A novel decomposition-based ensemble broad learning system for short-term load forecasting

Yihan Tian

School of Automation, Central South University, Changsha, 410083, China;

Yihan_Tian@163.com

Abstract. Load forecasting of the power system plays a key role in the production planning and actual operation scheduling of power systems. However, as the power system becomes larger and more complex, it is very important to propose a forecasting model with high accuracy and low computational cost. In this paper, a novel short-term load forecasting model is proposed, which combines the empirical wavelet transform (EWT) and the broad learning system (BLS). The advantage of EWT is that it can decompose the signal into multiple local frequency bands, select the local wavelet function adaptively, and overcome the modal mixing problem caused by the discontinuity of the signal time-frequency scale. While, the advantage of BLS is that it allows the model to learn more by stretching the width, and uses ridge regression to make it less time consuming and faster. To verify the effectiveness, the model is compared with some other state-of-the-art methods, and several performance estimations indicate that the model has high accuracy and low computational cost.

Keywords: Load Forecasting; Empirical Wavelet Transform; Broad Learning System.

1. Introduction

Electric power is the lifeblood of modern social development and the basic guarantee of economic development, people's life, and social stability [1][2][3]. Electric energy is difficult to store. To ensure the stable operation of the power network, the power generation side and load side of the power network must maintain a real-time power balance, which requires accurate prediction of future power demand. High-precision short-term load forecasting can not only maintain the reliable operation of the power system, but also reduce unnecessary power generation, improve energy efficiency, and reduce the operation cost of the power network. However, with the "dual carbon" target proposed [5], the improvement of the power network working mode, and the continuous complexity of the power system, the difficulty of short-term load forecasting greatly increases, which challenges the traditional load forecasting model [5][6].

With the increasing importance of load forecasting in the scientific community, domestic and foreign scholars have carried out a large number of theoretical research on load forecasting. They proposed a variety of prediction models and constantly verified and optimized them, so the prediction results have been greatly improved. According to the development history, short-term load forecasting can be divided into three categories: traditional forecasting methods, single forecasting models, and hybrid forecasting models.

Traditional forecasting methods are mainly classified into time series method, regression analysis method, and Kalman filter method. Scientists have proposed a variety of forecasting models using these traditional methods. For example, Orang et al. proposed a novel univariate time series forecasting technique, called randomized high order FCM models (R-HFCM). The obtained results confirm the proposed R-HFCM model is superior to the other methods [7]. Haida et al. proposed a regression-based daily peak load forecasting method and a conversion technique, which reduced the large forecasting error caused by the nonlinear characteristics [8]. Trudnowski et al. proposed a strategy for developing a very short-term load predictor using slow and fast Kalman estimators with an hourly forecaster. The design strategy presents better performance and sensitivity [9].

Although traditional methods can accurately predict linear stable series, they still have many drawbacks in some cases. For example, the time series method is suitable for cases where the data has continuous characteristics and changes slowly, but can’t be adjusted effectively in time when the load changes rapidly due to external disturbances. The regression analysis method is influenced by more factors, and can't accurately predict when the data fluctuates greatly. The Kalman filter method can get more ideal prediction results at the prediction point with little noise, but when the noise becomes large, the prediction accuracy fluctuates more.

Machine learning is the study of using experience or data to improve algorithms. The algorithm allows machines to learn and find patterns from a large amount of historical data, get some kind of model, and use this model to predict the future. The artificial neural network (ANN) is a kind of information-processing system simulating the human brain, which also has the functions of information processing, learning, and memory. It has a good effect on pattern recognition and processing of nonlinear problems. As scholars at home and abroad pay more and more attention to artificial intelligence, these methods have been applied to load forecasting. Dash et al. applied an adaptive Kalman-filter-based learning algorithm into ANN for forecasting weather-sensitive loads. The results show that the method has the advantages of short learning time, fast convergence, and self-adaptive capability [10]. Shi et al. obtained a smaller root mean square error than the autoregressive integrated moving average model (ARIMA) and support vector regression (SVR) by using a novel pooling-based deep recurrent neural network [11]. Kong et al. addressed the fluctuation problem with a deep learning prediction framework based on long short-term memory (LSTM). It is shown that prediction accuracy can be significantly improved by adding device measurements to the training data [12]. Deng et al. proposed a novel model multi-scale convolutional neural network with time-cognition (TCMS-CNN) and obtained good prediction results by integrating MS-CNN and periodic coding into the proposed TCMS-CNN model [13]. By using an advanced backpropagation algorithm, Soeb et al. proposed an advanced algorithm in load forecasting due to it can update the hidden layer by adaptation mechanism [14]. Sun et al. adopted a novel probabilistic day-ahead net load forecasting method to capture both epistemic uncertainty and aleatoric uncertainty using Bayesian deep learning [15]. Bacanin et al. proposed a tuned LSTM model for multivariate time series forecasting of electric loads. It is verified that parameter tuning leads to better results when using metaheuristics in all cases [16].

Single machine learning prediction models have achieved better prediction performance. But with the rapid development of various industries in society, the increasing electricity demand, and the power system becoming large and complex, has made load forecasting work more and more difficult. People's demand for the reliability of power supply has increased, and it is difficult to meet the requirements of load forecasting in today's society by relying on a single forecasting method. To further improve the prediction accuracy of the model, it is necessary to analyze the advantages of each prediction model and propose more practical and efficient prediction models.

The hybrid model algorithm is a new algorithm that combines at least two prediction models to achieve the goal of improving prediction accuracy. Many domestic and foreign scholars have also studied it and proposed innovative methods. Alipour et al. succeeded in improving the net-load prediction accuracy by a deep neural network model that performs a wavelet transform on the inputs of the model [17]. Khwaja et al. designed integrated machine learning based on ANNs to improve short-term load forecasting by training bagged-boosted ANNs [18]. Müller et al. obtained excellent results compared to traditional methods by using singular spectrum analysis (SSA) combined with fuzzy an adaptive resonance theory map (ARTMAP) ANN for noise removal [19]. Li et al. proposed an improved short-term load forecasting method, the main component of which is the frequency decomposition of the load followed by different methods for forecasting different frequency components [20]. Deng et al. designed a unified quantile regression deep neural network with time-cognition, which used a convolutional neural network(CNN) with multiscale convolution and a novel periodical coding method for forecasting [21]. Adil et al. finally obtained better prediction results by using a hybrid model of mutual information (MI) and feature weighting algorithms (ReliefF), combined with redundancy and regression processes [22]. Shi et al. proposed a very short-term bus load forecasting model based on phase space reconstruction (PSR) and deep belief network (DBN). The results show that it has better adaptability in the case of high load fluctuations [23]. Chen et al.

adopted a day-ahead aggregated load forecasting method based on two-terminal sparse coding and deep neural network fusion, which can overcome the challenges posed by high-dimensional data [24]. Ribeiro et al. proposed a framework for the construction of wavelet network ensembles for short-term load forecasting with high prediction accuracy [25]. Zahid et al. used deep learning and data mining techniques for power load and price forecasting, and the model has good performance [26]. Wang et al. proposed a method based on attention mechanism (AM), rolling update (RU), and bi-directional long short-term memory neural network (Bi-LSTM) to improve the prediction accuracy [27].

As can be seen, deep structural neural networks have achieved substantial improvements in prediction accuracy, but most of them are plagued by extremely time-consuming training processes. One of the main reasons is that all of these networks are structurally complex and involve a large number of hyperparameters. On the one hand, the complexity makes it extremely difficult to analyze the structure theoretically. On the other hand, the models have to continuously increase the number of network layers or adjust the number of parameters to obtain higher accuracy in the application. Thus, it is of great practical importance to propose a prediction model with high accuracy and low cost. In response to the above problems, a novel prediction model is proposed in this paper.

Specifically, Empirical wavelet transform (EWT) integrates the adaptive decomposition concept of empirical mode decomposition (EMD) and the tight support framework of wavelet transform theory and provides a new adaptive time-frequency analysis idea for signal processing. Compared with EMD, EWT can select the frequency band adaptively and overcome the modal mixing problem caused by the discontinuity of the signal time-frequency scale.

Furthermore, a broad learning system (BLS) is designed based on the idea of using mapped features as input to the random vector functional-link neural network (RVFLNN). BLS allows the model to learn more by stretching the width. Compared with deep learning, which is often trapped in a local optimum due to repeated training, the advantage of BLS is very obvious.

By combining EWT and BLS reasonably, this paper proposes a novel decomposition-based ensemble broad learning system for short-term load forecasting, named EWT-BLS. Experimental results show that EWT-BLS has higher prediction accuracy and lower computational cost compared with other comparative models.

The rest of the paper is described according to the following structure. In section 2, the framework and related theories of the proposed model are introduced. Section 3 is data analysis and parameter settings. Section 4 is the case study with discussions. Finally, section 5 gives a conclusion and future work.

2. Model Framework

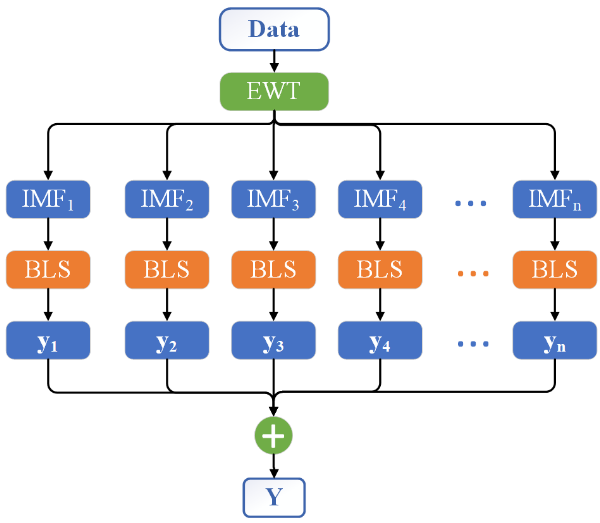

Fig. 1 The overall framework of the EWT-BLS model and its specific processes.

Fig. 1 shows the overall model framework of the EWT-BLS model proposed in this paper. Firstly, the original multiplexed data are decomposed into several subsequences using EWT, which decomposes the originally unbalanced and nonlinear electric load data into several stable and linear subsequences. Then, each subsequence is trained separately using BLS, and the prediction results of the respective sequence are derived. Finally, the obtained prediction results of each component are superimposed together as the final prediction result. More details about EWT and BLS are given as follows.

2.1 Empirical wavelet transform

The principle of the EWT is to divide the Fourier spectrum of the signal into successive intervals, then construct a wavelet filter set on each interval for filtering, and finally obtain a set of amplitude modulation (AM) and frequency modulation (FM) components by signal reconstruction.

2.1.1 EWT band division method

For a given signal $f(\mathrm{t})$, it is first Fourier transform and obtains a Fourier spectrum normalized in the range of 2π. According to the Shannon sampling theorem, only the signal characteristics on $[0,\pi]$ are discussed in the analysis process. Therefore, the Fourier spectrum support interval is defined in the interval from 0 to π.

Assuming that the signal is composed of N single-component components, the support interval of the Fourier spectrum is partitioned into N consecutive segments during processing, with a total of N + 1 boundaries. (ω_n denotes the boundaries between the segments, $\omega_0 = 0, \omega_N = \pi$)

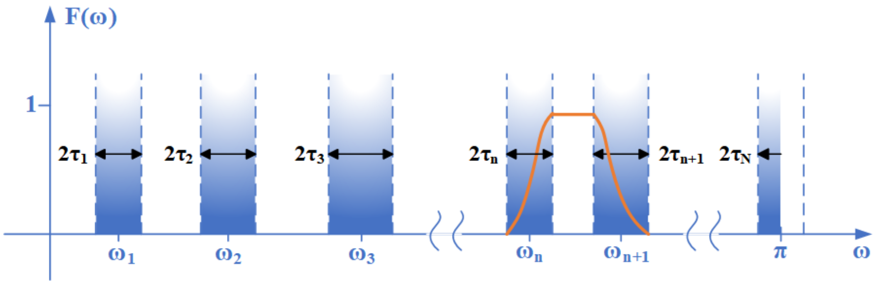

Fig. 2 Schematic diagram of Fourier axis segmentation.

Λ_n denotes the division result, which is expressed by the following equation:

$$\Lambda_n = [\omega_{n-1}, \omega_n], \quad n = 1,2,\ldots,N \tag{1}$$

The support interval of the entire Fourier axis is expressed as follows:

$$\sum_{n=1}^{N} \Lambda_n = [0,\pi] \tag{2}$$

The method of determining ω_n is: for the Fourier spectrum of signal $f(\mathrm{t})$, assume that the algorithm finds M maxima in the spectrum and arranges them in descending order, at which point Gilles points out the existence of two cases:

If $M \geq N$, it means that the number of selected maxima is large enough to provide a basis for spectrum segmentation. In this case, the first N-1 maxima are retained.

If $M < N$, it means the preset value of the number of mode components in the signal is large. In this case, the M maxima found are kept, and the parameter N is reset to M.

Thus, for the signal $f(\mathrm{t})$, let the first N maxima correspond to the angular frequency $\Omega_n (n = 1,\ldots,N)$ in the frequency domain, then the band division boundary of EWT is $\omega_n = \frac{\Omega_{n+1}+\Omega_n}{2}$. To facilitate the construction of the filter in the subsequent process, transition segments of width $T_n = 2\tau_n$ are defined, which are centered on the boundary ω_n. ($\tau_n = \gamma\omega_n (0 < \gamma < 1)$, $\gamma < \min_n \left(\frac{\omega_{n+1}-\omega_n}{\omega_{n+1}+\omega_n}\right)$)

In general, it is difficult to guess the number N of patterns for a segment of the signal for which no a priori information is available. For this problem, Gilles proposed a simple method for estimating the value of N: detect M maxima in the Fourier spectrum of the signal and form them into a set $\{M_i\}_{k=1}^{M}$. The M_i in the set are arranged in descending order, so that, $M_1 > M_2 > M_3 > \cdots > M_M$. The threshold value is set as $[M_M + \alpha(M_1 - M_M)]$ (α is the relative amplitude ratio). The number of maxima greater than the threshold is the value of N. Then, the above-mentioned spectrum segmentation is to obtain the results of the signal band division.

2.1.2 EWT algorithm theory

The empirical wavelets are band-pass filter banks defined on the interval Λ_n. It is carried out using the idea of constructing Littlewood-Paley and Meyer wavelets. For $n > 0$, the empirical wavelet function $\psi_n(\omega)$ and the empirical scale function $\varphi_n(\omega)$ are expressed by the following two equations:

$$\psi_n(\omega) = \begin{cases} 1, & \omega_n + \tau_n \le |\omega| \le \omega_{n+1} - \tau_{n+1} \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_{n+1}}(|\omega| - \omega_{n+1} + \tau_{n+1})\right)\right], & \omega_{n+1} - \tau_{n+1} \le |\omega| \le \omega_{n+1} + \tau_{n+1} \\ \sin\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}(|\omega| - \omega_n + \tau_n)\right)\right], & \omega_n - \tau_n \le |\omega| \le \omega_n + \tau_n \\ 0, & \text{others} \end{cases} \tag{3}$$

$$\varphi_n(\omega) = \begin{cases} 1, & |\omega| \le \omega_n - \tau_n \\ \cos\left[\frac{\pi}{2}\beta\left(\frac{1}{2\tau_n}(|\omega| - \omega_n + \tau_n)\right)\right], & \omega_n - \tau_n \le |\omega| \le \omega_n + \tau_n \\ 0, & \text{others} \end{cases} \tag{4}$$

In the equations, $\beta(x) = x^4(35 - 84x + 70x^2 - 20x^3)$

According to the above filter construction method, only the support interval of $[0, \pi]$is considered. The wavelet filter set shown in Fig. 3 can be built. The first interval is the filter determined by the empirical scale function, and the other intervals are the band-pass filters determined by the empirical wavelet function. The EWT is performed on this basis.

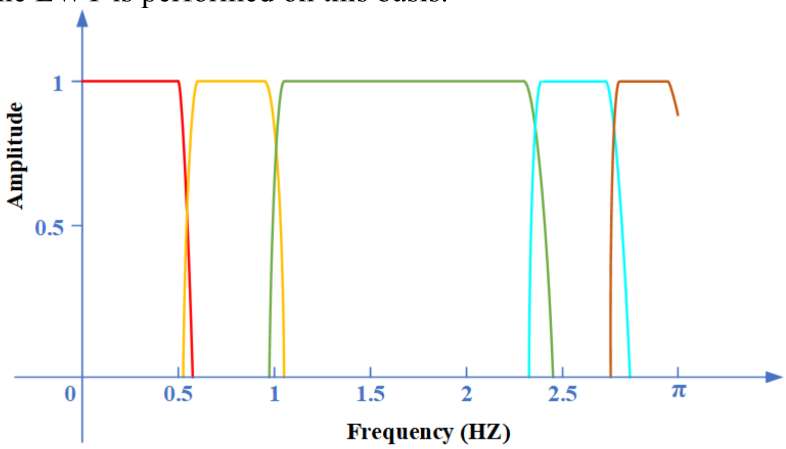

Fig. 3 Wavelet filter bank.

Drawing on the idea of the classical wavelet transform, the empirical wavelet coefficients constructed by Gilles are generated by the inner product. The detail coefficients $W_f^e(n,t)$ are generated by the inner product of the empirical wavelet function ψ_n and the signal $f(t)$. It can be written as the following equation:

$$W_f^e(n,t) = \langle f, \psi_n \rangle = \int f(\tau)\,\overline{\psi_n(\tau - t)}\,d\tau = F^{-1}\left[f(\omega)\overline{\psi_n(\omega)}\right] \tag{5}$$

The approximation coefficient $W_f^e(0,t)$, generated by the inner product of the empirical scale function φ_1 and the signal $f(t)$, can be written as:

$$W_f^e(0,t) = \langle f, \varphi_1 \rangle = \int f(\tau)\, \overline{\varphi_1(\tau - t)}\, d\tau = F^{-1}\left[f(\omega) \overline{\varphi_1(\omega)} \right] \tag{6}$$

In the equation, $\psi_n(\omega)$ and $\varphi_n(\omega)$ are the Fourier transform of $\psi_n(t)$ and $\varphi_n(t)$. The Fourier transform and inverse transform are denoted as $F[\cdot]$ and $F^{-1}[\cdot]$.

So that, the resulting reconfiguration expression for the signal $f(t)$ is:

$$f(t) = W_f^s(0,t) * \phi_1(t) + \sum_{n-1}^{N} W_f^s(n,t) * \varphi_n(t)$$

$$= \left(W_f^s(0,\omega)\phi_1(\omega) + \sum_{n-1}^{N} W_f^s(n,\omega) * \varphi_n(\omega) \right) \tag{7}$$

After EWT processing, the signal $f(t)$ is decomposed to obtain the AM/FM single-component $f_k(t)$ $(k = 1,2,3\ldots)$ with frequencies ranging from low to high:

$$f_0(t) = W_f^e(0,t) * \varphi_1(t) \tag{8}$$

$$f_k(t) = W_f^e(k,t) * \psi_k(t) \tag{9}$$

2.2 Broad learning system

2.2.1 Random vector functional-link neural network (RVFLNN)

The predecessor of BLS is the random vector functional-link neural network (RVFLNN) that has been studied for a long time. The RVFLNN is formed by adding a direct connection from the input layer to the output layer in a single layer feedforward neural network (SLFN).

a. RVFLNN network structure: the first layer is the input layer, the second layer is the enhancement layer, and the third layer is the output layer.

b. RVFLNN network connection: the input layer to the enhancement layer is nonlinear transformations, and the enhancement layer to the output layer and the input layer to the output layer have only linear transformations.

In RVFLNN, only the enhancement layer is a neural network unit in the real sense, because only it carries the activation function. All other parts of the network are linear.

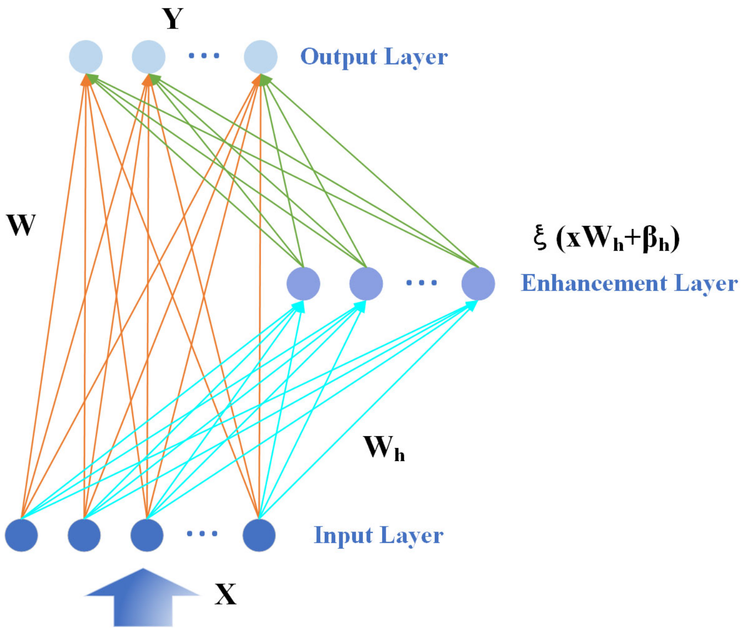

Fig. 4 RVFLNN network structure.

c. RVFLNN network calculation method:

When we line up the enhancement layer and the input layer, they are considered as one, and the network becomes a linear transformation from A to Y. The weight matrix W corresponding to the linear transformation is the linear connection between the "input layer plus the enhancement layer" and the "output layer". If we ignore the connections between the input layer and the enhancement layer, initialize these connections randomly, and fix the weights between the input layer and the enhancement layer, then the training of the whole network is to find the transformation W from A to Y:

$$W = A^{-1}Y \tag{10}$$

If the input X is known, the enhancement layer A is known. If the labels of the training data are known, Y is known.

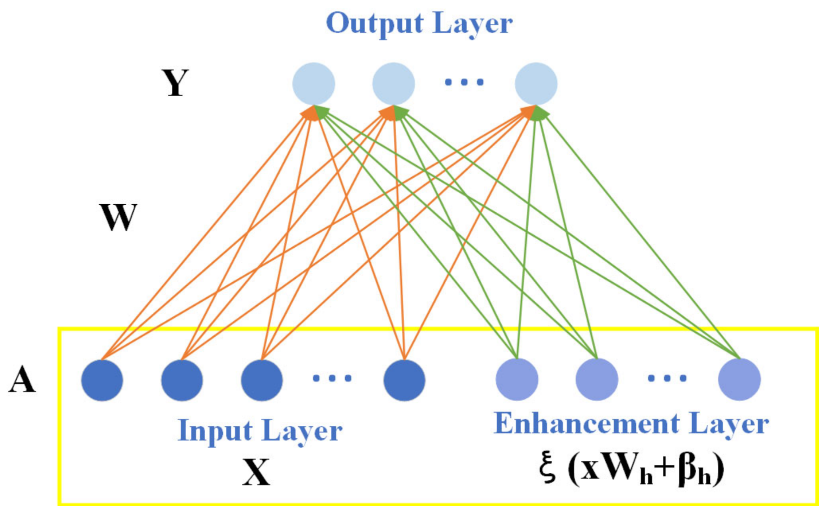

Fig. 5 RVFLNN network calculation method.

RVFLNN is proved to be used to approximate any continuous function on a tight set, and its nonlinear approximation ability is reflected in the nonlinear activation function of the enhancement layer. As long as the number of enhancement layer cells is large enough, the number of nonlinearities can be many.

2.2.2 Broad learning system (RVFLNN)

BLS makes a little improvement to the input layer based on RVFLNN. Instead of using the original data directly as the input layer, it first does some transformations on the data, which is equivalent to feature extraction, and uses the changed features as the input layer of the original RVFLNN. Broad learning can be trained using features extracted from other models. So, it can be assembled with other machine learning algorithms.

a. BLS network structure: the first layer is the feature layer, the second layer is the enhancement layer, and the third layer is the output layer.

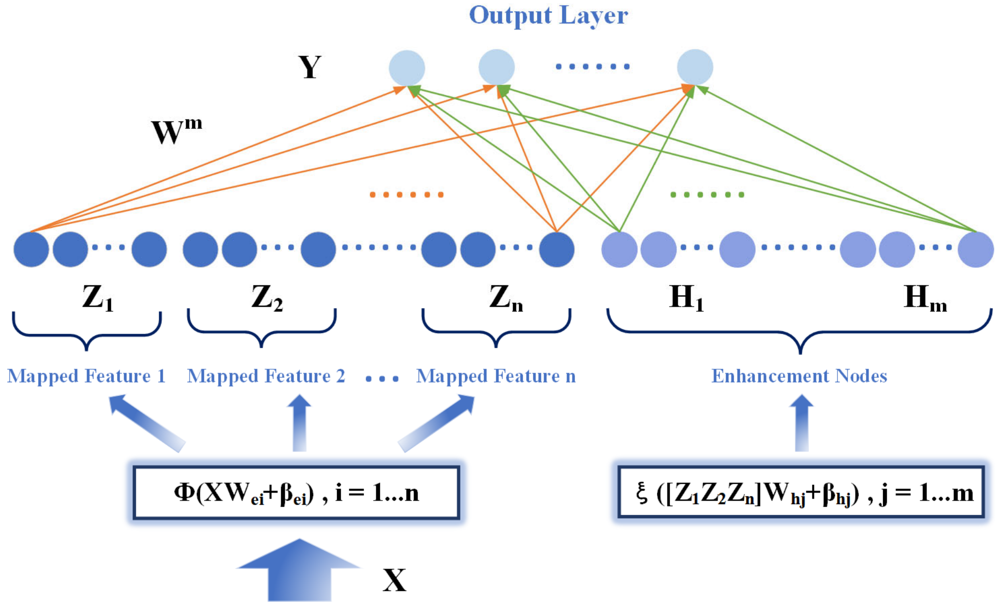

Fig. 6 BLS network structure.

b. BLS network calculation method:

When the features Z are given, it is straightforward to calculate the enhancement layer *H*. Then, the feature and enhancement layers are combined into $A = [Z|H]$ (the vertical line in it indicates the combination into one row). Since the labels *Y* of the training data are known, it is sufficient to calculate the weights *W*. For the actual calculation, ridge regression is used to solve the weight matrix. *W* can be solved by the following optimization problem: ($\sigma_1 = \sigma_2 = v = u = 2$)

$$\underset{W}{\arg\min}: \ \|AW - Y\|_v^{\sigma_1} + \lambda \, \|W\|_u^{\sigma_2} \tag{11}$$

So that,

$$W = (\lambda I + AA^T)^{-1}A^TY \tag{12}$$

Specifically, we have that:

$$A^+ = \lim_{\lambda \to 0}(\lambda I + AA^T)^{-1}A^T \tag{13}$$

The above process is a one-step process. When the data are fixed and the model structure is fixed, the optimal parameter *W* can be found directly.

3. Data analysis and parameter settings

3.1 Data sets

The data set, which is used as the experimental sample in this paper, is the historical load data of New South Wales (NSW), Australia, in 2009. The sampling interval of the data set is 30 min, so 48 load data samples are contained in one day. The data from January and February 2009 (2880 data in total) are used as the training data set, and the data from March 2009 (1392 data in total) are used as the test data set.

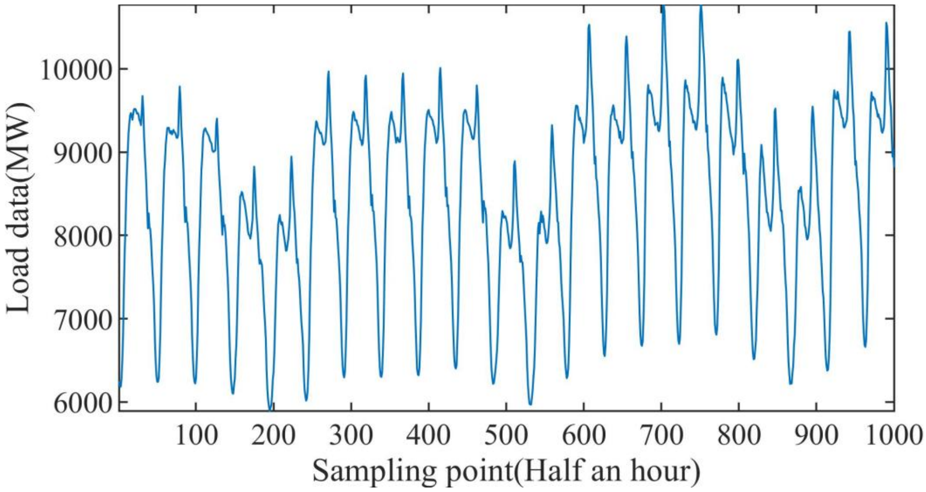

Fig. 7 Data of 1,000 sampling points.

Table 1. The statistical data of the historical load data of NSW.

	Max	Min	Mean	Std
NSW	14274.15	5498.36	8894.00	1409.05

Fig. 7 shows the data from 1,000 sampling points. From the figure, we can find that the load data fluctuate greatly at the peak. Table 1 shows the statistical data of this data set, including the maximum, minimum, average, and standard deviation. To avoid the negative impact of singular sample data on the prediction accuracy, the load data are normalized and restricted to the range of [0,1] before the experiment. The normalization formula is shown in (14).

$$\tilde{y}_i = \frac{y_{max} - y_i}{y_{max} - y_{min}} \tag{14}$$

Here, $\tilde{y}_i$ is the normalized result, y_i is the load data at a certain moment, and y_{max} and y_{min} are the maximum and minimum values in the load data, respectively.

3.2 Performance estimation

The accuracy of the prediction result needs to be evaluated by the evaluation function. This paper uses common evaluation methods in load forecasting to assess the prediction performance, including the root mean square error (RMSE), the mean square error (MSE), the mean absolute error (MAE), and the mean absolute percentage error (MAPE). They are defined in formulas (15), (16), (17), and (18).

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} \tag{15}$$

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 \tag{16}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \tag{17}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{18}$$

where $\hat{y}_i$ is the predicted data, y_i is the real data, and n is the total number of test samples. For these evaluation indicators, the smaller the value is, the higher the accuracy of the prediction.

3.3 Parameter settings

Table 2. The parameter settings of different models.

Model	Optimal parameters
ANN	$\mathrm{n_h} = 200, \mathrm{m_i} = 10, \mathrm{eta} = 0.01$
RBFNN	$\mathrm{f_{RBF}} = \mathrm{Gaussian}, \mathrm{s_{RBF}} = 50$
RVFL	$\mathrm{n_e} = 10{,}000, \mathrm{a_f} = \mathrm{Sigmoid}, \mathrm{DL} = \mathrm{ture}, \mathrm{r_m} = \mathrm{uniform}$
EMD-BLS [28]	$\mathrm{n_f} = 24, \mathrm{n_e} = 15$
SWT-LSTM [29]	$\mathrm{n_h} = 200, \mathrm{eta} = 0.01$
EWT-BLS	$\mathrm{n_f} = 24, \mathrm{n_e} = 15$

*Parameter Description:
n_h- the number of hidden nodes; m_i- the maximum number of iterations;
eta- learning rate;
f_{RBF}- radial basis functions; s_{RBF}- the spread of radial basis functions;
n_e- the number of enhancement nodes; a_f- activation function;
DL- whether to have the direct link between the input layer and output layer;
r_m- randomization methods; n_f- the number of feature nodes.

To verify the effectiveness of the model proposed in this paper, it is compared with some state-of-the-art models. Among them, the single forecasting models include ANN, the radial basis function neural network (RBFNN), and the random vector functional link network (RVFL). The decomposition-based hybrid models include EMD-BLS[28] and the stationary wavelet transform (SWT)-LSTM[29], which have been published in previous papers. We optimized the parameters for all these models, and the specific parameter settings for the comparison models and the EWT-BLS model are shown in Table 2. The number of feature nodes in the EWT-BLS model is set to 24, while the number of enhancement nodes is set to 15.

In addition, we are using the load data of a past day to predict the load data of a future half hour. The sampling interval of the data set is 30 minutes, while 48 load data samples are contained in one day, so the dimension of our model input data is 48 and the dimension of the output data is 1.

4. Case study

4.1 Results

The performance of each model is given in Table 3. Comparing the performance estimations of RMSE, MSE, MAE, and MAPE of different models, it can be found that the ANN model of the largest prediction error values, got 286.30, 81968.52, 2.76, and 228.98. The prediction error values of RVFL are slightly higher than those of RVFNN and SWT-LSTM. While, the EMD-BLS model got 43.27, 1872.42, 0.40, and 33.93, which indicates that its prediction error values are significantly reduced compared to those of other methods. So, as a whole, the hybrid models are better than the single models, in terms of load forecasting.

Although the performance of the existing hybrid models has been improved substantially, they still fall short of the ideal prediction accuracy requirements. However, the EWT-BLS model proposed in this paper has the smallest prediction error values and is significantly different from other hybrid models. Its performance indexes are 6.76, 45.72, 0.07, and 5.42. By comparing the accuracy of EMD-BLS and EWT-BLS, we can find that the decomposition method of EWT is significantly better than EMD.

In terms of total time, although the total time of EWT-BLS is slightly higher than some other methods, its response speed is significantly improved compared with SWT-LSTM. In addition, the advantages of high prediction accuracy and low cost of EWT-BLS can compensate for this drawback. In summary, considering various indicators, the EWT-BLS model proposed in this paper achieves the expected goal of high accuracy and low computational cost.

Table 3. Different models for load forecasting on NSW, Australia, in 2009 datasets. Performance estimation: RMSE, MSE, MAE, MAPE, and Total time.

	RMSE	MSE	MAE	MAPE	Total time
ANN	286.30	81968.52	2.76	228.98	0.88
RBFNN	124.63	15531.97	1.07	92.19	2.85
RVFL	178.02	31692.62	1.64	136.09	0.09
EMD-BLS [28]	43.27	1872.42	0.40	33.93	0.92
SWT-LSTM [29]	133.35	17781.74	1.29	108.54	73.07
EWT-BLS	6.76	45.72	0.07	5.42	8.31

Furthermore, the predictions of different models for load forecasting on the datasets are shown in Fig. 8. It clearly shows that the ANN model has the largest deviation with great inertia and very serious lag when the real data fluctuation is large. The other models have more or fewer deviations at the peaks and troughs of the waveform, and these models can hardly follow the data changes accurately and quickly. While the EMD-BLS model proposed in this paper can always follow real data fluctuation very well.

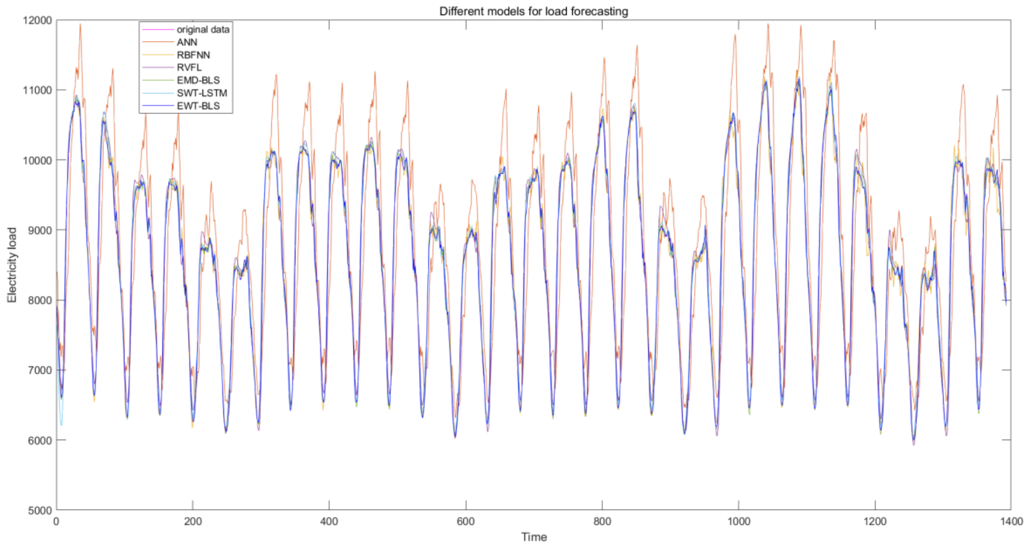

Fig. 8 Different models for load forecasting on NSW, Australia, in 2009 datasets.

4.2 Discussion

Through the prediction results of the datasets and model prediction curves, we can find that by innovatively combining two models with good performance to form a hybrid method based on decomposition, they can show their respective advantages well and consume less time to obtain more accurate prediction results.

Specially the reasons have been given as follows:

EWT has good localization, adaptivity, and robustness. EWT can decompose the signal into multiple local frequency bands, select the local wavelet function adaptively, overcome the modal mixing problem caused by the discontinuity of signal time-frequency scale, and handle the noise in the signal well.

BLS provides an alternative approach to deep learning networks. BLS allows the model to learn more by stretching the width. In addition, since BLS uses ridge regression instead of backpropagation, it consumes less time and is faster compared to deep learning.

The prediction results show that the EWT-BLS model has the smallest prediction error values and a significant gap compared with other methods. In addition, the EWT-BLS model can follow well when the data fluctuates a lot. Thus, the EWT-BLS model proposed in this paper has high accuracy and low computational cost.

5. Conclusion and Future Work

Electricity load forecasting directly affects the quality level of power system planning, and may also influence the formulation of strategic goals for energy development and utilization, so it has been a popular issue. At present, how to design a high-accuracy and low-cost model is a difficult problem. To solve this difficulty, this paper reasonably integrates two methods, EWT and BLS, to form a hybrid method based on decomposition, named EWT-BLS. Through experimental verification, the model has high accuracy and a low computational cost.

Although the EWT-BLS model already has a good performance, there is still room for optimization in terms of time consumption. The next step should focus on the parameter tuning of the model to get a faster model response speed. This will contribute to the stable operation of the power network and improve the economic and social benefits of the power system. Further, the model can be tested on additional datasets to validate the model performance and continue to optimize the model. For example, the model can be used to predict the spread of infectious diseases, which is important for early detection of disease trends, disease prevention and treatment, and health decision-making. This model can also be used to predict traffic flow, which is crucial for reducing urban traffic congestion, improving urban traffic efficiency, and promoting the synergistic development of smart transportation and smart cities.

References

[1] Niu, Dongxiao, Yongli Wang, and Desheng Dash Wu. "Power load forecasting using support vector machine and ant colony optimization." Expert systems with Applications 37.3 (2010): 2531-2539.

[2] Amjady, Nima. "Short-term hourly load forecasting using time-series modeling with peak load estimation capability." IEEE Transactions on power systems 16.3 (2001): 498-505.

[3] Talavera-Llames, R., et al. "MV-kWNN: A novel multivariate and multi-output weighted nearest neighbours algorithm for big data time series forecasting." Neurocomputing 353 (2019): 56-73.

[4] Wang, Yao, et al. "Carbon peak and carbon neutrality in China: Goals, implementation path and prospects." China Geology 4.4 (2021): 720-746.

[5] Zhang, Qian, et al. "Short-term load forecasting based on frequency domain decomposition and deep learning." Mathematical Problems in Engineering 2020 (2020): 1-9.

[6] Saviozzi, M., S. Massucco, and F. Silvestro. "Implementation of advanced functionalities for Distribution Management Systems: Load forecasting and modeling through Artificial Neural Networks ensembles." Electric Power Systems Research 167 (2019): 230-239.

[7] Orang, Omid, Petrônio Cândido de Lima Silva, and Frederico Gadelha Guimarães. "Introducing Randomized High Order Fuzzy Cognitive Maps as Reservoir Computing Models: A Case Study in Solar Energy and Load Forecasting." arXiv preprint arXiv:2201.02158 (2022).

[8] Haida, Takeshi, and Shoichi Muto. "Regression based peak load forecasting using a transformation technique." IEEE Transactions on Power Systems 9.4 (1994): 1788-1794.

[9] Trudnowski, Dan J., Warren L. McReynolds, and Jeffery M. Johnson. "Real-time very short-term load prediction for power-system automatic generation control." IEEE Transactions on Control Systems Technology 9.2 (2001): 254-260.

[10] Dash, P. K., A. C. Liew, and G. Ramakrishna. "Power-demand forecasting using a neural network with an adaptive learning algorithm." IEE proceedings. Generation, transmission and distribution 142 (1995).

[11] Shi, Heng, Minghao Xu, and Ran Li. "Deep learning for household load forecasting—A novel pooling deep RNN." IEEE Transactions on Smart Grid 9.5 (2017): 5271-5280.

[12] Kong, Weicong, et al. "Short-term residential load forecasting based on resident behaviour learning." IEEE Transactions on Power Systems 33.1 (2017): 1087-1088.

[13] Deng, Zhuofu, et al. "Multi-scale convolutional neural network with time-cognition for multi-step short-term load forecasting." IEEE Access 7 (2019): 88058-88071.

[14] Soeb, Md Janibul Alam, and Muhammad Rashed Al Mamun. "Application of advanced back propagation algorithm in electric load forecasting." J. Math. Stat. Sci. 6.3 (2020): 95-102.

[15] Sun, Mingyang, et al. "Using Bayesian deep learning to capture uncertainty for residential net load forecasting." IEEE Transactions on Power Systems 35.1 (2019): 188-201.

[16] Bacanin, Nebojsa, et al. "On the Benefits of Using Metaheuristics in the Hyperparameter Tuning of Deep Learning Models for Energy Load Forecasting." Energies 16.3 (2023): 1434.

[17] Alipour, Mohammadali, et al. "A novel electrical net-load forecasting model based on deep neural networks and wavelet transform integration." Energy 205 (2020): 118106.

[18] Khwaja, Ahmed Shaharyar, et al. "Joint bagged-boosted artificial neural networks: Using ensemble machine learning to improve short-term electricity load forecasting." Electric Power Systems Research 179 (2020): 106080.

[19] Müller, M. R., et al. "Electrical load forecasting in disaggregated levels using Fuzzy ARTMAP artificial neural network and noise removal by singular spectrum analysis." SN Applied Sciences 2.7 (2020): 1218.

[20] Li, Jian, et al. "A novel hybrid short-term load forecasting method of smart grid using MLR and LSTM neural network." IEEE Transactions on Industrial Informatics 17.4 (2020): 2443-2452.

[21] Deng, Zhuofu, et al. "Unified quantile regression deep neural network with time-cognition for probabilistic residential load forecasting." Complexity 2020 (2020): 1-18.

[22] Adil, Muhammad, et al. "Big data based electricity price forecasting using enhanced convolutional neural network in the smart grid." Web, Artificial Intelligence and Network Applications: Proceedings of the Workshops of the 34th International Conference on Advanced Information Networking and Applications (WAINA-2020). Springer International Publishing, 2020.

[23] Shi, Tian, et al. "Phase space reconstruction algorithm and deep learning-based very short-term bus load forecasting." Energies 12.22 (2019): 4349.

[24] Chen, Haiwen, et al. "Day-ahead aggregated load forecasting based on two-terminal sparse coding and deep neural network fusion." Electric Power Systems Research 177 (2019): 105987.

[25] Ribeiro, Gabriel Trierweiler, Viviana Cocco Mariani, and Leandro dos Santos Coelho. "Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting." Engineering Applications of Artificial Intelligence 82 (2019): 272-281.

[26] Zahid, Maheen, et al. "Electricity price and load forecasting using enhanced convolutional neural network and enhanced support vector regression in smart grids." Electronics 8.2 (2019): 122.

[27] Wang, Shouxiang, et al. "Bi-directional long short-term memory method based on attention mechanism and rolling update for short-term load forecasting." International Journal of Electrical Power & Energy Systems 109 (2019): 470-479.

[28] Zhu, L. , and C. Lian . "Wind Speed Forecasting Based on a Hybrid EMD-BLS Method." 2019 Chinese Automation Congress (CAC) 2019.

[29] Yan, Ke, et al. "A hybrid LSTM neural network for energy consumption forecasting of individual households." Ieee Access 7 (2019): 157633-157642.