of Y conditional on $x = 0.6$ for a sample size $n = 500$ and different bandwidth values.

sample size	y	$h_{i,cv}$	$h_{i,cv}/2$	$h_{i,cv}/6$	Y
100	1.43	1.63	1.47	**1.40**	1.33
500	0.40	0.57	0.45	**0.39**	0.37

Table 6: Mean squared error (×100) for the conditional mean function of Y for different sample sizes and different pre-smoothing steps, y stands for no smoothing at all the noisy data and Y denotes estimations based on non noisy curves. Bandwidth values for h_1 are selected by cross-validation (eq. 16).

sample size	y	$h_{i,cv}$	$h_{i,cv}/2$	$h_{i,cv}/6$	Y
100	3.85	3.64	**3.60**	3.69	3.10
500	0.31	0.31	**0.30**	**0.30**	0.26

Table 7: Mean squared error (×100) for the conditional estimation of Y in a 3-dimensional space ($q = 3$) for different sample sizes and different pre-smoothing steps, y stands for no smoothing at all the noisy data and Y denotes estimations based on non noisy curves. Bandwidth values for h_1 and h_2 are selected by cross-validation.