Autoregressive forecasting of some functional climatic variations

Philippe C. Besse, Hervé Cardot
Laboratoire de Statistique et Probabilités
Université Paul Sabatier — CNRS
Philippe.Besse@cict.fr

David B. Stephenson
Centre National de la Recherche Météorologique
Météo-France Toulouse

Abstract

Many variations such as the annual cycle in sea surface temperatures can be considered to be smooth functions and are appropriately described using methods from functional data analysis. This study defines a class of Functional AutoRegressive (FAR) models which can be used as robust predictors for making forecasts of entire smooth functions in the future. The methods are illustrated and compared with pointwise predictors such as SARIMA by applying them to forecasting the entire annual cycle of climatological El Niño-Southern Oscillation (ENSO) time series one year ahead. Forecasts for the period 1987-1996 suggest that the FAR functional predictors show some promising skill, compared to traditional scalar SARIMA forecasts which perform poorly.

Key words: functional prediction, nonparametric kernel predictor, SARIMA, climatological forecast, functional data analysis, ENSO, El Niño.

1 Introduction

Many dynamical processes and relationships in the real world obey smooth functional forms. For example, the position of a knee joint while walking traces a continuous differentiable curve in space (Ramsay, 1988). Another example is provided by the seasonal evolution of climatic temperatures caused by the annual march of the earth about the sun. Observations of such processes, however, result in a finite discrete times series, that are often simply treated as multivariate data. This multivariate approach completely ignores important information about the smooth functional behavior of the generating process. The smoothness can provide a very useful constraint for making regression problems well-posed (Green and Silverman, 1994).

Some early remarks about factorial analysis of functional data were made by Tucker (1958). Much later, Deville (1974) performed a Principal Component Analysis (PCA) of functional observations, and Dauxois and Pousse (1976) produced a sophisticated functional analytical exposition of PCA. Ramsay (1982) presented strong arguments for introducing functional analysis methods for dealing with the statistics of functional data, and Besse and Ramsay (1986) showed how functional PCA corresponds to using a modified norm. Ramsay and Dalzell (1991) discussed some tools for functional data analysis and applied them in the analysis of the annual cycle of some monthly mean Canadian climate data. Rice and Silverman (1991) discussed the estimation of means and covariances of functions and clearly discussed functional norms in terms of roughness penalties. Pezzulli and Silverman (1993) obtained perturbative expansions which showed that functional PCA could be advantageous. Silverman (1996) and Besse et al. (1997) introduced new ways of performing smoothed functional PCA by simply adding a roughness penalty to the norm. A comprehensive description of methods for exploring and estimating functional data can be found in the recent book by Ramsay and Silverman (1997).

The purpose of this study is to develop and compare different methods for forecasting functional data. It is possible that the smoothness of functional data may be exploitable to give improved forecasts. The data can be considered as discrete observations made of a functional process, in other words, a process taking values within a linear vector space of functions. Each realization can be assumed to give a regular smooth curve. Several predictors are well suited to this situation. Firstly, a vector valued nonparametric kernel predictor can be used as an approximation to the predicted curve. A more recent tool dealing with the spline approximation (Besse and Cardot 1996) of a functional predictor (Bosq 1991) can also be considered. If the functional process is first order autoregressive, one can also construct a functional extension of AR(1) models, referred to here as FAR(1) models, in which whole functions are forecast one step ahead. We propose in this paper a new hybrid method that is a local adaptation of the FAR(1) model by introducing a weighted kernel estimator of the covariance operator. In this article, these different methods will be compared by forecasting real climatic data, namely, the annual cycle, represented by 12 monthly means, of climate indices describing the El Niño-Southern Oscillation - one of the major climatic features controlling world climate. The functional forecasts will also be compared to more conventional approaches such as the parametric SARIMA model (Box and Jenkins 1976, Brockwell and Davis 1987), and a nonparametric kernel predictor (Collomb 1983, Györfi et al. 1989, Bosq 1996).

2 Forecasting methods

Consider a real-valued time series $(X_k)_{k\in\mathbb{Z}}$ observed p times a period for n periods $\{x_1, x_2, \ldots, x_{np}\}$. For example, 50 years of monthly mean climate data has $n = 50$ and $p = 12$. The time series can also usefully be considered as a sequence of n functions $\{y_i(t_j) = x_{(i-1)p+j}; i = 1, n\}$ of the parameter $\{t_j; j = 1, p\}$ describing the time during the period (e.g. calendar month).

2.1 Scalar kernel predictor

Firstly, consider the nonparametric prediction of a real process. Let

$$\mathbf{X}_{t,(r)} = (X_t, X_{t-1}, \ldots, X_{t-r+1}) \in \mathbb{R}^r$$

be the vector of lagged variables and s the forecast horizon ($0 < s \leq p$). The autoregression function is defined by

$$f_s(\mathbf{x}) = \mathbb{E}(X_{T+s} \mid \mathbf{X}_{T,(r)} = \mathbf{x}).$$

The kernel estimator based on $(x_1, \ldots, x_T)$ of this regression function f_s is then

$$f_{T,s}(\mathbf{x}) = \frac{\displaystyle\sum_{t=r}^{T-s} x_{t+s}\, K\left(\frac{\mathbf{x} - \mathbf{x}_{t,(r)}}{h_T}\right)}{\displaystyle\sum_{t=r}^{T-s} K\left(\frac{\mathbf{x} - \mathbf{x}_{t,(r)}}{h_T}\right)} \tag{1}$$

where h_T is the bandwidth window, and K, the kernel function, is a bounded symmetric r-dimensional density. Theoretical results show that the detailed choice of the kernel function does not strongly influence the asymptotic behavior of prediction (Bosq 1996). In this study, K is chosen to be the normal density kernel

$$K(\mathbf{x}) = (2\pi)^{-r/2} \exp\left(-\frac{\|\mathbf{x}\|^2}{2}\right), \;\; \mathbf{x} \in \mathbb{R}^r.$$

The horizon s prediction is then given by

$$\widehat{X}_{T+s|T} = f_{T,s}(\mathbf{x}_{T,(r)}).$$

The prediction is expressed as a weighted mean of past values, where the weights measure the similarity between $\mathbf{x}_{T,(r)}$ and $\mathbf{x}_{t,(r)}$, taking into account the process history.

Theoretical studies demonstrate the good asymptotic behavior of this predictor, which can reach the optimal rate of convergence under different sets of assumptions dealing with stationarity, mixing conditions, Markovian properties or ergodicity. Furthermore, Bosq (1996) demonstrated the robustness of the kernel predictors in the presence of certain nonstationarities. This justifies the common practice of dealing with the raw data without first removing any unknown deterministic component such as trend or seasonality which naturally take part of the process history.

The choice of the h_T value is crucial for prediction accuracy. It is optimized by means of the cross-validation criterion, as in Poggi (1994),

$$CV_s(h) = \sum_{k=r,\dots,T-pm-s} \left(x_{k+s} - f_{s,h,-k}(\mathbf{x}_{k,(r)})\right)^2 \tag{2}$$

where $f_{s,h,-k}$ is the horizon-s regression function based on a learning subset $\{x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_{T-mp}\}$ excluding the kth value. The m last periods is used for testing the quality of prediction. The optimal value of h_T is given by

$$h_T = \operatorname{argmin} \sum_{s=1}^{p} CV_s(h).$$

For simplicity, we have not chosen to make h_T depend on the horizon prediction s. The period of the seasonal component p is a natural choice for the order r. It can also be optimized by using cross-validation (Vieu, 1995).

2.2 Functional kernel prediction

Let $(Y_i)_{i\in\mathbb{Z}}$ be a second order stationary Hilbertian random process, which is assumed to be Markovian. It is a sequence of random functions belonging to the functional Hilbert space H. In our functional framework, we assume that Y_i is a "smooth" stochastic process. One way to control the smoothness and the regularity of observations is to let H to be a Sobolev space W^d, that is to say a collection of functions on the range $[t_1, t_p]$ which satisfy

$$\left\{f; f, f', \dots, f^{(d-1)} \text{ absolutely continuous }, f^{(d)} \in L^2([t_1, t_p])\right\}.$$

A natural approximation of a function when minimizing a norm in such a space is a smoothing spline interpolation. However if the above smoothness assumption is not valid the use of a spline approximation is no longer appropriate. Other families of functional spaces and basis functions such as wavelets can then sometimes be used to take into account singularities and discontinuities.

The forecasting method studied by Bosq (1983) considered the conditional expectation $\rho(y) = \mathbb{E}(Y_i|Y_{i-1} = y)$ where y_i is the i'th functional trajectory of the stochastic process. The operator ρ is not necessarily linear, and can be estimated by using a nonlinear kernel regression; $\rho(y)$ defines a regression between functions rather than vectors. For computational simplicity, the distance between two curves belonging to H is evaluated according to the $L^2[t_1, t_p]$ norm:

$$\|f - g\|_{L^2}^2 = \int_{t_1}^{t_p} (f(t) - g(t))^2 \; dt, \quad \forall (f,g) \in H \times H.$$

The functional kernel estimator is then defined as

$$\widehat{\rho}_{h_n}(y) = \frac{\displaystyle\sum_{i=1}^{n-1} y_{i+1} K\left(\frac{\|y_i - y\|_{L^2}}{h_n}\right)}{\displaystyle\sum_{i=1}^{n-1} K\left(\frac{\|y_i - y\|_{L^2}}{h_n}\right)},$$

where $K(.)$ is the usual Gaussian kernel.

In practice, the random curves y_i are only known for discretized time values $\{t_j; j = 1, \ldots, p\}$, and thus they must be approximated by some smooth $\widetilde{y}_i$ functions. Let $\widetilde{y}_i$ be the spline interpolation (Wahba, 1990) of the i'th trajectory based upon the vector of observations $\{y_i(t_1), \ldots, y_i(t_p)\}$. By definition, it is the solution of the optimization problem

$$\widetilde{y}_i = \quad \text{argmin} \, \|L\widetilde{y}_i\|_{L^2}^2 \qquad \text{subject to} \quad \widetilde{y}_i(t_j) = y_i(t_j), \; j = 1, \ldots p, \tag{3}$$

where L is an order d linear differential operator. In all examples, $L = D^2$, but more general Chebyshev splines (Wahba 1990, Ramsay and Silverman 1997) can also be used. We then consider the following approximation of the functional kernel estimator

$$\widehat{\rho}_{h_n}(y) = \frac{\displaystyle\sum_{i=1}^{n-1} \widetilde{y}_{i+1} K\left(\frac{\|\widetilde{y}_i - y\|_{L^2}}{h_n}\right)}{\displaystyle\sum_{i=1}^{n-1} K\left(\frac{\|\widetilde{y}_i - y\|_{L^2}}{h_n}\right)}, \tag{4}$$

giving the prediction

$$\widehat{y}_{n+1} = \widehat{\rho}_{h_n}(y_n)$$

where the function y is constrained to belong to the space of spline functions which are third degree piecewise polynomials. A cross-validation procedure is performed to optimize the bandwidth value h_n. It aims at minimizing a quadratic prediction error between a subset of r observed curves and their own prediction based on $\{y_i; i = 1, \ldots, n - r - m\}$. As above in the scalar kernel prediction case, the final m curves are used to test the quality of the forecast.

2.3 Prediction using a Functional Autoregressive model

Functional predictors for first order Functional Auto-Regressive FAR(1) processes ("Auto-Regressif Hilbertien d'ordre 1" in French) were introduced by Bosq (1991). Independently, Besse (1994a,b) and Besse et al. (1997) developed methods for simultaneous nonparametric estimation of several curves, that involved solving an optimization problem that has both a dimension reduction and regularity constraints (see equation 5 later). By merging the two approaches, Besse and Cardot (1996) then developed an improved method for predicting FAR processes that uses spline-smoothed functional PCA. After recalling the definition of the smooth FAR predictor, we introduce, in this section, a local version to reduce the sensitivity to the stationarity assumption.

We now consider $(Y_i)_{i\in\mathbb{Z}}$ to be a first order Functional Auto-Regressive (FAR) process satisfying

$$\forall i \in \mathbb{Z}, \quad Y_i = \rho Y_{i-1} + \epsilon_i.$$

with $\mathbb{E}(Y_i) = 0$ and $\mathbb{E}\,\|Y_i\|_H^2 < +\infty$. The autocorrelation operator ρ is compact and $\|\rho\| < 1$. The error terms $\{\epsilon_i\}$ are assumed to be zero-mean independent identically distributed random elements of H with $\mathbb{E}\,\|\epsilon_i\|_H^2 = \sigma^2 < +\infty$.

Denote by $\langle .,.\rangle_H$ the inner product of H and define $x \otimes y$ as the rank one operator which satisfies :

$$\forall (x,y,z) \in H \times H \times H, \quad [x \otimes y](z) = \langle x, z\rangle_H\, y.$$

The operators $\Gamma = \mathbb{E}(Y_i \otimes Y_i)$ and $\Delta = \mathbb{E}(Y_i \otimes Y_{i+1})$ are the simultaneous and lag-1 covariance operators of the process and ρ obeys the following relationship

$$\rho\Gamma = \Delta.$$

A natural estimator of ρ can be obtained from the empirical covariance operators

$$\widehat{\Gamma} = \frac{1}{n}\sum_{i=1}^{n} y_i \otimes y_i \quad \text{and} \quad \widehat{\Delta} = \frac{1}{n-1}\sum_{i=1}^{n-1} y_i \otimes y_{i+1}.$$

Smooth FAR(1) predictor

As above, the random curves y_i are only known for discretized time values, and must therefore be approximated by functions $\widehat{y}_i$. Different strategies can be considered, such as linear interpolations (Pumo, 1992), spline interpolation or spline smoothing, before considering the prediction problem.

Since Γ^{-1} is unbounded, Bosq (1991) suggested to approximate it by its restriction to the reduced subspace spanned by the leading eigenvectors of

$\widehat{\Gamma}$. Besse and Cardot (1996) proposed to directly approximate the random curves by functions that both obey a rank and a smoothness constraint. This leads us to adapt the classical spline smoothing definition by adding a rank constraint, and then to consider the following optimization problem

$$\min_{\hat{y}_i \in H_q} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{p} \sum_{j=1}^{p} (y_i(t_j) - \widehat{y}_i(t_j))^2 + \ell \left\| D^2 \widehat{y}_i \right\|_{L^2}^2 \right). \tag{5}$$

H_q is a q-dimensional functional subspace to be estimated. It is spanned by the smooth eigenvectors of $\widehat{\Gamma}$ associated with the q largest eigenvalues, relative to a specific Euclidean metric depending on the smoothing parameter. The solution $\{\widehat{y}_1, \ldots, \widehat{y}_n\}$ is obtained by means of a functional PCA (Besse 1994, Besse et al. 1997).

Let us now consider the estimators of the covariance operators defined by

$$\widehat{\Gamma}_{q,\ell} = \frac{1}{n} \sum_{i=1}^{n} \widehat{y}_i \otimes \widehat{y}_i \quad \text{and} \quad \widehat{\Delta}_{q,\ell} = \frac{1}{n-1} \sum_{i=1}^{n-1} \widehat{y}_i \otimes \widehat{y}_{i+1}. \tag{6}$$

An estimator of the operator ρ is then constructed by inverting $\widehat{\Gamma}_{q,\ell}$ in H_q

$$\widehat{\rho}_{q,\ell} = \widehat{\Delta}_{q,\ell} \left(\widehat{\Gamma}_{q,\ell} \right)^{-1} \tag{7}$$

and the smooth FAR(1) prediction is given by

$$\widehat{y}_{n+1} = \widehat{\rho}_{q,\ell} y_n. \tag{8}$$

In this new approach, two parameters must be tuned: The dimension q, the number of eigenvectors or principal components, and the smoothing parameter ℓ. Their values are jointly optimized by means of a cross-validation approach as above. Implicitly and for simplicity we assumed that all the observed curves share the same discretization design. Nevertheless it is easy to drop that assumption and to accept different discretization designs or missing data. Aiming at the same purpose within the functional PCA framework, Besse et al. (1997) use a hybrid spline approximation merging box splines and spline smoothing. This technique can easily be adopted within the functional prediction framework of the smooth FAR(1) model.

Local FAR(1) predictor

The FAR(1) prediction model is clearly based on the stationarity assumption for $(Y_j)_{j \in \mathbb{Z}}$. This means that Γ and Δ do not depend on j. As this is a very strong assumption, it could be interesting to consider a FAR(1) model that is robust with respect certain departures from stationarity. This is achieved by mimicking kernel predictors, and thus by defining local estimates of the covariance operators. These are weighted estimators whose weights depend on the proximity between the last observed curve and the passed curves.

The prediction of y_{n+1} knowing the sequence $\{y_i; i=1,\ldots,n\}$ leads us to consider the estimators

$$\widehat{\Delta}_{h_n} = \frac{\displaystyle\sum_{i=1}^{n-1} \widetilde{y}_i \otimes \widetilde{y}_{i+1} K\left(\frac{\|\widetilde{y}_i - \widetilde{y}_n\|_{L^2}}{h_n}\right)}{\displaystyle\sum_{i=1}^{n-1} K\left(\frac{\|\widetilde{y}_i - \widetilde{y}_n\|_{L^2}}{h_n}\right)}, \tag{9}$$

$$\widehat{\Gamma}_{h_n} = \frac{\displaystyle\sum_{i=1}^{n} \widetilde{y}_i \otimes \widetilde{y}_i K\left(\frac{\|\widetilde{y}_i - \widetilde{y}_n\|_{L^2}}{h_n}\right)}{\displaystyle\sum_{i=1}^{n-1} K\left(\frac{\|\widetilde{y}_i - \widetilde{y}_n\|_{L^2}}{h_n}\right)}. \tag{10}$$

The estimation of the operator ρ becomes

$$\widehat{\rho}_{q,h_n} = \widehat{P}_{q,h_n} \widehat{\Delta}_{h_n} \widehat{P}_{q,h_n} \widehat{\Gamma}_{h_n}^{-1} \widehat{P}_{q,h_n}$$

where $\widehat{P}_{q,h_n}$ is the eigenprojector associated with the q largest eigenvalues of $\widehat{\Gamma}_{h_n}$. Both q and h_n values are optimized by cross-validation.

Remarks

Formally we could define some more complex predictors in both cases of the functional kernel prediction and the local FAR model. In these two cases we could approximate each curve by mean of a smoothing spline rather than an interpolating one. This would lead to tune two smoothing parameters. It could be justified when dealing with nosier or more complex data, which does not appear to be necessary for the ENSO series used in this study. For simplicity and for making the predictors better comparable, we chose to tune each functional predictor by only one real parameter, either a bandwidth in the case of a kernel regressor or a smoothing parameter spline functions.

It is quite easy to find references dealing with asymptotic properties of kernel predictors among a very wide literature, but only partial results are known in the functional case. Bosq (1986) gives some results for the Hilbertian regression predictor, consistency and some convergence rates are proved for the theoretical FAR(1) model in Bosq (1991) and Cardot (1997) gives some results on its smooth approximation.

3 Matrix representation

Each of the above functional predictors require either a spline interpolation or a spline smoothing approximation of the raw data, and have a vectorial representation within a p-dimensional subspace S_p of the reference Sobolev space $H = W^2$. Different kinds of spline subspaces may be used according to the choice of spline basis. A reproducing kernel basis (Wahba, 1990) is easy to manipulate in the present simple case and can be computed in any more general functional framework (Ramsay and Silverman, 1997). It has been used in all further computations.

Let y be a function belonging to W^2, $\widetilde{y}$ its spline interpolation, solution of (3), fitting the values which are collected in the vector $\mathbf{y} = [y(t_1), \ldots, y(t_p)]'$. We denote by $\mathbf{A}_\ell$ the smoothing or hat matrix

$$\mathbf{A}_\ell = (\mathbf{M} + \ell\mathbf{N})^{-1}$$

where the matrices $\mathbf{M}$ and $\mathbf{N}$ are respectively the matrix representation of the induced L^2 norm and the W^2 semi-norm onto S_p

$$\begin{aligned} \|\widetilde{y}\|^2_{L^2} &= \mathbf{y}'\mathbf{M}\mathbf{y} = \|\mathbf{y}\|^2_{\mathbf{M}}\,, \\ \left\|D^2\widetilde{y}\right\|^2_{L^2} &= \mathbf{y}'\mathbf{N}\mathbf{y} = \|\mathbf{y}\|^2_{\mathbf{N}}\,. \end{aligned}$$

The usual spline smoothing of $\mathbf{y}$ is obtained by a spline interpolation of values contained in the vector $\mathbf{A}_\ell\mathbf{y}$.

Functional kernel

Let us denote by $\mathbf{y}_i$ the column vector containing the ith row of $\mathbf{Y}$, containing the observed values of the ith period of the time series, and by $\mathbf{W}_h$ the diagonal matrix of the weights

$$w_i = \frac{K\left((\mathbf{y}_i - \mathbf{y}_n)'\mathbf{M}(\mathbf{y}_i - \mathbf{y}_n)/h\right)}{\sum_{l=1}^{n} K\left((\mathbf{y}_i - \mathbf{y}_n)'\mathbf{M}(\mathbf{y}_i - \mathbf{y}_n)/h\right)}.$$

The functional kernel predictor is directly drawn from (4) and

$$\widehat{\mathbf{y}}_{n+1} = \sum_{i=1}^{n-1} w_i \mathbf{y}_{i+1}.$$

As for all the subsequent functional predictors, the functional forecast estimation is obtained by a spline interpolation of the values of $\widehat{\mathbf{y}}$.

Smooth FAR(1)

The data matrix $\mathbf{Y}$ was column-centered by subtracting the empirical mean $\bar{\mathbf{y}}$ and we aimed at forecasting climatic anomalies. The solution $\widehat{y}$ of the smoothing spline problem with a rank constraint (5) was deduced from the eigenanalysis of the "smoothed" covariance matrix

$$\mathbf{S} = \frac{1}{n}\mathbf{A}_\ell^{1/2}\mathbf{Y}'\mathbf{Y}\mathbf{A}_\ell^{1/2}.$$

Let $\mathbf{V}_q$ be the orthogonal matrix containing q eigenvectors associated with the q largest eigenvalues,

$$\widehat{\mathbf{y}}_i = \mathbf{A}_\ell^{1/2}\mathbf{V}_q\mathbf{V}_q'\mathbf{A}_\ell^{1/2}\mathbf{y}_i, \quad i = 1, \ldots, n.$$

The matrix approximation of the covariance operators of (6) becomes

$$\widehat{\mathbf{\Gamma}}_{q,\ell} = \frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{y}}_i\widehat{\mathbf{y}}_i'\mathbf{M}, \quad \text{and} \quad \widehat{\mathbf{\Delta}}_{q,\ell} = \frac{1}{n-1}\sum_{i=1}^{n-1}\widehat{\mathbf{y}}_{i+1}\widehat{\mathbf{y}}_i'\mathbf{M}.$$

Then the matrix approximation $\widehat{\boldsymbol{\rho}_{q,\ell}}$ of ρ is directly drawn from (6) which gives us a smooth functional forecast

$$\widehat{\mathbf{y}}_{n+1} = \widehat{\boldsymbol{\rho}_{q,\ell}}\mathbf{y}_n + \mathbf{A}_\ell\bar{\mathbf{y}}.$$

In finite dimensions, the operator $\widehat{\Gamma}$ is generally invertible. Nevertheless, it is often severly ill-conditioned and a dimension reduction that leads to a generalized inverse can often help to improve prediction quality. As in Besse et al. (1997), it is often found that the regularity of the eigenfunctions increases with their associated eigenvalues. This is similar to a Fourier decomposition where a dimension reduction can act, together with spline smoothing, to help denoise by filtering out high frequencies. In the case of a Wiener process, the eigenfunctions of the covariance are identical to a Fourier decomposition. More details can be found in Besse and Cardot (1996).

Local FAR(1)

The local approximation of the autocovariance operators are deduced from (9) and (10).

$$\begin{aligned}\widehat{\mathbf{\Gamma}}_h &= \frac{1}{n}\sum_{i=1}^{n} w_{i,h}\mathbf{y}_i\mathbf{y}_i'\mathbf{M} \\ &= \mathbf{Y}'\mathbf{W}_h\mathbf{Y}\mathbf{M};\end{aligned}$$

and

$$\widehat{\mathbf{\Delta}}_h = \frac{1}{n-1}\sum_{i=1}^{n-1} w_i\mathbf{y}_{i+1}\mathbf{y}_i'\mathbf{M} \ .$$

Then

$$\left(\widehat{\mathbf{P}}_q\widehat{\mathbf{\Gamma}}_h\widehat{\mathbf{P}}_q\right)^{-1} = \sum_{l=1}^{q}\frac{1}{\lambda_l}\mathbf{v}_l\mathbf{v}_l'\mathbf{M}$$

where $(\lambda_l, \mathbf{v}_l)$ are the eigenelements of the matrix $\widehat{\mathbf{\Gamma}}_h$ with respect to the metrics $\mathbf{M}$, and $\widehat{\boldsymbol{\rho}}_{q,h}$ may be written as follows:

$$\widehat{\boldsymbol{\rho}}_{q,h} = \left(\sum_{l=1}^{q}\mathbf{v}_l\mathbf{v}_l'\right)\mathbf{M}\widehat{\mathbf{\Delta}}_h\left(\sum_{l=1}^{q}\frac{1}{\lambda_l}\mathbf{v}_l\mathbf{v}_l'\right)\mathbf{M}. \qquad (11)$$

4 A climate example: ENSO

4.1 Data

Functional data methods of forecasting will be illustrated using real data, namely, climatological time series describing the El Niño-Southern Oscillation (ENSO). We prefer to test forecasts on real observed data rather than on simulated time series, which can easily be constructed to yield good performances. ENSO is a natural phenomenon arising from interactions between the atmosphere and the ocean in the tropical Pacific Ocean (Philander, 1990). El Niño (EN) is the ocean component of ENSO, involving major changes in tropical Pacific sea surface temperatures (especially around December time), and the Southern oscillation (SO) is its atmospheric counterpart. The Southern Oscillation causes correlations between the weather in widely separated regions of the globe, and was discovered at the beginning of this century by Sir Gilbert Walker while searching for suitable predictors for the Indian summer monsoon. Sir Gilbert Walker also developed statistical methods such as the Yule-Walker equations for forecasting such climatic indices. Most of the year-to-year variability in the tropics, as well as a part of the extra-tropical variability over both the Hemispheres, is related to this phenomenon. In 1983, an exceedingly warm ENSO event occurred which caused worldwide climate damage estimated to have cost more than 8 billion US dollars. Not surprisingly, there is much interest in forecasting ENSO and current climate models show some ability to do this up to about 6 months in advance (Latif et al., 1994).

The interannual variability of ENSO is strongly modulated by the annual cycle, and vice versa, and this interaction is known to play an important role in the predictability of ENSO anomalies (Chen et al., 1997). Using a functional forecasting approach, we can treat the annual cycle as an entity in its own right and will use this concept to make one year ahead forecasts of the entire annual cycle for both the El Niño and the Southern Oscillation indices. This approach has never been attempted in making forecasts of ENSO, which invariably attempt to forecast only one particular month some months in advance.

A fiducial and much used index of El Niño variability is provided by the sea surface temperatures averaged over the Niño-3 domain ($5°S - 5°N$, $150°W - 90°W$). Monthly mean values have been obtained for January 1950 to December 1996 from the gridded analyses made at the U.S. National Centers for Environmental Prediction (Reynolds and Smith 1994, Smith et al., 1996)[1]. The time series of this EN index is depicted in the upper panel of Figure 1, and shows marked interannual variations superimposed on a strong seasonal component.

[1]Data is freely available from http://nic.fb4.noaa.gov:80/data/cddb.

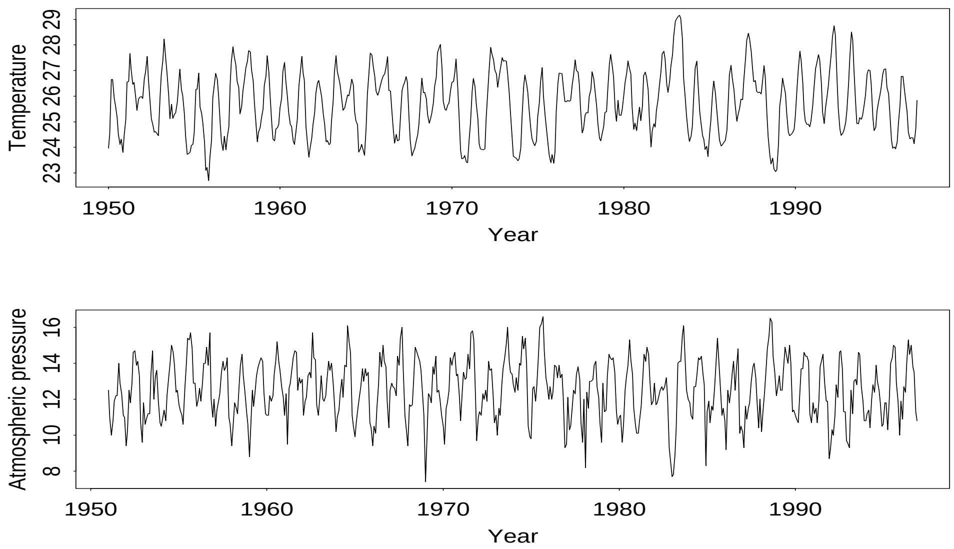

Figure 1: *Two monthly mean time series which provide a contracted description of ENSO: The monthly mean Nino-3 sea surface temperature index (°C) (upper panel), and the monthly mean sea level pressure at Tahiti (hPa) (lower panel). The sea-level pressures are relative to 1000hPa.*

The associated Southern Oscillation in sea-level pressure projects strongly on the sea-level pressure at Tahiti, and hence sea level pressure measurements at Tahiti can be used as an index for Southern Oscillation behavior. The Darwin pressure index has been extended back to 1866 by Ropelewski and Jones (1987) and its recent evolution is depicted in the lower panel of Figure 1. It also shows a marked seasonality but is noisier than the EN index due to it being an atmospheric rather than an oceanic variable. Both the El Niño (EN) and Southern Oscillation indices have probability distributions close to normal (Stephenson, 1997).

4.2 Predictions

The kernel predictor is easy to compute from (1) for different horizons up to the period $s = 12$. The autoregressive order is linked to this period ($r = 12$). Only the optimization of the bandwidth value h_T by minimizing (2) requires some computation effort.

To complete the comparison, classical parametric models (ARIMA) including seasonality (Brockwell and Davis, 1987) were fitted to the two series. These models are driven by seven integer parameters setting the degree of each polynomial respectively dealing with the autoregressive and the moving average part and removing polynomial or seasonal trends. The most parsi-

	Smooth FAR(1)	Local FAR(1)	Kernel	Functional kernel
EN	$q = 4$ $\ell = 1.6\text{e-}05$	$q = 4$ $h_n = 0.9$	$h_T = 0.9$	$h_n = 0.3$
SO	$q = 3$ $\ell = 8\text{e-}05$	$q = 2$ $h_n = 5$	$h_T = 1.5$	$h_n = 0.6$

Table 1: *Optimal parameter values for the different functional models*

monious convenient parametric model was $(0,1,1) \times (1,0,1)_{12}$ for the Sea Surface Temperature (EN) and $(1,1,1) \times (0,1,1)_{12}$ for the Sea Level Pressure (SO). The portmanteau test was performed to check that the residuals were not correlated and then validate these models.

The same cross-validation procedure was used for all the functional predictors. The first 36 years, from 1950, were considered as a learning subset. The optimal h_n, q or ℓ values minimized the MSE when predicting years from the 32nd to the 36th. These optimal values, displayed in Table 1, were used to tune the predictors in order to forecast the ten last years from the 37th to the 46th.

The data were pretty regular, and the smoothing had a minor effect. As a consequence, the optimal value of h_n was small, especially for the EN data. This would not be the case with noisy data as highway traffic (Besse and Cardot 1996). Figure 2 displays the observed data of the 37th year and its forecasts. A linear interpolation joins the raw data and scalar predictions whereas functional predictions are piecewise polynomials of degree 3.

4.3 State space model

We discuss in this section how the smooth FAR(1) predictor can be approximated by a vector autoregressive model (VAR), and then how the equivalent state space approach allows us to consider the goodness of fit of the FAR(1) model. This is achieved by considering the equivalent Markov representation with state variables of any multivariate ARMA model (Akaïke, 1976).

For simplicity, smoothing and dimension reduction may be split into two steps, and then the approximation of the FAR(1) model may be decomposed into the following procedure:

1. Smooth each observed curve by any convenient tool with one smoothing parameter ℓ. For instance by performing a functional PCA (Besse et al. 1997, Ramsay and Silverman 1997),

2. project the smoothed curves onto the subspace spanned by the first q eigenvectors of the covariance matrix to give the first q principal components c_j,

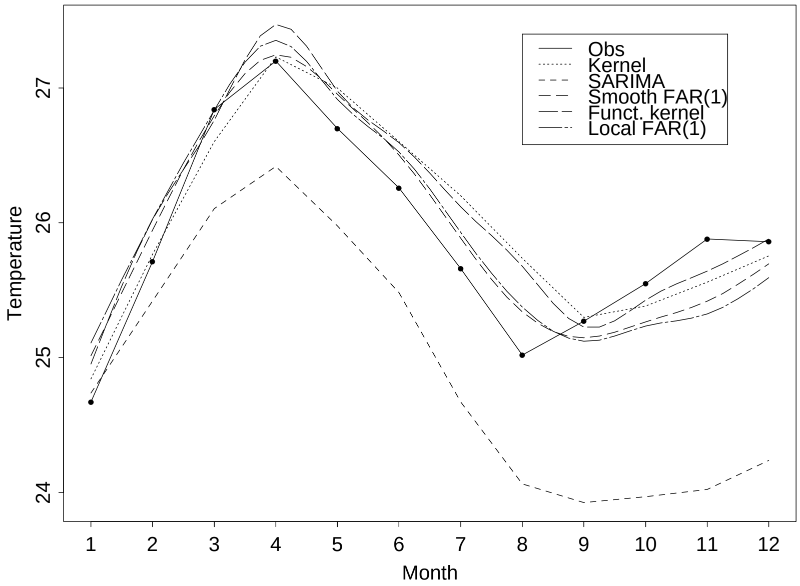

Figure 2: *The raw ENSO time series of Pacific surface temperatures during 1986 and its different forecasts.*

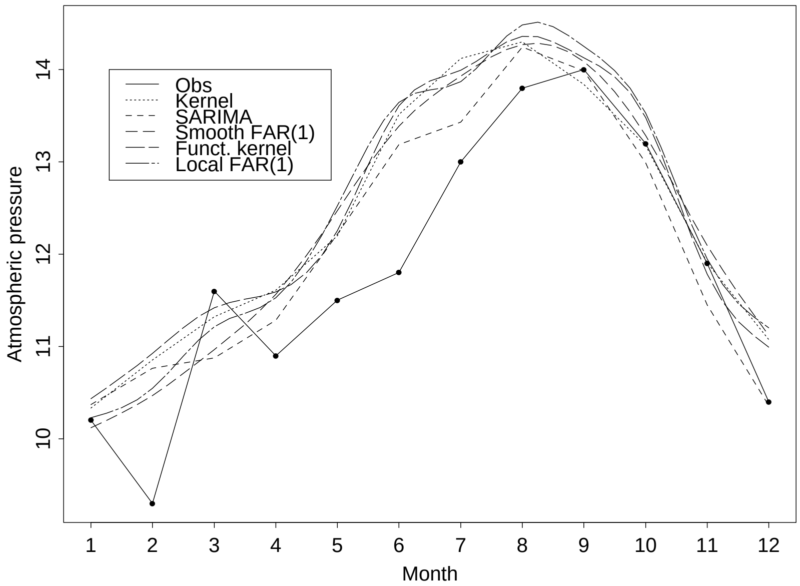

Figure 3: *Southern oscillation pressures in Tahiti during 1986 and its different forecasts.*

3. estimate the multivariate ARMA predictor on these principal components.

For the third step, the state-space procedure of the SAS/ETS (1989) software or S-PLUS functions available from Statlib can be used to compute estimates and forecasts according to these models. Iterating this procedure allows one to optimize a cross-validation index, and thus find optimal values for the parameters ℓ and q.

The main value of the state space approach is its capability to find the best model in the sense of the Akaïke's criterion. For both the cases of EN and SO data, the best model, in the sense of the Akaïke's criterion, is found to be order one autoregressive. Furthermore, since EN data looks very smooth, even without performing the first step, the prediction results are very similar with those obtained by the smooth FAR(1) model. In that case, the procedure just amounts to doing PCA together with a vector autoregressive model estimation.

If the FAR(1) model fails, this approach can nevertheless gives an heuristic approximation of more complex models. This is theoretically justified in the case of "higher" order autoregressive models (Mourid, 1995), but not necessarily when dealing with a moving average component. Further theoretical developments should study the consequences of projecting the innovation components onto the eigensubspace.

4.4 Assessment of the Forecasts

All the predictors used the data from 1950-86 (36 years) to estimate their parameters, and then were used to make 10 one year ahead forecasts for the independent period 1986-97. Various error statistics compiled over the years 1987-1996 are presented in Table 2 for the different forecasts:

The forecast quality has been assessed using both the often-used Mean Squared Error (MSE), and the Mean Relative Absolute Error (MRAE):

$$MRAE = \frac{1}{s}\sum_{t=1}^{s}\frac{|\widehat{X}_{T+t|T} - x_{T+t}|}{|x_{T+t}|}.$$

The MRAE accounts for non-centered quantities being much larger in certain seasons (eg. Winter) which then causes the forecast errors for these months to dominate the MSE. It can be seen that none of the forecasts are particularly skillful in forecasting the annual cycle of either EN or SO one year ahead. This is perhaps, however, not so surprising considering that present day forecasts of ENSO only show skill for leads of less than 6 months and sometimes even less (Latif et al., 1994). Forecasting the annual cycle one year ahead, therefore represents a more difficult exercise and provides a more exacting test for forecast models. Most of the models give smaller

Predictor	El Niño index MSE	El Niño index MRAE	S. Osc. index MSE	S. Osc. index MRAE
Climatology	0.73	2.5 %	0.91	6.3 %
SARIMA	1.45	3.7 %	0.95	6.2 %
Kernel	0.60	2.3%	0.87	6.1%
Functional kernel	0.58	2.2%	0.82	6.0%
Smooth FAR(1)	0.55	2.3%	**0.78**	**5.8%**
Smooth FAR(1) with $q = p = 12$	0.60	2.4%	0.91	6.5%
Local FAR(1)	**0.53**	**2.2%**	0.82	5.8%

Table 2: *Mean squared error (MSE) and Mean Relative Absolute Error (MRAE) for one-year ahead forecasts of El Niño and the Southern oscillation indices for the period 1987-1996. Climatology uses the time mean annual cycle to forecast the following year's annual cycle and represents a simple benchmark forecast with which to compare the other forecasts. Bold face indicates the best forecast for each index*

MSE and MRAE than does the reference climatological forecast, which is made by simply assuming that the following year annual cycle will be equal to the mean annual cycle estimated over all the previous years.

SARIMA is an exception which gives the worst predictions. This is perhaps due to a lake of stationarity, the variance appears to be seasonal, or to the modulation of the annual cycle by the interannual ENSO events (Gu and Philander, 1995), which leads to non-orthogonality between the seasonal and trend components (Franses, 1996). This is in agreement with Carbon and Delecroix (1993), who showed that a classical kernel predictor generally gives better results than a SARIMA model.

Examination of Table 1 suggests that functional FAR forecasts outperform the other methods. The local FAR(1) model forecasts the smoother EN data better than does the smooth FAR(1) model, yet the smooth FAR(1) method forecasts better the noisier SO data. It can also be seen that a dimension reduction improves the prediction. These preliminary results are encouraging and more work is currently in progress. Furthermore, new methods should also be developed for assessing the skill of functional forecasts, which take into account the functional nature of the forecast. For example, which norms are the most appropriate for judging the forecast function, and how one could assign a probability of success to such a forecast.

5 Concluding Remarks

In the absence of any universal unique predictor that will work best for all encountered cases, it is a wise strategy to experiment with a wide class of different estimators, and then retain the most suitable one. In appropriate cases and for a small horizon prediction, a classical ARIMA model can often be expected to give the best forecasts. However, in the case of real data and longer forecast horizons, strong departures from the important assumptions of linearity and stationarity can invalidate the use of ARIMA models. In such cases, more robust predictors can perform better and are worth investigating. Both local and smooth FAR predictors provide the best one year ahead forecasts of EN and SO, respectively (Table 2). On one hand, smooth FAR generally yields better forecasts for rough time series such as the SO index based on noisy pressures. On the other hand, the local FAR model is more robust to nonstationarity.

This study has introduced both smooth and local functional FAR forecast models and compared their ability against that of traditional scalar estimators such as SARIMA in forecasting one year ahead the annual cycle of real climate ENSO data. ENSO indices are strongly modulated by the annual cycle and are known to exhibit strong seasonal nonstationarities. Furthermore, a forecast of more than 6 months represents a long horizon forecast of ENSO and in such cases one might expect more robust predictors to perform better. This study has indeed shown this to be the case, with the FAR models outperforming the other predictors.

This study is a first attempt to develop functional data analytic methods for use in forecasting applications. The results appear promising and there are many potential situations where one might consider the data to be functional, for example, the smooth evolution of annual cycles. In such cases, forecasting an entire function using FAR models and their future extensions, offer robust useful new ways to forecast data, and therefore merit further attention.

References

Akaike, H. (1976). Canonical correlation analysis of time series and the use of an information criterion. In Mehra, R. and Lainiotis, D., editors, *Advances and Case Studies in System Identification*. Academic Press.

Besse, P. (1994a). Simultaneous non-parametric regressions. *Publication du laboratoire de Statistique et Probabilités de Toulouse*, 94(08).

Besse, P. (1994b). Models for multivariate analysis. In Dutter, R. and Grossman, W., editors, *Compstat 94*, pages 271–285. Physica Verlag.

Besse, P. and Cardot, H. (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Revue Canadienne de Statistique/Canadian Journal of Statistics*, 24:467–487.

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

Besse, P. and Ramsay, J. (1986). Principal component analysis of sampled curves. *Psychometrika*, 51:285–311.

Bosq, D. (1983). Non parametric prediction in stationary processes. *Lecture Notes in Statistics*, 16:69–84. Berlin-Heidelberg-New York.

Bosq, D. (1991). Modelization, non-parametric estimation and prediction for continuous time processes. In Roussas, G., editor, *Nonparametric Functional Estimation and Related Topics*, pages 509–529. Nato, ASI series.

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes, Estimation and Prediction.* Lecture Notes in Statistics. Springer-Verlag.

Box, G. and Jenkins, G. (1976). *Time series analysis.* Holden Day.

Brockwell, P. and Davis, A. (1987). *Times Series: Theory and Methods.* Springer-Verlag.

Carbon, M. and Delecroix, M. (1993). Non-parametric vs parametric forecasting in time series : a computational point of view. *Applied Stochastic Models ans Data Analysis*, 9:215–229.

Cardot, H. (1997). Convergence du lissage et de l'interpolation spline de la prévision des processus autorégressifs fonctionnels. *Comptes Rendus à l'Académie des Sciences*, t.326, Série 1326:755–758.

Chen, Y.-Q., Battisti, D., Palmer, T., Barsugli, J., and Sarachik, E. (1997). A study of the predictability of tropical pacific sst in a coupled atmosphere-ocean model using singular vector analysis: The role of the annual cycle and the enso cycle. *Monthly Weather Review*, 125:831–845.

Collomb, G. (1983). From non parametric regression to non parametric prediction: Survey on the mean square error and original results on the predictogram. *Lecture Notes in Statistics*, 16:182–204. Berlin-Heidelberg-New York.

Dauxois, J. and Pousse, A. (1976). *Les analyses factorielles en calcul des Probabilités et en Statistique : Essai d'étude synthétique.* PhD thesis, Université Paul Sabatier, Toulouse.

Deville, J. (1974). *Méthodes statistiques et numériques de l'analyse harmonique.* Ann Insee, 15.

Franses, P. (1996). *Periodicity and stochastic trends in economic time series.* Oxford University Press.

Green, P. and Silverman, B. (1994). *Nonparametric Regression and Generalized Linear Models.* Chapman and Hall.

Gu, D. and Philander, S. (1995). Secular changes of annual and interannual variability in the tropics during the past century. *J. of Climate*, 8:864–876.

Györfi, L., Härdle, W., Sarda, P., and Vieu, P. (1989). *Nonparametric Curve Estimation from Time Series.* Springer-Verlag, New York.

Latif, M., Barnett, T., Cane, M., Flugel, M., Graham, N., Xu, J., and Zebiak, S. (1994). A review of ENSO prediction studies. *Climate Dynamics*, 9:167–179.

Mourid, T. (1995). *Contribution à la statistique des processus autorégressifs à temps continu.* Thèse d'Etat, Université Paris VI.

Pezzulli, S. and Silverman, B. (1993). On smoothed principal components analysis. *Computational Statistics*, 8:1–16.

Philander, S. (1990). *El Niño, La Niña and the Southern Oscillation.* Academic Press, San Diego.

Poggi, J. (1994). Prévision non paramétrique de la consommation électrique. *Revue de Statistique Appliquée*, XLII(4):83–98.

Pumo, B. (1992). *Estimation et prévision de processus fonctionnels autorégressifs : Application aux processus à temps continu.* PhD thesis, Université Paris VI.

Ramsay, J. (1988). Monotone regression splines in action. *Statistical Sciences*, 3:425–461.

Ramsay, J. and Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society, B*, 53:539–572. with discussion.

Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis.* Springer-Verlag.

Reynolds, R. W. and Smith, T. M. (1994). Improved global sea surface temperature analyses using optimum interpolation. *J. of Climate*, 7:929–948.

Rice, J. and Silverman, B. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, B*, 53:233–243.

Ropelewski, C. and Jones, P. (1987). An extension of the Tahiti-Darwin southern oscillation index. *Mon. Wea. Rev.*, 115:2161–2165.

SAS (1989). *SAS/ETS User's Guide*, first edition. version 6.

Silverman, B. (1996). Smoothed functional principal components analysis by choice of norm. *The Annals of Statistics*, 24:1–24.

Smith, T. M., Reynolds, R., Livezey, R., and Stokes, D. (1996). Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. of Climate*, 9:1403–1420.

Stephenson, D. B. (1997). Correlation of spatial climate/weather maps and the advantages of using the mahalanobis metric in predictions. *Tellus*, 49A.

Tucker, L. R. (1958). Determination of parameters of a functional relationship by factor analysis. *Psychometrika*, 23:19–23.

Vieu, P. (1995). Order choice in nonlinear autoregressive models. *Statistics*, 26:307–328.

Wahba, G. (1990). *Spline Models for Observational Data.* SIAM.