ISSN 1440-771X

MONASH University

Department of Econometrics and Business Statistics

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

Nonparametric modeling and forecasting electricity demand: an empirical study

Han Lin Shang

October 2010

Working Paper 19/10

Nonparametric modeling and forecasting electricity demand: an empirical study

Han Lin Shang
Department of Econometrics and Business Statistics,
Monash University, Caulfield East,
Melbourne, VIC 3145,
Australia.
Email: HanLin.Shang@monash.edu.au

18 October 2010

JEL classification: C88, C63, C14, C22

Nonparametric modeling and forecasting electricity demand: an empirical study

Abstract

This paper uses half-hourly electricity demand data in South Australia as an empirical study of nonparametric modeling and forecasting methods for prediction from half-hour ahead to one year ahead. A notable feature of the univariate time series of electricity demand is the presence of both intraweek and intraday seasonalities. An intraday seasonal cycle is apparent from the similarity of the demand from one day to the next, and an intraweek seasonal cycle is evident from comparing the demand on the corresponding day of adjacent weeks. There is a strong appeal in using forecasting methods that are able to capture both seasonalities. In this paper, the forecasting methods slice a seasonal univariate time series into a time series of curves. The forecasting methods reduce the dimensionality by applying functional principal component analysis to the observed data, and then utilize an univariate time series forecasting method and functional principal component regression techniques. When data points in the most recent curve are sequentially observed, updating methods can improve the point and interval forecast accuracy. We also revisit a nonparametric approach to construct prediction intervals of updated forecasts, and evaluate the interval forecast accuracy.

Keywords: functional principal component analysis; functional time series; multivariate time series, ordinary least squares, penalized least squares; ridge regression; seasonal time series

1 Introduction

Forecasting electricity demand is becoming more and more important, as the costs of power generation increase, and market competition intensifies. Research on electricity demand forecasting usually consider three major problems: long-term forecasts for generator planning, medium-term forecasts for generator maintenance, and short-term forecasts for daily operation. Accurate forecasts of electricity demand are relevant to energy sector for scheduling generator planning and maintenance.

The rapid development in electricity demand forecasting has been reflected in many contributions in the special issue (vol 24, issue 4) of the *International Journal of Forecasting* on energy forecasting in 2008. Among many forecasting methods, the popular techniques include artificial neural network (Hippert et al. 2001), Bayesian approach (Cottet & Smith 2003), ARIMA models (Weron 2006), exponential smoothing state space models (Taylor 2003), principal component analysis (Taylor & McSharry 2007), regression models and least-squares (Bajay 1983), and unobserved components method (Harvey & Koopman 1993, Pedregal & Young 2008). In the more recent literature, Hyndman & Fan (2010) utilized a semiparametric regression to forecast long-term electricity peak demand, while Goia et al. (2010) forecasted medium-term electricity demand through the viewpoint of functional data analysis.

In this article, we revisit some nonparametric modeling and forecasting methods using a functional data analytic approach. In contrast to Goia et al. (2010), we focus on the issue of short-term electricity demand forecasting. We revisit some forecast updating methods, as the data points are sequentially observed. This situation arises most frequently when a seasonal univariate time series of electricity demand is sliced into segments and treated as a time series of curves (also known as functional time series (Hyndman & Shang 2009)). The idea of forming a functional time series from a seasonal univariate time series has been considered by several authors, including Aneiros-Pérez & Vieu (2008), Antoch et al. (2008), Antoniadis & Sapatinas (2003), Besse et al. (2000), Ferraty & Vieu (2006, Chapter.12). However, little attention has been given to the practical problem of forecasting when the data points in the most recent curve are incompletely observed, with exceptions of Shen & Huang (2008) and Shang & Hyndman (2010).

We demonstrate the methods using half-hourly electricity demand (in Megawatts) in South Australia, from 6/7/1997 to 31/3/2007. Since the intra-daily pattern of electricity demand varies, the data set is divided into seven weekly data sets of electricity demand from Monday to Sunday. As an example, let $\{Z_w, w \in [1, N]\}$ be a seasonal univariate time series of half-hourly electricity demand on Mondays from 7/7/1997 to 26/3/2007, which has been observed at N = 24384 discrete time points. To model and forecast the univariate time series, a nonparametric method is introduced by adapting the ideas from functional data analysis (Ramsay & Silverman 2005). We divide the observed 24384 discrete time points into $n = 508$ trajectories, and then consider each trajectory of length $p = 48$ as a curve. The functional time series is given by

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

The problem of interest is to forecast the data in week $n+h$, $y_{n+h}(x)$, from the historical curves $\{y_1(x), \dots, y_n(x)\}$.

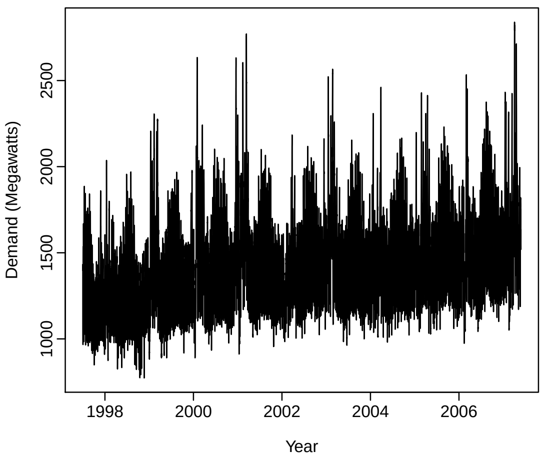
When $N = np$, all trajectories are complete, and forecasting is straightforward with several possible functional methods. These methods include the functional autoregressive of order 1 (Bosq 2000, Bosq & Blanke 2007), functional linear regression (Ramsay & Silverman 2005, Goia et al. 2010), functional kernel regression (Aneiros-Pérez & Vieu 2008, Ferraty & Vieu 2006), functional principal component regression (Hyndman & Ullah 2007, Hyndman & Booth 2008, Hyndman & Shang 2009), and functional partial least squares regression (Preda & Saporta 2005*a*,*b*, Hyndman & Shang 2009). Sections 3.1 and 3.2 present an example of applying functional principal component regression to model and forecast future curves.

In this article, the nonparametric modeling and forecasting methods are all based on functional principal component analysis (FPCA). Using FPCA, a time series of curves is decomposed into a number of functional principal components and their uncorrelated principal component scores. Using a univariate time series forecasting method, we can forecast principal component scores individually. Conditioning on the historical curves and the fixed functional principal components, the 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".

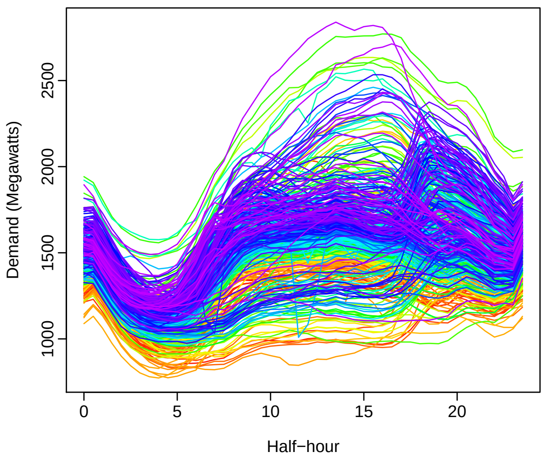
Section 2 introduces the motivated data set. In Section 3, we revisit the nonparametric method utilizing FPCA. Section 4 reviews briefly four updating methods of Shang & Hyndman (2010) to address the problem when the most recent curve is partially observed. These updating methods can also improve the point and interval forecast accuracy. This paper differs from Shang & Hyndman (2010), where the contribution is on the application of short-term electricity demand forecasting. In Section 5, we introduce a nonparametric method to construct prediction intervals for the updated forecasts. The evaluation and comparison of the point and interval forecast accuracy are given in Section 6. Conclusions are presented in Section 7.

2 Data set

The data set consists of half-hourly electricity demand in South Australia from 6/7/1997 to 31/3/2007. These data were obtained from Australian Energy Market Operator (http://www.aemo.com.au/). As a vehicle of illustration, we consider the half-hourly electricity demand on Mondays (In the data analyses, we consider the half hourly electricity demand from Mondays to Sundays). A univariate time series display of electricity demand on Mondays from 7/7/1997 to 26/3/2007 is presented in Figure 1a, with the same data shown in Figure 1b as a time series of curves.



(a) *A univariate time series display of electricity demand on Mondays. There are 24384 discrete time points. Each time point represents one dimension.*



(b) *A functional time series display of electricity demand on Mondays. There are 508 curves. Each curve has 48 dimensions.*

Figure 1: *Exploratory plots suggesting that both regular pattern and extreme electricity demand are presented in the Monday electricity demand data between 7/7/1997 and 26/3/2007.*

In Figure 1b, there are some weeks showing extreme electricity demand and are suspected to be outliers. Because the presence of outliers can seriously affect the performance of modeling and forecasting, we applied the outlier detection method of Hyndman & Shang (2010*b*). This outlier detection method applies functional principal component analysis to reduce the dimensionality of original curves down to two, and it detects an outlier if it is far from the center of bivariate (first two) principal component scores. As a surrogate of original curves, the bivariate principal component scores can be easily plotted via bivariate bagplot of Rousseeuw et al. (1999), from which outliers and inliers are separated. The detected outliers correspond to the following dates (15/11/1998, 14/1/2001, 18/2/2001, 19/1/2003, 15/2/2004, 28/11/2004, 22/1/2006, 5/3/2006, 10/12/2006, 4/2/2007, and 18/2/2007). These outliers reflect the extremely high electricity demand during the summer season from December to February and holiday period in South Australia. Consequently, they have been removed from further analyses.

3 Forecasting method

3.1 Functional principal component analysis (FPCA)

The forecasting method utilizes FPCA, which plays an important 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 & Vieu (2006) and Ramsay & Silverman (2002, 2005). Papers covering the development of FPCA include those of Hyndman & Shang (2009), Hyndman & Ullah (2007), Reiss & Ogden (2007), Rice & Silverman (1991), Shen (2009) and Silverman (1995, 1996). Significant treatments of the theory of FPCA are given by Cai & Hall (2006), Dauxois et al. (1982), Delaigle et al. (2009), Hall et al. (2006), Hall & Horowitz (2007) and Hall & Hosseini-Nasab (2006, 2009).

When all trajectories are complete, the forecasting method begins by subtracting the time-varying functional mean from the original functional data. The time-varying functional mean $\mu(x)$ is estimated by

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

where $\{y_1(x),\dots,y_n(x)\}$ is a time series of curves, which can be obtained using a linear interpolation method. If one seeks a robust estimator, then the L_1 median of data should be used, and is denoted by

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

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

Using FPCA, $\{y_1(x)-\hat{\mu}(x),\dots,y_n(x)-\hat{\mu}(x)\}$ can be approximated by the sum of orthogonal functional principal components and their associated principal component scores:

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

where $\{\phi_1(x),\cdots,\phi_K(x)\}$ represents a set of functional principal components, $\{\hat{\beta}_{1,t},\dots,\hat{\beta}_{K,t}\}$ represents a set of estimated principal component scores, $e_t(x)$ is the zero-mean residual function, and $K < n$ is the number of functional principal components. The optimal value of K in a given data set can be determined using a holdout method (see Section 4.5 for detail).

3.2 Point forecasts

Because the principal component scores are uncorrelated to each other, it is appropriate to forecast each series $\{\hat{\beta}_{k,1},\dots,\hat{\beta}_{k,n}; k=1,\dots,K\}$ using univariate time series models, such as the ARIMA models (Box et al. 2008). It is noteworthy that the lagged cross correlations are not necessarily zero, but they are likely to be small because the contemporaneous correlations are zero (Hyndman & Ullah 2007, Shen & Huang 2008).

Conditioning on the historical curves $\mathcal{I}$ and the fixed functional principal components $\mathbf{\Phi} = \{\phi_1(x),\dots,\phi_K(x)\}$, the forecasted curves are expressed as

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

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

4 Updating point forecasts

When the functional time series are segments of a seasonal 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 $y_{n+1}(x_e) = [y_{n+1}(x_1),\dots,y_{n+1}(x_{m_0})]^{'}$, we are interested in forecasting the data in the remainder of week $n+1$, denoted by $y_{n+1}(x_l)$. However, the TS method described in Section 3 does not utilize the most recent data, namely the sequentially observed data points in the most recent curve. Instead, using (2), the time series forecast of $y_{n+1}(x_l)$ is given by

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

where $\mathcal{I}_l$ denotes the historical curves corresponding to the remaining time periods; $\mathbf{\Phi}_l = \{\phi_1(x_l), \ldots, \phi_K(x_l)\}$ is a set of the functional principal components corresponding to the remaining time periods; $\hat{\mu}(x_l)$ is the time-varying mean function corresponding to the remaining time periods.

In order to improve point forecast accuracy, it is desirable to update the point forecasts for the rest of week $n+1$ by incorporating the partially observed data. To address this issue, we review briefly four updating methods recently proposed by Shang & Hyndman (2010), and apply them to the electricity demand data.

4.1 Block moving (BM) method

The BM method simply redefines the start and end points of our "week" (the time for a single trajectory). Because time is a continuous variable, we can change the support of our trajectories from $[1, p]$ to $[m_0+1, p] \cup [1, m_0]$.

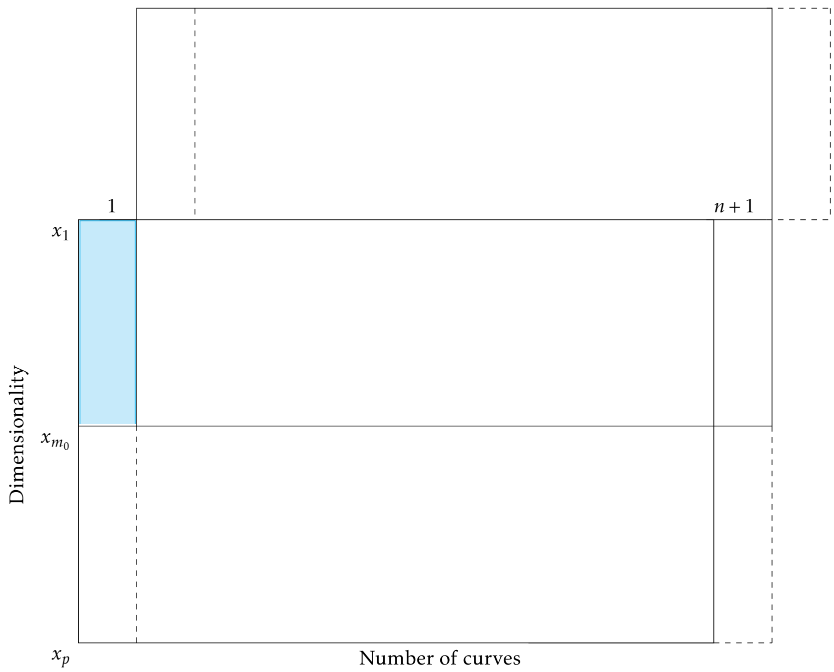


Figure 2: *Update via the block moving approach. The colored region shows the data loss in the first week. The forecasts for the rest of week $n+1$ can be updated by the forecasts using the TS method applied to the top block.*

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 loss in the first week. The loss of data in the first week will have minimal effect on the forecasts, if the number of curves is large.

The partially observed last trajectory under the old function support range completes the last trajectory under the new function support range. The forecasts can be obtained by applying the TS method to the new complete data block.

4.2 Ordinary least squares (OLS) method

We can model and forecast the remaining part of the last trajectory using a regression, based on the functional principal components obtained in (1). Let $\mathcal{F}_e$ be $m_0 * K$ matrix whose $(j,k)^{\text{th}}$ entry is $\phi_k(x_j)$ for $1 \leq j \leq m_0$ and $1 \leq k \leq K$. Let $\boldsymbol{\beta}_{n+1} = [\beta_{1,n+1},\dots,\beta_{K,n+1}]'$, and $\boldsymbol{\epsilon}_{n+1}(x_e) = [\epsilon_{n+1}(x_1),\dots,\epsilon_{n+1}(x_{m_0})]'$. 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

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

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

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

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 remaining time periods of week $n+1$, but it needs a sufficiently large number of observations (at least equal to K) in order for $\hat{\boldsymbol{\beta}}^{\text{OLS}}_{n+1} = [\hat{\beta}^{\text{OLS}}_{1,n+1},\dots,\hat{\beta}^{\text{OLS}}_{K,n+1}]'$ to be numerically stable. To address this problem, we adapt the ridge regression (RR) method of Hoerl & Kennard (1970) with the predictors being the corresponding functional principal components and the partially observed data being the responses. The advantage of RR method is that it uses a square penalty function, which is rotationally invariant hypersphere centered at the origin (Izenman 2008). Thus, the regression coefficient estimates of the RR method have a closed form.

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

$$\underset{\beta_{n+1}}{\operatorname{argmin}}\left\{[\hat{\boldsymbol{y}}^*_{n+1}(x_e)-\mathcal{F}_e\boldsymbol{\beta}_{n+1}]'[\hat{\boldsymbol{y}}^*_{n+1}(x_e)-\mathcal{F}_e\boldsymbol{\beta}_{n+1}]+\lambda\boldsymbol{\beta}_{n+1}^{'}\boldsymbol{\beta}_{n+1}\right\}, \tag{3}$$

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

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

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

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

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

4.4 Penalized least squares (PLS) method

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

$$\underset{\beta_{n+1}}{\arg\min}\left\{[\hat{\boldsymbol{y}}^*_{n+1}(x_e)-\mathcal{F}_e\hat{\boldsymbol{\beta}}_{n+1}]^{'}[\hat{\boldsymbol{y}}^*_{n+1}(x_e)-\mathcal{F}_e\hat{\boldsymbol{\beta}}_{n+1}]+\lambda(\hat{\boldsymbol{\beta}}_{n+1}-\hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1|n})^{'}(\hat{\boldsymbol{\beta}}_{n+1}-\hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1|n})\right\}. \tag{4}$$

The first term in (4) measures the "goodness of fit", while the second term penalized the departure of the regression coefficient estimates from the TS forecasted regression coefficient estimates. The $\hat{\boldsymbol{\beta}}^{\text{PLS}}_{n+1}$ 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 $\hat{\boldsymbol{\beta}}_{n+1}$ in (4), we obtain

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{n+1}^{\text{PLS}} &= (\mathcal{F}_e{}'\mathcal{F}_e + \lambda \boldsymbol{I}_K)^{-1}[\mathcal{F}_e{}'\hat{\boldsymbol{y}}^*_{n+1}(x_e) + \lambda \hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1|n}] \\
&= \frac{\mathcal{F}_e{}'\hat{\boldsymbol{y}}^*_{n+1}(x_e)}{\mathcal{F}_e{}'\mathcal{F}_e + \lambda \boldsymbol{I}_K} + \frac{\lambda \hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1|n}}{\mathcal{F}_e{}'\mathcal{F}_e + \lambda \boldsymbol{I}_K} \\
&= \frac{\mathcal{F}_e{}'\mathcal{F}_e}{\mathcal{F}_e{}'\mathcal{F}_e + \lambda \boldsymbol{I}_K}\hat{\boldsymbol{\beta}}^{\text{OLS}}_{n+1} + \frac{\lambda}{\mathcal{F}_e{}'\mathcal{F}_e + \lambda \boldsymbol{I}_K}\hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1|n} \\
&= \left(\boldsymbol{I}_K - \frac{\lambda \boldsymbol{I}_K}{\mathcal{F}_e{}'\mathcal{F}_e + \lambda \boldsymbol{I}_K}\right)\hat{\boldsymbol{\beta}}^{\text{OLS}}_{n+1} + \frac{\lambda \boldsymbol{I}_K}{\mathcal{F}_e{}'\mathcal{F}_e + \lambda \boldsymbol{I}_K}\hat{\boldsymbol{\beta}}^{\text{TS}}_{n+1|n}.
\end{aligned}$$

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

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

$$\hat{y}^{\text{PLS}}_{n+1}(x_l) = \text{E}[y_{n+1}(x_l)|\mathcal{I}_l, \boldsymbol{\Phi}_l] = \hat{\mu}(x_l) + \sum_{k=1}^{K} \phi_k(x_l)\hat{\beta}^{\text{PLS}}_{k,n+1}. \tag{5}$$

4.5 Selections of penalty parameter and number of components

We split the data into a training sample and a testing sample (including one-year electricity demand from $n-51$ to n weeks, where n denotes the total number of weeks, excluding the outliers). Within the training sample, we further split the data into a training set and a validation set (including electricity demand from $n-103$ to $n-52$ weeks, excluding the outliers).

For a set of possible number of principal components $K = 1, 2, \dots, 10$, we apply the TS method to the training set and obtain forecasts for the data in the validation set. The optimal number of component is determined by minimizing the mean absolute percentage error (MAPE) criterion within the validation set. In the Monday electricity demand data, the optimal number of components is $K = 3$.

The optimal values of penalty parameter λ for different updating periods are also determined by minimizing the MAPE criterion within the validation set. The MAPE criterion is the most widely used error measure in electricity demand forecasting. It can be expressed as

$$\text{MAPE} = \frac{1}{pq}\sum_{j=1}^{q}\sum_{i=1}^{p}\left|\frac{y_{m-j+1}(x_i) - \hat{y}_{m-j+1|m-j}(x_i)}{y_{m-j+1}(x_i)}\right| * 100,$$

where p represents the number of observations in each week, q represents the number of weeks in the validation set, and m denotes the index corresponding to the maximum number of weeks in the validation set (i.e., $m = n - 52$ in the above forecast setting). In Table 1, the optimal tuning parameters for different updating periods on Mondays are given for both the RR and PLS methods (Due to limited space and repetition, the optimal tuning parameters for different updating periods from Tuesdays to Sundays are availble upon request from the author).

Updating periods	RR	PLS	Updating periods	RR	PLS
1:00—23:30	0.3532	0.1378	12:30—23:30	0.4769	0.6998
1:30—23:30	0.3960	0.2038	13:00—23:30	0.4769	0.7376
2:00—23:30	0.3820	0.2566	13:30—23:30	0.6459	0.7716
2:30—23:30	0.3293	0.2469	14:00—23:30	0.6805	0.8115
3:00—23:30	0.3921	0.2548	14:30—23:30	0.3558	0.6912
3:30—23:30	0.4321	0.3240	15:00—23:30	0.3820	0.6669
4:00—23:30	0.3289	0.2016	15:30—23:30	0.3460	0.7600
4:30—23:30	0.2209	0.2196	16:00—23:30	0.2427	0.6908
5:00—23:30	0.2000	0.2409	16:30—23:30	0.1860	0.2510
5:30—23:30	0.3011	0.2574	17:00—23:30	0.3350	0.9572
6:00—23:30	0.4060	0.3620	17:30—23:30	0.0792	0.9947
6:30—23:30	0.4822	0.7108	18:00—23:30	0.0048	0.9937
7:00—23:30	0.4932	0.9874	18:30—23:30	0.0048	0.9951
7:30—23:30	0.3262	0.3475	19:00—23:30	0.4590	0.5149
8:00—23:30	0.2196	0.2196	19:30—23:30	0.2475	0.1510
8:30—23:30	0.1379	0.4080	20:00—23:30	0.3421	0.1090
9:00—23:30	0.1492	0.4438	20:30—23:30	0.3213	0.0192
9:30—23:30	0.4566	0.4803	21:00—23:30	0.0899	0.0048
10:00—23:30	0.4918	0.5147	21:30—23:30	0.0048	0.0048
10:30—23:30	0.4185	0.5573	22:00—23:30	0.0048	0.0048
11:00—23:30	0.5505	0.5917	22:30—23:30	0.0048	0.0048
11:30—23:30	0.5729	0.6262	23:00—23:30	0.0048	0.0048
12:00—23:30	0.4671	0.6656	23:30—23:30	0.0048	0.0048

Table 1: *For different updating periods, the optimal tuning parameters used in the RR and PLS methods are determined by minimizing the MAPE criterion within the validation set on Mondays.*

5 Interval 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 residual. In Sections 5.1 and 5.2, we describe a parametric method and a nonparametric method to construct prediction intervals for the TS and BM methods. In Section 5.3, we revisit a nonparametric bootstrap method to update prediction intervals, by incorporating newly observed data in the most recent curve.

5.1 Parametric method to construct 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):

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

where $\hat{\zeta}_{k,n+h|n} = \text{Var}(\hat{\beta}_{k,n+h}|\hat{\beta}_1,\dots,\hat{\beta}_n)$ 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 week $n+h$, $\hat{\epsilon}^2_{n+h}(x)$, for each x variable. Under the normality assumption, the $100(1-\alpha)\%$ prediction intervals of $y_{n+h}(x)$ are constructed as follows,

$$\left(\hat{y}_{n+h|n}(x) - z_\alpha\sqrt{\hat{\vartheta}_{n+h|n}(x)},\quad \hat{y}_{n+h|n}(x) + z_\alpha\sqrt{\hat{\vartheta}_{n+h|n}(x)}\right),$$

where z_α is the $(1-\alpha/2)$ standard normal quantile. This will also work for the BM method with appropriately defined function support range.

5.2 Nonparametric method to construct prediction intervals

We review a nonparametric method used in Hyndman & Shang (2009) and Shang & Hyndman (2010) to construct prediction intervals for the TS and BM methods. We can obtain one-step-ahead forecasts for the principal component scores $\{\beta_{k,1},\dots,\beta_{k,n}; k = 1,\dots,K\}$, using a univariate time series model. Let the h-step-ahead forecast errors be given by $\hat{\pi}_{k,j} = \hat{\beta}_{k,n-j+1|n-j} - \hat{\beta}_{k,n-j+1}$, for $j = 1,\dots,n-K$. These can then be sampled with replacement to give a bootstrap sample of

$\beta_{k,n+h}$:

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

where $\hat{\pi}^b_{k,*}$ 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 $\hat{\xi}^b_{n+h|n}(x)$ by sampling with replacement from the residual term $\{\hat{\xi}_1(x),\dots,\hat{\xi}_n(x)\}$.

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 $y_{n+h|n}(x)$,

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

Hence, the $100(1-\alpha)\%$ prediction intervals are defined as:

$$\left(\hat{y}^{b,\frac{\alpha}{2}}_{n+h|n}(x), \quad \hat{y}^{b,1-\frac{\alpha}{2}}_{n+h|n}(x)\right).$$

This will also work for the BM method with appropriately defined function support range.

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, $\hat{\beta}^{b,\text{TS}}_{n+1|n}$, 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

$$\hat{y}^{b,\text{PLS}}_{n+1}(x_l) = \hat{\mu}(x_l) + \sum_{k=1}^{K} \phi_k(x_l)\hat{\beta}^{b,\text{PLS}}_{k,n+1} + \hat{\xi}^b_{n+1}(x_l).$$

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

5.4 Evaluating interval forecasts

According to Baillie & Bollerslev (1992), McNees & Fine (1996) and Christoffersen (1998), the standard evaluation of interval forecasts proceeds by simply comparing the nominal coverage probability to the empirical (conditional) coverage probability. The evaluation of empirical coverage probability was preformed as follows: for each curve in the testing sample, prediction intervals of one-step-ahead forecasts were computed parametrically and nonparametrically at the 95% nominal coverage probability, 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 that fall 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

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

The narrower the average width of prediction intervals is, the more informative 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 is 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 & Jones (1991). This bandwidth selection method is based on choosing the bandwidth that minimizes estimates of the mean integrated squared error, which seems to be close to optimal and generally preferred (Venables & Ripley 2002).

6 Results

6.1 Point forecasts

The forecasting method decomposes a functional time series into a number of functional principal components and their associated principal component scores. In the top panel of Figure 3, we display and attempt to interpret the first three functional principal components.

Clearly, the mean function illustrates a strong seasonal pattern, with a peak at 18:00 and a trough at 5:00. The functional principal components are of second order effects, as indicated by much smaller scales. The first functional principal component models electricity demand in the afternoon and evening. While the second functional principal component models the contrast in electricity demand between morning and evening, the third functional principal component models the contrast in electricity demand between morning and afternoon.

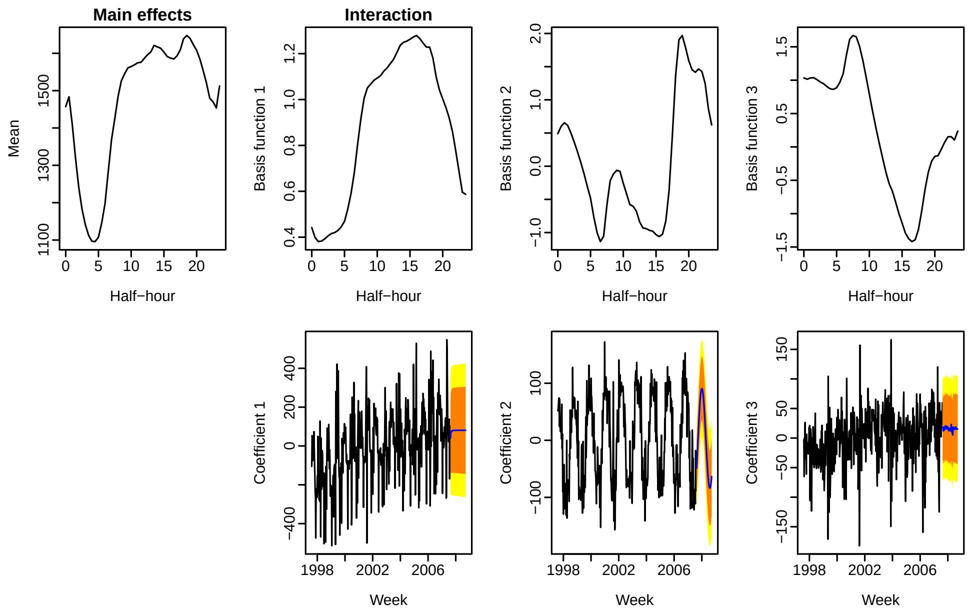


Figure 3: *The mean function, the first three functional principal components and their associated principal component scores for the Monday electricity demand from 7/7/1997 to 26/3/2007 (excluding the outliers). The 80% and 95% prediction intervals of the principal component scores are shown by the orange and yellow regions.*

The automatic ARIMA algorithm of Hyndman & Khandakar (2008) is a stepwise approach to select the optimal ARIMA model by minimizing Akaike Information Criterion and Bayesian Information Criterion. Using the automatic ARIMA algorithm, we obtained the optimal ARIMA model, from which the principal component scores are forecasted and their 80% and 95% prediction intervals are highlighted by the orange and yellow regions in the bottom panel of Figure 3.

By conditioning on the historical data and fixed functional principal components, the forecasts are obtained by multiplying the forecasted principal component scores with the fixed functional principal components. As an example, Figure 4 displays the forecasted Monday electricity demand in the last week of data (i.e., 26/3/2007), along with the 95% parametric and nonparametric prediction intervals. In this example, we found that the width of the parametric prediction intervals seems to be narrower than the nonparametric counterpart.

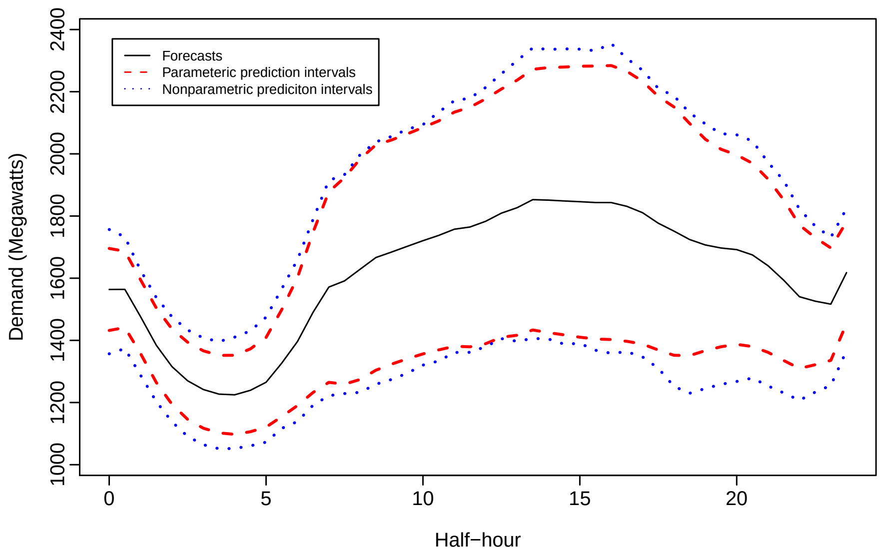


Figure 4: *Point forecasts of the Monday electricity demand in 26/3/2007, and the 95% prediction intervals constructed via the parametric and nonparametric approaches.*

6.2 Point forecast comparisons with some existing methods

By means of comparisons, we also investigate the point forecast accuracy of seasonal autoregressive moving average (SARIMA), random walk (RW), and mean predictor (MP) methods. The MP method consists in predicting values at week $t+1$ by the empirical mean value for each time variable from the first week to the t^{th} week. The RW approach predicts new values at week $t+1$ by the observations at week t. In the forecasting literature, SARIMA has been considered as a benchmark method for forecasting a seasonal time series (Besse et al. 2000, Ferraty et al. 2005). However, it requires the specification of the orders of seasonal components and non-seasonal components. The automatic ARIMA algorithm developed by Hyndman & Khandakar (2008) can be used to select the optimal orders for both seasonal and non-seasonal components.

To compare their point forecast accuracy, Table 2 shows the averaged MAPE of the 52 iterative one-step-ahead point forecasts for different updating periods in the testing sample. Using the data in the training set, we forecast the electricity demand for half-hour ahead and calculate the MAPE. By incorporating new observations into the training set, we successively forecast the electricity demand for half-hour ahead. The method that produces the minimal averaged MAPE across all time periods is considered to be the best one. As a result, the four updating methods

Updating periods	Non-dynamic updating methods				Dynamic updating methods			
	MP	RW	SARIMA	TS	BM	OLS	RR	PLS
1:00—23:30	9.9393	7.5067	6.4567	6.6380	6.3717	6.4030	6.4498	6.4658
1:30—23:30	10.0241	7.5869	6.5207	6.7136	6.3909	6.9219	6.8729	6.9087
2:00—23:30	10.1066	7.6658	6.5847	6.7903	6.4640	6.9013	6.9707	6.9871
2:30—23:30	10.1839	7.7447	6.6491	6.8665	6.5687	6.9425	6.8954	6.9066
3:00—23:30	10.2537	7.8235	6.7154	6.9415	6.6080	6.8430	6.8598	6.8891
3:30—23:30	10.3160	7.9072	6.7859	7.0169	6.6259	6.6180	6.5749	6.5989
4:00—23:30	10.3677	7.9955	6.8613	7.0926	6.7594	6.3496	6.3262	6.3400
4:30—23:30	10.4107	8.0871	6.9386	7.1681	6.8499	6.2161	6.2057	6.2122
5:00—23:30	10.4450	8.1816	7.0147	7.2402	6.9559	6.1238	6.1195	6.1226
5:30—23:30	10.4733	8.2761	7.0877	7.3086	7.0207	6.1269	6.1245	6.1265
6:00—23:30	10.4930	8.3617	7.1508	7.3662	7.1019	6.4182	6.7110	6.7301
6:30—23:30	10.5023	8.4420	7.2078	7.4126	7.1010	6.7486	6.7276	6.7384
7:00—23:30	10.4948	8.5095	7.2518	7.4442	7.1837	6.5580	6.5502	6.5561
7:30—23:30	10.4762	8.5693	7.2903	7.4666	7.2146	6.4237	6.4170	6.4201
8:00—23:30	10.4602	8.6223	7.3232	7.4958	7.1763	6.2799	6.2768	6.2784
8:30—23:30	10.4410	8.6686	7.3532	7.5267	7.2338	6.2303	6.2288	6.2298
9:00—23:30	10.4246	8.7111	7.3852	7.5592	7.2023	6.2327	6.2320	6.2326
9:30—23:30	10.4118	8.7559	7.4173	7.5946	7.2633	6.2085	6.2038	6.2069
10:00—23:30	10.3977	8.7951	7.4463	7.6252	7.3626	5.9750	5.9230	5.9370
10:30—23:30	10.3835	8.8267	7.4749	7.6492	7.4509	5.7481	5.7364	5.7458
11:00—23:30	10.3714	8.8509	7.5039	7.6654	7.2585	5.6179	5.6104	5.6177
11:30—23:30	10.3626	8.8691	7.5340	7.6743	7.0264	5.5516	5.5436	5.5506
12:00—23:30	10.3513	8.8818	7.5642	7.6806	6.8234	5.6612	5.6566	5.6613
12:30—23:30	10.3443	8.8844	7.5898	7.6805	6.7780	5.7994	5.7966	5.7999
13:00—23:30	10.3476	8.8761	7.6025	7.6655	6.7226	5.9145	5.9126	5.9151
13:30—23:30	10.3529	8.8479	7.6029	7.6337	6.7328	6.0473	6.0460	6.0479
14:00—23:30	10.3515	8.8088	7.5925	7.5919	6.5597	6.2160	6.2151	6.2165
14:30—23:30	10.3562	8.7542	7.5666	7.5324	6.3209	6.3974	6.3966	6.3978
15:00—23:30	10.3648	8.6846	7.5324	7.4562	6.0053	6.5974	6.5968	6.5977
15:30—23:30	10.3733	8.5930	7.4697	7.3573	5.7507	6.7292	6.7256	6.7266
16:00—23:30	10.3760	8.4770	7.3791	7.2308	5.5620	6.4529	6.4245	6.4253
16:30—23:30	10.3695	8.3352	7.2659	7.0752	5.4975	6.0517	6.0449	6.0444
17:00—23:30	10.3544	8.1706	7.1358	6.8912	5.4994	6.2606	6.2581	6.2576
17:30—23:30	10.3336	7.9876	7.0047	6.6822	5.5173	6.5472	6.5467	6.5463
18:00—23:30	10.2881	7.7883	6.8294	6.4493	5.6643	6.9187	6.9190	6.9186
18:30—23:30	10.1663	7.5652	6.6136	6.2043	5.6307	7.3121	7.3122	7.3121
19:00—23:30	9.9810	7.3339	6.4094	6.0053	5.4354	7.3216	7.3216	7.3217
19:30—23:30	9.7757	7.1090	6.2104	5.8395	5.2703	2.9426	2.9366	2.9356
20:00—23:30	9.5759	6.8838	6.0193	5.6792	5.0626	2.5905	2.5884	2.5882
20:30—23:30	9.3964	6.6490	5.8275	5.5161	4.8095	2.2986	2.2983	2.2979
21:00—23:30	9.2094	6.3991	5.6116	5.3385	4.5848	2.0978	2.0979	2.0976
21:30—23:30	8.9951	6.1284	5.3667	5.1449	4.4028	2.0618	2.0620	2.0618
22:00—23:30	8.7333	5.8180	5.0849	4.9302	4.1887	2.1685	2.1689	2.1687
22:30—23:30	8.4264	5.4437	4.7318	4.6839	4.0672	2.4173	2.4178	2.4176
23:00—23:30	8.2795	5.0244	4.3459	4.4646	3.8888	2.6172	2.6179	2.6175
23:30—23:30	8.9836	4.4998	3.9002	4.2196	4.5676	2.0160	2.0166	2.0162
Mean	10.0832	7.8848	6.7872	6.8089	6.1855	**5.5843**	5.5856	5.5911

Table 2: *Averaged MAPE of the 52 iterative one-step-ahead point forecasts from Monday to Sunday using the MP, RW, SARIMA, TS, BM, OLS, RR and PLS methods for different updating periods in the testing sample. The minimal value of the averaged MAPE is marked in bold.*

achieved better point forecast accuracy than the non-updating methods in general. Among the updating methods, the OLS, RR and PLS methods performed equally well for forecasting electricity demand.

We further carried out a pairwise t test to examine whether or not the difference of point forecast accuracy among methods is significant. Based on the MAPE across 46 different time periods, the pairwise t-test given in Table 3 indicates that the MP method differs significantly from all other methods, as is the RW method. The SARIMA method performs similarly with the TS and BM methods, but it differs significantly from other methods. The four updating methods differs significantly from other non-updating methods, but they perform similarly among each other.

	MP	RW	SARIMA	TS	BM	OLS	RR
RW	2.2e-14	-	-	-	-	-	-
SARIMA	< 2e-16	0.00039	-	-	-	-	-
TS	< 2e-16	0.00050	1.00000	-	-	-	-
BM	< 2e-16	5.4e-09	0.17832	0.16129	-	-	-
OLS	< 2e-16	1.5e-15	8.6e-05	7.2e-05	0.17832	-	-
RR	< 2e-16	1.5e-15	8.6e-05	7.2e-05	0.17832	1.00000	-
PLS	< 2e-16	1.7e-15	8.6e-05	7.2e-05	0.17832	1.00000	1.00000

Table 3: *Based on the MAPE across 46 different time periods, the p-values of paired t test statistics are calculated to test the difference in point forecast accuracy among methods.*

6.3 Updating interval forecasts

Supposing we observe the electricity demand from midnight to 18:30 in 26/3/2007, it is possible to dynamically update the interval forecasts for the remaining time periods using the BM and PLS methods. Based on the historical data on Mondays (excluding the outliers), we obtain the forecasted principal component scores using the automatic ARIMA models. Utilizing the relationship between $\hat{\beta}^{b,\text{TS}}_{n+1|n}$ and $\hat{\beta}^{b,\text{PLS}}_{n+1}$, the PLS prediction intervals for the updating time periods can be obtained from (5). As an example, Figure 5 presents the 95% prediction intervals obtained by the TS, BM and PLS methods for the electricity demand from 19:00 to 23:30.

From Figure 5, the PLS prediction intervals are comparably narrower than the prediction intervals of the TS and BM methods. Thus, they provide more informative evaluation of forecast uncertainty, subject to the same coverage probability. To compare the interval forecast accuracy, Table 4 shows the average coverage probability deviance and the average width of prediction intervals for different updating time periods in the testing sample.

Updating periods	Coverage probability deviance					Prediction interval width				
	Parametric		Nonparametric			Parametric		Nonparametric		
	TS	BM	TS	BM	PLS	TS	BM	TS	BM	PLS
1:00—23:30	0.017	0.016	0.014	0.037	0.047	562	560	647	643	625
1:30—23:30	0.018	0.017	0.016	0.033	0.045	569	567	653	653	633
2:00—23:30	0.018	0.017	0.013	0.028	0.047	576	574	662	662	639
2:30—23:30	0.019	0.018	0.015	0.035	0.049	583	582	667	667	646
3:00—23:30	0.018	0.018	0.016	0.039	0.047	591	589	680	674	648
3:30—23:30	0.018	0.018	0.014	0.036	0.048	599	597	681	683	649
4:00—23:30	0.018	0.017	0.015	0.036	0.052	607	605	697	688	660
4:30—23:30	0.018	0.017	0.014	0.038	0.048	616	614	705	697	670
5:00—23:30	0.018	0.017	0.014	0.041	0.049	625	623	714	707	680
5:30—23:30	0.019	0.018	0.014	0.039	0.048	634	631	722	715	690
6:00—23:30	0.019	0.018	0.016	0.034	0.052	643	640	736	723	693
6:30—23:30	0.019	0.020	0.015	0.035	0.048	651	647	746	736	705
7:00—23:30	0.020	0.020	0.017	0.034	0.048	658	654	755	733	704
7:30—23:30	0.019	0.019	0.014	0.035	0.049	663	659	761	740	713
8:00—23:30	0.020	0.018	0.016	0.038	0.053	669	663	770	743	696
8:30—23:30	0.019	0.019	0.020	0.035	0.049	673	667	771	745	691
9:00—23:30	0.020	0.019	0.017	0.039	0.054	677	672	781	748	676
9:30—23:30	0.020	0.019	0.017	0.043	0.055	681	678	782	758	669
10:00—23:30	0.021	0.020	0.017	0.045	0.050	684	682	791	761	673
10:30—23:30	0.021	0.021	0.016	0.047	0.059	686	683	790	756	661
11:00—23:30	0.021	0.022	0.020	0.036	0.060	688	678	792	738	654
11:30—23:30	0.020	0.020	0.023	0.037	0.060	689	670	795	722	650
12:00—23:30	0.020	0.023	0.019	0.034	0.055	689	660	794	712	646
12:30—23:30	0.019	0.024	0.023	0.042	0.054	688	651	791	703	629
13:00—23:30	0.020	0.023	0.020	0.038	0.055	686	640	793	691	628
13:30—23:30	0.022	0.023	0.022	0.045	0.057	683	626	787	683	618
14:00—23:30	0.024	0.024	0.021	0.050	0.055	679	607	781	683	612
14:30—23:30	0.024	0.031	0.021	0.052	0.057	673	584	770	681	602
15:00—23:30	0.024	0.027	0.017	0.065	0.053	666	556	758	694	593
15:30—23:30	0.024	0.025	0.022	0.054	0.054	658	532	748	709	582
16:00—23:30	0.025	0.029	0.023	0.042	0.050	647	522	732	719	572
16:30—23:30	0.025	0.030	0.022	0.039	0.052	634	523	713	722	535
17:00—23:30	0.025	0.031	0.022	0.038	0.062	620	528	698	725	501
17:30—23:30	0.026	0.031	0.021	0.041	0.054	604	534	679	706	493
18:00—23:30	0.029	0.028	0.025	0.033	0.062	586	535	658	685	428
18:30—23:30	0.029	0.029	0.019	0.037	0.076	567	529	637	661	367
19:00—23:30	0.029	0.024	0.019	0.065	0.039	548	516	609	619	450
19:30—23:30	0.029	0.026	0.024	0.051	0.048	528	493	584	583	367
20:00—23:30	0.029	0.024	0.026	0.042	0.054	509	474	564	546	340
20:30—23:30	0.030	0.021	0.020	0.037	0.121	489	451	540	513	261
21:00—23:30	0.030	0.019	0.023	0.027	0.138	467	426	509	489	208
21:30—23:30	0.031	0.017	0.031	0.025	0.136	443	396	443	479	157
22:00—23:30	0.029	0.020	0.029	0.027	0.159	417	365	417	473	109
22:30—23:30	0.029	0.023	0.029	0.033	0.144	391	338	391	467	110
23:00—23:30	0.029	0.023	0.029	0.037	0.157	363	317	363	452	111
23:30—23:30	0.030	0.059	0.030	0.044	0.120	340	300	340	469	152
Mean	0.023	0.023	**0.020**	0.039	0.066	600	566	678	664	**539**

Table 4: *Averaged empirical coverage probability deviance and width of the TS, BM and PLS prediction intervals constructed parametrically and nonparametrically for the 52 iterative one-step-ahead forecasts. The minimal mean coverage probability deviance and the minimal width are marked in bold.*

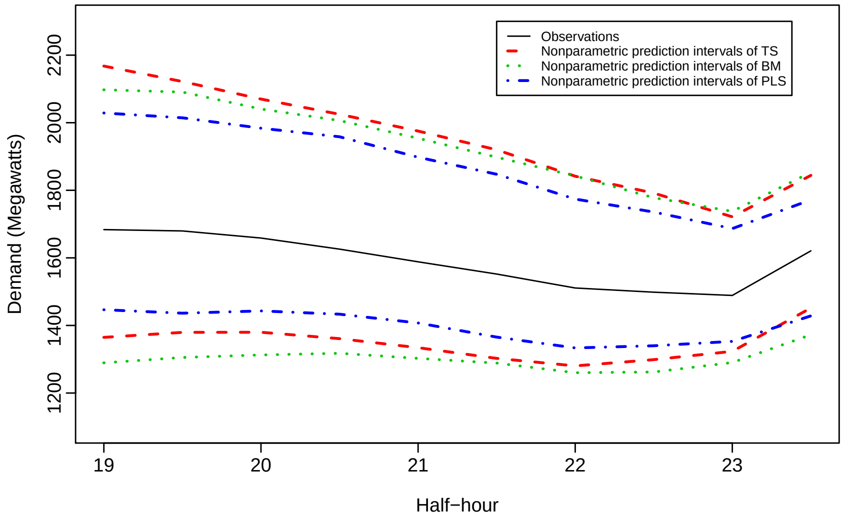


Figure 5: *The 95% prediction intervals of the Monday electricity demand from 19:00 to 23:30 in 26/3/2007. By incorporating the electricity demand from midnight to 18:30, the prediction intervals can be updated using the nonparametric bootstrap method.*

The narrowest width of prediction intervals obtained by the PLS method comes at a cost of the worst coverage probability deviance. The coverage probabilities of the TS and BM methods using the parametric approach are similar; but the BM method produces narrower prediction intervals than the TS method, thus it provides more informative evaluation of uncertainty.

An advantage of generating bootstrap samples is to provide density forecasts obtained using kernel density estimation. For example, Figure 6 displays the kernel density plots of Monday electricity demand in 26/3/2007 at different time periods, based on $B = 500$ replications.

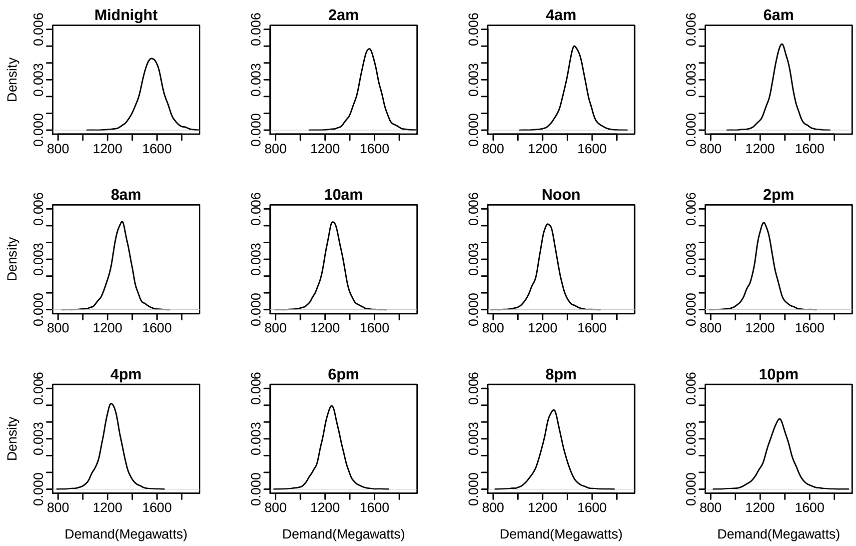


Figure 6: *Kernel density plots of the half-hourly Monday electricity demand in 26/3/2007. The bandwidth of kernel density plots is selected using a pilot estimation of derivatives (Sheather & Jones 1991).*

7 Conclusions

This paper uses half-hourly electricity demand data in South Australia as an empirical study of nonparametric modeling and forecasting methods for prediction from half-hour ahead to one year ahead. The nonparametric forecasting and updating approaches treat the historical data as a time series of curves. 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 all seven electricity demand data sets. The problem of forecasting future electricity demand has been overcame by forecasting $K = 3$ one-dimensional principal component scores. Conditioning on the historical data and fixed functional principal components, the 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, the four updating methods can not only improve point forecast accuracy, but they also eliminate the assumption, $N = np$, made in Aneiros-Pérez & Vieu (2008), Antoch et al. (2008), Antoniadis & Sapatinas (2003), Besse et al. (2000), and Ferraty & Vieu (2006, 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 functional principal components. It however may suffer from the singularity problem when the number of partially observed data points is 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 $\hat{\beta}^{\text{TS}}_{n+1|n}$. Based on the averaged MAPE over the 52 iterative one-step-ahead point forecasts in the testing sample, the OLS, RR and PLS methods perform equally the best among all of the methods investigated.

Furthermore, we used a nonparametric method to construct prediction intervals, and compared the empirical coverage probability to the parametric method. Although the coverage probability of the parametric and nonparametric methods for the TS and BM methods do not differ much, 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 between the BM and TS methods, the prediction interval width obtained by the BM method is narrower, thus the BM method provides more informative evaluation of forecast uncertainty than the TS method without updating. The narrowest width of prediction intervals obtained by the PLS method comes at a cost of the worst coverage probability.

The aforementioned approaches may seem complicated for calculating point forecasts, updating point forecasts, and constructing parametric and nonparametric prediction intervals, but their implementation is straightforward using the *ftsa* package of Hyndman & Shang (2010*a*).

Acknowledgement

The author would like to acknowledge the financial support of postgraduate publication award of Monash University. The manuscript is greatly benefitted from the comments by Professors Rob Hyndman and Vijay Nair. Thanks are also due to the High Performance Computing Center for access to the Monash Sun Grid Cluster for the data analyses.

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.

Baillie, R. T. & Bollerslev, T. (1992), 'Prediction in dynamic models with time-dependent conditional variances', *Journal of Econometrics* **51**(1-2), 91–113.

Bajay, S. V. (1983), 'Long-term electricity demand forecasting models: a review of methodologies', *Electric Power Systems Research* **6**(4), 243–257.

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*, 4th 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/CRC, Boca Raton, Florida.

Christoffersen, P. F. (1998), 'Evaluating interval forecasts', *International Economic Review* **39**(4), 841–862.

Cottet, R. & Smith, M. (2003), 'Bayesian modeling and forecasting of intraday electricity load', *Journal of the American Statistical Association* **98**(464), 839–849.

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', *Electronic Journal of Statistics* **3**, 865–885.

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.

Goia, A., May, C. & Fused, G. (2010), 'Functional clustering and linear regression for peak load forecasting', *International Journal of Forecasting* **26**(4), 700–711.

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. C. & Koopman, S. J. (1993), 'Forecasting hourly electricity demand using time-varying splines', *Journal of the American Statistical Association* **88**(424), 1228–1242.

Hippert, H. S., Pedreira, C. E. & Souza, R. C. (2001), 'Neural networks for short-term load forecasting: a review and evaluation', *IEEE Transactions on Power Systems* **16**(1), 44–55.

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. & Fan, S. (2010), 'Density forecasting for long-term peak electricity demand', *IEEE Transactions on Power Systems* **25**(2), 1142–1153.

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

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. & Shang, H. L. (2010*a*), *ftsa: Functional time series analysis*. R package version 1.6.
URL: *http://CRAN.R-project.org/package=ftsa*

Hyndman, R. J. & Shang, H. L. (2010*b*), 'Rainbow plots, bagplots, and boxplots for functional data', *Journal of Computational and Graphical Statistics* **19**(1), 29–45.

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.

McNees, S. K. & Fine, L. K. (1996), Forecast uncertainty: can it be measured?, *in* 'Paper presented at the Conference on Expectations in Economics,', Federal Reserve Bank of Philadelphia.

Pedregal, D. J. & Young, P. C. (2008), 'Development of improved adaptive approaches to electricity demand forecasting', *Journal of the Operational Research Society* **59**(8), 1066–1076.

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*, 2nd 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.

Rousseeuw, P., Ruts, I. & Tukey, J. (1999), 'The bagplot: A bivariate boxplot', *The American Statistician* **53**(4), 382–387.

Shang, H. L. & Hyndman, R. J. (2010), 'Nonparametric time series forecasting with dynamic updating', *Mathematics and Computers in Simulation* **in press**.

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 & 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.

Taylor, J. W. (2003), 'Short-term electricity demand forecasting using double seasonal exponential smoothing', *Journal of the Operational Research Society* **54**(8), 799–805.

Taylor, J. W. & McSharry, P. E. (2007), 'Short-term load forecasting methods: an evaluation based on European data', *IEEE Transactions on Power Systems* **22**(4), 2213–2219.

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

Weron, R. (2006), *Modeling and forecasting electricity loads and prices: A statistical approach*, Wiley, Chichester.