LINEAR PROCESSES FOR FUNCTIONAL DATA

André Mas, Besnik Pumo

LINEAR PROCESSES FOR FUNCTIONAL DATA

André Mas[a] and Besnik Pumo[b]
[a] I3M, Université Montpellier 2,
Place Eugène Bataillon, 34095 Montpellier, France
mas@math.univ-montp2.fr

[b] Agrocampus Ouest, Centre d'Angers,
2 rue Le Nôtre, 49045 Angers, France
besnik.pumo@agrocampus-ouest.fr

Abstract

Linear processes on functional spaces were born about fifteen years ago. And this original topic went through the same fast development as the other areas of functional data modeling such as PCA or regression. They aim at generalizing to random curves the classical ARMA models widely known in time series analysis. They offer a wide spectrum of models suited to the statistical inference on continuous time stochastic processes within the paradigm of functional data. Essentially designed to improve the quality and the range of prediction, they give birth to challenging theoretical and applied problems. We propose here a state of the art which emphasizes recent advances and we present some promising perspectives based on our experience in this area.

1 Introduction

The aim of this chapter is double. First of all we want to provide the reader with basic theory and application of linear processes for functional data. The second goal consists for us in giving a state of the art which complements the monograph by Bosq (2000). Many crucial theorems were given in this latter book to which we will frequently refer. Consequently, even if our work is self-contained we pay special attention to recent results, published from 2000 to 2008, and try to draw the lines of future and promising research in this area.

It is worth recalling now the approach that leads to modelizing and inferring from curves-data. We start from a continuous time stochastic process $(\xi_t)_{t\geq 0}$. The paths of ξ are cut into equally spaced pieces of trajectories. Each of these piece is then viewed as a random curve. With mathematical

symbols we set:

$$X_k(t) = \xi_{kT+t}, \quad 0 \leq t \leq T$$

where T is fixed. The function $X_k(\cdot)$ maps $[0,T]$ to $\mathbb{R}$ and is random. Observing ξ over $[0,nT]$ produces a n sample $X_1, ..., X_n$. Obviously the choice of T is crucial and is usually left to the practitioner and may be linked with seasonality (with period T). Dependence along the paths of ξ will create dependence between the X_i's. But this approach is not restricted to the whole path of stochastic process. One could as well imagine to model whole curves observed at discrete intervals: the interest rate curves at day k $I_k(\delta)$ is for instance a function linking duration δ (as an input) and the associated interest rates (as outputs). Observing these curves, whose random variations will depend on financial markets, along n days produces a sample similar in nature to the one described above, although there is no underlying continuous-time process in this situation, rather a surface $(k, \delta, I_k(\delta))$. We refer for instance to Kargin and Onatski (2008) for an illustration.

Statistical models will then be proposed and mimic or adapt the scalar or finite-dimensional approaches for time-series (see Brockwell, Davis (1987)). Each of these (random or not) functions will be viewed as a vector in a vector space of functions. This paradigm has been adopted for a long time in probability theory as will be seen through, for instance, Ledoux and Talagrand (1991) and references therein. But the first book entirely dedicated to the formal and applied aspects of statistical inference in this setting is certainly due to Ramsay and Silverman (1997), followed by Bosq (2000), Ramsay and Silverman (2002) again then Ferraty and Vieu (2006).

In the sequel we will consider centered processes with values in a Hilbert space of functions denoted H with inner product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|$. The Banach setting though more general has several drawbacks. Some references will be given yet throughout this section. The reason for privileging Hilbert spaces are both theoretic and practical. First many fundamental asymptotic theorems are stated under simple assumptions in this setting. The central limit theorem is a good example. Considering random variables with values in $C([0,1])$ or in Hölder spaces for instance lead to very specific assumptions to get the CLT and computations are often uneasy whereas in a Hilbert space moment conditions are usually both necessary and sufficient. The nice geometric features of Hilbert space allow us to consider denumerable bases, projections, etc in a framework that generalizes the euclidean space with few drawbacks. Besides, in practice, recovering curves from discretized observations by interpolation or smoothing techniques such as splines or wavalets yields functions in the Sobolev spaces, say $W^{m,2}$ (here m is an order

of differentiation connected with the desired smoothness of the output), are all Hilbert spaces. We refer to Ziemer (1989) or to Adams, Fournier (2003) for monographs on Sobolev spaces.

In statistical models, unknown parameters will be functions or linear operators (the counterpart of matrices of the euclidean space), the latter being of utter interest. We give now some basic facts about operators which will be of great use in the sequel.

Several monographs are dedicated to operator theory, which is a major theme within the mathematical science. Classical references are Dunford, Schwartz (1988) and Gohberg, Goldberg and Kaashoek (1991). The adjoint of the operator T is classically denoted T^*. The Banach space of compact operators $\mathcal{C}$ on a Hilbert space H is separable when endowed with the classical operator norm $\|\cdot\|_\infty$:

$$\|T\|_\infty = \sup_{x \in \mathcal{B}_1} \|Tx\|$$

where $\mathcal{B}_1$ denotes the unit ball of the Hilbert space H. The space $\mathcal{C}$ contains the set of Hilbert-Schmidt operators which is a Hilbert space and denoted $\mathcal{S}$. Let T and S belong to $\mathcal{S}$ the inner product between T and S and the norm of T are respectively defined by:

$$\langle S, T \rangle_{\mathcal{S}} = \sum_p \langle Se_p, Te_p \rangle ,$$
$$\|T\|_{\mathcal{S}}^2 = \sum_p \|Te_p\|^2$$

where $(e_p)_{p \in \mathbb{N}}$ is a complete orthonormal system in H. The inner product and the norm defined just above do not depend on the choice of the c.o.n.s.$(e_p)_{p \in \mathbb{N}}$. The nuclear (or trace-class) operators are another important family of operators for which the series:

$$\sum_p \|Te_p\| < +\infty.$$

It is plain that a trace class operator is Hilbert-Schmidt as well. Many of the asymptotic result mentioned from now on and involving random operators are usually obtained for the Hilbert-Schmidt norm, unless explicitly mentioned. It should be noted as well that this norm is thinner than the usual operator norm.

The next section is devoted to general linear processes. Then we will focus on the autoregressive model and its recent advances, which will be developed in the third section. We will conclude with some issues for future work.

2 General linear processes

The linear processes on function spaces generalize the classical scalar or vector linear processes to random elements which are curves or functions and more generally valued in an infinite-dimensional separable Hilbert space H.

Definition 1 *Let $(\epsilon_k)_{k\in\mathbb{N}}$ be a sequence of i.i.d. centered random elements in H and let $(a_k)_{k\in\mathbb{N}}$ be a sequence of bounded linear operators from H to H such that $a_0 = I$ and $\mu \in H$ be a fixed vector. If*

$$X_n = \mu + \sum_{j=0}^{+\infty} a_j\left(\epsilon_{n-j}\right), \tag{1}$$

$(X_n)_{n\in\mathbb{N}}$ is a linear process on H (denoted in the sequel H-linear process) with mean μ.

Unless explicitly mentioned the mean function μ will always be assumed to be null (and the process X is centered). Its seems that, after a collection of paper dating back to the end of the 90's-early 00's the model creates less inspiration in the community. We guess that the recent works by Bosq (2007) and the book by Bosq and Blanke (2007) may bring some fresh ideas. We state here some basic facts: invertibility and convergence of estimated moments.

2.1 Invertibility

When the sequence ϵ is a strong H-white noise, that is a sequence of i.i.d. random elements such that $E\left\|\epsilon\right\|^2 < +\infty$ and whenever

$$\sum_{j=0}^{+\infty} \left\|a_j\right\|_\infty^2 < +\infty \tag{2}$$

the series defining the process $(X_n)_{n\in\mathbb{N}}$ through (1) converges in square norm and almost surely through the $0-1$ law. The strict stationarity of X_n is ensured as well. The problem of invertibility is addressed in Merlevède (1995).

Theorem 1 *If $(X_n)_{n\in\mathbb{N}}$ is a linear process with values in H defined by (1) and such that :*

$$1 - \sum_{j=1}^{+\infty} z^j \left\|a_j\right\|_\infty \neq 0 \quad for\ \ |z| < 1 \tag{3}$$

then $(X_n)_{n\in\mathbb{N}}$ *is invertible:*

$$X_n = \epsilon_n + \sum_{j=1}^{+\infty} \rho_j \left(X_{n-j}\right)$$

where all the ρ_j*'s are bounded linear operators in* H *with* $\sum_{j=1}^{+\infty} \|\rho_j\|_\infty < +\infty$ *and the series converges in mean square and almost surely.*

Remark 1 *We deduce from the latter that* ϵ_n *is the innovation of the process* X *and that (1) coincides with the Wold decomposition of* X*.*

We give now some convergence theorems for the mean and the covariance of Hilbert-valued linear processes. These results are not completely new but essential.

2.2 Asymptotics

It is worth mentioning a general scheme for proving asymptotic results for linear processes. If several approaches are possible, it turns out that, up to the authors' opinion, one of the most fruitful relies on approximating the process X_n by truncated versions like:

$$X_{n,m} = \sum_{j=0}^{m} a_j \left(\epsilon_{n-j}\right)$$

where $m \in \mathbb{N}$. The sequence $X_{n,m}$ is for fixed m, blockwise independent: $X_{n+m+1,m}$ is indeed stochastically independent from $X_{n,m}$ if the ϵ_j's are. The outline of the proofs usually consists in proving asymptotic results for the m-dependent sequence $X_{n,m}$, then to let m tend to infinity with an accurate control of the residual $X_n - X_{n,m} = \sum_{j=m+1}^{+\infty} a_j \left(\epsilon_{n-j}\right)$.

2.2.1 Mean

Asymptotic results for the mean of a linear process may be found in Merlevède (1996) and Merlevède, Peligrad and Utev (1997). Even if the first is in a way more general, we focus here on the second article since it deals directly with the mean of the non-causal process indexed by $\mathbb{Z}$:

$$X_k = \sum_{j=-\infty}^{+\infty} a_j \left(\epsilon_{k-j}\right).$$

The authors obtain sharp conditions for the CLT of $S_n = \sum_{k=1}^{n} X_k$.

Theorem 2 *Let $(a_j)_{j\in\mathbb{Z}}$ be a sequence of operators such that:*

$$\sum_{j=-\infty}^{+\infty} \|a_j\|_\infty^2 < +\infty$$

Then

$$\frac{S_n}{\sqrt{n}} \to_w N\left(0, A\Gamma_\epsilon A^*\right)$$

where $N\left(0, A\Gamma_\epsilon A^*\right)$ is the H-valued centered gaussian random element with covariance operator $A\Gamma_\epsilon A^*$ where $\Gamma_\epsilon = E\left(\epsilon_0 \otimes \epsilon_0\right)$ is the covariance operator of ϵ_0 and $A = \sum_{j=-\infty}^{+\infty} a_j$.

Remind that if u and v are two vectors in H then notation $u \otimes v$ stands for the rank-one linear operator from H to H defined by: $(u \otimes v)(x) = \langle v, x\rangle u$.

This result is extended with additional assumptions to the case of strongly mixing ϵ_k's. Note that the problem of weak convergence for the mean of stationary Hilbertian process under mixing conditions had been addressed in the early 80's by Maltsev and Ostrovski (1982). A standard equi-integrability argument and classical techniques provide the following rates of convergence for S_n. Nazarova (2000) proved the same sort of theorem when X is a linear random field with values in a Hilbert space. Now we turn to the rate of convergence of the empirical mean in quadratic mean and almost surely. The following theorem may be found in Bosq (2000).

Proposition 1 *Let $X_k = \sum_{k=0}^{+\infty} a_j\left(\epsilon_{k-j}\right)$ and $S_n = \sum_{k=1}^{n} X_k$ then*

$$nE\left\|\frac{S_n}{n}\right\|^2 \to \sum_{k=-\infty}^{+\infty} E\left\langle X_0, X_k\right\rangle,$$

$$\frac{n^{1/4}}{(\log n)^{1/2+\epsilon}}\left\|\frac{S_n}{n}\right\| \to 0 \quad a.s.$$

for all $\epsilon > 0$.

We turn to covariance operators now.

2.2.2 Covariance operators

The situation is slightly more complicated than for the mean due to the tensor product.

Definition 2 *The theoretical covariance operator at lag $h \in \mathbb{N}$ of a process X is defined by:*

$$\Gamma_h = \mathbb{E}\left(X_h \otimes X_0\right).$$

The linear operator Γ_h is nuclear on H when the second order strong moments of the X is convergent. Its empirical counterpart based on the sample is:

$$\Gamma_{n,h} = \frac{1}{n}\sum_{t=1}^{n} X_{t+h} \otimes X_t.$$

The covariance operator of the process, $\Gamma_0 = \Gamma$ is selfadjoint, positive and nuclear hence Hilbert-Schmidt and compact.

It should be noted that Γ_h is not in general a symmetric operator conversely to the classical covariance operator Γ_0. The weak convergence of covariance operators for H-linear processes was addressed by Mas (2002). It is assumed that:

$$\mathbb{E}\left\|\epsilon_0\right\|^4 < +\infty$$
$$\sum_{k=-\infty}^{+\infty} \left\|a_k\right\|_\infty < +\infty$$

then the vector of the h covariance operators up to any fixed lag h is asymptotically gaussian in the Hilbert-Schmidt norm.

Theorem 3 *Let us consider the following linear and Hilbert space valued process*

$$X_t = \sum_{j=-\infty}^{+\infty} a_j\left(\epsilon_{t-j}\right)$$

then

$$\sqrt{n}\begin{pmatrix} \Gamma_{n,0} - \Gamma_0 \\ \Gamma_{n,1} - \Gamma_1 \\ \ldots \\ \Gamma_{n,h} - \Gamma_h \end{pmatrix} \underset{n\to+\infty}{\overset{w}{\to}} G_\Gamma$$

where $G_\Gamma = \left(G_\Gamma^{(0)}, ..., G_\Gamma^{(h)}\right)$ is a Gaussian centered random element with values in $\mathcal{S}^{h+1}$. Its covariance operator is $\Theta_\Gamma = \left(\Theta_\Gamma^{(p,q)}\right)_{0\le p,q\le h}$ which is a nuclear operator in $\mathcal{S}^{h+1}$ defined blockwise for all T in $\mathcal{S}$ by

$$\Theta_\Gamma^{(p,q)}(T) = \sum_h \Gamma_{h+p-q} T \Gamma_h + \sum_h \Gamma_{h+q} T \Gamma_{h-p} + A_q\left(\Lambda - \Phi\right) A_p(T) \qquad (4)$$

where Λ, Φ *and* A_p *are linear operators from* $\mathcal{S}$ *to* $\mathcal{S}$ *respectively defined by*

$$\Lambda(T) = \mathbb{E}\left((\epsilon_0 \otimes \epsilon_0) \widetilde{\otimes} (\epsilon_0 \otimes \epsilon_0)\right)(T)$$
$$\Phi(T) = C(T + T^*)C + \left(C \widetilde{\otimes} C\right)(T)$$

and $A_p(T) = \sum_i a_{i+p} T a_i^*$.

As by-products, weak convergence results for the eigenelements of $\Gamma_{n,0} - \Gamma_0$, that is for the PCA of the stationary process X, are derived. The reader interested with these developments should refer to the paper by Mas and Menneteau (2003a) which proposes a general method to derive asymptotics for the eigenvalues and eigenvectors of $\Gamma_{n,0}$ (the by-products of the functional PCA) from the covariance sequence itself. Perturbation theory is the main tool through a modified delta-method.

2.3 Perspectives and trends: towards generalized linear processes ?

It turns out that the literature based on inference methods for general linear processes is rather meager. Obviously estimating simultaneously many a_j's seems to be intricate and not necessarily needed as the functional AR process, which will be enlightened in the next section, is quite successful and easier to handle. However general linear processes are the starting point for very interesting theoretical problems where dependence plays a key-role. We mention at last the abstract papers by Merlevède and Dedecker (2003) especially section 2.4 dedicated to proving a conditional central limit theorem for linear processes under mild assumptions and to Merlevède and Dedecker (2007) whose section 3 deals with rates in the law of large numbers. These works may provide theoretical material to go further into the asymptotic study of these processes.

In a very recent article, Bosq (2007) introduces the notion of linear process in the wide sense. The definition remains essentially the same as in display (1) but the operators $(a_j)_{j \in \mathbf{N}}$ may then be unbounded; which finally generalizes the notion. A key role is played by linear closed spaces (LCS) introduced by Fortet. A LCS $\mathcal{G}$ is a subspace of L_H^2 -the space of random variables with values in H and finite strong second moment- such that:

(i) $\mathcal{G}$ is closed in H.

(ii) If $X \in \mathcal{G}$, $l(X) \in \mathcal{G}$ for all bounded linear operator l.

This theory -involving projection on LCS, weak and strong orthogonality, dominance of operators- allows Bosq to revisit and extend the notions of linear process, Wold decomposition, Markovian process when the bounded $(a_j)_{j\in\mathbf{N}}$ may be replaced with measurable mappings $(l_j)_{j\in\mathbf{N}}$. Several examples are given : derivatives of functional processes like in the MAH $X_n = \epsilon_n + c\epsilon'_n$, arrays of linear processes, truncated Ornstein-Uhlenbeck process... The personnal communication Bosq (2009) discusses these extensions to tensor products of linear processes and will certainly shed a new light at their covariance structure. We also refer to chapters 10 and 11 in the book by Bosq and Blanke (2007) for an exposition of these concepts.

3 Autoregressive processes

3.1 Introduction

The model generalizes the classical AR(1) for scalar or multivariate time series to functional data and was introduced for the first time in Bosq (1991). Let $X_1, \ldots, X_n$ be a sample of random curves for which a stochastic dependence is suspected (for instance the curve of temperature observed during n days at a given place). We assume that all the X_i's are valued in a Hilbert space H and set:

$$X_n = \rho\left(X_{n-1}\right) + \epsilon_n \tag{5}$$

where ρ is a linear operator from H to H and $(\epsilon_n)_{n\in\mathbb{N}}$ is a sequence of H valued centered random elements usually with common covariance operator. The model is simple, with a single unknown operator, leaving however the possibility to decline various assumptions either on the operator ρ (linear, compact, Hilbert-Schmidt, symmetric or not, etc) or on the dependence between the ϵ_n's. The latter are quite often independent and identically distributed but alternatives are possible (mixing or more naturally martingale differences). Bosq (2000) proved that assumption (2) comes down actually to the existence of $a > 0$, $b \in [0, 1[$such that for all $p \in \mathbb{N}$:

$$\left\|\rho^p\right\|_\infty \leq ab^p$$

which ensures that (5) admits a unique stationary solution. The process $(X_n)_{n\in\mathbb{N}}$ is Markov as soon as $E\left(\epsilon_n | X_{n-1}, \ldots, X_1\right) = 0$. As often noted the interest of the model relies in its predictive power. The estimation of ρ is usually the first and necessary step before deriving the statistical predictor given the new input X_{n+1}: $\widehat{\rho}\left(X_n\right)$. The prediction are often compared with ARMA model or with non-parametric smoothing techniques. The global

treatment of the trajectory as a function often ensures better long-run prediction but at the expense of more tedious numerical procedures.

3.1.1 Representation of stochastic processes by functional AR

Various real valued processes allow the ARH representation. We plotted on Figure 1 graphs of two simulated processes, the Ornstein-Uhlenbeck (O-U) process and the Wong process. The O-U process $(\eta_t, t \in R)$ is a real stationary Gaussian process :

$$\eta_t = \int_{-\infty}^{t} e^{-a(t-u)} dw_u, t \in R$$

where $(w_t)_{t\in R}$ is a bilateral standard Wiener process and a a positive constant. Bosq (1996) gives the ARH representation $X_n = \rho(X_{n-1}) + \epsilon_n$ with values in $L^2 := L^2[0,1]$ where $X_n(t) = \eta_{n+t}, t \in [0,1], n \in Z$ and ρ is a degenerated linear operator

$$\rho(x)(t) = e^{-at} x(1), t \in [0,1], x \in L^2$$

and

$$\epsilon_n(t) = \int_{n}^{n+t} e^{-a(n+t-s)} dw_s, t \in [0,1], n \in Z.$$

The Wong process is a mean-square differentiable stationary Gaussian process which is zero-mean and is defined for $t \in R$ by:

$$\xi_t = \sqrt{3}\, \exp\left(-\sqrt{3}\, t\right) \int_0^{\exp\left(2t/\sqrt{3}\right)} w_u \, du.$$

Cutting R in intervals of length 1 and defining $X_n(t) = \xi_{n+t}$ for $t \in [0,1]$, Mas and Pumo (2007) obtain ARH representation, $X_n = A(X_{n-1}) + \epsilon_n$ of this process with values in Sobolev space $W := W^{2,1} = \left\{u \in L^2, u' \in L^2\right\}$ and

$$\epsilon_n(t) = \sqrt{3}\, \exp[-\sqrt{3}\,(n-1+t)] \int_{\exp[2(n-1)/\sqrt{3}]}^{\exp[2(n-1+t)/\sqrt{3}]} \left[w_u - w_{\exp[2(n-1)/\sqrt{3}]}\right] du$$

for $t \in [0,1]$. The linear and degenerated operator A is given by $\phi + \Psi(D)$ where D is the ordinary differential operator and

$$[\phi(f)](t) = [\exp(-\sqrt{3}t) + \sqrt{3}c(t)]f(1), \quad [\Psi(D)(f)](t) = c(t)f'(1).$$

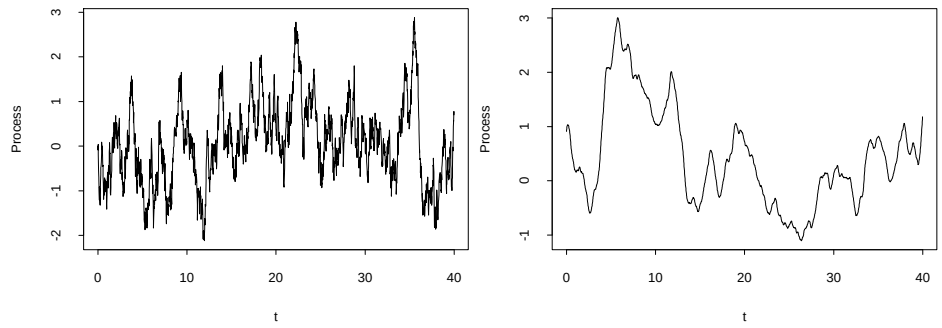

Figure 1: Ornstein-Uhlenbeck ($a = 1$) and Wong processes both evaluated at instants $t_i = 0.02 * i$.

with $c(t) = \frac{\sqrt{3}}{2} \cdot \exp(-\sqrt{3}t) \cdot \{\exp(2t/\sqrt{3}) - 1\}$.

Other examples are given in the paper of Bosq (1996) or in the classical book of Bosq (2000).

A major issue would be to infer deeper links between autoregressive functional processes and diffusion processes or stochastic differential equations with more general autocorrelation operators. But this remains an open question and the work by Ramsay (2000) about this topic should certainly deserve more attention to be extended.

3.1.2 Asymptotics for the mean and covariance

Obviously all the results obtained for general linear processes hold for the ARH(1): namely for the mean and the covariance operator. Some new results are stated below -they are new essentially with respect to Bosq (2000)- and are related with moderate deviations (Mas and Menneteau (2003b)) and laws of the iterated logarithm (Menneteau (2003)). Let η be a square integrable real valued random variable. We need to introduce the following notations (I_X and J_X are functions from H to H, J_Γ is a function from S

to S and K_Γ is a subset of S):

$$\begin{aligned} u_1 &= \rho(X_0) \otimes \epsilon_1 + \epsilon_1 \otimes \rho(X_0) + \epsilon_1 \otimes \epsilon_1 - \Gamma_\epsilon, \\ I_X(x) &= \sup_{h \in H} \{\langle h, x - \rho(x) \rangle - \mathbb{E} \exp \langle h, \epsilon_1 \rangle\}, \\ J_X(x) &= \frac{1}{2} \inf \left\{ E\eta^2 : x = E\left[\eta (I_H - \rho)^{-1}(\epsilon_1)\right] \right\}, \\ J_\Gamma(s) &= \frac{1}{2} \inf \left\{ E\eta^2 : s = E\left[\eta (I_S - R)^{-1}(u_1)\right] \right\}, \\ K_\Gamma &= \left\{ E\left[\eta (I_S - R)^{-1}(u_1)\right] : \eta \in L^2(P), E\eta^2 \leq 1 \right\} \end{aligned} \tag{6}$$

and R is a linear operator from S to S defined by $R(s) = \rho s \rho^*$.

We refer to Dembo and Zeitouni (1993) for an exposition on large and moderate deviations.

Theorem 4 *The empirical mean of the ARH(1) process follows the large deviation principle in H with speed n^{-1} and rate function I_X and the moderate deviation principle with rate function J_X.*
The covariance sequence of the ARH(1), $\Gamma_n - \Gamma$, follows the moderate deviation principle in the space of Hilbert-Schmidt operators with rate function J_Γ and the law of the iterated logarithm with limit set K_Γ.

The first results obtained on the covariance sequence are in Bosq (1991) but we mention here the interesting decomposition given in Bosq (1999).

Proposition 2 *Let X_n be an ARH(1) such that $E\|X_0\|^4 < +\infty$, then the tensorized process*

$$Z_i = X_i \otimes X_i - \Gamma$$

is an autoregressive process with values in $\mathcal{S}$ such that:

$$Z_i = R(Z_{i-1}) + u_i$$

where $R(S) = \rho S \rho^*$ and u_1 was defined at (6). The sequence u_i is a martingale difference with respect to the filtration $\sigma(\epsilon_i, \epsilon_{i-1}, ...)$.

3.2 Two issues related to the general estimation problem

3.2.1 Identifiability

The moment method provides the following normal equation:

$$\Delta = \rho \Gamma \tag{7}$$

where

$$\Gamma = \mathbb{E}\left(X_1 \otimes X_1\right),$$
$$\Delta = \mathbb{E}\left(X_2 \otimes X_1\right)$$

are the covariance operator (resp. the cross covariance operator of order one) of the process $(X_n)_{n\in\mathbb{Z}}$.

The first step consists in checking that the Yule-Walker equation (7) correctly defines the unknown parameter ρ.

Proposition 3 *When the inference on ρ is based on the moment equation (7), identifiability holds if* $\ker\Gamma = \{0\}$.

The proof of this proposition is plain since taking $\widetilde{\rho} = \rho + u \otimes v$ where v belongs to the kernel of Γ, whenever this set is non-empty we see that

$$\widetilde{\rho}\Gamma = \rho\Gamma + u \otimes \Gamma v = \rho\Gamma$$

hence that (7) holds for $\widetilde{\rho} \neq \rho$.
Consequently the injectivity of Γ is a basic assumption which can hardly be removed and which entails that the eigenvalues of Γ are infinite, strictly positive. These eigenvalues will be denoted $(\lambda_i)_{i\in\mathbb{N}}$ where one assumes once and for all that the λ_i's are arranged in a decreasing order with $\sum_{i\in\mathbb{N}} \lambda_i$ finite. The corresponding eigenvectors (resp. eigenprojectors) will be denoted $(e_i)_{i\in\mathbb{N}}$ (resp. $(\pi_i)_{i\in\mathbb{N}}$ where $\pi_i = e_i \otimes e_i$). Heuristically we should expect with (7) at hand that Γ^{-1} exists to estimate ρ and this inverse will not be defined if Γ is not one to one.

3.2.2 The inverse problem

Even if the identifiability is ensured estimating ρ is a difficult task due to an underlying inverse problem which stems from display (7). The notion of inverse (or ill-posed) problem is classical in mathematical analysis (see for instance Tikhonov, Arsenin (1977) or Groetsch (1993)). In our framework it could be explained by claiming that equation (7) will imply that any attempt to estimate ρ will result in a highly unstable estimate. This comes down with simple words, which will be developed below, from the inversion of Γ. A canonical example of an inverse problem is the numerical inversion of an ill-conditioned matrix (that is a matrix with eigenvalues close to zero).

The first stumbling stone comes from the fact that we cannot deduce from (7) that $\Delta\Gamma^{-1} = \rho$. We know that a sufficient condition for Γ^{-1} to

be defined as a linear mapping is: $\ker \Gamma = \{0\}$. Then Γ^{-1} is an unbounded symmetric operator on H. Some consequences are collected in the next proposition:

Proposition 4 *When Γ is injective Γ^{-1} may be defined. It is a linear measurable mapping defined on a dense domain in H, denoted $\mathcal{D}\left(\Gamma^{-1}\right)$ and defined by:*

$$\mathcal{D}\left(\Gamma^{-1}\right) = \mathrm{Im}\Gamma = \left\{ x = \sum_{p=1}^{+\infty} x_p e_p \in H,\ \sum_{p=1}^{+\infty} \frac{x_p^2}{\lambda_p^2} < +\infty \right\}.$$

This domain is dense in H. It is not an open set for the norm topology of H. The operator is unbounded which means that it is continuous at no point of $\mathcal{D}\left(\Gamma^{-1}\right)$. Besides $\Gamma^{-1}\Gamma = I_H$ but $\Gamma\Gamma^{-1} = I_{\mathcal{D}(\Gamma^{-1})}$ and $\Gamma\Gamma^{-1}$, which is not defined on the whole H, may be continuously extended to H.

For similar reasons (7) implies $\Delta\Gamma^{-1} = \rho_{|\mathrm{Im}\Gamma} \neq \rho$ and $\Delta\Gamma^{-1}$ may be continuously and formally extended to the whole H. In fact Γ hence Γ^{-1} are unknown. However would Γ be totally accessible we should find a way to regularize the odd mathematical object that is Γ^{-1}. Within the literature on inverse problems (see for instance Groetsch (1993)) one often replaces Γ^{-1} by a linear operator "close" to it but endowed with additional regularity (continuity/boundedness) properties, say $\Gamma^{\dagger}$. The Moore-Penrose pseudo inverse is an example of such an operator but many other techniques exist. Indeed starting from

$$\Gamma^{-1}(x) = \sum_{l \in \mathbb{N}} \frac{1}{\lambda_l} \pi_l(x)$$

for all x in $\mathcal{D}\left(\Gamma^{-1}\right)$ one may set for instance:

$$\Gamma^{\dagger}(x) = \sum_{l \leq k_n} \frac{1}{\lambda_l} \pi_l(x) \tag{8}$$

$$\Gamma^{\dagger}(x) = \sum_{l} \frac{1}{\lambda_l + \alpha_n} \pi_l(x) \tag{9}$$

$$\Gamma^{\dagger}(x) = \sum_{l} \frac{\lambda_l}{\lambda_l^2 + \alpha_n} \pi_l(x) \tag{10}$$

where k_n is an increasing and unbounded sequence of integers and α_n a sequence of positive real numbers decreasing to 0. The three operators in the display above are indexed by n, are all bounded with increasing norm

and are known as the spectral cut-off, penalized and Tikhonov regularized inverses of Γ. They share the following pointwise convergence property:

$$\Gamma^{\dagger} x \to \Gamma^{-1} x$$

for all x in $\mathcal{D}\left(\Gamma^{-1}\right)$.

In practice if Γ_n is a convergent estimator of Γ the regularizing methods introduced below can be applied to Γ_n which is usually not invertible (see below for an example). It should be noted at this point that the regularization for the inverse of the covariance operator appears in the linear regression model for functional variable:

$$y = \langle X, \varphi \rangle + \epsilon$$

when estimating the unknown φ (see Cardot et al. (2007)).

At last a general scheme to estimate ρ may be proposed with estimates of Γ and Δ at hand say Γ_n and Δ_n: compute $\Gamma_n^{\dagger}$ and take for the estimate and the predictor based on the new input X_{n+1} respectively:

$$\widehat{\rho}_n = \Delta_n \Gamma_n^{\dagger} \quad \text{and} \quad \widehat{\rho}_n \left(X_{n+1}\right).$$

Obviously examples of such estimates are the empirical covariance and cross-covariance operators

$$\Gamma_n = \frac{1}{n} \sum_{k=1}^{n} X_k \otimes X_k,$$

$$\Delta_n = \frac{1}{n-1} \sum_{k=1}^{n-1} X_{k+1} \otimes X_k$$

where the X_k's were reconstructed by interpolation techniques.

For instance the spectral cut-off version for Γ_n is

$$\Gamma_n^{\dagger} = \sum_{l \leq k_n} \frac{1}{\widehat{\lambda}_l} \widehat{\pi}_l$$

where the eigenvalues $\widehat{\lambda}_l$ and the eigenprojectors $\widehat{\pi}_l$ are by-products of the functional PCA of the sample $X_1, ..., X_n$.

Remark 2 *This inverse problem is the main serious abstract concern when infering on the ARH model. The considerations above are moreless exposed in all the articles dealing with it and we guess it will be of some interest to expose and sum up this issue and some of its solutions in this monograph.*

3.3 Convergence results for the autocorrelation operator and the predictor

As the data are of functional nature, the inference on ρ cannot be based on likelihood. Lebesgue's measure cannot be defined on infinite-dimensional spaces. However it must be mentioned that Mourid and Bensmaïn (2006) propose to adapt Grenander's theory of sieves (Grenander (1981) and Geman and Hwang (1982)) to this issue. They prove consistency in two very important cases: when ρ is a kernel operator and when ρ is Hilbert-Schmidt. In the former case ρ is identified with the associated kernel K, developed on a basis of trigonometric functions along the sieve:

$$\Theta_m = \left\{ K \in L^2 : K(t) = c_0 + \sum_{k=1}^{m} \sqrt{2} c_k \cos(2\pi k t),\ t \in [0,1],\ \sum_{k=1}^{m} k^2 c_k^2 \leq m \right\}.$$

This approach is truly original within the literature on functional data and could certainly be extended to other problems of linear or non linear regression.

The seminal paper dealing with the estimation of the operator ρ dates back to 1991 and is due to Bosq (1991). Several consistency results are carried out immediately relayed by Pumo's (1992) and Mourid's (1995) PhD thesis. Then Pumo (1998) focus on random functions with values in $C([0,1])$ with specific techniques. Besse and Cardot (1996), then Besse, Cardot and Stephenson (2000) implement spline and kernel methodology with application to climatic variations. Amongst several interesting ideas they introduce a local covariance estimate:

$$\widehat{\Gamma}_{h_n} = \frac{\sum_{i=1}^{n} [X_i \otimes X_i] K(\|X_i - X_n\| / h)}{\sum_{i=1}^{n-1} K(\|X_i - X_n\| / h)}$$

and a local cross-covariance estimate which emphasize data close to the last observation. This method make it possible to consider data with departures from the stationarity assumption. This issue of the estimation of ρ is also treated in Guillas (2001) and Mas (2004).

A recent paper by Antoniadis and Sapatinas (2003) carry out wavelet estimation and prediction in the ARH(1) model. The inverse problem is underlined through a class of estimates stemming from the deterministic literature on this topic. This class of estimates is compatible with wavelet techniques and lead to consistency of the predictor. The method is applied on the "El Nino" dataset which tends to become a benchmark for comparing the performances of the predictions.

Ruiz-Medina et al (2007) consider the functional principal oscillation pattern (POP) decomposition of the operator ρ as an alternative to functional PCA decomposition. They implement a Kalman filter to the state-space equation obtained at the preceding step and derive the optimal predictor. This original approach, illustrated by some simulations, seems to be suited to spatial functional data as well.

Kargin and Onatski (2008) introduce the notion of predictive factor which seems to be better suited than the PCA basis to project the data if one really focuses on the predictor (and not on the operator itself). A double regularization (penalization and projection) provides them with a rate of $O\left(n^{-1/6}\log^{\beta} n\right)$ (where $\beta > 0$) for the prediction mean square error.

In Mas (2007) the problem of weak convergence is addressed. The main results are given in the Theorem below:

Theorem 5 *It is impossible for $\widehat{\rho}_n - \rho$ to converge in distribution for the classical norm topology of operators. But under moment assumptions, if $\left\|\Gamma^{-1/2}\rho\right\|_{\infty} < +\infty$ and if the spectrum of Γ is convex then when $k_n = o\left(\dfrac{n^{1/4}}{\log n}\right)$,*

$$\sqrt{\frac{n}{k_n}}\left(\widehat{\rho}_n\left(X_{n+1}\right) - \rho\widehat{\Pi}_{k_n}\left(X_{n+1}\right)\right) \xrightarrow{w} \mathcal{G}$$

where $\mathcal{G}$ is a H-valued gaussian centered random variable with covariance operator Γ_ϵ and $\widehat{\Pi}_{k_n}$ is the projector on the k_n first eigenvectors of Γ_n.

Remark 3 *The first sentence of the Theorem above is quite surprising but is a direct consequence of the underlying inverse problem. Finally considering the predictor weakens the topology and has a smoothing effect on $\widehat{\rho}_n$. This phenomenon -which was exploited in Antoniadis, Sapatinas (2003)- appears as well in the linear regression model for functional data (see Cardot, Mas, Sarda (2007)).*

It should be noted that rates of convergence are difficult to obtain (see Guillas (2001) or Kargin and Onatski (2008), Theorem 3) and rather slow with respect to those obtained in the regression model. An exponential inequality appears at Theorem 8.8 in Bosq (2000) but it seems that a more systematic study of the mean square prediction error has not been carried out yet and that optimal bounds are not available.

3.3.1 Hypothesis testing

A very recent article by Horvath, Huskova and Kokoszka (2009) focuses on the stability of the autocorrelation operator against change-point alternatives. In fact the model (5) based on the sample $X_1, ..., X_n$ is slightly modified to:

$$X_n = \rho_n (X_{n-1}) + \epsilon_n$$

and the authors test

$$H_0 : \rho_1 = ... = \rho_n$$

against the alternative:

$$H_A : \text{ there exists } k^* \in \{1, ..., n\} : \rho_1 = ... = \rho_{k^*} \neq \rho_{k^*+1} = .. = \rho_n.$$

The test is based on the projection of the process X_n on the p first eigenvectors of the functional PCA and on an accurate approximation of the long-run covariance matrix. The asymptotic distribution is derived by means of empirical process techniques. The consistency of the test is obtained and a simulation/real case study dealing with credit card transaction time series is treated.

It turns out that Laukaïtis and Rackauskas (2002) considered the same sort of problem a few years sooner. They introduce a functional version of the partial sum process of estimated residuals:

$$S(t) = \sum_{k=2}^{\lfloor t \rfloor} [X_k - \widehat{\rho}(X_{k-1})]$$

and obtain weak convergence results for its normalized version to an H-valued Wiener process. This formal theorem yields different strategies (dyadic increment of partial sums or moving residual sums) to derive a test.

It seems however that the topic of hypothesis testing was rarely addressed yet quite promising even if serious theoretic and technical problems appear, once again in connection with the inverse problem mentioned earlier in this article.

3.4 Extension of ARH model

Various extensions have been proposed for ARH(1) model in order to improve the prediction performance of ARH(1) model. The first one is the natural extension autoregressive process of order p with $p > 1$, denoted ARH(p), defined by

$$X_n = \rho_1 X_{n-1} + \ldots + \rho_p X_{n-p} + \epsilon_n.$$

Using the Markov representation $Y_n = \rho' Y_{n-1} + \epsilon'_n$ where

$$\rho' = \begin{bmatrix} \rho_1 & \rho_2 & \cdots & \rho_n \\ I & 0 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & & & 1 \end{bmatrix}, Y_n = \begin{bmatrix} X_n \\ X_{n-1} \\ \vdots \\ X_{n-p+1} \end{bmatrix}, \text{and } \epsilon'_n = \begin{bmatrix} \epsilon_n \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

and I denotes the identity operator, Mourid (2003) obtain asymptotic results of projector estimators and predictors.

Damon and Guillas (2002) introduced autoregressive Hilbertian process with exogenous variables model, denoted ARHX(1), which intends to take into account the dependence structure of random curves under the influence of explanatory variables. The model is defined by the equation

$$X_n = \rho(X_{n-1}) + a_1(Z_{n,1}) + \ldots a_q(Z_{n,q}) + \epsilon_n, n \in Z$$

where $a_1, \cdots, a_q$ are bounded linear operators in H and $Z_{n,1}, \cdots, Z_{n,q}$ are ARH(1) exogenous variables; they suppose that the noises of the $q+1$ H−valued autoregressive processes are independent. They obtain some limit theorems, derive consistent estimators, present a simulation study in order to illustrate the accuracy of the estimation and compare the forecasts with other functional models.

Guillas (2002) consider a H-valued autoregressive stochastic sequence (X_n) with several regimes such that the underlying process (I_n) is stationary. Under some dependence assumptions on (I_n) he proves the existence of a unique stationary solution and state a law of large numbers and the consistency of the covariance estimator. Following the same idea in a recent work Mourid (2004) introduces and studies the autoregressive process with random operators $X_n = \rho_n X_{n-1} + \epsilon_n$ where $(\rho_n, n \in Z)$ is stationary and independent of (ϵ_n). Results similar to classical $ARH(1)$ are obtained.

A new model, denoted ARHD process, considering the derivative curves of an ARH(1) model was introduced by Marion and Pumo (2004). In a recent paper Mas and Pumo (2007) introduced and study a slightly new model:

$$X_n = \phi X_{n-1} + \Psi(X'_{n-1}) + \epsilon_n$$

where X_n are random function with values in the Sobolev space $W^{2,1} = \{u \in L^2[0,1], u' \in L^2[0,1]\}$, ϕ is a compact operator from W to W, Ψ is a compact operator from $L^2[0,1]$ to $W^{2,1}$ and $\|\phi h + \Psi h'\| \leq \|h\|$ for $h \in W^{2,1}$. Convergent estimates are obtained through an original double penalization method. Simulations on real data show that predictions are comparable to

those obtained by other classical methods based on ARH(1) modelization. Tests on the derivative part and models with higher derivatives may be interesting from both theoretical and practical point of view.

3.5 Numerical aspects

We present in this section some numerical aspects concerning the prediction when data are curves observed at discrete points. To our knowledge the prediction methods based on linear processes are limited to application of ARH(1) model since tractable algorithms using general linear processes in Hilbert spaces do not exist (see Merlevède (1997)). However some partial results are available for moving average processes in Hilbert spaces which will be briefly discussed in the next section.

The literature using ARH(1) model to make prediction is various and rich and concern different domains:

- Environment: Besse et al. (2000); Antoniadis and Sapatinas (2003); Mas and Pumo (2007); Fernández de Castro et al. (2005); Damon et Guillas (2002);
- Economy and finance: Kargin and Onatski (2008);
- Electricity consumption: Cavallini et al. (1994);
- Medical sciences: Marion and Pumo (2004); Glendinning and Fleet (2007)

From a technical point of view the different approaches for implementing an ARH proceed in two steps. The first step consists in decomposing data in some functional basis in order to reconstruct them on the whole observed interval. Most of the methods use spline or wavelet basis and suppose that curves belong to the Sobolev $W^{2,k}$ space of functions such that the k-th derivative is squared integrable. We invite the reader to refer to the papers by Besse and Cardot (1996), Pumo (1998) and Antoniadis and Sapatinas (2003) among others for detailed discussions about the use of splines and wavelets for numerical estimation and prediction using ARH(1) model and for the numerical results presented hereafter. The second step consists in choosing tuning parameters required by these methods, for example the dimension of the projection subspace for the projection estimators. A general method used by the precedent authors is based on cross-validation approach which gives satisfactory results in applications. Note at last that

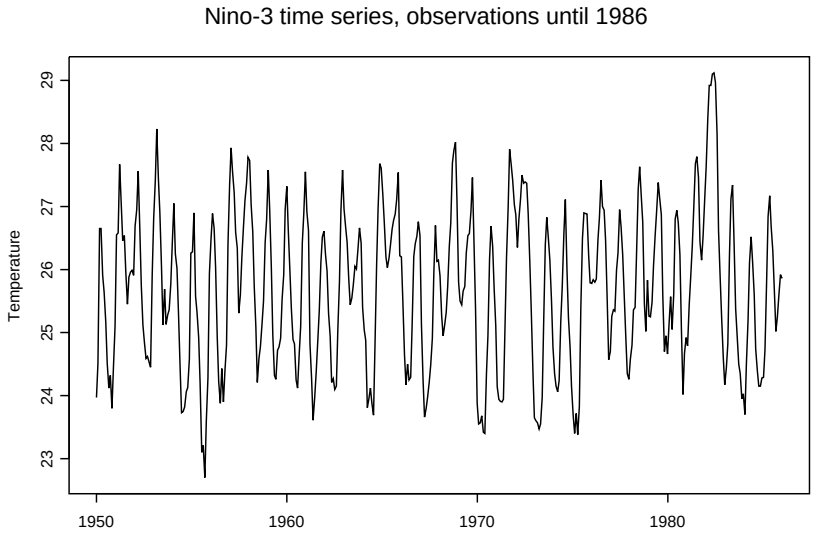

Figure 2: Monthly mean El Niño sea surface temperature index from January 1950 to December 1986

alternatives approaches of prediction based on ARH(1) modelization are proposed by Mokhtari and Mourid (2002) and Mourid and Bensmain (2005). In Mokhtari and Mourid (2002) the authors use a Parzen approximation on reproducing kernel spaces framework. Some simulation studies are presented in the recent paper published in 2008 by the same authors.

In order to compare methods described above we consider a climatological time series describing the El Niño-Southern Oscillation (see. for example Besse et al. (2000) or Smith et al. (1996) for a description of the data[1]). The series gives the monthly mean El Niño sea surface temperature index from January 1950 to December 1986 and is presented in figure 2. We compare the ARHD predictor with various functional prediction methods.

We compare the predictors of month temperature during 1986 knowing the data until 1985 by two-criteria: mean-squared error (MSE) and relative mean-absolute error (RMAE) defined by:

$$MSE = \frac{1}{12}\sum_{i=1}^{12}\left(X_n^i - \hat{X}_n^i\right)^2,\ RMAE = \frac{1}{12}\sum_{i=1}^{12}\frac{\left|X_n^i - \hat{X}_n^i\right|}{X_n^i}$$

[1]Data is freely available from http://www.cpc.ncep.noaa.gov/data/indices/index.html

Prediction method	MSE	RMAE (%)
Wavelet II	0.063	0.89
Splines	0.065	0.89
ARHD	0.167	1.25
ARH(1): linear spline	0.278	2.4
SARIMA	1.457	3.72

Table 1: Mean Squared Error (MSE) and RMAE errors for prediction of El Niño index during 1986

where X_n^i (resp. $\hat{X}_n^i$) denotes the i−th month observation (resp. prediction). The two-criteria for various functional predictors are given in Table 1. Results show that the best method are Wavelet II (one of the wavelet approaches proposed in Antoniadis and Sapatinas (2003)) and spline smoothing ARH(1). Globally the predictors obtained using ARH(1) model are better and numerically faster than the classical SARIMA $(0,1,1) \times (1,0,1)_{12}$ model (the best SARIMA model based on classical criteria).

4 Perspectives

In the precedent sections we insisted on two important statistical problems concerning H linear processes. The first discussed in §2.3 and in relation with inference on general or generalized linear processes. The estimation with the aim to make predictions with such processes seems to arise difficult technical problems. Some new results in this direction are obtained recently by Bosq (2006) by introducing the moving average process of order $q \geq 1$, $MAH(q)$. Some partial consistency results for the particular process MAH(1) are presented in a paper by Turbillon et al. (2008). A MAH(1) is a H valued process satisfying the equation $X_t = \epsilon_t + \ell(\epsilon_{t-1})$ where ℓ is a compact operator and (ϵ_t)) a strong white noise. It is simple from (3) to show that this process is invertible when the condition $\|\ell\| < 1$. The difficulty in estimating ℓ as for the real valued MA processes stems from the fact that the moment equation is not linear conversely to the ARH(1) process. Under mild conditions Turbillon et al. (2008) propose two types of estimators for ℓ and give consistency results.

The second direction concerns the ARH(1) model and his extensions. In §3.3.1 we recall some serious theoretical and technical problems with the topic of hypothesis testing. But the problem is very important in particular

from a practical point of view. As an example let us consider the ARHD model and the test addressing the significance of the derivative in the model. Another issue may be the characterization of real valued processes allowing an ARH representation or more generally linear processes. While some examples exists admitting an ARH or MAH representation (see the book by Bosq (2000)) a general approach to recognize real processes allowing such a representation is an issue for future works.

The above questions are important from a theoretical point of view, in particular for the research in statistics. For the people who analyze data that are discretized curves it's more and more necessary to dispose of analogue description tools as for the ARMA(p,q) real valued processes. In this direction a work by Hyndman and Shang (2008) for visualizing functional data and identifying functional outliers is an example.

Acknowledgement. The authors thank Frédéric Ferraty, Yves Romain and the whole group STAPH for initiating this work as well as for permanent and fruitful collaboration and are grateful to Professor Denis Bosq for helpful discussions and pointing out recent articles about functional linear processes.

References

Adams R.A., Fournier J.J.F. (2003). *Sobolev spaces*. Academic Press, 2nd ed.

Antoniadis A., Sapatinas T. (2003). Wavelet methods for continuous-time prediction using representations of autoregressive processes in Hilbert spaces. *J. of Multivariate Analysis*, **87** 133-158.

Besse, P. et Cardot, H. (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canad. J. Statist*, **24** 467-487.

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

Bosq, D. (1991). Modelization, nonparametric estimation and prediction for continuous time processes. in: *Roussas (Ed), Nato Asi Series C*, **335** 509-529.

Bosq, D. (1996). Limit theorems for banch-valued autoregressive processes. Applications to real continuous time processes. *Bull. Belg. Math. Sc*, **3** 537-555.

Bosq, D. (1999). Représentation autorégressive de l'opérateur de covariance empirique d'un ARH(1). Applications. *C.R. Acad.Sci.*, **329** Sér. I 531-534.

Bosq, D. (2000). *Linear processes in function spaces.* Lectures notes in statistics, Springer Verlag.
Bosq, D. (2007). General linear processes in Hilbert spaces and prediction. *J. Stat. Planning and Inference*, **137** 879-894.
Bosq, D., Blanke D. (2007). *Inference and prediction in large dimensions.* Wiley series in probability and statistics, John Wiley and Sons, Dunod.
Bosq, D. (2009). Tensor products of functional ARMA processes, submitted article.
Brockwell P. and Davis A. (1991). *Time series: Theory and methods.* Springer Verlag.
Cardot H, Mas A., Sarda P. (2007) CLT in functional linear regression models. Probability Theory and Related Fields, **138** 325-361.
Cavallini, A., Montanari G.C., Loggini, M., Lessi, O., Cacciari M., (1994). Nonparmetric prediction of harmonic levels in electrical networks. in: *Proceedings of IEEE ICHPS VI. Bologna*, 165-171.
Damon, J., Guillas, S. (2002). The inclusion of exogenous variables in functional autoregressive ozone forecasting. *Environmetrics*, **13** (7) 759-774.
Damon, J., Guillas, S. (2005). Estimation and Simulation of Autoregressive Hilbertian Processes with Exogenous Variables. *Stat. Infer. for Stoch. Proc.*, **8** (2) 185-204.
Dedecker, J., Merlevede, F. (2003). The conditional central limit theorem in Hilbert spaces. *Stochastic Process. Appl.*, **108** 229-262.
Dedecker, J., Merlevede, F. (2007). Convergence rates in the law of large numbers for Banach valued dependent variables. *Teor. Veroyatnost. i Primenen* **52** 562-587.
Dembo A., Zeitouni O. (1993). *Large deviations techniques and applications.* Jones and Bartlett, London.
Dunford, N., Schwartz, J.T. (1988). *Linear Operators*, Vol. I & II. Wiley Classics Library.
Fernández de Castro, B., Guillas, S., González Manteiga, W. (2005). Functional Samples and Bootstrap for Predicting Sulfur Dioxide Levels. *Technometrics* **47** (2) 212-222.
Ferraty F., Vieu P. (2006). *Nonparametric functional data analysis. Theory and Practice.* Springer-Verlag, New-York.
Geman, S., Hwang, C.R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* **10** (2) 401-414.
Glendinning, R.H., Fleet, S.L. (2007). Classifying functional time series. *Signal Processing*, **87** (1) 79-100.
Gohberg, I., Goldberg, S., Kaashoek,M.A. (1991). *Classes of linear operators Vol I & II. Operator Theory: advances and applications.* Birkhaüser

Verlag.
Grenander, U. (1981). *Abstract Inference*. Wiley, New York.
Groetsch, C. (1993). *Inverse Problems in the Mathematical Sciences*. Vieweg, Wiesbaden.
Guillas, S. (2001). Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes. *Statist. Probab. Lett.*, **55** (3) 281-291.
Guillas, S. (2002) Doubly stochastic Hilbertian processes. *J. Appl. Probab.*, **39** (3) 566-580.
Horvat L., Huskova M, Kokoszka P. (2009). Testing the stability of the functional autoregressive process. *J. of Multivariate Analysis*, **to appear**.
Hyndman R., Shang H.L. (2008). Bagplots, boxplots and outlier detection for Functional Data, in: *Functional and Operatorial Statistics*, 201-208, Physica-Verlag Heidelberg.
Kargin,V., Onatski A. (2008). Curve forecasting by functional autoregression. *J. of Multivariate Analysis*, **99** 2508-2526.
Labbas, A., Mourid, T. (2002). Estimation and prediction of a Banach valued autoregressive process. *C. R. Acad. Sci. Paris*, **335** (9) 767-772.
Laukaitis A., Rackauskas A. (2002). Functional data analysis of payment systems. *Nonlinear Analysis: Modeling and Control*, **7** 53-68.
Ledoux, M., Talagrand M. (1991). *Probability in Banach spaces - Isoperimetry and Processes*, Springer Verlag.
Maltsev V.V., Ostrovskii E.I. (1982): Central limit theorem for stationary processes in Hilbert space. *Theor. Prob. and its Applications*, **27** 357-359.
Marion J.M., Pumo B. (2004). Comparaison des modèles ARH(1) et ARHD(1) sur des données physiologiques, *Annales de l'ISUP*, **48** (3) 29-38.
Mas, A. (2002). Weak convergence for the covariance operators of a Hilbertian linear process. *Stoch Process. App*, **99** 117-135.
Mas, A. (2004). Consistance du prédicteur dans le modèle ARH(1): le cas compact. *Annales de l'Isup*, **48** 39-48.
Mas, A. (2007). Weak convergence in the functional autoregressive model. *J. of Multivariate Analysis*, **98** 1231-126.
Mas A., Menneteau L. (2003a). Perturbation appraoch applied to the asymptotic study of random operators. *Progress in Probability*, **55** 127-134.
Mas A., Menneteau L. (2003b). Large and moderate deviations for infinite-dimensional autoregressive processes. *J. of Multivariate Analysis,* **87** 241-260.
Mas, A., Pumo, B. (2007). The ARHD process. *J. of Statistical Planning and Inference*, **137** 538-553.

MENNETEAU, L (2005) Some laws of the iterated logarithm in Hilbertian autoregressive models. *J. of Multivariate Analysis*, **92** 405-425.
MERLEVÈDE, F. (1995). Sur l'inversibilité des processus linéaires à valeurs dans un espace de Hilbert. *C. R. Acad. Sci. Paris*, **321** Série I, 477-480.
MERLEVÈDE, F. (1996). Central limit theorem for linear processes with values in Hilbert space. *Stoch. Proc. and their Applications*, **65** 103-114.
MERLEVÈDE, F. (1997). Résultats de convergence presque sûre pour l'estimation et la prévision des processus linéaires hilbertiens. *C. R. Acad. Sci. Paris*, **324** Série I 573-576.
MERLEVÈDE, F., PELIGRAD M., UTEK, S. (1997). Sharp conditions for the CLT of linear processes in Hilbert space. *J. of Theoritical Probability*, **10** (3) 681-693.
MOKHTARI, F., MOURID, T. (2002) Prediction of autoregressive processes via the reproducing kernel spaces. *C. R. Acad. Sci. Paris, Ser.*, **334** 65-70.
MOURID T. (1995). Contribution à la statistique des processus autorégressifs à temps continu. PHD. Thesis, Univ. Paris VI.
MOURID, T. (2002). Estimation and prediction of functional autoregressive processes. *Statistics*, **36** (2) 125-138.
MOURID, T. (2004). The hilbertian autoregressive process with random oparator (in french). *Annales de l'ISUP*, **48** (3) 79-86.
MOURID, T., BENSMAIN, N. (2006). Sieves estimator of the operator of a functional autoregressive process. *Stat. and Probab. Letters*, **76** (1) 93-108.
NAZAROVA A.N. (2000). Normal approximation for linear stochastic processes and random fields in Hilbert space. Math. Notes, 68 363-369.
PUMO B. (1992). Estimation et prévision de processus autorégressifs fonctionnels. Applications aux processus à temps continu. *PHD Thesis*, Univ.Paris VI.
PUMO B. (1998). Prediction of continuous time processes by $C[0,1]$-valued autoregressive process. *Statist. Infer. for Stoch. Processes*, **3** (1) 297-309.
RAMSAY J.O.(2000). Differential equation models for statistical functions. *Canad. J. Statist*, **28** 225-240.
RAMSAY J.O., SILVERMAN B.W. (1997). *Functional Data Analysis*, Springer Series in Statistics.
RAMSAY J.O., SILVERMAN B.W. (2002). *Applied Functional Data Analysis*. Springer Series in Statistics, Springer-Verlag, New-York.
RUIZ-MEDINA M.D., SALMERON R., ANGULO J.M. (2007). Kalman filtering from POP-based diagonalization of ARH(1), *Comp. Statist. Data Anal.,* **51**, 4994–5008.
SMITH T.M., REYNOLDS R.W., LIVEZEY R.E., STOKES D.C. (1996).

Reconstruction of Historical Sea Surface Temperatures Using Empirical Orthogonal Functions. *Journal of Climate*, **9** (6) 1403-1420.
Tikhonov A.N., Arsenin V.Y. (1977). *Solutions of ill-posed problems.* V.H. Winstons and sons, Washington.
Turbillon, S., Bosq, D., Marion, J.M., Pumo, B. Parameter estimation of moving averages in Hilbert spaces, *C. R. Acad. Sci. Paris, Ser.*, **346** 347-350.
Wong, E. (1966). Some results concerning the zero-crossings of Gaussian noise, *SIAM J. Appl. Math.*, **14** 6 1246-1254.
Ziemer W.P. (1989). *Weakly differentiable functions. Sobolev spaces and functions of bounded variations.* Graduate Text in Mathematics 120, Springer-Verlag, New-York.