MONASH University

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Nonparametric time series forecasting with dynamic updating

Han Lin Shang and Rob J Hyndman

August 2009

Nonparametric time series forecasting with dynamic updating

Han Lin Shang
Department of Econometrics and Business Statistics,
Monash University, VIC 3800
Australia.
Email: HanLin.Shang@buseco.monash.edu.au

Rob J Hyndman
Department of Econometrics and Business Statistics,
Monash University, VIC 3800
Australia.
Email: Rob.Hyndman@buseco.monash.edu.au

28 August 2009

Nonparametric time series forecasting with dynamic updating

Abstract

We present a nonparametric method to forecast a seasonal univariate time series, and propose four dynamic updating methods to improve point forecast accuracy. Our methods consider a seasonal univariate time series as a functional time series. We propose first to reduce the dimensionality by applying functional principal component analysis to the historical observations, and then to use univariate time series forecasting and functional principal component regression techniques. When data in the most recent year are partially observed, we improve point forecast accuracy using dynamic updating methods. We also introduce a nonparametric approach to construct prediction intervals of updated forecasts, and compare the empirical coverage probability with an existing parametric method. Our approaches are data-driven and computationally fast, and hence they are feasible to be applied in real time high frequency dynamic updating. The methods are demonstrated using monthly sea surface temperatures from 1950 to 2008.

1 Introduction

We consider how to forecast a functional time series when the most recent curve is partially observed. This situation arises most frequently when a seasonal univariate time series is sliced into segments and treated as a time series of functions. The idea of forming a functional time series from a seasonal univariate time series has been considered by several authors, including Besse et al. (2000), Antoniadis & Sapatinas (2003), Ferraty & Vieu (2006, Chapter.12), Aneiros-Pérez & Vieu (2008) and Antoch et al. (2008). However, little attention has been given to the practical problem of forecasting when the final curve is incompletely observed.

Let $\{Z_w, w \in [1, N]\}$ be a seasonal univariate time series which has been observed at N equispaced times, where the seasonality is of length p. When the seasonal pattern is strong, one way to model the time series nonparametrically is to use ideas from functional data analysis (Ramsay & Silverman 2005). We divide the observed time series into n trajectories each of length p, and then consider each trajectory as a curve rather than as p distinct points. The functional time series is then given by

$$y_t(x) = \{Z_w, w = \in (p(t-1), pt]\}, \quad t = 1, \cdots, n, \quad 1 < x < p.$$

The usual problem of interest is to forecast $y_{n+h}(x)$, the data in year $n + h$, from the observed data, $\{y_1(x), \ldots, y_n(x)\}$. For example, in Section 2 we consider $\{Z_w\}$ to be monthly sea surface temperatures from 1950 to 2008, so that $p = 12$ and $N = 59 \times 12 = 708$, and we are interested in forecasting sea-surface temperatures in 2009 and beyond.

When $N = np$, all trajectories are complete, and forecasting is straightforward with several available methods. These techniques include the functional autoregressive of order 1 (Bosq 2000, Bosq & Blanke 2007), functional kernel regression (Ferraty & Vieu 2006, Aneiros-Pérez & Vieu 2008), functional principal component regression (Hyndman & Ullah 2007, Hyndman & Booth 2008, Hyndman & Shang 2009), and functional partial least squares regression (Preda & Saporta 2005a,b). In this article, we consider the problem of forecasting when the last trajectory is incomplete. We call this "dynamic updating" and we propose and compare four possible dynamic updating methods.

Our methods are all based on functional principal component analysis (FPCA), as described in Section 3. When all trajectories are complete, FPCA allows a decomposition of the historical data, $\{y_t(x), t = 1, \cdots, n\}$, into a number of functional principal components and their uncorrelated principal component scores. To forecast the principal component scores, one can employ univariate time series (TS) models on the historical principal component scores. Conditioning on the historical

observations and fixed functional principal components, the point forecasts are obtained by multiplying the forecasted principal component scores with the fixed functional principal components. Since this method uses univariate time series forecasts, we call it the "TS method".

We introduce four dynamic updating methods in Section 4 to deal with the situation when the most recent curve is partially observed. These are the called the block moving (BM), ridge regression (RR), ordinary least squares (OLS) and penalized least squares (PLS) methods.

Distributional forecasts are discussed in Section 5, including a new nonparametric approach to construct prediction intervals for the TS, BM and PLS methods. Our methods are illustrated using the monthly sea surface temperature data set described in Section 2.

Conclusions are discussed in Section 6, along with some thoughts on how the methods developed here might be further extended.

2 Data set

As a vehicle of illustration, we consider the monthly sea surface temperatures from January 1950 to December 2008, available online at http://www.cpc.noaa.gov/data/indices/sstoi.indices. These averaged sea surface temperatures are measured by moored buoys in the "Niño region" defined by the coordinate $0 - 10°$ South and $90 - 80°$ West. A univariate time series display is given in Figure 1a, with the same data shown in Figure 1b as a time series of functions.
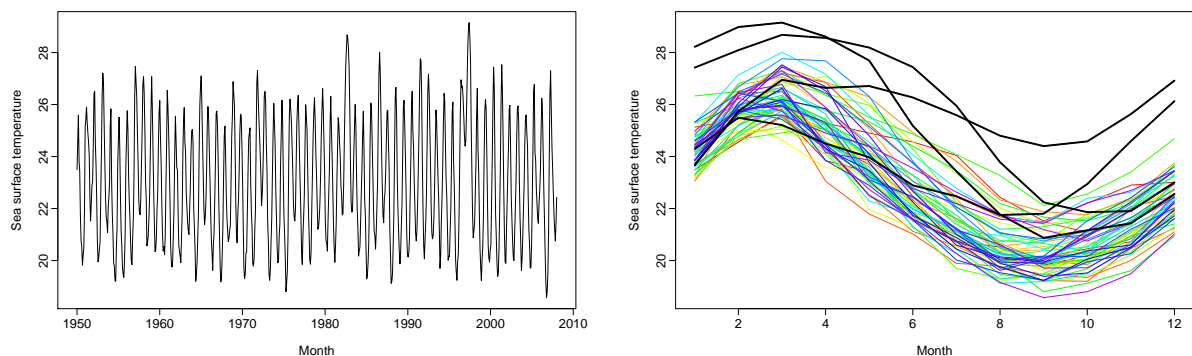
(a) *A univariate time series display of the monthly sea surface temperatures.*

(b) *A functional time series display of the monthly sea surface temperatures.*

Figure 1: *Exploratory plots suggesting that both predictive regularity and abnormality are presented in the sea surface temperature data set from Jan 1950 to Dec 2008 measured by moored buoys in the region defined by the coordinate $0 - 10°$ South and $90 - 80°$ West.*

From Figure 1b, there are some years showing extreme sea surface temperatures and are suspected

to be outliers. Since the presence of outliers can seriously affect the performance of modeling and forecasting, we applied the outlier detection method of Hyndman & Shang (2008) and identified four outliers. These outliers correspond to the years 1982–1983 and 1997–1998, highlighted by the thick black lines in Figure 1b. The sea surface temperatures during 1982–1983 began in June 1982 with a moderate increase, which was followed by abnormal increases between September 1982 and June 1983 (Moran et al. 2006, Timmermann et al. 1999). The sea surface temperatures during 1997–1998 were also unusual and became extreme in the latter half of 1997, and stayed high for the early part of 1998. Dioses et al. (2002) reported that the northern central region of Peru was strongly affected because warm waters with low salinity approached the coast, while the southern region was influenced more by oceanic waters. These detected outliers have consequently been removed from further analysis.

3 Forecasting method

Our forecasting method utilizes FPCA, which plays a central role in the development of functional data analysis. An account of the statistical properties of FPCA, along with applications of the methodology, are given by Ramsay & Silverman (2002), Ramsay & Silverman (2005) and Ferraty & Vieu (2006). Papers covering the development of FPCA include those of Rice & Silverman (1991), Silverman (1995), Silverman (1996), Reiss & Ogden (2007), Hyndman & Ullah (2007), Hyndman & Shang (2009) and Shen (2009). Significant treatments of the theory of FPCA are given by Dauxois et al. (1982), Cai & Hall (2006), Hall & Hosseini-Nasab (2006), Hall et al. (2006), Hall & Horowitz (2007), Hall & Hosseini-Nasab (2009), and Delaigle et al. (2009).

In this section, we assume that all trajectories are complete. Our forecasting method begins with decentralizing the functional data by subtracting the functional mean. The functional mean $\mu(x)$ is estimated by

$$\hat{\mu}(x) = \frac{1}{n} \sum_{t=1}^{n} y_t(x).$$

If one seeks a robust estimator, then the L_1 median of data should be used, and is denoted by

$$\hat{\mu}(x) = \arg\min_{\theta(x)} \sum_{t=1}^{n} \left\| y_t(x) - \theta_t(x) \right\|,$$

where $\|g(u)\| = (\int g^2(u)du)^{1/2}$. The algorithm of Hössjer & Croux (1995) can be used to compute $\hat{\mu}(x)$.

Using FPCA, $\{y_t; t = 1, \cdots, n\}$ can be approximated by the sum of orthogonal functional principal components and their associated principal component scores:

$$y_t(x) = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\phi}_k(x)\hat{\beta}_{k,t} + \epsilon_t(x), \tag{1}$$

where $\{\hat{\phi}_1(x), \cdots, \hat{\phi}_K(x)\}$ are the estimated functional principal components, $\{\hat{\beta}_{1,t}, \cdots, \hat{\beta}_{K,t}\}$ are uncorrelated principal component scores, $\epsilon_t(x)$ is the zero-mean residual function, and $K < n$ is the number of functional principal components.

3.1 Point forecasts

Because the principal component scores are uncorrelated to each other, it is appropriate to forecast each series $\{\hat{\beta}_{k,1}, \cdots, \hat{\beta}_{k,n}\}$ using univariate time series models, such as the state-space models (Harvey 1990), ARIMA models (Box et al. 2008), or exponential smoothing state-space models (Hyndman et al. 2008). It is noteworthy that although the lagged cross correlations are not necessarily zero, they are likely to be small because the contemporaneous correlations are zero (Hyndman & Ullah 2007, Shen & Huang 2008, Shen 2009).

Based on the historical observations ($\mathscr{I}$) and the functional principal components $\Phi = \{\hat{\phi}_1(x), \cdots, \hat{\phi}_K(x)\}$, the forecasted curves are expressed as

$$\hat{y}_{n+h|n}^{\text{TS}}(x) = \text{E}[y_{n+h}(x) \mid \mathscr{I}, \Phi] = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\phi}_k(x)\hat{\beta}_{k,n+h|n}^{\text{TS}}, \tag{2}$$

where $\hat{\beta}_{k,n+h|n}^{\text{TS}}$ denotes an h-step-ahead forecast of $\beta_{k,n+h}$.

3.2 Component selection

Hyndman & Booth (2008) found that the point forecasts are insensitive to the choice of K, provided that K is large enough. Although there is a computational difficulty in choosing a large K, a small K may result in a poor forecast accuracy. Consequently, in the analysis we choose $K = 6$, which should be larger than any of the components really require. In the context of overparametrized regression problems, Greenshtein & Ritov (2004) and Greenshtein (2006) described this phenomenon as "persistence in high-dimensional linear predictor selection".

4 Updating point forecasts

When the functional data are segments of a univariate time series, the most recent trajectory is observed sequentially. When we have observed the first m_0 time periods of $y_{n+1}(x)$, denoted by $y_{n+1}(x_e) = [y_{n+1}(x_1), \cdots, y_{n+1}(x_{m_0})]'$, we are interested in forecasting the data in the remainder of year $n + 1$, denoted by $y_{n+1}(x_l)$. However, the TS method described in Section 3 does not utilize the most recent data, namely the partially observed trajectory. Instead, using (2), the time series forecast of $y_{n+1}(x_l)$ is given by

$$\hat{y}_{n+1|n}^{\text{TS}}(x_l) = \text{E}[y_{n+1}(x_l) \mid \mathscr{I}_l, \Phi_l] = \hat{\mu}(x_l) + \sum_{k=1}^{K} \hat{\phi}_k(x_l)\hat{\beta}_{k,n+1|n}^{\text{TS}}, \quad \text{for} \quad m_0 < l \le p,$$

where $\hat{\mu}(x_l)$ is the mean function corresponding to the remaining time periods, $\mathscr{I}_l$ denotes the historical data corresponding to the remaining time periods, and $\Phi_l = \{\hat{\phi}_1(x_l), \cdots, \hat{\phi}_K(x_l)\}$ are the estimated functional principal components corresponding to the remaining time periods.

In order to improve point forecast accuracy, it is desirable to dynamically update the point forecasts for the remaining time periods of the year $n+1$ by using the information from the partially observed data. To address this issue, we shall introduce four dynamic updating methods.

4.1 Block moving method

The block moving (BM) method simply redefines the start and end points of our "year" (the time for a single trajectory). Because time is a continuous variable, we can change the support of our trajectories from $[1, p]$ to $[m_0 + 1, p] \cup [1, m_0]$. Then the first trajectory $y_1(x)$ becomes partially observed as it is only observed on $1 < x < m_0$.

The redefined data are shown diagrammatically in Figure 2 where the bottom box has moved to become the top box. The colored region shows the data lost in the first year. The partially observed last trajectory under the old "year" completes the last trajectory under the new year.

The TS method can be applied to the new complete data block. The loss of data will have minimal effect on the forecasts, if the number of curves is large.

4.2 Ordinary least squares

We can estimate the remaining part of the last trajectory using a regression based on the principal components obtained in (1). Let $\mathscr{F}_e$ be the $m_0 \times K$ matrix whose (j, k)th entry is $\hat{\phi}_k(x_j)$ for $1 \le j \le$

m_0 and $1 \leq k \leq K$. Let $\boldsymbol{\beta}_{n+1} = [\beta_{1,n+1}, \cdots, \beta_{K,n+1}]'$, and $\boldsymbol{\epsilon}_{n+1}(x_e) = [\epsilon_{n+1}(x_1), \cdots, \epsilon_{n+1}(x_{m_0})]'$. As the mean-adjusted $\hat{\boldsymbol{y}}_{n+1}^*(x_e) = \boldsymbol{y}_{n+1}(x_e) - \hat{\mu}(x_e)$ becomes available, we have a regression equation expressed as

$$\hat{\boldsymbol{y}}_{n+1}^*(x_e) = \mathscr{F}_e \boldsymbol{\beta}_{n+1} + \boldsymbol{\epsilon}_{n+1}(x_e).$$

The $\boldsymbol{\beta}_{n+1}$ can be estimated via ordinary least squares giving

$$\hat{\boldsymbol{\beta}}_{n+1}^{\text{OLS}} = \left(\mathscr{F}_e' \mathscr{F}_e\right)^{-1} \mathscr{F}_e' \hat{\boldsymbol{y}}_{n+1}^*(x_e).$$

The OLS forecast of $y_{n+1}(x_l)$ is then given by

$$\hat{y}_{n+1}^{\text{OLS}}(x_l) = \text{E}[y_{n+1}(x_l) \mid \mathscr{I}_l, \Phi_l] = \hat{\mu}(x_l) + \sum_{k=1}^{K} \hat{\phi}_k(x_l) \hat{\beta}_{k,n+1}^{\text{OLS}}.$$

4.3 Ridge regression

The OLS method uses the partially observed data in the most recent curve to improve point forecast accuracy for the remainder of year $n+1$, but it needs a sufficiently large number of observations (at least equal to K) in order for $\hat{\beta}_{n+1}^{\text{OLS}} = \{\hat{\beta}_{1,n+1}^{\text{OLS}}, \cdots, \hat{\beta}_{K,n+1}^{\text{OLS}}\}$ to be numerically stable. To address this
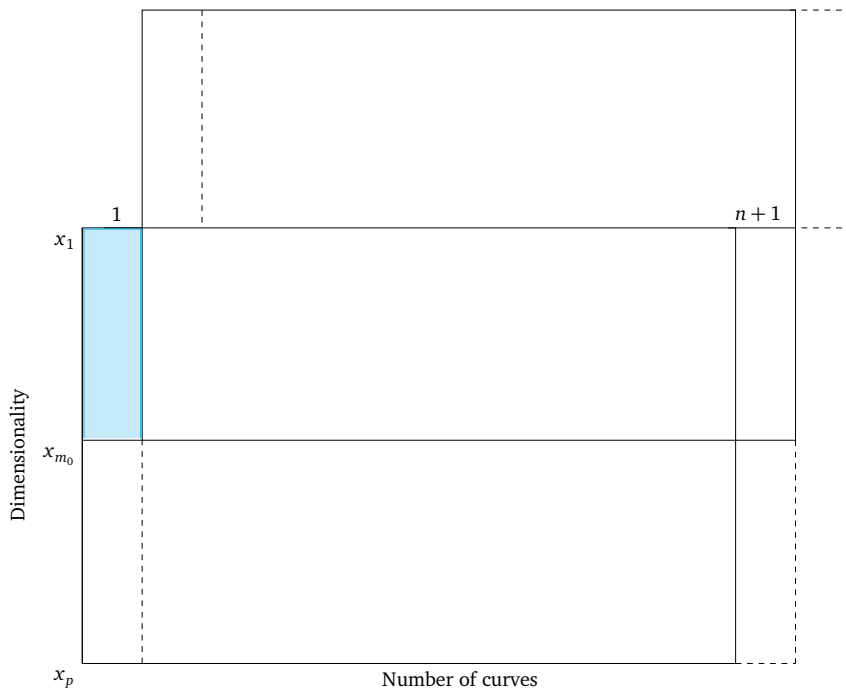

Figure 2: *Dynamic update via the block moving approach. The colored region shows the data loss in the first year. The forecasts for the remaining months in year $n+1$ can be updated by the forecasts using the TS method applied to the upper block.*

problem, we adapt the ridge regression (RR) method of Hoerl & Kennard (1970) with the predictors being the corresponding principal components and the partially observed data being the responses. The main advantage of RR is that it uses a square penalty function, which is a rotationally invariant hypersphere centered at the origin (Izenman 2008). Two-dimensional contours of the different penalty functions are presented in Figure 3.

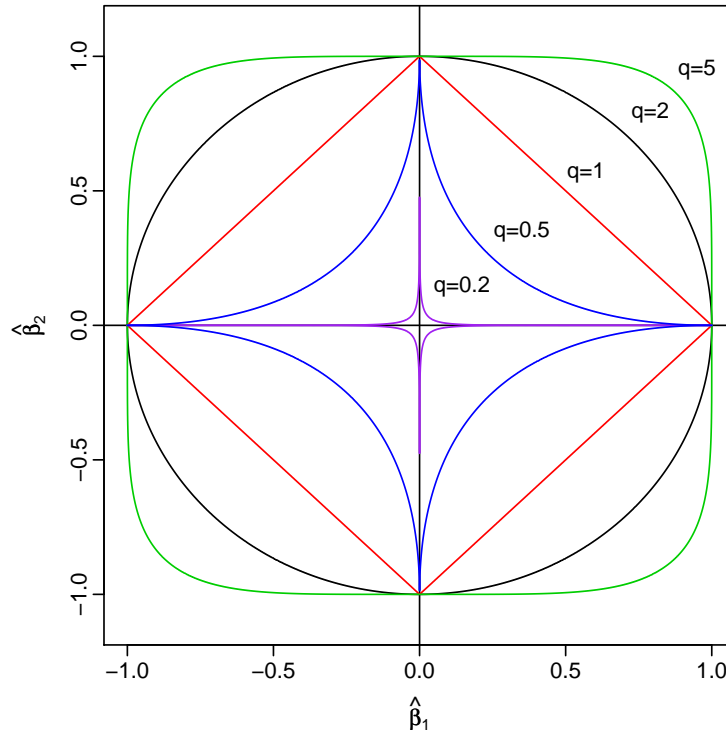**Figure 3:** *Two-dimensional contours of the symmetric penalty function $p_q(\boldsymbol{\beta}) = |\beta_1|^q + |\beta_2|^q = 1$ for $q = 0.2, 0.5, 1, 2, 5$. When $q = 2$, the square penalty function is rotationally invariant hypersphere centered at the origin.*

The RR method shrinks the regression coefficient estimates towards zero. The RR coefficient estimates are obtained by minimizing a penalized residual sum of squares

$$\arg\min_{\boldsymbol{\beta}_{n+1}} \left\{ (\hat{\boldsymbol{y}}^*_{n+1}(x_e) - \mathscr{F}_e \boldsymbol{\beta}_{n+1})'(\hat{\boldsymbol{y}}^*_{n+1}(x_e) - \mathscr{F}_e \boldsymbol{\beta}_{n+1}) + \lambda \boldsymbol{\beta}'_{n+1} \boldsymbol{\beta}_{n+1} \right\}, \tag{3}$$

where $\lambda > 0$ is a tuning parameter that controls the amount of shrinkage. By taking the first derivative with respect to $\boldsymbol{\beta}_{n+1}$ in (3), we obtain

$$\hat{\boldsymbol{\beta}}^{\text{RR}}_{n+1} = \left(\mathscr{F}_e' \mathscr{F}_e + \lambda \boldsymbol{I}_K \right)^{-1} \mathscr{F}_e' \hat{\boldsymbol{y}}^*_{n+1}(x_e),$$

where I_K is the $K \times K$ identity matrix. When $\lambda = 0$, this reduces to the unregularized OLS regression coefficient estimates, provided that $(\mathscr{F}_e'\mathscr{F}_e)^{-1}$ exists. As λ approaches infinity, the regression coefficient estimates tend toward zero.

The RR forecast of $y_{n+1}(x_l)$ is given by

$$\hat{y}^{\text{RR}}_{n+1}(x_l) = \text{E}[y_{n+1}(x_l) \mid \mathscr{I}_l, \Phi_l] = \hat{\mu}(x_l) + \sum_{k=1}^{K} \hat{\phi}_k(x_l)\hat{\beta}^{\text{RR}}_{k,n+1}.$$

4.4 Penalized least squares

Although the RR method solves the potential singularity problem, it does not take account of the TS forecasted regression coefficient estimates, $\hat{\beta}^{\text{TS}}_{n+1|n}$. This motivates the development of the PLS method (Shen & Huang 2008, Shen 2009), in which the regression coefficients are selected by shrinking them toward $\hat{\beta}^{\text{TS}}_{n+1|n}$. The PLS regression coefficients minimize a penalized residual sum of squares

$$\left(\hat{y}^*_{n+1}(x_e) - \mathscr{F}_e\hat{\beta}_{n+1}\right)'\left(\hat{y}^*_{n+1}(x_e) - \mathscr{F}_e\hat{\beta}_{n+1}\right) + \lambda\left(\hat{\beta}_{n+1} - \hat{\beta}^{\text{TS}}_{n+1|n}\right)'\left(\hat{\beta}_{n+1} - \hat{\beta}^{\text{TS}}_{n+1|n}\right). \tag{4}$$

The first term in (4) measures the goodness of fit, while the second term penalizes the departure of the regression coefficient estimates from the TS forecasted regression coefficient estimates. The $\hat{\beta}^{\text{PLS}}_{n+1}$ obtained can thus be seen as a tradeoff between these two terms, subject to a penalty parameter λ.

By taking the first derivative with respect to $\hat{\beta}_{n+1}$ in (4), we obtain

$$\hat{\beta}^{\text{PLS}}_{n+1} = \left(\mathscr{F}_e'\mathscr{F}_e + \lambda I_K\right)^{-1}\left(\mathscr{F}_e'\hat{y}^*_{n+1}(x_e) + \lambda\hat{\beta}^{\text{TS}}_{n+1|n}\right). \tag{5}$$

When the penalty parameter $\lambda \to 0$, $\hat{\beta}^{\text{PLS}}_{n+1}$ approaches $\hat{\beta}^{\text{OLS}}_{n+1}$; when $\lambda \to \infty$, $\hat{\beta}^{\text{PLS}}_{n+1}$ approaches $\hat{\beta}^{\text{TS}}_{n+1|n}$; when $0 < \lambda < \infty$, $\hat{\beta}^{\text{PLS}}_{n+1}$ is a weighted average between $\hat{\beta}^{\text{TS}}_{n+1|n}$ and $\hat{\beta}^{\text{OLS}}_{n+1}$.

The PLS forecast of $y_{n+1}(x_l)$ is given by

$$\hat{y}^{\text{PLS}}_{n+1}(x_l) = \text{E}[y_{n+1}(x_l) \mid \mathscr{I}_l, \Phi_l] = \hat{\mu}(x_l) + \sum_{k=1}^{K} \hat{\phi}_k(x_l)\hat{\beta}^{\text{PLS}}_{k,n+1}.$$

4.5 Penalty parameter selection

We split the data into a training sample (including sea surface temperatures from 1950 to 1992 excluding the outliers) and a testing sample (including sea surface temperatures from 1993 to 2008

excluding the outliers). Within the training sample, we further split the data into a training set (including sea surface temperatures from 1950 to 1970) and a validation set (including sea surface temperatures from 1971 to 1992 excluding the outliers). The optimal values of λ for different updating periods are determined by minimizing the MAE and MSE criteria, expressed as

$$\text{MAE} = \frac{1}{hp} \sum_{j=1}^{h} \sum_{i=1}^{p} \left| y_{n+j}(x_i) - \hat{y}_{n+j}(x_i) \right|,$$

$$\text{and} \qquad \text{MSE} = \frac{1}{hp} \sum_{j=1}^{h} \sum_{i=1}^{p} \left[y_{n+j}(x_i) - \hat{y}_{n+j}(x_i) \right]^2,$$

within the validation set. In Table 1, the optimal tuning parameters for different updating periods are given for both the PLS and RR methods.

Updating period	Minimum MSE		Minimum MAE	
	PLS	RR	PLS	RR
Mar–Dec	908.43	0.00	1118.58	0.00
Apr–Dec	335.40	3.11	197.66	3.35
May–Dec	233.53	8.99	245.64	7.34
Jun–Dec	111.85	11.00	138.92	8.29
Jul–Dec	7.47	6.23	4.86	4.91
Aug–Dec	27.90	11.62	18.42	7.61
Sep–Dec	279.05	15.59	197.80	10.50
Oct–Dec	9.01	4.60	7.77	5.41
Nov–Dec	3.29	0.73	4.82	1.44
Dec	3.60	1.74	8.25	2.33

Table 1: *For different updating periods, the optimal tuning parameters are determined by minimizing the MSE and MAE criteria within the validation set.*

5 Distributional forecast methods

Prediction intervals are a valuable tool for assessing the probabilistic uncertainty associated with point forecasts. As emphasized in Chatfield (1993, 2000), it is important to provide interval forecasts as well as point forecasts so as to

1. assess future uncertainty;

2. enable different strategies to be planned for a range of possible outcomes indicated by the interval forecasts;

3. compare forecasts from different methods more thoroughly; and

4. explore different scenarios based on different assumptions.

In our forecasting method, there are two sources of errors that need to be taken into account: errors in estimating the regression coefficient estimates and errors in the model residuals. First, in Sections 5.1 and 5.2, we describe two methods for constructing prediction intervals for the TS method. Then in Section 5.3 we show how the prediction intervals can be updated using the most recent data.

5.1 Parametric prediction intervals

Based on orthogonality and linear additivity, the total forecast variance for the TS method can be approximated by the sum of individual variances (Hyndman & Ullah 2007):

$$\hat{\vartheta}_{n+h|n}(x) = \text{Var}\left[y_{n+h}(x) \mid \mathscr{I}, \Phi\right] \approx \sum_{k=1}^{K} \hat{\phi}_k^2(x)\hat{\zeta}_{k,n+h|n} + \hat{v}_{n+h}(x),$$

where $\hat{\zeta}_{k,n+h|n} = \text{Var}(\hat{\beta}_{k,n+h}|\hat{\beta}_{k,1}, \cdots, \hat{\beta}_{k,n})$ can be obtained by a time series model, and the model residual variance $\hat{v}_{n+h}(x)$ is estimated by averaging model residual square in year $n + h$, $\hat{\epsilon}_{n+h}^2(x)$, for each x variable. Under the normality assumption, the $100(1 - \alpha)\%$ prediction intervals for $y_{n+h}(x)$ are constructed as usual. This will also work for the BM method with appropriately defined functions.

5.2 Nonparametric prediction intervals

We present a nonparametric bootstrap method used in Shen (2009) and Hyndman & Shang (2009) to construct prediction intervals for the TS method. We can obtain one- or multi-step-ahead forecasts for the principal component scores $\{\hat{\beta}_{k,1}, \ldots, \hat{\beta}_{k,n}\}$, using a univariate time series model. Let the h-step-ahead forecast errors be given by $\hat{\xi}_{k,t,h} = \hat{\beta}_{k,t|t-h} - \hat{\beta}_{k,t}$, for $t = h + 1, \cdots, n$ where $h < n - 1$. These can then be sampled with replacement to give a bootstrap sample of $\beta_{k,n+h}$:

$$\hat{\beta}_{k,n+h|n}^b = \hat{\beta}_{k,n+h|n} + \hat{\xi}_{k,*,h}^b, \quad \text{for} \quad b = 1, \cdots, B.$$

Assuming the first K functional principal components approximate the data relatively well, the model residual should contribute nothing but independent and identically distributed random noise. Consequently, we can bootstrap the model fit error $\hat{\epsilon}_{n+h|n}^b(x)$ by sampling with replacement from the residual term $\{\hat{\epsilon}_1(x), \cdots, \hat{\epsilon}_n(x)\}$.

Adding all possible components of variability and assuming that those components of variability do

not correlate to each other, we obtain B forecast variants of $y_{n+h|n}(x)$,

$$\hat{y}^b_{n+h|n}(x) = \hat{\mu}(x) + \sum_{k=1}^{K} \hat{\phi}_k(x)\hat{\beta}^b_{k,n+h|n} + \hat{\epsilon}^b_{n+h|n}(x).$$

Hence, the $100(1-\alpha)$% prediction intervals are defined as $\alpha/2$ and $(1-\alpha/2)$ empirical quantiles of $\hat{y}^b_{n+h|n}(x)$. This will also work for the BM method with appropriately defined functions.

5.3 Updating distributional forecasts

The prediction intervals of the nonparametric distributional forecasts can also be updated using a bootstrap method. First, we bootstrap B samples of the TS forecasted regression coefficient estimates, $\hat{\beta}^{b,\text{TS}}_{n+1|n}$, and these bootstrapped samples in turn lead to $\hat{\beta}^{b,\text{PLS}}_{n+1}$, according to (5). From $\hat{\beta}^{b,\text{PLS}}_{n+1}$, we obtain B replications of

$$\hat{y}^{b,\text{PLS}}_{n+1}(x_l) = \hat{\mu}(x_l) + \sum_{k=1}^{K} \hat{\phi}_k(x_l)\hat{\beta}^{b,\text{PLS}}_{k,n+1} + \hat{\varepsilon}^b_{n+1}(x_l). \tag{6}$$

Hence, the $100(1-\alpha)$% prediction intervals for the PLS method are defined as $\alpha/2$ and $(1-\alpha/2)$ empirical quantiles of $\hat{y}^{b,\text{PLS}}_{n+1}(x_l)$.

5.4 Evaluating distributional forecasts

To evaluate the empirical coverage probabilities of prediction intervals, we compare the calculated prediction intervals with the original observations in the testing set (including sea surface temperatures from 1993 to 2008). The calculation process was performed as follows: for each curve in the testing sample, prediction intervals were generated by the TS, BM and PLS methods, at the 90% and 95% nominal coverage probabilities, and were tested to check if the known values fall within the specific prediction intervals. The empirical coverage probability was calculated as the ratio between the number of observations falling in the calculated prediction intervals and the number of total observations. Furthermore, we calculated the coverage probability deviance, which is the difference between the empirical and nominal coverage probabilities as a performance measure. Subject to the same average width of prediction intervals, the smaller the coverage probability deviance is, the better the method is.

The average width of prediction intervals is a way to assess which approach gives narrower prediction intervals. It can be expressed as:

$$W = \frac{1}{hp} \sum_{j=1}^{h} \sum_{i=1}^{p} \left| \hat{y}_{n+j|n}^{b,(1-\alpha/2)}(x_i) - \hat{y}_{n+j|n}^{b,\alpha/2}(x_i) \right|.$$

The narrower the average width of prediction intervals is, the better the method is, subject to the empirical coverage probability being close to the nominal coverage probability.

5.5 Density forecasts

As a by-product of the nonparametric bootstrap method, we can produce kernel density plots for visualizing density forecasts using the bootstrapped forecast variants. This graphical display can be useful for visualizing the extremes and the median. As with the kernel density estimate, we select the bandwidth using a pilot estimation of derivatives proposed by Sheather & Jones (1991), which seems to be close to optimal and generally preferred (Venables & Ripley 2002).

6 Results

6.1 Point forecasts

Our forecasting method decomposes a functional data set into a number of functional principal components and their associated scores. For simplicity of presentation, we display and attempt to interpret only the first three functional principal components and their associated scores in Figure 4, although we used $K = 6$ in modeling. Clearly, the mean function illustrates a strong seasonal pattern, with a peak in March and a trough in September. The functional principal components are of second order effects, as indicated by much smaller scales. The first functional principal component models the mid-year sea surface temperatures. While the second functional principal component models the contrast in sea surface temperatures between March and September, the third functional principal component models the contrast in sea surface temperatures from September to February and from March to August. Using the exponential smoothing state-space models of Hyndman et al. (2008), we obtained the forecasted principal component scores, and their 90% and 95% prediction intervals highlighted by the dark and light gray regions.

By conditioning on the historical data and fixed functional principal components, the decentralized forecasts are obtained by multiplying the forecasted principal component scores with the fixed
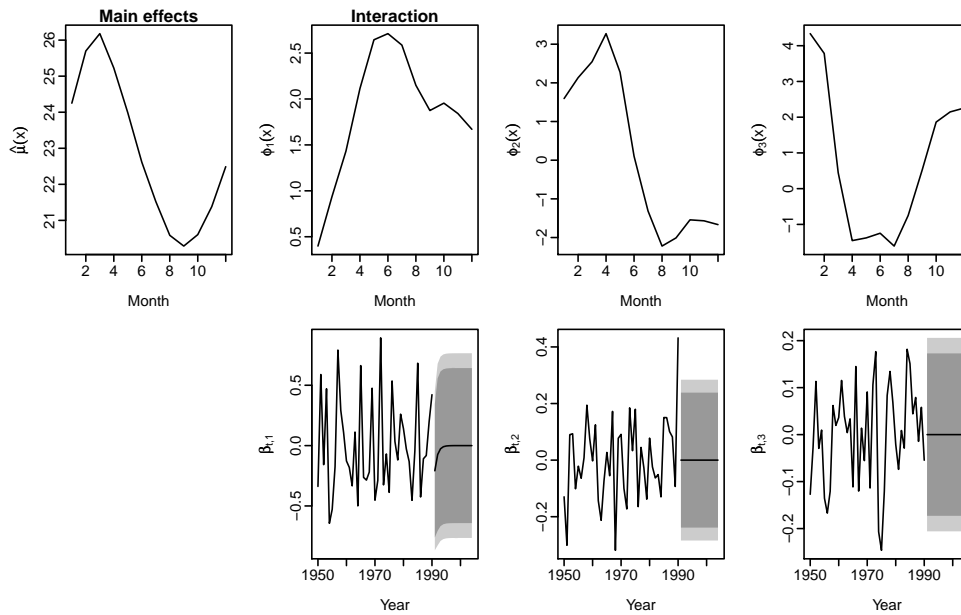
Figure 4: *The mean function, the first three functional principal components and their associated scores for the monthly sea surface temperatures from 1950 to 1992 (excluding the outliers). The 90% and 95% prediction intervals of the principal component scores are shown by the dark and light gray regions.*

functional principal components. For instance, Figure 5 displays the forecasted monthly sea surface temperatures in 2008, along with the 95% parametric and nonparametric prediction intervals. We found that the parametric prediction intervals seem to be similar to the nonparametric counterparts.
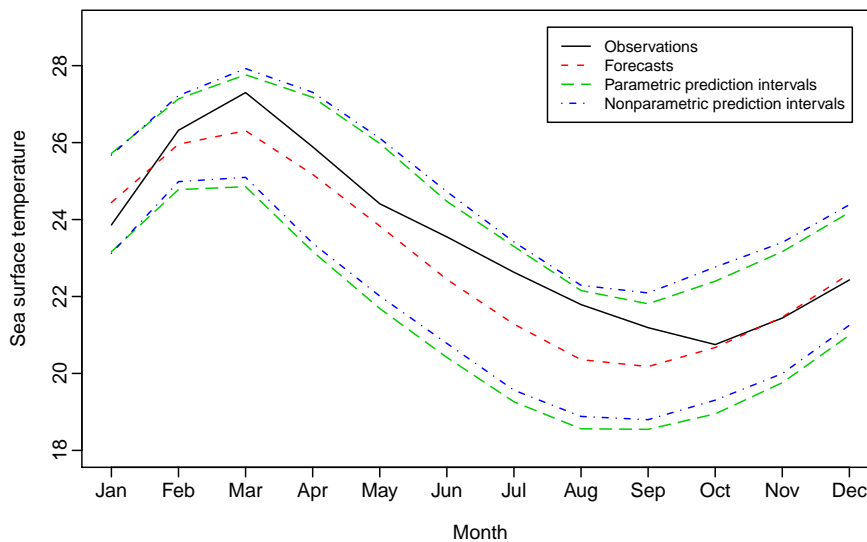

Figure 5: *One-step-ahead point forecasts of monthly sea surface temperatures in 2008, and the 95% prediction intervals constructed parametrically and nonparametrically.*

6.2 Comparisons with some existing methods

By means of comparison, we also investigate the point forecast accuracy of seasonal autoregressive integrated moving average (SARIMA), random walk (RW), and mean predictor (MP) methods. The MP method consists in predicting values at year $t + 1$ by the empirical mean values for each month from the first year to the tth year. The RW approach predicts new values at year $t + 1$ by the observations at year t. In the forecasting literature, SARIMA has been considered a benchmark method for forecasting a seasonal time series (Besse et al. 2000, Antoniadis & Sapatinas 2003, Ferraty et al. 2005). However, it requires the specification of the orders of seasonal components and non-seasonal components of an ARIMA model, which can be troublesome due to a large number of possible orders of seasonal and non-seasonal components. However, an automatic algorithm developed by Hyndman & Khandakar (2008) can be used to select the optimal orders for both seasonal and non-seasonal components. As a result, the optimal model selected is a SARIMA$(2,0,1)(0,1,0)_{12}$.

	Update month	MP	RW	SARIMA	TS	OLS	BM	PLS	RR
MAE	Mar-Dec	0.7166	0.8581	0.9604	0.7265	**0.6885**	0.7048	0.7131	**0.6885**
	Apr-Dec	0.7307	0.8659	0.9835	0.7364	0.6411	0.7262	0.6859	**0.6295**
	May-Dec	0.7099	0.8579	0.8767	0.7141	0.9346	0.7062	0.6675	**0.5974**
	Jun-Dec	0.7071	0.8421	0.8592	0.7119	0.9696	0.6973	0.6641	**0.5722**
	Jul-Dec	0.7189	0.8652	0.8602	0.7278	0.8026	0.6805	0.6023	**0.5378**
	Aug-Dec	0.7146	0.9063	0.8348	0.7379	1.1162	0.6875	0.6247	**0.5331**
	Sep-Dec	0.7062	0.9329	0.8380	0.7443	1.3868	0.7049	0.6910	**0.5928**
	Oct-Dec	0.7239	0.9607	0.5709	0.7819	0.6268	0.7381	0.5717	**0.5422**
	Nov-Dec	0.7235	0.9229	0.5211	0.7892	0.2598	0.7469	0.2698	**0.2445**
	Dec	0.6403	0.8307	**0.2110**	0.7115	0.2996	0.5857	0.2648	0.2601
	Mean	0.7092	0.8843	0.7516	0.7382	0.7726	0.6978	0.5755	**0.5198**
MSE	Mar-Dec	0.6928	1.3196	1.4155	0.7100	0.7967	**0.6895**	0.7036	0.8324
	Apr-Dec	0.7115	1.3607	1.4706	0.7296	**0.6161**	0.7180	0.6399	0.6233
	May-Dec	0.6822	1.3683	1.3195	0.7026	1.2242	0.6903	0.6173	**0.6142**
	Jun-Dec	0.6792	1.3710	1.1880	0.7035	1.3420	0.6803	0.6077	**0.5852**
	Jul-Dec	0.6984	1.4660	1.2089	0.7322	0.9243	0.6772	0.5237	**0.5135**
	Aug-Dec	0.7011	1.5726	1.1279	0.7541	2.0693	0.6835	0.5415	**0.5132**
	Sep-Dec	0.7056	1.6499	1.0624	0.7801	2.6651	0.7095	0.6284	**0.6112**
	Oct-Dec	0.7261	1.6972	0.5394	0.8263	0.5173	0.7443	**0.4537**	0.5759
	Nov-Dec	0.7112	1.5097	0.4244	0.8202	0.1296	0.7566	0.1096	**0.1091**
	Dec	0.5646	1.1353	**0.0676**	0.6615	0.1501	0.5093	0.1171	0.0997
	Mean	0.6873	1.4450	0.9824	0.7420	1.0435	0.6858	**0.4943**	0.5078

Table 2: *MAE and MSE of the point forecasts via the MP, RW, SARIMA, TS, OLS, BM, PLS, and RR methods with different updating months in the testing sample. The minimal values for each updating period are marked in bold.*

To compare the point forecast accuracy, we calculated the averaged MAE and MSE over the forecast horizon, shown in Table 2, for all methods investigated with different updating months in the testing

sample. Although the TS method performs better than the SARIMA and RW, it performs worse than the MP model for this data set. Among all dynamic updating methods, the RR performs the best with the minimum MAE, followed by the PLS, BM and OLS methods. Measured by the minimum MSE, the PLS method performs the best, followed by the RR, BM and OLS methods.

6.3 Distributional forecasts

Supposing we observe the sea surface temperatures from January to February 2008, it is possible to dynamically update the distributional forecasts for the remaining of 2008 using the BM and PLS methods. Based on the historical data from 1950 to 2007 (excluding the outliers), we obtain the forecasted principal component scores using the exponential smoothing state-space model. Utilizing the relationship between the $\hat{\beta}_{n+1|n}^{b,\text{TS}}$ and $\hat{\beta}_{n+1|n}^{b,\text{PLS}}$, the PLS prediction intervals for the updating periods can be obtained from (6). For instance, Figure 6 presents the 95% prediction intervals of the TS, BM and PLS methods for the sea surface temperatures from March to December 2008.
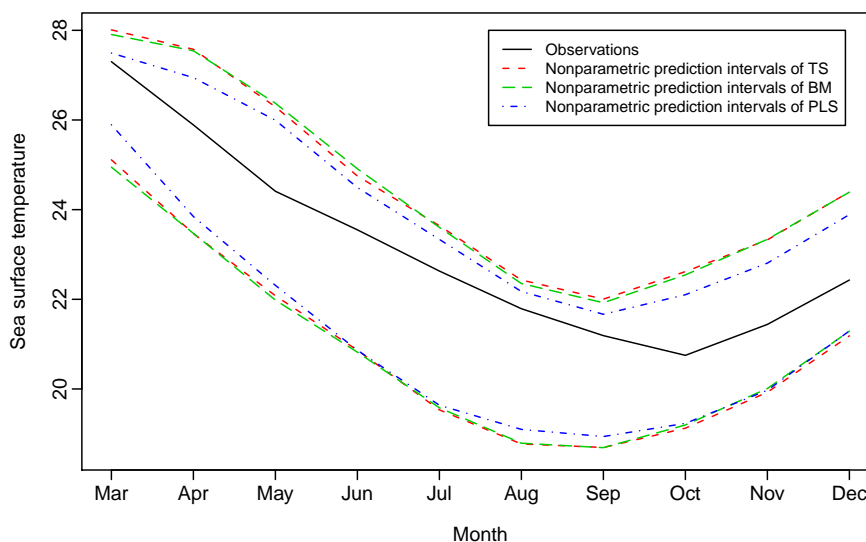

Figure 6: *Distributional forecasts of the sea surface temperatures during Mar-Dec 2008.*

From Figure 6, the PLS prediction intervals are comparably narrower, thus more informative than the TS and BM prediction intervals. Furthermore, we also examine the average coverage probability deviance and the average width of prediction intervals using different updating periods, shown in Tables 3 and 4 respectively.

An advantage of generating bootstrap samples is to provide density forecasts obtained using kernel density estimation. For example, Figure 7 displays the density plots of the monthly sea surface temperatures in 2008 based on $B = 1000$ replications.
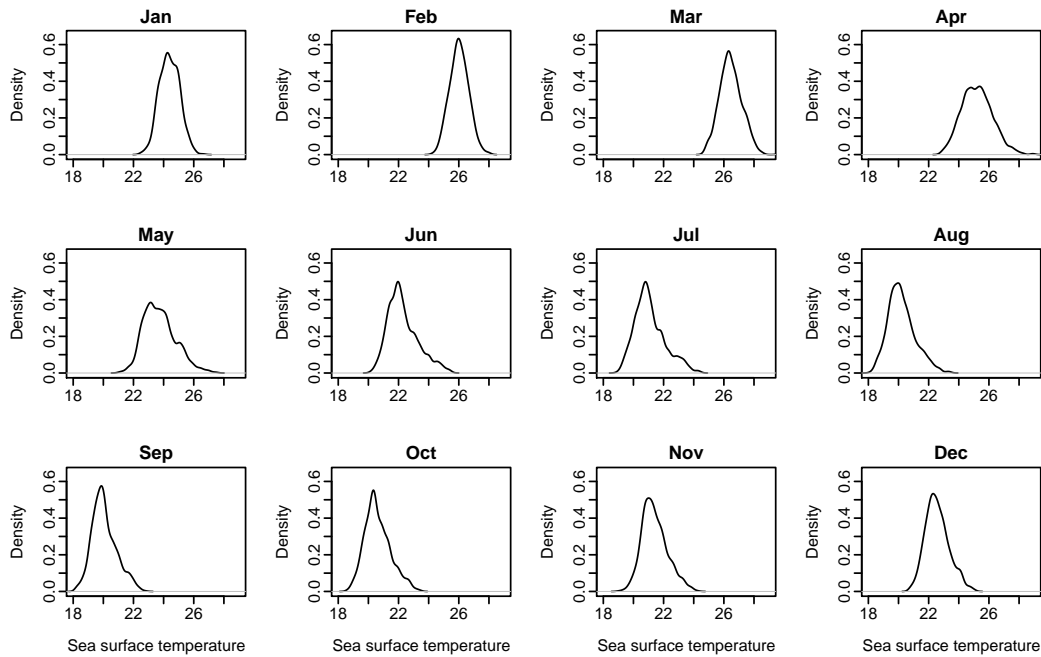
Figure 7: *Density plots of the monthly sea surface temperatures in 2008. The bandwidth is selected using a pilot estimation of derivatives.*

7 Conclusions

Our forecasting and updating approaches treat the historical data as a functional time series. Using FPCA, the dimensionality of data is effectively reduced, and the main features in the data are represented by a set of functional principal components, which explain more than 95% of the total variation in the monthly sea surface temperature data set.

The problem of forecasting future sea surface temperatures has been overcome by forecasting $K = 6$ one-dimensional principal component scores. Based on the historical data and the fixed functional principal components, the decentralized forecasts are obtained by multiplying the forecasted principal component scores with fixed functional principal components.

When partial data in the most recent curve are observed, four dynamic updating methods can not only update the point forecasts in order to improve point forecast accuracy, but also eliminate the assumption, $N = np$, made in Besse et al. (2000), Antoniadis & Sapatinas (2003), Ferraty & Vieu (2006, Chapter.12), Aneiros-Pérez & Vieu (2008) and Antoch et al. (2008). The BM approach rearranges the observations to form a complete data block, on which the TS method can still be applied. The OLS approach considers the partially observed data in the most recent curve as responses, and uses them to regress against the corresponding principal components. It however may suffer from the singularity problem when the number of partially observed data are less than the number of

principal components. To overcome this problem, the RR method heavily penalizes those regression coefficient estimates that deviate significantly from 0. However, the OLS and RR methods fail to consider all of the historical information. In contrast, the PLS method combines the TS forecasts and OLS forecasts by heavily penalizing for those regression coefficients that deviate significantly from $\hat{\beta}_{n+1|n}^{\text{TS}}$. Based on the the MAE and MSE in the testing sample, the RR and PLS show better forecast accuracy than other methods investigated and the difference between them is almost negligble.

Furthermore, we proposed a nonparametric method to construct prediction intervals, and compared the empirical coverage probability to a parametric method. Although the coverage probabilities of the parametric and nonparametric methods are similar, the nonparametric method is appropriate to produce density plots and to construct prediction intervals for updated forecasts. With a similar empirical coverage probability, the prediction interval width of the updated forecasts is narrower, thus the PLS and BM methods are more informative than the TS method without updating.

R code for calculating the point forecasts, updating point forecasts, and constructing parametric and nonparametric prediction intervals is available from the authors upon request.

References

Aneiros-Pérez, G. & Vieu, P. (2008), 'Nonparametric time series prediction: A semi-functional partial linear modeling', *Journal of Multivariate Analysis* **99**(5), 834–857.

Antoch, J., Prchal, L., De Rosa, M. R. & Sarda, P. (2008), Functional linear regression with functional response: Application to prediction of electricity consumption, *in* S. Dabo-Niang & F. Ferraty, eds, 'Functional and operatorial statistics', Springer, Heidelberg, pp. 23–29.

Antoniadis, A. & Sapatinas, T. (2003), 'Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes', *Journal of Multivariate Analysis* **87**(1), 133–158.

Besse, P. C., Cardot, H. & Stephenson, D. B. (2000), 'Autoregressive forecasting of some functional climatic variations', *Scandinavian Journal of Statistics* **27**(4), 673–687.

Bosq, D. (2000), *Linear processes in function spaces: Theory and applications*, Springer, Berlin.

Bosq, D. & Blanke, D. (2007), *Inference and prediction in large dimensions*, John Wiley, Chichester, England.

Box, G. E. P., Jenkins, G. M. & Reinsel, G. C. (2008), *Time series analysis: Forecasting and control*, 4 edn, John Wiley, Hoboken, New Jersey.

Cai, T. & Hall, P. (2006), 'Prediction in functional linear regression', *Annals of Statistics* **34**(5), 2159–2179.

Chatfield, C. (1993), 'Calculating interval forecasts', *Journal of Business & Economic Statistics* **11**(2), 121–135.

Chatfield, C. (2000), *Time series forecasting*, Chapman & Hall, Boca Raton.

Dauxois, J., Pousse, A. & Romain, Y. (1982), 'Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference', *Journal of Multivariate Analysis* **12**(1), 136–154.

Delaigle, A., Hall, P. & Apanasovich, T. V. (2009), Weighted least squares methods for prediction in the functional data linear model. arXiv:0902.3319v1 [stat.ME].
URL: *http://arxiv.org/abs/0902.3319v1*

Dioses, T., Dávalos, R. & Zuzunaga, J. (2002), 'El Niño 1982-1983 and 1997-1998: Effects on Peruvian Jack Mackerel and Peruvian Chub Mackerel', *Investigaciones marinas* **30**(1), 185–187.

Ferraty, F., Rabhi, A. & Vieu, P. (2005), 'Conditional quantiles for dependent functional data with application to the climatic El Niño phenomenon', *Sankhya: The Indian Journal of Statistics* **67**(2), 378–398.

Ferraty, F. & Vieu, P. (2006), *Nonparametric functional data analysis: Theory and practice*, Springer, New York.

Greenshtein, E. (2006), 'Best subset selection, persistence in high-dimensional statistical learning and optimization under l_1 constraint', *Annals of Statistics* **34**(5), 2367–2386.

Greenshtein, E. & Ritov, Y. (2004), 'Persistence in high-dimensional linear predictor selection and the virtue of overparameterization', *Bernoulli* **10**(6), 971–988.

Hall, P. & Horowitz, J. L. (2007), 'Methodology and convergence rates for functional linear regression', *Annals of Statistics* **35**(1), 70–91.

Hall, P. & Hosseini-Nasab, M. (2006), 'On properties of functional principal components analysis', *Journal of the Royal Statistical Society: Series B* **68**(1), 109–126.

Hall, P. & Hosseini-Nasab, M. (2009), 'Theory for high-order bounds in functional principal components analysis', *Mathematical Proceedings of the Cambridge Philosophical Society* **146**(1), 225–256.

Hall, P., Müller, H.-G. & Wang, J.-L. (2006), 'Properties of principal component methods for functional and longitudinal data analysis', *Annals of Statistics* **34**(3), 1493–1517.

Harvey, A. (1990), *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press, Cambridge.

Hoerl, A. E. & Kennard, R. W. (1970), 'Ridge regression: Biased estimation for nonorthogonal problems', *Technometrics* **12**(1), 55–67.

Hössjer, O. & Croux, C. (1995), 'Generalizing univariate signed rank statistics for testing and estimating a multivariate location parameter', *Journal of Nonparametric Statistics* **4**(3), 293–308.

Hyndman, R. J. & Booth, H. (2008), 'Stochastic population forecasts using functional data models for mortality, fertility and migration', *International Journal of Forecasting* **24**(3), 323–342.

Hyndman, R. J. & Khandakar, Y. (2008), 'Automatic time series forecasting: the forecast package for R', *Journal of Statistical Software* **27**(3).

Hyndman, R. J., Koehler, A. B., Ord, J. K. & Snyder, R. D. (2008), *Forecasting with exponential*

smoothing: the state space approach, Springer, Berlin.

Hyndman, R. J. & Shang, H. L. (2008), Bagplots, boxplots, and outlier detection for functional data, *in* S. Dabo-Niang & F. Ferraty, eds, 'Functional and Operatorial Statistics', Springer, Heidelberg, pp. 201–207.

Hyndman, R. J. & Shang, H. L. (2009), 'Forecasting functional time series (with discussion)', *Journal of the Korean Statistical Society* **38**(3), 199–221.

Hyndman, R. J. & Ullah, M. S. (2007), 'Robust forecasting of mortality and fertility rates: A functional data approach', *Computational Statistics & Data Analysis* **51**(10), 4942–4956.

Izenman, A. J. (2008), *Modern multivariate statistical techniques: Regression, classification, and manifold learning*, Springer, New York.

Moran, E. F., Adams, R., Bakoyéma, B., Stefano, F. T. & Boucek, B. (2006), 'Human strategies for coping with EL Niño related drought in Amazônia', *Climatic Change* **77**(3-4), 343–361.

Preda, C. & Saporta, G. (2005*a*), 'Clusterwise PLS regression on a stochastic process', *Computational Statistics & Data Analysis* **49**(1), 99–108.

Preda, C. & Saporta, G. (2005*b*), 'PLS regression on a stochastic process', *Computational Statistics & Data Analysis* **48**(1), 149–158.

Ramsay, J. O. & Silverman, B. W. (2002), *Applied functional data analysis: Methods and case studies*, Springer, New York.

Ramsay, J. O. & Silverman, B. W. (2005), *Functional data analysis*, 2 edn, Springer, New York.

Reiss, P. T. & Ogden, T. R. (2007), 'Functional principal component regression and functional partial least squares', *Journal of the American Statistical Association* **102**(479), 984–996.

Rice, J. A. & Silverman, B. W. (1991), 'Estimating the mean and covariance structure nonparametrically when the data are curves', *Journal of the Royal Statistical Society: Series B* **53**(1), 233–243.

Sheather, S. J. & Jones, M. C. (1991), 'A reliable data-based bandwidth selection method for kernel density estimation', *Journal of the Royal Statistical Society: Series B* **53**(3), 683–690.

Shen, H. (2009), 'On modeling and forecasting time series of smooth curves', *Technometrics* **51**(3), 227–238.

Shen, H. & Huang, J. Z. (2008), 'Interday forecasting and intraday updating of call center arrivals',

Manufacturing and Service Operations Management **10**(3), 391–410.

Silverman, B. W. (1995), 'Incorporating parametric effects into functional principal components ', *Journal of the Royal Statistical Society: Series B* **57**(4), 673–689.

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

Timmermann, A., Oberhuber, J., Bacher, A., Esch, M., Latif, M. & Roeckner, E. (1999), 'Increased El Niño frequency in a climate model forced by future greenhouse warming', *Nature* **398**(6729), 694–697.

Venables, W. N. & Ripley, B. D. (2002), *Modern applied statistics with S*, 4 edn, Springer, New York.

Table 3: *The 90% and 95% coverage probabilities of the TS, BM and PLS prediction intervals constructed parametrically and nonparametrically.*

| | Parametric | | | | Nonparametric | | | | | |
| | TS | | BM | | TS | | BM | | PLS | |
Update period	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
Mar-Dec	0.9643	0.9714	0.9571	0.9786	0.9357	0.9714	0.9500	0.9714	0.8929	0.9500
Apr-Dec	0.9603	0.9683	0.9444	0.9762	0.9286	0.9683	0.9365	0.9683	0.8968	0.9524
May-Dec	0.9554	0.9643	0.9554	0.9643	0.9286	0.9643	0.9464	0.9643	0.9196	0.9554
Jun-Dec	0.9490	0.9592	0.9490	0.9592	0.9286	0.9592	0.9388	0.9490	0.9082	0.9490
Jul-Dec	0.9405	0.9524	0.9167	0.9643	0.9167	0.9524	0.9167	0.9405	0.8690	0.9405
Aug-Dec	0.9286	0.9429	0.9000	0.9429	0.9143	0.9429	0.9429	0.9429	0.8571	0.9286
Sep-Dec	0.9107	0.9286	0.8750	0.9464	0.8929	0.9286	0.8929	0.9464	0.8810	0.9286
Oct-Dec	0.9048	0.9286	0.8810	0.9286	0.8810	0.9286	0.9048	0.9286	0.8929	0.9048
Nov-Dec	0.9286	0.9286	0.8929	0.9643	0.8929	0.9286	0.9286	0.9286	0.9286	0.9286
Dec	0.9286	0.9286	0.9286	1.0000	0.9286	0.9286	0.9286	0.9286	0.9286	0.9286
Mean coverage probability deviance	0.0371	0.0158	0.0302	0.0188	0.0214	0.0158	0.0302	**0.0140**	**0.0209**	0.0149

Table 4: *Width of the TS, BM and PLS prediction intervals at the 90% and 95% nominal coverage probabilities.*

Update period	Parametric						Nonparametric					
	TS		BM		TS		BM		PLS			
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%		
Mar-Dec	3.0633	3.6501	3.0559	3.6413	3.0525	3.5491	3.0497	3.5078	2.7000	3.1496		
Apr-Dec	3.1280	3.7272	3.1266	3.7256	3.1188	3.6212	3.1328	3.6584	2.7664	3.2141		
May-Dec	3.0938	3.6864	3.0956	3.6886	3.0840	3.5727	3.1270	3.6070	2.7722	3.2099		
Jun-Dec	3.0058	3.5816	3.0044	3.5800	2.9976	3.4731	2.9987	3.5024	2.6599	3.0511		
Jul-Dec	2.9162	3.4749	2.9063	3.4630	2.9110	3.3842	2.8409	3.4059	2.4866	2.8996		
Aug-Dec	2.8007	3.3373	2.7986	3.3347	2.8012	3.2588	2.8471	3.3734	2.2538	2.6135		
Sep-Dec	2.7352	3.2591	2.7391	3.2639	2.7385	3.1860	2.7432	3.2494	2.3376	2.8172		
Oct-Dec	2.7439	3.2695	2.7495	3.2762	2.7521	3.2013	2.7330	3.2316	2.4022	2.7802		
Nov-Dec	2.7128	3.2325	2.7172	3.2378	2.7117	3.1635	2.7364	3.2552	2.3347	2.6885		
Dec	2.6753	3.1878	2.6672	3.1781	2.6774	3.1220	2.7619	3.2979	2.0819	2.4772		
Mean width	2.8875	3.4406	2.8860	3.4389	2.8845	3.3532	2.8971	3.4089	**2.4795**	**2.8901**		