Forecasting with unequally spaced data by a functional principal component approach

Ana M. Aguilera
Departamento de Estadística e Investigación Operativa
Facultad de Ciencias, Universidad de Granada
18071 Granada, Spain. aaguiler@goliat.ugr.es
Francisco A. Ocaña and Mariano J. Valderrama
Departamento de Estadística e Investigación Operativa
Facultad de Farmacia, Universidad de Granada
18071 Granada, Spain. focana@platon.ugr.es valderra@platon.ugr.es

Abstract

The Principal Component Regression model of multiple responses is extended to forecast a continuous-time stochastic process. Orthogonal projection on a subspace of trigonometric functions is applied in order to estimate the principal components using discrete-time observations from a sample of regular curves. The forecasts provided by the this approach are compared with classical principal component regression on simulated data.

Key Words: Principal components, Karhunen-Loève expansion, Least-squares linear prediction, Orthogonal projection.
AMS subject classification: 60G25, 60G12, 65F15.

1 Introduction

There has been a good deal of discussion, in probabilistic and statistical literature, on the pure prediction problem of a continuous-time stochastic process in the future in terms of its recent past. In this line of study, Bosq (1991) has introduced the autoregressive Hilbert (ARH) processes as a generalization of the classical AR processes to random variables with values in a functional Hilbert space. In this paper, taking into account that there are many applications not verifying the hypothesis of these particular models, Functional Principal Component Analysis (FPCA) is applied to reduce the infinite number of predictor and response variables, and to simplify the

This research was supported by Project PB96-1436, Dirección General de Enseñanza Superior, Ministerio de Educación y Cultura, Spain.
Preprint published in **Test (1999) Vol. 8, No. 1, pp. 233-254**

solution. An interesting perspective on the analysis of functional data by PCA can be seen in the recent book by Ramsay and Silverman (1997).

Since Deville (1974), who set up the basic theory on PCA of continuous processes, a large number of papers have focused on the approximation of the FPCA from a set of independent sample paths, particularly in the case of discrete-time observations of these curves as it is usual in practice. An alternative approach is to perform a conventional PCA of the observed data. Castro et al. (1986) were concerned with the effect on the PCA results of unequally spaced sampling points on the sample paths. They used quadrature rules to approximate the FPCA and gave some rather dramatic examples to demonstrate that this functional approach produces stable and consistent estimates of the lead terms of the Karhunen-Loève expansion, contrary to classical PCA that provides unstable estimate of process variability. Besse and Ramsay (1986) and Besse (1991) used interpolating and approximating splines with either Green's functions or reproducing kernels and established a class of eigenfunction models of which conventional PCA is a special case. Based on these works, Besse and Cardot (1994) have developed a forecasting model for ARH(1) processes. Aguilera et al. (1996) have performed a simulation study to show that natural cubic spline interpolation of the sample paths between equally spaced observations gives an optimum PCA approximation and corrects the failings of the classical PCA.

The forecasting technique proposed in this article is based on linear regression of the principal components (p.c.'s) associated to the process in the future against the principal components in the past. It can be seen as an extension of the topic of Principal Component Regression of multiple responses (MPCR, see e.g., Jolliffe (1985)) to predict an infinite set of responses from an infinite set of predictors.

Section 2 presents a brief summary of the MPCR technique. In Section 3 we propose a forecasting model for a continuous-time stochastic process based on its PCA. Section 3 describes the estimation of such forecasting model from a set of independent sample paths. In Section 4, the approximation of the FPCA of the process from discrete-time observations is performed by projection on a finite subspace of trigonometric functions (Aguilera et al., 1995). Finally, in Section 5, model fitting and forecasting results by using this FPCA approach will be analyzed on simulated discrete-time data of the harmonic oscillator stochastic process.

2 Multivariate principal component regression

The MPCR technique solves the problem of predicting a set of response variables, denoted by the random vector $\underline{Y} = (Y_1, \ldots, Y_s)'$, from a set of predictors, denoted by $\underline{X} = (X_1, \ldots, X_m)'$ $(m, s \in \mathbb{Z}^+)$.

The principal feature of this technique is that a double-PCA is performed. That is, a classical PCA is carried out for each set of variables.

Let $\{\xi_1, \ldots, \xi_m\}$ denote the principal components associated to the random vector $\underline{X}$, given by

$$\xi_i = \sum_{j=1}^{m} \phi_{ji}(X_j - \mu_{X_j}) = \underline{\phi_i}' \, (\underline{X} - \underline{\mu}_X), \tag{2.1}$$

where $\underline{\phi_i}$ is the *ith* eigenvector of the covariance matrix of $\underline{X}$, and $\underline{\mu}_X$ is the mean vector of $\underline{X}$.

Similarly, let $\{\eta_1, \ldots, \eta_s\}$ denote the p.c.'s associated to the random vector $\underline{Y}$, defined as

$$\eta_k = \sum_{j=1}^{s} \delta_{jk}(Y_j - \mu_{Y_j}) = \underline{\delta_k}' \, (\underline{Y} - \underline{\mu}_Y), \tag{2.2}$$

where $\underline{\delta_k}$ is the *kth* eigenvector of the covariance matrix of $\underline{Y}$ and $\underline{\mu}_Y$ is the mean vector of $\underline{Y}$.

Then, a MPCR model for predicting each response variable Y_j $(j = 1, \ldots, s)$ is given by

$$\tilde{Y_j} = \mu_{Y_j} + \sum_{k=1}^{q} \hat{\eta}_k \delta_{jk}, \quad q \leq s, \tag{2.3}$$

where $\hat{\eta}_k$ $(k = 1, \ldots, s)$ is the least-squares linear estimate for the p.c. η_k on a subset of the set of p.c.'s $\{\xi_i\}_{i=1}^m$, and μ_{Y_j} is the mean of the variable Y_j.

If in equation (2.3) q equals s, and all the p.c.'s $\{\xi_i\}_{i=1}^m$ are used for estimating η_k, then MPCR will produce the same results as Multivariate Least-Squares Regression (MLSR) but with possibly more accuracy if the original covariance matrix of $\underline{X}$ has inversion problems.

3 Linear prediction based on PCA of continuous time random processes

Let $[T_1, T_2]$ and $[T_3, T_4]$ $(T_3 \geq T_2)$ denote the past and the future intervals, respectively. Let us now consider a second order and quadratic mean continuous random process, $\{X(t)\}$, whose sample functions have squares integrable over the intervals $[T_1, T_2]$ and $[T_3, T_4]$.

The problem is to find a linear estimate of the random variable $X(s)$ $(s \in [T_3, T_4])$ given the variables $\{X(t) : \ t \in [T_1, T_2]\}$. As the estimate proposed in this article will be a function of the principal components of $\{X(t)\}$, we will begin by setting the basic theory about PCA of random processes.

3.1 Basic theory on functional PCA

By analogy with the finite case, the *ith* principal component associated to the process $\{X(t)\}$ in the interval $[T_1, T_2]$ is defined as (Deville (1974))

$$\xi_i = \int_{T_1}^{T_2} (X(t) - \mu(t)) f_i(t) dt, \tag{3.1}$$

where f_i, called the *ith* principal factor (p.f.), is the normalized eigenfunction corresponding to the *ith* largest eigenvalue λ_i of the covariance operator $\mathcal{C}$ defined as

$$\mathcal{C}(f)(t) = \int_{T_1}^{T_2} C(t,s) f(s) ds \quad \forall f \in L^2[T_1, T_2], \tag{3.2}$$

whose kernel is the covariance function $C(t,s)$, and denoting by $\mu(t)$ the mean function of the process.

The p.c.'s defined above have indeed the same optimal properties that in the finite case. That is, ξ_i is the normalized generalized linear combination of the process variables having maximum variance, λ_i, out of all generalized linear combinations which are uncorrelated with $\{\xi_j\}_{j=1}^{i-1}$. Thus, the variance explained by the *ith* principal component is $V_i^P = \lambda_i / V^P$, with $V^P = \sum_i \lambda_i$ being the total variance of the process in the past interval $[T_1, T_2]$.

Then the Karhunen-Loève orthogonal expansion (see, for example, Todorovic (1992)) provides the following principal component decomposition for the process in the past interval:

$$X(t) = \mu(t) + \sum_{i=1}^{\infty} \xi_i f_i(t) \quad t \in [T_1, T_2], \tag{3.3}$$

where the infinite series in (3.3) converges in quadratic mean to X(t) uniformly in t.

Moreover, the series (3.3) truncated in the *qth* term is the best linear approximation of the process $X(t) - \mu(t)$, in the least-squares sense, by a sum of q quasi-deterministic terms (see, for example, Fukunaga (1990)), with $(\sum_{i=1}^{q} \lambda_i)/V^P$ being the variance explained by this linear model and $\sum_{i=q+1}^{\infty} \lambda_i$ being the minimum mean-square error.

Similarly, PCA gives the following orthogonal representation for the process in the future interval:

$$X(s) = \mu(s) + \sum_{j=1}^{\infty} \eta_j g_j(s) \quad s \in [T_3, T_4] \tag{3.4}$$

where g_j and η_j denote the principal factors and components in the interval $[T_3, T_4]$, respectively.

Finally, the total variance of the process in the future is given by $V^F = \sum_j \alpha_j$, where α_j denotes the variance of the p.c. η_j. Therefore, the variance explained by the *jth* p.c. η_j is given by the ratio $V_j^F = \alpha_j / V^F$.

3.2 Principal component prediction models

Let us assume without loss of generality that the process $\{X(t)\}$ is centred and defined on the probabilistic space (Ω, A, P). Let L_X^2 be the closed linear manifold spanned by the r.v.'s $\{X(t) : T_1 \leq t \leq T_2\}$, that is

$$L_X^2 = \overline{\text{Lin}\{X(t) : t \in [T_1, T_2]\}}.$$

Clearly, L_X^2 is a Hilbert subspace of the Hilbert space $L^2(\Omega)$ of the second order r.v.'s on (Ω, A, P).

Let us remember that we are concerned with the linear prediction of $X(s)$ $(s \in [T_3, T_4])$ given the values of the process in the period $[T_1, T_2]$.

The best linear predictor would be the element in L_X^2 closest (in the least-squares sense) to $X(s)$. Therefore, the linear mean-square estimate for $X(s)$ is defined as the random variable $\tilde{X}(s)$ verifying

$$E[\mid \tilde{X}(s) - X(s) \mid^2] = \inf\{E[\mid Z - X(s) \mid^2] : Z \in L_X^2\}.$$

In Hilbert space terminology, $\tilde{X}(s)$ is the orthogonal projection of X(s) on L_X^2, characterized as the element of L_X^2 such that

$$\left(\tilde{X}(s) - X(s)\right) \perp L_X^2,$$

which particularly implies that

$$E[X(s)X(t)] = E[\tilde{X}(s)X(t)] \quad \forall t \in [T_1, T_2].$$

The quantity

$$\epsilon^2(s) = E[\mid \tilde{X}(s) - X(s) \mid^2]$$

is called the mean-square error of the linear prediction for every $s \in [T_3, T_4]$.

It is known (see, for example, Todorovic (1992)) that the p.c.'s $\{\xi_i\}$, associated to the process in the past, make up a complete orthogonal family in L_X^2, verifying $E[\xi_i^2] = \lambda_i$. Therefore, the set $\{Z_i\}$ of normalized p.c.'s defined as $Z_i = \lambda_i^{-1/2}\xi_i$ is an orthonormal complete system in this Hilbert space. As a consequence, the orthogonal projection $P : L^2(\Omega) \to L_X^2$ admits the following expansion, convergent in quadratic mean, in terms of the p.c.'s in the past:

$$P(X) = \sum_{i=1}^{\infty} \frac{E[X\xi_i]}{\lambda_i}\xi_i \quad \forall X \in L^2(\Omega). \tag{3.5}$$

Due to the fact that P is a bounded linear operator on $L^2(\Omega)$ it is easy to prove, from equations (3.4) and (3.5), that the least-squares linear estimate $\tilde{X}$ can be expanded as

$$\tilde{X}(s) = P(X(s)) = \sum_{j=1}^{\infty} \tilde{\eta}_j g_j(s), \quad s \in [T_3, T_4], \tag{3.6}$$

where $\tilde{\eta}_j$ is given by

$$\tilde{\eta}_j = P(\eta_j) = \sum_{i=1}^{\infty} \frac{E[\eta_j\xi_i]}{\lambda_i}\xi_i = \sum_{i=1}^{\infty} \beta_i^j \xi_i \quad j = 1, \ldots. \tag{3.7}$$

Let us observe that $\tilde{\eta}_j$ represents the least-squares linear estimate for the p.c. η_j against the process variables $\{X(t) : T_1 \leq t \leq T_2\}$.

By truncating off each of the infinite series in equation (3.7), the following approximated linear prediction for each of the p.c.'s in the future is obtained

$$\tilde{\eta}_j^{p_j} = \sum_{i=1}^{p_j} \beta_i^j \xi_i. \tag{3.8}$$

Finally, we can construct the following Principal Component Prediction model for the process in the future:

$$\tilde{X}^q(s) = \mu(s) + \sum_{j=1}^{q} \tilde{\eta}_j^{p_j} g_j(s), \quad s \in [T_3, T_4], \tag{3.9}$$

where equation (3.4) has been restricted to the first q p.c.'s in the future whose cumulative variance $CV_j^F = \sum_{j=1}^{q} \alpha_j$ is as close to one as one desires. This model will be denoted by PCP$(q, p_1, \ldots, p_q)$.

Summarizing, from equations (3.8) and (3.9), it is clearly observed that the problem has been reduced to least-squares linear regression for some of the first p.c.'s in the future on some of the first p.c.'s in the past. Let us also observe the similarity between the PCP model (3.9) and the MPCR model (2.3).

Finally, the total mean-square prediction error of the PCP$(q, p_1, \ldots, p_q)$ model is obtained as

$$\begin{aligned} \epsilon^2 &= \int_{T_3}^{T_4} \mathrm{E}\left[\left(X(s) - \tilde{X}^q(s) \right)^2 \right] ds = \sum_{j=1}^{\infty} \alpha_j - \sum_{j=1}^{q} \sum_{i=1}^{p_j} \lambda_i (\beta_i^j)^2 \\ &= \sum_{j=1}^{q} \Upsilon_j^2 + \sum_{l=q+1}^{\infty} \alpha_l, \end{aligned} \tag{3.10}$$

where Υ_j^2 is the mean-square regression error associated to the linear model for $\tilde{\eta}_j^{p_j}$ and given by

$$\Upsilon_j^2 = \mathrm{E}\left[\left(\eta_j - \tilde{\eta}_j^{p_j} \right)^2 \right] = \alpha_j \left(1 - \sum_{i=1}^{p_j} r^2(\eta_j, \xi_i) \right) \in [0, 1].$$

Remark 1: As with the MPCR model, the main problem is to choose the optimum p_j p.c.'s ξ_i to be introduced in the PCP model as predictors for each of the first q p.c.'s η_j in the future. Although the usual practice consists in automatically dropping, as predictors, those p.c.'s associated to the smallest eigenvalues, some authors, as Hötelling and Jolliffe among others (see, for example, Jackson (1992)), pointed out that there is no reason for the p.c.'s with the largest variance to be the best predictors and gave examples where some of the smallest p.c.'s were highly correlated with the response variable. Because of this we will choose those p.c.'s in the past having the highest correlations as the best predictors for each p.c. in the future.

4 Fitting PCP models

Having defined the PCP model, we are now going to estimate it from N independent sample functions of the process in each interval, denoted as

$$\{X_w(t) : t \in [T_1, T_2]\} \quad \text{and} \quad \{X_w(s) : s \in [T_3, T_4]\}; \quad (w = 1, \ldots, N).$$

Then, the identification, estimation and diagnostic of the PCP models are summarized in the following steps:

1. We have to estimate the principal factors of the process in each interval. It is known that the natural estimators of the p.f.'s from a set of independent sample paths are the eigenfunctions of the sample covariance operator $\hat{\mathcal{C}}$ whose kernel is the sample covariance function defined as

$$\hat{C}(t,s) = \frac{1}{N-1} \sum_{w=1}^{N} \left(X_w(t) - \bar{X}(t)\right) \left(X_w(s) - \bar{X}(s)\right), \qquad (4.1)$$

   where $\bar{X}(t)$ is the usual unbiased estimate of the mean $\mu(t)$.

   Then, the sample principal values and factors associated to the process $\{X(t)\}$ in the past and denoted by $\hat{\lambda}_i$ and $\hat{f}_i$, respectively, are the solutions to the second kind integral equation

$$\int_{T_1}^{T_2} \hat{C}(t,s)\hat{f}_i(s)ds = \hat{\lambda}_i \hat{f}_i(t), \quad t \in [T_1, T_2]. \qquad (4.2)$$

The most important sampling properties of these estimates that have been studied in detail by Deville (1973), can be summarized as follows:

(a) $\hat{\mathcal{C}}$ is an unbiased estimate of $\mathcal{C}$.

(b) $\hat{\mathcal{C}}$ is convergent in quadratic mean to $\mathcal{C}$.

(c) For each i, the eigenvalue-eigenfunction pair $(\hat{\lambda}_i, \hat{f}_i)$ of $\hat{\mathcal{C}}$ is convergent in quadratic mean to the corresponding eigenvalue-eigenfunction pair (λ_i, f_i) of $\mathcal{C}$ once the eigenvalues λ_iÿ have been set up in decreasing order.

For simplicity we will suppose that all eigenvalues λ_iÿ are simple. However, the estimator of any multiple eigenvalue is obtained by averaging the corresponding eigenvalues of $\hat{\mathcal{C}}$. A detailed study about the asymptotic distributions of these estimators can be seen in Dauxois et al. (1982).

There are various ways of tackling the problem of solving the eigenequation (4.2). Some numerical techniques involve discrete approximations for which an adequate convergence theory exists. The simplest form involves choosing a sufficiently large number of equally spaced points in $[T_1, T_2]$ and approximating the integrals in (4.2) by the corresponding sums divided by the number of points. This way, the problem becomes identical to that of the classical multivariate PCA. However, more sophisticated quadrature procedures will generally require fewer points as shown in the simulation study using trapezoid quadrature performed by Aguilera et al. (1992). In the next section we are going to summarize briefly the orthogonal projection method for approximating the sample principal factors when the sample paths are only observed at a finite set of knots.

Once the principal factors have been estimated, the *ith* sample p.c. associated to the process in the past $[T_1, T_2]$ is given by

$$\hat{\xi}_{wi} = \int_{T_1}^{T_2} \left(X_w(t) - \bar{X}(t)\right) \hat{f}_i(t)dt, \quad w = 1, \ldots, N. \tag{4.3}$$

Similarly, $\hat{\alpha}_j$, $\hat{g}_j$ and $\hat{\eta}_j$ will denote the sample principal values, factors and components in the future $[T_3, T_4]$, respectively.

2. To choose the number q of p.c.'s $\hat{\eta}_j$ to be introduced in the PCP model as response variables, we will fix a cut-off (somewhere between

80 and 99 percent) and retain the first q p.c.'s whose percentage of cumulative variance is greater or equal to this cut-off.

3. To select the best predictors for each p.c. in the future, we will compute the sample linear correlation between each p.c. in the past and each p.c. in the future. Then, for each p.c. $\hat{\eta}_j \quad (j = 1, \ldots, p)$ the p.c.'s $\hat{\xi}_i$ will be entered in the regression model (3.8) in the order of magnitude of the square of their correlations with the response variable $\hat{\eta}_j$ by following a stepwise regression procedure.

4. Once the response and predictor p.c.'s have been identified, we will estimate, in the usual way, the linear regression model for each p.c. $\hat{\eta}_j$ $(j = 1, \ldots, q)$ against its predictors $\hat{\xi}_i$ $(i = 1, \ldots, p_j)$

$$\tilde{\hat{\eta}}_j^{p_j} = \sum_{i=1}^{p_j} \hat{\beta}_i^j \hat{\xi}_i$$

where the vector of regression coefficients $\underline{\hat{\beta}}^j$ is given by

$$\underline{\hat{\beta}}^j = \left[\boldsymbol{\Xi}'\boldsymbol{\Xi}\right]^{-1} \boldsymbol{\Xi}' \underline{\hat{\eta}}_j \tag{4.4}$$

and

- $\boldsymbol{\Xi} = \left(\underline{\hat{\xi}}_1, \ldots, \underline{\hat{\xi}}_{p_j}\right)$ is the $N \times p_j$ dimension regression matrix whose columns are the vectors

$$\underline{\hat{\xi}}_i = \left(\hat{\xi}_{1i}, \ldots, \hat{\xi}_{Ni}\right)' \quad (i = 1, \ldots, p_j).$$

- $\underline{\hat{\eta}}_j$ is the $N \times 1$ column vector whose components are the sample values of the p.c. $\hat{\eta}_j$. That is,

$$\underline{\hat{\eta}}_j = (\hat{\eta}_{1j}, \ldots, \hat{\eta}_{Nj})'.$$

Let us observe that the matrix $\boldsymbol{\Xi}'\boldsymbol{\Xi}$ is diagonal, so the elements of its inverse are merely the reciprocals of its diagonal elements.

5. We will estimate the identified PCP model as

$$\tilde{\hat{X}}^q(s) = \bar{X}(s) + \sum_{j=1}^{q} \tilde{\hat{\eta}}_j^{p_j} \hat{g}_j(s) \quad s \in [T_3, T_4]. \tag{4.5}$$

For every new individual w observed only in the past interval $[T_1, T_2]$, we obtain a forecasting $\tilde{X}_w^q(s)$ for every $s \in [T_3, T_4]$, after estimating its principal components in the past and replacing in equation (4.5).

6. Finally, to evaluate the adequacy of the estimated PCP model we estimate the mean-square error as

$$\hat{\epsilon}^2(s) = \frac{1}{N-1} \sum_{w=1}^{N} \left(X_w(s) - \tilde{X}^q(s) \right)^2 \quad s \in [T_3, T_4]. \tag{4.6}$$

5 Approximation from discrete data

In many applied contexts we have a data set corresponding to unequally spaced discrete-time observations, denoted by

$$\{X_w(t_j) : w = 1, \ldots, N; j = 0, 1, ..., m\},$$

from a collection of continuous functions, $\{X_w(t) : w = 1, \ldots, N\}$, which can be modelled as independent sample paths of a continuous-time stochastic process $\{X(t)\}$.

To forecast the process for $s \in [t_{k+1}, t_m]$, $(k : 1, \ldots, m-2)$, we will consider in the observed period $[t_0, t_m]$ two intervals, $[t_0, t_k]$ and $[t_{k+1}, t_m]$, representing the past and the future, respectively, and denoting $T_1 = t_0$, $T_2 = t_k$, $T_3 = t_{k+1}$ and $T_4 = t_m$.

As we have already discussed in the introduction, the sample principal factors have to be approximated in each interval by using efficient numerical techniques. In the simulation study performed in the next section the orthogonal projection method (OPM) proposed by Aguilera et al. (1995) is going to be applied. This numerical procedure approximates the principal factors by those of a stochastic process whose sample paths are the orthogonal projections of the original ones on a finite-dimension subspace E_n of $L^2[T_1, T_2]$.

Let $\{e_j\}_{j \in \mathbb{N}}$ be an orthonormal basis in the separable Hilbert space $L^2[T_1, T_2]$ and let E_n the subspace spanned by the first n functions e_j. The orthogonal projection method approximates the eigensystem of the sample covariance operator $\hat{\mathcal{C}}$ by means of the eigensystem of the sample covariance operator associated with the projected process and defined by

$$\hat{\mathcal{C}}_n = P_n \hat{\mathcal{C}} P_n,$$

where P_n denotes the orthogonal projection on E_n.

Then the *ith* approximated principal factor denoted as $\hat{f}_i^{(n)}$ is given by

$$\hat{f}_i^{(n)} = \sum_{j=1}^{n} \gamma_{ji} e_j \tag{5.1}$$

where the vector $\underline{\gamma}_i$, whose components are γ_{ji}, is a solution to the matrix eigenequation

$$\mathbf{R}\underline{\gamma}_i = \hat{\lambda}_i^{(n)} \underline{\gamma}_i,$$

with $\mathbf{R}$ being the $n \times n$ matrix whose elements are given by

$$R_{ij} = \int_{T_1}^{T_2} \int_{T_1}^{T_2} \hat{C}(t,s) e_i(t) e_j(s) dt ds = \frac{1}{N-1} \sum_{w=1}^{N} (Y_{wi} - \bar{Y}_i)(Y_{wj} - \bar{Y}_j)$$

and defining,

$$Y_{wj} = \int_{T_1}^{T_2} X_w(s) e_j(s) ds, \quad \forall j = 1 \ldots n, \tag{5.2}$$

$$\bar{Y}_j = \frac{1}{N} \sum_{w=1}^{N} Y_{wj} = \int_{T_1}^{T_2} \bar{X}(s) e_j(s) ds.$$

Finally, the approximated principal components are given by

$$\hat{\xi}_{wi}^{(n)} = \sum_{j=1}^{n} \gamma_{ji} \left(Y_{wj} - \bar{Y}_j \right) \quad w = 1, \ldots, N. \tag{5.3}$$

Let us observe that the approximated p.c.'s $\hat{\xi}^{(n)}$ are exactly those of the data matrix of integrals $\mathbf{Y}$ whose elements are defined in equation (5.2).

Remark 2: The properties of the sample covariance operator $\hat{\mathcal{C}}$ guarantee the convergence of the approximations given by the OPM approach. As $\{E_n\}_{n \in \mathbb{N}}$ approximates $L^2[T_1, T_2]$ in the sense of $lim_{n \to \infty} P_n(f) = f$ for all $f \in L^2[T_1, T_2]$, and $\hat{\mathcal{C}}$ is a compact, hermitian, and positive operator, then it is verified that for each i, there exists an eigenvalue-eigenfunction pair $(\hat{\lambda}_i^{(n)}, \hat{f}_i^{(n)})$ of $\hat{\mathcal{C}}_n$ such that the sequence $\{\hat{\lambda}_i^{(n)}\}$ converges to $\hat{\lambda}_i$, and the sequence $\{\hat{f}_i^{(n)}\}$ converges to $\hat{f}_i$ in norm (Riesz and Sz-Nagy, 1990).

Remark 3: The approximating subspace E_n has to be selected according to the nature of the process sample paths. For point processes, for example,

E_n will be the subspace of piecewise constant functions on a fixed partition in $[T_1, T_2]$. In the next section we are going to use a subspace of normalized trigonometrics functions for a stochastic process with regular sample paths.

Remark 4: The computations in the next section have been performed by using the SMCP^2 program coded in Turbo Pascal and developed by the authors. Let us observe that in order to apply the orthogonal projection method it is necessary to approximate the inner products given by expresion (5.2). SMCP^2 approximates these integrals by Romberg's integration after using natural cubic splines interpolating the original sample paths on the observed time points in order to compute their values at the knots of discretization. The cubic splines are computed by using a recursive definition of cubic B-splines which is numerically stable.

6 Comparing predictions on simulated data

In this section, we are going to test the accuracy of the proposed principal component forecasting approach on unequally spaced simulated data from the Gaussian harmonic oscillator stochastic process defined by

$$O(t) = R\cos[2\pi t + \theta] \tag{6.1}$$

where R and θ are independent random variables, R has Rayleigh distribution with parameter σ, and θ is uniformly distributed on the interval [0,2π].

The covariance kernel of this process in an interval $[T_1, T_2]$ has a unique eigenvalue with multiplicity two, given by $\lambda = (T_2 - T_1)\sigma/2$.

In addition, if T_1 and T_2 are integers then the following functions

$$\begin{aligned} f_1(t) &= \left(\frac{2}{T_2 - T_1}\right)^{1/2} \sin(2\pi t) \\ f_2(t) &= \left(\frac{2}{T_2 - T_1}\right)^{1/2} \cos(2\pi t) \end{aligned} \tag{6.2}$$

make up an orthonormal basis of the eigensubspace associated to λ.

In this paper our objective is to forecast the process $O(t)$ in the interval [3,5] from its evolution in the interval [0,3]. In order to build a PCP

model we have simulated thirty sample paths of this process at discrete-time points, twenty two have been used to fit the model and the final eight have been held back for post-sample forecasting comparison. These simulations have been performed from thirty observations randomly generated of a Rayleigh distribution with parameter 0.5 and an uniform distribution on $[0,2\pi]$. The thirty sample paths have been simulated at thirteen time points in the past interval [0,3] and nine time points in the future interval [3,5]. Two different sets of observation points have been considered: unequally spaced time points generated by an uniform distribution in each interval (case I) and unequally spaced knots obtained by adding a Gaussian noise, with mean cero and standard deviation 0.25, to equally spaced time points in each interval (case II).

PCP models are not directly comparable with MPCR due to the following reasons among others:

1. From a theoretic point of view, a MPCR model approximates the least-squares linear estimate for $O(s)$ on a finite set of variables (the discrete-time simulated observations in the past) whereas a PCP model approximates the least-squares linear estimate for $O(s)$ on an infinite set of variables (the continuous-time process values in the past).

2. The OPM approach projects the original sample paths on a subspace of trigonometric functions and performs a FPCA of the smoothed sample curves whereas classic PCA analyzes directly the crude observed data.

3. The goodness of fit of the linear regression models for the p.c.'s in the future against the p.c.'s in the past depends on the linear correlations between p.c.'s that could be very different for the two forecasting approaches as it is shown in the examples provided in this section.

However we are going to compare the predictions at the discretization time knots in the interval [3,5].

Firstly, we have approximated the sample principal factors and components, in each interval, by using OPM on a subspace of ten trigonometric functions for each case. We have also performed a classic PCA of the discrete data in each interval for each case. The exact and approximated

principal values in each period appear in Table 1. As the harmonic oscillator process has a double eigenvalue, its estimation is obtained by averaging the first two sample principal values. As each of the two principal components of the process $\{O(t)\}$ explains a 50% of its total variance, the PCP model must be constructed with the two principal components in the future as response variables.

Past Interval: [0,3]				
	Exact	Case	OPM	Classical PCA
$\hat{\lambda}_1$		I	0.72918	3.72522
		II	0.84012	5.08570
$\hat{\lambda}_2$		I	0.57407	3.32059
		II	0.55287	2.32550
$\hat{\lambda}$	0.75	I	0.65163	3.52291
		II	0.69650	3.70560
Future Interval: [3,5]				
	Exact	Case	OPM	Classical PCA
$\hat{\lambda}_1$		I	0.48660	2.98035
		II	0.51378	2.56250
$\hat{\lambda}_2$		I	0.44596	1.82405
		II	0.36600	2.30790
$\hat{\lambda}$	0.5	I	0.46628	2.40220
		II	0.43989	2.43520

Table 1: Exact and approximated principal values for the past and the future intervals.

Secondly, we have estimated the linear correlations between the principal components in the two periods. The squares of these correlations appear in Table 2 for the OPM and classic PCA approaches. Let us observe that the first and the second p.c.'s in the future are highly correlated with the first and the second p.c.'s in the past, respectively, for the OPM approach in the two cases whereas the correlations between p.c.'s for the classic PCA are very different.