Sankhyā : The Indian Journal of Statistics
Special Issue on Quantile Regression and Related Methods
2005, Volume 67, Part 2, pp 378-398

Conditional Quantiles for Dependent Functional Data with Application to the Climatic *El Niño* Phenomenon

Frédéric Ferraty
Université Paul Sabatier, Toulouse, France
Abbes Rabhi
Université Djillali Liabes, Sidi Bel Abbes, Algeria
Philippe Vieu
Université Paul Sabatier, Toulouse, France

Abstract

This paper deals with a scalar response conditioned by a functional random variable. The main goal is to estimate nonparametrically the quantiles of such a conditional distribution when the sample is considered as an α-mixing sequence. Firstly, a kernel type estimator for the conditional cumulative distribution function (*cond-cdf*) is introduced. Afterwards, we derive an estimate of the quantiles by inverting this estimated *cond-cdf*, and asymptotic properties are stated. This approach can be applied in time series analysis. For that, the whole observed time series has to be split into a set of functional data, and the functional conditional quantile approach can be used both to forecast and to build confidence prediction bands. The *El Niño* time series illustrates this.

AMS (2000) *subject classification.* Primary 62G05, 62G99; secondary 62M10.
Keywords and phrases. Conditional quantile, conditional cumulative distribution, *El Niño* series, functional random variable, nonparametric estimation, semi-metric, α-mixing.

1 Introduction

Estimating quantiles of any distribution is an important part of statistical analysis. This allows one to build confidence ranges and leads to applications in various fields (chemistry, geophysics, medicine, meteorology, etc.). Statistics for functional random variables are also becoming more and more important. The recent literature in this area shows the great potential of these new statistical methods for functional data. The most popular case of functional random variable corresponds to the situation when we observe random curves on different statistical units. Such data are called *Functional Data*. Many multivariate statistical techniques, mainly parametric in the functional model terminology, have been extended to functional data and good overviews on this topic can be found in Ramsay and Silverman (1997 and 2002) or Bosq (2000). More recently, nonparametric methods taking into account functional variables have been developed with very interesting practical motivations dealing with environmetrics (see Damon and Guillas 2002, Fernández *et al.* 2003, Aneiros *et al.* 2004), chemometrics (see Ferraty and Vieu 2002), meteorological science (see Besse *et al.* 2000, Hall and Heckman 2002), speech recognition problems (see Ferraty and Vieu 2003a), radar range profile (see Hall *et al.* 2001, Dabo-Niang *et al.* 2004), medical data (see Gasser *et al.* 1998), etc.. On the other hand, forecasting methods cover a large part of these statistical problems. Because a continuous time series can be viewed as a sequence of dependent functional random variables, the above mentioned functional methodology can be used for time-series forecasting (see for instance Ferraty *et al.*, 2002, for a functional forecasting approach of time-series based on conditional expectation estimation).

This paper proposes to put together the three previous statistical aspects in order to derive a method for estimating conditional quantiles in situations when the data are both dependent and of functional nature. More precisely, we focus on the nonparametric estimation of the conditional quantiles of a real random variable given a functional random variable under mixing assumption. We start by estimating the conditional distribution by means of a kernel estimator, and we derive estimates of the conditional quantiles (see Section 2). From a theoretical point of view, a crucial problem is related to the so-called *curse of dimensionality*. Indeed, in a nonparametric context, it is known that the rate of convergence decreases with the dimension of the space in which the conditional variable lies. But here, the conditional variable takes its values in an infinite dimensional space. One way to overcome this problem is to consider some concentration hypotheses acting on the distribution of the functional variable, which allows us to obtain asymptotic properties of our kernel estimates (see Section 3). This approach is used to derive a new method to forecast time series, and the *El Ninõ* time series is used to illustrates the method (see Section 4).

2 The Model and the Estimates

2.1. The functional nonparametric framework. We consider a random pair (X, Y) where Y is valued in $\mathbb{R}$ and X is valued in some infinite dimensional semi-metric vector space $(\mathcal{F}, d(.,.))$. Let $(X_i, Y_i)_{i=1,\ldots,n}$ be the statistical sample of pairs which are identically distributed like (X, Y), but not necessarily independent. From now on, X is called functional random variable *f.r.v.* Let x be fixed in $\mathcal{F}$ and let $F(.|x)$ be the conditional cumulative distribution function (*cond-cdf*) of Y given $X = x$, namely:

$$\forall y \in \mathbb{R},\ F(y|x) = P(Y \leq y|X = x).$$

Saying that, we are implicitly assuming the existence of a regular version for the conditional distribution of Y given X. Now, let t_γ be the γ-order quantile of the distribution of Y given $X = x$. From the *cond-cdf* $F(.|x)$, it is easy to give the general definition of the γ-order quantile:

$$t_\gamma = \inf\{t \in \mathbb{R} :\ F(t|x) \geq \gamma\}, \qquad \forall \gamma \in (0, 1).$$

In order to simplify our framework and to focus on the main interest of our paper (the functional feature of X), we assume that $F(.|x)$ is strictly increasing and continuous in a neighbourhood of t_α. This is insuring that the conditional quantile t_γ is uniquely defined by:

$$t_\gamma \ = \ F^{-1}(\gamma|x). \tag{1}$$

In the remaining of the paper, we wish to stay in a distribution-free framework. This will lead to assume only smoothness restrictions for the *cond-cdf* $F(.|x)$ through nonparametric modeling (see Section 2.4 below). The last point of our work is to consider dependent data. It will be assumed that $(X_i, Y_i)_{i\in\mathbb{N}}$ is an α-mixing sequence, which is one among the most general mixing structures. The α-mixing condition together with the functional approach allow us to deal with continuous time processes (see Section 4 for an example).

In finite dimensional settings, nonparametric modeling is mainly used for estimating functions (such as the *cond-cdf* $F(.|x)$), in such a way that the words *functional* and *nonparametric* are quite often used equivalently. In our infinite dimensional purpose, things have to be clarified. We use the terminology *functional nonparametric*, where the word *functional* refers to the infinite dimensionality of the data and where the word *nonparametric* refers to the infinite dimensionality of the model. Such *functional nonparametric* statistics is also called *doubly infinite dimensional* (see Ferraty and Vieu, 2003b, for larger discussion). We could also use the terminology *operatorial statistics* since the target object to be estimated (the *cond-cdf* $F(.|x)$) can be viewed as a nonlinear operator.

2.2. The estimators. The kernel estimator $\widehat{F}(.|x)$ of $F(.|x)$ is defined as follows:

$$\widehat{F}(y|x) = \frac{\sum_{i=1}^{n} K\left(h_K^{-1} d(x, X_i)\right) H\left(h_H^{-1}(y - Y_i)\right)}{\sum_{i=1}^{n} K\left(h_K^{-1} d(x, X_i)\right)}, \tag{2}$$

where K is a kernel function, H a cumulative distribution function and $h_K = h_{K,n}$ (resp. $h_H = h_{H,n}$) a sequence of positive real numbers. Note that using similar ideas, Roussas (1969) introduced some related estimate but in the special case when X is real, while Samanta (1989) produced previous asymptotic study.

As a by-product of (1) and (2), it is easy to derive an estimator $\widehat{t_\gamma}$ of t_γ:

$$\widehat{t_\gamma} \;=\; \widehat{F}^{-1}(\gamma|x). \tag{3}$$

As we will see later on, such an estimator is unique as soon as H is an increasing continuous function. Such an approach has been widely used before in situation when the variable X is of finite dimension (see *e.g* Whang and Zhao, 1999, Cai, 2002, Zhou and Liang, 2003 or Gannoun *et al.*, 2003). As far as we know, our paper is the first one about kernel conditional quantile with dependent and possibly infinite dimensional variable. From a theoretical point of view the main mathematical problem comes from the so-called *curse of dimensionality.* This phenomenon is well-known in the nonparametric statistical community, and it can be summarized by saying that the rate of convergence decreases with the dimension of the space $\mathcal{F}$. Because we consider here an infinite dimensional space $\mathcal{F}$, we have to keep in mind this point. We propose to solve this dimensionality problem by taking into account the distribution of the *f.r.v.* X through its small balls probabilities (see (H0), (H2) and discussion below).

2.3. Assumptions on the functional variable. Let N_x be a fixed neighbourhood of x and let $B(x,h)$ be the ball of center x and radius h, namely $B(x,h) = \{f \in \mathcal{F}/d(f,x) < h\}$. Then, we consider the following hypotheses:

(H0) $\forall h > 0,\ P\left(X \in B(x,h)\right) = \phi_x(h) > 0,$

(H1) $(X_i, Y_i)_{i\in\mathbb{N}}$ is an α-mixing sequence whose coefficients of mixture satisfy:

$$\exists a > 0,\ \exists c > 0 :\ \forall n \in \mathbb{N},\ \alpha(n) \leq cn^{-a}.$$

(H2) $0 < \sup_{i \neq j} P\left((X_i, X_j) \in B(x,h) \times B(x,h)\right) \;=\; O\left(\frac{(\phi_x(h))^{(a+1)/a}}{n^{1/a}}\right).$

Note that (H0) can be interpreted as a concentration hypothesis acting on the distribution of the *f.r.v.* X, whereas (H2) concerns the behaviour of the joint distribution of the pairs (X_i, X_j). In fact, this hypothesis is equivalent to assume that, for n large enough

$$\sup_{i \neq j} \frac{P\left((X_i, X_j) \in B(x,h) \times B(x,h)\right)}{P\left(X \in B(x,h)\right)} \leq C \left(\frac{\phi_x(h)}{n}\right)^{1/a}.$$

This is one way to control the local asymptotic ratio between the joint distribution and its margin. Remark that the upper bound increases with a. In other words, more the dependence is strong, more restrictive is (H2). The hypothesis (H1) specifies the asymptotic behaviour of the α-mixing coefficients.

2.4. The nonparametric model. As usual in nonparametric estimation, we suppose that the *cond-cdf* $F(.|x)$ satisfies some smoothness constraints. Let b_1 and b_2 be two positive numbers; we assume that $F(.|x)$ is such that:

(H3) $\forall (x_1, x_2) \in N_x \times N_x$, $\forall (y_1, y_2) \in \mathbb{R}^2$,
$|F(y_1|x_1) - F(y_2|x_2)| \leq C\left(d(x_1,x_2)^{b_1} + |y_1 - y_2|^{b_2}\right)$,

(H4) $F(.|x)$ is j-times continuously differentiable in some neighbourhood of t_γ,

(H5) $\forall (x_1, x_2) \in N_x \times N_x$, $\forall (y_1, y_2) \in \mathbb{R}^2$,
$|F^{(j)}(y_1|x_1) - F^{(j)}(y_2|x_2)| \leq C\left(d(x_1,x_2)^{b_1} + |y_1 - y_2|^{b_2}\right)$,
where, for any positive integer l, $F^{(l)}(z|x)$ denotes its *lth* derivative, *i.e.* $\left.\frac{\partial^l F(y|x)}{\partial y^l}\right|_{y=z}$.

Later on, (H3) is used to proof the almost complete convergence of $\widehat{t_\gamma}$ whereas (H4) and (H5) are needed to establish the rate of convergence.

3 Asymptotic Study

We start this theoretical section by giving the almost complete convergence (*a.co.*) of the estimated conditional quantile $\widehat{t_\gamma}$. Thereafter, we will focus on the rate of convergence. Concerning the notations, as soon as possible, C and C' will denote generic constants. Moreover, from now on, h_H (resp. h_K) is a sequence which tends to zero with n.

3.1. Almost complete convergence. Let us start with the statement of an almost complete convergence property[1]. Before giving the result, the following assumptions concerning the kernel estimator $\widehat{F}(.|x)$ are needed:

(H6) The restriction of H to the set $\{u \in \mathbb{R},\ H(u) \in (0,1)\}$ is a strictly increasing function,

(H7) $\forall (y_1, y_2) \in \mathbb{R}^2,\ |H(y_1) - H(y_2)| \le C|y_1 - y_2|$ and $\int |t|^{b_2} H^{(1)}(t)dt < \infty$, where, for all $l \in \mathbb{N}^*$, $H^{(l)}(t) = \left.\frac{d^l H(y)}{dy^l}\right|_{y=t}$,

(H8) K is a function with support $[0,1]$ such that $0 < C_1 < K(t) < C_2 < \infty$,

(H9) $\dfrac{\log n}{n\,\phi_x(h_K)} \underset{n\to\infty}{\to} 0.$

Note that (H7) insures the existence of $\widehat{t}_\gamma$, while (H6) insures its uniqueness.

THEOREM 3.1 *Under hypotheses (H0)-(H3) and (H6)-(H9) and if*

$$\exists \eta > 0,\ C\, n^{\frac{3-a}{a+1}+\eta} \le \phi_x(h_K) \le C' n^{\frac{1}{1-a}} \tag{4}$$

holds with $a > (5 + \sqrt{17})/2$*, we have:*

$$\widehat{t}_\gamma - t_\gamma \underset{n \to \infty}{\to} 0,\ a.co. \tag{5}$$

PROOF OF THEOREM 3.1. Because of (H6) and (H7), $\widehat{F}(.|x)$ is a continuous and strictly increasing function. So, we have:

$$\forall \epsilon > 0,\ \exists \delta(\epsilon) > 0, \forall y, |\widehat{F}^x(y) - \widehat{F}^x(t_\gamma)| \le \delta(\epsilon) \Rightarrow |y - t_\gamma| \le \epsilon.$$

This leads us to write

$$\begin{aligned} \forall \epsilon > 0,\ \exists \delta(\epsilon) > 0,\ P\left(|\hat{t}_\gamma - t_\gamma| > \epsilon\right) &\le P\left(|\widehat{F}^x(\hat{t}_\gamma) - \widehat{F}^x(t_\gamma)| \ge \delta(\epsilon)\right) \\ &= P\left(|F^x(t_\gamma) - \widehat{F}^x(t_\gamma)| \ge \delta(\epsilon)\right), \end{aligned}$$

[1]Recall that a sequence $(T_n)_{n\in\mathbb{N}}$ of random variables is said to converge almost completely to some variable T, if for any $\epsilon > 0$, we have $\sum_n P(|T_n - T| > \epsilon) < \infty$. This mode of convergence implies both almost sure and in probability convergence (see for instance Bosq and Lecoutre, 1987).

since (3) is implying that $\widehat{F}(\widehat{t}_\gamma|x) = \gamma = F(t_\gamma|x)$. Thus, it is clear that the proof of Theorem 3.1 is achieved as soon as we show the pointwise convergence of $\widehat{F}(.|x)$ at t_γ:

$$F(t_\gamma|x) - \widehat{F}(t_\gamma|x) \underset{n \to \infty}{\to} 0, \ a.co. \tag{6}$$

Consider now, for $i = 1, \ldots, n$, the following notations:

$$K_i(x) = K\left(h_K^{-1}\, d(x, X_i)\right), \qquad H_i(t_\gamma) = H\left(h_H^{-1}\,(t_\gamma - Y_i)\right),$$

$$\widehat{F}_N(t_\gamma|x) = \frac{1}{n\,EK_1(x)} \sum_{i=1}^{n} K_i(x)\, H_i(t_\gamma) \text{ and } \widehat{F}_D(x) = \frac{1}{n\,EK_1(x)} \sum_{i=1}^{n} K_i(x).$$

Because some quantities are divided by $EK_1(x)$, it is important to remark once for all that (H0) and (H8) imply that

$$0 < C\phi_x(h_K) < EK_1(x) < C'\phi_x(h_K).$$

By using the following decomposition

$$\begin{aligned}\widehat{F}(t_\gamma|x) - F(t\gamma|x) &= \frac{1}{\widehat{F}_D(x)} \left\{ \left(\widehat{F}_N(t_\gamma|x) - E\widehat{F}_N(t_\gamma|x) \right) \right. \\ &\quad \left. - \left(F(t_\gamma|x) - E\widehat{F}_N(t_\gamma|x) \right) \right\} \\ &\quad + \frac{F(t_\gamma|x)}{\widehat{F}_D(x)} \left\{ E\widehat{F}_D(x) - \widehat{F}_D(x) \right\},\end{aligned} \tag{7}$$

the proof of (6) comes from next lemmas:

Lemma 3.2 *Under the conditions of Theorem 3.1, we have*

$$|F(t_\gamma|x) - E\widehat{F}_N(t_\gamma|x)| = O\left(h_K^{b_1}\right) + O\left(h_H^{b_2}\right). \tag{8}$$

Lemma 3.3 *Under the assumptions of Theorem 3.1, we have:*

i) $\widehat{F}_D(x) - E\widehat{F}_D(x) = O\left(\sqrt{\dfrac{\log n}{n\,\phi_x(h_K)}}\right), \quad a.co.,$

ii) $\widehat{F}_N(t_\gamma|x) - E\widehat{F}_N(t_\gamma|x) = O\left(\sqrt{\dfrac{\log n}{n\,\phi_x(h_K)}}\right), \quad a.co.$

Our theorem will be proved as soon as these two lemmas are verified. This will be done in the final section. □

3.2. Almost complete rate of convergence. In this section we study the rate of convergence[2] of our conditional quantile estimator $\widehat{t}_\gamma$. Because this kind of result is stronger than the previous one, we have to introduce some additional assumptions. As it is usual in conditional quantiles estimation, the rate of convergence can be linked with the flatness of the *cond-cdf* $F(.|x)$ around the conditional quantile t_γ. This is one reason why we introduced hypotheses (H4) and (H5). But a complementary way to take into account this local shape constraint is to assume that:

(H10) $\exists j > 0,\ \forall l,\ 1 \le l < j,\ F^{(l)}(t_\gamma|x) = 0$ and $\left|F^{(j)}(t_\gamma|x)\right| > 0.$

Because we focus on the local behaviour of $F(.|x)$ around t_γ via its derivatives, that leads us to consider the successive derivatives of $\widehat{F}(.|x)$ and subsequently some assumptions on the successive derivatives of the cumulative kernel H:

(H11) The support of $H^{(1)}$ is compact and $\forall l \ge j$, $H^{(l)}$ exists and is bounded.

(H12) $\forall i \ne i'$, the conditional density of $(Y_i, Y_{i'})$ given $(X_i, X_{i'})$ is continuous at (t_γ, t_γ).

THEOREM 3.4 *If the conditions (H0)-(H12) hold, and if the following inequalities*

$$\exists \eta > 0,\ C\, n^{\frac{3-a}{a+1}+\eta} \le h_H\, \phi_x(h_K) \text{ and } \phi_x(h_K) \le C' n^{\frac{1}{1-a}} \tag{9}$$

are satisfied with $a > (5+\sqrt{17})/2$*, we have*

$$\hat{t}_\gamma - t_\gamma = O\left(h_K^{\frac{b_1}{j}} + h_H^{\frac{b_2}{j}}\right) + O\left(\left(\frac{\log n}{n\, \phi_x(h_K)}\right)^{\frac{1}{2j}}\right), \quad a.co. \tag{10}$$

PROOF OF THEOREM 3.4. The proof is based on the Taylor expansion of $\widehat{F}(.|x)$ at t_γ and on the use of (H10):

$$\begin{aligned}\widehat{F}(t_\gamma|x) - \widehat{F}(\widehat{t}_\gamma|x) &= \sum_{l=1}^{j-1} \frac{(t_\gamma - \widehat{t}_\gamma)^l}{l!} \widehat{F}^{(l)}(t_\gamma|x) + \frac{(t_\gamma - \widehat{t}_\gamma)^j}{j!} \widehat{F}^{(j)}(t^*|x),\\ &= \sum_{l=1}^{j-1} \frac{(t_\gamma - \widehat{t}_\gamma)^l}{l!} \left(\widehat{F}^{(l)}(t_\gamma|x) - F^{(l)}(t_\gamma|x)\right) + \frac{(t_\gamma - \widehat{t}_\gamma)^j}{j!} \widehat{F}^{(j)}(t^*|x),\end{aligned}$$

[2]Recall that a sequence $(T_n)_{n\in\mathbb{N}}$ of random variables is said to be of order of complete convergence u_n, if there exists some $\epsilon > 0$ for which $\sum_n P(|T_n| > \epsilon u_n) < \infty$. This is denoted by $T_n = O(u_n)$, *a.co.* (or equivalently by $T_n = O_{a.co.}(u_n)$).

where, for all $y \in \mathbb{R}$,

$$\widehat{F}^{(j)}(y|x) = \frac{h_H^{-j} \sum_{i=1}^n K\left(h_K^{-1} d(x, X_i)\right) H^{(j)}\left(h_H^{-1}(y - Y_i)\right)}{\sum_{i=1}^n K\left(h_K^{-1} d(x, X_i)\right)},$$

and where $\min(t_\gamma, \widehat{t}_\gamma) < t^* < \max(t_\gamma, \widehat{t}_\gamma)$. Suppose now that we have the following result.

Lemma 3.5 *Suppose that the hypotheses (H0)-(H8), (H10)-(H12) are satisfied, (9) holds, and*

$$\lim_{n\to\infty} \frac{\log n}{n\, h_H^{2j-1}\, \phi_x(h_K)} = 0.$$

Then we have:

$$\widehat{F}^{(j)}(t_\gamma|x) - F^{(j)}(t_\gamma|x)| \;=\; O(h_K^{b_1}) + O(h_H^{b_2}) + O\left(\sqrt{\frac{\log n}{n\, h_H^{2j-1}\, \phi_x(h_K)}}\right), \quad a.co.$$

Because of Theorem 3.1, Lemma 3.5 and (H10), we have:

$$\widehat{F}^{(j)}(t^*|x) \underset{n\to\infty}{\rightarrow} F^{(j)}(t_\gamma|x) \;>\; 0,\; a.co.,$$

and we derive

$$(t_\gamma - \widehat{t}_\gamma)^j = O\left(\widehat{F}(t_\gamma|x) - F(t_\gamma|x)\right) + O\left(\sum_{l=1}^{j-1} (t_\gamma - \widehat{t}_\gamma)^l (\widehat{F}^{(l)}(t_\gamma|x) - F^{(l)}(t_\gamma|x))\right), a.co. \qquad (11)$$

Now, comparing the convergence rates given in Lemmas 3.3 and 3.5, we get

$$(t_\gamma - \hat{t}_\gamma)^j \;=\; O\left(\widehat{F}(t_\gamma|x) - F(t_\gamma|x)\right),\; a.co.$$

Thus, Lemmas 3.2 and 3.3 allow us to get the claimed result. It remains to prove Lemma 3.5 which is done in the final section. □

4 Time Series Forecasting and Application to *El Niño*

This section deals with a functional approach for time series forecasting, based on the splitting of the observed time series into several continuous functional data. We start by describing the *El Niño* time series, and then we show how it can be viewed as a set of functional dependent variables. Thereafter, we will explain how a forecasting method can be built from the estimation of the conditional quantiles defined in Section 2. Finally, we will show how this forecasting approach behaves on the real *El Niño* data-set.

4.1. El Niño data. Much information and several data-sets can be found about the phenomenon called *El Niño*. Our study concerns the monthly time series of the Sea Surface Temperature (SST) from June, 1950 up to May, 2004, available on line at: "http://www.cpc.ncep.noaa.gov/data/indices/" (see also: "http://kdd.ics.uci.edu/databases/el_nino/el_nino.data.html" for relevant discussion on these data). These temperatures are measured by moored buoys in the "Niño region" defined by the coordinates 0-10°South and 90°West-80°West. Our SST time series comes from the average of the monthly temperatures over the moored buoys in this area. Finally, the statistical sample is of size 648. The graphical display is given in Figure 1.

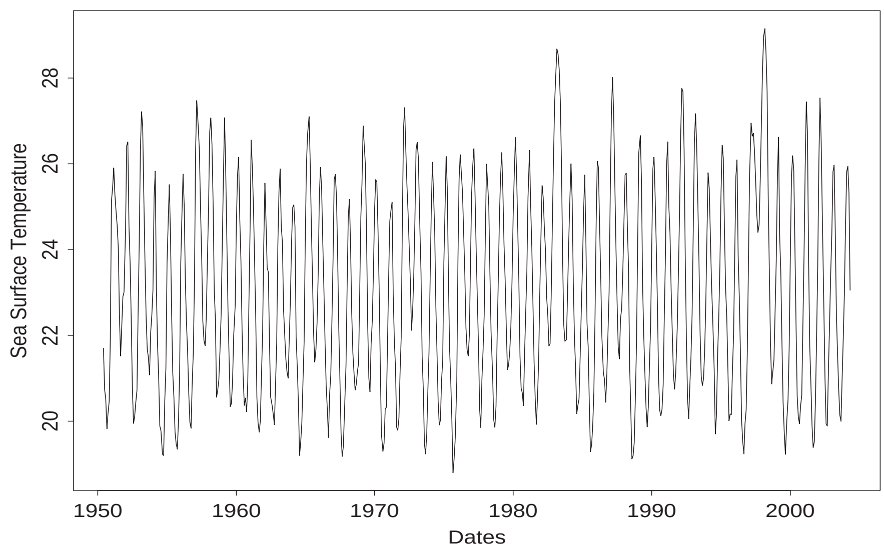

Figure 1. El Niño monthly Sea Surface Temperature (0-10°South and 90°West-80°West)

4.2. Splitting El Niño time series into functional data. A useful way to display such a time series consists in cutting it into 54 pieces or 54 "annual curves" (see Figure 2). More precisely, let $\{Z(k)\}_{k=1,\ldots,648}$ be our El Niño time series. We can build, for $i = 1, \ldots, 54$, the following subsequences:

$$\forall t \in \{1, 2, \ldots, 12\}, \qquad z_i(t) = Z(12 * (i-1) + t),$$

$\boldsymbol{z}_i = (z_i(1), \ldots, z_i(12))$ corresponding to the variations of the SST at the i^{th} year. Because the climatic phenomenon is changing continuously over time, there is evidence for considering each annual curve as a continuous path (i.e. $Z_i = \{Z(12 * (i-1) + t),\ t \in [0; 12)\}$). Of course, this continuous yearly curve will be observed only at some discretized points (here, at 12 discretized points). Finally, the time series can be viewed as a sample of 54 dependent functional data, namely $Z_1, \ldots, Z_{54}$.

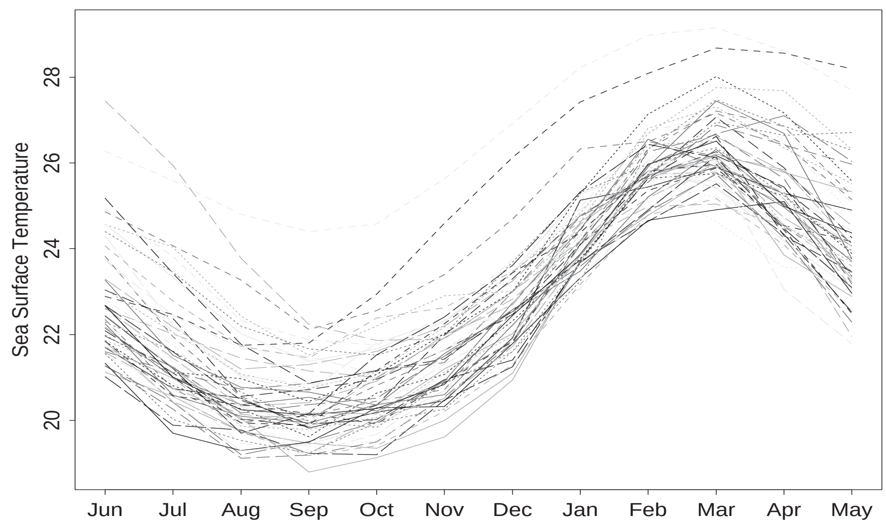

Figure 2. El Niño monthly Sea Surface Temperature displayed Year by Year

As discussed before, the main advantage of using such continuous path for the past of the time series (rather than using some multivariate discretized past vector) is to be unsensitive to the curse of dimensionality. More precisely, the standard nonparametric methods (see Cai, 2002 or Gannoun *et al.*, 2003) are either considering only one single past value (with obvious loss of information) or more than one past value (but being then sensitive to the curse of dimensionality). Our approach is able to capture much information in the past of the time series, but still using for the past a single continuous object (exactly one whole year is taken into account) and avoiding for the dimensionality effects. To see this more precisely, let us suppose for instance that the time series could be measured p-times each year with $p > 12$ (actually, it is true with $p = 12$). In this case, our functional method using the whole continuous past year will have the same asymptotic behaviour (independently of p). Having more than 12 yearly data measurements would provide a gain in terms of prediction, since the observation of the curves would be more precise. This is not really the case with classical un-functional nonparametric methods. To take into account a whole past year, any non-functional method will have to deal with a p-dimensional vector and the rates of convergence of the estimate will decrease exponentially with p, loosing from one side what could be gained in an other hand by having larger statistical sample.

4.3. Choosing the parameters of the estimate. As usual in nonparametric kernel smoothing, the choice of the shape of the kernels is of less importance than the choice of the bandwidths. For this application, we have constructed our kernel from the standard Epanechnikov's parabolic kernel:

$$G(u) \;=\; \frac{3}{4}(1-u^2)\,1_{(-1,1)}(u).$$

More precisely, we have used:

$$K(u) \;=\; 2G(u)1_{(0,1)}(u) \quad \text{and} \quad H(u) \;=\; \int_{-\infty}^{u} G(t)dt.$$

The choice of the smoothing parameters h_H and h_K is much more crucial for the performance of any kernel approach and should be data-driven as described in the next section.

Another important point for ensuring a good behaviour of the method, is to use a semi-metric that is well adapted to the kind of data we have to deal with. Here, we used some semi-metric based on the q first terms of the Functional Principal Components Analysis (FPCA) of the data (see Appendix). Indeed, we do not choose a semi-metric at the beginning of the study but only a family of semi-metrics, and the key question is more to select the best semi-metric inside of the family than to choose the family itself. The key parameter is the order q of the FPCA expansion[3], which should also be chosen in a data-driven way.

4.4. Data-driven step and forecasting procedure. We have at hand a set of 54 functional data. However, to show the performance of our method, we will ignore the 54th year and we will predict it from the 53's previous ones. Moreover, we will build our statistical method only on the 52 previous data, the 53rd being used as a learning step to select the parameters in a data-driven way. For $i = 1, \ldots, 52$ and for any fixed δ in $\{1, \ldots, 12\}$, we take

$$Y_i(\delta) \;=\; z_{i+1}(\delta).$$

[3]Other choices of semi-metrics are possible, like for instance those based on some L_2 errors between higher order derivatives of the curves, and in this case the key parameter would be the order of the derivative (see Ferraty and Vieu, 2003a, for discussion).

Thus, we build a sample of 52 pairs $(Y_i(\delta), Z_i)_{i=1,\ldots,52}$, considered as a sample of dependent pairs, the $Y_i(\delta)$'s being real r.v. and the Z_i's being functional r.v. According to Section 2, one can estimate $Y_{52}(\delta)$ knowing Z_{52} by estimating the median of the conditional distribution:

$$\widehat{Y_{52}(\delta)} \;=\; \widehat{t}_{\frac{1}{2}} \;=\; \widehat{F}^{-1}\left(\frac{1}{2}\middle| Z_{52}\right),$$

where $\widehat{F}(.|Z_{52})$ is the estimated distribution of $Y_1(\delta)$ given Z_{52}. Repeating this step for $\delta = 1, \ldots, 12$, we can estimate the variations of the SST for the 53rd year. The bandwidths[4] h_K and h_H, as well as the parameter q of the semi-metric, are chosen in order to minimize $\sum_{\delta=1}^{n}(\widehat{Y_{52}(\delta)} - Y_{52}(\delta))^2$ which corresponds to a prediction error over the 53rd year. Thereafter, given $(Y_i(\delta), Z_i)_{i=1,\ldots,52}$, and given these optimally selected parameters h_H, h_K and q, we estimate $\widehat{F}(.|Z_{53})$ and derive via the conditional median an estimation of $Z_{54}(\delta)$ for $\delta = 1, \ldots, 12$. We estimate $\{Z_{54}(\delta)\}_{\delta=1,\ldots,12}$ by using only the 53's first years, and so this is clearly a forecasting procedure. Of course, all the steps described before can also be used for estimating other conditional quantiles than the median. This could provide confidence prediction bands for the forecasted 54th year.

4.5. The results on El Niño time series. The bandwidths h_H and h_K have been chosen automatically by the learning sample procedure described above, as well as the value $q = 3$ for the FPCA semi-metric. Figure 3 gives the results by using conditional median $\hat{t}_{1/2}$ as prediction tool (discontinuous line $- - - -$) and conditional quantiles $\hat{t}_{.05}$ and $\hat{t}_{.95}$ to build pointwise 90% confidence prediction band (dotted lines $\cdots\cdots$). In Figure 3, the vertical line means that the first 12 values (the 53rd year) are estimated along the learning step, whereas the last 12 values (the 54th year) are forecasted. The continuous line (——) shows the observed values.

Undoubtedly, Figure 3 shows the good behaviour of our functional forecasting procedure (at least for these data). Of course, there is need for comparing this method with other ones (such as un-functional quantiles approaches or such as functional conditional expectation approaches), as well on this *El Nino* time series as on some other ones. This is outside the scope of this paper and will be the object of further works. Note that our S+

[4]To speed up our algorithm, the set of possible bandwidths among which the optimal ones are selected is chosen to be finite and is computed by the k-nearest neighbours type bandwidths.

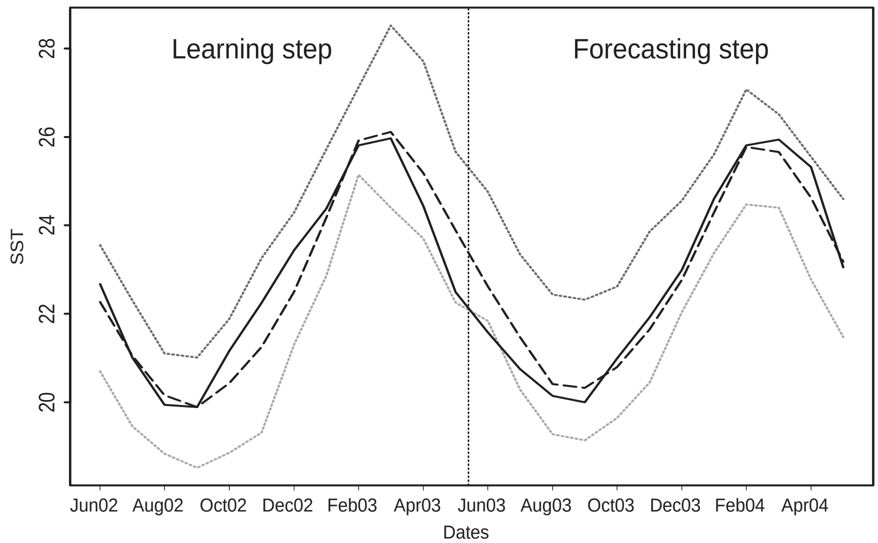

Figure 3. Functional quantile forecasting for *El Niño* time series

procedures are available on request, in such a way that any user could treat other time series data-sets and any researcher could use it for participating to further developments in this field.

5 Proofs of Technical Lemmas

To save space and to emphasize on the main contributions of our paper (i.e. α-mixing and functional variables) some details are omitted, and are available on request.

PROOF OF LEMMA 3.2. The asymptotic behaviour of bias term is standard, in the sense that it is not affected by the dependence structure of the data. We have

$$E\widehat{F}_N(t_\gamma|x)-F(t_\gamma|x) \;=\; \frac{1}{EK_1(x)}E\left(K_1(x)\left\{E\left(H_1(t_\gamma)/X\right)-F(t_\gamma|x)\right\}\right) \quad (12)$$

and by noting that

$$E\left(H_1(t_\gamma)|X\right) \;=\; \int_{\mathbb{R}} H^{(1)}(t)F^X(t_\gamma-h_H t)dt,$$

we can write, because of (H3) and (H7):

$$|E\left(H_1(t_\gamma)|X\right) - F(t_\gamma|x)| \;\leq\; C_x \int_{\mathbb{R}} H^{(1)}(t)\left(h_K^{b_1} + |t|^{b_2} h_H^{b_2}\right) dt.$$

Combining this last result with (12) allows us to achieve this proof. □

Proof of Lemma 3.3 i). Following the ideas used in regression (see Ferraty and Vieu, 2004), the main point consists in using a pseudo-exponential inequality taking into account the α-mixing structure. We start by writing

$$\widehat{F}_D(x) - E\widehat{F}_D(x) = \frac{1}{n\,EK_1(x)} \sum_{i=1}^{n} \Delta_i(x),$$

where $\Delta_i(x) = K_i(x) - EK_i(x)$. The Fuk-Nagaev's inequality (Rio 1999, formula 6.19b) allows one to get, for all $\lambda > 0$ and $r > 1$:

$$P\left(|\widehat{F}_D(x) - E\widehat{F}_D(x)| \;>\; 4\lambda\right)$$
$$\leq \;\; C\left\{\underbrace{\frac{n}{r}\left(\frac{r}{\lambda\, n\, EK_1(x)}\right)^{a+1}}_{Q_1} + \underbrace{\left(1 + \frac{\lambda^2\, n^2\, \left(EK_1(x)\right)^2}{r\, s_n}\right)^{r/2}}_{Q_2}\right\},$$

where

$$s_n = \sum_{i=1}^{n}\sum_{j=1}^{n} Cov\left(\Delta_i(x), \Delta_j(x)\right).$$

By taking

$$r = C(\log n)^2 \text{ and } \lambda = \lambda_0 \frac{\sqrt{n\,\phi_x(h_K)\,\log n}}{n\,EK_1(x)}, \tag{13}$$

and by using the left part of inequality (4), it follows that:

$$Q_1 \;\leq\; C\, n^{-1-\nu}. \tag{14}$$

Before we focus on Q_2, we have to study the asymptotic behaviour of

$$s_n \;=\; \underbrace{\sum_{i\neq j} Cov\left(\Delta_i(x), \Delta_j(x)\right)}_{s_n^{cov}} + \underbrace{\sum_{i=1}^{n} Var\left(\Delta_i(x)\right)}_{s_n^{var}}. \tag{15}$$

On one hand, we have by using successively (H8), (H0), (H2) and the right part of (4):

$$|Cov(\Delta_i(x), \Delta_j(x))| \;=\; O\left(\left(\frac{\phi_x(h_K)}{n}\right)^{1/a} \phi_x(h_K)\right). \qquad (16)$$

On the other hand, these covariances can be controlled by means of the usual Davydov's covariance inequality for mixing processes (see Rio 1999, formula 1.12a). Together with (H1), this inequality leads to:

$$\forall i \neq j,\; |Cov(\Delta_i(x), \Delta_j(x))| \;\leq\; C\,|i-j|^{-a}. \qquad (17)$$

Thus, by using the following classical technique (see Bosq, 1998), we can write

$$s_n^{cov} \;=\; \sum_{0<|i-j|\leq u_n} |Cov(\Delta_i(x), \Delta_j(x))| + \sum_{|i-j|>u_n} |Cov(\Delta_i(x), \Delta_j(x))|\,.$$

By choosing $u_n = \left(\frac{\phi_x(h_K)}{n}\right)^{-1/a}$, and using (16) (resp. (17)) to treat the first (resp. second) covariance term, we get:

$$s_n^{cov} \;=\; O\left(n\,\phi_x(h_K)\right). \qquad (18)$$

The computation of the variance terms can be done by following the same arguments as those invoked to get (16), and we arrive at:

$$Var\left(\Delta_i(x)\right) \;=\; O\left(\phi_x(h_K)\right). \qquad (19)$$

Finally, (18) and (19) lead directly to:

$$s_n \;=\; O\left(n\,\phi_x(h_K)\right). \qquad (20)$$

This is enough to study the quantity Q_2, since (13) and (20) allow us to write that, for n and λ_0 large enough:

$$\exists \nu' > 0, \qquad Q_2 \;\leq\; C\,n^{-1-\nu'}. \qquad (21)$$

Finally, put together (14) and (21) and use (H8) to achieve the proof of Lemma (3.3) *i).* □

Proof of Lemma 3.3 ii). It proceeds along the same steps and by invoking the same arguments, just changing the variables $\Delta_i(x)$ into the following ones:

$$\Gamma_i(x) = H_i(t_\gamma)\,K_i(x) - E\,H_i(t_\gamma)\,K_i(x).$$

Because H is a cumulative kernel, we have $H_i(t_\gamma) \leq 1$. By using systematically this fact to bound the variables H_i, all the calculus made previously with the variables $\Delta_i(x)$ remain valid with the variables $\Gamma_i(x)$. □

Proof of Lemma 3.5. We use again the same kind of decomposition as in (7):

$$\begin{aligned} \widehat{F}^{(j)}(t_\gamma|x) - F^{(j)}(t_\gamma|x) &= \frac{1}{\widehat{F}_D(x)} \left\{ \left(\widehat{F}_N^{(j)}(t_\gamma|x) - E\widehat{F}_N^{(j)}(t_\gamma|x) \right) \right. \\ & \left. - \left(F^{(j)}(t_\gamma|x) - E\widehat{F}_N^{(j)}(t_\gamma|x) \right) \right\} \\ & + \frac{F^{(j)}(t_\gamma|x)}{\widehat{F}_D(x)} \left\{ E\widehat{F}_D(x) - \widehat{F}_D(x) \right\}. \end{aligned} \qquad (22)$$

This proof is very similar to the one of Theorem 3.1. Firstly, we consider the bias term $F^{(j)}(t_\gamma|x) - E\widehat{F}_N^{(j)}(t_\gamma|x)$. Using the same arguments as along the proof of Lemma 3.2, replacing $F(t_\gamma|x)$ (resp. $\widehat{F}(t_\gamma|x)$) with $F^{(j)}(t_\gamma|x)$ (resp. $\widehat{F}_N^{(j)}(t_\gamma|x)$) and considering in addition hypotheses (H5), (H7) and (H11) we get:

$$F^{(j)}(t_\gamma|x) - E\widehat{F}_N^{(j)}(t_\gamma|x) = O\left(h_K^{b_1}\right) + O\left(h_K^{b_2}\right). \qquad (23)$$

Now, we focus on the term $\widehat{F}_N^{(j)}(t_\gamma|x) - E\widehat{F}_N^{(j)}(t_\gamma|x)$. To get the asymptotic behaviour of this quantity, we follow the framework of the proof of Lemma 3.3. To do that, once again replace $F(t_\gamma|x)$ (resp. $\widehat{F}(t_\gamma|x)$) with $F^{(j)}(t_\gamma|x)$ (resp. $\widehat{F}_N^{(j)}(t_\gamma|x)$). Note that (H11) and (H12) allow us to show that

$$E\left(H^{(j)}(h_H^{-1}(t_\gamma - Y_i))H^{(j)}(h_H^{-1}(t_\gamma - Y_{i'}))|(X_i, X_{i'})\right) = O(h_H^2),$$

while (H1) and (H5) imply that

$$E\left(H^{(j)}(h_H^{-1}(t_\gamma - Y_i))|X_i\right) = O(h_H).$$

Moreover, the right part of (9) implies that

$$Cov\left(\Gamma_i^*(x), \Gamma_{i'}^*(x)\right) = O\left(h_H^2 \left(\frac{\phi_x(h_K)}{n}\right)^{1/a} \phi_x(h_K)\right),$$

where

$$\Gamma_i^*(x) = H^{(j)}(h_H^{-1}(t_\gamma - Y_i))K_i(x) - E\left(H^{(j)}(h_H^{-1}(t_\gamma - Y_i))K_i(x)\right).$$

So, by using similar arguments as those invoked in the proof of Lemma 3.3, we deduce easily that for any $u_n > 0$:

$$\sum_{i=1}^{n}\sum_{i'=1}^{n} Cov\left(\Gamma_i^*(x), \Gamma_{i'}^*(x)\right) = O\left(n\, u_n\, h_H^2 \left(\frac{\phi_x(h_K)}{n}\right)^{1/a} \phi_x(h_K)\right) + O\left(n^2\, h_H^2\, u_n^{-a}\right) + O\left(n\, h_H\, \phi_x(h_K)\right).$$

It suffices now to take $u_n = h_H^{-1}\left(\frac{\phi_x(h_K)}{n}\right)^{-1/a}$ to get the following expression for the sum of the covariances:

$$\sum_{i=1}^{n}\sum_{i'=1}^{n} Cov\left(\Gamma_i^*(x), \Gamma_{i'}^*(x)\right) = O\left(n\, h_H\, \phi_x(h_K)\right).$$

Finally, because we have:

$$P\left(\left|\widehat{F}^{(j)}(t_\gamma|x) - E\widehat{F}^{(j)}(t_\gamma|x)\right| > \lambda\right) = P\left(\left|\sum_{i=1}^{n}\Gamma_i^*(x)\right| > \lambda\, n\, h_H^j\, EK_1(x)\right),$$

the application of Fuk-Nagaev's inequality with $\lambda = \lambda_0 \frac{\sqrt{n\, h_H\, \phi_x(h_K)\, \log n}}{n\, h_H^j\, EK_1(x)}$, leads directly to the result of Lemma 3.5. □

Appendix

In many multivariate situations, the classical Principal Components Analysis (PCA) is considered as a useful tool for displaying data in a reduced dimensional space. Following the same idea, the functional PCA is also a good tool for computing proximities between curves in a reduced dimensional space. As long as $E\int X^2(t)dt$ is finite, the Functional Principal Components Analysis (FPCA) (see Dauxois *et al.*, 1982) of the functional r.v. X allows one to obtain the following expansion of X:

$$X = \sum_{k=1}^{\infty}\left(\int X(t)v_k(t)dt\right) v_k,$$

v_1, v_2,..., being the orthonormal eigenfunctions of the covariance operator

$$\Gamma_X(s,t) = E\left(X(s)X(t)\right)$$

associated with the eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$. Now, let

$$\widetilde{X}^{(q)} = \sum_{k=1}^{q} \left(\int X(t) v_k(t) dt \right) v_k$$

be a truncated version of the previous expansion of X. Thus, for all $(x_1, x_2) \in \mathcal{F}^2$ we can deduce a parametrized family of semi-metrics in the following way:

$$\sqrt{\sum_{k=1}^{q} \left(\int [x_1(t) - x_2(t)] v_k(t) dt \right)^2}.$$

Here, q is not really a smoothing parameter but rather a tuning parameter indicating the resolution level at which the problem is considered. Note that in practice Γ_X is unknown and then the $v_k's$ too, but the covariance function can be well approximated by its empirical version

$$\Gamma_X^n(s,t) = 1/n \sum_{i=1}^{n} X_i(s) X_i(t),$$

and the eigenfunctions $v_{k,n}(.)$ of Γ_X^n are consistent estimators of those of Γ_X (for more details, see Cardot *et al.*, 1999). So, the semi-metrics used in our practical study are defined as:

$$d_q(x_1, x_2) = \sqrt{\sum_{k=1}^{q} \left(\int [x_1(t) - x_2(t)] v_{k,n}(t) dt \right)^2}.$$

Acknowledgments. All the participants of the group STAPH on Functional and Operatorial Statistics in the Department of Statistics and Probability of Toulouse are greatly acknowledged for their numerous comments and suggestions. Recent activities of this group can be found in Staph (2004). The authors are also grateful to one anonymous referee for his(her) comments which have contributed to nicer presentation of this work.

References

Aneiros Pérez, G., Cardot, H., Estévez Pérez, G. and Vieu, P. (2004). Maximum ozone forecasting by functional nonparametric approach. *Environmetrics*, **15**, 675-685.

Besse, P., Cardot, H. and Stephenson, D. (2000). Autoregressive forecasting of some functional climatic variations. *Scand. J. Statist.*, **27**, 673-687.

BOSQ, D. (1998). *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction* (Second edition). Lecture Notes in Statistics **110**, Springer-Verlag, Berlin.

BOSQ, D. (2000). *Linear Processes in Function Spaces.* Lecture Notes in Statistics **149**, Springer-Verlag, Berlin.

BOSQ, D. and LECOUTRE, J.P. (1987). *Théorie de l'Estimation Fonctionnelle* (in French). Economica, Paris.

CAI, Z. (2002). Regression quantiles for time series. *Econometric Theory*, **18**, 169-192.

CARDOT, H., FERRATY, F. and SARDA, P. (1999). Linear Functional Model. *Statist. Probab. Lett.*, **45**, 11-22.

DABO-NIANG, S., FERRATY, F. and VIEU, P. (2004). Nonparametric unsupervised classification of satellite wave altimeter forms. In *Computational Statistics 2004*, J. Antoch, ed., Physica Verlag, Berlin, 879-887.

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 application to statistical inference. *J. Multivariate Anal.*, **12**, 136-154.

FERNÁNDEZ DE CASTRO, B., GUILLAS, S. and GONZÁLEZ MANTEIGA, W. (2003). Functional samples and bootstrap for the prediction of SO2 levels. *Preprint*.

FERRATY, F., GOIA, A. and VIEU, P. (2002). Functional nonparametric model for times series: A fractal approach for dimensional reduction. *TEST*, **11**, 317-344.

FERRATY, F. and VIEU, P. (2002). The functional nonparametric model and application to spectrometric data. *Computat. Statist.*, **17**, 545-564.

FERRATY, F. and VIEU, P. (2003a). Curves discrimination: a nonparametric functional approach. *Comput. Statist. Data Anal.*, **44**, 161-173.

FERRATY, F. and VIEU, P. (2003b). Functional nonparametric statistics: a double infinite dimensional framework. In *Recent Advances and Trends in Nonparametric Statistics*, M. Akritas and D. Politis, eds., Elsevier, Amsterdam.

FERRATY, F. and VIEU, P. (2004). Nonparametric methods for functional data, with applications in regression, time-series prediction and curves discrimination. *J. Nonparametr. Statist.*, **16**, 11-126.

GANNOUN, A., SARACCO, J. and YU, K. (2003). Nonparametric prediction by conditional median and quantiles. *J. Statist. Plann. Inference*, **117**, 207-223.

GASSER, T., HALL, P. and PRESNELL, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *J. Roy. Statist. Soc. Ser. B*, **60**, 681-691.

HALL, P. and HECKMAN, N. (2002). Estimating and depicting the structure of the distribution of random functions, *Biometrika*, **89**, 145-158.

HALL, P., POSKITT, P. and PRESNELL, D. (2001). A functional data-analytic approach to signal discrimination. *Technometrics*, **35**, 140-143.

RAMSAY, J.O. and SILVERMAN, B.W. (1997). *Functional Data Analysis.* Springer, Berlin.

RAMSAY, J.O. and SILVERMAN, B.W. (2002). *Applied Functional Data Analysis.* Springer-Verlag, New York.

RIO, E. (2000). *Théorie Asymptotique des Processus Aléatoires Dépendants* (in French). Mathématiques et Applications **31**, Springer, Berlin.

ROUSSAS, G. (1969). Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Math. Statist.*, **40**, 1386-1400.

SAMANTA, M. (1989). Nonparametric estimation of conditional quantiles. *Statist. Probab. Lett.*, **7**, 407-412.

STAPH WORKING GROUP (2004). Statistique fonctionnelle et opératorielle. Partie V (in French). *Publ. du Laboratoire de Statistique et Probabilités de Toulouse*, LSP 2004-06. (Available on request).

WANG, H. and ZHAO, Y. (1999). A kernel estimator for conditional t-quantiles for mixing samples and its strong uniform convergence (in Chinese). *Math. Appl.* (Wuhan), **12**, 123-127.

ZHOU, Y. and LIANG, H. (2003). Asymptotic properties for L_1 norm kernel estimator of conditional median under dependence. *J. Nonparametr. Statist.*, **15**, 205-219.

FRÉDÉRIC FERRATY AND PHILIPPE VIEU
LABORATOIRE DE STATISTIQUE ET PROBABILITÉS
UNIVERSITÉ PAUL SABATIER
31062 TOULOUSE, FRANCE
E-mail: ferraty@cict.fr
vieu@cict.fr

ABBES RABHI
UNIVERSITÉ DJILLALI LIABES
SIDI BEL ABBES, ALGERIA
E-mail: rabhi_abbes@yahoo.fr

Paper received: July 2004; revised April 2005.