Research and Application of a Hybrid Forecasting Model Based on Data Decomposition for Electrical Load Forecasting

Authors:

Yuqi Dong, Xuejiao Ma, Chenchen Ma, Jianzhou Wang

Abstract:

Accurate short-term electrical load forecasting plays a pivotal role in the national economy and people's livelihood through providing effective future plans and ensuring a reliable supply of sustainable electricity. Although considerable work has been done to select suitable models and optimize the model parameters to forecast the short-term electrical load, few models are built based on the characteristics of time series, which will have a great impact on the forecasting accuracy. For that reason, this paper proposes a hybrid model based on data decomposition considering periodicity, trend and randomness of the original electrical load time series data. Through preprocessing and analyzing the original time series, the generalized regression neural network optimized by genetic algorithm is used to forecast the short-term electrical load. The experimental results demonstrate that the proposed hybrid model can not only achieve a good fitting ability, but it can also approximate the actual values when dealing with non-linear time series data with periodicity, trend and randomness.

Research and Application of a Hybrid Forecasting Model Based on Data Decomposition for Electrical Load Forecasting

Yuqi Dong [1], Xuejiao Ma [2,*], Chenchen Ma [3] and Jianzhou Wang [2,3]

[1]   College of Law, Guangxi Normal University, Guilin 541004, China; 18074845560@163.com
[2]   School of Statistics, Dongbei University of Finance and Economics, Dalian 116023, China; wangjz@dufe.edu.cn
[3]   School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China; chenchenma133@163.com
*   Correspondence: xuejiaomadufe@163.com; Tel.: +86-186-2443-8689

Abstract: Accurate short-term electrical load forecasting plays a pivotal role in the national economy and people's livelihood through providing effective future plans and ensuring a reliable supply of sustainable electricity. Although considerable work has been done to select suitable models and optimize the model parameters to forecast the short-term electrical load, few models are built based on the characteristics of time series, which will have a great impact on the forecasting accuracy. For that reason, this paper proposes a hybrid model based on data decomposition considering periodicity, trend and randomness of the original electrical load time series data. Through preprocessing and analyzing the original time series, the generalized regression neural network optimized by genetic algorithm is used to forecast the short-term electrical load. The experimental results demonstrate that the proposed hybrid model can not only achieve a good fitting ability, but it can also approximate the actual values when dealing with non-linear time series data with periodicity, trend and randomness.

Keywords: electrical load forecasting; data decomposition; genetic algorithm; generalized regression neural network

1. Introduction

The electric power industry plays a pivotal role in the national security, social stability and all aspects of people's life. As is known to all, electricity, as one of the most important energy resources, is difficult to store. A great variety of instability factors can affect the electric system, such as emergencies, holidays, population changes, the weather and more [1]. Therefore, there is a high demand for the generation, transmission and sales of electricity, because excess supply can result in wasted energy resources and in case of excess demand the need for electricity cannot be satisfied. Therefore, performing load forecasting based on the historical data has been a basic task in the operation of electric systems [2]. With the rapid development of society and continuous improvement of economic levels, people have gradually shown a higher desire for electricity, which poses a huge challenge to the forecasting accuracy. A higher accuracy can improve the electric energy usage, enhance the safety and reliability of power grid and have a big impact on all sections in the electric power system. Accurate forecasting of electrical load plays a significant role, which can be reflected in the following aspects:

◆   Improve the social and economic benefits. The electrical power sector is supposed to ensure a good social benefit through providing safe and reliable electricity and improving the economic benefits considering the cost problems. Thus, the electrical load forecasting is beneficial for electrical power system to achieve the economic rationality of power dispatching.

♦ Ensure the reliability of electricity supply. Whether the power generation or supply, equipment needs periodical overhauls to ensure the safety and reliability of electricity. However, when to overhaul or replace the equipment should be based on accurate electrical load forecasting results.

♦ Plan for electrical power construction. The construction of electrical power production sites cannot stay unchanged, and should be adjusted and perfected, to satisfy the demands of a constantly changing future with the progress of society and development of the economy.

There are a great number of methods to forecast the electrical load, and in general the electrical load forecasting can be divided into three types, according to the applied field and forecasting time:

♦ Long-term electrical load forecasting. This means a time interval above five years and is usually conducted during the planning and building stage of the electrical system, which considers the characteristics of the electrical system and the development tendencies of the national economy and society;

♦ Middle-term electrical load forecasting. It is mainly applied in the operation stage of the electrical power system, for direction of the scientific dispatch of power, arrangement of overhauling and so on;

♦ Short-term electrical load forecasting. It plays a pivotal role in the whole electrical system and is the most important part, for it is the basis of long- and middle-term electrical load forecasting. Besides, it can ensure the stable and safe operation of the electrical power system based on the forecasting data.

Electrical load forecasting is a very complicated work. On the one hand, the electrical power system itself is complex and of large size. On the other hand, the electrical market closely combines the electrical power system with the whole society. Therefore, to properly monitor changes of the electrical load has become increasingly crucial for utilities so as to secure a steady power supply and make a suitable plans for investing in power facilities [3]. On the contrary, the inaccurate electrical load forecasting would be counterproductive. The overestimated future electrical load will result in an unnecessary generation of electrical power; while the underestimated forecasting would lead to trouble in offering sufficient electrical power, resulting in high losses for per peaking unit [4,5]. In addition, the inaccurate electrical load forecasting would also directly increase the operating costs. Therefore, to develop a better forecasting method and improve the forecasting ability has been more and more imperative, which is a both significant and challenging task [6].

In recent years, the study of short-term electrical load time series forecasting has mainly included four aspects, which are classic forecasting methods, modern forecasting methods, combined forecasting methods and hybrid forecasting methods [7].

The classic forecasting models refer to regression analysis, time series analysis and so on. The regression analysis models regard the influencing factors of time series as independent variables, and the historical data as the dependent variable, ensuring the relationship between the series and influencing factors. These methods are based on the analysis of historical data, so they can better model the history, however, as time goes by, the forecasting effect of regression analysis models will become weaker and weaker. The regression analysis process is easy, and the parameter estimation methods are complete; however, when dealing with non-linear time series data, the forecasting quality is bad and the forecasting accuracy is low. Another drawback is that it is difficult to select the influencing factors owing to the complexity of the objective data [8]. Time series forecasting aims to construct mathematical models based on the statistics of historical data, and it requires relatively small datasets and achieves a fast analysis speed, which can capture the variation trends of the recent data. However, it has a high requirement for stability, so when the influence of random factors is strong, the model will achieve a bad forecasting effect and low forecasting accuracy.

The modern forecasting methods include artificial intelligence neural networks [9,10], chaotic time series methods [11], expert system forecasting methods [12], grey models [13,14], support vector

machines [15,16], fuzzy systems [17], self-adaptable models [18], optimization algorithms and so on. The artificial neural networks (ANNs) can simulate the human brain to realize the intelligent dealing, and it can obtain a good forecasting performance when addressing the non-structural and non-linear time series data owing to their ability of self-adaptability, self-learning and memory. In 1991, Park [19] first applied ANNs in electrical load forecasting, proving the good performance of the model and at the same time concluding that ANNs were applicable in electrical load forecasting. Since then a large number of researchers have utilized many types of ANNs to forecast the time series [20–22]; however, ANNs also have its own limitations and disadvantages: (1) It is difficult to determine scientifically the number of layers and neurons of a network structure; (2) ANNs have a relatively slow self-learning convergence rate, which makes it easy to fall into a local minimum; (3) The ability to express the fuzzy awareness of human brain is not strong. Therefore, other methods, such as support vector machine (SVM) and evolution algorithms (EA), are used to overcome the dependence of ANNs on the samples, enhance the extrapolation power, and reduce the learning time. Pandian [23] and Pai [24] applied ANNs in electrical load forecasting systems. The optimization algorithms are enlightened by the biological evolution, which is effective in dealing with complicated problems. Optimization algorithms are usually combined with other forecasting methods, with the aim of selecting and recognizing parameters. For example, in the aspect of ANNs, optimization algorithms do not depend on subjective experience to determine parameters; instead, it can select more reasonable parameters through objective algorithms.

In view of the limitations and accuracy errors of single algorithms, they cannot be adapted to all situations; therefore, the combined models have gradually become the development tendency currently [25]. The combined forecasting models were initially proposed by Bates and Granger who proved that the linear combination of two forecasting models could obtain better forecasting results than the single models alone. Xiao et al. [26] and Wang et al. [27] also proved that the forecasting accuracy of the combined model were higher than that of a single model. The basic principles of the combined forecasting methods are to integrate the forecasting output results of different single models based on certain weights, narrowing the value range of the forecasting down to a smaller scale. A problem is supposed to be studied from different angles instead of a single angle, and this is why the combined forecasting model is needed. The information obtained from each single forecasting method is not the same, and a weight is necessary to express the outputs of each single model more comprehensively in order to retain the original valuable information. Recently the combined forecasting models have been commonly used to solve forecasting issues, but how to select the single model properly and distribute the weight reasonably is a challenging task.

The theory of hybrid algorithms can get over the shortcomings of the single forecasting model through integrating two or more than two single models. As discussed above, the single models have their own advantages and disadvantages when dealing with different forecasting problems. In comparison, the hybrid forecasting methods can increase the forecasting accuracy through determining an optimal combination and putting the advantages of single models into full play. In other words, the hybrid algorithms can integrate many different forecasting techniques to solve practical problems in practice. For example, the blind number theory can be applied in middle- and long-term electrical load forecasting to build a hybrid model, which can enhance the forecasting effects well due to the irregular nature of electrical load time series.

Affected by many factors, the complexity of time series continues improving, and several techniques are utilized to solve the forecasting problems of time series. Azimi et al. [28] built a novel hybrid model to forecast the short-term electrical load, because a single model cannot figure out the characteristics of the time series data. Khashei and Bijari [29] considered that there was no a single model that could ensure the real process of the data generation. Shukur and Lee [30] proposed a hybrid model, including ANN and auto regressive integrated moving average (ARIMA), taking full advantage of the linear and non-linear advantages of the two models. Considerable experimental results demonstrate that the forecasting accuracy of the hybrid model represents a great improvement

when compared with other single models. Aiming to improve the forecasting quality, Niu [31] built a new hybrid ANN model and combined some statistical methods to conduct forecasting. Lu and Wang [32] developed a growing hierarchical self-organizing map (SOM) with support vector machine (SVM) to forecast the product demand. Okumus and Dinler [33] integrated the adaptive neuro-fuzzy inference system and ANNs to forecast the wind power and their experimental results proved that the proposed hybrid model was better than applying the single model. Che and Wang [34] put forward the SVMARIMA hybrid model with SVM and ARIMA to forecast both the linear and non-linear trends more accurately. Meng et al. [35] developed a hybrid model for short-term wind speed forecasting by applying wavelet packet decomposition, crisscross optimization algorithm and artificial neural networks, and their experimental results showed that the proposed hybrid model had the minimum mean absolute percentage error, regardless of whether one-step, three-step or five-step prediction was used. Elvira [36] selected five forecasting methods to forecast the electrical load in summer and winter in the southeastern region of Oklahoma respectively. The empirical results showed that there was no one model that could always perform the best in all conditions, and differences in the original time series data and the evaluation metrics used to measure errors would both have an impact on the selection of the optimal model. Wu et al. [37] proposed a hybrid forecasting method based on seasonal index adjustment, and applied it in the forecasting of short-term wind speed and electrical load. The experimental results indicated that compared with the method without seasonal index adjustment, the proposed hybrid model could achieve a better forecasting result.

As discussed above, the single modela cannot satisfy the requirementa for forecasting accuracy in practice, and there is no one model applicable in any situation. Given that the actual data will be affected by various factors, which are difficult to recognize and measure, and it is not possible to take every related factor into consideration, the model is supposed to be built based on some key factors that can be extracted. The establishment of the hybrid model has become the mainstream currently. Therefore, this paper proposes a hybrid forecasting model considering periodicity, trend and randomness for electrical load time series. The contributions of the model are summarized as follows:

(1) The time series data have the characteristics of continuity, periodicity, trend and randomness, and considerable work has been done to select suitable models and the optimize the model parameters; however, few studies focus on building forecasting models based on the characteristics of the time series data. Therefore, the initial contribution of this paper is to decompose the time series data. Based on the traditional additive model, the layer-upon-layer decomposition and reconstitution method is applied to improve the forecasting accuracy. Then according to the data features after decomposition, suitable models could be found to perform the forecasting. Through effective decomposition of the data and selection of reasonable model, the forecasting quality and accuracy could be improved to a great degree.

(2) This paper uses the generalized regression neural network (GRNN) to improve the forecasting performance. The data after decomposition have noises, so the empirical mode decomposition (EMD) is applied to reduce the noise in the data. Then the genetic algorithm (GA) is utilized to optimize the GRNN to conduct the forecasting to enhance the forecasting accuracy of the single model.

(3) The practical application of the proposed hybrid model in this paper is to forecast the short-term electrical load in New South Wales of Australia, and compare it with the single models and models without decomposition. The forecasting results demonstrate that the proposed model has a strong non-linear fitting ability and good forecasting quality for electrical load time series. Both the simulation results and the forecasting process could fully show that the hybrid model based on the data decomposition has the features of small errors and fast speed. The algorithm applied in the electrical power system is not only applicable, but also effective.

The rest of this paper is organized as follows: Section 2 describes the method and Section 3 introduces the detailed steps of the hybrid model, respectively. The experimental results are shown in Section 4. Section 5 presents the conclusions.

2. Methods

Conducting an accurate electrical load forecasting needs better developed forecasting methods and it is imperative to have improved forecasting abilities. This paper proposes a hybrid model to perform short-term electrical load forecasting, and this part introduces the fundamental methods, including additive model of time series, moving average model, cycle adjustment model, empirical mode decomposition and generalized regression neural network.

2.1. Additive and Multiplicative Model of Time Series

In general, a time series can be decomposed into two types of models through data transformation, including the additive model and the multiplicative model, as shown in Equations (1) and (2):

$$Y_t = S_t + T_t + C_t + R_t \tag{1}$$

$$Y_t = S_t \times T_t \times C_t \times R_t \tag{2}$$

where S_t is a seasonal item, indicating the law of transformation of time series with the season, which exists objectively. Actually, the electrical load time series always shows a seasonal cycle fluctuation; that is to say, the sequence will change repeatedly and continuously with time, showing a periodicity rule. Therefore, this paper classifies the seasonal item into a periodic item considering the clarity of expression. T_t is a trend item, denoting the law of transformation of time series with the trend. It mainly represents a long-term changing rule, because the time series will keep increasing, decreasing or remain stable. C_t is a periodic item and it indicates a periodic and non-seasonal law of transformation of time series with time. The number of a cycle fluctuation periods is expressed as h. R_t is a random item, which indicates the random change. Through decomposition, the original time series could be transformed into a stationary time series, which could achieve a good fitting and forecasting result.

2.2. Moving Average Model

The original time series will show the features of continuity, periodicity, trend and randomness. In order to eliminate the features and obtain a smoother time series, the moving average model will be applied. The algorithm principle is to calculate the average of the historical data, and the average is regarded as the next forecasting value until the final forecasting goal is realized. In other words, a new value will replace the old value, among which the number of items of the moving average is fixed. The detailed calculation equation is described as follows:

$$M_t^{(1)} = \frac{y_t + y_{t-1} + \cdots y_{t-N-1}}{N} = M_{t-1}^{(1)} + \frac{y_t - y_{t-N}}{N}, t \geq N \tag{3}$$

where $X = \{y_1, y_2, \cdots y_t\}$ is the original time series, N is the number of average, $M_t^{(1)}$ is the moving average in the t-th period, y_t is the observed value in the t-th period and N is the number of fixed items. The forecasting equation is:

$$\hat{y}_{t+1} = M_t^{(1)} \tag{4}$$

2.3. Periodic Adjustment Model

The essence of the cycle adjustment is to summarize the cycle variation law based on the periodic historical data. Assume that a group of periodic data $\{c_t, t \in \{1, 2, \cdots T\}\}$, it is divided into l groups and the number of data in each group is h ($T = l \times h$). The data series can be defined as:

Definition 1. *The time series data $\{c_t, t \in \{1, 2, \cdots T\}\}$ is decomposed into $\{c_{11}, c_{12}, \ldots, c_{1s}, \ldots, c_{1h}\}$, $\{c_{21}, c_{22}, \ldots, c_{2s}, \ldots, c_{2h}\}$, $\{c_{k1}, c_{k2}, \ldots, c_{ks}, \ldots, c_{km}\}, \ldots, \{c_{l1}, c_{l2}, \ldots, c_{ls}, \ldots, c_{lh}\}$ ($k = 1, 2, \ldots, l$; $s = 1, 2, \ldots, h$). c_{ks} means s-th data in the k-th period.*

The average of each group can be used to approximate the periodic average [38]. The *s*-th average period is:

$$\bar{c}_s = (c_{1s} + c_{2s} + \cdots c_{ls})/l \quad (s = 1, 2, \ldots h) \tag{5}$$

The average of all data is:

$$Z = (\bar{c}_1 + \bar{c}_2 + \cdots \bar{c}_h)/h \tag{6}$$

The periodic value after adjustment is:

$$\hat{c}_s = \bar{c}_s - Z(s = 1, 2, \ldots, h) \tag{7}$$

Equations (5)–(7) represent the periodic variation law.

2.4. Empirical Mode Decomposition

The empirical mode decomposition, initially proposed in 1998, belongs to the data mining methods, which play a crucial role in dealing with the non-linear data Currently, it has been applied in many fields, such as geography [39], economics [40] and so on. EMD is a type of new method to divide the same non-stationary into different frequencies. The sequence of the composed different signal scales is called intrinsic mode function (IMF), which is the non-linear and stationary signal. IMF has an obvious feature that the wave amplitude changes with time. For given signal $x(t) \in R^t$, the detailed steps of EMD are described as follows (as shown in Figure 1I):

Step 1. Find all the local extreme points of $x(t)$.

Step 2. For all local extreme points of $x(t)$, build the envelope function of the signal, respectively, which can be denoted as $e_{max}(t)$ and $e_{min}(t)$.

Step 3. Calculate the average of the envelope function:

$$e_m(t) = \frac{e_{min}(t) + e_{max}(t)}{2} \tag{8}$$

Step 4. Calculate the differential function between signal $x(t)$ and the envelope average function

$$h(t) = x(t) - e_m(t) \tag{9}$$

Step 5. Replace the original signal $x(t)$ with $h(t)$, and repeat above steps from Step 2 to Step 4 until all averages of envelope function tends to zero. In this way an IMF $c_1(t)$ is decomposed.

Step 6. $c_1(t)$ represents the component with the highest frequency, so the low frequency of the original signal is $r_1(t)$:

$$r_1(t) = x(t) - c_1(t) \tag{10}$$

$$r_2(t) = r_1(t) - c_2(t) \tag{11}$$

$$r_n(t) = r_{n-1}(t) - c_n(t) \tag{12}$$

Step 7. For $x_1(t)$, repeat Step 2, Step 3 and Step 4, and the second IMF $c_2(t)$ can be obtained until the differential function $r_n(t)$ is a constant function or monotone function. Finally, the original signal $x(t)$ can be represented by IMF $c_j(t), j = 1, 2, \cdots, n$ and $r_n(t)$ as shown in Equation (13):

$$x(t) = \sum_{j=1}^{n} c_j(t) + r_n(t) \tag{13}$$

The EMD steps of the time series are shown in Figure 1I, and the pseudo code of EMD is described in Algorithm 1 below.
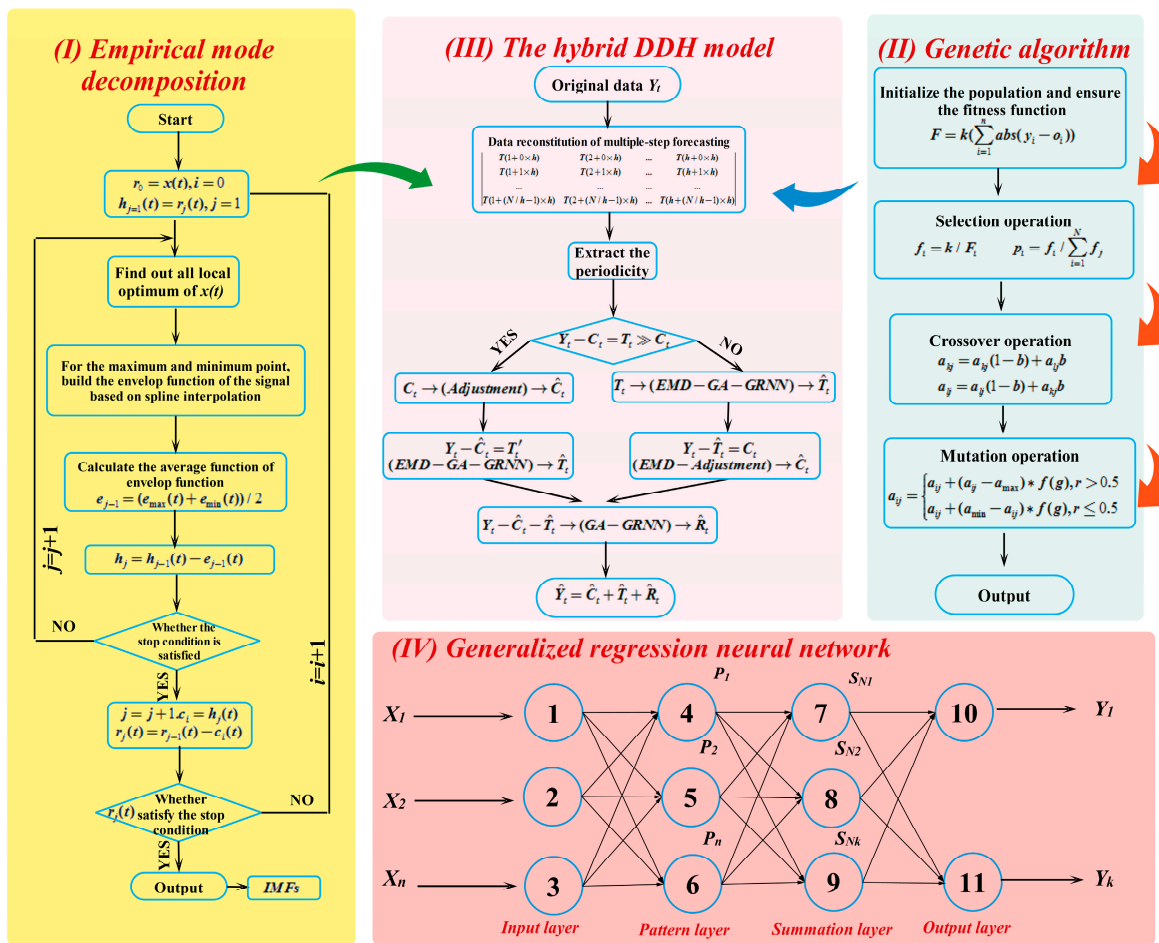
Figure 1. Steps of the main methods and proposed hybrid model in this paper.

Algorithm 1: Pseudo code of Empirical Mode Decomposition

Input: $x_s^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$—a sequence of sample data.

Output: $\hat{x}_s^{(0)} = (\hat{x}^{(0)}(l+1), \hat{x}^{(0)}(l+2), \ldots, \hat{x}^{(0)}(l+n))$—a sequence of denoising data.

Parameters:

　　δ—represent a random number in the algorithm with the value between 0.2 and 0.3.

　　T—a parameter describing the length of the original *electrical load* time series data.

1: /***Initialize** residue $r_0(t) = x(t), i = 1, j = 0$; Extract local maxima and minima of $r_{i-1}(t)$.*/

2: FOR EACH ($j = j + 1$) **DO**

3: FOR EACH ($i = 1 : n$) **DO**

4: WHILE (Stopping Criterion $SD_j = \sum\limits_{t=0}^{T} \dfrac{\left| h_{i,j-1}(t) - h_{i,j}(t) \right|^2}{[h_{i,j-1}(t)]^2} > \delta$) **DO**

5: Calculate the upper envelope $U_i(t)$ and $L_i(t)$ via cubic spline interpolation.

6: $m_i(t) = \dfrac{U_i(t) + L_i(t)}{2}$ /* Mean envelope */; $h_i(t) = r_{i-1}(t) - m_i(t)$ /* ith component */

7: /***Let** $h_{i,j}(t) = h_i(t)$, with $m_{i,j}(t)$ being the mean envelope of $h_{i,j}(t)$*/

8: END WHILE

9: Calculate $h_{i,j}(t) = h_{i,j-1}(t) - m_{i,j-1}(t)$

10: /***Let** the jth IMF be $IMF_i(t) = h_{i,j}(t)$; Update the residue $r_i(t) = r_{i-1}(t) - IMF_i(t)$*/

11: END DO

12: END DO

13: Return $x(t) = \sum\limits_{j=1}^{n} c_j(t) + r_n(t)$ /* The noise reduction process is finished */

2.5. Generalized Regression Neural Network (GRNN)

The generalized regression neural network, first proposed by Specht in 1991, is a type of radial basis function neural network (RBF). The theory of GRNN is based on non-linear regression analysis, and in essence, the purpose of GRNN is to calculate y with the biggest probability value based on the regression analysis of dependent variable Y and independent variable x. Assume that joint probability density function of the random variable x and y is $f(x, y)$, and the observed value x is known as X, so the regression of y about x is:

$$\hat{Y} = E(y/X) = \frac{\int_{-\infty}^{\infty} y f(X, y) dy}{\int_{-\infty}^{\infty} f(X, y) dy} \tag{14}$$

The density function $f(X, y)$ can be estimated from the sample data set $\{x_i, y_i\}_{i=1}^{n}$ by applying Parzen non-parametric estimation:

$$\hat{f}(X, y) = \frac{1}{n(2\pi)^{\frac{p+1}{2}} \sigma^{p+1}} \sum_{i=1}^{n} \exp[-\frac{(X - X_i)^T(X - X_i)}{2\sigma^2}] \exp[-\frac{(X - Y_i)^2}{2\sigma^2}] \tag{15}$$

where X_i and Y_i is the sample observed value of x and y, n is the sample size, p is the number of dimension of random variable x and σ is the smoothing factor. $\hat{f}(X, y)$ can replace $f(X, y)$ of Equation (15), so the function after transformation is:

$$\hat{Y}(X) = \frac{\sum\limits_{i=1}^{n} \exp[-\frac{(X - X_i)^T(X - X_i)}{2\sigma^2}] \int_{-\infty}^{\infty} y \exp[-\frac{(Y - Y_i)^2}{2\sigma^2}] dy}{n \sum\limits_{i=1}^{n} \exp[-\frac{(X - X_i)^T(X - X_i)}{2\sigma^2}] \int_{-\infty}^{\infty} \exp[-\frac{(Y - Y_i)^2}{2\sigma^2}] dy} \tag{16}$$

For $\int_{-\infty}^{\infty} z e^{-z^2} dz = 0$, after calculating the two integration, the output of GRNN can be $\hat{Y}(X)$ obtained as follows:

$$\hat{Y}(X) = \frac{\sum\limits_{i=1}^{n} y \exp[-\frac{(X - X_i)^T(X - X_i)}{2\sigma^2}]}{\sum\limits_{i=1}^{n} \exp[-\frac{(X - X_i)^T(X - X_i)}{2\sigma^2}]} \tag{17}$$

After obtaining the training samples of GRNN, the training process of the network involves optimizing the smoothing parameter σ. In order to improve the fitting ability of GRNN, σ needs to be optimized, which indicates the importance of optimizing the smoothing parameter σ in GRNN.

As for the structure of GRNN, it is similar to that of RBF, including input layer, pattern layer, summation layer and output layer. The corresponding network input is $X = [x_1, x_2, \ldots, x_n]$, and its output is $Y = [y_1, y_2, \ldots, y_n]^T$, which are described below.

(1) Input layer

The number of neuron of the input layer is the same as the dimension number of input variable, which plays a role in transferring signals.

(2) Pattern layer

The number of neuron of the pattern layer is the same as the number of learning samples, and the transfer function is

$$P_i = \exp[-\frac{(X - X_i)^T(X - X_i)}{2\sigma^2}], i = 1, 2, \ldots, n \tag{18}$$

where X is the input variable of the network, and X_i is the learning sample of ith neuron.

(3) Summation layer

Two methods can be applied to calculate the neuron. One is shown in Equation (10):

$$\sum_{i=1}^{n} \exp\left[-\frac{(X-X_i)^T(X-X_i)}{2\sigma^2}\right] \tag{19}$$

where the arithmetic sum of each neuron is calculated, the link weight is 1, and the transfer function is:

$$S_D = \sum_{i=1}^{n} P_i \tag{20}$$

The other method is:

$$\sum_{i=1}^{n} Y_i \exp\left[-\frac{(X-X_i)^T(X-X_i)}{2\sigma^2}\right] \tag{21}$$

where the weighted arithmetic sum of each neuron is calculated, and the link weight between the *i*-th neuron and *j*-th molecular sum neurons is the *j*-th element of *i*-th output sample Y_j. The transfer function is:

$$S_{Nj} = \sum_{i=1}^{n} y_{ij} P_i j, j = 1, 2, \ldots, k \tag{22}$$

(4) Output layer

The number of neuron of output layer is the same as the dimension number *k* of output variable. The output of summation layer is divided by each neuron as shown in Equation (23):

$$y_j = \frac{S_{Nj}}{S_D} \tag{23}$$

Then there are some weights in GRNN to connect different layers, and the least mean squares and differential chain rule are applied to adjust them. Initially, we define the least mean square of each neuron in the output layer:

$$E_k = [d_k(X) - F_k(W, X)]^2 / 2, k = 1, 2, \ldots, K \tag{24}$$

where $d_k(X)$ is the expected output, $F_k(W, X)$ is the actual output. E_k can arrive at the smallest value through adjusting the weights according to Equation (25) by using the least mean squares method:

$$\Delta w_{ki}(n) = \eta_k \left(-\frac{\partial E_k}{\partial w_{ki}}\right), i = 1, 2, \ldots, M; k = 1, 2, \ldots, K \tag{25}$$

where η_k is the learning rate. Therefore, the key to realizing the least square mean is to solve $(-\partial E_k / \partial w_{ki})$, so by using the differential chain rule, we can get:

$$-\frac{\partial E_k}{\partial w_{ki}} = -\frac{\partial E_k}{\partial F_k(W, X)} \frac{\partial F_k(W, X)}{\partial w_{ki}} \tag{26}$$

where $-\partial E_k / \partial F_k(W, X) = d_k(X) - F_k(W, X)$, which can be denoted as δ_k. Then we can get $(-\partial E_k / \partial w_{ki}) = \delta_k y_i$ according to Equation (27):

$$\frac{\partial F_k(W, X)}{\partial w_{ki}} = \frac{\partial}{\partial w_{ki}} \left(\sum_{i=1}^{M} w_{ki} y_{ki}\right) = y_i \tag{27}$$

so $\Delta w_{ki}(n) = \eta_k \delta_k y_i$, where y_i is the output of *i*-th neuron in the hidden layer, and the input of *k*th neuron in the output layer. The detailed structure of GRNN is described in Figure 1IV.

3. The Proposed Hybrid Model

In the proposed data decomposition hybrid model (DDH), we initially remove the periodicity in the original series, and then the EMD-GA-GRNN is applied to forecast the electrical load time series without periodicity. After that the periodicity is added to the forecasted time series by using the additive model. This part will introduce the basic ideas of both DDH and EMD-GA-GRNN.

3.1. Genetic Algorithm

The genetic algorithm is based on the natural selection rule and biological evolution principle, and its basic idea is to generate a set of initial solutions (population) in the problem space. Each group of solutions is regarded as the individuals in the population, which is defined as a chromosome. In the searching process, the adaptive value of chromosomes is the standard used to evaluate and select individuals. In the next generation, new individuals are generated through crossover and mutation operations, becoming a new generation of the population [41]. The above steps are repeated so that the chromosome can converge to a desired optimum value and solution. GA is applied in this paper to optimize GRNN, and the detailed steps are described as follows (as shown in the pseudo code of Algorithm 2 and Figure 1II):

Step 1. Initialize the population. Each individual in the population is a real number, with a known net structure, the initial values can form a neural network with structure, weight value and threshold value.

Step 2. Ensure the fitness function. The fitness value *F* is the absolute error values between the forecasting output and expected output calculated by Equation (28):

$$F = k(\sum_{i=1}^{n} abs(y_i - o_i)) \tag{28}$$

where *n* is the number of the output node of the network, y_i is the expected output of *i*th node, o_i is the forecasting output of *i*th node, and *k* is the coefficient.

Step 3. Selection operation. This operation is based on the proportion of the fitness, and the selection probability of each individual *i* is *p*:

$$f_i = \frac{k}{F_i} \tag{29}$$

$$p_i = \frac{f_i}{\sum\limits_{i=1}^{n} f} \tag{30}$$

where F_i is the fitness of individual *i*, and the smaller fitness is better. Before the selection operation, the reciprocal of fitness should be calculated. *k* is the coefficient and *N* is the number of individual in the population.

Step 4. Crossover operation. The individual is coded by using the real number, and the crossover operation in the *j*th position between *k*th chromosome a_k and a_l *l*th chromosome a$_l$:

$$a_{kj} = a_{kj}(1 - b) + a_{lj}b \tag{31}$$

$$a_{lj} = a_{lj}(1 - b) + a_{kj}b \tag{32}$$

where *b* is a random number of [0,1].

Step 5. Mutation operation. Select the *j*-th gene of *i*-th individual to conduct the mutation operation, and the method is:

$$a_{kj} = a_{kj}(1-b) + a_{lj}b$$

$$a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{max}) * f(g), r > 0.5 \\ a_{ij} + (a_{min} - a_{ij}) * f(g), r \leq 0.5 \end{cases} \quad (33)$$

where a_{max} is the upper bound of gene a_{ij}, a_{min} is the lower bound of gene a_{ij}, $f(g) = r_2(1 - g/G_{max})^2$, r_2 is a random number, g is the current iteration number, G_{max} is the maximum iteration number and r is a random of [0,1].

Algorithm 2: Pseudo Code of the genetic algorithm

Input: $x_s^{(0)} = (x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(n))$—a sequence of training data

$\hat{x}_s^{(0)} = (\hat{x}^{(0)}(l+1), \hat{x}^{(0)}(l+2), \ldots, \hat{x}^{(0)}(l+n))$—a sequence of verifying data

Output: ***fitness_value x_b***—the value with the best fitness value in the population of populations

Parameters:

Gen_{max}—the maximum number of iterations; *n*—the number of individuals

F_i—the fitness function of the individual *i*; *x_i*—the population *i*

g—the current iteration number of GA; *d*—the number of dimension

1: /***Initialize** the population of *n* individuals which are $x_i \backslash (i = 1, 2, ..., n)$ randomly.*/

2: /***Initialize** the parameters of **GA**: Initial probabilities of crossover p_c and mutation p_m.*/

3: **FOR EACH (***i*: $1 \leq i \leq n$**) DO**

4: **Evaluat**e the corresponding fitness function **F_i** $fitness\_popu(best(idx, 1), 1)$

5: **END FOR**

6: **WHILE (***g* < Gen_{max}**) DO FOR EACH (***I* = 1:*n*) **DO**

7: **IF (***p_c* > *rand***) THEN**

8: /***Conduct** the crossover operation*/ $a_{kj} = a_{kj}(1-b) + a_{lj}b$ and $a_{lj} = a_{lj}(1-b) + a_{kj}b$

9: **END IF**

10: **IF (***p_m* > *rand***) THEN**

11: /***Conduct the** Mutate operation*/ $a_{ij} = \begin{cases} a_{ij} + (a_{ij} - a_{max}) * f(g), r > 0.5 \\ a_{ij} + (a_{min} - a_{ij}) * f(g), r \leq 0.5 \end{cases}$

12: **END IF END FOR**

13: **FOR EACH (***i*: $1 \leq i \leq n$**) DO**

14: **Evaluate** the corresponding fitness function F_i $fitness\_popu(best(idx, 1), 1)$

15: **END FOR**

16: /***Update** the best nest *x_p* of the *d* generation in the genetic algorithm.*/

17: **FOR EACH (***i*: $1 \leq i \leq n$**) DO IF (***F_p*< *F_b***) THEN**

18: /* The global best solution can be obtained to replace the local optimal $x_b \leftarrow x_p$*/

19: **END IF END FOR END WHILE**

20: **RETURN** x_b/* The optimal solution in the global space has been obtained.*/

3.2. Data Decomposition Hybrid (DDH) Model

The time series always changes as time goes by, and such change has the features of continuity, periodicity, trend, and a certain randomness. In the previous research, no matter which models, including single model, combined model or hybrid model, they are all applied in forecasting the whole time series. Unlike the previous research, this paper proposes a data decomposition hybrid model (DDH) based on the periodicity, trend and randomness in the time series. The basic idea of DDH is to decompose the times series based on the main influencing factors. On the basis of decomposition and recombination of traditional additive model, the layer-upon-layer decreasing is applied to improve the

forecasting accuracy. Then suitable models are selected to conduct the forecasting according to the data characteristics and features. The effective decomposition of data and proper forecasting models for each part can enhance the fitting performance of the model and decrease the forecasting errors to a great degree compared with conventional single forecasting methods. The detailed steps of DDH are described below (as shown in Figure 1III):

Step 1. Observe whether the time series Y_t contains trend, periodicity and randomness, and judge the applicability of the additive model and multiplicity model. In general, compared to the additive model, the multiplicity model is more suitable for time series with large fluctuations [42]. The electrical load time series have a relatively stable fluctuation range; therefore, the additive model is chosen, and the following discussion is based on it.

Step 2. Apply the moving average method or other methods to extract the periodicity C_t.

Step 3. Without the periodicity C_t, the rest of the data can be defined as trend T_t. If T_t is far larger than C_t, a periodic adjustment of C_t should be conducted to obtain the estimated periodicity $\hat{C}_t$, and this is because if we firstly forecast larger data, there will be much noise in the latter data, which will affect the forecasting accuracy. Then the new trend T_t' can be obtained ($T'_t = Y_t - \hat{C}_t$). Finally, EMD-GA-GRNN can be utilized to forecast T_t', and the forecasting value is $\hat{T}_t$. On the contrary, if C_t is far larger than T_t, EMD-GA-GRNN is used to forecast the trend T_t, and get the forecasting value $\hat{T}_t$. Then the periodicity data C_t can be obtained. Finally, the estimated value $\hat{C}_t$ is obtained through the periodic adjustment.

Step 4. The original randomness R_t is calculated ($R_t = Y_t - \hat{C}_t - \hat{T}_t$). We forecast the randomness after decomposition by applying GA-GRNN to get the forecasting value $\hat{R}_t$. The randomness after decomposition is nearly stable, so EMD is unnecessary.

Step 5. Utilize the additive model to get the final forecasting values of the time series: $\hat{Y}_t = \hat{C}_t + \hat{T}_t + \hat{R}_t$.

3.3. The EMD-GA-GRNN Forecasting Model

In the model of DDH, EMD-GA-GRNN is proposed, which is based on the data state after applying the layer-upon-layer decreasing method. However, data after the layer-upon-layer decreasing method may include some noise due to the forecasting accuracy in the former forecasting methods. Thus, it is pivotal to apply a proper method to remove the noise in the decomposed data. This paper chooses the empirical mode decomposition method considering its advantages in dealing with non-linear time series data. Then the GRNN is utilized to forecast the dealt data, because it performs well in fitting non-stationary data. The training process of GRNN is actually to ensure the optimum *s*, and the specific steps of the hybrid model EMD-GA-GRNN are listed as follows (Pseudo code of Algorithm 3):

Algorithm 3: Pseudo code of the hybrid model of EMD-GA-GRNN

Input: $x_s^{(0)} = \left(x^{(0)}(1), x^{(0)}(2), \ldots, x^{(0)}(q) \right)$—a sequence of training data

　　　　$x_p^{(0)} = \left(x^{(0)}(q+1), x^{(0)}(q+2), \ldots, x^{(0)}(q+d) \right)$—a sequence of verifying data

Output: $\hat{y}_z^{(0)} = \left(\hat{y}^{(0)}(q+1), \hat{y}^{(0)}(q+2), \ldots, \hat{y}^{(0)}(q+d) \right)$—forecasting electrical load from GRNN

Fitness function: $fitness = 1/\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{K} \left(Y_j(i) - \overline{Y}_j(i) \right)^2$ /*The objective fitness function*/

Parameters:

　　Gen_{max}—the maximum number of iterations; *n*—the number of individuals

　　F_i—the fitness function of individual *i*; x_i—the total population *i*

　　G—the current iteration number; *d*—the number of dimension

1: /* **Process** original electrical load time series data with the noise reduction method EMD */

2: /***Initialize** the population of *n* individuals x_i (*i* = 1, 2, ..., *n*) randomly.*/

3: /*Initialize the original parameters: Initial probabilities of crossover p_c and mutation p_m.*/

Algorithm 3: *Cont.*

4: FOR EACH (*i*: $1 \leq i \leq n$) **DO**

5: Evaluate the corresponding fitness function F_i $fitness = 1/\sum\limits_{i=1}^{N} \sum\limits_{j=1}^{K} (Y_j(i) - \overline{Y}_j(i))^2$

6: END FOR

7: WHILE ($g < Gen_{max}$) **DO**

8: FOR EACH ($i = 1$:*n*) **DO IF** ($p_c > rand$) **THEN**

9: Conduct the crossover operation of **GA** to optimize the smoothing factor of **GRNN**

10: END IF

11: IF ($p_m > rand$) **THEN**

12: Conduct the mutate operation of **GA** to optimize the smoothing factor of **GRNN**

13: END IF END FOR

14: FOR EACH (*i*: $1 \leq i \leq n$) **DO**

15: Evaluate the corresponding fitness function F_i $fitness = 1/\sum\limits_{i=1}^{N} \sum\limits_{j=1}^{K} (Y_j(i) - \overline{Y}_j(i))^2$

16: END FOR

17: /*Update best nest x_p of the *d* generation to replace the former local optimal solution.*/

18: FOR EACH (*i*: $1 \leq i \leq n$) **DO IF** ($F_p < F_b$) **THEN** $x_b \leftarrow x_p$;

19: END IF END FOR END WHILE

20: RETURN x_b /* Set the weight and threshold of the **GRNN** according to x_b.*/

21: Use x_t to train the **GRNN** and update the weight and threshold of the **GRNN** and **input** the historical data into **GRNN** to obtain the forecasting value $\hat{y}$.

Step 1. Data addressed by layer-upon-layer decreasing method would include some noises, affecting the forecasting accuracy; therefore, the first step is to denoise the composed data by using EMD method.

Step 2. Standardize and code the time series after the denoising.

Step 3. Generate the initial population $P(t)$, and the evolutionary generation is $t = 0$.

Step 4. Code the chromosome, and get the parameters of GRNN, which can be used to train the network structure.

Step 5. Set the individual evaluation standard according to the fitness function in Equation (34):

$$fitness = \frac{1}{\sum\limits_{i=1}^{N} \sum\limits_{j=1}^{K} (Y_j(i) - \overline{Y}_j(i))^2} \tag{34}$$

where $Y_j(i)$ is the output of GRNN and $\overline{Y}_j(i)$ is the output.

Step 6. Apply the optimum strategy based on the values of fitness function.

Step 7. Judge whether the fitness value meets the accuracy requirement. If so, the process ends; or move to the next step.

Step 8. Judge whether the current iteration t gets to the maximum iteration. If so, the process ends; or go to the next step.

Step 9. Perform the selection, crossover and mutation operation for the current population.

Step 10. Generate the new generation of the population, and the iteration *t* becomes *t* + 1, return Step 3.

4. Experiments

With the rapid development of technology and science, the electrical power system in each country tends to develop fast as well. Similarly, the power grid management has become more complicated. The forecasting is the premise and basis of decision and control; therefore, the premise

and the most vital step of electrical load management is to conduct the electrical load forecasting. The accurate forecasting can not only help the electrical power system operate safely based on reasonable maintenance schedules, but it can also decrease the grid costs and maximize the profits.

4.1. Model Evaluation

To conduct the model evaluation can lead to a clear and direct understanding of the forecasting accuracy, and it is helpful to analyze the reasons causing errors to enhance the forecasting performance. The main reasons are listed below:

(1) *Selection of influencing factors when constructing mathematical models*. In truth, the time series is affected by various factors, and it is difficult to master all of them. Therefore, errors between forecast values and actual values cannot be avoided.

(2) *Improper algorithms*. For forecasting, we just build a relatively appropriate model, so if the algorithms are chosen wrongly, the errors would become larger.

(3) *Inaccurate or incomplete data*. The forecasting should be based on the historical data, so inaccurate or incomplete data can result in forecasting errors.

When there are abnormal values, we are supposed to find the reasons causing the errors and correct each step of the model. The forecasting accuracy plays a crucial role in assessing a forecasting algorithm, and two types of evaluation metrics are chosen to evaluate the forecasting accuracy: the accuracy of forecasting a single point and the overall accuracy of forecasting multiple points. Two evaluation metrics are applied to examine a single point forecasting accuracy, which are absolute error (AE) and relative error (RE). Then we select four evaluation metrics, including mean absolute error (MAE), root mean square error (RMSE), mean absolute percentage error (MAPE) and mean error (ME), to evaluate the model performance more comprehensively. MAPE is a generally accepted metric for forecasting accuracy, and MAE and RMSE can measure the average magnitude of the forecast errors; however, RMSE imposes a greater penalty on a large error than several small errors [43].

For a group of time series x_t (t = 1, 2, ... , T), the corresponding forecasting output is $\hat{x}_t$ and detailed description of evaluation metrics is shown in Table 1.

Table 1. The evaluation metrics.

Name of Metrics	Equation	No.	Name of Metrics	Equation	No.
MAE	$MAE = \frac{1}{T} \sum\limits_{t=1}^{T} |x_t - \hat{x}_t|$	(35)	ME	$ME = \frac{1}{T} \sum\limits_{t=1}^{T} (x_t - \hat{x}_t)$	(38)
RMSE	$RMSE = \sqrt{\frac{1}{T} \sum\limits_{t=1}^{T} (x_t - \hat{x}_t)^2}$	(36)	AE	$AE = x_t - \hat{x}_t$	(39)
MAPE	$MAPE = \frac{1}{T} \sum\limits_{t=1}^{T} \left| \frac{x_t - \hat{x}_t}{x_t} \right| \times 100\%$	(37)	RE	$RE = \frac{x_t - \hat{x}_t}{x_t}$	(40)

The smaller values of the six metrics are, the higher forecasting accuracy is. Therefore, the evaluation metrics can both reflect the forecasting results and its accuracy clearly and directly and provide a reference base for decisions, which is beneficial to improving the model and conducting the analysis. Thus, the significance of the evaluation metrics is very large.

4.2. Experimental Setup

This paper uses the 30-min interval data of New South Wales, Australia in April 2011 to verify the effectiveness of the proposed hybrid DDH model based on data decomposition. In the first experiment, the data size is 1440, and data in the first 29 days are the training set, and the testing set includes data in the 30th day. The detailed ideas of the proposed electrical load hybrid model is summarized as follows (as shown in Figure 2):

(1) The original electrical load time series data Y_t has an obvious trend and periodicity. Initially, the moving average method is conducted to extract the periodicity C_t. For the periodicity C_t, conduct the periodic adjustment and obtain $\hat{C}_t$.

(2) Subtract the periodicity of the original time series data, and get the original trend $T_t(T_t = Y_t - \hat{C}_t)$. For the original data without periodicity, EMD needs to be initially applied to eliminate the noises and improve the forecasting accuracy. Then the genetic algorithm could be used to optimize GRNN to obtain the forecasting trend item $\hat{T}_t$.

(3) Finally, the randomness can be obtained through $R_t = Y_t - \hat{C}_t - \hat{T}_t$, then the GRNN optimized by the genetic algorithms is utilized to forecast the randomness and the forecasting value is obtained. The trend tends to be steady; therefore, there is no need to eliminate noises.

(4) The final forecasting is performed by the additive model of time series $\hat{Y}_t = \hat{C}_t + \hat{T}_t + \hat{R}_t$.
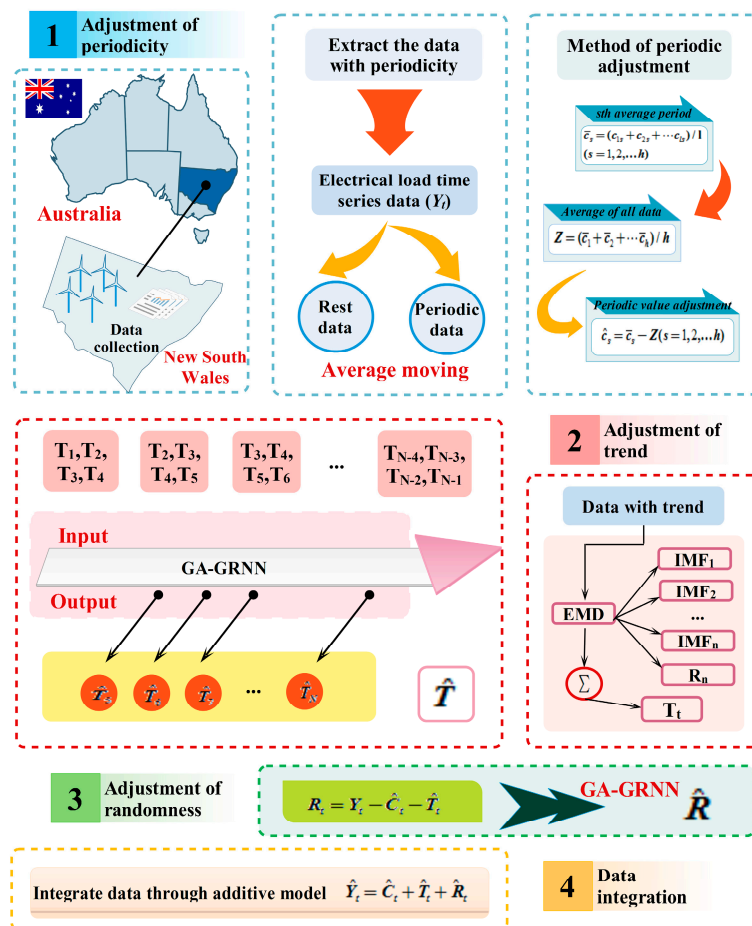

Figure 2. The process of electrical load forecasting for New South Wales.

4.3. Empirical Results

The model performance is evaluated based on the upper data, and the results are obtained by using MATLAB®(2015a), which was implemented under Windows 8.1 with a 2.5 GHz Intel Core i5 3210 M, 64 bit CPU with 4 GB RAM. Figure 3 shows the data decomposition process.
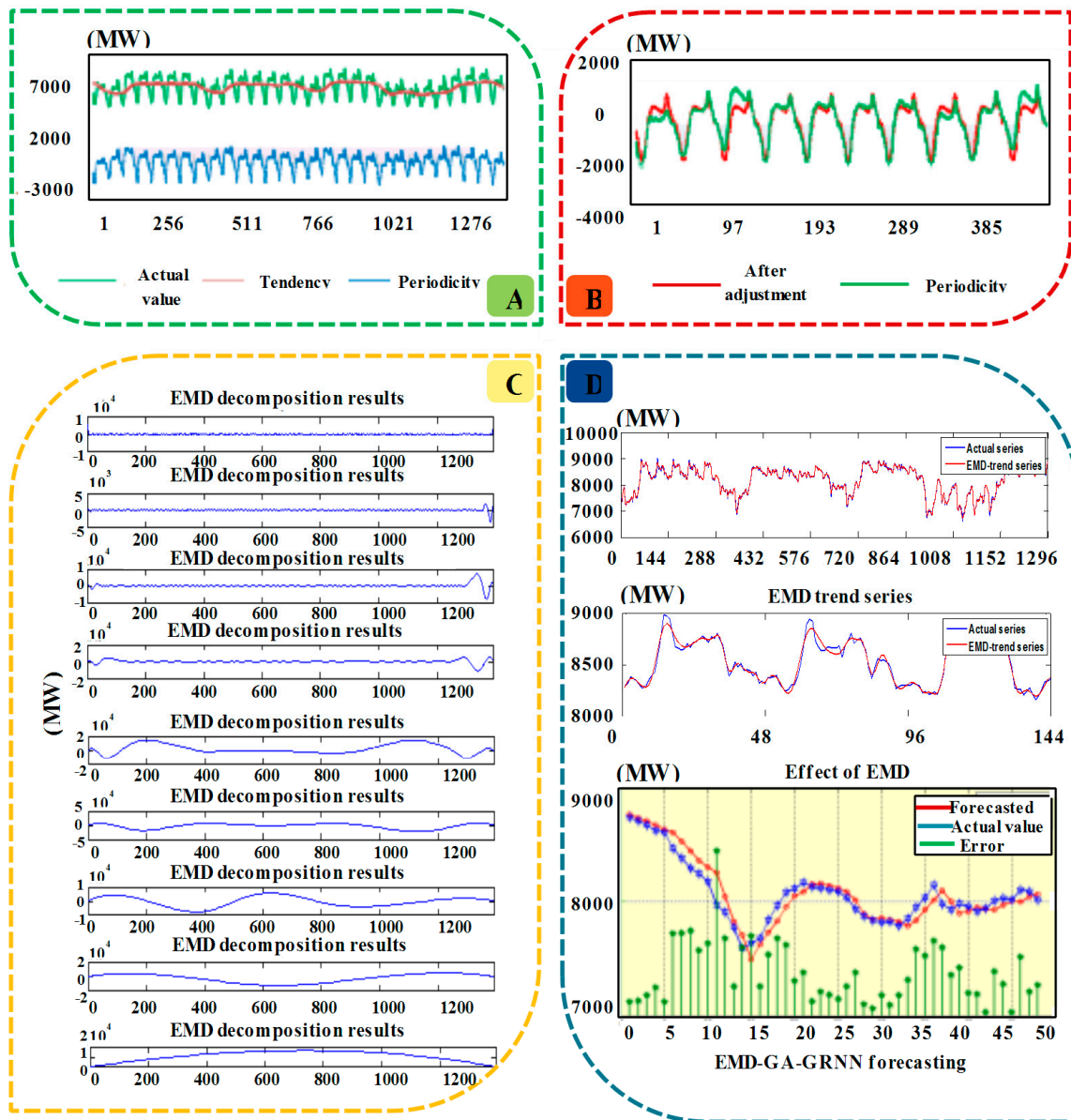
Figure 3. The forecasting effects. (**A**) The original electrical load time series; (**B**) Electrical load time series data after adjustment; (**C**) EMD decomposition results; (**D**) EMD trend series, effect of EMD and results of EMD-GA-GRNN forecasting.

(1)　Figure 3A shows the results after decomposition by moving average, from which it can be seen that the original electrical load data contains a certain periodicity, and the variation of the period is roughly equal, so the additive model is more suitable. The length of the period $h = 48$ can be ensured based on the data distribution. Thus the moving average method is used to decompose the electrical load data into two parts, which are periodicity and trend. Besides, from the decomposed results, it can be known that the level of trend is nearly ten times the periodicity. This is because that the moving average method can demonstrate the large trend of the development, eliminating the fluctuation factors such as season. Therefore, the periodic adjustment should be conducted through extracting the periodicity.

(2)　Figure 3B is the electrical load data after periodic adjustment, from which is can be known that the electrical load data after the periodic adjustment have periodic sequence and basis trend characteristics.

(3)   Figure 3C demonstrates the output results of trend data after EMD. It shows that nine components are obtained, including IMF_1, IMF_2, ..., IMF_8 and R_n, after EMD data decomposition. The high-frequency data in highest component is removed, and the rest data are regarded as the new trend time series data.

(4)   Figure 3D clearly reveals the trend data after EMD decomposition by removing the high frequency component, and it can be obviously seen that the data denoised by EMD are smoother than the original data.

Next, data after removing the high frequency component by EMD is fitted and forecast by GRNN. The genetic algorithm is applied to optimize the smoothing factor σ in GRNN. The hybrid electrical load forecasting model EMD-GA-GRNN constructed in this paper is applied to forecast the trend value in the next time point by using the historical data in the past time point. In this experiment, the trend value of the former four time points are used to forecast the trend value of the 5th time point. For the given data, the data need initially to be divided into the training sample and testing sample. Take the training sample for example, x_1, x_2, x_3, x_4, x_5 is the first sample group, and x_1, x_2, x_3, x_4 are independent variables, and x_5 is the objective function value. Similarly, x_2, x_3, x_4, x_5, x_6 is the second sample group, x_2, x_3, x_4, x_5 are independent variables, and x_6 is the objective function value. By that analogy, the final training matrix is:

$$
\begin{pmatrix}
x_1 & x_2 & x_3 & \cdots & x_{1292} \\
x_2 & x_3 & x_4 & \cdots & x_{1293} \\
x_3 & x_4 & x_5 & \cdots & x_{1294} \\
x_4 & x_5 & x_6 & \cdots & x_{1295} \\
x_5 & x_6 & x_7 & \cdots & x_{1296}
\end{pmatrix}
\tag{41}
$$

where each column is a sub-sample sequence, and the last row is the expected output. The training sample is used to train GA-GRNN, after that the network after training is obtained. The forecasting effects can be clearly seen from Figure 3D that EMD-GA-GRNN has a better fitting effect, and MAPE between network output and real value is 2.11%. The training model in Figure 4 is shown as follows.
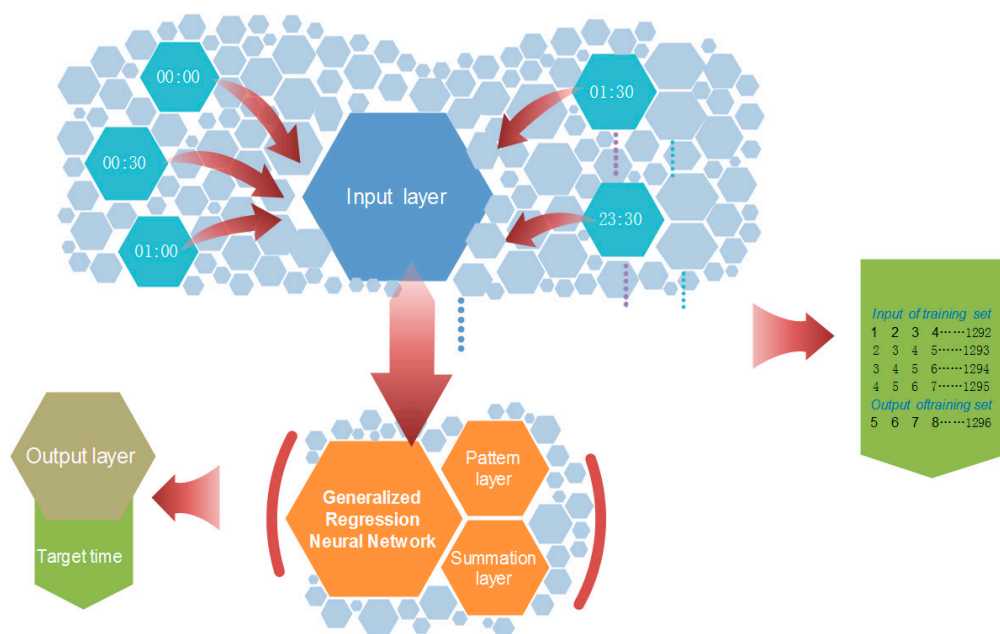

Figure 4. The generalized regression neural network model.

To the best of our knowledge, a great variety of forecasting approaches can achieve good performance in dealing with non-linear time series; therefore, in this paper we compared the proposed GRNN with three other well-known and commonly used methods, including wavelet neural network (WNN), the secondary exponential smoothing method (SES) and auto regressive integrated moving average (ARIMA). The forecasting results are compared as shown in Figure 5, from which it can be known that:

(1)　The speed to forecast the nonlinear time series data by using WNN is fast, with a better ability of generalization and a higher accuracy; however, the stability is weak.

(2)　The advantages of SES are the simple calculation, strong adaptability and stable forecasting results, but the ability to address nonlinear time series data is weak.

(3)　ARIMA performs well with a relatively higher accuracy when forecasting the electrical load data. However, as time goes by, the forecasting errors would gradually become larger and larger, which is only suitable for short-term forecasting.

(4)　On the whole, compared with other methods, GRNN can obtain a better and more stable forecasting result, as it deals with the non-linear data well and can fit and forecast the electrical load data well.
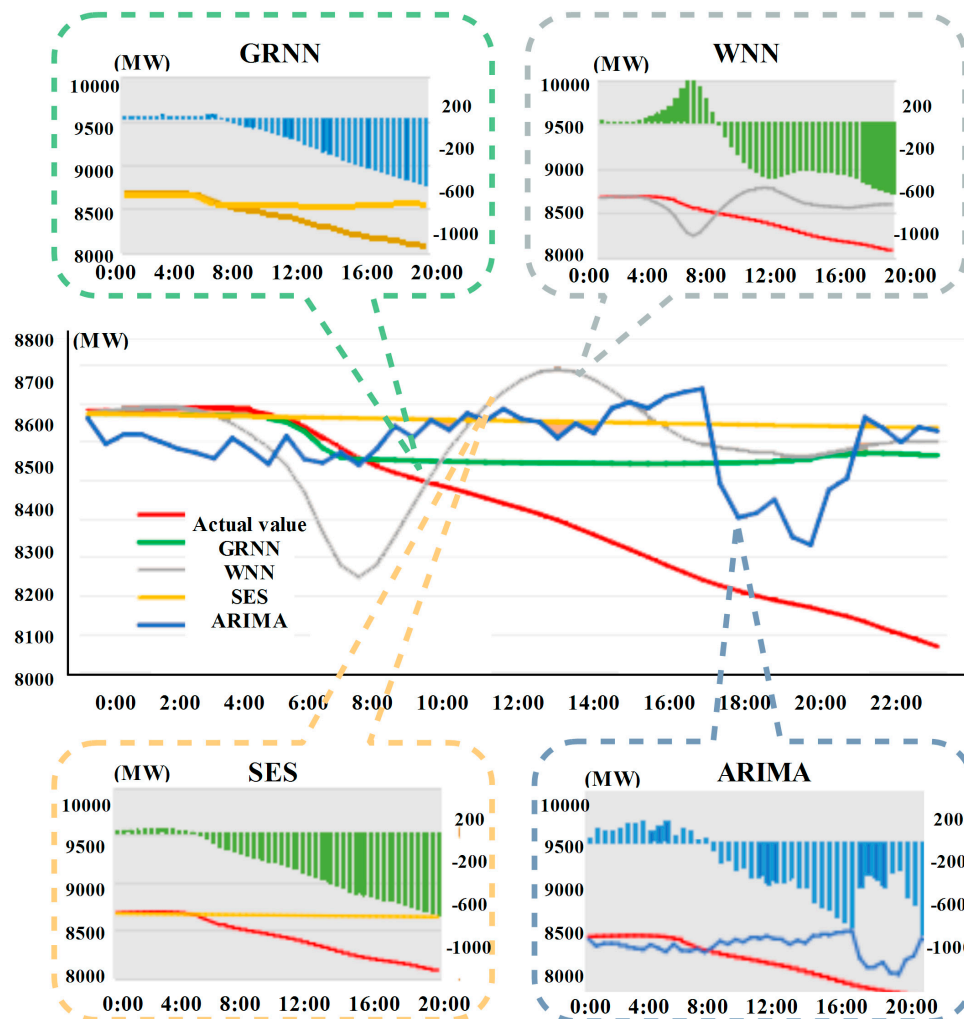

Figure 5. Forecasting results for trend of each model after removing the periodicity.

Next the randomness is obtained by $R_t = Y_t - \hat{C}_t - \hat{T}_t$. Because it tends to be stationary, we can only apply GA-GRNN to get the forecasting value $\hat{R}_t$. The forecasting results of DDH can be calculated

by the additive model $\hat{Y}_t = \hat{C}_t + \hat{T}_t + \hat{R}_t$, and results are shown in Figure 6. Figure 6II demonstrates that the forecasting error in the 11th time point is the largest with an MAPE within 5%, and this results is satisfactory.
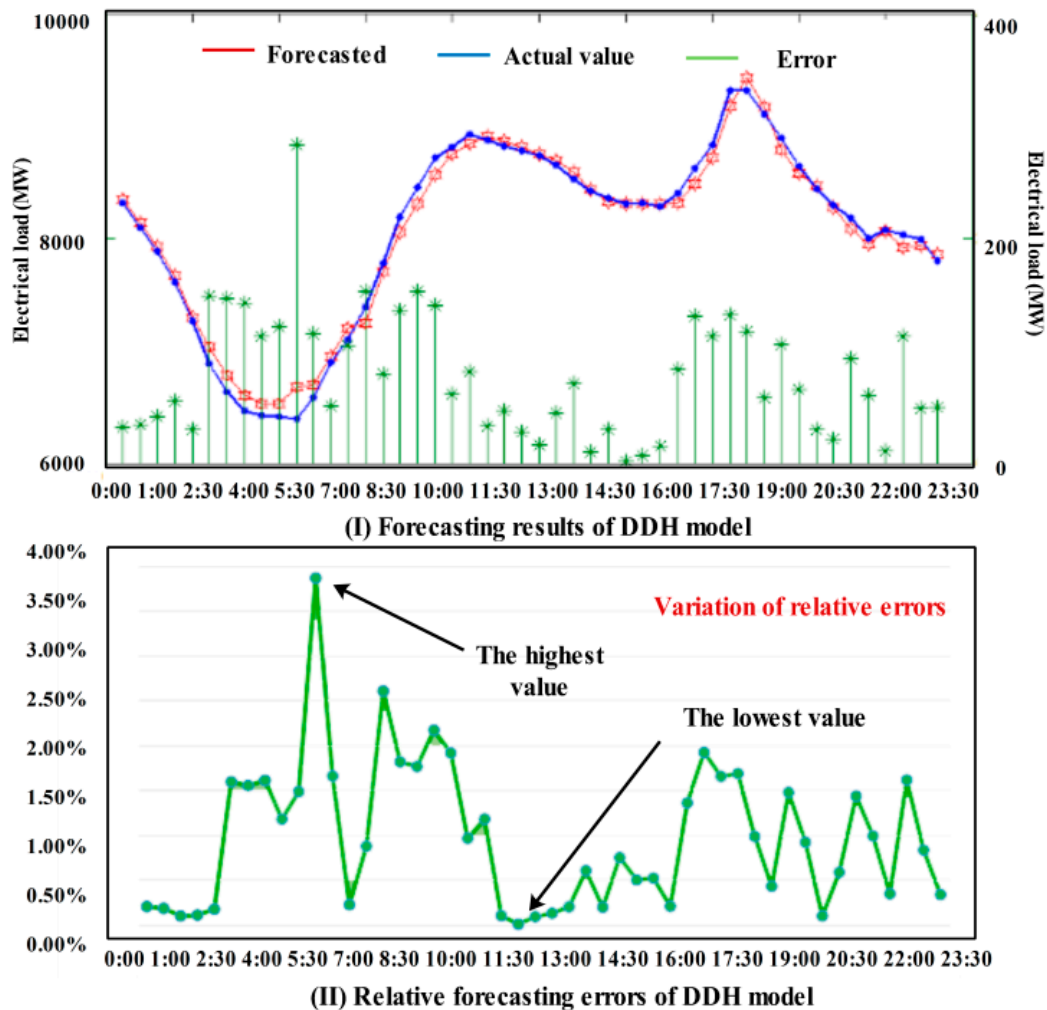
Figure 6. Forecasting results and MAPE of DDH model.

4.4. Comparative Analysis

In order to prove the good performance of the proposed DDH model in this paper, three other hybrid models are compared with it, which are EMD-GA-WNN, GA-GRNN and EMD-GA-GRNN. The comparison results are shown in Table 2.

(1)   From Figure 7, it can be seen that EMD-GA-WNN does not perform well when forecasting the electrical load data, and the relative errors of some parts even exceed 5%. This may be caused by the weak forecasting stability of WNN, and although GA can optimize its parameters, the effect to improve its stability is weak.

(2)   As for GA-GRNN and EMD-GA-GRNN, MAPEs are all within 5%, which indicates that the two forecasting models have better performance. In detail, the forecasting effect of EMD-GA-GRNN is much better than that of GA-GRNN, proving the function of EMD in improving the forecasting accuracy.

(3)　The DDH model based on the data decomposition put forward in this paper can control the MAPE at 4%; thus, it can be known that it has a very strong fitting ability for non-linear data and forecasting ability for the electrical load time series. Both the simulation results and the forecasting process demonstrate that the proposed model can have a good performance when forecasting the non-linear time series data with periodicity, trend and randomness.

(4)　From the evaluation metrics in Figure 7, it can be known that the forecasting ability of GRNN is better than WNN, which is because that GRNN can deal well with the data such as electrical load time series; therefore, this paper also establishes the model based on GRNN. The proposed forecasting model EMD-GA-GRNN and EMD-GA-GRNN based on WNN and GRNN can improve the forecasting accuracy well. However, in comparison, GRNN is more suitable for the nonlinear time series data, and MAPEs of EMD-GA-WNN and EMD-GA-GRNN are 2.22% and 1.53%, respectively. Certainly, EMD can reduce the forecasting errors in some degree. Besides, MAPE decreases from 1.62% of GA-GRNN to 1.53% of EMD-GA-GRNN. However, DDH model can reduce MAPE within 1%.

(I) Comparison of forecasting values and actual values of four models
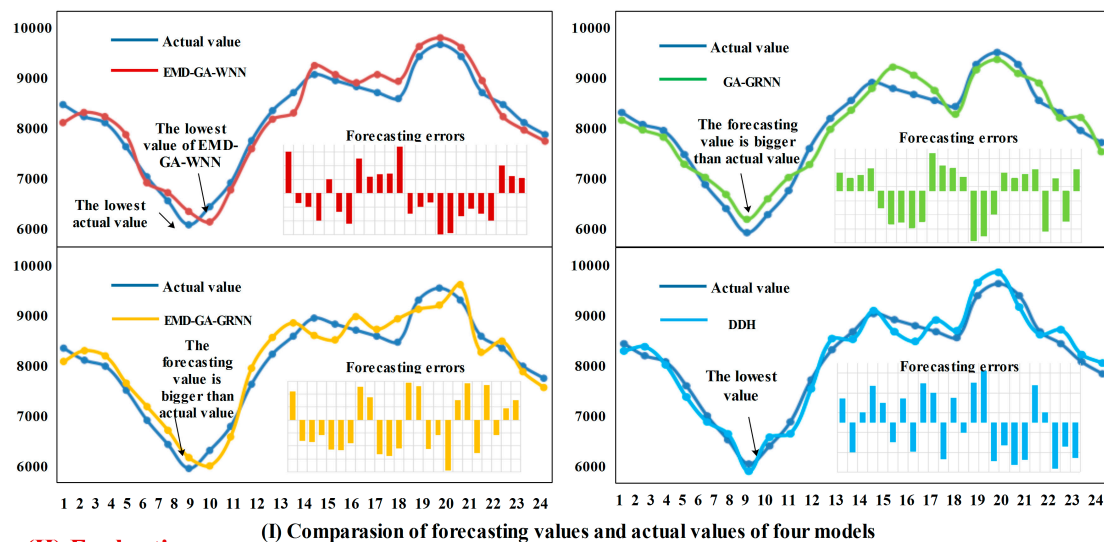
(II) Evaluation metrics

$$MAE = \frac{1}{T}\sum_{t=1}^{T}\left|x_t - \hat{x}_t\right|$$

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(x_t - \hat{x}_t)^2}　ME = \frac{1}{T}\sum_{t=1}^{T}\left|x_t - \hat{x}_t\right|$$

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{x_t - \hat{x}_t}{x_t}\right|\times100\%　RE = \frac{(x_t - \hat{x}_t)}{x_t}$$

$$AE = x_t - \hat{x}_t$$

(III) MAE, RMSE, MAPE and ME of four models

	MAE	RMSE	MAPE	ME
EMD-GA-WNN	168.1870	217.1040	2.22%	-147.5324
GA-GRNN	130.9404	167.7759	1.62%	-59.7562
EMD-GA-GRNN	124.0530	163.6047	1.53%	-26.5351
DDH	77.0542	97.7689	0.98%	28.6760

Figure 7. Forecasting results of each model.

Table 2. The forecasting output of each model.

Time	Actual Value	Forecasting Output of the Model				Time	Actual Value	Forecasting Output of the Model			
		EMD-GA-WNN	GA-GRNN	EMD-GA-GRNN	DDH			EMD-GA-WNN	GA-GRNN	EMD-GA-GRNN	DDH
0:00	8314.34	8316.79	8485.70	8448.82	8297.61	12:00	8731.01	8911.29	8740.80	8749.89	8719.90
0:30	8097.56	8120.24	8324.54	8277.18	8082.89	12:30	8649.94	8869.53	8629.02	8653.62	8667.12
1:00	7881.03	7928.45	8120.26	8077.17	7873.46	13:00	8520.88	8820.83	8534.06	8544.21	8572.19
1:30	7613.53	7687.64	7880.49	7856.92	7621.62	13:30	8419.95	8746.33	8436.37	8418.62	8403.67
2:00	7265.94	7330.03	7586.18	7558.97	7253.44	14:00	8359.06	8668.88	8332.43	8324.83	8296.61
2:30	6884.16	7067.63	7116.65	7134.47	6993.73	14:30	8304.55	8570.31	8261.18	8263.58	8263.13
3:00	6638.55	6870.98	6725.93	6750.43	6741.38	15:00	8312.49	8529.19	8219.80	8205.82	8269.39
3:30	6464.02	6743.91	6519.60	6508.65	6567.98	15:30	8285.25	8513.52	8219.06	8196.43	8268.39
4:00	6421.67	6707.56	6392.63	6404.61	6497.16	16:00	8401.56	8545.67	8210.85	8185.83	8287.56
4:30	6412.23	6724.73	6419.99	6451.49	6507.20	16:30	8616.59	8660.40	8297.24	8257.93	8451.27
5:00	6394.72	6873.82	6536.38	6556.23	6641.75	17:00	8829.22	8806.80	8600.25	8503.72	8683.46
5:30	6582.09	7033.42	6598.89	6605.74	6691.06	17:30	9308.85	9216.90	9000.54	8890.66	9152.10
6:00	6899.25	7399.88	6802.04	6814.45	6914.60	18:00	9307.55	9429.90	9249.77	9200.67	9398.89
6:30	7093.06	7642.45	7200.95	7193.05	7155.01	18:30	9106.91	9271.13	9328.58	9330.35	9145.80
7:00	7395.69	7715.40	7589.45	7524.90	7203.21	19:00	8893.08	8987.73	9292.04	9240.84	8762.06
7:30	7783.28	7973.56	7850.94	7754.65	7642.06	19:30	8641.32	8747.35	9042.83	8970.91	8562.08
8:00	8193.10	8145.44	8135.61	8026.69	8048.81	20:00	8437.27	8565.72	8596.66	8541.77	8445.35
8:30	8454.25	8325.99	8475.43	8493.17	8270.93	20:30	8297.38	8389.37	8407.47	8394.00	8249.05
9:00	8710.86	8616.62	8718.43	8606.26	8544.31	21:00	8174.13	8189.12	8278.29	8260.20	8057.12
9:30	8806.93	8824.25	8975.41	8860.02	8722.24	21:30	8004.14	7998.97	8129.02	8116.55	7925.10
10:00	8920.64	8984.50	8994.28	8924.99	8815.63	22:00	8077.99	8098.08	7964.69	7949.84	8050.27
10:30	8872.57	9023.17	9034.39	8910.56	8863.76	22:30	8033.84	7965.71	7977.29	7939.60	7904.37
11:00	8816.19	8981.31	8953.15	8840.24	8815.61	23:00	7989.55	7951.66	8016.42	7970.00	7923.18
11:30	8777.05	8946.75	8844.91	8810.79	8769.22	23:30	7803.73	7841.56	8020.16	7972.34	7829.81

The summary is concluded in Remark 1.

Remark 1. *It can be concluded that compared to the single forecasting model, DDH model is more suitable for forecasting the electrical load time series data with a higher fitting ability and better forecasting capacity.*

The analysis above only shows results of three models in one experiment, but it cannot comprehensively and fully demonstrate the model performance. Each model will be trained 10 times with the same iteration numbers to make the forecasting results more stable. The obtained forecasting quality and results are shown in Figure 8 and Table 3. The two figures both indicate that DDH model based on the data decomposition perform well when measured by different evaluation metrics. A smaller MAE means a higher forecasting accuracy, a lower RMSE indicates a better fitting degree of electrical load, and MAPE is an index to assess the forecasting ability of the model. At present, for the data of New South Wales, the best standard is about 1%. From the average of MAE in ten experiments, DDH has the smallest value, indicating the best forecasting accuracy. What is more, the smallest RMSE cannot only mean that DDH can fit the electrical load time series well, but it can also prove that the forecasting results of the model are stable.
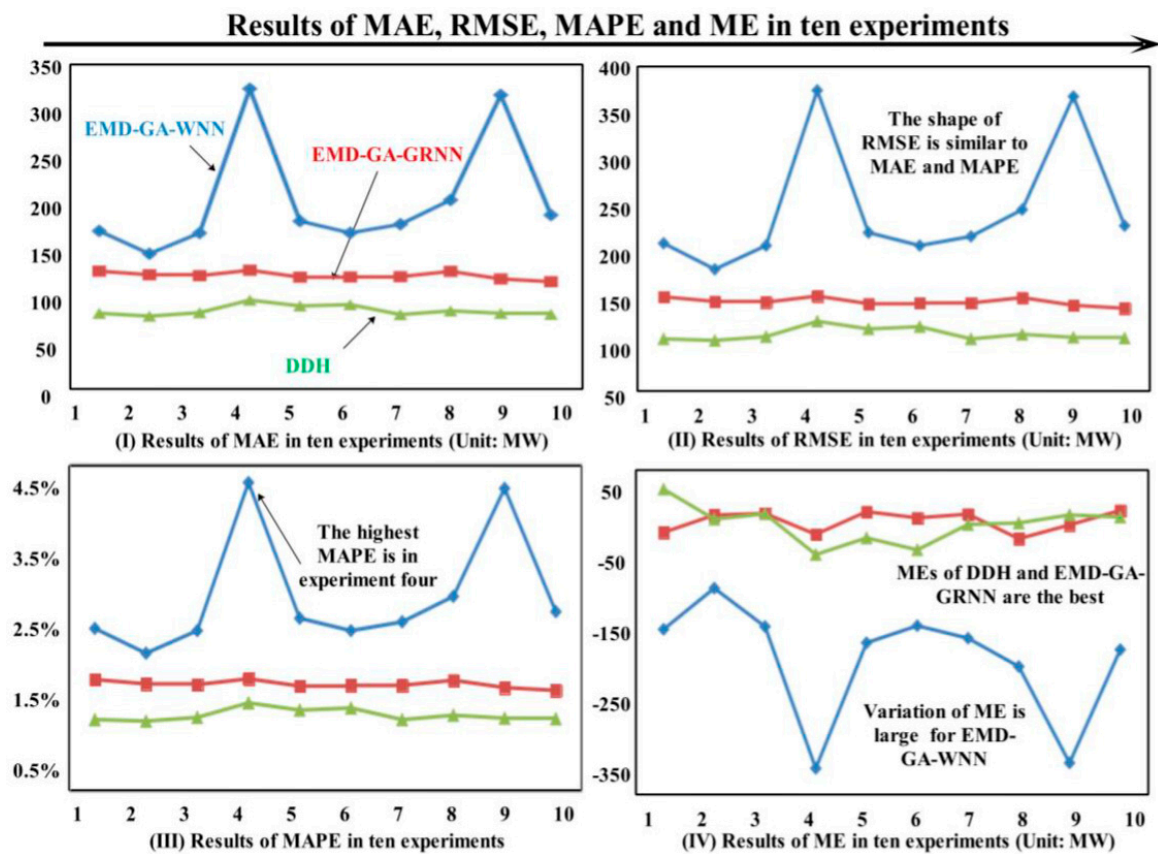

Figure 8. Model evaluation of three forecasting models.

Furthermore, MAPE of DDH also shows that DDH model based on the data decomposition put forward in this paper can reach the best forecasting standard currently.

Table 3. Forecasting performance evaluation results.

No.	EMD-GA-WNN				EMD-GA-GRNN				DDH			
	MAE	RMSE	MAPE	ME	MAE	RMSE	MAPE	ME	MAE	RMSE	MAPE	ME
1	168.9702	218.9832	0.0223	−150.3171	124.0530	163.6047	0.0153	−26.5351	77.0542	97.7687	0.0098	28.6760
2	143.4750	187.9665	0.0189	−97.7305	120.0837	160.1870	0.0147	−4.4056	73.8617	90.7383	0.0096	−9.6550
3	166.6531	215.6456	0.0220	−146.7813	119.0237	161.9871	0.0146	−2.3599	77.7576	96.1427	0.0101	−2.7978
4	327.7693	366.7733	0.0423	−327.7693	125.1301	160.2820	0.0154	−28.9499	91.7547	110.8149	0.0121	−54.7912
5	179.9683	231.0331	0.0237	−167.4704	117.3492	158.6897	0.0144	−0.0239	85.2639	101.5193	0.0111	−33.4767
6	166.7894	215.3270	0.0220	−145.6936	117.6313	157.8023	0.0145	−7.9405	86.7251	105.4555	0.0114	−48.7363
7	176.4214	228.1589	0.0232	−161.6078	117.9699	158.3187	0.0145	−3.1824	75.6047	90.9116	0.0098	−16.3685
8	203.5435	250.7716	0.0267	−197.7634	123.8518	163.8211	0.0152	−34.5862	79.7625	96.8084	0.0104	−14.4799
9	320.8268	358.9548	0.0415	−320.8268	115.4209	157.6318	0.0142	−16.9704	77.0379	94.9811	0.0100	−4.2135
10	186.7493	239.4212	0.0246	−176.1809	112.3399	155.0562	0.0138	1.6029	75.9955	93.4155	0.0099	−6.9212
Mean	204.1166	251.3035	0.0267	−189.2141	119.2854	159.7381	0.0147	−12.3351	80.0818	97.8556	0.0104	−16.2764

4.5. Further Experiments

Initially, in order to further prove the effectiveness of the proposed DDH hybrid model, we expand our sample size by using the data in 89 days to forecast the data in the 90th day. That is to say, the first 89th days are the training set, and the testing set include data in the 90th day. The experiment results of both working days and weekends are shown in Table 4. Besides, experiments of days in different seasons are also done to examine the effectiveness and robustness of the proposed hybrid model, which are listed in Table 4 and detailed analysis are as follows:

(1) As for the weekly analysis, it can be seen that the average MAPE of DDH in one week is 1.01%, which is lower than EMD-GA-WNN and EMD-GA-GRNN. About other indexes, including MAE, RMSE and ME, DDH all obtain the best forecasting results. When comparing the working days with weekends, the proposed hybrid model can both have a high forecasting accuracy, which proves the effectiveness of the model.

(2) Table 5 shows the forecasting results of days in different seasons. Based on the comparison, it can be concluded that DDH is superior to the other two models with the values of MAPE 0.96%, 1.18%, 1.18% and 1.13% in spring, summer, autumn and winter, respectively. The results can validate that the proposed hybrid DDH model has a high degree of robustness and forecasting accuracy.

The summary is concluded in Remark 2.

Remark 2. *The performance of the DDH model is stable and good when forecasting the electrical load data in one week and different seasons.*

Table 4. Forecasting performance evaluation results of one week with larger training set.

	Monday	**Tuesday**	**Wednesday**	**Thursday**	**Friday**	**Saturday**	**Sunday**	**Week**
				EMD-GA-WNN				
MAE	117.6839	129.4867	142.1135	279.4283	225.6017	118.0649	206.1598	174.0770
RMSE	288.4620	237.0900	218.0418	211.3369	202.1987	245.0624	289.4222	241.6591
MAPE	0.0249	0.0286	0.0273	0.0263	0.0275	0.0219	0.0226	0.0256
ME	−129.1364	−18.7732	−130.2645	−88.3712	−32.1989	−60.3024	−115.4213	−82.0668
				EMD-GA-GRNN				
MAE	117.6850	126.4213	110.0889	112.1325	129.0178	157.3144	138.0976	127.2511
RMSE	141.7699	168.3712	155.2626	148.0987	161.5546	132.1019	168.3174	153.6395
MAPE	0.0144	0.0137	0.0149	0.0155	0.0158	0.0159	0.0142	0.0149
ME	−98.6273	−87.0125	−110.2455	−136.4188	−65.231	−21.0987	−33.4685	−78.8718
				DDH				
MAE	76.4219	88.1348	76.1653	79.0187	84.315	69.1083	70.4245	77.6555
RMSE	97.6681	102.4269	99.8349	105.1917	112.3416	108.1947	98.1032	103.3944
MAPE	0.0101	0.0094	0.0112	0.0095	0.0098	0.0106	0.0103	0.0101
ME	−13.0719	−2.0715	−4.3728	18.1605	12.1004	−34.5671	−10.0628	−4.8407

Table 5. Forecasting performance evaluation results of different seasons with a larger training set.

Evaluation Index	EMD-GA-WNN	EMD-GA-GRNN	DDH	Evaluation Index	EMD-GA-WNN	EMD-GA-GRNN	DDH
	Spring				Summer		
MAE	119.4287	117.0216	76.0138	MAE	137.0345	108.417	60.1837
RMSE	292.0655	140.3726	97.0138	RMSE	213.0418	156.1783	94.1296
MAPE	0.0231	0.0158	0.0096	MAPE	0.0274	0.0162	0.0118
ME	−112.0659	−78.4257	−16.1076	ME	−97.3125	−52.1035	−20.0244
	Autumn				Winter		
MAE	125.0638	100.0246	78.1025	MAE	112.0605	100.4629	73.1068
RMSE	213.1294	148.7329	89.4237	RMSE	200.0217	158.0376	971136
MAPE	0.0219	0.0143	0.0118	MAPE	0.0212	0.0158	0.0113
ME	−101.0137	−25.4269	−11.0036	ME	−94.1346	−36.0599	−11.0217

In addition, we also compare the forecasting performance of the proposed DDH model in this paper to the models in the literature, including [1,4,44,45]. As shown in Table 6, the model in this paper improves the forecasting accuracy by 0.089% compared to the HS-ARTMAP network. The MAPEs of the combined model based on BPNN, ANFIS and diff-SARIMA and hybrid model based on WT, ANN and ANFIS are 1.654% and 1.603%, respectively. In the compared models, the combined model based on BPNN, RBFNN, GRNN and GA-BPNN has the lowest MAPE, which is 1.236%. Therefore, in summary, the DDH model outperforms the other compared models in the literature. The superior performance of DDH is because that the model can deal with both trend and periodicity in the original time series, which can greatly enhance the forecasting accuracy. Besides, compared to conventional BPNN and ARIMA, GRNN has a strong ability of generalization, robustness, fault tolerance and convergence ability.

Table 6. Comparison of MAPE with models in the literature.

Model	Period	MAPE (%)	Ref.
Combined model based on BPNN, ANFIS and diff-SARIMA	Data from May to June 2011	1.654	[1]
Combined model based on BPNN, RBFNN,GRNN and GA-BPNN	Data from February 2006 to February 2009	1.236	[4]
HS-ARTMAP network	Data in the head days in January from 1999 to 2009	1.900	[44]
Hybrid model based on WT, ANN and ANFIS	Data from 12 July to 31 July 2004	1.603	[45]
The proposed DDH	Data from April to June 2011	1.010	/

4.6. Discussion on Model Features

As discussed above, the major model in DDH model is GRNN which is optimized by GA. The experimental results also demonstrate their effectiveness in forecasting the short-term electrical load time series. This part will discuss the advantages of GRNN and GA further and more deeply. As shown in Table 7, GRNN has four obvious features:

1. It has a relatively low requirement for the sample size during the model building process, which can reduce the computing complexity;
2. The human error is small. Compared with the back propagation neural network (BPNN), GRNN is different. During the training process, the historical samples will directly control the learning process without adjusting the connection weight of neurons. What is more, parameters like learning rate, training time and the type of transfer function, need to be adjusted. Accordingly, there is only one parameter in GRNN that needs to be set artificially, which is the smoothing factor;
3. Strong self-learning ability and perfect nonlinear mapping ability. GRNN belongs to a branch of RBF neural networks with strong nonlinear mapping function. To apply GRNN in electrical load forecasting can better reflect the nonlinear mapping relationship;
4. Fast learning rate. GRNN uses BP algorithm to modify the connection weight of the relative network, and applies the Gaussian function to realize the internal approximation function, which can help arrive at an efficient learning rate. The above features of GRNN play a pivotal role in performing the electrical load forecasting when the original data are fluctuating and non-linear.

Table 7. Comparison of GRNN with grey model, regression model and BP model.

Model	Nonlinear Mapping Ability	Set parameters Artificially	Generalization	Robustness	Fault Tolerance	Structural Interpretability	Convergence Ability	Sample Size
Grey model	Middle	Middle	Weak	Weak	Weak	Good interpretability for internal structure	——	Large
Regression model	Weak	Large	Middle	Weak	Weak	Good interpretability for internal structure	——	Large
Back Propagation (BP) model	Strong	Large	Middle	Middle	Weak	No structural interpretability	Weak	Large
GRNN model	Strong	Only smoothing factor parameter	Strong	Strong	Strong	No structural interpretability	Strong	Low requirement

The genetic algorithm is utilized to optimize the only one parameter in GRNN, and it is a type of algorithm that works without limiting the field or type of the problem. That is to say, it does not depend on detailed problems, and can provide a universal framework to solve problems. Compared to the traditional optimization, it has the following advantages:

- Self-adaptability. When solving problems, GA deals with the chromosome individuals through coding. During the process of evolution, GA will search the optimal individuals based on the fitness function. If the fitness value of chromosome is large, it indicates a stronger adaptability. It obeys the rules of survival of the fittest; meanwhile, it can keep the best state in a changing environment;
- Population search. The conventional methods usually search for single points, which is easily trapped into a local optimum if a multimodal distribution exists in the search space. However, GA can search from multiple starting points and evaluate several individuals at the same time, which makes it achieve a better global searching;
- Need for a small amount of information. GA only uses the fitness function to evaluate the individuals without referring to other information. It has a small dependence or limitation conditions to the problems, so it has a wider applicability;
- Heuristic random search. GA highlights the probability transformation instead of the certain transformation rule;
- Parallelism. On the one hand, it can search multiple individuals in the solution space; on the other hand, multiple computers can be applied to perform the evolution calculation to choose the best individuals until the computation ends. The above advantages make GA widely used in many fields, such as function optimization, production dispatching, data mining, forecasting for electrical load and so on.

5. Conclusions

The electrical load forecasting can not only provide the electricity supply plans for regions in a timely and reliable way, but it can also help maintain normal social production and life. Thus, to improve the forecasting accuracy of electrical load can lower risks, improve the economic benefits, decrease the costs of generating electricity, enhance the safety of electrical power systems and help policy makers make better action plans. Therefore, how to forecast the changing trends and features of electrical loads in the power grid accurately and effectively has become a both significant and challenging problem. This paper proposes a Data Decomposition Hybrid (DDH) model based on the data decomposition that can deal well with the task, and it mainly contains two key steps:

The first one is to decompose the data based on the main factors of electrical load time series data. On the basis of decomposition and reconstitution of traditional time series additive model, the layer-upon-layer decreasing decomposition is applied for the reconstitution to enhance the forecasting accuracy. Then according to the characteristics of the decomposed data, suitable forecasting models are found to fit and forecast the sub-sequence. Through the effective decomposition of electrical load time series data and selection of proper forecasting models, the fitting ability and forecasting capacity can be well improved.

The second idea is to improve the forecasting accuracy of Generalized regression neural network (GRNN). The major forecasting model in this paper is GRNN, and genetic algorithm is utilized to optimize parameters in GRNN. Before that EMD is applied to eliminate the noises in the data. Thus, with the help of EMD and GA, the forecasting performance of GRNN can be greatly enhanced.

The experimental results show that compared with EMD-GA-WNN, GA-GRNN and EMD-GA-GRNN, the proposed hybrid model has a good forecasting effect for electrical load time series data with periodicity, trend and randomness. In practice, the DDH model based on data decomposition can reach a high forecasting accuracy, becoming a promising method in the future. Besides, if the time series show an obvious periodicity, trend and randomness, the hybrid model can

be applied commonly and effectively in other forecasting fields, such as product sales forecasting, tourism demand forecasting, warning and forecasting of flood, wind speed forecasting, traffic flow forecasting and so on.

However, with the development of technology and information, there are still many problems existing in the forecasting field. This paper mainly focuses on the study of a hybrid forecasting model based on time series decomposition and how to improve the forecasting accuracy, and further analysis can be conducted in the following aspects: (1) This paper ignores the influences of other factors on the electrical time series owing to the limitations of data collection; therefore, how to design a forecasting model and algorithm of multiple variables is a problem worth studying; (2) The forecasting techniques continue to improve, and there is no a perfect forecasting model that can deal well with all time series forecasting problems. Thus, it is necessary to develop new algorithms to achieve the future forecasting work; (3) Denoising of time series. The EMD method applied in this paper is just one type of denoising method, and other algorithms, such as Kalman filtering and wavelet packet decomposition, should be compared to EMD to select a better one.

Author Contributions: Yuqi Dong proposed the concept of this research and made overall guidance; Xuejiao Ma completed the whole paper; Chenchen Ma made tables and figures; Jianzhou Wang provided related data.

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

Abbreviations

ANNs	Artificial neural networks
SVM	Support vector machine
EA	Evolution algorithms
ARIMA	Auto regressive integrated moving average
SOM	Self-organizing map
GRNN	Generalized regression neural network
EMD	Empirical mode decomposition
IMF	Intrinsic mode function
WNN	Wavelet neural network
SES	Secondary exponential smoothing
ANFIS	Adaptive network-based fuzzy inference system
RBFNN	Radial basis function neural network
HS-ARTMAP	Hyper-spherical ARTMAP network
RBF	Radial basis function
GA	Genetic algorithm
DDH	Data Decomposition Hybrid Model
MAE	Mean absolute error
RMSE	Root mean square error
MAPE	Mean absolute percentage error
ME	Mean error
AE	Absolute error
RE	Relative error
BPNN	Back propagation neural network
diff-SARIMA	Difference seasonal autoregressive integrated moving average
ART	Adaptive resonance theory
WT	Wavelet transform

References

1. Yang, Y.; Chen, Y.H.; Wang, Y.C.; Li, C.H.; Li, L. Modelling a combined method based on ANFIS and neural network improved by DE algorithm: A case study for short-term electricity demand forecasting. *Appl. Soft Comput.* **2015**, *49*, 663–675. [CrossRef]
2. Li, S.; Goel, L.; Wang, P. An ensemble approach for short-term load forecasting by extreme learning machine. *Appl. Energy* **2016**, *170*, 22–29. [CrossRef]
3. Takeda, H.; Tamura, Y.; Sato, S. Using the ensemble Kalman filter for electricity load forecasting and analysis. *Energy* **2016**, *104*, 184–198. [CrossRef]

4.   Xiao, L.Y.; Wang, J.Z.; Hou, R.; Wu, J. A combined model based on data pre-analysis and weight coefficients optimization for electrical load forecasting. *Energy* **2015**, *82*, 524–549. [CrossRef]

5.   Ren, Y.; Suganthan, P.N.; Srikanth, N.; Amaratunga, G. Random vector functional link network for short-term electricity load demand forecasting. *Inf. Sci.* **2016**, *1*, 1078–1093. [CrossRef]

6.   Xiao, L.Y.; Shao, W.; Liang, T.L.; Wang, C. A combined model based on multiple seasonal patterns and modified firefly algorithm for electrical load forecasting. *Appl. Energy* **2016**, *167*, 135–153. [CrossRef]

7.   Xiao, L.Y.; Shao, W.; Wang, C.; Zhang, K.Q.; Lu, H.Y. Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting. *Appl. Energy* **2016**, *180*, 213–233. [CrossRef]

8.   Jiang, P.; Ma, X.J. A hybrid forecasting approach applied in the electrical power system based on data preprocessing, optimization and artificial intelligence algorithms. *Appl. Math. Model.* **2016**, *40*, 10631–10649. [CrossRef]

9.   Azadeh, A.; Ghaderi, S.F.; Sohrabkhani, S. A simulated-based neural network algorithm for forecasting electrical energy consumption in Iran. *Energy Policy* **2008**, *36*, 37–44. [CrossRef]

10.   Chang, P.C.; Fan, C.Y.; Lin, J.J. Monthly electricity demand forecasting based on a weighted evolving fuzzy neural network approach. *Electr. Power Syst. Res.* **2011**, *33*, 17–27. [CrossRef]

11.   Wang, J.Z.; Chi, D.Z.; Wu, J.; Lu, H.Y. Chaotic time serious method combined with particle swarm optimization and trend adjustment for electricity demand forecasting. *Expert Syst. Appl.* **2011**, *38*, 19–29. [CrossRef]

12.   Ghanbari, A.; Kazami, S.M.; Mehmanpazir, F.; Nakhostin, M.M. A cooperative ant colony optimization-genetic algorithm approach for construction of energy demand forecasting knowledge-based expert systems. *Knowl. Based Syst.* **2013**, *39*, 194–206. [CrossRef]

13.   Zhao, H.R.; Guo, S. An optimized grey model for annual power load forecasting. *Energy* **2016**, *107*, 272–286. [CrossRef]

14.   Zeng, B.; Li, C. Forecasting the natural gas demand in China using a self-adapting intelligent grey model. *Energy* **2016**, *112*, 810–825. [CrossRef]

15.   Kavaklioglu, K. Modeling and prediction of Turkeys electricity consumption using Support Vector Regression. *Appl. Energy* **2011**, *88*, 68–75. [CrossRef]

16.   Wang, J.Z.; Zhu, W.J.; Zhang, W.Y.; Sun, D.H. A trend fixed on firstly and seasonal adjustment model combined with the SVR for short-term forecasting of electricity demand. *Energy Policy* **2009**, *37*, 1–9. [CrossRef]

17.   Kucukali, S.; Baris, K. Turkeys shorts-term gross annual electricity demand forecast by fuzzy logic approach. *Energy Policy* **2010**, *38*, 38–45. [CrossRef]

18.   Nguyen, H.T.; Nabney, I.T. Short-term electricity demand and gas price forecasts using wavelet transforms and adaptive models. *Energy* **2010**, *35*, 13–21. [CrossRef]

19.   Park, D.C.; Sharkawi, M.A.; Marks, R.J. Electric load forecasting using a neural network. *IEEE Trans. Power Syst.* **1991**, *6*, 442–449. [CrossRef]

20.   Wang, J.Z.; Liu, F.; Song, Y.L.; Zhao, J. A novel model: Dynamic choice artificial neural network (DCANN) for an electricity price forecasting system. *Appl. Soft Comput.* **2016**, *48*, 281–297. [CrossRef]

21.   Yu, F.; Xu, X.Z. A short-term load forecasting model of natural gas based on optimized genetic algorithm and improved BP neural network. *Appl. Energy* **2014**, *134*, 102–113. [CrossRef]

22.   Hernandez, L.; Baladron, C.; Aguiar, J.M. Short-term load forecasting for microgrids based on artificial neural networks. *Energies* **2013**, *6*, 1385–1408. [CrossRef]

23.   Pandian, S.C.; Duraiswamy, K.D.; Rajan, C.C.; Kanagaraj, N. Fuzzy approach for short term load forecasting. *Electr. Power Syst. Res.* **2006**, *76*, 541–548. [CrossRef]

24.   Pai, P.F. Hybrid ellipsoidal fuzzy systems in forecasting regional electricity loads. *Energy Convers. Manag.* **2006**, *47*, 2283–2289. [CrossRef]

25.   Liu, L.W.; Zong, H.J.; Zhao, E.D.; Chen, C.X.; Wang, J.Z. Can China realize its carbon emission reduction goal in 2020: From the perspective of thermal power development. *Appl. Energy* **2014**, *124*, 199–212. [CrossRef]

26.   Xiao, L.; Wang, J.Z.; Dong, Y.; Wu, J. Combined forecasting models for wind energy forecasting: A case stud in China. *Renew. Sustain. Energy Rev.* **2015**, *44*, 271–288. [CrossRef]

27.   Wang, J.J.; Wang, J.Z.; Li, Y.N.; Zhu, S.L.; Zhao, J. Techniques of applying wavelet de-noising into a combined model for short-term load forecasting. *Int. J. Electr. Power* **2014**, *62*, 816–824. [CrossRef]

28. Azimi, R.; Ghofrani, M.; Ghayekhloo, M. A hybrid wind power forecasting model based on data mining and wavelets analysis. *Energy Convers. Manag.* **2016**, *127*, 208–225. [CrossRef]

29. Khashei, M.; Bijari, M. An artificial neural network model for time series forecasting. *Expert Syst.* **2010**, *37*, 479–489. [CrossRef]

30. Shukur, O.B.; Lee, M.H. Daily wind speed forecasting though hybrid KF-ANN model based on ARIMA. *Renew. Energy* **2015**, *76*, 637–647. [CrossRef]

31. Niu, D.; Shi, H.; Wu, D.D. Short-term load forecasting using bayesian neural networks learned by hybrid Monte Carlo algorithm. *Appl. Soft Comput.* **2012**, *12*, 1822–1827. [CrossRef]

32. Lu, C.J.; Wang, Y.W. Combining independent component analysis and growing hierarchical self-organizing maps with support vector regression in product demand forecasting. *Int. J. Prod. Econ.* **2010**, *128*, 603–613. [CrossRef]

33. Okumus, I.; Dinler, A. Current status of wind energy forecasting and a hybrid method for hourly predictions. *Energy Convers. Manag.* **2016**, *123*, 362–371. [CrossRef]

34. Che, J.; Wang, J. Short-term electricity prices forecasting based on support vector regression and auto-regressive integrated moving average modeling. *Energy Convers. Manag.* **2010**, *51*, 1911–1917. [CrossRef]

35. Meng, A.B.; Ge, J.F.; Yin, H.; Chen, S.Z. Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm. *Energy Convers. Manag.* **2016**, *114*, 75–88. [CrossRef]

36. Elvira, L.N. Annual Electrical Peak Load Forecasting Methods with Measures of Prediction Error 2002. Ph.D. Thesis, Arizona State University, Tempe, AZ, USA, 2002.

37. Wang, J.Z.; Ma, X.L.; Wu, J.; Dong, Y. Optimization models based on GM (1,1) and seasonal fluctuation for electricity demand forecasting. *Int. J. Electr. Power* **2012**, *43*, 109–117. [CrossRef]

38. Guo, Z.H.; Wu, J.; Lu, H.Y.; Wang, J.Z. A case study on a hybrid wind speed forecasting method using BP neural network. *Knowl. Based Syst.* **2011**, *2*, 1048–1056. [CrossRef]

39. Peng, L.L.; Fan, G.F.; Huang, M.L.; Hong, W.C. Hybridizing DEMD and quantum PSO with SVR in electric load forecasting. *Energies* **2016**, *9*. [CrossRef]

40. Fan, G.F.; Peng, L.L.; Hong, W.C.; Sun, F. Electric load forecasting by the SVR model with differential empirical mode decomposition and auto regression. *Neurocomputing* **2016**, *173*, 958–970. [CrossRef]

41. Liu, H.; Tian, H.Q.; Liang, X.F.; Li, Y.F. New wind speed forecasting approaches using fast ensemble empirical model decomposition, genetic algorithm, mind evolutionary algorithm and artificial neural networks. *Renew. Energy* **2015**, *83*, 1066–1075. [CrossRef]

42. Niu, D.X.; Cao, S.H.; Zhao, L.; Zhang, W.W. *Methods for Electrical Load Forecasting and Application*; China Electric Power Press: Beijing, China, 1998.

43. Liu, L.; Wang, Q.R.; Wang, J.Z.; Liu, M. A rolling grey model optimized by particle swarm optimization in economic prediction. *Comput. Intell.* **2014**, *32*, 391–419. [CrossRef]

44. Yuan, C.; Wang, J.Z.; Tang, Y.; Yang, Y.C. An efficient approach for electrical load forecasting using distributed ART (adaptive resonance theory) & HS-ARTMAP (Hyper-spherical ARTMAP network) neural network. *Energy* **2011**, *36*, 1340–1350.

45. Hooshmand, R.A.; Amooshahi, H.; Parastegari, M. A hybrid intelligent algorithm based short-term load forecasting approach. *Int. J. Electr. Power* **2013**, *45*, 313–324. [CrossRef]