Varying-Coefficient Functional Linear Regression Models

Hervé Cardot[1] and Pascal Sarda[2]

1) Institut de Mathématiques de Bourgogne, UMR CNRS 5584
Université de Bourgogne
9, avenue Alain Savary - B.P. 47 870, 21078 Dijon, France
email: Herve.Cardot@u-bourgogne.fr

2) Institut de Mathématiques de Toulouse, UMR CNRS 5219
Université Paul Sabatier,
31062 Toulouse cedex 09, France.
email: sarda@cict.fr

Abstract

The paper considers a generalization of the functional linear regression in which an additional real variable influences smoothly the functional coefficient. We thus define a varying-coefficient regression model for functional data. We propose two estimators based respectively on conditional functional principal regression and on local penalized regression splines and prove their pointwise consistency. We check, with the prediction one day ahead of ozone concentration in the city of Toulouse, the ability of such non linear functional approaches to produce competitive estimations.

Short title : Varying-Coefficient Functional Regression

Keywords : Conditional Principal Components Regression, Conditional Covariance Function, Ill-Posed Problem, B-Splines, Penalization.

Corresponding author : Hervé Cardot.

1 Introduction

A well-known and commonly used tool to describe relations between functional predictors and scalar reponses is the so-called Functional Linear Regression introduced by Ramsay and Dalzell (1991) which has been studied recently by numerous authors: see among others Marx and Eilers (1996), Cardot *et al.* (1999), Cuevas *et al.* (2002), Müller and Stadtmüller (2005), Cai and Hall (2006), Hall and Horowitz (2007) and Crambes, Kneip and Sarda (2007) for estimation procedures as well as theoretical results. Linear modelling is attractive from the point of view of interpretability. The drawback is that for some applications, the linear approach may not be flexible enough to capture all the variability of the data. A way to increase the flexibility is to replace the linear functional with a parameter-free unknown functional leading to functional nonparametric regression models as introduced by Ferraty and Vieu (2006): see for a review their recent monograph. We propose here another generalization of the functional linear regression model in which the functional coefficient may vary according to the values of other (scalar) inputs.

For instance, in the example developed below, the aim is to predict the maximum of ozone over a day (the response) explained by the curve of ozone of the preceding day (functional predictor). Prediction of the peak of ozone can be achieved by means of usual functional linear regression. However, one can think that the ozone indicator also depends on other variables such as the wind speed or concentration of other chemical components in the atmosphere. In order to incorporate the effects of an additional variable such as the maximum of hourly NO concentrations during a whole day we present in section 4 an illustration of the ability of the *varying-coefficient model* introduced below to predict ozone concentration one day ahead in the city of Toulouse. Some comparisons with other approaches are made according to the prediction skill.

We come now more precisely to the description of the context of our study. We thus consider a regression problem in which the response Y is a real valued random variable and the predictor X is functional, that is to say the random variable X takes values in some space of functions. In the following we assume that this space is the Hilbert space of square integrable functions, say $H = L^2(T)$, where T is a compact interval of $\mathbb{R}$. Although other (Hilbert) spaces of functions such as Sobolev spaces can be considered, this formalization has the advantage to provide a quite general presentation of the problem.

We propose to consider a generalization of the functional linear regression model which takes the form

$$Y \ = a(Z) + \int_T \alpha(Z, t) \ X(t) \ dt \ + \ \varepsilon, \tag{1}$$

where Z and ϵ are real random variables such that

$$\mathbb{E}\left(\varepsilon|Z\right) \ = \ 0, \qquad \mathbb{E}\left(X\varepsilon|Z\right) \ = \ 0 \quad \text{and} \quad Var(\varepsilon|Z) \ = \ \sigma^2.$$

given a value z of Z, our aim is to estimate the functional coefficient $\alpha(z, .) = \alpha_z(.)$ belonging to H and the mean parameter $a(z) = a_z$. In order to simplify further developments, we choose to consider only an univariate Z noting that the generalization to a multivariate Z is straightforward.

The main difference between model (1) and the functional linear regression introduced by Ramsay and Dalzell (1991) is that the regression function α is now allowed to vary with Z taking into account nonparametrically this new information. Model (1) can be seen as a direct extension of the *varying-coefficient regression model* proposed by Hastie and Tibshirani (1993), considering functions instead of vectors. Let us note that model (1) includes more specific models which appear to be particular cases:

- If $\alpha(Z, t) = \alpha_0(t)$ we get an *additive model*, separating the effects of Z and X on Y, similar to the one introduced by Damon and Guillas (2002).

- If $\alpha(Z, t) = \alpha_0(t)$ and if moreover $a(Z) = a_0$, then we come back to the *functional linear regression model.*

- If $X(t) = 1$, the regression takes the form $a(Z) + \int_T \alpha(Z, t)dt$ which is a kind of nonparametric regression model with Z as the predictor and a time-varying effect.

- If $X(t) = 1$ and $\alpha(Z, t) = \alpha_1(Z)$, we have the usual nonparametric model of regression with Z as a predictor.

Let us introduce some notations. In the following we denote by $\langle \phi, \psi \rangle = \int_T \phi(t)\psi(t)dt$ the usual inner product in H and by $\|\phi\|$ the induced norm. Let us assume at first that X has a finite second conditional moment:

$\mathbb{E}(\|X\|^2|Z = z) < +\infty$. Then we consider the conditional expectations of X and Y denoted by $\eta_z = \mathbb{E}(X|Z = z)$ and by $\mu_z = \mathbb{E}(Y|Z = z)$.

One can easily check that

$$Y - \mu_z = \int \alpha(z,t)(X(t) - \eta_z(t))dt + \varepsilon,$$

and, taking expectations given $Z = z$, this leads to the following moment equation

$$\Gamma_z \alpha_z = \Delta_z, \tag{2}$$

where Γ_z is the conditional covariance operator

$$\Gamma_z = \mathbb{E}\left(\langle X - \eta_z, .\rangle(X - \eta_z)|Z = z\right),$$

and Δ_z the conditional cross-covariance operator

$$\Delta_z = \mathbb{E}\left((X - \eta_z)(Y - \mu_z)|Z = z\right).$$

We note at this point that equations (1) and (2) define an ill-posed problem with the consequence that parameter α_z is not always identifiable. As a matter of fact, it can be seen from (2) that the functional parameter α_z may be identified only in $(Ker(\Gamma_z))^\perp$, so that we suppose from now on, without loss of generality, that $Ker(\Gamma_z)$ is reduced to zero. Since $\mathbb{E}(\|X\|^2|Z = z) < +\infty$, the operator Γ_z is compact and has no continuous inverse. As a consequence, the existence of a unique solution to (1)-(2) is not always insured unless the following necessary and sufficient condition holds (see Cardot *et al.*, 2003, for a justification in the unconditional case).

Condition 1. For every z, the random variables X and Y satisfies

$$\sum_{j=1}^{+\infty} \frac{\langle \Delta_z, v_j(z)\rangle^2}{\lambda_j(z)^2} < +\infty,$$

where $(\lambda_j(z), v_j(z))_j$ are the eigenelements of Γ_z, $\Gamma_z v_j(z) = \lambda_j(z)\ v_j(z)$, with $\lambda_1(z) \geq \lambda_2(z) \geq \cdots \geq 0$ and $v_1(z), v_2(z), \cdots$ form an orthonormal basis of $L^2(T)$. Under Condition 1, there is a unique solution to (1)-(2) in $(Ker(\Gamma_z))^\perp$ given by

$$\alpha_z = \sum_{j=1}^{+\infty} \frac{\langle \Delta_z, v_j(z)\rangle}{\lambda_j(z)} v_j(z). \tag{3}$$

Consider now a sample $(X_i, Z_i, Y_i), i = 1, \ldots, n$ of independent and identically distributed realizations of (X, Z, Y). Assuming implicitly that Condition 1 is fullfilled, we propose in section 2 estimation procedures for α based either on a regression on conditional functional principal components or on a penalized least squares B-splines expansion. Some consistency results are also given. Practical aspects of implementation of these estimators are discussed in section 3. In section 4, we show the benefits of using a varying-coefficient model in order to predict pollution events in the area of Toulouse. Comparison of prediction skills are made with other functional approaches that have already been proposed for such studies. Finally, section 5 is devoted to concluding remarks and possible extensions of this work. Proofs are gathered in an Appendix.

2 Estimation of the functional coefficient

2.1 Conditional Functional Principal Components Regression

Borrowing ideas from Cardot (2007) on conditional functional principal components analysis, we can build consistent estimators of the two conditional operators Γ_z and Δ_z with kernel smoothers. For that purpose let us introduce the following weights

$$w_i(z,h) \quad = \quad \frac{K\left((Z_i - z)/h\right)}{\sum_{i=1}^{n} K\left((Z_i - z)/h\right)} \ , \tag{4}$$

where the kernel K is a positive, symmetric around zero and bounded function with compact support and $h = h_n > 0$ is a generic bandwidth whose asymptotic behavior will be described later.

Let us define the estimator $\widehat{\eta}_z$ of η_z as follows

$$\widehat{\eta}_z \quad = \quad \sum_{i=1}^{n} w_i(z, h_x) X_i \ , \tag{5}$$

where h_x is a bandwidth and consider the following estimator $\widehat{\Gamma}_z$ of Γ_z

$$\widehat{\Gamma}_z \quad = \quad \sum_{i=1}^{n} w_i(z, h_1) \langle X_i - \widehat{\eta}_z, . \rangle \left(X_i - \widehat{\eta}_z\right), \tag{6}$$

where again h_1 is a specific bandwidth.

Consider now the cross covariance operator Δ_z, and its estimator defined as follows

$$\widehat{\Delta}_z = \sum_{i=1}^{n} w_i(z, h_2) \, (X_i - \widehat{\eta}_z) \, (Y_i - \widehat{\mu}_z) \, , \tag{7}$$

where $\widehat{\mu}_z$ is the classical kernel regression estimator of $\mu(z)$, defined by

$$\widehat{\mu}_z = \sum_{i=1}^{n} w_i(z, h_y) Y_i \; , \tag{8}$$

using a bandwidth h_y.

Note that for defining the above estimators we have always taken the same kernel. We could have considered different kernels at the expense of heavy additional notations and a poor gain (in terms of performance of the estimators) since it is well known that in nonparametric kernel estimation the most important tuning parameter is the bandwidth value.

Considering now a positive integer K_n, the functional principal components regression consists, in this context, in expanding an estimator of α_z in the K_n dimensional sub-space of H spanned by the K_n eigenfunctions, $\widehat{v}_1(z,.)$, $\ldots, \widehat{v}_{K_n}(z,.)$ of $\widehat{\Gamma}_z$ associated to the largest eigenvalues, $\widehat{\lambda}_1(z) \geq \ldots \geq \widehat{\lambda}_{K_n}(z) \geq 0$.

Introducing now the generalized inverse of $\widehat{\Gamma}_z$ in the space spanned by its first K_n eigenfunctions,

$$\widehat{\Gamma}_z^{\dagger} = \sum_{j=1}^{K_n} \frac{\widehat{v}_j(z) \otimes \widehat{v}_j(z)}{\widehat{\lambda}_j(z)}, \tag{9}$$

and inverting equation (2) lead to the following CFPCR estimator of α_z,

$$\begin{aligned} \widehat{\alpha}_{z,PCR} &= \widehat{\Gamma}_z^{\dagger} \widehat{\Delta}_z \\ &= \sum_{j=1}^{K_n} \frac{< \widehat{v}_j(z), \widehat{\Delta}_z > \widehat{v}_j(z)}{\widehat{\lambda}_j}, \end{aligned} \tag{10}$$

which is analogous to (3). Equation (10) clearly shows that this estimator is a direct generalization of the principal components regression estimator studied in Cardot *et al.* (1999), considering the conditional covariance operators instead of the empirical ones.

2.2 Penalized local least squares splines

For notation convenience we assume, from now on and whithout loss of generality, that $T = [0, 1]$. It is quite easy to show that α_z defined in (1) is solution to the minimization problem

$$\min_{\beta_z \in H} \mathbb{E}\left(\left(Y - \mu_z - \int_0^1 \beta_z(t)(X - \eta_z(t))dt\right)^2 |Z = z\right). \tag{11}$$

The penalized local least squares estimator is based on an expansion of the functional coefficient in B-splines basis which minimizes a penalized empirical counterpart of (11). For a fixed non negative integer q and a given non negative integer $k = k_n$ depending on the sample size n, we denote by $S_{k,q}$ the set of spline functions defined on $[0, 1]$ with order q and k equispaced interior knots (de Boor, 1978). A function s in $S_{k,q}$ is a polynomial of degree q on each subinterval defined by the knots and is $q - 1$ times continuously differentiable on $[0, 1]$. We consider a basis of $S_{k,q}$ composed of B-splines, $\{B_{k,j}, j = 1, \ldots, k + q\}$, and define $\mathbf{B}_k = (B_{k,1}, \ldots, B_{k,k+q})$.

Thus, we look for $\boldsymbol{\theta}_z \in \mathbb{R}^{k+q}$ minimizing the following penalized criterion

$$\min_{\boldsymbol{\theta} \in \mathbb{R}^{k+q}} \sum_{i=1}^{n} w_i(z, h) \left(Y_i - \hat{\mu}_z - \sum_{j=1}^{q+k} < \theta_j B_{k,j}, X_i - \hat{\eta}_z >\right)^2 + \ell \left\| \mathbf{B}_k^{(m)'} \boldsymbol{\theta} \right\|^2, \tag{12}$$

where ℓ is a smoothing parameter. The first term in (12) is a usual kernel nonparametric estimator of the conditional expectation (11) where β_z has been replaced by a spline function. It is well known however that the minimizer of this first term would not be a consistent estimator. Indeed, the eigenvalues of the matrix $\widehat{\mathbf{C}}_z$ defined below decreases rapidly to zero when $k = k_n$ grows up. Mimicing what is usually done for ill-posed inverse problems, the solution is regularized by the addition of a penalty term in (12). The penalty is proportional to the norm of a derivative of the functional parameter which means that ℓ controls the trade-off between data adjustments and regularity of the solution. The estimator defined by (12) is similar to the one introduced in Cardot *et al.* (2003) with the difference that in our case we are estimating the solution of the conditional expectation (11) while Cardot *et al.* (2003) consider the unconditioned counterpart of (11). This results in considering in our case a weighted sum of squares in (12).

Now, the solution $\widehat{\boldsymbol{\theta}}_z$ of (12) is given by

$$\widehat{\boldsymbol{\theta}}_z \;=\; \widehat{\mathbf{C}}_{z,\ell}^{-1}\widehat{\mathbf{b}}_z \;=\; \left(\widehat{\mathbf{C}}_z + \ell\mathbf{G}_k\right)^{-1}\widehat{\mathbf{b}}_z, \qquad (13)$$

where $\widehat{\mathbf{C}}_z$ is the $(q+k)\times(q+k)$ matrix with elements $\sum_{i=1}^n w_i(z,h) < B_{k,j}, X_i - \hat{\eta}_z >< B_{k,l}, X_i - \hat{\eta}_z > = < \widehat{\Gamma}_z B_{k,j}, B_{k,l} >$, $\widehat{\mathbf{b}}_z$ is the vector in $\mathbb{R}^{q+k}$ with elements $\sum_{i=1}^n w_i(z,h) < B_{k,j}, X_i > (Y_i - \widehat{\mu}_z) = \widehat{\Delta}_z B_{k,j}$ and where $\mathbf{G}_k$ is the $(q+k)\times(q+k)$ matrix with elements $< B_{k,j}^{(m)}, B_{k,l}^{(m)} >$. Note that for this approach $\widehat{\Gamma}_z$ and $\widehat{\Delta}_z$ are built using the same bandwidth $h_1 = h_2 = h$. Finally, the penalized splines estimator of α is

$$\widehat{\alpha}_{z,PS} \;=\; \mathbf{B}_k'\widehat{\boldsymbol{\theta}}_z. \qquad (14)$$

2.3 Some consistency properties

We study the performance of estimators $\widehat{\alpha}_{z,PCR}$ and $\widehat{\alpha}_{z,PS}$ in terms of prediction error. This leads to consider the criterion error $\|.\|_z$ defined as follows $\|\beta\|_z^2 = \langle \Gamma_z\beta, \beta\rangle$, and it is easily seen (see Cardot and Sarda, 2005, for a discussion) that this criterion is a conditional prediction error,

$$\|\beta\|_z^2 = \mathbb{E}\left(\langle X, \beta\rangle^2 | Z = z\right).$$

Thus, we are looking for the behavior of the prediction errors $\|\widehat{\alpha}_{z,PCR} - \alpha_z\|_z$ and $\|\widehat{\alpha}_{z,PS} - \alpha_z\|_z$ for large values of the sample size n: we show that these quantities converge to zero respectively almost surely and in probability.

Theorem 2.1 *Under assumptions given in the appendix, we have that, as n tends to infinity*

$$\|\alpha_z - \widehat{\alpha}_{z,PCR}\|_z \;\longrightarrow\; 0\ ,\ a.s. \qquad (15)$$

and

$$\|\alpha_z - \widehat{\alpha}_{z,PS}\|_z^2 \;\longrightarrow\; 0\ ,\ \textit{in probability}. \qquad (16)$$

The conditions (given in the Appendix) in theorem above involve Lipschitz regularity assumptions on the various conditional moments. They also imply, for the principal components regression estimator, that the dimension K_n must tend slowly enough to infinity with rates depending on the behavior of the eigenvalues as n increases.

As far as the penalized estimator is concerned, one must assume that the linear functional is regular and the smoothing parameter value ℓ must not tend too rapidly to zero. On the other hand, one can notice that there are no strong conditions on the eigenvalues.

3 Some comments on practical implementation

3.1 Discretized data

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

$$\mathbf{X}_i = (X_i(t_{i1}), X_i(t_{i2}), \ldots, X_i(t_{ip_i})), \quad i = 1, \ldots, n, \tag{17}$$

where the discretization points are not necessarily equispaced but are just supposed to satisfy $0 \leq t_{i1} < t_{i2} < \cdots < t_{ip_i} \leq 1$.

There are several methods proposed in the literature to deal with the discretization issue which depend on the sparsity of the design points. If the grid of points is sufficiently dense, one can perform a basis functions expansion of the trajectories (B-splines, wavelets, *etc*) and then operate on the coordinates instead of the observed data (see Ramsay and Silverman, 2005). On the other hand for sparse designs direct estimation procedures of the covariance function can be performed using kernel smoothers (Staniswalis and Lee, 1998) or local polynomials (Yao *et al.* 2005). Basis expansion are also considered by James *et al.* (2000) with estimation procedures relying on the EM algorithm. Note that for sparse designs computational procedures are definitely heavier.

3.2 Choosing values for the tuning parameters

The first step of the estimation procedure consists in the estimation of the conditional expectations of X and Y given z by kernel smoothers with smoothing parameters h_x and h_y as defined in section 2. The estimator of μ_z is clearly an usual nonparametric regression kernel estimator and thus the choice of h_y can be performed by data-driven selectors of the bandwidth such as cross-validation (see Härdle, 1990). The specificity of the estimation of η_z is that the output of the regression is a curve: the cross-validation criterion can be adapted to this situation in order to select the bandwidth h_x in the kernel estimator $\hat{\eta}_z$ as follows (see Hart and Wehrly, 1993)

$$CV_x(h_x) \;=\; \sum_{i=1}^{n} \int_T \left(X_i(t) - \widehat{\eta_{Z_i}^{-i}}(t) \right)^2 dt,$$

where $\widehat{\eta_{Z_i}^{-i}}(t)$ is the leave-one-out estimator (based on the sub-sample in which the observation (X_i, Z_i) has been eliminated) evaluated at point Z_i.

Then, the CFPCR estimator is tuned by three parameters, the bandwidths h_1, h_2 and the dimension K_n. They all act as smoothing parameters even if they have really different roles. One needs to choose a reasonable values for the bandwidths h_1 and h_2 which controls the "dependency" between Y, X and Z and the dimension parameter K_n which controls the "smoothness" of the estimator of α. Again a cross-validation criterion is a natural candidate to provide a data-driven selection procedure for these tuning parameters, minimizing

$$\mathrm{CV}_{CFPCR}(h_1, h_2, K_n) = \frac{1}{n}\sum_{i=1}^{n}\Big(Y_i - \widehat{\mu}_{Z_i} - < \widehat{\alpha}_{Z_i}^{-i}, X_i - \widehat{\eta}_{Z_i} >\Big)^2. \qquad (18)$$

where $\widehat{\alpha}_{z_i}^{-i}$ is the estimator of α_z obtained by the CFPCR procedure once the observation (X_i, Z_i, Y_i) has been left out the initial sample.

Now the Penalized Splines estimator is essentially controled by two tuning parameters, ℓ for the smoothness of the estimator and the bandwidth h. Indeed, the degree of the Splines functions m takes generally a value between 2 and 4. It is also known that the number of knots k plays a less important role than the parameter ℓ which controls the smoothness of the estimator provided that one considers reasonably large values of k (Eilers and Marx, 1996, Besse *et al.*, 1997). Thus, values of ℓ and the bandwidth h can be chosen by minimizing the following cross-validation criterion

$$\mathrm{CV}_{PS}(h, \ell) = \frac{1}{n}\sum_{i=1}^{n}\Big(Y_i - \widehat{\mu}_{Z_i} - < \widehat{\alpha}_{Z_i,PS}^{-i}, X_i - \widehat{\eta}_{Z_i} >\Big)^2. \qquad (19)$$

The selection procedures described above are operational but appear to be quite expensive from a computational point of view. Looking for efficient approximations to these criterions certainly deserves further investigation.

4 Ozone pollution forecasting

The prediction of pollution events in the atmosphere is of great interest in the scientific community. Ozone creation is the result of complex chemical processes and due to the complexity of the phenomenon non linear and nonparametric statistical approaches have been widely used to give predictions of alerts. These statistical models are generally based on threshold time series (Mélard and Roy, 1988), generalized additive models (Davis and

Speckman, 1999), regression trees (Ghattas, 1999) or neural networks (Yi and Prybutok 1996, Gardner and Dorling 1999).

We have data that are collected in Toulouse by the ORAMIP (Observatoire de l'air de la Région Midi-Pyrénées) air quality authority for the Midi-Pyrénées area in South West of France. A detailed description of the data can be found in Aneiros *et al.* (2004). They consist in hourly measurements of five variables, ozone concentration (O_3), NO concentration (NO) and NO_2 concentration (NO_2), wind speed (WS), and wind direction (WD), during the summers (15th May - 15th September) of the years 1997-2000. Eliminating missing data due to transmission problems or measurement apparatus failures we finally get a sample of n=474 days with hourly measurements at times points $t_1 < \ldots < t_{24}$. We aim at forecasting daily maximum O_3 concentrations with the help of the functional covariates measured the day before until 5 pm. For day i, the response is the maximum of the observed concentration in ozone during day i, $\mathrm{O}_{3,i}(t)$, and is defined by

$$Y_i \quad = \max_{j=1,2,\ldots,24} \mathrm{O}_{3,i}(t_j) \tag{20}$$

whereas the covariates are the 24 hourly measurements of the five variables above mentioned for time t varying from t=6pm of day $i-2$ to t=5pm of day $i-1$.

Our data are clearly functional even if they are discretized versions of the continuous underlying phenomenon. Let us fist note that not much work has taken this important feature into account in such a context. Damon and Guillas (2002) proposes linear functional autoregressive models including covariates and Cardot *et al.* (2007) extended the functional linear model with several functional covariates. From a completely nonparametric point of view, Aneiros *et al.* (2004) have built nonparametric functional predictors that can take many functional covariates into account extending to multivariate functional data the procedures developed in Ferraty and Vieu (2006).

The varying functional linear model which allows the functional coefficient of regression to vary nonparametrically according to another covariate of interest seems to be a good comprise between completely nonparametric approaches and linear functional models. These statistical models can be compared for the prediction of the maximum of O_3 concentration one day ahead in the city of Toulouse.

We split the original data in a learning and a test sample as in Aneiros *et al.* (2004) and Cardot *et al.* (2007) in order to be able to provide a comparison of the different approaches for predicting maximum ozone concentration one day ahead. The original sample is split randomly in a test sample, say I_T, composed of $n_T = 142$ observations and a learning sample, say I_L, with $n_L = 332$ observations.

Ozone prediction with varying-coefficient functional linear regression models

As noticed in Cardot *et al* (2007) the most informative functional covariate for predicting maximum O_3 one day ahead (day i+1) at 5pm is the ozone curve measured for time t varying from t=6pm of day $i-1$ to t=5pm of day i. This functional covariate is denoted by X_i from now on and will be the functional covariate in our varying-coefficient functional linear models. Note that we only have discrete trajectories of ozone evolution, with equispaced design points. For computational facilities, we have considered the penalized spline estimator and expand the functional coefficients in B-splines basis of order $q = 3$ with $k = 6$ equispaced interior knots. Then, estimators are computed by minimizing criterion (12) where all integrals are approximated with quadrature rules. We consider as potential real additional covariates Z the maximum values between t=6pm of day $i-1$ to t=5pm of day i as well as the value at time t=5pm of the available functional covariates. For each candidate Z the smoothing parameters are selected by minimizing the cross validation criterion on the learning sample defined in (19) and then the best covariate Z is chosen by considering the model whose cross-validation prediction error in the learning sample is minimum.

We finally get, according to this model selection procedure, a varying coefficient model with $Z_i = \max_t \mathrm{NO}(t)$ where t varies from t=6pm of day $i-1$ to t=5pm of day i. The smoothing parameters values are $\ell_{CV} = 0.035$ and $h_{CV} = 0.2$.

Comparison with other functional approaches

We first consider for comparison two naive models which consists in predicting an observation of the test sample by the empirical mean in the learning sample (model M_0) and predicting a new value at day i by the observed value of maximum ozone concentration at day $i-1$ (model named *persistency*). As in Aneiros *et al.* (2004), these naive predictors serve as benchmarks.

As far as the functional linear model is concerned, Cardot *et al* (2007) consider both the univariate case, *i.e.* with only one functional covariate, and the multivariate case where an additive functional linear model are proposed. They note that the best prediction in the univariate case (denoted by FLM) was obtained when considering the ozone curves of the preceding days as covariate and that prediction skill could be improved by considering a multivariate functional linear model (say MFLM) with O_3, NO, N2 and WS curves of the preceding day as covariates.

Dealing now with nonparametric functional predictors, the best model for prediction found by Aneiros *et al.* (2004) is an additive model whose covariates are the observed curves of O_3, NO, WS and WD of the preceding day. This model is denoted by FAM.

The criterions used on the test sample to compare the skill of the different models are based on the quadratic errors

$$E_1(i) = \frac{\left(Y_i - \widehat{Y_i}\right)^2}{\mathrm{Var}_{I_T}(Y)},$$

where $\widehat{Y_i}$ is the prediction given by the model and $\mathrm{Var}_{I_T}(Y)$ is the empirical variance of Y_i in the test sample, as well as the absolute errors

$$E_2(i) = \left|Y_i - \widehat{Y_i}\right|.$$

Prediction errors for the different models discussed before are given in Table (1).

The first interesting thing to be noticed is that the two naive models (namely M_0 and M_0') can be considerably improved by incorporating additional information. Then, one can note that among the functional approaches, the varying coefficient model gives the best prediction errors according to the quadratic error criterion even if it does not take into account all the potential covariates as this is the case for the MFLM and FAM. This means that even if it is apparently less flexible it allows to take into account interactions processes between O_3 creation and NO level and in some ways threshold effects by allowing the functional regression coefficient to vary in a non linear way with the NO concentration. These first results are encouraging and should lead to further developments, aiming at understanding what

Model	Quadratic Error	Absolute Error
M_0	0.995	18.71
Persistency	0.917	17.44
FLM	0.416	12.62
MFLM	0.391	11.87
FAM	0.378	**11.40**
VFLM	**0.376**	**11.40**

Table 1: Mean value of the criterion errors on the test sample for the different predictors.

can be the chemical process in action when ozone concentration levels are high.

5 Concluding remarks

The model presented in this paper allows to extend functional linear regression by taking into account the effect of an additional variable. The estimators presented can be implemented in statistical softwares quite easily, have good asymptotic behaviors and can be useful for empirical studies as seen for the ozone prediction. One can imagine many extensions to this study which are beyond the scope of this paper.

At first let us note that one could consider a different estimation approach by expanding directly function $\alpha_z(t)$ in a two-dimensional splines basis and thus minimizing a penalized least squares criterion. This leads to a tensor product splines estimator: we refer for instance to Eilers and Marx (2003) or Ramsay and Silverman (2005) for introducing tensor product splines estimators in a slightly different context than the one considered here.

A natural extension to consider are Generalized Linear Models to cover for instance the situation where the response Y is binary. A conditional version of the procedure proposed by Cardot and Sarda (2005) and based on splines approximation can be considered. It is also possible to derive conditional principal components regression estimators in this context by adapting ideas from Müller and Stadtmüller (2005).

Model (1) can also be extended to a multivariate framework. Firstly one can consider a vectorial variable Z instead of a real one. Estimators proposed here can be extended readily to such a multivariate framework by considering multivariate kernel smoothers. These approaches will clearly suffer from the curse of dimensionality and developing dimension reduction approaches would be of real interest in such a context. It would also be interesting to study the situation in which one has several curves as predictors (as it is the case in several applications) and modify our model in an additive model with functional covariates. Then, specific estimation procedures would need to be developed: one can think for instance at the functional version of the algorithm based on PLS introduced by Hastie and Tibshirani (1993).

There are some situations in which the (functional) variables are measured with errors due for instance to the measurement apparatus. In such situations, estimation procedures that do not take into account measurement errors may fail to give accurate and even consistent estimators . One can adopt several strategies to make the estimation more efficient. The most natural, but maybe also the most time consuming, is to simply denoise the predictive curves X_i by using a common smoother (kernel, splines, ...). This leads to consider new curves $\tilde{X}_i$ that are smooth and then apply the estimation procedures as defined above with X_i replaced by $\tilde{X}_i$. Another possibility is to correct directly the estimators defined in (14) in order to take into account measurement errors. Such correction procedures are proposed in Crambes *et al.* (2007) and they consist in adding to the matrix $\widehat{\mathbf{C}}_{z,\ell}$ a denoising term proportional to the variance of the error-in-variable term. While this procedure can certainly be applied for ozone forecasting with minor modifications (in that case this would lead to correct in some way the variables Z_i), further theoretical studies on the behavior of the corrected estimators have to be done.

Finally a quite important problem in practice is to test the influence of a covariate Z on the regression function α_z, that is to say to consider the null hypothesis

$$H_0 : \alpha(z,t) = \alpha(t), \quad \forall z \ .$$

We believe it is possible to develop a test procedure based on permutations similar to those proposed in Cardot *et al.* (2004) . This issue is the subject of works in progress.

Appendix : proofs

Assumptions and preliminary results

We first present some preliminary results on the conditional mean of X and Y given $Z = z$, where z is a fixed point of $\mathbb{R}$ belonging to the support of the distribution of Z. We also give some asymptotic results for the estimators of the conditional covariance and cross-covariance operators. The variables X, Y and Z are assumed to satisfy the following conditions.

$$(H.1) \qquad \|X\| \ \le \ C_1 \ < \ +\infty, \ a.s.$$

$$(H.2) \qquad |Y| \ \le \ C_2 \ < \ +\infty, \ a.s.$$

$(H.3)$ The distribution of the variable Z is absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}$ and its density f_Z is such that $f_Z(z) > 0$ and is Lipschitz continuous with coefficient τ_z in a neighborhood of z.

As usual in nonparametric estimation the underlying functions are supposed to satisfy some regularity condition. Let us define the second order moment functions:

$r_1(z,s,t) = \mathbb{E}(X(s)X(t)|Z = z)$ and $r_2(z,s) = \mathbb{E}(X(s)Y|Z = z)$.

$(H.4)$ The functions μ, η, $r_1(z,.,.)$ and $r_2(z,.)$ are Lipschitz continuous in a neighborhood of z with respective coefficients τ_y, τ_x, τ_1 and τ_2.

Let h be one of the bandwidth h_x, h_y, h_1 or h_2. We assume that

$$h \longrightarrow 0 \text{ and } \log n/(nh) \longrightarrow 0 \text{ as } n \text{ tends to infinity.}$$

The estimator of the conditional mean of Y given $Z = z$ is the classical kernel one and we have (see *e.g.* Sarda and Vieu, 2000)

$$|\mu_z - \widehat{\mu}_z| \quad = \quad O(h_y^{\tau_y}) + O\left(\frac{\log n}{nh_y}\right)^{1/2}, \quad a.s. \tag{21}$$

From Cardot (2007) we can deduce the following results

$$\|\eta_z - \widehat{\eta}_z\| = O(h_x^{\tau_x}) + O\left(\frac{\log n}{n h_x}\right)^{1/2}, \quad a.s, \tag{22}$$

and under Lipschitz conditions (H.4), we have

$$\left\|\Gamma_z - \widehat{\Gamma}_z\right\| = O(h_1^{\tau_1}) + O(h_x^{\tau_x}) + O\left(\frac{\log n}{n \min(h_1, h_x)}\right)^{1/2}, \quad a.s., \tag{23}$$

where $\|.\|$ stands for the Hilbert-Schmidt norm for operators. Using the same technics as in Cardot (2007), we can also show that

$$\left\|\Delta_z - \widehat{\Delta}_z\right\| = O(h_2^{\tau_2}) + O(h_x^{\tau_x}) + O(h_y^{\tau_y}) + O\left(\frac{\log n}{n \min(h_2, h_x, h_y)}\right)^{1/2}, \quad a.s.$$

Asymptotic properties of the conditional principal components regression estimator

Let us first have a look at the existence of the principal components regression estimator $\widehat{\alpha}_{z,PCR}$ which is uniquely determined provided that the first K_n eigenvalues are distinct and $\widehat{\lambda}_{K_n}(z)$ is strictly positive, the latter condition ensuring the existence of the generalized inverse $\widehat{\Gamma}_z^{\dagger}$. By well known inequalities for positive compact operators, we have

$$|\lambda_{K_n}(z) - \widehat{\lambda}_{K_n}(z)| \leq \|\Gamma_z - \widehat{\Gamma}_z\|, \tag{24}$$

and thus, by (23),

$$\widehat{\lambda}_{K_n}(z) = \lambda_{K_n}(z) + o(\lambda_{K_n}(z)), \quad a.s., \tag{25}$$

if

$$\frac{1}{\lambda_{K_n}(z)}\left(h_1^{\tau_1} + h_x^{\tau_x} + \left(\frac{\log n}{n \min(h_1, h_x)}\right)^{1/2}\right) \to 0,$$

as n tends to infinity, noting that $\lambda_{K_n}(z) > 0$. Then, equation (25) means that $\widehat{\lambda}_{K_n}(z)$ is strictly positive on an event whose probability tends to one as n tends to infinity and this ensures existence of $\widehat{\Gamma}_z^{\dagger}$ and $\widehat{\alpha}_{z,PCR}$. Assuming moreover that the eigenvalues of Γ_z are distinct we get uniqueness of the solution.

Let us define $\Pi_{K_n}(z) = \sum_{j=1}^{K_n} v_j(z) \otimes v_j(z)$: it is the projection onto the eigenspace generated by the first K_n eigenfunctions of Γ_z. With (H.1), we have

$$\|\alpha_z - \widehat{\alpha}_{z,PCR}\|_z \leq C_1 \left(\|\alpha_z - \Pi_{K_n}\alpha_z\| + \|\Pi_{K_n}\alpha_z - \widehat{\alpha}_{z,PCR}\|\right), \tag{26}$$

where C_1 is a strictly positive constant. We first get using straighforward arguments

$$\|\alpha_z - \Pi_{K_n}\alpha_z\|^2 \leq \sum_{j>K_n} < \alpha_z, v_j(z) >^2, \tag{27}$$

which proves that the approximation error due to the projection $\Pi_{K_n}\alpha_z$ tends to zero when K_n tends to infinity.

The second part of the demonstration is based on Lemma 5.1 in Cardot *et al.* (1999) which states that

$$\|\Pi_{K_n}\alpha_z - \widehat{\alpha}_{z,PCR}\| \leq \gamma_n \left\|\Gamma_z - \widehat{\Gamma}_z\right\| + \frac{1}{\widehat{\lambda}_{K_n}(z)} \left\|\Delta_z - \widehat{\Delta}_z\right\|, \tag{28}$$

where γ_n is defined by

$$\gamma_n = \|\Delta_z\| \left(\frac{1}{\widehat{\lambda}_{K_n}(z)\lambda_{K_n}(z)} + 2\left(\frac{1}{\widehat{\lambda}_{K_n}(z)} + \frac{1}{\lambda_{K_n}(z)} \sum_{j=1}^{K_n} a_j \right)\right) \tag{29}$$

with

$$\begin{cases} a_1 = 2\sqrt{2}(\lambda_1(z) - \lambda_2(z))^{-1}, \\ a_j = 2\sqrt{2}\left[\min(\lambda_{j-1}(z) - \lambda_j(z), \lambda_j(z) - \lambda_{j+1}(z))\right]^{-1}, \ j > 1. \end{cases}$$

Let us study in details the first term in γ_n, the other upper bounds being obtained with similar manipulations.

Using inequality (24), we get with (23),

$$\frac{\lambda_{K_n}(z)}{\widehat{\lambda}_{K_n}(z)} - 1 \quad = \quad O(h_1^{\tau_1}) + O(h_x^{\tau_x}) + O\left(\frac{\log n}{n \ \min(h_1, h_x)}\right)^{1/2}, \quad a.s. \tag{30}$$

and

$$\frac{1}{\widehat{\lambda}_{K_n}(z)\lambda_{K_n}(z)} \quad = \quad \frac{1}{\lambda_{K_n}^2(z)} + o\left(\frac{1}{\lambda_{K_n}^2(z)}\right), \quad a.s. \tag{31}$$

Then,

$$\frac{\left\|\Gamma_z - \widehat{\Gamma}_z\right\|}{\widehat{\lambda}_{K_n}(z)\lambda_{K_n}(z)}$$

$$= \frac{1}{\lambda_{K_n}^2(z)} \left(O(h_1^{\tau_1}) + O(h_x^{\tau_x}) + O\left(\frac{\log n}{n \ \min(h_1, h_x)}\right)^{1/2}\right), \ a.s. \ , \tag{32}$$

which ensures the almost sure converge towards zero of this term provided that $\frac{1}{\lambda_{K_n}^2(z)}\left(h_1^{\tau_1} + h_x^{\tau_x} + \left(\frac{\log n}{n \ \min(h_1, h_x)}\right)^{1/2}\right) \to 0$, as n tends to infinity. Repeating the same procedure for the other terms in (28) and (29), we get as n tends to infinity $\|\alpha_z - \widehat{\alpha}_{z,PCR}\|_z \to 0$, *a.s.* when the following set of assumptions is fulfilled

- $\dfrac{h_1^{\tau_1}}{\lambda^2_{K_n}(z)} \to 0, \quad \dfrac{h_x^{\tau_x}}{\lambda^2_{K_n}(z)} \to 0, \quad \dfrac{\log n}{\lambda^2_{K_n}(z)\; n \min(h_1, h_x)} \to 0,$
- $\dfrac{\sum_j^{K_n} a_j h_1^{\tau_1}}{\lambda_{K_n}(z)} \to 0, \quad \dfrac{\sum_j^{K_n} a_j h_x^{\tau_x}}{\lambda_{K_n}(z)} \to 0, \quad \dfrac{\log n \sum_j^{K_n} a_j}{\lambda_{K_n}(z)\; n \min(h_1, h_x)} \to 0,$
- $\dfrac{h_2^{\tau_2}}{\lambda_{K_n}(z)} \to 0, \quad \dfrac{h_y^{\tau_y}}{\lambda_{K_n}(z)} \to 0, \quad \dfrac{\log n}{\lambda_{K_n}(z)\; n \min(h_2, h_y)} \to 0,$

These conditions mean that K_n must tend slowly enough to infinity to get consistent estimators, depending on the regularity of the various conditional moments and the shape of the eigenvalues. On the other hand the different bandwidths have to be large enough to ensure that $\dfrac{h^\tau}{\lambda^2_{K_n}(z)} \to 0.$

Asymptotic properties of the weighted penalized splines estimator

The behavior of the error of prediction for the estimator α depends on regularity assumption on α_z: for some integer p', the function α_z is assumed to have p' derivatives with $\alpha_z^{(p')}$ satisfying

$$(H.5) \qquad |\alpha_z^{(p')}(y_1) - \alpha_z^{(p')}(y_2)| \;\le\; C_3 |y_1 - y_2|^\nu, \; C_3 > 0, \; \nu \in [0,1].$$

We note $p = p' + \nu$ and assume that $q \ge p$. From Cardot (2002), we can deduce that

$$C_6 k^{-1} \|\mathbf{u}\|^2 \;\le\; \mathbf{u}'\mathbf{G}_k\mathbf{u}, \; \mathbf{u} \in \mathcal{K}(\mathbf{G}_k)^\perp,$$

and

$$\mathbf{u}'\mathbf{G}_k\mathbf{u} \;\le\; C_7 k^{2m-1} \|\mathbf{u}\|^2, \; \mathbf{u} \in \mathbb{R}^{q+k},$$

where $\mathcal{K}(\mathbf{G}_k)$ is the null space of $\mathbf{G}_k$ and C_6 and C_7 are two positive constants.

Define the matrix $\overline{\mathbf{C}}_z$ as the $(q+k) \times (q+k)$ matrix with elements $< \Gamma_z B_{k,j}, B_{k,l} >$ and $\overline{\mathbf{C}}_{z,\ell} = \overline{\mathbf{C}}_z + \ell \mathbf{G}_k$. Then using the same developments as in Lemma 6.2 in Cardot, Ferraty and Sarda (2003) we can show that the eigenvalues of $\overline{\mathbf{C}}_{z,\ell}$ lie between $C_7 \ell k^{-1}$ and $C_8 k^{-1}$ and that

$$\|\widehat{\mathbf{C}}_{z,\ell} - \overline{\mathbf{C}}_{z,\ell}\| \;=\; O_P(k^{-1}(h^{\tau_1} + h_x^{\tau_x} + (\log n / (n \, \min(h, h_x)))^{1/2})).$$

From this last result and taking $\ell = o(h^{\tau_1} + h_x^{\tau_x} + (\log n/(n \, \min(h, h_x)))^{1/2})$, we get that $\widehat{\mathbf{C}}_{z,\ell}$ is non singular except on an event whose probability tends to

zero as n tends to infinity and this allows us to deduce that a unique solution $\widehat{\alpha}_{z,PS}$ to the problem (12) exists except on an event whose probability goes to zero as n tends to infty.
Now, let us consider the following quantity

$$\begin{aligned} & \Lambda_{z,\ell}(a) \\ = \; & E\left(< a, X-\eta_z > (Y-\mu_z) - \frac{< a, X-\eta_z >^2}{2} | Z=z\right) \; - \; \frac{1}{2}\ell\|a^{(m)}\|^2, \end{aligned}$$

and its empirical version

$$\begin{aligned} & \widehat{\Lambda}_{z,\ell}(a) \\ = \; & \sum_{i=1}^{n} w_i(z,h)\left(< a, X_i-\hat{\eta}_z >)(Y_i-\hat{\mu}_z) - \frac{< a, X_i >^2}{2})\right) \; - \; \frac{1}{2}\ell\|a^{(m)}\|^2. \end{aligned}$$

Note that the solution $\widehat{\alpha}_{z,PS}$ of (12) also satisfies

$$\widehat{\alpha}_{z,PS} \;=\; \arg\max_{\beta\in S_{kq}} \widehat{\Lambda}_{z,l}(\beta).$$

With the same arguments as in Lemma 5.1 in Cardot and Sarda (2005), one can show that under the assumptions outlined above and if moreover $\ell^{-1}k^{-2p}+\ell k^{2(m-p)} = O(1)$ there is a unique $\widetilde{\alpha}_{z,PS}\in S_{kq}$ such that

$$\widetilde{\alpha}_{z,PS} \;=\; \arg\max_{\beta\in S_{kq}} \Lambda_{z,\ell}(\beta),$$

and that

$$\|\alpha_z-\widetilde{\alpha}_{z,PS}\|_z^2 \;=\; O(k^{-2p})+O(\ell k^{2(m-p)})+O(\ell). \tag{33}$$

Let us introduce $\beta\in S_{kq}$ and write $\beta(t) = \sum_{j=1}^{q+k}\theta_j B_{kj}(t) = \boldsymbol{\theta}'\mathbf{B}_k(t)$ and $\widetilde{\alpha}_{z,PS}(t) = \sum_{j=1}^{q+k}\widetilde{\theta}_j B_{kj}(t) = \widetilde{\boldsymbol{\theta}}'\mathbf{B}_k(t)$. The score $s_n(\boldsymbol{\theta})$ is given by

$$\begin{aligned} s_n(\boldsymbol{\theta}) \;&=\; \frac{\partial\widehat{\Lambda}_{z,\ell}(\beta)}{\partial\boldsymbol{\theta}} \\ &=\; \sum_{i=1}^{n} w_i(z,h) < \mathbf{B}_k, X_i-\hat{\eta}_z > (Y_i-\hat{\mu}_z - < \beta, X_i-\hat{\eta}_z >) - \ell\mathbf{G}_k\boldsymbol{\theta}, \end{aligned}$$

where $< \mathbf{B}_k, X_i-\hat{\eta}_z >$ is the vector with generic element $< B_{kj}, X_i-\hat{\eta}_z >$, $j=1,\ldots,k+q$. The second derivative satisfies

$$\frac{\partial^2\widehat{\Lambda}_{z,\ell}(\beta)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'} \;=\; -\,\widehat{\mathbf{C}}_{z,l}.$$

Let us write now the solution of (12) as $\widehat{\alpha}_{z,PS} = \widehat{\boldsymbol{\theta}}' \boldsymbol{B_k}$. By definition of $\widehat{\alpha}_{z,PS}$, $s_n(\widehat{\boldsymbol{\theta}}) = 0$ and then a Taylor expansion of the score gives us

$$s_n(\widetilde{\boldsymbol{\theta}}) = \widehat{\mathbf{C}}_{z,l}\left(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}\right). \tag{34}$$

Since $\widehat{\mathbf{C}}_{z,l}$ is a strictly positive matrix except on an event whose probability tends to zero with n, one has equivalently

$$\widehat{\mathbf{C}}_{z,l}^{-1/2} s_n(\widetilde{\boldsymbol{\theta}}) = \widehat{\mathbf{C}}_{z,l}^{1/2}\left(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}\right). \tag{35}$$

Using the same arguments as for showing (38) in Cardot and Sarda (2005), we obtain, except on an event whose probability tends to zero with n

$$\left\|\widehat{\mathbf{C}}_{z,l}(\boldsymbol{\theta}^*)^{1/2}\left(\widehat{\boldsymbol{\theta}} - \widetilde{\boldsymbol{\theta}}\right)\right\|^2$$

$$\geq C_9\left(\|\widetilde{\alpha}_{z,PS} - \widehat{\alpha}_{z,PS}\|_z^2 + \ell\left\|(\widetilde{\alpha}_{z,PS} - \widehat{\alpha}_{z,PS})^{(m)}\right\|^2\right) + o_P(\iota), \tag{36}$$

where $\iota = \iota_n$ is a sequence of positive reals such that ι/ℓ is bounded. Defining

$$s(\boldsymbol{\theta}) = \frac{\partial \Lambda_{z,\ell}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = I\!\!E\left(\langle \boldsymbol{B}_k, X - \eta_z\rangle\left(Y - \mu_z - \langle \beta, X - \eta_z\rangle\right)|Z = z\right) + \ell \boldsymbol{G}_k \boldsymbol{\theta},$$

and noticing that $s(\widetilde{\boldsymbol{\theta}}) = 0$,. we have using the facts that $\|\boldsymbol{B}_k\|^2 = 0(1)$ and $\|\widetilde{\alpha}_{z,PS}\|^2$ is bounded by a positive constant

$$\begin{aligned}
\|s_n(\widetilde{\boldsymbol{\theta}})\|^2 &= \|s_n(\widetilde{\boldsymbol{\theta}}) - s(\widetilde{\boldsymbol{\theta}})\|^2 \\
&= \left\|\langle \boldsymbol{B}_k, \widehat{\Delta}_z - \Delta_z\rangle + \langle \boldsymbol{B}_k, (\widehat{\Gamma}_z - \Gamma)\widetilde{\alpha}_{z,PS}\rangle\right\|^2 \\
&= O_P\left(h^{2\tau_1}\right) + O_P\left(h^{2\tau_2}\right) + O_P\left(h_x^{2\tau_x}\right) + O_P\left(h_y^{2\tau_y}\right) \\
&\quad + O_P\left(\frac{\log n}{n\ \min(h, h_x, h_y)}\right).
\end{aligned}$$

This gives us with (36) and since $\|\widehat{C}_{z,l}^{-1}\| = O_p(k/\ell)$

$$\|\widetilde{\alpha}_{z,PS} - \widehat{\alpha}_{z,PS}\|_z^2 = O_P\left(\frac{k}{\ell}\left(h^{2\tau_1} + h^{2\tau_2} + h_x^{2\tau_x} + h_y^{2\tau_y} + \frac{\log n}{n\, min(h, h_x, h_y)}\right)\right).$$

Finally, we obtain with (33)

$$\begin{aligned}\|\alpha_z - \widehat{\alpha}_{z,PS}\|_z^2 &= O(k^{-2p}) + O(\ell k^{2(m-p)}) + O(\ell) \\ &\quad + O_P\left(\frac{k}{\ell}\left(h^{2\tau_1} + h^{2\tau_2} + h_x^{2\tau_x} + h_y^{2\tau_y} + \frac{\log n}{n \min(h, h_x, h_y)}\right)\right) \\ &= o_P(1),\end{aligned}$$

for well chosen values of k and ℓ.

Acknowledgements. We would like to thank the members of the STAPH working group in Toulouse for helpful discussions.

Bibliography

Aneiros-Perez, G., Cardot, H., Estevez-Perez, G. and P. Vieu. (2004). Maximum ozone concentration forecasting by functional nonparametric approaches. *Environmetrics*, **15**, 675-685.

Besse, P.C., Cardot, H. and Ferraty, F. (1997). Simultaneous nonparametric regressions of unbalanced longitudinal data. *Computational Statistics and Data Analysis*, **24**, 255-270.

Cai, T.T. and Hall, P. (2006). Prediction in Functional Linear Regression *Ann. Statist.*, **34**, 2159-2179.

Cardot, H. (2007). Conditional functional principal components analysis. *Scand. J. Statist.*, **34**, 317-335

Cardot, H., Crambes, C. and P. Sarda. (2007). Ozone Pollution Forecasting using Conditional Mean and Conditional Quantiles with Functional Covariates. In *Statistical Methods for Biostatistics and Related Fields,* Härdle, W., Mori, Y. and P. Vieu (Eds.), Springer Verlag, New York, 221-244.

Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica.* **13**, 571-591.

Cardot, H., Goia, A. and Sarda, P. (2004). Testing for no effect in functional regression models, some computational approaches. *Communications in Statistics - Simulations and Computations*, **33**, 179-199.

Cardot, H. and Sarda, P. (2005). Estimation in generalized linear models for functional data via penalized likelihood. *J. Multivariate Analysis*, **92**, 24-41.

Crambes, C., Kneip, A. and Sarda, P. (2007). Smoothing Splines Estimators for Functional Linear Regression. *Ann. Statist.*, To appear.

Cuevas, A., Febrero, M. and Fraiman, R. (2002). Linear functional regression: The case of fixed design and functional response. *Canadian Journal of Statistics*, **30**, 285-300.

De Boor, C., (1978). *A practical guide to Splines.* Springer-Verlag, New-York.

Damon, J. and Guillas, S. (2002). The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics*, **13**, 759-774.

Davis, J.M. and Speckman, P. (1999). A model for predicting maximum and 8 hour average ozone in Houston. *Atmospheric Environment*, **33**, 2487-2500.

Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties (with discussion). *Statistical Science*, **11**, 89-121.

Eilers, P. and Marx, B. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemometrics and Intelligent Lab. Systems*, **66**, 159-174.

Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis, theory and practice.* Springer Series in Statistics, Springer-Verlag.

Gardner, M.W. and Dorling, S.R. (1999). Neural network modelling and prediction of hourly NO and NO_2 concentrations in urban air in London. *Atmospheric Environment*, **33**, 709-719.

Ghattas, B. (1999). Prévisions des pics d'ozone par arbres de régression simples et agrégés par bootstrap. *Revue de Statistique Appliquée*, **XLVII**, 2, 61-80.

Hall, P. and Horowitz, J.L. (2007). Methodology and Convergence Rates for Functional Linear Regression. *Ann. Statist.*, To appear.

Hart, J. and Werhly, T. (1993). Consistency of cross-validation when data are curves. *Stochastic Processes and Their Applications.* **45**, 351-361.

Hastie, T. and Tibshirani, R. (1993). Varying coefficient models (with discussion). *J. Royal Satist. Soc. Ser.B.*, **55**, 757-796.

Hoover, D.R., Rice, J.A., Wu, C.O. and Yang, L.P. (1998). Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, **85**, 809-822.

James, G.M., Hastie, T.J. and Sugar, C.A. (2000). Principal component models for sparse functional data. *Biometrika*, **87**, 587-602.

Mélard, G. and Roy, R. (1988). Modèles de séries chronologiques avec seuils. *Revue de Statistique Appliquée*, **XXXVI**, 4, 5-24.

Müller, H.G. and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.* , **33**, 774-805.

Ramsay, J.O. and Dalzell, C.J. (1991). Some tools for functional data analysis (with discussion). *J. Roy. Statist. Soc. Ser.* **B 53**, 539-572.

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

Ramsay, J. O. and Silverman, B.W. (2002). *Applied Functional Data Analysis: Methods and Case Studies.* Springer-Verlag, New-York.

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.

Yi, J. and Prybutok, V.R. (1996). A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. *Environmental Pollution*, **3**, 349-357.