Available online at www.sciencedirect.com
ScienceDirect
Mathematics and Computers in Simulation 81 (2011) 1310–1324
MATHEMATICS AND COMPUTERS IN SIMULATION
www.elsevier.com/locate/matcom

Nonparametric time series forecasting with dynamic updating

Han Lin Shang *, Rob.J. Hyndman

Department of Econometrics & Business Statistics, Monash University, Wellington Road, Clayton, Melbourne, VIC 3800, Australia

Received 22 September 2009; received in revised form 23 March 2010; accepted 19 April 2010
Available online 11 May 2010

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 by 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.

Keywords: Functional principal component analysis; Functional time series; Penalized least squares; Ridge regression; Seasonal time series

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 Aneiros-Pérez and Vieu [1], Antoch et al. [2], Antoniadis and Sapatinas [3], Besse et al. [4], Ferraty and Vieu [15, Chapter 12]. 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. When the seasonal pattern is strong, one way to model the time series nonparametrically is to use ideas from functional data analysis (Ramsay and Silverman [38]). We divide the observed time series into n trajectories, and then consider each trajectory of length p as a curve rather than as p distinct points. The functional time series is then given by

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

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.

* Corresponding author. Tel: +61 3 99050547; fax: +61 3 99055474.
E-mail address: HanLin.Shang@buseco.monash.edu.au (H.L. Shang).

0378-4754/$36.00 © 2010 IMACS. Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.matcom.2010.04.027

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 [5], Bosq and Blanke [6]), functional kernel regression (Aneiros-Pérez and Vieu [1], Ferraty and Vieu [15]), functional principal component regression (Hyndman and Booth [25], Hyndman and Shang [29], Hyndman and Ullah [32]), and functional partial least squares regression (Preda and Saporta [35], Preda and Saporta [36]). 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, $\{y_t(x), t = 1, \ldots, n\}$, 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 called the block moving (BM), ridge regression (RR), ordinary least squares (OLS) and penalized least squares (PLS) methods.

Interval 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. Sea surface temperature 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 Fig. 1a, with the same data shown in Fig. 1b as a time series of functions.

From Fig. 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 and Shang [31] and identified four outliers. These outliers correspond to the years 1982–1983 and 1997–1998, highlighted by the thick black lines in Fig. 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. [34], Timmermann [46]). 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. [13] reported that the northern central region of Peru was strongly affected because warm waters with low salinity

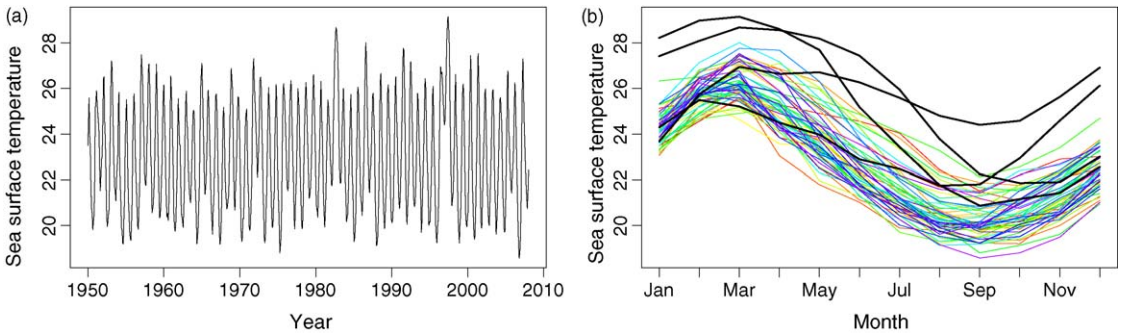

Fig. 1. Exploratory plots suggesting that both predictive regularity and abnormality are presented in the sea surface temperature data between January 1950 and December 2008 measured by moored buoys in the region defined by the coordinate 0–10° South and 90–80° West. (a) A univariate time series display of the monthly sea surface temperatures. (b) A functional time series display of the monthly sea surface temperatures.

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 Ferraty and Vieu [15], Ramsay and Silverman [37], Ramsay and Silverman [38]. Papers covering the development of FPCA include those of Hyndman and Shang [29], Hyndman and Ullah [32], Reiss and Ogden [39], Ramsay and Silverman [40], Shen [42], Silverman [44], Silverman [45]. Significant treatments of the theory of FPCA are given by Cai and Hall [8], Dauxois et al. [11], Delaigle et al. [12], Hall and Horowitz [18], Hall and Hosseini-Nasab [19], Hall and Hosseini-Nasab [20], Hall et al. [21].

In this section, we assume that all trajectories are complete. Our forecasting method begins by subtracting the functional mean from the original functional data. 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

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

where $\|g(u)\| = (\int g^2(u)du)^{1/2}$. The algorithm of Hössjer and Croux [24] can be used to compute $\hat{\mu}(x)$.

Using FPCA, $\{y_t; t = 1, \ldots, n\}$ 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 $\{\hat{\phi}_1(x), \ldots, \hat{\phi}_K(x)\}$ represents a set of the estimated functional principal components, $\{\hat{\beta}_{1,t}, \ldots, \hat{\beta}_{K,t}\}$ represents a set of the 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 $\{\hat{\beta}_{k,1}, \ldots, \hat{\beta}_{k,n}; k = 1, \ldots, K\}$ using univariate time series models, such as the ARIMA models (Box et al. [7]), damped Holt's models (Hyndman et al. [28]), state space models (Harvey [22]), or exponential smoothing state space models (Hyndman et al. [27]). 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 and Ullah [32], Shen and Huang [43]).

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

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

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

3.2. Component selection

Hyndman and Booth [25] 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 poor forecast

accuracy. In this analysis, we chose $K = 6$, which should be larger than any of the components required. In the context of overparametrized regression problems, Greenshtein [16], Greenshtein and Ritov [17] described this phenomenon as "persistence in high-dimensional linear predictor selection".

It is noteworthy that the optimal K in a data set can be determined using a cross-validation method, such as 5- or 10-fold cross validation.

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 $\boldsymbol{y}_{n+1}(x_e) = [y_{n+1}(x_1), \ldots, y_{n+1}(x_{m_0})]'$, 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

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

where $\mathcal{I}_l$ denotes the historical data corresponding to the remaining time periods; $\boldsymbol{\Phi}_l = \{\hat{\phi}_1(x_l), \ldots, \hat{\phi}_K(x_l)\}$ is a set of the estimated functional principal components corresponding to the remaining time periods; $\hat{\mu}(x_l)$ is the mean function corresponding to the remaining time periods; $\hat{\phi}_k(x_l)$ is the k th functional principal component corresponding to the remaining time periods.

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

4.1. Block moving (BM) method

The 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 $[m_0 + 1, p] \cup [1, m_0]$.

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

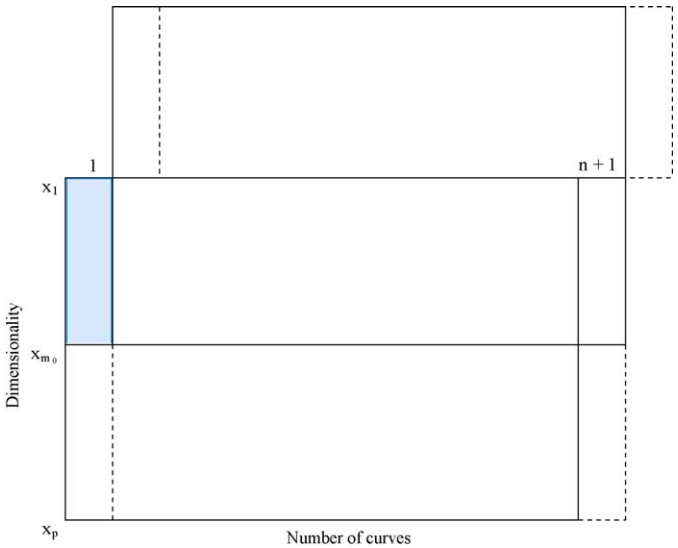

Fig. 2. Dynamic update via the block moving approach. The colored region shows the data loss in the first year. The forecasts for the rest of year $n + 1$ can be updated by the forecasts using the TS method applied to the top block.

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

4.2. Ordinary least squares (OLS) method

We can estimate the remaining part of the last trajectory using a regression based on the principal components obtained in (1). Let $\mathcal{F}_e$ 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 $1 \le k \le K$. Let $\boldsymbol{\beta}_{n+1} = [\beta_{1,n+1}, \ldots, \beta_{K,n+1}]'$, and $\boldsymbol{\epsilon}_{n+1}(x_e) = [\epsilon_{n+1}(x_1), \ldots, \epsilon_{n+1}(x_{m_0})]'$. 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

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

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

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

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

$$\hat{y}^{\mathrm{OLS}}_{n+1}(x_l) = \mathrm{E}[y_{n+1}(x_l)|\mathcal{I}_l, \boldsymbol{\Phi}_l] = \hat{\mu}(x_l) + \sum_{k=1}^{K} \hat{\phi}_k(x_l) \hat{\beta}^{\mathrm{OLS}}_{k,n+1}.$$

4.3. Ridge regression (RR) method

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 $\hat{\boldsymbol{\beta}}^{\mathrm{OLS}}_{n+1} = \{\hat{\beta}^{\mathrm{OLS}}_{1,n+1}, \ldots, \hat{\beta}^{\mathrm{OLS}}_{K,n+1}\}$ to be numerically stable. To address this problem, we adapt the ridge regression (RR) method of Hoerl and Kennard [23] with the predictors being the corresponding principal components and the partially observed data being the responses. The main advantage of RR method is that it uses a square penalty function, which is a rotationally invariant hypersphere centered at the origin (Izenman [33]). Two-dimensional contours of the different penalty functions are presented in Fig. 3.

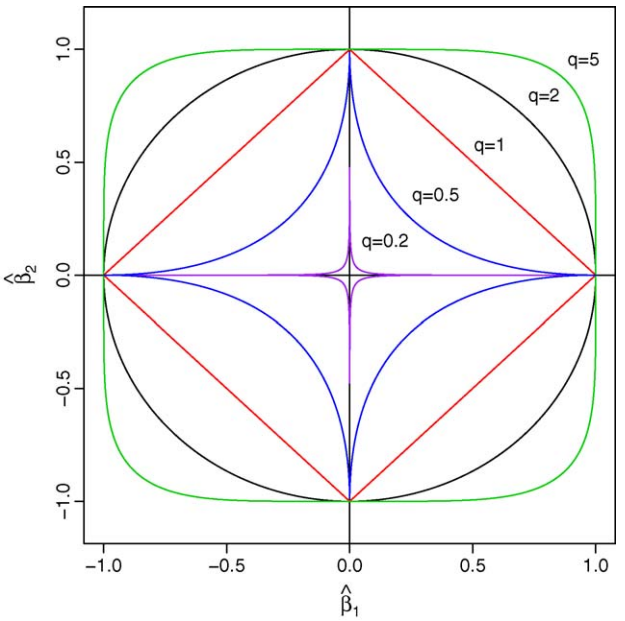

Fig. 3. Two-dimensional contours of the symmetric penalty function $p_q(\boldsymbol{\beta}) = |\boldsymbol{\beta}_1|^q + |\boldsymbol{\beta}_2|^q = 1$ for $q = 0.2, 0.5, 1, 2, 5$. When $q = 2$, the square penalty function is rotationally invariant hypersphere centered at the origin.

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

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

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:

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

where $\boldsymbol{I}_K$ is the $K \times K$ identity matrix. When the penalty parameter $\lambda \to 0$, $\hat{\boldsymbol{\beta}}^{\mathrm{RR}}_{n+1}$ approaches $\hat{\boldsymbol{\beta}}^{\mathrm{OLS}}_{n+1}$, provided that $\left(\mathcal{F}'_e \mathcal{F}_e \right)^{-1}$ exists; when $\lambda \to \infty$, $\hat{\boldsymbol{\beta}}^{\mathrm{RR}}_{n+1}$ approaches 0; when $0 < \lambda < \infty$, $\hat{\boldsymbol{\beta}}^{\mathrm{RR}}_{n+1}$ is a weighted average between 0 and $\hat{\boldsymbol{\beta}}^{\mathrm{OLS}}_{n+1}$.

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

$$\hat{y}^{\mathrm{RR}}_{n+1}(x_l) = \mathrm{E}[y_{n+1}(x_l) | \mathcal{I}_l, \boldsymbol{\Phi}_l] = \hat{\mu}(x_l) + \sum_{k=1}^{K} \hat{\phi}_k(x_l) \hat{\beta}^{\mathrm{RR}}_{k,n+1}.$$

4.4. Penalized least squares (PLS) method

Although the RR method solves the potential singularity problem, it does not take account of the TS forecasted regression coefficient estimates, $\hat{\boldsymbol{\beta}}^{\mathrm{TS}}_{n+1|n}$. This motivates the development of the PLS method (Shen [42], Shen and Huang [43]), in which the regression coefficient estimates are selected by shrinking them toward $\hat{\boldsymbol{\beta}}^{\mathrm{TS}}_{n+1|n}$. The PLS regression coefficient estimates minimize a penalized residual sum of squares:

$$\arg\min_{\boldsymbol{\beta}_{n+1}} \{ (\hat{\mathbf{y}}^*_{n+1}(x_e) - \mathcal{F}_e \hat{\boldsymbol{\beta}}_{n+1})' (\hat{\mathbf{y}}^*_{n+1}(x_e) - \mathcal{F}_e \hat{\boldsymbol{\beta}}_{n+1}) + \lambda (\hat{\boldsymbol{\beta}}_{n+1} - \hat{\boldsymbol{\beta}}^{\mathrm{TS}}_{n+1|n})' (\hat{\boldsymbol{\beta}}_{n+1} - \hat{\boldsymbol{\beta}}^{\mathrm{TS}}_{n+1|n}) \}. \tag{4}$$

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 $\hat{\boldsymbol{\beta}}^{\mathrm{PLS}}_{n+1}$ 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 $\hat{\boldsymbol{\beta}}_{n+1}$ in (4), we obtain:

$$\hat{\boldsymbol{\beta}}^{\mathrm{PLS}}_{n+1} = (\mathcal{F}'_e \mathcal{F}_e + \lambda \boldsymbol{I}_K)^{-1} (\mathcal{F}'_e \hat{\mathbf{y}}^*_{n+1}(x_e) + \lambda \hat{\boldsymbol{\beta}}^{\mathrm{TS}}_{n+1|n}). \tag{5}$$

When the penalty parameter $\lambda \to 0$, $\hat{\boldsymbol{\beta}}^{\mathrm{PLS}}_{n+1}$ approaches $\hat{\boldsymbol{\beta}}^{\mathrm{OLS}}_{n+1}$, provided that $\left(\mathcal{F}'_e \mathcal{F}_e \right)^{-1}$ exists; when $\lambda \to \infty$, $\hat{\boldsymbol{\beta}}^{\mathrm{PLS}}_{n+1}$ approaches $\hat{\boldsymbol{\beta}}^{\mathrm{TS}}_{n+1|n}$; when $0 < \lambda < \infty$, $\hat{\boldsymbol{\beta}}^{\mathrm{PLS}}_{n+1}$ is a weighted average between $\hat{\boldsymbol{\beta}}^{\mathrm{TS}}_{n+1|n}$ and $\hat{\boldsymbol{\beta}}^{\mathrm{OLS}}_{n+1}$.

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

$$\hat{y}^{\mathrm{PLS}}_{n+1}(x_l) = \mathrm{E}[y_{n+1}(x_l) | \mathcal{I}_l, \boldsymbol{\Phi}_l] = \hat{\mu}(x_l) + \sum_{k=1}^{K} \hat{\phi}_k(x_l) \hat{\beta}^{\mathrm{PLS}}_{k,n+1}.$$

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

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

Updating period	Minimum MSE		Minimum MAE	
	PLS	RR	PLS	RR
March–December	908.43	0.00	1118.58	0.00
April–December	335.40	3.11	197.66	3.35
May–December	233.53	8.99	245.64	7.34
June–December	111.85	11.00	138.92	8.29
July–December	7.47	6.23	4.86	4.91
August–December	27.90	11.62	18.42	7.61
September–December	279.05	15.59	197.80	10.50
October–December	9.01	4.60	7.77	5.41
November–December	3.29	0.73	4.82	1.44
December	3.60	1.74	8.25	2.33

λ for different updating periods are determined by minimizing the MAE and MSE criteria within the validation set. These criteria can be expressed as

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

and

$$\text{MSE} = \frac{1}{pq}\sum_{j=1}^{q}\sum_{i=1}^{p}[y_{n-j+1}(x_i) - \hat{y}_{n-j+1|n-j}(x_i)]^2,$$

where p represents the number of observations in each year, and q represents the number of years in the validation set. In Table 1, the optimal tuning parameters for different updating periods are given for both the PLS and RR methods.

5. Interval forecast methods

Prediction intervals are a valuable tool for assessing the probabilistic uncertainty associated with point forecasts. As emphasized in Chatfield [9] and Chatfield [10], 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 to construct prediction intervals for the TS and BM methods. Then in Section 5.3 we show how the prediction intervals can be updated using the nonparametric bootstrap method.

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 and Ullah [32]):

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

where $\hat{\zeta}_{k,n+h|n} = \mathrm{Var}(\hat{\beta}_{k,n+h}|\hat{\beta}_{k,1},\ldots,\hat{\beta}_{k,n})$ 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 $n+h$, $\hat{\epsilon}^2_{n+h}(x)$, for each x variable. Under the normality assumption, the $100(1-\alpha)\%$ prediction intervals of $y_{n+h}(x)$ 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 Hyndman and Shang [29] and Shen [42] to construct prediction intervals for the TS and BM methods. We can obtain one- or multi-step-ahead forecasts for the principal component scores $\{\hat{\beta}_{k,1},\ldots,\hat{\beta}_{k,n}; k=1,\ldots,K\}$, using a univariate time series model. Let the h-step-ahead forecast errors be given by $\hat{\xi}_{k,j} = \hat{\beta}_{k,n-j+1|n-j} - \hat{\beta}_{k,n-j+1}$, for $j=1,\ldots,n-K$. These can then be sampled with replacement to give a bootstrap sample of $\beta_{k,n+h}$:

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

where $\hat{\xi}^b_{k,*}$ denotes the bootstrap samples, and B is the number of bootstrap replications.

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 residual $\hat{\epsilon}^b_{n+h|n}(x)$ by sampling with replacement from the residual term $\{\hat{\epsilon}_1(x),\ldots,\hat{\epsilon}_n(x)\}$.

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 interval forecasts

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

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

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

5.4. Evaluating interval forecasts

The evaluation of empirical coverage probability was performed as follows: for each curve in the testing sample, prediction intervals of one-step-ahead forecasts were computed parametrically and nonparametrically at the 90 and 95% nominal coverage probabilities, and were tested to check if the holdout data points 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 total number of observations in the testing sample. 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

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

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 quantile. As with the kernel density estimate, we select the bandwidth using a pilot estimation of derivatives proposed by Sheather and Jones [41], which seems to be close to optimal and generally preferred (Venables and Ripley [47]).

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 in the top panel of Fig. 4, although we used $K = 6$ in modeling and forecasting. 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 April and August, the third functional principal component models

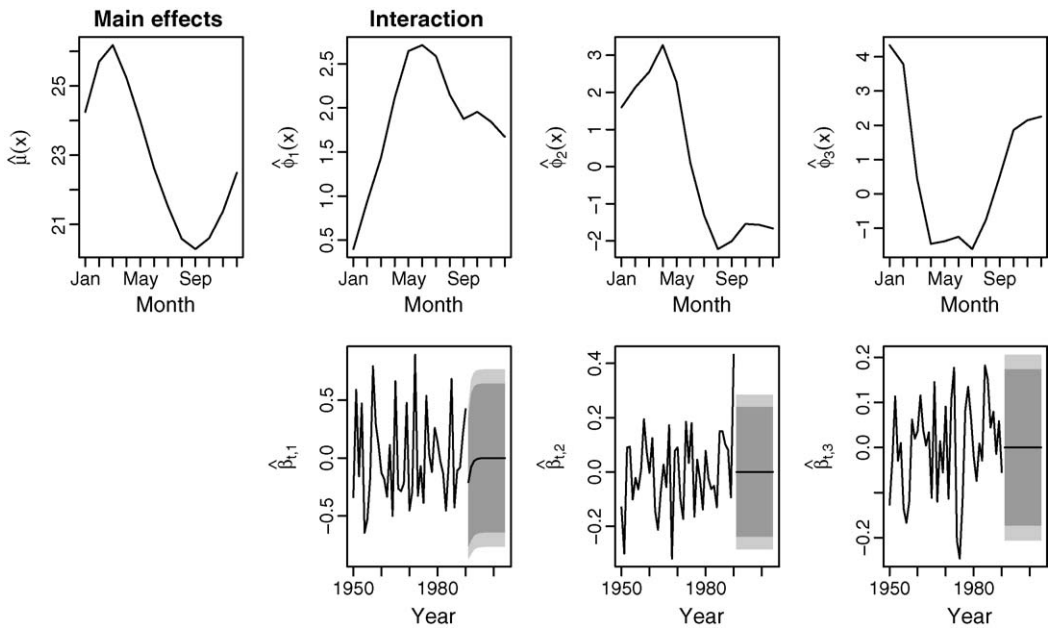

Fig. 4. The mean function, the first three functional principal components and their associated scores for the monthly sea surface temperatures from 1950 to 1992 (excluding the outliers). The 90 and 95% prediction intervals of the principal component scores are shown by the dark and light gray regions.

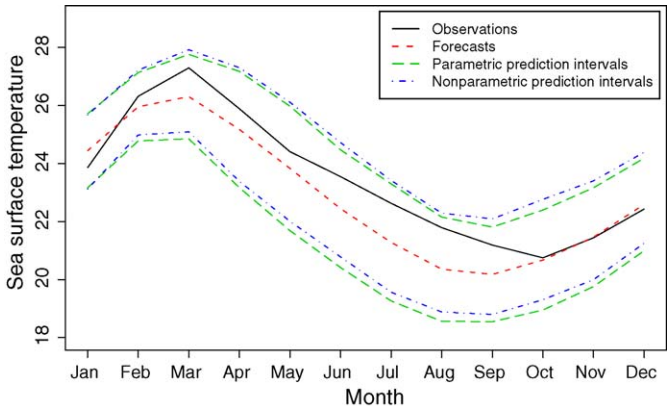

Fig. 5. Point forecasts of the monthly sea surface temperatures in 2008, and the 95% prediction intervals constructed via the parametric and nonparametric approaches.

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. [27], we obtained the forecasted principal component scores, and their 90 and 95% prediction intervals highlighted by the dark and light gray regions in the bottom panel of Fig. 4.

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 functional principal components. For instance, Fig. 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.

6.2. *Point forecast comparisons with some existing methods*

By means of comparisons, 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 value for each month from the first year to the t th 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 as a benchmark method for forecasting a seasonal time series (Antoniadis and Sapatinas [3], Besse et al. [4], Ferraty et al. [14]). 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 and Khandakar [26] can be used

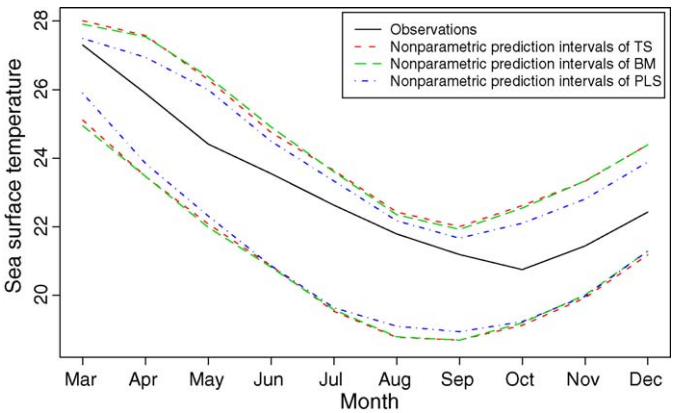

Fig. 6. Prediction intervals of the sea surface temperatures between March and December 2008. By incorporating the sea surface temperatures between January and February, the prediction intervals can be updated using the bootstrap method.

Table 2
Average 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 of the MAE and MSE for each updating period are marked in bold.

	Non-dynamic updating methods				Dynamic updating methods			
Update month	MP	RW	SARIMA	TS	OLS	BM	PLS	RR
MAE								
March–December	0.72	0.86	0.96	0.73	**0.69**	0.70	0.71	**0.69**
April–December	0.73	0.87	0.98	0.74	0.64	0.73	0.69	**0.63**
May–December	0.71	0.86	0.88	0.71	0.93	0.71	0.67	**0.60**
June–December	0.71	0.84	0.86	0.71	0.97	0.70	0.66	**0.57**
July–December	0.72	0.87	0.86	0.73	0.80	0.68	0.60	**0.54**
August–December	0.71	0.91	0.83	0.74	1.12	0.69	0.62	**0.53**
September–December	0.71	0.93	0.84	0.74	1.39	0.70	0.69	**0.59**
October–December	0.72	0.96	0.57	0.78	0.63	0.74	0.57	**0.54**
November–December	0.72	0.92	0.52	0.79	0.26	0.75	0.27	**0.24**
December	0.64	0.83	**0.21**	0.71	0.30	0.59	0.26	0.26
Mean	0.71	0.89	0.75	0.74	0.77	0.70	0.57	**0.52**
MSE								
March–December	**0.69**	1.32	1.42	0.71	0.80	**0.69**	0.70	0.80
April–December	0.71	1.36	1.47	0.73	0.62	0.72	0.64	**0.60**
May–December	0.68	1.37	1.32	0.70	1.22	0.69	0.62	**0.59**
June–December	0.68	1.37	1.19	0.70	1.34	0.68	0.61	**0.58**
July–December	0.70	1.47	1.21	0.73	0.92	0.68	0.52	**0.47**
August–December	0.70	1.57	1.13	0.75	2.07	0.68	0.54	**0.50**
September–December	0.71	1.65	1.06	0.78	2.67	0.71	0.63	**0.61**
October–December	0.73	1.70	0.54	0.83	0.52	0.74	0.45	**0.40**
November–December	0.71	1.51	0.42	0.82	0.13	0.76	**0.11**	0.12
December	0.56	1.14	**0.07**	0.66	0.15	0.51	0.12	0.14
Mean	0.69	1.45	0.98	0.74	1.04	0.69	0.49	**0.48**

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)$_{12}$.

To compare the point forecast accuracy, we calculated the averaged MAE and MSE shown in Table 2, for all methods investigated with different updating months in the testing sample. Although the TS method performs better than the

Table 3
Empirical coverage probabilities of the TS, BM and PLS prediction intervals constructed parametrically and nonparametrically. The minimal mean coverage probability deviance is marked in bold, respectively.

Update period	Parametric				Nonparametric					
	TS		BM		TS		BM		PLS	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
March–December	0.9643	0.9714	0.9571	0.9786	0.9357	0.9714	0.9500	0.9714	0.8929	0.9500
April–December	0.9603	0.9683	0.9444	0.9762	0.9286	0.9683	0.9365	0.9683	0.8968	0.9524
May–December	0.9554	0.9643	0.9554	0.9643	0.9286	0.9643	0.9464	0.9643	0.9196	0.9554
June–December	0.9490	0.9592	0.9490	0.9592	0.9286	0.9592	0.9388	0.9490	0.9082	0.9490
July–December	0.9405	0.9524	0.9167	0.9643	0.9167	0.9524	0.9167	0.9405	0.8690	0.9405
August–December	0.9286	0.9429	0.9000	0.9429	0.9143	0.9429	0.9429	0.9429	0.8571	0.9286
September–December	0.9107	0.9286	0.8750	0.9464	0.8929	0.9286	0.8929	0.9464	0.8571	0.9286
October–December	0.9048	0.9286	0.8810	0.9286	0.8810	0.9286	0.9048	0.9286	0.8810	0.9048
November–December	0.9286	0.9286	0.8929	0.9643	0.8929	0.9286	0.9286	0.9286	0.8929	0.9286
December	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	0.0189	0.0214	0.0158	0.0302	**0.0140**	**0.0210**	0.0149

Table 4
Width of the TS, BM and PLS prediction intervals at the 90% and 95% nominal coverage probabilities. The minimal mean prediction interval width is marked in bold, respectively.

Update period	Parametric				Nonparametric					
	TS		BM		TS		BM		PLS	
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
March–December	3.06	3.65	3.06	3.64	3.05	3.55	3.05	3.51	2.70	3.15
April–December	3.13	3.73	3.13	3.73	3.12	3.62	3.13	3.66	2.77	3.21
May–December	3.09	3.69	3.10	3.69	3.08	3.57	3.13	3.61	2.77	3.21
June–December	3.01	3.58	3.00	3.58	3.00	3.47	3.00	3.50	2.66	3.05
July–December	2.92	3.47	2.91	3.46	2.91	3.38	2.84	3.41	2.49	2.90
August–December	2.80	3.34	2.80	3.33	2.80	3.26	2.85	3.37	2.25	2.61
September–December	2.74	3.26	2.74	3.26	2.74	3.19	2.74	3.25	2.34	2.82
October–December	2.74	3.27	2.75	3.28	2.75	3.20	2.73	3.23	2.40	2.78
November–December	2.71	3.23	2.72	3.24	2.71	3.16	2.74	3.26	2.33	2.69
December	2.68	3.19	2.67	3.18	2.68	3.12	2.76	3.30	2.08	2.48
Mean prediction interval width	2.89	3.44	2.89	3.44	2.88	3.35	2.90	3.41	**2.48**	**2.89**

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 averaged MAE and MSE, followed by the PLS, BM and OLS methods.

6.3. Updating interval forecasts

Supposing we observe the sea surface temperatures from January to February 2008, it is possible to dynamically update the interval forecasts for the rest 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

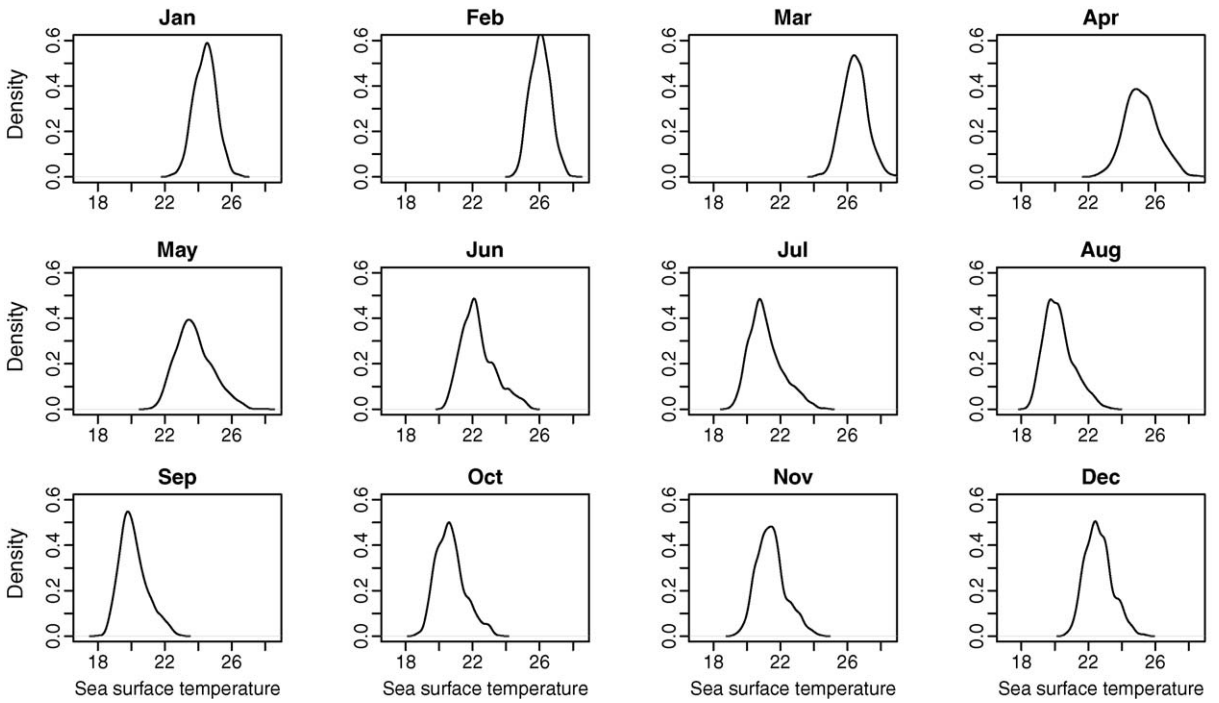

Fig. 7. Kernel density plots of the monthly sea surface temperatures in 2008. The bandwidth of kernel density plots is selected using a pilot estimation of derivatives.

smoothing state space model. Utilizing the relationship between the $\hat{\boldsymbol{\beta}}^{b,\mathrm{TS}}_{n+1|n}$ and $\hat{\boldsymbol{\beta}}^{b,\mathrm{PLS}}_{n+1|n}$, the PLS prediction intervals for the updating periods can be obtained from (6). For instance, Fig. 6 presents the 95% prediction intervals using the TS, BM and PLS methods for the sea surface temperatures from March to December 2008.

From Fig. 6, the PLS prediction intervals are comparably narrower than the TS and BM prediction intervals. Thus, they provide more informative evaluation of forecast uncertainty, subject to the same coverage probability. Furthermore, we also examine the average coverage probability deviance and the average width of prediction intervals using different updating months, which are 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, Fig. 7 displays the kernel density plots of the monthly sea surface temperatures in 2008 based on $B = 1000$ replications.

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. 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 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 Aneiros-Pérez and Vieu [1], Antoch et al. [2], Antoniadis and Sapatinas [3], Besse et al. [4], Ferraty and Vieu [15, Chapter 12]. 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 may, however, suffer from the singularity problem when the number of partially observed data points are less than the number of functional 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 coefficient estimates that deviate significantly from $\hat{\boldsymbol{\beta}}^{\mathrm{TS}}_{n+1|n}$. Based on the averaged MAE and MSE in the testing sample, the RR and PLS methods show better point forecast accuracy than other methods investigated and the difference between them is almost negligible.

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 kernel density plots and to construct prediction intervals for the updated forecasts. With a similar empirical coverage probability, the prediction interval width of the updated forecasts is narrower, thus the PLS and BM methods provide more informative evaluation of forecast uncertainty than the TS method without updating.

A natural direction for future research is to incorporate the smooth property of functions into the proposed nonparametric forecasting method and dynamic updating methods. One way of doing this is to apply a nonparametric smoothing technique to smooth out the noise component in a data set, prior to applying the proposed methods. This may result in a better point forecast accuracy and interval forecast accuracy. Another direction for future research is to apply the proposed methods to other univariate and multivariate seasonal time series problems, although we demonstrated our methods using the sea surface temperature data set in this paper.

The aforementioned approaches may seem complicated for calculating point forecasts, updating point forecasts, and constructing parametric and nonparametric prediction intervals, but their implementation is rather easy using the *ftsa* package of Hyndman and Shang [30].

Acknowledgements

This paper is an extended and revised version of the paper presented at the 18th World IMACS Congress and MODSIM09 International Congress on Modeling and Simulation. We thank the participants of the Congress, two

reviewers and editors of *Mathematics and Computers in Simulation* for their comments, which have improved the paper.

References

[1] G. Aneiros-Pérez, P. Vieu, Nonparametric time series prediction: a semi-functional partial linear modeling, Journal of Multivariate Analysis 99 (5) (2008) 834–857.
[2] J. Antoch, L. Prchal, M.R. De Rosa, P. Sarda, 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, 2008, pp. 23–29.
[3] A. Antoniadis, T. Sapatinas, Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes, Journal of Multivariate Analysis 87 (1) (2003) 133–158.
[4] P.C. Besse, H. Cardot, D.B. Stephenson, Autoregressive forecasting of some functional climatic variations, Scandinavian Journal of Statistics 27 (4) (2000) 673–687.
[5] D. Bosq, Linear Processes in Function Spaces: Theory and Applications, Springer, Berlin, 2000.
[6] D. Bosq, D. Blanke, Inference and Prediction in Large Dimensions, John Wiley, Chichester, England, 2007.
[7] G.E.P. Box, G.M. Jenkins, G.C. Reinsel, Time Series Analysis: Forecasting and Control, 4th ed., John Wiley, Hoboken, NJ, 2008.
[8] T. Cai, P. Hall, Prediction in functional linear regression, Annals of Statistics 34 (5) (2006) 2159–2179.
[9] C. Chatfield, Calculating interval forecasts, Journal of Business & Economic Statistics 11 (2) (1993) 121–135.
[10] C. Chatfield, Time Series Forecasting, Chapman & Hall, Boca Raton, 2000.
[11] J. Dauxois, A. Pousse, Y. Romain, Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference, Journal of Multivariate Analysis 12 (1) (1982) 136–154.
[12] A. Delaigle, P. Hall, T.V. Apanasovich, Weighted least squares methods for prediction in the functional data linear model, Electronic Journal of Statistics 3 (2009) 865–885.
[13] T. Dioses, R. Dávalos, J. Zuzunaga, El Niño 1982–1983 and 1997–1998: effects on Peruvian Jack Mackerel and Peruvian Chub Mackerel, Investigaciones Marinas 30 (1) (2002) 185–187.
[14] F. Ferraty, A. Rabhi, P. Vieu, Conditional quantiles for dependent functional data with application to the climatic El Niño phenomenon, Sankhya, The Indian Journal of Statistics 67 (2) (2005) 378–398.
[15] F. Ferraty, P. Vieu, Nonparametric Functional Data Analysis: Theory and Practice, Springer, New York, 2006.
[16] E. Greenshtein, Best subset selection, persistence in high-dimensional statistical learning and optimization under l_1 constraint, Annals of Statistics 34 (5) (2006) 2367–2386.
[17] E. Greenshtein, Y. Ritov, Persistence in high-dimensional linear predictor selection and the virtue of overparameterization, Bernoulli 10 (6) (2004) 971–988.
[18] P. Hall, J.L. Horowitz, Methodology and convergence rates for functional linear regression, Annals of Statistics 35 (1) (2007) 70–91.
[19] P. Hall, M. Hosseini-Nasab, On properties of functional principal components analysis, Journal of the Royal Statistical Society: Series B 68 (1) (2006) 109–126.
[20] P. Hall, M. Hosseini-Nasab, Theory for high-order bounds in functional principal components analysis, Mathematical Proceedings of the Cambridge Philosophical Society 146 (1) (2009) 225–256.
[21] P. Hall, H.-G. Müller, J.-L. Wang, Properties of principal component methods for functional and longitudinal data analysis, Annals of Statistics 34 (3) (2006) 1493–1517.
[22] A. Harvey, Forecasting, Structural Time Series Models and the Kalman Filter, Cambridge University Press, Cambridge, 1990.
[23] A.E. Hoerl, R.W. Kennard, Ridge regression: biased estimation for nonorthogonal problems, Technometrics 12 (1) (1970) 55–67.
[24] O. Hössjer, C. Croux, Generalizing univariate signed rank statistics for testing and estimating a multivariate location parameter, Journal of Nonparametric Statistics 4 (3) (1995) 293–308.
[25] R.J. Hyndman, H. Booth, Stochastic population forecasts using functional data models for mortality, fertility and migration, International Journal of Forecasting 24 (3) (2008) 323–342.
[26] R.J. Hyndman, Y. Khandakar, Automatic time series forecasting: the forecast package for R, Journal of Statistical Software 27 (3) (2008).
[27] R.J. Hyndman, A.B. Koehler, J.K. Ord, R.D. Snyder, Forecasting with Exponential Smoothing: the State Space Approach, Springer, Berlin, 2008.
[28] R.J. Hyndman, A.B. Koehler, R.D. Snyder, S. Grose, A state space framework for automatic forecasting using exponential smoothing methods, International Journal of Forecasting 18 (3) (2002) 439–454.
[29] R.J. Hyndman, H.L. Shang, Forecasting functional time series (with discussion), Journal of the Korean Statistical Society 38 (3) (2009) 199–221.
[30] R.J. Hyndman, H.L. Shang, ftsa: Functional time series analysis, R package version 1.6 (2010). URL http://CRAN.R-project.org/package=ftsa.
[31] R.J. Hyndman, H.L. Shang, Rainbow plots, bagplots, and boxplots for functional data, Journal of Computational and Graphical Statistics 19 (1) (2010) 29–45.
[32] R.J. Hyndman, M.S. Ullah, Robust forecasting of mortality and fertility rates: A functional data approach, Computational Statistics & Data Analysis 51 (10) (2007) 4942–4956.
[33] A.J. Izenman, Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning, Springer, New York, 2008.
[34] E.F. Moran, R. Adams, B. Bakoyéma, F.T. Stefano, B. Boucek, Human strategies for coping with El Niño related drought in Amazônia, Climatic Change 77 (3–4) (2006) 343–361.
[35] C. Preda, G. Saporta, Clusterwise PLS regression on a stochastic process, Computational Statistics & Data Analysis 49 (1) (2005) 99–108.
[36] C. Preda, G. Saporta, PLS regression on a stochastic process, Computational Statistics & Data Analysis 48 (1) (2005) 149–158.

[37] J.O. Ramsay, B.W. Silverman, Applied Functional Data Analysis: Methods and Case Studies, Springer, New York, 2002.
[38] J.O. Ramsay, B.W. Silverman, Functional Data Analysis, 2nd ed., Springer, New York, 2005.
[39] P.T. Reiss, T.R. Ogden, Functional principal component regression and functional partial least squares, Journal of the American Statistical Association 102 (479) (2007) 984–996.
[40] J.A. Rice, B.W. Silverman, Estimating the mean and covariance structure nonparametrically when the data are curves, Journal of the Royal Statistical Society: Series B 53 (1) (1991) 233–243.
[41] S.J. Sheather, M.C. Jones, A reliable data-based bandwidth selection method for kernel density estimation, Journal of the Royal Statistical Society: Series B 53 (3) (1991) 683–690.
[42] H. Shen, On modeling and forecasting time series of smooth curves, Technometrics 51 (3) (2009) 227–238.
[43] H. Shen, J.Z. Huang, Interday forecasting and intraday updating of call center arrivals, Manufacturing and Service Operations Management 10 (3) (2008) 391–410.
[44] B.W. Silverman, Incorporating parametric effects into functional principal components, Journal of the Royal Statistical Society: Series B 57 (4) (1995) 673–689.
[45] B.W. Silverman, Smoothed functional principal components analysis by choice of norm, Annals of Statistics 24 (1) (1996) 1–24.
[46] A. Timmermann, J. Oberhuber, A. Bacher, M. Esch, M. Latif, E. Roeckner, Increased El Niño frequency in a climate model forced by future greenhouse warming, Nature 398 (6729) (1999) 694–697.
[47] W.N. Venables, B.D. Ripley, Modern Applied Statistics with S, 4th ed., Springer, New York, 2002.