



Modeling of electricity demand forecast for power system

Ping Jiang¹ · Ranran Li¹ · Haiyan Lu² · Xiaobo Zhang¹

Received: 30 October 2018 / Accepted: 15 March 2019 / Published online: 23 March 2019

© Springer-Verlag London Ltd., part of Springer Nature 2019

Abstract

The emerging complex circumstances caused by economy, technology, and government policy and the requirement of low-carbon development of power grid lead to many challenges in the power system coordination and operation. However, the real-time scheduling of electricity generation needs accurate modeling of electricity demand forecasting for a range of lead times. In order to better capture the nonlinear and non-stationary characteristics and the seasonal cycles of future electricity demand data, a new concept of the integrated model is developed and successfully applied to research the forecast of electricity demand in this paper. The proposed model combines adaptive Fourier decomposition method, a new signal preprocessing technology, for extracting useful element from the original electricity demand series through filtering the noise factors. Considering the seasonal term existing in the decomposed series, it should be eliminated through the seasonal adjustment method, in which the seasonal indexes are calculated and should multiply the forecasts back to restore the final forecast. Besides, a newly proposed moth-flame optimization algorithm is used to ensure the suitable parameters of the least square support vector machine which can generate the forecasts. Finally, the case studies of Australia demonstrated the efficacy and feasibility of the proposed integrated model. Simultaneously, it can provide a better concept of modeling for electricity demand prediction over different forecasting horizons.

Keywords Electricity demand forecasting · Adaptive Fourier decomposition · Seasonal adjustment · Moth-flame optimization algorithm · Machine learning method

1 Introduction

Electricity demand forecasting is a strategic and momentous technology, and it can provide many operating decisions, such as energy transactions in competitive electric power markets [1, 2]. Alternatively, producers and consumers can adjust their production schedule and select the best bidding strategy based on the forecasting information. Inaccurate demand forecasting may raise the operating cost of a utility company, which means considerable money cannot be saved. For example, the operating cost will increase by 10 million pounds every year for every 1%

forecasting error increases [3]. It is effective early warning information of electricity demand that ensures the balance between supply and demand. Therefore, precise electricity demand forecasting is of great importance for society and economic development.

Electricity demand series usually display several significant characteristics such as highly volatility and seasonality. Modeling electricity demand is complicated by the nonlinear and seasonal link to various factors. It is of importance to obtain and analyze the inner patterns and variation for the electricity demand forecasting. Furthermore, reliability and quality of the modeling should be enhanced significantly with the credibility of forecast. Based on the accurate information of demand forecasting, the power generation can be scheduled timely and thus unexpected and precarious load shedding events would be avoided.

The objectives of this work are to model the electricity demand and then use the proposed model to address the limitation of the baseline models, thus improving the

✉ Ranran Li
lirandufe@163.com

¹ School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China

² School of Software, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia

forecast performance. More specifically, this paper makes a fourfold contribution in the literature as follows.

1. A new concept of the integrated model is introduced for electricity demand prediction over different forecasting horizons. It can provide insight for forecasting electricity demand with noise information and seasonal cycles to include in a method.
2. Considering removing the noise information in electricity demand data, an efficient method-based adaptive Fourier decomposition technology is first conducted to extract the useful element through cutting the noise part of time series.
3. There is no guarantee that the seasonal cycles in a electricity demand data would improve the forecasting accuracy. So an important method in preprocessing input variable is seasonal adjustment, which can eliminate the seasonal factor in the decomposed series.
4. In order to ensure the accuracy and robustness, the moth-flame optimization algorithm is used to tune the parameters of forecast engine, i.e., regularization parameter γ and kernel parameter σ^2 . Through the experimental design and simulation, it proves that the proposed integrated model can guarantee the high accuracy of electricity demand forecasting.

After first reviewing a comprehensive literature (Sect. 2), the methodology supported in this paper is presented in Sect. 3. Then, the proposed integrated model is introduced in Sect. 4. Section 5 illustrates the data preparation and analysis. The results of the proposed model and baseline models are provided in Sect. 6. And Sect. 7 gives conclusions and future advances.

2 Literature review

In the literature, several researchers have studies around the theme of forecasting of electricity demand/load, and numerous methods have been used for short timeframes. These methods can be categorized as traditional models and artificial intelligence models.

Traditionally, these methods are based on statistical theory, including Box-Jenkins method, regression analysis, exponential smoothing, grey forecast model, etc. Box-Jenkins method has been widely used in the short-term electricity demand forecasting [4, 5]. But this model requests that the time series satisfy stationary condition; otherwise, it may generate poor electric power forecasting accuracy in variation [6]. Although the grey forecast model does not need statistical assumption, its forecasting accuracy is proportional to the degree of dispersion of the raw time series [7]. In the past few years, some methods based

on artificial intelligence techniques, such as support vector regression (SVR), artificial neural network (ANN), fuzzy system, knowledge-based expert systems and other methods, have been implemented for electricity demand forecasting. And there are many studies mentioned about next-hour load forecasting. With the deepening of research and complexity, the accuracy of the forecasting methods based on artificial intelligence model has markedly improved. For example, Wang et al. [8] applied SARIMA, seasonal exponential smoothing model and SVM to combine for electricity demand forecasting, and its results are promising. Xia et al. [9] used radial basis function neural networks for short-term load forecasting, and the accuracy and stability of the developed are verified by experimental results. Bercu et al. [10] proposed the properties of a dynamic coupled modeling to the intraday prediction of energy consumption and illustrated on a real individual load curve. Hsu and Chen [11] employed backpropagation neural network (BPNN) to forecast Taiwan's annual regional peak load, and the forecast results of the proposed model are promising. However, the BPNN method has several drawbacks, such as slow convergence and often being trapped in a local minimum [12]. Zhang et al. [13] embed a recurrent neural network into an encoder-decoder framework, which can provide some ideas for sequence-to-sequence learning in electricity forecasting. Besides, the forecast performance is mainly influenced by the approximation degree of it parameter values for the forecast engines. Inappropriate parameter values of forecast engines may lead to over-fitting or under-fitting, and thus a main process to improve the forecasting accuracy is how to choose the proper parameter values of forecast engines. However, there is still not a general principle for choosing the best parameter values of forecast engine models [14].

Over the past few years, all kinds of intelligent optimization algorithms have already been developed rapidly. Meta-heuristics algorithms can be classified into individual-based and population-based algorithms, which have achieved competitive contributions when solving optimization problems including parameter optimization problem. Some intelligent algorithms are based on the animals foraging behavior, such as particle swarm optimization algorithm (PSO) [15, 16], genetic algorithm (GA) [17, 18] and ant colony optimization algorithm [19, 20], which are not only applied to solve the optimization problems, but also used to tune the parameter values of forecasting engine models. Recently, many human-based intelligence algorithms have been developed to efficiently resolve the complex optimization problems [21–23]. The popular approaches-human-based include teaching learning-based approach (TLBA), tabu search algorithm (TSA) [24], imperialist competitive algorithm (ICA) [25], simulated annealing (SA) [26], etc. These intelligent

optimization algorithms can perform the parameter optimization problems through maintaining proper balance between exploitation and exploration processes. For instance, Piltan et al. [27] applied PSO and GA to attain parameters of the models in order to forecast and analyze energy in the Iranian metal industry, and the results indicate significant improvement. Barman et al. [28] introduced a new technique, called grasshopper optimization algorithm (GOA) to tune suitable parameters of support vector machine (SVM), and the results can indicate the superiority of the proposed model. Zhang et al. [29] designed a deep convolutional neural network (CNN) for object-level video advertising, which demonstrates the effectiveness of the proposed model.

A key point in electricity demand/load forecasting is the appropriate treatment of seasonal behavior of the time series [30–35]. Since seasonality is a major feature of electricity demand, it should not be ignored in the process of modeling. Therefore, the methods for processing seasonal elements have been a non-negligible issue in electricity demand analysis. According to Zhang et al. [36], deseasonalization can dramatically reduce forecasting errors in many seasonal time series forecasting. It suggests that more attention should be focused on the estimation of current and current trend levels. The method used for removing the seasonal elements is generally the simple moving average. However, these methods have often been criticized when they are utilized alone due to lacking an explicit model concerning the decomposition of the original series [37]. Besides, their estimates for observations in the most recent seasonal patterns do not always keep the same degree of reliability as those of central observations. In reality, the seasonal component adjustment method can be used to eliminate the seasonal factor through separating the seasonal variable and trend variable from the time series, which can make short-term forecasting more efficient. Therefore, the seasonal component adjustment method is considered to preprocess the seasonal electricity demand data and then the data-eliminated seasonal component is used for forecasting in this paper.

The seasonal component prediction can be carried out by the seasonal patterns adjustment method. In addition, the electricity demand is always influenced by various factors, such as social and economic environments, electrical power unit, dynamic electricity prices, etc. While above-mentioned approaches improve the forecasting accuracy to some extent, they use the original electricity demand series directly to model without considering the inherent characteristics of the data. It can increase the convergence difficulty due to noise component caused by unstable factors, which lead to bad forecasting accuracy. The recent tools in signal analysis seem promising, such as ensemble mode decomposition (EMD) family. They can be

applied to non-stationary time series of electricity demand, but the disadvantage of EMD is that the number of intrinsic mode functions can be changed according to harmonic content of signals [38]. And the tool based on wavelet transform suffers from the problems of different mother wavelets and the required number of decomposition levels. In order to overcome the drawbacks of EMD family, this study applied adaptive Fourier decomposition (AFD) technique for electricity demand data preprocessing.

3 Methodology

In this section, we introduce the detailed methodologies applied for electricity demand forecasting.

3.1 Adaptive Fourier decomposition

As a generalization of the conventional Fourier decomposition method combined with the greedy algorithm, adaptive Fourier decomposition (AFD) is suitable for separating a signal with noise by overlapping frequency ranges [39]. The AFD method contains the adaptive decomposition of a given analytic signal $\mathbf{G}(t)$ that is H^2 space into a series of mono-components, $s_n(t)$ and a standard remainder $R_N(t)$, which can be shown as follows.

$$\mathbf{G}(t) = \sum_{k=0}^{\infty} c_k e^{ikt} = \sum_{n=1}^N s_n(t) + R_N(t), \sum_{k=0}^{\infty} |c_k|^2 < \infty \quad (1)$$

The basic functions of AFD are the rational orthogonal system $\{\mathbf{B}_n\}_{n=1}^{\infty}$ which can be shown as follows.

$$\mathbf{B}_n(e^{it}) = \frac{\sqrt{1 - |a_n|^2}}{1 - \bar{a}_n e^{it}} \prod_{k=1}^{n-1} \frac{e^{it} - a_k}{1 - \bar{a}_k e^{it}} \quad (2)$$

The characteristics of $\mathbf{B}_n(e^{it})$ only depend on the array of a_n . The main procedure of the AFD is to find suitable a_n array which can get high energy convergence rate.

The AFD extracts sequentially mono-components from high-energy components to low-energy components. In order to find the energy relationship easily, the reduced remainders \mathbf{G}_n can be defined by:

$$\mathbf{G}_n(e^{it}) = R_{n-1}(e^{it}) \prod_{l=1}^{n-1} \frac{1 - \bar{a}_l e^{it}}{e^{it} - a_l} \quad (3)$$

where R_{n-1} are the standard remainders of \mathbf{G}_n .

Using the reduced remainders \mathbf{G}_n , Eq. (1) can be calculated as follows.

$$\mathbf{G}(t) = \sum_{n=1}^N \langle \mathbf{G}_n, e_{\{a_n\}} \rangle B_n(e^{jt}) + G_{N+1}(e^{jt}) \prod_{n=1}^N \frac{e^{jt} - a_n}{1 - \bar{a}_n e^{jt}} \quad (4)$$

where $e_{\{a_n\}}(e^{jt})$ is the evaluator at a_n which can be considered as a dictionary constituting the elementary function:

$$e_{\{a_n\}} e^{jt} = \frac{\sqrt{1 - |a_n|^2}}{1 - \bar{a}_n e^{jt}} \quad (5)$$

Based on Eq. (4), the energy of $\mathbf{G}(t)$ is expressed by:

$$\|\mathbf{G}(t)\|^2 = \sum_{n=1}^N |\langle \mathbf{G}_n, e_{\{a_n\}} \rangle|^2 + \|G_{N+1}(e^{jt})\|^2 \quad (6)$$

To realize a high energy convergence rate, the energy of the standard remainder $\|G_{N+1}(e^{jt})\|^2$ should be kept minimal. Thus, the maximal projection principle is used to find a_n as follows.

$$a_n = \arg \max \left\{ |\langle \mathbf{G}_n, e_{\{a_n\}} \rangle|^2 \right\} \quad (7)$$

AFD decomposes signals based on their energy distribution instead of frequency analysis.

Before decomposing, the mean value of the measured noisy signal $s(t)$ is removed to prevent the DC offset effect. The noisy signal can be projected to H^2 space through the Hilbert transform [40].

$$H\{s(t)\} = \frac{1}{\pi} \int_{-\infty}^{\infty} s(t) \frac{1}{t - \tau} dt \quad (8)$$

Considering the analytic signal, the analytic representation of the noisy signal can be shown in Eq. (9), which is regarded as the input of the AFD.

$$G(t) = s(t) + jH\{s(t)\} \quad (9)$$

In the AFD technique, a stop rule for the iterative decomposition process is based on the estimated signal-to-noise ratio (SNR) of the noisy signal. The aim of denoising is to make the mean square error (MSE) as small as possible.

$$MSE = L^{-1} E \|\hat{h} - h\|^2 \quad (10)$$

where L is the number of total data. The reconstructed signal should be minimized as follows.

$$\frac{\|s(t)\|^2}{\|\hat{h}(t)\|^2} - \left(1 + \frac{1}{10^{\text{SNR}_e/10}} \right) \quad (11)$$

where SNR_e is the estimated signal-to-noise ratio of the noisy signal, which can be defined by

$$\text{SNR}_e = 10 \log \left(\frac{\|h(t)\|^2}{\|w(t)\|^2} \right) \quad (12)$$

3.2 Seasonal component adjustment

Generally, electricity demand datasets are more influenced by the seasonal factors. For seasonal electricity demand, some methods should be applied to identify the window of seasonal components and eliminate the seasonal components. Based on the window of seasonal components, season adjustment method is used to address the seasonality that existed in the original data.

A set of data y_1, y_2, \dots, y_T ($T = ms$) is expressed as $y_{11}, y_{12}, \dots, y_{1s}; y_{21}, y_{22}, \dots, y_{2s}; \dots; y_{m1}, y_{m2}, \dots, y_{ms}$, where m and s are the number of cycles and data items in each cycle, respectively. Then, the average value can be calculated as:

$$\bar{y}_k = (y_{k1} + y_{k2} + \dots + y_{ks})/s \quad k = 1, 2, \dots, m \quad (13)$$

Normalize the data items y_{ms} , and the specific form is defined as:

$$I_{ks} = \frac{y_{kl}}{\bar{y}_k} \quad (k = 1, 2, \dots, m; l = 1, 2, \dots, s) \quad (14)$$

Then, the i -th seasonal index values can be calculated as follows.

$$I_i = \frac{I_{1i} + I_{2i} + \dots + I_{mi}}{m} \quad (i = 1, 2, \dots, s) \quad (15)$$

Using the seasonal indexes, the series without the seasonal effect can be calculated based on Eq. (15) as follows.

$$y'_{ki} = \frac{y_{ki}}{I_i} \quad k = 1, 2, \dots, m; i = 1, 2, \dots, s \quad (16)$$

Finally, new data without the seasonal effect are realized through rearranging the data items $y'_{11}, y'_{12}, \dots, y'_{1s}, y'_{21}, y'_{22}, \dots, y'_{2s}, \dots, y'_{m1}, y'_{m2}, \dots, y'_{ms}$ to y'_1, y'_2, \dots, y'_T .

3.3 Moth-flame optimization technique

Inspired by the transverse orientation navigation behaviors of moths, MFO could be utilized to solve optimization problems with constrained and unknown search spaces [41]. Similar to other optimization algorithms, MFO possesses local and global search capabilities. In MFO algorithm, the key components are moths and flames where both of them are carried out to represent solutions. But they differ in the way of their treatment and their updating each iteration. The moths are actual search agents that move around the search space. And flames are the best position of moths which have been obtained so far [42].

Since the MFO algorithm is a population-based algorithm, the set of moths could be represented in a matrix as follows:

$$\mathbf{M} = \begin{bmatrix} m_{1,1} & m_{1,2} & \cdots & m_{1,d} \\ m_{2,1} & m_{2,2} & \cdots & m_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ m_{n,1} & m_{n,2} & \cdots & m_{n,d} \end{bmatrix} \quad (17)$$

where n is the number of moths and d is the number of variables or dimension.

For all the moths, the corresponding fitness function values can be stored in an array as follows:

$$\text{OM} = (\text{OM}_1 \quad \text{OM}_2 \quad \cdots \quad \text{OM}_n)^T \quad (18)$$

where n is the number of moths. Equation (18) is the fitness function value for each moth. And the position vector (e.g., the first row in Eq. 17) of each moth is passed to the fitness function, and the output of the objective function is assigned to the corresponding moth as its objective value (e.g., OM_1 in Eq. 18).

Hence, the flames could be represented a matrix similar to Eq. (17) as follows:

$$\mathbf{F} = \begin{bmatrix} F_{1,1} & F_{1,2} & \cdots & F_{1,d} \\ F_{2,1} & F_{2,2} & \cdots & F_{2,d} \\ \vdots & \vdots & \vdots & \vdots \\ F_{n,1} & F_{n,2} & \cdots & F_{n,d} \end{bmatrix} \quad (19)$$

So there is also an array for storing the corresponding fitness function values as follows:

$$\text{OF} = (\text{OF}_1 \quad \text{OF}_2 \quad \cdots \quad \text{OF}_n)^T \quad (20)$$

where n is the number of moths.

The MFO generally contains three-tuple approximation functions that can be defined as follows:

$$\text{MFO} = (I, P, T) \quad (21)$$

where I is an initialization function that creates a random population of moths and their correspond fitness values are defined as follows:

$$M(i, j) = (ub(i) - lb(i)) * rand + lb(i) \quad (22)$$

$$\text{OM} = \text{FitnessFunction}(M) \quad (23)$$

where ub and lb represent the upper and lower bounds of the variables, respectively.

Then, a logarithmic spiral function is employed to update the position of a moth, which is defined as follows.

$$S(M_i, F_j) = D_i \cdot e^{bt} \cdot \cos(2\pi t) + F_j \quad (24)$$

where S is the spiral function, M_i represents the i -th moth, and F_j indicates the j -th flame. $D_i = |F_j - M_i|$ is the distance between the i -th moth and the j -th flame, while b is a

constant and t is a random number in $[r, 1]$. The spiral equation allows a moth to explore around a flame and not necessarily in the space between them. In addition, each moth updates its position with respect to a specific flame to avoid local optimum and increase global exploration. So, the positions of the moths are updated by the updated flames, thus exploring the search space more effectively.

To balance between the global exploration and local exploitation of the moth and flame, the number of flames is adaptively decreased as follows.

$$\text{flame NO} = \text{round}\left(N - l \times \frac{N - 1}{T}\right) \quad (25)$$

where N and T indicate the maximum number of flames and iterations, respectively, while l is the current number of iteration.

Overall, the logarithmic spiral search is employed in MFO. In Eq. (25), the flame decrement strategy is carried

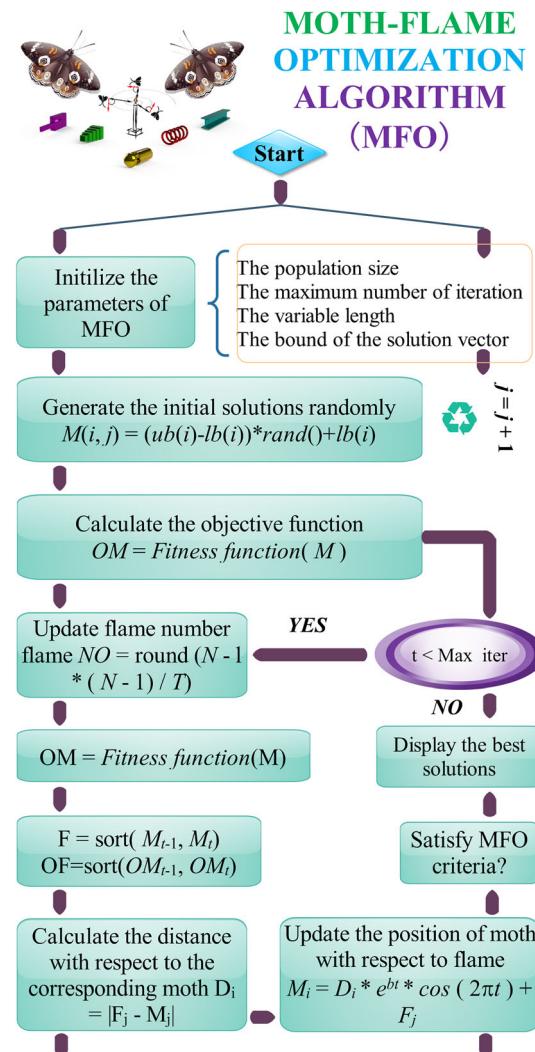


Fig. 1 The flowchart figure of the moth-flame optimization technique

out to balance between exploration and exploitation to achieve global optimality. Figure 1 shows the flowchart of the moth-flame optimization technique. The MFO shows superior capabilities of solving the optimization problems like multimodal and unimodal optimization functions [43]. And the pseudo-code of the MFO algorithm is depicted in “Appendix A.”

3.4 Least squares support vector machine

LSSVM, a modification of standard SVM, overcomes the shortcomings of time consuming when analyzing huge data, which uses the least square linear system at the loss function and the equality constraints [44].

Define that there is a given training set $\{x_i, y_i\}_{i=1}^n$, where x_i is the input data and y_i is the output data. In the modeling of LSSVM, the approximation error ξ_i is taken as the loss function. The optimization problem of LSSVM can be defined as follows.

$$\min_{\omega, b, \xi} J(\omega, \xi) = \frac{1}{2} \omega^T \omega + \gamma \frac{1}{2} \sum_{i=1}^n \xi_i^2, \quad \gamma > 0 \quad (26)$$

subject to

$$y_i = \omega^T \varphi(x_i) + b + \xi_i, \quad i = 1, 2, \dots, n \quad (27)$$

LSSVM model can be denoted as follows:

$$f(x) = \omega^T \varphi(x) + b \quad (28)$$

Different from the standard SVM, the optimization problem in the LSSVM is introduced as follows:

$$L(\omega, b, \xi, \alpha) = J(\omega, \xi) - \sum \alpha_i [\omega^T \varphi(x_i) + b + \xi_i - y_i] \quad (29)$$

where α_i is the Lagrange multiplier.

Through the Karush–Kuhn–Tucker (KKT) conditions, the solutions could be acquired by partially differentiating with respect to w , b , e_i and α_i :

$$\left\{ \begin{array}{l} \frac{\partial L}{\partial w} = 0 \rightarrow w = \sum_{i=1}^l \alpha_i \varphi(x_i) \\ \frac{\partial L}{\partial b} = 0 \rightarrow \sum_{i=1}^l \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \rightarrow \alpha_i = \gamma e_i \\ \frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T \varphi(x_i) + b + e_i - y_i = 0 \end{array} \right. \quad (30)$$

When the variables w and e are removed, it can be described as a linear system [45, 46].

$$\begin{bmatrix} 0 & 1 & & \cdots & 1 \\ 1 & K(x_1, x_1) + \frac{1}{\gamma} & \cdots & K(x_1, x_n) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & K(x_n, x_n) & \cdots & K(x_n, x_n) + \frac{1}{\gamma} \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \quad (31)$$

where $\alpha = [\alpha_1, \dots, \alpha_n]^T$ and the Mercer's condition $K(x_i, x_j) = \varphi(x_i)^T \varphi(x_j)$ is called kernel function. Therefore, the function estimating the LSSVM model becomes

$$f(x) = \sum_{i=1}^n \alpha_i K(x, x_i) + b \quad (32)$$

where α_i, b are the solutions of Eq. (30), and $K(x, x_i)$ is the kernel function.

In general, kernel function of SVM could be divided into the whole kernel function and the local kernel function [47]. Besides above-mentioned two functions, there are some other ones such as

1. *The linear kernel $K(x, y) = x^T y$*
2. *The poly kernel $K(x, y) = (1 + x^T y / \sigma^2)^d$*
3. *The radial basis function (RBF) $K(x, y) = \exp(-\|x - y\|^2 / 2\sigma^2)$*

The RBF kernel used in this paper is frequently used for various problems for its high resolution power [48]. A major advantage of LSSVM approach lies in the capacity of capturing the nonlinear patterns hidden in the power load demand [49].

4 The new concept of the integrated model

In this paper, the proposed integrated model (AFD-S-OLSSVM) is developed to forecast the electricity demand. The flowchart of the proposed model is shown in Fig. 2, which is composed of data preprocessing module, seasonality adjustment module, parameters optimization module, forecast module, and restore and evaluation module. The detailed physical interpretations of each module can be presented as follows:

Step 1 Data preprocessing module. Due to the high noise and fluctuation, it is not easy to achieve the forecast of electric power curve accurately. In this study, the electricity demand time series is decomposed through the adaptive Fourier decomposition technology. The objective of this step is to filter the inherent noise factor and extract the useful component from the original electricity demand series in order to improve the effectiveness and accuracy.

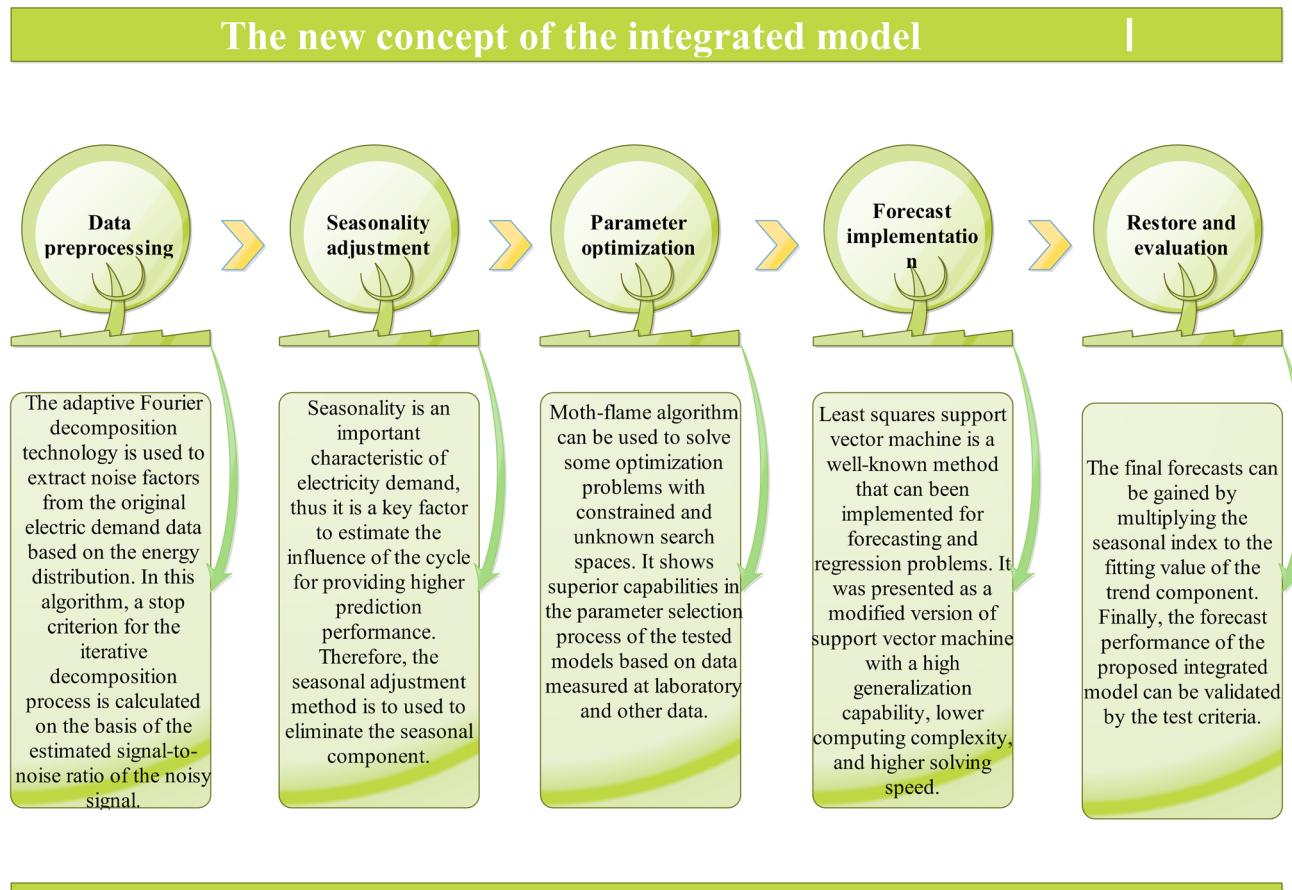


Fig. 2 The flowchart of the proposed integrated model

Step 2 Seasonality adjustment module. Electricity demand data are usually influenced by various factors and thus often present periodicity and seasonal characteristic. The reason why data preprocessing module is the first step is that forecasting the electricity demand directly may lead to poor forecasting accuracy. After eliminating the noise signal, the seasonal aspects of the electricity demand data should be considered. Seasonal adjustment method is used to eliminate the seasonal components, after which the seasonal indexes can be calculated.

Step 3 Parameters optimization module. In an effort to achieve a high level of the accuracy and stability of prediction, the MFO algorithm is utilized to determine the optimal parameters of forecast engine in this section because the initial values are random. Through the optimization, the values of the regularization parameter γ and kernel parameter σ^2 can be confirmed as 160.3516 and 1.9101, respectively.

Step 4 Forecast module. The forecast values of the electricity demand can be obtained through the least squares

support vector machine method optimized by the optimization algorithm. It should be noted that the forecast performance depends on the parameters of the forecast engine, so it is key to carry out the parameters optimization module.

Step 5 Restore and evaluation module. In order to obtain the ultimate forecasts, it is essential to multiply the seasonal index to the forecasting value of the trend component. And the performance of the proposed integrated model can be evaluated quantitatively by the test criteria.

5 Data illustration and analysis

This section presents the data sources in this research and the preprocessing of variables.

5.1 Data preparation

Three historic electricity demand time series are collected from New South Wales (NSW) and Queensland (QLD)

Table 1 Statistical indicators of the electricity demand samples in NSW, QLD and SA markets

Calendar mode of markets	Data set	Numbers	Statistical indicator (*units: MW)						
			Mean*	Max*	Min*	Med*	SD	Skew	Kurt
NSW	All samples	1344	8121.61	11,095.83	5730.32	8276.46	1182.87	-0.17	2.07
	Training	1104	8093.51	11,095.83	5730.32	8233.21	1187.90	-0.11	2.10
	Testing	240	8250.88	10,113.55	5954.51	8567.08	1153.09	-0.48	2.05
QLD	All samples	1344	6012.26	8077.46	4520.73	6118.91	855.55	0.13	2.24
	Training	1104	6013.79	8077.46	4520.73	6077.94	875.52	0.23	2.25
	Testing	240	6005.23	7197.43	4627.38	6308.41	758.55	-0.55	1.85
SA	All samples	1488	1416.15	2178.98	891.94	1415.78	257.25	0.20	2.46
	Training	1248	1434.05	2178.98	891.94	1435.75	261.65	0.18	2.37
	Testing	240	1323.08	1734.40	936.01	1367.93	210.18	-0.20	2.04

over February 1st–February 28th, 2014, and from South Australia (SA) over August 1st–August 31st, 2013, in Australia's market (www.aemo.com.au). Thus, three simulation experiments were used to evaluate the effectiveness and universality of the proposed model.

The time gap of observation series is half an hour which means 1 day has 48 observation values. To further confirm the forecasting performance of the proposed model under different data types, electricity time series is also divided into two sets: the training data sampling including 1104/1104/1248 data points and the testing data sampling including 240 data points. And statistical measures [i.e., mean, median value, maximum (Max), minimum (Min), median (Med), standard deviation (SD), skewness (Skew.) and kurtosis (Kurt.)] of the datasets are calculated and shown in Table 1.

5.2 Input variable preprocessing

In the proposed model, the adaptive Fourier decomposition technology is used to decompose the original electricity demand series. The detailed mechanism for the decomposition process in the adaptive Fourier decomposition method is presented in Sect. 3.1. The raw electricity demand signal is decomposed to series mono-components in view of the energy distribution. And the desired signal and the noise factors are separated by the adaptive Fourier decomposition technology. In order to illustrate the process of decomposition, the noise factors were extracted from the raw electricity demand. Taken NSW electricity market as an example, Fig. 3 shows the raw electricity demand and the noise factors produced by the data decomposition. From the curve of noise factors, it presents the noise information which is redundant for the time series. In order to avoid the poor forecasting accuracy, the noise information should be eliminated.

After the process of filtering noise, the seasonal components are significant. Thus, the seasonal tendency should be extracted from the filtering signal through the above-mentioned seasonal adjustment method. From the view of each half-hour period in load data, the seasonal coefficients should belong to half an hour period. In the theory of seasonal adjustment method, the seasonal index values can be obtained and the final forecast should be achieved by multiplying the seasonal indexes to restore because the forecasting results are carried out by the trend components. Through the seasonal adjustment method, their corresponding seasonal indexes in different markets can be calculated, as shown in Table 2. And the curve after eliminating the seasonal factor is presented also shown in Fig. 3 and it can be shown that there is no obvious seasonal tendency.

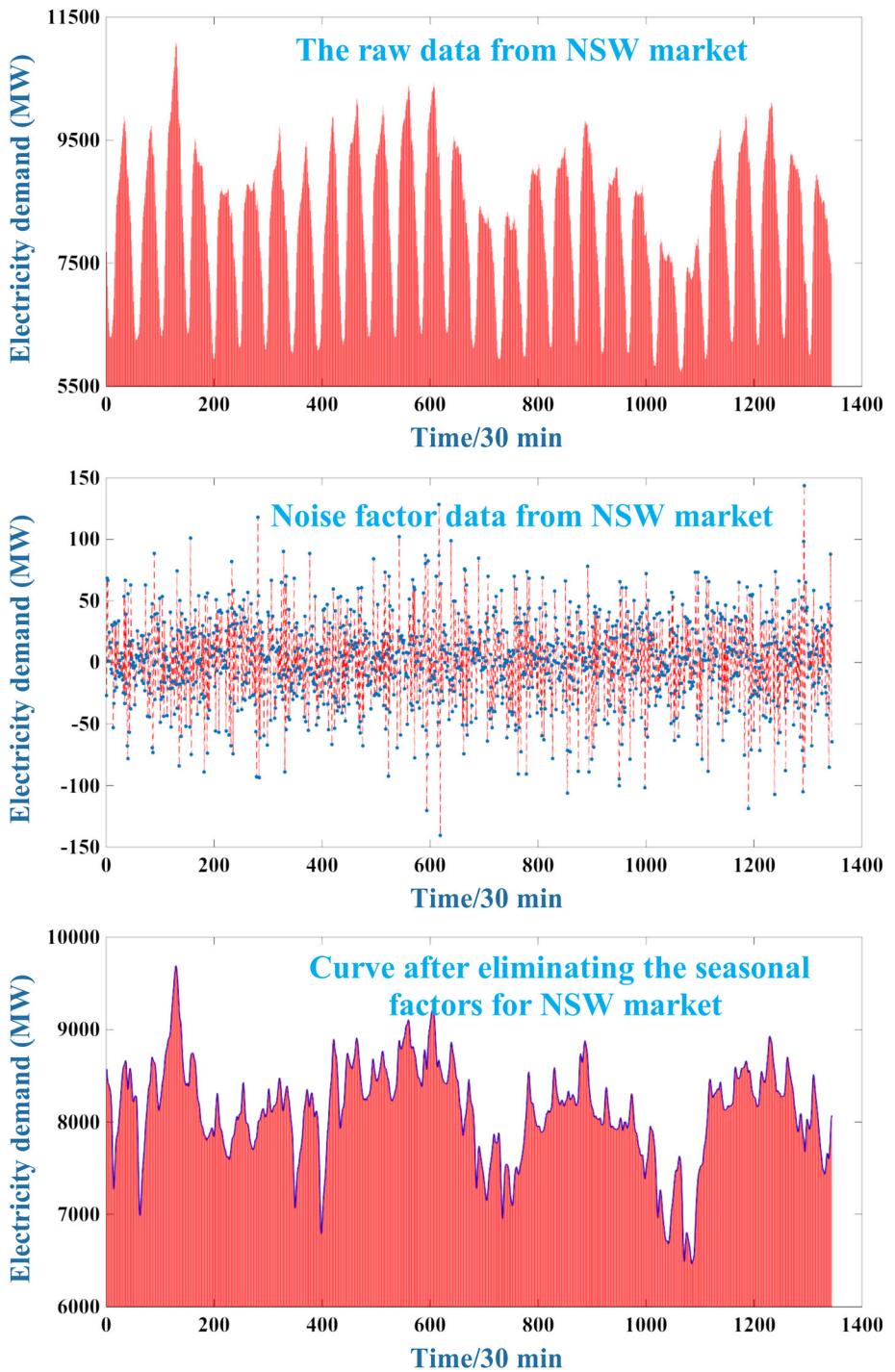
6 Case studies and discussion

This section presents the performance of the proposed integrated model for electricity demand forecasting. And the simulation results and analysis are implemented in MATLAB 2014a, LS-SVM lab v1.8 toolbox, Photoshop, self-written Matlab programs, and a computer with an Inter(R) Core(TM) i5-4590 CPU @3.30 GHz and windows 7 professional operating system.

6.1 Model performance evaluation

Several validation criteria are used to establish comprehensively the accuracy of the proposed integrated model and the benchmark models. These error measure indexes include the mean absolute error (MAE), the root mean square error (RMSE), the mean absolute percentage error (MAPE), the standard deviation error (SDE), the Theil

Fig. 3 The electricity demand curve of after the process of filtering noise and eliminating seasonality



inequality (THI) coefficient and the directional symmetry (DS). And the definition is described in Table 3.

6.2 Case 1: NSW market

In this case study, the electricity demand data collected from NSW market sampled per 30 min from February 1st to February 28th, 2014, were used to perform multi-step

forecasting (+ 1, + 2, + 3 point forecast). It should be noted that the + 2, + 3 point is forecasted independently using the values of previous + 1, + 2 point cumulative forecast. In order to evaluate the performance of the proposed AFD-S-OLSSVM model, seven benchmark models are used for validation purpose. The benchmark models include BPNN (backpropagation neural network) model, LSSVM model, OLSSVM (LSSVM model optimized by

Table 2 The seasonal factors for each NSW, QLD and SA market

Time	Seasonal index			Time	Seasonal index		
	NSW	QLD	SA		NSW	QLD	SA
0:00	0.895484	0.906666	1.114292	12:00	1.107111	1.066699	0.962568
0:30	0.869996	0.876789	1.10565	12:30	1.111022	1.072146	0.952094
1:00	0.839876	0.848804	1.04682	13:00	1.115473	1.076969	0.942523
1:30	0.808679	0.823991	0.96882	13:30	1.119231	1.082671	0.935992
2:00	0.782584	0.813215	0.896507	14:00	1.122058	1.089487	0.933938
2:30	0.765491	0.804425	0.840525	14:30	1.125567	1.099475	0.934728
3:00	0.757736	0.800652	0.80205	15:00	1.131484	1.103399	0.937315
3:30	0.758034	0.796042	0.777879	15:30	1.138838	1.109115	0.944052
4:00	0.766568	0.79697	0.763525	16:00	1.144118	1.117291	0.958841
4:30	0.785927	0.80473	0.755645	16:30	1.14369	1.129437	0.98538
5:00	0.818268	0.816076	0.754069	17:00	1.135592	1.141108	1.026871
5:30	0.86121	0.835355	0.761976	17:30	1.120086	1.139047	1.084919
6:00	0.908242	0.858764	0.784948	18:00	1.101156	1.137996	1.154785
6:30	0.953166	0.899075	0.828071	18:30	1.085323	1.134978	1.222282
7:00	0.991835	0.943639	0.890306	19:00	1.076455	1.144978	1.267858
7:30	1.022573	0.986127	0.959519	19:30	1.070476	1.144101	1.279339
8:00	1.046745	1.00998	1.019418	20:00	1.058438	1.126943	1.263221
8:30	1.066588	1.024524	1.058412	20:30	1.036908	1.104211	1.235331
9:00	1.082489	1.034609	1.073407	21:00	1.011488	1.080254	1.204007
9:30	1.093547	1.043593	1.068007	21:30	0.98882	1.06229	1.166652
10:00	1.099608	1.048157	1.04827	22:00	0.970127	1.026862	1.120193
10:30	1.102317	1.048215	1.021245	22:30	0.952732	0.990555	1.072474
11:00	1.103549	1.052244	0.99519	23:00	0.934272	0.95733	1.045296
11:30	1.10474	1.056942	0.975703	23:30	0.91428	0.933071	1.059089

the moth-flame optimization algorithm) model, AFD-OLSSVM (OLSSVM model preprocessed by the adaptive Fourier decomposition technology) model, S-OLSSVM (OLSSVM model preprocessed by seasonal adjustment method) model, EEMD-OLSSVM (OLSSVM preprocessed by ensemble empirical mode decomposition) model and EEMD-S-OLSSVM model. The multi-step forecasting results of the proposed integrated model and the above-mentioned benchmark models are summarized in Table 4. Figure 4 visualizes the absolute forecasted error comparison of different prediction methods.

It can be seen from Table 4 that the proposed AFD-S-OLSSVM model has the best forecasting performance with the minimum value of MAE as 26.998, RMSE as 35.28, MAPE as 0.333 and SDE as 35.28 over one-step forecast, the minimum value of MAE as 37.343, RMSE as 49.374, MAPE as 0.463 and SDE as 49.368 over two-step forecast, and the minimum value of MAE as 50.542, RMSE as 70.385, MAPE as 0.621 and SDE as 70.358 over three-step forecast. The best one of the compared benchmark models is AFD-OLSSVM model with the MAPE value of one-step forecast, two-step forecast and three-step forecast as 0.353, 0.48 and 0.811, respectively, while the worst one is BPNN model with the MAPE value of one-step forecast, two-step

forecast and three-step forecast as 0.788, 1.092 and 1.162, respectively. Through comparing the AFD-OLSSVM model with AFD-S-OLSSVM model, S-OLSSVM model with OLSSVM model, and EEMD-OLSSVM model with EEMD-S-OLSSVM model, it can be seen that the method of seasonal adjustment is necessary and significant. For the electricity demand data with clear seasonal trend, the seasonal adjustment is an essential method to improve the forecasting performance. As a result, the seasonal adjustment method makes the overestimated and underestimated forecast ones much closer to the actual.

As shown in Table 4, OLSSVM model performs better than LSSVM model, and it can be seen that the value of improvement in MAPE percentages from one-step, two-step and three-step of LSSVM model by OLSSVM model is up to 2.10%, 6.30% and 1.94%, respectively. Obviously, the MFO algorithm can make contribution to the forecasting performance. In order to show the forecasting effectiveness, the absolute errors of different models are vividly depicted in Fig. 4. From Fig. 4, we can conduct a conclusion that the proposed model has better accuracy than other models. So, it indicates that the proposed model can beat the benchmark models in terms of performance.

Table 3 Descriptions related to the validation criteria

Criterion	Interpretation	Equation
MAE	The mean absolute error	$MAE = \frac{1}{T} \sum_{t=1}^T \hat{y}_t - y_t $
RMSE	The root mean square error	$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2}$
MAPE	The mean absolute percentage error	$MAPE = \frac{1}{T} \sum_{t=1}^T \left \frac{\hat{y}_t - y_t}{y_t} \right \times 100\%$
SDE	The standard deviation error	$SDE = \sqrt{\frac{1}{T} \sum_{t=1}^T \left[y_t - \hat{y}_t - \sum_{t=1}^T ((y_t - \hat{y}_t)/T) \right]^2}$
THI	The Theil inequality coefficient	$THI = \frac{\sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2}}{\sqrt{\frac{1}{T} \sum_{t=1}^T y_t^2} + \sqrt{\frac{1}{T} \sum_{t=1}^T \hat{y}_t^2}}$
DS	The directional symmetry	$DS = \frac{1}{T} \sum_{t=1}^T d_t, \quad d_t = \begin{cases} 1 & (y_t - \hat{y}_{t-1})(\hat{y}_t - y_{t-1}) \geq 0 \\ 0 & (y_t - \hat{y}_{t-1})(\hat{y}_t - y_{t-1}) < 0 \end{cases}$

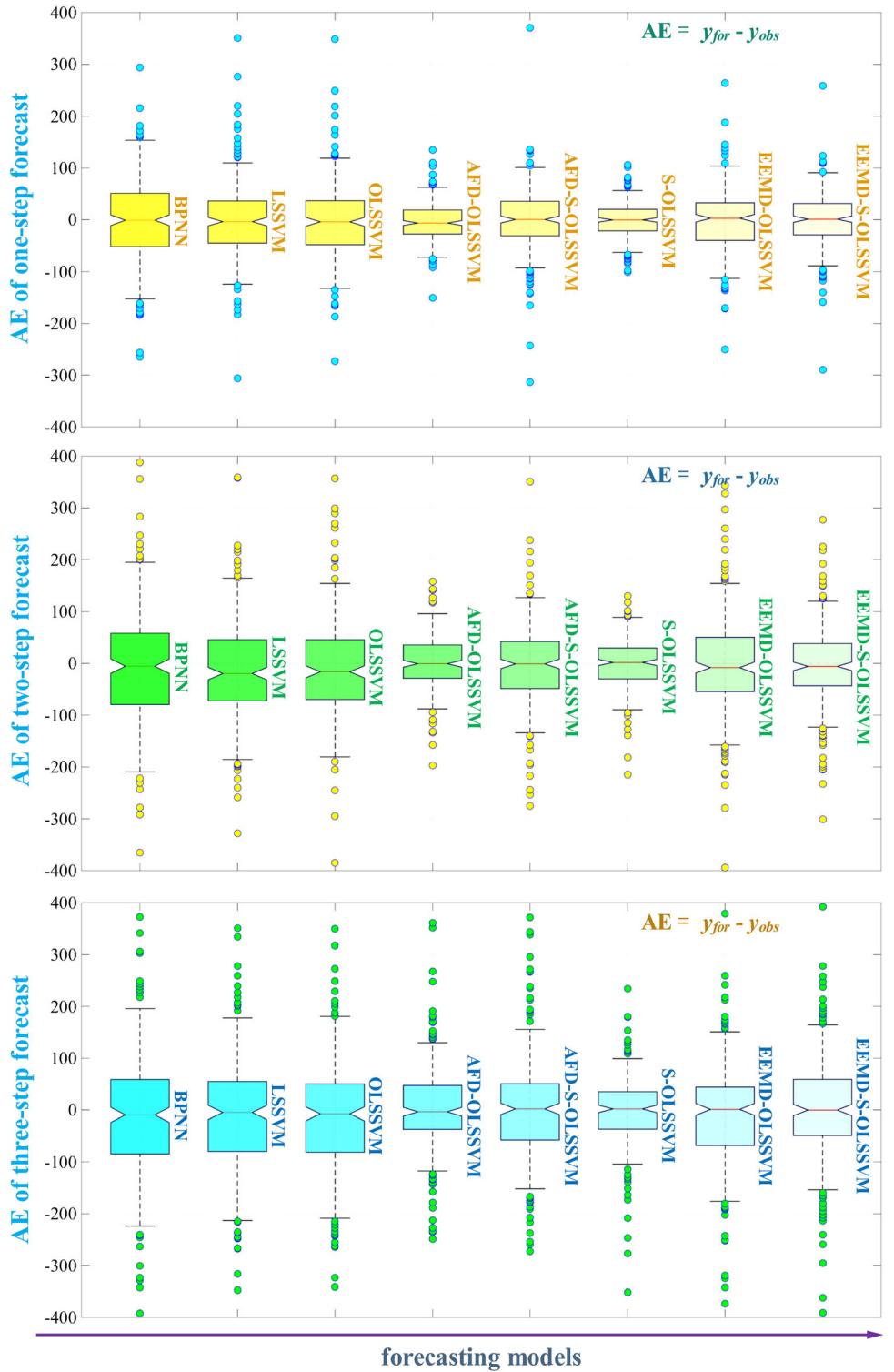
\hat{y}_t and y_t are the forecast and the true time series values of t times, respectively. T is the number of forecasting samples

Table 4 Performance evaluation of forecasting models using electricity demand data in NSW market

Forecasting models	Step	MAE	RMSE	MAPE	SDE	THI	DS
BPNN	1	63.801	82.148	0.788	82.070	1.18E–06	0.779
	2	88.686	117.320	1.092	116.768	1.69E–06	0.763
	3	93.986	123.886	1.162	123.234	1.78E–06	0.788
LSSVM	1	56.025	76.340	0.695	76.328	1.10E–06	0.783
	2	78.386	106.109	0.971	104.913	1.53E–06	0.742
	3	90.770	125.465	1.138	125.363	1.81E–06	0.788
OLSSVM	1	55.634	74.776	0.689	74.726	1.08E–06	0.775
	2	74.016	98.647	0.916	98.304	1.42E–06	0.771
	3	89.294	123.126	1.118	122.933	1.77E–06	0.783
AFD-OLSSVM	1	28.643	36.908	0.353	36.743	5.32E–07	0.875
	2	38.658	50.439	0.480	50.356	7.27E–07	0.863
	3	66.350	98.465	0.811	98.169	1.42E–06	0.808
AFD-S-OLSSVM	1	26.998	35.280	0.333	35.280	5.08E–07	0.883
	2	37.343	49.374	0.463	49.368	7.11E–07	0.825
	3	50.542	70.385	0.621	70.358	1.01E–06	0.829
S-OLSSVM	1	45.676	65.207	0.561	65.200	9.39E–07	0.808
	2	63.784	93.760	0.781	93.595	1.35E–06	0.779
	3	78.638	110.195	0.970	110.191	1.59E–06	0.821
EEMD-OLSSVM	1	45.990	61.882	0.574	61.792	8.92E–07	0.850
	2	70.837	96.745	0.875	96.737	1.39E–06	0.783
	3	77.201	105.424	0.960	105.038	1.52E–06	0.825
EEMD-S-OLSSVM	1	38.765	54.052	0.485	54.037	7.79E–07	0.858
	2	60.177	88.217	0.742	88.056	1.27E–06	0.796
	3	76.858	105.045	0.942	105.0282	1.51E–06	0.817

The best results are shown in bold

Fig. 4 Boxplots of the absolute forecasted error over the testing period in NSW market



6.3 Case 2: QLD market

In an effort to evaluate comprehensively the proposed integrated model, the electricity demand data collected from QLD market are applied in this subsection to make

comparisons. Moreover, the above-mentioned benchmark models are used to highlight the effectiveness and applicability of the proposed model. And the corresponding forecasting results of these models over one-step, two-step and three-step forecasts are provided in Table 5. The

Table 5 Performance evaluation of forecasting models using electricity demand data in QLD market

Forecasting models	Step	MAE	RMSE	MAPE	SDE	THI	DS
BPNN	1	46.500	59.179	0.766	58.978	1.62E–06	0.804
	2	60.243	82.501	0.987	81.958	2.25E–06	0.733
	3	63.808	86.366	1.067	85.586	2.36E–06	0.779
LSSVM	1	44.648	59.098	0.742	59.058	1.61E–06	0.829
	2	57.270	80.061	0.947	80.051	2.18E–06	0.758
	3	63.451	91.902	1.043	91.828	2.51E–06	0.804
OLSSVM	1	42.927	57.230	0.712	57.223	1.56E–06	0.821
	2	56.452	79.040	0.932	79.039	2.16E–06	0.767
	3	63.412	91.498	1.042	91.450	2.50E–06	0.800
AFD-OLSSVM	1	27.787	35.077	0.464	35.006	9.57E–07	0.888
	2	40.593	53.951	0.672	53.771	1.47E–06	0.821
	3	48.809	71.298	0.802	71.298	1.95E–06	0.808
AFD-S-OLSSVM	1	21.235	25.864	0.360	25.725	7.06E–07	0.875
	2	27.381	35.973	0.465	35.592	9.82E–07	0.842
	3	37.125	52.239	0.618	52.188	1.43E–06	0.838
S-OLSSVM	1	36.388	47.001	0.618	46.449	1.28E–06	0.838
	2	47.116	61.393	0.799	60.106	1.68E–06	0.838
	3	58.756	80.090	0.995	78.862	2.19E–06	0.842
EEMD-OLSSVM	1	24.141	31.307	0.403	31.271	8.54E–07	0.850
	2	27.332	37.892	0.465	37.852	1.03E–06	0.821
	3	41.662	58.533	0.691	58.451	1.60E–06	0.825
EEMD-S-OLSSVM	1	22.526	29.956	0.373	29.950	8.18E–07	0.838
	2	34.941	49.394	0.578	49.137	1.35E–06	0.829
	3	53.101	76.474	0.881	75.897	2.09E–06	0.821

The best results are shown in bold

values of MAPE are combined visually in the bottom of Fig. 5. Besides, the curve of forecast and the scatter diagram given by the proposed model and benchmark models are shown in Fig. 5.

The ensemble forecasting results of each model present that the proposed integrated model (AFD-S-OLSSVM) can improve the forecasting accuracy significantly based on a comparison with BPNN, LSSVM, OLSSVM, AFD-OLSSVM, S-OLSSVM, EEMD-OLSSVM and EEMD-S-OLSSVM models according to six evaluation criteria. Particularly, the MAPE values of the proposed model are 0.618, 0.799 and 0.995 over one-step, two-step and three-step forecasts, respectively. From Table 5, the proposed decomposition-ensemble model shows the significant improvement through comparing the AFD-OLSSVM model with OLSSVM model and the AFD-S-OLSSVM model with AFD-OLSSVM model. In addition, the ensemble model based on AFD technology obtains higher performance metrics than that based on EEMD technology. So, the proposed integrated model obviously outperforms all the benchmark models.

To vividly express the forecasting effectiveness, the statistical histograms of MAPE values are shown in Fig. 5,

which indicates that the proposed integrated model is superior and robust for electricity demand forecasting. From Fig. 5, the proposed model shows apparently a better curve fitting of the actual electricity demand over one-step and multi-step forecasts and the scatter plots of measured and predicted electricity demands for the one-step, two-step and three-step forecasting horizons using the proposed integrated model are also depicted in Fig. 5. *R* value represents the relationship between the forecasted and observed data. Although the *R* value is generally goes down as more forecasting steps, it is clear that the proposed model can yield high forecasting accuracy.

6.4 Case 3: SA market

This subsection also focuses on the effectiveness of the proposed integrated model. And the electricity demand data collected from SA market in August 1st–August 31st, 2013 were used. Table 6 displays the forecasting results of the proposed model and the benchmark model. From the results of Table 6, it shows that the proposed model also ranks first for one-step and multi-step electricity demand forecasting with the minimum value of the forecasting

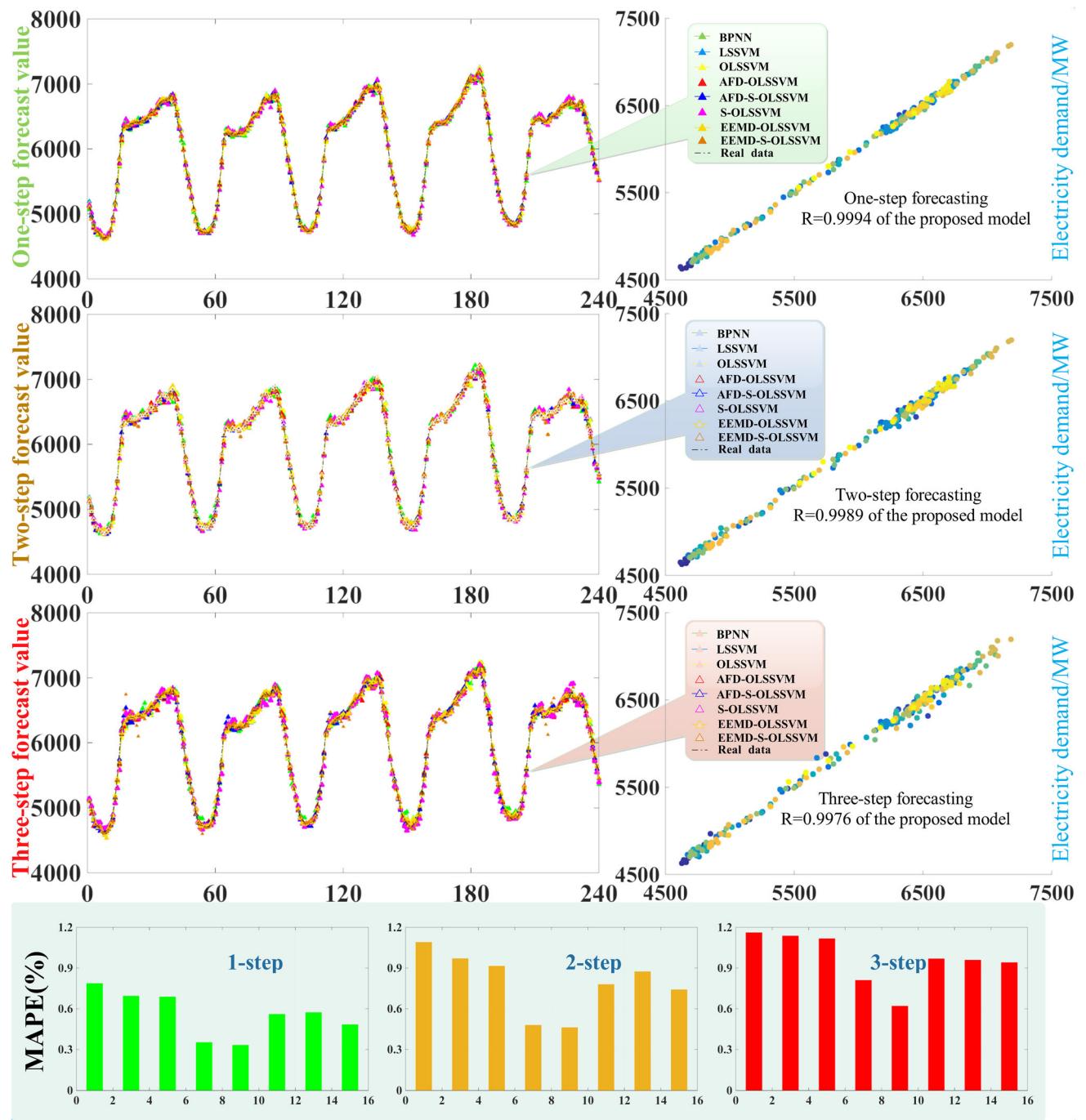


Fig. 5 Measured versus predicted electricity demand output at first, second and third forecast steps

performance indices, such as MAE, RMSE, MAPE and SDE. Besides, the Theil inequality (THI) coefficient of the AFD-S-OLSSVM model is $1.46\text{E}-05$, $2.09\text{E}-05$ and $2.66\text{E}-05$ for one-step, two-step and three-step forecasts. The THI index is a relative quantity, which indicates that the closer is closed to 0, the smaller the error is. Generally,

the range of the THI coefficient is $[0, 1]$. And the values of the DS index are also presented in Table 6.

In addition, Fig. 6 shows the performance comparison of different models in terms of MAPE. As shown in Fig. 6, among the error bars of all the models, the proposed model has the lowest forecasting value over different horizon

Table 6 Performance evaluation of forecasting models using electricity demand data in SA market

Forecasting models	Step	MAE	RMSE	MAPE	SDE	THI	DS
BPNN	1	28.808	42.689	2.165	42.335	2.38E–05	0.779
	2	36.746	53.005	2.779	52.938	2.95E–05	0.733
	3	53.641	73.811	4.046	71.573	4.11E–05	0.738
LSSVM	1	27.355	39.930	2.046	39.434	2.22E–05	0.792
	2	38.023	52.163	2.841	51.233	2.90E–05	0.750
	3	46.356	64.111	3.476	62.196	3.57E–05	0.763
OLSSVM	1	27.224	39.693	2.034	39.178	2.21E–05	0.792
	2	37.601	52.020	2.812	51.074	2.90E–05	0.754
	3	46.153	63.888	3.452	62.141	3.56E–05	0.783
AFD-OLSSVM	1	13.392	21.404	1.013	21.340	1.19E–05	0.871
	2	19.876	28.760	1.523	27.790	1.60E–05	0.808
	3	29.459	42.291	2.228	41.321	2.35E–05	0.783
AFD-S-OLSSVM	1	12.338	18.989	0.926	18.875	1.06E–05	0.904
	2	17.774	25.149	1.354	24.561	1.40E–05	0.833
	3	24.763	34.649	1.907	33.674	1.93E–05	0.821
S-OLSSVM	1	20.727	26.182	1.608	25.480	1.46E–05	0.850
	2	27.778	37.481	2.144	36.238	2.09E–05	0.792
	3	35.734	47.748	2.757	46.465	2.66E–05	0.800
EEMD-OLSSVM	1	25.737	33.201	1.987	32.994	1.85E–05	0.825
	2	35.562	49.568	2.650	48.723	2.76E–05	0.717
	3	45.878	63.590	3.470	63.053	3.54E–05	0.733
EEMD-S-OLSSVM	1	20.727	26.182	1.608	25.480	1.46E–05	0.850
	2	27.778	37.481	2.144	36.238	2.09E–05	0.792
	3	35.734	47.748	2.757	46.465	2.66E–05	0.800

The best results are shown in bold

forecasting. Meanwhile, the MAPE values of the proposed model are the lowest as 1.608, 2.144 and 2.757.

Remark The forecasting results made by the models with adaptive Fourier decomposition are superior to those from the other models without the decomposition process. Besides, the results of the models with seasonal adjustment method have the great accuracy than those models without seasonal adjustment method. That is to say, the adaptive Fourier decomposition and seasonal adjustment method play an important role in improving the forecasting performance for the data with complex nonlinear characteristic and an obvious seasonal tendency. Besides, the proposed model takes obvious advantage in one-step, two-step and three-step forecasting horizons. It should be noted that the optimization algorithm can provide the suitable parameters for forecast engine.

7 Conclusions and future advances

Forecasting of electricity demand is of crucial importance for daily operation and management of the power system. In recent years, the accuracy and effectiveness of electricity forecasting technology have been considered in order to capture the nonlinear and non-stationary characteristics of electricity demand. Accurate electricity demand forecasting plays a significant role in demand side for optimal economic dispatch to maximize renewable and minimize the operational cost. With respect to the forecast model, it may fail to produce day-to-day forecast if the necessary factors are not considered.

In this research study, a new concept of the integrated model is developed for electricity demand forecasting. The nonlinear and non-stationary characteristics and the seasonal cycles in electricity demand data can be captured by input variable preprocessing method. The adaptive Fourier decomposition technology is firstly applied to process the

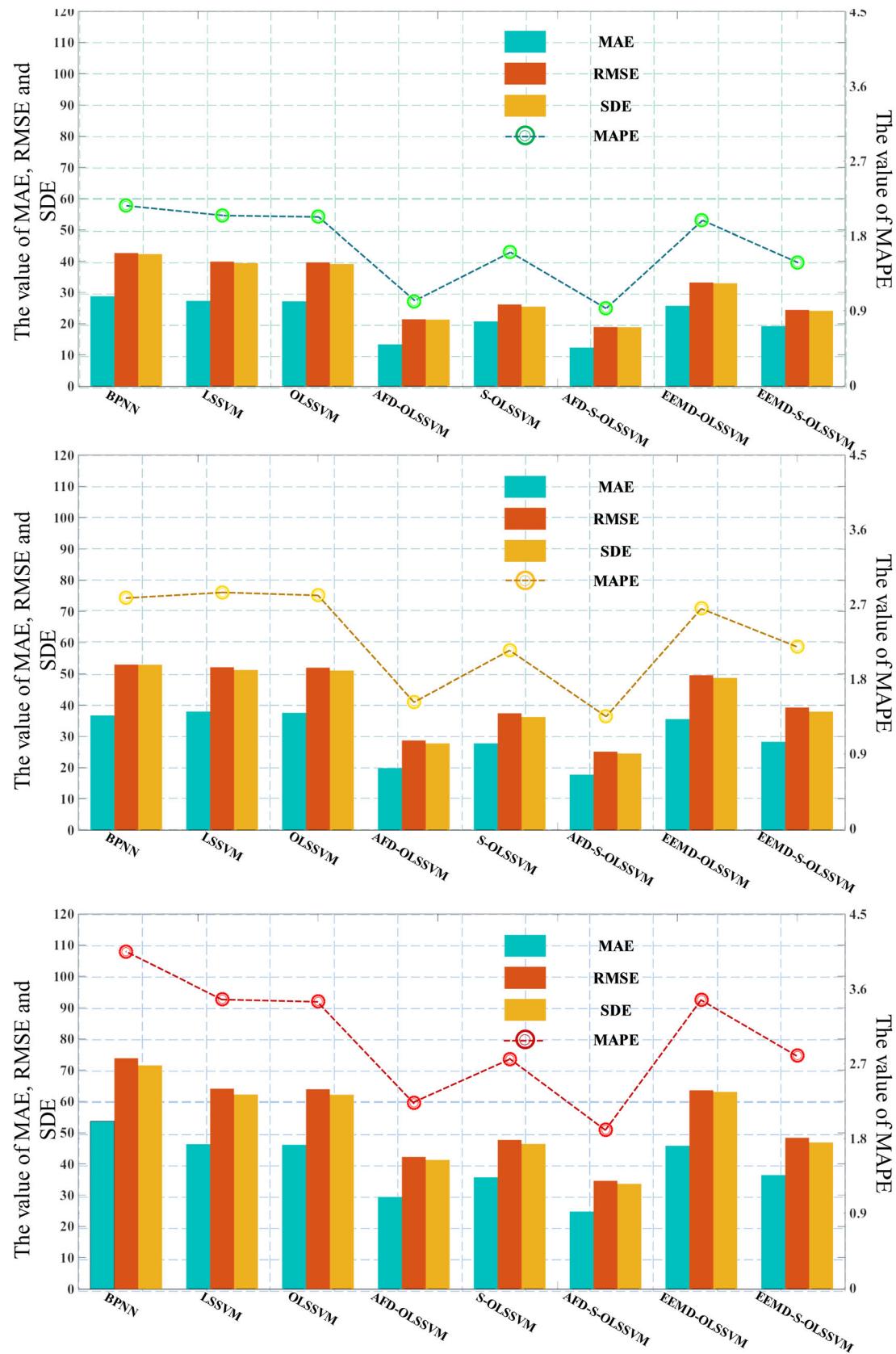


Fig. 6 Performance comparison of different models in terms of MAPE in SA market

electricity demand data, after which the seasonality in the processing series is removed by the seasonal adjustment method before conducting prediction. In addition, moth-flame optimization algorithm is used to tune the parameters of the forecast engine. Through the experimental results of the three markets of Australia, it shows that the proposed integrated model outperforms the benchmark model. Meanwhile, based on the adaptive Fourier decomposition method and seasonal adjustment method, the results indicate that this input variable preprocessing strategy can significantly improve the accuracy of electricity demand forecasting. Thus, the proposed model can provide an effective and powerful mining tool to support forecasting of electricity demand.

However, the challenges in Smart Grids from a stochastic and data-driven point of view are processing continually. Precise electricity demand forecasts are potential for demand management to balance power supply in the grid. Demand side is a key component toward sustainability and efficiency in smart grid because the ultimate goal of smart grid is to energy supply and balance effectively. Importantly, smart grid enables the participants and decision maker of electricity market to adjust and make their bidding strategies. In this regard, the accurate forecast of electricity demand in smart grids is an unavoidable problem as growing recognition of electricity grid modernization. Besides, with the change in inherent character of data in power engineering, prediction in smart grid environment becomes more complex than in the conventional energy market.

This paper mainly focuses on an efficient approach for short-term electricity demand forecast. Besides, long-term prediction for the electricity demand is also very useful for power scheduling. The most important factors in long-term

prediction are the predictor variables and vanishing gradient problem. When using the developed model to perform long-term demand forecasting, the results may be sensitive to the initial value and the model may lead to poor adaptability for emergencies. When dealing with long-term problems, the hidden dependencies/patterns existing in electricity demand should be fully considered in the forecasting system, so the accuracy and stability of the demand forecasting improved. In addition, long-term electricity demand forecasting can rely on the combination of social economic development, econometric and artificial intelligence methods.

From the perspective of superior performance, some sequential modeling methods would be widely used, such as recurrent neural network and long short-term memory units. In the power system coordination and operation, the factors that exert an impact on electricity demand forecasting are diversifying. It is potential to develop an ideal deep learning method for electricity demand forecasting. One future challenge will be the relevance between precise measurement of electricity and appropriate learning framework. It would be also be interesting to extend our work to deep learning methods.

Acknowledgements This work was supported by the National Natural Science Foundation of China [Grant Number 71573034].

Compliance with ethical standards

Conflict of interest All the authors declare that there is no any possible conflict of interest regarding the publication of this paper.

Appendix A

Algorithm: Pseudo-code of moth-flame optimization algorithm**Input:**

$x_s = (x(1), x(2), \dots, x(q))$ —training data.
 $x_p = (x(q+1), x(q+2), \dots, x(q+d))$ —verifying data

Output:

x_b —the best fitness value as the best estimation for the global optimum

Parameters:

dim —number of search agents lb —Lower Bound of Variables
 N —number of search agents ub —Upper Bound of Variables
 $Iter_{max}$ —the maximum number of iterations

```

1: /*Read and reshape training data and verifying data*/
2: /*input/output data normalization*/
3: /*Initialize parameters of MFO, such as search agents, bounds and iterations*/
4: /*Load details of the selected benchmark function*/
5: /*Initialize the positions of moths*/
6: WHILE ( $Iter < Max\_Iter$ ) DO
7:   /*Update flame number using  $flame\_no = round\left(N - l * \frac{N-1}{T}\right)$ */
8:   FOR EACH  $i: 1 \leq i \leq N$  DO
9:     /* Check if moths go out of the search space and bring it back */
10:    /* Calculate the fitness of moths, OM = Fitness function ( $M$ ) */
11:   END FOR
12:   IF  $Iter == 1$  Then
13:     /* Sort the first population of moths, F = sort(M) */
14:     /* Update the flames */
15:   ELSEIF
16:     /* Sort the first population of moths */
17:     /* F = sort( $M_{i-1}, M_i$ ), OF = sort( $M_{i-1}, M_i$ ) */
18:     /* Update the flames */
19:   END IF
20:   /* Update the position best flame obtained so far */
21:   /* a linearly decreases from -1 to -2 to calculate t */
22:    $S(M_i, F_j) = D_i \cdot e^{bt} \cdot \cos(2\pi t) + F_j$ 
23:   FOR EACH  $i: 1 \leq i \leq N$  DO
24:     FOR EACH  $j: 1 \leq j \leq N$  DO
25:       IF  $(i \leq flame\_no)$  Then
26:         /* Update the moth position concerning corresponding flame */
27:          $D_i = |F_j - M_i|$ 
28:          $S(M_i, F_j) = D_i \cdot e^{bt} \cdot \cos(2\pi t) + F_j$ 
29:       END IF
30:       IF  $(i > flame\_no)$  Then
31:         /* Update the moth position concerning one flame */
32:          $D_i = |F_j - M_i|$ 
33:          $S(M_i, F_j) = D_i \cdot e^{bt} \cdot \cos(2\pi t) + F_j$ 
34:       END IF
35:       Calculate Distance with respect to the corresponding moth
36:       Update  $M(i, j)$  with respect to the corresponding moth
37:     END EACH
38:   END EACH
39: END WHILE
40: Subject to the Cost function =  $\min(F(x))$ .
41: Stopping searching and obtain the best flame (score)
42: Return the  $x_b$ 

```

References

- Fan S, Chen LN (2006) Short-term load forecasting based on an adaptive hybrid method. *IEEE Trans Power Syst* 21:392–401
- Sheikhan M, Mohammadi N (2013) Time series prediction using PSO-optimized neural network and hybrid feature selection algorithm for IEEE load data. *Neural Comput Appl* 23(3–4):1185–1194
- Ying LC, Pan MC (2008) Using adaptive network based fuzzy inference system to forecast regional electricity loads. *Energy Convers Manag* 49(2):205–211
- Pappas SS, Ekonomou L, Karamousantas DC, Chatzarakis GE, Katsikas SK, Liatsis P (2008) Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models. *Energy* 33(9):1353–1360
- Fang TT, Lahdelma R (2016) Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Appl Energy* 179:544–552
- Huang YJ, Wang H, Khajepour A, He H, Ji J (2017) Model predictive control power management strategies for HEVs: a review. *J Power Sources* 341:91–106
- Xie NM, Yuan CQ, Yang YJ (2015) Forecasting China's energy demand and self-sufficiency rate by grey forecasting model and Markov model. *Int J Electr Power Energy Syst* 66:1–8
- Wang J, Zhu S, Zhang W, Lu H (2010) Combined modeling for electric load forecasting with adaptive particle swarm optimization. *Energy* 35(4):1671–1678
- Xia C, Wang J, McMenemy K (2010) Short, medium and long term load forecasting model and virtual load forecaster based on radial basis function neural networks. *Int J Electr Power Energy Syst* 32(7):743–750
- Bercu S, Proia F (2013) A SARIMAX coupled modelling applied to individual load curves intraday forecasting. *J Appl Stat* 40(6):1333–1348
- Hsu CC, Chen CY (2003) Regional load forecasting in Taiwan—applications of artificial neural networks. *Energy Convers Manag* 44(12):1941–1949
- Qiu XH, Suganthan PN, Amaralunga GAJ (2018) Ensemble incremental learning Random Vector Functional Link network for short-term electric load forecasting. *Knowl Based Syst* 145:182–196
- Zhang H, Li J, Ji Y, Yue H (2016) Understanding subtitles by character-level sequence-to-sequence learning. *IEEE Trans Ind Inform* 13(2):616–624
- Xiao L, Wang JZ, Dong Y, Wu J (2015) Combined forecasting models for wind energy forecasting: a case study in China. *Renew Sustain Energy Rev* 44:271–288
- Wang JZ, Chi DZ, Wu J, Lu HY (2011) Chaotic time series method combined with particle swarm optimization and trend adjustment for electricity demand forecasting. *Expert Syst Appl* 38(7):8419–8429
- El-Telbany M, El-Karmi F (2008) Short-term forecasting of Jordanian electricity demand using particle swarm optimization. *Electr Power Syst Res* 78(3):425–433
- Azadeh A, Ghaderi SF, Tarverdian S, Saberi M (2007) Integration of artificial neural networks and genetic algorithm to predict electrical energy consumption. *Appl Math Comput* 186(2):1731–1741
- Ghanbari A, Kazemi SMR, Mehmanpazir F, Nakhostin MM (2013) A cooperative ant colony optimization-genetic algorithm approach for construction of energy demand forecasting knowledge-based expert systems. *Knowl Based Syst* 39:194–206
- Hong WC (2010) Application of chaotic ant swarm optimization in electric load forecasting. *Energy Policy* 38(10):5830–5839

20. Kiran MS, Ozceylan E, Gunduz M, Paksoy T (2012) A novel hybrid approach based on particle swarm optimization and ant colony algorithm to forecast energy demand of Turkey. *Energy Convers Manag* 53(1):75–83
21. Reddy KS, Panwar LK, Panigrahi BK, Kumar R (2018) Solution to unit commitment in power system operation planning using binary coded modified moth flame optimization algorithm (BMMFOA): a flame selection based computational technique. *J Comput Sci* 25:298–317
22. Zhao J et al (2016) An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed. *Appl Energy* 162:808–826
23. Jiang P, Li RR, Zhang KQ (2018) Two combined forecasting models based on singular spectrum analysis and intelligent optimized algorithm for short-term wind speed. *Neural Comput Appl* 30:1–19
24. Pothiya S, Ngamroo I, Kongprawechnon W (2008) Application of multiple tabu search algorithm to solve dynamic economic dispatch considering generator constraints. *Energy Convers Manag* 49(4):506–516
25. Hadji MM, Vahidi B (2012) A solution to the unit commitment problem using imperialistic competition algorithm. *IEEE Trans Power Syst* 27(1):117–124
26. Simopoulos DN, Kavatza SD, Vournas CD (2006) Unit commitment by an enhanced simulated annealing algorithm. *IEEE Trans Power Syst* 21(1):68–76
27. Piltan M, Shiri H, Ghaderi SF (2012) Energy demand forecasting in Iranian metal industry using linear and nonlinear models based on evolutionary algorithms. *Energy Convers Manag* 58:1–9
28. Barman M, Choudhury NBD, Sutradhar S (2018) A regional hybrid GOA-SVM model based on similar day approach for short-term load forecasting in Assam, India. *Energy* 145:710–720
29. Zhang H, Cao X et al (2016) Object-level video advertising: an optimization framework. *IEEE Trans Ind Inform* 13(2):520–531
30. Yadav V, Srinivasan D (2011) A SOM-based hybrid linear-neural model for short-term load forecasting. *Neurocomputing* 74(17):2874–2885
31. Soares LJ, Souza LR (2006) Forecasting electricity demand using generalized long memory. *Int J Forecast* 22(1):17–28
32. Bessec M, Fouquau J (2018) Short-run electricity load forecasting with combinations of stationary wavelet transforms. *Eur J Oper Res* 264(1):149–164
33. Jiang P, Liu F, Song YL (2017) A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting. *Energy* 119:694–709
34. Andersen FM, Larsen HV, Boomsma TK (2013) Long-term forecasting of hourly electricity load: identification of consumption profiles and segmentation of customers. *Energy Convers Manag* 68(3):244–252
35. Han H, Zhong Z, Wen C, Sun H (2018) Agricultural environmental total factor productivity in China under technological heterogeneity: characteristics and determinants. *Environ Sci Pollut Res* 25:32096–32111
36. Zhang GP, Qi M (2005) Neural network forecasting for seasonal and trend time series. *Eur J Oper Res* 160(2):501–514
37. Wang JZ, Zhu WJ, Zhang WY, Sun DH (2009) A trend fixed on firstly and seasonal adjustment model combined with the ε-SVR for short-term forecasting of electricity demand. *Energy Policy* 37(11):4901–4909
38. Abdoos AA (2016) A new intelligent method based on combination of VMD and ELM for short term wind power forecasting. *Neurocomputing* 203:111–120
39. Wang Z, Wan F, Wong CM, Zhang LM (2016) Adaptive Fourier decomposition based ECG denoising. *Comput Biol Med* 77:195–205
40. Benitez D, Gaydecki PA, Zaidi A, Fitzpatrick AP (2001) The use of the Hilbert transform in ECG signal analysis. *Comput Biol Med* 31(5):399–406
41. Seyedali M (2015) Moth-flame optimization algorithm: a novel nature-inspired heuristic paradigm. *Knowl Based Syst* 89:228–249
42. Allam D, Yousri DA, Eteiba MB (2016) Parameters extraction of the three diode model for the multi-crystalline solar cell/module using moth-flame optimization algorithm. *Energy Convers Manag* 123:535–548
43. Hassanien AE, Gaber T, Mokhtar U, Hefny H (2017) An improved moth flame optimization algorithm based on rough sets for tomato diseases detection. *Comput Electron Agric* 136:86–96
44. Lin KP, Pai PF (2016) Solar power output forecasting using evolutionary seasonal decomposition least-square support vector regression. *J Clean Prod* 134:456–462
45. Ahmadi MA, Masoumi M, Askarinezhad R (2015) Evolving Smart model to predict the combustion front velocity for in situ combustion. *Energy Technol* 3(2):128–135
46. Moradi MH, Abedini M, Tousi SMR, Hosseiniyan SM (2015) Optimal siting and sizing of renewable energy sources and charging stations simultaneously based on differential evolution algorithm. *Int J Electr Power Energy Syst* 73:1015–1024
47. Li B, Li DY, Zhang ZJ, Yang SM, Wang F (2015) Slope stability analysis based on quantum-behaved particle swarm optimization and least squares support vector machine. *Appl Math Model* 39(17):5253–5264
48. Mahani AS, Shojaee S, Salajegheh E, Khatibinia M (2015) Hybridizing two-stage meta-heuristic optimization model with weighted least squares support vector machine for optimal shape of double-arch dams. *Appl Soft Comput* 27:205–218
49. Zhu BZ, Wei YM (2013) Carbon price forecasting with a hybrid ARIMA and least squares support vector machines methodology. *Omega Int J Manag Sci* 41(3):517–524

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.