



## The ARHD model

André Mas, Besnik Pumo

### ► To cite this version:

André Mas, Besnik Pumo. The ARHD model. Journal of Statistical Planning and Inference, 2007, 137 (2), pp.538-553. 10.1016/j.jspi.2005.12.006 . hal-00004252

**HAL Id: hal-00004252**

**<https://hal.science/hal-00004252>**

Submitted on 14 Feb 2005

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The ARHD model

André Mas<sup>1,\*</sup>, Besnik Pumo<sup>2</sup>

<sup>(1)</sup>Département de Mathématiques, UMR 5149, Université Montpellier 2,

<sup>(2)</sup>Unité de Statistiques, UMR A462 SAGAH, INH Angers

## Abstract

We introduce and study a new model for functional data. The ARHD is an autoregressive model in which the first order derivative of the random curves appears explicitly. Convergent estimates are obtained through an original double penalization method. The prediction method is applied to a real set of data already studied in the literature.

*Key words* : ARHD model, Functional data, Continuous-time prediction, Wong process, ENSO, Sobolev space.

## 1 Introduction

Usually time series may be viewed as the discretized observations obtained from an underlying stochastic process  $(\xi(t), t \in \mathbb{R})$ . Models and statistical inference on such processes aim at providing the best possible predictor. Let us assume that the process  $\xi$  is observed on an interval  $[0, T]$ . We divide  $[0, T]$  into  $n$  subintervals  $[i\delta, (i+1)\delta]$ ,  $i = 0, \dots, n-1$  with  $\delta = T/n$ . This approach is clearly justified in the case when  $\xi$  is periodic with period  $\delta$  but may be generalized to processes that are stationary or not. In the following we consider the functional-valued process  $X = (X_i, i \in \mathbb{Z})$  defined by:

$$X_{i+1}(t) = \xi(i\delta + t), \quad 0 \leq t \leq \delta, \quad i \in \mathbb{Z}.$$

For a review on statistical analysis of functional data we refer to Ramsey and Silverman (1997). In this paper we consider the prediction problem of the process  $\xi$  on an entire time-interval  $[T, T+\delta]$ , or equivalently the prediction of  $X_{n+1}$  knowing  $X_1, \dots, X_n$ . To deal with the prediction problem Bosq (1991) introduced and studied an *H-valued autoregressive process of order one*, denoted ARH in the following, where  $(H, \langle \cdot, \cdot \rangle)$  is a suitable Hilbert

---

\*Corresponding author : Département de Mathématiques, CC 051, Université Montpellier 2, Place Eugène Bataillon, 34095 Montpellier Cedex 5.  
mas@math.univ-montp2.fr

space of function with inner product  $\langle \cdot, \cdot \rangle$  such that  $X_i \in H$  (typically  $H = L^2[0, \delta]$  the space of square integrable functions on  $[0, \delta]$ ). The model tends to generalize to functional data the classical and celebrated  $AR(1)$  model. Then the ARH process admits the presentation

$$X_i = \rho(X_{i-1}) + \epsilon_i$$

where  $(\epsilon_i; i \in \mathbb{Z})$  is the  $H$ -valued innovation process and  $\rho$  is a bounded compact linear operator such that  $\|\rho\|_\infty < 1$ . Note that  $\|\rho\|_\infty$  stands for the classical operator norm for  $\rho$  and is defined as

$$\|\rho\|_\infty = \sup_{h \in H, \|h\| \leq 1} \|\rho(h)\|.$$

The ARH process is stationary and the best possible prediction of  $X_{n+1}$  is  $\hat{X}_{n+1} = \rho(X_n)$  whenever  $\mathbb{E}(\epsilon_n | X_{n-1}) = 0$ . If  $\hat{\rho}_n$  is a consistent estimate for  $\rho$ , the prediction is made through  $\hat{\rho}_n(X_n)$ . Bosq (2000) proposes such a predictor and proves the consistency under mild conditions.

Considering some regularity conditions on the sample paths, one may obtain similar results for autoregressive spaces with values in other functional spaces (see Pumo (1992) for results on  $C[0, 1]$ , or Mourid (1995) for results on general Banach spaces). Alternative approaches to solve the prediction problem based on ARH modelization are proposed by Besse et al. (1996, 2000)) by means of *spline smoothing* and more recently Antoniadis and Sapatinas (2003) who implemented *Wavelet techniques*. Furthermore numerical studies show that the predictors obtained by this alternative methods are better then those obtained by *linear interpolation ARH predictor* (Pumo (1998)).

Generalizing to the functional data the classical multivariate models is a new and fruitful trend in modern statistics : linear or non linear regression, high-dimensional or functional ANOVA,  $MA(\infty)$  processes all have their "functional" counterparts. But usually and up to the authors' knowledge, these classical statistical models for functional data never involve the derivatives of the random curves rebuilt from discretized data. However the opportunity to compute explicit first (or higher) order derivatives is one of the main feature differencing truly functional from multivariate data. We refer to Silverman (1996) in the framework of principal component analysis, then Ferraty and Vieu (2003) for regression models. These authors underline the specific amount of information contained in the derivatives of curves rebuilt from functional data as well as their practical interests.

The aim of this paper is to use the functional properties of smooth sample paths in order to improve the predictor. More precisely we suppose that the sample paths belongs to the Sobolev space  $W^{2,1}$  (defined in the next section). We study an autoregressive model, initially introduced by Marion and Pumo (2004), whose definition explicitly involves the first order derivative of the data. This "ARHD" model is detailed below. Wong process, that is a stationary Gaussian process with continuously differentiable paths may be represented this way. We propose to estimate the two unknown parameters of the model by an

original method. This technique is inspired by the ridge regression method and involves two overlapping penalization through two parameters depending on each other.

The paper is organized as follows. In the following paragraph we introduce the ARHD(1) model, simply denoted ARHD in the sequel, and show that it is strictly stationary. In section 4 we give conditions for the unknown operators  $\varphi$  and  $\Psi$  to be identifiable and provide estimates as well as asymptotic results. Section 6 is devoted to technical details about the numerical calculation of ARHD predictors and to the comparison of the ARHD predictors with various functional methods in two cases: the first one is a simulated example (the Wong process which allows an ARHD presentation) ; the second is a real data study concerning the El Niño-Southern Oscillation (ENSO) time series. Proofs of asymptotic results are postponed to section 7.

## 2 The model

It was introduced by Pumo and Marion (2004). Let  $X_i$  be a sample of random curves. We introduced the model above :

$$X_{i+1} = \phi(X_i) + \Psi(X'_i) + \varepsilon_{i+1} \quad (1)$$

where  $\phi$  and  $\Psi$  are linear operators.

Now we suppose that for all  $i$ ,  $X_i$  takes its values in the Sobolev space  $W^{2,1}[0, 1]$ .

$$W^{2,1} = \{u \in L^2[0, 1], u' \in L^2[0, 1]\}.$$

The space  $W^{2,1}$  is a separable Hilbert space endowed with scalar product :

$$\langle u, v \rangle_W = \int_0^1 u(t) v(t) dt + \int_0^1 u'(t) v'(t) dt.$$

We refer to Ziemer (1989) or to Adams and Fournier (2003) for monographs dedicated to Sobolev spaces. In the sequel  $W^{2,1}$  will be denoted  $W$  and  $W^{2,0} = L^2$  will be denoted  $L$  for the sake of simplicity. Obviously if we set  $Du = u'$  then  $D$  maps  $W$  onto  $L$  ( $D$  is the ordinary differential operator). Furthermore Sobolev's imbedding theorem ensures that (see Adams and Fournier (2003) Theorem 4.12 p.85)

$$\|Du\|_L \leq C \|u\|_W$$

(where  $C$  is some constant which does not depend on  $u$ ) i.e.  $D$  is a bounded operator from  $W$  to  $L$ .

From now on we assume that  $\phi$  is a compact operator from  $W$  to  $W$  and  $\Psi$  is a compact operator from  $L$  to  $W$ . For a review on compact operators we refer to Dunford-Schwartz

(1988) or Gohberg, Goldberg, Kaashoek (1991).

### 3 ARW representation of the ARHD process.

From the above paragraph we know that  $\phi + \Psi D$  is a well defined operator on  $W$  and onto  $W$ , that  $\Psi D$  is a compact operator as the product of a bounded and a compact operator and consequently that  $\phi + \Psi D$  is itself compact as the sum of two compact operators.

We can rewrite (1) :

$$X_{i+1} = A(X_i) + \varepsilon_{i+1} \quad (2)$$

where  $A = \phi + \Psi D$ .

Finally the ARHD process may be rewritten as a special ARH(1) process with values in  $W$ .

The trouble with (2) is the following : the parameters  $\phi$  and  $\Psi$  are hidden behind  $A$  and we are not willing to infer on the latter. Obviously we are going to face two issues :

- Studying the identifiability of  $\phi$  and  $\Psi$  in the model above.
- Providing a consistent estimation procedure for  $\phi$  and  $\Psi$  before forecasting.

From now on we suppose that

$$\mathbf{H1} : \|A\|_{\infty} < 1,$$

$$\mathbf{H2} : \|X\|_W < +\infty \quad a.s.$$

The first assumption is crucial for the stationarity of the process. The second is quite restrictive but could be alleviated to mild moment assumptions but it will make the proofs of the main result more easily readable. This assumption appears for instance in Cardot, Ferraty, Sarda (1999) for the same reasons. We assess the first property of the process, which will be useful in the sequel.

**Proposition 3.1** *When assumptions **H1** and **H2** hold,  $(X_i)_{i \in \mathbb{Z}}$  and  $(X'_i)_{i \in \mathbb{Z}}$  are strictly stationary sequences on  $W$  and  $L$  respectively.*

The stationarity of  $(X_i)_{i \in \mathbb{Z}}$  is a simple consequence of the representation of equation (2) and of previous results obtained for instance by Bosq (2000), Chapter 3. The continuity of  $D$  on  $W$  implies the stationarity of the sequence  $(X'_i)_{i \in \mathbb{Z}}$ .

## 4 Estimation procedure

### 4.1 The moment method

From a practical point of view the Sobolev setting is not really a constraint. It is well known that either splines or wavelets will provide standard reconstruction method (from the discretized data) yielding functions in  $W^{2,1}$ .

The model is purely functional : we cannot invoke any likelihood-based technique since the "density of a random curve" makes non sense (Lebesgue's measure does not exist on infinite dimensional spaces). We propose to start from a classical moment method and to adapt it to our setting.

By  $\mathcal{C}_H$  (resp.  $\mathcal{C}_{HH'}$ ) we denote the space of compact operators on the Hilbert space  $H$  (resp. mapping the Hilbert space  $H$  onto  $H'$ ). Some finite rank operators are defined by means of the tensor product : if  $u$  and  $v$  belong to  $H$  and  $H'$  respectively  $u \otimes_H v$  is the operator defined on  $H$  by : for all  $h \in H$

$$(u \otimes_H v)(h) = \langle u, h \rangle_H v.$$

We start from a sample  $(X_i, X'_i)_{1 \leq i \leq n}$  and we denote

$$\begin{aligned} \Gamma &= E(X_0 \otimes_W X_0), \quad \Gamma' = E(X_0 \otimes_W X'_0), \\ \Gamma'^* &= E(X'_0 \otimes_L X_0), \quad \Gamma'' = E(X'_0 \otimes_L X'_0), \\ \Delta &= E(X_0 \otimes_W X_1), \quad \Delta' = E(X'_0 \otimes_L X_1). \end{aligned}$$

Under assumption **H2** all these operators belong either to  $\mathcal{C}_W$ ,  $\mathcal{C}_{WL}$ ,  $\mathcal{C}_{LW}$  or  $\mathcal{C}_L$ . In fact assumption **H2** could be replaced by  $E\|X\|_W^2 < +\infty$ .

By  $\Gamma_n, \Gamma'_n, \dots, \Delta'_n$  we denote the empirical counterpart of these operators based on the sample  $(X_i, X'_i)_{1 \leq i \leq n}$ . For example :

$$\begin{aligned} \Gamma_n &= \frac{1}{n} \sum_{k=1}^n X_k \otimes_W X_k, \\ \Delta'_n &= \frac{1}{n-1} \sum_{k=1}^{n-1} X'_k \otimes_L X_{k+1}. \end{aligned} \tag{3}$$

**Remark 4.1** *The notation  $\Gamma'^*$  is not ambiguous :  $\Gamma'^*$  is truly the adjoint operator of  $\Gamma'$ .*

**Remark 4.2** *Conversely, if the random function  $X'$  is truly the derivative of  $X$ , this is no more the case as far as linear operators are concerned :  $\Gamma'$  is not the derivative of  $\Gamma$ . The  $'$  is just a notation in this setting ; this would make no sense anyway. However, it should be remarked that for all  $u$  in  $L$*

$$(\Gamma'^*(u))' = (E(\langle u, X' \rangle_L X))' = \Gamma''(u)$$

and for all  $v$  in  $W$

$$(\Gamma(v))' = (E(\langle v, X \rangle_W X))' = \Gamma'(v).$$

Quite naturally, from (1) -multiply with  $\langle X_i, \cdot \rangle$  and  $\langle X'_i, \cdot \rangle$  successively then take expectation- we easily deduce both moment equations :

$$\mathcal{S} = \begin{cases} \Delta = \phi\Gamma + \Psi\Gamma' \\ \Delta' = \phi\Gamma'^* + \Psi\Gamma'' \end{cases} \quad (4)$$

Resolving this system is apparently easy but we should be aware of two facts :

- Operators (here,  $\Delta, \phi, \Gamma, \dots$ ) do not commute !
- The inverse operators of  $\Gamma$ , and  $\Gamma''$  do not necessarily exist and when they do, they are unbounded, i.e. not continuous (remind that  $\Gamma$ , and  $\Gamma''$  are compact operators and that compact operators have no bounded inverses).

At this point, before trying to solve (4) we need to study identifiability of the unknown infinite dimensional parameter  $(\phi, \Psi) \in \mathcal{C}_W \times \mathcal{C}_{LW}$  in our statistical problem.

## 4.2 Identifiability

We set  $\mathcal{E} = \mathcal{C}_W \times \mathcal{C}_{LW}$ . If both equations in (4) are the starting point we should make sure that solutions to these equations are well and uniquely defined. Suppose for instance that  $\text{Ker}\Gamma \neq \{0\}$  and take  $h$  in it. Now set  $\tilde{\phi} = \phi + h \otimes_W h$ . Then

$$\tilde{\phi}\Gamma = \phi\Gamma + (h \otimes_W h)\Gamma$$

but  $(h \otimes_W h)\Gamma = 0$ . So  $\tilde{\phi}\Gamma = \phi\Gamma$  and  $\phi$  is not unique (there are even infinitely many solutions in the space  $\phi + \text{Ker}\Gamma$ ). The next assumption is

$$\mathbf{H3} : \text{Ker}\Gamma = \text{Ker}\Gamma'' = \{0\}.$$

In other words we suppose that both operators above are one to one.

Now turning back to (4) we rewrite the system. Equivalently :

$$\mathcal{S} = \begin{cases} \Delta = \phi\Gamma + \Psi D\Gamma \\ \Delta' = \phi\Gamma D^* + \Psi D\Gamma D^* \end{cases} \quad (5)$$

$$\iff \begin{cases} (\Delta, \Delta') = (\phi, \Psi) \Lambda \\ \Lambda = \begin{pmatrix} \Gamma & \Gamma D^* \\ D\Gamma & D\Gamma D^* \end{pmatrix} \end{cases} \quad (6)$$

We are now ready to solve the identification problem.

**Proposition 4.1** *The couple  $(\phi, \Psi) \in \mathcal{E}$  is identifiable for the moment method proposed in (4) if and only if  $(\phi, \Psi) \notin \mathcal{N}$  where  $\mathcal{N}$  is the vector subspace of  $\mathcal{E}$  defined by*

$$\mathcal{N} = \{(U, V) \in \mathcal{E} : U + VD = 0\}.$$

*Note that  $\mathcal{N}$  is a closed set in  $\mathcal{E}$ .*

In other words if, for all  $u$  in  $W$ ,  $\phi u + \Psi u' = 0$ , the parameter  $(\phi, \Psi)$  cannot be identified.

The Proposition is proved at the beginning of the last section of the paper.

## 5 Definition of the estimates and convergence

Since the unknown parameters  $\phi$  and  $\Psi$  are operators estimating them means dealing with random operators based on the double sample  $(X_i, X'_i)_{1 \leq i \leq n}$ . We refer for instance to (3) above for examples of these available operators.

The estimates stem from (4) which is a highly non invertible system. We are classically going to add a small perturbation to regularize it and make it invertible. We solve :

$$\mathcal{S}' = \begin{cases} \Delta = \phi(\Gamma + \alpha I_W) + \Psi \Gamma' \\ \Delta' = \phi \Gamma'^* + \Psi(\Gamma'' + \alpha I_L) \end{cases}$$

where  $\alpha$  is a positive real number and  $I_W$  denotes the identity operator on  $W$ . Now the operators  $(\Gamma + \alpha I_W)$  and  $(\Gamma'' + \alpha I_L)$  are no more compact but have bounded inverses. Basic algebra gives :

$$\mathcal{S}' = \begin{cases} \phi[(\Gamma + \alpha I_W) - \Gamma'^*(\Gamma'' + \alpha I_L)^{-1} \Gamma'] = \Delta - \Delta'(\Gamma'' + \alpha I_L)^{-1} \Gamma' \\ \Psi[(\Gamma'' + \alpha I_L) - \Gamma'(\Gamma + \alpha I_W)^{-1} \Gamma'^*] = \Delta' - \Delta(\Gamma + \alpha I_W)^{-1} \Gamma'^*. \end{cases} \quad (7)$$

Which is then once more approximated by :

$$\mathcal{S}'' = \begin{cases} \phi[\Gamma - \Gamma'^*(\Gamma'' + \alpha I_L)^{-1} \Gamma'] = \Delta - \Delta'(\Gamma'' + \alpha I_L)^{-1} \Gamma' \\ \Psi[\Gamma'' - \Gamma'(\Gamma + \alpha I_W)^{-1} \Gamma'^*] = \Delta' - \Delta(\Gamma + \alpha I_W)^{-1} \Gamma'^*. \end{cases}$$

We just dropped  $\alpha I_W$  on the first line and  $\alpha I_L$  on the second to get  $\mathcal{S}''$ . Take the first line in the above display. The operator

$$S_\phi = \Gamma - \Gamma'^*(\Gamma'' + \alpha I_L)^{-1} \Gamma' \quad (8)$$

is a selfadjoint compact operator. (Indeed  $\Gamma'^*(\Gamma'' + \alpha I_L)^{-1} \Gamma'$  is a compact operator because  $\Gamma'$  and  $\Gamma'^*$  are). We may deduce from this fact that  $S_\phi$  has real eigenvalues (not necessarily positive) and furthermore that once again  $S_\phi$  has no bounded inverse. The



same remarks hold for

$$S_\Psi = \Gamma'' - \Gamma' (\Gamma + \alpha I_W)^{-1} \Gamma'^*. \quad (9)$$

However we can provide an approximate solution to  $\mathcal{S}''$  by regularizing  $S_\phi$  and  $S_\Psi$  once more by a penalization method. Finally the pseudo solutions we propose to solve  $\mathcal{S}''$  hence  $\mathcal{S}$  are based on a second strictly positive parameter  $\beta$  and are denoted  $\tilde{\phi}$  and  $\tilde{\Psi}$  :

$$\begin{cases} \tilde{\phi} = [\Delta - \Delta' (\Gamma'' + \alpha I_L)^{-1} \Gamma'] (S_\phi + \beta I)^{-1} \\ \tilde{\Psi} = [\Delta' - \Delta (\Gamma + \alpha I_W)^{-1} \Gamma'^*] (S_\phi + \beta I)^{-1}. \end{cases} \quad (10)$$

This new system defines relations from which we propose to deduce estimates. From now on -in order to alleviate the notations- by  $S^\dagger$  we denote the operator defined by  $(S + \alpha_n I_L)^{-1}$  where  $\alpha_n$  is a non increasing sequence of positive numbers decaying to zero. We set :

$$S_{n,\phi} = \Gamma_n - \Gamma_n'^* (\Gamma_n''^\dagger) \Gamma_n', \quad (11)$$

$$S_{n,\Psi} = \Gamma_n'' - \Gamma_n' (\Gamma_n^\dagger) \Gamma_n'^*, \quad (12)$$

$$T_{n,\phi} = \Delta_n - \Delta_n' (\Gamma_n''^\dagger) \Gamma_n', \quad (13)$$

$$T_{n,\Psi} = \Delta_n' - \Delta_n (\Gamma_n^\dagger) \Gamma_n'^*. \quad (14)$$

Taking  $\beta_n \downarrow 0$  we obtain the following

**Definition 5.1** *The estimate of the couple  $(\phi, \Psi)$  is  $(\phi_n, \Psi_n)$  based on (10) and defined by :*

$$\begin{cases} \phi_n = T_{n,\phi} (S_{n,\phi} + \beta_n I)^{-1} \\ \Psi_n = T_{n,\Psi} (S_{n,\phi} + \beta_n I)^{-1}. \end{cases} \quad (15)$$

The next Theorem is the main theoretical result of this article. It provides the convergence of our estimates when the sample size goes to infinity.

**Theorem 5.1** *When **H1** – **3** hold and if  $\alpha_n \rightarrow 0$ ,  $\beta_n \rightarrow 0$  with  $\sqrt{n}\alpha_n^2\beta_n^2 \rightarrow +\infty$  and  $\sqrt{\alpha_n}/\beta_n \rightarrow 0$ ,*

$$\begin{aligned} \phi_n &\xrightarrow{\mathbb{P}} \phi, \\ \Psi_n &\xrightarrow{\mathbb{P}} \Psi. \end{aligned}$$

The convergence is understood in the  $\|\cdot\|_\infty$  norm for bounded operators.

Note that Theorem 5.1 holds whenever  $\alpha_n = n^{-a}$  and  $\beta_n = n^{-b}$  with  $b < a/2$  and  $2b + 2a < 1/2$ .

**Remark 5.1** *Originally the linear model (1) is subject to serious multicollinearity troubles since  $X'_n = DX_n$ . Even if the curve  $X'_n$  usually looks quite different from  $X_n$ , there*

is a total stochastic dependence between them. The method used in this article to tackle this problem (as well as the intrinsic "inverse problem" aspects related to the inversion of the covariance operators  $\Gamma$  and  $\Gamma''$ ) is new up to the authors' knowledge. As it can be seen through above at display (10) or in the proofs below, it relies on a double penalization technique first by the index  $\alpha_n$  then by  $\beta_n$  linking both indexes in order to asymptotically suppress the bias terms.

## 6 A numerical study and application: ENSO

In this section we illustrate the ARHD method of prediction proposed in this paper by some numerical studies for two examples. We give first some technical results to carry out numerical calculations. The first application is connected to Wong's process (see Wong (1966)) which admits an ARHD presentation. We compare the ARHD predictor with various predictors based on the notion of ARH process that is, *linear interpolation ARH predictor* (Pumo (1998)), *Fourier interpolation ARF predictor* and *ARW predictor* based on the presentation (2), by two statistical criteria: *mean-squared error* (MSE) and *relative mean-absolute error* (RMAE) defined by :

$$MSE = \frac{1}{m} \sum_{j=1}^m (X_n(t_j) - \hat{X}_n(t_j))^2, \quad RMAE = \frac{1}{m} \sum_{j=1}^m \frac{|X_n(t_j) - \hat{X}_n(t_j)|}{|X_n(t_j)|}$$

where  $m$  is the number of discretized points.

The second example concerns real data, namely climatological time series describing the El Niño-Southern Oscillation (ENSO). We compare our predictor with predictor based on similar approaches found in the literature : *spline smoothing FAR predictor* (Besse and Cardot (1996)), *Local FAR predictor* (Besse et al. (2000)) and *wavelet based predictor* (Antoniadis and Sapatinas (2003)).

### 6.1 Some technical details about simulations

Consider the Fourier basis on  $L^2[0, \delta]$  and denote  $e_0(t) = 1/\delta$  and  $e_{2j-1}(t) = \cos(2j\pi t/\delta)$ ,  $e_{2j}(t) = \sin(2j\pi t/\delta)$  for  $j \geq 1$ . Then a simple calculation shows that

$$\mathbf{w} = \{e_0, [1 + 4j^2\pi^2/\delta^2]^{-1/2} \cdot e_{2j-1}, [1 + 4j^2\pi^2/\delta^2]^{-1/2} \cdot e_{2j}, j \geq 1\}$$

is an orthonormal basis for  $W$ . Let  $f = \sum_{i=0,\infty} c_j e_j$ , where  $c_j = \langle f, e_j \rangle_{L^2}$ , be the Fourier series of a continuously differentiable function  $f$ . Then  $f' = \sum_{i=0,\infty} c_j e'_j$ . Furthermore the decomposition  $\sum_{j=0,\infty} \langle f, w_j \rangle_W w_j$  on  $W$  of  $f$  is given by  $c_0 + \sum_{j=1,\infty} \langle f, w_j \rangle_W w_j$  where  $\langle f, w_{2j-1} \rangle_W = [1 + 4j^2\pi^2/\delta^2]^{-1/2} \cdot \langle f, e_{2j-1} \rangle_{L^2}$  and  $\langle f, w_{2j} \rangle_W = [1 + 4j^2\pi^2/\delta^2]^{-1/2} \cdot \langle f, e_{2j} \rangle_{L^2}$ .

In order to calculate the covariance operators given in section 4 denote  $\mathbf{w}_N$  (resp.  $\mathbf{e}_N$ )  $N$ -vectors of the basis on  $W$  (resp.  $L^2$ ) that is  $\mathbf{w}_N = {}^t(w_0, w_1 \dots w_{N-1})$  (resp.  $\mathbf{e}_N = {}^t(e_0, e_1, \dots, e_{N-1})$ ) and suppose that  $N$  is an odd and positive number. Denote  $\mathbf{X}$  and  $\mathbf{X}'$  the  $N \times n$  matrices :

$$\begin{aligned}\mathbf{X} &= (\langle X_i, w_{k-1} \rangle_W, k = 1, \dots, N; i = 1, \dots, n, \\ \mathbf{X}' &= (\langle X'_i, e_{k-1} \rangle_{L^2}, k = 1, \dots, N; i = 1, \dots, n.\end{aligned}$$

As noted above the coefficients  $\langle X_i, w_{k-1} \rangle_W$  and  $\langle X'_i, e_{k-1} \rangle_{L^2}$  are obtained directly from the Fourier decomposition of  $X_i$ , for  $i = 1, \dots, n$ .

It follows that covariance operators  $\Gamma_n, \Gamma'_n, \Gamma_n^*, \Gamma_n'', \Delta_n, \Delta'_n$  can be approximated by :

$$\begin{aligned}C_n &= (1/n) \cdot \mathbf{X} ({}^t\mathbf{X}), C'_n = (1/n) \cdot \mathbf{X}' ({}^t\mathbf{X}), \\ C_n^{*'} &= (1/n) \cdot \mathbf{X} ({}^t\mathbf{X}'), C_n^{''} = (1/n) \cdot \mathbf{X}' ({}^t\mathbf{X}'), \\ D_n &= (1/[n-1]) \cdot \mathbf{X}_{-1} ({}^t\mathbf{X}_{-n}), D'_n = (1/[n-1]) \cdot \mathbf{X}_{-1} ({}^t\mathbf{X}'_{-n})\end{aligned}$$

where  $\mathbf{X}'_{-n}$  (resp.  $\mathbf{X}_{-1}$ ) is the matrix  $\mathbf{X}'$  (resp.  $\mathbf{X}$ ) without the column  $n$  (resp. 1). So in order to obtain the estimators given in the second section it suffices to substitute the covariance operators in (11-14) and (15) by their approximations given above and choosing suitable values for  $\alpha_n$  and  $\beta_n$ .

## 6.2 Wong process

This process is defined for  $u \in R$  by:

$$\xi_u = \sqrt{3} \exp(-\sqrt{3}u) \int_0^{\exp(2u/\sqrt{3})} W_s ds.$$

Wong process is a mean-square differentiable stationary Gaussian process which is zero-mean and with variance 1. Let  $\delta > 0$  and  $X_i \in W$  given by  $X_{i+1}(t) = \xi_{i \cdot \delta + t}$  for  $t \in ]0, \delta]$ . Let  $\epsilon_{i+1}$  be a squared differentiable r.v. with values in  $W^{2,1}[0, \delta]$  :

$$\epsilon_{i+1}(t) = \sqrt{3} \exp(-\sqrt{3}(i+t)) \int_{\exp(2i/\sqrt{3})}^{\exp(2(i+t)/\sqrt{3})} (W_s - W_{\exp(2i/\sqrt{3})}) ds. \quad (16)$$

Then the process  $(X_i, i \in Z)$  can be written as:

$$X_{i+1} = [\phi + \Psi(D)]X_i + \epsilon_{i+1} \quad (17)$$

where  $c(t) = \frac{\sqrt{3}}{2} \cdot \exp(-\sqrt{3}t) \cdot \{\exp(2t/\sqrt{3}) - 1\}$  and:

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

Furthermore  $\epsilon_{i+1}$  is independent of  $X_i, X'_i$  and a direct calculation shows that assumption **H1** is satisfied.

Using a method presented in Blanke and Pumo (2003) we simulated a Wong process  $\xi(t)$  on  $[0, 192.65]$ , that corresponds to  $n = 105$  intervals of length  $\delta = 1.8348$ , each known at  $m = 50$  equidistant points  $t_j, j = 1, \dots, 50$ . A simulated process is presented in Figure 1. The associated process  $(X_i)$  is a  $W^{2,1}[0, 1.8348]$  valued process.

**Please insert here Figure 1**

Fifty Wong processes were simulated and for each of them we calculated the MSE and RMAE criteria. The mean values for the two criteria for the 50 simulations and various predictors are presented in Table 1. Figure 2 presents the different predictions for one of the simulations.

For the calculation of ARHD predictors we consider two values for the parameter  $\alpha_n$ , that are 0.1 and 0.3. The corresponding values for  $\beta_n$  are 0.65 and 0.5. In the calculation of ARH, ARF and ARW predictor we consider  $k_n = 1$ , that is the projection subspace for observation is equal to one (see Pumo (1998) for details). Simulations show that the ARW and ARF predictor are very similar and when  $m$  (this is the case for example when  $m = 50$ ) is large they give similar results to linear interpolation ARH predictor. But the three predictors are less better than the ARHD predictor. Notice also that the choice of the optimal values for the parameters  $\alpha_n$  and  $\beta_n$  may be done by a cross-validation procedure.

**Please insert here Table 1**

**Please insert here Figure 2**

### 6.3 Example SST: Sea Surface Temperature

The second example concerns 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<sup>1</sup>). The series gives the monthly mean El Niño sea surface temperature index from January 1950 to December 1969, that is  $m = 12$ , and is presented in figure 3. We compare the ARHD predictor with various functional prediction methods.

**Please insert here Figure 3**

In the first numerical study we compare the prediction of the temperature during 1986 knowing the data until 1985. We calculated the ARHD predictor with  $\alpha_n = 0.4$  and 0.1 and  $\beta_n = 0.8$  and 0.4. The MSE and RMAE criteria for various functional predictors are given in Table 2. Results show that the best method are Wavelet II (one of the

---

<sup>1</sup>Data is freely available from <http://www.cpc.ncep.noaa.gov/data/indices/index.html>

wavelet approaches proposed in Antoniadis and Sapatinas) and spline smoothing FAR. Nevertheless our predictor is better than the other predictors or the classical SARIMA  $(0, 1, 1) \times (1, 0, 1)_{12}$  model (see for example Brockwell and Davis (1987)). Figure 3 displays the observed data during 1986 and its predictors by some of the predictors discussed above. Notice that the ARF or ARW predictors are not satisfactory as  $m = 12$ .

In the second numerical study we make 10 one year ahead forecasts for the period 1986-97. The statistical criteria for various functional methods are presented in Table 3. The reader may notice that the ARHD method gives a similar prediction as the Local FAR method which is the best functional prediction method appearing in Besse et al. (2000). Note finally that as described in the introduction of this section the computational effort to obtain an ARH predictor is comparable to that of the calculation of an ARH predictor.

**Please insert here Table 2**

**Please insert here Figure 4**

**Please insert here Table 3**

## 7 Proofs

### Proof of Proposition 4.1:

The couple  $(\phi, \Psi)$  will be identified whenever, for any other couple  $(\tilde{\phi}, \tilde{\Psi})$ , if  $(\tilde{\phi}, \tilde{\Psi}) \Lambda = (\phi, \Psi) \Lambda$ ,  $(\tilde{\phi}, \tilde{\Psi}) = (\phi, \Psi)$ . This will be true if

$$\{(U, V) \in \mathcal{E} : (U, V) \Lambda = 0\} = \{0\}.$$

But  $\Lambda$  may be decomposed as the product of three operators, namely :

$$\Lambda = \begin{pmatrix} I \\ D \end{pmatrix} \Gamma(I, D^*).$$

As  $\Gamma$  is one to one by assumption **H3**, and since

$$(T, TD^*) = 0 \text{ iff } T = 0,$$

it is readily seen that

$$(U, V) \Lambda = 0 \text{ iff } U + VD = 0$$

which finishes the proof of the Proposition.

We begin with five Lemmas needed to prove Theorem 5.1:

**Lemma 7.1**

$$\begin{aligned}\|\Gamma_n - \Gamma\|_\infty &= O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right), \|\Gamma'_n - \Gamma'\|_\infty = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right), \\ \|\Gamma_n^* - \Gamma'^*\|_\infty &= O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right), \|\Gamma''_n - \Gamma''\|_\infty = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right).\end{aligned}$$

**Proof :** Since  $X_n$  is an ARH(1) process (with autocorrelation operator  $A$ ), we can invoke for instance Theorem 4.1 p.98 in Bosq (2000) to get the first rate of decay. All the other results above are due to the boundedness (in our framework) of the differential operator  $D$ . Indeed for instance  $\Gamma''_n = D\Gamma_n D^*$ .

**Lemma 7.2**

$$\begin{aligned}\|\Gamma''^\dagger\|_\infty &= \|\Gamma_n''^\dagger\|_\infty = \frac{1}{\alpha}, \\ \|\Gamma''^\dagger - \Gamma_n''^\dagger\|_\infty &= O_{\mathbb{P}}\left(\frac{1}{\alpha^2\sqrt{n}}\right).\end{aligned}$$

**Proof :** We prove the first bound

$$\|\Gamma''^\dagger\|_\infty = \left\|(\Gamma'' + \alpha I)^{-1}\right\|_\infty$$

and as  $\Gamma''$  is a positive compact operator, the norm of operator  $(\Gamma'' + \alpha I)^{-1}$ , which is known as the resolvent operator of  $\Gamma''$ , is non random and evaluated at  $\alpha^{-1}$ . The same is true with  $\Gamma''_n$  replacing  $\Gamma''$ .

Using  $B^{-1} - A^{-1} = A^{-1}(A - B)B^{-1}$  for two invertible matrices  $A$  and  $B$ , we get :

$$\Gamma''^\dagger - \Gamma_n''^\dagger = (\Gamma''_n + \alpha I)^{-1} (\Gamma''_n - \Gamma'') (\Gamma'' + \alpha I)^{-1}$$

which entails

$$\begin{aligned}\|\Gamma''^\dagger - \Gamma_n''^\dagger\|_\infty &\leq \left\|(\Gamma''_n + \alpha I)^{-1}\right\|_\infty \|\Gamma''_n - \Gamma''\|_\infty \left\|(\Gamma'' + \alpha I)^{-1}\right\|_\infty \\ &= \frac{1}{\alpha^2} \|\Gamma''_n - \Gamma''\|_\infty = O_{\mathbb{P}}\left(\frac{1}{\alpha^2\sqrt{n}}\right)\end{aligned}$$

by Lemma 7.1.

**Lemma 7.3** *Let  $S_{n,\phi}$  and  $S_\phi$  defined respectively by (11) and (8). Then*

$$\|S_{n,\phi} - S_\phi\|_\infty = O_{\mathbb{P}}\left(\frac{1}{\alpha^2\sqrt{n}}\right).$$

**Proof :** From equations (8) and (9) we obtain  $S_{n,\phi} - S_\phi = \Gamma_n - \Gamma + \Gamma'^* (\Gamma''^\dagger) \Gamma' - \Gamma_n'^* (\Gamma_n''^\dagger) \Gamma_n'$  and

$$\|S_{n,\phi} - S_\phi\|_\infty \leq \|\Gamma_n - \Gamma\|_\infty + \|\Gamma'^* (\Gamma''^\dagger) \Gamma' - \Gamma_n'^* (\Gamma_n''^\dagger) \Gamma_n'\|_\infty.$$

We look for a bound for

$$\begin{aligned} \|\Gamma'^* (\Gamma''^\dagger) \Gamma' - \Gamma_n'^* (\Gamma_n''^\dagger) \Gamma_n'\|_\infty &\leq \|\Gamma'^* (\Gamma''^\dagger) \Gamma' - \Gamma_n'^* (\Gamma''^\dagger) \Gamma'\|_\infty \\ &\quad + \|\Gamma_n'^* (\Gamma''^\dagger) \Gamma' - \Gamma_n'^* (\Gamma_n''^\dagger) \Gamma_n'\|_\infty + \|\Gamma_n'^* (\Gamma_n''^\dagger) \Gamma_n' - \Gamma_n'^* (\Gamma''^\dagger) \Gamma_n'\|_\infty \end{aligned}$$

Obviously the two first terms above may be bounded in probability by

$$\|\Gamma''^\dagger\|_\infty \|\Gamma_n'\|_\infty \|\Gamma' - \Gamma_n'\|_\infty = O_{\mathbb{P}}\left(\frac{1}{\alpha\sqrt{n}}\right)$$

since  $\|\Gamma_n'\|_\infty = \|\Gamma_n'^*\|_\infty$ . The remaining term may be bounded by

$$\|\Gamma_n'\|^2 \|\Gamma''^\dagger - \Gamma_n''^\dagger\|_\infty = O_{\mathbb{P}}(\|\Gamma''^\dagger - \Gamma_n''^\dagger\|_\infty)$$

and Lemma 7.2 finishes the proof.

**Lemma 7.4** *The operator  $S_\phi$  is positive hence*

$$\|(S_\phi + \beta I)^{-1}\|_\infty \leq \frac{1}{\beta}.$$

**Proof :** Before starting the proof it is worth reminding the following fact to the reader. It  $T$  is a compact operator from a Hilbert space  $H_1$  to a Hilbert space  $H_2$  it admits a Schmidt representation

$$T = \sum_{i=1}^{+\infty} s_i (u_i \otimes v_i)$$

where the  $s_i$ 's are the positive eigenvalues of  $T^*T$  (i.e. of  $TT^*$ ) and where  $u_i$  (resp.  $v_i$ ) denote a complete orthonormal system of  $H_1$  (resp.  $H_2$ ). We refer for instance to Theorem 1.1 page 96 in Gohberg, Goldberg, Kaashoek (1991). Now we turn to

$$S_\phi = \Gamma - \Gamma'^* \Gamma''^\dagger \Gamma'$$

We set  $U = D\Gamma^{1/2}$  (remind that  $\Gamma' = D\Gamma$ ). The operator  $U$  is a compact from  $W$  to  $L$  since  $D$  is bounded and  $\Gamma^{1/2}$  is compact like  $\Gamma$ . Then  $\Gamma'' = UU^*$  and  $\Gamma''^\dagger = (UU^* + \alpha I)^{-1}$  and we rewrite

$$S_\phi = \Gamma^{1/2} (I - U^* (UU^* + \alpha I)^{-1} U) \Gamma^{1/2} \quad (18)$$

Now let us write the Schmidt decomposition of  $U$

$$U = \sum_{i=1}^{+\infty} s_i (u_i \otimes_W v_i) \quad u_i \in W, v_i \in L$$

Easy computations lead to

$$(I - U^* (UU^* + \alpha I)^{-1} U) = \sum_{i=1}^{+\infty} \frac{\alpha}{\alpha + s_i^2} u_i \otimes_W u_i$$

From (18) we deduce that for all  $x$  in  $W$   $\langle S_\phi x, x \rangle \geq 0$  hence the announced result.

**Lemma 7.5**

$$\|(S_{n,\phi} + \beta I)^{-1} - (S_\phi + \beta I)^{-1}\|_\infty = O_{\mathbb{P}} \left( \frac{1}{\alpha^2 \beta^2 \sqrt{n}} \right).$$

**Proof :**

$$(S_{n,\phi} + \beta I)^{-1} - (S_\phi + \beta I)^{-1} = (S_\phi + \beta I)^{-1} (S_\phi - S_{n,\phi}) (S_{n,\phi} + \beta I)^{-1} \quad (19)$$

hence

$$(I - (S_\phi + \beta I)^{-1} (S_\phi - S_{n,\phi})) (S_{n,\phi} + \beta I)^{-1} = (S_\phi + \beta I)^{-1}. \quad (20)$$

Since

$$\|(S_\phi + \beta I)^{-1} (S_\phi - S_{n,\phi})\|_\infty \xrightarrow{\mathbb{P}} 0$$

the probability that  $I - (S_\phi + \beta I)^{-1} (S_\phi - S_{n,\phi})$  is an invertible operator tends to 1. It suffices indeed that

$$\|(S_\phi + \beta I)^{-1} (S_\phi - S_{n,\phi})\|_\infty < 1$$

to write from (20) :

$$(S_{n,\phi} + \beta I)^{-1} = (I - (S_\phi + \beta I)^{-1} (S_\phi - S_{n,\phi}))^{-1} (S_\phi + \beta I)^{-1}. \quad (21)$$

We set  $H_n = (S_\phi + \beta I)^{-1} (S_\phi - S_{n,\phi})$  then

$$(I - H_n)^{-1} = I + \sum_{p=1}^{+\infty} H_n^p.$$

At last from (21) and (19) :

$$(S_{n,\phi} + \beta I)^{-1} - (S_\phi + \beta I)^{-1} = \sum_{p=1}^{+\infty} H_n^p (S_\phi + \beta I)^{-1}$$



and

$$\begin{aligned} \|(S_{n,\phi} + \beta I)^{-1} - (S_\phi + \beta I)^{-1}\|_\infty &\leq \|(S_\phi + \beta I)^{-1}\|_\infty \sum_{p=1}^{+\infty} \|H_n\|_\infty \\ &= O_{\mathbb{P}}\left(\frac{1}{\alpha^2 \beta^2 \sqrt{n}}\right) \end{aligned}$$

by Lemmas 7.3 and 7.4.

**Proof of Theorem 5.1 :** We prove the Proposition for  $\phi_n$  since the same technique would lead to an analogous result for  $\Psi$ . Developing the expression of  $\Delta_n$  and  $\Delta'_n$  yields

$$\begin{aligned} \Delta_n &= \phi \Gamma_n + \Psi \Gamma'_n + U_n, \\ \Delta'_n &= \phi \Gamma_n^* + \Psi \Gamma''_n + U'_n, \end{aligned}$$

with

$$U_n = \frac{1}{n} \sum_{k=1}^n X_k \otimes \varepsilon_{k+1}, \quad U'_n = \frac{1}{n} \sum_{k=1}^n X'_k \otimes \varepsilon_{k+1}.$$

Hence

$$\begin{aligned} T_{n,\phi} &= \Delta_n - \Delta'_n (\Gamma_n^{\prime\prime\prime}) \Gamma'_n \\ &= \phi \Gamma_n + \Psi \Gamma'_n - \phi \Gamma_n^* (\Gamma_n^{\prime\prime\prime}) \Gamma'_n - \Psi \Gamma''_n (\Gamma_n^{\prime\prime\prime}) \Gamma'_n + U_n - U'_n (\Gamma_n^{\prime\prime\prime}) \Gamma'_n \\ &= \phi S_{n,\phi} + \Psi [\Gamma'_n - \Gamma''_n (\Gamma_n^{\prime\prime\prime}) \Gamma'_n] + [U_n - U'_n (\Gamma_n^{\prime\prime\prime}) \Gamma'_n] \end{aligned}$$

At last

$$\begin{aligned} \phi_n - \phi &= \beta \phi (S_{n,\phi} + \beta I)^{-1} + \Psi [\Gamma'_n - \Gamma''_n (\Gamma_n^{\prime\prime\prime}) \Gamma'_n] (S_{n,\phi} + \beta I)^{-1} \\ &\quad + [U_n - U'_n (\Gamma_n^{\prime\prime\prime}) \Gamma'_n] (S_{n,\phi} + \beta I)^{-1}. \end{aligned} \tag{22}$$

The proof will be achieved if we prove that the three terms in the display above tend to zero in probability. The three next Propositions namely Propositions 7.1, 7.2 and 7.3 are devoted to this goal. We begin with the last one involving  $U_n$  and  $U'_n$ . We need two auxiliary Lemmas.

**Lemma 7.6**

$$\|U_n\| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right), \tag{23}$$

$$\|U'_n\| = O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right), \tag{24}$$

$$\|U'_n (\Gamma_n^{\prime\prime\prime}) \Gamma'_n\|_\infty = O_{\mathbb{P}}\left(\frac{1}{\alpha \sqrt{n}}\right). \tag{25}$$

**Proof :** The proof of (23) and (24) is obvious since  $U_n$  and  $U'_n$  are sums of uncorrelated

random operators (here uncorrelated means that the cross covariance operator between two distinct random elements is the null operator). Then

$$\|U'_n (\Gamma''_n)^\dagger \Gamma'_n\|_\infty \leq \|U'_n\|_\infty \|\Gamma''_n\|_\infty \|\Gamma'_n\|_\infty$$

where the last term on the right side is bounded in probability, the first is an  $O_{\mathbb{P}}(n^{-1/2})$  and the norm of the second is almost surely bounded by  $\alpha^{-1}$  and (25) is proved.

**Proposition 7.1**

$$[U_n - U'_n (\Gamma''_n)^\dagger \Gamma'_n] (S_{n,\phi} + \beta I)^{-1} = O_{\mathbb{P}}\left(\frac{1}{\alpha\beta\sqrt{n}}\right)$$

**Proof :** The proof of the Proposition is a consequence of Lemmas 7.4 and 7.6.

We turn to the first term in (22).

**Proposition 7.2** *If  $\beta \rightarrow 0$  and  $\frac{1}{\alpha^2\beta^2\sqrt{n}} \rightarrow 0$ ,*

$$\beta\phi (S_{n,\phi} + \beta I)^{-1} \xrightarrow{\mathbb{P}} 0$$

**Proof :** We invoke Lemma 7.5 to claim that it suffices to drop the index  $n$  in the Proposition and to prove that :

$$\beta\phi (S_\phi + \beta I)^{-1} \xrightarrow{\mathbb{P}} 0.$$

In fact Lemma 7.5 links the asymptotic behavior of  $(S_{n,\phi} + \beta I)^{-1}$  and  $(S_\phi + \beta I)^{-1}$ . Remember that  $\phi$  is a compact operator from  $W$  to  $W$ . This fact is crucial. It implies that we just have to prove that

$$\beta (S_\phi + \beta I)^{-1} x \rightarrow 0$$

By Lemma 7.4 this fact is straightforward. Indeed it was then proved that  $S_\phi$  is a self-adjoint positive operator hence admits the spectral decomposition

$$S_\phi = \sum_{i=1}^{+\infty} \mu_i (t_i \otimes_W t_i)$$

where the  $\mu_i$ 's are the positive eigenvalues of  $S_\phi$  arranged in a decreasing order and the  $t_i$ 's are the associated eigenvectors. Then if  $x = \sum x_i t_i$  where  $x_i = \langle x, t_i \rangle_W$  we easily get

$$\|\beta (S_\phi + \beta I)^{-1} x\|_W^2 = \sum_{i=1}^{+\infty} \frac{\beta^2 \langle x, t_i \rangle_W^2}{(\mu_i + \beta)^2}$$

For fixed  $i$ ,  $\frac{\beta^2}{(\mu_i + \beta)^2} \rightarrow 0$  as  $\beta \rightarrow 0$ ,  $\sup_i \left\{ \frac{\beta^2}{(\mu_i + \beta)^2} \right\} \leq 1$  and as  $\sum_{i=1}^{+\infty} \langle x, t_i \rangle_W^2 < +\infty$ ,

applying Lebesgue's dominated convergence Theorem yields  $\beta (S_\phi + \beta I)^{-1} x \rightarrow 0$  in  $W$ .

Let us deal with the second term in (22).

**Proposition 7.3** *if  $\beta$  and  $\sqrt{\alpha}/\beta$  decays to zero,*

$$\|\Psi [\Gamma'_n - \Gamma''_n (\Gamma''_n)^\dagger \Gamma'_n] (S_{n,\phi} + \beta I)^{-1}\|_\infty \xrightarrow{\mathbb{P}} 0.$$

**Proof :** Once more it suffices to prove that

$$\frac{\|[\Gamma'_n - \Gamma''_n (\Gamma''_n)^\dagger \Gamma'_n]\|_\infty}{\beta} \xrightarrow{\mathbb{P}} 0.$$

since

$$\|[\Gamma'_n - \Gamma''_n (\Gamma''_n)^\dagger \Gamma'_n] (S_{n,\phi} + \beta I)^{-1}\|_\infty \leq \left( \frac{\|[\Gamma'_n - \Gamma''_n (\Gamma''_n)^\dagger \Gamma'_n]\|_\infty}{\beta} \right). \quad (26)$$

We keep on replacing the random operators based on the sample by their limits.

$$\begin{aligned} \|[\Gamma'_n - \Gamma''_n (\Gamma''_n)^\dagger \Gamma'_n]\|_\infty &= \|\alpha (\Gamma''_n)^\dagger \Gamma'_n\|_\infty \\ &\leq \|\alpha (\Gamma''_n)^\dagger\|_\infty \|\Gamma'_n - \Gamma'\|_\infty + \|\alpha (\Gamma''_n)^\dagger - \alpha (\Gamma')^\dagger\|_\infty \|\Gamma'\|_\infty \end{aligned}$$

By Lemma 7.1 the first term is an  $O_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$ , the second is an  $O_{\mathbb{P}}\left(\frac{1}{\alpha\sqrt{n}}\right)$ .

The last term is totally deterministic and we are going to prove that it is an  $o(\sqrt{\alpha})$ . Once again we introduce the compact operator  $U = D\Gamma^{1/2}$ . We see that

$$\alpha \Gamma''^\dagger \Gamma' = \alpha (\Gamma''^\dagger)^{1/2} (UU^* + \alpha I)^{-1/2} U \Gamma^{1/2}$$

since  $\Gamma'' = UU^*$ . First we show that  $(UU^* + \alpha I)^{-1/2} U$  is a class of operators uniformly bonded with respect to  $\alpha$ . In fact introducing the Schmidt representation for  $U$  from Lemma 7.4 we get

$$(UU^* + \alpha I)^{-1/2} U = \sum_{i=1}^{+\infty} \frac{s_i}{\sqrt{s_i^2 + \alpha}} u_i \otimes u_i$$

and

$$\sup_{\alpha \geq 0} \|(UU^* + \alpha I)^{-1/2} U\|_\infty \leq \sup_{\alpha, i} \left\{ \frac{s_i}{\sqrt{s_i^2 + \alpha}} \right\} \leq 1$$

At last noting that  $\|\alpha (\Gamma''^\dagger)^{1/2}\|_\infty = o(\sqrt{\alpha})$  and taking into account (26), the proof of Proposition 7.3 is finished.

## References

- [1] Adams R.A. and Fournier J.J.F., 2003. Sobolev spaces, Academic Press, 2nd ed.

- [2] Antoniadis A., Sapatinas T., 2003. Wavelet methods for continuous-time prediction using representations of autoregressive processes in Hilbert spaces, *J. Mult. Anal.*, 87, 133–158.
- [3] 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.
- [4] Besse, P., Cardot, H. and Stephenson, D., 2000. Autoregressive forecasting of some climatic variations, *Scand. J. Statist*, 27, 673-687.
- [5] Blanke D., Pumo B., 2003. Optimal Sampling for Density Estimation in Continuous Time, *Journal of Time Series Analysis.*, 24, No. 1, 1–23.
- [6] Bosq, D., 1991. Modelization, nonparametric estimation and prediction for continuous time processes. In: Roussas (Ed), *Nato Asi Series C*, 335, 509-529.
- [7] Bosq, D., 2000. Linear processes in function spaces. *Lectures notes in statistics*. Springer Verlag.
- [8] Brockwell P. and Davis A., 1987. *Time series: Theory and methods*, Springer-Verlag.
- [9] Cardot H., Ferraty F., Sarda P., 1999. Functional linear model. *Statist. Probab. Lett.* 45, 11-22.
- [10] Dunford, N. and Schwartz, J.T., 1988. *Linear Operators*, Vol. I & II. Wiley Classics Library.
- [11] Ferraty,F., Vieu P., 2003. The functional nonparametric model and application to spectrometric data. *Comput. Statist.* 17 n°4, 545-564.
- [12] Gohberg, I., Goldberg, S. and Kaashoek,M.A., 1991. *Classes of linear operators Vol I & II. Operator Theory : advances and applications*, Birkhäuser Verlag.
- [13] Hoel, P.G., Port S.C., Stone C.J., 1972. *Introduction to stochastic processes* Waveland Press, Inc.
- [14] Mas, A., 1999. Normalité asymptotique de l'estimateur empirique de l'opérateur d'autocorrélation d'un processus ARH(1). *C.R. Acad.Sci.*, t.329, Sér. I, 899-902.
- [15] Mas A., Menneteau L., 2003a. Perturbation approach applied to the asymptotic study of random operators, *Progress in Probability*, 55, 127-134.
- [16] Mas A., Menneteau L., 2003b. Large and moderate deviations for infinite-dimensional autoregressive processes, *Journal of Multivariate Analysis* 87, 241-260.
- [17] Marion J.M., Pumo B., 2004. Comparaison des modèles ARH(1) et ARHD(1) sur des données physiologiques, *Annales de l'ISUP*, 48, 3, pp. 29-38.

- [18] Mourid T., 1995. Contribution à la statistique des processus autorégressifs à temps continu. PHD. Thesis, Univ. Paris VI.
- [19] Pumo B., 1992. Estimation et prévision de processus autorégressifs fonctionnels. Applications aux processus à temps continu., PHD Thesis, Univ.Paris VI.
- [20] Pumo B., 1998. Prediction of continuous time processes by  $C[0, 1]$ -valued autoregressive process. Statist. Infer. for Stoch. Processes, 3,V ol. 1, 297-309.
- [21] Ramsay J.O., Silverman B.W., 1997. Functional Data Analysis, Springer.
- [22] Silverman B.W., 1996. Smoothed functional principal component analysis by choice of norm, Ann. Stat., 24, No 1, 1-24.
- [23] 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, No. 6, pp. 1403–1420.
- [24] Tikhonov A.N., Arsenin V.Y., 1977. Solutionsn of ill-posed problems. V.H. Winstons and sons, Washington..
- [25] Wong, E., 1966. Some results concerning the zero-crossings of Gaussian noise, SIAM J. Appl. Math., 14, 6, 1246-1254.
- [26] Ziemer W.P., 1989. Weakly differentiable functions. Sobolev spaces and functions of bounded variations. Graduate Text in Mathematics 120. Springer-Verlag, New-York.

Figure 1: Wong process simulated on  $[0, 192.65]$

	ARH approach			ARHD approach	
	$ARH$	$ARF$	$ARW$	$\alpha = 0.3, \beta = 0.65$	$\alpha = 0.1, \beta = 0.5$
$MSE$	0.624	0.622	0.623	0.327	0.323
$RMAE$	1.580	1.599	1.599	1.223	1.125

Table 1: Mean of MSE and RMAE error for the 50 simulations.

Error of predictions 86		
Predictor	MSE	RMAE
Wavelet II	0.063	0.89%
FAR	0.065	0.89%
ARHD $\alpha = 0.1, \beta = 0.4$	0.167	1.25%
Wavelet III	0.191	1.20%
ARHD $\alpha = 0.4, \beta = 0.8$	0.219	1.33%
ARH(1) $k_n = 1$	0.278	1.60%
SARIMA	1.457	3.72%

Table 2: MSE and RMAE for the prediction of El Niño surface temperatures during 1986 for various methods.

Figure 2: Prediction of the 105-th sample path by  $ARH$ ,  $ARF$ ,  $ARW$  and  $ARHD$  method.

Figure 3: The monthly mean Niño-3 sea surface temperature index from 1950 until 1996.

Figure 4: Evolution of Niño-3 surface temperature during 1986 and its various predictions.

Mean Error of predictions 87-96		
Predictor	MSE	RMAE
ARHD $\alpha = 0.4, \beta = 0.8$	0.53	2.1%
Local FAR	0.53	2.2%
ARHD $\alpha = 0.1, \beta = 0.4$	0.53	2.2%
FAR	0.55	2.3%
ARH(1) $k_n = 1$	0.68	2.4%
SARIMA	1.45	3.7%

Table 3: Mean value of MSE and RMAE errors for prediction of SST from 1987 to 1996.