

Article

An Ensemble Model Based on Machine Learning Methods and Data Preprocessing for Short-Term Electric Load Forecasting

Yanbing Lin ¹, Hongyuan Luo ^{1,*}, Deyun Wang ^{1,2}, Haixiang Guo ^{1,2} and Kejun Zhu ¹

¹ School of Economics and Management, China University of Geosciences, Wuhan 430074, China; lin.yanbing@cug.edu.cn (Y.L.); zhukejun@cug.edu.cn (K.Z.)

² Mineral Resource Strategy and Policy Research Center, China University of Geosciences, Wuhan 430074, China; wangdy@cug.edu.cn (D.W.); ghx@cug.edu.cn (H.G.)

* Correspondence: luohongyuan@cug.edu.cn; Tel.: +86-027-6784-8527

Received: 14 July 2017; Accepted: 7 August 2017; Published: 11 August 2017

Abstract: The experience with deregulated electricity market has shown the increasingly important role of short-term electric load forecasting in the energy producing and scheduling. However, because of nonlinear, stochastic and nonstable characteristics associated with the electric load series, it is extremely difficult to precisely forecast the electric load. This paper aims to establish a novel ensemble model based on variational mode decomposition (VMD) and extreme learning machine (ELM) optimized by differential evolution (DE) algorithm for multi-step ahead electric load forecasting. The proposed model is novel in the sense that VMD is firstly applied to decompose the original electric load series into a set of components with different frequencies in order to effectively eliminate the stochastic fluctuation characteristic so as to improve the overall prediction accuracy. The proposed ensemble model is tested using two electric load series collected from New South Wales (NSW) and Queensland (QLD) in the Australian electricity market. The experimental results show that: (1) the data preprocessing by VMD can effectively decrease the stochastic fluctuation characteristics that existed in the electric load series, consequently improving the whole forecasting accuracy, and; (2) the proposed forecasting model performs better than all other benchmark models for both one-step and multi-step ahead electric load forecasting.

Keywords: electric load forecasting; variational mode decomposition; extreme learning machine; differential evolution

1. Introduction

Electric power is a clean and efficient energy which plays an increasingly vital role in our daily life. Compared with traditional energy such as natural gas, coal and oil, electric power is more suitable and efficient for the requirement of environment-friendly society. In the field of power system planning, the accuracy of electric load forecasting is of great importance for energy generating capacity scheduling and power system management [1,2]. An overestimation may waste many energy resources and significantly improve the operational costs, and an underestimation will decrease the reliability of the power system and cannot meet the requirement of electricity utilization [3]. Therefore, accurate electric load forecasting is essential and significant for power systems. However, because the electric load series is affected by many complicated influencing factors, it is really a challenging job to accurately predict the electric load.

Over the past several decades, various interesting models for the electric load forecasting have been established, and they can be generally divided into the following two types. The first type is the multi-factor forecasting method, which needs to search the causal relationships between different

influencing factors and forecasting values [4]. The other one is time series forecasting method which is based on the historical series. As is mentioned above, the electric load series is affected by various complicated and nonobjective factors which are very difficult to be controlled in practical application and, consequently, it is a very challenging job to establish an accurate forecasting model by utilizing the multi-factor forecasting method. Based on the above considerations, lots of researchers turn to utilizing the time series forecasting method to forecast electric load [5]. Compared to multi-factor forecasting method, the time series forecasting method is much easier and quicker. The most frequently and widely used time series forecasting models can be further divided into the following three subcategories: statistical models; machine learning models, and; hybrid models [6]. As for the statistical models, the auto-regressive moving average (ARMA), auto-regressive integrated moving average (ARIMA), generalized autoregressive conditional heteroskedasticity (GARCH), vector auto-regression (VAR), and Kalman filters methods are widely used in the electric load forecasting areas. For example, Pappas et al. [7] established an ARMA model for short-term electric load forecasting, and the results showed the good performance of the proposed model. Kavousi-Fard and Kavousi-Fard [8] proposed a new hybrid model based on ARIMA for short-term load forecasting. Takiyar et al. [9] developed a GARCH-based hybrid model for short-term electric load forecasting. Garcia-Ascanio and Mate [10] utilized VAR-based interval time series model to forecast electric load forecasting. Takeda et al. [11] developed an ensemble model based on Kalman filter for short-term electric load forecasting, and the experiment indicated that the predicting accuracy of the proposed model is obviously better than that of the present state-of-the-art models.

With the fast development of machine learning method, many of them have been widely used in various forecasting issues, such as artificial neural network (ANN), extreme learning machine (ELM) and support vector machine (SVM). Yolcu et al. [12] developed a new ANN model based on both linear and nonlinear structures for time series predicting. Gutierrez-Corea et al. [13] focused on the usage of ANN in short-term global solar irradiance forecasting. Li et al. [14] proposed a novel model based on modified artificial bee colony (MABC) and ELM for short-term electric load forecasting, and the experimental results showed that the proposed hybrid model owned the best forecasting ability with comparison to the other benchmark models adopted in their paper. Zhou et al. [15] developed a hybrid model based on SVM for short-term wind speed forecasting. Chen and Lee [16] established a weighted least square support vector machine (LSSVM) model based on learning system for time series forecasting, and the experimental results demonstrated the effectiveness of the proposed model.

By integrating the advantages of different single forecasting models, hybrid forecasting models generally performs better than the single forecasting models and therefore are widely used in many forecasting areas. For example, Liu and Shi [17] proposed an ARMA-GARCH model for predicting short-term electricity prices. Voyant et al. [18] developed an original technique to forecast global radiation based on a hybrid ARMA-ANN model for numerical weather forecasting. Ismail et al. [19] developed a hybrid self-organizing maps (SOM)-LSSVM model combining SOM and LSSVM for time-series predicting, and the experiment in the paper showed that SOM-LSSVM outperforms the single LSSVM model. Even though the hybrid model performs better than the single forecasting model, they still cannot effectively deal with the nonlinear and nonstationary characteristics associated with majority of time series in practical life. Therefore, many researchers have integrated various data preprocessing techniques into the forecasting models in order to effectively decrease the forecasting errors. The most widely used data preprocessing techniques includes empirical mode decomposition (EMD), ensemble empirical mode decomposition (EEMD), wavelet packet transform (WPT) and variational mode decomposition (VMD). Liu [20] predicted short-term wind speed by a hybrid model combining wavelet transform (WT) and SVM optimized by genetic algorithm, and the simulation results demonstrated that the proposed hybrid model is more efficient than the other comparison models adopted in their paper. Wang et al. [21] developed a hybrid forecasting model based on WPT, particle swarm optimization (PSO), simulated annealing (SA), phase space reconstruction (PSR) and LSSVM for multi-step ahead wind speed forecasting, and the case studies showed that the proposed

model outperformed all the other comparison models. Zhang et al. [22] investigated the usage of WT-LSSVM model in the time series forecasting of fair-weather atmospheric electric field, and the experimental results showed that the proposed WT-LSSVM model is a superior method compared to the single LSSVM and ANN model. Wang et al. [23] developed a two-layer decomposition model for multi-step ahead electricity price forecasting, where VMD is specifically applied to decompose the high frequency sub-signals generated by fast ensemble empirical mode decomposition (FEEMD), and the experimental results illustrated the superior performances of the proposed model. For more details on different forecasting models, the reader can refer to the references [24–26], among others.

Although VMD has been utilized in many forecasting issues such as forecasting of wind power [27] and financial time series [28], the advantages of VMD have not been confirmed in the electric load forecasting area. Based on the above considerations, this paper aims to establish an ensemble model combining VMD and an improved ELM for short-term electric load forecasting. The process of the proposed model can be summarized into the following three steps: first, VMD is used to decompose the original electric load series into a set of components with different frequencies in order to effectively decrease the stochastic fluctuation characteristics; then, each component is forecasted using the ELM model whose initial weights and thresholds between the input layer and hidden layer are optimized by DE algorithm; finally, the ultimate forecasting series of electric load is obtained by aggregating the forecasting value of each component. Two real-world electric load series collected from New South Wales (NSW) and Queensland (QLD) located in Australia are used to test the effectiveness of the proposed ensemble model.

Based on aforementioned researches, the main novelties and contributions of this paper can be summarized to the following four aspects: (1) VMD technique is firstly applied to preprocess the electric load series in order to improve the overall forecasting accuracy; (2) VMD technique is combined with machine learning methods to develop a novel ensemble model for short-term electric load forecasting; (3) DE algorithm is employed to optimize the initial weights and thresholds of ELM in order to improve its function approximation ability; (4) the effectiveness of the proposed model is also examined by comparing with the forecasting models combining different decomposition techniques such as EMD and WT.

The rest of this paper is organized as follows. Section 2 firstly describes in detail the fundamental methods used in the paper, and then develops the hybrid VMD-DE-ELM model. Section 3 provides the data source and preprocessing results. Section 4 is the empirical study where the performance evaluating criteria, parameter settings and experimental results are presented. Section 5 concludes the paper.

2. Methodology

2.1. Empirical Mode Decomposition

Empirical mode decomposition (EMD) which is firstly proposed by Huang [29] is widely used signal decomposition method. Compared with other decomposition methods such as Fourier transformation and wavelet analysis, EMD owns many better temporal and frequency resolutions [30]. Due to the advantages of EMD, it has been widely used as a data preprocessing technique in many forecasting issues. EMD can effectively decrease the nonlinear, nonstable and stochastic characteristics associated with the signal through decomposing the signal into a set of components, that is, intrinsic mode functions (IMFs) and residual component. Each IMF generated by EMD has to meet the following two conditions: (1) for all IMFs, the number of extrema and the number of passing through zero must be equal or differ at most by one; (2) at any point, the mean value of the envelopes must be zero. Based on the above explanations, a time series denoted by signal $p(t)$ could be disassembled through the following three steps.

Step 1: Find out all local maxima and minima of a signal $p(t)$. Connect all maxima by a cubic spline line to generate the upper envelop, and connect all minima by another cubic spline line to

generate the lower envelop. The mean value of the upper envelop and the lower envelop is defined as m , and the difference of $p(t)$ and m is defined as h , which is illustrated in Equation (1).

$$h = p(t) - m \quad (1)$$

Step 2: Consider h as a new signal $p(t)$, and repeat Step (1) for enough times until h satisfies the conditions of IMF. The criterion (2) is used to decide if h is an IMF.

$$D_m = \frac{\sum_{t=0}^T |h_{(m-1)}(t) - h_m(t)|^2}{\sum_{t=0}^T |h_{(m-1)}(t)|^2} \quad (2)$$

where, if D_m is smaller than the predetermined value which can take the values between 0.2 and 0.3, h_m can be an IMF. The first IMF is named as c_1 , which is the highest frequency component. The other components can be expressed as the Equation (3).

$$r_1 = P(t) - c_1 \quad (3)$$

Step 3: Consider r_1 as the new signal $P(t)$, repeat the operations in Step (1) and Step (2) for times until r_j is smaller than the predetermined threshold or r_j is a monotone function. Thus, all the IMFs and the residue r can be obtained.

2.2. Variational Mode Decomposition

Variational mode decomposition (VMD), a new effective signal decomposition method proposed by Dragomiretskiy and Zosso [31] in 2014, has been frequently used in many practical applications. Many previous researches have shown the good ability of VMD in the signal denoising field against other decomposition methods such as EMD and WT. In the decomposition process of VMD, each mode μ_k can be compressed around a center pulsation ω_k . The procedure for obtaining each mode's bandwidth includes the following three steps: (1) for each mode, calculate the associated analytic signal with the benefit of Hilbert transform to obtain a unilateral frequency spectrum; (2) mix with an exponential tuned to the respective estimated center frequency in order to shift the mode's frequency spectrum to baseband; (3) estimate the bandwidth of each model through Gaussian smoothness of the demodulated signal. Then, the constrained variational problem can be illustrated as follows:

$$\min_{\mu_k, \omega_k} = \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * \mu_k(t) \right] e^{-j\omega_k t} \right\|_2 \right\} \quad (4)$$

Subject to

$$\sum_k \mu_k = f \quad (5)$$

where f represents the signal, μ denotes its mode, ω means the frequency, δ is the Dirac distribution, t is time script, k is the total number of modes, and $*$ represents the convolution function.

Then, the solution to the constrained variational problem (see Equation (4)) is the saddle point of the following augmented Lagrangian (L) expression:

$$L(\mu_k, \omega_k, \lambda) = \alpha \sum_k \left\| \partial_t \left[\delta(t) + \frac{j}{\pi t} * \mu_k(t) \right] \right\|_2^2 + \left\| f - \sum_k \mu_k \right\|_2^2 + \langle \lambda, f - \sum_k \mu_k \rangle \quad (6)$$

where λ and α represents the Lagrange multiplier and the balancing parameter of the data-fidelity constraint, respectively. Consequently, the solutions for u and ω can be obtained according to the following two formulas:

$$\mu_n^{n+1} = \left(f - \sum_{i \neq k} \mu_i + \frac{\lambda}{2} \right) \frac{1}{1 + 2\alpha(\omega - \omega_k)^2} \quad (7)$$

$$\omega_n^{n+1} = \frac{\int_0^\infty \omega |\mu_k(\omega)|^2 d\omega}{\int_0^\infty |\mu_k(\omega)|^2 d\omega} \quad (8)$$

where n represents the number of iterations.

2.3. The DE-ELM Model

2.3.1. Extreme Learning Machine

Extreme learning machine (ELM), which was first proposed by Huang [32], is a single-hidden layer feed-forward neural network. In the ELM model, the weights of hidden layer nodes are randomly selected, and the output weights of ELM are determined by a least-squares solution. In the following section, we will introduce the ELM model in detail.

Let $D = \{(x_i, t_i), i = 1, 2, \dots, N\}$ be a set with N different samples, where $x_i = [x_{i1}, x_{i2}, \dots, x_{im}]^T \in R$ is the inputs of sample, and $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T \in R$ is the outputs of the sample. The activation function of ELM is $g(x)$, the threshold value of the hidden layer is b_i and the output of ELM can be calculated by the following formula:

$$y_j = \sum_{i=1}^l \beta_i g_i(x_j) = \sum_{i=1}^l \beta_i g_i(w_i x_j + b_i) \quad j = 1, 2, \dots, n \quad (9)$$

where $w_i = [w_{i1}, w_{i2}, \dots, w_{im}]$ is the weights between the input layer and the hidden layer and $\beta_i = [\beta_{i1}, \beta_{i2}, \dots, \beta_{im}]$ is the weights between the hidden layer and the output layer.

Formula (9) can also be denoted as $Y = H\beta$, where Y represents the output of ELM and H means the output matrix of the hidden layer. The output matrix H can be formulated as follows:

$$H(w_1, \dots, w_l; b_1, \dots, b_l, x_1, \dots, x_n) = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & g(w_2 \cdot x_1 + b_2) & \dots & g(w_l \cdot x_1 + b_l) \\ g(w_1 \cdot x_2 + b_1) & g(w_2 \cdot x_2 + b_2) & \dots & g(w_l \cdot x_2 + b_l) \\ \vdots & \vdots & \dots & \vdots \\ g(w_1 \cdot x_n + b_1) & g(w_2 \cdot x_n + b_2) & \dots & g(w_l \cdot x_k + b_l) \end{bmatrix}_{n \times l} \quad (10)$$

2.3.2. Differential Evolution Algorithm

Differential Evolution (DE) algorithm, a simple population-based stochastic evolutionary algorithm first developed by Storn and Price [33], has been utilized for solving the combinatorial optimization problems in many application areas. Compared to other optimization algorithms, DE algorithm is easier and quicker, and only requires few control parameters. Generally, DE algorithm contains the following four steps: (1) initialization; (2) mutation; (3) crossover, and; (4) selection. The detailed information for each step is provided as follows:

Step 1: Initialize the population. The population of DE algorithm is initialized using the following formula.

$$x_{j,i,0} = x_{j,\min} + \text{rand}_{i,j}(0,1) \times (x_{j,\max} - x_{j,\min}) \quad (11)$$

where $x_{j,i,0}$ represents the value of i th individual in the 0th generation and j th dimension.

Step 2: Mutation. Randomly select the following three indices, m, i and $j, m \neq i \neq j$, then, a mutant vector $V_{k,G}$ can be obtained according to the following formula.

$$V_{k,G} = X_{m,G} + F(X_{i,G} - X_{j,G}) \quad (12)$$

where $k \neq m \neq i \neq j$, F represents a scaling factor and $F \in [0, 1]$, $X_{m,G}$ denotes the base vector.

Step 3: Crossover. The purpose of crossover operation is to increase the multiplicity of the perturbed parameter vectors. The trial point $U_{j,k,G+1}$ is established from its parents $V_{j,k,G+1}$ and $X_{j,k,G}$ by the following formula.

$$U_{j,k,G+1} = \begin{cases} V_{j,k,G+1} & \text{if } (\text{rand } b(j) \leq C_R) \text{ or } j = \text{rnbr}(i) \\ X_{j,k,G} & \text{if } (\text{rand } b(j) > C_R \text{ and } j \neq \text{rnbr}(i)) \end{cases} \quad j = 1, \dots, D, \quad (13)$$

where C_R is crossover probability and $C_R \in [0, 1]$, $\text{rnbr}(i)$ is a randomly selected index in the set of $\{1, 2, 3, \dots, D\}$, which ensures that $U_{j,k,G+1}$ obtains at least one parameter from $V_{j,k,G+1}$. The trial vector is formed of both current parameter vectors and mutant vector parameters.

Step 4: Selection. The trial vector $X_{i,G+1}$ can be obtained by comparing the fitness value of the vector obtained through mutation and crossover, and the process can be denoted as follows:

$$X_{i,G+1} = \begin{cases} U_{i,G} & \text{if } f(U_{i,G}) \leq f(X_{i,G}) \\ X_{i,G} & \text{otherwise} \end{cases} \quad (14)$$

Step 5: Repeat the above operations and stop the DE algorithm if the result satisfies the error requirement or the maximum number of iterations is reached. Otherwise, return to Step (2).

2.3.3. The DE-ELM Model

In order to reduce the non-stationarity and non-linearity characteristics of ELM forecasting model, DE algorithm is applied to optimize the initial weights and thresholds between input layer and hidden layer of ELM model. When the DE algorithm is stopped in the training process, the ELM model is established based on the optimized weights and thresholds determined by DE algorithm, and can be used to forecast the electric load series directly in the testing process.

$$F_{fitness} = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{X}(t) - X(t))^2} \quad (15)$$

The fitness function of DE algorithm is the root mean square error (RMSE) of forecasting which is defined as follows:

- Step 1: Initialize the individuals and the parameters of DE algorithm;
- Step 2: Set the iterative variable and calculate the fitness value of each individual by using Equation (15). Stop the DE algorithm when the stop criterion is satisfied. Go to Step (4);
- Step 3: Update the population. Go to Step (2);
- Step 4: When DE algorithm is stopped, the optimal individual is taken as the initial connection weights and thresholds of ELM;
- Step 5: Train and test ELM based on the training and testing samples.

2.4. The VMD-DE-ELM Forecasting Model

In this paper, an ensemble model based on machine learning methods and data preprocessing is proposed for short-term electric load forecasting. The basic structure of the ensemble method is based on decomposition and ensemble, which is shown in Figure 1 where Mode_1' , Mode_2' , ..., Mode_n' respectively represent the forecasting values of Mode_1 , Mode_2 , ..., Mode_n .

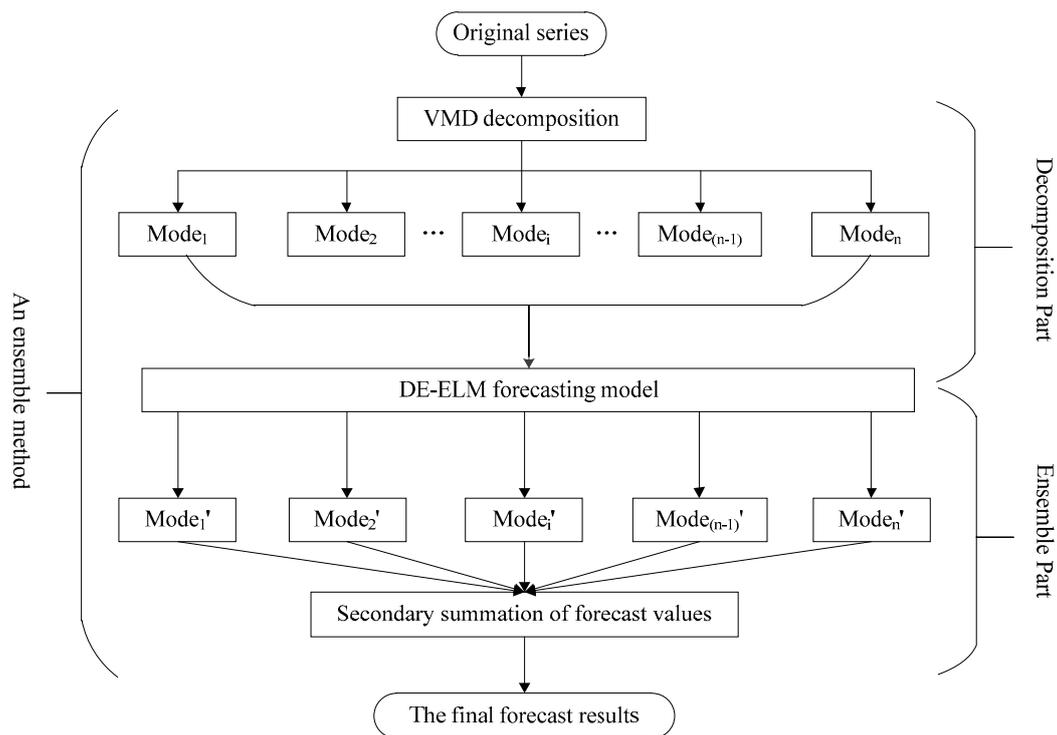


Figure 1. The flowchart of VMD-DE-ELM model.

The steps of the proposed ensemble model are shown as follows:

Step 1: Data preprocessing. VMD is used to decompose the original electric load series into several sub-series with different frequencies, namely, $Mode_1, Mode_2, \dots, Mode_n$;

Step 2: Individual mode forecasting. Each mode is forecasted using the ELM model improved by the DE algorithm, and the forecasting result of each mode is obtained, namely, $Mode'_1, Mode'_2, \dots, Mode'_n$;

Step 3: Equal-weight aggregation. The modes generated by VMD are equally weighted and then aggregated in order to obtain the final forecasting results based on the following formula.

$$\text{Final result} = Mode'_1 + Mode'_2 + \dots + Mode'_n \quad (16)$$

3. Data Description and Preprocessing

In this paper, two half-hour electric load series collected from NSW and QLD [34] are adopted to test the effectiveness of the proposed model. In Australia's electricity market, there are 48 observation data points every day, which means the time gap of observation values is half an hour. In this paper, each electric load series contains 1488 observation data points (from 1 January 2017 to 31 January 2017), see Figure 2. Moreover, in each case, the 1st–1200th observation points are taken as training sets, and the rest, the 1201st–1488th observation points, are taken as testing sets. The descriptive statistics of the two electric load series is shown in Table 1.

Table 1. Descriptive statistics of the electric load series.

Site	Descriptive Statistics						
	N	Mean	Standard Deviation	Minimum Value	Maximum Value	Coefficient of Skewness	Coefficient of Kurtosis
NSW	1488	8588.11	1812.13	5767.31	13947.70	0.786	0.024
QLD	1488	7025.13	964.55	5254.92	9357.09	0.171	−0.989

As is shown in Table 1, the mean of electric load series of NSW is 8588.11, and the minimum value and maximum value of electric load series of NSW is 5767.31 and 13947.70, respectively. It is obvious that the maximum value of electric load series of NSW is larger than twice of the minimum value, which illustrates that the electric load series of NSW owns notable stochastic fluctuation characteristic. In QLD, the mean of electric load series is 7025.13, and the minimum value and maximum value of electric load series is respectively 5254.92 and 9357.09. The stochastic fluctuation characteristic of electric load series of QLD is also obvious. In addition, it should be noted that all numerical experiments in this paper are coded in MATLAB R2010a.

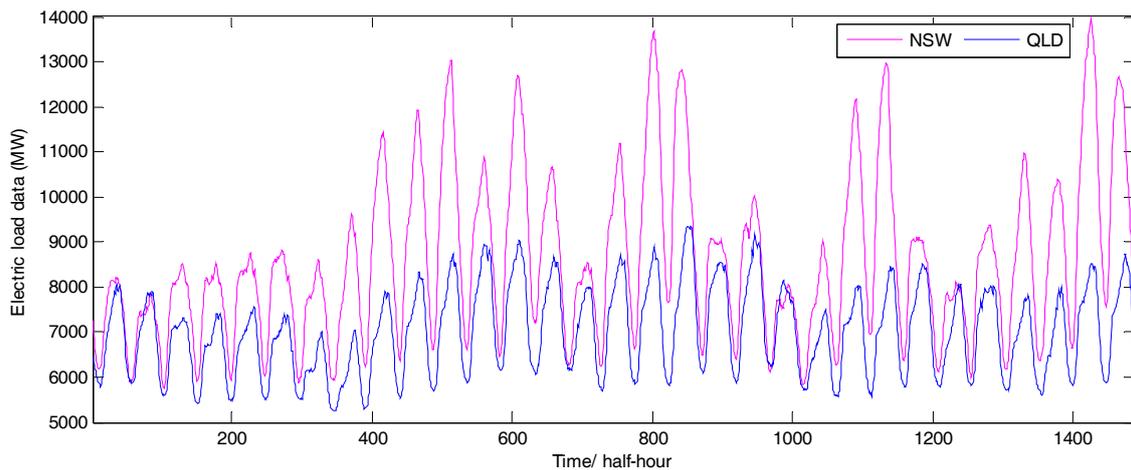


Figure 2. The original electric load series of NSW and QLD.

In order to make the forecasting results own more practical significance, this paper adopts multi-step ahead electric load forecasting. Compared with one-step ahead forecasting, multi-step ahead forecasting can supply more information for electricity market participants. Since multi-step ahead forecasting has to deal with many other additional complications such as the accumulation of errors, increased uncertainty and reduced accuracy [35], it is therefore more difficult to obtain precise forecasting results. This paper aims to propose an ensemble model for electric load forecasting over different horizons. Different horizons have different practical meanings. For instance, as for the half-hour data source, four-step ahead represents two-hour ahead, eight-step ahead represents four-hour ahead, and twelve-step ahead represents six-hour ahead. Furthermore, as shown in Figure 3, we take four-step ahead forecasting as an example to clearly interpret the multi-step ahead forecasting process. As shown in Figure 3, in the training process, the input and output datasets of DE-ELM model for four-step ahead forecasting are respectively $\{x_1, x_2, \dots, x_8\}, \{x_2, x_3, \dots, x_9\}, \dots, \{x_{1189}, x_{1190}, \dots, x_{1196}\}$ and $\{x_{12}, x_{13}, \dots, x_{1200}\}$, and in the testing process, the input and output datasets of DE-ELM model are respectively $\{x_{1190}, x_{1191}, \dots, x_{1197}\}, \{x_{1191}, x_{1192}, \dots, x_{1198}\}, \dots, \{x_{1477}, x_{1478}, \dots, x_{1484}\}$ and $\{x_{1201}, x_{1202}, \dots, x_{1488}\}$.



Figure 3. The input and output data selection process.

4. Empirical Study

4.1. Performance Criteria of Forecasting Accuracy

This paper selects the following three error measures to evaluate the performances of the proposed ensemble model: mean absolute error (MAE), root mean square error (RMSE) and mean absolute percentage error (MAPE). The computational formulas of these three error measures are provided as follows:

$$MAE = \frac{1}{N} \sum_{t=1}^N |\hat{X}(t) - X(t)| \quad (17)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^N (\hat{X}(t) - X(t))^2} \quad (18)$$

$$MAPE = \frac{1}{N} \sum_{t=1}^N \left| \frac{\hat{X}(t) - X(t)}{X(t)} \right| \quad (19)$$

where N represents the number of observations, $X(t)$ means the actual electric load value at time t , and $\hat{X}(t)$ denotes the forecasting value at time t .

4.2. Multi-Step Ahead Electric Load Forecasting in NSW

4.2.1. Data Preprocessing of the Original Electric Load Series

The electric load series is preprocessed using VMD in order to effectively decrease the stochastic fluctuations existing in the data series. Much previous research has shown that the number of components generated by VMD have significant effects on the characteristics of modes. In the decomposition process of VMD, more components will result in more stationary modes; however, more components might cause decomposition information loss and thus low forecasting accuracy. Based on the above considerations, after extensive numerical experiments, this study decomposes the data series into nine modes in order to ensure good forecasting accuracy. The decomposition results are shown in Figure 4 where the nine components are denoted by Mode1, Mode2, ..., Mode9, respectively.

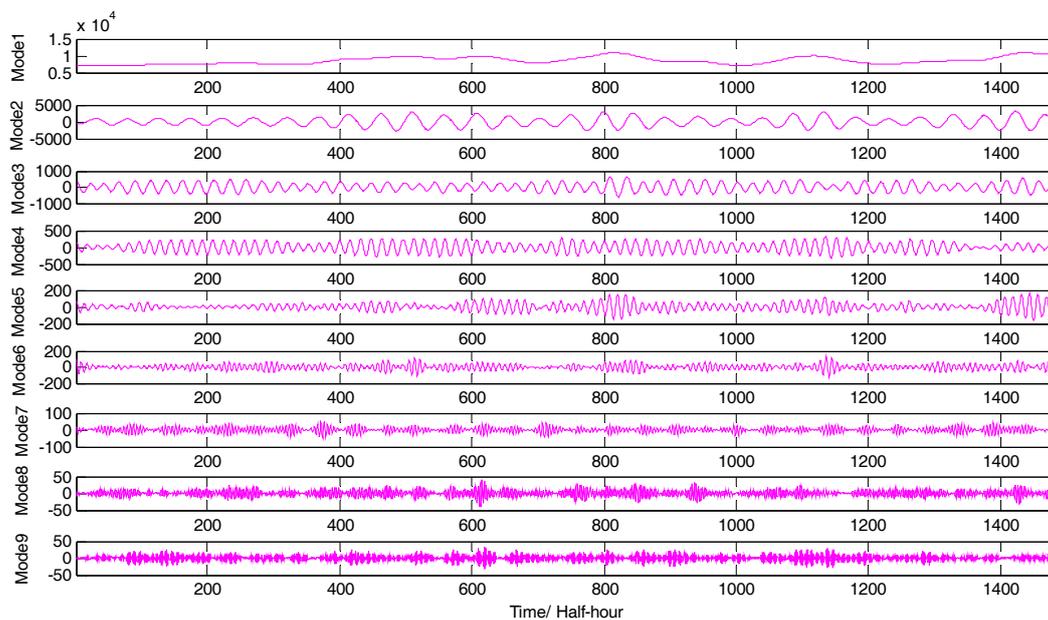


Figure 4. The decomposed series of electric load series in NSW.

Then, the forecasting issue turns to the forecasting of each mode using the ELM model improved by the DE algorithm. In this study, the inputs of DE-ELM model are determined using the rolling technique proposed in the reference [36], that is, every eight continuous electric load data points are employed to predict the ninth one, the detailed procedure is provided in Figure 3. The parameter settings of DE algorithm including iteration number (Gen), population size (Pop), mutation control parameter (F) and crossover probability (Cr) are all listed as follows: $Gen = 100$, $Pop = 30$, $F = 0.9$ and $Cr = 0.5$. The above input determination method of ELM model and parameter settings are utilized in all of the models adopted in this paper to ensure fair and effective comparisons among different models.

4.2.2. Forecasting Results, Comparative Analysis and Discussion

The effectiveness of the proposed forecasting is examined by comparing with the following two groups of benchmark models. The first group of models including ELM and DE-ELM models are adopted to confirm the influences of VMD and DE algorithm on the ELM model. The second group of models containing ARIMA, EMD-DE-ELM and WT-MABC-ELM models are selected to verify the reliability, practicability and portability of the developed model. The forecasting performances and the forecasting errors of all the above models are provided in Figure 5 and Table 2, respectively.

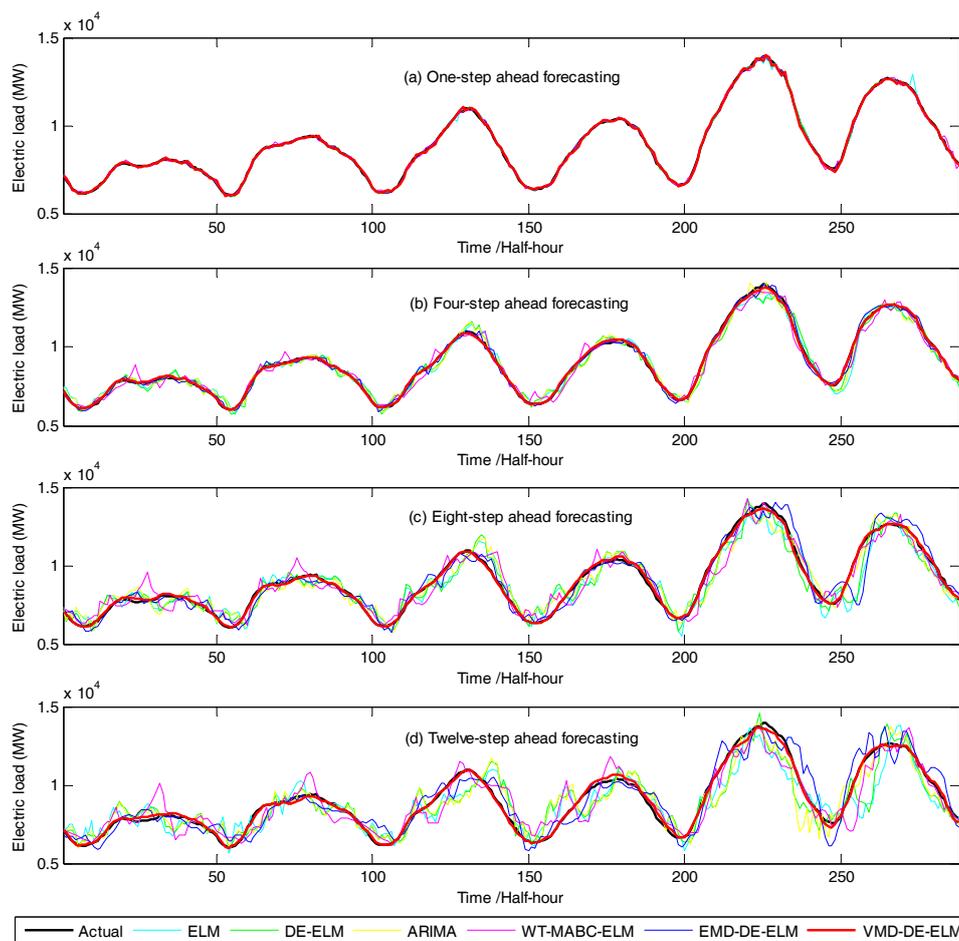


Figure 5. Forecasting results of different models in NSW.

In order to show the comparison more clearly, three histograms based on the MAE, RMSE and MAPE values of all the considered models in the paper are also presented in Figure 6. As shown in Table 2 and Figure 6, the MAE, RMSE and MAPE values of the proposed model are all smaller than the other comparison models, which confirms the superior forecasting performance of the proposed model for both one-step and multi-step ahead electric load forecasting.

Table 2. Forecasting errors of different models (NSW).

Prediction Horizon	Index	ELM	DE-ELM	ARIMA	WT-MABC-ELM	EMD-DE-ELM	VMD-DE-ELM
One-step ahead	MAE	74.477	66.385	64.283	51.991	33.652	26.809
	RMSE	111.127	83.838	83.691	78.671	47.185	34.212
	MAPE (%)	0.826	0.765	0.729	0.595	0.399	0.306
Four-step ahead	MAE	268.927	253.366	236.422	227.829	142.189	54.471
	RMSE	350.613	329.449	303.286	302.815	184.695	71.585
	MAPE (%)	3.090	2.876	2.731	2.588	1.616	0.590
Eight-step ahead	MAE	571.469	523.137	480.196	443.308	393.016	84.301
	RMSE	724.253	665.499	592.378	569.010	587.119	109.534
	MAPE (%)	6.546	5.946	5.549	5.147	4.364	0.918
Twelve-step ahead	MAE	753.415	718.724	684.807	633.498	626.593	116.900
	RMSE	975.893	918.407	880.956	831.308	797.494	152.374
	MAPE (%)	8.341	7.949	7.717	7.206	7.061	1.311

Note: The smallest value of each row is marked in boldface.

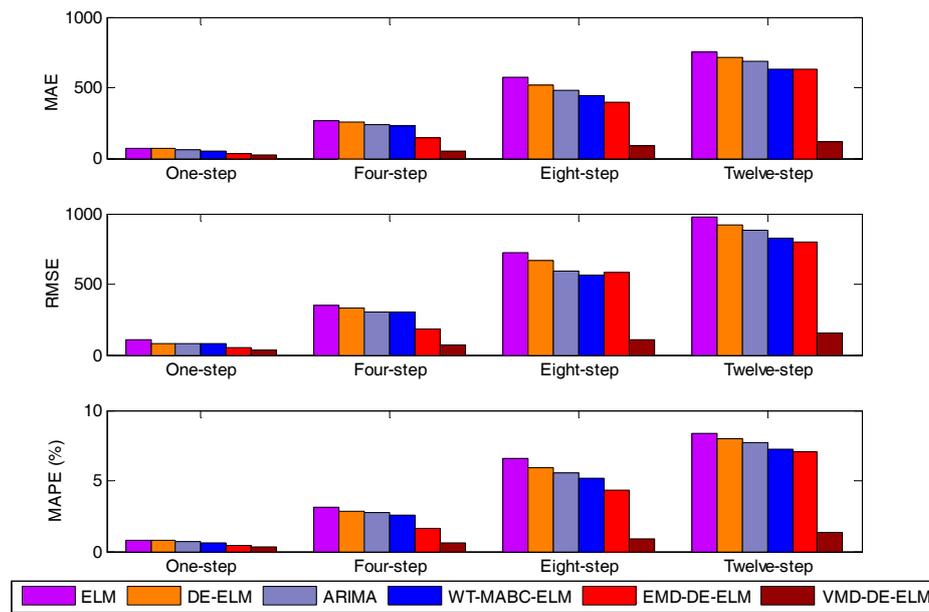


Figure 6. Comparison of different models in terms of MAE, RMSE and MAPE (NSW).

In order to well investigate why the proposed model performs better than other comparison models, this section conducts the following three comparative analysis: (1) VMD-DE-ELM and EMD-DE-ELM; (2) VMD-DE-ELM and DE-ELM; (3) ELM and DE-ELM. In order to more effectively compare the forecasting abilities of different models, the differences between the forecasting errors of different models are calculated according to the following formula: $DF = (e_2 - e_1) / e_2 \times 100\%$, where DF is the error difference; e_1 and e_2 are the MAE, RMSE or MAPE value of the first and second models. The results of the three above comparative analysis are all shown in Table 3. Based on the results shown in Table 3, the following three findings can be derived: (1) the proposed model owns the better forecasting accuracy than EMD-DE-ELM, which confirms the good performance of VMD against EMD; (2) the forecasting errors of ELM is significantly decreased by combining the VMD, which illustrates that the data preprocessing by VMD can effectively decrease the stochastic fluctuation characteristics existed in the electric load series, and consequently improves the whole forecasting accuracy, and; (3) the forecasting ability of ELM model is effectively improved through integrating the DE algorithm, which is used to optimize the initial weight and thresholds, thus it can be concluded that the DE algorithm has positive influences on the performance of ELM model.

Table 3. Differences between the forecasting errors of different models (NSW).

Prediction Horizon	Index (%)	DF				
		VMD-DE-ELM vs. ARIMA	VMD-DE-ELM vs. EMD-DE-ELM	VMD-DE-ELM vs. WT-MABC-ELM	VMD-DE-ELM vs. DE-ELM	DE-ELM vs. ELM
One-step ahead	MAE	58.30	20.33	48.44	59.62	10.87
	RMSE	59.12	27.49	56.51	59.19	24.56
	MAPE	58.02	23.31	48.57	60.00	7.38
Four-step ahead	MAE	76.96	61.69	76.09	78.50	5.79
	RMSE	76.40	61.24	76.36	78.27	6.04
	MAPE	78.40	63.49	77.20	79.49	6.93
Eight-step ahead	MAE	82.44	78.55	80.98	83.89	8.46
	RMSE	81.51	81.34	80.75	83.54	8.11
	MAPE	83.46	78.96	82.16	84.56	9.17
Twelve-step ahead	MAE	82.93	81.34	81.55	83.74	4.60
	RMSE	82.70	80.89	81.67	83.41	5.89
	MAPE	83.01	81.43	81.81	83.51	4.70

4.3. Multi-Step Ahead Electric Load Forecasting in QLD

In this section, the proposed model is tested using another real-world electric load series collected from QLD in order to verify its reliability and practicability. The forecasting results, including MAE, RMSE and MAPE values of all the models adopted in the paper, are provided in Tables 4 and 5. Based on the results shown in Tables 4 and 5, similar conclusions to that in the case of NSW are derived. It should be noted that because of different regional scales, population, geographical positions, climatic characteristics and industrial structures associated with NSW and QLD located in Australia, the two electric load series have significantly different complexities of irregularity, randomness and non-stationarity. In addition, the proposed VMD-DE-ELM model works only on the historical data series, consequently making different forecasting accuracies in the above two study cases. Nevertheless, all the experiment results in the two above cases can demonstrate the superiority of the proposed model.

Table 4. Forecasting errors of different models (QLD).

Prediction Horizon	Index	ELM	DE-ELM	ARIMA	WT-MABC-ELM	EMD-DE-ELM	VMD-DE-ELM
One-step ahead	MAE	54.319	52.142	47.230	36.690	26.595	24.257
	RMSE	72.347	67.914	62.768	45.993	34.996	30.673
	MAPE (%)	0.766	0.742	0.673	0.537	0.377	0.346
Four-step ahead	MAE	185.559	165.310	154.155	132.181	96.670	33.801
	RMSE	236.005	209.209	204.769	171.577	134.903	43.935
	MAPE (%)	2.620	2.345	2.178	1.884	1.395	0.476
Eight-step ahead	MAE	369.909	327.546	293.851	253.762	239.131	50.352
	RMSE	475.637	433.910	391.159	329.213	328.414	67.260
	MAPE (%)	5.281	4.688	4.151	3.579	3.454	0.703
Twelve-step ahead	MAE	506.554	490.289	480.470	419.567	394.999	78.343
	RMSE	643.787	629.534	627.282	559.602	513.964	99.880
	MAPE (%)	7.289	7.034	6.888	5.936	5.821	1.099

Note: The smallest value of each row is marked in boldface.

Table 5. Differences between the forecasting errors of different models (QLD).

Prediction Horizon	Index (%)	DF				
		VMD-DE-ELM vs. ARIMA	VMD-DE-ELM vs. EMD-DE-ELM	VMD-DE-ELM vs. WT-MABC-ELM	VMD-DE-ELM vs. DE-ELM	DE-ELM vs. ELM
One-step ahead	MAE	48.64	8.79	33.89	53.48	4.01
	RMSE	51.13	12.35	33.31	54.84	6.13
	MAPE	48.59	8.22	35.57	53.37	3.13
Four-step ahead	MAE	78.07	65.03	74.43	79.55	10.91
	RMSE	78.54	67.43	74.39	79.00	11.35
	MAPE	78.15	65.88	74.73	79.70	10.50
Eight-step ahead	MAE	82.86	78.94	80.16	84.63	11.45
	RMSE	82.80	79.52	79.57	84.50	8.77
	MAPE	83.06	79.65	80.36	85.00	11.23
Twelve-step ahead	MAE	83.69	80.17	81.33	84.02	3.21
	RMSE	84.08	80.57	82.15	84.13	2.21
	MAPE	84.04	81.12	81.49	84.38	3.50

5. Conclusions

In the deregulated electricity market, since the short-term electric load forecasting can provide lots of vital trading information for all of the market participants, thus it plays an increasingly important role for energy production, consumption and scheduling. However, various complicating influencing factors make the electric load series own severe stochastic fluctuation characteristics, which significantly increases the forecasting difficulties. This paper aims to establish an ensemble model based on VMD and ELM improved by the DE algorithm. In this proposed model, the electric load

series is firstly preprocessed by VMD into a number of components in order to effectively decrease the stochastic fluctuation characteristics, and then each component is forecasted over different horizons using the optimized ELM model. Two real-world electric load series collected from NSW and QLD are used to test the proposed forecasting model, and comparative analysis based on two groups of benchmark models are conducted in this paper. Based on the experimental results, the following three conclusions can be derived: (1) the proposed forecasting model owns the better forecasting ability than all other comparison models, which illustrates the superior performances of the proposed model for both one-step and multi-step ahead electric load forecasting; (2) the forecasting errors of ELM are significantly decreased by combining the VMD, which illustrates that the data preprocessing by VMD can effectively decrease the stochastic fluctuation characteristics existing in the electric load series, and consequently improves the whole forecasting accuracy; (3) the forecasting ability of ELM model is effectively improved through integrating the DE algorithm, which is used to optimize the initial weight and thresholds, thus it can be concluded that the DE algorithm has positive influences on the performance of the ELM model.

Acknowledgments: This research was supported by the National Natural Science Foundation of China (Grant Nos. 71473232, 71573237); the Science Foundation of Mineral Resource Strategy and Policy Research Center, China University of Geosciences (Grant No. H2017011B).

Author Contributions: Yanbing Lin and Hongyuan Luo designed the experiments and collected the data. Deyun Wang conducted the case studies and analyzed the results. Haixiang Guo and Kejun Zhu provided critical review and manuscript editing. Yanbing Lin and Hongyuan Luo wrote the paper. All authors read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

VMD	Variational mode decomposition
EMD	Empirical mode decomposition
EEMD	Ensemble empirical mode decomposition
FEEMD	Fast ensemble empirical mode decomposition
IMF	Intrinsic mode function
WPT	Wavelet packet transform
WT	Wavelet transform
ARMA	Auto-regressive moving average
ARIMA	Auto-regressive integrated moving average
GARCH	Generalized autoregressive conditional heteroskedasticity
VAR	Vector auto-regression
ANN	Artificial neural network
SVM	Support vector machine
LSSVM	Least square support vector machine
ELM	Extreme learning machine
DE	Differential evolution
PSO	Particle swarm optimization
SA	Simulated annealing
ABC	Artificial bee colony
MABC	Modified artificial bee colony
SOM	Self-organizing maps
PSR	Phase space reconstruction
RMSE	Root mean square error
MAE	Mean absolute error
MAPE	Mean absolute percentage error
NSW	New south wales
QLD	Queensland

References

1. Abosedra, S.; Dah, A.; Ghosh, S. Electricity consumption and economic growth, the case of Lebanon. *Appl. Energy* **2009**, *86*, 429–432. [[CrossRef](#)]
2. Adam, N.R.B.; Elahee, M.K.; Dauhoo, M.Z. Forecasting peak electricity demand in Mauritius using the non-homogeneous Gompertz diffusion process. *Energy* **2013**, *36*, 6763–6769. [[CrossRef](#)]
3. Kavaklioglu, K.; Ceylan, H.; Ozturk, H.K.; Canyurt, O.E. Modeling and prediction of Turkey's electricity consumption using Artificial Neural Networks. *Energy Convers. Manag.* **2009**, *50*, 2719–2727. [[CrossRef](#)]
4. Chase, R.B.; Jacobs, F.R.; Aquilano, N.J. *Operations Management for Competitive Advantage*; McGraw-Hill/Irwin: New York, NY, USA, 2001.
5. He, K.J.; Yu, L.A.; Tang, L. Electricity price forecasting with a BED (Bivariate EMD Denoising) methodology. *Energy* **2015**, *91*, 601–609. [[CrossRef](#)]
6. Lei, M.; Luan, S.J.; Jiang, C.W.; Liu, H.L.; Zhang, Y. A review on the forecasting of wind speed and generated power. *Renew. Sustain. Energy. Rev.* **2009**, *13*, 915–920. [[CrossRef](#)]
7. Pappas, S.S.; Ekonomou, L.; Karampelas, P.; Karamousantas, D.C.; Katsikas, S.K.; Chatzarakis, G.E.; Skafidas, P.D. Electricity demand load forecasting of the Hellenic power system using an ARMA model. *Electr. Power Syst. Res.* **2010**, *80*, 256–264. [[CrossRef](#)]
8. Kavousi-Fard, A.; Kavousi-Fard, F. A new hybrid correction method for short-term load forecasting based on ARIMA, SVR and CSA. *J. Exp. Theor. Artif. Intell.* **2013**, *25*, 559–574. [[CrossRef](#)]
9. Takiyar, S.; Upadhyay, K.G.; Singh, V. Fuzzy ARTMAP and GARCH-based hybrid model aided with wavelet transform for short-term electricity load forecasting. *Energy Sci. Eng.* **2016**, *4*, 14–22. [[CrossRef](#)]
10. Garcia-Ascanio, C.; Mate, C. Electric power demand forecasting using interval time series: A comparison between VAR and iMLP. *Energy Policy* **2010**, *38*, 715–725. [[CrossRef](#)]
11. Takeda, H.; Tamura, Y.; Sato, S. Using the ensemble Kalman filter for electricity load forecasting and analysis. *Energy* **2016**, *104*, 184–198. [[CrossRef](#)]
12. Yolcu, U.; Egrioglu, E.; Aladag, C.H. A new linear and nonlinear artificial neural network model for time series forecasting. *Decis. Support Syst.* **2013**, *54*, 1340–1347. [[CrossRef](#)]
13. Gutierrez-Corea, F.V.; Manso-Callejo, M.A.; Moreno-Regidor, M.P.; Manrique-Sancho, M.T. Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations. *Sol. Energy* **2016**, *134*, 119–131. [[CrossRef](#)]
14. Li, S.; Wang, P.; Goel, L. Short-term load forecasting by wavelet transform and evolutionary extreme learning machine. *Electr. Power Syst. Res.* **2015**, *122*, 96–103. [[CrossRef](#)]
15. Zhou, J.Y.; Jing, S.; Gong, L. Fine tuning support vector machines for short-term wind speed forecasting. *Energy Convers. Manag.* **2011**, *52*, 1990–1998. [[CrossRef](#)]
16. Chen, T.T.; Lee, S.J. A weighted LSSVM based learning system for time series forecasting. *Inf. Sci.* **2015**, *299*, 99–116. [[CrossRef](#)]
17. Liu, H.; Shi, J. Applying ARMA-GARCH approaches to forecasting short-term electricity prices. *Energy Econ.* **2013**, *37*, 152–166. [[CrossRef](#)]
18. Voyant, C.; Muselli, M.; Paoli, C.; Nivet, M. Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation. *Energy* **2012**, *39*, 341–355. [[CrossRef](#)]
19. Ismail, S.; Shabri, A.; Samsudin, R. A hybrid model of self-organizing maps (SOM) and least square support vector machine (LSSVM) for time-series forecasting. *Expert Syst. Appl.* **2011**, *38*, 10574–10578. [[CrossRef](#)]
20. Liu, D.; Niu, D.X.; Wang, H.; Fan, L.L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renew. Energy* **2014**, *62*, 592–597. [[CrossRef](#)]
21. Wang, J.Z.; Wang, Y.; Jiang, P. The study and application of a novel hybrid forecasting model—A case study of wind speed forecasting in China. *Appl. Energy* **2015**, *143*, 472–488. [[CrossRef](#)]
22. Zhang, Y.; Li, H.; Wang, Z.H.; Li, J. A preliminary study on time series forecast of fair-weather atmospheric electric field with WT-LSSVM method. *J. Electrostat.* **2015**, *75*, 85–89. [[CrossRef](#)]
23. Wang, D.Y.; Luo, H.Y.; Grunder, O.; Lin, Y.B.; Guo, H.H. Multi-step ahead electricity price forecasting using a hybrid model based on two-layer decomposition technique and BP neural network optimized by firefly algorithm. *Appl. Energy* **2017**, *190*, 390–407. [[CrossRef](#)]
24. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **2014**, *30*, 1030–1081. [[CrossRef](#)]

25. Cincotti, S.; Gallo, G.; Ponta, L.; Raberto, M. Modelling and forecasting of electricity spot-prices: Computational intelligence vs classical econometrics. *AI Commun.* **2014**, *27*, 301–314.
26. Amjady, N.; Keynia, F. Day ahead price forecasting of electricity markets by a mixed data model and hybrid forecast method. *Int. J. Electr. Power* **2008**, *30*, 533–546. [[CrossRef](#)]
27. Abdoos, A.A. A new intelligent method based on combination of VMD and ELM for short term wind power forecasting. *Neurocomputing* **2016**, *203*, 111–120. [[CrossRef](#)]
28. Lahmiri, S. A variational mode decomposition approach for analysis and forecasting of economic and financial time series. *Expert Syst. Appl.* **2016**, *55*, 268–273. [[CrossRef](#)]
29. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary Time Series Analysis. *Proc. R. Soc. Lond. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [[CrossRef](#)]
30. Huang, N.E.; Wu, M.C.; Qu, W.; Long, S.R.; Shen, Z. Application of Hilbert-Huang Transform to Non-stationary Financial Time Series Analysis. *Appl. Stoch. Model. Bus. Ind.* **2003**, *19*, 245–268. [[CrossRef](#)]
31. Dragomireskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [[CrossRef](#)]
32. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: A new learning scheme of feedforward neural networks. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks, Budapest, Hungary, 25–29 July 2004; Volume 2, pp. 985–990.
33. Storn, R.; Price, K. Differential evolution: A simple and efficient adaptive scheme for global optimization over continuous spaces. *J. Glob. Optim.* **1995**, *23*, 341–359.
34. Australian Energy Market Operator (AEMO). Available online: www.aemo.com.au (accessed on May 2017).
35. Taieb, S.B.; Bontempi, G.; Atiya, A.F.; Sorjamaa, A. A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition. *Expert Syst. Appl.* **2011**, *39*, 7067–7083. [[CrossRef](#)]
36. Wang, S.X.; Zhang, N.; Wu, L.; Wang, Y.M. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* **2016**, *94*, 629–636. [[CrossRef](#)]



© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).